CONCEPTS OF NETWORK AND DEFORMATION ANALYSIS

W. F. CASPARY

MONOGRAPH 11

SCHOOL OF
GEOMATIC ENGINEERING

THE UNIVERSITY OF NEW SOUTH WALES UNSW SYDNEY NSW 2052 AUSTRALIA

CONCEPTS OF NETWORK AND DEFORMATION ANALYSIS

W. F. CASPARY

Edited by

J. M. Rüeger

Third (corrected) Impression

August 2000

Monograph 11
School of Geomatic Engineering
The University of New South Wales,
UNSW SYDNEY NSW 2052
Australia

School of Geomatic Engineering
The University of New South Wales
UNSW SYDNEY NSW 2052
Australia

First published: March 1987
Second (corrected) Impression: August 1988
Third (corrected) Impression: August 2000

National Library of Australia
Card Number and ISBN
0-85839-044-2

CONTENTS

FOREWORD

The availability of fast and efficient computers and the introduction of statistical testing procedures in surveying opened the way for more sophisticated design and analysis procedures in deformation surveys during the last two decades. Prior to this change, the design and analysis of deformation measurements was essentially based on semigraphic methods. The lack of a suitable textbook may have been a reason why the newer techniques have found limited usage in practice. So far, the topic of deformation analysis has been dealt with primarily in articles in journals and in congress papers.

This monograph provides a comprehensive introduction to the design and analysis of networks in general and of deformation measurements in particular. The related topics of "Outlier Detection", "Reliability" and "Variance Component Estimation" are also presented in separate chapters. The textbook is aimed at final year undergraduate and at postgraduate students, wishing to specialize in this area as well as at practising surveyors and geodesists actively involved in deformation measurements and crustal movement studies by geodetic methods. Previous knowledge in linear algebra and least squares adjustments is an advantage. Some key reference books in these subjects are mentioned in the text for the benefit of readers.

The book purposely does not cover all analysis methods which are being promoted by the members of the FIG ad hoc Committee on the Analysis of Deformation Surveys. This Committee was established in 1978 for the purpose of designing a generally acceptable standard approach to deformation analysis. After eight years (and after an increase of the number of analysis procedures from about four to ten) the Committee concludes that "it would be difficult to completely unify all the approaches into one general set of guidelines for practising surveyors" and that "the final choice of the approach should be left to the user" (CHRZANOWSKI and CHEN, 1986). Because of the multitude of proposed methods, the author had to select a few appropriate methods for a detailed discussion. The book emphasizes simple, efficient and transparent approaches for the detection of single point movements from two epoch analyses although some alternative approaches (including robust techniques) are also discussed.

The most important concepts are demonstrated with the aid of numerical examples. These examples will be highly appreciated by all readers who wish to implement some of the computational procedures. SI-units have been used in these examples, with two exceptions: the four hundredth part of a circle is called "grad" rather than "gon" (or centesimal degree), and the ten thousandth part of a grad is defined as a centesimal second (1cc) rather than 0.1 mgon (one second of arc (1") corresponds to 3cc or 0.0003 grad).

The editor is greatly indebted to B. R. Harvey and W. Haen for their assistance in proof-reading as well as for their valuable remarks and to K. Kovacs for the careful layout of the book. Comments on errors in English expression or on printing errors are welcome and should be addressed to the editor.

It is the editor's belief that this monograph will greatly assist students in their studies on network design, network analysis and deformation measurements and that it will enable practising surveyors to upgrade the analysis of existing deformation measurement schemes to present standards.

J. M. Rüeger

January 1987

AUTHOR'S NOTES AND ACKNOWLEDGEMENTS

While on sabbatical leave from the Universität der Bundeswehr (University of the Defence Forces), Munich, Federal Republic of Germany, the author had the pleasure to join the School of Surveying, The University of New South Wales, Sydney during the first half of 1985. As a part of the visiting arrangements he gave an elective course on the subject of "Deformation Measurements with Geodetic Methods". The first draft of this monograph evolved from the lecture notes for this course and was completed in Sydney. In the process the author was encouraged and supported by the stimulating climate of School and city and by the fruitful discussions with students. The final work on this monograph was carried out in Munich and completed by 1986. The author is very much indebted to Jean Rüeger who kindly smoothed the English expression and made invaluable comments on numerous aspects of the first draft. Beyond that he unselfishly supervised the word processing and the preparations of many of the figures and thus rendered the publication really possible.

Harald Borutta kindly assisted in taking charge of the computations for the numerical examples and in thoroughly checking all equations. He made valuable suggestions with respect to the presentation of some approaches and did help much to improve the work.

Awraham Perelmuter, whilst on sabbatical leave at the Universität der Bundeswehr, took the trouble to read through the first draft. His constructive criticism and his worthy advice is gratefully acknowledged.

The author is greatly indebted to K. Kovacs and L. Tullis for the word processing of the manuscript, to C. Rusu for the preparation of some figures and to A.H.W. Kearsley for his efforts in the production of this book.

W. Caspary

June 1986

Author's Address:

Prof. Dr.-Ing. W. Caspary
Insitut für Geodäsie
Fakultät für Bauingenieur- und Vermessungswesen
Universität der Bundeswehr München
D-85577 NEUBIBERG
Germany

1. INTRODUCTION

The theory of deformation analysis has become popular during the last decade. However, the demand for deformation analysis is much older. First applications date back to the early twenties of this century. Geodetic methods for the determination of displacements and deformations of large dams were first applied in Switzerland. Extensive literature exists on measuring methods and equipment. A rigorous analysis of the measurements became feasible only recently with the advent of modern computers. These permit data processing using sophisticated mathematical models and are capable of dealing simultaneously with large numbers of observational data. The development of new and improved instruments of significantly higher accuracy has also opened new fields of application.

Today, deformation analysis is mainly applied to:

- recent crustal movements
- slope creep studies
- glacier and shelf ice movements
- ground subsidence
- deformation of man-made structures

The purpose of the monitoring of deformations usually is:

- establishment and/or verification of hypotheses in geophysics, geology, glaciology and engineering sciences
- assessment of safety and performance of engineering structures
- protection of the population from hazards caused by rock slides, slope creep and engineering structures
- determination of the responsibility for damages caused by mining, tunnelling and similar activities

This monograph treats methods of monitoring deformations based on repeatedly observed geodetic networks with emphasis on modelling, processing and evaluation of geodetic observations. Other methods using special instrumentation (pendulums, clinometers, extensometers, thermometers, pressure or strain gauges) or close-range and aerial photogrammetry are important as well. The analysis techniques are similar to those outlined here, so that no specific reference must be made.

The object or area under investigation is usually represented by a number of points which are monumented or marked durably. Geodetic observations transform the cluster of points into a geodetic network. The selection of points is, in most cases, governed by the topography or the structure of the object. The question of the number of points necessary strongly depends on the object and on the deformations anticipated. It is not possible to establish a general rule. An interdisciplinary approach to this problem should always be aimed at. Usually, two types of points are considered, namely reference points and object points. Two typical examples are depicted in Figures 1.1 and 1.2. In the first case, the network consists of a reference and an object part. The determination of displacements of the object points relative to the reference is the goal of the analysis. In the second case, only object points exist. The network is divided into two sub-nets, by a geological feature. In this case, only relative displacements between the blocks of object points can be determined.

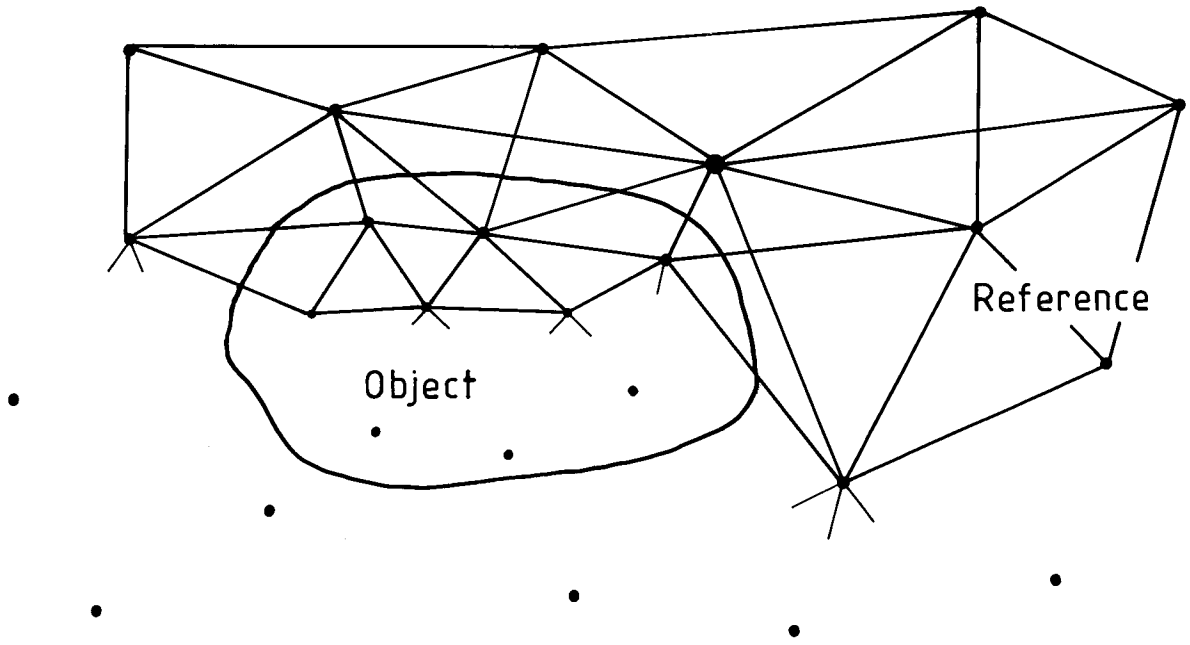


Figure 1.1: Monitoring network consisting of a reference network and an object network.

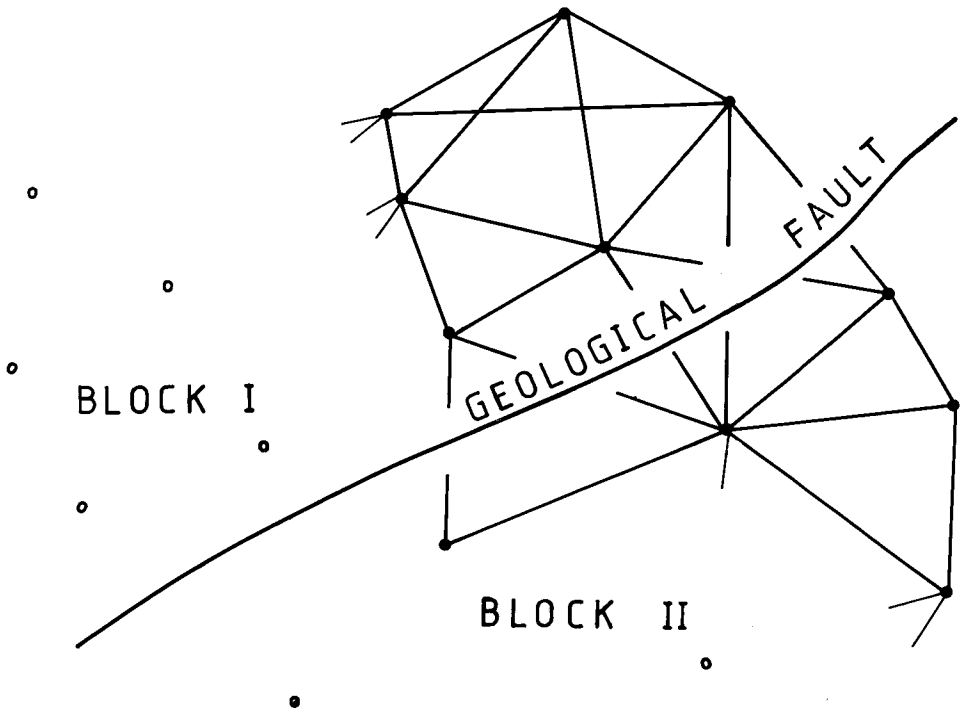


Figure 1.2: Monitoring network consisting of object points only, separated into two blocks by a geological fault.

To investigate the deformations of an object or an area the geodetic observations are repeated at different epochs of time. The observations of each epoch are adjusted independently. From the coordinate differences between the epochs the parameters of a deformation model are estimated and conclusions on the object deformations are drawn.

Theoretically, all observations of one epoch are carried out simultaneously. Since it is impossible to meet this condition, deformations taking place during the observation period need to be accounted for. If the rate of the deformations is small and the period of observation is short as compared with the time between the epochs, then the effect can be neglected in most cases. Otherwise, and especially in the presence of fast deformations, the observations must be centred to a common date and/or time.

The deformation analysis is carried out in consecutive steps which depend on the expected deformation pattern and on the type of network.

If the network comprises reference and object points (Fig. 1.1), the main steps are:

1. Screening of the reference points and elimination of unstable points.
2. Screening of the object points for single point movements, omission of the respective points or modeling of their displacements.
3. Design of a deformation model comprising rigid body movements and object deformations.
4. Verification of the model by statistical tests.

If solely object points are given (Fig. 1.2) or if, by the nature of the problem, no clear distinction between reference and object points is possible, the procedure starts with step 2. In this case it is not possible to model absolute rigid body movements in step 3.

After the first repetition of the observations, when two sets of geodetic data are available, a **two-epoch analysis** is carried out. The models used in this case are either static or dynamic. The former adopts a pure geometric point of view and provides the result in form of displacement vectors. The dynamic model links the displacements to their underlying forces. The result of this analysis is usually the basis for decisions on the suitability of the established network and on the frequency of future measurement epochs. Further two-epoch analyses are usually carried out after each new epoch.

When three or more measurement epochs are available a **multi-epoch analysis** can be carried out. The applied models are either kinematic or dynamic. The kinematic model does not include forces. It describes the deformations by means of displacement velocities and accelerations.

The deformation analysis is a powerful tool in investigating the stability of objects. But one must not forget that the results are always derived from carefully and accurately taken observations.

As the signal to noise ratio is small in most problems one has to be very careful in assessing the quality of the geodetic observations and in specifying the mathematical models for the network adjustments and for the deformations. All inaccuracies of the model, especially undetected systematic and gross errors in the observations, as well as incorrectly evaluated a priori variances will lead to apparent deformations contaminating the result. Therefore, it is of prime importance to employ all known methods which can contribute to the development of a realistic model.

Accordingly, the next Chapters will present in detail the main tools and techniques of design, optimization and verification of the pertinent mathematical models after the algebraic and statistical foundations have been introduced. This sequence frees the subsequent treatment of the deformation problem from distracting mathematical and statistical derivations.

2. THE GAUSS-MARKOV MODEL (GMM)

2.1 Basic Assumptions

The GMM is a linear mathematical model consisting of functional and stochastic relations. It relates the stochastic observations l_i to fixed parameters x_j. In matrix notation it takes the following form:

$$E(l) = Ax \quad \text{or} \quad l = Ax + \varepsilon$$
$$E(\varepsilon\varepsilon^t) = \Sigma = \sigma_0^2 Q \qquad (2\text{-}1)$$

To apply this theoretical model to the estimation of parameters from real data, it has to be rewritten in the form of a sample model:

$$l + v = A\hat{x}, \qquad P = Q^{-1} \qquad (2\text{-}2)$$

The symbols above have the following meaning:

l	n - vector of observations
$E(\cdot)$	expectation operator
x	u - vector of unknown parameters
A	n x u - matrix of known coefficients
ε	n - vector of true (real) errors
Σ	n x n - covariance matrix
σ_0^2	a priori variance factor
Q	n x n - cofactor matrix of observations
P	n x n - weight matrix of observations
v	n - vector of residuals

The original observation equations in geodetic networks are usually non-linear, e.g.:

distance: $E(l) = \sqrt{\Delta x^2 + \Delta y^2} + C_d$

azimuth: $E(l) = \arctan(\Delta y/\Delta x) + C_a$

direction: $E(l) = \arctan(\Delta y/\Delta x) + o + C_b$

To obtain the linear form as required in the GMM, a linearization (Taylor series) is carried out. With $a_{ij} = \partial l_i/\partial x_j$ the observation equations read:

$$l_i + v_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{iu}x_u, \quad i \in [1,n] \qquad (2\text{-}3)$$

where the orientation (o) and correction (C) parameters have been eliminated. The derivatives a_{ij} are computed with approximate values of the parameters. Thus, the vectors of observations and unknown parameters in Eq. (2-3) are computed as:

l_i = observed minus computed value of observation

x_j = estimated minus approximate parameter

The weights of the observations are calculated from:

$$p_i = \sigma_o^2/\sigma_i^2 \qquad (2\text{-}4)$$

where σ_i^2 is the a priori variance of l_i. Correlations between observations are hard to estimate realistically. Therefore they are usually ignored.

2.2 Parameter Estimation

It is known from experience that all observations contain errors. Therefore, it is impossible to get the true values of the parameters of Eq. (2-1). It is assumed that the errors are random with expectation zero. The number of observations is usually much larger than the number of parameters of the GMM. Equation (2-2) then represents an overdetermined system of equations and it is possible to introduce certain criteria in order to get optimal estimates of the parameters. There are different possibilities to define the optimality of statistical estimates. In the most frequently used approach, which shall be adopted here, an estimate is considered as optimal in the statistical sense, if it is unbiased and has minimum variance:

$$E(\hat{x}_j) = x_j \qquad \text{(unbiasedness)}$$

$$\forall\, j \in (1, u) \qquad (2\text{-}5)$$

$$s_{\hat{x}_j}^2 = \text{min.} \qquad \text{(minimum variance)}$$

(Note: The upside down A means "for all".)

The mathematical derivation, which can be found in textbooks on statistics (e.g. KOCH (1980, Section 32), RAO (1973, Section 4a), SEARLE (1971, Section 3)), leads to the estimation function:

$$\hat{x} = (A^tPA)^{-1}A^tPl \qquad (2\text{-}6)$$

being referred to as BLUE (best linear unbiased estimator). Under the same optimality criteria the variance estimator yields:

$$s_o^2 = v^tPv/(n-u) \qquad (2\text{-}7)$$

being the BIQUE (best invariant quadratic unbiased estimator) of the variance factor σ_o^2.

The estimation method dates back to the early 19th century and was developed independently by Gauss and Legendre. It is usually termed least squares (LS) method, as it minimizes the sum of the squares of the residuals. The same estimators are obtained by application of the maximum likelihood principle under the assumption that the observations are normally distributed, which is symbolically expressed by Eq. (2-8):

$$l \sim N(Ax, \Sigma) \qquad (2\text{-}8)$$

where Ax is the expectation of l and Σ its covariance matrix.

Later, the following cofactor matrices will be needed:

$$Q_{\hat{x}} = (A^tPA)^{-1}$$

$$Q_v = Q - A(A^tPA)^{-1}A^t \qquad (2\text{-}9)$$

$$Q_{\hat{l}} = A(A^tPA)^{-1}A^t = Q - Q_v$$

where:

$$\hat{l} = l + v$$

The derivations of these matrices are not given. They may be found in many textbooks on adjustments (e.g. BJERHAMMAR (1973, Section 21), KOCH (1980, Section 32), RAO (1973, Section 4a)) or derived by the reader using the following rules:

To obtain the cofactor matrix of a vector of functions y of a random vector l with known cofactor matrix Q_l :

1. express y as a linear (linearized) function of l :

$$y = Bl \qquad (2\text{-}10)$$

2. apply the propagation law of variances:

$$Q_y = BQ_lB^t \qquad (2\text{-}11)$$

3. simplify the right hand side of Eq. (2-11) if possible.

Upon completion of the computations it is essential to check whether the numerical results conform with the model or not.

2.3 Global Test of the Model

The principles of testing hypotheses are treated in Chapter 5 in some detail. In this context only the global model test is introduced without giving the underlying statistical theory.

The criterion of the so-called null hypothesis of the global test is: "The model is correct and complete". This can be expressed as:

$$H_o : E(s_o^2) = \sigma_o^2 \qquad (2\text{-}12)$$

and leads to the test statistic:

$$T = v^tPv/\sigma_o^2 \qquad (2\text{-}13)$$

which, under H_o, has a χ^2-distribution with expectation $(n - u)$:

$$T \sim \chi^2(n-u) \mid H_o \qquad (2\text{-}14)$$

where:

$\sim$ = is distributed as

$\mid$ = under the condition that (H_o is true)

After selection of an error probability, typically $\alpha = 5\%$, the value of the χ^2-distribution with $n - u$ degrees of freedom at $\alpha\%$ is read from a table. This value $\chi_\alpha{}^2(n - u)$ is then compared with T of Eq. (2-13). If:

$$T \leq \chi_\alpha{}^2(n - u) \qquad (2\text{-}15)$$

the test does not indicate contradictions between the observations and the mathematical model. A test cannot prove the validity of the model or the correctness of the observations!

In the opposite case ($T > \chi^2$) it needs to be investigated why the model or the observations or both are wrong. Further statistical tests can help in these investigations. Possible error sources are:

(a) In the functional model:

- map projection (coordinate system)
- instrument (calibration parameters)
- gravity field
- refraction model
- time factor

Previously unmodelled effects can be considered by reduction or correction of the observations or absorbed by an introduction of additional parameters.

(b) In the stochastic model:

- a priori variances
- correlations

A more realistic a priori covariance matrix might be required.

(c) In the observational data:

- gross measurement errors
- booking errors
- mistakes in identification of points
- unstable monuments
- centring errors

Screening of observations and residuals in combination with statistical tests can assist in cleaning the data.

(d) In the computations:

- programming errors
- input errors
- numerical stability of matrix inversions
- accumulation of rounding-off errors

The use of independent programs can indicate whether problems of this nature exist.

The detection of error sources is usually rather cumbersome. But every effort should be made to arrive at a mathematical model which approximates the physical reality with an accuracy considerably better than the precision of the observations.

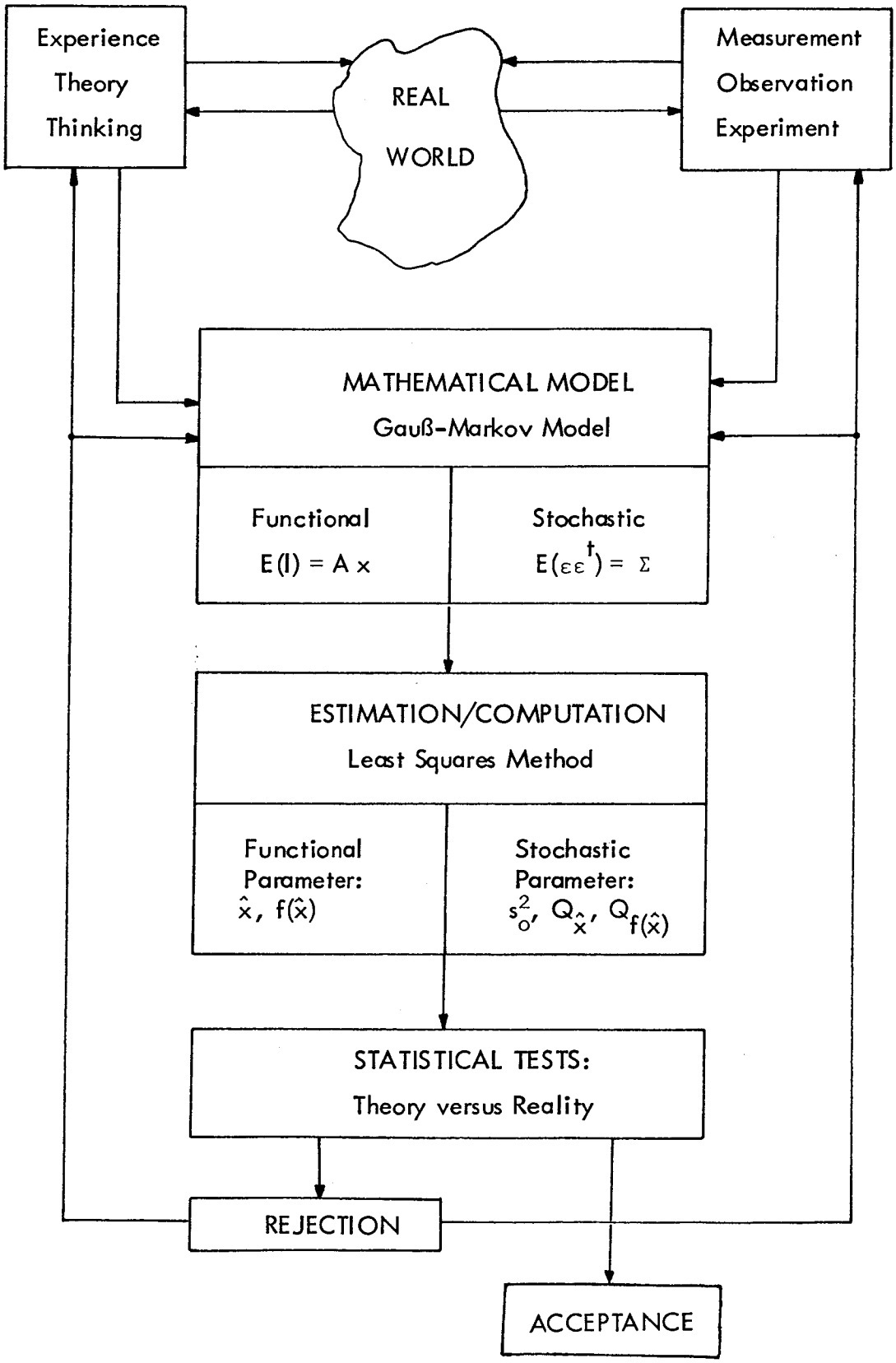


Figure 2.1: Design and Verification of the Gauss-Markov Model.

3. GEODETIC DATUM AND INVARIANT QUANTITIES

From a mathematical point of view the geodetic datum problem can be considered as the result of improper modelling. When using the functional model:

$$l + v = A\hat{x} \qquad (3\text{-}1)$$

parameters need to be chosen which are computable from the observations. It is trivial that no information on absolute heights can be expected if solely height differences are observed. Everybody knows that it is impossible to produce coordinates, when the observations comprise distances and angles only.

Nevertheless the unknown parameters in the GMM of geodetic networks are usually chosen to be heights and coordinates and this is done with good reasons:

- the observation equations are straightforward and easily programmed for computer calculation

- the covariance matrix of the parameters is a by-product of the parameter calculation and thus easily obtained

- the result is clear and easily visualised, and it is suitable for subsequent processing and documentation

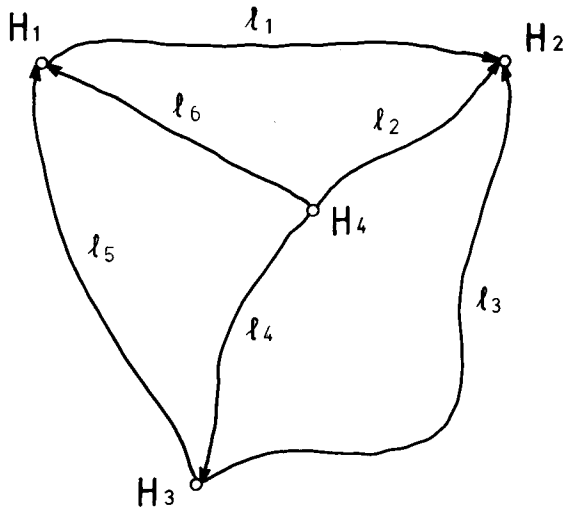
In consequence it is necessary to look carefully at the so-called geodetic datum problem in order to find out how the shortcomings of the model can be overcome.

Prior to the advent of modern computers the geodetic datum problem did not appear in the context of adjustment theory. The condition method of adjustment was preferred because of the typically smaller number of equations to be solved for. But the condition method proved to be less suitable for programming, particularly, if distance measurements are involved. The datum problem, of course, always existed but it was dealt with independently, outside of the adjustment calculations.

Before going into details, two simple examples are presented to illustrate the situation. Figure 3.1 depicts the datum problem for a one dimensional (1d) network. Six height differences are measured. The heights of the four points are chosen as the parameters of the GMM. As a consequence the coefficient matrix A only contains three linearly independent columns. The fourth column is a linear combination of the other three, i.e., there exists a vector $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)^t$ with at least one $\lambda_i \neq 0$, such that:

$$A\lambda = 0 \qquad (3\text{-}2)$$

This equation is the definition of linear dependency. The 4 x 4 - matrix of the normal equations A^tPA has a rank identical to the number of independent columns of A, i.e. $r(A^tPA) = 3$, hence this matrix is singular. It follows that no inverse exists and that it is impossible to estimate the heights from Eq. (2-6). The rank deficiency (or rank defect) d of A^tPA is the difference between the dimension and the rank of the square matrix; here $d = 1$. The deficiency can be eliminated by introducing information on the height system, i.e. by defining the geodetic datum.



Geometric Levelling

GMM: $l + v = A\hat{x}, \quad P = I$

Observation Equations:

$$
\begin{aligned}
l_1 + v_1 &= -H_1 + H_2 \\
l_2 + v_2 &= + H_2 - H_4 \\
l_3 + v_3 &= + H_2 - H_3 \\
l_4 + v_4 &= + H_3 - H_4 \\
l_5 + v_5 &= + H_1 - H_3 \\
l_6 + v_6 &= + H_1 - H_4
\end{aligned}
$$

Coefficient Matrix A

$$
\begin{bmatrix}
-1 & +1 & \cdot & \cdot \\
\cdot & +1 & \cdot & -1 \\
\cdot & +1 & -1 & \cdot \\
\cdot & \cdot & +1 & -1 \\
+1 & \cdot & -1 & \cdot \\
+1 & \cdot & \cdot & -1
\end{bmatrix}
$$

6 rows (observations)

4 columns (unknowns)

$\Rightarrow$ 6 x 4 - matrix A

$A\lambda = 0$ for $\lambda = (1, 1, 1, 1)^t$

$\Rightarrow$ $r(A) = 3, \quad d = 4 - 3 = 1$

Normal Equation: $A^tA\hat{x} - A^tl = 0$ for $P = I$

Matrix of Normals: $A^tA = N$

$$
\begin{bmatrix}
+3 & -1 & -1 & -1 \\
-1 & +3 & -1 & -1 \\
-1 & -1 & +3 & -1 \\
-1 & -1 & -1 & +3
\end{bmatrix}
$$

4 x 4 - matrix N

$N\lambda = 0$ for $\lambda = (1, 1, 1, 1)^t$

$\Rightarrow$ $r(N) = 3, \quad d = 4 - 3 = 1$

There is no matrix B such that: $NB = BN = I$.

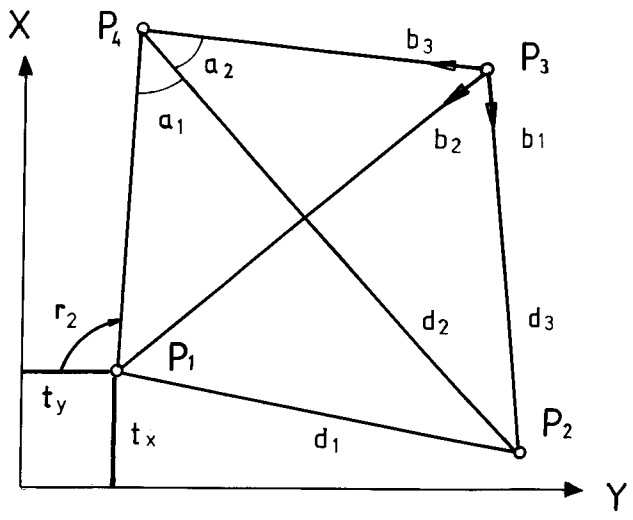
$\Rightarrow$ N does not possess an inverse.

$\Rightarrow$ The normals cannot be solved for $\hat{x}$.

Definition of a Reference System:

One translation in height.

Figure 3.1: The geodetic datum problem in a 1d-network: The definition of the height system is missing.



Braced Quadrilateral

Observations:

a. 8 angles: a_i

b. 6 distances: d_j

c. 12 bearings: b_k

Parameters:

8 coordinates

GMM: $l + v = A\hat{x}, \quad P = Q^{-1}$

a. 8×8 coefficient matrix A, $r(A) = 4 \Rightarrow d = 8 - 4 = 4$

b. 6×8 coefficient matrix A, $r(A) = 5 \Rightarrow d = 8 - 5 = 3$

c. 8 coordinates plus 4 orientation unknowns:

$\Rightarrow$ 12×12 coefficient matrix A, $r(A) = 8 \Rightarrow d = 12 - 8 = 4$

Normal Equation: $A^tPA\hat{x} - A^tPl = 0, \quad A^tPA = N$

$r(A) = r(N) \Rightarrow$ N does not possess an inverse.

$\Rightarrow$ The normals cannot be solved for $\hat{x}$ in the usual way.

Definition of a Reference System:

two translations: t_x and t_y

one rotation about z: r_z

one scale if no distances are measured: s

Figure 3.2: The geodetic datum problem in a 2d-network: The definition of the coordinate system is missing.

The second example, given in Figure 3.2, illustrates a two dimensional (2d) network. If the basic braced quadrilateral is considered with eight observed angles and if the coordinates of the four points are the parameters of the GMM, then the coefficient matrix A has the order of 8 x 8 with the rank being only r(A) = 4. The rank of A^tPA is 4 as well. In this case the rank deficiency is d = 4, hence it is necessary to introduce four independent quantities in order to define the geodetic datum. In a 2d-network these datum quantities are two translations which locate the coordinate system, one rotation to orientate it and, possibly, the scale. If the distances in Figure 3.2 are measured, then the scale can be defined by them. In the same way the orientation of the network can be introduced by measured azimuths.

The generalization to a three dimensional (3d) network is easily carried out. The maximal rank deficiency is d = 7 in this case. Seven independent quantities are required to define a proper reference system. This number is reduced, if the scale is taken from distance measurements and the orientation of the z-axis from zenith angle observations.

The examples illustrate the meaning of the geodetic datum and they underline the importance of properly defining a reference system if coordinates are chosen as parameters in the GMM.

In national survey control networks this was done when the first observations of the first order control were processed; in Central Europe more than one hundred years ago. All subsequent observations extending or densifying the national networks are related to the national datum by observations connecting new points with already fixed ones. Since the coordinates of the old points do not appear in the GMM, they are kept fixed and thus provide for the geodetic datum. The model, therefore, is of full rank, and apparently one has not to deal with the datum problem. It should be mentioned that this traditional geodetic procedure has been questioned during the last decade. Some geodesists advocate a so-called dynamic network in which all positions are allowed to change when new observations are adjusted. Thus the datum is variable (weak) and has to be defined in each particular case.

The situation is similar when a special purpose network (e.g. for deformation analysis) is established. The required accuracy is usually much higher than in the existing national control. To adopt the national datum would therefore degrade the results. Independent networks are therefore required. The datum is chosen in an optimal way depending on the purpose of a particular network.

3.1 Conventional Selection of Datum

To define a reference system some datum parameters have to be fixed. The number of these parameters equals the datum defect, as shown in the examples of Figures 3.1 and 3.2. The conventional approach is as follows:

1d-network: A surface of equal potential is the reference. It is introduced by arbitrarily fixing the height (gravity) of **one** point. In most cases the scale is taken from measured height (gravity) differences. If the scale is also considered as a datum parameter **one** height (gravity) difference is kept fixed, in addition to the fixed point. Other special 1d-problems are an arc of directions to unknown stations, where the direction of an arbitrarily selected station is set zero, and EDM-calibration lines where all distances refer to one pillar, usually selected at one end of the line.

2d-network: A Cartesian coordinate system is defined by fixing the **two** coordinates of one point and **one** bearing to a second point. If no distances are available, or if the scale shall not be taken from distance measurements, additionally, **one** distance is to be kept fixed. In the latter case **four** coordinates of two points may be fixed instead.

3d-network: Here, a Cartesian coordinate system is defined by fixing **three** coordinates of one point, **one** bearing and **two** zenith angles. If the scale is a datum parameter then **one** distance is also to be fixed. Alternatively, the reference system can be defined by the **six** coordinates of two points plus **one** additional element.

If a non-Cartesian reference system is used its definition has to be looked at carefully. A variety of different datum parameters may be necessary. This problem is not considered any further in this context.

One important condition of the datum definition is that it must not affect the geometry of the net. The relative position of the points shall be defined solely by the geodetic observations. The datum shall not impose any strain.

3.2 Conventional Datum Transformation

A change of the reference system is easily possible by the application of a transformation which retains the geometry of the network. Again, the number of parameters of such a transformation must not exceed the datum defect. It also has to be a so-called similarity transformation. The maximum number of parameters and their geometrical meanings are:

1d-network: d = 1 (2)
t_z = shift in z-direction
(s = scale, in special cases only)

Note: The numbers in brackets refer to the datum defect if scale is also a parameter.

2d-network: d = 3 (4)
t_x = translation in x-direction
t_y = translation in y-direction
r_z = rotation about z-axis
(s = scale, if not derived from distances)

3d-network: d = 6 (7)
t_x = translation in x-direction
t_y = translation in y-direction
t_z = translation in z-direction
r_x = rotation about x-axis
r_y = rotation about y-axis
r_z = rotation about z-axis
(s = scale, if not taken from distances)

So far, the scale factor has been treated differently from the other parameters, because the distance measurements contain information on the scale. A thorough investigation of geodetic observables, which usually are used to establish a 2d-network, shows that some other observables also contain information suitable for the definition of a reference system. The results are compiled in Table 3.1.

Observable	Datum Parameter t_x	t_y	r_z	s
distance measurements	–	–	–	x
horizontal angles	–	–	–	–
arcs of directions	–	–	–	–
azimuth (astron., gyro)	–	–	x	–
positions (astron., GPS)	x	x	x	x
position differences (GPS, inertial)	–	–	x	x

Table 3.1: Datum information contained in geodetic observables.

3.3 Zero-Variance Computational Base

The classical way of defining the datum of a geodetic network is to delete those columns of the design matrix A of Eq. (3-1) which refer to the parameters being kept fixed. If the model is partitioned according to this concept, assuming that the parameters in question are in the last positions of the parameter vector, the model takes the form:

$$l + v = (A_1 : A_2) \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} \tag{3-3}$$

where the d components of $\hat{x}_2$ shall define the reference system. The columns of the matrix A_2 are linear combinations of those of A_1. Their number equals the rank deficiency d. The linear dependency means that there exists a certain matrix L, such that:

$$A_1 L = A_2 \tag{3-4}$$

Rearrangement of Eq. (3-4) yields:

$$A_1 L - A_2 = (A_1 : A_2) \begin{pmatrix} L \\ -I_d \end{pmatrix} = 0$$

which is a generalization of Eq. (3-2), with I_d being a d x d - identity matrix.

Substitution of Eq. (3-4) in Eq. (3-3) gives:

$$l + v - A_1 L \hat{x}_2 = A_1 \hat{x}_1 \tag{3-5}$$

which is a model of full rank, showing that the approximate coordinates of $\hat{x}_2$ remain unchanged. The well known least squares solution of Eq. (3-5) for $\hat{x}_2 = 0$ reads:

$$\begin{aligned} \hat{x}_1^{\,o} &= (A_1^t P A_1)^{-1} A_1^t P l, & \hat{x}_2 &= 0 \\ Q_{\hat{x}_1}^{\,o} &= (A_1^t P A_1)^{-1}, & Q_{\hat{x}_2} &= 0 \\ Q_{\hat{x}_1 \hat{x}_2}^{\,o} &= 0 \end{aligned} \tag{3-6}$$

Sometimes it is desirable to leave the values of $\hat{x}_2$ undefined which leads to the general solution:

$$\left.\begin{aligned} \{\hat{x}_1\} &= (A_1^tPA_1)^{-1}A_1^tP(l - A_1L\hat{x}_2) \\ \{\hat{x}_1\} &= \hat{x}_1^o - L\hat{x}_2 \\ \hat{x}_2 &= \hat{x}_2 \end{aligned}\right\} \quad \text{for any } \hat{x}_2 \qquad (3\text{-}7)$$

The cofactor matrices $Q_{\hat{x}_1}$, $Q_{\hat{x}_2}$ and $Q_{\hat{x}_1\hat{x}_2}$ are independent of the selection of values for $\hat{x}_2$; they are the same as in Eq. (3-6). But, of course, the covariances strongly depend on the selection of points whose coordinates serve as datum parameters. These points are sometimes called the zero-variance computational base of the model. All elements of the covariance matrix $\Sigma_{\hat{x}_1}$ are to be considered relative variances in respect to the computational base. This approach has the advantage of leading to a reduced order of the matrix of normal equations and thus saving computer time and storage; but it is restricted to the special case of defining the datum by determination of d components of the vector x.

A more general approach consists of constraining the GMM by d condition equations.

3.4 Datum Constraints

The constrained GMM has the form:

$$\begin{aligned} A\hat{x} &= l + v \\ R^t\hat{x} &= c, \qquad c = \text{constant} \end{aligned} \qquad (3\text{-}8)$$

and leads to the well known normal equations:

$$\begin{pmatrix} A^tPA & R \\ R^t & 0 \end{pmatrix}\begin{pmatrix} \hat{x} \\ k \end{pmatrix} = \begin{pmatrix} A^tPl \\ c \end{pmatrix} \qquad (3\text{-}9)$$

with k being a vector of Lagrangian multipliers or correlates. The u x u - matrix A^tPA is singular with rank deficiency d = u – r(A). R^t has to consist of d linear independent constraints. For a 2d-network with computational base given by points P_i and P_j, according to Section 3.3, the classical approach takes the form:

$$\hat{x} = (x_1 \;\; y_1 \;\; x_2 \;\; y_2 \;\; \dots \;\; x_i \;\; y_i \;\; \dots \;\; x_j \;\; y_j \;\; \dots \;\; x_{u/2} \;\; y_{u/2})^t$$

$$R^t = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & +1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & +1 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & +1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & +1 & \dots & 0 & 0 \end{pmatrix}$$

If, as usual, the approximate coordinates of P_i and P_j are kept fixed, then c is a zero vector. This approach has the advantage that it can be generalized easily. The only condition to be met by the constraints is that they must provide a unique solution for the parameter vector $\hat{x}$ of Eq. (3-9). This means, algebraically, that the bordered normal equations must be made regular through suitable selection of R^t:

$$r\begin{pmatrix} A^tPA & R \\ R^t & 0 \end{pmatrix} = u+d$$

In this case the usual inverse exists, yielding the following parameter vector:

$$\begin{pmatrix} \hat{x} \\ k \end{pmatrix} = \begin{pmatrix} A^tPA & R \\ R^t & 0 \end{pmatrix}^{-1} \begin{pmatrix} A^tPl \\ c \end{pmatrix} \qquad (3\text{-}10)$$

The direct inversion of the normals given by Eq. (3-10) is numerically inefficient. Because of that and also to achieve a better understanding of this approach to the definition of the geodetic datum, a detailed derivation of the inverse block matrix is warranted.

From the definition of an inverse, $MM^{-1} = M^{-1}M = I$, follows, with $A^tPA = N$:

$$\begin{pmatrix} N & R \\ R^t & 0 \end{pmatrix} \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \begin{pmatrix} I_u & 0 \\ 0 & I_d \end{pmatrix} \qquad (3\text{-}11)$$

where the inverse Q of the normals and the identity matrix I have been partitioned accordingly. The multiplications in Eq. (3-11) yield the following four equations defining the blocks Q_{ij} of the inverse:

$$\begin{aligned} &\text{a.} && NQ_{11} + RQ_{21} = I_u \\ &\text{b.} && NQ_{12} + RQ_{22} = 0 \\ &\text{c.} && R^tQ_{11} = 0 \\ &\text{d.} && R^tQ_{12} = I_d \end{aligned} \qquad (3\text{-}12)$$

Since the n x u - matrix A has a rank deficiency of d, there exists an u x d - matrix S of rank d such that $AS = 0$. Refer to the definition of linear dependency in Eqs (3-2) and (3-4) in this context. The matrix S is not uniquely defined because:

$$A\bar{S} = 0 \text{ holds for any } \bar{S} = ST$$

where T is an arbitrary matrix with d rows.

One selection of S is related to the matrix L of Eq. (3-4). The same partitioning is used as before, namely $A = (A_1 : A_2)$, where A_1 consists of r independent columns of A, while the remaining $d = u - r$ columns, being linearly dependent on A_1, are combined in A_2. The following relations hold:

$$\begin{aligned} A_1 &= A\begin{pmatrix} I_r \\ 0 \end{pmatrix} \\ A_2 &= A\begin{pmatrix} 0 \\ I_d \end{pmatrix} = A_1L = A\begin{pmatrix} L \\ 0 \end{pmatrix} \\ A_1L - A_2 &= A\begin{pmatrix} L \\ 0 \end{pmatrix} - A\begin{pmatrix} 0 \\ I_d \end{pmatrix} = A\begin{pmatrix} L \\ -I_d \end{pmatrix} = AS = 0 \end{aligned} \qquad (3\text{-}13)$$

Equations (3-13) show that:

$$S = \begin{pmatrix} L \\ -I_d \end{pmatrix}$$

is one possible selection of the required matrix. For the following derivations it is not necessary to know S explicitly; only its existence must be established. On the other hand, Eq. (3-4) can be solved for L, yielding:

$$L = (A_1{}^tA_1)^{-1}A_1{}^tA_2$$

so that S can be specified easily if so desired.

Premultiplication of Eq. (3-12a) by S^t results in:

$$S^tRQ_{21} = S^t \quad \Rightarrow \quad Q_{21} = (S^tR)^{-1}S^t \tag{3-14}$$

The inverse of S^tR exists because S and R have (by definition) full rank d. Consequently the d x d - product matrix is regular.

Premultiplication of Eq. (3-12b) by S^t yields:

$$S^tRQ_{22} = 0 \quad \Rightarrow \quad Q_{22} = 0 \tag{3-15}$$

as S^tR is of full rank.

Postmultiplication of Eq. (3-14) by R gives:

$$Q_{21}R = (S^tR)^{-1}S^tR = I_d = R^tQ_{21}{}^t \tag{3-16}$$

Comparison with Eq. (3-12d) shows that:

$$Q_{12} = Q_{21}{}^t = S(R^tS)^{-1} \tag{3-17}$$

Premultiplication of Eq. (3-16) by R yields $R = RR^tQ_{12}$. Combining with $AS = 0 = NS$ leads to an equation for Q_{12} which does not depend on the matrix S:

$$(N + RR^t)Q_{12} = R \quad \Rightarrow \quad Q_{12} = (N + RR^t)^{-1}R \tag{3-18}$$

Considering Eq. (3-12c), Eq. (3-12a) expands to:

$$(N + RR^t)Q_{11} = I_u - RQ_{21} \tag{3-19}$$

By definition, R consists of d linear independent vectors which are not linear combinations of the rows of A. Hence the sum $N + RR^t$ has rank r(A) + d and is regular. Therefore the inverse of $N + RR^t$ exists. Premultiplication of Eq. (3-19) by this inverse together with Eq. (3-18) yields:

$$\begin{aligned} Q_{11} &= (N + RR^t)^{-1}(I_u - RQ_{21}) \\ Q_{11} &= (N + RR^t)^{-1} - Q_{12}Q_{21} \end{aligned} \tag{3-20}$$

To obtain a more favourable expression for Q_{11}, Eq. (3-12a) is postmultiplied by N leading to:

$$NQ_{11}N = N \tag{3-21}$$

where $S^tA^t = S^tN = 0$ and Eq. (3-14) have been used. Equation (3-21) shows that Q_{11} is a generalized inverse of N. (Refer to Section 3.11 for a definition of generalized inverses.). Considering Eq. (3-12c) it is possible to expand Eq. (3-21) as follows:

$$(N + RR^t)Q_{11}(N + RR^t) = N \tag{3-22}$$

which, after pre- and postmultiplication by $(N + RR^t)^{-1}$, gives an expression for Q_{11}. The expression for the matrix Q_{11} is independent of S and demonstrates the symmetry of the result:

$$Q_{11} = (N + RR^t)^{-1}N(N + RR^t)^{-1} \tag{3-23}$$

The definition of a geodetic datum by use of the constraints $R^tx = c$ leads to the general solution of Eq. (3-10) provided that R^t meets the two conditions:

i. R^t consists of $d = u - r(A)$ independent rows

ii. the rows of R^t are not linear combinations of the rows of A, so that $r(N + RR^t) = r(A) + r(R) = u$.

The solution yields:

$$\begin{pmatrix} \hat{x} \\ k \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & 0 \end{pmatrix} \begin{pmatrix} A^tPl \\ c \end{pmatrix}$$

$$\hat{x} = Q_{11}A^tPl + Q_{12}c \tag{3-24}$$

$$k = Q_{21}A^tPl = (S^tR)^{-1}S^tA^tPl = 0$$

with:

$$Q_{11} = (N + RR^t)^{-1}N(N + RR^t)^{-1} \tag{3-23}$$

$$Q_{12} = Q_{21}{}^t = (N + RR^t)^{-1}R \tag{3-18}$$

From the propagation law of variances follows:

$$Q_{\hat{x}} = Q_{11}A^tPQPAQ_{11} = Q_{11}NQ_{11} = Q_{11} \tag{3-25}$$

3.5 Constraints for the Conventional Datum

To illustrate the results of the previous section the general solutions are now specialized for the classical approach of selecting a computational base as discussed in Section 3.3. Arranging the parameters so that the last d elements of $\hat{x}$ refer to the datum points, and partitioning of A and R^t accordingly, yields:

$$A = (A_1 : A_2) \qquad R^t = (O : I_d) \qquad A^tPA = N = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix} \tag{3-26}$$

where $N_{ij} = A_i{}^tPA_j$. In conformity with Eq. (3-10) the general solution for the parameter vector is:

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ k \end{pmatrix} = \begin{pmatrix} N_{11} & N_{12} & 0 \\ N_{21} & N_{22} & I_d \\ 0 & I_d & 0 \end{pmatrix}^{-1} \begin{pmatrix} A_1{}^t P l \\ A_2{}^t P l \\ c \end{pmatrix} \qquad (3\text{-}27)$$

The block matrices of the inverse of Eq. (3-27) are easily computed from the relationships given in Eq. (3-28), which is a modification of Eq. (3-11):

$$\begin{pmatrix} N_{11} & N_{12} & 0 \\ N_{21} & N_{22} & I_d \\ 0 & I_d & 0 \end{pmatrix} \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{pmatrix} = \begin{pmatrix} I_r & 0 & 0 \\ 0 & I_d & 0 \\ 0 & 0 & I_d \end{pmatrix} \qquad (3\text{-}28)$$

Carrying out the multiplications according to Eq. (3-12) leads immediately to the result:

$$\begin{pmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{pmatrix} = \begin{pmatrix} N_{11} & N_{12} & 0 \\ N_{21} & N_{22} & I_d \\ 0 & I_d & 0 \end{pmatrix}^{-1} = \begin{pmatrix} N_{11}{}^{-1} & 0 & -N_{11}{}^{-1}N_{12} \\ 0 & 0 & I_d \\ -N_{21}N_{11}{}^{-1} & I_d & 0 \end{pmatrix} \qquad (3\text{-}29)$$

Substitution in Eq. (3-27) yields in agreement with Eq. (3-7):

$$\begin{aligned} \hat{x}_1 &= N_{11}{}^{-1}A_1{}^t P l - N_{11}{}^{-1}N_{12}c \\ \hat{x}_2 &= c, \qquad k = 0 \\ Q_{\hat{x}_1} &= N_{11}{}^{-1}, \qquad Q_{\hat{x}_2} = 0, \qquad Q_{\hat{x}_1\hat{x}_2} = 0 \end{aligned} \qquad (3\text{-}30)$$

This conventional solution for the geodetic datum problem is encountered regularly in the context of special purpose networks. It has the shortcoming that the parameter vector $\hat{x}$ and its cofactor matrix $Q_{\hat{x}}$ depend entirely on the arbitrarily selected datum points.

Nevertheless, this approach is widely used because the numerical computations are easily carried out and the interpretation of the result is straightforward. The estimation principles as introduced in Eq. (2-5) are:

$$\left.\begin{aligned} E(\hat{x}_i) &= x_i \\ s^2_{\hat{x}_i} &= \min \end{aligned}\right\} \quad \forall\, i$$

Examining the results of a so-called "free network" as defined by Eq. (3-24) and using $E(l) = Ax$ from Eq. (2-1), shows that the estimates are biased:

$$E(\hat{x}) = Q_{11}Nx + Q_{12}c \neq x \qquad (3\text{-}31)$$

and that the cofactor matrix:

$$Q_{\hat{x}} = Q_{11} = (N + RR^t)^{-1}N(N + RR^t)^{-1} \qquad (3\text{-}32)$$

obviously depends on the choice of the constraints.

3.6 Invariant Functions

In spite of the biasedness of the derived estimates there exist functions of $\hat{x}$ which are invariant with respect of the constraints R and therefore BLUEs in the sense of Section 2.2. It is easily shown that these functions, say $f = Bx$, must meet certain conditions. Substituting Eq. (3-24) for x yields for the estimate $\hat{f}$ of $f = Bx$:

$$\hat{f} = B\hat{x} = BQ_{11}A^tPl + BQ_{12}c$$

$$E(\hat{f}) = BE(\hat{x}) = BQ_{11}Nx + BQ_{12}c$$

Hence,

$$BQ_{11}N = B \quad \text{and} \quad BQ_{12} = 0 \tag{3-33}$$

are two conditions B has to fulfil so that f become invariant estimable.

If B is selected with its rows being linear combinations of the rows of A, i.e. $B = GA$, for some matrix G, then both conditions are met.

$$GAQ_{11}N = GA = B$$

$$GAQ_{12} = GAS(R^tS)^{-1} = 0$$

where $Q_{11}N = I_u - Q_{12}R$ from Eq. (3-12a) and $AS = 0$ by the definition of S.

Two important examples for invariant functions are the adjusted observations and the residuals. The derivation is again based on Eq. (3-12a) and the definition of S:

$$\hat{l} = l + v = A\hat{x} = AQ_{11}A^tPl$$

$$E(\hat{l}) = AQ_{11}A^tPE(l) = AQ_{11}Nx = Ax$$

$$Q_{\hat{l}} = AQ_{11}NQ_{11}A^t = AQ_{11}A^t$$

$$v = A\hat{x} - l = (AQ_{11}A^tP - I)l$$

$$E(v) = (AQ_{11}A^tP - I)E(l) = (AQ_{11}N - A)x = 0$$

$$Q_v = (AQ_{11}A^tP - I)Q(AQ_{11}A^tP - I)^t$$

$$Q_v = Q - AQ_{11}A^t$$

The cofactor matrices:

$$Q_{\hat{l}} \text{ and } Q_v$$

are independent of the selection of the constraints R because products of the form $HAQA^tK$ (for any H and K) are not influenced by the selection of Q (compare Eq. (3-21)) if Q is a generalized inverse of A^tPA (for more details see Section 3.11).

Because the adjusted observations and the residuals are independent of R, if R is selected according to the conditions given at the end of Section 3.4, the quadratic form v^tPv and the variance estimate s_o^2 are also independent of R. Thus these quantities are invariants of the GMM.

3.7 Minimum Trace Datum

The trace of a square matrix M is defined as the sum of the diagonal elements of M, written as tr M or sometimes tr(M). If M is a covariance matrix, then tr M is the sum of all variances and can be interpreted as a measure of the overall accuracy of the associated vector of random variates. Since the least squares estimate of the parameters in a free network adjustment is neither unbiased nor of minimum variance in general - in fact $\{\hat{x}\} = Q_{11}A^tPl + Q_{12}c$ defines a manifold of vectors - the question arises whether there are reasons to prefer one of all possible vectors of $\{\hat{x}\}$.

In the conventional approach, as outlined in Sections 3.3 and 3.5, the computational base is selected on the basis of geometrical considerations or arbitrarily. Considering the general solution of Section 3.4 it is possible to investigate if at least one unique vector exists in $\{\hat{x}\}$ with minimum trace of the associated cofactor matrix. Thus the criterion below can be established:

$$\text{tr } Q_{\hat{x}} = \text{tr } (N + RR^t)^{-1}N(N + RR^t)^{-1} = \min \tag{3-34}$$

where $N = A^tPA$ and $\hat{x}$ is any element of $\{\hat{x}\}$. In Eq. (3-34) the matrix R of the datum constraints can be varied within the range defined at the end of Section 3.4 to achieve the minimum of tr $Q_{\hat{x}}$. It turns out that the matrix S defined in Eq. (3-13) leads to the desired result. This can be proved in the following way.

Let: $S^tx = b$

with $AS = 0$
$r(S) = d$

be the datum constraints. Then, according to Eqs (3-20) and (3-23), the estimates are:

$$\underline{\hat{\bar{x}}} = \bar{Q}_{11}A^tPl + \bar{Q}_{12}b \tag{3-35}$$

$$Q_{\underline{\hat{\bar{x}}}} = \bar{Q}_{11} = (N + SS^t)^{-1} - \bar{Q}_{12}\bar{Q}_{21} \tag{3-36}$$

$$Q_{\underline{\hat{x}}} = (N + SS^t)^{-1} - S(S^tS)^{-1}(S^tS)^{-1}S^t \tag{3-37}$$

where the bars denote quantities which refer to the special constraints $R \equiv S$.

Postmultiplying Eq. (3-37) by $(N + SS^t)$ yields:

$$Q_{\underline{\hat{x}}}N = I - S(S^tS)^{-1}S^t \tag{3-39}$$

where $Q_{\hat{x}}S = 0$ from Eq. (3-12c) and $AS = 0$. On the other hand, $NQ_{\hat{x}}N = N$ (Eq. (3-21)) can be expanded to $(N + SS^t)Q_{\hat{x}}N = N$ which gives after premultiplication by $(N + SS^t)^{-1}$ and substitution in Eq. (3-39):

$$I - S(S^tS)^{-1}S^t = (N + SS^t)^{-1}N \tag{3-40}$$

Considering Eq. (3-23) $Q_{\hat{x}} = (N + SS^t)^{-1}N(N + SS^t)^{-1}$ and Eq. (3-21) $NQ_{\hat{x}}N = N$ indicates that the following relationship holds:

$$Q_{\underline{\hat{x}}} = (N + SS^t)^{-1}NQ_{\hat{x}}N(N + SS^t)^{-1} \tag{3-41}$$

where $Q_{\hat{x}}$ refers to any properly selected set of constraints $R^t\hat{x} = c$. Comparing Eqs (3-40) and (3-41) leads to the result:

$$Q_{\underline{\hat{x}}} = (I - S(S^tS)^{-1}S^t)Q_{\hat{x}}(I - S(S^tS)^{-1}S^t) \tag{3-42}$$

which expresses $Q_{\underline{\hat{x}}}$ as a function of S and $Q_{\hat{x}}$.

$$Q_{\underline{\hat{x}}} = Q_{\hat{x}} + S(S^tS)^{-1}S^tQ_{\hat{x}}S(S^tS)^{-1}S^t - Q_{\hat{x}}S(S^tS)^{-1}S^t - S(S^tS)^{-1}S^tQ_{\hat{x}} \tag{3-43}$$

Ensuring that:

$$\text{tr ABC} = \text{tr BCA} = \text{tr CAB} \tag{3-44}$$

holds if the respective products exist, application of the trace operator to Eq. (3-43) yields:

$$\text{tr } Q_{\underline{\hat{x}}} = \text{tr } Q_{\hat{x}} - \text{tr } (S^tS)^{-1}S^tQ_{\hat{x}}S \tag{3-45}$$

In order to prove that $\text{tr } Q_{\underline{\hat{x}}} \leq \text{tr } Q_{\hat{x}}$ for any $Q_{\hat{x}}$ it is sufficient to show that:

$$\text{tr } (S^tS)^{-1}S^tQ_{\hat{x}}S \geq 0$$

It follows from the definition of S that S^tS is a regular symmetrical matrix. Therefore, a decomposition (e.g. Cholesky) exists such that:

$$S^tS = (S^tS)^{1/2}[(S^tS)^{1/2}]^t \tag{3-46a}$$

and, correpondingly:

$$(S^tS)^{-1} = [(S^tS)^{-1/2}]^t(S^tS)^{-1/2} \tag{3-46b}$$

Using Eq. (3-46b) it is possible to rearrange Eq. (3-45) as follows:

$$\begin{aligned} \text{tr } Q_{\underline{\hat{x}}} &= \text{tr } Q_{\hat{x}} - \text{tr } (S^tS)^{-1/2}S^tQ_{\hat{x}}S[(S^tS)^{-1/2}]^t \\ &= \text{tr } Q_{\hat{x}} - \text{tr } F^tQ_{\hat{x}}F \end{aligned} \tag{3-47}$$

where $F^t = (S^tS)^{-1/2}S^t$. The second term on the right hand side of Eq. (3-47) has the form of a cofactor matrix of certain functions $F^t\hat{x}$ of the parameter vector $\hat{x}$. Consequently the diagonal elements of $F^tQ_{\hat{x}}F$ are all positive or zero and the trace must be positive. This gives the final proof, that $\text{tr } Q_{\underline{\hat{x}}}$ is smaller than or equal to $\text{tr } Q_{\hat{x}}$ for all cofactor matrices $Q_{\hat{x}}$ according to Eq. (3-25) with equality only for R = S.

The matrix S which yields this optimal solution is not unique. Any matrix of rank d, fulfilling the condition AS = 0 is a suitable choice.

In the context of geodetic networks the matrix S is usually found by geometrical considerations. An algebraical method more appropriate for computer solutions is based on the eigenvalues of the u x u - matrix $N = A^tPA$ of the normal equations.

$$(N - \lambda_iI)s_i = 0 \tag{3-48}$$

In Eq. (3-48) λ_i is called the eigenvalue or latent root of N and s_i is the associated eigenvector of N. It is well known from linear algebra, that a square matrix of type N with rank r possesses r non-zero eigenvalues associated with r linear independent eigenvectors s_i. If N has rank deficiency d, then a d-fold eigenvalue zero exists with d linear independent eigenvectors s_i, which can be merged in an u x d - matrix S, yielding one selection of constraints for the minimum trace datum. For $\lambda = 0$ from Eq. (3-48) the relation $NS = 0$ follows.

Some of the above equations become simpler and the numerical computations easier and more stable if the matrix S is normalized to S_n:

$$S_n = S(S^tS)^{-1/2} \quad \Rightarrow \quad S_n^tS_n = I_d \tag{3-49}$$

Eigenvalue (or spectral) decomposition routines are usually contained in algebra computer program libraries. They provide the normalized eigenvectors. (Normalization is explained in Section 3.10.)

So far, it has been proved that the cofactor matrix $Q_{\hat{x}}$ of minimum trace exists and is unique. But the corresponding parameter estimate, Eq. (3-35):

$$\hat{\underline{x}} = \bar{Q}_{11}A^tPl + \bar{Q}_{12}b$$

still depends on the arbitrary selection of b, namely the right hand side constants of the constraints $S_n^t\hat{x} = b$. To select one set of constants, say b_n, by a meaningful method, the criterion:

$$\|\hat{\underline{x}}\| = (\hat{\underline{x}}^t\hat{\underline{x}})^{1/2} = \min \tag{3-50}$$

is introduced. This approach is reasonable if the approximate coordinates available are very close to the optimal estimates. In this case the coordinate unknowns will be small.

The Euclidean norm (or length) of $\hat{x}$ in Eq. (3-50) becomes a minimum if $b_n = 0$ is chosen:

$$\begin{aligned}\hat{\underline{x}}^t\hat{\underline{x}} &= (\bar{Q}_{11}A^tPl + \bar{Q}_{12}b)^t(\bar{Q}_{11}A^tPl + \bar{Q}_{12}b) \\ &= l^tPA\bar{Q}_{11}\bar{Q}_{11}A^tPl + 2l^tPA\bar{Q}_{11}\bar{Q}_{12}b + b^t\bar{Q}_{21}\bar{Q}_{12}b\end{aligned}$$

From Eq. (3-17) follows $\bar{Q}_{12} = \bar{Q}_{21}^t = S(S^tS)^{-1}$ which leads after substitution to:

$$\hat{x}^t\hat{x} = l^tPA\bar{Q}_{11}\bar{Q}_{11}A^tPl + 2l^tPA\bar{Q}_{11}S(S^tS)^{-1}b + b^t(S^tS)^{-1}b$$

Use of Eq. (3-12c) $Q_{11}S = 0$ simplifies the quadratic form to:

$$\hat{x}^t\hat{x} = l^tPA\bar{Q}_{11}\bar{Q}_{11}A^tPl + b^t(S^tS)^{-1}b$$

Since both terms of the right hand side are strictly positive, the minimum with respect to b is attained, when the second term is zero. For the positive definite form matrix $(S^tS)^{-1}$ this requires $b_n = 0$.

Given $b_n = 0$ the norm of $\hat{x}$ depends on the selection of the constraints $R^t\hat{x} = 0$. It can be shown that the criterion $\text{tr } Q_{\hat{x}} = \text{min}$ as introduced in Eq. (3-34) is equivalent to the condition $\| \hat{x} \| = \text{min}$. Hence, both criteria require R = S and then hold simultaneously.

Summary:

In the GMM of a free network adjustment:

$$E(l) = Ax, \quad \Sigma = \sigma_o^2 Q, \quad P = Q^{-1}$$

the least squares (LS):

$$v^tPv = \text{min}$$

minimum norm:

$$\| \hat{x} \| = \text{min}$$

estimate of the parameters with minimum trace of the cofactor matrix:

$$\text{tr } Q_{\hat{x}} = \text{min}$$

is obtained by introducing the following constraints on the parameters:

$$S^t\hat{x} = 0$$

where $AS = 0$
$r(S) = d$

yielding:

$$\hat{x} = Q_{11}A^tPl \qquad (3\text{-}51)$$

$$Q_{\hat{x}} = Q_{11} = (N + SS^t)^{-1}N(N + SS^t)^{-1}$$

where $N = A^tPA$ and $v = A\hat{x} - l$. The columns of S are the d linear independent eigenvectors associated with the d-fold zero eigenvalue of N.

It should be noted that the matrix $Q_{\hat{x}} = Q_{11}$ is not needed to compute the parameter vector, as it follows from Eqs (3-14), (3-20) and from $S^tA^t = 0$ that:

$$Q_{11}A^t = (N + RR^t)^{-1}A^t$$

Therefore, it suffices to invert $(N + RR^t)^{-1}$ if no cofactor matrix is required.

3.8 General Datum Transformation

In Section 3.2 the conventional datum transformation has been addressed without going into detail. It has been pointed out, that a similarity transformation with d properly selected parameters does not deform the shape of the network and, thus, is suitable to change the reference system. In this section the transition from one datum, i.e. set of constraints, say $R_1^t\hat{x} = 0$ to another one, $R_2^t\hat{x} = 0$, is investigated. A linear transformation, namely the so-called S-transformation shall be established, which provides the datum transformation without requiring a repetition of the inversion of the u x u - normal matrix.

Let $R_1 = B_1S$ and $R_2 = B_2S$, where $AS = 0$ and $S^tS = I_d$. Further let B_1 and B_2 be selection matrices which are diagonal and have coefficients of +1 or 0 only, with at least d elements being +1. B obviously fills zeros in those rows of S which correspond to zeros of B and retains all other rows leading to a special structure of R. Thus by simply changing B a variety of constraints can be produced, for example:

1. $B = I \Rightarrow R = S \Rightarrow$ the minimum norm least squares solution.

2. $B = \begin{pmatrix} 0 & \\ & I_d \end{pmatrix} \Rightarrow$ the conventional datum with the last d coordinates serving as zero-variance computational base.

3. $B = \begin{pmatrix} 0 & & \\ & I_k & \\ & & 0 \end{pmatrix}$, $k > d \Rightarrow$ a solution which minimizes the partial norm $\| \hat{x}_k \|$ consisting of those k unknowns, which are indicated by the block matrix I_k of B. At the same time the sum of the corresponding k variances, i.e. the partial trace of the cofactor matrix, attains a minimum.

It is possible to define any reasonable datum for a network adjustment by correspondingly designing B.

The parameter estimate $\hat{x}_1$, associated with the constraints $R_1^tx = 0$, possesses according to Eq. (3-23) the cofactor matrix:

$$Q_{\hat{x}_1} = (N + R_1R_1^t)^{-1}N(N + R_1R_1^t)^{-1} \qquad (3\text{-}23)$$

A different set of constraints, e.g. $R_2^tx = 0$, yields the new parameter vector $\hat{x}_2$ with cofactor matrix $Q_{\hat{x}_2}$. These two solutions can be connected by virtue of the relationship $NQ_{\hat{x}}N = N$ of Eq. (3-21), which is correct for any $Q_{\hat{x}}$:

$$Q_{\hat{x}_1} = (N + R_1R_1^t)^{-1}NQ_{\hat{x}_2}N(N + R_1R_1^t)^{-1} \qquad (3\text{-}52)$$

Since N and $(N + R_1R_1^t)^{-1}$ are symmetrical matrices, Eq. (3-52) can be written as:

$$Q_{\hat{x}_1} = K_1Q_{\hat{x}_2}K_1^t \qquad (3\text{-}53)$$

with the substitution:

$$K_1 = (N + R_1R_1^t)^{-1}N \qquad (3\text{-}54)$$

Equation (3-53) resembles the result of the general law of propagation of variances if applied to the linear function:

$$\hat{x}_1 = K_1\hat{x}_2 \qquad (3\text{-}55)$$

In fact, it can be proved that this relationship exists. To demonstrate that Eq. (3-55) holds it is sufficient to show that:

$$\hat{x}_1 = Q_{\hat{x}_1} A^t P l = K_1 Q_{\hat{x}_2} A^t P l \qquad (3\text{-}56)$$

is valid for any vector l of observations. Considering that P is a regular matrix, Eq. (3-56) is equivalent to:

$$Q_{\hat{x}_1} A^t = K_1 Q_{\hat{x}_2} A^t \qquad (3\text{-}57)$$

If Eq. (3-57) is premultiplied by $(N + R_1 R_1{}^t)$ and the relationships $R_1 = B_1 S$ and $S^t A^t = 0$ are employed then:

$$N Q_{\hat{x}_1} A^t = N Q_{\hat{x}_2} A^t \qquad (3\text{-}58)$$

is the result which is correct for any matrix $Q_{\hat{x}} = Q_{11}$ as defined in Eqs (3-12) and (3-21). Thus the S-transformation is established which can be used to transform any solution $\hat{x}$, $Q_{\hat{x}}$ of the general GMM into another solution based on the datum constraints matrix $R_1 = B_1 S$. It is not necessary to know the constraints which have been used for the original solution. But they must be of type R = BS as introduced in this section. The computation employs the equations:

$$\hat{x}_1 = K_1 \hat{x} \qquad (3\text{-}55)$$

$$Q_{\hat{x}_1} = K_1 Q_{\hat{x}} K_1{}^t \qquad (3\text{-}53)$$

where:

$$K_1 = (N + R_1 R_1{}^t)^{-1} N = I - S(R_1{}^t S)^{-1} R_1{}^t \qquad (3\text{-}54)$$

The second solution for K_1 on the right hand side of Eq. (3-54) is computationally more convenient than the first one. It only involves the inversion of a d x d - matrix while an u x u - matrix $N + R_1 R_1{}^t$ is to be inverted in the other case. This expression for K_1 follows from Eqs (3-20) and (3-23), in combination with Eq. (3-17).

In the case where a second change of datum is required to meet the new set of constraints (for example: $R_2{}^t x = 0$, $R_2 = B_2 S$), it suffices to compute the new transformation matrix K_2 and to apply Eqs (3-55) and (3-53) on $\hat{x}_1$ or on $\hat{x}$ yielding the same result:

$$\hat{x}_2 = K_2 \hat{x}_1 = K_2 \hat{x}$$

$$Q_{\hat{x}_2} = K_2 Q_{\hat{x}_1} K_2{}^t = K_2 Q_{\hat{x}} K_2{}^t$$

This outcome is due to the general relationship:

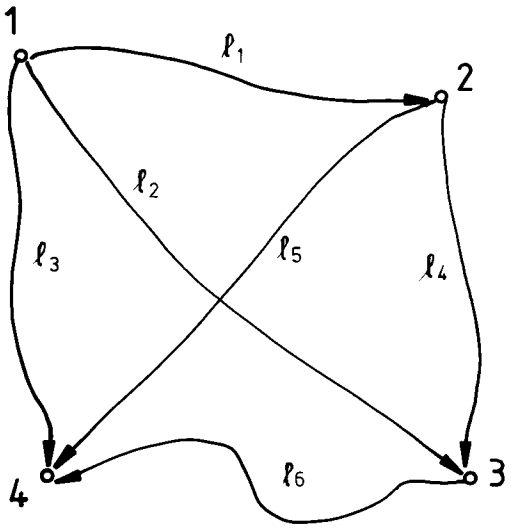
$$K_i K_j = K_i$$

which can easily be proved for any K_i and K_j for which the definition of Eq. (3-54) applies.

On completion of the discussion of the datum problem, a numerical example follows.

3.9 Example

In order to illustrate the results of the previous sections a simple example is presented. The various solutions should be carefully studied by the reader. A levelling network similar to that of Fig. 3.1 is selected.



$$l_1 = H_2 - H_1 = 1.2 \text{ mm}$$
$$l_2 = H_3 - H_1 = 1.6 \text{ mm}$$
$$l_3 = H_4 - H_1 = 1.7 \text{ mm}$$
$$l_4 = H_3 - H_2 = 1.2 \text{ mm}$$
$$l_5 = H_4 - H_2 = 2.1 \text{ mm}$$
$$l_6 = H_4 - H_3 = 1.3 \text{ mm}$$

Figure 3.3: Geometric Levelling

As all points lie on a horizontal plane, they should have the same heights. The levelling lines are nearly of the same length. The sample model according to Eq. (2-2) is:

GMM: $l + v = A\hat{x}, \quad \Sigma = \sigma_o^2 I \Rightarrow P = I$

The vector x contains the heights of the u = 4 points, hence the order of the design matrix A is o(A) = 6 x 4. It is known from Fig. 3.1 that A has rank r(A) = 3. Thus the rank deficiency of A is d = u − r = 4 − 3 = 1. As it is assumed that all n = 6 observations have weight $P_i = 1$ the LS normal equations take the form:

$$A^tA\hat{x} - A^tl = 0$$

with:

$$A = \begin{bmatrix} -1 & +1 & 0 & 0 \\ -1 & 0 & +1 & 0 \\ -1 & 0 & 0 & +1 \\ 0 & -1 & +1 & 0 \\ 0 & -1 & 0 & +1 \\ 0 & 0 & -1 & +1 \end{bmatrix}, \quad A^tA = N = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}$$

As pointed out several times, it is not possible to solve the normal equations for $\hat{x}$ without defining a height system. Five approaches are discussed:

a. Conventional datum definition by keeping the approximate height of one point fixed. Refer to Sections 3.1 and 3.3. If point P_1 is chosen as the datum point, then the first column of A, referring to x_1, is dropped yielding the 6 x 3 - matrix A_a and the 3 x 3 - matrix N_a of the normal equations with rank deficiency d = 0.

$$A_a = \begin{bmatrix} +1 & 0 & 0 \\ 0 & +1 & 0 \\ 0 & 0 & +1 \\ -1 & +1 & 0 \\ -1 & 0 & +1 \\ 0 & -1 & +1 \end{bmatrix}, \quad N_a = \begin{bmatrix} +3 & -1 & -1 \\ -1 & +3 & -1 \\ -1 & -1 & +3 \end{bmatrix}, \quad A_a^t l = \begin{bmatrix} l_1 & -l_4 & -l_5 \\ l_2 & +l_4 & -l_6 \\ l_3 & +l_5 & +l_6 \end{bmatrix}$$

$$N_a^{-1} = {}^1/_4 \begin{bmatrix} +2 & +1 & +1 \\ +1 & +2 & +1 \\ +1 & +1 & +2 \end{bmatrix}, \quad A_a^t l = \begin{bmatrix} -2.1 \\ +1.5 \\ +5.1 \end{bmatrix} \text{ mm}$$

$$\hat{x}_a = N_a^{-1} A_a^t l$$

$$x_{1a} = 0 \qquad = 0 \text{ (datum!)}$$

$$\hat{x}_{2a} = {}^1/_4 (2l_1 + l_2 + l_3 - l_4 - l_5 \quad \bullet \;) = +0.6 \text{ mm}$$

$$\hat{x}_{3a} = {}^1/_4 (\; l_1 + 2l_2 + l_3 + l_4 \quad \bullet \quad - l_6 \;) = +1.5 \text{ mm}$$

$$\hat{x}_{4a} = {}^1/_4 (\; l_1 + l_2 + 2l_3 \quad \bullet \quad + l_5 + l_6 \;) = +2.4 \text{ mm}$$

$$v = A_a \hat{x}_a - l$$

$$v^t = (-0.6 \quad -0.1 \quad +0.7 \quad -0.3 \quad -0.3 \quad -0.4) \text{ mm}$$

$$v^t v = 1.2 \text{ mm}^2$$

$$s_o^2 = v^t v/(n-u+d) = \frac{1.2}{6-4+1} = 0.4 \text{ mm}^2$$

$$Q_{\hat{x}_a} = {}^1/_4 \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & +2 & +1 & +1 \\ 0 & +1 & +2 & +1 \\ 0 & +1 & +1 & +2 \end{bmatrix}, \quad \text{tr } Q_{\hat{x}_a} = {}^3/_2, \quad \Sigma_{\hat{x}_a} = s_o^2 Q_{\hat{x}_a}$$

$$\| \hat{x}_a \| = (\hat{x}_a^t \hat{x}_a)^{1/2} = 2.89 \text{ mm}$$

b. Zero-variance computational base by introducing the datum constraint $R^t x = 0$. Refer to Sections 3.4 and 3.5. If the height of point P_1 is chosen as the zero-variance computational base, then the constraint referring to x_1 takes the form: $R_b = (+1 \; 0 \; 0 \; 0)^t$. Naturally, the result using this datum definition must be identical to that of the previous subsection (a), as P_1 serves as the only datum point in both cases. A comparison with Section 3.8 shows that R_b is of the form: $R_b = B_b S$, where:

$$B_b = \begin{pmatrix} +1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} \quad \text{and} \quad S^t = (1 \; 1 \; 1 \; 1)$$

as pointed out in Fig. 3.1. It follows from $AS = 0$ and Eq. (3-48) that S is an eigenvector associated with the zero eigenvalue of N.

Eq. (3-23): $Q_{\hat{x}_b} = (N + R_b R_b^t)^{-1} N (N + R_b R_b^t)^{-1}$

$$N + R_b R_b^t = \begin{bmatrix} +4 & -1 & -1 & -1 \\ -1 & +3 & -1 & -1 \\ -1 & -1 & +3 & -1 \\ -1 & -1 & -1 & +3 \end{bmatrix}, \quad (N + R_b R_b^t)^{-1} = {}^1/_4 \begin{bmatrix} 4 & 4 & 4 & 4 \\ 4 & 6 & 5 & 5 \\ 4 & 5 & 6 & 5 \\ 4 & 5 & 5 & 6 \end{bmatrix}$$

$$A^t l = \begin{bmatrix} -l_1 & -l_2 & -l_3 \\ +l_1 & -l_4 & -l_5 \\ +l_2 & +l_4 & -l_6 \\ +l_3 & +l_5 & +l_6 \end{bmatrix} = \begin{bmatrix} -4.5 \\ -2.1 \\ +1.5 \\ +5.1 \end{bmatrix} \text{mm}$$

Eq. (3-24): $\hat{x}_b = Q_{\hat{x}_b} A^t l$

It is easily verified that: $\hat{x}_b \equiv \hat{x}_a$ and $Q_{\hat{x}_b} \equiv Q_{\hat{x}_a}$. Furthermore, v and s_o^2 are identical for (a) and (b).

c. Minimum norm: $\| \hat{x} \| = \min$, minimum trace: $\text{tr } Q_{\hat{x}} = \min$ solution. Refer to Section 3.7. As mentioned before, N possesses one zero eigenvalue due to d = 1. The associated eigenvector is $S^t = (1\ 1\ 1\ 1)$ yielding NS = 0 and AS = 0. Normalization of S, as defined in Eq. (3-49), results in:

$$S_c = S(S^t S)^{-1/2} = {}^1/_2 S = {}^1/_2 (1\ 1\ 1\ 1)^t \Rightarrow S_c^t S_c = 1$$

Eq. (3-37): $Q_{\hat{x}_c} = (N + S_c S_c^t)^{-1} - S_c S_c^t$

$$(N + S_c S_c^t) = {}^1/_4 \begin{bmatrix} +13 & -3 & -3 & -3 \\ -3 & +13 & -3 & -3 \\ -3 & -3 & +13 & -3 \\ -3 & -3 & -3 & +13 \end{bmatrix}, \quad (N + S_c S_c^t)^{-1} = {}^1/_{16} \begin{bmatrix} 7 & 3 & 3 & 3 \\ 3 & 7 & 3 & 3 \\ 3 & 3 & 7 & 3 \\ 3 & 3 & 3 & 7 \end{bmatrix}$$

$$Q_{\hat{x}_c} = {}^1/_{16} \begin{bmatrix} +3 & -1 & -1 & -1 \\ -1 & +3 & -1 & -1 \\ -1 & -1 & +3 & -1 \\ -1 & -1 & -1 & +3 \end{bmatrix}, \quad \text{tr } Q_{\hat{x}_c} = {}^3/_4$$

$$\hat{x}_c = Q_{\hat{x}_c} A^t l = {}^1/_4 \begin{bmatrix} -l_1 & -l_2 & -l_3 \\ +l_1 & -l_4 & -l_5 \\ +l_2 & +l_4 & -l_6 \\ +l_3 & +l_5 & +l_6 \end{bmatrix} = \begin{bmatrix} -1.125 \\ -0.525 \\ +0.375 \\ +1.275 \end{bmatrix} \text{mm}, \quad \|\hat{x}_c\| = 1.82 \text{ mm}$$

The solution $\hat{x}_c$ meets the constraint $S_c^t \hat{x}_c = 0$. A comparison with datum (a) shows that the minimum norm and minimum trace solution in fact produces smaller values of norm and trace:

$$\|\hat{x}_c\| < \|\hat{x}_a\| \text{ and } \text{tr } Q_{\hat{x}_c} < \text{tr } Q_{\hat{x}_a}$$

d. Datum definition by minimizing the partial norm $(\hat{x}_d^t B \hat{x}_d)^{1/2} = \min$ and, simultaneously, the partial trace $\mathrm{tr}\, BQ_{\hat{x}_d} B = \min$. Refer to Section 3.8, case 3.

If the points P_2 and P_3 are chosen as datum points, then the selection matrix B reads:

$$B_d = \begin{pmatrix} 0 & I_2 & 0 \end{pmatrix} \quad \text{yielding:} \quad R_d = B_d S = (0\ 1\ 1\ 0)^t$$

The datum constraint $R_d^t \hat{x} = 0$ yields $\hat{x}_2 + \hat{x}_3 = 0$.

Eq. (3-23): $Q_{\hat{x}_d} = (N + R_d R_d^t)^{-1} N (N + R_d R_d^t)^{-1}$

$$N + R_d R_d^t = \begin{bmatrix} +3 & -1 & -1 & -1 \\ -1 & +4 & 0 & -1 \\ -1 & 0 & +4 & -1 \\ -1 & -1 & -1 & +3 \end{bmatrix}, \qquad (N + R_d R_d^t)^{-1} = {}^1/_8 \begin{bmatrix} 5 & 2 & 2 & 3 \\ 2 & 3 & 1 & 2 \\ 2 & 1 & 3 & 2 \\ 3 & 2 & 2 & 5 \end{bmatrix}$$

$$Q_{\hat{x}_d} = {}^1/_8 \begin{bmatrix} +3 & 0 & 0 & +1 \\ 0 & +1 & -1 & 0 \\ 0 & -1 & +1 & 0 \\ +1 & 0 & 0 & +3 \end{bmatrix}, \qquad \begin{aligned} &\mathrm{tr}\, Q_{\hat{x}_d} = 1 \\ &\text{partial tr } Q_{\hat{x}_d} = {}^1/_4 \end{aligned}$$

$$\hat{x}_d = Q_{\hat{x}_d} A^t l = \begin{bmatrix} -1.05 \\ -0.45 \\ +0.45 \\ +1.35 \end{bmatrix} \text{mm}, \qquad \begin{aligned} &\|\hat{x}_d\| = 1.83 \\ &\text{partial } \|\hat{x}_d\| = 0.64 \end{aligned}$$

The result meets the constraint. The comparison with datum (a), (b) and (c) shows, that the partial trace, i.e. the sum of the diagonal elements of $Q_{\hat{x}_d}$, referring to $\hat{x}_2$ and $\hat{x}_3$ is less than for the other solutions, while the complete trace exceeds the corresponding value of (c). The same applies for the norm of the parameter vector.

A note at the end of Section 3.7 states that:

$$\hat{x} = Q_{\hat{x}} A^t P l = (N + RR^t)^{-1} A^t P l$$

It is suggested that the reader verify this relationship by simple computations for the datum definitions (a) through (d).

e. Datum transformations according to Section 3.8. If the results $\hat{x}$ and $Q_{\hat{x}}$ for one datum are available then any other datum can be introduced by simple transformation.

Transformation from (a) to (c):

Eq. (3-55): $\hat{x}_c = K_c \hat{x}_a$

Eq. (3-53): $Q_{\hat{x}_c} = K_c Q_{\hat{x}_a} K_c^t$

Eq. (3-54): $K_c = I - S(R_c^t S)^{-1} R_c^t$

The constraints for datum (c) are $S_c \equiv R_c$, $R_c^t = {}^1/_2(1\ 1\ 1\ 1)$, the matrix S repeatedly used in these examples is $S = (1\ 1\ 1\ 1)^t$, hence:

$$R_c^tS = {}^4/_2 = 2, \quad (R_c^tS)^{-1} = {}^1/_2$$

$$K_c = {}^1/_4 \begin{bmatrix} +3 & -1 & -1 & -1 \\ -1 & +3 & -1 & -1 \\ -1 & -1 & +3 & -1 \\ -1 & -1 & -1 & +3 \end{bmatrix}, \quad K_c\hat{x}_a = \begin{bmatrix} -1.125 \\ -0.525 \\ +0.375 \\ +1.275 \end{bmatrix} \text{ mm}$$

Eq. (3-53) can be readily verified. Similarly, it can be demonstrated, that the result of the transformation only depends on R_c, thus:

$$\hat{x}_c = K_c\hat{x}_d \quad \text{and} \quad Q_{\hat{x}_c} = K_cQ_{\hat{x}_d}K_c^t$$

It is suggested that the reader transforms the results of (a) and (c) into (d) using the transformation matrix K_d. Furthermore the reader is encouraged to establish the relations:

$$K_aK_d = K_a, \qquad K_dK_a = K_d$$

and

$$K_aK_cK_d = K_a, \qquad K_cK_dK_a = K_c$$

3.10 Datum Constraints for Geodetic Networks

The previous sections have shown that datum definitions for geodetic networks can always be introduced by means of constraints. Furthermore, it has been pointed out (in Sections 3.7 and 3.8) that geometrically meaningful datum selections are of the form $R^tx = 0$, where $R = BS$, with the matrix B selected according to the definition in Section 3.8 and the matrix S so that $AS = NS = 0$. In this section some previously mentioned methods and some special cases of establishing S shall be presented.

The matrices S required for common surveying networks are well known. For a **1d-network**, as demonstrated in Section 3.9, one constraint is required due to the rank deficiency of $d = 1$. The constraint has the form $S^t = (1\ 1\ 1\ \ldots\ 1)a$, where a is any constant. To achieve numerical stability for computer solutions, the matrix S is to be normalized, i.e. the length (norm) of the rows are to be made equal to unity. The general method of normalization is:

$$S_n^t = (S^tS)^{-1/2}S^t \Rightarrow S_n^tS_n = I \tag{3-59}$$

where:

$$(S^tS)^{-1/2} = [(S^tS)^{1/2}]^{-1} \quad \text{and} \quad S^tS = (S^tS)^{1/2}[(S^tS)^{1/2}]^t \tag{3-60}$$

The factorization of S^tS according to Eq. (3-60) exists for all regular square matrices. (Only square matrices are considered in this context.). The standard method is the Cholesky factorization, available from major program libraries. In the special case of the one-dimensional network the normalization yields:

$$S_n^t = \frac{1}{\sqrt{u}}(1\,1\,1\,\ldots\,1) \qquad (3\text{-}61a)$$

where u is the number of non-zero elements of S. For the datum of a **2d-network** with a rank deficiency of d = 4, four constraints are required. If the arrangement of the coordinates of the p points is:

$$x^t = (x_1,\ y_1,\ x_2,\ y_2,\ \ldots\ x_p,\ y_p)$$

then:

$$S^t = \begin{bmatrix} +1 & 0 & +1 & 0 & \ldots \\ 0 & +1 & 0 & +1 & \ldots \\ +\bar{y}_1 & -\bar{x}_1 & +\bar{y}_2 & -\bar{x}_2 & \ldots \\ +\bar{x}_1 & +\bar{y}_1 & +\bar{x}_2 & +\bar{y}_2 & \ldots \end{bmatrix} \begin{matrix} \Rightarrow \Sigma\hat{x}_i = 0 \\ \Rightarrow \Sigma\hat{y}_i = 0 \\ \Rightarrow \Sigma(\bar{y}_i\hat{x}_i - \bar{x}_i\hat{y}_i) = 0 \\ \Rightarrow \Sigma(\bar{x}_i\hat{x}_i + \bar{y}_i\hat{y}_i) = 0 \end{matrix} \qquad (3\text{-}61b)$$

$\bar{x}_i$ and $\bar{y}_i$ are the approximate coordinates while $\hat{x}_i$ and $\hat{y}_i$ are the small corrections to be estimated. The same arguments as for the 1d-network require a normalization of S. Usually, this is done in two steps. The coordinate system is shifted first to make the origin coincide with the centre of gravity of the network.

$$\bar{x}_i = x_i^o - \Sigma x_i^o/p, \quad \bar{y}_i = y_i^o - \Sigma y_i^o/p \qquad (3\text{-}62)$$

where x_i^o and y_i^o denote the coordinates in the original system. As a by-product of this translation, the rows s_j^t of S^t become orthogonal, i.e.:

$$s_i^t s_k = 0 \text{ for } i \neq k \qquad (3\text{-}63)$$

Hence S^tS is a diagonal matrix in this case. The normalization is therefore achieved very easily. The matrix $(S^tS)^{-1/2}$ is a diagonal matrix with the reciprocals of the square roots of the elements of S^tS as elements. The matrix S_n^t computed from Eq. (3-59) is thus orthogonal and normalized or, to use another term, orthonormal.

Geometrically, the two first rows of S^t take care of the translations in x and y direction. The third row defines the rotation about the vertical axis and the fourth row adjusts the scale of the observations to that given by the approximate coordinates. This interpretation coincides with the datum transformation according to Section 3.2. If some observations contain datum information (see Table 3.1), then the corresponding rows of S^t are omitted.

This matrix of constraints S^t defines the datum such that the network determined by the observations is positioned relative to the cluster of the approximate positions of the points and so that the criterion $\hat{x}^t\hat{x} = \min$ is satisfied. The overdetermined similarity transformation, sometimes called "Helmert-transformation" satisfies the same criterion; instead of $\hat{x}$ the discrepancies in the identical points are introduced.

In this version all approximate coordinates contribute equally to the geodetic datum. If certain points are to be excluded from contributing to the datum, the corresponding coefficients of S^t are simply replaced by zeros. This is most efficiently done by use of a selection matrix B according to Section 3.8.

A **3d-network** has the maximum rank deficiency of d = 7, hence a 7 x u - matrix S^t of constraints is required. Under the assumption of the coordinates being arranged as:

$$x^t = (x_1,\ y_1,\ z_1,\ x_2,\ y_2,\ z_2,\ \ldots\ x_p,\ y_p,\ z_p)$$

S^t has the form:

$$S^t = \begin{bmatrix} +1 & 0 & 0 & +1 & 0 & 0 & \dots \\ 0 & +1 & 0 & 0 & +1 & 0 & \dots \\ 0 & 0 & +1 & 0 & 0 & +1 & \dots \\ 0 & +\bar{z}_1 & -\bar{y}_1 & 0 & +\bar{z}_2 & -\bar{y}_2 & \dots \\ -\bar{z}_1 & 0 & +\bar{x}_1 & -\bar{z}_2 & 0 & +\bar{x}_2 & \dots \\ +\bar{y}_1 & -\bar{x}_1 & 0 & +\bar{y}_2 & -\bar{x}_2 & 0 & \dots \\ +\bar{x}_1 & +\bar{y}_1 & +\bar{z}_1 & +\bar{x}_2 & +\bar{y}_2 & +\bar{z}_2 & \dots \end{bmatrix} \begin{matrix} \Rightarrow \Sigma\hat{x}_i = 0 \\ \Rightarrow \Sigma\hat{y}_i = 0 \\ \Rightarrow \Sigma\hat{z}_i = 0 \\ \Rightarrow \Sigma(\bar{z}_i\hat{y}_i - \bar{y}_i\hat{z}_i) = 0 \\ \Rightarrow \Sigma(\bar{x}_i\hat{z}_i - \bar{z}_i\hat{x}_i) = 0 \\ \Rightarrow \Sigma(\bar{y}_i\hat{x}_i - \bar{x}_i\hat{y}_i) = 0 \\ \Rightarrow \Sigma(\bar{x}_i\hat{x}_i + \bar{y}_i\hat{y}_i + \bar{z}_i\hat{z}_i) = 0 \end{matrix} \quad (3\text{-}64)$$

The approximate coordinates $\bar{x}_i$, $\bar{y}_i$ and $\bar{z}_i$ are again to refer to the centre of gravity of the network followed by a normalization of S^t to achieve numerical stability. The computations are along the lines described for the 2d-network.

The first three rows of S^t dispose of the translations along x, y and z, respectively. The next three rows define the free rotations about x, y and z, respectively, and the last row defines the scale of the network.

The remarks on the meaning and the possible modification of the datum selection given with the 2d-case apply here as well.

In photogrammetry and special fields of geodesy (e.g. satellite geodesy), the geometry of networks may not be as clear as in the cases considered so far. Quite often it is not even possible to find the rank deficiency d from geometrical considerations, let alone the matrix S^t. In these cases the singular value decomposition (SVD) of the design matrix A is the most efficient method of finding d and S^t. A description of the SVD is commonly given in linear algebra and is beyond the scope of this monograph. The interested reader will find the necessary information in textbooks on linear algebra. Computer routines for the SVD are contained in several program libraries.

An alternative to the SVD is the eigenvalue decomposition of $N = A^tPA$ as introduced in Eq. (3-48). This method is numerically less stable than SVD and should be applied in the case of small matrices only.

A further method of determining S^t is based on Eqs (3-3) and (3-4). The matrix $A = (A_1 : A_2)$ is partitioned such that $r(A_1) = r(A)$ and $A_1L = A_2$, where the $u - r = d$ columns of A_2 are linear combinations of the r columns of A_1, expressed by a certain matrix L. This can be expressed as:

$$A_1L - A_2 = A \begin{pmatrix} L \\ -I_d \end{pmatrix} = AS = 0 \quad (3\text{-}65)$$

yielding:

$$S = \begin{pmatrix} L \\ -I_d \end{pmatrix} \quad (3\text{-}66)$$

L is not uniquely determined, but can be derived from $A_1L = A_2$:

$$L = (A_1{}^tGA_1)^{-1}A_1{}^tGA_2 \quad (3\text{-}67)$$

The matrix G can be chosen arbitrarily, provided that $r(A_1) = r(A_1{}^tGA_1)$ holds. A reliable approach is to replace G by the weight matrix P. This leads to:

$$S = \begin{pmatrix} N_{11}{}^{-1}N_{12} \\ -I_d \end{pmatrix} \tag{3-68}$$

which can easily be computed from the bordered normal equations (3-27) yielding the result of Eq. (3-29).

3.11 Geodetic Datum Problem in the Context of Linear Algebra

In the previous Chapters the datum problem has been treated on the basis of geometrical considerations wherever possible. The algebraic point of view has been neglected. This approach is perfectly adequate for a geodesist who is mainly interested in applications. Application oriented geodesists may therefore skip this section. This section will provide a broader view for those who are interested in the wider mathematical background and who are familiar with the symbols and terms of linear algebra.

1. The Regular (Cayley) Inverse

Let A be a real matrix with $o(A) = n \times n$, $r(A) = n \Rightarrow \exists\ A^{-1}$, $o(A^{-1}) = n \times n$, $r(A^{-1}) = n$ such that:

$$AA^{-1} = A^{-1}A = I_n \tag{3-69}$$

A^{-1} is the unique regular inverse of A. (The symbol $\exists$ is short for "there exists a ...".).

2. The One-Sided Inverse

Let A be a real matrix with $o(A) = n \times u$, $r(A) = \min\{n, u\}$:

i. If $r(A) = u < n \Rightarrow \exists\ A_L{}^{-1}$, $o(A_L{}^{-1}) = u \times n$, $r(A_L{}^{-1}) = u$ such that:

$$A_L{}^{-1}A = I_u \tag{3-70}$$

$A_L{}^{-1}$ is a left-inverse of A, which is not unique, since:

$$A_L{}^{-1} = (A^tPA)^{-1}A^tP \tag{3-71}$$

holds for any n x n - matrix P of rank n.

ii. If $r(A) = n < u \Rightarrow \exists\ A_R{}^{-1}$, $o(A_R{}^{-1}) = u \times n$, $r(A_R{}^{-1}) = n$ such that:

$$AA_R{}^{-1} = I_n \tag{3-72}$$

$A_R{}^{-1}$ is a right-inverse of A, which is not unique since:

$$A_R{}^{-1} = PA^t(APA^t)^{-1} \tag{3-73}$$

holds for any u x u - matrix P of rank u.

3. The Generalized Inverse

Let A be a real matrix with $o(A) = n \times u$, $r(A) = r, r \leq \min\{n, u\}$, $\Rightarrow \exists\ A^-$, $o(A^-) = u \times n$, $r \leq r(A^-) \leq \min\{n, u\}$ such that:

$$AA^-A = A \tag{3-74}$$

A^- is a g-inverse of A, which is not unique. Let $A = QR$ be a rank factorization, so that $o(Q) = n \times r$, $o(R) = r \times u$, $r(Q) = r(R) = r$, then:

$$A^- = R_R^{-1} Q_L^{-1} \tag{3-75}$$

is one selection of A^-.

4. Linear Equations

Let A be a real matrix with $o(A) = n \times u$ and A^- any g-inverse of A, then the general solution:

i. of the homogeneous equations $Ax = 0$ is:

$$x = (I - A^-A)z, \text{ with } z \text{ arbitrary} \tag{3-76}$$

ii. of the consistent non-homogeneous equations $Ax = y$ is:

$$x = A^-y + (I - A^-A)z, \text{ with } z \text{ arbitrary} \tag{3-77}$$

The proof follows from premultiplication by A.

5. The Reflexive g-Inverse

Let A be a real matrix with $o(A) = n \times u$ and $r(A) = r, r \leq \min\{n, u\}$, then a g-inverse A^- of A is called reflexive g-inverse of A if:

$$r(A) = r(A^-) \tag{3-78}$$

For reflexive g-inverses the following statements hold:

$$AA^-A = A \text{ and } A^-AA^- = A^- \tag{3-79}$$

Any g-inverse of A is reflexive if it can be expressed as:

$$A^- = A_i^- A A_j^- \tag{3-80}$$

where A_i^- and A_j^- are arbitrary g-inverses. Conversely, if two not necessarily different g-inverses are used to form the product of Eq. (3-80) a reflexive g-inverse originates.

6. LS-Solution of Inconsistent Linear Equations

Let A be a real matrix with $o(A) = n \times u$, $r(A) = r \leq u \leq n$ and let:

$$Ax \doteq y \qquad (\doteq \text{ means 'approximately'}) \tag{3-81}$$

be a set of inconsistent equations, then $\hat{x}$ is a LS-solution of Eq. (3-81) if:

$$\| A\hat{x} - y \| = [(A\hat{x} - y)^t (A\hat{x} - y)]^{1/2} = \min \tag{3-82}$$

Assuming $Gy = \hat{x}$ to be a LS-solution, then by definition of Eq. (3-82):

$$\| AGy - y \| \leq \| Ax - y \| \qquad \forall\, x \in \mathbb{R}^u,\ y \in \mathbb{R}^n$$

(The symbols $\forall$,$\in$ and $\mathbb{R}^i$ are short for "for all", "elements of ..." and "domain of all real vectors of dimension i", respectively.).

Now, substitute $x = \hat{x} + w \Rightarrow$

$$\| AGy - y \| \leq \| A\hat{x} + Aw - y \|$$

$$\leq \| (AGy - y) + Aw \|$$

$$\| AGy - y \|^2 \leq (AGy - y)^t(AGy - y) + w^tA^tAw + 2(AGy - y)^tAw$$

This inequality is correct if the last term on the right hand side disappears.

$$(AGy - y)^tAw = 0 \qquad \forall\, w \in \mathbb{R}^u,\ y \in \mathbb{R}^n$$

Hence, G must be such that:

$$G^tA^tA = A \tag{3-83}$$

which is equivalent to the two relationships:

$$AGA = A \text{ and } (AG)^t = AG \tag{3-84}$$

and which show that G is a g-inverse of A. G is usually called LS-inverse of A. Any matrix:

$$G = (A^tA)^-A^t \tag{3-85}$$

belongs to the class of LS-inverses of A. G is not unique, as any g-inverse of A^tA can be substituted in Eq. (3-85). Consequently the LS-solution $\hat{x}$ is also not unique. In fact, any vector of the set:

$$\{\hat{x}\} = Gy + (I - GA)z, \quad z \text{ arbitrary} \tag{3-86}$$

where G is given by Eq. (3-84), is a LS-solution of Eq. (3-81). The vectors $A\hat{x}$, $A\hat{x} - y$ and the residual norm $\| A\hat{x} - y \|$ are unique, which follows directly from Eq. (3-82). The set of all LS-inverses of A, forming a subset of all g-inverses, is given by:

$$\{G\} = G_o + (I - G_oA)U, \ U \text{ arbitrary} \tag{3-87}$$

where G_o is any element of {G} and U a matrix of suitable order.

7. Minimum Norm LS-Solution of Inconsistent Linear Equations

Let A be a real matrix with $o(A) = n \times u$, $r(A) = r \leq u \leq n$. Further, let:

$$Ax \doteq y$$

be a set of inconsistent linear equations and {G} the set of LS-inverses of A. To obtain the shortest solution vector ($\| \hat{x} \| = \min$) from the vectors of Eq. (3-86) (being LS-solutions of Eq. (3-81)), a specific LS-inverse $\bar{G}$ from the set of all LS-inverses (Eq. (3-87)) must be selected meeting the condition:

$$\| \bar{G}y \| \leq \| \hat{x} \| \qquad \forall\ \hat{x} \in \{\hat{x}\} \text{ and } y \in \mathbb{R}^n \qquad (3\text{-}88)$$

Substitution of Eq. (3-86) yields:

$$\| \bar{G}y \| \leq \| Gy + (I - GA)z \| \qquad \forall\ y \in \mathbb{R}^n,\ z \in \mathbb{R}^u$$

$$\Rightarrow \quad \| \bar{G}y \|^2 \leq \| Gy \|^2 + \| (I - GA)z \|^2 + 2y^t G^t (I - GA)z$$

This inequality is fulfilled if the last term on the right hand side disappears.

$$y^t G^t (I - GA)z = 0 \qquad \forall\ z \in \mathbb{R}^u,\ y \in \mathbb{R}^n$$

Hence, $\bar{G}$ must be such that:

$$\bar{G}^t = \bar{G}^t \bar{G} A \qquad (3\text{-}89)$$

which is equivalent to the two relationships:

$$\bar{G}A\bar{G} = \bar{G} \text{ and } (\bar{G}A)^t = \bar{G}A \qquad (3\text{-}90)$$

As $\bar{G}$ is restricted to being an element of the set of LS-inverses of Eq. (3-87), it meets the conditions of Eq. (3-84). Thus $\bar{G}$, usually denoted by A^+, is defined by the four well-known Moore-Penrose conditions:

$$\begin{aligned} AA^+A &= A \quad , \quad A^+AA^+ = A^+ \\ (AA^+)^t &= AA^+ \quad , \quad (A^+A)^t = A^+A \end{aligned} \qquad (3\text{-}91)$$

and termed Moore-Penrose inverse or pseudo-inverse of A.

A^+ is uniquely defined by Eq. (3-91) and can be calculated from:

$$A^+ = A^tA(A^tAA^tA)^-A^t = A^t(A^tAA^t)^-A^t \qquad (3\text{-}92)$$

where the involved g-inverses can be any element of the set of g-inverses as defined by Eq. (3-74).

8. Application of g-Inverses to the GMM

Let the GMM be defined as in Eq. (2-1):

$$\bar{A}x + \varepsilon = l, \quad \Sigma = E(\varepsilon\varepsilon^t) \qquad (2\text{-}1)$$

where the n x u - matrix $\bar{A}$ has rank $r \leq u \leq n$ and the n x n covariance matrix Σ is of full rank.

In order to apply the results of paragraph 7 of this section, Eq. (2-1) needs to be transformed into the form of Eq. (3-81), where y is a random vector with covariance-matrix $\Sigma_y \equiv I$. The transformation is based on the factorization of the positive definite matrix Σ analogous to that defined by Eq. (3-60):

$$\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^t, \qquad \Sigma^{-1} = (\Sigma^{-1/2})^t\Sigma^{-1/2} \qquad (3\text{-}93)$$

Premultiplication of the Eq. (2-1) by $\Sigma^{-1/2}$ and application of the law of propagation of variances yield:

$$\Sigma^{-1/2}\bar{A}x + \Sigma^{-1/2}\varepsilon = \Sigma^{-1/2}l$$

$$Ax + \bar{\varepsilon} = y$$

$$E(\bar{\varepsilon}\,\bar{\varepsilon}^t) = \Sigma_y = \Sigma^{-1/2}E(\varepsilon\varepsilon^t)(\Sigma^{-1/2})^t$$

$$\Sigma_y = \Sigma^{-1/2}\Sigma(\Sigma^{-1/2})^t = I$$

and hence the homogenized sample model:

$$A\hat{x} = y + w, \qquad \Sigma_y = I \tag{3-94}$$

Eq. (3-94) can be solved for by the Moore-Penrose inverse A^+:

$$\begin{aligned} \hat{x} &= A^+y, \qquad w = AA^+y - y \\ Q_{\hat{x}} &= A^+(A^+)^t \end{aligned} \tag{3-95}$$

The parameter vector $\hat{x}$ of the homogenized model given by Eq. (3-94) is identical to that of the original model in Eq. (2-1). The same applies for the quadratic form:

$$w^tw = v^t(\Sigma^{-1/2})^t\Sigma^{-1/2}v = v^t\Sigma^{-1}v$$

where $\sigma_o^2 = 1$. A comparison of the criteria for the solution of the linear model of paragraph 7 of this section with those of Section 3.7, as compiled in the summary shows that Eqs (3-95) are nothing else but the least squares solution for the minimum trace datum.

This section demonstrates that the datum problem is not restricted to geodetic networks. It is a more general problem of the singular GMM or of inconsistent linear equations with a singular coefficient matrix. The general solution in Eq. (3-95) has the disadvantage of being less easily visualized than the solutions of Sections 3.4 and 3.7 which are based on constraints. A proper interpretation is therefore difficult.

9. Computation of a g-Inverse

Let A be a real matrix with o(A) = n x u and r(A) = r, $r \leq u \leq n$, then a simple method of computing a g-inverse of A consists of three steps:

i. partition A such that A_{11} is a r x r - matrix of rank r:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

ii. compute A_{11}^{-1}

iii. form the matrix:

$$A^{-} = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

of order u x n, which is a reflexive g-inverse of A.

In practice, step (i) does not have to be carried out at the beginning, if an inversion algorithm with pivot strategy is used; in such a case, the rearrangement of A is done during the computations.

10. Exercises

To understand how the geodetic datum problem fits into the general context, show that:

i. $Q_{11}A^t$ is a LS-inverse of A for any properly chosen set of constraints $R^t\hat{x} = 0$.

ii. $Q_{11}A^t = A^+$ for the constraints $S^t\hat{x} = 0$, $AS = 0$.

iii. $Q_{11} = N^+$ for the constraints $S^t\hat{x} = 0$, $AS = 0$.

iv. Compute A^+, N^+ and the set $\{\hat{x}\}$ of LS-solutions and the set $\{G\}$ of LS-inverses for the numerical example of Section 3.9.

v. Compute the g-inverses N_1^- by rank factorization (sub-section 4), N_2^- according to sub-section 9 and show that $N_3^- = N_1^-NN_2^-$ is reflexive.

vi. Show that the matrices $G_i = N_i^-A^t$, N_i from (v), are elements of the set $\{G\}$ of LS-inverses of A.

4. MEASURES OF ACCURACY IN GEODETIC NETWORKS

In many cases the deformations to be monitored by geodetic methods are small; they often are of the same magnitude as the observational errors. The network design and the assessment of the achieved accuracy therefore play an important role in deformation analysis.

Practical considerations require coordinates of points to be selected as the parameters in the GMM of the adjustment of geodetic networks. The previous section demonstrated that the coordinates comprised in the vector $\hat{x}$ as well as the corresponding variance-covariance matrix $\Sigma_{\hat{x}}$ are not invariant quantities of the model; they depend on the selection of the geodetic datum. The fact that these results are not unique causes some problems in assessing the precision of the adjusted positions of points.

The u x u-matrix of variances and covariances contains the information about the accuracy of a network:

$$\Sigma_{\hat{x}} = \sigma_o^2 Q_{\hat{x}}, \qquad r(\Sigma_{\hat{x}}) = r = u - d \tag{4-1}$$

where $Q_{\hat{x}}$ is defined by Eqs (3-25) and (3-23). In a free network $\Sigma_{\hat{x}}$ has rank deficiency d, hence d columns of $\Sigma_{\hat{x}}$ are redundant. There is an infinite number of ways of overcoming the redundancy, each resulting in a different base of the corresponding r-dimensional vector space.

Many methods have been developed to cope with the problem of non-uniqueness, but none is entirely satisfying. The situation is better if a comparison of two covariance matrices is aimed at. There exist relative measures which are datum independent.

An additional but minor problem is the selection of the variance factor in Eq. (4-1). If the variances of the observations are known and if the GMM is correct, then σ_o^2 is of course the factor to be used. But quite often the a priori information is vague and the model is only the best available approximation of reality. Then s_o^2 as defined in Eq. (2-7) is preferred, provided that the model has enough degrees of freedom, i.e. the ratio $(n - r)/r$ is greater than 0.5.

It follows from Eq. (4-1) that there are three possibilities of influencing the accuracy:

i. The variance factor can be controlled by the selection of instruments and the repetition number of multiple observations.

ii. The matrix A depends on the geometry of the network, i.e. the relative position of points and the connecting observations.

iii. The weight matrix P contains the a priori weights which are functions of the type of observables and the relative precision.

Chapter 9 is devoted to the optimal design of monitoring networks. There, these problems will be addressed in more detail.

It is not practical to assess the accuracy of a network by inspecting or comparing the covariance matrices, as the number of elements of the u x u-matrix $\Sigma_{\hat{x}}$ is too large. Therefore, the mass of information must be compressed into one or few representative measures which can be assessed easily. Alternatively, graphical methods can be developed which display the information comprised in $\Sigma_{\hat{x}}$, or at least an essential part of it. Unfortunately, both approaches lead to a loss of information. Thus the purpose of the network must be considered, to ensure that the desired information is not lost.

4.1 Eigenvalue Decomposition

The fundamental approach to the analysis of covariance matrices is the eigenvalue decomposition of $\Sigma_{\hat{x}}$.

If x is the u-vector of parameters of the GMM, $\hat{x}$ the corresponding least squares estimates and $\Sigma_{\hat{x}}$ the covariance matrix of rank r, then the quadratic form:

$$(x-\hat{x})^t\Sigma_{\hat{x}}^-(x-\hat{x}) = c \qquad (4\text{-}2)$$

is the equation of a r-dimensional hyperellipsoid with centre in x. $\Sigma_{\hat{x}}^-$ is a generalized inverse of $\Sigma_{\hat{x}}$ as defined in Eq. (3-74). Considering Eq. (3-21) the simplest selection of the generalized inverse is:

$$\Sigma_{\hat{x}}^- = 1/\sigma_o^2 N, \qquad N = A^tPA \qquad (4\text{-}3)$$

Under the assumption of normality (refer to Eq. (2-8)):

$$\hat{x} \sim N(x, \Sigma_{\hat{x}}) \qquad (4\text{-}4)$$

for σ_o^2 known, the quadratic form of Eq. (4-2) has a χ^2-distribution with r degrees of freedom:

$$(x-\hat{x})^t\Sigma_{\hat{x}}^-(x-\hat{x}) \sim \chi^2(r) \qquad (4\text{-}5)$$

After selecting a specific probability, e.g. $1-\alpha = 95\%$, the corresponding value $\chi^2_{1-\alpha}(r)$ of the probability function can be calculated. This value will be exceeded only in $\alpha\%$ of all cases provided that the underlying assumptions are correct. Thus the quadratic form:

$$(x-\hat{x})^t\Sigma_{\hat{x}}^-(x-\hat{x}) = \chi^2_{1-\alpha}(r) \qquad (4\text{-}6)$$

can be interpreted statistically as the equation of a $(1-\alpha)$-confidence hyperellipsoid in the r-dimensional space. This hyperellipsoid is defined by its r semi-axes and their directions in space. Since the lengths of these axes are proportional to the square roots of the eigenvalues of $\Sigma_{\hat{x}}$ and since the directions are given by the eigenvectors, the importance of the eigenvalue decomposition in this context is obvious. A more detailed interpretation of Eq. (4-5) follows in later sections.

The basic eigenvalue equation is:

$$(\Sigma_{\hat{x}} - \lambda_i I)s_i = 0, \qquad i \in \{1, 2, ..., u\} \qquad (4\text{-}7)$$

where λ_i is the i-th eigenvalue or latent root of $\Sigma_{\hat{x}}$ and s_i its associated eigenvector. Since $\Sigma_{\hat{x}}$ is symmetric and positive semi-definite by definition, there exist $r(\Sigma_{\hat{x}}) = r$ different positive eigenvalues and a d-fold ($d = u - r$) eigenvalue zero. Associated with these eigenvalues exist u different mutual orthogonal eigenvectors s_i. The homogeneous equation (Eq. (4-7)) has a non-trivial solution s_i if and only if the matrix $(\Sigma_{\hat{x}} - I\lambda_i)$ is singular, i.e. if the determinant is zero.

$$\det(\Sigma_{\hat{x}} - I\lambda_i) = 0 \qquad (4\text{-}8)$$

Equation (4-8) is called the characteristic equation of $\Sigma_{\hat{x}}$. The elements of this matrix are as follows:

$$\det \begin{bmatrix} (\sigma_{11}-\lambda_i) & \sigma_{12} & \cdots & \sigma_{1u} \\ \sigma_{21} & (\sigma_{22}-\lambda_i) & \cdots & \sigma_{2u} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u1} & \sigma_{u2} & \cdots & (\sigma_{uu}-\lambda_i) \end{bmatrix} = 0 \qquad (4\text{-}8)$$

The development of the determinant leads to a polynomial of u-th order in λ_i the so-called characteristic polynomial. The solution for λ_i is simple for small number of u only. But efficient algorithms for large u are available for all modern computers. Details on direct and iterative methods for the solution of Eq. (4-8) can be found in textbooks on linear algebra.

Once the eigenvalues are computed the linear homogeneous set of equations (Eq. (4-7)) can be solved for s_i by standard elimination methods. Since s_i is defined except for an arbitrary constant, its normalized form is usually given:

$$\| s_i \| = (s_i^t s_i)^{1/2} = 1 \qquad (4\text{-}9)$$

If a diagonal matrix Λ of the eigenvalues is formed and a u x u-matrix S of eigenvectors:

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_u \end{pmatrix}, \qquad S = (s_1, s_2, \ldots, s_u) \qquad (4\text{-}10)$$

then the n equations (Eq. (4-7)) can be replaced by a single one:

$$\Sigma_{\hat{x}} S = S\Lambda \qquad (4\text{-}11)$$

Under the assumption of the eigenvectors being normalized according to Eq. (4-9) and considering the orthogonality of the eigenvectors, the relation:

$$S^t S = S S^t = I \qquad (4\text{-}12)$$

exists, yielding in conjunction with Eq. (4-11), a very compact form of the eigenvalue problem:

$$S^t \Sigma_{\hat{x}} S = \Lambda, \qquad \Sigma_{\hat{x}} = S \Lambda S^t \qquad (4\text{-}13)$$

One property of the eigenvalues is of great interest in the present context, namely their invariance with respect to similarity transformations. Thus conventional datum transformations according to Section 3.2 do not change the eigenvalues and hence retain all measures of accuracy which are based on them. A note of caution may be appropriate with regard to the general datum transformation of Section 3.9; this transformation is not a similarity transformation in the sense of linear algebra and consequently alters the eigenvalues.

Usually, the eigenvalues are listed in decreasing order, i.e. $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ which simplifies the analyses which are often based on the extreme values only. Each pair of $\lambda_i s_i$ defines one axis of the hyperellipsoid of Eq. (4-2) by means of the length of the semi-axis $a_i = \sqrt{c\lambda_i}$ and its direction s_i.

More details follow in Sections 4.3 and 4.4. Some further properties of the eigenvalues λ_i of special interest for the accuracy analysis, are given below.

$$\sum_{i=1}^{u} \lambda_i = \operatorname{tr} \Sigma_{\hat{x}} ; \qquad \prod_{i=1}^{u} \lambda_i = \det \Sigma_{\hat{x}} \tag{4-14}$$

$$\frac{s_i^t \Sigma_{\hat{x}} s_i}{s_i^t s_i} = \lambda_i , \qquad s_i = \text{eigenvector of } \Sigma_{\hat{x}} \tag{4-15}$$

$$\lambda_{min} \leq \frac{f^t \Sigma_{\hat{x}} f}{f^t f} \leq \lambda_{max} \qquad \forall f \in \mathbb{R}^u \tag{4-16}$$

Equation (4-14) states that the sum of all eigenvalues equals the trace of the corresponding covariance matrix. The product of the eigenvalues symbolized by capital pi in Eq. (4-14) equals the determinant, and Eq. (4-16), the so-called Rayleigh relation, serves to estimate the extreme values of certain functions $f^t x$ of the parameter vector.

4.2 Generalized Eigenvalue Decomposition

The generalized eigenvalue problem can serve as a basis for the comparison of two covariance matrices. Therefore, a short presentation may help the understanding of its use in Section 4.5.

If A and B are two quadratic matrices, the generalized eigenvalues λ_i and the generalized eigenvectors s_i of A in respect of B are defined analogous to Eq. (4-7) by:

$$(A - \lambda_i B) s_i = 0 \tag{4-17}$$

For B = I Eq. (4-17) is reduced to the ordinary eigenvalue equation. If Eq. (4-17) is premultiplied by B^{-1}, supposing that B is regular, then the result:

$$(B^{-1}A - \lambda_i) s_i = 0 \tag{4-18}$$

shows that the generalized eigenvalue of A with respect to B is the same as the ordinary eigenvalue of $B^{-1}A$.

The same applies to the generalized eigenvector s_i.

4.3 Local Measures of Accuracy

The term accuracy as a criterion of quality of a geodetic network is a rather vague one. In a general purpose network, such as a national control network, another measure of accuracy is required than in an engineering network serving a special purpose, such as the monitoring of deformations. A variety of measures is in use reflecting different aspects of accuracy, all being derived from the previously introduced covariance matrix (Eq. (4-1)) $\Sigma_{\hat{x}} = \sigma_o^2 Q_{\hat{x}}$, where:

$$Q_{\hat{x}} = \begin{bmatrix} Q_{11} & Q_{12} & \cdots & Q_{1i} & \cdots & Q_{1j} & \cdots & Q_{1p} \\ Q_{21} & Q_{22} & \cdots & Q_{2i} & \cdots & Q_{2j} & \cdots & Q_{2p} \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ Q_{i1} & Q_{i2} & \cdots & Q_{ii} & \cdots & Q_{ij} & \cdots & Q_{ip} \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ Q_{j1} & Q_{j2} & \cdots & Q_{ji} & \cdots & Q_{jj} & \cdots & Q_{jp} \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ Q_{p1} & Q_{p2} & \cdots & Q_{pi} & \cdots & Q_{pj} & \cdots & Q_{pp} \end{bmatrix} \qquad (4\text{-}19)$$

is partitioned (in 2d-networks) in 2 x 2-block matrices, each referring to one point or a pair of points according to:

$$Q_{kk} = \begin{bmatrix} {q_{xx}}^{k} & {q_{xy}}^{k} \\ {q_{yx}}^{k} & {q_{yy}}^{k} \end{bmatrix} \quad \text{or} \quad Q_{kl} = \begin{bmatrix} {q_{xx}}^{kl} & {q_{xy}}^{kl} \\ {q_{yx}}^{kl} & {q_{yy}}^{kl} \end{bmatrix} \qquad (4\text{-}20)$$

In the following a 2d-network in an x,y-coordinate system is considered; a generalization to other frames or dimensions is straightforward. An immediate local measure of accuracy is the **standard deviation of the coordinates.** In case of a known population variance σ_o^2,

$$\sigma_x = \sigma_o\sqrt{q_{xx}}\ , \qquad \sigma_y = \sigma_o\sqrt{q_{yy}} \qquad (4\text{-}21)$$

are readily computed, where the cofactors q are the pertinent diagonal elements of $Q_{\hat{x}}$. Whenever σ_o is unknown, the sample variance s_o^2 is computed according to Eq. (2-7) yielding:

$$s_x = s_o\sqrt{q_{xx}}\ , \qquad s_y = s_o\sqrt{q_{yy}} \qquad (4\text{-}22)$$

These standard deviations are often called mean square errors of the coordinates. Since the assessment of the standard deviations depends on the redundancy of the GMM, it is strongly recommended to replace Eqs (4-21) and (4-22) by the respective **confidence intervals.** To this end the probability distributions of the population and of the ratio $(x - \hat{x})/s_x$ is needed.

If a normal population is assumed (refer to Eqs (2-8) and (4-4)), then the difference between the unknown parameter and its estimate:

$$(x - \hat{x}) \sim N(0,\ \sigma_x) \qquad (4\text{-}23)$$

is normally distributed according to Eq. (4-23), and the probability relationship:

$$P\{-C_{\alpha/2} < (x - \hat{x})/\sigma_x < C_{\alpha/2}\} = 1 - \alpha \qquad (4\text{-}24)$$

states, that the parameter x is (with a probability of $1 - \alpha$) inside the interval given by:

$$P\{\hat{x} - C_{\alpha/2}\sigma_x < x < \hat{x} + C_{\alpha/2}\sigma_x\} = 1 - \alpha \qquad (4\text{-}25)$$

A typical value is $\alpha = 5\%$ for which the $\alpha/2$ point $C_{\alpha/2} = 1.96$ from tables of the normal distribution, such that:

$$P\{\hat{x} - 1.96\sigma_x < x < \hat{x} + 1.96\sigma_x\} = 95\% \qquad (4\text{-}26)$$

defines the 95%-confidence interval for $\hat{x}$, replacing the less informative measure of Eq. (4-21).

If, under the same assumptions as before, only the sample variance s_x^2 is known, a confidence interval can be derived from the t-distribution:

$$(x - \hat{x})/s_x \sim t(n - r) \qquad (4\text{-}27)$$

with $n - r$ degrees of freedom. Analogous to Eqs (4-24) and (4-25) the probability statements can be given as:

$$P\{-t_{\alpha/2} < (x - \hat{x})/s_x < t_{\alpha/2}\} = 1 - \alpha \qquad (4\text{-}28)$$

$$P\{\hat{x} - t_{\alpha/2}s_x < x < \hat{x} + t_{\alpha/2}s_x\} = 1 - \alpha \qquad (4\text{-}29)$$

where the value $t_{\alpha/2}$ depends on the confidence level $1 - \alpha$ and the redundancy of the underlying GMM. Tables of the t-distribution can be found in textbooks on statistics.

$n - r$	2	4	6	8	10	15	20	25	30	40	120	∞
$\alpha = 5\%$	4.30	2.78	2.45	2.31	2.23	2.13	2.09	2.06	2.04	2.02	1.98	1.96
$\alpha = 1\%$	9.92	4.60	3.71	3.36	3.17	2.95	2.84	2.79	2.75	2.70	2.62	2.58

Table 4.1: Values $t_{\alpha/2}$ for selected combinations of α and $n - r$. The last column lists the corresponding value $C_{\alpha/2}$ of the normal distribution.

In special purpose networks certain functions are often of special interest. Examples are the azimuth and length of an axis for a bridge or a dam, or selected directions and distances being sensitive to expected deformations. These functions f(x) of the parameters (x) must be linearized to yield a form:

$$\begin{aligned} f(x) &= f_0(x) + f_1x_1 + f_2x_2 + \ldots + f_nx_n \\ &= f_0(x) + f^tx \end{aligned} \qquad (4\text{-}30)$$

which is suitable for the application of the law of variance propagation:

$$q_f = f^tQ_{\hat{x}}f$$

and:

$$\sigma_f = \sigma_o\sqrt{q_f} \qquad \text{or} \qquad s_f = s_o\sqrt{q_f} \qquad (4\text{-}31)$$

The population and sample standard deviations of the function are denoted by σ_f and s_f, respectively. The more informative confidence intervals of f(x) are easily established using Eqs (4-25) or (4-29).

In practice it is most popular to display the precision of a network by its 2-dimensional **standard** or **confidence point error ellipses**. The terminology shall be clarified first: Standard ellipses are generalizations of the standard deviation (refer to Eqs (4-21) and (4-22)) while confidence ellipses are the 2-dimensional equivalent of confidence intervals (refer to Eqs (4-25) and (4-29)). The ellipses are termed absolute if they refer to one point and relative if they refer to the position difference of two points.

In Section 4.1 the quadratic form:

$$(x-\hat{x})^t\Sigma_{\hat{x}}^{-}(x-\hat{x}) = \chi^2{}_{1-\alpha}(r) \qquad (4\text{-}6)$$

has been represented as a $(1-\alpha)$-confidence hyperellipsoid in the r-dimensional space. Now, specializing Eq. (4-6) for the 2-dimensional position of a single point P_k and multiplying by σ_o^2 yields:

$$(x-\hat{x})^t Q_{kk}^{-1}(x-\hat{x}) = \sigma_o^2\chi^2{}_{1-\alpha}(2) \qquad (4\text{-}32)$$

This is the equation of an error ellipse with axes being functions of the eigenvalues and eigenvectors of Q_{kk}.

According to Eq. (4-7) the eigensystem is:

$$(Q_{kk}-\lambda_i I)s_i = 0\ , \qquad i = 1,2 \qquad (4\text{-}33)$$

and can be solved for λ_i by using the characteristic polynomial:

$$\det\begin{bmatrix} q_{xx}{}^k-\lambda & q_{xy}{}^k \\ q_{yx}{}^k & q_{yy}{}^k-\lambda \end{bmatrix} = 0$$

$$(q_{xx}{}^k-\lambda)(q_{yy}{}^k-\lambda) - q_{xy}{}^k q_{yx}{}^k = 0$$

which has the two roots:

$$\lambda_1 = 1/2(q_{xx}{}^k + q_{yy}{}^k + z)$$

$$\lambda_2 = 1/2(q_{xx}{}^k + q_{yy}{}^k - z)$$

where:

$$z^2 = (q_{xx}{}^k - q_{yy}{}^k)^2 + 4q_{xy}{}^k q_{xy}{}^k \qquad (4\text{-}34)$$

Substitution in Eq. (4-33) results in the equations of the orthogonal eigenvectors $s_i = (s_{i1},\ s_{i2})^t$:

$$\begin{aligned} (q_{xx}{}^k-\lambda_i)s_{i1} + q_{xy}{}^k s_{i2} &= 0 \\ q_{yx}{}^k s_{i1} + (q_{yy}{}^k-\lambda_i)s_{i2} &= 0 \end{aligned} \qquad i = 1,2 \qquad (4\text{-}35)$$

from which the eigendirections follow:

$$\tan \alpha_i = \frac{s_{i2}}{s_{i1}} \text{ with } \alpha_2 = \alpha_1 + 90° \tag{4-36}$$

More convenient is the expression:

$$\tan 2\alpha = 2q_{xy}^k/(q_{xx}^k - q_{yy}^k) \tag{4-37}$$

where α is the direction (bearing) of the major semi-axis of the ellipse.

It has been assumed so far that the population variance σ_o^2 is known, hence Eq. (4-6) gives directly the confidence ellipse, defined by the semi-axes:

$$\begin{aligned} a &= \sigma_o \sqrt{\lambda_1 \chi^2_{1-\alpha}(2)} \\ b &= \sigma_o \sqrt{\lambda_2 \chi^2_{1-\alpha}(2)} \end{aligned} \tag{4-38}$$

and the direction α from Eq. (4-37). For the most frequently used significance levels of 95% and 99% the values of $\chi^2_{1-\alpha}(2)$ are 5.99 and 9.21, respectively.

The standard ellipse is derived from Eq. (4-32) by replacing the value of the χ^2-distribution by unity. The definition then reads:

$$(x - \hat{x})^t Q_{kk}^{-1} (x - \hat{x}) = \sigma_o^2 \tag{4-39}$$

This simplification does not affect the eigensystem. It only alters the semi-axes into:

$$a = \sigma_o \sqrt{\lambda_1} \qquad \text{and} \qquad b = \sigma_o \sqrt{\lambda_2} \tag{4-40}$$

The probability α that the true position of P_k is within the ellipse as defined by Eq. (4-39) is only 39.4%.

Usually, however, the a posteriori variance or sample variance s_o^2 is the best estimate of the dispersion of the observations. In this case the distributional assumption of Eq. (4-5) no longer applies. The quadratic form now has a Fisher-(F-)distribution with two and (n – r) degrees of freedom. This leads to the following definition of the confidence ellipse:

$$(x - \hat{x})^t Q_{kk}^{-1} (x - \hat{x}) = 2 s_o^2 F_{1-\alpha}(2,\ n - r) \tag{4-41}$$

where $F_{1-\alpha}(2, n - r)$ is the $(1 - \alpha)$-point of the F-distribution. A comparison with Eq. (4-32) shows that the same eigenvalues and eigenvectors apply; only the length of the axes of the ellipses differ. The standard ellipse is again a simplification of Eq. (4-41), where $2F_{1-\alpha}$ is set to one. The elements of the sample ellipses are:

confidence ellipse:

$$a = s_o\sqrt{2\lambda_1 F_{1-\alpha}(2,\ n-r)}$$

$$b = s_o\sqrt{2\lambda_2 F_{1-\alpha}(2,\ n-r)}$$

$$\tan 2\alpha = 2q_{xy}/(q_{xx} - q_{yy}) \qquad (4\text{-}42)$$

standard ellipse:

$$a = s_o\sqrt{\lambda_1}\ , \qquad b = s_o\sqrt{\lambda_2} \qquad (4\text{-}43)$$

The probability, that the true point is inside the standard ellipse of the estimated point, is a function of the degrees of freedom (n – r) and reaches its maximum of 39.4% for (n – r) = ∞.

Sometimes the ellipses are approximated by **circular regions** simplifying the required computations. There is a variety of definitions of error circles derived from standard deviations or confidence intervals either based on the population or the sample variance. The most commonly used error circles have the following radii:

i. circular standard error, probability ≈ 39%:

$$s_p = 1/2(s_x + s_y) \qquad (4\text{-}44)$$

ii. circular probable error, probability ≈ 50%:

$$s_p = 0.59(s_x + s_y) \qquad (4\text{-}45)$$

iii. mean square positional error, probability depends on the ratio s_x/s_y (Helmert's point error):

$$s_p = \sqrt{s_x^2 + s_y^2} = \sqrt{a^2 + b^2} = s_o\sqrt{\lambda_1 + \lambda_2} = s_o\sqrt{\text{tr } Q_{kk}} \qquad (4\text{-}46)$$

iv. generalized point error, probability depends on the ratio s_x/s_y (Werkmeister's point error):

$$s_p = a \cdot b/s_o = s_o\sqrt{\lambda_1\lambda_2} = s_o\sqrt{\det Q_{kk}} \qquad (4\text{-}47)$$

The absolute (point) error ellipses and the related circular regions have the disadvantage of neglecting the inter-point covariances. This problem is partially overcome by using **relative error ellipses** which express the relative accuracy of a pair of, mostly adjacent, points. The elements of these ellipses are computed as outlined for the absolute ellipses with the only difference that the cofactor matrix refers to the coordinate differences of the involved points:

$$Q_{kl}^{\Delta} = Q_{kk} + Q_{ll} - Q_{kl} - Q_{lk} \qquad (4\text{-}48)$$

Equations (4-39) through (4-43) apply for the computation of the axes and orientations of the relative ellipses which are usually plotted around the midpoint of the line connecting the involved points.

All measures presented so far suffer from being datum dependent. In the case of a conventional datum the datum defining points are error-free, i.e. standard deviations, ellipses and circles vanish. For all other points these measures increase with the distance from the datum points. This effect is less pronounced with the relative ellipses and can be completely avoided with standard deviations of functions, if solely invariant functions according to Section 3.6 are considered. The situation changes completely if a minimum trace datum (see Section 3.7) is used. Then, all coordinates and points feature measures of accuracy which grow only slowly from the centre of the network to the boundary. They represent the so-called inner precision of the network.

The interpretation of the introduced local measures of accuracy is difficult and mostly not clear at all. They can indicate weak zones in the network design and are an important tool for the comparison between different designs and methods. The accuracy of a position, however, is only meaningful as an accuracy relative to the geodetic datum. Therefore, a clear understanding is only possible if the datum has been defined conventionally (Sections 3.1 and 3.3). In this case the standard deviations, error ellipses and circles reflect the accuracy relative to the zero-variance computational base.

Figures 4.1 to 4.5 show a selection of local measures of accuracy for different selections of the geodetic datum. All plots are based on the same network design and the same set of observational data. The extreme changes of the size of some ellipses underline how doubtful the commonly used measures of positional accuracy are if no clear statement specifying the datum is added. Since no better measures are available, geodesist have to live with this problem and should exercise due care when assessing accuracy.

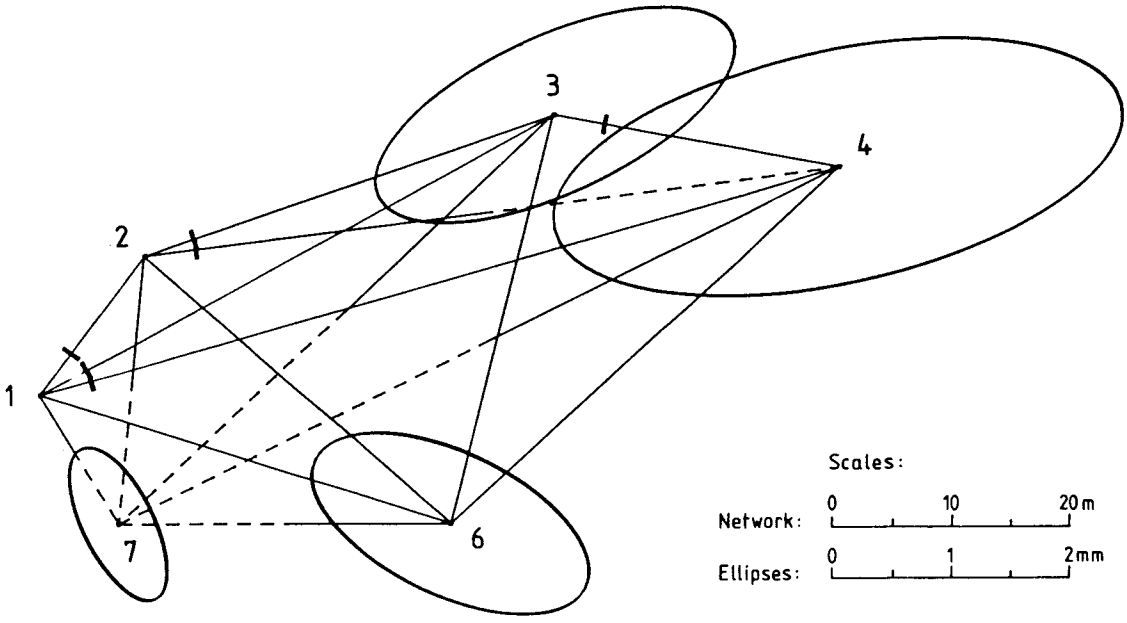


Figure 4.1: Reference points of the monitoring network Montsalvens (compare Section 6.5). 30 observations, 18 unknowns, 4 rank deficiencies. 95%-confidence point error ellipses from an adjustment with zero-variance computational base (Section 3.3) P_1, P_2.

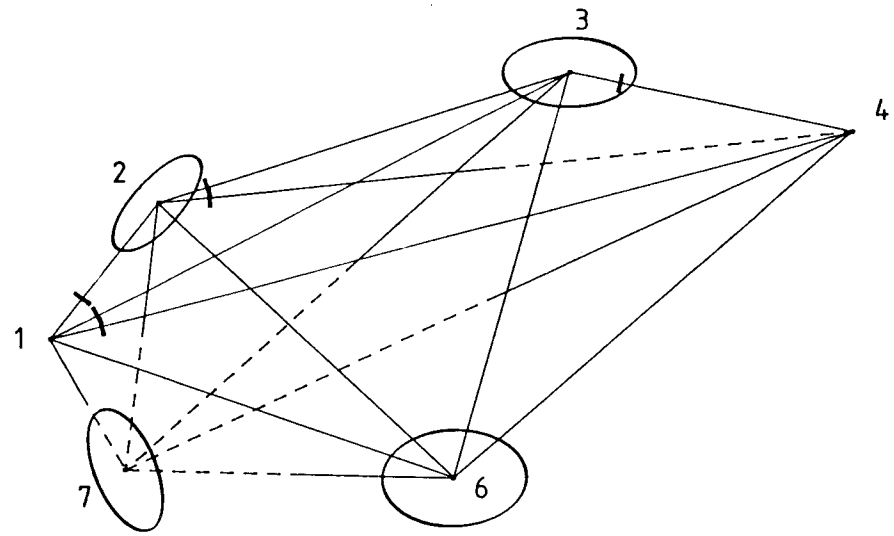


Figure 4.2: 95%-confidence ellipses of the network of Figure 4.1 from an adjustment with zero-variance computational base P_1 and P_4. (Same scale as Figure 4.1.)

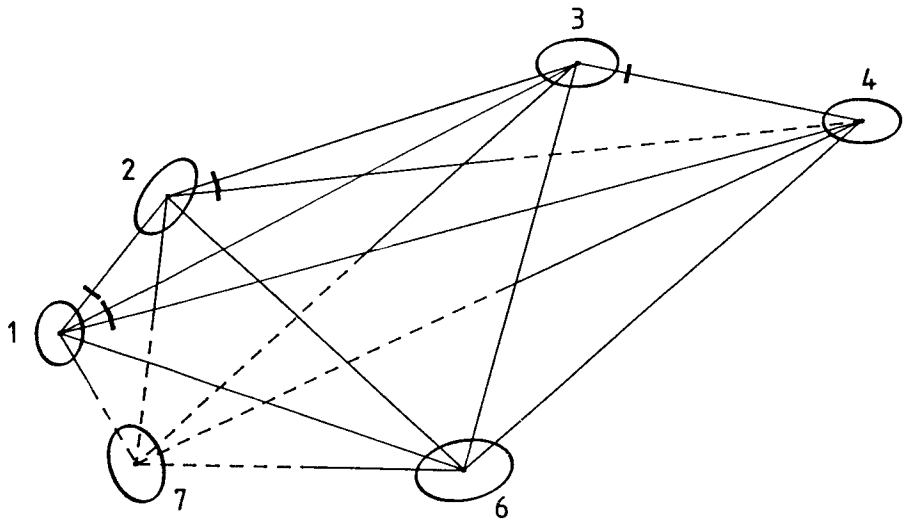
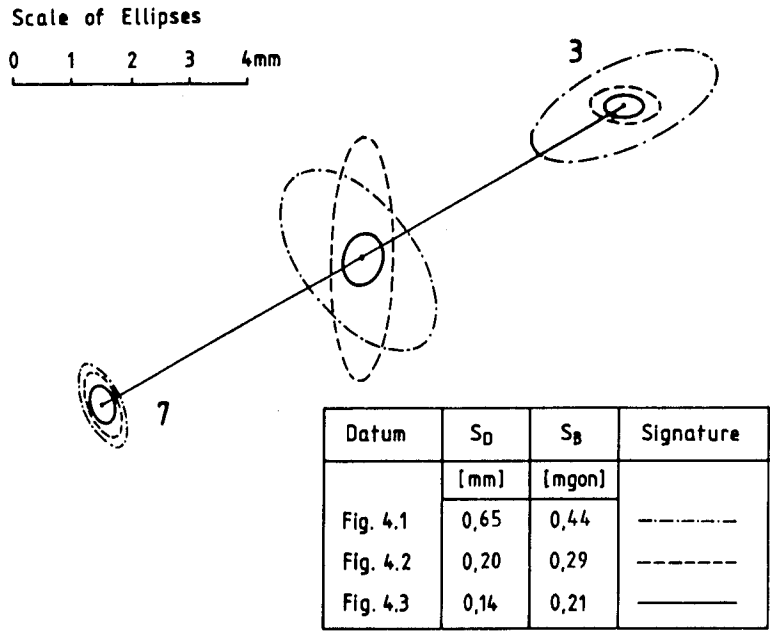


Figure 4.3: 95%-confidence ellipses of the network of Figure 4.1 from an adjustment according to Section 3.7 (tr $Q_{\hat{x}}$ = min). (Same scale as Figure 4.1.)



Datum	S_D	S_B	Signature
	[mm]	[mgon]	
Fig. 4.1	0,65	0,44	-·-·-·-
Fig. 4.2	0,20	0,29	- - - - - - -
Fig. 4.3	0,14	0,21	———

Figure 4.4: Absolute and relative confidence ellipses (95%) of P_3 and P_7 of the network of Figure 4.1 for the three geodetic datum as defined in Figures 4.1, 4.2 and 4.3. The table shows the standard deviations of the distance (s_D) and the bearing (s_B) of the line P_3P_7.

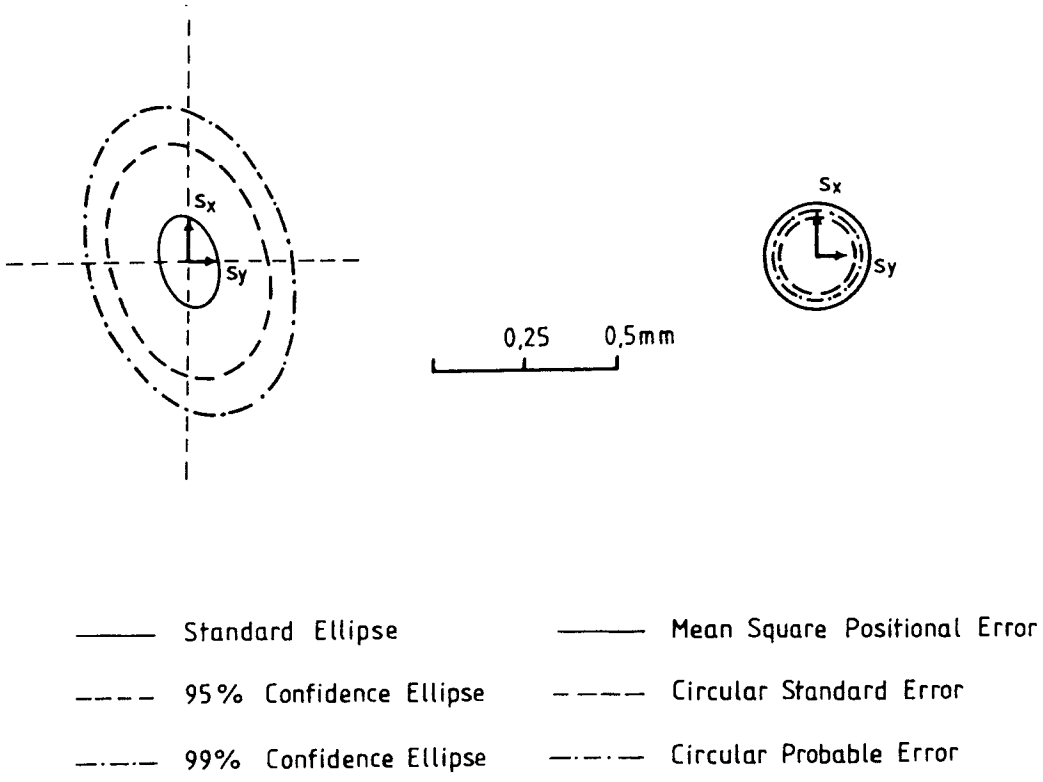


Figure 4.5: Different measures for the positional error of point P_7 from the adjustment according to Figure 4.3.

4.4 Global Measures of Accuracy

The most common global measure of accuracy is the average of all variances, namely the **mean variance**:

$$\bar{s}_{\hat{x}}^2 = \frac{1}{u} \operatorname{tr} \Sigma_{\hat{x}} \tag{4-49}$$

sometimes replaced by the **mean point error**:

$$\bar{s}_p = \bar{s}_{\hat{x}} \sqrt{2}$$

Unfortunately these measures, like the local ones, depend on the geodetic datum. The minimum trace datum (Section 3.7) is preferred here, yielding an appreciably smaller number, characterizing the inner precision. Nevertheless, for the comparison of network designs the mean variance proves useful, if the adjustments are based on the same datum. The loss of information by averaging can however lead to wrong conclusions. This problem is demonstrated with the aid of an extreme example which also reveals the poor defintion of the term accuracy in the present context. For the determination of a single point two observation schemes A and B are considered. The scheme leading to the highest accuracy shall be carried out. For a decision based on objective criteria the mean variance according to Eq. (4-49) has been calculated for both schemes. The result is:

$$1/2 \operatorname{tr} \Sigma_A = 1/2 \operatorname{tr} \Sigma_B = \bar{s}_{\hat{x}}^2 = 25$$

The conclusion, that both schemes lead to the same accuracy may be premature. The eigenvalue decomposition of Σ_A and Σ_B could show, for instance, that for A the semi-axes of the confidence ellipse are equal: a = b = 5, and for B very different: a = 7, b = 1. Figure 4.6 clearly shows that the two schemes are certainly not equivalent irrespective of the same mean variance. But which scheme is the better one? Another global measure of accuracy, favoured by many statisticians, is the **generalized variance** which is based on the determinant of $\Sigma_{\hat{x}}$:

$$\bar{\bar{s}}_{\hat{x}}^2 = \sqrt[u]{\det \Sigma_{\hat{x}}} \tag{4-50}$$

It is defined as the u-th root of its value.

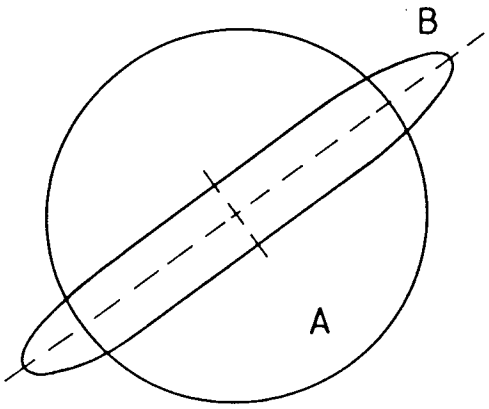


Figure 4.6: Ellipse A (a = b = 5) and Ellipse B (a = 7, b = 1) associated with the same mean variance $\bar{s}_{\hat{x}}^2 = 25$.

Apart from being datum dependent, this measure is mainly ruled by the smallest eigenvalue of $\Sigma_{\hat{x}}$ (see Eq. 4-15). This property contradicts the general geodetic philosophy of worst case thinking. For a singular $\Sigma_{\hat{x}}$ the determinant disappears and Eq. (4-50) becomes zero. In this case the determinant is usually replaced by the product of the non-zero eigenvalues. The generalized variance has a simple geometric meaning, since the determinant is proportional to the volume of the ellipsoid or area of the ellipse.

The application of the generalized variance to the example introduced above (Fig. 4.6) leads to the clear conclusion that scheme B is better than A, since $\overline{\overline{s}}_{\hat{x}}^2(A) = 25$ while $\overline{\overline{s}}_{\hat{x}}^2(B) = 7$.

A measure of accuracy often recommended is the **largest eigenvalue of** $\Sigma_{\hat{x}}$. The motivation for this measure stems from the following relationships. If:

$$g = f^t x = f_1 x_1 + f_2 x_2 + \ldots + f_u x_u \qquad (4\text{-}51)$$

is a linear invariant function of the GMM, with the variance computed according to the law of propagation of variances:

$$s_{\hat{g}}^2 = f^t \Sigma_{\hat{x}} f \qquad (4\text{-}52)$$

then the maximum value, which $s_{\hat{g}}^2$ can reach for any vector f of coefficients, is restricted by Eq. (4-16):

$$s_{\hat{g}}^2 \leq f^t f \lambda_{max} \qquad (4\text{-}53)$$

where λ_{max} is the maximum eigenvalue of $\Sigma_{\hat{x}}$.

Since the worst case, namely the upper limit of the variance of any function, is usually of major interest there are good reasons to introduce λ_{max} as a measure of global accuracy, which is datum dependent like all other measures.

If this measure is applied to the example introduced above (Fig. 4.6), it follows from $\lambda_{max}(B) = 49$ and $\lambda_{max}(A) = 25$, that scheme A is better than B.

If the purpose of the network requires a homogeneous accuracy, then the difference $\lambda_{max} - \lambda_{min}$ can be used to judge the achievement of this goal. A small difference indicates that the error hyperellipsoid is close to a hypersphere.

Datum	tr $Q_{\hat{x}}$	det $Q_{\hat{x}}$	λ_{max}	λ_{min}
Figure 4.1	0.802	1.43 E-14	0.641	0.0033
Figure 4.2	0.102	2.66 E-17	0.038	0.0021
Figure 4.3	0.061	8.14 E-19	0.020	0.0011

Table 4.2: Global measures of accuracy of the network defined in Figure 4.1. The geodetic datum of the three adjustments are shown in Figures 4.1, 4.2 and 4.3, respectively.

4.5 Comparison of Variance-Covariance Matrices

The scalar measures of global accuracy are necessarily rather course. Especially in the context of optimization of geodetic networks, it is sometimes preferred to construct a **criterion matrix** which represents the desired structure of the variance-covariance matrix. The comparison of the actual matrix $\Sigma_{\hat{x}}$ with the artificial criterion matrix, for example on the basis of error ellipses, leads to an assessment of the quality of the network.

For a general purpose network the criterion matrix is constructed so that all absolute ellipses are circles of the same radius and that the relative ellipses, being circles as well, increase according to a reasonably selected function of the distance between the two points involved. This homogeneous and isotropic structure (Taylor-Karman Structure) is achieved either empirically or by use of the theory of statistical vector fields. The datum problem appears here as well. It is essential that the matrices being compared have the same geodetic datum. This can be achieved easily by application of an S-transformation as outlined in Section 3.8 (GRAFAREND, 1984). The comparison of variance-covariance matrices on the basis of local measures of accuracy (such as error ellipses) is not satisfactory because the correlations are not taken into account. A more general approach is based on the generalized eigenvalue decomposition of Section 4.2.

If $\hat{x}_A$, Q_A and $\hat{x}_B$, Q_B are two sets of estimates of the parameter vector x with the associated cofactor matrices belonging to different sets of observations or to two different network designs or where one of the cofactor matrices is a criterion matrix, then a rule is desirable which allows the analyst to decide which one of the possible relations between A and B is true:

$$\text{precision (A)} \quad \text{smaller or equal or larger than precision (B)} \tag{4-55}$$

The decision is simple if only one function $y = f^t x$ is of interest. A comparison of the pertinent cofactors f^tQ_Af and f^tQ_Bf solves the problem. In the general case, where the overall precision needs to be compared, all possible functions $y_i = f_i^t x$ must be considered. This leads to the definition that A is better than or as good as B if:

$$f^tQ_Af \le f^tQ_Bf \qquad \forall f \in \mathbb{R}^u \tag{4-56}$$

holds. Rearrangement of Eq. (4-56):

$$f^t(Q_A - Q_B)f \le 0 \qquad \forall f \in \mathbb{R}^u \tag{4-57}$$

shows that Eq. (4-56) requires the difference of the cofactor matrices $Q_A - Q_B$ to be negative semi-definite, from which follows that all eigenvalues of the difference matrix are to be negative or zero. A necessary but not a sufficient condition of this property is that all diagonal elements of $Q_A - Q_B$ are negative. It is important to note that the relationship $\text{tr } A \le \text{tr } B$ does not mean that A is better than B in the sense of Eq. (4-56). In general it is necessary to compute the largest eigenvalue λ_{max} of $Q_A - Q_B$. Equation (4-56) is correct if λ_{max} is less than or equal to zero.

Another possibility is to consider the quotient, which is always positive:

$$k_f = \frac{f^tQ_Af}{f^tQ_Bf} \tag{4-58}$$

under the condition that $f^t Q_B f > 0$. If:

$$k_f \leq 1 \qquad \forall f \mid f^t Q_B f > 0 \tag{4-59}$$

then the relationship of Eq. (4-56) is obviously valid and $\hat{x}_A$ is more precise than $\hat{x}_B$.

Equation (4-58) is equivalent to:

$$f^t Q_A f - k_f f^t Q_B f = f^t (Q_A - k_f Q_B) f = 0 \tag{4-60}$$

This shows in comparison with Eqs (4-8) and (4-17) that Eq. (4-60) is only fulfilled if:

$$\det (Q_A - k_f Q_B) = 0$$

and, hence, if k_f is a generalized eigenvalue of Q_A with respect to Q_B (see Section 4.2). This property of k leads to the following generalization of Eqs (4-15) and (4-16):

$$\frac{s_i^t Q_A s_i}{s_i^t Q_B s_i} = k_i \tag{4-61}$$

$$k_{min} \leq \frac{f^t Q_A f}{f^t Q_B f} \leq k_{max} \qquad \forall f \in \mathbb{R}^u \mid f^t Q_B f > 0 \tag{4-62}$$

where s_i is the generalized eigenvector of Q_A with respect to Q_B associated with the generalized eigenvalue k_i. Equation (4-62) immediately shows that $\hat{x}_A$ with cofactor matrix Q_A is more accurate than $\hat{x}_B$ with Q_B if the maximum eigenvalue of Q_A with respect to Q_B is less than one.

For more details on the construction of criterion matrices and on the comparison of variance-covariance matrices the reader is referred to the literature, e.g. GRAFAREND (1984).

5. GENERAL LINEAR HYPOTHESIS

The GMM as introduced in Chapter 2:

$$l + v = A\hat{x}, \qquad \Sigma = \sigma_o^2 Q, \qquad P = Q^{-1} \tag{2-2}$$

is a set of functional and stochastic relationships based on observations, experience and hypotheses, which represent the physical reality. For ease of computation and interpretation the most simple model is usually aimed at. Quite often only a simple model is possible because of lack of knowledge. Therefore it is always essential to consider and question the model assumptions and to check how the model conforms with reality.

The global model test (Section 2.3) has been discussed previously as an example. It is based on the comparison of s_o^2 and σ_o^2 (see Eq. (2-12)) from which a χ^2-distributed test statistic has been derived. In this chapter a verification or improvement of the parameter vector of Eq. (2-2) is aimed at. This objective is very important, especially in deformation analysis, since errors in the model tend to be interpreted as deformations and may lead to wrong and even dangerous conclusions. Statistical testing is used as a guide for decision making.

The main questions to be answered are: is the parameter vector complete? Does it contain insignificant components? Do the parameters meet certain conditions known in advance? Some obvious examples are:

i. The vector of observations contains distances and zenith angles. Is it reasonable to consider a coefficient of refraction and an additive constant?

ii. During a measurement epoch observations were taken on a site at two different dates. At the second date the monument looked damaged. Has the control point changed its position?

iii. Two epochs of deformation measurements have been carried out. The adjustments result in slightly different sets of coordinates. Are the differences caused by significant deformations?

iv. A number of points in an engineering network are to be situated on a straight line, a construction axis for example. Do the estimated point positions meet the required alignment precision?

As the testing of hypotheses is a statistical concept, knowledge about the statistical properties of the estimates of the GMM is required. It is usually assumed that the observations are normally distributed, which can be expressed as:

$$l \sim N(Ax, \Sigma) \tag{2-8}$$

since the estimators are linear with l as derived in Sections 3.6 and 3.7, Eqs (3-32), (3-33) and (3-51) the following distributions are assumed:

$$\begin{aligned} &\hat{x} \sim N(x, \Sigma_{\hat{x}}) \\ &\Sigma_{\hat{x}} = s_o^2 (N + RR^t)^{-1} N (N + RR^t)^{-1} \end{aligned} \tag{5-1}$$

$$v = N(0, \Sigma_v), \qquad \Sigma_v = \Sigma - A\Sigma_{\hat{x}}A^t \tag{5-2}$$

$$l + v = N(Ax, \Sigma_{l+v}), \qquad \Sigma_{l+v} = A\Sigma_{\hat{x}}A^t \tag{5-3}$$

Prior to discussing the mathematical details a short summary of the fundamentals of statistical testing and some necessary properties of the distributions of quadratic forms are given.

5.1 Null Hypothesis and Alternatives

All hypotheses considered in this chapter can be put in the same form, so that a single testing procedure can be applied to all cases. The general linear hypothesis as applied to the parameter vector has the form:

$$H_o : H^t x - g = 0 \qquad (5\text{-}4)$$

where H^t is a m x u-matrix with $r(H) \leq m$, $1 \leq m \leq u$. Equation (5-4) represents m linear functions:

$$h_i^t x - g_i = 0, \qquad i \in \{1, 2, ..., m\} \qquad (5\text{-}5)$$

which have to be invariant in the sense of Section 3.6. Hence the rows h_i^t of H^t have to be linear combinations of the rows of the design matrix A of the GMM, which, in turn, requires the existence of some matrix L such that $LA = H^t$. The vector g of Eq. (5-4) contains constants given by the hypothesis. The formulation of a specific null hypothesis is the first step in the testing procedure, followed by a transformation to the general form of Eq. (5-4). For example:

$$H_o : x_i - x_k = 0$$

From:

$$x = (x_1, x_2, ..., x_i, x_j, x_k, ..., x_u)^t$$

follows:

$$h^t = (0, 0, ..., +1, 0, -1, 0, ..., 0) \qquad \text{and} \qquad g = 0$$

yielding in accordance with Eq. (5-4):

$$H_o : h^t x = 0$$

Consideration of a specific null hypothesis implies naturally an idea of what is true if H_o fails. This is formulated as the alternative hypothesis, which is not always written down explicity but is kept in mind by the person applying the test. Typical forms of alternative hypotheses are:

$H_a : h^t x \neq g$ — two-tailed alternative

$h^t x > g$, $h^t x < g$ — one-tailed alternatives

$h^t x = \bar{g} \neq g$ — definite alternative

5.2 Test Statistics

The second step of the test procedure consists of the calculation of a suitable test statistic from the GMM using the actual observations. The statistic is selected in such a way that its distributional properties are known by definition if H_o is true and that it is as sensitive as possible to departures from H_o. There is a large number of test statistics. When testing linear hypotheses in the GMM, only four different types need to be considered, which are the most efficient ones under the present assumptions.

If: $y \sim N(\eta, \Sigma_y)$

$z \sim N(\xi, \Sigma_z)$

are two normally distributed random vectors with expectation η and ξ respectively, then:

$$T_1 = (y_i - \eta_i)/\sigma_{y_i} \sim N(0, 1) \tag{5-6}$$

is normally distributed with expectation zero and variance 1,

$$T_2 = (y_i - \eta)/s_{y_i} \sim t(f) \tag{5-7}$$

exhibits Student's t-distribution with f being the degrees of freedom of the underlying GMM, provided that y_i and s_{y_i} are statistically independent:

$$T_3 = (y - \eta)^t \Sigma_y^{-1} (y - \eta) \sim \chi^2(r) \tag{5-8}$$

is χ^2-distributed with $r = r(\Sigma_y)$ degrees of freedom, and:

$$T_4 = \frac{(y - \eta)^t \Sigma_y^{-1} (y - \eta)/r_y}{(z - \xi)^t \Sigma_z^{-1} (z - \xi)/r_z} \sim F(r_y, r_z) \tag{5-9}$$

has a Fisher-distribution with $r_y = r(\Sigma_y)$ and $r_z = r(\Sigma_z)$ degrees of freedom if numerator and denominator are statistically independent.

5.3 Risk Level and Critical Value

In a third step the very delicate selection of the risk level α follows. While the whole testing procedure follows strict mathematical rules and, therefore, looks like a totally objective and rigorous method, the selection of α is completely arbitrary. There is neither a right nor a wrong risk level, and no satisfactory criteria or rules exist which can be applied. A decision on the risk level must be made and must be made at this stage of the procedure; otherwise, the results of the subsequent computations could influence the decision. Typical risk levels are α = 10%, α = 5%, α = 1% and α = 0.1%, but any other number in this range can be justified equally. The fourth step of the test procedure is the determination of the critical value of the test statistic for the chosen risk level α. This is usually done by extracting the relevant value from tables or graphs of the distribution of the test statistic. As shown in Fig. 5.1 it is important to distinguish between one-tailed and two-tailed tests. The critical value is compared with the computed test statistic, followed by either a rejection of the null hypothesis at risk α or, in the opposite case, to no rejection.

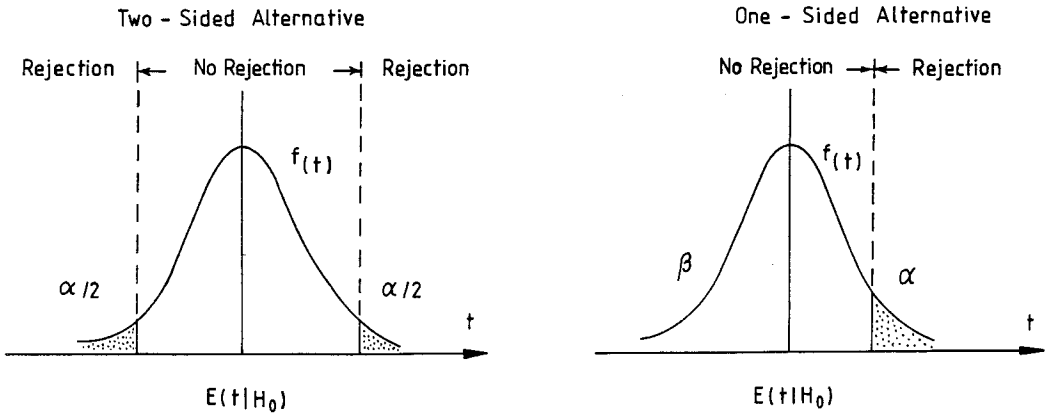


Figure 5.1: Rejection and non-rejection regions of one- and two-tailed tests with statistics according to Eqs (5-6) and (5-7).

It should be stressed that no test can prove a hypothesis. (The definitions of tests are not always specific on this point.) The result of a test is either rejection or no rejection. In the latter case it is not possible to prove a specific objection to the hypothesis. This does not testify that there exist no other objections which have not yet been considered. If the null hypthesis is true, the determined value of T will cause a rejection of the true hypothesis in α of all cases (see Fig. 5.1). This erroneous outcome of the test is often called type I error of the test, which occurs with probability α. On the other hand it can happen that an alternative hypothesis is true although the calculated value of the test statistic T falls in the region of no rejection (see again Fig. 5.1). This erroneous result is called type II error of the test and occurs with probability β. This probability depends on the density function f(t), the risk level α and on the alternative hypothesis H_a. Figure (5.2) shows the relationship between α and β.

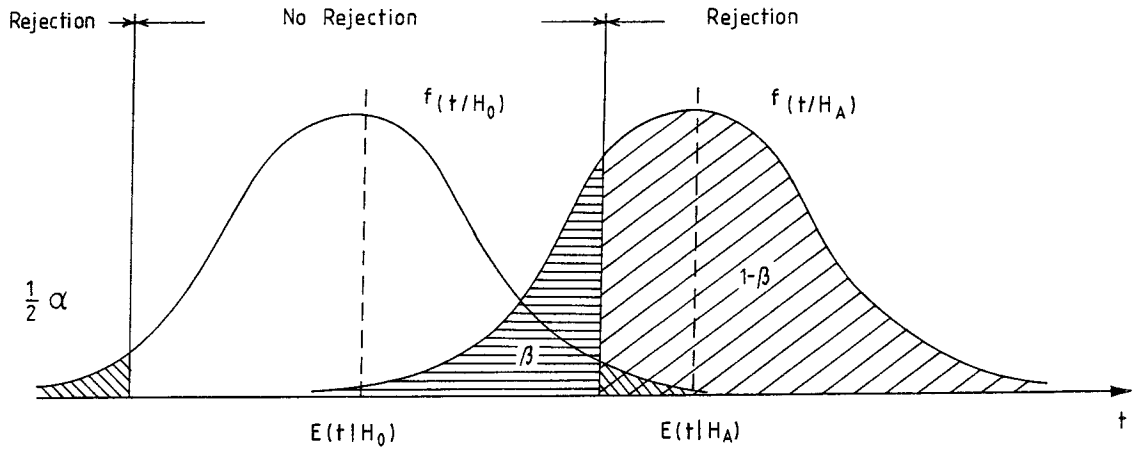


Figure 5.2: Two-tailed test with risk level α and probability β of type II error for a definite H_a.

5.4 Some Properties of Distributions of Quadratic Forms

The distributions of the test statistics given in Eqs (5-7), (5-8) and (5-9) depend on the statistical properties of the quadratic forms involved. Although a rigorous derivation of these distributions is outside the scope of this text, it is sometimes useful to have a rigorous definition of the underlying theorems at hand. Therefore the basic theorems are given. Their proofs can be found in textbooks on linear models such as RAO (1973) or SEARLE (1971).

Theorem 1: Let $y \sim N(\eta, \Sigma)$, then a necessary and sufficient condition that the quadratic form q:

$$q = (y-\eta)^t A(y-\eta) \sim \chi^2(r)$$

is: $$\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$$

where $r = r(A\Sigma)$. If $r(\Sigma) = n$ then the condition reduces to:

$$A\Sigma A = A$$

Theorem 2: Let $y \sim N(\eta, \Sigma)$, then a necessary and sufficient condition that:

$$(y-\eta)^t A_1 (y-\eta) \quad \text{and} \quad A_2 y$$

are statistically independent is:

$$\Sigma A_2 \Sigma A_1 = 0$$

Theorem 3: Let $y \sim N(\eta, \Sigma)$, then a necessary and sufficient condition that the quadratics:

$$q_1 = (y-\eta)^t A_1 (y-\eta) \quad \text{and} \quad q_2 = (y-\eta)^t A_2 (y-\eta)$$

are statisitically independent is:

$$\Sigma A_2 \Sigma A_1 \Sigma = 0$$

The simplication for $r(\Sigma) = n$ yields for the two last theorems:

$$A_2 \Sigma A_1 = 0$$

Theorem 4: Let $y \sim N(\eta, \Sigma)$ and $q_1 = (y-\eta)^t A_1 (y-\eta)$ and $q_2 = (y-\eta)^t A_2 (y-\eta)$. If:

$$q = (y-\eta)^t A(y-\eta) = q_1 + q_2$$

and: $$\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$$

and: $$\Sigma A_1 \Sigma A_1 \Sigma = \Sigma A_1 \Sigma$$

then: $$q \sim \chi^2(r), \quad q_1 \sim \chi^2(r_1), \quad q_2 \sim \chi^2(r_2)$$

where: $$r = r(A\Sigma), \quad r_1 = r(A_1\Sigma) \quad \text{and} \quad r_2 = r(A_2\Sigma) = r - r_1$$

Note: Theorem 1 gives the basic condition for the χ^2-distribution of a quadratic form and, therefore, refers to the statistic T_3 of Eq. (5-8). Theorem 2 provides the tool for testing the required independence of numerator and denominator of T_2 in Eq. (5-7). Theorem 3 serves the same purpose for T_4 of Eq. (5-9). Theorem 4 applies to the decomposition of quadratic forms which plays an important role in testing linear hypotheses. As an example for the application of the theorems the distribution of the quadratic form $q = v^tPv$ (see Eq. (2-7)) shall be investigated. According to Eq. (5-2) v has the Gaussian distribution $v \sim N(0, \Sigma_v)$, where $\Sigma_v = \Sigma - A\Sigma_{\hat{x}}A^t$, $r(\Sigma_v) = n - r(A)$. Since $P = Q^{-1}$ and $\Sigma = \sigma_o^2 Q$ the quadratic form can be expanded to:

$$q/\sigma_o^2 = v^t\Sigma^{-1}v \qquad (5\text{-}10)$$

It follows now from Theorem 1 that $q/\sigma_o^2 \sim \chi^2$ if:

$$\Sigma_v\Sigma^{-1}\Sigma_v\Sigma^{-1}\Sigma_v = \Sigma_v\Sigma^{-1}\Sigma_v$$

which is easily verified by multiplication, hence proving:

$$\frac{v^tPv}{\sigma_o^2} \sim \chi^2(f) \qquad (5\text{-}11)$$

with: $f = r(\Sigma^{-1}\Sigma_v) = n - r(A)$.

5.5 Linear Hypotheses in the GMM

The general linear hypothesis as introduced in Eq. (5-4) must be considered as a component of an extended GMM:

$$l = Ax + \varepsilon, \qquad \Sigma = \sigma_o^2 Q$$

$$g = H^tx$$

If H_o is true, then the relationship $H^tx - g = 0$ applies. Since the least squares method only provides the estimate $\hat{x}$ of the parameter vector x, there will be discrepancies between $H^t\hat{x} = \hat{g}$ and the constants g. If these discrepancies:

$$w = H^t\hat{x} - g \qquad (5\text{-}12)$$

have an expectation of zero ($E(w) = 0$) then there is no reason to reject H_o. If in the opposite case H_o is false, then significant deviations ($E(w) \neq 0$) from zero are expected.

When considering a single hypothesis $H_o: h^tx - g = 0$ the corresponding test statistic is computed from the estimate $\hat{g} = h^t\hat{x}$ which is distributed normally:

$$\hat{g} \sim N(g, \sigma_{\hat{g}}^2), \qquad \sigma_{\hat{g}}^2 = h^t\Sigma_{\hat{x}}h \qquad (5\text{-}13)$$

Two cases need to be distinguished:

i. The a priori variance factor σ_o^2 is the true variance of unit weight. The test statistic follows from Eq. (5-6):

$$T = (\hat{g} - g)/\sigma_{\hat{g}} \sim N(0, 1) \qquad (5\text{-}14)$$

and the critical values are taken from the standardized normal distribution at a risk level α.

ii. The estimate s_o^2 of the variance factor is considered as a more realistic value and is used for the test. Using Theorem 2 the independence of v^tPv and $h^t\hat{x}$ can be proved. Hence, the conditions of Eq. (5-7) are met leading to:

$$T = (\hat{g} - g)/s_{\hat{g}}, \qquad s_{\hat{g}}^2 = s_o^2 h^t Q_{\hat{x}} h$$

$$T \sim t(n - r(A)) \qquad (5\text{-}15)$$

The critical value is taken from Student's t-distribution with $n - r(A)$ degrees of freedom using a risk level α.

If a composite hypothesis $H_o: H^tx - g = 0$ is considered then a test statistic of type T_4 (Eq. (5-9)) must be computed. It would be wrong to split H_o into single hypotheses and to repeatedly apply Eq. (5-14) or Eq. (5-15).

The estimate $\hat{g} = H^t\hat{x}$ of the given vector g has the variance-covariance matrix $\Sigma_{\hat{g}} = H^t\Sigma_{\hat{x}}H$. If H_o is true, then the estimate is unbiased and distributed according to $\hat{g} \sim N(g, \Sigma_{\hat{g}})$. The distribution of the quadratic form follows from Theorem 1 in Section 5.4:

$$q_\Delta = (\hat{g} - g)^t Q_{\hat{g}}^{-1} (\hat{g} - g) \sim \sigma_o^2 \chi^2(m) \qquad (5\text{-}16)$$

with $m = r(H)$. Substitution of $\hat{g} - g = H^t(\hat{x} - x)$ yields:

$$q_\Delta = (\hat{x} - x)^t H (H^t Q_{\hat{x}} H)^{-1} H^t (\hat{x} - x) \qquad (5\text{-}17)$$

It can be readily verified with $A = H(H^tQ_{\hat{x}}H)^{-1}H^t/\sigma_o^2$ that the condition $\Sigma_{\hat{x}}A\Sigma_{\hat{x}}A\Sigma_{\hat{x}} = \Sigma_{\hat{x}}A\Sigma_{\hat{x}}$ holds. Furthermore, using Theorem 3, it can be shown that the quotients q_Δ/σ_o^2 of Eq. (5-16) and q/σ_o^2 of Eq. (5-10) are distributed independently and thus meet the conditions of the test statistic of Eq. (5-9):

$$T = \frac{q_\Delta/m}{q/f} \sim F(m, f) \qquad (5\text{-}18)$$

The numerator and denominator are defined in Eqs (5-16) and (5-11), respectively.

If H_o is true, then the test statistic of Eq. (5-18) follows the specified Fisher-distribution. Hence, the decision is based on the comparison of T with the critical value $F_\alpha(m, f)$ for a risk level α as taken from a table or graph of the F-distribution at m and f degrees of freedom.

In this version of the composite test it is not necessary to know the true variance factor σ_o^2. Should it be known then a more efficient test based on the statistic T_3 of Eq. (5-8) can be employed:

$$T = q_\Delta/\sigma_o^2 \sim \chi^2(m) \tag{5-19}$$

with q_Δ and m as defined above. Here, the critical value of risk level α and m degrees of freedom of the χ^2-distribution applies.

These tests of the composite null hypothesis H_o : $H^t x - g = 0$ are based on the following probability relationships:

$$\begin{aligned} P(T > F_\alpha(m, f) \mid H_o) &= \alpha \\ P(T > \chi^2(m) \mid H_o) &= \alpha \end{aligned} \tag{5-20}$$

The above probabilities refer to the one-tailed test according to Fig. 5.1, i.e. the probability P, that T is greater than the corresponding critical value in the case that H_o is true, equals the type I error probability α.

5.6 Computation of Quadratic Forms

The composite null hypthesis $H^t x - g = 0$ can be considered as a component of an extended GMM as already stated in Eq. (5-12). It has the form of a set of conditions which must be met by the parameters. Later, the influence of these conditions on the solution shall be investigated and the algorithms derived which are required for the computation of the quadratic forms and the test statistics according to Eqs (5-18) and (5-19).

Note: The meaning of the conditions $H^t x = g$ is conceptionally different from that of the constraints $R^t x = c$ of Chapter 3, which only serve to fix the geodetic datum. The difference in terms of linear algebra is that the rows of H^t are linear combinations of the rows of A, while the rows of R^t are a complement of the rows of A in $\mathbb{R}^u$. The conditions $H^t x = g$ represent error-free information about the parameter vector. The constraints $R^t x = c$ define a reference frame in which the parameter vector is located.

The normal equations of the extended GMM of Eq. (5-12) are:

$$\begin{aligned} N\hat{x}_H + Hk &= A^t P l\,, \qquad N = A^t P A \\ H^t \hat{x}_H &= g \end{aligned} \tag{5-21}$$

where $\hat{x}_H$ is the parameter vector under the conditions $H^t x = g$. Using generalized inverses, the first equation of Eqs (5-21) has the solution:

$$\hat{x}_H = N^- (A^t P l - Hk) \tag{5-22}$$

Since the parameter estimate of the ordinary GMM without conditions is $\hat{x} = N^- A^t P l$, Eq. (5-22) is equivalent to:

$$\hat{x}_H = \hat{x} - N^- Hk \tag{5-23}$$

Substitution in the second equation of Eqs (5-21) yields:

$$H^t\hat{x} - H^tN^{-}Hk = g \tag{5-24}$$

Solution for k, leads to:

$$-k = (H^tN^{-}H)^{-1}(g - H^t\hat{x}) \tag{5-25}$$

Hence, Eq. (5-23) becomes:

$$\hat{x}_H = \hat{x} + N^{-}H(H^tN^{-}H)^{-1}(g - H^t\hat{x}) \tag{5-26}$$

and clearly shows the influence of the conditions on the parameter estimate.

The residuals of the GMM with conditions are:

$$v_H = A\hat{x}_H - l \tag{5-27}$$

leading to the quadratic form:

$$q_H = (A\hat{x}_H - l)^tP(A\hat{x}_H - l) \tag{5-28}$$

The expansion of v_H into:

$$v_H = A\hat{x}_H - l + A\hat{x} - A\hat{x} = A\hat{x} - l + A(\hat{x}_H - \hat{x})$$

$$v_H = v + A(\hat{x}_H - \hat{x}) \tag{5-29}$$

demonstrates the contribution of the conditions to the residuals. Substitution of Eq. (5-29) in Eq. (5-28) yields:

$$\begin{aligned} q_H &= v^tPv + (\hat{x}_H - \hat{x})^tN(\hat{x}_H - \hat{x}) \\ q_H &= q + q_\Delta \end{aligned} \tag{5-30}$$

The second term on the right hand side is caused by the conditions. Using Eq. (5-26) it can be shown that it is identical with Eq. (5-17). Rearrangement yields the computationally most convenient equation for q_Δ:

$$q_\Delta = (g - H^t\hat{x})^t(H^tN^{-}H)^{-1}(g - H^t\hat{x}) \tag{5-31}$$

A solution for the extended GMM is not required for the testing of H_o if the solution $\hat{x}$ of the ordinary model is known already. Equation (5-31) together with Eq. (5-18) or Eq. (5-19) establishes a very efficient method of testing a composite null hypothesis.

5.7 Example

Consider the levelling network of Section 3.9. Using the solution of version (a) based on a conventional datum two different tests shall be demonstrated.

a. The null hypothesis states that points 1 and 4 have the same height:

$$H_o : x_1 = x_4 \Rightarrow x_1 - x_4 = 0$$

$$h^t x = 0 \Rightarrow h^t = (1, 0, 0, -1); \quad g = 0$$

$$H_a : h^t x \neq 0 \Rightarrow \text{two-tailed test}$$

Eq. (5-13): $\hat{g} = h^t \hat{x} = -2.4\text{mm}$

Since σ_o^2 is not known, the estimate $s_o^2 = 0.4\,\text{mm}^2$ has to be used. The test statistic follows from Eq. (5-15) with $s_{\hat{g}}^2 = s_o^2 h^t Q_{\hat{x}} h = 0.4(2/4) = 0.2$; $s_{\hat{g}} = \sqrt{0.2}$:

$$T = (\hat{g} - g)/s_{\hat{g}} = -2.4/0.45 = -5.37$$

If H_o is true, then T has a t-distribution with 3 degrees of freedom. Some critical values are (two-tailed):

$$\alpha = 10\%, \qquad t_\alpha(3) = \pm 2.35$$

$$\alpha = 5\%, \qquad t_\alpha(3) = \pm 3.18$$

$$\alpha = 1\%, \qquad t_\alpha(3) = \pm 5.84$$

The test statistic clearly exceeds the 5%-point but is just below the 1%-point. This result stresses the importance of deciding on α before the computations are carried out.

b. The null hypothesis assumes that all points have the same height:

$$H_o : H^t x = 0 \Rightarrow g = 0$$

$$H^t = \begin{pmatrix} -1 & +1 & 0 & 0 \\ -1 & 0 & +1 & 0 \\ -1 & 0 & 0 & +1 \end{pmatrix}$$

$$H_a : H^t x \neq 0$$

Eq. (5-31): $H^t \hat{x} = (+0.6, +1.5, +2.4)^t\,\text{mm}$

$$H^tN^{-}H = {}^1/_4 \begin{pmatrix} +2 & +1 & +1 \\ +1 & +2 & +1 \\ +1 & +1 & +2 \end{pmatrix}, \quad (H^tN^{-}H)^{-1} = \begin{pmatrix} +3 & -1 & -1 \\ -1 & +3 & -1 \\ -1 & -1 & +3 \end{pmatrix}$$

$$q_\Delta = 13.23$$

Eq. (5-18): $$T = \frac{13.23}{1.2} = 11.025$$

If H_o is true, then T has a F-distribution with m = 3 and f = 3 degrees of freedom. Some critical values are:

$$\alpha = 10\% \qquad F_\alpha(3, 3) = 5.39$$

$$\alpha = 5\%, \qquad F_\alpha(3, 3) = 9.28$$

$$\alpha = 1\%, \qquad F_\alpha(3, 3) = 29.46$$

Again, the test statistic falls between the critical points for 5% and 1%.

Note: The tests do not depend on the geodetic datum, since H^tx has the form of an invariant function.

The matrix H^t for test (b) is not uniquely defined by H_o. An alternative selection would be:

$$H^t = \begin{pmatrix} -1 & +1 & 0 & 0 \\ 0 & -1 & +1 & 0 \\ 0 & 0 & -1 & +1 \end{pmatrix}$$

The outcome of the test is independent of the particular definition of H^t provided that H^t represents the null hypothesis.

The following null hypothesis is a typical application in deformation analysis.

Consider two epochs of observations of a monitoring network, and the respective free network adjustments of Epoch i (i = 1, 2):

$$l_i + v_i = A_i x_i, \qquad \Sigma_i = \sigma_o^2 Q_i, \qquad P_i = Q_i^{-1}$$

$$\hat{x}_i = Q_{\hat{x}_i} A_i^t P_i l_i, \qquad \Sigma_{\hat{x}_i} = s_{oi}^2 Q_{\hat{x}_i}$$

$$s_{oi}^2 = v_i^t P_i v_i / f_i, \qquad f_i = n_i - r(A_i)$$

$$H_o : x_1 = x_2 \Rightarrow H^tx = g$$

with:

$$H^t = (I : -I), \qquad x = (x_1, x_2)^t, \qquad g = 0$$

To compute the test statistic after Eq. (5-18) a common variance estimate $s_o^2 = q/f$ and a common cofactor matrix $Q_{\hat{x}}$ is required. Provided that the two single epoch adjustments have been carried out using the same a priori variance factor σ_o^2 and have been based on the same geodetic datum, the variance can be pooled:

$$s_o^2 = \frac{v_1^t P_1 v_1 + v_2^t P_2 v_2}{f_1 + f_2} = \frac{v^t P v}{f}$$

and the cofactor matrices can be merged to give:

$$Q_{\hat{x}} = \begin{pmatrix} Q_{\hat{x}_1} & 0 \\ 0 & Q_{\hat{x}_2} \end{pmatrix}$$

The values of s_o^2 and $Q_{\hat{x}}$ are identical with the corresponding quantities of a combined adjustment of both epochs.

The quadratic form q_Δ can be computed from Eq. (5-31) as follows:

$$q_\Delta = (\hat{x}_1 - \hat{x}_2)^t [(I: -I)^t Q_{\hat{x}} (I: -I)]^- (\hat{x}_1 - \hat{x}_2)$$

$$q_\Delta = (\hat{x}_1 - \hat{x}_2)^t (Q_{\hat{x}_1} + Q_{\hat{x}_2})^- (\hat{x}_1 - \hat{x}_2)$$

where any generalized inverse of the form matrix may be used. Under the condition that H_o is true (no deformations between the epochs) the test statistic of Eq. (5-18):

$$T = \frac{q_\Delta / m}{s_o^2} \sim F(m, f)$$

has a Fisher-distribution with $m = r(Q_{\hat{x}_1} + Q_{\hat{x}_2})$ and $f = f_1 + f_2$.

6. OUTLIER DETECTION

The classical theory of errors distinguishes random errors, systematic errors and gross errors.

The **random errors** are the unavoidable, usually small differences between the observations and their expectations. They are unpredictable and follow statistical rules. Thus, only the average behaviour can be expressed by equations and not the actual occurrence. For geodetic estimation and testing procedures it is generally assumed that the random errors obey the normal distribution.

The **systematic error** or bias is defined as the difference between the functional model and reality. Typical examples are insufficient reductions of the observations due to refraction, instrument constant or map projection and omission of nuisance parameters. Theoretically it is possible to eliminate the bias by refining the mathematical model. This can be done in two different ways:

i. The bias is considered as a strictly functional part of the model, which is expressed by fixed parameters or removed by reductions.

ii. The bias is considered as a stochastic effect which is accounted for by random parameters or by a priori correlations between the observations.

The **gross error** is the result of a malfunctioning of either the instrument or the surveyor. Typical examples are the incorrect reading or incorrect recording of results and the failure of the instrument due to weak power supply or extreme environmental conditions. At least theoretically, gross errors can be avoided by due care or they can be detected by carefully designed observation schemes.

This error philosophy is the key to the comprehension of adjustment models, but it is hardly helpful in developing a strategy for the detection of gross errors in the GMM. The reason for this shortcoming of the theory is the inability of separating the errors according to their classification in a real world application. The estimation process provides residuals, which are a mixture of all error types. Certain assumptions on the stochastic properties of the residuals are required in order to deal with this problem. An **outlier** is defined as a residual which, according to some test rule, is in contradiction to the assumption. This operational definition of an outlier allows for a test strategy and a clear statistical concept, but it is only a relative definition, depending on the selected risk level, the assumed distribution and the test procedure. Despite this fundamental difference between the definitions of outliers and gross errors, it is naturally expected that the detected outliers are caused by gross errors.

Whenever outlying residuals are detected, it is necessary to thoroughly check the records of the observations in order to find out if a gross error can be traced. If no gross error can be found, the corresponding observation is deleted and, if necessary, remeasured.

The approach to outlier detection is based on the well-known Gauss-Markov model. The first step is usually a global model test. If this global test fails, procedures to flag erroneous measurements follow.

6.1 Global Model Test

The global test of the model, as introduced in Section 2.3, is reviewed here in a concise form for easy reference and completeness of the detection procedure.

GMM: $l + v = A\hat{x}, \quad \Sigma = \sigma_o^2 Q, \quad P = Q^{-1}$

$o(A) = n \times u, \quad r(A) = r \leq u, \quad N = A^t P A$ (2-2)

The LS-estimates depend according to Chapter 3 on the geodetic datum. They are given by:

$$\hat{x} = N^{-} A^t P l, \qquad Q_{\hat{x}} = N^{-} N (N^{-})^t \tag{6-1}$$

$$v = (A N^{-} A^t P - I) l, \qquad Q_v = Q - A Q_{\hat{x}} A^t \tag{6-2}$$

where N^{-} is any g-inverse defined by Eq. (3.74) and

$$v^tPv = l^tPQ_vPQ_vPl = l^tPQ_vPl$$
$$s_o^2 = v^tPv/f, \qquad f = n-r \tag{6-3}$$

Since all tests are based on statistical distributions, a reasonable stochastic concept must be introduced. Normally the observations are assumed to be normally distributed:

$$l \sim N(Ax, \Sigma)$$

The above equation assumes the absence of systematic and gross errors. Solely random errors are considered. From the law of propagation of variances and the linearity of the Eqs (6-1) and (6-2) follows:

$$\hat{x} \sim N(x, \sigma_o^2 Q_{\hat{x}}) \tag{6-4}$$

$$v \sim N(0, \sigma_o^2 Q_v) \tag{6-5}$$

$$v^tPv \sim \sigma_o^2 \chi^2(f) \tag{6-6}$$

The global model test questions the assumptions by comparing the a posteriori variance factor s_o^2 with σ_o^2:

H_o : The model is correct and complete; the distributional assumptions meet the reality.

The test statistic follows from Eq. (2-13):

$$T = v^tPv/\sigma_o^2 = fs_o^2/\sigma_o^2 \sim \chi^2(f) \tag{2-13}$$

6.2 Modelling Errors in the Observations

Should the global test indicate discrepancies between model and reality, and should these be caused by the functional part of the model, then they could always be taken into account by adding a vector Δ to the observations, leading to the corrected vector $\bar{l}$:

$$\bar{l} = l + \Delta \tag{6-7}$$

which brings the model in accordance with reality. The estimates in this corrected GMM are:

$$\hat{\bar{x}} = N^{-}A^tP(l+\Delta) = \hat{x} + N^{-}A^tP\Delta \tag{6-8}$$

$$\bar{v} = v - Q_vP\Delta \tag{6-9}$$

$$\bar{v}^tP\bar{v} = v^tPv + \Delta^tPQ_vPQ_vP\Delta - 2\Delta^tPQ_vPv \tag{6-10}$$

Realizing that:

$$PQ_v \cdot PQ_v = PQ_v, \qquad E(\bar{v}) = 0, \qquad A^tPv = 0$$

and assuming that $\bar{v}$ and Δ are independent leads after substitution of $v = \bar{v} + Q_v P\Delta$ into the most right term of Eq. (6-10) to:

$$\bar{v}^t P \bar{v} = v^t P v - \Delta^t P Q_v P \Delta - 2\Delta^t P Q_v P \bar{v}$$

$$E(\bar{v}^t P \bar{v}) = E(v^t P v) - E(\Delta^t P Q_v P \Delta)$$

$$\sigma_o^2 = E(s_o^2) - E(\Delta^t P Q_v P \Delta)/f \qquad (6\text{-}11)$$

The expectation of s_o^2, namely the variance factor computed from the incorrect model, is always greater than or equal to the true variance σ_o^2, with equality for:

$$\Delta = 0 \quad \text{or} \quad Q_v P \Delta = 0$$

that is:

$$E(s_o^2) \geq \sigma_o^2 \qquad \forall\, \Delta \in \mathbb{R}^n \qquad (6\text{-}12)$$

The case of $Q_v P \Delta = 0$ occurs if Δ belongs to the null space of $Q_v P = (I - AQ_{\hat{x}}A^t P)$, i.e. if $\Delta = Ag$ for some vector g.

Thus the global model test cannot detect all errors of the functional model. It can never prove the adequacy of the model! If Δ is assumed to be a deterministic vector then $v^t Pv/\sigma_o^2$ has a non-central χ^2-distribution with f = n - r degrees of freedom and a non-centrality parameter λ:

$$H_a : \quad v^t Pv/\sigma_o^2 \sim \chi(f, \lambda) \qquad (6\text{-}13)$$

$$\lambda = \Delta^t P Q_v P \Delta/\sigma_o^2 \qquad (6\text{-}14)$$

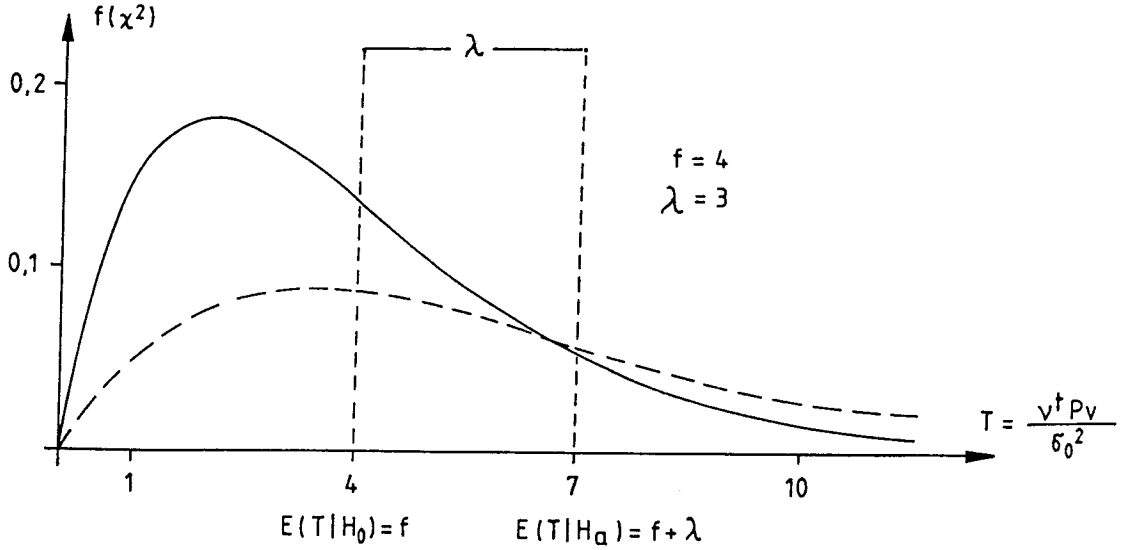


Figure 6.1: Density function $f(\chi^2)$ under the hypotheses H_o and H_a for f degrees of freedom.

Full line: central χ^2-distribution for f = 4 and λ = 0

Dashed line: non-central χ^2-distribution for f = 4 and λ = 3

6.3 Conventional Alternative Hypothesis

So far no specific alternative hypothesis to H_o has been considered. From a large number of possibilities, the so-called conventional alternative is selected. The conventional alternative hypothesis is the only one for which a satisfactory but simple statistical concept exists.

H_a : **One** observation has caused the rejection of H_o.

In this case the vector Δ takes the form:

$$\Delta = e_i\Delta_i \tag{6-15}$$

if the i-th observation is considered. The i-th unit vector (i.e. a vector with 1 in the i-th position and zeros in all others) and the constant gross error of observation l_i are denoted by e_i and $-\Delta_i$, respectively. Now Eq. (6-7) becomes:

$$\bar{l} = l + \Delta = l + e_i\Delta_i \tag{6-16}$$

where:

$$\bar{l} = (\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_i, \ldots, \bar{l}_n)^t \quad \text{correct observations}$$

$$l = (\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_i - \Delta_i, \ldots, \bar{l}_n)^t \quad \text{gross error } -\Delta_i \text{ in } l_i$$

$$e_i\Delta_i = (0, 0, \ldots, \Delta_i, \ldots, 0)^t \quad \text{correction}$$

If the general estimation equations of the corrected model, namely Eqs (6-8), (6-9) and (6-10), are specialized for this particular H_a, then the following equations are obtained:

$$\hat{\bar{x}} = \hat{x} + N^{-}A^tPe_i\Delta_i = \hat{x} + N^{-}a_ip_i\Delta_i \tag{6-17}$$

with a_i being the i-th column of A^t and p_i the weight of l_i based on the usual stochastic model:

$$P = \text{diag}\,(p_1, p_2, \ldots, p_i, \ldots, p_n)$$

The vector of residuals yields:

$$\bar{v} = v - Q_vPe_i\Delta_i = v - q_{v_i}p_i\Delta_i \tag{6-18}$$

where q_{v_i} is the i-th column of Q_v.

Equation (6-18) clearly shows the reason why the detection of gross errors is so difficult. The whole vector of residuals is contaminated by Δ_i and it is even possible that the effect of Δ_i on residuals other than v_i is greater than that on v_i.

The quadratic form changes to:

$$\bar{v}^tP\bar{v} = v^tPv - \Delta_i^2p_i^2q_{v_iv_i} - 2\Delta_ip_iq_{v_i}^t\bar{v}_i$$

$$\sigma_o^2 = E(s_o^2) - \Delta_i^2p_i^2q_{v_iv_i}/f \tag{6-19}$$

$$v^tPv/\sigma_o^2 \sim \chi^2(f, \lambda_i) \quad \text{with} \quad \lambda_i = \Delta_i^2p_i^2q_{v_iv_i}/\sigma_o^2 \tag{6-20}$$

where $q_{v_iv_i}$ is the diagonal element of Q_v referring to v_i.

The gross error $-\Delta_i$ can be expressed as a function of the non-centrality parameter λ_i of Eq. (6-20):

$$\Delta_i = \frac{\sigma_o}{p_i}\sqrt{\frac{\lambda_i}{q_{v_i v_i}}} \qquad (6\text{-}21)$$

After these preliminaries three different methods of screening the residuals will be presented.

6.4 Baarda's Data Snooping

If the type I error is fixed at $\alpha = \alpha_o$ and the type II error at $\beta = \beta_o$ (refer to Section 5.3) for the test of the residuals, then the non-centrality parameter λ of the χ^2-distribution is solely a function of the degrees of freedom f. The non-centrality parameter can be computed or obtained from tables or nomograms of the non-central χ^2-distribution (see Figure 6.3). It is the offset of the expectation of the test statistic under H_a corresponding to the critical value:

$$\chi_{\alpha_o}{}^2(f)$$

In other words, λ is the offset of the expectation which the test statistic has to attain, in order that the sample value exceeds the critical value with probability $1 - \beta_o$. Denoting the boundary value λ by λ_o this probability is defined by Eq. (6-22):

$$P\{\chi^2(f, \lambda_o) > \chi_{\alpha_o}{}^2(f)\} = 1 - \beta_o \qquad (6\text{-}22)$$

and depicted in Figure 6.2:

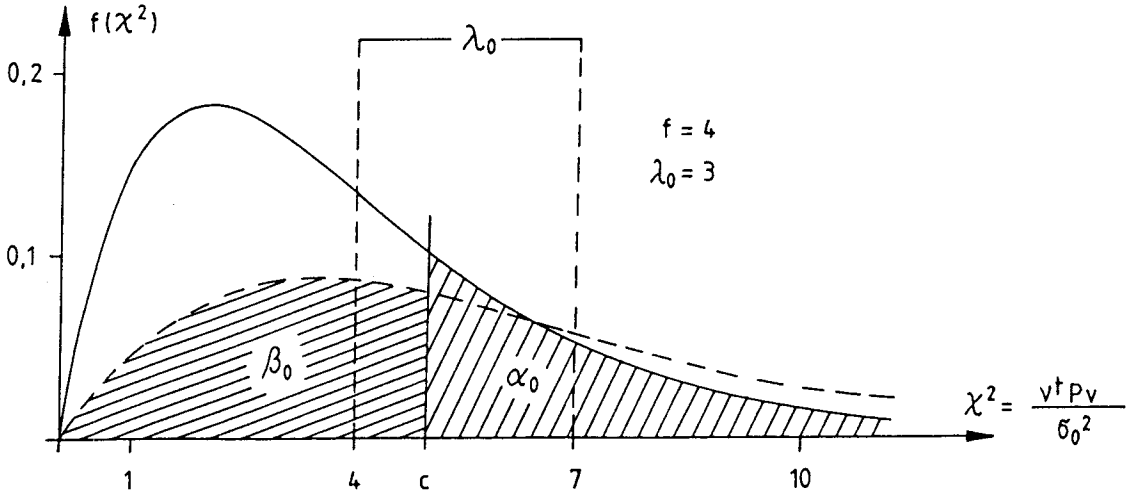


Figure 6.2: Central and non-central χ^2-distribution for given α_o, β_o and f:

$$1 - \alpha_o = \int_o^c f(\chi^2(f))d\chi^2, \qquad 1 - \beta_o = \int_c^\infty f(\chi^2(f, \lambda_o))\, d\chi^2, \qquad c = \chi_{\alpha_o}{}^2(f) = \chi_{1-\beta_o}{}^2(f, \lambda_o)$$

The conventional alternative hypothesis assumes that **one** gross error exists. Hence, from λ_o for f = 1 a critical value Δ_{oi} can be computed for each observation. If a gross error reaches this value Δ_{oi}, then the probability of detecting this error is $1 - \beta_o$ while the probability of rejecting a good observation is α_o.

$$\Delta_{oi} = \frac{\sigma_o}{p_i}\sqrt{\frac{\lambda_o}{q_{v_i v_i}}} \tag{6-23}$$

This critical error affects the vector of residuals:

$$\Delta v_o = q_{v_i} p_i \Delta_{oi} = q_{v_i}\sigma_o\sqrt{\frac{\lambda_o}{q_{v_i v_i}}} \tag{6-24}$$

and the corresponding residual:

$$\Delta v_{oi} = \sigma_o\sqrt{\lambda_o q_{v_i v_i}} \tag{6-25}$$

BAARDA's test assumes that the residuals are normally distributed and that σ_o is known. It is based on the standardized residuals:

$$u_i = \frac{v_i}{\sigma_{v_i}} = \frac{v_i}{\sigma_o\sqrt{q_{v_i v_i}}} = \frac{\bar{v}_i}{\sigma_o\sqrt{q_{v_i v_i}}} + \frac{p_i\sqrt{q_{v_i v_i}}}{\sigma_o}\Delta_i \tag{6-26}$$

This equation can be written as:

$$u_i = \bar{u}_i + \Delta u_i$$

with obvious definitions of $\bar{u}_i$ and Δu_i. The distribution of u_i is given by:

$$u_i \mid H_o \sim N(0, 1); \qquad u_i \mid H_a \sim N(\Delta u_i, 1) \tag{6-27}$$

The critical value:

$$u_{\alpha_o}$$

of the standardized normal distribution depends only on α_o. It will be exceeded with a probability of $1 - \beta_o$ if a gross error Δ_{oi} according to Eq. (6-23) occurs.

Some values of:

$$\sqrt{\lambda_o}$$

for selected risk levels α_o and β_o and for f = 1 degree of freedom (**one** gross error is assumed) are given in Table 6.1.

β_o	α_o in per cent							
in %	0.001	0.005	0.01	0.05	0.1	1.0	2.5	5
10	5.6	5.3	5.2	4.8	4.6	3.9	3.5	3.2
20	5.3	4.9	4.7	4.3	4.1	3.4	3.1	2.8
30	4.9	4.6	4.3	4.0	3.8	3.1	2.8	2.5

Table 6.1: Values of $\sqrt{\lambda_o}$ for α_o, β_o and f = 1. Refer to Figure 6.2 for definition.

For a better understanding of how data snooping works, it is helpful to express the gross error Δ_i as a multiple of the standard deviation of the corresponding observation:

$$\Delta_i = k_i\sigma_i = k_i\sigma_o / \sqrt{p_i} \qquad (6\text{-}28)$$

Substituting Eq. (6-21) yields:

$$k_i = \sqrt{\frac{\lambda_i}{p_i q_{v_i v_i}}}$$

and, considering Eq. (6-23):

$$k_{oi} = \sqrt{\frac{\lambda_o}{p_i q_{v_i v_i}}} \; ; \; \Delta_{oi} = k_{oi}\sigma_i \qquad (6\text{-}29)$$

Therefore, a gross error of an observation l_i has to reach a value of $k_{oi}\sigma_i$ to be detected with probability 1 - β_o. The value k_{oi} solely depends on:

$$\sqrt{\lambda_o}$$

from Table 6.1, the weight p_i of the observation and the cofactor $q_{v_i v_i}$ of the corresponding residual.

It is possible to compute these values prior to the execution of any observation to investigate the feasibility of detecting gross errors in the respective GMM.

The risk level α of the f-dimensional global test of Section 6.1 must relate to the one-dimensional test with type I error α_o and type II error β_o in such a way that a gross error Δ_{oi} causes the test to fail with a probability of $1 - \beta_o$. This is achieved by using the same H_a, i.e. the non-centrality parameter λ_o must be the same for both tests. This is symbolically expressed in Eq. (6-30):

$$\lambda_o = \lambda(\alpha, \beta_o, f) = \lambda(\alpha_o, \beta_o, 1) \qquad (6\text{-}30)$$

Solution for α yields the type I error probability of the global test. Since the explicit form of Eq. (6-30) is rather involved, it is usually preferred to use nomograms. Figure 6.3 depicts a nomogram for $\beta_o = 20\%$. Additional nomograms may be found in BAARDA (1968).

$$\lambda_o = \lambda\{\alpha_o, \beta_o = 0.20, 1\} = \lambda\{\alpha, \beta_o = 0.20, f\}$$

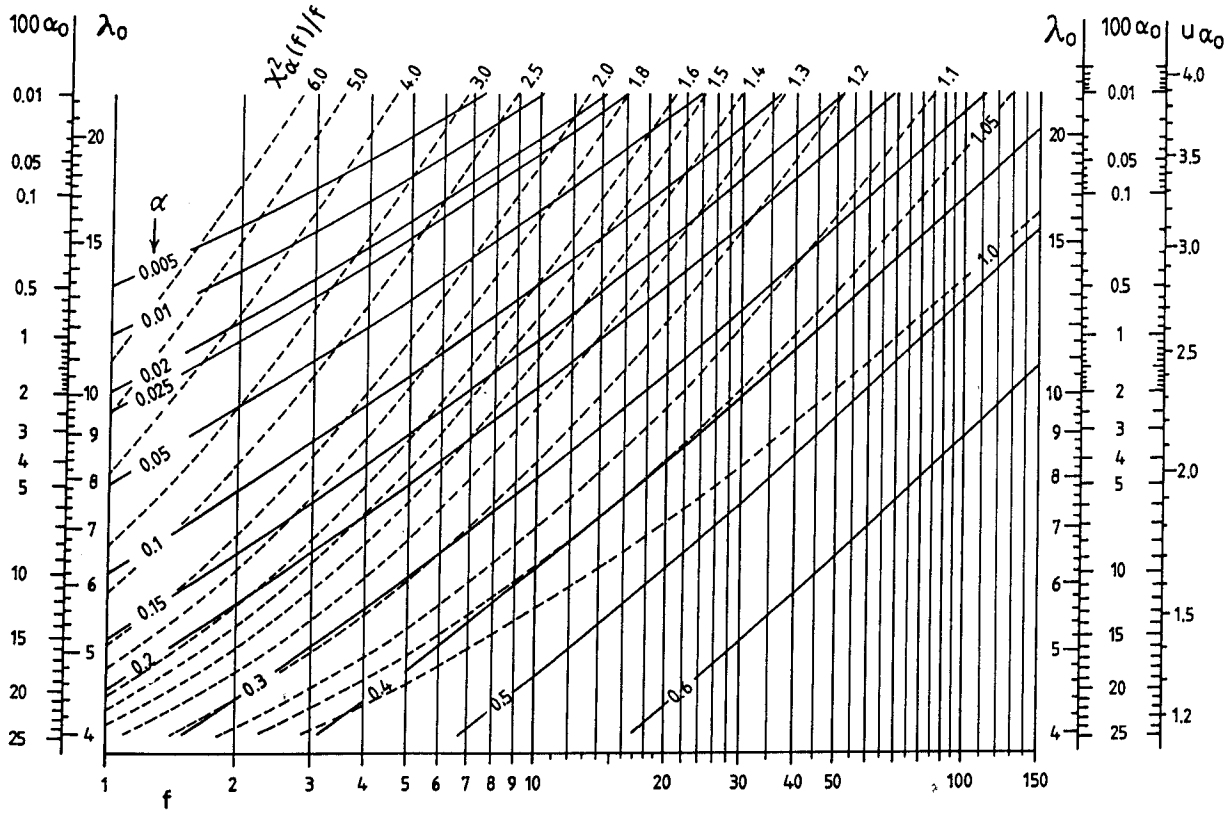


Figure 6.3: Nomogram for the determination of λ_o and u_{α_o} as functions of α_o, and of α and $\chi_\alpha^2(f)$ as functions of α_o and f, for $\beta_o = 0.20$. Similar nomograms for $\beta_o = 0.10$ and $\beta_o = 0.30$ may be found in BAARDA (1968).

Example: $100\alpha_o = 0.1 \Rightarrow \lambda_o = 17.0, \quad u_{\alpha_o} = 3.29$

$100\alpha_o = 0.1, \quad f = 20 \Rightarrow \alpha = 0.11, \quad \chi_\alpha^2(f) = 20 \cdot 1.4 = 28.0$

Based on 20 years of experience with data snooping (in the Netherlands, data snooping is a standard requirement of the cadastral surveying regulations), BAARDA recommends that probabilities of $\beta_o = 20\%$ and $\alpha_o = 0.1\%$ be used. A risk level $\alpha = 1\%$ corresponds roughly to these values considering a degree of freedom of f = 20. Larger networks should be partitioned into subnetworks of suitable size for this test.

6.5 Pope's Tau Method

Different to BAARDA's data snooping philosophy, POPE considers the true variance factor σ_o^2 to be unknown in practice. Consequently, he does not apply the global test of Section 6.1. POPE's approach is based on the Studentized residuals:

$$v_i / s_{v_i}$$

Unfortunately,

$$v_i \text{ and } s_{v_i}$$

are usually estimated from the same data and are therefore statistically dependent. This means that the t-distribution does not apply to this ratio. Instead, the ratio:

$$v_i / s_{v_i}$$

is governed by the rather involved τ-distribution with f = n – r degrees of freedom:

$$T_i = \frac{|v_i|}{s_{v_i}} = \frac{|v_i|}{s_o \sqrt{q_{v_i v_i}}} \sim \tau_{(f)} \qquad (6\text{-}31)$$

Tables of the τ-distribution are not as easily accessible as those of the t-distribution. It is therefore sometimes convenient to transform a τ-variate into a t-variate and vice-versa (POPE, 1976):

$$\tau_{(f)} = \frac{\sqrt{f}\, t_{(f-1)}}{\sqrt{f - 1 + t_{(f-1)}^2}} \qquad (6\text{-}32)$$

$$t_{(f-1)} = \sqrt{\frac{(f-1)\,\tau_{(f)}^2}{f - \tau_{(f)}^2}} \quad \text{for } \tau^2 < f \qquad (6\text{-}33)$$

Nomograms or tables of the t-distribution can then be used.

The null hypothesis of the τ-test assumes that all observations are normally distributed with $E(l) = Ax$, so that the residuals of the LS-estimation in the GMM have an expectation of zero:

$$H_o : \quad E(v_i) = 0 \qquad \forall\, i \in \{1, 2, ..., n\} \qquad (6\text{-}34)$$

The conventional alternative hypothesis is:

H_a : **one** residual is an outlier.

The probability of the type I error of the test, which consists of n individual tests, is usually chosen at $\alpha = 5\%$. A rigorous computation of the risk level α_o of the n one-dimensional tests is not possible, as the residuals and consequently the tests are statistically dependent.

An approximate value for α_o, ignoring this dependency, can be given as:

$$\alpha_o \approx 1 - (1 - \alpha)^{1/n} \tag{6-35}$$

According to the rules of the test, H_o is rejected for a residual v_k, if the inequality:

$$T_k > \tau_{\alpha_o/2}(f) \tag{6-36}$$

holds, which is a two-tailed test as defined in Section 5. The observation corresponding to the tested residual is by definition an outlier and therefore a candidate for further investigation.

Data snooping as well as the τ-method of testing are based on the assumption that just **one** observation of the GMM is affected by a gross error. The whole theory breaks down, if the observations include two or more gross errors. The following pragmatic approach, which usually but not always gives the correct results, is recommended, if more than one test statistic T exceeds the critical value of the one-dimensional test. The observation with the greatest test statistic is discarded. The adjustment is repeated with the remaining n – 1 observations, leading to new residuals and a new estimate s_o^2 of the variance factor. Then a new α_o is computed and the test is applied again with f – 1 degrees of freedom. This process is repeated until all outliers are flagged.

Both test strategies of Sections 6.4 and 6.5 have been applied extensively and proved to be very useful in detecting and eliminating blunders in geodetic observations. This success is somewhat surprising, considering the assumptions under which the methods have been developed:

i. the observations are normally distributed,

ii. the GMM is perfect except for **one** gross observational error,

iii. the model is such, that the gross error maximises the test statistic of the corresponding residual,

iv. the n one-dimensional tests are statistically independent.

6.6 Danish Method

The Danish method of treating outlying observations has been developed after a proposal by KRARUP (KRARUP, JUHL and KUBIK, 1980).

It has been successfully applied in geodetic and photogrammetric adjustments for many years. The basic idea is that large residuals indicate less accurate observations and vice-versa. Therefore after a conventional LS-estimation of the parameters of the GMM, the a priori weights are replaced by new ones being functions of the residuals. Then a new estimation is carried out and leads to new residuals, from which again new weights are computed. This process of estimation and modification of weights is repeated until convergence is achieved. Typically 5 - 10 iterations are necessary. The Danish method is purely heuristic; there is no underlying probabilistic or statistical theory, no assumptions are introduced with respect to the stochastic property of the observations and, consequently, no statistical tests are carried out.

Various reweighting functions have been proposed, which derive from the set of equations (Eq. (6-37)) in most cases:

$$p_{v+1} = p_v f(v_v), \qquad v = 1, 2, \ldots \tag{6-37}$$

$$f(v_v) = \begin{cases} 1 & \text{for } \dfrac{|v_v|\sqrt{p_1}}{s_o} < c \\ \exp - \dfrac{|v_v|\sqrt{p_1}}{cs_o} & \text{else} \end{cases}$$

Equation (6-37) defines an interval:

$$-c < v\sqrt{p_1}\,/s_o < c$$

in which the a priori weights are maintained. The weights of all observations with residuals outside this interval are reduced. The constant c is usually selected between 2 and 3 depending on the redundancy of the GMM and the quality of the data. The objective of the weighting strategy is to reduce the influence of outlying observations on the estimates of the parameters and to obtain results which are closer to their expectation.

All residuals greater than a selected boundary value are considered to indicate gross errors. The corresponding observations are inspected thoroughly and, if justified, corrected. Obvious blunders are deleted and, if necessary, remeasured.

This very effective and simple method of treating gross errors in the GMM deserves much more attention than it has received in the past. It is closely related to the method of robust estimation as advocated by HUBER (1964, 1981).

There are two alternatives for the final parameter estimation:

i. after discarding all outliers a new LS-adjustment with the a priori weights is carried out,

ii. the parameters of the last iteration step are regarded as the final and best estimates. No observations are discarded.

It is stressed that none of the methods of outlier detection should be used as a black box which automatically cleans the observational data if implemented in an adjustment program. All methods should be used in an interactive mode with the computer flagging the suspicious observations, and the surveyor deciding on what to do with them and how to proceed.

6.7 Example

The monitoring network previously introduced in Section 4.3 is used to demonstrate the methods of outlier detection and localization. The monitoring network Montsalvens is characterized by:

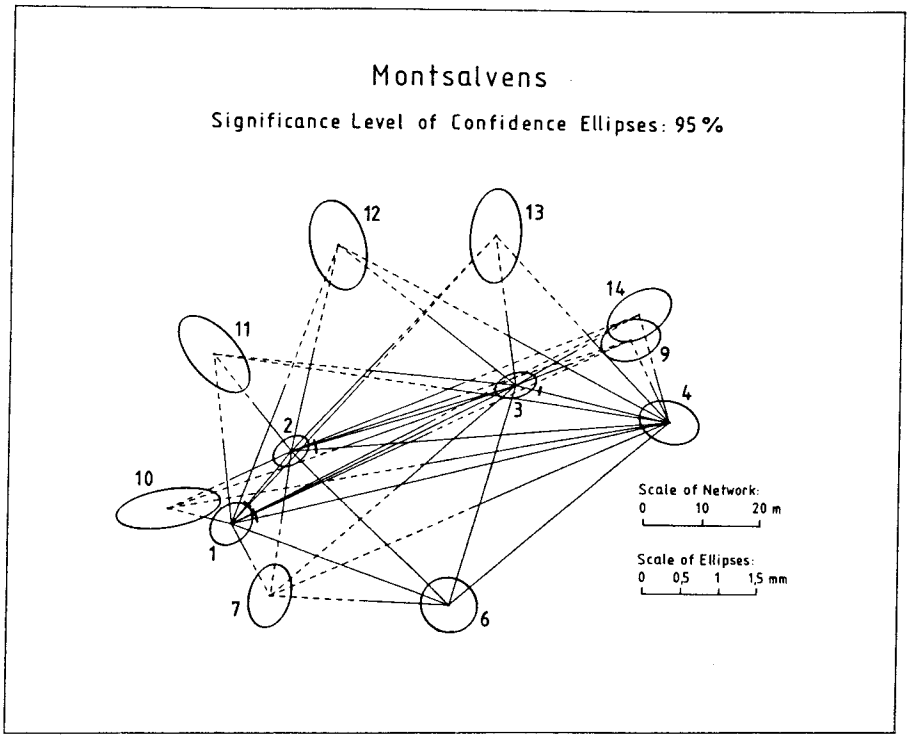


Figure 6.4 : Monitoring Network Montsalvens:

49 directions, $\sigma = 3^{cc}$ (The dimension decimal second (1^{cc} = 0.1 mgon ≈ 0.3") is used here for convenience.)

6 distances, σ = 0.3 mm;

24 coordinates; 5 orientation unknowns; 29 degrees of freedom; 3 rank deficiencies; $\sigma_o = 1^{cc}$.

weights: directions $p = \frac{1}{9}$, distances $p = \frac{1}{0.09}\left[\frac{cc}{mm}\right]^2$

The adjustments are based on a datum according to Section 3.7 (tr $Q_{\hat{x}}$ = min). The observations are assumed to be uncorrelated.

i. **Adjustment with properly generated normally distributed observations (without gross errors):**

$$v^tPv = 22.97, \qquad f = 29, \qquad s_o^2 = 0.79$$

BAARDA's data snooping:

α_o = 0.1%, β_o = 20% chosen for a single test. From the nomogram (Figure 6.3) is taken:

$$\sqrt{\lambda_o} = 4.1, \qquad u_{\alpha_o} = 3.29, \qquad \chi^2(29) = 36.5, \qquad \alpha = 0.15$$

Global test:

$$T = \frac{v^t P v}{\sigma_o^2} = 22.97, \qquad \chi_{0.15}^2(29) = 36.5$$

Since $T < \chi_\alpha^2(f)$ the null hypothesis is not rejected and the assumptions of the mathematical model are confirmed by the test. Hence no further data snooping is required.

POPE's data screening:

Type I error $\alpha = 5\%$ is chosen for the simultaneous test of $n = 55$ observations for outliers. The type I error of the single tests is:

$$\alpha_o \approx 1 - (1 - 0.05)^{1/55} = 0.00093 \approx 0.09\%$$

For $f - 1 = 28$ and $\alpha_o = 0.09\%$ the value of the t-distribution is interpolated in tables for the two-tailed t-test yielding $t(f - 1) = 3.69$. The critical value $\tau(f)$ is calculated according to Eq. (6-32) resulting in:

$$\tau = 3.08$$

The test stastic T_i is computed for all observations (Column 8 of Table 6.2) and compared with τ. Since none of the values exceeds τ the absence of outliers is confirmed.

Danish method:

The Danish method of Section 6.6 has been applied to the same data set with $c = 3$ as the constant of Eq. (6-37). The residuals of this adjustment are listed in Column 9 of Table 6.2. The comparison with Column 3, where the residuals of the ordinary LS-solution are compiled, shows a perfect coincidence, confirming that the Danish method yields classical LS-results if no gross errors are present.

Smallest detectable errors:

Independent of the actual occurrence of gross errors it is possible to compute boundary values of errors which are just detectable if BAARDA's data snooping is applied. Column 4 of Table 6.2 gives these values for the network Montsalvens using the risk levels $\alpha_o = 0.1\%$ and $\beta_o = 20\%$. Their effect on the corresponding residuals can be seen from comparison with Column 5. In Column 6 the values k_{oi} express the critical errors as multiples of the standard deviation. The interpretation of these columns in conjunction with Figure 6.4 is obvious and indicates those observations where gross errors are hardly detectable.

ii. **Adjustment using the same observations as under (i) except for two gross errors:**

Observation No. 21 : $\Delta = 10\sigma = 30^{cc}$

Observation No. 51 : $\Delta = 7\sigma = 2.1$ mm

Table 6.3 comprises all results of the three applied methods of data screening.

No	from	to	LS v_i	(6-23) Δ_{oi}	(6-25) Δv_{oi}	(6-29) k_{oi}	BAARDA (6-26) u_i	POPE (6-31) T_i	DANISH METHOD v_i
1	2		3	4	5	6	7	8	9
Directions in cc									
1	1	2	0.76	16.40	9.34	5.5	0.34	0.38	0.76
2	1	13	0.40	15.76	9.72	5.3	0.17	0.19	0.40
3	1	14	1.90	13.99	10.95	4.7	0.72	0.81	1.90
4	1	3	-2.02	13.65	11.22	4.5	0.74	0.84	-2.02
5	1	9	-3.51	13.91	11.01	4.6	1.32	1.48	-3.51
6	1	4	0.42	14.07	10.89	4.7	0.16	0.18	0.42
7	1	6	-2.07	16.16	9.48	5.4	0.90	1.01	-2.07
8	1	7	1.70	35.05	4.37	11.7	1.60	1.80	1.70
9	1	10	-0.04	236.01	0.65	78.6	0.27	0.30	-0.04
10	1	11	-0.56	28.39	5.39	9.5	0.43	0.48	-0.56
11	1	12	3.03	16.91	9.06	5.6	1.38	1.55	3.03
12	2	1	0.67	19.41	7.89	6.5	0.35	0.39	0.67
13	2	10	0.21	41.00	3.74	13.7	0.24	0.27	0.21
14	2	11	0.49	28.38	5.40	9.5	0.37	0.42	0.49
15	2	12	-2.57	22.63	6.77	7.5	1.57	1.76	-2.57
16	2	13	-0.69	19.09	8.02	6.4	0.36	0.40	-0.69
17	2	14	-1.50	14.44	10.61	4.8	0.58	0.65	-1.50
18	2	9	0.95	14.26	10.74	4.8	0.36	0.41	0.95
19	2	3	2.32	14.38	10.65	4.8	0.90	1.01	2.32
20	2	4	1.80	14.31	10.70	4.8	0.69	0.78	1.80
21	2	6	1.66	17.82	8.60	5.9	0.80	0.90	1.66
22	2	7	-3.35	20.22	7.57	6.7	1.82	2.05	-3.35
23	3	1	-1.14	14.27	10.73	4.8	0.44	0.49	-1.14
24	3	10	-1.49	14.24	10.76	4.7	0.57	0.64	-1.49
25	3	2	2.00	14.51	10.56	4.8	0.78	0.88	2.00
26	3	11	-3.29	18.91	8.10	6.3	1.67	1.88	-3.29
27	3	12	2.15	24.88	6.16	8.3	1.44	1.62	2.15
28	3	13	0.34	27.76	5.52	9.2	0.26	0.29	0.34
29	3	14	-0.01	30.96	4.95	10.3	0.01	0.01	-0.01
30	3	9	0.67	34.20	4.48	11.4	0.61	0.69	0.67
31	3	4	1.09	20.16	7.60	6.7	0.59	0.67	1.09
32	3	6	-1.66	18.20	8.41	6.1	0.81	0.91	-1.66
33	3	7	1.34	15.04	10.19	5.0	0.54	0.61	1.34
34	4	1	0.92	13.72	11.17	4.6	0.34	0.38	0.92
35	4	10	1.55	13.66	11.21	4.6	0.57	0.64	1.55
36	4	2	-3.51	13.60	11.26	4.5	1.29	1.44	-3.51
37	4	11	4.05	14.92	10.26	5.0	1.63	1.83	4.05
38	4	3	-1.29	15.49	9.89	5.2	0.54	0.61	-1.29
39	4	12	-2.78	16.06	9.54	5.3	1.20	1.35	-2.78
40	4	13	-0.46	21.19	7.23	7.1	0.26	0.29	-0.46
41	4	9	0.05	673.18	0.23	224.3	0.98	1.10	0.05
42	4	14	-0.05	304.46	0.50	101.4	0.38	0.43	-0.05
43	4	6	-0.64	15.84	9.67	5.3	0.27	0.31	-0.64
44	4	7	2.16	14.30	10.71	4.8	0.83	0.93	2.16
45	6	1	-2.42	16.11	9.51	5.4	1.05	1.18	-2.42
46	6	7	0.91	21.55	7.11	7.2	0.53	0.60	0.91
47	6	4	-2.09	18.50	8.28	6.2	1.04	1.17	-2.09
48	6	3	3.00	17.02	9.00	5.7	1.38	1.55	3.00
49	6	2	0.60	16.29	9.40	5.4	0.26	0.29	0.60
Distances in mm									
50	1	2	-0.31	1.32	1.16	4.4	1.10	1.24	-0.31
51	1	3	0.38	1.46	1.05	4.9	1.51	1.70	0.38
52	1	4	-0.25	1.59	0.96	5.3	1.08	1.21	-0.25
53	2	3	-0.23	1.38	1.11	4.6	0.85	0.96	-0.23
54	2	4	0.17	1.50	1.02	5.0	0.68	0.76	0.17
55	3	4	0.07	1.38	1.10	4.6	0.25	0.28	0.07

Table 6.2: Residuals v_i of adjustment (i), least detectable error Δ_{oi} and its effect Δv_{oi} on the corresponding residual, factor k_{oi} of Eq. (6-29), the test statistics u_i of BAARDA and T_i of POPE and the residuals of the Danish Method.

Baarda's data snooping:

The procedure is the same as outlined in paragraph (i) and commences with a LS-adjustment.

Step 1: Adjustment results:

$$\bar{v}^t P \bar{v} = 79.43, \qquad f = 29, \qquad s_o^2 = 2.74$$

Global test:

$$T = 79.43, \qquad \chi_{0.15}^2(29) = 36.5$$

Since $T > \chi_\alpha^2(f)$ the null hypothesis is rejected and the observations have to be screened. The residuals $\bar{v}_i$ (Column 3) are normalized (Eq. 6-26) yielding $\bar{u}_i$ (Column 6) and the critical value:

$$u_{\alpha_o} = 3.29$$

is interpolated in Figure 6.3. The test statistics $\bar{u}_i$ of observations 21, 47, 49 and 51 exceed the critical value but only observation 21 with the maximum value of $\bar{u}_i$ is deleted.

Step 2: The LS-adjustment is repeated without observation 21:

$$\bar{v}^t P \bar{v} = 39.85, \qquad f = 28, \qquad s_o^2 = 1.42$$

Global test:

$$T = 39.85, \qquad \chi_{0.15}^2(28) = 35.6$$

Since $T > \chi_\alpha^2(f)$ the null hypothesis has to be rejected again. The residuals $\bar{v}_i$ (Column 4) are normalized yielding $\bar{u}_i$ (Column 7). The critical value is independent of f again:

$$u_{\alpha_o} = 3.29$$

The test statistic $\bar{u}_i$ of observation 51 exceeds the critical value and is deleted.

Step 3: The LS-adjustment is repeated without observations 21 and 51:

$$v^t P v = 20.11, \qquad f = 27, \qquad s_o^2 = 0.74$$

Global test:

$$T = 20.11, \qquad \chi_{0.15}^2(27) = 34.7$$

No	from	to	LS - Residuals $\bar{v}_i$	$\bar{\bar{v}}_i$	v_i	BAARDA Test $\bar{u}_i$	$\bar{\bar{u}}_i$	u_i	POPE Test $\bar{T}_i$	$\bar{\bar{T}}_i$	T_i	Danish Method v_i	p_i
1	2		3	4	5	6	7	8	9	10	11	12	13
Directions in cc													
1	1	2	0.32	1.13	0.78	0.14	0.50	0.34	0.08	0.42	0.40	0.77	0.111
2	1	13	-0.15	1.19	0.47	0.07	0.51	0.20	0.04	0.42	0.23	0.46	0.111
3	1	14	1.02	1.97	1.83	0.38	0.74	0.69	0.23	0.62	0.80	1.82	0.111
4	1	3	-3.55	-2.18	-2.19	1.30	0.80	0.81	0.79	0.67	0.94	-2.19	0.111
5	1	9	-4.56	-3.69	-3.64	1.71	1.38	1.36	1.03	1.16	1.58	-3.64	0.111
6	1	4	-0.54	-0.51	0.18	0.20	0.19	0.07	0.12	0.16	0.08	0.18	0.111
7	1	6	4.31	-2.52	-1.53	1.88	1.24	0.76	1.13	1.04	0.88	-1.51	0.111
8	1	7	-0.30	1.07	1.41	0.28	1.03	1.36	0.17	0.86	1.58	1.40	0.111
9	1	10	0.05	-0.01	-0.03	0.34	0.06	0.18	0.20	0.05	0.21	-0.02	0.111
10	1	11	0.36	0.03	-0.38	0.28	0.02	0.29	0.17	0.02	0.34	-0.37	0.111
11	1	12	3.03	3.53	3.10	1.38	1.61	1.42	0.83	1.35	1.64	3.10	0.111
12	2	1	1.82	0.81	0.80	0.95	0.43	0.42	0.57	0.36	0.49	0.80	0.111
13	2	10	0.83	0.34	0.29	0.91	0.38	0.32	0.55	0.31	0.37	0.29	0.111
14	2	11	-0.44	-0.11	0.31	0.34	0.08	0.23	0.20	0.07	0.27	0.30	0.111
15	2	12	-2.51	-2.83	-2.61	1.53	1.73	1.59	0.92	1.45	1.84	-2.60	0.111
16	2	13	0.56	0.10	-0.45	0.29	0.05	0.23	0.17	0.04	0.27	-0.44	0.111
17	2	14	0.50	-1.06	-1.24	0.19	0.41	0.48	0.12	0.35	0.56	-1.23	0.111
18	2	9	2.97	1.18	1.18	1.14	0.45	0.45	0.69	0.38	0.53	1.18	0.111
19	2	3	4.31	2.39	2.52	1.67	0.93	0.98	1.01	0.78	1.14	2.52	0.111
20	2	4	4.44	1.40	1.99	1.71	0.55	0.78	1.03	0.46	0.91	1.99	0.111
21	2	6	-13.11	-	-	6.29*	-	-	3.80*	-	-	-26.68	0.000*
22	2	7	0.65	-2.21	-2.79	0.35	1.24	1.57	0.21	1.04	1.82	-2.78	0.111
23	3	1	-3.49	-1.57	-1.43	1.34	0.61	0.55	0.81	0.51	0.64	-1.43	0.111
24	3	10	-2.87	-1.60	-1.64	1.10	0.62	0.63	0.66	0.52	0.73	-1.64	0.111
25	3	2	1.57	2.01	1.96	0.61	0.78	0.77	0.37	0.66	0.89	1.95	0.111
26	3	11	-2.65	-2.92	-3.17	1.35	1.49	1.61	0.81	1.25	1.87	-3.17	0.111
27	3	12	1.75	1.52	2.02	1.17	1.02	1.35	0.71	0.86	1.57	2.01	0.111
28	3	13	-0.39	-0.89	0.08	0.29	0.67	0.06	0.18	0.56	0.07	0.08	0.111
29	3	14	-0.52	-0.20	-0.09	0.43	0.17	0.08	0.26	0.14	0.09	-0.09	0.111
30	3	9	0.27	0.64	0.62	0.24	0.59	0.57	0.15	0.49	0.67	0.62	0.111
31	3	4	2.33	2.15	1.38	1.26	1.17	0.75	0.76	0.98	0.87	1.37	0.111
32	3	6	4.79	-0.45	-0.86	2.35	0.24	0.46	1.42	0.20	0.54	-0.84	0.111
33	3	7	-0.79	1.32	1.14	0.32	0.54	0.46	0.19	0.45	0.54	1.13	0.111
34	4	1	-0.71	-0.02	0.61	0.26	0.01	0.23	0.16	0.01	0.26	0.61	0.111
35	4	10	0.51	1.02	1.37	0.19	0.37	0.50	0.11	0.31	0.58	1.36	0.111
36	4	2	-3.14	-3.70	-3.50	1.15	1.35	1.28	0.69	1.14	1.49	-3.50	0.111
37	4	11	4.96	4.70	4.24	1.99	1.89	1.70	1.20	1.58	1.97	4.23	0.111
38	4	3	0.09	0.03	-0.96	0.04	0.01	0.40	0.02	0.01	0.46	-0.95	0.111
39	4	12	-2.15	-1.71	-2.55	0.93	0.74	1.11	0.56	0.62	1.28	-2.55	0.111
40	4	13	0.52	1.16	-0.11	0.30	0.66	0.07	0.18	0.55	0.08	-0.11	0.111
41	4	9	0.10	0.06	0.06	1.79	1.09	1.08	1.08	0.92	1.25	0.05	0.111
42	4	14	0.04	-0.02	-0.04	0.31	0.20	0.29	0.18	0.17	0.33	-0.03	0.111
43	4	6	-0.61	-2.54	-0.94	0.26	1.09	0.41	0.16	0.92	0.47	-0.93	0.111
44	4	7	0.41	1.03	1.82	0.16	0.40	0.70	0.09	0.33	0.81	1.81	0.111
45	6	1	-2.75	-2.90	-2.53	1.19	1.26	1.10	0.72	1.06	1.27	-2.52	0.111
46	6	7	-0.23	0.52	0.74	0.14	0.30	0.43	0.08	0.25	0.50	0.74	0.111
47	6	4	-8.10	-4.02	-2.96	4.03*	2.11	1.57	2.44	1.77	1.82	-2.97	0.111
48	6	3	3.56	4.46	3.28	1.63	2.05	1.52	0.99	1.72	1.76	3.28	0.111
49	6	2	7.52	1.95	1.47	3.30*	0.93	0.70	1.99	0.78	0.81	1.47	0.111
Distances in mm													
50	1	2	0.08	-0.07	0.24	0.28	0.27	0.85	0.17	0.22	0.98	-0.23	11.111
51	1	3	-1.17	-1.13	-	4.60*	4.44*	-	2.78	3.73*	-	-1.57	0.001*
52	1	4	0.34	0.28	-0.11	1.46	1.20	0.52	0.88	1.01	0.60	-0.11	11.111
53	2	3	-0.02	0.15	-0.15	0.09	0.55	0.58	0.05	0.46	0.67	-0.15	11.111
54	2	4	0.48	0.51	0.25	1.94	2.07	1.05	1.17	1.74	1.21	0.25	11.111
55	3	4	0.20	0.05	0.08	0.76	0.20	0.29	0.46	0.16	0.33	0.07	11.111

Table 6.3: Results of three different methods of data screening applied to the network of Figure 6.4 with two simulated gross errors. The test statistics marked with an asterisk exceed the critical value or are weighted down drastically.

Since $T < \chi_\alpha(f)$ the null hypothesis is not rejected, confirming that all erroneous observations have been deleted. The final residuals v_i are shown in Column 5. None of the test statistics u_i exceeds the critical value:

$$u_{\alpha_o} = 3.29$$

Pope's data screening:

Step 1: A LS-adjustment is carried out yielding the residuals $\bar{v}_i$ of Column 3 which are used to compute the test statistics $\bar{T}_i$ of Column 9 according to Eq. (6-31). The critical value for f = 29 is computed in the same way as described in (i) yielding $\tau_\alpha(f) = 3.08$. Observation 21 exceeds this value and is deleted.

Step 2: A new LS-adjustment without observation 21 is carried out. The residuals $\bar{v}_i$ and the corresponding test statistics $\bar{T}_i$ are given in Columns 4 and 10, respectively. The critical value for f = 28 is $\tau_\alpha(f) = 3.07$. Observation 51 exceeds this value and is deleted.

Step 3: A new LS-adjustment without observations 21 and 51 is carried out. The residuals v_i and the corresponding test statistics T_i are given in Columns 5 and 11, respectively. None of the statistics exceeds the critical value $\tau_\alpha(27) = 3.06$. The efficiency of the method is confirmed since both simulated gross errors have been detected and deleted.

Danish method:

The Danish method with c = 3 yields the final result after 4 iterations. The residuals v_i are contained in Column 12; they coincide with the residuals of the third step of the other methods. The final weights of the observations, given in Column 13, show that the erroneous observations 21 and 51 are of no influence in the model.

All three methods have correctly detected the two gross errors. In general, when an unknown (usually larger) number of outlying observations exist, the Danish method is faster than the other two, since it arrives at the final result in one computational step without repeatedly requiring statistical tests. Another argument in favour of the Danish method is that it does not require the deletion of outlying observations. Instead, these are weighed down according to their departure from the bulk of the data. Furthermore, the method is not based on questionable assumptions but developed heuristically.

Whenever difficult problems are encountered, it may be wise to use all three methods in order to get as much information as possible.

7. CONCEPT OF RELIABILITY

Since deformation analysis using geodetic observations is based on the comparison of estimated positions of points of monitoring networks, at different epochs of time, the quality of the network is of paramount concern to the engineer in charge. The traditional way of describing the quality of a geodetic network by means of measures of accuracy has been outlined in Chapter 4. Its result is not really satisfactory, because it is based on datum dependent quantities, and neglects the aspect of reliability.

Consider for example a local trilateration network observed with an EDM instrument exhibiting an undetected frequency error. If no other errors occur, all measures of accuracy will indicate a good result, and yet the adjusted coordinates are biased. The only way of detecting this error is through an independent piece of information on the scale of the network. If the observation scheme is not designed to acquire this information, then the network design and, as a consequence, the estimated positions are unreliable. A deformation analysis based on biased estimates can lead to completely wrong conclusions.

In this simple example the meaning of reliability is rather clear, but in general it is as indistinct as the term accuracy. There is no absolute reliability and different measures refer to different aspects of the concept. Furthermore, some measures depend on the geodetic datum, thus causing the same problems as discussed in the context of accuracy.

Nevertheless, the concept of reliability is important for the assessment of geodetic networks. Measures of accuracy and measures of reliability form together a sufficient basis for the assessment and comparison of the quality of geodetic networks.

In quality control, the reliability of manufacturing processes and of products has always been of great concern. The different measures used in this field are usually functions of the time of continuous proper functioning of a device or a part thereof. The most commonly used measure, playing an important role in specifications for military equipment, is the so-called Mean Time Between Failures, MTBF. Of course, geodesists cannot adopt this concept without modifications. But the fundamental ideas, which have proved useful in quality control, have stimulated geodesists to develop a concept of reliability for the GMM. The first suggestions were published by BAARDA (1968). Since then numerous other geodesists have contributed to the concept. However, a lot of further research is necessary.

7.1 General Aspects and Definitions

Generally speaking, reliability of geodetic operations requires:

i. well-trained, responsible and careful, that is, reliable surveying professionals,

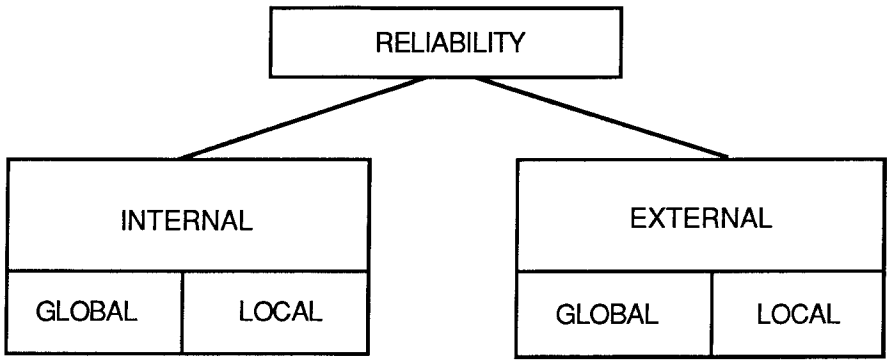
ii. precise and self-checking instruments, which rather cease measuring than give uncertain results, for example, when the environmental conditions or the supply voltage become critical,

iii. self-checking measurement procedures, for example, forward and backward runs, face left and face right observations, closed traverses and levelling loops all provide a field check for gross errors and, possibly, systematic errors,

iv. self-checking network designs, that is, designs in which all observations are checked by other ones.

Below, measures of reliability are aimed at which are suitable for a rigorous mathematical formulation. They shall be based on the GMM and on specific assumptions on the distributional properties of the observables.

Following a suggestion of Baarda, the main criteria of reliability of geodetic networks can be classified as:



The **internal reliability** refers to the desired property of the GMM of facilitating the detection of systematic errors and the localization of gross errors without requiring additional information (self-checking model).

The **external reliability** of the GMM measures the response of the model to undetected systematic and gross errors or, in other words, the effect on the parameter estimation.

The measures or criteria are usually different for local and global considerations.

There is a close relationship to the theory of robust estimation. Refer to Section 6.6, Chapter 10 and HUBER (1964) as far as the objective is concerned. The approaches are different however. The concept of robustness aims at the design of estimators, which are insensitive to deviations from the underlying model, while the reliability concept retains the LS-estimator and aims at a correction or refinement of the model.

The underlying Gauss-Markov model (GMM) has been defined in Chapter 2:

$$\begin{aligned} l + v &= A\hat{x}, \qquad \Sigma = \sigma_o^2 Q, \qquad P = Q^{-1} \\ E(l) &= Ax \\ o(A) &= n \times u, \qquad r(A) = r \leq u < n \end{aligned} \tag{2-2}$$

7.2 Global Measures of Internal Reliability

The measures of internal reliability are equivalent to the probability of detecting deviations from the model, especially gross and systematic errors. The term global means that the detection as such is considered, independent of the possibility of localization.

An obvious and simple measure is the number of redundant observations, if evenly distributed in the network, since it parallels the probability of error detection.

A more sophisticated criterion of reliability is related to the global model test. It has been pointed out in the previous paragraph that an error in the functional model of the GMM can always be compensated by a vector Δ which is added to the erroneous vector l of observations to yield the conforming vector $\bar{l}$:

$$\begin{aligned} \bar{l} &= l + \Delta \\ E(\bar{l}) &= Ax, \qquad E(l) = Ax - \Delta \end{aligned} \tag{6-7}$$

The quadratic form of the vector v of residuals:

$$\frac{v^t P v}{\sigma_o^2} \sim \chi^2(f, \lambda) \qquad (6\text{-}13)$$

has a non-central χ^2-distribution with $f = n - r$ degrees of freedom and a non-centrality parameter:

$$\lambda = \Delta^t P Q_v P \Delta / \sigma_o^2 \qquad (6\text{-}14)$$

The power of the global test, namely, the probability of detecting the existence of the error vector $-\Delta$, depends directly on the value of λ. An analysis of Eq. (6-14) shows that vectors Δ yielding $\lambda = 0$ cannot be detected by the test. The form matrix of Eq. (6-14), PQ_vP, is orthogonal to the design matrix A:

$$PQ_vPA = 0 \qquad (7\text{-}1)$$

Thus any error vector Δ of the form:

$$\Delta = Ab \qquad (7\text{-}2)$$

for some b, yields $\lambda = 0$.

The geometrical interpretation of this result is depicted in Fig. 7.1. An error parallel to $A\hat{x}$, i.e. $\Delta \in S(A)$, does not change the residuals and is therefore undetectable. An error parallel to v and orthogonal to S(A) creates a maximum value of λ. Since, in practice, the vector Δ cannot be controlled, one has to be aware of the fact, that the GMM is entirely unreliable with respect to errors $\Delta \in S(A)$. An example of this unfavourable property is the previously mentioned case of distance measurements with an EDM instrument having a constant frequency error.

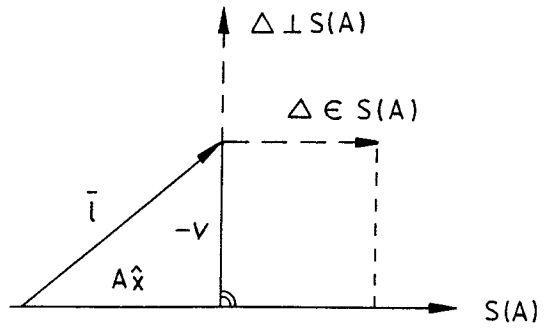


Figure 7.1: Effect of Δ parallel to, and orthogonal to, S(A).

As a basis for the development of a measure of reliability with respect to an error $\Delta \notin S(A)$ the Rayleigh inequality of Eq. (4-16) might be considered:

$$\Lambda_{min} \leq \frac{g^t M g}{g^t g} \leq \Lambda_{max}, \qquad \forall\, g \qquad (4\text{-}16)$$

where Λ represents the eigenvalues of the matrix M and g any vector of suitable order. Comparing Eq. (4-16) with Eq. (6-14) leads, for $M := PQ_vP$ and $\Delta := g$, to the relationship:

$$\lambda = \Delta^t PQ_vP\Delta/\sigma_o^2 \leq \Delta^t\Delta\, \Lambda_{max}/\sigma_o^2 \qquad (7\text{-}3)$$

This provides an upper limit for the non-centrality parameter λ, being proportional to the maximum eigenvalue Λ_{max} of PQ_vP. Since the reliability increases with λ, as has been pointed out, Λ_{max} may be utilized as a global measure of internal reliability. This means that for a given error vector $\Delta \not\in S(A)$ the probability of detecting its existence is related to the magnitude of $\Lambda_{max}(PQ_vP)$. As PQ_vP is datum independent the same applies to Λ_{max}.

But the usefulness of Λ_{max} must not be overrated since it is not clear at all how close λ comes to the upper limit of Eq. (7-3). If all observations are of equal weight, that is $P = pI$, Λ_{max} may become useless as a measure of reliability, because the matrix PQ_vP gets an r-fold ($r = r(A)$) eigenvalue zero and an f-fold ($f = n - r$) eigenvalue p. Hence $\Lambda_{min} = 0$ and $\Lambda_{max} = p$; other eigenvalues do not exist. An increase in the number of observations will definitely increase the probability of error detection but will not change the magnitude of Λ_{max}. In this case Eq. (7-3) only shows that $\lambda \leq \Delta^t\Delta/\sigma^2$ if the weight function $p = \sigma_o^2/\sigma^2$ is used.

An alternative measure may be derived from the general relationship between conforming matrix and vector norms:

$$\| Ax \| \leq \| A \| \cdot \| x \| \qquad (7\text{-}4)$$

Setting $x := \Delta$ and $A := Q_v^{1/2}P$ and using the Euclidean norm, which is defined as:

vector: $\| x \| = (x^t x)^{1/2}$

matrix: $\| A \| = (\mathrm{tr}\, A^t A)^{1/2}$

yields after squaring:

$$\| Q_v^{1/2}P\Delta \|^2 \leq \| Q_v^{1/2}P \|^2 \cdot \| \Delta \|^2$$

or, explicitly:

$$\Delta^t PQ_vP\Delta \leq \Delta^t\Delta(\mathrm{tr}\, PQ_vP) \qquad (7\text{-}5)$$

Substituting Eq. (6-14) yields the inequality:

$$\lambda \leq \frac{\Delta^t\Delta}{\sigma_o^2}\, \mathrm{tr}\, PQ_vP \qquad (7\text{-}6)$$

The equation indicates that $\mathrm{tr}\, PQ_vP$, for a given error vector Δ, defines the upper limit of the non-centrality parameter λ. Thus, $\mathrm{tr}\, PQ_vP$ may serve as a measure of reliability in the same sense as Λ_{max}.

A comparison of Eqs (7-6) and (7-3), together with the relationship tr M = $\Sigma\lambda$ of Eq. (4-14), shows that Λ_{max} is the closer limit for λ and hence should be preferred. On the other hand, the computational effort favours the trace. In the case of equal weight for all observations as considered above, the trace becomes:

$$\text{tr } PQ_vP = p \text{ tr } Q_vP = pf$$

Additional observations increase the number of degrees of freedom f and hence the magnitude of the trace, which indicates the increased probability of error detection.

Both measures are solely suitable for the comparison of model designs and should be used with due care. It seems that further research is still required in this field.

7.3 Local Measures of Internal Reliability

The concept of local reliability is closely related to the probability of outlier detection. The greater the probability of locating erroneous observations the higher is the degree of internal reliability of the model.

The shift Δv of the residual vector, caused by a single gross error Δ_i in the observation l_i, is:

$$\Delta v = -q_{v_i} p_i \Delta_i \qquad (7\text{-}7)$$

(compare Eq. (6-24)), where p_i is the a priori weight of l_i and q_{v_i} the i-th column of the cofactor matrix Q_v. The i-th component of Δv:

$$\Delta v_i = -q_{v_i v_i} p_i \Delta_i = -f_i \Delta_i \qquad (7\text{-}8)$$

is the bias of v_i, the residual of the affected observation.

The greater Δv_i for a single error Δ_i, the greater is the probability of localizing the erroneous observation l_i. Also considering Eq. (7-8), the cofactor matrix Q_v should have strongly dominant diagonal elements $q_{v_i v_i}$. Since for a given error Δ_i the bias Δv_i is governed by:

$$q_{v_i v_i} p_i = f_i \qquad (7\text{-}9)$$

the so-called redundancy contribution, this expression can be used directly as a local measure of internal reliability.

To assess the local reliability of a model, the values of f_i are computed for all observations. This is done before the observations are actually carried out. Screening the values of f_i, exposes in terms of reliability the weak parts of the model which need to be strengthened by additional observations.

Further interpretation and a justification of the term 'redundancy contribution' for f_i of Eq. (7-9) can be derived from the general relationship of Eq. (6-9):

$$\bar{v} - v = \Delta v = -Q_vP\Delta \qquad (7\text{-}10)$$

where:

$$Q_vP = I - AQ_{\hat{x}}A^tP \qquad (7\text{-}11)$$

Since Q_vP is idempotent, i.e. $Q_vP \cdot Q_vP = Q_vP$, the rank of Q_vP equals the trace of this product:

$$\mathrm{tr}\,(Q_vP) = r(Q_vP) = r(I) - r(AQ_{\hat{x}}A^tP)$$

hence:

$$\mathrm{tr}\,(Q_vP) = n - r(A) = f \qquad (7\text{-}12)$$

For a diagonal matrix P (independent observations) Q_vP has diagonal elements:

$$q_{v_iv_i}p_i = f_i$$

Thus:

$$\mathrm{tr}\,(Q_vP) = \Sigma f_i = f \qquad (7\text{-}13)$$

The sum of the redundancy contributions f_i of all observations equals the degrees of freedom (redundancy) of the model. Each individual value of f_i is the contribution of the corresponding observation l_i to the redundancy of the model.

Substituting the cofactor matrix of the adjusted observations:

$$Q_{\hat{l}} = AQ_{\hat{x}}A^t, \qquad \hat{l} = l + v$$

in Eq. (7-11) yields:

$$Q_vP = I - Q_{\hat{l}}P \qquad (7\text{-}14)$$

This allows Eq. (7-9) to be rewritten as:

$$f_i = 1 - q_{\hat{l}_i\hat{l}_i}p_i = 1 - p_i/p_{\hat{l}_i} \qquad (7\text{-}15)$$

where:

$$q_{\hat{l}_i\hat{l}_i}$$

is the i-th diagonal element of the cofactor matrix $Q_{\hat{l}}$, and its inverse the weight $p_{\hat{l}_i}$ of the adjusted observation $\hat{l}_i$. Since the a posteriori weight $p_{\hat{l}_i}$ of an observation is always greater than or equal to the a priori weight p, the redundancy contribution is restricted by Eq. (7-15) to:

$$0 \leq f_i \leq 1 \qquad (7\text{-}16)$$

The lower limit is attained when:

$$p_{\hat{l}_i} = p_i$$

In this case the corresponding l_i is not checked by other observations. Also, certain parameters of the model cannot be computed without l_i. The upper limit is reached when:

$$p_{\hat{l}_i} = \infty$$

This means that l_i is perfectly checked by the model and that the variance of the adjusted observation is zero.

Typical redundancy contributions in geodetic networks are:

Traverse networks:	f_i	= 0.1 – 0.2
Trilateration networks:	f_i	= 0.3 – 0.6
Combined networks:	f_i	= 0.5 – 0.8
Levelling networks:	f_i	= 0.2 – 0.5

Most authors agree that the minimum redundancy contribution for observations in geodetic networks should be around $f_i = 0.3$.

The close relationship of the measure f_i of internal reliability with Baarda's data snooping, as outlined in the previous Chapter, follows from Eq. (6-29), which can be rewritten as:

$$k_{oi} = \sqrt{\lambda_o / f_i} \qquad (7\text{-}17)$$

For the usual choice of $\alpha_o = 0.1\%$ and $\beta_o = 20\%$, Eq. (7-17) takes the form:

$$k_{oi} = 4.1/\sqrt{f_i}$$

In combination with the typical values of f_i given above this shows that:

$$5 \leq k_{oi} \leq 13$$

where $k_{oi}\sigma_i$ is the minimum value of a gross error which can just be detected with the probability of $1 - \beta_o$.

The average redundancy contributions:

$$\bar{f} = f/n$$

is a reasonable global measure of reliability, provided that the f_i are distributed evenly in the network. The measures f_i and $\bar{f}$ of internal reliability do not depend on the geodetic datum.

7.4 Global External Reliability

In spite of all the sophisticated testing procedures, there will never be certainty about having detected all gross and systematic errors of the GMM. Furthermore, some small errors must be expected which are just below the established boundary values. Therefore, one of the crucial points of model analysis is to gain full knowledge of the effect of these model errors on the parameter estimation.

A mathematical model is said to have a high external reliability, if it responds insignificantly to undetected errors.

The bias of the parameter vector caused by the error vector Δ follows from Eq. (6-8):

$$\hat{\bar{x}} - \hat{x} = \Delta\hat{x} = N^{-}A^{t}P\Delta \qquad (7\text{-}18)$$

where N^{-} is a generalized inverse of $N = A^{t}PA$.

Generally, when Δ is considered to be an unknown systematic error affecting all or some of the observations, Eq. (7-18) is not very informative. But it is very useful to trace the effect of certain biases, which can be modelled reasonably well as for example frequency errors or meteorological effects in EDM measurements, uncertain refractive indices for zenith angles or systematic errors in levelling.

Eq. (7-18) shows that $\Delta\hat{x}$ depends on the geodetic datum as does $\hat{x}$. An invariant average measure of the biasing effect of Δ can be derived from the quadratic form:

$$q_{\Delta\hat{x}} = \Delta\hat{x}^t Q_{\hat{x}}^{-} \Delta\hat{x} \tag{7-19}$$

which is datum independent and reasonably small for a reliable model. Expansion of Eq. (7-19) with $Q_{\hat{x}}^{-} = N = A^tPA$ yields:

$$\begin{aligned} q_{\Delta\hat{x}} &= \Delta^tPA(N^{-})^tNN^{-}A^tP\Delta \\ &= \Delta^tPAN^{-}A^tP\Delta = \Delta^tPQ_l\hat{}P\Delta \end{aligned} \tag{7-20}$$

Based on Eq. (7-3), the reliability criterion can be established from the inequality:

$$\Delta^t\Delta \cdot \Lambda_{max} \geq \Delta^tPQ_l\hat{}P\Delta, \qquad \forall\, \Delta \in \mathbb{R}^n \tag{7-21}$$

where Λ_{max} is the maximum eigenvalue of $PQ_l\hat{}P$. Thus, a smaller Λ_{max} yields a smaller maximal possible effect of an error Δ on the parameter estimation and, consequently, a more reliable (robust) model.

It should be noted that the interpretation of Λ_{max} of $PQ_l\hat{}P$ is as unclear as the interpretation of Λ_{max} of PQ_vP. The latter has been shown in Section 7.2.

Since $Q_l\hat{}P$ is idempotent its eigenvalues are either all zero or unity and, in the case of equally weighted observations, $PQ_l\hat{}P$ has r-fold eigenvalues p and $n - r$ = f-fold eigenvalues zero.

Additional observations of weight p do not change Λ_{max} or the trace of $PQ_l\hat{}P$; they only increase the number f of zero eigenvalues.

7.5 Local External Reliability

If the general Eq. (7-18) is specialized for the effect of a gross error in **one** observation according to Eq. (6-15) then:

$$\Delta\hat{x}_i = N^{-}A^tPe_i\Delta_i = N^{-}a_ip_i\Delta_i \tag{7-22}$$

where a_i is the i-th column of A^t. All parameters are biased by a single error Δ_i. In practice, the parameters appearing in the i-th observation equation with coefficient vector a_i are expected to be more biased than the other ones, thus the effect of **one** error can be regarded as local. But Eq. (7-22) is not readily interpreted, since $\Delta\hat{x}_i$ depends on the geodetic datum.

If, for lack of better measures, the quadratic form of $\Delta\hat{x}_i$ is used again, then the conclusion should be drawn that small values of:

$$q_{\Delta\hat{x}_i} = \Delta\hat{x}_i^t Q_{\hat{x}}^- \Delta\hat{x}_i = \Delta_i^2 p_i^2 a_i^t Q_{\hat{x}} a_i \qquad (7\text{-}23)$$

indicate a good local reliability.

Since Δ_i cannot be controlled, the requirement for reliability can be put in the form:

$$p_i^2 a_i^t Q_{\hat{x}} a_i = p_i(1 - f_i) = \min, \qquad \forall\, i \in \{1, 2, \ldots, n\} \qquad (7\text{-}24)$$

The criterion of Eq. (7-24) can be computed for all observables. It allows to check the reliability of the model in the design stage of the network, when improvements are still possible. An average measure of local reliability can be calculated from the sum of all individual values according to Eq. (7-24), yielding:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} p_i^2 a_i^t Q_{\hat{x}} a_i &= \frac{1}{n}\,\mathrm{tr}\,(PAN^- A^t P) \\ &= \frac{1}{n}\,\mathrm{tr}\,(PQ_l\hat{} P) \end{aligned} \qquad (7\text{-}25)$$

Equation (7-25) should be a minimum. It represents an alternative to Eq. (7-21) as a global measure.

7.6 Synopsis of Measures of Reliability

It is not surprising that the global measures of internal and external reliability as compiled in Table 7.1 are very similar. It follows from Eq. (7-14), after premultiplication with the weight matrix P, that:

$$PQ_v P = P - PQ_l\hat{} P \qquad (7\text{-}26)$$

For a given weight matrix P, the minimization of:

$$\mathrm{tr}\,(PQ_l\hat{} P)$$

is equivalent to a maximization of $\mathrm{tr}\,(PQ_vP)$. The same applies to all properly selected matrix norms and hence to the maximum eigenvalue.

	Internal	External
Global	$\Lambda_{max}(PQ_vP) = max$ $tr\,(PQ_vP) = max$	$\Lambda_{max}(PQ_{\hat{l}}P) = min$ $tr\,(PQ_{\hat{l}}P) = min$
Local, $\forall i$	$f_i = p_i q_{v_i v_i} = max$	$p_i^2 a_i^t Q_{\hat{x}} a_i = min$
Local average = global	$\bar{f} = f/n = max$	$\frac{1}{n} tr\,(PQ_{\hat{l}}P) = min$

Table 7.1: Useful measures of internal and external reliability in a GMM.

The result can be interpreted as follows: If the parameters cannot absorb the error vector the latter causes a large shift in the residuals. In this case the bias of the parameters is smaller and the probability of detecting the errors is great.

The derived measures of reliability are mainly used in pre-analyses of Gauss-Markov models for the purpose of optimization of the design of the model, particularly for geodetic control. Hence, only the relative magnitude is relevant. This avoids the need for an interpretation of the absolute values which is rather unclear, especially as far as the global measures are concerned.

It should be stressed that the concept of reliability, as outlined here, is based on Baarda's data snooping. A proper interpretation of the results has always to consider this context. If Pope's τ-method of outlier detection were chosen to establish a reliability concept, the results would be similar. As mentioned before, robust estimation methods provide an alternative to the concept of reliability. Properly selected robust estimators ignore small deviations from the assumed model and yield "reliable" results. The price for this property is a loss in efficiency of about 5-10% in comparison with least squares estimation if no deviations from the assumptions exist.

7.7 Example

The monitoring network as defined in Section 6.5 is used to illustrate the local measures of reliability. The a priori estimate of the standard deviation is $\sigma = 3^{cc}$ (= 0.0003 grad = 1.0") for the directions and $\sigma = 0.3$ mm for the distances. With $\sigma_o = 1^{cc}$ the weights become:

directions: $p = \frac{1}{9}$

distances: $p = \frac{1}{0.09} \left[\frac{cc}{mm}\right]^2$

No	from	to	f_i (7 - 15)			$p_i^2 a_i^t Q_{\hat{x}\hat{x}} a_i$ (7 - 24)		
			a	b	c	a	b	c
Directions						Dimension [mm/cc]^2		
1	1	2	0.569	0.551	0.548	0.038	0.037	0.038
2	1	13	0.616	0.594	0.583	0.032	0.033	0.034
3	1	14	0.782	-	-	0.014	-	-
4	1	3	0.822	0.761	0.757	0.010	0.014	0.015
5	1	9	0.791	-	-	0.013	-	-
6	1	4	0.774	0.726	0.715	0.015	0.018	0.019
7	1	6	0.586	0.579	0.560	0.036	0.034	0.036
8	1	7	0.124	0.130	0.125	0.087	0.084	0.085
9	1	10	0.002*	0.040	0.040	0.101*	0.094	0.094
10	1	11	0.190	0.185	0.168	0.080	0.078	0.080
11	1	12	0.535	0.526	0.519	0.041	0.040	0.041
12	2	1	0.406	0.417	0.413	0.056	0.055	0.055
13	2	10	0.091	0.026	0.026	0.091	0.098	0.098
14	2	11	0.190	0.185	0.168	0.080	0.080	0.082
15	2	12	0.299	0.294	0.290	0.068	0.068	0.069
16	2	13	0.420	0.419	0.409	0.054	0.054	0.056
17	2	14	0.734	0.722	0.719	0.019	0.021	0.021
18	2	9	0.753	0.746	0.745	0.017	0.018	0.018
19	2	3	0.740	0.733	0.732	0.019	0.019	0.020
20	2	4	0.748	0.747	0.737	0.018	0.018	0.019
21	2	6	0.482	0.495	0.480	0.047	0.046	0.048
22	2	7	0.374	0.378	0.360	0.059	0.059	0.061
23	3	1	0.752	0.728	0.724	0.017	0.019	0.020
24	3	10	0.755	-	-	0.017	-	-
25	3	2	0.727	0.683	0.679	0.020	0.024	0.024
26	3	11	0.428	0.418	0.404	0.053	0.053	0.055
27	3	12	0.247	0.246	0.240	0.073	0.073	0.073
28	3	13	0.198	0.193	0.169	0.079	0.078	0.081
29	3	14	0.159	0.161	0.161	0.083	0.082	0.082
30	3	9	0.130	0.145	0.145	0.086	0.084	0.084
31	3	4	0.376	0.387	0.381	0.059	0.057	0.058
32	3	6	0.462	0.465	0.446	0.050	0.048	0.050
33	3	7	0.677	0.664	0.657	0.026	0.026	0.027
34	4	1	0.814	0.818	0.811	0.011	0.010	0.011
35	4	10	0.820	0.814	0.812	0.010	0.011	0.011
36	4	2	0.827	0.827	0.821	0.009	0.009	0.010
37	4	11	0.687	0.687	0.679	0.025	0.025	0.026
38	4	3	0.638	0.646	0.634	0.030	0.029	0.030
39	4	12	0.594	0.590	0.575	0.035	0.035	0.037
40	4	13	0.341	0.332	0.290	0.063	0.064	0.069
41	4	9	0.000*	0.018	0.018	0.101*	0.099	0.099
42	4	14	0.001*	0.019	0.019	0.101*	0.099	0.099
43	4	6	0.610	0.629	0.579	0.033	0.031	0.037
44	4	7	0.749	0.751	0.740	0.018	0.017	0.019
45	6	1	0.590	0.700	0.696	0.023	0.019	0.020
46	6	7	0.329	0.348	0.345	0.052	0.058	0.059
47	6	4	0.447	0.605	0.574	0.039	0.030	0.033
48	6	3	0.528	0.647	0.604	0.030	0.025	0.030
49	6	10	0.577	0.676	0.674	0.025	0.022	0.022
50	6	14	-	0.723	0.723	-	0.017	0.017
51	6	9	-	0.701	0.701	-	0.019	0.019
52	6	2	-	0.616	0.584	-	0.029	0.032
Distances						Dimension [·]		
53	1	2	0.874	0.876	-	1.395	1.372	-
54	1	3	0.719	0.719	-	3.118	3.114	-
55	1	4	0.605	0.605	-	4.383	4.384	-
56	2	3	0.806	0.806	-	2.151	2.153	-
57	2	4	0.682	0.683	-	3.524	3.522	-
58	3	4	0.798	0.797	-	2.245	2.256	-

Table 7.2: Measures of local reliability for three versions of the monitoring network Montsalvens (Figure 7.2). The asterisks indicate the least reliable observations.

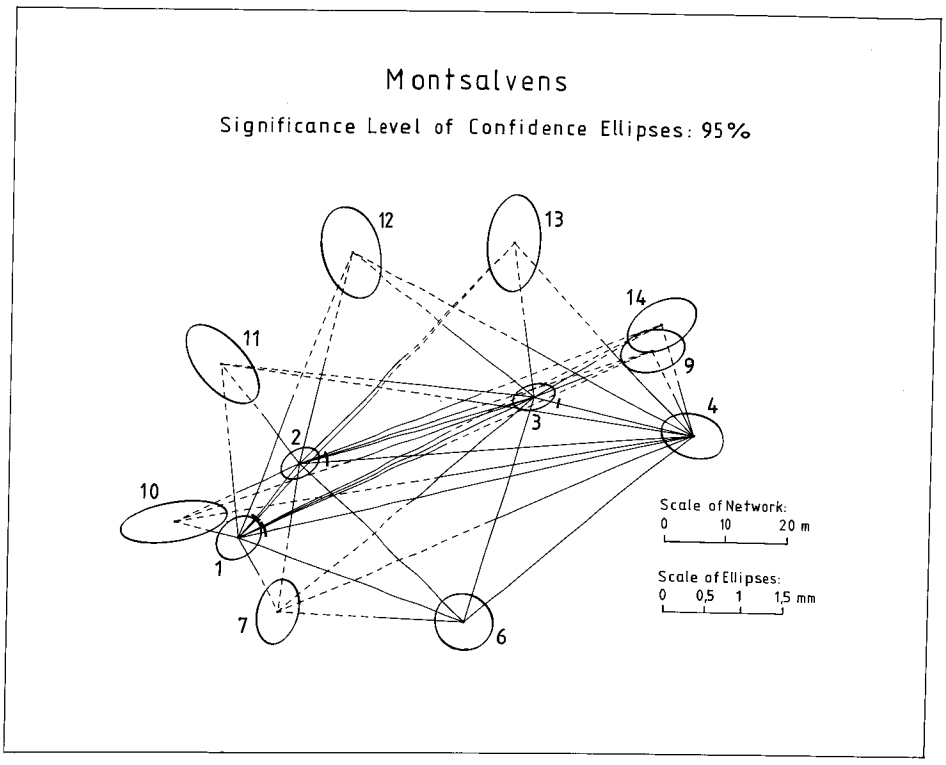


Figure 7.2: Monitoring network Montsalvens

Three different data sets have been generated. The corresponding redundancy contributions f_i as measures of internal reliability and the values of $p_i^2 a_i^t Q_{\hat{x}} a_i$ as measures of external reliability are listed in Table 7.2.

The data set of **version a** is identical with the set on which Table 6.2 is based. The observations 9, 41 and 42 have redundancy contributions close to zero. Hence, gross errors in these observations would be undetectable and would cause a maximum bias in the parameter vector as seen from $p_i^2 a_i^t Q_{\hat{x}} a_i$. A correct interpretation of the last numbers has to consider the dimensions. If for instance the influence of an error of magnitude one sigma is asked for, i.e. $\Delta = 3^{cc}$ or $\Delta = 0.3$ mm, then the values referring to the directions must be multiplied by $\Delta^2 = 9$ and the values referring to the distances by $\Delta^2 = 0.09$ to yield the quadratic form of Eq. (7-23) in mm^2 for both cases.

The data set of **version b** is an attempt to improve the case of version a without increasing the number of observations. Three new observations (50, 51 and 52) are introduced replacing the observations 3, 5 and 24. The selection of these observations is based on the geometry of the network shown in Fig. 7.2. The redundancy contributions of the observations 9, 41 and 42 very much exceed those obtained in version a but are still insufficient in terms of magnitude. This fact can only be remedied by a change of the geometry of the net.

The data set of **version c** has been derived from version b by omitting the distances. The change of the measures of reliability listed in Table 7.2 is only marginal, indicating that the distances contribute little to the reliability of the network.

The computation of the global measures of reliability of Table 7.1 does not make sense for this particular network because all observations have the same weight. This represents the trivial case mentioned in the previous sections where all eigenvalues are either p or zero.

8. ESTIMATION OF VARIANCE COMPONENTS

It has been stressed several times that the establishment of a proper GMM for the adjustment of observations is the primary goal of the first step of a deformation analysis, in order to avoid a misinterpretation of the imperfections of the model as deformations of the object. While the previous chapters are mainly concerned with the functional part of the GMM, this chapter focuses on the stochastic model. So far, it has been taken for granted, that the stochastic properties of the observations are exhaustively modelled by the variance-covariance matrix:

$$\Sigma = E([l - Ax][l - Ax]^t) = \sigma_o^2 Q \qquad (2\text{-}1)$$

The matrix Q is a known coefficient matrix and σ_o^2 the a priori variance factor, which is checked and, usually, replaced by the estimate of the a posteriori variance factor.

$$s_o^2 = v^t P v / f \qquad (2\text{-}7)$$

This is a simplification which is unlikely to reflect reality.

Consider, for example, a modern geodetic network with distances and angles measured with different instruments under different conditions by different surveyors. Who is able to write the proper variances and covariances of this inhomogeneous batch of observations into the matrix Σ of Eq. (2-1)? Or consider a combined adjustment of first, second and third order levelling. Are the relative accuracies known well enough to guarantee a realistic stochastic model? Consider finally a multi-epoch adjustment for a deformation analysis. In the course of time new instruments with higher precision are employed, new measurement methods are applied. How shall one know the proper variances for the selection of the coefficients of Σ?

The first proposals to replace the simple model of Eq. (2-1) by a more flexible and sophisticated one were made by F.R. HELMERT (1924). His method was not adopted at that time since the equations are rather involved and the computational task is considerable. Unaware of Helmert's work, C.A. RAO (1970) developed a new stochastic model which proved to be identical to Helmert's model, if normal distribution is assumed. In the following years a flood of papers appeared on this topic, authored by geodesists and statisticians. Today, improved and more flexible methods of dealing with the stochastic model are available. Modern computers and efficient approximation methods enable the geodetic community to readily adopt the method of estimation of variance components. It seems that this breakthrough has not occurred yet.

8.1 The Stochastic Model

The Gauss-Markov model as introduced in Eq. (2-1) is expanded into the form:

$$E(l) = Ax, \qquad l = Ax + \varepsilon$$

$$E(\varepsilon\varepsilon^t) = \Sigma = \sum_{i=1}^{k} \sigma_{oi}^2 Q_i \qquad (8\text{-}1)$$

where the covariance matrix Σ is considered as a sum of k variance components σ_{oi}^2, each multiplied by the respective cofactor matrix Q_i. The problem is how to estimate the parameters x as well as the k variance components σ_{oi}^2. This is a real generalization of the classical GMM of Eq. (2-1), which is the special case for k = 1. The definition of Σ in Eq. (8-1) can be established in the following way: The vector ε of true errors is a sum of linear functions of various small independent primary errors ε_i. Each ε_i belongs to a certain error source and is associated with a known coefficient matrix G_i. This leads to the model:

$$\varepsilon = \sum_{i=1}^{k} G_i \varepsilon_1 \tag{8-2}$$

From Eq. (8.1) follows:

$$\Sigma = E(\varepsilon\varepsilon^t) = \sum_{i=1}^{k} G_i E(\varepsilon_i \varepsilon_i^t) G_i^t \tag{8-3}$$

which can be simplified under the assumptions:

$$\begin{aligned} E(\varepsilon_i \varepsilon_i^t) &= \sigma_{oi}^2 I, & E(\varepsilon_i \varepsilon_j^t) &= 0 \text{ for } i \neq j \\ G_i G_i^t &= Q_i, & E(\varepsilon_i) &= 0 \end{aligned} \tag{8-4}$$

to yield the final form of the stochastic model:

$$\Sigma = \sigma_{01}^2 Q_1 + \sigma_{02}^2 Q_2 + \ldots + \sigma_{ok}^2 Q_k \tag{8-5}$$

8.2 Examples of Variance Component Models

To illustrate the practical relevance of the expanded stochastic model, two geodetic examples are given:

i. Geometric Levelling

Three main error sources contribute to the total error: a reading error not depending on distance, a distance dependent error term, and a scale error, proportional to the levelled height difference. The j-th component of the error vector ε takes the form:

$$\varepsilon_j = \varepsilon_{1j} + \sqrt{D_j}\, \varepsilon_{2j} + \Delta H_j \varepsilon_{3j}$$

and hence:

$$\varepsilon = G_1\varepsilon_1 + G_2\varepsilon_2 + G_3\varepsilon_3$$

with coefficient matrices according to Eq. (8-2):

$$G_1 = \begin{pmatrix} 1 & & \\ & 1 & \\ & \ddots & \\ & & 1 \end{pmatrix}$$

$$G_2 = \begin{pmatrix} \sqrt{D_1} & & & \\ & \sqrt{D_2} & & \\ & & \ddots & \\ & & & \sqrt{D_n} \end{pmatrix}$$

$$G_3 = \begin{pmatrix} \Delta H_1 & & & \\ & \Delta H_2 & & \\ & & \ddots & \\ & & & \Delta H_n \end{pmatrix}$$

From Eq. (8-4) follows:

$$Q_1 = G_1 G_1^t = I$$

$$Q_2 = G_2 G_2^t = \text{diag}\,(D_1, D_2, ..., D_n)$$

$$Q_3 = G_3 G_3^t = \text{diag}\,(\Delta H_1^2, \Delta H_2^2, ..., \Delta H_n^2)$$

and yields the covariance matrix:

$$\Sigma = \sigma_{01}^2 Q_1 + \sigma_{02}^2 Q_2 + \sigma_{03}^2 Q_3$$

comprising three unknown variance components to be estimated.

ii. Three dimensional network with heterogeneous observables. The coordinates of the network stations shall be estimated from a GMM comprising four different types of observables, namely:

- p - direction observations
- q - distance measurements
- r - zenith angles
- s - GPS vectors (differential positioning)

Each group of observations has its own error characteristic and has to be modelled separately. The n x n - covariance matrix, n = p + q + r + s, can be written according to Eq. (8-5) as:

$$\Sigma = \sigma_{op}^2 Q_p + \sigma_{oq}^2 Q_q + \sigma_{or}^2 Q_r + \sigma_{os}^2 Q_s$$

If the observations are grouped in the order given above, then the cofactor matrices take the form:

$$Q_p = \begin{pmatrix} \tilde{Q}p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q_q = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \tilde{Q}_q & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$Q_r = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{Q}_r & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad Q_s = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{Q}_s \end{pmatrix}$$

where the coefficient matrices $\tilde{Q}_p$, $\tilde{Q}_q$, $\tilde{Q}_r$ and $\tilde{Q}_s$ are given by the stochastic behaviour of the respective observations.

Example (i) is a model for additive effects while the second example (ii) is a model of group variances. Models combining additive and group effects are possible as well.

It is important to note that the estimation procedure is only successful if the GMM allows the separation of the variance components. This requires in the case (i) that the effects are far from being colinear, and in the case (ii) that each group has sufficient redundancy and is strongly correlated with the other groups by virtue of the functional model.

8.3 Three Lemmas on Traces and Quadratic Forms

The a posteriori variance factor of the GMM is estimated according to Eq. (6-3) from the quadratic form:

$$q = v^tPv = l^tPQ_vPl = l^tBl \qquad (8\text{-}6)$$

with:

$$B = P - PAN^-A^tP, \qquad \text{where: } N = A^tPA \qquad (8\text{-}7)$$

where N^- is a g-inverse of N, as defined in Eq. (3-74).

The estimators of the k variance components (Eq. (8-5)) must fulfil certain optimality criteria and shall be quadratic functions of the observations. This is necessary since the variance is defined as the expectation of the square of the errors, namely $E(\varepsilon_i^2) = \sigma_i^2$.

To facilitate the derivations below, this section is devoted to three pertinent lemmas of matrix algebra.

Lemma 1: Let l be a random vector of n elements with $E(l) = Ax$ and $E([l - Ax][l - Ax]^t) = \Sigma$ and let B be any real positive semidefinite n x n - matrix. Then the quadratic form $q = l^tBl$ has the expectation:

$$E(q) = \operatorname{tr} B\Sigma + x^tA^tBAx \qquad (8\text{-}8)$$

Proof: Let $l - Ax = \varepsilon$, then $\varepsilon\varepsilon^t = Axx^tA^t - 2Axl^t + ll^t$ and

$$E(\varepsilon\varepsilon^t) = Axx^tA^t - 2AxE(l^t) + E(ll^t) = E(ll^t) - Axx^tA^t = \Sigma.$$

On the other hand (refer to Eq. (3-44)):

$$E(l^tBl) = E(\operatorname{tr} l^tBl) = E(\operatorname{tr} Bll^t) = \operatorname{tr} BE(ll^t)$$

with $E(ll^t) = \Sigma + Axx^tA^t$ provides the proof:

$$E(l^tBl) = \operatorname{tr} B(\Sigma + Axx^tA^t) = \operatorname{tr} B\Sigma + \operatorname{tr} BAxx^tA^t$$

$$E(l^tBl) = \operatorname{tr} B\Sigma + x^tA^tBAx$$

Lemma 1 holds for all random vectors since no distributional assumptions have been made.

Lemma 2: Let l be a normally distributed random vector with $E(l) = Ax$ and $\Sigma = E([Ax - l][Ax - l]^t)$, i.e. $l \sim N(Ax, \Sigma)$.

Further, let B be any real positive semidefinite matrix of order n x n and q the quadratic form $q = l^tBl$. The variance of q is then given by:

$$\sigma_q^2 = 2\text{tr } B\Sigma B\Sigma + 4x^tA^tB\Sigma BAx \qquad (8\text{-}9)$$

Proof: $\sigma_q^2 = E(q - E(q))^2$ by definition of the variance. An expansion using Lemma 1 yields:

$$\sigma_q^2 = E(q - \text{tr } B\Sigma - x^tA^tBAx)^2$$

$$= E(q - \text{tr } B\Sigma - (l - \varepsilon)^tB(l - \varepsilon))^2$$

$$= E(q - \text{tr } B\Sigma - l^tBl - \varepsilon^tB\varepsilon + 2\varepsilon^tBl)^2$$

which becomes after substitution of $l = Ax + \varepsilon$:

$$\sigma_q^2 = E(\varepsilon^tB\varepsilon - \text{tr } B\Sigma + 2\varepsilon^tBAx)^2$$

$$= E(\varepsilon^tB\varepsilon\varepsilon^tB\varepsilon) + (\text{tr } B\Sigma)^2 + 4E(\varepsilon^tBAx\varepsilon^tBAx) - 2\text{tr } B\Sigma E(\varepsilon^tB\varepsilon) + 4x^tA^tBE(\varepsilon\varepsilon^tB\varepsilon) - 4\text{tr } B\Sigma x^tA^tBE(\varepsilon)$$

For normally distributed random errors the following equations are easily proven by integration of the respective expectation functions:

$$E(\varepsilon_i) = 0, \qquad E(\varepsilon_i^3) = 0, \qquad E(\varepsilon_i\varepsilon_j\varepsilon_k) = 0$$

Thus the right hand side of σ_q^2 can be simplified to:

$$\sigma_q^2 = E(\varepsilon^tB\varepsilon\varepsilon^tB\varepsilon) + (\text{tr } B\Sigma)^2 + 4E(\varepsilon^tBAx\varepsilon^tBAx) - 2\text{tr } B\Sigma E(\varepsilon^tB\varepsilon)$$

Applying the expectation operator the right hand side terms take the forms:

$$E(\varepsilon^tB\varepsilon\varepsilon^tB\varepsilon) = E(\text{tr } B\varepsilon\varepsilon^tB\varepsilon\varepsilon^t) = (\text{tr } B\Sigma)^2 + 2\text{tr } B\Sigma B\Sigma$$

$$E(\varepsilon^tBAx\varepsilon^tBAx) = x^tA^tBE(\varepsilon\varepsilon^t)BAx = x^tA^tB\Sigma BAx$$

$$\text{tr } B\Sigma E(\varepsilon^tB\varepsilon) = (\text{tr } B\Sigma)^2$$

and lead to the final result:

$$\sigma_q^2 = 2\text{tr } B\Sigma B\Sigma + 4x^tA^tB\Sigma BAx$$

Lemma 3: Let A, B and C be n x n - matrices and let f(A, B, C) be a scalar function of these matrices. Further, the matrix of partial derivatives of f with respect to the elements of A be denoted by $\partial/\partial A\, f(A, B, C)$. If f is the trace of the products of the matrices, then the following equations apply:

$$\frac{\partial}{\partial A} \operatorname{tr} AB = B^t \tag{8-10}$$

$$\frac{\partial}{\partial A} \operatorname{tr} ABAC = (BAC + CAB)^t \tag{8-11}$$

$$\frac{\partial}{\partial A} \operatorname{tr} ABA^tC = C^tAB^t + CAB \tag{8-12}$$

The proof is straightforward and requires simple calculations only.

8.4 Invariant Quadratic Unbiased Estimation (IQUE)

Let σ be a vector comprising the k variance components σ_{oi}^2:

$$\sigma = (\sigma_{01}^2, \sigma_{02}^2, \ldots, \sigma_{ok}^2)^t$$

and let p be a vector of k coefficients. The general procedure is to estimate the parameter vector x and, additionally, the linear function $p^t\sigma$ of the variance components of the expanded GMM of Eq. (8-1).

As established in Section 8.3, the required estimator must be a quadratic function of the observations, hence the name quadratic estimator.

The criterion of unbiasedness implies the existence of a real symmetrical matrix B which yields:

$$E(l^tBl) = p^t\sigma$$

Substitution of the expectation of the quadratic form, as given in Lemma 1, results in:

$$E(l^tBl) = \operatorname{tr} B\Sigma + x^tA^tBAx = p^t\sigma \tag{8-13}$$

Expansion of Σ according to Eq. (8-5) leads to:

$$\operatorname{tr} B\sum_{i=1}^{k} \sigma_{oi}^2 Q_i + x^tA^tBAx = p^t\sigma$$

$$\sum_{i=1}^{k} \operatorname{tr} BQ_i\sigma_{oi}^2 + x^tA^tBAx = p^t\sigma$$

Thus, the criterion of unbiasedness requires:

$$\begin{aligned} &\operatorname{tr} BQ_i = p_i \quad \text{for} \quad i \in \{1, 2, \ldots, k\} \\ &x^tA^tBAx = 0 \quad \Rightarrow \quad A^tBA = 0 \end{aligned} \tag{8-14}$$

The linear function $p^t\sigma$ is invariantly estimable, if a matrix B exists, and provides a quadratic estimator which is independent (invariant) with respect to the parameter vector x. This requirement translates into:

$$l^tBl = (l - Ax)^tB(l - Ax) = \varepsilon^tB\varepsilon \qquad \forall\, x \in \mathbb{R}^u \tag{8-15}$$

and leads to the condition:

$$A^tB = BA = 0 \tag{8-16}$$

Since Eq. (8-16) implies $A^tBA = 0$, the results of Section 8.4 can be summarized:

$$\begin{aligned} &\text{For B real symmetrical, } l^tBl \text{ is IQUE} \\ &\text{of } p^t\sigma \text{ if } A^tB = 0 \text{ and } \operatorname{tr} BQ_i = p_i \end{aligned} \tag{8-17}$$

To illustrate the contents of this section, it shall be verified that the usual a posteriori variance factor s_o^2 is an IQUE (invariant quadratic unbiased estimator) of σ_o^2.

From Eq. (2-7) follows:

$$s_o^2 = v^tPv/f, \qquad f = n - r(A)$$

Substituting $\hat{x} = N^-A^tPl$ into $v = A\hat{x} - l$ yields $v = (AN^-A^tP - I)l$. Thus, the quadratic form can be expressed as:

$$\begin{aligned} v^tPv &= l^t(PAN^-A^t - I)P(AN^-A^tP - I)l \\ &= l^t(P - PAN^-A^tP)l = l^tBl \end{aligned}$$

The matrix $B = P - PAN^-A^tP$ is symmetrical, $A^tB = 0$ and $\operatorname{tr} BQ = \operatorname{tr}(I - PAN^-A^t) = \operatorname{tr} I - \operatorname{tr}(PAN^-A^t)$. Since PAN^-A^t is idempotent, it follows that $\operatorname{tr}(PAN^-A^t) = r(PAN^-A^t) = r(A)$, hence $\operatorname{tr} BQ = n - r(A) = f$. Finally $v^tPv = fs_o^2$ is an IQUE of $p\sigma_o^2$ with $p = n - r(A)$.

8.5 Best Invariant Quadratic Unbiased Estimation (BIQUE)

Since the IQUE of $p^t\sigma$ of Section 8.4 is not unique, an additional criterion is required to define a unique element of the set of IQUEs. For this purpose the following minimum variance criterion is introduced:

$$\sigma^2(l^tBl) = E(l^tBl - p^t\sigma)^2 = \min \tag{8-18}$$

for l^tBl being IQUE according to Eq. (8-17).

In agreement with Eq. (8-9) of Lemma 2, the variance is:

$$\sigma^2(l^tBl) = 2\operatorname{tr} B\Sigma B\Sigma + 4x^tA^tB\Sigma BAx \tag{8-9}$$

Since the minimum variance (best) estimator is restricted to the class of IQUEs, Eq. (8-16) must hold. This reduces the minimum problem to:

$$\operatorname{tr} B\Sigma B\Sigma \quad \Rightarrow \quad \min \tag{8-19}$$

with the boundary conditions $BA = 0$ and $\operatorname{tr} BQ_i = p_i$. Hence the Lagrange function of the problem yields:

$$L = \text{tr}\, B\Sigma B\Sigma - 2\text{tr}\, BA\Lambda^t - 2\sum_{i=1}^{k} \lambda_i(\text{tr}\, BQ_i - p_i) \tag{8-20}$$

where Λ is an n x u - matrix and $\lambda = (\lambda_1, \lambda_2, ..., \lambda_k)t$ a k-vector of Lagrangean multipliers. The solution is found by equating the partial derivatives of L with respect to B, Λ and λ to zero. The use of Lemma 3 yields:

$$\frac{\partial L}{\partial B} = 2\Sigma B\Sigma - 2\Lambda A^t - 2\sum_{i=1}^{k} \lambda_i Q_i \tag{8-21}$$

and leads to the set of equations:

$$\begin{aligned} &\Sigma B\Sigma - \Lambda A^t - \sum_{i=1}^{k} \lambda_i Q_i = 0 \\ &A^t B = 0, \qquad \text{tr}\, BQ_i - p_i = 0 \end{aligned} \tag{8-22}$$

An auxiliary matrix R is introduced to solve Eqs (8-22) for B:

$$R = \Sigma^{-1} - \Sigma^{-1} AN\text{–}At\Sigma^{-1} \tag{8-23}$$

and is selected to meet the conditions:

$$R\Sigma B\Sigma R = B \quad \text{and} \quad AtR = 0$$

Pre- and postmultiplication of the first equation of Eq. (8-22) with R yields:

$$B = \sum_{i=1}^{k} \lambda_i RQ_i R \tag{8-24}$$

and subsequent postmultiplication by Q_j results in:

$$BQ_j = \sum_{i=1}^{k} \lambda_i RQ_i RQ_j \tag{8-25}$$

Application of the trace operator and comparison with Eq. (8-14) leads to:

$$\text{tr}\, BQ_j = \sum_{i=1}^{k} \lambda_i \text{tr}\, RQ_i RQ_j = p_j \tag{8-26}$$

Let $h_{ij} = \text{tr}\, RQ_i RQ_j$, $i, j \in \{1, 2, ..., k\}$, then a k x k - matrix H can be composed of the elements h_{ij}, and enables Eq. (8-26) to be written in the concise form:

$$H\lambda = p \tag{8-27}$$

with the k-vector λ and p defined as before. Since the matrix H may be of less than full rank, a generalized inverse (Eq. (3-74)) is used to solve Eq. (8-27) for λ, the vector of Lagrangean multipliers:

$$\lambda = H^{-}p = Gp, \qquad G = H^{-} \tag{8-28}$$

and

$$\lambda_i = g_i^t p = p^t g_i$$

where g_i^t is the i-th row of the matrix $G = H^{-}$. Substitution into Eq. (8-24) yields:

$$B = \sum_{i=1}^{k} p^t g_i R Q_i R \tag{8-29}$$

and results in the estimator for $p^t\sigma$:

$$l^t B l = \sum_{i=1}^{k} p^t g_i l^t R Q_i R l = p^t \hat{\sigma} \tag{8-30}$$

This equation can be put in a more suitable form, if $f_i = l^t R Q_i R l$ is defined as a vector $f = (f_1, f_2, ..., f_k)^t$:

$$l^t B l = p^t G f = p^t \hat{\sigma}$$

The final result is obtained as:

$$\hat{\sigma}_{oi}^2 = g_i^t f \qquad i \in \{1, 2, ..., k\} \tag{8-31}$$

with $p = e_i$ being the i-th unit vector.

An iterative numerical procedure must be employed because the matrix R, defined in Eq. (8-23) and required for the estimator of Eq. (8-30), contains the true covariance matrix Σ, i.e. the vector σ of covariance components to be estimated. Starting with a reasonable vector of a priori variance components a first estimation is carried out. The resulting variances σ_{oi}^2 are multiplied with the corresponding cofactor matrices Q_i leading to improved covariance matrices. These serve as input for the next iteration. This process is repeated until σ converges towards the vector:

$$\hat{\sigma} = (1, 1, ..., 1)^t$$

More computational details follow in Section 8.6.

Certain difficulties arise whenever r(H) < k. In such cases only r(H) components of σ are estimable. If according to Eq. (8-31) $p_i = e_i$ is chosen, where e_i is the i-th unit vector, then:

$$e_i^t \hat{\sigma} = \hat{\sigma}_{oi}^2$$

is only BIQUE if $e_i \in S(H)$.

The BIQUE $l^t B l$ is based on the real symmetrical matrix B. No constraints have been imposed in relation to the positive definiteness of B. Therefore, B may not be positive definite occasionally, and may lead, possibly, to negative variance components. Such results are useless. An improvement of the GMM through the addition of observations, reduction of the number of parameters and/or variance components can remedy the shortcoming.

Despite these problems, the method outlined is a very valuable tool for the design of a realistic mathematical model, particularly for inhomogeneous geodetic observations. It requires a considerable amount of computations but yields reasonable estimates in all fairly well planned networks which feature a sufficient number of redundancies. In case of normally distributed observations the method is equivalent to Helmert's estimator. A generalization to the estimation of covariance components is easily possible, but probably not relevant to geodesy.

8.6 Numerical Operations

This section summarizes the equations for the iterative numerical computation. The equations are arranged in a form to provide a computational recipe.

Mathematical model:

$$l = Ax + \varepsilon = Ax + G_1\varepsilon_1 + G_2\varepsilon_2 + \ldots + G_k\varepsilon_k$$

$$E(\varepsilon\varepsilon^t) = \Sigma = \sigma_{01}^2 Q_1 + \sigma_{02}^2 Q_2 + \ldots + \sigma_{ok}^2 Q_k \qquad \text{(8-1) - (8-5)}$$

i. Compute the initial cofactor matrices:

$$Q_1^0 = \sigma_{01}^2 Q_1, \quad Q_2^0 = \sigma_{02}^2 Q_2, \quad \ldots, \quad Q_k^0 = \sigma_{ok}^2 Q_k$$

ii. For $\nu = 0, 1, 2, \ldots$, compute Σ from Eq. (8-5) and R from Eq. (8-23):

$$\Sigma_\nu = \sum_{i=1}^{k} Q_i^\nu$$

$$R_\nu = \Sigma_\nu^{-1} - \Sigma_\nu^{-1} A (A^t \Sigma_\nu^{-1} A)^- A^t \Sigma_\nu^{-1}$$

iii. Form H and f on the basis of Eqs (8-27) and (8-30):

$$H_\nu = (h_{ij}^\nu), \qquad h_{ij}^\nu = \operatorname{tr} R_\nu Q_i^\nu R_\nu Q_j^\nu$$

$$f_\nu = (f_1^\nu, f_2^\nu, \ldots, f_k^\nu)^t, \qquad f_i^\nu = l^t R_\nu Q_i^\nu R_\nu l$$

iv. Compute G from Eq. (8-28) and σ from Eq. (8-31) with the (generalized) inverse $H_\nu^{-1}(H_\nu^-)$ $\Rightarrow G_\nu$:

$$(\hat{\sigma}_{01}^2)^{\nu+1} = e_1^t G_\nu f_\nu, \qquad (\hat{\sigma}_{02}^2)^{\nu+1} = e_2^t G_\nu f_\nu, \qquad \ldots$$

$$\hat{\sigma}_{\nu+1} = ((\hat{\sigma}_{01}^2)^{\nu+1}, \quad (\hat{\sigma}_{02}^2)^{\nu+1}, \quad \ldots, \quad (\hat{\sigma}_{ok}^2)^{\nu+1})^t$$

v. If $\hat{\sigma}_{\nu+1} \neq (1, 1, \ldots, 1)^t$ compute new cofactor matrices:

$$Q_1^{\nu+1} = (\hat{\sigma}_{01}^2)^{\nu+1} Q_1^\nu, \qquad Q_2^{\nu+1} = (\hat{\sigma}_{02}^2)^{\nu+1} Q_2^\nu, \qquad \ldots$$

Return to (ii).

If $\hat{\sigma}_{v+1} = (1, 1, ..., 1)^t$ compute the final variance components:

$$\hat{\sigma}_{01}^2 = \prod_{\alpha=0}^{m} (\hat{\sigma}_{01}^2)_\alpha, \quad \hat{\sigma}_{02}^2 = \prod_{\alpha=0}^{m} (\hat{\sigma}_{02}^2)_\alpha, \quad ...; \quad m = v+1$$

$$\hat{\sigma} = s = (\hat{\sigma}_{01}^2, \hat{\sigma}_{02}^2, ..., \hat{\sigma}_{ok}^2)^t$$

The number m of iterations required strongly depends on the initial values of σ and upon the strength of the functional relations of the elements of the GMM. For large models up to 20-100 iterations may be required which represents a considerable computational effort.

8.7 Simplified Estimators

Because of the difficult mathematical background and the computational effort already required for moderate sized models, approximate methods have been developed. They yield the same numerical results whilst using fewer mathematical operations, provided the BIQUE exists as defined in the previous sections.

The estimator as derived in this section only applies to group effects as outlined in example (ii) of Section 8.2. Since this type of problem is the most frequent one in geodesy, the results are of practical importance.

The derivation begins with the special GMM:

$$l = Ax + \varepsilon, \qquad \Sigma = \sigma_o^2 I \tag{8-32}$$

where all observations are independent and of equal variance. This does not restrict the generality since any general GMM (2-1) can be transformed into the form of Eq. (8-32) according to Eqs (3-93) and (3-94). In the model of Eq. (8-32) a unique BIQUE for σ_o^2 exists, namely:

$$s_o^2 = v^t v/(n - r(A)) \tag{8-33}$$

If Eq. (8-32) represents the model of a geodetic network with inhomogeneous observables comprising k groups, each with a particular variance factor σ_{oi}^2, then the vectors l and v can be partitioned as follows:

$$l^t = (l_1^t, \quad l_2^t, \quad ..., \quad l_k^t)$$

$$v^t = (v_1^t, \quad v_2^t, \quad ..., \quad v_k^t)$$

$$\sigma = (\sigma_{01}^2, \quad \sigma_{02}^2, \quad ..., \quad \sigma_{ok}^2)$$

The quadratic form of Eq. (8-33) can be written as:

$$q = v^t v = v_1^t v_1 + v_2^t v_2 + ... + v_k^t v_k \tag{8-34}$$

where each term on the right hand side refers to one group of observations.

In order to establish the relationship between the subforms $v_i^t v_i$ and the variance factors σ_{oi}^2 the mathematical expectations of the subforms are required. From $v = A\hat{x} - l$ and $\varepsilon = l - Ax$ follows:

$$v + \varepsilon = A(\hat{x} - x) \tag{8-35}$$

Premultiplication with A^t and $(A^tA)^- = N^-$ yields:

$$A^t\varepsilon = A^tA(\hat{x} - x)$$

$$\hat{x} - x = N^-A^t\varepsilon \tag{8-36}$$

where $A^tv = 0$ and $N^-N(\hat{x} - x) = (\hat{x} - x)$ have been used. Substitution of Eq. (8-36) in Eq. (8-35) results in the relation between the residual vector v and the vector of true errors ε:

$$v = AN^-A^t\varepsilon - \varepsilon \tag{8-37}$$

The subvector v_i is given by:

$$v_i = A_iN^-A^t\varepsilon - \varepsilon_i \tag{8-38}$$

Considering $N_i = A_i^tA_i$ the above equation yields the subform:

$$v_i^tv_i = \varepsilon_i^t\varepsilon_i - 2\varepsilon_i^tA_iN^-A^t\varepsilon + \varepsilon^tAN^-N_iN^-A^t\varepsilon$$

The subform has the expectation:

$$E(v_i^tv_i) = n_i\sigma_{oi}^2 - 2E(\varepsilon_i^tA_iN^-A^t\varepsilon) + E(\varepsilon^tAN^-N_iN^-A^t\varepsilon) \tag{8-39}$$

which can be simplified if $E(\varepsilon_i) = 0$, $E(\varepsilon_i^t\varepsilon_i) = n_i\sigma_{oi}^2$ and $E(\varepsilon_i^t\varepsilon_j) = 0$ for $i \neq j$ is considered.

The second term on the right hand side of Eq. (8-39) yields after rearrangement:

$$2E(\varepsilon_i^tA_iN^-A^t\varepsilon) = 2\text{tr } A_iN^-A^tE(\varepsilon\varepsilon_i^t)$$

$$= 2\sigma_{oi}^2\text{tr } A_iN^-A_i^t = 2\sigma_{oi}^2\text{tr } N^-N_i$$

and similarly the third term:

$$E(\varepsilon^tAN^-N_iN^-A^t\varepsilon) = \text{tr } AN^-N_iN^-A^tE(\varepsilon\varepsilon^t)$$

$$= \sum_{j=1}^{k}\sigma_{oj}^2\text{tr } N^-N_iN^-N_j$$

Substitution in Eq. (8-39) results in:

$$E(v_i^tv_i) = n_i\sigma_{oi}^2 - 2\sigma_{oi}^2\text{tr } N^-N_i + \sum_{j=1}^{k}\sigma_{oj}^2\text{tr } N^-N_iN^-N_j \tag{8-40}$$

The computations are iterative, and the estimates of each step are used to re-transform the model to get the special form of Eq. (8-32). The iteration continues until the estimates of all variance components are unity, i.e. $\sigma_{oi}^2 = 1 \ \forall$ i. This justifies to set $\sigma_{oj}^2 = \sigma_{oi}^2$ in Eq. (8-40), so that the third term on the right hand side can be replaced by:

$$\sum_{j=1}^{k} \sigma_{oj}^2 \mathrm{tr}\, N^{-}N_i N^{-}N_j = \sigma_{oi}^2 \mathrm{tr}\, N^{-}N_i N^{-} \sum_{j=1}^{k} N_j$$

$$= \sigma_{oi}^2 \mathrm{tr}\, N^{-}N_i N^{-}N = \sigma_{oi}^2 \mathrm{tr}\, N^{-}N_i$$

This leads to the following expression for the expectation of $v_i^t v_i$:

$$E(v_i^t v_i) = n_i \sigma_{oi}^2 - \sigma_{oi}^2 \mathrm{tr}\, N^{-}N_i$$

$$= \sigma_{oi}^2 (n_i - \mathrm{tr}\, A_i N^{-} A_i^t) \qquad (8\text{-}41)$$

From:

$$l_i + v_i = \hat{l}_i = A_i \hat{x}$$

follows that the matrix product under the trace operator of Eq. (8-41) is nothing else than the cofactor matrix of the group i of the adjusted observations, $Q_{\hat{l}_i}$.

Since, on the other hand, n_i equals the trace of a $n_i \times n_i$ - unit matrix I_{n_i}, Eq. (8-41) can be rewritten as:

$$E(v_i^t v_i) = \sigma_{oi}^2 \mathrm{tr}\, (I_{n_i} - Q_{\hat{l}_i}) \qquad (8\text{-}42)$$

Comparison with Eq. (7-14) and considering that $P = I$ applies in the applied special model according to Eq. (8-32), enables the bracket under the trace to be replaced by:

$$I_{n_i} - Q_{\hat{l}_i} = I_{n_i} - Q_{\hat{l}_i} P_i = Q_{v_i} P_i$$

This leads to:

$$E(v_i^t v_i) = \sigma_{oi}^2 \mathrm{tr}\, (Q_{v_i} P_i) \qquad (8\text{-}43)$$

and results in the final expression after use of Eq. (7-13):

$$E(v_i^t v_i) = \sigma_{oi}^2 f_i \qquad (8\text{-}44)$$

The redundancy contribution of the i-th group of observations in the sense of Section 7.3 is denoted by f_i. The iteration scheme follows easily:

$$(s_{oi}^2)_\nu = \left(\frac{v_i^t v_i}{f_i}\right)_{\nu-1} \qquad i = 1, 2, ..., k \qquad (8\text{-}45)$$

$$(l_i + v_i = A_i\hat{x})_\nu = \frac{1}{s_{oi}} (l_i + v_i = A_i\hat{x})_{\nu-1}$$

$$\nu = 1, 2, \ldots \text{ until } (s_{oi}^2)_\nu = 1 \qquad \forall i$$

$$s_{oi}^2 = (s_{oi}^2)_1 \cdot (s_{oi}^2)_2 \cdot (s_{oi}^2)_3 \quad \ldots$$

Despite the simplicity of Eq. (8-45) the computing effort is usually considerable but less than in the case of rigorous estimators of Section 8.5.

8.8 Example

The monitoring network Montsalvens which comprises 49 directions and 6 distances is used to demonstrate the efficiency of the simplified method of Section 8.7. The data set has been generated on the computer. All observations are normally distributed. The population variance of the directions is $\sigma_{01}^2 = (3^{cc})^2$ and of the distances $\sigma_{02}^2 = (0.3mm)^2$. ($3^{cc} = 3.10^{-4}$ grads = 1 second of arc.)

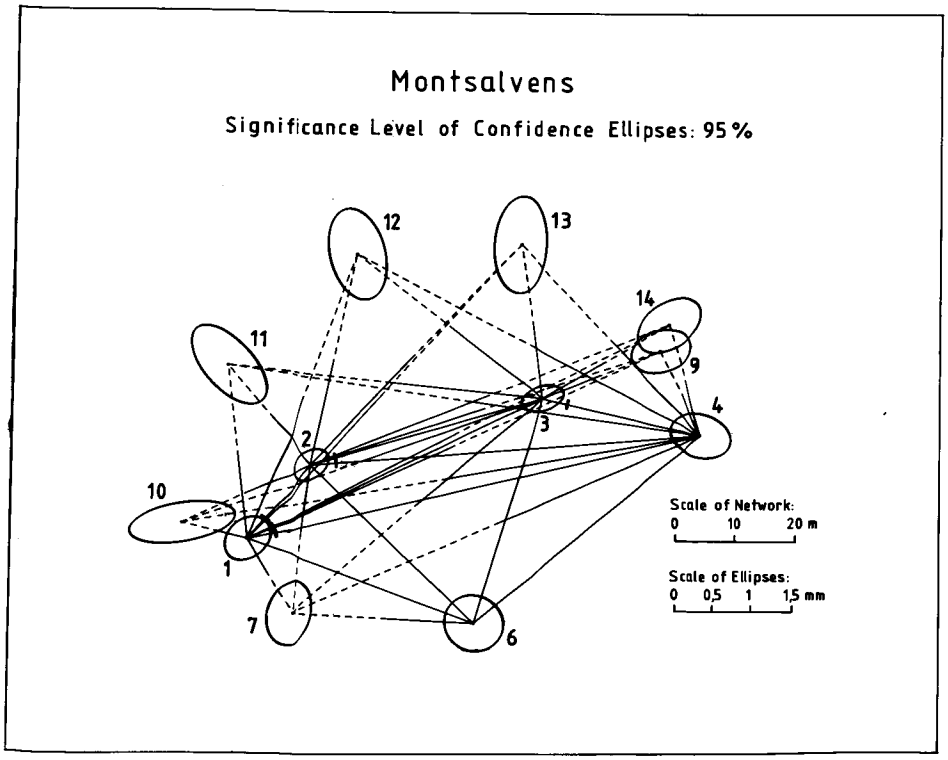


Figure 8.1: Monitoring network Montsalvens

The estimation of the variance components is carried out using Eq. (8-45). As a priori variances for the first estimation step the poor estimates $s_{01}^2 = (1^{cc})^2$ and $s_{02}^2 = (1mm)^2$ are introduced yielding the weight matrix $P = I$. After three iterations the aim $(s_{01}^2)_3 = (s_{02}^2)_3 = 1$ is achieved.

The computations are summarized in Table 8.1. The weights are computed with $\sigma_o^2 = 1$ so that:

$$p_1 = \frac{1}{s_{01}^2} \quad \text{and} \quad p_2 = \frac{1}{s_{02}^2}$$

The current estimates in columns 6 and 7 are defined as:

$$s_{oi}^2 = (s_{oi}^2)_1 \cdot (s_{oi}^2)_2 \cdot \ldots$$

and s_o^2 is the common a posteriori variance factor.

Iteration Step ν	$(s_{01})_\nu$	$(s_{02})_\nu$	$(p_1)_\nu$	$(p_2)_\nu$	s_{01}	s_{02}	s_0
1	2	3	4	5	6	7	8
1	1.00	1.00	1.00	1.00	1.00	1.00	
2	2.64	0.29	0.143	11.89	2.64	0.29	2.40
3	0.99	1.01	0.147	11.89	2.61	0.29	0.99
4	1.003	1.013	0.147	11.89	2.61	0.29	1.005

Table 8.1: Summary of the computations of the variance components for directions and distances of the network of Figure 8.1 using the simplified method of Eq. (8-45).

9. DESIGN OF DEFORMATION NETWORKS

The optimization of the design of a geodetic network is a theoretically very interesting and complex problem. The body of literature on this subject is immense and hard to digest. Recent reviews including numerous references have been given by CROSS (1983a), GRAFAREND (1981) and SCHMITT (1982). The book by GRAFAREND and SANSO (1985) gives the current state of research. In practice, apart from some special projects initiated or supervised by university researchers, none of the mathematical optimization methods have found wide applications. The main reasons for the lack of acceptance are that the methods:

- have a difficult mathematical background,
- require expensive computer facilities,
- often depend on assumptions which are hardly met in practice,
- sometimes yield unreasonable results due to unclear or awkwardly selected objective functions,
- overrate accuracy measures and do not pay enough attention to reliability and feasibility.

The average practising surveyor rather relies on his experience.

Nevertheless, the optimization methods are of more than theoretical interest. They give an excellent insight into the structure of a network model and into the connections between the various measures of quality. Therefore, they can serve as a tool to study geodetic networks and, in due course, help to gain the experience which is needed to do designs without the use of optimization methods.

9.1 Objectives and Variables

The objectives of the optimization of a deformation network are:

- to meet a predetermined accuracy goal,
- to establish a self-checking and reliable mathematical model,
- to yield sufficient sensitivity in respect to certain a priori known functions of the parameters,
- to design an observation scheme which is feasible under practical and financial constraints.

The problems related to the first objective, namely **accuracy**, are easily recognized following the detailed treatment in Chapter 4. All optimization results stemming from datum dependent objective functions are questionable. Only the accuracy of invariant functions is suitable as a basis of an optimization procedure. In monitoring networks it may be possible to work along this line if sufficient a priori knowlege about the deformations is available.

The second criterion, namely **reliability**, has only recently been considered as an objective of its own (van MIERLO, 1981; TESKY and GRUENDIG, 1985). In the past it was treated as a type of boundary condition in accuracy optimization. Strict mathematical approaches towards reliability optimization have not yet been finalized.

The objective of **sensitivity** is of special interest in the context of monitoring networks. If it is possible to predict certain deformation patterns then it is possible to formulate the pertinent hypotheses in the design phase of the net. In this case the optimization aims at a high probability of identifying the true deformation. This criterion is sometimes called testability of pre-formulated hypotheses. See for example NIEMEIER, TESKY and LYALL (1982).

The last (and by scientists often underrated) objective is to control the **cost** of the observations and to work out a feasible observation scheme. As far as cost is concerned there have been many attempts to establish a mathematical model for this part of the problem, but it proved impossible to include all important factors and to realistically assess cost and time implications.

The variables of the optimization methods are the **geometry** of the network and the **weights** of the observables. A direct mathematical approach dealing simultaneously with these two variables in order to achieve all four objectives is not feasible because of the complexity of the problem. All direct methods which have proved to be useful combine just one objective with one variable. Not all combinations are possible, since the geometry as a variable is very difficult to deal with. All attempts to cope with this variable have turned out to be too restrictive to be of practical use. Most of the scientific work, so far, has concentrated on the maximization of some measures of accuracy for a given geometry by selecting the weights. For this special problem a number of efficient algorithms are available (SCHMITT, 1982). A second direct method leads to an observation scheme which produces a cofactor matrix of the parameters being just better than a previously selected criterion matrix (GRAFAREND and KRUMM, 1983). Refer to Section 4.5.

9.2 Computer Simulation Methods

The most promising approach to the network design problem is the computer simulation method. The method is flexible enough to consider all objectives quasi-simultaneously and to vary geometry and observational weights. A criticism heard occasionally claiming that simulation methods and particularly those of the trial and error type were pseudo-scientific and were never producing the true absolute optimum, does not have any practical importance. The limitation of this method lies in the computer equipment as such and the availability of programs. A good example is described in MEPHAM and KRAKIWSKY (1984) while CROSS (1983b) provides a directory of existing software.

The design procedure consists of iterations of the following steps:

a. Selection of the locations of the points.

b. Specification of all possible observations.

c. Estimation of the accuracy of the observations based on the instruments considered and on the number of measurements per observation.

d. Formulation of the GMM and computation of the project relevant measures of accuracy, reliability and sensitivity; assessment of cost.

e. Computation of the influence of each observation on all four measures listed in step d and ordering of the observations according to their influence on the objective functions.

f. Comparison of the results of step d with the given criteria and search for the minimum cost solution with just sufficient accuracy, reliability and sensitivity by:

   f.1 deletion of low ranking observations,
   f.2 consideration of a change of instruments,
   f.3 increase of the repetition number of suitable observations.

   After major changes of the design the steps d to f are to be repeated.

g. After obtaining a satisfactory result, the complete process should be repeated using alternative point locations.

The extent of the computations when using this method must not be underestimated. Efficient algorithms for matrix operations and the consequent use of sequential least squares techniques are required. Interactive graphics may assist the designer in order to accelerate the decision making process between trials.

9.3 Some Details of the Design Steps

The steps listed in the previous section will be reviewed here in more detail.

Step a. The actual users of networks are frequently geophysists, geologists, glaciologists, civil engineers or mining engineers. These specialists may anticipate a certain deformation pattern of the object or the area to be monitored. At least they should be able to indicate the most important part of the structure and where reference points are best established. This information is used as a guide in the selection of sites under the constraint of intervisibility. The density of the network is usually a question of available funds.

In **Step b** it is suggested to introduce all possible observations into the model for the first iteration. In the following steps the number of observations is successively reduced until an optimum is reached. Alternatively, some designers favour a start with a minimum number of observations and selection of those additional elements after each iteration, which cause a minimal increase in cost together with a maximal gain with respect to the other criteria. This alternative, however, is computationally less convenient and yields a result, which unneccessarily depends on the selected observables at the beginning of the simulation.

The a priori estimation of accuracy in **Step c** should be conservative. The usual assumption, based on independent observations, that the variance of the observations is proportional to the reciprocal of their repetition number is very questionable. Practice indicates that all observations taken at one site are dependent (physically correlated) or, in the case of distance measurements, even strongly dependent. Thus it does not make much sense to repeat the observations more than three or four times. Of course, the repetition number of the readings within an arc of directions has to be the same for all targets. This practical rule is usually violated by direct optimization methods. The selection of instruments is an effective way to control accuracy.

The most important point of **Step d** is the formulation of appropriate objective functions. Similar to step a, the expected deformation pattern plays an important role. All the datum related difficulties in assessing the accuracy, as outlined in Chapter 4, are to be considered. If suitable invariant functions can be found, these should be used to establish the criteria (refer to Chapter 7). Some guidelines for the sensitivity criterion can be found in Chapter 5. A cost model estimates the expense of the current design. The number of measures to be reviewed in this step should be moderate but sufficient. Computer graphics can be of great help.

In **Step e** the change of the objective functions caused by the deletion of each single observation is calculated. According to the results of this calculation the observations are grouped with respect to their influence upon the optimization. For this ordering certain rules must be established as the defined criteria are contradicting. This step in the most expensive one in terms of computations. Sequential algorithms are imperative to limit the expenditure. If the designer has enough experience he can select certain candidates for deletion or enhancement. The computations can then be restricted to these observables, thus reducing the cost of computing.

The decisions are made in **Step f.** If the design exceeds the requirements, then the observation scheme is reduced by deleting observables of minor significance. Theoretically, the whole optimization process to this step has to be repeated after each deletion. But a skilled designer will be able to estimate the simultaneous influence of a couple of observations thus speeding up the iteration. If the set criteria are not met, additional observations or an upgrading of the accuracy of selected observations is necessary. The selection of observables for this purpose is again guided by the results (in descending order) of the previous step. When, after some iterations, a satisfactory result has been found, the whole process can be repeated beginning with a different and feasible first design. Since all optimization processes lead to a workable design with respect to the criteria, the final decision is often based on cost arguments.

When the optimal design of the network has been established, the selected point locations are to be monumented. It is important to make every effort to establish marks which will remain stable over a long period of time, and which facilitate the centring of instruments. If high accuracy is required, it is often necessary to build observation pillars of reinforced concrete, equipped with a forced centring system.

9.4 Non-Geodetic Observables

Frequently, the object to be monitored is instrumented with a variety of sensors. Especially large concrete dams are usually equipped with inclinometers, floating plumb lines, extensiometers, strain gauges, temperature sensors and water pore pressure gauges. The readings of these instruments are recorded at short intervals allowing a quasi-permanent monitoring of the object. The observations are only relative and usually affected by time dependent systematic errors. Therefore it is necessary to calibrate the installed devices at regular intervals and to relate the readings to absolute observations. If this is one of the purposes of the geodetic network, then the selection of the point locations has to consider the position of the installed sensors to allow the correlation of the data. The processing is usually done in separate models for the geodetic and for the non-geodetic data. This procedure is justified by the different frequencies of measurement epochs and by the two dissimilar data sets. The calibration or updating of the non-geodetic results at the epochs of geodetic observations is carried out by taking the geodetic results as a reference and by adjusting the others to meet the geodetic results.

Recently models have been suggested to deal simultaneously with all data. This approach is attractive from a theoretical point of view. But in practice large and complex models cause a lot of difficulties. Particularly, the detection of outliers, the detection of systematic errors and the assessment of the a priori accuracies are easier in moderately sized models.

10. TWO-EPOCH ANALYSIS

The simultaneous analysis of two epochs of observations of a monitoring network is of great importance in all deformation studies. It is usually carried out between any two consecutive epochs and, additionally, between the first and the current one. The main objectives of this analysis are:

i. To confirm the stability of reference points and to detect single point movements. The latter are considered as discontinuities with respect to time and locality, thus not conforming to continuous deformation models.

ii. To provide a plot of deformation vectors which assists in developing a suitable deformation model. This geometrical aid is particularly useful for the detection of trends, which may not be detectable by statistical tests.

iii. To inform about the most recent deformations, which may be important for quick decisions which cannot wait until the entire material is analysed.

In many engineering applications two-epoch comparisons provide completely sufficient information and are all that is needed. On the other hand, in scientific projects in general and in geophysical research in particular a full exploitation of the geodetic observations is imperative. The smaller the signal to noise ratio, the more effort is required for the evaluation of the data.

From a didactic point of view, it is much easier to understand the complex approaches to multi-epoch analyses if a comprehensive study of the two-epoch case has preceded. In the latter, the notations and equations are simpler and clearer.

The deformation analysis based on geodetic methods consists usually of four steps:

- The **first step** is to establish a monitoring network and to develop an observation scheme meeting all the requirements dictated by the anticipated deformation pattern and by the specified accuracy (see Chapter 9).

- The **second step** is a thorough analysis of the geodetic data including outlier detection (Chapter 6) and variance component estimation (Chapter 8). This has to be carried out for each epoch separately and must lead to a proper GMM of the network. The importance of this step cannot be stressed enough, since any undetected model errors will be interpreted as deformations in the subsequent steps. Figure 10.1 shows the first two steps in more detail. The flow-chart is self-explanatory in pointing out the links between and the interactions of the various steps and substeps having been treated at length in the previous chapters.

- In the **third step** follows a combined adjustment of the epochs yielding the information required to detect single point movements, to establish a set of reference points and to recognize rigid body displacements of the object or of parts thereof. This step results in a vector of deformations of the object points.

- The **fourth step** aims at a description of the deformation pattern by a suitable deformation model, the parameters of which are estimated and verified statistically.

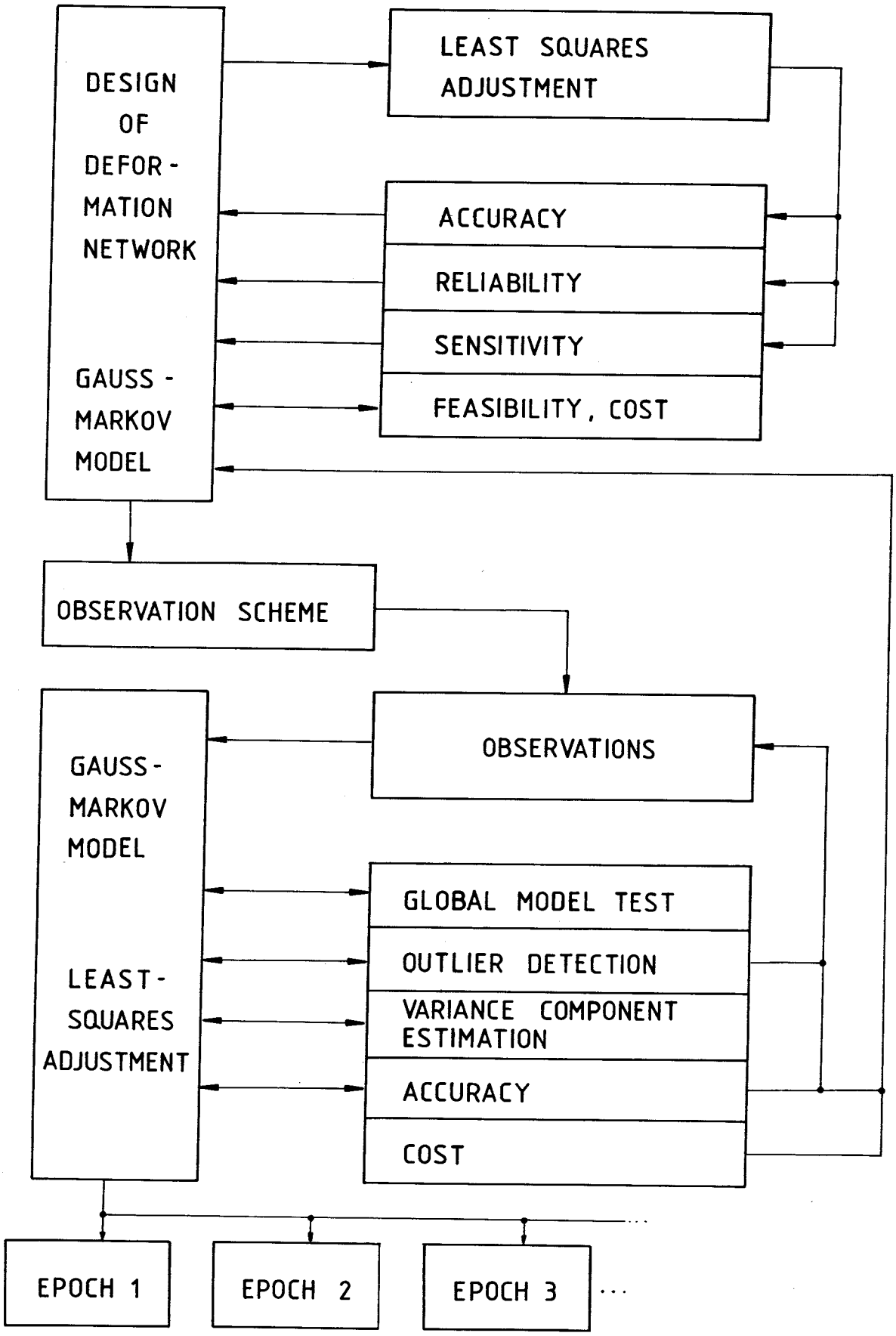


Figure 10.1: Network design and single epoch adjustment.

The last two steps are shown in more detail in Fig. 10.2 in form of a flow-chart. The figure does not require further explanations.

Recent reviews of deformation analysis by geometric methods have been published by HECK et al (1982) and CHRZANOWSKI et al (1985).

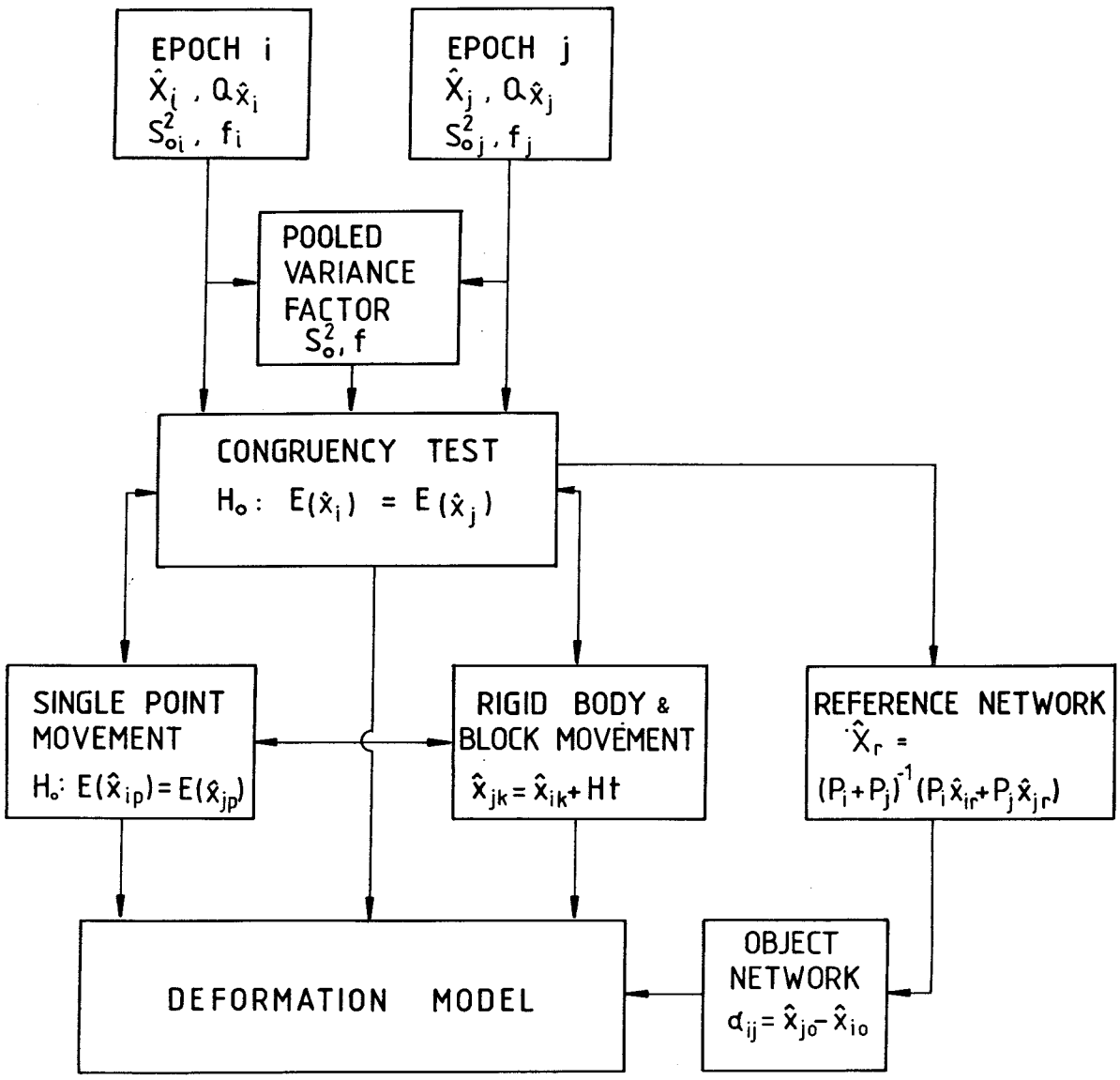


Figure 10.2: Two-epoch analysis.

10.1 Single Epoch Adjustment

The main objective of the single epoch adjustment is to establish an optimal mathematical model of the network. This is achieved by:

i. Screening the data for gross errors,

ii. accounting for systematic errors by a proper selection of additional unknown ("nuisance") parameters,

iii. estimating the a priori accuracy of the observations.

Although this objective is quite common to all estimation methods in geodesy, it is of particular importance in monitoring networks, where an erroneous model will contaminate the estimated deformation parameters. This fact has already been stressed in the introduction to this chapter.

The adjustment is carried out in a sequence of steps which partly form iterative loops and are depicted in Fig. 10.3. The theoretical background has been given in the previous chapters.

a. The Gauss-Markov Model (GMM) is developed and comprises three parts:

- functional part $E(l) = Ax$ (10-1)
- stochastic part $\Sigma_l = \sum_{i=1}^{k} \sigma_{oi}^2 Q_i = \sum_{i=1}^{k} \bar{Q}_i$ (10-2)
- datum constraints $R^t x - c = 0$ (10-3)

It is usually advantageous to supplement all reasonable nuisance parameters to the coordinate parameters in a first tentative model of Eq. (10-1). Omission from the model of insignificant parameters in the subsequent steps changes the estimates less than an addition of significant parameters does, see also step (d).

The decomposition of Σ_l of Eq. (10-2) into a sum of matrices follows along the lines discussed in Chapter 8. The number of variance components k should be moderate in order to avoid numerical difficulties.

The selection of datum constraints according to Eq. (10-3) is not crucial in this adjustment, as it is always possible to transform the single epochs to a common datum (see Section 3.8), before the deformation analysis is carried out. The decision of whether or not the scale is to be considered as a free datum parameter has a large effect on the analysis. Many examples show that a scale parameter can absorb a large portion of the existent deformations. Being aware of this effect, it is advisable in uncertain cases to compute both alternatives. This allows to assess the influence on the problem at hand.

b. A conventional LS-estimation is executed followed by a screening of the observational data for outliers. Depending on the selected method (see Chapter 6), the calculations are either iterative using a cycle of testing and deleting single observations followed by new adjustments or in a re-weighting and re-estimation process.

c. After cleaning the data the variance components are estimated. The stochastic model is adjusted to the new "a priori weights" and the computations are repeated. This iterative process comes to an end when all variance components are stabilizing at $s_{oi}^2 = 1$. In unfavourable cases it may be necessary to return to step (b), because great changes in the original variances can alter the critical statistics of the tests for outliers.

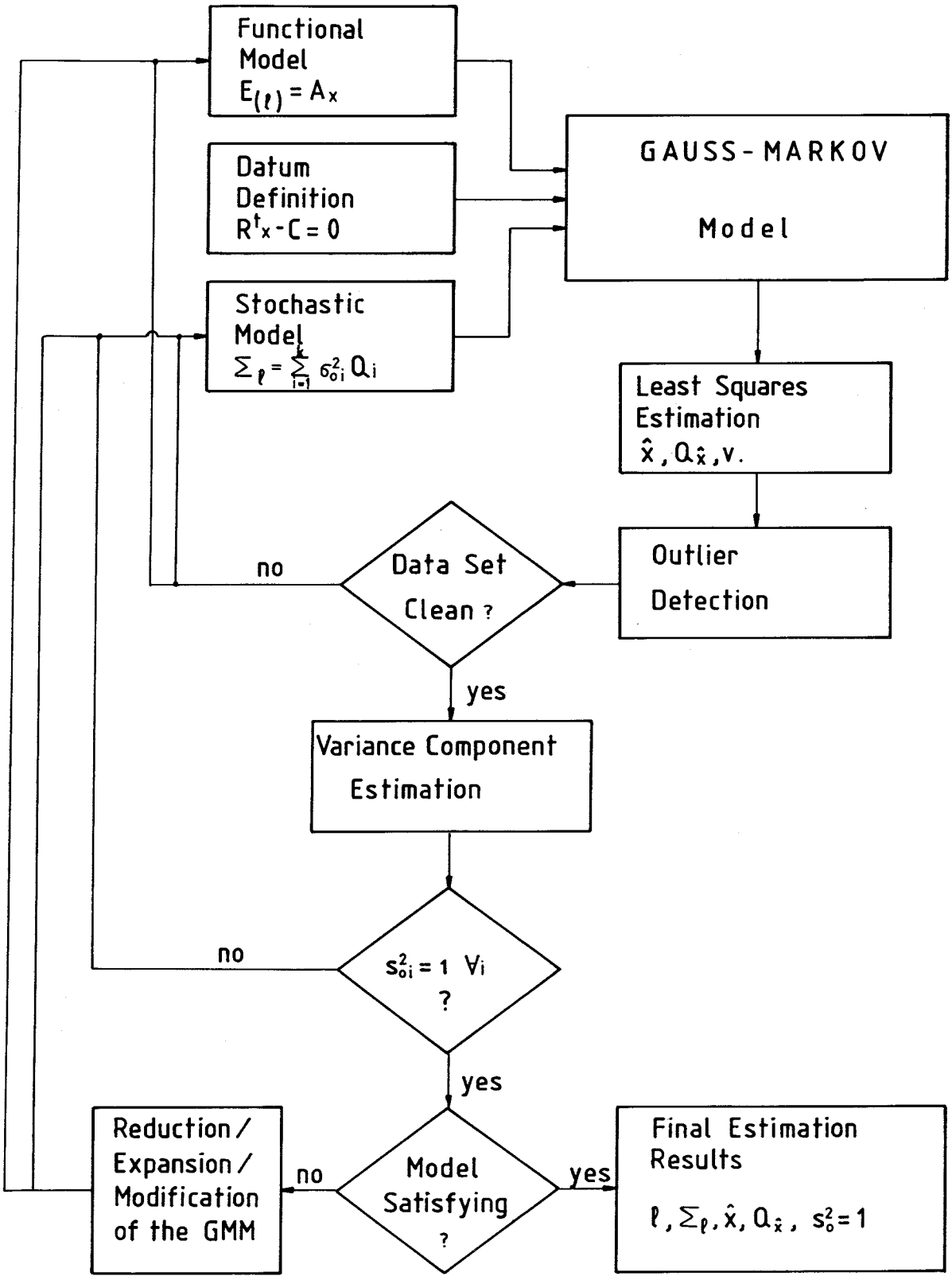


Figure 10.3: Flow-chart of a single epoch adjustment.

d. The final model should contain only significant parameters but all of them. Therefore, the nuisance parameters are checked thoroughly in this step. Statistical tests along the lines of Chapter 5 support the decisions. The exclusion of insignificant parameters will only slightly change the results of the previous steps. Alternatively, if additional nuisance parameters are required then all previous results are rendered obsolete. In this case the whole estimation procedure has to be repeated. It is generally advisable to proceed very carefully with this step since certain types of parameters tend to absorb deformations and thus might spoil the following analysis. Therefore, not only statistics but also a good deal of experience and a sensible attitude are required to achieve a realistic model.

After the model has been finalized the nuisance parameters are usually eliminated in a next step, in order to facilitate the subsequent computations which solely deal with coordinates. For this reason the parameter vector is partitioned into the subvectors x_1 of coordinates and x_2 of nuisance parameters, $x = (x_1, x_2)^t$. If the design matrix A is partitioned accordingly, the model reads:

$$E(l) = (A_1 A_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{10-4}$$

and yields the normal equations:

$$\begin{aligned} A_1^t P A_1 \hat{x}_1 + A_1^t P A_2 \hat{x}_2 - A_1^t P l &= 0 \\ A_2^t P A_1 \hat{x}_1 + A_2^t P A_2 \hat{x}_2 - A_2^t P l &= 0 \end{aligned} \tag{10-5}$$

Provided that the inverse exists, the second equation can be premultiplied by $A_1^t P A_2 (A_2^t P A_2)^{-1}$ which results in:

$$A_1^t P A_2 (A_2^t P A_2)^{-1} A_2^t P A_1 \hat{x}_1 + A_1^t P A_2 \hat{x}_2 - A_1^t P A_2 (A_2^t P A_2)^{-1} A_2^t P l = 0$$

This equation is now subtracted from the first equation of Eq. (10-5) yielding:

$$(A_1^t P A_1 - A_1^t P A_2 (A_2^t P A_2)^{-1} A_2^t P A_1)\hat{x}_1 - (A_1^t P - A_1^t P A_2 (A_2^t P A_2)^{-1} A_2^t P) l = 0 \tag{10-6}$$

This represents the new normal equations which contain only the parameters of interest. For standard cases such as elimination of orientation unknowns, additive constants or scale factors very efficient algorithms exist, which take advantage of the special structure of the coefficient matrix. Modern textbooks on adjustment discuss these cases (e.g. KOCH (1980, Section 32), LAWSON and HANSON (1974, Section 25.5)).

The final results of this step, namely:

$$l, \quad \Sigma_l, \quad \hat{x}, \quad Q_{\hat{x}} = \Sigma_{\hat{x}}, \quad s_{0i}^2 = 1 \tag{10-7}$$

will be subjected to further processing together with the results of other epochs.

10.2 Combined Adjustment

A combined adjustment of two epochs is only necessary in case that correlations between epochs exist. Many practical reasons as well as experience support the presumption of correlations between epochs. But, at present, no proven method is known which would reasonably estimate such correlations. Therefore, it is usually preferred to ignore them.

Also if no actual combined adjustment is carried out, some calculations are required to make the epochs comparable.

Let $l_i + v_i = A_i x_i$, Σ_i and $l_j + v_j = A_j x_j$, Σ_j represent the Gauss-Markov models of epoch i and j, respectively. Then the model for the combined adjustment can be expressed as:

$$\begin{pmatrix} l_i \\ l_j \end{pmatrix} + \begin{pmatrix} v_i \\ v_j \end{pmatrix} = \begin{pmatrix} A_i & 0 \\ 0 & A_j \end{pmatrix} \begin{pmatrix} x_i \\ x_j \end{pmatrix}, \quad \begin{pmatrix} \Sigma_i & 0 \\ 0 & \Sigma_j \end{pmatrix} = \sigma_o^2 \begin{pmatrix} Q_i & 0 \\ 0 & Q_j \end{pmatrix} \tag{10-8}$$

or in concise form:

$$l + v = A x, \quad \Sigma = \sigma_o^2 Q$$

The only difference to a simple merger of the single epoch adjustments is that, here, one variance factor σ_o^2 common to both epochs results. The estimate of this variance factor can also be computed from the two single epoch adjustments provided that the same a priori factor σ_o^2 has been used or, more precisely, the estimates have the same expectation:

$$E(s_{oi}^2) = E(s_{oj}^2) = \sigma_o^2 \tag{10-9}$$

This can be verified by the usual F-test. If the variance-component estimation (Chapter 8) has been applied to the single epochs, then the equality of the variance factors is guaranteed. The common or pooled variance estimate is computed from:

$$\begin{aligned} s_o^2 &= \frac{f_i s_{oi}^2 + f_j s_{oj}^2}{f_i + f_j} = \frac{q}{f} \\ q &= (v^t P v)_i + (v^t P v)_j; \quad f = f_i + f_j \end{aligned} \tag{10-10}$$

where f_i, f_j and f are the respective degrees of freedom of single epochs and combined adjustment, respectively.

Two other prerequisites for the comparison of epochs are that:

i. both of the models are based on the same geodetic datum, and that

ii. the same approximate coordinates for the common stations have been used.

While the latter requirement is self-evident and easily fulfilled in the process of Taylorization of the initially non-linear observation equations, the former requires some explanations.

The outcome of the congruency check of the epochs is not influenced by the selection of the geodetic datum. The zero-variance computational base (Section 3.3) and the minimum trace datum (Section 3.7) are equally appropriate. The only condition is that the parameters of the points, which are involved in the datum definition, are present in both epochs. Models which are initially incompatible are easily transformed to a common datum by use of the equations of Section 3.8.

10.3 The Congruency Test

The objective of the congruency test is to detect whether or not the point group considered has remained stable. The point group can consist of all points common to the epochs, of the points establishing the reference net or of those belonging to a specified block of the network.

In the first case the null hypothesis of the global congruency test reads:

H_o: The common points of both epochs are stable and thus have the same expectation of the estimated positions:

$$E(\hat{x}_i) = E(\hat{x}_j) = x \tag{10-11}$$

For a correct mathematical model and normally distributed observations the quadratic form q/σ_o^2 computed from the model of Eq. (10-8) or from Eq. (10-10) has a central χ^2-distribution. Now, if H_o is true, then Eq. (10-8) as extended by the condition equations:

$$x_i - x_j = H^t x = 0$$

is a correct model, as well. Thus the quadratic form q_H/σ_o^2 of this model, where:

$$q_H = v_H^t P v_H \sim \sigma_o^2 \chi^2(f_H); \quad f_H = f + f_\Delta \tag{10-12}$$

is also χ^2-distributed with f_H degrees of freedom, where f is defined in Eq. (10-10) and f_Δ is the number of independent estimates of x_i or x_j, i.e. the rank of the design matrices A_i, A_j of Eq. (10-8). The inclusion of H_o in the model results in the difference q_Δ of the quadratic form:

$$q_\Delta = q_H - q, \quad f_\Delta = r(H) \tag{10-13}$$

Using Theorems 3 and 4 of Section 5.4 it can be easily demonstrated that q_Δ/σ_o^2 is χ^2-distributed with f_Δ degrees of freedom and that q and q_Δ are stochastically independent. Hence, the ratio (refer to Eq. (5-18)):

$$T = \frac{q_\Delta/f_\Delta}{q/f} \sim F(f_\Delta, f) \tag{10-14}$$

is a suitable test statistic with a central Fisher-distribution if H_o is true.

Two different but equivalent methods can be used for the computation of q_Δ. Which method suits best depends on the software available.

i. The model of Eq. (10-8) is modified in such a way that the common points of the epochs are introduced only once. If the corresponding coordinates are denoted by x^c and the associated part of the design matrix by A^c, the modified model takes the form:

$$\begin{pmatrix} l_i \\ l_j \end{pmatrix} + \begin{pmatrix} v_i \\ v_j \end{pmatrix} = \begin{pmatrix} A_i^c & A_i^n & 0 \\ A_j^c & 0 & A_j^n \end{pmatrix} \begin{pmatrix} x^c \\ x_i^n \\ x_j^n \end{pmatrix} \tag{10-15}$$

where the subvectors x^n represent the non-common points or nuisance parameters and A^n the corresponding blocks of the design matrix. The quadratic form of the residuals in this model equals q_H of Eq. (10-12), thus q_Δ follows from Eq. (10-13).

ii. The model of Eq. (10-8) is extended by the condition equations:

$$H^t x = (I, -I)\begin{pmatrix} x_i \\ x_j \end{pmatrix} = 0 \qquad (10\text{-}16)$$

representing the standard case of Chapter 5. In this chapter it has been pointed out that the corresponding adjustment must not be carried out, since q_Δ can be computed directly from Eq. (5-31):

$$q_\Delta = (\hat{x}_i - \hat{x}_j)^t (Q_{\hat{x}_i} + Q_{\hat{x}_j})^- (\hat{x}_i - \hat{x}_j) \qquad (10\text{-}17)$$

If the models of the epochs contain non-common points or nuisance parameters, these have to be eliminated from the cofactor matrices $Q_{\hat{x}}$ prior to the summation according to Eq. (10-17). This elimination is achieved by deleting all rows and columns of the matrices $Q_{\hat{x}}$ referring to the parameters to be eliminated followed by a re-ordering of the matrices.

In case that the monitoring network comprises reference points and object points or the network is split into blocks conforming to some expected deformation pattern, the procedure is usually different. The reference block or the part with the presumably highest relative stability is selected as the basis for the calculation of the deformation of all other points. In the sequel this basis is termed reference block of the monitoring network or, in short, reference block. The parameter vector is partitioned accordingly into the subvector x^r pertaining to the reference block and x^o referring to the object points:

$$x = (x^r, x^o)^t$$

The problem of investigating the stability of the reference block is solved by a test of the null hypothesis:

H_o: All points of the reference block are stable and have the same expectation of the position in both epochs:

$$E(\hat{x}_i^r) = E(\hat{x}_j^r) = x^r \qquad (10\text{-}18)$$

The quadratic form q_Δ can be calculated as outlined for the general case. The points of the reference block replace the common points in Eq. (10-15) and the object points take the position of the non-common ones, hence:

$$\begin{pmatrix} l_i \\ l_j \end{pmatrix} + \begin{pmatrix} v_i \\ v_j \end{pmatrix} = \begin{pmatrix} A_i^r & A_i^o & 0 \\ A_j^r & 0 & A_j^o \end{pmatrix} \begin{pmatrix} x^r \\ x_i^o \\ x_j^o \end{pmatrix} \qquad (10\text{-}19)$$

The quadratic form q_H of this model has:

$$f_H = f_i + f_j + f_\Delta$$

degrees of freedom, where f_Δ equals the number of elements of x^r minus the datum defect d. The test statistic is formed with $q_\Delta = q_H - q$ and f_Δ according to Eq. (10-14).

Alternatively, the quadratic form can be computed by explicitly introducing the conditions associated with Eq. (10-18):

$$H^t x = x_i^r - x_j^r = 0$$

Equation (10-17) is then employed to compute the quadratic form.

The result of the F-test of congruency is either a rejection of H_o, leading to further analysis as outlined in the next section, or a non-rejection in which case it might be possible to skip the next section. But such a decision should be taken with due care, as a non-rejection does not mean that the hypothesis is proved. The type I error probability of this test is usually α = 5% (see related remarks in Chapter 5).

10.4 Single Point Diagnosis

If the null hypothesis described in the previous section fails, it is necessary to investigate the underlying reasons. The techniques for this investigation are outlined in this section and apply also to cases where single points do not fit into a generally satisfying deformation model or, in general, when the estimate of the variance factor under a certain hypothesis is suspiciously large.

The alternative hypothesis to H_o of the congruency test of the previous section reads as follows:

H_a: All points except **one** conform to the considered deformation model.

The deformation model used is either the model of Section 10.3 where "all points of the reference block are stable" and the reference block may contain all common points or any other model discussed in subsequent sections.

Two basically different methods have been developed to test H_o against H_a: a conventional approach assuming normally distributed observations and employing least squares estimation and a robust method which suits a larger range of distributions.

The conventional approach is more widely accepted. It exists in some variations which all lead to identical results if applied rigorously. The basic idea is to split the quadratic form q_Δ of Eq. (10-17) into as many parts as there are points involved in H_o. The splitting method, where each point forms a part, can be interpreted as providing the contribution of such a point to the total form q_Δ. The point with the largest part is considered as non-conformingly deformed in the sense of H_a.

10.4.1 Decomposition by Implicitly Formulated Null Hypothesis

The method of decomposing q_Δ is the most direct approach. It is very transparent but requires extensive computations. The evaluation of the model of Eq. (10-19) yields the quadratic form q_Δ and the degrees of freedom f_Δ. The contribution of a point P to this quadratic form can be computed indirectly from the model of Eq. (10-20):

$$\begin{pmatrix} l_i \\ l_j \end{pmatrix} + \begin{pmatrix} v_i \\ v_j \end{pmatrix} = A_p x_p \qquad (10\text{-}20)$$

$$x_p = (\bar{x^r},\ x_i^p,\ x_i^o,\ x_j^p,\ x_j^o)$$

$$A_p = \begin{pmatrix} \bar{A_i^r} & A_i^p & A_i^o & 0 & 0 \\ \bar{A_j^r} & 0 & 0 & A_j^p & A_j^o \end{pmatrix}$$

where x^p comprises the coordinates of point P being separated from the subvector x^r and treated like the parameters in the subvector x^o. This means that the condition $\hat{x}_i^p = \hat{x}_j^p$ has been removed and leads to the reduced quadratic form q_H'. The difference:

$$q_\Delta^P = q_H - q_H' = (q + q_\Delta) - (q + q_\Delta') = q_\Delta - q_\Delta' \qquad (10\text{-}21)$$

is exactly the contribution of the hypothesis $E(\hat{x}_i^P - \hat{x}_j^P) = 0$ to the quadratic form q_Δ. The evaluation of Eq. (10-20) has to be carried out for every point of the reference block separately. In other words, all points have to play the role of point P in turn. This leads to a set of quadratic forms with two degrees of freedom each (in case of a horizontal network).

10.4.2 Successive Decomposition of the Quadratic Form

To facilitate the derivation, the Equation (10-17) is rewritten as:

$$q_\Delta = \Delta^t Q_\Delta^- \Delta, \; f_\Delta = r(Q_\Delta) \qquad (10\text{-}22)$$

with:

$$\Delta = \hat{x}_i - \hat{x}_j, \qquad Q_\Delta = Q_{\hat{x}_i} + Q_{\hat{x}_j}$$

Successively, each point is considered as having changed its position and all others as being stable. This corresponds to the partitioning of Δ into two subvectors:

$$\Delta = (\Delta_n^t \; \Delta_p^t)^t$$

where Δ_p contains the coordinate differences of the considered point P while Δ_n comprises the differences of all other "non-changed" points.

The form matrix $Q_\Delta^- = P_\Delta$ of Eq. (10-22) is partitioned accordingly, yielding:

$$\Delta = \begin{pmatrix} \Delta_n \\ \Delta_p \end{pmatrix}, \quad Q_\Delta^- = P_\Delta = \begin{pmatrix} P_{nn} & P_{np} \\ P_{pn} & P_{pp} \end{pmatrix} \qquad (10\text{-}23)$$

The quadratic form q_Δ can be decomposed into statistically independent subforms by a transformation of Δ. This has to be carried out in such a way that two orthogonal subvectors result referring to point P and all other points, respectively. This transformation is obtained by:

$$\bar{\Delta} = \begin{pmatrix} I & 0 \\ P_{pp}^{-1} P_{pn} & I \end{pmatrix} \Delta = \begin{pmatrix} \Delta_n \\ \bar{\Delta}_p \end{pmatrix} \qquad (10\text{-}24)$$

To express $P_{pp}^{-1} P_{pn}$ in terms of the corresponding blocks of the cofactor matrix Q_Δ the g-inverse is required. For the block matrix the generalized inverse has the form:

$$P_\Delta = Q_\Delta^- = \begin{pmatrix} Q_{nn} & Q_{np} \\ Q_{pn} & Q_{pp} \end{pmatrix}^- = \begin{pmatrix} P_{nn} & P_{np} \\ P_{pn} & P_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} Q_{nn}^- + Q_{nn}^- Q_{np} B^{-1} Q_{pn} Q_{nn}^- & -Q_{nn}^- Q_{np} B^{-1} \\ -B^{-1} Q_{pn} Q_{nn}^- & B^{-1} \end{pmatrix} \qquad (10\text{-}25)$$

where:

$$B = Q_{pp} - Q_{pn} Q_{nn}^- Q_{np}$$

From Eq. (10-25) follows $P_{pp} = B^{-1}$ and $P_{pn} = -B^{-1}Q_{pn}Q_{nn}^{-}$, hence:

$$P_{pp}^{-1}P_{pn} = -Q_{pn}Q_{nn}^{-}$$

which can be substituted in Eq. (10-24). Application of the law of variance propagation to Eq. (10-24):

$$Q_{\Delta}^{-} = \begin{pmatrix} I & 0 \\ -Q_{pn}Q_{nn}^{-} & I \end{pmatrix} \begin{pmatrix} Q_{nn} & Q_{np} \\ Q_{pn} & Q_{pp} \end{pmatrix} \begin{pmatrix} I & (-Q_{pn}Q_{nn}^{-})^{t} \\ 0 & I \end{pmatrix}$$

yields:

$$Q_{\Delta}^{-} = \begin{pmatrix} Q_{nn} & 0 \\ 0 & Q_{pp} - Q_{pn}Q_{nn}^{-}Q_{np} \end{pmatrix} \qquad (10\text{-}26)$$

The statistical independence of the subvectors Δ_n and $\bar{\Delta}_p$ is thus demonstrated.

The quadratic form for $\bar{\Delta}$:

$$\bar{\Delta}^{t}Q_{\Delta}^{-}\bar{\Delta} = \Delta_n^{t}Q_{nn}^{-}\Delta_n + \bar{\Delta}_p(Q_{pp} - Q_{pn}Q_{nn}^{-}Q_{np})^{-1}\bar{\Delta}_p$$

is thus decomposed in two independent subforms. Comparison with Eq. (10-25) leads to the identities:

$$Q_{nn}^{-} = P_{nn} - P_{np}P_{pp}^{-1}P_{pn}$$

$$P_{pp}^{-1} = Q_{pp} - Q_{pn}Q_{nn}^{-}Q_{np} \qquad (10\text{-}27)$$

and to the final decomposed form of q_{Δ}:

$$\Delta^{t}Q_{\Delta}^{-}\Delta = \bar{\Delta}^{t}Q_{\Delta}^{-}\bar{\Delta} = \Delta_n^{t}Q_{nn}^{-}\Delta_n + \bar{\Delta}_p^{t}P_{pp}\bar{\Delta}_p$$

$$q_{\Delta} = q_{\Delta}^{n} + q_{\Delta}^{p} \qquad (10\text{-}28)$$

Since the transformation of Eq. (10-24) is regular, it does not change the value of q_{Δ}. The subform q_{Δ}^{P} is the contribution of point P to the form q_{Δ}. It is identical with q_{Δ}^{P} of Eq. (10-21).

This decomposition is carried out for every point, yielding a set of quadratic forms which serve as a base for further investigations.

In statistical terminology $\bar{\Delta}_p$ is the conditional estimate of Δ_p for Δ_n given.

10.4.3 Cholesky Decomposition

Through an application of the Cholesky decomposition, as fully outlined in LAWSON and HANSON (1974, Chapters 19 and 25), the form matrix $P_\Delta = Q_\Delta^-$ of the quadratic form of Eq. (10-22) can be factorized in the product of a lower triangular matrix C and its transposed yielding:

$$Q_\Delta^- = P_\Delta = CC^t \tag{10-29}$$

Compare also Eqs (3-60) and (3-93).

The triangular matrix C can be computed by reduction of P_Δ using the usual Gauss elimination algorithm for symmetrical matrices with subsequent division of each row by the square root of the diagonal element. The quadratic form can then be written as:

$$q_\Delta = \Delta^t Q_\Delta^- \Delta = \Delta^t CC^t \Delta = d^t d \tag{10-30}$$

with:

$$d = C^t \Delta$$

The cofactor matrix of the vector d becomes:

$$Q_d = C^t Q_\Delta C = C^t (CC^t)^- C = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \tag{10-31}$$

If the reference block under investigation comprises m points, then Δ contains 2m differences in the 2-dimensional case. The rank of Q_Δ is f_Δ, thus Q_Δ has a rank deficiency of $2m - f_\Delta$. Hence the same number of rows of C and the corresponding elements of the vector d vanish. The identity matrix of Eq. (10-31) is of order $f_\Delta \times f_\Delta$.

It is advantageous to consider Eq. (10-30) as a sum of independent quadratic forms:

$$q_\Delta = q_\Delta p1 + q_\Delta p2 + \ldots + q_\Delta p f_\Delta/2 \tag{10-32}$$

where each subform refers to one point. The Cholesky factorization has two properties which are of special interest in this context. Firstly, the value of the form $q_\Delta pi$ depends on the position of the corresponding differences $\Delta\hat{x}_i$, $\Delta\hat{y}_i$ in the vector Δ. Secondly, if $q_\Delta pi$ is once calculated it is not affected any more by position changes of points being positioned behind it in the vector.

The pivoting strategy discussed below leads to a decomposition of q_Δ so that the first subform $q_\Delta p1$ refers to the point with the greatest contribution to q_Δ. The next form $q_\Delta p2$ is the greatest contribution to the reduced form $q_\Delta - q_\Delta p1$, then follows the third greatest contribution and so on. This ordering is achieved by cyclic exchange of the rows and columns of q_Δ and of the corresponding elements of Δ and the subsequent computation of the current $q_\Delta p1$. The point with the maximal value of $q_\Delta p1$ gets the leading position in Δ, and the reduction is completed for this point p_1. In the next cycle the remaining points are shifted successively into the second position where the correspondent value of $q_\Delta p2$ is computed. The point maximizing this value gets the second position and the reduction continues. After $f_\Delta/2$ cycles the decomposition is completed and the result of Eq. (10-32) available.

It is important to realize that the sets of subforms computed by the methods of Sections 10.4.1 and 10.4.2 are identical but different from the set obtained in this Section, 10.4.3. They correspond to a set of subforms $q_\Delta{}^{P1}$ which results when all points successively take the leading position during the first cycle of a decomposition according to Section 10.4.3.

10.4.4 Statistical Tests

The conventional decomposition methods outlined provide the statistics for the testing of the alternative hypothesis, that one point in the block considered is unstable. Various test procedures are in use, which differ in the error probabilities applied and in the rigour of the statistical basis:

i. The failure of the congruency test at type I error probability α leads to the acceptance of H_a which gives reason for the decomposition of q_Δ. The point with the greatest contribution to q_Δ is eliminated from the reference block and a new congruency test with the same risk α is applied to the remaining reduced reference block. The corresponding reduced quadratic form equals q_H' of Eq. (10-21), $q_\Delta{}^n$ of Eq. (10-28) or $q_\Delta - q_\Delta{}^{P1}$ of Eq. (10-32) depending on the decomposition method.

If the test fails again, the reduced quadratic form is decomposed and the point with the greatest contribution is considered as unstable and excluded from the reference block. This sequence of congruency test, decomposition and localization is repeated until the test statistic falls into the region where H_o is accepted. The partial network under investigation (the reduced reference block or the reduced network of common stations) contains then only statistically verified stable points.

ii. Following the data snooping approach of outlier detection (see Section 6.4) the variance factor $\sigma_o{}^2$ replaces the denominator in Eq. (10-14) yielding a $F(f_\Delta, \infty)$-distributed test statistic for the congruency test of the reference block. A rejection of the null hypothesis gives rise to a test for single point movements. The test statistic:

$$T = \frac{0.5 q_\Delta{}^P}{\sigma_o{}^2} \sim F(2, \infty) \qquad (10\text{-}33)$$

is calculated for each point, where $q_\Delta{}^P$ is computed from Eq. (10-21) or Eq. (10-28). The critical value for the m (number of points involved) single point tests is taken from the F-distribution for the error probabilities of $\alpha_o = 0.1\%$ and $\beta_o = 20\%$ of type I and type II errors, respectively. These tests are related to the global congruency test by the common type II error probability $\beta = \beta_o$ and the common noncentrality parameter $\lambda = \lambda_o$. The risk α of committing a type I error in the global test is defined by Eq. (6-30):

$$\lambda_o = \lambda(\alpha, \beta_o, f_\Delta) = \lambda(\alpha_o, \beta_o, 2) \qquad (10\text{-}34)$$

and taken from graphs or tables. For more details, the reader is referred to KOK(1982).

The point with the largest statistic according to Eq. (10-33) is considered as unstable and transferred from the block of reference points to the block of object points. The test procedure consisting of the congruency test of the reduced reference block, the computation of the statistics of Eq. (10-33) and the localization and exclusion of the most probable unstable point is repeated until a complete separation of stable and unstable points is accomplished. The result is a confirmed reference block for subsequent stages of the deformation analysis.

iii. If the quadratic form has been decomposed by the Cholesky method, then $f_\Delta/2 = k(f_\Delta = rQ_\Delta)$ independent quadratic subforms $q_\Delta P_i$ are available, which are ordered by the pivot strategy described. The null hypothesis for each point:

$$E(0.5q_\Delta P_i) - \sigma_o^2 = 0$$

taking point P_i as a stable one, leads to $f_\Delta/2$ test statistics:

$$T_i = \frac{0.5q_\Delta P_i}{s_o^2}, \qquad i = 1, 2, ..., f_\Delta/2 = k \tag{10-35}$$

which are F(2, f)-distributed according to Eq. (5-18). Since the T_i are independent the k tests can be carried out simultaneously. To tune the test with respect to type I error probability α of the global congruency test, the critical value of the F-distribution is computed at:

$$\bar{\alpha} = 1 - (1 - \alpha)^{1/k}$$

All T_i greater than the critical value indicate single point movements and lead to an exclusion of the corresponding points. The remaining reference block is considered stable and serves as a basis for further stages of the analysis. (For details see CASPARY and SCHWINTZER (1981).)

The tests (i) and (ii) are similar as far as the procedure is concerned: an iterated congruency test is executed. But the philosophy of fixing the error probability is different as is the computation of the test statistics which is based on σ_o^2 or s_o^2, respectively. In the procedure (iii) single point tests are performed simultaneously using a type I error probability different from both (i) and (ii). No wonder that the tests may give different results in marginal cases when the test statistics are near the critical values. This should not be considered as a shortcoming of the test methods, but rather as an indication that a test must never be more than an aid in decision making.

From a computational point of view a combination of successive decompositions (see Section 10.4.2) with test (ii) is most efficient if only few unstable points are expected in the reference block. For a greater number of unstable points (4 or more) the Cholesky factorization (Section 10.4.3) together with test (iii) may be the best choice.

10.4.5 Robust Method

The robust method of single point screening is comparable with the robust outlier detection method of Section 6.6. The difference vector $\Delta = \hat{x}_i - \hat{x}_j$ is considered as a vector of observations of a deformation field, which can be modelled in some way. To keep in line with the conventional methods in this chapter, this model is viewed as a similarity transformation, assuming that the points can be brought to coincidence by relative translations, rotations and a scale adjustment. Thus only the datum parameters (see Section 3.8) must be changed to obtain congruency of the networks considered. The corresponding model is given by the equations of a similarity (or Helmert) transformation.

Let $m = u/2$ be the number of points, Δ the vector of observed coordinate differences, δ the residual vector, H the u x 4 design matrix and t the vector of transformation parameters. Then:

$$\Delta + \delta = Ht \tag{10-36}$$

with:

$$\Delta = (\Delta_1^t, \quad \Delta_2^t, \quad \ldots, \quad \Delta_m^t)^t$$

$$\delta = (\delta_1^t, \quad \delta_2^t, \quad \ldots, \quad \delta_m^t)^t$$

$$\Delta_k = \begin{pmatrix} \hat{x}_{ik} - \hat{x}_{jk} \\ \hat{y}_{ik} - \hat{y}_{jk} \end{pmatrix}, \quad \delta_k = \begin{pmatrix} \delta x_k \\ \delta y_k \end{pmatrix}$$

$$H = (H_1^t, \quad H_2^t, \quad \ldots \quad H_m^t)^t \qquad H_k^t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\bar{y}_k & \bar{x}_k \\ \bar{x}_k & \bar{y}_k \end{pmatrix}$$

$$t = (t_x, \quad t_y, \quad r_z, \quad s)$$

The components of t are explained in Section 3.2. The coefficient matrix H is identical with the matrix S of Eq. (3-61). The subvectors Δ_k and δ_k refer to the point P_k. Obviously, the components δx_k and δy_k of δ_k depend on the orientation of the coordinate system. In contrast the length of the residual discrepancy vector is independent of coordinate system's orientation:

$$d_k = \sqrt{\delta x_k^2 + \delta y_k^2} = \sqrt{\delta_k^t \delta_k} \qquad (10\text{-}37)$$

The parameter vector t is estimated by minimizing the sum of the lengths of the residual discrepancies:

$$\sum_{k=1}^{m} d_k = \min \qquad (10\text{-}38)$$

The minimum of the objective function in Eq. (10-38) is obtained by forming its derivatives with respect to t and setting them to zero:

$$\frac{\partial}{\partial t} \sum_{k=1}^{m} d_k = 0 \qquad (10\text{-}39)$$

Equation (10-36) can be expressed as:

$$\delta_k = H_k t - \Delta_k \qquad (10\text{-}40)$$

and substituted into Eq. (10-39):

$$\frac{\partial}{\partial t} \sum_{k=1}^{m} [(H_k t - \Delta_k)^t (H_k t - \Delta_k)]^{1/2} = 0 \qquad (10\text{-}41)$$

which yields:

$$\sum_{k=1}^{m} H_k^t \delta_k / d_k = 0 \qquad (10\text{-}42)$$

The final estimation equation is obtained by expressing the above equation in matrix notation:

$$H^tD^{-1}H\hat{t} - H^tD^{-1}\Delta = 0 \qquad (10\text{-}43)$$

where the matrix D is defined by:

$$D = \text{diag}\,(d_1, d_1, d_2, d_2, \ldots, d_m, d_m) \qquad (10\text{-}44)$$

The numerical solution for the parameter estimate $\hat{t}$ and the discrepancies δ_k is carried out through an iteratively reweighted least squares process, using:

$$t_\nu = (H^tD_{\nu-1}{}^{-1}H)^{-1}H^tD_{\nu-1}{}^{-1}\Delta$$

$$\nu = 1, 2, \ldots \qquad (10\text{-}45)$$

$$\delta_\nu = \Delta - Ht_\nu, \qquad D_o{}^{-1} = I$$

The computation is initialized with the "weights" $D_o{}^{-1} = I$. Thus, the first step yields the usual least squares estimate of t. The algorithm of Eq. (10-45) converges after 15 to 20 iterations. The final residual discrepancies d_k are estimates of point deformations.

The estimation method is robust in the sense that large deformations of single points do not affect the parameter estimate $\hat{t}$. The deformations emerge as residuals with full magnitude and do not contaminate the residuals of the stable points. Unfortunately, the covariance matrix of the residuals Q_δ cannot be computed in the familiar way since the relation between Δ and t is non-linear according to Eq. (10-43); thus the law of variance propagation does not apply. A conservative estimate of Q_δ is the original cofactor matrix $Q_\Delta = Q_{\hat{x}_i} + Q_{\hat{x}_j}$; see Eq. (10-22). The separation of stable and unstable points should be based on or supported by statistical tests. An approximate test can be developed from the quadratic form q_k referring to point P_k:

$$q_k = \delta_k{}^tQ_{\delta_k}{}^{-1}\delta_k \qquad (10\text{-}46)$$

where $Q_{\delta_k} = Q_{\Delta_k}$ is the 2 x 2-submatrix of Q_Δ corresponding to point P_k. The following expression can serve as a test statistic:

$$T_k = \frac{q_k}{\sigma_o{}^2} \sim \chi^2(2) \qquad (10\text{-}47)$$

The statistic is $\chi^2(2)$-distributed if H_o (P_k is stable) is true. The alternative statistic:

$$T_k' = \frac{q_k/2}{s_o{}^2} \sim F(2, f) \qquad (10\text{-}48)$$

is F(2, f)-distributed under the same assumption; compare Eqs (5-18) and (5-19).

The type I error probability $\bar{\alpha}$ of the m tests should be computed from:

$$\bar{\alpha} = 1 - (1 - \alpha)^{1/m}$$

where $1 - \alpha$ is the overall significance level of, for example, 95%.

10.4.6 Final Adjustment and Graphics

Based on one of the outlined methods of single point stability diagnosis, an undeformed reference block will eventually be identified. A final combined adjustment is then executed, in which one set of coordinates is estimated for the stable points, while the object points (including the points removed from the reference block) are parameterized for each epoch separately. The model has therefore the form of Eq. (10-19). The geodetic datum is introduced by minimizing the trace of the submatrix of the cofactor matrix $Q_{\hat{x}}$ which refers to the established reference block; for details see Section 3.8. The position differences of the object points, as computed from this adjustment, form the basis for all further deformation analyses.

In practice it is not required to actually carry out this adjustment, since the results can be derived from a transformation similar to the decomposition method of Section 10.4.2. If the difference vector Δ and the weight matrix P_Δ are partitioned as shown in Eq. (10-23):

$$\Delta = \begin{pmatrix} \Delta_r \\ \Delta_o \end{pmatrix}, \quad Q_\Delta = P_\Delta^- = \begin{pmatrix} P_{rr} & P_{ro} \\ P_{or} & P_{oo} \end{pmatrix} \tag{10-23a}$$

where the subscripts r and o refer to the reference and object points respectively, then it follows from Eq. (10-24) that:

$$\bar{\Delta}_o = P_{oo}^{-1} P_{or} \Delta_r + \Delta_o \tag{10-24a}$$

and, from Eq. (10-26):

$$Q_{\bar{\Delta}_o} = Q_{oo} - Q_{or} Q_{rr}^- Q_{ro} = P_{oo}^{-1} \tag{10-26a}$$

The quantities $\bar{\Delta}_o$ and $Q_{\bar{\Delta}_o}$ are exactly the inputs of the analysis required and are identical to the results of a final adjustment as specified above.

The methods of stability diagnosis can be exemplified by supporting the numbers with graphical means. The contribution of point P to the total quadratic form q_Δ is computed, for example, from Eq. (10-28):

$$q_\Delta^p = \bar{\Delta}_p^t P_{pp} \bar{\Delta}_p \tag{10-49}$$

This equation defines a two-dimensional quadratic form which can be visualized as an ellipse. The semi-axes and their orientation are given by the square roots of the eigenvalues and by the eigenvectors, respectively, of the inverse of the form matrix:

$$P_{pp}^{-1} = Q_{pp} - Q_{pn} Q_{nn}^- Q_{np} \tag{10-50}$$

According to the F-test applicable the axes of these ellipses are to be multiplied by:

$$s_o \sqrt{2F_\alpha(2, f)}$$

For details of the computation refer to Eqs (4-32) to (4-42). The resulting confidence ellipse is sometimes called deformation ellipse (HECK et al (1982)) and can be plotted at point P together with the deformation vector $\bar{\Delta}$. If the tip of the vector falls outside of the ellipse, then the point is unstable. Since this is done for all points, a very good picture of the deformation field and a better understanding of the result of the single point analysis is obtained. Furthermore, the plot contains valuable information for the design of a deformation model.

All equations and descriptions in this chapter refer to two-dimensional (2d) networks. The necessary modifications for 1d- or 3d-networks are readily established.

10.5 Rigid Body Displacement

After the localization of points exhibiting a deformation feature which does not conform with the general pattern of the surrounding, three different cases need to be considered for a further analysis depending on the type of the monitoring problem:

i. The stability of the points of the reference block is verified. For these points only one set of coordinates is estimated. The deformation vectors as depicted on a plan of the network show the relative movements of the object points in respect to the reference block between the epochs.

ii. The network consists of two or more blocks, separated by geological features like faults or by structural properties like cracks or crevices. Points exhibiting non-typical deformations are eliminated from each block. One of the blocks is selected as the reference block and plays the same role as the reference network in paragraph (i).

iii. No reference block is defined. The whole network consists of object points only. Single points showing non-typical deformations are localized. Deformation vectors are plotted for all points. They refer to a geodetic datum defined by one of the methods discussed in Chapter 3.

In this stage of the analysis rigid body displacements are defined. These displacements can only be modelled in a relative sense. Thus for case (i) the displacement of the object points with respect to the reference block is estimable. Similarly, in case (ii), the movements of entire blocks relative to the reference block can be modelled and estimated. In case (iii) no rigid body displacement can be determined.

The model of a rigid body displacement is the same as for a similarity transformation which has already been used in Sections 3.8 and 10.4.5. Following Eq. (10-36) the model may be expressed as:

$$\Delta + \delta = Ht, \qquad \Sigma_\Delta = s_o^2 Q_\Delta \tag{10-51}$$

The vector Δ contains the "observations", i.e. the coordinate differences of the m points of the partial network under investigation. The residuals of the model form the vector δ. The parameter vector t comprises the two translation unknowns t_x and t_y, the rotation term r_z and the scale parameter s.

$$t = (t_x,\ t_y,\ r_z,\ s)^t \tag{10-52}$$

It should be noted that the use of a scale parameter in two epoch comparisons is rejected by many experts, and its role is questionable indeed. The design matrix H has already been given in the context of datum transformations (Eq. (3-61)). The coefficients $\bar{x}_i$ and $\bar{y}_i$ are the coordinates of the m points of the block to be analysed in a special coordinate system with its origin in the centre of gravity of the block:

$$H = \begin{pmatrix} 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 1 \\ -\bar{y}_1 & \bar{x}_1 & -\bar{y}_2 & \bar{x}_2 & \dots & -\bar{y}_m & \bar{x}_m \\ \bar{x}_1 & \bar{y}_1 & \bar{x}_2 & \bar{y}_2 & \dots & \bar{x}_m & \bar{y}_m \end{pmatrix}^t \tag{10-53}$$

The covariance matrix Σ_Δ results from the preceding combined adjustment of epochs i and j:

$$Q_\Delta = Q_{\hat{x}_i}^{\,o} + Q_{\hat{x}_j}^{\,o} - Q_{\hat{x}_i \hat{x}_j}^{\,o\ o} - Q_{\hat{x}_j \hat{x}_i}^{\,o\ o} \tag{10-54}$$

The superscript "o" refers to the object points as opposed to the reference points which are indicated by the superscript "r".

The parameter vector t is estimated by the usual least squares method:

$$\hat{t} = (H^t Q_\Delta^{-1} H)^{-1} H^t Q_\Delta^{-1} \Delta$$

$$Q_{\hat{t}} = (H^t Q_\Delta^{-1} H)^{-1} \qquad (10\text{-}55)$$

Alternatively, the estimation of $\hat{t}$ can be integrated in the combined adjustment which follows after the single epoch analyses. The model of Eq. (10-8) would then have to be modified in the following way:

$$\begin{pmatrix} l_i \\ l_j \end{pmatrix} + \begin{pmatrix} v_i \\ v_j \end{pmatrix} = \begin{pmatrix} A_i^r & A_i^o & 0 \\ A_j^r & A_j^o & -A_j^o H \end{pmatrix} \begin{pmatrix} \hat{x}^r \\ \hat{x}^o \\ \hat{t} \end{pmatrix} \qquad (10\text{-}56)$$

The extension to three and more blocks with different sets of transformation parameters $\hat{t}$ is straightforward. At this stage the question arises whether or not the parameter vector $\hat{t}$ is significant. The answer can be based on the usual F-test.

Under the null hypothesis:

$$H_o : E(\hat{t}) = t = 0$$

that the parameters are insignificant and assuming normal distribution of the "observations" Δ it follows from Theorem 1 of Section 5.4 that the quadratic form:

$$q_{\hat{t}} = \hat{t}^t Q_{\hat{t}}^{-1} \hat{t} \sim \sigma_o^2 \chi^2(4) \qquad (10\text{-}57)$$

has a central χ^2-distribution with 4 degrees of freedom.

Since $q_{\hat{t}}$ is independent of the quadratic form q of Eq. (10-10) the variance ratio:

$$T = \frac{0.25 q_{\hat{t}}}{q/f} \sim F(4, f) \qquad (10\text{-}58)$$

is an F-distributed test statistic. The probability of the test statistic exceeding a critical value $F_\alpha(4, f)$ is α. Upon completion of the estimation of the rigid body displacement a new plot of the remaining and not yet modelled deformation vectors is useful for the decision on further steps of the analysis. The question of whether or not the model applied so far suffices can be answered by inspection of the vector plot and by another statistical test.

The null hypothesis reads now:

H_o : All deformations not yet modelled are insignificant.

The numerator of the corresponding variance ratio test statistic:

$$T = \frac{q_\Delta / f_\Delta}{q/f} \sim F(f_\Delta, f) \qquad (10\text{-}59)$$

can be calculated either from the models of Eqs (10-56) or (10-51). Equation (10-56) can be regarded as a model which includes the hypothesis that all points are undeformed apart of the rigid body displacement of a block. Thus the quadratic form of the residuals of this model follows Eq. (10-12). The associated model without the application of the hypothesis is given in Eq. (10-19), thus:

$$\begin{aligned} q_\Delta &= q_{(10\text{-}56)} - q_{(10\text{-}19)} \\ f_\Delta &= f_{(10\text{-}56)} - f_{(10\text{-}19)} \end{aligned} \qquad (10\text{-}60)$$

On the other hand q_Δ may be calculated from the model of Eq. (10-51) applying the usual equations:

$$q_\Delta = \delta^t Q_\Delta^{-1} \delta, \qquad f_\Delta = 2m - 4 \qquad (10\text{-}61)$$

The denominator of Eq. (10-59) follows from Eq. (10-10) as before.

It should be pointed out that the numerous tests applied so far are not independent. The true type I error probability is therefore different from the selected value of α. Unfortunately, there is no rigorous statistical method to handle this problem. Thus the results of the tests should not be overvalued. The decisions for further analysis should also be based on experience and a careful assessment of the plotted vector field.

10.6 Strain Model

In numerous applications of deformation analyses and, particularly, in crustal movement studies, the final aim is a representation of the deformations in terms of strain parameters. The basic principles of strain analysis, as developed in the theory of elasticity, are applicable if the area or the object covered by the monitoring network can be considered as a continuum deforming as such under stress. Thus the deformation is continuous by definition. Since observations are only available at discrete points of the network, the unknown and usually rather involved deformation is to be interpolated over the entire object. This can be done either by continuous position dependent functions, fitted to the observations by least squares, or by locally continuous functions like straight lines and splines connecting the observed points. Considering again 2d-networks and neglecting the error term the general equation is:

$$\hat{x}_i - \hat{x}_j = \Delta = Ht + g(x) \qquad (10\text{-}62)$$

where Ht is the rigid body displacement of the previous section but without an unknown scale parameter. The function g(x) models the actual deformations. In all geodetic applications considered here, the deformations are small in comparison with the size of the network as soon as the rigid body displacements have been removed. Thus the deformations may be modelled by a differential relationship in the vicinity of the network points:

$$\Delta = Ht + Edx \qquad (10\text{-}63)$$

where:

$$E = \frac{\partial \Delta}{\partial x} = \begin{pmatrix} \frac{\partial \Delta x}{\partial x} & \frac{\partial \Delta x}{\partial y} \\ \frac{\partial \Delta y}{\partial x} & \frac{\partial \Delta y}{\partial y} \end{pmatrix} = \begin{pmatrix} e_{xx} & e_{xy} \\ e_{yx} & e_{yy} \end{pmatrix} \qquad (10\text{-}64)$$

is the infinitesimal **deformation tensor** with Δx, Δy being the components of the deformation vector Δ. Strictly, the Eq. (10-63) is valid only in a differential vicinity of the points. In order to apply the tensor in the entire area or object under investigation it is necessary to assume that the deformations are homogeneous, i.e. that g(x) is linear in x. In general, this is a very tight restriction which makes strain analysis unsuitable for many deformation problems. Fortunately it is possible to relax the restriction in practice by splitting the network into finite elements. The deformations of three points only are required for the solution of Eq. (10-63) hence the network can be decomposed into triangular elements. The requirement of homogeneity is then confined to the interior of the triangles.

The non-symmetric matrix E of Eq. (10-64) is usually decomposed into the sum of a symmetric matrix ε and a skew symmetric matrix ω:

$$E = 1/2(E + E^t) + 1/2(E - E^t) = \varepsilon + \omega \qquad (10\text{-}65)$$

with:

$$\varepsilon = \begin{pmatrix} e_{xx} & (e_{xy} + e_{yx})/2 \\ (e_{xy} + e_{yx})/2 & e_{yy} \end{pmatrix} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} \\ \varepsilon_{yx} & \varepsilon_{yy} \end{pmatrix} \qquad (10\text{-}66)$$

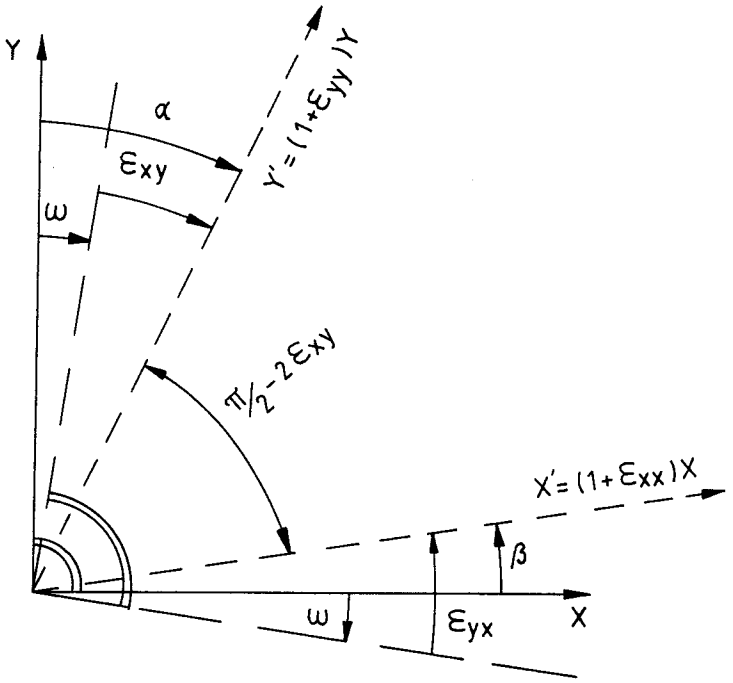
and:

$$\omega = \begin{pmatrix} 0 & (e_{xy} - e_{yx})/2 \\ (e_{yx} - e_{xy})/2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \omega_{xy} \\ \omega_{yx} & 0 \end{pmatrix} \qquad (10\text{-}67)$$

The matrix ε of Eq. (10-66) represents the so called **strain tensor**. The diagonal elements of ε give the **extensional strain** in x- and y-direction, respectively. The expression $1 + \varepsilon_{xx}$ $(1 + \varepsilon_{yy})$ is the scale factor applying to all lines parallel to the x-axis (y-axis). The double of the off-diagonal element $2\varepsilon_{xy} = 2\varepsilon_{yx}$ is equivalent to the angular distortion of a right angle which was originally parallel to the axes of the coordinate system. This effect is called **shear strain**.

The rigid body rotation is given by the angle ω_{xy} of Eq. (10-67). For the 2d-case discussed here a geometrical interpretation of the parameters of differential homogeneous strain is given in Figure 10.4.

Homogeneity means that straight lines remain straight and parallel lines remain parallel under the deformation model. The strain tensor is independent of translations but its elements refer to the coordinate system used. For further reading MARGRAVE and NYLAND (1980), CHEN (1983), BRUNNER et al (1981), SCHNEIDER (1982) and WELSCH (1981) are recommended.



X, Y Original System

X', Y' Deformed System

$\alpha = e_{xy}$

$\beta = e_{yx}$

$\omega = 1/2(\alpha - \beta)$

$\varepsilon_{xy} = 1/2(\alpha + \beta)$

Figure 10.4: Geometrical interpretation of the parameters of homogeneous strain.

Let R_φ be a rotation matrix, where φ is the angle between the original system (x, y) and the rotated system (ξ, η):

$$R_\varphi = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$$

The deformation vector $\bar{\Delta}$ of Eq. (10-63):

$$\bar{\Delta} = \Delta - Ht = Edx$$

takes the form of δ after rotation by φ:

$$\delta = R_\varphi \bar{\Delta} = R_\varphi E dx \qquad (10\text{-}68)$$

From $\xi = R_\varphi x$ follows $d\xi = R_\varphi dx$. Since R_φ is orthogonal, i.e. $R_\varphi{}^t R_\varphi = R_\varphi R_\varphi{}^t = I$, the above Equation (10-68) can be rewritten as:

$$\delta = R_\varphi E R_\varphi{}^t d\xi = E_\varphi d\xi$$

with:

$$E_\varphi = R_\varphi E R_\varphi{}^t \qquad (10\text{-}69)$$

The general relationship is thus established, where E_φ is the deformation tensor in the new system after rotation by the angle φ.

For any symmetrical matrix A the following eigenvalue decomposition (also refer to Section 4.1 and, particularly, Eq. (4-13)) exists:

$$A = S\Lambda S^t, \qquad S^tAS = \Lambda \tag{10-70}$$

with:

$$\Lambda = \text{diag}\,(\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_n)$$

being the diagonal matrix of eigenvalues and S the associated matrix of eigenvectors (rotation matrix). Substitution of Eq. (10-70) in Eq. (10-69) leads to:

$$E_\varphi = R_\varphi S\Lambda S^t R_\varphi{}^t \tag{10-71}$$

This demonstrates that the off-diagonal elements of E disappear and the diagonal elements coincide with the eigenvalues, if $R_\varphi = S^t$ is used, i.e. for a rotation by the angle Θ, as given by the eigenvalue decomposition of E:

$$E_\Theta = \Lambda = R_\Theta E R_\Theta{}^t \tag{10-72}$$

This transformation of E into its principal axes yields the semi-axes and the orientation of the **strain ellipse**, which may be plotted and interpreted as the error ellipse of Chapter 4. The axes are orthogonal and give the maximum and the minimum extensional strain. They have a bearing of Θ with respect to the original coordinate system. Since the shear strain is zero in the system rotated by Θ, a right angle with sides parallel to the axes of the strain ellipse remains undistorted in this model.

The linear deformation model of Eq. (10-63), which is usually applied for the estimation or computation of the parameters of homogeneous strain, can be applied to point P_k, after separation of the rigid body displacement:

$$\Delta_k - H_k t = \bar{\Delta}_k = \begin{pmatrix} \bar{\Delta}x \\ \bar{\Delta}y \end{pmatrix}_k = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}_k \tag{10-73}$$

Substitution of Eq. (10-65) yields:

$$\bar{\Delta}_k = \begin{pmatrix} \varepsilon_{xx}x_k + \varepsilon_{xy}y_k - \omega y_k \\ \varepsilon_{xy}x_k + \varepsilon_{yy}y_k + \omega x_k \end{pmatrix} \tag{10-74}$$

If the rigid body displacement shall be computed simultaneously with the strain parameters, then the two rotational terms merge and Eq. (10-74) needs only an extension by the translation unknowns t_x and t_y. This is the general approach to modelling deformations in terms of homogeneous strain, where six parameters (ε_{xx}, ε_{xy}, ε_{yy}, ω, t_x, t_y) are determined requiring a minimum of three points.

Depending on the nature of the object under investigation and on the structure of the deformation field it may be appropriate to compute either one set of parameters for the whole network, one set for each block or one set for each triangular finite element. Usually the result is transformed into the system of principal axes according to Eq. (10-72), and the strain ellipses or their axes are plotted in the network plan.

If more than one set of parameters is computed, it should be noted that the sets are correlated, since points on boundaries between blocks or between elements enter into two or more computations. When applying statistical tests of significance this fact would have to be considered.

10.7 Polynomial Deformation Models

A generalization of the rigid body and strain models leads to the polynomial approach. The contribution of point P_k to the model has the form:

$$\Delta_k = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}_k = \begin{pmatrix} a_o + a_1 x_k + a_2 y_k + a_3 x_k^2 + a_4 x_k y_k + a_5 y_k^2 + \ldots \\ b_o + b_1 x_k + b_2 y_k + b_3 x_k^2 + b_4 x_k y_k + b_5 y_k^2 + \ldots \end{pmatrix} \quad (10\text{-}75)$$

For $a_o = t_x$, $b_o = t_y$, $a_1 = b_2 = s$, $b_1 = -a_2 = r_z$, $a_i = b_i = 0$ for all $i > 2$ the rigid body displacement model of Eq. (10-51) results. If the parameters are selected as $a_o = t_x$, $b_o = t_y$, $a_1 = e_{xx}$, $b_1 = e_{yx}$, $a_2 = e_{xy}$, $b_2 = e_{yy}$, $a_i = b_i = 0$ for all $i > 2$, then the strain model of Eq. (10-63) is obtained.

For the parameter estimation in the model of Eq. (10-75) the usual least squares method can be employed (NEY, 1975).

As a rule, the use of a polynomial model of second or higher order should be questioned, since the parameters do not have a clear physical meaning. This makes the interpretation of the results difficult and does not lead to an understanding of the phenomena observed. Such models also contradict the principle of always selecting the simplest model possible. Furthermore, the well known drawbacks of higher order polynomials such as numerical instability and unwanted oscillations between data points apply as well.

Despite these severe deficiencies there are applications where a polynomial model is suitable. Altogether these special cases may even be in the majority. It should be kept in mind that deformation analysis typically deals with special cases, despite all generalizing approaches. Usually a model has to be tailored for a particular case. The models outlined in this section are by no means exhaustive; they may serve as examples on how to approach a practical task, where a special polynomial model may be appropriate.

Ground subsidence in areas where the mining industry exploits minerals or where oil, gas or water is pumped shows similar features. A cross-section through the area has the form of a trough. The lines of equal deformation (settlement) close, are of similar shape and resemble contour lines. The changes in height are usually greater than those in horizontal position. A polynomial of second or even higher order in a suitably selected local coordinate system may be most adequate.

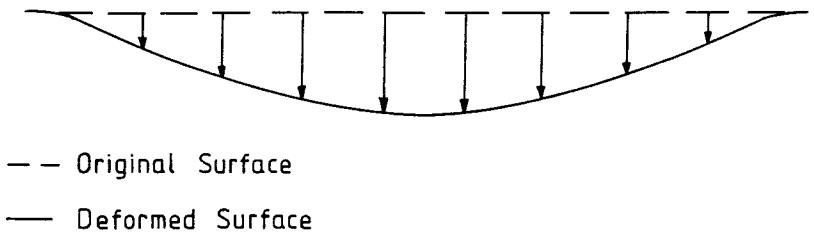


Figure 10.5: Cross-section of a subsidence trough.

Large concrete dams are usually crescent-shaped in plan view and are locked at both ends into the bedrock. The changes in water level and temperature cause regular deformations of the dam which can be approximated by a parabola. The deformation model has to consider this regularity. Only deformations deviating from the predicted values are of interest. If they are too large dangerous conditions may arise.

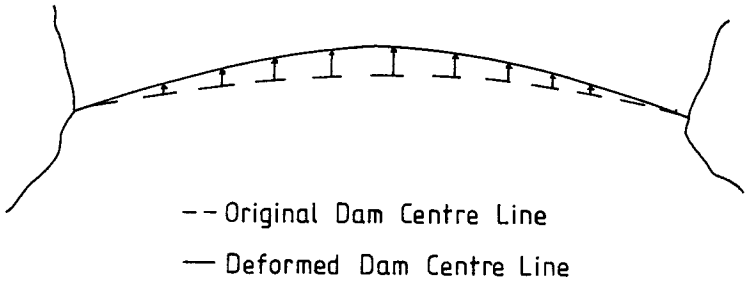


Figure 10.6: Sketch of a concrete dam with regular deformations.

Man-made slopes and dams of water reservoirs, large mining pits and along mountain roads are often hazards requiring permanent monitoring. Certain deformations follow known patterns and are non-critical. The general pattern of expected subsidence as depicted in Fig. 10.7 is similar to that of Fig. 10.5. The horizontal deformations resemble those of Fig. 10.6. Again, a polynomial model may be suitable to describe the pattern.

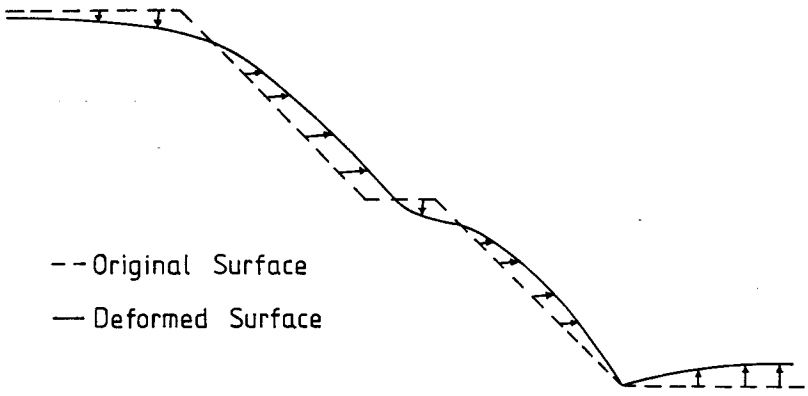


Figure 10.7: Bank of an open brown coal mining pit.

10.8 Graphical Representation of Deformation Patterns

It has been frequently mentioned in previous sections that a plot of vectors of position differences is a very helpful mean to obtain a first impression of existing deformations. Unfortunately, the vectors depend on the geodetic datum, so that their interpretation has to be done with caution. Nevertheless, in many practical applications these vectors represent the result, particularly where it is obvious at first glance that nothing has happened between the epochs which could give reason for concern. Some simple examples of such plots are depicted in the Figures 10.5 to 10.7.

A more sophisticated description of deformation vectors is outlined in Section 10.4.6. A display of the single point displacements together with conditional confidence ellipses (refer to Eqs (10-49) and (10-50)) conveys all available information of a two-epoch comparison in a very concise form. This facilitates the interpretation.

A different method of graphical representation is adequate for monitoring relatively fast movements as, for example, those of large glaciers. A plot of equally spaced lines across the glacier which have been straight in the first observation epoch is usually chosen in this case. These lines resembling contour lines give a fairly good picture of the ice flow. They are often preferred to a mathematic analysis based on a set of polynomial coefficients of unclear meaning.

Plots of lines of equal deformations have been the sole and entirely sufficient result of subsidence monitoring for many decades. These iso-lines are easily constructed from the observation results and have the advantage of being intuitively understood and interpreted by the user, who is not always accustomed to sophisticated mathematical models.

These examples and those of the previous section raise the question whether or not a mathematical model of deformations is really useful and necessary in all cases. It seems to be the trend of our times to mathematize all problems and to accept only abstract approaches as being scientific. The proven simple methods, as for example the graphical representation, are too often despised, albeit being the most suitable approaches in many applications.

10.9 Mixed Model Approach

All methods of deformation analysis discussed previously are based on the Gauss-Markov model, which can always be put into the form:

$$\Delta + \delta = By, \qquad \Sigma_\Delta = \Sigma_\delta = \sigma_o^2 Q_\Delta \tag{10-76}$$

where $\Delta = \hat{x}_i - \hat{x}_j$ denotes the vector of coordinate differences, δ the vector of residuals, B the design matrix and y the vector of unknown deformation parameters. The cofactor matrix of the "observations" Δ is given as:

$$Q_\Delta = Q_{\hat{x}_i} + Q_{\hat{x}_j} - Q_{\hat{x}_i\hat{x}_j} - Q_{\hat{x}_j\hat{x}_i}$$

and the variance factor σ_o^2 is estimated by the pooled variance according to Eq. (10-10). Examples are the model of a rigid body displacement (Eq. (10-51)) and the model of homogeneous strain (Eq. (10-63)).

The functional part By = E(Δ) of Eq. (10-76) can only be an approximation of the true and usually complicated deformation pattern. Therefore, it has to be assumed that the residual vector δ comprises observational errors as well as model deficiencies. Based on this aspect the mixed model approach into deformation analysis has been developed by SCHWINTZER (1982, 1984).

Let ε be the vector representing the random observational errors, characterized by:

$$E(\varepsilon) = 0, \qquad \Sigma_\varepsilon = \sigma_o^2 Q_\Delta \tag{10-77}$$

The effect of model deficiencies may be expressed as Cz, i.e. as linear combinations, defined by the matrix C, of some stochastic variables z. The residual vector δ can now be written as:

$$\delta = \varepsilon - Cz, \qquad \Sigma_\delta = \sigma_o^2 (Q_\Delta + C Q_z C^t) \tag{10-78}$$

where ε and z are assumed statistically independent. The magnitude of the term Cz depends on how well the linear function By describes reality. As simple models are usually preferred, in order to allow physically meaningful interpretations, the model's residual vector δ might be considerably greater than the observational error vector ε.

Nevertheless, it shall be assumed that the model is appropriate. The variables z can then be considered as random variables with zero mean. By virtue of the central limit theorem of statistics it is further assumed that z is normally distributed, hence being completely defined by the moments of first and second order:

$$E(z) = 0, \qquad \Sigma_z = \sigma_o^2 Q_z \tag{10-79}$$

These considerations lead to an expansion of model (10-76) into the mixed model defined as:

$$\begin{aligned} &\Delta + \varepsilon = By + Cz \\ &\varepsilon \sim N(0,\ \sigma_o^2 Q_\Delta), \qquad z \sim N(0,\ \sigma_o^2 Q_z) \end{aligned} \tag{10-80}$$

The deterministic (parameter) vector y and the stochastic vector z need to be estimated.

The estimators $\hat{y}$ of the parameters of the model of Eq. (10-80) are derived under the principle of best linear unbiased estimation (BLUE), as outlined in Chapter 2. For this purpose the stochastic terms of Eq. (10-80) are combined in the vector δ as shown in Eq. (10-78). This simplifies the mixed model to:

$$\Delta + \delta = By, \quad \Sigma_\delta = \sigma_o^2 (Q_\Delta + CQ_z C^t) = \sigma_o^2 Q_\delta \tag{10-81}$$

being equivalent to the usual GMM of Eq. (10-76) apart from the special structure of the cofactor matrix Q_δ.

The previously derived estimators can be adopted, if adjusted to the special covariance matrix of Eq. (10-81). The result is the following set of estimation equations for the mixed model:

$$\begin{aligned} \hat{y} &= (B^t P B)^{-1} B^t P \Delta, & \delta &= B\hat{y} - \Delta \\ Q_{\hat{y}} &= (B^t P B)^{-1}, & s_o^2 &= \delta^t P \delta / f \\ \hat{\varepsilon} = v &= -Q_\Delta P (\Delta - B\hat{y}), & \hat{z} &= Q_z C^t P (\Delta - B\hat{y}) \end{aligned} \tag{10-82}$$

with:

$$P = Q_\delta^{-1} = (Q_\Delta + CQ_z C^t)^{-1}$$

The application of the mixed model approach to deformation analysis requires two additional assumptions in order to set up the matrices C and Q_z of the model of Eq. (10-78).

Since, in practice, only a vague idea exists usually about the random variate z, the product Cz may as well be considered as the actual model deficiency. This admissable simplification makes it possible to put $C = I$ and leads to a new form of the weight matrix:

$$P = (Q_\Delta + Q_z)^{-1} \tag{10-83}$$

The prior determination of the cofactor matrix Q_z of the vector of random parameters z is comparable to the estimation of the a priori weight matrix of an ordinary GMM and, hence, affected by the same uncertainties. Usually, the observations as such do not contain sufficient information for a selection of Q_z. Therefore experience and reasonable guesses play an important role.

In most applications a smooth distance dependent correlation between the observed deformations is assumed. The correlation decreases more or less quickly and vanishes beyond a certain distance. It can be modelled by a simple positive definite function of the type:

$$r = \exp\{-(cd)^2\} \tag{10-84}$$

for example, where c is a parameter governing the steepness of the decrease of r and d defines the correlation length. Good results have also been achieved with a diagonal matrix for Q_z, i.e. for $r = 0$.

For the testing of the significance of the vector $\hat{y}$ of deterministic deformation parameters and for a separation of random observation errors and model residuals, the quadratic form $\delta^t P \delta$ of Eq. (10-82) must be decomposed. This is achieved by considering the following system of equations which is completely equivalent to Eq. (10-81):

$$\begin{aligned} \Delta + \varepsilon &= By + Cz, & \Sigma_\Delta &= \sigma_o^2 Q_\Delta \\ 0_z + \nu &= z, & \Sigma_z &= \sigma_o^2 Q_z \end{aligned} \tag{10-85}$$

The pseudo observation vector $0_z = E(z)$ consists of zero elements only. The parameter vector $(y^t z^t)$ may be estimated using the usual Gauss-Markov model, i.e. minimizing the quadratic form of the residuals. The analogy is obvious if Eq. (10-85) is written in the form:

$$\begin{pmatrix} \Delta \\ 0_z \end{pmatrix} + \begin{pmatrix} \varepsilon \\ \nu \end{pmatrix} = \begin{pmatrix} B & C \\ 0 & I \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}, \quad P = \begin{pmatrix} Q_\Delta^{-1} & 0 \\ 0 & Q_z^{-1} \end{pmatrix} \tag{10-86}$$

The minimized quadratic form:

$$q = \hat{\varepsilon}^t Q_\Delta^{-1} \hat{\varepsilon} + \hat{\nu}^t Q_z^{-1} \hat{\nu} \tag{10-87}$$

can now be decomposed in the following way:

$$\begin{aligned} q &= \Delta^t Q_\Delta^{-1} \Delta - \hat{\bar{y}}^t N_{\hat{y}} \hat{\bar{y}} - \hat{z}^t Q_{\hat{z}}^{-1} \hat{z} \\ &= q_1 - q_2 - q_3 \end{aligned} \tag{10-88}$$

where $N_{\hat{y}}$ and $Q_{\hat{z}}$ are the corresponding blocks of the normal equation matrix and its inverse, respectively, of the model of Eq. (10-86). The vector $\hat{\bar{y}}$ is the result of the model of Eq. (10-76), i.e. of the ordinary Gauss-Markov model without stochastic parameters. The subform q_3 indicates the reduction of the quadratic form q due to the use of the mixed model approach.

The degree of freedom of $\delta^t P \delta$ in the model of Eq. (10-86) is $f = n - u$, assuming full ranks of the n x u-matrix B and the n x n-matrix $P = Q_\Delta^{-1}$.

Equation (10-86) shows that this degree of freedom is distributed between the subvectors Δ and 0_Z of the total observation vector.

The deformation analysis using the mixed model consists of three consecutive steps:

i. **Variance component estimation:** the technique of variance component estimation as outlined in Chapter 8 is applied to the model of Eq. (10-86) in order to estimate a scaling factor for the matrix Q_Z. This is required for the correct relative weighing of the "observations" 0_Z with respect to the coordinate differences vector Δ. The cofactor matrix Q_Δ is assumed to be known. It is kept fixed in the estimation procedure. This step requires the computation of the redundancy contributions of the subvectors Δ and 0_Z of Eq. (10-86).

ii. **Outlier detection:** the estimated residuals $\hat{v}_i$ (being identical to the estimates $\hat{z}_i$) are screened for outliers. The associated cofactor matrix is computed from:

$$Q_{\hat{v}} = Q_Z - Q_{\hat{z}}$$

Outliers are located by applying one of the methods of Section 6.4 or 6.5. Detection of an outlier indicates that the corresponding point does not fit the adopted deformation model By. If a non-conformingly deformed point is identified, it is deleted from the model and the analysis is repeated starting again with step (i). This iterative outlier detection method is repeated until all remaining points conform with the model.

iii. **Significance testing:** the estimated parameters $\hat{y}$ are statistically tested for significance using the resulting cofactor matrix $Q_{\hat{y}}$ and the estimated variance factor s_0^2 of Eq. (10-82). The parameters identified as not contributing to the model are eliminated and, eventually, the final adjustment is carried out solely considering those parameters, which are relevant for the deformation model.

The mixed model approach stems from the sound principle that errors should be separated and treated according to their different sources, wherever this is possible. Unfortunately, the resulting equations are rather complicated and the numerical computations required are repellingly extensive and expensive. Therefore, despite its appeal from a theoretical point of view, this approach is unlikely to be used extensively.

10.10 Example: Dam Monitoring Network

The monitoring network Montsalvens serves again as a basis for an example of a two-epoch analysis. All observations and deformations have been simulated. Figure 10.8 shows the plan of the network and the observations, the 95%-confidence ellipses and the simulated deformations of points 3, 11, 12 and 13. The points 10 through 14 are located on the crest of the dam; they represent the object of investigation. The points 1, 2, 3, 4 and 6 form the reference network reinforced by the two targets 7 and 9. The geodetic datum is defined by minimizing the partial trace of the cofactor matrix corresponding to this group of reference points (see Section 3.8).

The network simulated for this example is a simplified version of the real network Montsalvens (Schweizerische Talsperrenkommission, 1946). The scale and the a priori variances are realistic. Tables 10.1 and 10.2 give the complete data of the two epochs of observation. The reader is encouraged to carry out his own analysis and to use the data for testing of the methods of the previous sections.

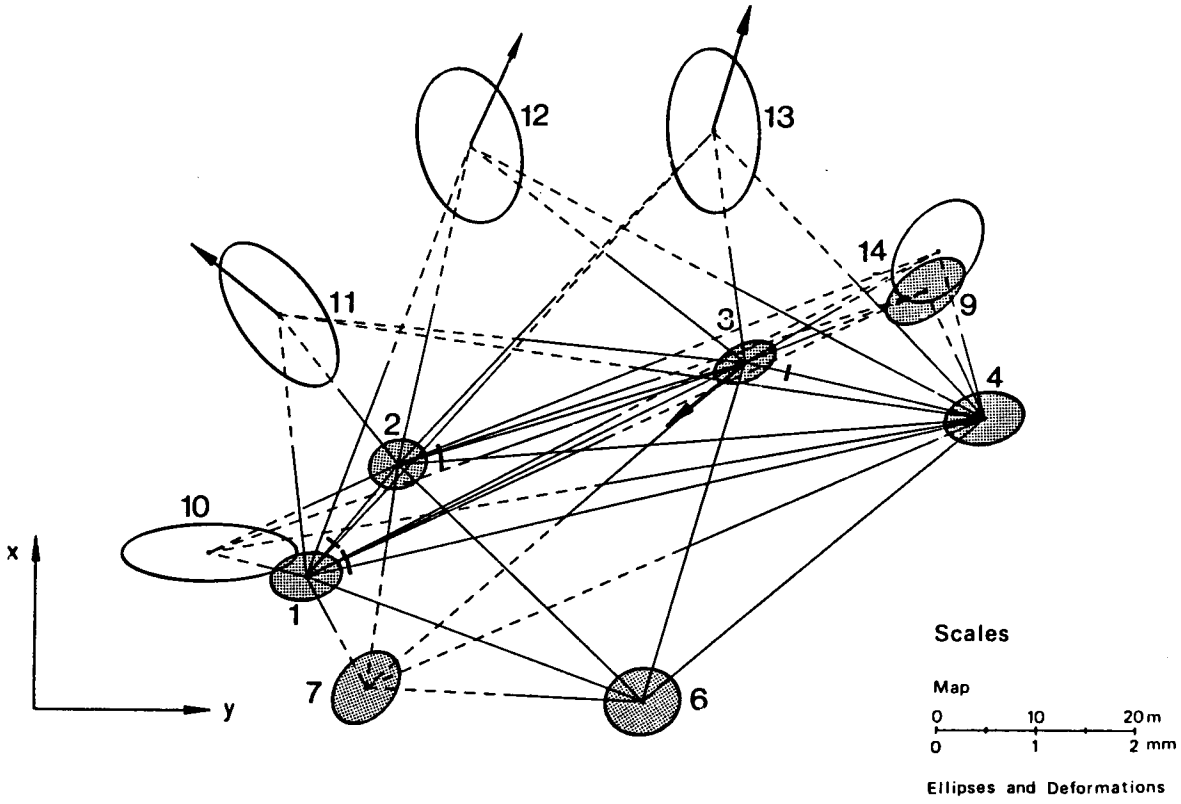


Figure 10.8: Monitoring Network Montsalvens showing the simulated observations and deformations together with the 95%-confidence ellipses. The geodetic datum is derived from the reference points (shaded ellipses).

The simulated observations consist of five arcs of directions with an a priori standard deviation of σ = 0.3mgon = 3^{cc} = 1" and six distances with σ = 0.3mm. The same GMM is valid for both epochs. It contains n = 55 observations and u = 29 unknowns (24 coordinates and 5 orientation unknowns). The rank deficiency of the model is d = 3, hence the geodetic datum has to dispose of three degrees of freedom: two translations and one rotation. This is achieved by minimizing that partial trace of the cofactor matrix of the unknowns, which refers to the coordinates of the reference points (numbers 1 – 9). The network has f = n − u + d = 29 redundancies. The dimensions for direction related quantities are either grad (400th part of a full circle) or centesimal seconds (1 grad = 10 000^{cc}). The length units used are metre and millimetre. Tables 10.1 and 10.2 list the input data and parts of the results of the two single epoch adjustments. The residuals v_i, Column (7), their standard deviations s_{v_i} and the test statistic T_i of Eq. (6-31) are tabled to facilitate a screening of the data for outliers. The critical value τ for a type I error probability of 5% and f = 29 is computed from Eq. (6-32) and yields τ = 3.09. The largest statistics T_i of epochs 1 and 2 are T_{20} = 2.385 and T_{37} = 2.856, respectively. Since $T_{20} < \tau$ and $T_{37} < \tau$ all observations are acceptable. Column (10) contains the redundancy contributions as defined by Eq. (7-15). The values indicate that the observations 9, 41 and 42 are hardly checked by other observations. Gross errors in these observations would be virtually undetectable. They would change the positions of points 10, 9 and 14 greatly, as a brief look at Fig. 10.8 will tell. This problem could only be overcome by selection of additional reference points.

No	from	to	"observed"	computed	l_i	weight	v_i	S_{v_i}	$T_i = v_i/S_{v_i}$	f_i
			directions in grad		cc		cc	cc		
1	2		3	4	5	6	7	8	9	10
1	1	2	0.00000	42.2524	0.27	0.111	0.863	2.363	0.365	0.5695
2	1	13	3.92813	46.1805	0.30	0.111	-1.558	2.459	0.634	0.6168
3	1	14	26.65768	68.9100	0.76	0.111	-1.277	2.769	0.461	0.7825
4	1	3	28.14169	70.3942	-1.08	0.111	1.230	2.839	0.433	0.8223
5	1	9	29.46106	71.7135	0.11	0.111	-1.590	2.786	0.571	0.7919
6	1	4	42.51533	84.7675	2.69	0.111	-2.959	2.754	1.074	0.7740
7	1	6	79.96743	122.2199	-0.49	0.111	0.190	2.398	0.079	0.5865
8	1	7	125.68087	167.9333	0.07	0.111	0.858	1.105	0.776	0.1247
9	1	10	272.63544	314.8878	0.19	0.111	0.064	0.164	0.389	0.0027
10	1	11	350.12485	392.3775	-1.91	0.111	1.961	1.365	1.437	0.1900
11	1	12	379.70302	21.9555	-0.91	0.111	2.218	2.291	0.968	0.5355
12	2	1	0.00000	242.2524	2.82	0.111	-1.494	1.996	0.749	0.4066
13	2	10	28.95076	271.2038	-3.87	0.111	1.264	0.945	1.338	0.0911
14	2	11	114.46143	356.7139	2.37	0.111	-1.935	1.365	1.418	0.1901
15	2	12	170.86868	13.1211	2.53	0.111	-1.951	1.712	1.140	0.2990
16	2	13	205.16361	47.4165	-2.54	0.111	0.757	2.030	0.373	0.4203
17	2	14	233.14019	75.3929	-0.43	0.111	-0.082	2.684	0.030	0.7348
18	2	9	236.90870	79.1612	2.02	0.111	-3.723	2.717	1.370	0.7534
19	2	3	239.06410	81.3169	-1.35	0.111	1.715	2.695	0.636	0.7408
20	2	4	252.26800	94.5213	-6.44	0.111	6.458	2.708	2.385	0.7481
21	2	6	306.12031	148.3730	-0.63	0.111	0.031	2.174	0.014	0.4824
22	2	7	365.52605	207.7782	5.52	0.111	-1.040	1.916	0.543	0.3745
23	3	1	0.00000	270.3942	-0.90	0.111	2.345	2.715	0.864	0.7521
24	3	10	7.24369	277.6373	5.08	0.111	-4.615	2.722	1.696	0.7557
25	3	2	10.92332	281.3169	5.39	0.111	-3.913	2.671	1.465	0.7278
26	3	11	35.65047	306.0446	-0.27	0.111	2.844	2.049	1.388	0.4284
27	3	12	71.45038	341.8449	-4.35	0.111	1.962	1.557	1.260	0.2475
28	3	13	119.97324	390.3675	-1.70	0.111	0.381	1.396	0.273	0.1987
29	3	14	195.22762	65.6219	-1.11	0.111	0.434	1.251	0.347	0.1597
30	3	9	204.68251	75.0772	-6.14	0.111	1.730	1.133	1.527	0.1309
31	3	4	243.91905	114.3133	-0.88	0.111	1.642	1.922	0.855	0.3767
32	3	6	347.23966	217.6338	0.13	0.111	-1.629	2.128	0.765	0.4621
33	3	7	383.60808	254.0017	4.74	0.111	-1.181	2.577	0.458	0.6774
34	4	1	0.00000	284.7675	-2.45	0.111	2.255	2.825	0.798	0.8141
35	4	10	3.71152	288.4789	-1.86	0.111	1.217	2.835	0.429	0.8203
36	4	2	9.75390	294.5213	-1.82	0.111	1.728	2.848	0.607	0.8277
37	4	11	24.10203	308.8692	1.10	0.111	-0.413	2.596	0.159	0.6877
38	4	3	29.54640	314.3133	3.70	0.111	-4.159	2.501	1.663	0.6384
39	4	12	45.87412	330.6413	0.13	0.111	-2.629	2.413	1.089	0.5941
40	4	13	66.81895	351.5862	0.26	0.111	-0.482	1.829	0.264	0.3412
41	4	9	86.84505	371.6120	3.32	0.111	-0.044	0.058	0.756	0.0003
42	4	14	97.16398	381.9311	1.31	0.111	-0.021	0.127	0.163	0.0017
43	4	6	370.06912	254.8368	-4.32	0.111	1.881	2.447	0.769	0.6108
44	4	7	388.29153	273.0587	0.64	0.111	0.667	2.710	0.246	0.7493
45	6	1	0.00000	322.2199	2.66	0.111	-2.436	2.405	1.013	0.5902
46	6	7	380.49429	302.7141	3.02	0.111	1.300	1.798	0.723	0.3297
47	6	4	132.61652	54.8368	-1.04	0.111	-0.958	2.094	0.457	0.4475
48	6	3	95.41328	17.6338	-3.31	0.111	1.031	2.276	0.453	0.5284
49	6	2	26.15275	348.3730	-1.33	0.111	1.062	2.379	0.447	0.5773
			Distances in m		mm		mm	mm		
50	1	2	14.5970	14.5966	0.42	11.111	-0.428	0.293	1.461	0.8744
51	1	3	49.2300	49.2302	-0.22	11.111	0.078	0.265	0.294	0.7194
52	1	4	69.9970	69.9972	-0.16	11.111	-0.064	0.244	0.262	0.6055
53	2	3	36.5730	36.5737	-0.71	11.111	0.569	0.281	2.025	0.8064
54	2	4	59.2300	59.2302	-0.20	11.111	-0.047	0.259	0.181	0.6828
55	3	4	24.6210	24.6207	0.34	11.111	-0.454	0.280	1.623	0.7980

Table 10.1: Input data of the GMM of **Epoch 1** and part of the results. Column (5) is defined as (5) = (4) - (3) + (orientation). The weights are based on $\sigma_0 = 1$. Columns (7) and (8) give the residuals after the adjustment and their standard deviations, respectively. POPE's test statistic T_i of Eq. (6-31) is compiled in Column (9). Column 10 shows the redundancy contributions according to Eq. (7-15).

No	from	to	"observed"	computed	l_i	weight	v_i	S_{v_i}	$T_i = v_i/S_{v_i}$	f_i
			directions in grad		cc		cc	cc		
1	2		3	4	5	6	7	8	9	10
1	1	2	0.00000	42.2524	1.43	0.111	1.935	2.319	0.835	0.5695
2	1	13	3.92707	46.1805	-9.15	0.111	3.482	2.413	1.443	0.6168
3	1	14	26.65771	68.9100	2.21	0.111	-0.239	2.718	0.088	0.7825
4	1	3	28.14160	70.3942	-0.83	0.111	4.972	2.786	1.785	0.8223
5	1	9	29.46127	71.7135	3.36	0.111	-1.456	2.734	0.532	0.7919
6	1	4	42.51552	84.7675	5.75	0.111	-3.578	2.703	1.324	0.7740
7	1	6	79.96779	122.2199	4.27	0.111	0.426	2.353	0.181	0.5875
8	1	7	125.68102	167.9333	2.72	0.111	-0.766	1.085	0.706	0.1247
9	1	10	272.63537	314.8878	0.65	0.111	0.120	0.161	0.744	0.0027
10	1	11	350.12351	392.3775	-14.15	0.111	-2.568	1.339	1.918	0.1900
11	1	12	379.70337	21.9555	3.74	0.111	-2.329	2.248	1.036	0.5355
12	2	1	0.00000	242.2524	5.43	0.111	-2.696	1.959	1.376	0.4066
13	2	10	28.95084	271.2038	-0.46	0.111	0.840	0.927	0.906	0.0911
14	2	11	114.45988	356.7139	-10.52	0.111	2.666	1.340	1.990	0.1901
15	2	12	170.86840	13.1211	2.34	0.111	1.922	1.680	1.144	0.2990
16	2	13	205.16268	47.4165	-9.23	0.111	-0.311	1.992	0.156	0.4203
17	2	14	233.14009	75.3929	1.18	0.111	-0.812	2.634	0.308	0.7348
18	2	9	236.90861	79.1612	3.73	0.111	-3.394	2.667	1.273	0.7534
19	2	3	239.06490	81.3169	9.26	0.111	-3.955	2.644	1.496	0.7408
20	2	4	252.26813	94.5213	-2.53	0.111	3.061	2.658	1.152	0.7481
21	2	6	306.12038	148.3730	2.68	0.111	0.909	2.134	0.426	0.4824
22	2	7	365.52505	207.7782	-1.87	0.111	1.770	1.880	0.941	0.3745
23	3	1	0.00000	270.3942	-4.86	0.111	2.141	2.665	0.803	0.7521
24	3	10	7.24348	277.6373	-0.97	0.111	-2.023	2.671	0.757	0.7557
25	3	2	10.92269	281.3169	-4.86	0.111	3.945	2.621	1.505	0.7278
26	3	11	35.65159	306.0446	6.98	0.111	1.202	2.011	0.598	0.4284
27	3	12	71.45451	341.8449	33.00	0.111	-1.314	1.528	0.860	0.2475
28	3	13	119.97659	390.3675	27.84	0.111	-2.089	1.370	1.525	0.1987
29	3	14	195.22732	65.6219	-8.07	0.111	0.393	1.228	0.320	0.1597
30	3	9	204.68209	75.0772	-14.29	0.111	1.579	1.112	1.420	0.1309
31	3	4	243.91734	114.3133	-21.93	0.111	-0.018	1.886	0.009	0.3767
32	3	6	347.23917	217.6338	-8.73	0.111	-1.760	2.089	0.843	0.4621
33	3	7	383.60759	254.0017	-4.11	0.111	-2.056	2.529	0.813	0.6774
34	4	1	0.00000	284.7675	-3.70	0.111	2.049	2.772	0.793	0.8141
35	4	10	3.71172	288.4789	-1.11	0.111	-1.166	2.783	0.419	0.8203
36	4	2	9.75375	294.5213	-4.57	0.111	1.920	2.795	0.687	0.8277
37	4	11	24.10290	308.8692	8.56	0.111	-7.277	2.548	2.856	0.6877
38	4	3	29.54442	314.3133	-17.34	0.111	-1.559	2.455	0.635	0.6384
39	4	12	45.87526	330.6413	10.29	0.111	1.607	2.368	0.679	0.5941
40	4	13	66.82025	351.5862	12.01	0.111	2.707	1.795	1.508	0.3412
41	4	9	86.84463	371.6120	-2.13	0.111	-0.040	0.056	0.701	0.0003
42	4	14	97.16418	381.9311	2.06	0.111	-0.042	0.125	0.335	0.0017
43	4	6	370.06951	254.8368	-1.67	0.111	1.459	2.401	0.608	0.6108
44	4	7	388.29135	273.0587	-2.41	0.111	0.342	2.660	0.129	0.7493
45	6	1	0.00000	322.2199	2.30	0.111	-0.006	2.360	0.003	0.5902
46	6	7	380.49413	302.7141	1.06	0.111	-0.342	1.764	0.194	0.3297
47	6	4	132.61674	54.8368	0.80	0.111	0.400	2.055	0.195	0.4475
48	6	3	95.41305	17.6338	-5.97	0.111	-0.061	2.233	0.027	0.5284
49	6	2	26.15310	348.3730	1.81	0.111	0.009	2.335	0.004	0.5773
			Distances in m		mm		mm	mm		
50	1	2	14.5960	14.5966	-0.58	11.111	0.414	0.287	1.442	0.8744
51	1	3	49.2290	49.2302	-1.22	11.111	0.145	0.260	0.557	0.7194
52	1	4	69.9970	69.9972	-0.16	11.111	-0.130	0.239	0.545	0.6055
53	2	3	36.5730	36.5737	-0.71	11.111	-0.197	0.276	0.715	0.8064
54	2	4	59.2300	59.2302	-0.20	11.111	-0.001	0.254	0.004	0.6828
55	3	4	24.6210	24.6207	0.34	11.111	0.130	0.274	0.473	0.7980

Table 10.2: Input data of the GMM of **Epoch 2** and part of the results. Column (5) is defined as (5) = (4) - (3) + (orientation). The weights are based on $\sigma_0 = 1$. Columns (7) and (8) give the residuals after the adjustment and their standard deviations, respectively. POPE's test statistic T_i of Eq. (6-31) is compiled in Column (9). Column 10 shows the redundancy contributions according to Eq. (7-15).

No	Approximate Coordinates X	Y	Simulated Deformations ΔX	ΔY	Eigen values λ of $Q_{\hat{x}}$	
1	100.1030	100.0110	0.00	0.00	0.0000	0.0010
2	111.6010	109.0030	0.00	0.00	0.0018	0.0040
3	122.1810	144.0130	-0.50	-0.60	0.0066	0.0107
4	116.6920	168.0140	0.00	0.00	0.0054	0.0086
6	87.6610	134.1990	0.00	0.00	0.0324	0.0070
7	88.8550	106.2100	0.00	0.00	0.0523	0.0150
9	129.5510	161.8670	0.00	0.00	0.0010	0.0020
10	102.4480	90.1670	0.00	0.00	0.0000	0.0021
11	126.6760	96.8140	0.60	-0.75	0.0170	0.0000
12	143.9770	115.7720	1.10	0.50	0.0029	0.1291
13	145.6870	140.4290	1.00	0.30	0.0358	0.0219
14	133.6100	163.0790	0.00	0.00	0.0294	0.0102

$tr Q_{\hat{x}}$ = 0.396 f = 29; P = 0.95 → F(2, 29) = 3.33, τ(29) = 3.09

No	Adjusted Coordinates X	Y	Parameter dX	dY	Standard Deviation S_x	S_y	S_p	0.95 - Conf. Ellipses a	b	α
1*	100.1030	100.0111	0.00	0.08	0.07	0.10	0.12	0.37	0.24	87.45
2*	111.6010	109.0031	-0.02	0.10	0.07	0.08	0.11	0.30	0.25	78.84
3*	122.1809	144.0130	-0.10	-0.03	0.06	0.09	0.11	0.34	0.19	70.70
4*	116.6919	168.0139	-0.07	-0.14	0.07	0.11	0.13	0.41	0.27	89.02
6*	87.6610	134.1991	-0.03	0.14	0.09	0.11	0.14	0.39	0.34	85.53
7*	88.8552	106.2100	0.17	-0.04	0.10	0.09	0.14	0.42	0.29	42.39
9*	129.5510	161.8669	0.05	-0.10	0.10	0.11	0.14	0.46	0.25	57.17
10	102.4480	90.1673	-0.03	0.26	0.08	0.25	0.26	0.90	0.29	99.18
11	126.6760	96.8141	0.02	0.09	0.20	0.17	0.26	0.87	0.41	159.54
12	143.9766	115.7721	-0.40	0.06	0.22	0.15	0.27	0.81	0.52	183.62
13	145.6869	140.4289	-0.09	-0.11	0.23	0.13	0.27	0.85	0.47	195.99
14	133.6099	163.0789	-0.07	-0.09	0.15	0.13	0.19	0.60	0.38	39.24

Epoch 1: s_0 = 1.043

1*	100.1032	100.0113	0.15	0.27	0.07	0.10	0.12	0.36	0.24	87.45
2*	111.6010	109.0032	0.00	0.20	0.07	0.08	0.10	0.29	0.25	78.84
3*	122.1805	144.0124	-0.49	-0.61	0.06	0.08	0.10	0.33	0.18	70.70
4*	116.6920	168.0140	0.04	-0.01	0.07	0.11	0.13	0.40	0.26	89.02
6*	87.6611	134.1989	0.12	-0.12	0.09	0.10	0.14	0.38	0.34	85.53
7*	88.8552	106.2103	0.16	0.26	0.10	0.09	0.14	0.41	0.28	42.39
9*	129.5510	161.8670	0.01	0.00	0.09	0.11	0.14	0.45	0.24	57.17
10	102.4481	90.1674	0.09	0.44	0.08	0.24	0.26	0.89	0.29	99.18
11	126.6765	96.8134	0.50	-0.56	0.20	0.16	0.26	0.85	0.40	159.54
12	143.9782	115.7726	1.18	0.61	0.21	0.15	0.26	0.80	0.51	183.62
13	145.6883	140.4293	1.26	0.29	0.23	0.13	0.26	0.84	0.46	195.99
14	133.6100	163.0791	0.05	0.11	0.14	0.12	0.19	0.59	0.37	39.24

Epoch 2: s_0 = 1.024

Pooled Variance: s_0^2 = 1.067, f = 58

m	f_Δ	q_Δ	T	F(11, 58)	
7	11	54.131	4.608	1.944	T > F → rejection

Decomposition of q_Δ (Sect 10.4.2):

q_Δ^1	q_Δ^2	q_Δ^3	q_Δ^4	q_Δ^6	q_Δ^7	q_Δ^9	
1.46	4.13	43.52	16.12	1.10	2.10	0.02	Point 3 unstable

m^1	f_Δ^1	q_Δ^1	T^1	F(9, 58)	
6	9	10.611	1.104	2.030	T < F → no rejection

Table 10.3: Adjustment results, measures of accuracy for both epochs and single point diagnosis in the reference network: Point 3 is unstable (P = 0.95). Datum points are marked with an asterisk.

The approximate coordinates of the stations, the simulated deformations between the epochs and the eigenvalues of the cofactor matrix of the adjusted coordinates are given in the upper part of Table 10.3. For a degree of freedom of f = 29 and a probability of P = 95% the value of the F-distribution for the scaling of the confidence ellipses is F = 3.33 (refer to Eq. (4-42)).

The middle part of Table 10.3 contains the adjusted coordinates, the estimated parameters (differences between approximate and final coordinates), the standard deviations of the coordinates and the mean square positional errors (Eq. 4-46) for both epochs separately. The last three columns give the axes and the orientation of the 95%-confidence ellipses according to Eq. (4-42). Since both a posteriori variances s_0^2 are close to the a priori value $\sigma_0^2 = 1$ the results can be adopted without a global model test.

The lower part of Table 10.3 lists the results of the single point diagnoses for the reference part of the monitoring network. The hypothesis of Eq. (10-18) is tested for m = 7 reference points, where q_Δ has been computed from Eq. (10-17). The test statistic of Eq. (10-14) exceeds the critical value F(11, 58). The test fails therefore which indicates that not all points are stable. The decomposition of q_Δ is carried out along the lines of Section 10.4.2. The partial form q_Δ^3 of point 3 is by far the largest. Hence, it is concluded that this point is unstable. It is removed from the group of reference points. With the remaining m = 6 reference points, the reduced quadratic form $q_\Delta' = q_\Delta - q_\Delta^3$ and $f_\Delta' = f_\Delta - 2$ are used to compute the new test statistic T'. As the statistic is smaller than the new critical value F(9, 58) the hypothesis, that the reduced reference network contains only stable points, is accepted with an approximate type I error probability of 5%.

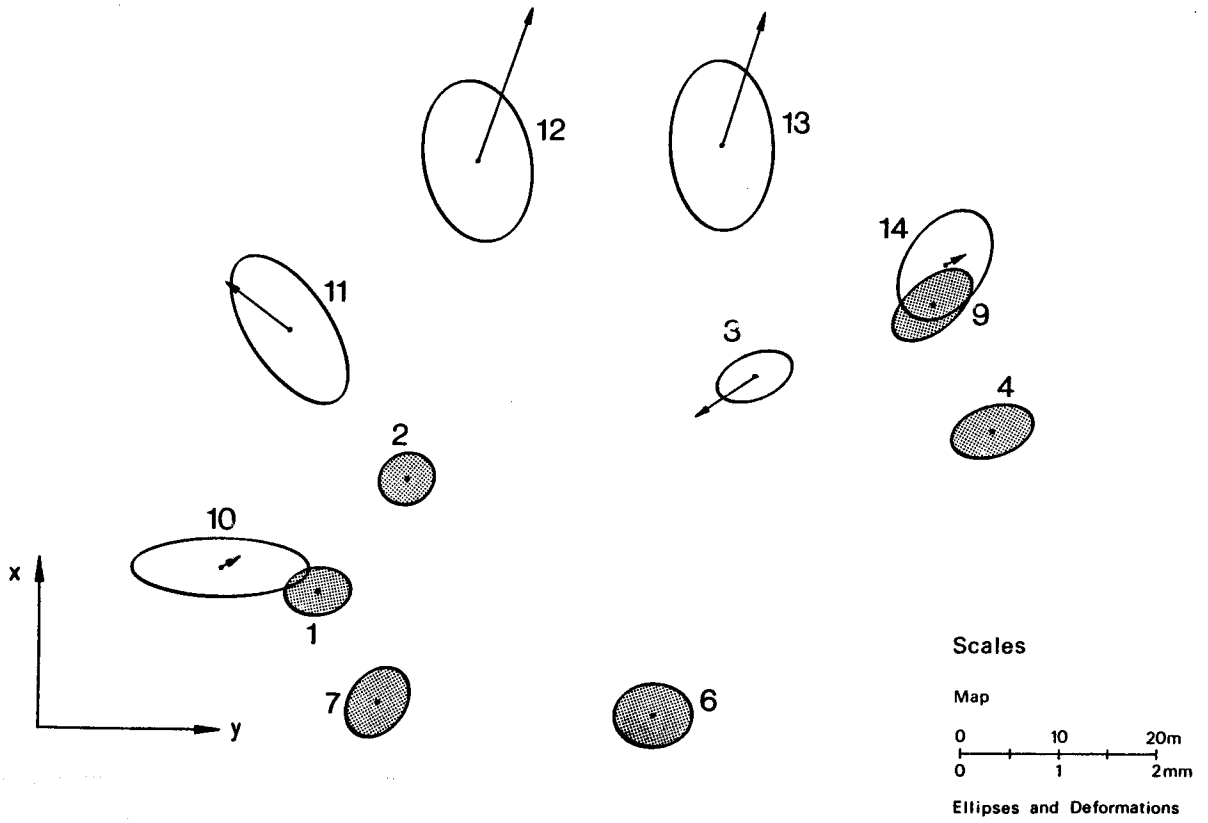


Figure 10.9: Monitoring Network Montsalvens showing the estimated deformations together with their 95%-confidence ellipses. The geodetic datum is derived from the stable reference points (shaded).

The reference network for the analysis of the object points is now established. It comprises the points 1, 2, 4, 6, 7 and 9. Since point 3 belongs to the group of points, which has been used to define the geodetic datum of the previous single epoch adjustments, these adjustments must be repeated with a new geodetic datum based on the stable reference points only. It is not necessary to actually carry out these adjustments, as the transition to the new datum can be achieved by a simple S-transformation, as defined by Eqs (3-53), (3-54) and (3-55). The adjustment results, referring to the new datum, are given in Figure 10.9 and Table 10.4. The differences with respect to the previous results are very small in this particular case. This must not lead to the wrong conclusion, that the transformation to the new datum is generally unnecessary.

	Differences 2nd Adjustment		Differences 3rd Adjustment		Standard Deviation	
No.	ΔX [mm]	ΔY [mm]	ΔX [mm]	ΔY [mm]	$S_{\Delta X}$ [mm]	$S_{\Delta Y}$ [mm]
1	2	3	4	5	6	7
1*	+0.15	+0.19	-	-	-	-
2*	+0.03	+0.09	-	-	-	-
3	−0.39	−0.58	−0.43	−0.61	0.08	0.13
4*	+0.11	+0.13	-	-	-	-
6*	+0.14	−0.25	-	-	-	-
7*	−0.01	+0.31	-	-	-	-
9*	−0.03	+0.11	-	-	-	-
10	+0.13	+0.18	−0.00	−0.15	0.07	0.25
11	+0.48	−0.65	+0.52	−0.83	0.25	0.19
12	+1.58	+0.54	+1.56	+0.44	0.28	0.19
13	+1.34	+0.40	+1.29	+0.34	0.29	0.19
14	+0.12	+0.20	+0.04	+0.08	0.18	0.13

Table 10.4: Coordinate differences between epoch 1 and 2 relative to the confirmed reference points (marked by an asterisk) and coordinate differences and their standard deviations of the object points after the 3rd (final) adjustment.

A new test of congruency of the reference network, relative to the new datum, confirms the previous result, that the points marked by an asterisk are stable. The corresponding test statistic T = 1.101 is less than the critical value F = 2.030. The plot of the deformation vectors of the reference points (Fig. 10.9) does not indicate any systematic effects exceeding the range of the uncertainty of point determination.

The final (3rd) adjustment combines the epochs in such a way that the reference points get identical coordinates and the object points different coordinates in both epochs. The resulting coordinate differences and the corresponding standard deviations are listed in Columns (4) to (7) of Table 10.4. These differences are the final estimates of the deformations of the object points. This third adjustment has not been actually carried out. The transformation as defined by Eqs (10-24a) and (10-26a) has been applied instead yielding the required result as demonstrated in Section 10.4.6.

A comparison of the estimated deformations with values predicted by an engineering model or other sources of information would follow as a next step. If the discrepancies are small, then the analysis can be regarded as completed. In uncertain cases the differences between estimated and predicted deformations can be considered as "deformations" requiring further analysis. These quantities then replace the entries of Columns (4) and (5) of Table 10.4 and the cofactor matrix computed in the 3rd adjustment is allocated to this difference vector. A global congruency test and a subsequent single point diagnosis along the lines of Sections 10.3 and 10.4 may clarify the question.

Based on the results of the single epoch adjustments of Table 10.3 another deformation analysis has been carried out applying the robust method of Section 10.4.5. This has been done to demonstrate the power of this approach. The model of Eq. (10-36) has been defined for the reference points only, namely points 1 to 9. The results are given in Table 10.5 and have been obtained after $\nu = 19$ iterations of the algorithm of Eq. (10-45). The residual deformations:

$$\delta = H\hat{t} - \Delta$$

(see Eq. (10-36)) can be compared with the differences of Table 10.4 and with the simulated deformations given in the upper part of Table 10.3. The results of both approaches are satisfactory; the robust method required, however, considerably less computations.

No.	Robust Method		Standard Deviation Simulated		Standard Deviation before Adjustments		Statistic
	δX [mm]	δY [mm]	$S_{\delta X}$ [mm]	$S_{\delta Y}$ [mm]	$S_{\Delta X}$ [mm]	$S_{\Delta Y}$ [mm]	$\delta^t Q_\delta^{-1} \delta$ [mm^2]
1	2	3	4	5	6	7	8
1*	+0.07	+0.04	0.08	0.05	0.10	0.14	1.47
2*	−0.03	−0.05	0.10	0.08	0.11	0.11	0.52
3*	−0.40	−0.68	0.08	0.14	0.08	0.12	43.56
4*	+0.11	+0.07	0.14	0.09	0.10	0.16	1.07
6*	+0.08	−0.34	0.12	0.20	0.13	0.15	3.64
7*	−0.10	+0.18	0.13	0.10	0.14	0.13	3.29
9*	−0.01	+0.02	0.08	0.08	0.13	0.16	0.09
10	+0.04	+0.01	0.11	0.32	0.11	0.36	0.18
11	+0.44	−0.83	0.28	0.19	0.28	0.23	20.71
12	+1.58	+0.38	0.27	0.20	0.30	0.21	35.56
13	+1.37	+0.28	0.28	0.18	0.33	0.18	28.18
14	+0.14	+0.12	0.19	0.15	0.21	0.18	1.10

Table 10.5: Results of the robust deformation analysis. Columns (2) and (3) give the final deformations, Columns (4) and (5) the standard deviations estimated from 50 repetitions of the simulation and the Columns (6) and (7) the standard deviations from the single epoch adjustments. The quadratic form of Column (8) is used as significance indicator.

The deformations resulting from the robust method in a single step are listed in Columns (2) and (3) of Table 10.5. In order to get a reliable estimate of the covariance matrix of the deformations, the computations have been repeated with 50 sets of simulated normally distributed observations. The resulting standard deviations, compiled in Columns (4) and (5), differ only slightly from the corresponding values in Columns (6) and (7) which are derived from the single epochs for the difference vector Δ. The statement of Section 10.4.5 that the cofactor matrix Q_Δ can be considered as an estimate of Q_δ is thus verified. The quadratic forms according to Eq. (10-46) are listed in Column (8). They may serve as indicators for significant deformations. An approximate critical value can be derived from Eq. (10-48). With a degree of freedom of $f = 58$ of the pooled variance $s_o^2 = 1.067$, and an overall type I error probability of $\alpha = 5\%$ leading to:

$$\bar{\alpha} = 1-(1-\alpha)^{1/12} = 0.43\%$$

the critical value becomes:

$$(\delta^t Q_\delta^{-1}\delta) = 2s_o^2 F_{\bar{\alpha}}(2, 58) = 12.87$$

Column (8) leads to the distinct result that points 3, 11, 12 and 13 are significantly deformed while all others remained stable within the range of the accuracy of the observations. This result is emphasized by the plot of the deformation vectors together with the 95%-confidence ellipses in Figure 10.10.

The final results of this example are reviewed in Table 10.6. The simulated deformations together with the estimated values of the least squares and the robust approach show that the employed methods of analysis work reliably. When assessing the differences between the methods their standard deviations have to be considered. These amount to approximately square root two times the values given in Columns (8) and (9).

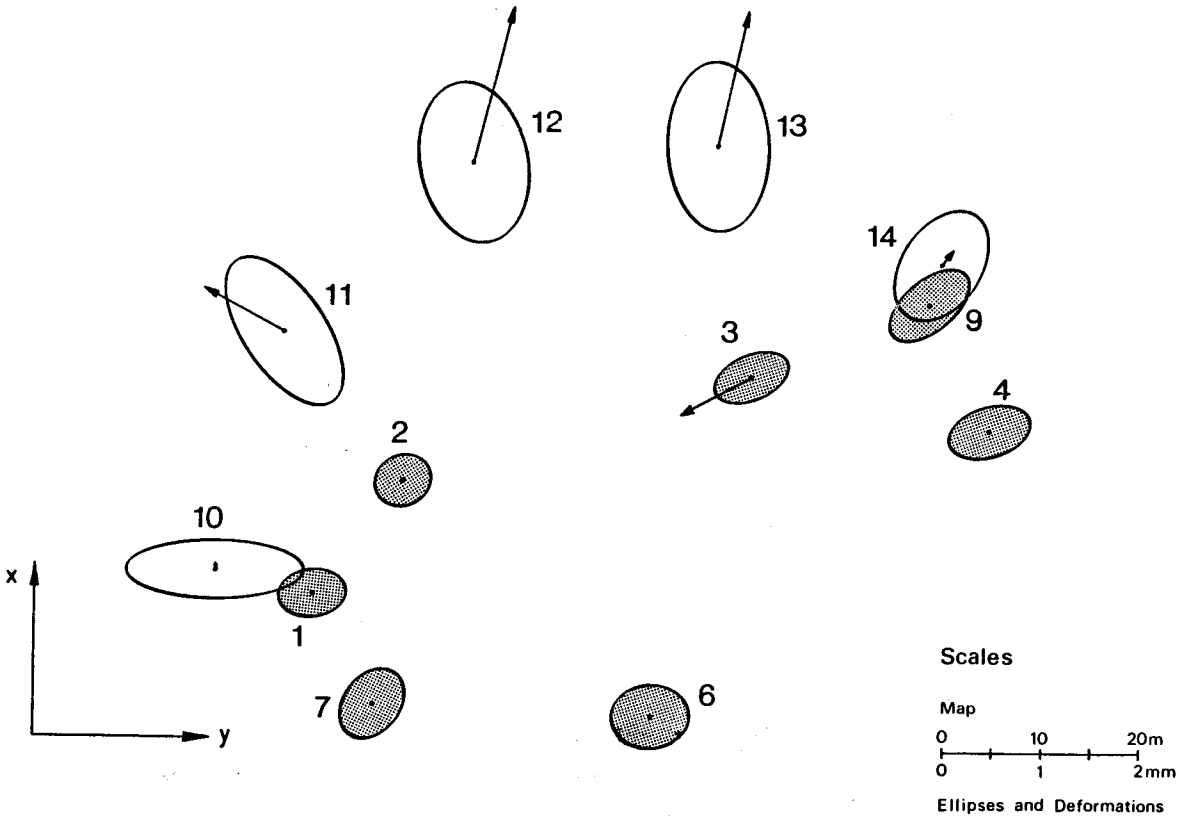


Figure 10.10: Plot of the deformations estimated by the robust method of Section 10.4.5. The 95%-confidence ellipses are based on the single epoch adjustments. The datum of the robust adjustment is based on the shaded points.

No.	Simulated Deformations		Conventional Estimates		Robust Estimates		95%-Confidence Intervals of Estimated Deformations	
	dX	dY	ΔX	ΔY	δX	δY	Δx	Δy
1	2	3	4	5	6	7	8	9
1	–	–	–	–	+0.07	+0.04	0.20	0.26
2	–	–	–	–	−0.03	−0.05	0.20	0.23
3	−0.50	−0.60	−0.43	−0.61	−0.40	−0.68	0.20	0.29
4	–	–	–	–	+0.11	+0.07	0.20	0.31
6	–	–	–	–	+0.08	−0.34	0.26	0.31
7	–	–	–	–	−0.10	+0.18	0.29	0.26
9	–	–	–	–	−0.01	+0.02	0.29	0.31
10	–	–	−0.00	−0.15	+0.04	+0.01	0.23	0.69
11	+0.60	−0.75	+0.52	−0.83	+0.44	−0.83	0.57	0.46
12	+1.10	+0.50	+1.56	+0.44	+1.58	+0.38	0.63	0.43
13	+1.00	+0.30	+1.29	+0.34	+1.37	+0.28	0.66	0.40
14	–	–	+0.04	+0.08	+0.14	+0.12	0.43	0.37

Table 10.6: Summary of results: Simulated and estimated deformations. The 95%-confidence intervals are listed in Columns (8) and (9). All quantities are given in millimetre.

10.11 Example: Crustal Movement Monitoring Network

The selected monitoring network can be considered as part of a larger first-order control network. The design resembles that of the relative network, which has been used by a FIG Working Group to check and compare different approaches into deformation analysis. Refer to the report by HECK et al (1982) and the proceedings DEFORMATIONSANALYSEN '83. The simulation of the observations of two epochs is based on a standard deviation of one centesimal second (0.3 second of arc) for all directions and 8cm for all distances. A geological fault divides the network into two blocks. The northern block comprises six points which are identical in both epochs except for point 39, for which a deformation has been simulated. All 10 points of the southern block show simulated deformations. The deformation field consists of a functional part simulated by a strain model and two local irregularities at points 39 and 43 defined by additional single point movements. Figure 10.11 shows the network, the observations, the simulated deformations and the 95%-confidence ellipses referring to a geodetic datum based on all points. The model parameters and the derived deformations as generated for this example are listed in Table 10.7. The Tables 10.8a and 10.8b give the simulated observations of both epochs as well as the approximate coordinates. The interested reader is therefore able to carry out his own analysis to exercise the outlined methods or to try other approaches.

The first single epoch adjustments were carried out using the minimum trace datum of Section 3.7 involving all points. The residuals of these adjustments and the test statistics for POPE's test of outlier detection (Section 6.5) are listed in Tables 10.8a and 10.8b. For $f = 61$ and $\alpha = 5\%$ type I error probability the critical value of the τ-distribution is $\tau = 3.36$. The maximum values of T_i are $T_g = 2.96$ and $T_{go} = 2.84$ in epoch 1 and epoch 2, respectively. Since both values are less than the critical value no observations need be rejected. The a posteriori variances of both adjustments (see Table 10.8b) are so close to the a priori variance factor $\sigma_o^2 = 1$ that no global model test is necessary. The redundancy contributions f_i of the observations of the monitoring network (see Tables 10.8a and 10.8b) range between 0.31 and 0.89 indicating that the network possesses a high degree of reliability. Altogether the GMMs of the single epoch adjustments are satisfactory.

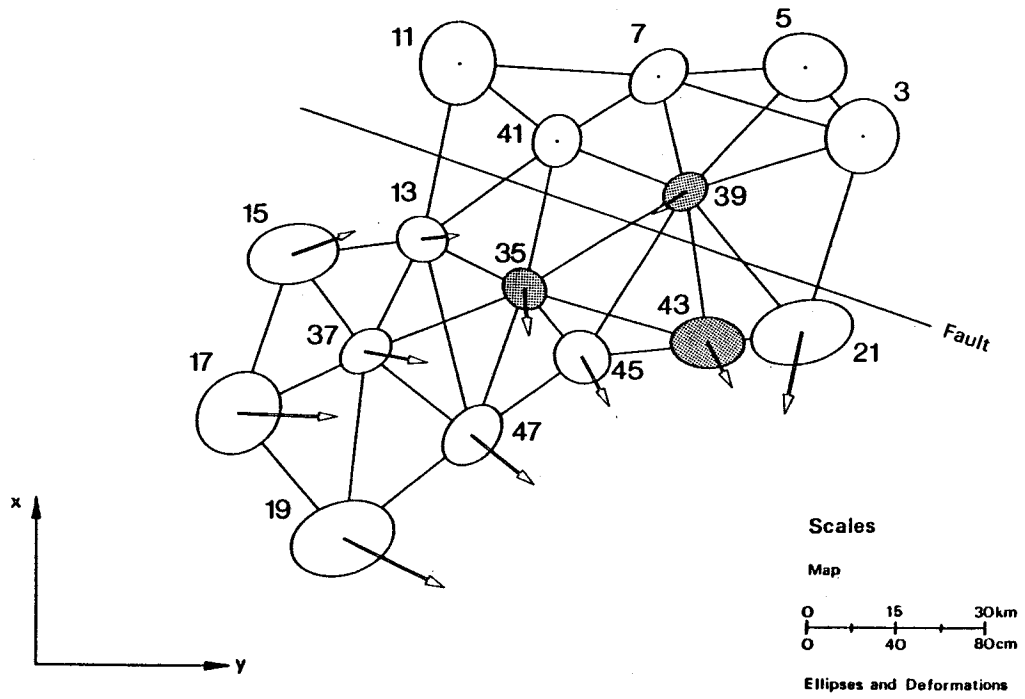


Figure 10.11: Monitoring network comprising 16 points divided into two blocks by a geological fault. The geodetic datum minimizes the trace of the cofactor matrix. The plot shows the 95%-confidence ellipses and the simulated deformations. The shaded points are non-conformingly deformed.

Simulated Deformation Field						Adjustment Results	
Single Point Deformations			Deformation Components			Coordinate Differences	
No.	ΔX	ΔY	No.	ΔX	ΔY	ΔX	ΔY
35	−10.0	−10.0	3	–	–	+8.39	−0.86
39	−10.0	−15.0	5	–	–	+4.79	−2.52
43	+10.0	+10.0	7	–	–	+10.14	−3.93
			11	–	–	+21.25	−6.84
			39	−10.0	−15.0	−5.80	−22.31
Strain of the Southern Block:			41	–	–	+12.21	−5.41
ε_{xx} = −5.5 microstrain			13	+1.79	+16.27	+14.52	−3.41
ε_{yy} = +4.5 microstrain			15	+10.47	+27.82	+22.51	+4.35
ε_{xy} = +4.5 microstrain			17	−1.14	+44.44	+8.10	+17.63
			21	−36.69	−6.39	−31.22	−11.68
ω = 0			35	−21.12	+2.14	−10.79	−9.73
			37	−4.98	+29.65	+5.93	+9.17
t_x = +10 cm			43	−20.49	+11.78	−12.69	+2.42
t_y = 0			45	−22.30	+12.68	−15.79	+1.93
			47	−21.53	+27.53	−17.20	+15.64
			19	−21.55	+45.85	−14.33	+15.55

Table 10.7: Simulated and estimated deformations in [cm].

No	from	to	Epoch 1	Epoch 2	f_i Eq(7-15)	Epoch 1		Epoch 2	
						v_i	T_i	v_i	T_i
			Directions (grads)			cc		cc	
1	2		3	4	5	6	7	8	9
1	3	5	0.00000	0.00000	0.3599	0.577	0.939	0.215	0.409
2		7	361.29765	361.29750	0.6078	0.408	0.511	-0.012	0.017
3		39	323.23110	323.23091	0.5784	-1.027	1.320	0.386	0.578
4		21	261.04079	261.04068	0.3981	0.043	0.066	-0.589	1.064
5	5	7	0.00000	0.00000	0.4322	1.115	1.657	0.030	0.052
6		39	352.76366	352.76347	0.5138	-0.553	0.753	0.036	0.058
7		3	261.80550	261.80562	0.3057	-0.562	0.994	-0.066	0.137
8	7	11	0.00000	0.00000	0.4526	-0.984	1.429	-0.257	0.436
9		41	359.69627	359.69668	0.5073	2.160	2.964	-0.077	0.124
10		39	282.23366	282.23324	0.4956	-0.309	0.429	-0.030	0.048
11		3	214.60760	214.60787	0.5739	-0.157	0.202	-0.329	0.495
12		5	191.50469	191.50468	0.4554	-0.711	1.029	0.694	1.171
13	11	13	0.00000	0.00000	0.3864	-0.244	0.384	-1.488	2.726
14		41	330.15086	330.15042	0.5059	-0.028	0.039	0.363	0.582
15		7	291.11635	291.11585	0.4531	0.272	0.395	1.125	1.903
16	13	15	0.00000	0.00000	0.3658	-0.468	0.756	-0.527	0.993
17		37	336.27306	336.27361	0.5307	-0.439	0.589	0.175	0.273
18		47	290.81403	290.81445	0.6280	0.101	0.125	-0.543	0.780
19		35	236.92919	236.92853	0.5213	0.116	0.158	0.392	0.618
20		41	166.45696	166.45727	0.5383	1.602	2.134	0.150	0.232
21		11	120.03977	120.03991	0.4123	-0.912	1.389	0.354	0.628
22	15	17	0.00000	0.00000	0.3556	-0.702	1.150	-0.435	0.830
23		37	339.41639	339.41589	0.4733	0.285	0.406	0.045	0.074
24		13	273.00007	272.99954	0.3439	0.416	0.694	0.390	0.758
25	17	19	0.00000	0.00000	0.3047	-0.047	0.083	0.545	1.126
26		37	315.56894	315.56943	0.4599	-0.246	0.354	-0.456	0.766
27		15	264.10554	264.10636	0.3703	0.292	0.469	-0.089	0.167
28	19	47	0.00000	0.00000	0.4137	-0.815	1.238	0.996	1.764
29		37	349.95888	349.95939	0.5324	0.817	1.094	-0.385	0.600
30		17	298.63016	298.63009	0.3774	-0.002	0.004	-0.612	1.134
31	21	3	0.00000	0.00000	0.3833	0.997	1.574	0.137	0.252
32		39	338.03640	338.03608	0.4929	-0.441	0.614	0.270	0.439
33		43	274.75303	274.75227	0.3161	-0.556	0.966	-0.407	0.825
34	35	39	0.00000	0.00000	0.6158	-1.993	2.482	0.002	0.002
35		41	348.68556	348.68571	0.5377	0.851	1.135	-0.752	1.168
36		13	265.50773	265.50676	0.5050	-0.971	1.335	0.678	1.086
37		37	211.62875	211.62863	0.5836	-0.534	0.684	-0.878	1.309
38		47	156.58445	156.58512	0.5072	1.645	2.257	1.311	2.097
39		45	90.00839	90.00860	0.4763	-0.652	0.924	-0.545	0.900
40		43	54.00849	54.00866	0.5958	1.655	2.095	0.185	0.274
41	37	19	0.00000	0.00000	0.4248	-0.454	0.681	0.397	0.694
42		47	334.85504	334.85449	0.4796	-0.644	0.908	0.135	0.222
43		35	268.82727	268.82697	0.5787	-0.496	0.637	0.023	0.035
44		13	222.04981	222.05035	0.4929	1.612	2.244	-0.338	0.548
45		15	152.19326	152.19295	0.3551	-0.447	0.733	0.015	0.028
46		17	64.24003	64.24007	0.3738	0.428	0.684	-0.232	0.432
47	39	3	0.00000	0.00000	0.4903	1.277	1.782	-0.121	0.197
48		5	367.72739	367.72697	0.4636	-0.409	0.587	-0.489	0.818
49		7	305.69288	305.69191	0.4671	-0.140	0.200	0.281	0.469
50		41	243.73986	243.73907	0.5353	0.389	0.519	-0.422	0.657
51		35	183.72443	183.72425	0.6403	-0.025	0.030	1.103	1.570
52		45	153.46066	153.46083	0.6131	-0.317	0.395	0.144	0.210
53		43	109.45283	109.45312	0.4675	-0.284	0.406	-0.435	0.724
54		21	75.84623	75.84576	0.3708	-0.491	0.789	-0.062	0.117
55	41	35	0.00000	0.00000	0.5439	-0.441	0.585	1.551	2.396
56		39	311.32964	311.32922	0.5038	-0.573	0.788	-0.320	0.514
57		7	250.74544	250.74550	0.4216	-0.332	0.500	0.336	0.589
58		11	130.08325	130.08342	0.3385	0.524	0.879	-0.906	1.773
59		13	46.34978	46.35013	0.5106	0.822	1.124	-0.661	1.054
60	43	21	0.00000	0.00000	0.2736	0.647	1.209	0.168	0.366
61		39	296.89002	296.89131	0.4480	0.469	0.685	-0.926	1.576
62		35	225.17056	225.17111	0.6046	-0.123	0.155	0.695	1.018
63		45	199.72996	199.73055	0.4739	-0.993	1.409	0.063	0.105

Table 10.8a: Simulated directions, the residuals v_i, POPE's test statistics $T_i = v_i/s_{v_i}$ for both epochs and the redundancy contributions f_i of the observations of the GMM.

No	from	to	Epoch 1	Epoch 2	f_i Eq(7-15)	Epoch 1 v_i	Epoch 1 T_i	Epoch 2 v_i	Epoch 2 T_i
			Directions (grads)			cc		cc	
1	2		3	4	5	6	7	8	9
64	45	43	0.00000	0.00000	0.4085	0.247	0.377	-0.724	1.290
65		39	341.16811	341.16827	0.5535	-0.524	0.688	0.866	1.325
66		35	261.44042	261.44043	0.4303	-0.391	0.583	-0.123	0.213
67		47	167.23344	167.23358	0.3353	0.668	1.128	-0.019	0.038
68	47	45	0.00000	0.00000	0.4655	-0.965	1.382	-0.446	0.745
69		35	360.78310	360.78339	0.5996	-0.327	0.413	1.107	1.628
70		13	323.59092	323.59096	0.6588	0.042	0.051	-0.561	0.788
71		37	281.85486	281.85440	0.5294	0.446	0.598	-0.771	1.207
72		19	197.04076	197.04054	0.3358	0.804	1.355	0.672	1.321
			Distances in m			cm		cm	
73	3	5	15822.626	15822.504	0.7193	-7.895	1.151	5.925	1.006
74	3	39	32261.868	32261.535	0.7836	-8.195	1.144	-0.067	0.011
75	3	21	36246.786	36246.380	0.6907	4.646	0.691	4.134	0.717
76	5	7	25458.444	25458.387	0.7386	0.392	0.056	4.774	0.800
77	5	39	30444.032	30443.871	0.8041	4.281	0.590	-1.087	0.175
78	7	41	21106.654	21106.462	0.8415	-5.025	0.677	13.688	2.150
79	7	39	20792.677	20792.756	0.8854	17.175	2.256	-2.080	0.318
80	7	11	34763.631	34763.688	0.6907	-2.988	0.444	-12.400	2.149
81	11	13	31514.879	31514.828	0.7221	4.149	0.604	3.268	0.554
82	11	41	21700.352	21700.303	0.7986	-6.628	0.917	-8.179	1.318
83	13	15	22989.885	22989.978	0.7562	-4.673	0.664	-5.252	0.870
84	13	37	22320.009	22320.066	0.8668	0.175	0.023	-7.820	1.210
85	13	47	35874.440	35874.151	0.8340	9.361	1.267	2.834	0.447
86	13	35	19982.358	19982.266	0.8940	-5.254	0.687	-0.632	0.096
87	13	41	28987.077	28987.111	0.8225	4.858	0.662	4.485	0.712
88	15	17	29540.163	29540.100	0.7078	3.154	0.463	-0.410	0.070
89	15	37	21750.955	21750.772	0.8376	-3.195	0.432	-0.814	0.128
90	17	19	28723.708	28723.422	0.6965	3.391	0.502	16.471	2.843
91	17	37	24493.897	24493.982	0.7977	0.331	0.046	0.507	0.082
92	19	47	28934.874	28934.931	0.6876	-1.922	0.287	-5.430	0.943
93	19	37	32928.815	32928.521	0.7606	-0.606	0.086	9.331	1.541
94	21	39	32339.620	32339.419	0.7697	4.954	0.698	-1.536	0.252
95	35	41	26856.765	26856.518	0.8567	-1.394	0.186	0.214	0.033
96	35	37	29804.167	29804.241	0.8424	-11.829	1.593	5.272	0.827
97	35	47	27132.980	27133.016	0.8437	2.960	0.398	1.157	0.181
98	35	45	15742.373	15742.093	0.8905	-9.669	1.267	6.124	0.935
99	35	43	33263.543	33263.564	0.7766	2.693	0.378	-12.468	2.038
100	35	49	41375.007	41374.845	0.8657	-3.330	0.442	-4.786	0.741
101	37	47	23978.105	23978.008	0.8662	1.976	0.262	-7.159	1.108
102	39	41	23948.918	23949.122	0.8633	3.963	0.527	-7.494	1.162
103	39	45	33973.168	33973.158	0.8260	-0.642	0.087	3.568	0.565
104	39	43	27140.901	27140.746	0.8592	-2.128	0.284	2.323	0.361
105	43	45	21680.018	21680.126	0.7633	7.441	1.053	-4.424	0.730
106	45	47	23576.305	23576.317	0.7856	-16.678	2.326	-7.278	1.183

Approximate Coordinates (m)

No	X	Y
3	3710.0000	91680.0000
5	15980.0000	81690.0000
7	13860.0000	56320.0000
11	15600.0000	21600.0000
39	-6390.0000	61040.0000
41	2240.0000	38700.0000
13	-15350.0000	15660.0000
15	-18220.0000	-7150.0000
17	-46450.0000	-15850.0000
21	-31170.0000	81820.0000
35	-24130.0000	33610.0000
37	-35500.0000	6060.0000
43	-33140.0000	65630.0000
45	-35850.0000	44120.0000
47	-49930.0000	25210.0000
19	-68270.0000	2830.0000

Stochastic Model (a priori)

Variance factor: $\sigma_0 = 1$

Directions: $\sigma = 1$ cc $\rightarrow$ $p = 1$

Distances: $\sigma = 8$ cm $\rightarrow$ $p = 1/64$

A posteriori: Epoch 1 $s_0 = 1.02$
Epoch 2 $s_0 = 0.88$
Pooled $s_0 = 0.95$

Redundancies: $f_1 = f_2 = 61$, $f = 122$

Rank Deficiency: $d = 3$

Table 10.8b: **Upper part:** simulated directions and distances, the residuals v_i, POPE's test statistics $T_i = v_i/s_{v_i}$ for both epochs and the redundancy contributions f_i.

Lower part: approximate coordinates, a priori and a posteriori variances, rank specifications of the GMM.

The coordinate differences between the two adjustments are given in the last two columns of Table 10.7. The congruency test indicates that the differences (deformations) are highly significant. For an analysis of the differences additional information is required. In this example, this may exist in form of the following assumptions:

i. the northern block is most probably stable,

ii. the southern block is possibly displaced relative to the former, and

iii. internally deformed.

These assumptions lead to the following step-by-step procedure.

Firstly a datum transformation (Section 3.8) is used to base the geodetic datum of both adjustments on the northern block only. The transformed coordinates and their standard deviations are listed in Table 10.9.

No	Epoch 1		Epoch 2		S_x	S_y	S_p
	X (m)	Y (m)	X (m)	Y (m)	cm	cm	cm
1	2	3	4	5	6	7	8
3*	3710.0172	91680.0336	3709.9399	91679.9596	1.93	3.27	3.80
5*	15980.0008	81690.0137	15979.9819	81689.9858	1.85	3.07	3.58
7*	13860.0025	56319.9885	13859.9903	56319.9733	1.72	2.10	2.71
11*	15600.0333	21599.9618	15599.9956	21599.9824	1.62	3.70	4.04
39*	-6390.0362	61039.9983	-6389.9001	61040.1190	1.80	1.73	2.50
41*	2239.9825	38700.0040	2239.9924	38699.9800	1.38	2.41	2.78
13	-15350.0270	15659.9298	-15349.9830	15659.8417	3.79	3.37	5.08
15	-18220.0625	-7150.0583	-18220.0386	-7150.2318	6.32	5.14	8.14
17	-46450.0713	-15850.0411	-46449.8802	-15850.4221	8.02	6.46	10.30
21	-31169.9635	81820.0231	-31169.6212	81819.9758	4.00	4.56	6.07
35	-24130.0146	33609.9813	-24129.7666	33609.9420	3.07	3.18	4.43
37	-35500.0516	6059.9617	-35499.8977	6059.6963	5.14	4.65	6.92
43	-33139.9637	65630.0672	-33139.7687	65629.8773	3.50	3.96	5.28
45	-35850.0108	44120.0219	-35849.7336	44119.8323	3.35	4.01	5.23
47	-49930.0410	25210.0498	-49929.7059	25209.6850	4.38	5.46	7.00
19	-68270.0590	2829.9345	-68269.6945	2829.5173	6.55	8.33	10.60

Table 10.9: Adjusted coordinates of the single epoch models. The geodetic datum refers to the points marked by an asterisk (northern block). The standard deviations of x and y and the circular point errors are given in Columns (6), (7) and (8).

Prior to using the northern block as the reference block during further analysis its stability has to be verified. The hypothesis according to Eq. (10-18) leads to the quadratic form $q_\Delta = 73.77$ with $f_\Delta = 12 - 3 = 9$ degrees of freedom. The test statistic of Eq. (10-14) yields $T = 9.08$ when using the pooled variance of Table 10.8b in the denominator. It has to be compared with the critical value $F_\alpha(9, 122) = 1.94$ of the F-distribution for $\alpha = 5\%$ type I error probability. Since $T > F_\alpha$, the stability of the reference block has to be questioned.

Considering the prior assumptions, a local instability could be responsible for the failure of the test. Therefore a single point diagnosis for the northern block is carried out. The decomposition of q_Δ, using the procedure of Section 10.4.2 leads to the following subforms:

$$q_\Delta^3 = 24.43, \qquad q_\Delta^7 = 11.92, \qquad q_\Delta^{39} = 66.16$$

$$q_\Delta^5 = 1.83, \qquad q_\Delta^{11} = 1.54, \qquad q_\Delta^{41} = 12.45$$

According to the philosophy of Section 10.4.2, point 39 is considered as unstable. The reduced quadratic form $q_\Delta' = 7.61$ and $f_\Delta' = 7$ lead to the new test statistic $T' = 1.21$ being well below the

critical value of $F_\alpha(7, 122) = 2.07$. Hence the stability of all points but 39 of the northern block, is established.

The estimated deformations are given in Columns (2) and (3) of Table 10.10 and are relative to the verified reference block. Figure 10.12 depicts these deformations together with the corresponding 95%-confidence ellipses.

A tentative continuation of the analysis could be a single point diagnosis of the southern (object) block including point 39. The results of an analysis along the lines of Section 10.4.2 are given in the Columns (4), (5), (6) and (7) of Table 10.10. Following the sequence of Column (9), one point after the other is deleted as unstable until only one point is left. Hence the applied method is unsuitable. As an alternative the Cholesky decomposition of Section 10.4.3 can be applied. The sequence of independent quadratic forms, according to Eq. (10-32) is given in Column (8). If a type I error probability of $\alpha = 5\%$ is selected, then the critical value of the F(2, 122)-distribution must be computed at:

$$\bar{\alpha} = 1 - (1 - \alpha)^{1/k} = 0.47\%$$

where $k = 11$ is the number of points involved in the multiple test, yielding:

$$F_{\bar{\alpha}} = 5.6$$

It follows from Eq. (10-35) that the critical form:

$$q_\Delta^{p_i} = 2s_o^2 F_{\bar{\alpha}} = 10.1$$

is less than the subform of point 17 and that all points but 13 are unstable. This confirms the previous result that a single point method is not appropriate to model the deformations of the object block.

Estimated Deformations			Single Point Diagnosis					
No.	ΔX	ΔY	q_Δ	f_Δ	T	$F_\alpha(f_\Delta, 122)$	q_Δ^p	No.
1	2	3	4	5	6	7	8	9
13	−1.24	+7.06	397.17	22	20.00	1.62	104.94	39
15	+3.76	+15.18	292.23	20	16.19	1.65	58.93	21
17	−11.55	+32.45	233.30	18	14.36	1.68	50.50	35
21	−36.59	−6.41	182.80	16	12.66	1.72	36.35	47
35	−23.11	+0.30	146.45	14	11.59	1.76	31.42	19
37	−10.59	+22.26	115.03	12	10.62	1.82	29.16	45
43	−19.59	+12.06	85.87	10	9.51	1.90	37.45	43
45	−25.96	+13.32	48.42	8	6.71	2.00	23.15	15
47	−30.28	+30.22	25.27	6	4.67	2.16	9.81	37
19	−30.89	+33.32	15.46	4	4.28	2.42	12.05	17
39	−14.21	−16.64	3.41	2	1.89	3.04	3.04	13

Table 10.10: Object point analysis. The deformations [cm] of the object points are given in Columns (2) and (3). The successively reduced quadratic forms q_Δ and the corresponding degrees of freedom are contained in Columns (4) and (5). The test statistics T of Eq. (10-14) and the critical value of the F-distribution for $\alpha = 5\%$ follow in Columns (6) and (7). Column (8) comprises the quadratic subforms of Eq. (10-32) of the Cholesky decomposition, referring to the points as ordered in Column (9), being also the sequence by which the points were removed from the quadratic forms of Column (4).

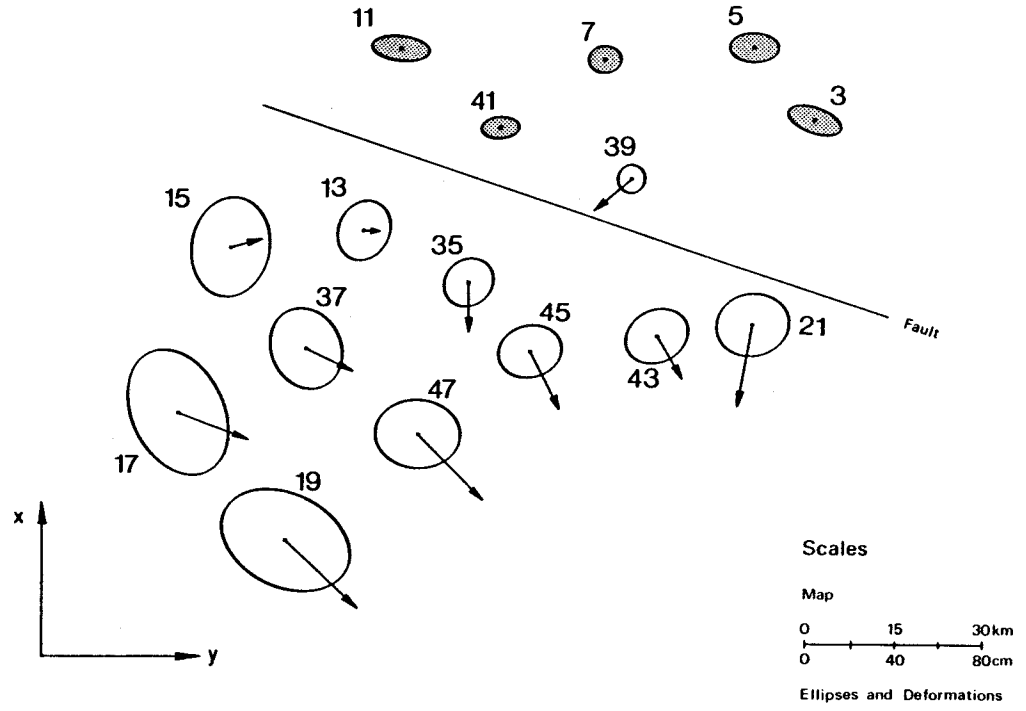


Figure 10.12: Estimated deformations of the object points and 95%-confidence ellipses relative to the verified reference block (shaded).

For another trial of a deformation model the rigid body model of Section 10.5 is adopted. The inter-epoch differences Δ of the object points of Table 10.10, Columns (2) and (3), are used as the "observations" of the model of Eq. (10-51). The parameter vector t is defined by Eq. (10-52) with the scale parameter s taken as zero. The estimated parameters and the residuals of the least squares adjustment are listed in Table 10.11. The significance test for the parameters leads, according to Eq. (10-58), to the test statistic T = 19.4. Comparison with the critical value $F_\alpha(4, 122)$ = 3.9 for α = 5% of the Fisher distribution indicates that the model parameters are significant. This does not prove the validity of the model, however. The residuals of the model yield the quadratic form q_δ = 247.2 with f_δ = 17. The test statistic according to Eq. (10-59) for the test of the null hypothesis ('all not modelled deformations are insignificant') is T = 14.8 and the corresponding critical F-value $F_\alpha(17, 122)$ = 1.7. Hence the test fails, meaning that a more suitable model has to be found.

The strain model of Section 10.6 with six parameters according to Eqs (10-63) and (10-73) is tested next in order to model the deformations of the object block. The estimated deformations of Table 10.10 and their covariance matrix form the input of the adjustment. The results are given in Table 10.11, under the heading 'Strain Model 1'. It turns out that the parameter vector is significant, but that the model test fails. The latter is carried out in the same way as for the rigid body model. The residuals obviously contain systematic effects which have not yet been modelled. An inspection of the residuals does not help since no clear tendency is apparent.

There are two possible strategies to solve the problem. The first one is a sophistication and extension of the functional model. This would cause difficulties in the interpretation of the result and a reduction of the redundancy of the model. The second strategy retains the strain model and excludes single points from the object block as non-conformingly deformed. This strategy is adopted.

	Rigid Body Model	Strain Model 1	Strain Model 2
t_x	+2.28 ± 1.06 cm	+4.90 ± 1.29 cm	+5.24 ± 1.33 cm
t_y	−4.97 ± 1.21 cm	−4.44 ± 1.15 cm	−4.03 ± 1.29 cm
ω, r_z	$-1.35 \pm 0.24 \cdot 10^{-2}$ grad	$+0.29 \pm 0.30 \cdot 10^{-2}$ grad	$-0.20 \pm 0.32 \cdot 10^{-2}$ grad
ε_{xx}	–	−5.9 ± 1.0 μ strain	−7.0 ± 1.0 μ strain
ε_{yy}	–	+5.6 ± 0.5 μ strain	+4.9 ± 0.5 μ strain
ε_{xy}	–	+1.5 ± 0.9 μ strain	+2.0 ± 0.9 μ strain

No.	Residuals [cm]		Residuals [cm]		Residuals [cm]	
	δx	δy	δx	δy	δx	δy
13	+16.8	−9.9	+2.4	+3.1	+0.4	+0.1
15	+19.9	−1.5	+0.3	+5.7	−1.6	+1.9
17	+3.7	+17.8	−1.8	+0.6	−0.8	+0.4
21	−14.9	−14.8	+0.1	−4.9	+0.5	−2.1
35	−4.7	−14.3	−8.5	−5.9	–	–
37	+6.2	+7.0	+0.8	+2.1	+0.9	+1.4
43	+0.2	−0.3	+11.5	+5.6	–	–
45	−7.5	−0.2	−1.1	+0.3	−0.6	+1.7
47	−12.9	+16.5	−3.9	+2.7	−2.1	+5.1
19	−14.8	+20.3	−1.7	−11.9	+1.9	−8.0
	q_δ = 247.42		q_δ = 82.94		q_δ = 6.92	
	f_δ = 17		f_δ = 14		f_δ = 10	
	T = 14.82		T = 6.03		T = 0.70	
	$F_\alpha(17, 122)$ = 1.70		$F_\alpha(14, 122)$ = 1.76		$F_\alpha(10, 122)$ = 1.90	

Table 10.11: Analysis of the southern (object) block. The rigid body model and the strain model fail the model test. The strain model 2 disregards the points 35 and 43, which show non-conforming deformations. This model is "verified" by the F-test (α = 5%).

The residuals of strain model 1 given in Table 10.11 are screened for outliers. The procedure parallels that of Section 10.4. The residuals are interpreted as deformations, which have expectation zero except for one point. The decomposition of the quadratic form, according to Section 10.4.2, indicates that point 43 does not fit the model. The quadratic form q_δ = 82.94 is reduced by q_Δ^{43} = 53.54 yielding the new test statistic T' = 2.71. It still exceeds the critical value $F_\alpha(12, 122)$ = 2.54. The decomposition of the reduced quadratic form indicates that point 35 does not conform to the model. The next step of reduction produces a test statistic which is smaller than the critical value. The strain model is therefore suitable for the approximation of the object deformations provided that the points 43 and 35 are excluded and considered as independently deformed.

The final step of the analysis is a new adjustment in order to estimate the strain parameters from the eight points of the object block fitting the model. The results of this adjustment are given as 'Strain Model 2' in Table 10.11. A concluding model test confirms the suitability of the adopted analysis procedure.

The "observations" of the deformation model (Column (3) of Table 10.12) and the corresponding covariance matrix is derived from the initial single epoch adjustments (Table 10.7). To simplify matters, the strain model known from the simulation is employed as the sole model. However, no measures are taken to account for non-conforming single point movements. A total of 21 iterations are required to get the final result of the estimation as listed in Column (5) of Table 10.12. The estimates of the parameters of the model are as close to the true (simulated) values as the final results of the least squares approach. The quadratic forms $\delta^t Q_\delta^{-1} \delta$ of the points, already used frequently as indicators for non-conformities (see Eq. (10-46)), clearly reveal that the points 39, 35 and 43 do not fit the model. The critical value is approximately $\sigma_o^2 \chi^2(2) = 5.99$ for $\alpha = 5\%$.

The final deformations obtained using this approach (refer to Column (5) of Table 10.12) are in very good agreement with the simulated values. The residual differences are insignificant and similar to those of the least squares estimation. Considering that the estimated differences of Column (3) constitute the sole input to the robust estimation process, the results are remarkably good. Obviously, the robustness of the method is strong enough to ignore the three non-fitting points. A one-step estimation procedure is thus possible in contrast to involved multi-step analyses of the conventional type.

Figure 10.13 exhibits the final results of the analyses together with the true deformations and the confidence ellipses of a single epoch adjustment with minimum trace datum according to Section 3.7. When comparing the methods it is important to consider the accuracy of the observations characterized by the confidence ellipses of Figure 10.13. The simplicity of the procedure and the reduced computational effort clearly favour the robust approach.

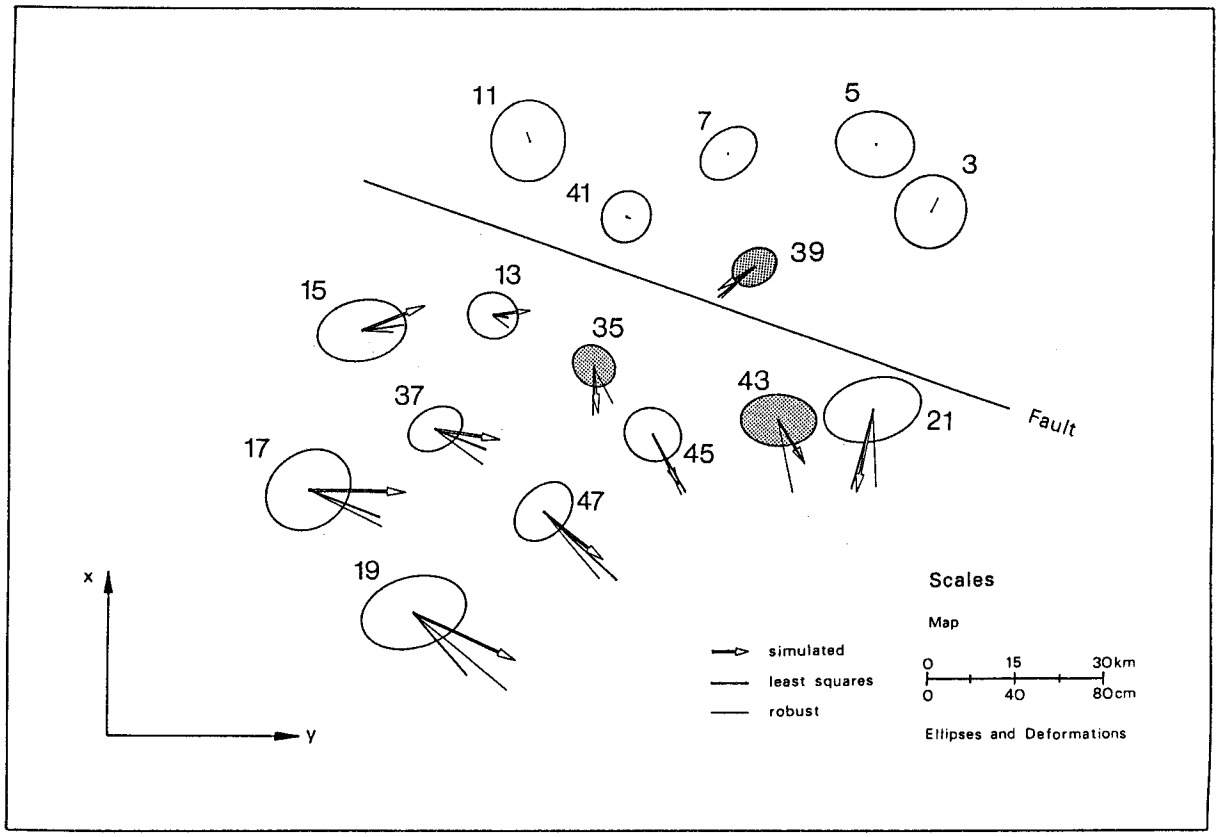


Figure 10.13: Simulated deformations and final results of the least squares analysis using strain model 2 and of the robust analysis. The 95%-confidence ellipses are based on the minimum trace datum. The shaded points are non-conformingly deformed.

In Table 10.12 the final results of the analysis are listed next to the simulated deformations. It appears that the superposition of the deformations by observational errors renders the identification of the true model difficult. When assessing the outcome, the standard deviations of the estimated coordinate differences should be considered. These amount to square root two times the quantities listed in Columns (6) and (7) of Table 10.9. It is surprising how well the true situation is approximated by the final model. All discrepancies can be explained with the random errors of the observations. It is doubtful, of course, whether it would have been possible to identify the true deformation model without knowledge of the simulated input deformations. Deformation models applied to real world examples usually exhibit a worse fit.

To conclude the series of different solutions, an analysis based upon the robust estimation method of Section 10.4.5 is carried out.

	Simulated Deformations		Estimated Coordinate Differences		Strain Model 2 Least Squares		Strain Model Robust Estimate		
1	2		3		4		5		
t_x	+10.0 cm				+5.2 ± 1.3 cm		+14.8 cm		
t_y	–				–4.0 ± 1.3 cm		–5.1 cm		
ω	–				$-0.2 \pm 0.3 \cdot 10^{-2}$ grad		$-1.3 \cdot 10^{-2}$ grad		
ε_{xx}	–5.5 μ strain				–7.0 ± 1.0 μ strain		–6.7 μ strain		
ε_{yy}	+4.5 μ strain				+4.9 ± 0.5 μ strain		+4.5 μ strain		
ε_{xy}	+4.5 μ strain				+2.0 ± 0.9 μ strain		+2.3 μ strain		
No.	ΔX [cm]	ΔY [cm]	ΔX [cm]	ΔY [cm]	ΔX [cm]	ΔY [cm]	ΔX [cm]	ΔY [cm]	$\delta^t Q_\delta^{-1} \delta$
3	–	–	+8.4	–0.9	–	–	+5.3	+5.1	3.3
5	–	–	+4.8	–2.5	–	–	–0.3	+1.0	0.0
7	–	–	+10.1	–3.9	–	–	+0.0	–0.0	0.0
11	–	–	+21.2	–6.8	–	–	+4.2	–3.3	1.4
39*	–10.0	–15.0	–5.8	–22.3	–14.2	–16.6	–15.0	–14.4	30.4*
41	–	–	+12.2	–5.4	–	–	–1.4	+0.8	0.9
13	+1.8	+16.3	+14.5	–3.4	–0.8	+7.2	–5.3	+7.7	0.4
15	+10.5	+27.8	+22.5	+4.4	+2.2	+18.1	–0.2	+14.6	0.0
17	–1.1	+44.4	+8.1	+17.6	–12.4	+32.8	–16.4	+33.5	0.0
21	–36.7	–6.4	–31.2	–11.7	–36.1	–8.5	–36.0	+2.0	0.0
35*	–21.1	+2.1	–10.8	–9.7	–23.1	+0.3	–16.6	+8.8	12.0*
37	–5.0	+29.6	+5.9	+9.2	–9.7	+23.7	–15.7	+21.9	1.3
43*	–20.5	+11.8	–12.7	+2.4	–19.6	+12.1	–32.6	+6.8	16.6*
45	–20.3	+12.7	–15.8	+1.9	–26.6	+15.0	–27.6	+13.4	0.5
47	–21.5	+27.5	–17.2	+15.6	–32.4	+35.3	–31.2	+26.2	2.8
19	–21.6	+45.8	–14.3	+15.6	–29.0	+24.3	–36.6	+42.3	1.9

Table 10.12: Simulated deformations, coordinate differences of the single epoch adjustments (input of analyses) and final results of the analyses using least squares and robust estimation methods. Non-conformingly deformed points are marked with an asterisk.

11. MULTI-EPOCH ANALYSIS

The methods of multi-epoch deformation analysis have not been developed to the same extent as the two-epoch analysis. This may be due to the fact that there are only few really new aspects, the mathematical expressions become rather tedious, the computing effort increases considerably and a sequence of two-epoch analyses is quite adequate for the majority of problems.

The first well documented examples of multi-epoch studies can be found in SCHWEIZERISCHE TALSPERRENKOMMISSION (1946) and in SWISS NATIONAL COMMITTEE ON LARGE DAMS (1964). The behaviour of large Swiss dams was (and still is) monitored regularly. The two reports describe how the observations were evaluated by graphical methods and depict the deformations in a large number of very illustrative and remarkable diagrams.

Mathematical modelling of deformations based on multiple epochs began in the early seventies (SAVAGE and BURFORD (1970), DORRER (1971), GHITAU (1973), HOLDAHL (1978), PAPO and PERELMUTER (1981) and SCHNEIDER (1982)). The changes of the object are expressed as velocities and accelerations of movements, particularly in the vertical direction, or as strain accumulations in the horizontal. The time can appear directly as a factor of the model or indirectly via temperature, water level, progress of exploitation of mineral resources, etc. The methods of modelling deformations in the geometry domain as developed in Chapter 10 are now to be extended to the time domain.

Basically two computational approaches can be used. The first one is based on a simultaneous adjustment of all epochs. This straightforward method is always applicable, but it has the disadvantage of requiring enormous computer memory space even for medium sized problems. Thus it is usually necessary to employ matrix blocking techniques which increase the amount of programming effort and processing time.

The second approach is based on sequential updating methods and therefore is suitable for relatively small computers. It assumes, however, that correlations between epochs do not exist.

11.1 Multi-Epoch Model

It is assumed that the geodetic data of k epochs of observations are available, that single-epoch adjustments according to Section 10.1 have been carried out (all with the same geodetic datum and the same a priori variance factor) and that the results are available as k sets:

$$\hat{x}_i,\ Q_{\hat{x}_i},\ f_i,\ s_{oi}^2, \qquad \text{for } i = 1, 2, ..., k \tag{11-1}$$

Furthermore the parameter vectors $\hat{x}_i$ are assumed to be partitioned into three subvectors:

$$\hat{x}_i = (\hat{x}_i^r,\ \hat{x}_i^o,\ \hat{x}_i^n) \tag{11-2}$$

where the reference points, the object points and the nuisance parameters, including points appearing solely in one epoch, are denoted by the superscripts r, o and n, respectively. If some reference or object points do not appear in one of the epochs, their positions in $\hat{x}_i$ are filled by dummy variables with zero weight.

The contribution of epoch i to the total model is given by:

$$\hat{x}_i + v_i = A_i y_i, \qquad Q_{\hat{x}_i}^{-} = N_i = P_i \tag{11-3}$$

or in partitioned form using the same superscripts as before:

$$\begin{pmatrix} \hat{x}^r \\ \hat{x}^o \\ \hat{x}^n \end{pmatrix}_i + v_i = \begin{pmatrix} y^r \\ y^o + B_i p \\ y_i^n \end{pmatrix}; \quad P_i = \begin{pmatrix} p^{rr} & p^{ro} & p^{rn} \\ p^{or} & p^{oo} & p^{on} \\ p^{nr} & p^{no} & p^{nn} \end{pmatrix}_i \tag{11-4}$$

This shows that A_i is of the form:

$$A_i = \begin{pmatrix} I^r & 0 & 0 & 0 \\ 0 & I^o & B_i & 0 \\ 0 & 0 & 0 & I_i^n \end{pmatrix} \qquad (11\text{-}5)$$

and that:

$$E(\hat{x}_i^r) = E(\hat{y}^r)$$

$$E(\hat{x}_1^o) = E(y^o)$$

$$E(\hat{x}_i^n) = E(\hat{y}_i^n)$$

where:

- y^r is the vector of coordinates of the reference points averaged over all k epochs
- y^o is the vector of coordinates of the object points at epoch 1 (reference epoch)
- p is the vector of deformation parameters
- B_i is the coefficient matrix of p associated with epoch i
- y_i^n is the vector of nuisance parameters of epoch i

The weight matrix P_i is identical with the normals N_i of the single-epoch adjustment and is partitioned in accordance to Eq. (10-2). I^r, I^o and I_i^n are properly defined identity matrices. The part of the model representing the required deformations of the object is:

$$E(\hat{x}_i^o) = y^o + B_i p \qquad (11\text{-}6)$$

The second term on the right hand side of Eq. (11-6) is the actual linear or linearized deformation model. The general form for point m reads, if a polynomial expression is employed:

$$\begin{pmatrix} \hat{x}_m \\ \hat{y}_m \end{pmatrix}_i + \begin{pmatrix} v_m \end{pmatrix}_i = \begin{pmatrix} x^o \\ y^o \end{pmatrix}_m + \begin{pmatrix} p_0 + p_2\bar{x} + p_4\bar{y} + p_6 t_i + p_8\bar{x}^2 + \ldots \\ p_1 + p_3\bar{x} + p_5\bar{y} + p_7 t_i + p_9\bar{x}^2 + \ldots \end{pmatrix}_m \qquad (11\text{-}7)$$

For a linear model 8 parameters (p_0 to p_7) are required. The number increases to 20 and to 40 for polynomials of second and third order, respectively. The coefficients in Eq. (11-7) are point coordinates $\bar{x}_m$, $\bar{y}_m$ with reference to the centre of gravity of the modelled block and the elapsed time t_i since the first (reference) epoch. The polynomial model can be replaced without problems by another linear (linearized) model of deformations such as $c \cdot e^{-dt}$ or $\cos((T/2\pi)t + \phi)$.

By imposing constraints on the parameters, Eq. (11-7) can be specialized to model deformations similar to similarity or affine transformations in the geometry domain. This corresponds to the rigid body model of Eq. (10-51) and the strain model of Eq. (10-63) as developed in Sections 10.5 and 10.6. Simple arguments lead to a generalization of the model of Eq. (11-4) which considers two or more blocks with different deformation patterns or with relative rigid body displacements but with the same deformation pattern. If a reference block does not exist, the design matrix block I^r gets the dimension zero. If a multi-epoch global congruency test is considered in such a case the matrices I^o and B_i do not appear in Eq. (11-5).

There are only a few cases where the structure of the deformation model is definitely known in advance. More frequently it is part of the objectives of the analysis to establish a suitable model. Then a tentative estimation is carried out, for example, with a full set of parameters for a second order polynomial according to Eq. (11-7). Statistical inference is applied to identify those parameters which are significant. A new adjustment of the reduced model is required to obtain the final values of the significant unknowns.

The selection of a deformation model is not unique. It is likely that more than one model will fit the observed data. The selection of the most suitable model should be guided by criteria such as:

i. the final model should be fairly simple

ii. a physical interpretation of the parameters should be possible

iii. the associated quadratic form of the residuals should correspond to the pooled variance of the single-epoch adjustments.

For details see CHEN (1983).

To clarify the matrices involved in the multi-epoch model, the complete equations for k = 3 are given. For simplicity of presentation it has been assumed that the epochs are uncorrelated. This is, however, not a condition since the equations are easily extended to the more general case.

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} I^r & 0 & 0 & 0 & 0 & 0 \\ 0 & I^o & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_1{}^n & 0 & 0 \\ I^r & 0 & 0 & 0 & 0 & 0 \\ 0 & I^o & B_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_2{}^n & 0 \\ I^r & 0 & 0 & 0 & 0 & 0 \\ 0 & I^o & B_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_3{}^n \end{pmatrix} \begin{pmatrix} y^r \\ y^o \\ p \\ y_1{}^n \\ y_2{}^n \\ y_3{}^n \end{pmatrix} \qquad (11\text{-}8)$$

$$\underset{n\,x\,1}{\hat{x}} + \underset{n\,x\,1}{v} = \underset{n\,x\,u}{A}\ \underset{u\,x\,1}{y}$$

$$P = \begin{pmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & P_3 \end{pmatrix}, \qquad n = \sum_{i=1}^{k} n_i$$

$$u = 2m + u_p + \sum_{i=1}^{k} \bar{u}_i$$

with:

n_i number of parameters of epoch i

m number of points (reference plus object points)

u_p number of deformation parameters

$\bar{u}_i$ number of nuisance parameters of epoch i

The matrix A is of full rank u, but P has the rank defect of $d' = k \cdot d$, where d is the datum defect of the single-epoch adjustments. Since the parameters of Eq. (11-8) are again coordinates, except for the additional elements of the subvector p, the normals A^tPA have the same rank defect d as the single-epoch models. Hence, a geodetic datum must be selected in order to define the coordinate system of the solution of Eq. (11-8). In line with previous considerations the geodetic datum is selected so that the partial trace of the cofactor matrix referring to the reference points is minimized. The corresponding constraints are selected according to Section 3.8. The generalized inverse of the normals follows from Eq. (3-18).

$$
\begin{aligned}
&Ry = 0, \qquad R = \begin{pmatrix} I^r \\ 0 \end{pmatrix} S, \qquad P_i S = N_i S = 0 \\
&Q_{\hat{y}} = (A^tPA + RR^t)^{-1} A^tPA (A^tPA + RR^t)^{-1} \\
&\hat{y} = Q_{\hat{y}} A^tP\hat{x}, \qquad f = r(P) - u + d \\
&v = A\hat{y} - \hat{x}, \qquad q = v^tPv = \hat{x}^tP\hat{x} - \hat{y}^tA^tPA\hat{y}
\end{aligned}
\tag{11-9}
$$

The matrix S is required to develop the matrix R of constraints; it is defined in Eq. (3-61).

The structure of the matrix of normal equations is given in Eq. (11-10). All summations run from i = 1 to i = 3 in this example:

$$
A^tPA = \left(\begin{array}{ccc|ccc}
\sum_i P_i^{rr} & \sum_i P_i^{ro} & \sum_i P_i^{ro}B_i & P_1^{rn} & P_2^{rn} & P_3^{rn} \\
\sum_i P_i^{or} & \sum_i P_i^{oo} & \sum_i P_i^{oo}B_i & P_1^{on} & P_2^{on} & P_3^{on} \\
\sum_i B_i^tP_i^{or} & \sum_i B_i^tP_i^{oo} & \sum_i B_i^tP_i^{oo}B_i & 0 & B_2^tP_2^{on} & B_3^tP_3^{on} \\
\hline
P_1^{nr} & P_1^{no} & 0 & P_1^{nn} & 0 & 0 \\
P_2^{nr} & P_2^{no} & P_2^{no}B_2 & 0 & P_2^{nn} & 0 \\
P_3^{nr} & P_3^{no} & P_3^{no}B_3 & 0 & 0 & P_3^{nn}
\end{array}\right)
\tag{11-10}
$$

The P-matrix is partitioned according to Eq. (11-4). Each additional epoch would lead to an increase of the size of A^tPA. This can be prevented by eliminating the nuisance parameters $\hat{x}_i^n$ from the single-epoch models according to Eqs (10-4) to (10-6), so that only the partly reduced leading block of Eq. (11-10) remains. The right hand side vector of the normals is:

$$
A^tP\hat{x} = \left(\sum_i\sum_j P_i^{rj}\hat{x}_i^j,\ \sum_i\sum_j P_i^{oj}\hat{x}_i^j,\ \sum_i\sum_j B_i^tP_i^{oj}\hat{x}_i^j,\ \sum_j P_1^{nj}\hat{x}_1^j,\ \sum_j P_2^{nj}\hat{x}_2^j,\ \sum_j P_3^{nj}\hat{x}_3^j\right)^t
\tag{11-11}
$$

where: $i = 1, 2, 3$ and $j = r, o, n$.

As an alternative to the model of Eq. (11-8) it is possible to return to the original observations l_i. The multi-epoch model can then be formulated as a generalization of Eq. (10-56) where the matrix H has been replaced by matrix B of Eq. (11-6):

$$\begin{pmatrix} l_1 \\ l_2 \\ \cdot \\ \cdot \\ l_k \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_k \end{pmatrix} = \begin{pmatrix} A_1^r & A_1^o & 0 & A_1^n & 0 & \dots & 0 \\ A_2^r & A_2^o & -A_2^oB_2 & 0 & A_2^n & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ A_k^r & A_k^o & -A_k^oB_k & 0 & 0 & & A_k^n \end{pmatrix} \begin{pmatrix} y^r \\ y^o \\ p \\ y_1^n \\ \cdot \\ \cdot \\ y_k^n \end{pmatrix} \qquad (11\text{-}12)$$

This formulation is easier to understand as, in particular, the role of the geodetic datum is much clearer, but is computationally less convenient. An elimination of the nuisance parameters y_1^n, y_2^n, ..., y_k^n prior to the combined adjustment of Eq. (11-12) would considerably simplify the computations.

11.2 Parameter Estimation

The obvious method of calculating the parameter vector $\hat{y}$ of models of Eqs (11-8) and (11-12) is through the inversion of the matrix A^tPA, where the constraints as given in Eq. (11-9) must be considered in both cases. The solution equations of the observational model of Eq. (11-12) have the same form as those given in Eq. (11-9). But there are two differences:

i. the residuals v_i refer to the observations l_i, and

ii. the number n in the equation for the degrees of freedom f is now the number of all observations l.

All computations must be repeated whenever the data of a new epoch become available. Generally more than one estimation is necessary, since the new data usually require a modification of the model being achieved after several trial adjustments.

A reduction in the amount of computations involved in the estimation and savings in storage space are achieved by application of sequential methods consisting of updating the matrix of normal equations and the vector of absolute values. This method, however, cannot be used if correlations between the epochs exist. Consider the estimation in the model of Eq. (11-8) covering k epochs as completed and the results available:

$$\hat{y}_k, \quad Q_{\hat{y}_k}, \quad P_{\hat{y}_k} = (A^tPA)_k, \quad q_k, \quad f_k \qquad (11\text{-}13)$$

Further let it be assumed that the nuisance parameters $\hat{x}_i^n$ have been eliminated from Eq. (11-2) in the single-epochs adjustments, using the procedure outlined in Section 10.1. If a new epoch, say epoch l, has been observed and adjusted, yielding:

$$\hat{x}_l, \quad Q_{\hat{x}_l}, \quad P_{\hat{x}_l} = (A^tPA)_l$$

then the results of Eq. (11-13) can be updated by evaluating the model:

$$\begin{pmatrix} \hat{y}_k \\ \hat{x}_l \end{pmatrix} + \begin{pmatrix} v_k \\ v_l \end{pmatrix} = \begin{pmatrix} I \\ A_l \end{pmatrix} y_l \;, \qquad \begin{pmatrix} P_{\hat{y}_k} & 0 \\ 0 & P_{\hat{x}_l} \end{pmatrix} \tag{11-14}$$

$$\hat{x} \quad + \quad v \quad = \quad A \quad y \;, \qquad\qquad P$$

where the parameter vector after adding epoch l is denoted by y_l and where v_k is the vector of corrections to the previous parameter vector $\hat{y}_k$ caused by the new epoch. The matrix of the normal equations of Eq. (11-14):

$$(A^tPA)_l \;=\; (P_{\hat{y}_k} + A_l{}^tP_{\hat{x}_l}A_l) \;=\; (A^tPA)_k + A_l^tP_{\hat{x}_l}A_l \tag{11-15}$$

is formed by addition of $A_l{}^tP_{\hat{x}_l}A_l$ to the previous normals. The new absolute term is the sum of two vectors according to:

$$(A^tP\hat{x})_l \;=\; P_{\hat{y}_k}\hat{y}_k \;+\; A_l{}^tP_{\hat{x}_l}\hat{x}_l \tag{11-16}$$

The updated parameter vector follows from Eqs (11-15) and (11-16):

$$\hat{y}_l \;=\; (A^tPA)_l^{-}(A^tP\hat{x})_l$$

where: (11-17)

$$Q_{\hat{y}_l} \;=\; (A^tPA)_l^{-} \quad , \quad P_{\hat{y}_l} \;=\; (A^tPA)_l$$

The g-inverse is selected according to the datum definition of the single-epoch adjustments as outlined by the set of Eqs (11-9). The residuals and the variance estimate are given by:

$$\begin{pmatrix} v_k \\ v_l \end{pmatrix} = \begin{pmatrix} I \\ A_l \end{pmatrix} \hat{y}_l - \begin{pmatrix} \hat{y}_k \\ \hat{x}_l \end{pmatrix}$$

$$\bar{q}_l \;=\; (v_k,\; v_l) \begin{pmatrix} P_{\hat{y}_k} & 0 \\ 0 & P_{\hat{x}_l} \end{pmatrix} \begin{pmatrix} v_k \\ v_l \end{pmatrix} \tag{11-18}$$

$$\bar{f}_l \;=\; r(P_{\hat{y}_k}) + r(P_{\hat{x}_l}) - u + d$$

The degrees of freedom of the model of Eq. (11-14) depend on the rank of the weight matrix concerned, the number u of parameters and the datum defect d.

The quadratic form of the combined adjustment of all l epochs is computed from:

$$q_l = \sum_{i=1}^{l} \hat{x}_i^t P_{\hat{x}_i} \hat{x}_i - \hat{y}_l^t (A^t P A)_l \hat{y}_l \tag{11-19}$$

with:

$$f_l = \sum_{i=1}^{l} r(P_{\hat{x}_i}) - u + d \tag{11-20}$$

being identical to the results of Eq. (11-9).

Similarly the quadratic form of Eq. (11-18) can be expressed as:

$$\bar{q}_l = \hat{y}_k^t P_{\hat{y}_k} \hat{y}_k + \hat{x}_l^t P_{\hat{x}_l} \hat{x}_l - \hat{y}_l^t (A^t P A)_l \hat{y}_l \tag{11-21}$$

Solving Eq. (11-21) for $\hat{y}_l^t (A^t P A)_l \hat{y}_l$ and substituting the results in Eq. (11-19) yields:

$$q_l = \sum_{i=1}^{l} \hat{x}_i^t P_{\hat{x}_i} \hat{x}_i + \bar{q}_l - \hat{y}_k^t P_{\hat{y}_k} \hat{y}_k - \hat{x}_l^t P_{\hat{x}_l} \hat{x}_l$$

since:

$$\sum_{i=1}^{l} \hat{x}_i^t P_{\hat{x}_i} \hat{x}_i - \hat{x}_l^t P_{\hat{x}_l} \hat{x}_l = \sum_{i=1}^{k} \hat{x}_i^t P_{\hat{x}_i} \hat{x}_i$$

and the quadratic form q_k of Eq. (11-13) can be expressed analogous to Eq. (11-19):

$$q_k = \sum_{i=1}^{k} \hat{x}_i^t P_{\hat{x}_i} \hat{x}_i - \hat{y}_k^t (A^t P A)_k \hat{y}_k$$

The total form of l epochs yields the simple sum:

$$q_l = q_k + \bar{q}_l \tag{11-22}$$

and, similarly, the associated degrees of freedom:

$$f_l = f_k + \bar{f}_l \tag{11-23}$$

Hence, in the sequential multi-epoch approach, the current quadratic form of Eq. (11-9) of the simultaneous model is computed by accumulating the forms $\bar{q}$ of Eq. (11-21) of the single steps. The same procedure applies to the computation of the current degrees of freedom. Also, it can be easily shown that:

$$\bar{q}_l / \sigma_o^2 \quad \text{and} \quad q_k / \sigma_o^2$$

are independently χ^2-distributed, if the global model hypothesis of Section 2.3 is true.

11.3 Statistical Tests and Model Adjustments

Statistical tests guide the professional in his effort to establish the most realistic deformation model. But, as stressed repeatedly before, the test results should not be followed blindly. A number of additional aspects not covered by the relatively simple test procedures need to be taken into account.

In the sequel it will be assumed that the sequential estimation approach has been adopted. The modifications required for using the simultaneous adjustment approach are easily established. The following null hypotheses are usually considered:

H_{01} : The new epoch is in agreement with the already established deformation model.

H_{02} : All points of the new epoch conform to the already established deformation model.

H_{03} : Certain parameters of the deformation model are insignificant.

The null hypothesis H_{01} is identical to that of the global congruency test outlined in Section 10.3. Basically the same equations apply; however, epoch i is to be replaced by the joint result of the first k epochs according to Eq. (11-13) and for epoch j the results of Eq. (11-18) are to be substituted. Thus the relevant test statistic of Eq. (10-14) takes the form:

$$T_1 = \frac{\bar{q}_l/\bar{f}_l}{q_k/f_k} \sim F(\bar{f}_l,\ f_k) \qquad (11\text{-}24)$$

T_1 is centrally F-distributed if H_{01} is true.

If the test fails several alternatives can be considered. Similar to the analysis of single point movements in Section 10.4 a search for non-conformities in the geometry domain might be successful. H_{02} is the corresponding null hypothesis. Based on the partitioning of the model as given in Eq. (11-4) first the reference part of the network is screened followed by the object part. If it is possible to explain the failure of H_{01} by one or few single point movements, then these points are flagged, removed from their block and added to the subset of nuisance parameters.

The test procedure is based on a decomposition of $\bar{q}_l$ of Eq. (11-18), applying one of the methods outlined in Section 10.4, and leads to the test statistics T_{2i} as given in Section 10.4.4. Refer to Eq. (10-35), for example. In subsequent epochs the flagged single points need special attention. If they prove again to be non-conforming, it may be appropriate to model this effect by an extension or modification of the deformation model and to transfer them back to their initial position in the $\hat{x}$-vector. If the non-conformity is also a singular event in the time domain then the point remains in the group of nuisance parameters for the current epoch only.

If single point non-conformities do not explain the failure of H_{01} sufficiently, then a general revision of the deformation model might be necessary. Certain effects, which can be expressed by a linear model in the geometry or time domain when solely a few epochs are considered, may later prove to be of a more involved nature, thus requiring higher order terms of the model. In this case a new adjustment of all epochs with new tentative models is necessary until a sufficient agreement is obtained.

Another way of explaining the failure of H_{01} may be based on the assumption that the object under investigation (or a part of the area covered by the monitoring network) has experienced a sudden shift, which can be modelled by an extra set of parameters for a rigid body displacement of the corresponding block according to Section 10.5. The results of subsequent epochs will demonstrate whether or not this discontinuity in the time domain is restricted to one epoch only. If this is the case, the rigid body shift must be reversed in the next epoch. In the other case the discontinuity is followed by a new continuous behaviour.

These evaluations can be effectively supported by graphical representations of the deformation field. After establishing a satisfying model the significance of the model parameters is checked with the aid of the null hypothesis H_{03}. Obviously non-contributing parameters should be ignored. But the decisions should be made with due care, since moderately significant parameters of the model might become highly significant, when subsequent epochs are added. As a rule, it is better to consider too many parameters than too few. The test procedure is explained in Section 5. It is based on one of the test statistics given in Eqs (5-6) to (5-9).

Altogether, the establishment of a deformation model is a trial and error process. Only general guidelines can be given and no ready-for-use models are available. The statistical inference applied in this process strongly depends on the underlying model. Thus any change of the assumptions on which a model is based changes the inference as well.

11.4 Special Methods and Final Remarks

It has been stressed during the discussion of the two-epoch analysis that the strategies for deformation analysis are strongly project orientated and that the multitude of tailored approaches cannot be presented exhaustively in this monograph. Some short indications shall be given only. The interested reader finds further information in the references.

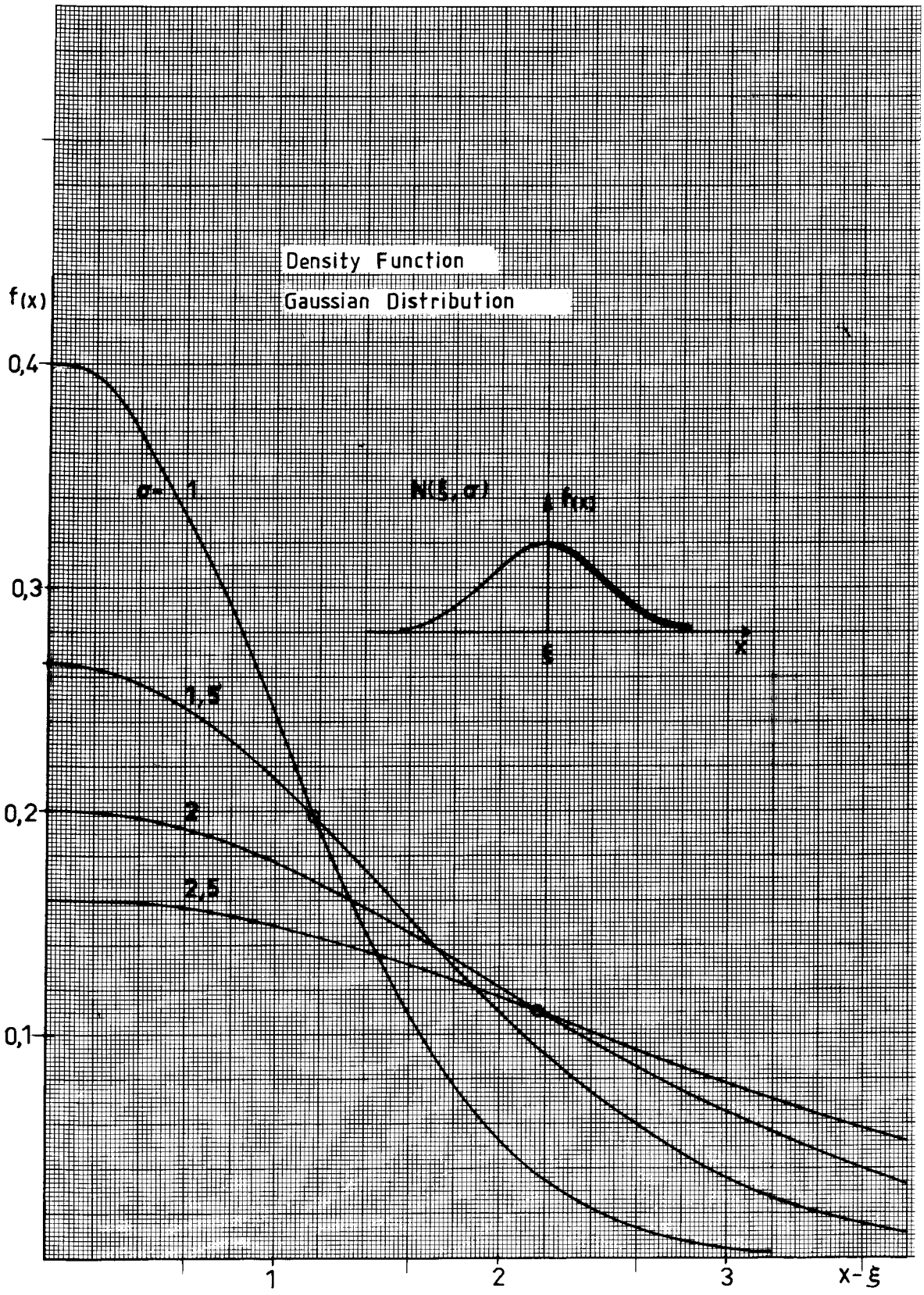
If the same observation scheme and the same parameter vector are used in all epochs, then the same design matrix and the same weight matrix result. In this case a multi-variate model can be formulated, which has certain advantages as to the amount of computations required and to the power of the applicable statistical tests, since the congruency test can be based on the Wishart-distribution (KOCH and FRITSCH, 1981).

With the aid of a special selection of the parameters of the general model of Eq. (11-7) strain accumulation or deformation velocities can be estimated. These parameters are of particular interest in geophysical and glaciological applications of deformation analysis. Apart from some new terms these special approaches do not contain new elements.

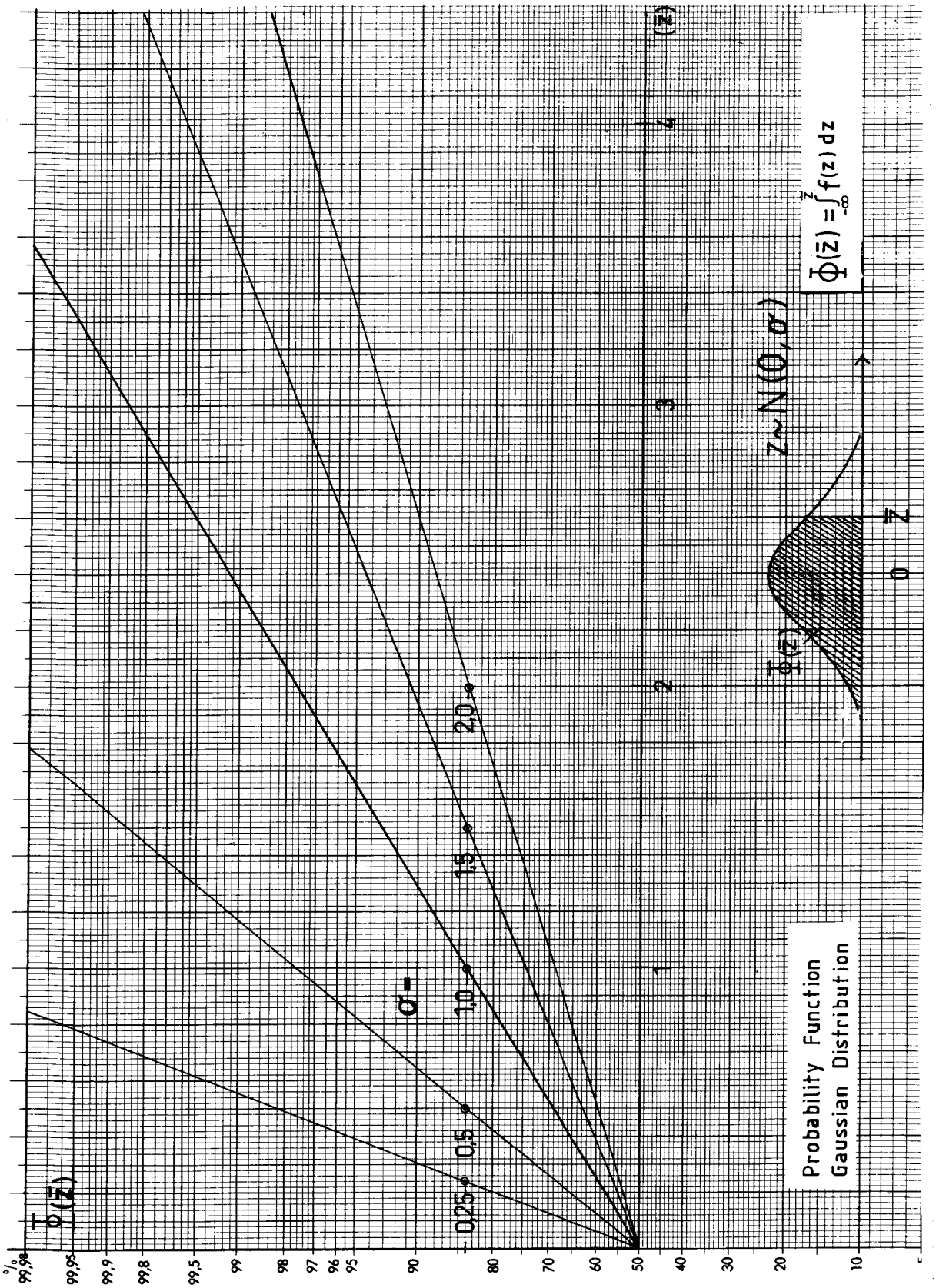
For many projects a good graphical display of the deformations is worthier than a numerical analysis. The plot of movements-versus-time or subsidence-versus-time diagrams are typical examples. All graphical techniques of visualizing three-dimensional processes can be used for this purpose. The general merits of graphical methods as discussed in Section 10.8 in the context of two-epoch analyses apply here as well.

Finally it should be remembered that it is always worthwhile to consider whether or not a sequence of two-epoch comparisons serves the purpose of a particular project better than a sophisticated and expensive multi-epoch approach. It must not be overlooked that a complicated method and a large size model can veil effects which could be detected easily in two-epoch comparisons. Therefore separate analyses in geometry and time domain may yield more realistic results in many cases. In general, simple and medium sized models have obvious advantages in practical applications over mathematically rigorous but oversized models.

12. NOMOGRAMS OF DISTRIBUTION FUNCTIONS

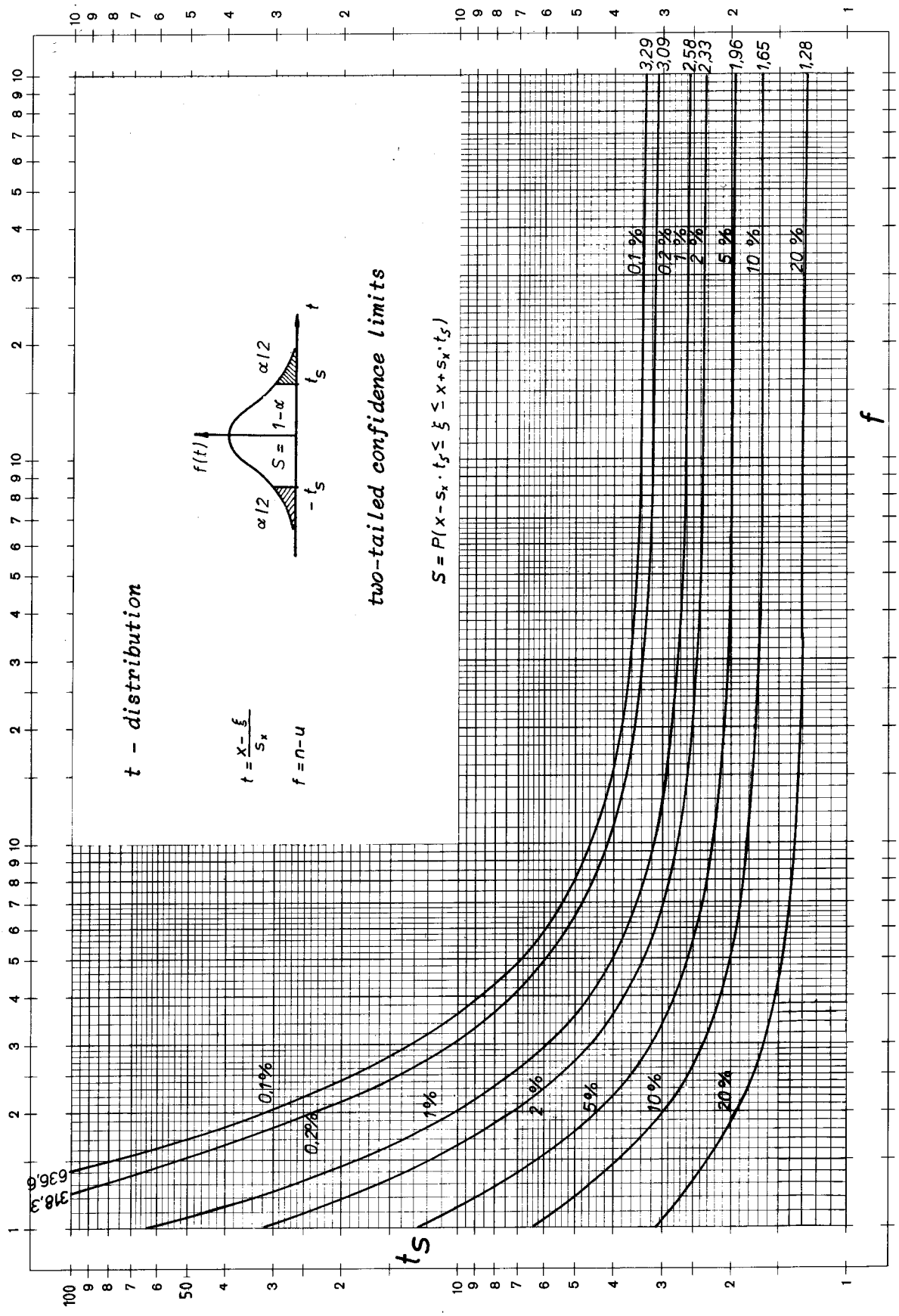


Nomogram 1: Density function of the Gaussian distribution. $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2}(\frac{x-\xi}{\sigma})^2$.

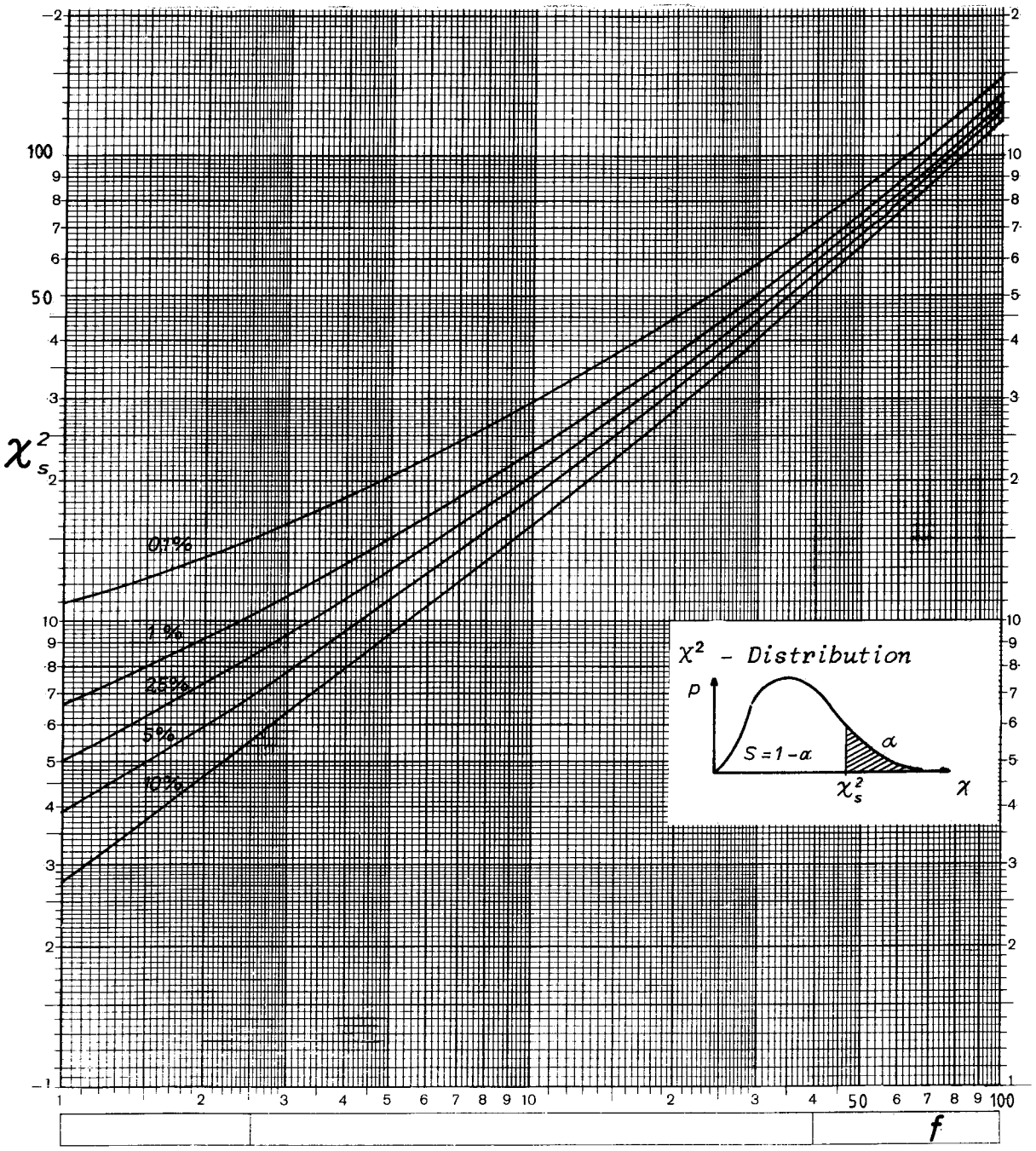


Nomogram 2: Probability function of the Gaussian distribution. Example: $\sigma = 1$, $\alpha = 5\%$ for a two-tailed test $\Rightarrow$ critical value:

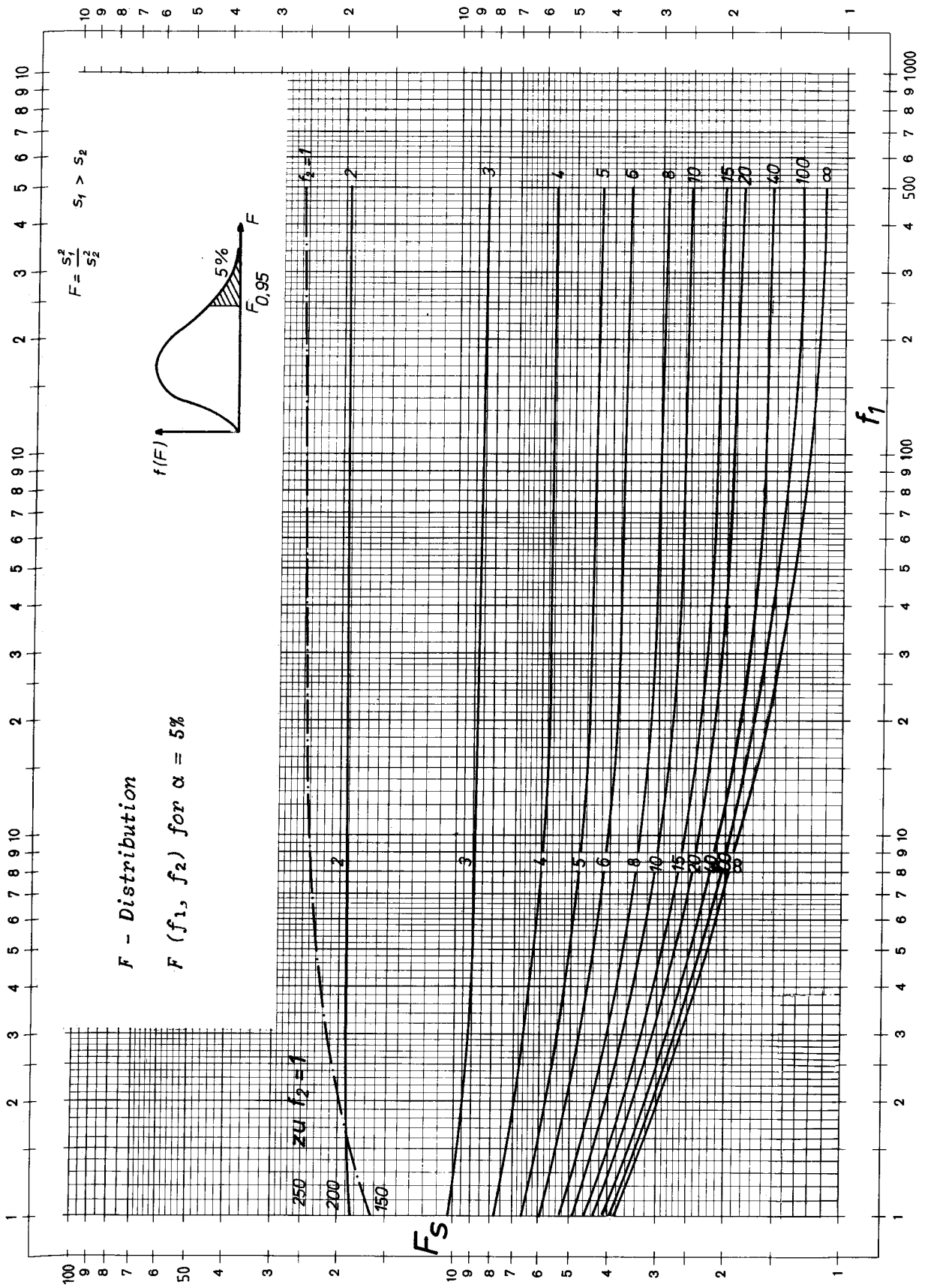
$$u_\alpha = \bar{z}_{97.5} = 1.96$$



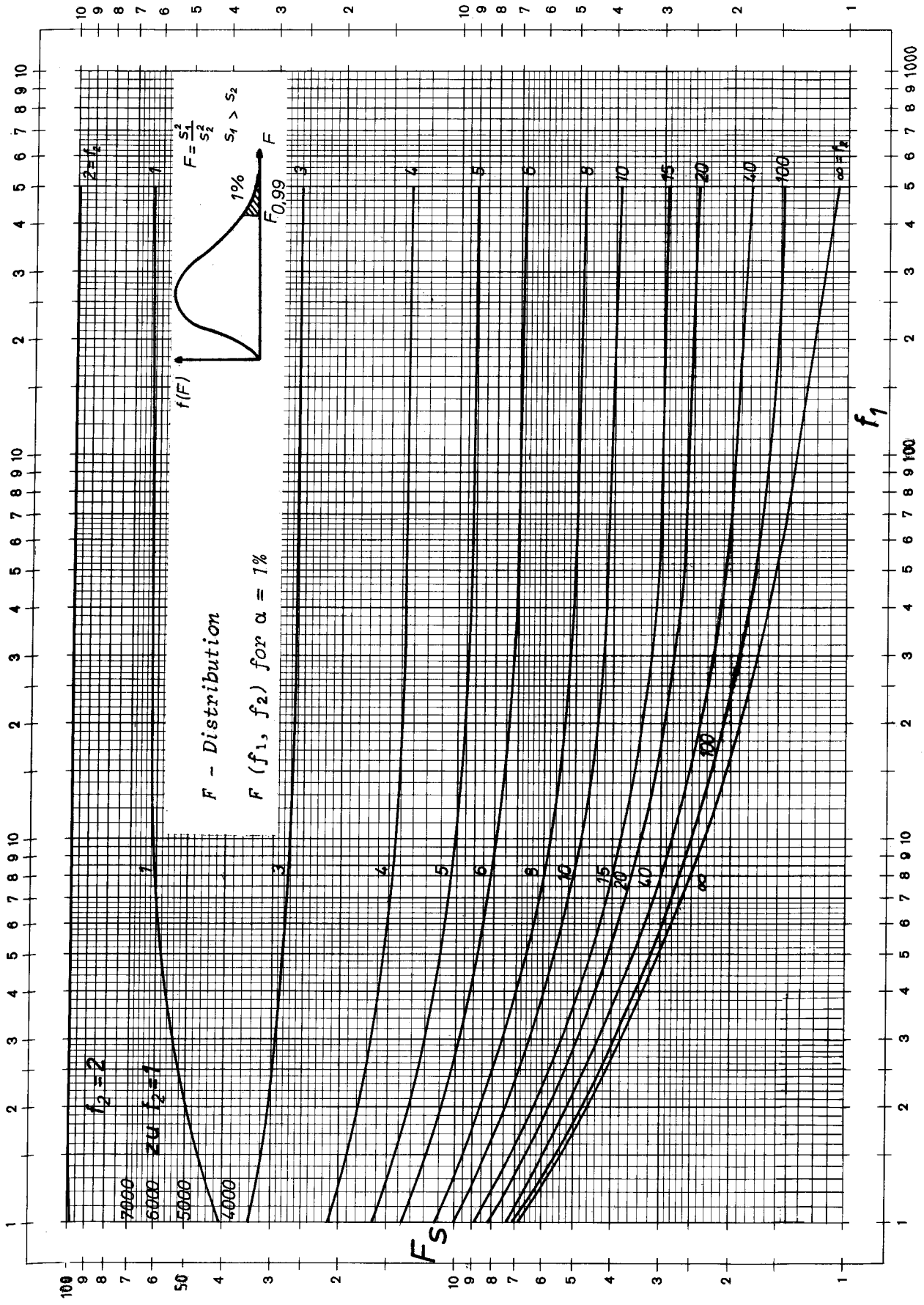
Nomogram 3: Two-tailed confidence limits for Student's t-distribution. Example: α = 5%, f = 10 for a two-tailed test $\Rightarrow$ critical value t_α = 2.23.



Nomogram 4: Upper confidence limits χ_s^2 of the χ^2-distribution. $\chi^2(f) = u_1^2 + u_2^2 + \ldots + u_f^2$ for $u_i \sim N(0, 1)$. Example: Test statistic $T = s^2/\sigma^2$, $\alpha = 5\%$, $f = 10 \Rightarrow$ critical value $\chi_\alpha^2 = 18.3$.



Nomogram 5: Upper confidence limits of Fisher's F-distribution with f_1 and f_2 degrees of freedom for α = 5%. Example: Test statistic $T = s_1^2/s_2^2$, $s_1 > s_2$, $f_1 = 10$, $f_2 = 20$, $\alpha = 5\% \Rightarrow$ critical value $F_\alpha(f_1, f_2) = 2.35$.



Nomogram 6: Upper confidence limits of Fisher's F-distribution with f_1 and f_2 degrees of freedom for α = 1%. Example: Test statistic $T = s_1^2/s_2^2$, $s_1 > s_2$, $f_1 = 10$, $f_2 = 20$, α = 1% $\Rightarrow$ critical value $F_\alpha(f_1, f_2) = 3.37$.

13. REFERENCES

Since the monograph is written in English for the English speaking community, preference has been given to English references. The author is well aware of the fact that many of the basic ideas have been first published in German. However, preference is given to good English references, if available.

13.1 Textbooks

BJERHAMMAR, A. (1973), Theory of Errors and Generalised Matrix Inverses, Elsevier Scientific Publishing Company, Amsterdam, London, New York.

HELMERT, F.R. (1924), Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate, 3. Aufl., Leipzig/Berlin.

HUBER, P.J. (1981), Robust Statistics, John Wiley & Sons, New York.

KOCH, K.R. (1980), Parameterschätzung und Hypothesentests in linearen Modellen, Dümmler-Verlag, Bonn.

LAWSON, C.L. and HANSON, R.J. (1974), Solving Least Squares Problems, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

RAO, C.R. (1973), Linear Statistical Inference and its Application, John Wiley & Sons, New York.

SEARLE, S.R. (1971), Linear Models, John Wiley & Sons, New York.

13.2 Proceedings and Collected Papers

FIG International Symposia on Deformation Measurements by Geodetic Methods.

1st Cracow, Poland, 1975

2nd Bonn, Germany, 1978, published by Konrad Wittwer Verlag, Stuttgart, 1981

3rd Budapest, Hungary, 1982

4th Kattowice, Poland, 1985

Fourth Canadian Symposium on Mining Surveying and Deformation Measurements, Banff, 1982, published by The University of Calgary, Dept. of Surveying Engineering.

Grafarend, E.W. and Sanso, F. (eds.), 1985, Optimization and Design of Geodetic Networks, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1985.

Kovari, K. (ed.), 1984, Field Measurements in Geomechanics, Int. Symp. Zurich 1983, publ. by A.A. Balkema, Rotterdam, 1984.

Schweizerische Talsperren Kommission, 1946, Messungen, Beobachtungen und Versuche an Schweizerischen Talsperren 1919-1946, Eidgenössisches Oberbauinspektorat, Bern, 1946.

Seminare über Deformationsanalysen, Schriftenreihe Wiss. Studiengang Vermessungswesen, Universität der Bundeswehr, München, Heft 4: Deformationsanalysen, 1979, Heft 9: Deformationsanalysen, 1983.

Swiss National Committee on Large Dams, Behaviour of Large Swiss Dams, Bern, 1964.

Symposium on the Surveillance of Engineering Structures, Dept. of Surv., University of Melbourne, 1983.

13.3 Monographs

AESCHLIMANN, H. (1971), Zur Genauigkeit geodätischer Verschiebungsmessungen, Mitteilungen aus dem Institut für Geodäsie und Photogrammetrie an der ETH Zürich, Nr.19.

BAARDA, W. (1967), Statistical Concepts in Geodesy, Neth. Geod. Comm., Publ. on Geod., New Series 2, No. 4, Delft.

BAARDA, W. (1968), A Testing Procedure for Use in Geodetic Networks, Neth. Geod. Comm., Publ. on Geod., New Series 2, No. 5, Delft.

BAARDA, W. (1973), S-Transformations and Criterion Matrices, Neth. Geod. Comm., Publ. on Geod., New Series 5, No. 1, Delft.

CHEN, Y.Q. (1983), Analysis of Deformation Surveys - A Generalised Method, Techn. Rep. No. 94, UNB, Dept. Surv. Eng., Fredericton, Canada.

KAVOURAS, M. (1982), On the Detection of Outliers and the Reliability of Geodetic Networks, Techn. Rep. No. 87, UNB, Dept. Surv. Eng., Fredericton, Canada.

LANG, W. (1929), Deformationsmessungen an Staumauern nach den Methoden der Geodäsie, Verlag der Abteilung für Landestopographie, Bern.

NIEMEIER, W. (1979), Zur Kongruenz mehrfach beobachteter geodätischer Netze, Wiss. Arb. Fachr. Verm., Universität Hannover, Nr. 88.

PELZER, H. (1971), Zur Analyse geodätischer Deformationsmessungen, Deutsche Geod. Komm., Reihe C., Heft 164, München.

POPE, A.J. (1976), The Statistics of Residuals and the Detection of Outliers, U.S. Dept. of Com., PB-258 428, National Ocean Survey, Rockville, Md.

SCHNEIDER, D. (1982), Complex Crustal Strain Approximation, Inst. f. Geod. u. Photogr. ETH Zurich, Nr. 33 and Tech. Rep. No. 91, UNB, Dept. of Surv. Eng., Fredericton, Canada.

SCHWINTZER, P. (1984), Analyse geodätisch gemessener Punklageänderungen mit gemischten Modellen, Schriftenreihe Wiss. Studiengang Vermessungswesen, Universität der Bundeswehr, München, Heft 12.

SNAY, R.A. and CLINE, M.W. (1980), Crustal Movement Investigation at Tejon Ranch, California, NOAA Techn. Rep. NOS87, NGS18, Rockville, Md.

13.4 Articles

ALBERDA, J.E. (1980), A Review of Analysis Techniques for Engineering Survey Control Schemes, Proc. Industrial and Engineering Survey Conference, London, Paper 6.2.

ARCA, S., BONASIA, V., DE NATALE, G., PINGUE, F. and SCARPA, R. (1985), A Multiple Fault System as the Fracture Mechanism for the 1976-1977 Friuli Earthquakes from Inversion of Geodetic Data, Boll. di Geod. e Sci. Affini, Vol. 44, 101-123.

BRUNNER, F.K., COLEMAN, R. and HIRSCH, B. (1981), A Comparison of Computation Methods for Crustal Strains from Geodetic Measurements, Tectonophysics (71), 281-298.

CASPARY, W. and SCHWINTZER, P. (1981), Bestimmung von Einzelpunktbewegungen und von Relativbewegungen zweier Netzteile in geodätischen Deformationsnetzen, Zeitschr. f. Verm., Vol. 106, (ZfV) 277-288.

CASPARY, W. (1982), Some Aspects Concerning the Datum of Geodetic Networks for Deformation Analysis, Proc. 3rd Int. on Symp. Deformation Measurements, Budapest, Vol. 3, 123-135.

CASPARY, W. (1984), Deformation Analysis Using a Special Similarity Transformation, Proc. FIG Int. Eng. Surv. Conf., Washington, 145-151 (published by American Congress on Surveying and Mapping, ACSM).

CASPARY, W. and BORUTTA, H. (1985), Robust Estimation as Applied to Deformation Analysis, Proc. 4th FIG Int. Symp. Geod. Meas. of Deformations, Kattowice, 283-294.

CHRZANOWSKI, A. and CHEN, Y.Q. (1986), Report of the Ad-hoc Committee on the Analysis of Deformation Surveys, Proc. 17th FIG Int. Congr., Toronto, Paper 608.1.

CROSILLA, F. and MARCHESINI, C. (1983), Geodetic Network Optimisation for the Detection of Crustal Movements Using a Mekometer, Boll. di Geod. e Sci. Affini, Vol. 42, 301-314.

CROSS, P.A. (1983a), Computer Aided Design of Geodetic Networks, Proc. 18th Gen. Assembly of IUGG (IAG), Hamburg, SSG 1.59.

CROSS, P.A. (1983b), A Directory of Software for the Computer Aided Design of Geodetic Networks, Proc. 18th Gen. Assembly of IUGG (IAG), Hamburg, SSG 1.59.

DORRER, E. (1971), Movement of the Ward Hunt Ice Shelf, Ellesmere Island, N.W.T., Canada, Journ. of Glac., Vol. 10, 211-225.

FOERSTNER, W. (1979), On the Internal and External Reliability of Photogrammetric Coordinates, Proc. 45th Annual Meeting, American Society of Photogrammetry, 18-24 March 1979, Washington, D.C., Vol. 1, 294-310.

FOERSTNER, W. (1981), On the Reliability of 3rd and 4th Order Network Densifications, Proc. IAG Symp. on Geod. Netw. and Comp., Munich, publ. by Deutsche Geod. Komm., Reihe B, Heft 258/V, 68-84.

FRASER, C.S. (1984), Network Design Considerations for Non-topographic Photogrammetry, Photogr. Eng. and Rem. Sens., Vol. 50, 1115-1126,

GHITAU, D. (1973), Über den Aufbau eines allgemeinen Modells zur Beschreibung von Landhebungen auf Grund von Wiederholungs-Nivellements, Zeitschr. f. Verm. (ZfV), Vol. 98, 28-36.

GRAFAREND, E.W. (1981), Optimisation of Geodetic Networks, Proc. IAG Symp. on Geod. Netw. and Comp., Munich, publ. by Deutsche Geod. Komm., Reihe B, Heft 258/III, 69-81

GRAFAREND, E.W. and KRUMM, F. (1983), Criterion Matrices of Heterogeneously Observed Three-dimensional Networks, Proc. 18th Gen. Assembly of IUGG (IAG), Hamburg.

GRAFAREND, E.W. (1984), Criterion Matrices for Deforming Networks, in: Grafarend & Sanso (eds.) Optimisation and Design of Geodetic Networks, Springer-Verlag, Berlin.

GRUENDIG, L. and BAHNDORF, J. (1985), Sequential Optimisation of Geodetic Networks with Respect to Accuracy and Reliability, Proc. 7th Int. Symp. on Geod. Comp., Cracow, Poland (in print).

HECK, B., KOK, J., WELSCH, W., BAUMER, R., CHRZANOWSKI, A., CHEN, Y.Q. and SECORD, J.M. (1982), Report of the FIG Working Group on the Analysis of Deformation Measurements, Proc. 3rd FIG Symp. on Def. Meas., Budapest, Vol. 3, 217-260.

HOLDAHL, S.H. (1978), Models for Extracting Vertical Crustal Movements from Levelling Data, Proc. 9th GEOP Res. Conf., Columbus, Ohio, 183-190.

HUBER, P.J. (1964), Robust Estimation of a Location Parameter, Ann. Math. Stat. (35), 73-101.

IMAMURA, A. (1934) , Further Notes on the Northward Movement of Crustal Deformation Along the Western Boundary of Kwanto Plain with Special Reference to the Block Movement Responsible for the Disastrous Earthquake of September 21, 1931, Jap. Journ. Astron. and Geoph. (XI), 95-112.

JENKINS, D.M. and FUNNEL, K.W. (1971), Dam Deformation Surveys in Tasmania, Proc. 14th Australian Survey Congress, The Inst. of Surv. Australia, Sydney, 120-138.

KOBOLD, F. (1958), Geodätische Methoden zur Bestimmung von Geländebewegungen und von Deformationen an Bauwerken, Schweiz. Bauzeitung, Vol. 76, 182-187.

KOBOLD, F. (1962), Measurement of Displacement and Deformation by Geodetic Methods, Trans. ASCE, IV, 163-192.

KOBOLD, F. (1969), Neues über Deformationsmessungen an Staumauern, Zeitschr. f. Verm., Vol. 94, (ZfV), 496-504.

KOCH, K.R. and FRITSCH, D. (1981), Multivariate Hypothesis Tests for Detecting Recent Crustal Movements, Tectonophysics (71), 301-313.

KOK, J.J. (1982), Statistical Analysis of Deformation Problems Using Baarda's Testing Procedures, Daar heb ik veertig jaar over nagedacht., Feestbundel ter gelegenheid van de 65 ste verjaardag van Professor Baarda, Delft, Deel II, 470-488.

KRARUP, T., JUHL, J. and KUBIK, K. (1980), Götterdämmerung over Least Squares Adjustment, Proc. 14th ISP Congr., Hamburg, 369-378.

LAZZARINI, T. (1966), Die gegenwärtigen polnischen Methoden und Erfahrungen bei geodätischen Deformationsmessungen an Bauwerken, Deutsche Geod. Komm., Reihe B, Heft 123, München, 1-17.

MARGRAVE, G.F. and NYLAND, E. (1980), Strain from Repeated Geodetic Surveys by Generalised Inverse Methods, Can. Journ. Earth Sci., Vol. 17, 1010-1020.

MEPHAM, M.P. and KRAKIWSKY, E.J. (1984), CANDSN: A Computer Aided Network Design and Adjustment System, Can. Surv., Vol. 38, 99-114.

van MIERLO, J. (1981), Second Order Design: Precision and Reliability Aspects, Allg. Verm. Nachr. (AVN), Vol. 88, 95-101.

van MIERLO, J. (1982), Difficulties in Defining the Quality of Geodetic Networks, in: Survey Control Network, Proc. Meeting FIG Study Group 5B, Aalboog, Denmark, publ. Schriftenreihe Wiss. Studiengang Vermessungswesen, Universität der Bundeswehr, München, Heft 7, 259-274.

NEY, B. (1975), Polynomial Model of Vector Dislocations Field of Horizontal Geodetic Network, Proc. 1st FIG Symp. on Deformation Measurements, Cracow, Poland, 225-237

NIEMEIER, W. (1981), Statistical Tests for Detecting Movements in Repeatedly Measured Geodetic Networks, Tectonophysics, Vol. 71, 335-351.

NIEMEIER, W., TESKY, W.F. and LYALL, R.G. (1982), Precision, Reliability and Sensitivity Aspects of an Open Pit Monitoring Network, Proc. 4th Can. Symp. Mining Surv. and Def. Meas., Banff, 409-430.

PAPO, H. and PERELMUTER, A. (1981), Kinematic Analysis of Deformations, Personal Paper, 16th FIG Congr., Montreux, Switzerland.

PAPO, H. and PERELMUTER, A. (1982), Deformations as Reflected in the Kinematics of a Network of Points, Proc. 3rd FIG Symp. on Def. Meas., Budapest, Vol. 1, 123-135.

PENMAN, A.D.M. and KENNARD, M.F. (1982), Long-term Monitoring of Embankment Dams in Britain, Water Power & Dam Constr., Nov. 82, Vol. 11, 19-26.

PENEV, P. (1983), Optimisation of Geodetic Networks for Defining Deformations of Engineering Structures, Proc. 17th FIG Congr., Sofia, Paper 608.4.

POPE, A.J., STEARN, J.L. and WHITTEN, C.A. (1966), Surveys for Crustal Movement Along the Hayward Fault, Bull. Seism. Soc. Am., Vol. 56, 317-323.

PRESCOTT, W.H., SAVAGE, J.C. and KINOSHITA, W.T. (1979), Strain Accumulation Rates in the Western United States Between 1970 and 1978, Journ. of Geoph. Res. (84) 5423-5435.

RAO, C.R. (1970), Estimation of Heteroscedastic Variances in Linear Models, Journ. Am. Stat. Ass. (63), 161-172.

RICHARDUS, P. (1964), Triangulation oder Polygonierung im Lichte der mathematischen Statistik, Geodätische Messungen der horizontalen Komponenten von Geländedebewegungen und Deformationen an Bauwerken, Zeitschr. f. Verm. (ZfV), Vol. 89, 378-387 and 461-465.

SAVAGE, J.C. and BURFORD, R.O. (1970), Accumulation of Tectonic Strain in California, Bull. Seism. Soc. Am., (60), 1877-1896.

SCHMITT, G. (1982), Optimisation of Geodetic Networks, Rev. of Geoph. and Space Phys. (20), 877-884.

SCHWINTZER, P. (1982), Zur Generalisierung von Deformationsvektoren mit gemischten Modellen, Proc. 3rd FIG Symp. on Def. Meas., Budapest, Vol. 1, 183-194.

SNAY, R.A. and GERGEN, J.G. (1978), Monitoring Regional Crustal Deformation with Horizontal Geodetic Data, Proc. 9th GEOP Res. Conf., Columbus, Ohio, 87-92.

TESKY, W.F. and GRUENDIG, L. (1985), Improving the Quality of Traverses, Can. Surv. (39), 211-222.

ULBRICH, K. (1956), Geodätische Deformationsmessungen an österreichischen Staumauern und Grossbauwerken, Oester. Zeitschr. f. Verm., Sonderheft 17.

VANICEK, P., ELLIOT, M.R. and CASTLE, R.O. (1979), Four-Dimensional Modelling of Recent Vertical Movements in the Area of the Southern California Uplift, Tectonophysics (52), 287-300.

WELSCH, W.M. (1981), Description of Homogeneous Horizontal Strains and some Remarks to Their Analysis, Proc. IAG Symp. on Geod. Netw. and Comp., Munich, publ. by Deutsche Geod. Komm. Reihe B, Heft 258/V, 188-205.

WELSCH, W.M. (1985), Some Aspects of the Analysis of Geodetic Strain Observations in Kinematic Models, Proc. Int. Symp. on Recent Crustal Movements in Central and South America, Maracaibo.

WHITTEN, C.A. (1967), Geodetic Network VersusTime, Bull.Geod. (84), 109-116.

Publications from the

SCHOOL OF GEOMATIC ENGINEERING

THE UNIVERSITY OF NEW SOUTH WALES
ABN 57 195 873 179

To order, write to:
Publications Officer, School of Geomatic Engineering
The University of New South Wales, UNSW SYDNEY NSW 2052, AUSTRALIA

NOTE: ALL ORDERS MUST BE PREPAID. CREDIT CARDS ARE ACCEPTED.
ASK FOR OUR CREDIT CARD ORDER FORM.

UNISURV REPORTS - S SERIES

(Prices effective August 2000)

Australian Prices:	S8 - S20		$11.00
(include postage and GST)	S29 onwards	Individuals	$27.50
		Institutions	$33.00
Overseas Prices:	S8 - S20		$10.00
(include postage by surface mail)	S29 onwards	Individuals	$25.00
(Air mail rates on application.)		Institutions	$30.00

S8. A. Stolz, "Three-D Cartesian Co-ordinates of Part of the Australian Geodetic Network by the Use of Local Astronomic Vector Systems", Unisurv Rep. S8, 182 pp, 1972.

S10. A. J. Robinson, "Study of Zero Error and Ground Swing of the Model MRA101 Tellurometer", Unisurv Rep. S10, 200 pp, 1973.

S12. G. J. F. Holden, "An Evaluation of Orthophotography in an Integrated Mapping System", Unisurv Rep. S12, 232 pp, 1974.

S14. E. G. Anderson, "The Effect of Topography on Solutions of Stokes` Problem", Unisurv Rep. S14, 252 pp, 1976.

S16. K. Bretreger, "Earth Tide Effects on Geodetic Observations", Unisurv S16, 173 pp, 1978.

S17. C. Rizos, "The Role of the Gravity Field in Sea Surface Topography Studies", Unisurv S17, 299 pp, 1980.

S18. B. C. Forster, "Some Measures of Urban Residential Quality from LANDSAT Multi-Spectral Data", Unisurv S18, 223 pp, 1981.

S19. R. Coleman, "A Geodetic Basis for Recovering Ocean Dynamic Information from Satellite Altimetry", Unisurv S19, 332 pp, 1981.

S20. D. R. Larden, "Monitoring the Earth's Rotation by Lunar Laser Ranging", Unisurv Report S20, 280 pp, 1982.

S29. G. S. Chisholm, "Integration of GPS into Hydrographic Survey Operations", Unisurv S29, 190 pp, 1987.

S30. G. A. Jeffress, "An Investigation of Doppler Satellite Positioning Multi-Station Software", Unisurv S30, 118 pp, 1987.

S31. J. Soetandi, "A Model for a Cadastral Land Information System for Indonesia", Unisurv S31, 168 pp, 1988.

S33. R. D. Holloway, "The Integration of GPS Heights into the Australian Height Datum", Unisurv S33, 151 pp,1988.

S34. R. C. Mullin, "Data Update in a Land Information Network", Unisurv S34, 168 pp, 1988.

S35. B. Merminod, "The Use of Kalman Filters in GPS Navigation", Unisurv S35, 203 pp, 1989.

S36. A. R. Marshall, "Network Design and Optimisation in Close Range Photogrammetry", Unisurv S36, 249 pp, 1989.

S37. W. Jaroondhampinij, "A model of Computerised Parcel-Based Land Information System for the Department of Lands, Thailand," Unisurv S37, 281 pp, 1989.

S38. C. Rizos (Ed.), D. B. Grant, A. Stolz, B. Merminod, C. C. Mazur "Contributions to GPS Studies", Unisurv S38, 204 pp, 1990.

S39. C. Bosloper, "Multipath and GPS Short Periodic Components of the Time Variation of the Differential Dispersive Delay", Unisurv S39, 214 pp, 1990.

S40. J. M. Nolan, "Development of a Navigational System Utilizing the Global Positioning System in a Real Time, Differential Mode", Unisurv S40, 163 pp, 1990.

S41. R. T. Macleod, "The Resolution of Mean Sea Level Anomalies along the NSW Coastline Using the Global Positioning System", 278 pp, 1990.

S42. D. A. Kinlyside, "Densification Surveys in New South Wales - Coping with Distortions", 209 pp, 1992.

S43. A. H. W. Kearsley (Ed.), Z. Ahmad, B. R. Harvey and A. Kasenda, "Contributions to Geoid Evaluations and GPS Heighting", 209 pp, 1993.

S44. P. Tregoning, "GPS Measurements in the Australian and Indonesian Regions (1989-1993)", 134 + xiii pp, 1996.

S45. W.-X. Fu, "A Study of GPS and Other Navigation Systems for High Precision Navigation and Attitude Determinations", 332 pp, 1996.

S46. P. Morgan et al, "A Zero Order GPS Network for the Australia Region", 187 + xii pp, 1996.

S47. Y. Huang, "A Digital Photogrammetry System for Industrial Monitoring", 145 + xiv pp, 1997.

S48. K. Mobbs, "Tectonic Interpretation of the Papua New Guinea Region from Repeat Satellite Measurements", 256 + xc pp, 1997.

S49. S. Han, "Carrier Phase-Based Long-Range GPS Kinematic Positioning", 185 + xi pp, 1997.

S50. M. D. Subari, "Low-cost GPS Systems for Intermediate Surveying and Mapping Accuracy Applications", 179 + xiii pp, 1997.

S51. L.-S. Lin, "Real-Time Estimation of Ionospheric Delay Using GPS Measurements", 199 + xix pp, 1997.

S52. M. B. Pearse, "A Modern Geodetic Reference System for New Zealand", 324 + xviii pp, 1997.

S53. D. B. Lemon, "The Nature and Management of Positional Relationships within a Local Government Geographic Information System", 273 + xvi pp, 1997.

S54. C. Ticehurst, "Development of Models for Monitoring the Urban Environment Using Radar Remote Sensing", 282 + xix pp, 1998.

S55. S. S. Boey, "A Model for Establishing the Legal Traceability of GPS Measurements for Cadastral Surveying in Australia",186 + xi pp, 1999.

S56. P. Morgan and M. Pearse, "A First-Order Network for New Zealand", 134 + x pp, 1999.

S57. P. N. Tiangco, "A Multi-Parameter Radar Approach to Stand Structure and Forest Biomass Estimation", 319 + xxii pp, 2000.

MONOGRAPHS

Australian and overseas prices include postage by surface mail.
Air mail rates for overseas orders on application.
Australian prices include GST.

(Prices effective August 2000)

		Price Australia (incl.GST)	**Price** Overseas
M1.	R. S. Mather, "The Theory and Geodetic Use of some Common Projections", (2nd edition), 125 pp, 1978.	$ 16.50	$ 15.00
M2.	R. S. Mather, "The Analysis of the Earth's Gravity Field", 172 pp, 1971.	$ 8.80	$ 8.00
M3.	G. G. Bennett, "Tables for Prediction of Daylight Stars", 24 pp, 1974.	$ 5.50	$ 5.00
M4.	G. G. Bennett, J. G. Freislich & M. Maughan, "Star Prediction Tables for the Fixing of Position", 200 pp, 1974.	$ 8.80	$ 8.00
M8.	A. H. W. Kearsley, "Geodetic Surveying", 96 pp, 1988.	$ 13.20	$ 12.00
M11.	W. F. Caspary, "Concepts of Network and Deformation Analysis",183 pp, 2000.	$ 27.50	$ 25.00
M12.	F. K. Brunner, "Atmospheric Effects on Geodetic Space Measurements", 110 pp, 1988.	$ 17.60	$ 16.00
M13.	B. R. Harvey, "Practical Least Squares and Statistics for Surveyors", (2nd edition), 319 pp, 1994.	$ 33.00	$ 30.00
M14.	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management", 269 pp, 1991. (Proceedings LIM Conference, July 1991).	$ 22.00	$ 20.00
M15/1	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management - Geographic Information Systems - Advance Remote Sensing Vol. 1", 295 pp, 1993 (Proceedings of LIM & GIS Conference, July 1993).	$ 33.00	$ 30.00
M15/2	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management - Geographic Information Systems - Advance Remote Sensing Vol. 2", 376 pp, 1993 (Proceedings of Advanced Remote Sensing Conference, July 1993).	$ 33.00	$ 30.00
M16.	A. Stolz, "An Introduction to Geodesy", 112 pp, 1994.	$ 22.00	$ 20.00
M17.	C. Rizos, "Principles and Practice of GPS Surveying", 565 pp, 1997.	$ 46.20	$ 42.00

Credit Card Orders / Payment

from / to

School of Geomatic Engineering

The University of New South Wales

Name and postal address of ordering person/organisation

_____

_____

_____

_____

_____

Date : _____

Day time phone number : _____

Credit Card debit for

_____ $

_____ $

_____ $

_____ $

Total $ _____

Please debit my credit card: ☐ Mastercard ☐ Visa ☐ Bankcard

Card Name: _____ Signature: _____

Card No: | | | | | | | | | | | | | | | | Expire Date : _____

Note: UNSW requires that you attach a photocopy of your credit card to this order