Network quality control

2nd edition

Teunissen, P.J.G.

DOI
10.59490/tb.100

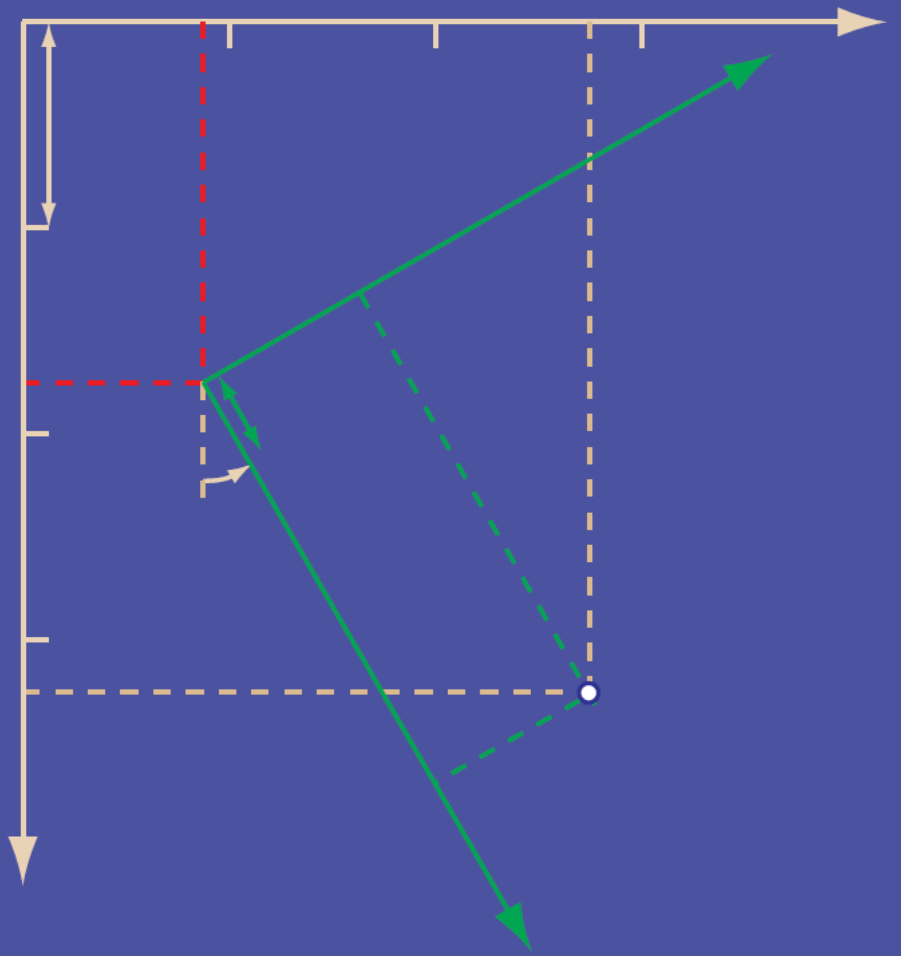
Publication date
2024

Document Version
Final published version

Citation (APA)
Teunissen, P. J. G. (2024). *Network quality control: 2nd edition*. TU Delft OPEN Publishing. https://doi.org/10.59490/tb.100

Network quality control

Peter J.G. Teunissen

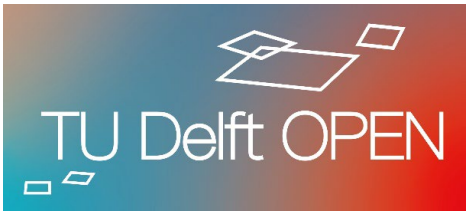


NETWORK QUALITY CONTROL

P.J.G. Teunissen

Series on Mathematical Geodesy and Positioning

ISBN: 978-94-6366-950-4 (Ebook)
ISBN: 978-94-6366-949-8 (Paperback/softback)
DOI: https://10.59490/tb.100



Keywords: precision, testing, reliability, free networks, connected networks, constrained networks

Preface 2nd Edition

To promote open access, this new edition of *Network Quality Control* is published by TU Delft Open Publishing instead of Delft Academic Press. The book builds on the foundations of adjustment theory and testing theory to enable precise and reliable designs of geodetic networks and their connections. As such, the book is a natural follow-on of the books *Adjustment Theory* (2nd Ed. 2024) and *Testing Theory* (3rd Ed. 2024), both with TU Delft Open Publishing.

September, 2024

Peter J.G. Teunissen

Foreword

This book is the result of a series of lectures and courses the author has given on the topic of network analysis. During these courses it became clear that there is a need for reference material that integrates network analysis with the statistical foundations of parameter estimation and hypothesis testing. **Network quality control** deals with the qualitative aspects of network design, network adjustment, network validation and network connection, and as such conveys the necessary knowledge for computing and analysing networks in an integrated manner.

In completing the book, the author received valuable assistance from Ir. Hedwig Verhoef, Dr. Ir. Dennis Odijk and Ria Scholtes. Hedwig Verhoef has also been one of the lecturers and took care of editing a large portion of the book. This assistance is greatly acknowledged.

P.J.G. Teunissen
December 2006

Contents

Chapter 1

An overview

This introductory chapter gives an overview of the material presented in the book. The book consists of three parts. A first part on estimation theory, a second part on testing theory and a third part on network theory. The first two parts are of a more general nature. The material presented therein is in principle applicable to any geodetic project where measurements are involved. Most of the examples given however, are focussed on the network application. In the third part, the computation and validation of geodetic networks is treated. In this part, we make a frequent use of the material presented in the first two parts. In order to give a bird's eye view of the material presented, we start with a brief overview of the three parts.

ADJUSTMENT: The need for an adjustment arises when one has to solve an inconsistent system of equations. In geodesy this is most often the case, when one has to solve a redundant system of observation equations. The adjustment principle used is that of least-squares. A prerequisite for applying this principle in a proper way, is that a number of basic assumptions need to be made about the input data, the measurements. Since measurements are always uncertain to a certain degree, they are modeled as sample values of a random vector, the m-vector of observables $\underline{y}$ (*note:* the underscore will be used to denote random variables). In case the vector of observables is normally distributed, its distribution is uniquely characterized by the first two (central) moments: the expectation (or mean) $E\{\underline{y}\}$ and the dispersion (or variance) $D\{\underline{y}\}$. Information on both the expectation and dispersion needs to be provided, before any adjustment can be carried out.

Functional model: In case of geodetic networks, the observables contain information on the relative geometry of the network points. Examples are: height differences, angles, distances, baselines, etc. Knowing the information content of the observables, allows one to link them to the parameters which are used for describing the geometry of the network. These parameters, which are often coordinates, are to be determined from the adjustment. The link between the observables and the n-vector of unknown parameters x, is established by means of the system of m observation equations

$$E\{\underline{y}\} = Ax$$

This system is referred to as the functional model. It is given once the design matrix A of order $m \times n$ is specified.

The system as it is given here, is linear in x. Quite often however, the observation equations are nonlinear. In that case a linearization needs to be carried, to make the system linear again. The parameter vector x usually consists of coordinates and possibly, additional nuisance parameters, such as for instance orientation unknowns in case of theodolite measurements. The coordinates could be of any type. For instance, they could

be Cartesian coordinates or geographic coordinates. The choice of the type of coordinates is not essential for the adjustment, but is more a matter of convenience and depends on what is required for the particular application at hand.

Stochastic model: Measurements are intrinsically uncertain. Remeasurement of the same phenomenon under similar circumstances, will usually give slightly different results. This variability in the outcomes of measurements is modelled through the probability density function of $\underline{y}$. In case of the normal distribution, it is completely captured by its dispersion. In order to properly weigh the observables in the adjustment process, the dispersion needs to be specified beforehand. It is given as

$$D\{\underline{y}\} = Q_y$$

This is the stochastic model, with Q_y being the $m \times m$ variance matrix of the observables. In these lecture notes, we will assume that Q_y is known. Hence, unknown variance components and their estimation are not treated.

Since the variance matrix describes the variability one can expect of the measurements when they are repeated under similar circumstances, it is said to describe the precision of the observables. In order to be able to specify Q_y correctly, a good understanding of the measurement equipment and the measurement procedures used, is needed. Quite often the variance matrix Q_y can be taken as a diagonal matrix. This is the case, when the measurements have been obtained independently from one another. The variance matrix becomes full (nondiagonal) however, when for instance, the measurements themselves are the result of a previous adjustment. This is the case when connecting geodetic networks.

Least-squares: Once the measurements have been collected and the functional model and the stochastic model have been specified, the actual adjustment can be carried out. The least-squares estimator of the unknown parameter vector x, is given as

$$\underline{\hat{x}} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y}$$

It depends on the design matrix A, the variance matrix Q_y and the vector of observables $\underline{y}$. With $\underline{\hat{x}}$, one can compute the adjusted observables as $\underline{\hat{y}} = A\underline{\hat{x}}$ and the least- squares residuals as $\underline{\hat{e}} = \underline{y} - \underline{\hat{y}}$.

The above expression for the least-squares estimator is based on a functional model which is linear. In the nonlinear case, one will first have to apply a linearization before the above expression can be applied. For the linearization one will need approximate values for the unknown parameters. In case already an approximate knowledge on the geometry of the network is available, the approximate coordinates of the network points can be obtained from a map. If not, a minimum set of the observations themselves will have to be used for computing approximate coordinates. In case the approximate values of the unknown parameters are rather poor, one often will have to iterate the least-squares solution.

Quality: Every function of a random vector, is itself a random variable as well. Thus $\underline{\hat{x}}$ is a random vector, just like the vector of observables $\underline{y}$ is. And when $\underline{\hat{x}}$ is linearly related to $\underline{y}$, it will have a normal distribution whenever $\underline{y}$ has one. In that case also the

distribution of $\underline{\hat{x}}$ can be uniquely characterized by means of its expectation and dispersion. Its expectation reads

$$E\{\underline{\hat{x}}\} = x$$

Thus the expectation of the least-squares estimator equals the unknown, but sought for parameter vector x. This property is known as unbiasedness. From an empirical point of view, the equation implies, that if the adjustment would be repeated, each time with measurements collected under similar circumstances, then the different outcomes of the adjustment would on the average coincide with x. It will be clear, that this is a desirable property indeed.

The dispersion of $\underline{\hat{x}}$, describing its precision, is given as

$$D\{\underline{\hat{x}}\} = (A^T Q_y^{-1} A)^{-1}$$

This variance matrix is independent of $\underline{y}$. This is a very useful property, since it implies that one can compute the precision of the least-squares estimator without having the actual measurements available. Only the two matrices A and Q_y need to be known. Thus once the functional model and stochastic model have been specified, one is already in a position to know the precision of the adjustment result. It also implies, that if one is not satisfied with this precision, one can change it by changing A and/or Q_y. This is typically done at the design stage of a geodetic project, prior to the actual measurement stage. Changing the geometry of the network and/or adding/deleting observables, will change A. Using different measurement equipment and/or different measurement procedures, changes Q_y.

TESTING: Applying only an adjustment to the observed data is not enough. The result of an adjustment and its quality rely heavily on the validity of the functional and stochastic model. Errors in one of the two, or in both, will invalidate the adjustment results. One therefore needs, in addition to the methods of adjustment theory, also methods that allow one to check the validity of the assumptions underlying the functional and stochastic model. These methods are provided for by the theory of statistical testing.

Model errors: One can make various errors when formulating the model needed for the adjustment. The functional model could have been misspecified, $E\{\underline{y}\} \neq Ax$. The stochastic model could have been misspecified, $D\{\underline{y}\} \neq Q_y$. Even the distribution of $\underline{y}$ need not be normal. In these lecture notes, we restrict our attention to misspecifications in the functional model. These are by far the most common modelling errors that occur in practice. Denoting the model error as b, we have $E\{\underline{y}\} = Ax + b$. If it is suspected that model errors did indeed occur, one usually, on the basis of experience, has a fair idea what type of model error could have occurred. This implies that one is able to specify the vector b in the form of equations like

$$b = C\nabla$$

where C is a matrix of order $m \times q$ and ∇ is a q-vector. The vector ∇ is still unknown, but the matrix C is then known. This matrix specifies how the vector of observables is

assumed to be related to the unknown error vector ∇. A typical example of modelling errors that can be captured through this description are blunders in the measurements. In case of a single blunder in one of the measurements, the C-matrix reduces to a unit vector, having a one at the entry that corresponds with the corrupted measurement.

Test statistic: It will be intuitively clear that the least-squares residual vector $\underline{\hat{e}}$, must play an important role in validating the model. It is zero, when the measurements form a perfect match with the functional model, and it departs from zero, the more the measurements fail to match the model. A test statistic is a random variable that measures on the basis of the least-squares residuals, the likelihood that a model error has occurred. For a model error of the type $C\nabla$, it reads

$$\underline{T}_q = \underline{\hat{e}}^T Q_y^{-1} C (C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C)^{-1} C^T Q_y^{-1} \underline{\hat{e}}$$

It depends, apart from the least-squares residuals, also on the matrix C, on the design matrix A (through $Q_{\hat{e}}$) and on the variance matrix Q_y. The test statistic has a central Chi-squared distribution, with q degrees of freedom, $\chi^2(q,0)$, when the model error would be absent. When the value of the test statistic falls in the right tail-area of this distribution, one is inclined to belief that the model error indeed occurred. Thus the presence of the model error is believed to be likely, when $T_q > \chi^2_{\alpha_q}(q,0)$, where α_q is the chosen level of significance.

Testing procedure: In practice it is generally not only one model error one is concerned about, but quite often many more than one. In order to take care of these various potential modelling errors, one needs a testing procedure. It consists of three steps: detection, identification and adaptation. The purpose of the detection step is to infer whether one has any reason to belief that the model is wrong. In this step one still has no particular model error in mind. The test statistic for detection, reads

$$\underline{T}_{m-n} = \underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}}$$

One decides to reject the model, when $T_{m-n} > \chi^2_{\alpha_{m-n}}(m-n,0)$.

When the detection step leads to rejection, the next step is the identification of the most likely model error. The identification step is performed with test statistics like $\underline{T}_q$. It implies that one needs to have an idea about the type of model errors that are likely to occur in the particular application at hand. Each member of this class of potential model errors is then specified through a matrix C. In case of one dimensional model errors, such as blunders, the C-matrix becomes a vector, denoted as c. In that case $q = 1$ and the test statistic $\underline{T}_q$ simplifies considerably. One can then make use of its square-root, which reads

$$\underline{w} = \frac{c^T Q_y^{-1} \underline{\hat{e}}}{\sqrt{c^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c}}$$

This test statistic has a standard normal distribution $N(0,1)$ in the absence of the model error. The particular model error that corresponds with the vector c, is then said to have occurred with a high likelihood, when $\mid w \mid > N_{\frac{1}{2}\alpha_1}(0,1)$. In order to have the model error

detected and identified with the same probability, one will have to relate the two levels of significance, α_{m-n} and α_1. This is done by equating the power and the noncentrality parameters of the above two test statistics $\underline{T}_{m-n}$ and $\underline{w}$.

Once certain model errors have been identified as sufficiently likely, the last step consists of an adaptation of the data and/or model. This implies either a remeasurement of the data or the inclusion of additional parameters into the model, such that the model errors are accounted for. In both cases one always should check again of course, whether the newly created situation is acceptable or not.

Quality: In case a model error of the type $C\nabla$ occurs, the least-squares estimator $\underline{\hat{x}}$ will become biased. Thus $E\{\underline{\hat{x}}\} \neq x$. The dispersion or precision of the estimator however, remains unaffected by this model error. The bias in $\underline{\hat{x}}$, due to a model error $C\nabla$, is given as

$$\nabla\hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C\nabla$$

The purpose of testing the model, is to minimize the risk of having a biased least-squares solution. However, one should realize that the outcomes of the statistical tests are not exact and thus also prone to errors. It depends on the 'strenght' of the model, how much confidence one will have in the outcomes of these statistical tests. A measure of this confidence is provided for by the concept of reliability. When the above w-test statistic is used, the size of the model error that can be found with a probability γ, is given by the Minimal Detectable Bias (MDB). It reads

$$|\nabla| = \sqrt{\frac{\lambda(\alpha_1, 1, \gamma)}{c^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c}}$$

where $\lambda(\alpha_1, 1, \gamma)$ is a known function of the level of significance α_1 and the detection probability (power) γ. The set of MDB's, one for each model error considered, is said to describe the internal reliability of the model.

As it was the case with precision, the internal reliability can be computed once the design matrix A and the variance matrix Q_y are available. Changing A and/or changing Q_y, will change the MDB's. In this way one can thus change (e.g. improve) the internal reliability. Substitution of $C\,|\nabla|$ for $C\nabla$ in the above expression for $\nabla\hat{x}$, will show by how much the least-squares solution becomes biased, when a model error of the size of the MDB occurs. The bias vectors $\nabla\hat{x}$, one for each model error considered, is then said to describe the external reliability of the model.

NETWORKS: The theory of adjustment and testing, is in principle applicable to any geodetic project where measurements are involved for the determination of unknown parameters. But in case of a project like computing a geodetic network, some additional considerations need to be taken into account as well. The aim of computing a geodetic network is to determine the geometry of the configuration of a set of points. The set of points usually consists of: (1) newly established points, of which the coordinates still need to be determined, and (2) already existing points, the so-called control points, of which the coordinates are known. By means of a network adjustment the relative geometry of the new points is determined and integrated into the geometry of the existing control

points. The determination and validation of the overall geometry is usually divided in two phases: (1) the free network phase, and (2) the connected network phase.

Free network phase: In this phase, the known coordinates of the control points do not take part in the determination of the geometry. It is thus free from the influence of the existing control points. The idea is that a good geodetic network should be sufficiently precise and reliable in itself, without the need of external control. It implies, when in the second phase, the connected network phase, rejection of the model occurs, that one has good reason to belief that the cause for rejection should be sought in the set of control points, instead of in the geometry of the free network.

As with any geodetic project, the three steps involved in the free network phase are: design (precision and reliability), adjustment (determination of geometry) and testing (validation of geometry). With free networks however, there is one additional aspect that should be considered carefully. It is the fundamental non-uniqueness in the relation between geodetic observables and coordinates. This implies, that when computing coordinates for the free network, additional information in the form of minimal constraints are needed, to eliminate the non-uniqueness between observables and coordinates. The minimal constraints however, are not unique. There is a whole set from which they can be chosen. This implies that the set of adjusted coordinates of the free network, including their variance matrix and external reliability, are not unique as well. This on its turn implies, that one should only use procedures for evaluating the precision and reliability, that are guaranteed to be invariant for the choise of minimal constraints. If this precaution is not taken, one will end up using an evaluation procedure of which the outcome is dependent on the arbitrary choice of minimal constraints.

Connected network phase: The purpose of this second phase is to integrate the geometry of the free network into the geometry of the control points. The observables are the coordinates of the free network and the coordinates of the control points. Since the coordinates of the two sets are often given in different coordinate systems, the connection model will often be based on a coordinate transformation from the coordinate system of the free network to that of the control network.

In contrast to the free network phase, the design, adjustment and testing are now somewhat nonstandard. First of all there is not much left to design. Once the free network phase has been passed, the geometry of the free network as well as that of the control points are given. This implies that already at the design stage of the free network, one should take into account the distribution of the free network points with respect to the distribution of the control points.

Secondly, the adjustment in the connected network phase is not an ordinary least-squares adjustment. In most applications, it is not very practical to see the coordinates of the control points change everytime a free network is connected to them. This would happen however, when an ordinary adjustment would be carried out. Thus instead, a constrained adjustment is applied, with the explicit constraints that the coordinates of the control points remain fixed.

For testing however, a constrained adjustment would not be realistic. After all, the

coordinates of the control points are still samples from random variables and therefore not exact. Thus for the validation of the connected geometry, the testing is based on the least-squares residuals that follow from an ordinary adjustment and not from a constrained adjustment.

Chapter 2

Estimation and precision

2.1 Introduction

This chapter is an introduction to least-squares estimation theory and it is written at a level which should be sufficient for understanding the chapters following. The terms 'estimation' and 'adjustment' are often used interchangingly. Adjustment is the process of making an inconsistent system of equations consistent, whereas estimation refers to the process of determining values for the unknown parameters that occur in the system of equations.

Loosely speaking, the need for an adjustment arises when there are more observations than unknowns in a system of equations. The system of equations is then usually inconsistent. That is, without an adjustment, the given observations will not be compatible with the system of equations, they will not fit these equations. Adjustment is thus the process of making an inconsistent system of equations, consistent.

In section 2, we start off by discussing the concepts of consistency and uniqueness in more detail. Inconsistency of a system of equations can only occur when the number of equations, m, exceeds the rank of the system's matrix A, rank A. The difference $m - \text{rank}\, A$ is known as the *redundancy* of the system. The system is always consistent when the redundancy equals zero. In that case, there is no need for an adjustment and the system will always have a solution. The solution however, may not be unique. The solution is only unique when the system's matrix is of full rank, that is, when rank A equals the total number of unknowns in the system.

The system of equations will generally be inconsistent when the redundancy is unequal to zero. In section 3 we show how an inconsistent system of equations can be made consistent. The method employed is that of least- squares. Both the deterministic as well as stochastic formulation of the least-squares method are discussed. For the stochastic formulation we rely on the concept of random variables. The first two (central) moments of a random variable, the mean and the variance, together with the very important propagation laws, are discussed in the appendix.

In section 4, we generalize the least-squares method and show how it can be used when the observation equations are nonlinear. In applications, one will find that the observation equations are more often than not, nonlinear. Using Taylor's theorem, it is shown how the nonlinear observation equations can be linearized and how, by means of an iteration process, the standard least-squares approach can be employed.

In section 5, we introduce the concept of condition equations. The formulation in terms of condition equations is the dual to the formulation in terms of observation equations. It is much like the equation of a circle, $x^2 + y^2 = R$, is the dual to the two parametric equations $x = R\cos\phi$ and $y = R\sin\phi$. It is shown how the least- squares formulae need to be formulated, when one is dealing with condition equations. Both linear as well as

nonlinear condition equations are considered.

In section 6, we discuss three least-squares procedures which are of use for the material discussed in the later parts of these lecture notes. First we discuss recursive least-squares estimation, then constrained least-squares estimation and finally, minimally constrained least-squares estimation. Minimally constrained least-squares estimation is needed when the systems's matrix is less than of full rank. In most geodetic applications this is more the rule than the exception. When discussing minimally constrained least-squares, particular attention is given to the dependency of the variance matrix on the chosen set of minimal constraints.

The chapter is concluded with a section on quality control. In this section, only one aspect of quality is discussed, namely that of precision. Precision as expressed by the variance matrix, can be said to describe the variability of sample values of an unbiased estimator around its mean. In this last section, we also present some ways of testing the precision against a given criterium.

2.2 Consistency and uniqueness

Assume that we want to determine n unknown parameters $x_\alpha \in R$, $\alpha = 1,\ldots,n$, from a given set of m measurements $y_i \in R$, $i = 1,\ldots,m$. If the measurements bear a known *linear* relationship with the unknown parameters, we may write

$$y_i = \sum_{\alpha=1}^{n} a_{i\alpha} x_\alpha \ , \ i = 1,\ldots m \tag{2.1}$$

In this equation the known scalars $a_{i\alpha}$ determine how the measurements are related to the unknown parameters. By introducing the matrix A and the vectors y and x as

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} , \ y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} , \ x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

equation (2.1) can be written in matrix-vector form as

$$\underset{m\times 1}{y} \quad = \quad \underset{m\times n\, n\times 1}{Ax} \tag{2.2}$$

This is a system of m linear equations in n unkown parameters. In order to be able to solve this linear system, it is first of interest to know under what conditions a solution to the linear system *exists* and if so, whether it is *unique* or not.

It will be clear that a solution to the linear system (2.2) exists if and only if the m-vector y can be written as a linear combination of the column vectors of matrix A. If this is the case, the vector y is said to be an element of the column space or *range space* of matrix A. The range space of A is denoted as $R(A)$. Thus a solution to (2.2) exists if and only if

$$y \in R(A) \tag{2.3}$$

Systems of equations for which this holds true, are called *consistent*. A system is said to be *inconsistent* if it is not consistent. In this case the vector y can not be written as a linear combination of the column vectors of matrix A and hence no vector x exists such that (2.2) holds true.

Since $y \in R^m$, it follows from (2.3) that consistency is guaranteed if $R(A) = R^m$, that is, if the range space of matrix A equals the m-dimensional space of real numbers. But $R(A) = R^m$ only holds true, if the dimension of $R(A)$ equals the dimension of R^m, which is m. It follows therefore, since the dimension of $R(A)$ equals the *rank* of matrix A (*note*: the rank of a matrix equals the total number of linear independent columns of this matrix), that consistency is guaranteed if and only if

$$\text{rank } A = m \tag{2.4}$$

In all other cases, rank $A < m$, the linear system may or may not be consistent.

Let us now assume that the system is indeed consistent. The next question one may ask is whether the solution to (2.2) is *unique* or not. That is, whether the information content of the measurements collected in the vector y is sufficient to determine the parameter vector x uniquely. The solution is only unique if all the column vectors of matrix A are linear independent. Hence, the solution is unique if the rank of matrix A equals the number of unknown parameters,

$$\text{rank } A = n \tag{2.5}$$

To see this, assume x and $x \neq x'$ to be two different solutions of (2.2). Then $Ax = Ax'$ or $A(x - x') = 0$. But this can only be the case if some of the columns of matrix A are lineary dependent, which contradicts the assumption (2.5) of full rank. Thus the solution is unique when rank $A = n$ and it is nonunique when rank $A < n$.

Unless otherwise stated, we will assume from now on that the matrix A of the linear system (2.2) is of full rank.

With rank $A = n$ and the fact that the rank of a matrix is always equal or less than the number of its rows or its columns, follows that we can discriminate between the following two cases

$$m = n = \text{rank } A \quad \text{or} \quad m > n = \text{rank } A \tag{2.6}$$

In the first case, both (2.4) and (2.5) are satisfied. This implies that the system is both consistent and unique. Thus a solution exists and this solution is also unique. The unique solution, denoted by $\hat{x}$, is given as

$$\hat{x} = A^{-1} y \tag{2.7}$$

where A^{-1} denotes the *inverse* of matrix A.

Example 1

Consider the linear system

$$\underbrace{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2.8}$$

In this case we have: $m = 2$, $n = 2$ and rank $A = 2$. Thus, the system is consistent since rank $A = m = 2$, and the system has a unique solution since rank $A = n = 2$. The unique solution of (2.8) reads: $\hat{x} = (5/7\,,\, 3/7)^T$. □

In the second case, only (2.5) is satisfied. This implies that a unique solution exists, provided that the system is consistent. If the system is consistent, the unique solution can be obtained by inverting n out of the $m > n$ linear equations. Hence, we first partition (2.2) as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} x$$

where y_1 is an n-vector, y_2 is an $m-n$-vector and A_1 and A_2 are of order $n \times n$ and $(m-n) \times n$ respectively. The unique solution $\hat{x}$ follows then as

$$\hat{x} = A_1^{-1} y$$

Note that y_2 is not used in computing $\hat{x}$. This is due to the fact that in the present situation, y_2 is consistent with y_1 and hence, it does not contain any additional information.

Example 2

Consider the linear system

$$\underbrace{\begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 1 & 2 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2.9}$$

In this case we have: $m = 3$, $n = 2$ and rank $A = 2$. Since $m = 3 >$rank $A = 2$, consistency of the system is not automatically guaranteed. A closer look at the measurementvector y of (2.9) shows however that

$$\begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix} = 1. \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 1. \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$$

This shows that y can be written as a linear combination of te columnvectors of A. Therefore $y \in R(A)$, showing that the system is consistent. And since $n =$rank $A = 2$, its solution is also unique.

If we partition (2.9) as

$$\begin{bmatrix} -2 \\ 3 \\ \cdots \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & -1 \\ \cdots\cdots & \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

the unique solution follows as

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = -\frac{1}{7} \begin{bmatrix} -1 & -3 \\ -2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

The system (2.9) may also be partitioned as

$$\begin{bmatrix} -2 \\ \cdots \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ \cdots\cdots \\ 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The unique solution follows then as

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = -\frac{1}{5} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

□

As we have seen, a full rank system of m equations in n unknowns (rank $A = n$), can only be inconsistent when $m - n > 0$. For a less than full rank system (rank $A < n$), this only holds true when $m - \text{rank } A > 0$. This number, which is a very important one in adjustment theory, is referred to as the *redundancy* of the system of observation equations. Thus

$$\text{redundancy } = m - \text{rank } A \tag{2.10}$$

The question that remains is: *What to do when the linear system of equations is inconsistent?* In that case, we first need to make the system consistent before a solution can be computed. There are however many ways in which an inconsistent system can be made consistent. In the next section we will give a way, which is intuitively appealing. And lateron we will show that this approach of finding a solution to an inconsistent system also has some optimality properties, in particular in a probabilistic context.

2.3 The Linear A-model: observation equations

2.3.1 Least-squares estimates

An inconsistent system, that is, a system for which $y \notin R(A)$ holds, can be made consistent by introducing an errorvector e as (see figure 2.1):

$$\underset{m\times 1}{y} = \underset{m\times n\, n\times 1}{Ax} + \underset{m\times 1}{e}, \quad m > n = \text{rank } A \tag{2.11}$$

In (2.11), y and A are given, whereas x and e are unknown. From the geometry of figure 2.1 it seems intuitively appealing to estimate x as $\hat{x}$ such that $A\hat{x}$ is as close as possible to the given measurement- or observationvector y. In other words, the idea is to

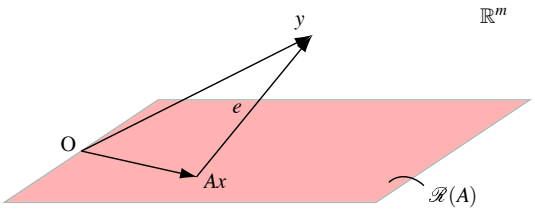


Figure 2.1 The geometry of $y = Ax + e$.

find that value of x that minimizes the length of the vector $e = y - Ax$. This idea leads to the following minimization problem:

$$\min_x (y - Ax)^T (y - Ax) \tag{2.12}$$

From calculus we know that $\hat{x}$ is a solution of (2.12) if $\hat{x}$ satisfies:

$$\frac{\partial F}{\partial x}(\hat{x}) = 0 \quad \text{and} \quad \frac{\partial^2 F}{\partial x^2}(\hat{x}) \ \text{ positive} - \text{definite} \tag{2.13}$$

where $F(x)$ is given as

$$F(x) = (y - Ax)^T (y - Ax) = y^T y - 2y^T Ax + x^T A^T Ax \tag{2.14}$$

Taking the first-order and second-order partial derivatives of $F(x)$ gives

$$\frac{\partial F}{\partial x}(x) = -2A^T y + 2A^T Ax \quad \text{and} \quad \frac{\partial^2 F}{\partial x^2}(x) = 2A^T A \tag{2.15}$$

Equating the first equation of (2.15) to zero shows that $\hat{x}$ satisfies the *normal equations*

$$A^T A\hat{x} = A^T y \tag{2.16}$$

Since rank $A^T A$ =rank $A = n$, the system is consistent and has a unique solution. Through an inversion of the *normalmatrix* $A^T A$ the unique solution of (2.16) is found as

$$\hat{x} = (A^T A)^{-1} A^T y \tag{2.17}$$

That this solution $\hat{x}$ is indeed the minimizer of (2.14) follows from the fact that the matrix $\partial^2 F / \partial x^2$ of (2.15) is indeed positive-definite. The vector $\hat{x}$ is known as the *least-squares estimate* of x, since it produces the *least* possible value of the sum-of-*squares* function $F(x)$.

From the normal equations (2.16) follows that $A^T (y - A\hat{x}) = 0$. This shows that the vector $\hat{e} = y - A\hat{x}$, which is the least-squares estimate of e, is orthogonal to the range space of matrix A (see figure 2.2):

$$A^T \hat{e} = 0 \quad , \quad \text{with } \hat{e} = y - A\hat{x} \tag{2.18}$$

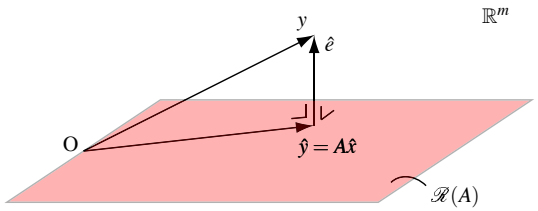


Figure 2.2 The geometry of Least-Squares.

Example 3

Let us consider the next problem

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

With $A = (1,1)^T$ the normal equation $A^T A\hat{x} = A^T y$ reads

$$2\hat{x} = y_1 + y_2$$

Hence the least-squares estimate $\hat{x} = (A^T A)^{-1} A^T y$ is

$$\hat{x} = \frac{1}{2}(y_1 + y_2)$$

Thus, to estimate x, one adds the measurements and divides by the number of measurements. Hence, the least-squares estimate equals in this case the arithmetic average.

The least-squares estimates of the obserations and observation errors follow from $\hat{y} = A\hat{x}$ and $\hat{e} = y - \hat{y}$ as

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} y_1 + y_2 \\ y_1 + y_2 \end{bmatrix}, \text{ and } \begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} y_1 - y_2 \\ y_2 - y_1 \end{bmatrix}$$

Finally the square of the length of $\hat{e}$ follows as

$$\hat{e}^T \hat{e} = \frac{1}{2}(y_1 - y_2)^2$$

□

So far we have discussed the *unweighted* least-squares principle. The least-squares principle can be generalized however by introducing a positive-definite $m \times m$ *weight matrix* W. This is done by replacing (2.12) by the following minimization problem

$$\min_x (y - Ax)^T W (y - Ax) \tag{2.19}$$

Table 2.1 Weighted Least-Squares.

Consistent Linear System
$y = Ax + e,\ y, e \in R^m,\ x \in R^n,\ m > n = \text{rank } A$
Weighted Least-Squares Principle
$\min_x (y - Ax)^T W (y - Ax),\ W = \text{positive definite}$
Weighted Least-Squares Estimates
parametervector : $\hat{x} = (A^T W A)^{-1} A^T W y$
observationvector : $\hat{y} = A\hat{x}$
errorvector : $\hat{e} = y - \hat{y}$

The solution of (2.19) can be derived along lines which are similar as the ones used for solving (2.12). The solution of (2.19) reads

$$\hat{x} = (A^T W A)^{-1} A^T W y \tag{2.20}$$

This is the *weighted least-squares estimate* of x. In case of weighted least-squares the normal equations read: $A^T W A\hat{x} = A^T W y$. This shows that the vector $\hat{e} = y - A\hat{x}$, which is the weighted least-squares estimate of e, satisfies

$$A^T W \hat{e} = 0 \quad , \quad \text{with } \hat{e} = y - A\hat{x} \tag{2.21}$$

If the *inner product* of the observation space R^m is defined as $(a, b) = a^T W b,\ \forall a, b \in R^m$, (2.21) can also be written as $(Ax, \hat{e}) = 0,\ \forall x \in R^m$. This shows that also in the case of weighted least-squares, the vector $\hat{e}$ can be considered to be orthogonal to the rangspace of A.

A summary of the least-squares algorithm is given in Table 2.1.

Example 4

Consider again the problem

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

As weight matrix we take

$$W = \begin{bmatrix} w_{11} & 0 \\ 0 & w_{22} \end{bmatrix}$$

With $A = (1, 1)^T$ the normal equation $A^T W A\hat{x} = A^T W y$ reads

$$(w_{11} + w_{22})\hat{x} = w_{11} y_1 + w_{22} y_2$$

Hence the weighted least-squares estimate $\hat{x} = (A^T W A)^{-1} A^T W y$ is

$$\hat{x} = \frac{w_{11} y_1 + w_{22} y_2}{w_{11} + w_{22}}$$

Thus instead of the arithmetic average of y_1 and y_2, as we had with $W = I$, the above estimate is a *weighted* average of the data. This average is closer to y_1 than to y_2 if $w_{11} > w_{22}$.

The weighted least-squares estimate of the observations, $\hat{y} = A\hat{x}$, is

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} = \frac{1}{w_{11} + w_{22}} \begin{bmatrix} w_{11} y_1 + w_{22} y_2 \\ w_{11} y_1 + w_{22} y_2 \end{bmatrix}$$

and the weighted least-squares estimate of $\hat{e}$, $\hat{e} = y - \hat{y}$ is

$$\begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \end{bmatrix} = \frac{1}{w_{11} + w_{22}} \begin{bmatrix} w_{22}(y_1 - y_2) \\ w_{11}(y_2 - y_1) \end{bmatrix}$$

Note that $|\hat{e}_2| > |\hat{e}_1|$ if $w_{11} > w_{22}$.

Finally the weighted sum of squares $\hat{e}^T W \hat{e}$ is

$$\hat{e}^T W \hat{e} = \frac{w_{11} w_{22}}{w_{11} + w_{22}} (y_1 - y_2)^2$$

□

2.3.2 A stochastic model for the observations

In the previous section the principle of least-squares was introduced. The least-squares principle enables us in case of inconsistent systems, to obtain an intuitively appealing estimate $\hat{x}$ of the parametervector x. But although the least-squares estimate $\hat{x}$ is intuitively appealing, no *quality measures* as yet can be attached to the estimate. That is, we know how to compute the estimate $\hat{x}$, but we are not able yet to say how good the estimate really is. Of course, the numerical value of the least value of the sum of squares, $\hat{e}^T W \hat{e}$, does indicate something about the quality of $\hat{x}$. If $\hat{e}^T W \hat{e}$ is small one is inclined to have more confidence in the estimate $\hat{x}$, than when $\hat{e}^T W \hat{e}$ is large. But how small is small? Besides, $\hat{e}^T W \hat{e}$ is identically zero if the linear system is consistent. Would this then automatically imply that the estimate $\hat{x}$ has good quality? Not really, since the observations may still be subject to measurement errors.

In order to obtain quality measures for the results of least-squares estimation, we start by introducing a qualitative description of the input, that is of the observations. This description will be of a *probabilistic* nature. The introduction of a probabilistic description is motivated by the experimental fact that the variability in the outcome of measurements, when repeated under similar circumstances, can be described to a sufficient degree by *stochastic or random variables*. We will therefore assume that the observation vector y, which contains the numerical values of the measurements, constitutes a *sample* of the

random vector of observables $\underline{y}$ (Note: the underscore indicates that we are dealing with a random variable). It is furthermore assumed that the vector of observables $\underline{y}$ can be written as the sum of a deterministic functional part Ax and a random residual part $\underline{e}$:

$$\underline{y} = Ax + \underline{e} \tag{2.22}$$

Although a random vector is completely described by its probability density function, we will restrict ourselves to the first two moments of random variables. That is, we will restrict ourselves to the *mean* and to the *variance matrix*. Properties of the mean and of the variance matrix together with the *error propagation laws*, are given in the appendix.

If we assume that $\underline{e}$ models the probabilistic nature of the variability in the measurements, it seems acceptable to assume that this variability is zero "on the average" and therefore that the mean of $\hat{e}$ is zero:

$$E\{\underline{e}\} = 0 \tag{2.23}$$

where $E\{.\}$ stands for the mathematical expectation operator. The measurement variability itself is modelled through the dispersion- or variance matrix of $\underline{e}$. We will assume that this matrix is known and denote it by Q_y:

$$D\{\underline{e}\} = Q_y \tag{2.24}$$

where $D\{.\}$ stands for the dispersion operator. It is defined in terms of $E\{.\}$ as

$$D\{.\} = E\{(. - E\{.\})(. - E\{.\})^T\}$$

With (2.23) and (2.24) we are now in the position to determine the mean and variance matrix of the vector of observables $\underline{y}$. Application of the law of propagation of means and the law of propagation of variances to (2.22) gives with (2.23) and (2.24):

$$E\{\underline{y}\} = Ax \quad ; \quad D\{\underline{y}\} = Q_y \tag{2.25}$$

This will be our model for the vector of observables $\underline{y}$. As the results of the next section show, model (2.25) enables us to describe the quality of the results of least-squares estimation in terms of the mean and the variance matrix.

2.3.3 Least-squares estimators

Functions of random variables are again random variables. It follows therefore, that if the vector of observables is assumed to be a random vector $\underline{y}$ and when it is substituted for y in the formulae of table 1 in section 2.3.1, the results are again random variables:

$$\begin{cases} \underline{\hat{x}} &= (A^T W A)^{-1} A^T W \underline{y} \\ \underline{\hat{y}} &= A\underline{\hat{x}} \\ \underline{\hat{e}} &= \underline{y} - \underline{\hat{y}} \end{cases} \tag{2.26}$$

These random vectors will be called least-squares *estimators*. And if $\underline{y}$ is replaced by its sample or measurement value y, we speak of least-squares *estimates*. The *quality* of the above estimators can now be deduced from the first two moments of $\underline{y}$.

The first moment; the mean: Together with $E\{\underline{y}\} = Ax$, an application of the propagation law of means to (2.26) gives

$$\begin{cases} E\{\underline{\hat{x}}\} &= x \\ E\{\underline{\hat{y}}\} &= E\{\underline{y}\} \\ E\{\underline{\hat{e}}\} &= E\{\underline{e}\} = 0 \end{cases} \tag{2.27}$$

These results show, that under the assumption that (2.25) holds, the least-squares estimators are *unbiased* estimators. Note that this property of unbiasedness is independent of the choice for the weightmatrix W.

The second moment; the variance matrix: Together with $D\{\underline{y}\} = Q_y$, an application of the propagation law of variances and covariances to (2.26) gives

$$\begin{cases} Q_{\hat{x}} &= (A^T WA)^{-1} A^T W Q_y WA (A^T WA)^{-1} \\ Q_{\hat{y}} &= AQ_{\hat{x}} A^T \\ Q_{\hat{e}} &= [I - A(A^T WA)^{-1} A^T W] Q_y [I - A(A^T WA)^{-1} A^T W]^T \end{cases} \tag{2.28}$$

and

$$\begin{cases} Q_{\hat{x}\hat{y}} &= Q_{\hat{x}} A^T \\ Q_{\hat{x}\hat{e}} &= (A^T WA)^{-1} A^T W Q_y - Q_{\hat{x}} A^T \\ Q_{\hat{y}\hat{e}} &= AQ_{\hat{x}\hat{e}} \end{cases} \tag{2.29}$$

The above variance matrices enable us now to give a complete precision description of any arbitrary linear function of the estimators. Consider for instance the linear function $\underline{\hat{\theta}} = a^T \underline{\hat{x}}$. Application of the propagation law of variances gives then for the precision of $\underline{\hat{\theta}}$: $\sigma_{\hat{\theta}}^2 = a^T Q_{\hat{x}} a$.

The above results enable us to describe the quality of the results of least-squares estimation in terms of the mean and the variance matrix. The introduction of a stochastic model for the vector of observables $\underline{y}$ enables us however also to judge the merits of the least-squares principle itself. Recall that the least-squares principle was introduced on the basis of intuition and not on the basis of probabilistic reasoning. With the mathematical model (2.24) one could now however try to develop an estimation procedure that produces estimators with certain well defined probabilistic *optimality* properties. One such procedure is based on the principle of "Best Linear Unbiased Estimation (BLUE)".

Assume that we are interested in estimating a parameter θ which is a linear function of x:

$$\underset{1\times 1}{\theta} = \underset{1\times n}{a^T} \underset{n\times 1}{x} \tag{2.30}$$

The estimator of θ will be denoted as $\underline{\hat{\theta}}$. Then according to the BLUE's criteria, the estimator $\underline{\hat{\theta}}$ of θ has to be a *linear* function of $\underline{y}$,

$$\underset{1\times 1}{\underline{\hat{\theta}}} = \underset{1\times m}{l^T} \underset{m\times 1}{\underline{y}} \tag{2.31}$$

such that it is *unbiased*,

$$E\{\underline{\hat{\theta}}\} = \theta \tag{2.32}$$

and such that it is best in the sense of *minimum variance*,

$$\sigma_{\hat{\theta}}^2 \rightarrow \text{minimum} \tag{2.33}$$

The objective is thus to find a vector $l \in R^m$ such that with (2.31), the conditions (2.32) and (2.33) are satisfied. It can be shown that the solution to the above problem is given by

$$l^T = a^T (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$$

If we substitute this into (2.31) we get

$$\underline{\hat{\theta}} = a^T (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} \tag{2.34}$$

This is the best linear unbiased estimator of θ. The important result (2.34) shows that the best linear unbiased estimator of x is given by

$$\underline{\hat{x}} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} \tag{2.35}$$

A comparison between (2.26) and (2.35) shows that the BLUE of x is identical to the weighted least-squares estimator of x if the weight matrix W is taken to be equal to the inverse of the variance matrix of $\underline{y}$:

$$W = Q_y^{-1} \tag{2.36}$$

This is an important result, because it shows that the weighted least-squares estimators are best in the probabilistic sence of having minimal variance if (2.36) holds. The variances and covariances of these estimators follow if the weightmatrix W is replaced in (2.28) and (2.29) by Q_y^{-1}.

From now on we will always assume, unless stated otherwise that the weightmatrix W is chosen to be equal to Q_y^{-1}. Consequently no distinction will be made any more in these lecture notes between weighted least-squares estimators and best linear unbiased estimators. Instead we will simply speak of least-squares estimators.

2.3.4 Summary

In Table 2.2 an overview is given of the main results of Least-Squares Estimation.

The table shows the linear model of observation equations, the linear A-model. Based on A, Q_y and $\underline{y}$, the least-squares estimator of the unknown parameter vector x can be determined. And from it, one can determine the adjusted vector of observables, $\underline{\hat{y}}$, the least-squares residual vector, $\underline{\hat{e}}$, and the least-squares estimator of an arbitrary function $\theta = a^T x$, as $\underline{\hat{\theta}} = a^T \underline{\hat{x}}$. As shown in the table, each of these random vectors have their own mean and variance matrix. Some of the random vectors are correlated, such as $\underline{\hat{x}}$ and $\underline{\hat{y}}$, and some of them are not, such as $\underline{\hat{x}}$ and $\underline{\hat{e}}$.

2.4 The Nonlinear A-model: observation equations

2.4.1 Nonlinear observation equations

Up to this point the development of the theory was based on the assumption that the m-vector $E\{\underline{y}\}$ is *linearly* related to the n-vector of unknown parameters x. In geodetic

Table 2.2 Least-Squares Estimation.

THE LINEAR A-MODEL

$$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n\, n\times 1}{Ax} \quad , \quad \underset{m\times m}{D\{\underline{y}\}} = \underset{m\times m}{Q_y} \quad , \quad m \geq n = \text{rank } A$$

LEAST-SQUARES ESTIMATORS

$$\begin{array}{lllllll} \underline{\hat{x}} & = & (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} & & \underline{\hat{e}} & = & \underline{y} - \underline{\hat{y}} \\ \underline{\hat{y}} & = & A\underline{\hat{x}} & & \underline{\hat{\theta}} & = & a^T \underline{\hat{x}} \end{array}$$

MEAN

$$\begin{array}{llll} E\{\underline{\hat{x}}\} = x & & E\{\underline{\hat{e}}\} = E\{\underline{e}\} = 0 \\ E\{\underline{\hat{y}}\} = E\{\underline{y}\} = Ax & & E\{\underline{\hat{\theta}}\} = \theta = a^T x \end{array}$$

VARIANCES AND COVARIANCES

$$\begin{array}{lllllll} Q_{\hat{x}} & = & (A^T Q_y^{-1} A)^{-1} & & & Q_{\hat{e}} & = & Q_y - Q_{\hat{y}} \\ Q_{\hat{y}} & = & A Q_{\hat{x}} A^T & & & \sigma^2_{\hat{\theta}} & = & a^T Q_{\hat{x}} a \\ Q_{\hat{x}\hat{y}} & = & Q_{\hat{x}} A^T \ , & Q_{\hat{x}\hat{e}} & = 0 \ , & Q_{\hat{y}\hat{e}} & = & 0 \end{array}$$

applications there are however only a few cases where this assumption truly holds. A typical example is levelling. In the majority of applications however the m-vector $E\{\underline{y}\}$ is *nonlinearly* related to the n-vector of unknown parameters x. This implies that instead of the linear A-model (2.25), we are generally dealing with a nonlinear model of observation equations:

$$E\{\underline{y}\} = A(x) \quad ; \quad D\{\underline{y}\} = Q_y \tag{2.37}$$

where $A(.)$ is a nonlinear vectorfunction from R^n into R^m. The following two simple examples should make this clear.

Example 5

Consider the configuration of figure 2.3. The x,y coordinates of the three

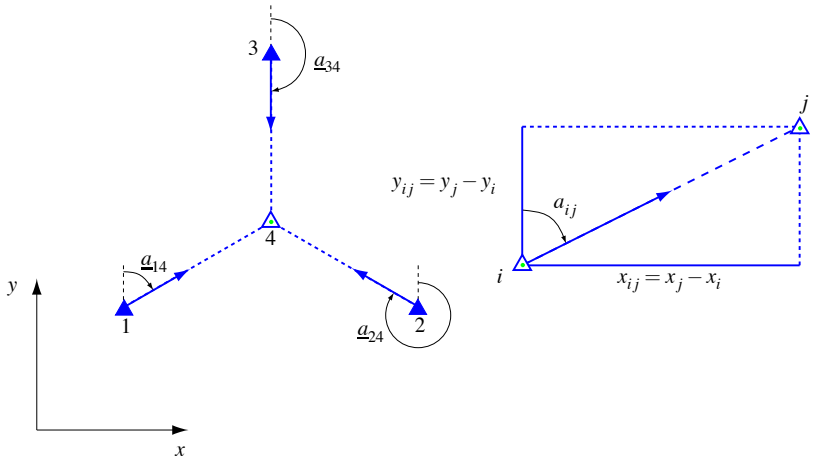


Figure 2.3 Azimuth resection.

points 1, 2 and 3 are known and the coordinates x_4 and y_4 of point 4 are unknown. The observables consist of the three azimuth variates $\underline{a}_{14}$, $\underline{a}_{24}$, and $\underline{a}_{34}$. Since azimuth and coordinates are related as

$$\tan a_{ij} = \frac{x_j - x_i}{y_j - y_i} = \frac{x_{ij}}{y_{ij}}$$

The model of observation equations for the configuration of figure 2.3 reads

$$E\{\begin{bmatrix} \underline{a}_{14} \\ \underline{a}_{24} \\ \underline{a}_{34} \end{bmatrix}\} = \begin{bmatrix} \arctan[x_{14}/y_{14}] \\ \arctan[x_{24}/y_{24}] \\ \arctan[x_{34}/y_{34}] \end{bmatrix}$$

This model consists of three *nonlinear* observation equations in the two unknown parameters x_4 and y_4. □

Example 6

Consider the situation of figure 2.4. It shows two cartesian coordinate systems: the x,y-system and the u,v-system. The two systems only differ in their orientation. This means that if the coordinates of a point i are given in the u,v-system, (u_i, v_i), a *rotation* through an angle α is needed to obtain the coordinates of the same point i in the x,y-system, (x_i, y_i):

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} u_i \\ v_i \end{bmatrix} \tag{2.38}$$

Let us now assume that we have at our disposal the coordinate observables of two points in both coordinate systems: $(\underline{x}_i, \underline{y}_i)$ and $(\underline{u}_i, \underline{v}_i)$, $i = 1, 2$. Using (2.38), our model reads then

$$E\{\begin{bmatrix} \underline{x}_i \\ \underline{y}_i \end{bmatrix}\} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} E\{\begin{bmatrix} \underline{u}_i \\ \underline{v}_i \end{bmatrix}\} \quad , \quad i = 1, 2 \tag{2.39}$$

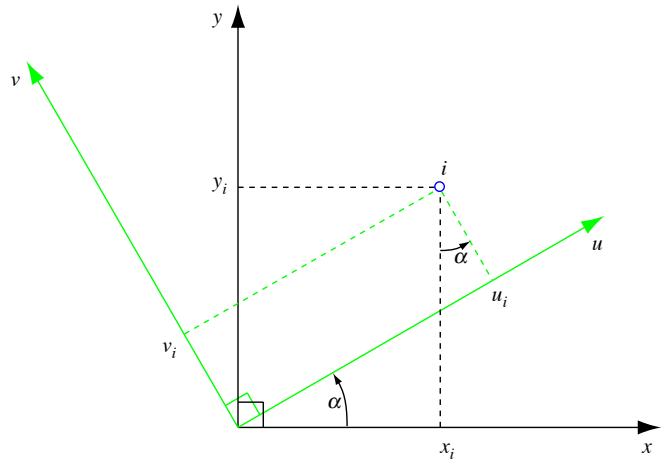


Figure 2.4 A coordinate transformation.

This model is however still not in the form of observation equations. If we consider the orientation angle α and the coordinates of the two points in the u,v-system as the unknown parameters, (2.39) can be written in terms of observation equations as

$$E\{\begin{bmatrix} \underline{x}_1 \\ \underline{y}_1 \\ \underline{x}_2 \\ \underline{y}_2 \\ \underline{u}_1 \\ \underline{v}_1 \\ \underline{u}_2 \\ \underline{v}_2 \end{bmatrix}\} = \begin{bmatrix} u_1\cos\alpha - v_1\sin\alpha \\ u_1\sin\alpha + v_1\cos\alpha \\ u_2\cos\alpha - v_2\sin\alpha \\ u_2\sin\alpha + v_2\cos\alpha \\ u_1 \\ v_1 \\ u_2 \\ v_2 \end{bmatrix} \tag{2.40}$$

This model consists of eight observations in five unknown parameters. Note that the first four observation equations are nonlinear. □

2.4.2 The linearized observation equations

We know how to compute least-squares estimators in case of a linear A-model. But how are we now going to compute least-squares estimators if the model of observation equations is nonlinear? For the majority of nonlinear problems the solution is to approximate the originally nonlinear A-model with a linear one. In order to show how this can be done, we first recall the celebrated theorem of Taylor. This theorem reads:

Taylor's Theorem: Let $f(x)$ be a function from R^n into R which is smooth enough. Let $x_0 \in R^n$ be an approximation to $x \in R^n$ and define $\Delta x = x - x_0$, and $\theta = x_0 + t(x - x_0)$ with

$t \in R$. Then a scalar $t \in (0,1)$ exists such that

$$\begin{aligned} f(x) = f(x_0) + \sum_{\alpha=1}^{n} \partial_\alpha f(x_0)\Delta x_\alpha + \frac{1}{2}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \partial^2_{\alpha\beta} f(x_0)\Delta x_\alpha \Delta x_\beta + \dots \\ \dots + \frac{1}{(q-1)!}\sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_{q-1}=1}^{n} \partial^{q-1}_{\alpha_1 \dots \alpha_{q-1}} f(x_0)\Delta x_{\alpha_1} \dots \Delta x_{\alpha_{q-1}} + R_q(\theta, \Delta x) \end{aligned} \tag{2.41}$$

with the remainder

$$R_q(\theta,\Delta x) = \frac{1}{q!}\sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_q=1}^{n} \partial^q_{\alpha_1 \dots \alpha_q} f(\theta)\Delta x_{\alpha_1} \dots \Delta x_{\alpha_q} \tag{2.42}$$

In (2.41) and (2.42), $\partial^q_{\alpha_1 \dots \alpha_q} f(x)$ denotes the qth-order partial derivative of $f(x)$ evaluated at x. For the case $q = 2$, it follows from (2.41) and (2.42) that

$$f(x) = f(x_0) + \sum_{\alpha=1}^{n} \partial_\alpha f(x_0)\Delta x_\alpha + \frac{1}{2}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \partial^2_{\alpha\beta} f(\theta)\Delta x_\alpha \Delta x_\beta \tag{2.43}$$

If we introduce the *gradientvector* and *Hessianmatrix* of $f(x)$ respectively as

$$\partial_x f(x) = \begin{bmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{bmatrix} \quad \text{and} \quad \partial^2_{xx} f(x) = \begin{bmatrix} \partial^2_{11} f(x) & \cdots & \partial^2_{1n} f(x) \\ \vdots & & \vdots \\ \partial^2_{n1} f(x) & \cdots & \partial^2_{nn} f(x) \end{bmatrix}$$

then equation (2.43) may be written in the more compact matrix-vector form as

$$f(x) = f(x_0) + \partial_x f(x_0)^T \Delta x + \frac{1}{2}\Delta x^T \partial^2_{xx} f(\theta)\Delta x \tag{2.44}$$

This important result shows that a nonlinear function $f(x)$ can be written as a sum of three terms. The first term in this sum is the *zero-order term* $f(x_0)$. The zero-order term depends on x_0 but is independent of x. The second term in the sum is the *first-order term* $\partial_x f(x_0)^T \Delta x$. It depends on x_0 and is *linearly* dependent on x. Finally, the third term in the sum is the second-order remainder $R_2(\theta,\Delta x)$.

A consequence of Taylor's Theorem is that the remainder $R_2(\theta,\Delta x)$ can be made arbitrarily small by choosing the approximation x_0 close enough to x. Now assume that the x_0 approximation is chosen such that the second-order remainder can indeed be neglected. Then, instead of (2.44) we may write to a sufficient approximation:

$$f(x) = f(x_0) + \partial_x f(x_0)^T \Delta x \tag{2.45}$$

Hence, if x_0 is sufficiently close to x, the *nonlinear* function $f(x)$ can be approximated to a sufficient degree by the function $f(x_0) + \partial_x f(x_0)^T \Delta x$ which is *linear* in x. This function is the linearized version of $f(x)$. A geometric interpretation of this linearization is given in figure 2.5 for the case $n = 1$. Let us now apply the above linearization to our nonlinear observation equations

$$E\{\underline{y}\} = A(x) = \begin{bmatrix} a_1(x) \\ \vdots \\ a_m(x) \end{bmatrix} \tag{2.46}$$

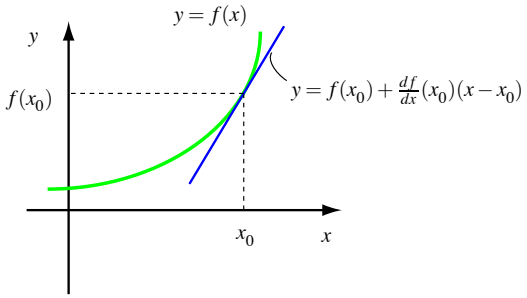


Figure 2.5 The nonlinear curve $y = f(x)$ and its linear tangent $y = f(x_0) + \frac{d}{dx}f(x_0)(x - x_0)$.

Each nonlinear observation equation $a_i(x)$, can now be linearized according to (2.45). This gives

$$\underset{m\times 1}{\begin{bmatrix} a_1(x) \\ \vdots \\ a_m(x) \end{bmatrix}} = \underset{m\times 1}{\begin{bmatrix} a_1(x_0) \\ \vdots \\ a_m(x_0) \end{bmatrix}} + \underset{m\times n}{\begin{bmatrix} \partial_x a_1(x_0)^T \\ \vdots \\ \partial_x a_m(x_0)^T \end{bmatrix}} \underset{n\times 1}{\Delta x} \tag{2.47}$$

If we denote the $m \times n$ matrix of (2.47) as $\partial_x A(x_0)$, and substitute (2.47) into (2.46) we get

$$E\{\underline{y}\} = A(x_0) + \partial_x A(x_0)\Delta x \tag{2.48}$$

If we bring the constant m-vector $A(x_0)$ to the lefthandside of the equation and define $\Delta\underline{y} = \underline{y} - A(x_0)$, we finally obtain our linearized model of observation equations

$$E\{\Delta\underline{y}\} = \partial_x A(x_0)\Delta x \quad ; \quad D\{\Delta\underline{y}\} = Q_y \tag{2.49}$$

This is the *linearized A-model*. Compare (2.49) with (2.37) and (2.25). Note when comparing (2.49) with (2.25) that in the linearized A-model $\Delta\underline{y}$ takes the place of $\underline{y}$, $\partial_x A(x_0)$ takes the place of A and Δx takes the place of x. Since the linearized A-model is linear, our standard formulae of least-squares can be applied again. This gives for the least-squares estimator $\underline{\hat{x}} = x_0 + \Delta\underline{\hat{x}}$ of x:

$$\underline{\hat{x}} = x_0 + [\partial_x A(x_0)^T Q_y^{-1} \partial_x A(x_0)]^{-1} \partial_x A(x_0)^T Q_y^{-1} \Delta\underline{y} \tag{2.50}$$

And application of the propagation law of variances to (2.50) gives

$$Q_{\hat{x}} = [\partial_x A(x_0)^T Q_y^{-1} \partial_x A(x_0)]^{-1} \tag{2.51}$$

It will be clear that the above results, (2.50) and (2.51), are approximate in the sense that the second-order remainder is neglected. But these approximations are good enough if the second-order remainder can be neglected to a sufficient degree. In this case also the optimality conditions of least-squares (unbiasedness, minimal variance) hold to a sufficient degree. A summary of the linearized least-squares estimators is given in Table 2.3.

Table 2.3 Linearized Least-Squares estimation.

THE NONLINEAR A-MODEL

$$E\{\underline{y}\} = A(x) \quad ; \quad D\{\underline{y}\} = Q_y \quad ; \quad A(.): R^n \rightarrow R^m$$

THE LINEARIZED A-MODEL

$$\underset{m\times 1}{E\{\Delta\underline{y}\}} = \underset{m\times n\, n\times 1}{\partial_x A(x_0)\Delta x} \quad ; \quad \underset{m\times m}{D\{\Delta\underline{y}\}} = \underset{m\times m}{Q_y} \quad , \quad m \geq n = \text{rank}\ \partial_x A(x_0)$$

LEAST-SQUARES ESTIMATORS

$$\begin{aligned} \underline{\hat{x}} &= x_0 + [\partial_x A(x_0)^T Q_y^{-1} \partial_x A(x_0)]^{-1} \partial_x A(x_0)^T Q_y^{-1} [\underline{y} - A(x_0)] \\ \underline{\hat{y}} &= A(\underline{\hat{x}}) \\ \underline{\hat{e}} &= \underline{y} - \underline{\hat{y}} \end{aligned}$$

VARIANCES

$$\begin{aligned} Q_{\hat{x}} &= [\partial_x A(x_0)^T Q_y^{-1} \partial_x A(x_0)]^{-1} \\ Q_{\hat{y}} &= \partial_x A(x_0) Q_{\hat{x}} \partial_x A(x_0)^T \\ Q_{\hat{e}} &= Q_y - Q_{\hat{y}} \end{aligned}$$

Example 7

Consider the configuration of figure 2.6. The x,y coordinates of the three points 1, 2 and 3 are known and the two coordinates x_4 and y_4 of point 4 are unknown. The observables consist of the three distance variates $\underline{l}_{14}$, $\underline{l}_{24}$, $\underline{l}_{34}$. Since distance and coordinates are related as

$$l_{ij} = (x_{ij}^2 + y_{ij}^2)^{1/2}$$

the model of observation equations for the configuration of figure 2.6 reads

$$E\{\begin{bmatrix} \underline{l}_{14} \\ \underline{l}_{24} \\ \underline{l}_{34} \end{bmatrix}\} = \begin{bmatrix} (x_{14}^2 + y_{14}^2)^{1/2} \\ (x_{24}^2 + y_{24}^2)^{1/2} \\ (x_{34}^2 + y_{34}^2)^{1/2} \end{bmatrix} \tag{2.52}$$

This model consists of three *nonlinear* observation equations in the two unknown parameters x_4 and y_4.

In order to linearize (2.52) we need approximate values for the unknown coordinates x_4 and y_4. These approximate values will be denoted as x_4^0 and

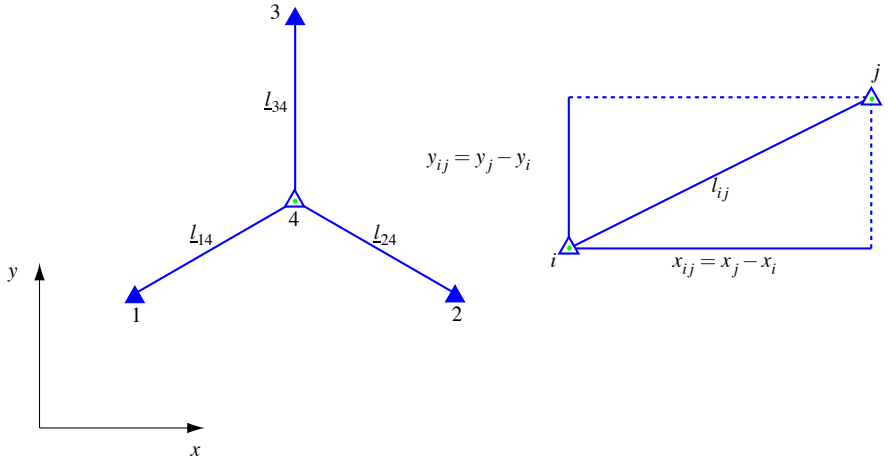


Figure 2.6 Distance resection.

y_4^0. With these approximate values a linearization of (2.52) gives

$$E\{\underbrace{\begin{bmatrix} \Delta\underline{l}_{14} \\ \Delta\underline{l}_{24} \\ \Delta\underline{l}_{34} \end{bmatrix}}_{\Delta\underline{y}}\} = \underbrace{\begin{bmatrix} (x_4^0 - x_1)/l_{14}^0 & (y_4^0 - y_1)/l_{14}^0 \\ (x_4^0 - x_2)/l_{24}^0 & (y_4^0 - y_2)/l_{24}^0 \\ (x_4^0 - x_3)/l_{34}^0 & (y_4^0 - y_3)/l_{34}^0 \end{bmatrix}}_{\partial_x A(x_0)} \underbrace{\begin{bmatrix} \Delta x_4 \\ \Delta y_4 \end{bmatrix}}_{\Delta x} \tag{2.53}$$

where

$$\begin{cases} \Delta\underline{l}_{i4} = \underline{l}_{i4} - l_{i4}^0\,, & l_{i4}^0 = [(x_4^0 - x_i)^2 + (y_4^0 - y_i)^2]^{1/2}\,, \quad i = 1,2,3 \\ \Delta x_4 = x_4 - x_4^0\,, & \Delta y_4 = y_4 - y_4^0 \end{cases}$$

Model (2.53) is the linearized version of the nonlinear A-model (2.52). □

Example 8

Consider the nonlinear *A*-model (2.40) of example 6. The unknown parameters are α and u_i and v_i for $i = 1,2$. The approximate values of these parameters will be denoted as α^0, u_i^0, v_i^0, for $i = 1,2$. Linearization of (2.40) gives then

$$E\{\underbrace{\begin{bmatrix} \Delta\underline{x}_1 \\ \Delta\underline{y}_1 \\ \Delta\underline{x}_2 \\ \Delta\underline{y}_2 \\ \Delta\underline{u}_1 \\ \Delta\underline{v}_1 \\ \Delta\underline{u}_2 \\ \Delta\underline{v}_2 \end{bmatrix}}_{\Delta\underline{y}}\} = \underbrace{\begin{bmatrix} u_1^0\sin\alpha^0 - v_1^0\cos\alpha^0 & \cos\alpha^0 & -\sin\alpha^0 & 0 & 0 \\ u_1^0\cos\alpha^0 - v_1^0\sin\alpha^0 & \sin\alpha^0 & \cos\alpha^0 & 0 & 0 \\ -u_2^0\sin\alpha^0 - v_2^0\cos\alpha^0 & 0 & 0 & \cos\alpha^0 & -\sin\alpha^0 \\ u_2^0\cos\alpha^0 - v_2^0\sin\alpha^0 & 0 & 0 & \sin\alpha^0 & \cos\alpha^0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\partial_x A(x_0)} \underbrace{\begin{bmatrix} \Delta\alpha \\ \Delta u_1 \\ \Delta v_1 \\ \Delta u_2 \\ \Delta v_2 \end{bmatrix}}_{\Delta x} \quad (2.54)$$

where:

$$\begin{cases} \Delta\underline{x}_i = \underline{x}_i - x_i^0\,, & x_i^0 = u_i^0\cos\alpha^0 - v_i^0\sin\alpha^0 \\ \Delta\underline{y}_i = \underline{y}_i - y_i^0\,, & y_i^0 = u_i^0\sin\alpha^0 + v_i^0\cos\alpha^0 \\ \Delta\underline{u}_i = \underline{u}_i - u_i^0\,, & \Delta u_i = u_i - u_i^0\,, \text{for } i = 1,2 \\ \Delta\underline{v}_i = \underline{v}_i - v_i^0\,, & \Delta v_i = v_i - v_i^0\,, \text{for } i = 1,2 \\ \Delta\alpha = \alpha - \alpha^0 \end{cases}$$

□

2.4.3 Least-Squares iteration

Up to this point it was assumed that the second-order remainder was sufficiently small and that x_0 was a good enough approximation. If this is not the case, then $\hat{x}$ as computed by (2.50) is not the least-squares estimate and hence an unacceptable error is made. In order to repair this situation, we need to improve upon the approximation x_0. It seems reasonable to expect that the estimate

$$x_1 = x_0 + [\partial_x A(x_0)^T Q_y^{-1} \partial_x A(x_0)]^{-1} \partial_x A(x_0)^T Q_y^{-1}(y - A(x_0))$$

is a better approximation than x_0. That is, it seems reasonable to expect that x_1 is closer to the true least-squares estimate than x_0. In fact one can show that this is indeed the case for most practical applications. But if x_1 is a better approximation than x_0, a further improvement can be expected if we replace x_0 by x_1 in the linearization of the nonlinear model. The recomputed linearized least-squares estimate reads then

$$x_2 = x_1 + [\partial_x A(x_1)^T Q_y^{-1} \partial_x A(x_1)]^{-1} \partial_x A(x_1)^T Q_y^{-1}(y - A(x_1))$$

By repeating this process a number of times, one can expect that finally the solution *converges* to the actual least-squares estimate $\hat{x}$. This is called the *least-squares iteration* process. The iteration is usually terminated if the difference between successive solutions is negligible. A flowdiagram of the least-squares iteration process is shown in Figure 2.7.

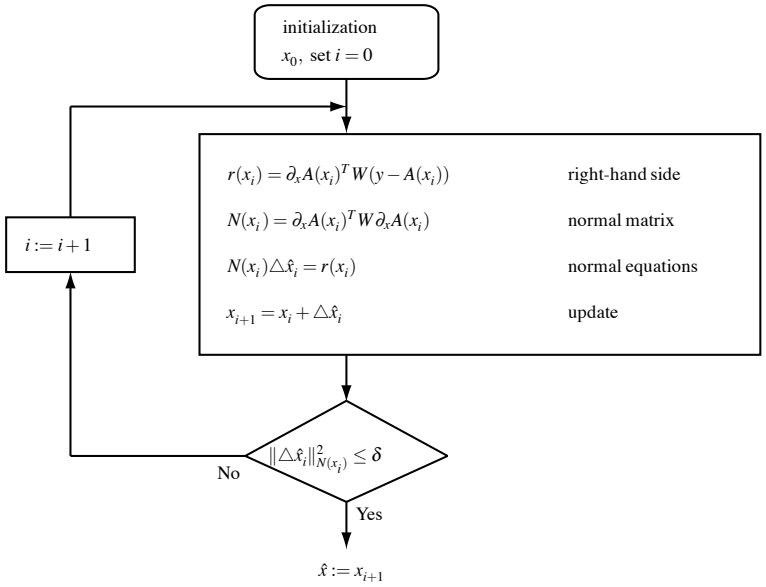


Figure 2.7 Least-squares iteration.

2.5 The *B*-Model: condition equations

2.5.1 Linear condition equations

In the previous sections we considered the model of observation equations. In this section and the next we briefly review the model of condition equations. As our starting point we take the linear A-model:

$$\underset{m\times 1}{y} = \underset{m\times n\, n\times 1}{Ax} + \underset{m\times 1}{e}, \quad m \geq n = \text{rank } A \tag{2.55}$$

This linear model is uniquely solvable if $m = n$, i.e. if the number of observables equals the number of unknown parameters. In this case A is a square matrix which is invertible because of rank $A = n$. If $m = n$ no conditions can be imposed on the observables. If $m > n =$rank A, then more observables are available then strictly needed for the determination of the n unknown parameters. In this case an $(m-n)$-number of redundant observables exist. Each seperate redundant observable gives rise to the possibility of formulating a condition equation. Thus the total number of independent condition equations that can be formulated equals

$$r = m - n \tag{2.56}$$

This number is also referred to as the *redundancy* of the model. We will now show how one can construct the condition equations, given the linear A-model (2.55). Each of the column vectors of matrix A is an element of the observationspace R^m. Together, the n-number of linearly independent column vectors of A span the range space of A. This range space has dimension n and it is a linear subspace of R^m : $R(A) \subset R^m$. Since dim$R(A) = n$

and $\dim R^m = m$, exactly $(m-n)$-number of linearly independent vectors can be found that are orthogonal to $R(A)$. Let us denote these vectors as: $b_i \in R^m$, $i = 1,\ldots,(m-n)$. Then

$$b_i \perp R(A) \quad \text{or} \quad A^T b_i = 0\ ,\ i = 1,\ldots,(m-n)$$

From this follows, if the $(m-n)$-number of linearly independent vectors b_i are collected in an $m \times (m-n)$ matrix B as

$$\underset{m\times(m-n)}{B} \quad = \quad (b_1, b_2, \ldots, b_{m-n})$$

that

$$\underset{(m-n)\times m}{B^T}\underset{m\times n}{A} \quad = \quad \underset{(m-n)\times n}{0} \quad ; \quad \operatorname{rank} B \quad = \quad m-n \tag{2.57}$$

This result may now be used to obtain the model of condition equations from (2.55). Premultiplication of the linear system of observation equations in (2.55) by B^T gives together with (2.57), the following linear model of condition equations:

$$\underset{n\times m}{B^T}\underset{m\times 1}{E\{\underline{y}\}} \quad = \quad \underset{n\times 1}{0} \quad ; \quad \underset{m\times m}{D\{\underline{y}\}} \quad = \quad \underset{m\times m}{Q_y} \quad ; \quad \operatorname{rank} B = r = m-n \tag{2.58}$$

Example 9

Consider the following linear A-model:

$$E\{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \end{bmatrix}\} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x \ ; \ D\{\underline{y}\} = Q_y \tag{2.59}$$

Since $m = 3$, $n = 1$ and rank $A = 1 = n$, the redundancy equals $m - n = 2$. Hence two linearly independent condition equations can be formulated. The two vectors

$$b_1 = (1,-1,0)^T \text{ and } b_2 = (0,1,-1)^T$$

are linearly independent and are both orthogonal to the column vector of matrix A in (2.59). Hence the with (2.59) corresponding linear model of condition equations reads

$$\underbrace{\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}}_{B^T} E\{\underbrace{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \end{bmatrix}}_{\underline{y}}\} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ ; \ D\{\underline{y}\} = Q_y \tag{2.60}$$

□

Table 2.4 Least-Squares estimation.

THE LINEAR B-MODEL

$$\underset{n\times m}{B^T}\,\underset{m\times 1}{E\{\underline{y}\}} = \underset{n\times 1}{0} \quad ; \quad \underset{m\times m}{D\{\underline{y}\}} = \underset{m\times m}{Q_y} \quad ; \quad \text{rank } B = r = m-n$$

LEAST-SQUARES ESTIMATORS

$$\underline{\hat{y}} = [I - Q_y B(B^T Q_y B)^{-1} B^T]\underline{y} \quad ; \quad \underline{\hat{e}} = \underline{y} - \underline{\hat{y}}$$

VARIANCES AND COVARIANCES

$$Q_{\hat{y}} = Q_y B(B^T Q_y B)^{-1} B^T Q_y \quad ; \quad Q_{\hat{e}} = Q_y - Q_{\hat{y}} \quad ; \quad Q_{\hat{y}\hat{e}} = 0$$

Now that we have the disposal of the linear B-model (2.58), how are we going to compute the corresponding least-squares estimators? We know how to compute the least-squares estimators for the linear A-model. The corresponding formulae are however all expressed in terms of the A-matrix. What is needed is therefore to transform these formulae such that they are expressed in terms of the B-matrix. This is possible with the following important matrix identity:

$$A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} = I - Q_y B(B^T Q_y B)^{-1} B^T \tag{2.61}$$

The proof of this matrix identity goes as follows. We define two matrices C and $\bar{C}$ as:

$$C = \left[A \vdots Q_y B\right] \quad \text{and} \quad \bar{C} = \begin{bmatrix} (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \\ \cdots\cdots\cdots \\ (B^T Q_y B)^{-1} B^T \end{bmatrix} \tag{2.62}$$

Since both matrices C and $\bar{C}$ are of the order $m \times m$ and since both can be shown to be of full rank, it follows that both matrices are invertible. From (2.62) follows with the help of (2.57) that $\bar{C}C = I_m$. Hence $\bar{C} = C^{-1}$ and therefore $C\bar{C} = I_m$. Substitution of (2.62) into this last expression proofs (2.61).

With (2.61) and the least-squares results of table 2 of section 2.3.4 we are now in the position to derive the expressions for the least-squares estimators in terms of the matrix B. The results are summarized in table 2.4.

For some applications it may happen, when formulating the condition equations, that the linear combinations of the observables do not sum up to a zero vector, but instead to a *known* nonzero vector b. In that case the model of condition equations reads

$$B^T E\{\underline{y}\} = b \quad ; \quad D\{\underline{y}\} = Q_y$$

Due to the fact that $b \neq 0$, the solution for $\underline{\hat{y}}$ now takes a somewhat different form. It reads

$$\underline{\hat{y}} = \underline{y} - Q_y B(B^T Q_y B)^{-1}(B^T \underline{y} - b)$$

Note that this solution reduces to the one given earlier, when $b = 0$.

2.5.2 Nonlinear condition equations

Just like in case of the *A*-model, there are very few geodetic applications for which the model of condition equations is linear. In most cases the model of condition equations is nonlinear. The nonlinear *B*-model reads:

$$B^T(E\{\underline{y}\}) = 0 \quad ; \quad D\{\underline{y}\} = Q_y \tag{2.63}$$

where $B^T(.)$ is a nonlinear vectorfunction from R^m into R^{m-n}. The relation between the nonlinear *B*-model and the nonlinear *A*-model is given by

$$B^T(A(x)) = 0 \ , \forall\, x \in R^n \tag{2.64}$$

This is the *nonlinear* generalization of (2.57). If we take the partial derivative with respect to x of (2.64) and apply the chainrule, we get

$$[\partial_y B(y_0)]^T [\partial_x A(x_0)] = 0 \quad ; \quad y_0 = A(x_0) \tag{2.65}$$

This is the *linearized* version of (2.64). Compare (2.65) with (2.57). With (2.65) we are now in the position to construct the linearized *B*-model from the linearized *A*-model (2.49). Premultiplication of (2.49) with the matrix $[\partial_y B(y_0)]^T$ gives together with (2.65) the result

$$[\partial_y B(y_0)]^T E\{\Delta\underline{y}\} = 0 \quad ; \quad D\{\Delta\underline{y}\} = Q_y \tag{2.66}$$

This is the *linearized B-model*. With (2.66) we are now in the position again to apply our standard least-squares estimation formulae.

2.6 Special Least-Squares procedures

In this section three special cases of least-squares estimation will be discussed. They are: recursive least-squares, constrained least-squares and minimally constrained least-squares. In particular the last two will be needed when we discuss the adjustment and testing of geodetic networks. Constrained least-squares, which can be seen as a particular form of recursive least-squares, deals with the problem of solving a model of observation equations, when there are explicit constraints on the unknown parameters. As it is shown, the solution can be obtained in two steps. In the first step the model is solved without the

constraints, while in the second step, the final solution is obtained by using the results of the first step in a formulation based on condition equations.

Minimally constrained least-squares is needed when the design matrix fails to be of full rank. This typically occurs in cases of free network adjustments. Due to the rank defect of the design matrix, a set of necessary and sufficient constraints need to be imposed on the parameter vector in order to be able to compute a solution. The set of constraints that can be imposed however, is not unique. This implies that a whole family of solutions can be computed, of which each member is characterized by a particular set of minimal constraints.

2.6.1 Recursive least-squares

In this section we will partition the observation equations into two sets and show how the least-squares solution can be obtained using a step-wise or recursive approach.

Let us assume that the model of observation equations $E\{\underline{y}\} = Ax$, $D\{\underline{y}\} = Q_y$ is partitioned as

$$E\{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix}\} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} x \ , \ D\{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix}\} = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} \tag{2.67}$$

Note that $\underline{y}_1$ and $\underline{y}_2$ are assumed to be *uncorrelated*. This assumption is essential in order to be able to formulate a recursive solution. We will denote the least-squares solution which is based on the first set of observation equations as $\underline{\hat{x}}_{(1)}$ and the least- squares solution which is based on both sets, as $\underline{\hat{x}}_{(2)}$. The least-squares solution which is based on the first set of observation equations reads

$$\underline{\hat{x}}_{(1)} = (A_1^T Q_1^{-1} A_1)^{-1} A_1^T Q_1^{-1} \underline{y}_1 \ , \ Q_{\hat{x}_{(1)}} = (A_1^T Q_1^{-1} A_1)^{-1} \tag{2.68}$$

and the least-squares solution which is based on the complete set of observation equations, reads

$$\begin{cases} \underline{\hat{x}}_{(2)} = (A_1^T Q_1^{-1} A_1 + A_2^T Q_2^{-1} A_2)^{-1} (A_1^T Q_1^{-1} \underline{y}_1 + A_2^T Q_2^{-1} \underline{y}_2) \\ Q_{\hat{x}_{(2)}} = (A_1^T Q_1^{-1} A_1 + A_2^T Q_2^{-1} A_2)^{-1} \end{cases} \tag{2.69}$$

Comparing the two solutions shows that the second solution can also be written as

$$\underline{\hat{x}}_{(2)} = (Q_{\hat{x}_{(1)}}^{-1} + A_2^T Q_2^{-1} A_2)^{-1} (Q_{\hat{x}_{(1)}}^{-1} \underline{\hat{x}}_{(1)} + A_2^T Q_2^{-1} \underline{y}_2) \ , \ Q_{\hat{x}_{(2)}} = (Q_{\hat{x}_{(1)}}^{-1} + A_2^T Q_2^{-1} A_2)^{-1} \tag{2.70}$$

This result shows that the solution of the partitioned model can be obtained in two steps. First one solves for the first set of observation equations. This gives $\underline{\hat{x}}_{(1)}$ and $Q_{\hat{x}_{(1)}}$. Then in the second step this result is used together with $\underline{y}_2$ to obtain $\underline{\hat{x}}_{(2)}$ and $Q_{\hat{x}_{(2)}}$. This recursive procedure has an important implication in practice. It implies that if new observations, say $\underline{y}_2$, become available one does not need to save the old observations $\underline{y}_1$ to compute $\underline{\hat{x}}_{(1)}$. One can compute $\underline{\hat{x}}_{(2)}$ from $\underline{y}_2$ and the solution $\underline{\hat{x}}_{(1)}$ of the first step. It will be clear that one can extent the recursion to more than two steps by using a partitioning of the model of observation equations into more than two sets.

Note that the first equation of (2.70) can also be written as

$$\underline{\hat{x}}_{(2)} = \underline{\hat{x}}_{(1)} + (Q^{-1}_{\hat{x}_{(1)}} + A_2^T Q_2^{-1} A_2)^{-1} A_2^T Q_2^{-1} (\underline{y}_2 - A_2 \underline{\hat{x}}_{(1)}) \tag{2.71}$$

The expression now clearly shows how $\underline{\hat{x}}_{(2)}$ is found from updating $\underline{\hat{x}}_{(1)}$. The correction to $\underline{\hat{x}}_{(1)}$ depends on the difference $\underline{y}_2 - A_2\underline{\hat{x}}_{(1)}$. Since $A_2\underline{\hat{x}}_{(1)}$ can be interpreted as the *prediction* of $E\{\underline{y}_2\}$ based on $\underline{y}_1$, the difference $\underline{y}_2 - A_2\underline{\hat{x}}_{(1)}$ is called the *predicted residual* of $E\{\underline{y}_2\}$. Note that the predicted residual is *not* the same as the least-squares residual. The least-squares residual of $E\{\underline{y}_2\}$ reads namely $\underline{y}_2 - A_2\underline{\hat{x}}_{(2)}$.

Note that in the above expressions we need to invert a matrix having an order which is equal to the number of entries in the parameter vector x. One can however also derive an expression for $\underline{\hat{x}}_{(2)}$ in which a matrix needs to be inverted which has an order equal to the dimension of $\underline{y}_2$. From the matrix identity

$$(Q^{-1}_{\hat{x}_{(1)}} + A_2^T Q_2^{-1} A_2)^{-1} A_2^T Q_2^{-1} = Q_{\hat{x}_{(1)}} A_2^T (Q_2 + A_2 Q_{\hat{x}_{(1)}} A_2^T)^{-1}$$

follows that the solution (2.70) may also be written as

$$\left\{ \begin{array}{lcl} \underline{\hat{x}}_{(2)} & = & \underline{\hat{x}}_{(1)} + Q_{\hat{x}_{(1)}} A_2^T (Q_2 + A_2 Q_{\hat{x}_{(1)}} A_2^T)^{-1} (\underline{y}_2 - A_2 \underline{\hat{x}}_{(1)}) \\ Q_{\hat{x}_{(2)}} & = & Q_{\hat{x}_{(1)}} - Q_{\hat{x}_{(1)}} A_2^T (Q_2 + A_2 Q_{\hat{x}_{(1)}} A_2^T)^{-1} A_2 Q_{\hat{x}_{(1)}} \end{array} \right. \tag{2.72}$$

Both expressions (2.70) and (2.72) give identical results. But the second expression is more advantageous than the first, when the dimension of $\underline{y}_2$ is small compared to the dimension of x. In that case the order of the matrix that needs to be inverted is smaller.

The above expression shows that the correction to $\underline{\hat{x}}_{(1)}$ is small if the predicted residual is small. This is also what one would expect. Also note that the correction is small if the variance of $\underline{\hat{x}}_{(1)}$ is small. This is also understandable, because if the variance of $\underline{\hat{x}}_{(1)}$ is small one has more confidence in it and one therefore would like to give more weight to it than to $\underline{y}_2$.

The above expression also shows how the variance matrix gets updated. Because of the minus sign, the precision of the estimator gets better. This is understandable, since by including $\underline{y}_2$ more information is available to estimate x.

The above results show how the least-squares solution of the parameter vector x can be updated in recursive form. But this is not the only solution that can be updated recursively. Also the weighted sum-of- squares of the least-squares residuals, can be updated recursively. It can be shown that

$$\underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}} = \underline{\hat{e}}_1^T Q_1^{-1} \underline{\hat{e}}_1 + \underline{v}_2^T Q_{v_2}^{-1} \underline{v}_2 \tag{2.73}$$

where $\underline{\hat{e}}_1 = \underline{y}_1 - A_1\underline{\hat{x}}_{(1)}$ is the least-squares residual vector of the first step and $\underline{v}_2 = \underline{y}_2 - A_2\underline{\hat{x}}_{(1)}$ is the predicted residual which comes available in the second step.

2.6.2 Constrained least-squares

It may happen for a particular application that we know, when formulating the model of observation equations, that some strict relations exist between the entries of the pa-

rameter vector. In that case, we are dealing with a model of observation equations, with *constraints* on the parameter vector. Our model reads then

$$E\{\underline{y}\} = Ax \ , \ D\{\underline{y}\} = Q_y \ \text{ with } \ B^T x = b \tag{2.74}$$

The equations of $B^T x = b$ constitute the constraints, where matrix B and vector b are assumed known. In order to solve the above model in a least-squares sense, we can make use of the results of the previous section. First note that we may write (2.74) also as

$$E\{\begin{bmatrix} \underline{y} \\ \underline{b} \end{bmatrix}\} = \begin{bmatrix} A \\ B^T \end{bmatrix} x \ , \ D\{\begin{bmatrix} \underline{y} \\ \underline{b} \end{bmatrix}\} = \begin{bmatrix} Q_y & 0 \\ 0 & Q_b = 0 \end{bmatrix} \tag{2.75}$$

This is again a model of observation equations which has been partitioned into two sets. The only difference with the model treated in the previous section is that the variance matrix corresponding to the second set of observation equations is zero ($Q_b = 0$). This variance matrix is set to zero, since it is assumed that the relations $B^T x = b$ are strictly valid. Thus $\underline{b}$, with sample value b, is interpreted as an observable having a zero variance matrix.

Based on the results of the previous section, it follows that the solution of the above model can also be obtained in two steps. The solution of the first step will be denoted as $\underline{\hat{x}}$ and the solution of the second step as $\underline{\hat{x}}_b$. The solution of the first step reads then

$$\underline{\hat{x}} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} \ , \ Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1} \tag{2.76}$$

This is the solution one would get when the constraints are *not* taken into account. The solution of the second step reads

$$\begin{cases} \underline{\hat{x}}_b &= \underline{\hat{x}} + Q_{\hat{x}} B (B^T Q_{\hat{x}} B)^{-1} (b - B^T \underline{\hat{x}}) \\ Q_{\hat{x}_b} &= Q_{\hat{x}} - Q_{\hat{x}} B (B^T Q_{\hat{x}} B)^{-1} B^T Q_{\hat{x}} \end{cases} \tag{2.77}$$

This solution directly follows from using (2.72). Note that A_2 and Q_2 of the previous section, correspond with B^T and $Q_b = 0$ of the present section.

The above result shows that also the solution of a constrained model of observation equations can be obtained in two steps. The constraints are disregarded in the first step. In the second step, the results of the first step are used together with the constraints. Note that the solution of the second step is in fact the solution one would obtain when solving for the model of *condition equations* $B^T E\{\underline{\hat{x}}\} = b \ , \ D\{\underline{\hat{x}}\} = Q_{\hat{x}}$.

As in the previous section, also the weighted sum-of-squares of the least-squares residuals can be written as a sum. If we denote the least- squares residual vector for the constrained model as $\underline{\hat{e}}_b$ and its counterpart when the constraints are disregared as $\underline{\hat{e}}$, it follows that

$$\underline{\hat{e}}_b^T Q_y^{-1} \underline{\hat{e}}_b = \underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}} + (b - B^T \underline{\hat{x}})^T (B^T Q_{\hat{x}} B)^{-1} (b - B^T \underline{\hat{x}}) \tag{2.78}$$

Remark: In some applications it may happen that we perform the adjustment with constraints on the parameters, although we know that these constraints are in fact not deterministic but instead stochastic. Thus although we know that $Q_b \neq 0$, we then still perform the adjustment as if $Q_b = 0$. This approach of course, will give a result which is less

optimal. Sometimes however, such an approach is dictated by what is *practically* feasible. A good example in this respect is the connection of a geodetic network to existing control points. Since it is not practical to see the coordinates of the control points change, everytime a new network is connected to them, the adjustment is carried out in such a way that the coordinates of the existing control points remain fixed. Hence in the adjustment process, their variance matrix is set to zero. But this does not mean that we are allowed to set this variance matrix to zero as well when applying the error propagation law. In this case, the second equation of (2.77) is thus not valid anymore. This equation was namely obtained from applying the error propagation law to the first equation, under the assumption that the constraints are nonstochastic. To obtain a proper precision description, we need to take Q_b into account as well. If this is done, we obtain instead of the second equation of (2.77), the following expression for the variance matrix

$$Q_{\hat{x}_b} = Q_{\hat{x}} - Q_{\hat{x}}B(B^T Q_{\hat{x}}B)^{-1}[B^T Q_{\hat{x}}B - Q_b](B^T Q_{\hat{x}}B)^{-1}B^T Q_{\hat{x}} \tag{2.79}$$

This result reduces to that of (2.77) when $Q_b = 0$, showing that $Q_{\hat{x}_b} < Q_{\hat{x}}$. This inequality is not guaranteed anymore however, when $Q_b \neq 0$. It then depends on whether $B^T Q_{\hat{x}}B > Q_b$ holds true or not. This explains why it is of importance when connecting networks to existing control, to have an existing control which is of a better precision than the precision of the network connected to it. If this is not the case, then $B^T Q_{\hat{x}}B < Q_b$ and thus $Q_{\hat{x}_b} > Q_{\hat{x}}$, showing that the precision of the network after the connection is *poorer* than it was before the connection.

2.6.3 Minimally constrained least-squares

Up to this point, all the matrices that we worked with were assumed to be of full rank. We will now investigate what can be done when the A-matrix of the model of observation equations

$$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{A}\ \underset{n\times 1}{x}\quad , \quad \underset{m\times 1}{D\{\underline{y}\}} = \underset{m\times m}{Q_y} \tag{2.80}$$

is less than of full rank. Let us assume that the rank of matrix A is given as

$$\text{rank } A = r < n$$

Hence, the *rank defect* equals $(n-r)$. This implies that there exist $(n-r)$ linear independent combinations of the column vectors of matrix A that produce the zero vector. These linear combinations are said to span the *null space* of matrix A. The null space of A is defined as

$$N(A) = \{x \in R^n \mid Ax = 0\}$$

It is a subspace of the parameter space R^n and its dimension equals $(n-r)$. Let us assume that the columns of the $n \times (n-r)$ matrix G span the null space of A. Then

$$R(G) = N(A) \text{ and } AG = O$$

Thus the range space of G equals the null space of A and the columns of matrix G form the linear combinations that need to be taken of the columns of matrix A to obtain the zero vector. Since $AG = O$, it follows that

$$E\{\underline{y}\} = Ax = A(x + G\gamma) \quad , \text{ with } \gamma \in R^{n-r}$$

This shows that $E\{\underline{y}\}$ remains unchanged, when we add the vector $G\gamma$, with $\gamma \in R^{n-r}$ arbitrary, to the vector x. Hence, $E\{\underline{y}\}$ is *invariant* to these type of changes of the parameter vector. It will now be clear, since $E\{\underline{y}\}$ is insensitive to the above changes of the parameter vector, that one can not expect the rank defect model of observation equations to have a unique solution. The information content of the measurements is simply not sufficient enough to determine x uniquely.

In practice one will meet such situations for instance, when adjusting so-called *free networks*. Imagine the simple example of adjusting a single levelling loop. In this case the levelled height differences constitute the measurements and the heights of the points in the loop constitute the unknown parameters. It will be clear that one can not determine heights from observed height differences only. The height differences will not change when one adds a constant value to all the heights of the points in the levelling loop. This shows that we need some additional information, in order to be able to solve for the heights. For this simple example, the height of one of the points in the levelling loop would suffice already.

Also for the general case, the lack of information in $E\{\underline{y}\}$ to determine x uniquely, implies that additional information is needed. We will formulate the additional information as constraints on the parameter vector x. Thus instead of (2.80), we will consider the constrained model

$$E\{\underline{y}\} = Ax \ , \ D\{\underline{y}\} = Q_y \ \text{ with } \ B^T x = 0 \tag{2.81}$$

(*note*: For reasons of simplicity we have set the value of the constant vector b in the constraints, equal to zero.) The matrix B is assumed to be of full rank. Thus the rows of the matrix B^T are assumed to be linearly independent.

When introducing the constraints, it is of importance to understand, that they are merely used as a tool to allow us to be able to compute a solution for x (*note*: there are other, but equivalent, ways to deal with a rank defect model of observation equations, for instance by using the theory of *generalized inverses*. This however, is beyond the scope of the present lecture notes). Since the constraints are only there to allow us to compute a solution for x, the constraints should not interfere with the *intrinsic* properties of the adjustment itself. In other words, the constraints should contain the information which is *minimally* needed to eliminate the lack of information in $E\{\underline{y}\}$. This implies that the contraints should satisfy two conditions. First, the constraints should be such that indeed a solution for x can be computed. This however, is not the only condition the constraints should satisfy. If it would be the only condition, then $B = I$ would be an admissible choice. But this choice can not be allowed of course, since it *overconstraints* the solution. In fact when B is chosen equal to the identity matrix, no adjustment would be necessary anymore and the measurements would not contribute to the solution. This clearly, is not acceptable. Thus the constraints should also satisfy a second condition, which is that the

least-squares solution of the measurements, $\hat{\underline{y}}$, is invariant to the particular choice of the constraints.

If we translate the above conditions in terms of linear algebra, it follows that the constraints are admissible, if and only if the matrix B satisfies

$$R^n = R(B) \oplus R(A^T) \tag{2.82}$$

Thus the range space of matrix B should be *complementary* to the range space of matrix A^T. Two spaces are said to be complementary, when the two spaces together span the whole space of real numbers and their intersection only contains the zero vector. Since the dimension of R^n equals n and the dimension of $R(A^T)$ equals the rank of A, which is rank $A = r$, it follows that the dimension of $R(B)$ must equal

$$\text{dimension } R(B) = n - r$$

This shows that the number of linearly independent constraints that are admitted, equals $(n-r)$. This is also understandable if one considers the dimension of the null space of A. That is, one needs as many constraints as there are dimensions in the null space of A.

An alternative, but equivalent formulation of the above condition (2.82) can be formulated as follows. Let the linear independent columns of the $n \times r$ matrix $B^\perp$ span the null space of matrix B^T. The above condition is then equivalent to

$$R^n = N(A) \oplus R(B^\perp) \tag{2.83}$$

Thus the range space of matrix $B^\perp$ should be complementary to the null space of matrix A. A direct consequence of this condition is that the matrix G, of which the columns span $N(A)$, and the matrix $B^\perp$, together form a square and full rank matrix. Thus the partitioned matrix $(B^\perp, G)$ is square and invertible. Since it is square and invertible, it may be used to reparametrize the parameter vector x as

$$x = (B^\perp, G) \begin{bmatrix} \beta \\ \gamma \end{bmatrix} \tag{2.84}$$

This equation establishes a one-to-one relation between on the one hand x and on the other hand β and γ. When we substitute (2.84) into (2.81), we obtain

$$E\{\underline{y}\} = AB^\perp\beta \ , \ D\{\underline{y}\} = Q_y \ \text{ with } \ B^T G\gamma = 0 \tag{2.85}$$

since $AG = O$ and $B^T B^\perp = O$. Note that this reparametrization shows which part of the unknown parameters can be determined from the measurements and which part is determined by the constraints. Since the matrix $B^T G$ is square and invertible, it directly follows from the constraints that $\gamma = 0$ (*note*: the invertibility of $B^T G$ is a direct consequence of (2.82) or (2.83)). Thus γ is determined by the constraints and β is the part that still needs to be determined in a least-squares sense from the measurements. The least-squares solution for β reads

$$\hat{\underline{\beta}} = [(B^\perp)^T A^T Q_y^{-1} A(B^\perp)]^{-1} (B^\perp)^T A^T Q_y^{-1} \underline{y}$$

If we substitute this together with $\gamma = 0$ into (2.84), we obtain the least-squares solution of the minimally constrained model (2.81) as

$$\begin{cases} \underline{\hat{x}}_b &= (B^\perp)[(B^\perp)^T A^T Q_y^{-1} A(B^\perp)]^{-1}(B^\perp)^T A^T Q_y^{-1}\underline{y} \\ Q_{\hat{x}_b} &= (B^\perp)[(B^\perp)^T A^T Q_y^{-1} A(B^\perp)]^{-1}(B^\perp)^T \end{cases} \tag{2.86}$$

This is the unique least-squares solution of the *minimally constrained* model (2.81) and it is one of the infinitely many least-squares solutions of the *rank defect* model (2.80). The whole family of least-squares solutions of model (2.80) reads then

$$\underline{\hat{x}} = \underline{\hat{x}}_b + G\gamma \tag{2.87}$$

where γ is still undetermined.

An important property of these least-squares solutions is that they all produce the same least-squares solution for the measurements. Thus $\underline{\hat{y}} = A\underline{\hat{x}}$ is invariant for the choice of the constraints $B^T x = 0$, as long as they are minimal, that is, as long as matrix B satisfies the condition (2.82). In terms of the example of the levelling loop this corresponds to the fact that the least-squares solution for the height differences will be invariant to the choice which one of the points in the loop is assumed to be constrained in height.

S-transformations: Above we have shown how an arbitrary least-squares solution $\underline{\hat{x}}$ of model (2.80) can be obtained from a particular solution $\underline{\hat{x}}_b$ and an undetermined part $G\gamma$. But also the inverse is possible. That is, one can also obtain a particular solution from any arbitrary solution. To see this, we premultiply (2.87) with B^T. This gives $B^T\underline{\hat{x}} = B^T G\gamma$, since $B^T\underline{\hat{x}}_b = 0$ by definition. Hence, $\underline{\gamma} = (B^T G)^{-1}B^T\underline{\hat{x}}$. From substituting this into (2.87) follows then

$$\begin{cases} \underline{\hat{x}}_b &= [I - G(B^T G)^{-1}B^T]\underline{\hat{x}} \\ Q_{\hat{x}_b} &= [I - G(B^T G)^{-1}B^T]Q_{\hat{x}}[I - G(B^T G)^{-1}B^T]^T \end{cases} \tag{2.88}$$

The second equation has been obtained from applying the error propagation to the first equation. This result shows that we only need to know the transformation matrix

$$S_b = [I - G(B^T G)^{-1}B^T] \tag{2.89}$$

and thus the matrices B and G, in order to obtain $\underline{\hat{x}}_b$ from $\underline{\hat{x}}$. This transformation matrix is a very famous one in geodesy and is known as the *S-transformation*. With the *S*-transformation we are thus able to transform any least-squares solution to a minimally constrained solution, specified by the matrix B. The matrix G, of which the columns span the null space of A, depends on A and thus on the type of measurements which are included in the model. For a levelling network for instance, the undetermined part will correspond with a constant shift in the height of all points. For a triangulation network however, the undetermined part will correspond to a translation, a rotation and a scale change of the network. Angles do namely not contain information on the position, orientation and size of the network.

One may wonder how the above results correspond to the results which were obtained in the previous section for the case of constrained least- squares. In the previous section, where matrix A was assumed to be of full rank, we showed that the least-squares solution of the constrained model could be obtained in two steps. The unconstrained solution of the first step, was used as input for the second step to finally come up with the constrained solution. These two steps can also be recognized in case of the minimally constrained solution. The first step corresponds then with $\underline{\hat{x}}$, which is an arbitrary least-squares solution of model (2.80), and the second step corresponds then with (2.88).

Unbiasedness: In one of our earlier sections we claimed that the least-squares estimator of the parameter vector is unbiased and thus that $E\{\underline{\hat{x}}\} = x$ holds true. At that point however, we still assumed the matrix A to be of full rank. So, what happens when the matrix A is not of full rank? We know that in that case the measurements fail to contain enough information to determine x uniquely. We may therefore suspect that in that case the least-squares estimators will also fail to be unbiased estimators of the parameter vector x. And indeed, when we take the expectation of (2.86), we obtain

$$\begin{aligned} E\{\underline{\hat{x}}_b\} &= (B^\perp)[(B^\perp)^T A^T Q_y^{-1} A(B^\perp)]^{-1}(B^\perp)^T A^T Q_y^{-1} E\{\underline{y}\} \\ &= (B^\perp)[(B^\perp)^T A^T Q_y^{-1} A(B^\perp)]^{-1}(B^\perp)^T A^T Q_y^{-1} Ax \neq x \end{aligned}$$

Thus $\underline{\hat{x}}_b$ is *not* an unbiased estimator of x. If the minimally constrained least-squares estimator is not an unbiased estimator of x, what is it then an unbiased estimator of? To see this, we substitute (2.84) into the above equation. This gives $E\{\underline{\hat{x}}_b\} = B^\perp \beta = x - G\gamma$, or when $\gamma = (B^T G)^{-1} B^T x$ is substituted in it,

$$E\{\underline{\hat{x}}_b\} = [I - G(B^T G)^{-1} B^T]x = S_b x \tag{2.90}$$

This shows that $\underline{\hat{x}}_b$ is an unbiased estimator of $S_b x$. In terms of our levelling example it means that the minimally constrained least-squares estimators of the heights are not unbiased estimators of absolute heights, but instead are unbiased estimators of the heights which are defined through the minimal constraint, i.e. the constraint height of one of the points in the levelling network.

2.7 Quality control: precision

Now that we have developed most of the adjustment theory that will be needed in these lecture notes, let us pause for a few moments and reflect on the various steps involved when an adjustment is carried out.

Formulate model: Before any adjustment can be carried out, one first must have a clear idea of: (1) the observables that are going to be used, (2) their assumed relation to the unknown parameters and (3) their assumed precision. Thus first one must be able to formulate the model of observation equations

$$E\{\underline{y}\} = Ax \ , \ D\{\underline{y}\} = Q_y \tag{2.91}$$

This model consists of the *functional* model $E\{\underline{y}\} = Ax$, which is specified through the $m \times n$ design matrix A, and of the *stochastic* model $D\{\underline{y}\} = Q_y$, which is specified through

the $m \times m$ variance matrix Q_y. In the functional model the link is established between the measurements and the unknown parameters. It captures the geometry of the network. In the stochastic model, one specifies the precision of the measurements. It depends on the type of measurement equipment used and on the measurement procedures used.

Adjustment: Based on the above model and on an available sample of $\underline{y}$, the measurements, one can commence with the actual adjustment. Using the principle of least-squares, the estimator for the unknown parameter vector x reads

$$\underline{\hat{x}} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} \tag{2.92}$$

Here it has been assumed that the design matrix is of full rank. If this is not the case, then the approach of using minimal constraints needs to be used.

If we assume that $\underline{y}$ is normally distributed, then its probability density function is completely specified by its first two moments, being its expectation and its dispersion. This then also holds true for the probability density function of the estimator $\underline{\hat{x}}$. The *quality* of this estimator can thus be characterized by its expectation $E\{\underline{\hat{x}}\}$ and its dispersion $D\{\underline{\hat{x}}\}$. For its expectation we have

$$E\{\underline{\hat{x}}\} = x \tag{2.93}$$

Loosely speaking, this implies that if the adjustment is repeated a sufficient number of times, each time with a new sample from $\underline{y}$, that the various outcomes for $\hat{x}$ would on the average coincide with x. This property of *unbiasedness* is an important one and it is automatically fulfilled when the least-squares principle is used, *provided* of course that the model which forms the basis of the adjustment is correct. In the next chapter we will see what happens when the model is specified incorrectly.

Precision: For the moment we will assume that the model used is correct and thus that the estimator $\underline{\hat{x}}$ is unbiased. In that case, the quality of the estimator is completely captured by its dispersion, being the variance matrix $Q_{\hat{x}}$. Thus under the provision that the least-squares estimator is unbiased, we may judge its quality on the basis of its variance matrix

$$Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1} \tag{2.94}$$

The variance matrix is said to describe the precision of the estimator. It describes the amount of variability in samples of $\underline{\hat{x}}$ around its mean.

Since the variance matrix depends on A and on Q_y, one can change $Q_{\hat{x}}$ by changing A and/or Q_y. This thus gives us a way of improving the precision of the least-squares solution. For instance, if for a certain application the precision turns out to be not good enough, one may decide to use more precise measurement equipment. In that case Q_y changes for the better and also $Q_{\hat{x}}$ will then change for the better. In many applications however, one will not have too much leeway in choosing from different sets or types of measurement equipment. In that case one depends on A for improving the precision (*note*: for that reason, matrix A is also often referred to as the design matrix). There are two ways in which A can be changed. One can change the dimension of the matrix and/or one can change its structure. For instance, when the precision of a geodetic network is not good

enough one can decide to add more observations to the network. In that case, matrix A is extended row wise. However, it may also be possible that a mere change in the geometry of the network already suffices to improve its precision. In that case, it is the structure of A that changes, while its dimension stays the same.

Precision testing: Although we know how to change $Q_{\hat{x}}$ for the better or the worse, we of course still need a way of deciding when the precision, as expressed by the variance matrix, is good enough. It will be clear that this depends very much on the particular application at hand. What is important though, is that one has a *precision criterium* available by which the precision of the least-squares solution can be judged. The following are some approaches that can be used for testing the precision of the solution.

It may happen in a particular application, that one is only interested in one particular function of the parameters, say $\theta = f^T x$. In that case it becomes very easy to judge its quality. One then simply has to compare its variance with the given criterium

$$\sigma^2_{\hat{\theta}} = f^T Q_{\hat{x}} f < \text{criterium} \tag{2.95}$$

The situation becomes more difficult though, when the application at hand requires that the precision of more than one function needs to be judged upon. Let us assume that the functions of interest are given as $\theta = F^T x$, where F is a matrix. The corresponding variance matrix is then given as $Q_{\hat{\theta}} = F^T Q_{\hat{x}} F$. One way to judge the precision in this case, is by inspecting the variances and covariances of the matrix $Q_{\hat{\theta}}$. An alternative way would be to use the average precision as precision measure. In that case one relies on the trace of the variance matrix

$$\tfrac{1}{p}\text{trace}\,(F^T Q_{\hat{x}} F) < \text{criterium} \tag{2.96}$$

where p is the dimension of the vector θ. When using the trace one has to be aware of the fact that one is not taking all the information of the variance matrix $Q_{\hat{\theta}}$ into account. It depends only on the variances and not on the covariances. Hence, the correlation between the entries of $\hat{\theta}$ are then not taken into account. When using the trace one should also make sure that it makes sense to speak of an average variance. In other words the entries of θ should contain the same type of variables. It would not make sense to use the trace, when θ contains completely different variables, each having their own physical dimension.

The trace of $Q_{\hat{\theta}}$ equals the sum of its eigenvalues. Instead of the sum of eigenvalues, one might decide that it suffices to consider the largest eigenvalue λ_{max} only,

$$\lambda_{max}(F^T Q_{\hat{x}} F) < \text{criterium} \tag{2.97}$$

In that case one is thus testing whether the function of θ which has the poorest precision, still passes the precision criterium. When this test is passed successfully, one knows that all other functions of θ will also have a precision which is better than the criterium. For some applications this may be an advantage, but it could be a disadvantage as well. It could be a disadvantage in the sense that the above test could be overly conservative. That is, when the function having the poorest precision passes the above test, all other

functions, which by definition have a better precision, may turn out to be unnecessarily precise.

So far we assumed that the precision criterium was given in scalar form. But this need not be the case. The precision criterium could also be given in matrix form. In that case one is working with a *criterium matrix*, which we will denote as C_x. The precision test amounts then to testing whether the precision as expressed by the actual variance matrix $Q_{\hat{\theta}} = F^T Q_{\hat{x}} F$ is better than or as good as the precision expressed by the criterium matrix $F^T C_x F$. Also this test can be executed by means of solving an eigenvalue problem, but now it will be a *generalized* eigenvalue problem

$$| F^T Q_{\hat{x}} F - \lambda F^T C_x F | = 0 \tag{2.98}$$

Note that when the matrix $F^T C_x F$ is taken as the identity matrix, the largest eigenvalue of (2.98) reduces to that of (2.97). Using the generalized eigenvalue problem is thus indeed more general than the previous discussed approaches. It is characterized by the fact that it allows one to compare the actual variance matrix $Q_{\hat{\theta}}$ directly with its criterium. The two matrices $F^T Q_{\hat{x}} F$ and $F^T C_x F$ are identical, when all eigenvalues equal one, and all functions of $\hat{\theta}$ have a better precision than the criterium when the largest generalized eigenvalue is less than one.

So far we assumed the matrix A to be of full rank. In many geodetic applications however, this is not the case. We know from our section on minimally constrained least-squares, that the variance matrix $Q_{\hat{x}}$ will not be unique, when A has a rank defect. In that case the variance matrix depends on the chosen set of minimal constraints. Since these constraints should not affect our conclusion when evaluating the precision, we have to make sure that our procedure of precision testing is invariant for the minimal constraints. This is possible when we make use of the S-transformations. Let C_x be the criterium matrix and $Q_{\hat{x}_b}$ the variance matrix of a minimally constrained solution. The eigenvalues of the generalized eigenvalue problem

$$| Q_{\hat{x}_b} - \lambda S_b C_x S_b^T | = 0 \tag{2.99}$$

where S_b is the S-transformation that corresponds to the minimal constraints of $Q_{\hat{x}_b}$, are then invariant to the chosen set of minimal constraints. Thus when the largest eigenvalue is less than one, it is guaranteed that all functions of $\underline{\hat{x}}_b$ have a precision which is better than what they would have were $Q_{\hat{x}_b}$ be replaced by the criterium matrix.

Chapter 3

Testing and reliability

3.1 Introduction

In the previous chapter we considered the model of observation equations

$$E\{\underline{y}\} = Ax \ , \ D\{\underline{y}\} = Q_y \tag{3.1}$$

and showed how a solution for the unknown parameter vector x, based on the least-squares principle, could be obtained. The least-squares solution $\underline{\hat{x}}$ is optimal in the sense that it is unbiased and that it is of minimal variance in the class of linear unbiased estimators. These optimality properties only hold true however, when the model (3.1) is correct. There are many ways in which the model could have been misspecified. The functional model could be wrong, in which case $E\{\underline{y}\} \neq Ax$. As a consequence, the least-squares estimator becomes biased, $E\{\underline{\hat{x}}\} \neq x$. Or, the stochastic model could be wrong, in which case $D\{\underline{y}\} \neq Q_y$. As a consequence, the property of minimal variance is lost.

The topic of the present chapter is to present ways of checking the validity of the above model. We start off in section 3.2, by discussing the basic concepts of hypothesis testing. We consider the general form of a null hypothesis and an alternative hypothesis, show what a test between the two implies and discuss the two type of errors one can make. Following these basic concepts, we move on in section 3.3 to the testing of the above model against the alternative model

$$E\{\underline{y}\} = Ax + C\nabla \ , \ D\{\underline{y}\} = Q_y \tag{3.2}$$

Note that the two models only differ in their functional model. Hence we restrict ourselves in these lecture notes to misspecifications in the mean of $\underline{y}$. In geodetic practice these are by far the most frequently occurring type of model errors (e.g. measurement blunders, unaccounted systematic effects). In section 3.3, we present and discuss the test statistic $\underline{T}_q$ through which the model (3.1) can be tested against the alternative (3.2).

In most practical applications, it is usually not only one model error one is concerned about, but quite often many more than one. To each of these model errors there belongs a vector $C\nabla$, with the matrix C specifying how the model error is related to the vector of observables. As a result one is not dealing with only one alternative hypothesis, but with as many alternative hypotheses as there are model errors one is willing to consider. This implies that one needs a *testing procedure* for handling the various alternative hypotheses. In section 3.4 such a procedure is presented. It consists of a detection step, an identification step and an adaptation step.

Just like the results of an adjustment are not exact (that is, not deterministic, but stochastic), also the results of the statistical tests are not exact. The confidence one will have in the outcomes of the statistical tests depends in a large part on the 'strength' of the model. A practical way of diagnosing this confidence is provided for by the concept of

reliability. It is introduced in section 3.5. Reliability together with precision, can then be considered to describe the quality which one can expect of the results of an adjustment and testing. They are both considered in the last section of this chapter.

3.2 Basic concepts of hypothesis testing

3.2.1 Statistical hypotheses

Many social, technical and scientific problems result in the question whether a particular theory or hypothesis is true or false. In order to answer this question one can try to design an experiment such that its outcome can also be predicted by the postulated theory. After performing the experiment, one can then confront the experimental outcome with the theoretically predicted value and on the basis of this comparison try to conclude whether the postulated theory or hypothesis should be rejected. That is, if the outcome of the experiment disagrees with the theoretically predicted value, one could conclude that the postulated theory or hypothesis should be rejected. On the other hand, if the experimental outcome is in agreement with the theoretically predicted value, one could conclude that as yet no evidence is available to reject the postulated theory or hypothesis.

Example 10

According to the postulated theory or hypothesis the three points 1, 2 and 3 of figure 3.1 lie on one straight line. In order to test or verify this hypothesis

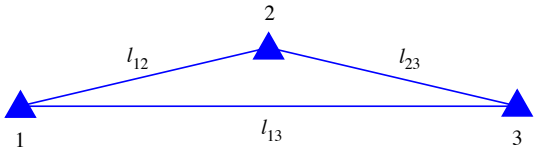


Figure 3.1 Three distances between three known points

we need to design an experiment such that its outcome can be compared with the theoretically predicted value.

If the postulated hypothesis is correct, the three distances l_{12}, l_{23} and l_{13} should satisfy the relation:

$$l_{13} = l_{12} + l_{23}$$

Thus, under the assumption that the hypothesis is correct we have:

$$H:\ l_{12} + l_{23} - l_{13} = 0 \tag{3.3}$$

To denote a hypothesis, we will use a capital H followed by a colon that in turn is followed by the assertion that specifies the hypothesis.

As an experiment we can now measure the three distances l_{12}, l_{23} and l_{13}, compute $l_{12} + l_{23} - l_{13}$ and verify whether this computed value agrees or disagrees with the theoretically predicted value of H. If it agrees, we are inclined

to accept the hypothesis that the three points lie on one straight line. In case of disagreement we are inclined to reject hypothesis H. □

It will be clear that in practice the testing of hypotheses is complicated by the fact that experiments (in particular experiments where measurements are involved) in general do not give outcomes that are exact. That is, experimental outcomes are usually affected by an amount of uncertainty, due for instance to measurement errors. In order to take care of this uncertainty, we will, in analogy with our derivation of Adjustment Theory in the previous chapter, model the uncertainty by making use of the results from the theory of random variables. The verification or testing of postulated hypotheses will therefore be based on the testing of hypotheses of random variables of which the probability distribution depends on the theory or hypothesis postulated. From now on we will therefore consider statistical hypotheses.

Statistical hypothesis: A statistical hypothesis is an assertion or conjecture about the probability distribution of one or more random variables, for which it is assumed that a random sample (mostly through measurements) is available.

The structure of a statistical hypothesis H is in general the following:

$$H: \underline{y} \sim p_{\underline{y}}(y|x) \tag{3.4}$$

with x fully or partially specified. This statistical hypothesis should be read as follows: According to H the scalar or vector observable random variable $\underline{y}$ has a probability density function given by $p_{\underline{y}}(y|x)$. The scalar, vector or matrix parameter x used in the notation of $p_{\underline{y}}(y|x)$ indicates that the probability density function of $\underline{y}$ is known except for the unknown parameter x. Thus, by specifying (either fully or partially) the parameter x, an assertion or conjecture about the density function of $\underline{y}$ is made. In order to see how a statistical hypothesis for a particular problem can be formulated, let us continue with our example 10.

Example 10 (continued)

We know from experience that in many cases the uncertainty in geodetic measurements can be *adequately* modeled by the normal distribution. We therefore model the three distances between the three points 1, 2 and 3 as normally distributed random variables [1] If we also assume that the three distances are uncorrelated and all have the same known variance $\frac{1}{3}\sigma^2$, the simultaneous probability density function of the three distance observables becomes:

$$\begin{bmatrix} \underline{l}_{13} \\ \underline{l}_{12} \\ \underline{l}_{23} \end{bmatrix} \sim N\left(\begin{bmatrix} E\{\underline{l}_{13}\} \\ E\{\underline{l}_{12}\} \\ E\{\underline{l}_{23}\} \end{bmatrix}, Q \right) \text{ with } Q = \frac{1}{3} I_3 \tag{3.5}$$

The notation $N(a, B)$ is a shorthand notation for the normal distribution, having a as mean and B as variance matrix. Statement (3.5) could already be

[1] Note that strictly speaking distances can never be normally distributed. A distance is always nonnegative, whereas the normal distribution, due to its infinite tails, admits negative sample values.

considered a statistical hypothesis, since it has the same structure as (3.4). Statement (3.5) asserts that the three distance observables are indeed normally distributed with unknown mean, but with known variancematrix Q. Statement (3.5) is however not yet the statistical hypothesis we are looking for. What we are looking for is a statistical hypothesis of which the probability density function depends on the theory or hypothesis postulated. For our case this means that we have to incorporate in some way the hypothesis that the three points lie on one straight line. We know *mathematically* that this assertion implies that

$$l_{12} + l_{23} - l_{13} = 0 \tag{3.6}$$

However, we cannot make this relation hold for the random variables $\underline{l}_{12}, \underline{l}_{23}$ and $\underline{l}_{13}$. This is simply because of the fact that random variables cannot be equal to a constant. Thus, a statement like: $\underline{l}_{12} + \underline{l}_{23} - \underline{l}_{13} = 0$ is nonsensical. What we can do is assume that relation (3.6) holds for the *expected* values of the random variables $\underline{l}_{12}, \underline{l}_{23}$ and $\underline{l}_{13}$:

$$E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} = 0 \tag{3.7}$$

For the hypothesis considered this relation makes sense. It can namely be interpreted as stating that if the measurement experiment were to be repeated a great number of times, then on the average the measurements will satisfy (3.7). With (3.5) and (3.7) we can now state our statistical hypothesis as:

$$H: \begin{bmatrix} \underline{l}_{13} \\ \underline{l}_{12} \\ \underline{l}_{23} \end{bmatrix} \sim N(\begin{bmatrix} E\{\underline{l}_{13}\} \\ E\{\underline{l}_{12}\} \\ E\{\underline{l}_{23}\} \end{bmatrix}, \frac{1}{3} I_3) \tag{3.8}$$

$$\text{with } E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} = 0$$

This hypothesis has the same structure of (3.4) with the three means playing the role of the parameter x. □

In many hypothesis-testing problems two hypotheses are discussed: The first, the hypothesis being tested, is called the *null hypothesis* and is denoted by H_0. The second is called the *alternative hypothesis* and is denoted by H_A. The thinking is that if the null hypothesis H_0 is false, then the alternative hypothesis H_A is true, and vice versa. We often say thatH_0 is tested against, or versus, H_A.

In studying hypotheses it is also convenient to classify them into one of two types by means of the following definition: if a hypothesis completely specifies the distribution, that is, if it specifies its functional form as well as the values of its parameters, it is called a *simple hypothesis*; otherwise it is called a *composite hypothesis*.

Example 10 (continued)

In our example, (3.8) is the hypothesis to be tested. Thus, the null hypothesis reads in our case:

$$H_0: \begin{bmatrix} \underline{l}_{13} \\ \underline{l}_{12} \\ \underline{l}_{23} \end{bmatrix} \sim N(\begin{bmatrix} E\{\underline{l}_{13}\} \\ E\{\underline{l}_{12}\} \\ E\{\underline{l}_{23}\} \end{bmatrix}, \frac{1}{3} I_3) \tag{3.9}$$

$$\text{with } E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} = 0$$

Since we want to find out whether $E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} = 0$ or not, we could take as alternative the inequality $E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} \neq 0$. However, we know from the geometry of our problem that the left hand side of the inequality can never be negative. The alternative should therefore read: $E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} > 0$. Our alternative hypothesis takes therefore the form:

$$H_A: \begin{bmatrix} \underline{l}_{13} \\ \underline{l}_{12} \\ \underline{l}_{23} \end{bmatrix} \sim N(\begin{bmatrix} E\{\underline{l}_{13}\} \\ E\{\underline{l}_{12}\} \\ E\{\underline{l}_{23}\} \end{bmatrix}, \frac{1}{3} I_3) \tag{3.10}$$

$$\text{with } E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} > 0$$

When comparing (3.9) and (3.10) we see that the type of the distribution of the observables and their variance matrix are not in question. They are assumed to be known and identical under both H_0 and H_A.

Both of the above hypotheses, H_0 and H_A, are examples of composite hypotheses. The above null hypothesis H_0 would become a simple hypothesis if the individual expectations of the observables were assumed known. □

3.2.2 Test of statistical hypotheses

After the statistical hypotheses H_0 and H_A have been formulated, one would like to test them in order to find out whether H_0 should be rejected or not.

Test of a statistical hypothesis: A test of a statistical hypothesis

$$H_0: \underline{y} \sim p_{\underline{y}}(y|x)$$

with x fully or partially specified, is a rule or procedure, in which a random sample of $\underline{y}$ is used for deciding whether to reject or not reject H_0. A test of a statistical hypothesis is completely specified by the so-called critical region, which will be denoted by K.

Critical region K: The critical region K of a test is the set of sample values of $\underline{y}$ for which H_0 is to be rejected. Thus, H_0 is rejected if $y \in K$.

It will be obvious that we would like to choose a critical region so as to obtain a test with desirable properties, that is, a test that is "best" in a certain sense. But let us first have a look at a simple testing problem for which, on more or less intuitive grounds, an acceptable critical region can be found.

Example 11

Let us assume that a geodesist measures a scalar variable, and that this measurement can be modeled as a random variable $\underline{y}$ with density function

$$\underline{y} \sim \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}(y - E\{\underline{y}\})^2] \tag{3.11}$$

Thus, it is assumed that $\underline{y}$ has a normal distribution with unit variance. Although this assumption constitutes a statistical hypothesis, it will not be tested here because the geodesist is quite certain of the validity of this assumption. The geodesist is however not certain about the value of the expectation of $\underline{y}$. His assumption is that the value of $E\{\underline{y}\}$ is x_0. This assumption is the statistical hypothesis to be tested. Denote this hypothesis by H_0. Then,

$$H_0 : E\{\underline{y}\} = x_0 \tag{3.12}$$

Let H_A denote the alternative hypothesis that $E\{\underline{y}\} \neq x_0$. Then:

$$H_A : E\{\underline{y}\} \neq x_0 \tag{3.13}$$

Thus the problem is one of testing the simple hypothesis H_0 against the composite hypothesis H_A. To test H_0, a single observation on the random variable $\underline{y}$ is made. In real-life problems one usually takes several observations, but to avoid complicating the discussion at this stage only one observation is taken here. On the basis of the value of $\underline{y}$ obtained, denoted by y, a decision will be made either to accept H_0 or reject it. The latter decision, of course, is equivalent to accepting H_A. The problem then is to determine what values of $\underline{y}$ should be selected for accepting H_0 and what values for rejecting H_0. If a choice has been made of the values of $\underline{y}$ that will correspond to rejection, then the remaining values of $\underline{y}$ will necessarily correspond to acceptance. As defined above, the rejection values of $\underline{y}$ constitute the critical region K of the test. Figure 3.2 shows the distribution of $\underline{y}$ under H_0 and under two possible alternatives H_{A_1} and H_{A_2}. Looking at this figure, it seems reasonable to reject H_0 if the observation y is remote enough from $E\{\underline{y}\}$. If H_0 is true, the

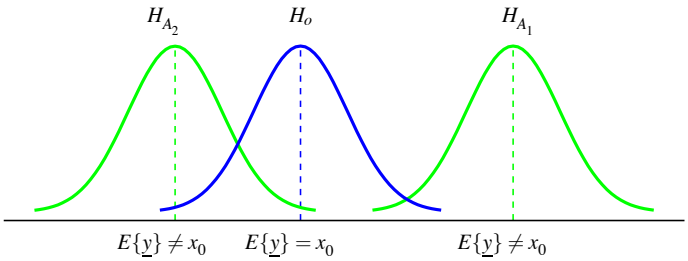


Figure 3.2 $H_0 : E\{\underline{y}\} = x_0$ *versus* $H_A : E\{\underline{y}\} \neq x_0$.

probability of a sample of $\underline{y}$ falling in a region remote from $E\{\underline{y}\}$ is namely small. And if H_A is true, this probability may be large. Thus the critical region K should contain those sample values of $\underline{y}$ that are remote enough from $E\{\underline{y}\}$. Also, since the alternative hypothesis can be located on either side of $E\{\underline{y}\}$, it seems obvious to have one portion of K located in the left tail of H_0 and one portion of K located in the right tail of H_0. Finally, one can argue that since the distribution is symmetric about its mean value, also the critical region K should be symmetric about $E\{\underline{y}\}$. This as a result gives the form of the critical region K as shown in figure 3.3. □

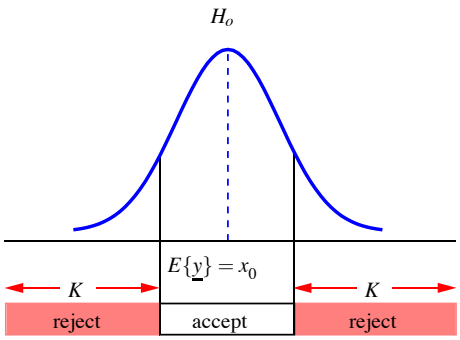


Figure 3.3 Critical region K for testing $H_0 : E\{\underline{y}\} = x_0$ *versus* $H_A : E\{\underline{y}\} \neq x_0$.

3.2.3 Two types of errors

We have seen that a test of a statistical hypothesis is completely specified once the critical region K of the test is given. The null hypothesis H_0 is rejected if the sample value or observation falls in the critical region, i.e. if $y \in K$. Otherwise the null hypothesis H_0 is accepted, i.e. if $y \notin K$. With this kind of thinking two types of errors can be made:

Type I error: Rejection of H_0 when in fact H_0 is true.

Type II error: Acceptance of H_0 when in fact H_0 is false.

Table 3.1 shows the decision table with the type I and II errors.

The *size* of a type I error is defined as the probability that a sample value of $\underline{y}$ falls in the critical region when in fact H_0 is true. This probability is denoted by α and is called the *size of the test* or the *level of significance* of the test. Thus:

$$\alpha = P(\text{type I error}) = P(\text{rejection of } H_0 \text{ when } H_0 \text{ is true})$$

or

$$\alpha = P(\underline{y} \in K | H_0) = \int_K p_{\underline{y}}(y|H_0)dy \tag{3.14}$$

The size of the test, α, can be computed once the critical region K and the probability density function of $\underline{y}$ is known under H_0. The *size* of a type II error is defined as the

Table 3.1 Decision table with type I and type II error.

	H_0 true	H_0 false
Reject H_0 $y \in K$	Wrong Type I error	Correct
Accept H_0 $y \notin K$	Correct	Wrong Type II error

probability that a sample value of $\underline{y}$ falls outside the critical region when in fact H_0 is false. This probability is denoted by β. Thus:

$$\beta = P(\text{type II error}) = P(\text{acceptance of } H_0 \text{when } H_0 \text{ is false})$$

or

$$\beta = P(\underline{y} \notin K|H_A) = 1 - \int_K p_{\underline{y}}(y|H_A)dy \tag{3.15}$$

The size of a type II error, β, can be computed once the critical region K and the probability density function of $\underline{y}$ is known under H_A.

Example 12

Assume that $\underline{y}$ is distributed as

$$\underline{y} \sim N(E\{\underline{y}\}, \sigma^2) \tag{3.16}$$

with known variance σ^2.

The following two *simple* hypotheses are considered

$$H_0 : E\{\underline{y}\} = x_0 \tag{3.17}$$

and

$$H_A : E\{\underline{y}\} = x_A > x_0 \tag{3.18}$$

The situation is sketched in figure 3.4.

Since the alternative hypothesis H_A is located on the right of the nullhypothesis H_0, it seems intuitively appealing to choose the critical region K right-sided. Figure 3.5a and 3.5b show two possible right-sided critical regions K. They also show the size of the test, α, which corresponds to the area under the graph of the distribution of $\underline{y}$ under H_0 for the interval of the critical region K.

The size of the test, α, can be computed once the probability density function of $\underline{y}$ under H_0 is known and the *form* and *location* of the critical region K is

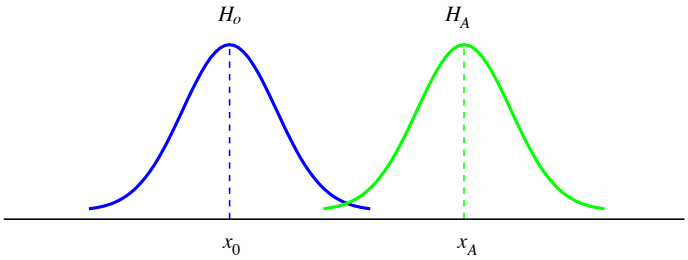


Figure 3.4 Two simple hypotheses: $H_0: E\{\underline{y}\} = x_0$ *and* $H_A: E\{\underline{y}\} = x_A > x_0$.

known. In the present example the form of the critical region has been chosen right-sided. Its location is determined by the value of k_α, the so-called critical value of the test. Thus, for the present example the size of the test can be computed as

$$\alpha = \int_{k_\alpha}^{\infty} p_{\underline{y}}(y|x_0)dy$$

or, since

$$p_{\underline{y}}(y|x_0) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_0)^2]$$

as

$$\alpha = \int_{k_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_0)^2]dy \tag{3.19}$$

When one is dealing with one-dimensional normally distributed random variables, one can usually compute the size of the test, α, from tables given for the *standard* normal distribution. In order to compute (3.19) with the help of such a table, we first have to apply a transformation of variables to (3.19). Since $\underline{y}$ is normally distributed under H_0 with mean x_0 and variance σ^2, it follows that the random variable $\underline{z}$, defined as

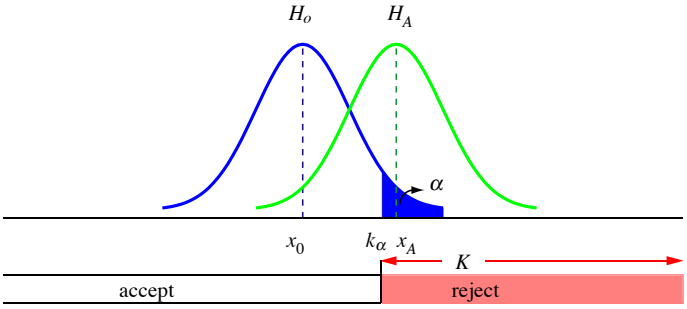
$$\underline{z} = \frac{\underline{y} - x_0}{\sigma} \tag{3.20}$$

is *standard* normally distributed under H_0. And since

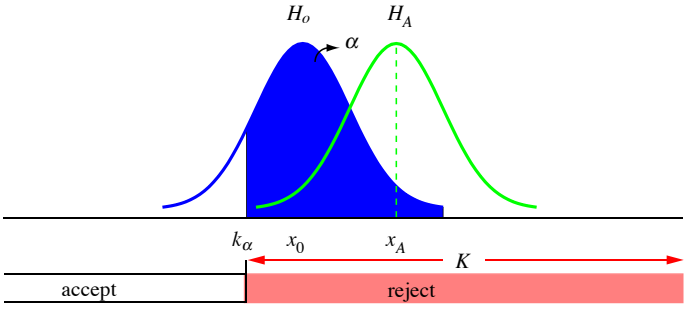
$$\alpha = P(\underline{y} > k_\alpha|H_0) = P(\underline{z} > \frac{k_\alpha - x_0}{\sigma}|H_0) \tag{3.21}$$

we can use the last expression of (3.21) for computing α. Application of the change of variables (3.20) to (3.19) gives

$$\alpha = \int_{\frac{k_\alpha - x_0}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}z^2]dz \tag{3.22}$$



(a)



(b)

Figure 3.5 Critical region K and size of test, α.

We can now make use of the table of the standard normal distribution. Table 2.6 shows some typical values of the α and k_α for the case that $x_0 = 1$ and $\sigma = 2$.

As we have seen the location of the critical region K is determined by the value chosen for k_α, the critical value of the test. But what value should we choose for k_α? Here the geodesist should base his judgement on his experience. Usually one first makes a choice for the size of the test, α, and then by using (3.22) or Table 3.2 determines the corresponding critical value k_α. For instance, if one fixes α at $\alpha = 0.01$, the corresponding critical value k_α (for the present example with $x_0 = 1$ and $\sigma = 2$) reads $k_\alpha = 5.64$. The choice of α is based on the probability of a type I error one is willing to accept. For instance, if one chooses α as $\alpha = 0.01$, one is willing to accept that 1 out of a 100 experiments leads to rejection of H_0 when in fact H_0 is true.

Let us now consider the size of a type II error, β. Figure 3.6 shows for the present example the size of a type II error, β. It corresponds to the area under the graph of the distribution of $\underline{y}$ under H_A for the interval complementary to

Table 3.2 Test size α, critical value k_α for $x_0 = 1$ and $\sigma = 2$.

α	$\frac{k_\alpha - x_0}{\sigma}$	k_α
0.1	1.28	3.56
0.05	1.65	4.30
0.01	2.32	5.64
0.001	2.98	6.96

the critical region K. The size of a type II error, β, can be computed once the

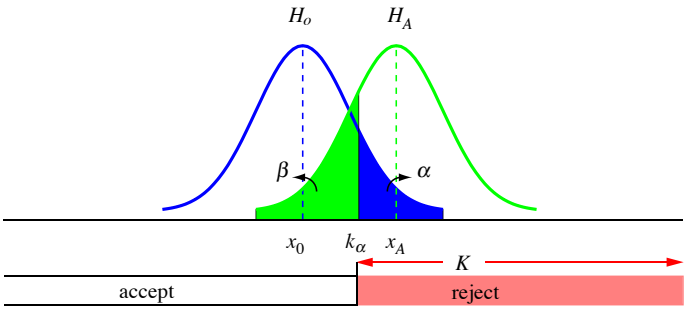


Figure 3.6 The sizes of type I and type II error, α and β, for testing $H_0 : E\{\underline{y}\} = x_0$ versus $H_A : E\{\underline{y}\} = x_A > x_0$.

probability density function of $\underline{y}$ under H_A is known and the critical region K is known. Thus, for the present example the size of the type II error can be computed as

$$\beta = \int_{-\infty}^{k_\alpha} p_{\underline{y}}(y|x_A)dy$$

or, since

$$p_{\underline{y}}(y|x_A) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_A)^2]$$

as

$$\beta = \int_{-\infty}^{k_\alpha} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_A)^2]dy \tag{3.23}$$

Also this value can be computed with the help of the table of the *standard* normal distribution. But first some transformations are needed. It will be clear that the probability that a sample or observation of $\underline{y}$ falls in the critical

region K when H_A is true, is identical to 1 minus the probability that the sample does not fall in the critical region when H_A is true. Thus,

$$\beta = P(\underline{y} \notin K|H_A) = 1 - P(\underline{y} \in K|H_A) \tag{3.24}$$

Since for the present example,

$$P(\underline{y} \in K|H_A) = \int_{k_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_A)^2]dy$$

substitution into (3.24) gives

$$1 - \beta = \int_{k_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_A)^2]dy \tag{3.25}$$

This formula has the same structure as (3.19). The value $1-\beta$ can therefore be computed in exactly the same manner as the size of the test, α, was computed. And from $1-\beta$ it is trivial to compute β, the size of the type II error. □

3.2.4 General steps in testing hypotheses

Thus far we have discussed the basic concepts underlying most of the hypothesis-testing problems. The same concept and guidelines will provide the basis for solving more complicated hypothesis-testing problems as treated in the next sections. Here we summarize the main steps of testing hypotheses about a general probability model.

a From the nature of the experimental data and the consideration of the assertions that are to be examined, identify the appropriate null hypothesis and alternative hypothesis:

$$H_0 : \underline{y} \sim p_{\underline{y}}(y|x_0) \text{ versus } H_0 : \underline{y} \sim p_{\underline{y}}(y|x_A)$$

In most geodetic applications, the probability density function can be considered to be normal under the null hypothesis as well as under the alternative hypothesis. Since the normal distribution is uniquely characterized by its first two (central) moments, the expectation (mean) and dispersion (variance), the two hypotheses can then only differ with respect to the mean and/or variance of $\underline{y}$. In these lecture notes, we will only consider differences in the mean of $\underline{y}$.

b Determine the critical region K. In practice this implies that one determines a function of $\underline{y}$, the test statistic, of which the values should always be larger than a chosen critical value. The values of $\underline{y}$ which make this happen, form the critical region.

c Specify the size of the type I error, α, that one wishes to assign to the testing process. Use tables to determine the location of the critical region K from

$$\alpha = P(\underline{y} \in K|H_0) = \int_K p_{\underline{y}}(y|x_0)dy$$

For many distributions, like the normal distribution or the Chi-squared distribution, these tables can be found in standard textbooks on statistics.

d Compute the size of the type II error, β

$$\beta = P(\underline{y} \notin K | H_A) = 1 - \int_K p_{\underline{y}}(y|x_A)dy$$

to ensure that there exists a reasonable protection against type II errors. Its complement, $\gamma = 1 - \beta$, is known as the detection probability or as the power of the test.

e After the test has been explicitly formulated, determine whether the sample or observation y of $\underline{y}$ falls in the critical region K or not. Reject H_0 if $y \in K$, and accept H_0 if $y \notin K$. Never claim however that the hypotheses have been *proved* false or true by the testing.

3.3 Test statistics for the linear(ized) model

3.3.1 The null and alternative hypothesis

Before any start can be made with the application of the theory of statistical testing, one needs to have a clear idea of the null hypothesis H_o and of the alternative hypothesis H_a. In the previous chapter on adjustment theory, we have been working with the model of observation equations

$$E\{\underline{y}\} = Ax \ , \ D\{\underline{y}\} = Q_y$$

We have seen, that in order to be able to apply the least-squares principle, only the first two moments of the random vector of observables ,$\underline{y}$, need to be specified. The first moment (mean) $E\{\underline{y}\} = Ax$ and the second moment (variance matrix) $D\{\underline{y}\} = Q_y$. In the case of statistical testing however, this is not sufficient. In addition, one will have to specify the type of probability function of $\underline{y}$ as well. Since most observational data in geodesy can be modelled as samples drawn from a *normal distribution*, we will assume that $\underline{y}$ has the normal probability density function

$$p_{\underline{y}}(y \mid x) = (2\pi)^{-\frac{m}{2}} \mid Q_y \mid^{-\frac{1}{2}} \exp[-\frac{1}{2}(y - Ax)^T Q_y^{-1}(y - Ax)]$$

Thus we assume that $\underline{y}$ is Normally distributed, with mean Ax and with the variance matrix Q_y. In shorthand notation

$$H_o : \quad \underline{y} \sim N(Ax, Q_y) \tag{3.26}$$

This will be our null hypothesis.

In order to test the null hypothesis against an alternative hypothesis, we need to know what type of misspecifications one can expect in the null hypothesis. Of course, every part of the null hypothesis could have been specified incorrectly. The mean Ax can be wrong, the variance matrix Q_y can be wrong and/or $\underline{y}$ could have a distribution other than the

normal distribution. In these lecture notes we will restrict ourselves to misspecifications in the mean. Experience has shown, that these are by large the most common errors that occur when formulating the model. As alternative hypothesis , we will therefore use

$$H_a: \quad \underline{y} \sim N(Ax + C\nabla, Q_y) \tag{3.27}$$

where C is a known matrix of order $m \times q$ and ∇ is an unknown vector of order $q \times 1$. With H_a we thus oppose the null hypothesis H_o to a more relaxed alternative hypothesis, in which more explanatory parameters, namely ∇, are introduced. These additional parameters are then supposed to model those effects which were assumed absent in H_o. For instance, through $C\nabla$ one may model the presence of one or more blunders in the data, or the the presence of refraction, or any other systematic effect which was not taken into account in H_o. The relation between $E\{\underline{y}\}$ and ∇ is specified through the matrix C.

The purpose of testing H_o against H_a is now, to infer whether the data supports H_o or whether the data rejects H_o on the basis of H_a.

3.3.2 Residuals and model errors

It will be intuitively clear that the data and thus the observable $\underline{y}$ must be instrumental for the testing of H_o against H_a. In this section we will discuss which function of $\underline{y}$ is a likely candidate to be used in the testing of H_o versus H_a.

The least-squares residual: If we write

$$\underline{y} = Ax + \underline{e} \tag{3.28}$$

then

$$H_o: \; E\{\underline{e}\} = 0 \quad , \quad H_a: \; E\{\underline{e}\} = C\nabla \neq 0 \tag{3.29}$$

Thus the mean of the residual vector $\underline{e}$ will be zero when H_o is true and unequal to zero when H_o is false. This shows that if we would have a sample or measurement of $\underline{e}$ available, we could use it to decide on the validity of H_o. Would the sample be close to zero, we would be inclined to accept H_o and would it differ greatly from zero, we would be inclined to reject H_o. Unfortunately, no sample of $\underline{e} = \underline{y} - Ax$ is available, since x is unknown. Instead of considering $\underline{e}$, let us therefore consider its estimator. The least-squares solution of x and $\underline{e}$ under H_o, reads

$$\begin{aligned} \underline{\hat{x}} &= (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} \\ \underline{\hat{e}} &= [I_m - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}] \underline{y} \end{aligned}$$

With (3.28), this can also be written as

$$\begin{aligned} \underline{\hat{x}} &= x + (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{e} \\ \underline{\hat{e}} &= [I_m - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}] \underline{e} \end{aligned}$$

This shows, when we take the expectation and use (3.29), that

$$H_o: \begin{cases} E\{\underline{\hat{x}}\} = x \\ E\{\underline{\hat{e}}\} = 0 \end{cases} , \; H_a: \begin{cases} E\{\underline{\hat{x}}\} \neq x \\ E\{\underline{\hat{e}}\} = C\nabla \neq 0 \end{cases} \tag{3.30}$$

Thus apart from the mean of $\underline{e}$, also the mean of $\underline{\hat{e}}$ is zero when H_o is true and nonzero when H_o is false. Note that in the first case, $\underline{\hat{x}}$ is an unbiased estimator of x, while it becomes biased when H_o is false. Contrary to $\underline{e}$, we do have a sample available of $\underline{\hat{e}}$, since it is a function of $\underline{y}$. Hence, instead of using $\underline{e}$, we could use $\underline{\hat{e}}$ to decide on the validity of H_o. If the sample of $\underline{\hat{e}}$ is close to zero, we are inclined to accept H_o and if it differs greatly from zero, we are inclined to reject H_o.

The least-squares model error: Instead of using the least-squares residual vector $\underline{\hat{e}}$, it also seems intuitively clear that the model error ∇ itself must be instrumental in deciding on the validity of H_o. Under H_a, we have

$$H_a: \ E\{\underline{y}\} = Ax + C\nabla \tag{3.31}$$

The model error ∇ itself is unknown of course. Let us therefore consider its least-squares estimator $\underline{\hat{\nabla}}$. Since it is a function of $\underline{y}$, we do have a sample value of it available. On the basis of this value we could also decide on the validity of H_o. If the sample value of $\underline{\hat{\nabla}}$ is small, we are inclined to belief that the model error is absent and thus that H_o is true. On the other hand, if the sample value of $\underline{\hat{\nabla}}$ turns out be significant, we will certainly not be inclined to belief H_o and rather have more faith in H_a.

From the above discussion, it seems that both $\underline{\hat{e}}$ and $\underline{\hat{\nabla}}$ can be used for the testing of H_o against H_a. One can therefore expect that the two estimators must be related in some way. And this is indeed the case. To show this, we first solve for ∇. The normal equations that belong to H_a of (3.31), read

$$\begin{bmatrix} A^T Q_y^{-1} A & A^T Q_y^{-1} C \\ C^T Q_y^{-1} A & C^T Q_y^{-1} C \end{bmatrix} \begin{bmatrix} \underline{\hat{x}}_a \\ \underline{\hat{\nabla}} \end{bmatrix} = \begin{bmatrix} A^T Q_y^{-1} \underline{y} \\ C^T Q_y^{-1} \underline{y} \end{bmatrix} \tag{3.32}$$

where $\underline{\hat{x}}_a$ is given the subindex $_a$ to indicate that it is the least-squares estimator of x under H_a and not the least-squares estimator of x under H_o. From the above normal equations, the reduced normal equations for $\underline{\hat{\nabla}}$ follow as

$$C^T Q_y^{-1} [I_m - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}] C \, \underline{\hat{\nabla}} = C^T Q_y^{-1} [I_m - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}] \underline{y}$$

In this expression, we recognize

$$\begin{array}{rcl} \underline{\hat{e}} & = & [I_m - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}] \underline{y} \\ Q_{\hat{e}} Q_y^{-1} & = & [I_m - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}] \end{array}$$

The least-squares estimator of the model error and its variance matrix follows therefore as

$$\underline{\hat{\nabla}} = [C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C]^{-1} C^T Q_y^{-1} \underline{\hat{e}} \ , \ Q_{\hat{\nabla}} = [C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C]^{-1} \tag{3.33}$$

Thus $\underline{\hat{\nabla}}$ is indeed a function of $\underline{\hat{e}}$. Under H_o and H_a, we have

$$H_o: \begin{cases} E\{\underline{\hat{\nabla}}\} = 0, & D\{\underline{\hat{\nabla}}\} = Q_{\hat{\nabla}} \\ E\{\underline{\hat{e}}\} = 0, & D\{\underline{\hat{e}}\} = Q_{\hat{e}} \end{cases} , H_a: \begin{cases} E\{\underline{\hat{\nabla}}\} = & \nabla \neq 0, & D\{\underline{\hat{\nabla}}\} = Q_{\hat{\nabla}} \\ E\{\underline{\hat{e}}\} = & C\nabla \neq 0, & D\{\underline{\hat{e}}\} = Q_{\hat{e}} \end{cases} \tag{3.34}$$

3.3.3 Significance of model error

Up to this point we spoke about $\hat{\nabla}$ being small or large, with reference to the acceptance or rejection of H_o. But when do we consider $\hat{\nabla}$ small or large? Here we need an objective way to measure the significance of $\hat{\nabla}$.

The one dimensional case: Let us first consider the one dimensional case. In that case, $q = 1$ and $\hat{\nabla}$ becomes a scalar instead of a vector. The significance of $\underline{\hat{\nabla}}$ can now be measured by using the precision of the estimator, thus by using its standard deviation $\sigma_{\hat{\nabla}}$. We therefore divide $\underline{\hat{\nabla}}$ by its standard deviation $\sigma_{\hat{\nabla}}$ and define the random variable

$$\underline{w} = \frac{\underline{\hat{\nabla}}}{\sigma_{\hat{\nabla}}} \tag{3.35}$$

This random variable has a standard normal distribution under H_o. Thus under H_o, it has a zero mean, with a variance of one. Under H_a however, it will have a nonzero mean, but again with a variance of one. Thus

$$H_o : \underline{w} \sim N(0,1) \quad \text{and} \quad H_a : \underline{w} \sim N(\frac{\nabla}{\sigma_{\hat{\nabla}}}, 1) \tag{3.36}$$

Since the distribution of $\underline{w}$ is completely known under H_o, we are now in a position to test the significance of the model error. The model error is said to be significant, if

$$\mid w \mid > N_{\frac{1}{2}\alpha}(0,1) \tag{3.37}$$

where $N_{\frac{1}{2}\alpha}(0,1)$ is the critical value of the standard normal distribution, based on the level of significance α (the test is two-sided). Note that instead of working with the absolute value of $\underline{w}$, one can also work with its square. In that case, the test reads

$$w^2 > \chi^2_\alpha(1,0) \tag{3.38}$$

where $\chi^2_\alpha(1,0)$ is the critical value of the central Chi- squared distribution having one degree of freedom.

The higher dimensional case: One can not use the above test when $q > 1$. In that case we need to take the complete variance matrix $Q_{\hat{\nabla}}$ into account, to measure the significance of $\hat{\nabla}$. Instead of using the one dimensional test statistic $\underline{w}^2$, we therefore use the q-dimensional test statistic

$$\underline{T}_q = \underline{\hat{\nabla}}^T Q^{-1}_{\hat{\nabla}} \underline{\hat{\nabla}} \tag{3.39}$$

Note that $\underline{T}_q = \underline{w}^2$, when $q = 1$. In the higher dimensional case, the model error is said to be significant, if

$$T_q > \chi^2_\alpha(q,0) \tag{3.40}$$

where $\chi^2_\alpha(q,0)$ is the critical value of the central Chi- square distribution, having q degrees of freedom.

Using the least-squares residual: We have seen that the least- squares estimator of the model error, $\underline{\hat{V}}$, can be written as a function of the least-squares residual vector $\underline{\hat{e}}$. This implies that the above test statistic $\underline{T}_q$ can also be expressed in terms of $\underline{\hat{e}}$. If we substitute (3.33) into (3.39), we get

$$\underline{T}_q = \underline{\hat{e}}^T Q_y^{-1} C [C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C]^{-1} C^T Q_y^{-1} \underline{\hat{e}} \tag{3.41}$$

Although the two test statistics (3.39) and (3.41) are identical, the second expression is often the more practical one, since the least-squares residual vector $\underline{\hat{e}}$ is often already available from the adjustment based on H_o.

3.4 Detection, identification and adaptation

In the previous section we gave the test statistic for testing the null hypothesis H_o against a particular alternative hypothesis H_a. In most practical applications however, it is usually not only one model error one is concerned about, but quite often many more than one. This implies that one needs a *testing procedure* for handling the various alternative hypotheses. In this subsection we will discuss a way of structuring such a testing procedure. It will consist of the following three steps: detection, identification and adaptation.

3.4.1 Detection

Since one usually first wants to know whether one can have any confidence in the assumed null hypothesis without the need to specify any particular alternative hypothesis, the first step consists of a check on the *overall* validity of H_o. This implies that one opposes the null hypothesis to the most relaxed alternative hypothesis possible. The most relaxed alternative hypothesis is the one that leaves the observables completely free. Hence, under this alternative hypothesis no restrictions at all are imposed on the observables. We therefore have the situation

$$H_o: E\{\underline{y}\} = Ax \quad \text{versus} \quad H_a: E\{\underline{y}\} \in R^m \tag{3.42}$$

Since $E\{\underline{y}\} \in R^m$ implies that matrix (A,C) is square and invertible, it follows that matrix C has $q = m - n$ columns and that its range space is complementary to the range space of A. Thus $R^m = R(A) \oplus R(C)$. It can be shown that in this case, the test statistic of (3.41) simplifies to

$$\underline{T}_{m-n} = \underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}} \tag{3.43}$$

The appropriate test statistic for testing the null hypothesis against the most relaxed alternative hypthesis, is thus equal to the weighted sum-of-squares of the least-squares residuals. The null hypothesis will then be rejected when

$$T_{m-n} > \chi^2_\alpha(m-n, 0) \tag{3.44}$$

The $\hat{\sigma}^2$ test: In the literature one often sees the above *overall model test* also formulated in a slightly different way. Let us use the factorization $Q_y = \sigma^2 G_y$, where σ^2 is the

variance factor of unit weight and where G_y is the corresponding cofactor matrix. It can be shown that

$$\underline{\hat{\sigma}}^2 = \frac{\underline{\hat{e}}^T G_y^{-1} \underline{\hat{e}}}{m-n}$$

is an *unbiased* estimator of σ^2. Thus $E\{\underline{\hat{\sigma}}^2\} = \sigma^2$. The test (3.44) can now also be formulated as

$$\frac{\hat{\sigma}^2}{\sigma^2} > \frac{\chi^2_\alpha(m-n,0)}{m-n} = F_\alpha(m-n,\infty,0)$$

where $F(m-n,\infty,0)$ is the central F-distribution having $m-n$ and ∞ degrees of freedom.

3.4.2 Identification

In the detection phase, one tests the overall validity of the null hypothesis. If this leads to a rejection of the null hypothesis, one has to search for possible model misspecifications. That is, one will then have try to identify the model error which caused the rejection of the null hypothesis. This implies that one will have to specify through the matrix C, the type of likely model errors. This specification of possible alternative hypotheses is application dependent and is one of the more difficult tasks in hypothesis testing. It depends namely very much on ones experience, which type of model errors one considers to be likely.

The 1-dimensional case: In case the model error can be represented by a scalar, $q = 1$ and matrix C reduces to a vector which will be denoted by the lowercase character c. This implies that the alternative hypothesis takes the form

$$H_a : E\{\underline{y}\} = Ax + c\nabla \tag{3.45}$$

The alternative hypothesis is specified, once the vector c is specified. The appropriate test statistic for testing the null hypothesis against the above alternative hypothesis H_a follows when the vector c is substituted for the C-matrix in (3.41). It gives

$$\underline{T}_{q=1} = \underline{w}^2 = \frac{[c^T Q_y^{-1} \underline{\hat{e}}]^2}{c^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c}$$

or when the square-root is taken

$$\underline{w} = \frac{c^T Q_y^{-1} \underline{\hat{e}}}{\sqrt{c^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c}} \tag{3.46}$$

This test statistic has a standard normal distribution $N(0,1)$ under H_o. The evidence on whether the model error as specified by (3.45) did or did not occur, is based on the test

$$|\, w \,| > N_{\frac{1}{2}\alpha_1}(0,1) \tag{3.47}$$

Data snooping: Apart from the possibility of having a one dimensional test as (3.47), it is standard practice in geodesy to always first check the individual observations for potential *blunders*. This implies that the alternative hypotheses take the form

$$H_{a_i} : E\{\underline{y}\} = Ax + c_i\nabla \quad i = 1,\ldots,m \tag{3.48}$$

with

$$c_i = (0,\ldots,0,1,0,\ldots,0)^T$$

Thus c_i is a unit vector having the 1 as its ith entry. The additional term $c_i\nabla$ models the presence of a blunder in the ith observation. The appropriate test statistic for testing the null hypothesis against the above alternative hypothesis H_{a_i} is again of the general form of (3.46), but now with the c-vector chosen as c_i,

$$\underline{w}_i = \frac{c_i^T Q_y^{-1}\underline{\hat{e}}}{\sqrt{c_i^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_i}} \tag{3.49}$$

This test statistic has of course also a standard normal distribution $N(0,1)$ under H_o. By letting i run from 1 up to and including m, one can screen the whole data set on the presence of potential blunders in the individual observations. The test statistic $\underline{w}_i$ which returns the in absolute value largest value, then pinpoints the observation which is most likely corrupted with a blunder. Its significance is measured by comparing the value of the test statistic with the critical value. Thus the jth observation is suspected to have a blunder, when

$$|w_j| \geq |w_i| \quad \forall i \text{ and } |w_j| > N_{\frac{1}{2}\alpha_1}(0,1) \tag{3.50}$$

This procedure of screening each individual observation for the presence of a blunder, is known as *data snooping*.

In many applications in practice, the variance matrix Q_y is diagonal. If that is the case, the expression of the above test statistic simplifies considerably. With a diagonal Q_y-matrix, we have

$$\underline{w}_i = \frac{\underline{\hat{e}}_i}{\sigma_{\hat{e}_i}}$$

The appropriate test statistic is then thus equal to the least-squares residual of the ith observation divided by the standard deviation of the residual.

The higher dimensional case: It may happen that a particular model error can not be represented by a single scalar. In that case $q > 1$ and ∇ becomes a vector. The appropriate test statistic is then the one we met earlier, namely

$$\underline{T}_q = \underline{\hat{e}}^T Q_y^{-1} C[C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C]^{-1} C^T Q_y^{-1}\underline{\hat{e}} \tag{3.51}$$

It is through the matrix C that one specifies the type of model error.

3.4.3 Adaptation

Once one or more likely model errors have been identified, a corrective action needs to be undertaken in order to get the null hypothesis accepted. Here, one of the two following approaches can be used in principle. Either one replaces the data or part of the data with new data such that the null hypothesis does get accepted, or, one replaces the original

null hypothesis with a new hypothesis that does take the identified model errors into account. The first approach amounts to a remeasurement of (part of) the data. This approach is feasible for instance, when in case of datasnooping some individual observations are identified as being potentially corrupted by blunders. These are then the observations which get remeasured. In the second approach no remeasurement is undertaken. Instead the model of the null hypothesis is enlarged by adding additional parameters such that all identified model errors are taken care off. Thus with this approach, the identified alternative hypothesis becomes the new null hypothesis.

Once the adaptation step is completed, one of course still has to make sure whether the newly created situation is acceptable or not. This at least implies a repetition of the detection step. When adaptation is applied, one also has to be aware of the fact that since the model may have changed, also the 'strength of the model' may have changed. In fact, when the model is adapted through the addition of more explanatory parameters, the model has become weaker in the sense that the test statistics will now have less detection and identification power. That is, the reliability has become poorer. It depends on the particular application at hand, whether this is considered acceptable or not.

3.5 Reliability

In the previous section we considered a testing procedure for the detection, identification and adaptation of model errors. Hence, we now know how to search for potential model errors and how to test their significance. But what we do not know yet, is how well these tests will perform. In particular we would like to know how the tests perform in terms of their power of detecting the model errors.

3.5.1 Power of tests

When applying the various tests, one should be aware of the fact that the outcomes of the tests may be erroneous as well. This is due to the fact that the test statistics which are used for testing, are random variables. They are random variables, which have a probability density function that depends on which of the hypotheses is true. For our test statistic $\underline{T}_q$ we have

$$H_o : \underline{T}_q \sim \chi^2(q,0) \quad \text{and} \quad H_a : \underline{T}_q \sim \chi^2(q,\lambda) \tag{3.52}$$

with the *noncentrality parameter*

$$\begin{aligned} \lambda &= \nabla^T C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C \nabla \\ &= \nabla^T C^T Q_y^{-1} [Q_y - A(A^T Q_y^{-1} A)^{-1} A^T] Q_y^{-1} C \nabla \end{aligned} \tag{3.53}$$

Thus $\underline{T}_q$ has a central Chi-square distribution with q degrees of freedom, when H_o is true, but a noncentral Chi-square distribution with q degrees of freedom and a noncentrality parameter λ, when the alternative hypothesis H_a is true.

Recall from one of the earlier sections that, when testing, two type of errors can be made. A type I error is made, when one decides to reject the null hypothesis, while in fact it is true. The probability of such an error is given by the level of significance α and reads

$$\alpha = P[\underline{T}_q > \chi^2_\alpha(q,0) \mid H_o] = \int_{\chi^2_\alpha(q,0)}^{\infty} p_{\chi^2}(x \mid q,0)dx \tag{3.54}$$

where $p_{\chi^2}(x \mid q,0)$ is the probability density function of the central Chi-square distribution, having q degrees of freedom.

A type II error is made, when one decides to accept the null hypothesis, while in fact it is false. The probability of such an error is given as $\beta = P[\underline{T}_q \leq \chi^2_\alpha(q,0) \mid H_a]$. Instead of working with β, one can also work with its complement $\gamma = 1-\beta$, which is known as the *power of the test*. It is the probability of correctly rejecting the null hypothesis. Hence, it is the probability

$$\gamma = P[\underline{T}_q > \chi^2_\alpha(q,0) \mid H_a] = \int_{\chi^2_\alpha(q,0)}^{\infty} p_{\chi^2}(x \mid q,\lambda)dx \tag{3.55}$$

Note that the power γ depends on the three parameters α, q and λ. Using a shorthand notation, we write

$$\gamma = \gamma(\alpha,q,\lambda) \tag{3.56}$$

When testing, we of course would like to have a sufficiently high probability of correctly detecting a model error when it occurs. One can make γ larger by increasing α, or, by decreasing q, or, by increasing λ. This can be seen as follows. A larger α implies a smaller critical value $\chi^2_\alpha(q,0)$ and via the integral (3.55) thus a larger power γ. Thus if we want a smaller probability for the error of the first kind (α smaller), this will go at the cost of a smaller γ as well. That is, one can not simultaneously decrease α and increase γ.

The power γ also gets larger, when q gets smaller. This is also understandable. When q gets smaller, the less additional parameters are used in H_a and therefore the more "information" is used in formulating H_a. For such an alternative hypothesis, one would expect that if it is true, the probability of accepting it will be higher than for an alternative hypothesis that contains more additional parameters. Finally, the power γ also gets larger, when λ gets larger. This is understandable when one considers (3.53). For instance, one would expect to have a higher probability of correctly rejecting the null hypothesis, when the model error gets larger. And when ∇ gets larger, also the noncentrality parameter λ gets larger.

Using α and/or q as tuning parameters to increase γ, does not make much sense however. The parameter q can not be changed at will, since it depends on the type of model error one is considering. And increasing α, also does not make sense, since it would lead to an increased probability of an error of the first kind. Hence, this leaves us with λ. According to (3.53), the noncentrality parameter depends on

1. the model error $C\nabla$
2. the variance matrix Q_y
3. The design matrix A

Since one also can not change the model error at will, it is through changes in the variance matrix Q_y and/or in the design matrix A that one can increase λ, thereby improving the detection power of the test. Recall that it is also through these two matrices that one can improve the *precision* of the least-squares solution. Thus the detection power γ of the tests

can, if needed, be improved by using more precise measurement equipment, by adding more obervations to the model and/or by changing the structure of the design matrix A.

3.5.2 Internal reliability

The power γ is the probability with which a model error $C\nabla$ can be found with the appropiate test. Thus given the model error $C\nabla$, we can compute the noncentrality parameter as

$$\lambda = \nabla^T C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C\nabla \tag{3.57}$$

and from it, together with α and q, the power as

$$\gamma = \gamma(\alpha, q, \lambda) \tag{3.58}$$

We can however also follow the inverse route. That is, given the power γ, the level of significance α and the dimension q, the noncentrality parameter can be computed as

$$\lambda = \lambda(\alpha, q, \gamma) \tag{3.59}$$

This function is the inverse of (3.58). If we combine (3.57) with (3.59), we get a *quadratic equation* in ∇

$$\lambda(\alpha, q, \gamma) = \nabla^T C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C\nabla \tag{3.60}$$

This equation is said to describe the *internal reliability* of the null hypothesis $E\{\underline{y}\} = Ax$ with respect to the alternative hypothesis $E\{\underline{y}\} = Ax + C\nabla$. Each model error $C\nabla$ that satisfies the quadratic equation, can be found with a probability γ. Since it is often more practical to know the size of the model error that can be found with a certain probability, than knowing the probability with which a certain model error can be found, we will use the inverse of (3.60).

Minimal Detectable Biases (MDB's): In the one dimensional case $q = 1$ and the matrix C reduces to the vector c. Inverting (3.60) becomes then rather straightforward. As a result we get for the size of the model error

$$|\nabla| = \sqrt{\frac{\lambda(\alpha_1, 1, \gamma)}{c^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c}} \tag{3.61}$$

The variate $|\nabla|$ is known as the *Minimal Detectable Bias (MDB)*. It is the size of the model error that can just be detected with a probability γ, using the appropriate $\underline{w}$-test statistic. Larger errors will be detected with a larger probability and smaller errors with a smaller probability.

In order to guarantee that the model error $c\nabla$ is detected with the same probability by both the w-test and the overall model test, one will have to relate the critical values of the two tests. This can be done by equalizing both the powers of the two tests and their noncentrality parameters. Thus if $\lambda(\alpha_{m-n}, m-n, \gamma_{m-n})$ is the noncentrality parameter of the $\underline{T}_{m-n}$-test statistic and $\lambda(\alpha_1, 1, \gamma_1)$ the noncentrality parameter of the $\underline{T}_1$-test statistic, we have

$$\lambda(\alpha_{m-n}, m-n, \gamma) = \lambda(\alpha_1, 1, \gamma) \tag{3.62}$$

From this relation, α_{m-n} and thus the critical value of the overall model test, can be computed, once the power γ and the level of significance α_1 is chosen. Common values in case of geodetic networks, are $\alpha_1 = 0.001$ and $\gamma = 0.80$.

Data snooping: Note that the MDB can be computed for each alternative hypothesis, once the vector c of that particular alternative hypothesis has been specified. In case of data snooping, the MDB's of the individual observations are computed as

$$| \nabla_i | = \sqrt{\frac{\lambda(\alpha_1, 1, \gamma)}{c_i^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_i}} \qquad i = 1, \ldots, m \tag{3.63}$$

As with the w_i-test statistic, also this expression simplifies considerably when the variance matrix Q_y is diagonal. In that case we have

$$| \nabla_i | = \sigma_{y_i} \sqrt{\frac{\lambda(\alpha_1, 1, \gamma)}{1 - \sigma_{\hat{y}_i}^2 / \sigma_{y_i}^2}} \tag{3.64}$$

where $\sigma_{y_i}^2$ is the *a priori* variance of the ith observation and $\sigma_{\hat{y}_i}^2$ is the *a posteriori* variance of this observation. We thus clearly see that a better precision of the observation as well as a larger amount in which its precision gets improved by the adjustment, will improve the internal reliability, that is, will result in a smaller MDB.

The higher dimensional case: When $q > 1$, the inversion is a bit more involved. In this case ∇ is a vector, which implies that its direction needs to be taken into account as well. We use the factorization $\nabla = \| \nabla \| d$, where d is a unit vector ($d^T d = 1$). If we substitute the factorization into (3.60) and then invert the result, we get

$$\| \nabla \| = \sqrt{\frac{\lambda(\alpha_q, q, \gamma)}{d^T C^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C d}} \qquad (d = \text{unit vector}) \tag{3.65}$$

The size of the model error now depends on the chosen direction vector d. But by letting d vary over the unit sphere in R^q, one can obtain the whole range of MDB's that can be detected with a probability γ.

3.5.3 External reliability

With the MDB's one describes the internal reliability. In practical applications however, one often not only wants to now the size of the model errors that can be detected with a certain probability, but also what their impact would be on the estimated parameters. After all, it are the estimated parameters, or functions thereof, which constitute the final result of an adjustment. The *external reliability* describes the influence of model errors of the size of the MDB's, on the final result of the adjustment.

For the expectation of the least-squares estimator $\underline{\hat{x}}$ under the alternative hypothesis H_a, we have

$$E\{\underline{\hat{x}} \mid H_a\} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} E\{\underline{y} \mid H_a\}$$

Substitution of $E\{\underline{y} \mid H_a\} = Ax + C\nabla$, gives

$$E\{\underline{\hat{x}} \mid H_a\} = x + (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C\nabla$$

This shows that the *bias* in $\underline{\hat{x}}$, due to the presence of the model error $C\nabla$, is

$$\nabla\hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C\nabla \tag{3.66}$$

Of course we do not know the actual model error $C\nabla$. But we do know the size of the model error that can be detected with a probability of γ. Hence, if we use the MDB and replace $C\nabla$ by $c \mid \nabla \mid$ in case $q = 1$ or by $Cd \parallel \nabla \parallel$ in case $q > 1$, the vector $\nabla\hat{x}$ will show us by how much the least-squares estimator $\underline{\hat{x}}$ gets biased, when a model error of the size of the MDB would occur. Note that there are as many MDB's as there are alternative hypotheses. Hence, there are also as many vectors $\nabla\hat{x}$. This whole set is said to describe the external reliability.

In certain applications, one may not be interested in the whole parameter vector x, but only in particular functions of it, say $\theta = F^T x$. In that case the external reliability is described by

$$\nabla\hat{\theta} = F^T (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C\nabla \tag{3.67}$$

A particular case occurs, when one is only interested in the x_1 - part of the parameter vector $x = (x_1^T, x_2^T)^T$. In that case $F^T = (I, 0)$. The bias-vector $\nabla\hat{x}_1$ can then be shown to equal

$$\nabla\hat{x}_1 = (\bar{A}_1^T Q_y^{-1} \bar{A}_1)^{-1} \bar{A}_1^T Q_y^{-1} C\nabla \tag{3.68}$$

where, with the partitioning $A = (A_1^T, A_2^T)^T$, the matrix $\bar{A}_1^T Q_y^{-1} \bar{A}_1$ is the *reduced* normal matrix and $\bar{A}_1 = [I - A_2(A_2^T Q_y^{-1} A_2)^{-1} A_2^T Q_y^{-1}] A_1$.

Bias-to-Noise Ratios (BNR's): In order to measure the significance of the external reliability, one can compare the bias-vectors $\nabla\hat{x}$ with the precision, or variance matrix $Q_{\hat{x}}$, of $\underline{\hat{x}}$. This can be done by using the *Bias-to-Noise Ratios (BNR)*

$$\begin{array}{lcll} \lambda_{\hat{x}} & = & \nabla\hat{x}^T Q_{\hat{x}}^{-1} \nabla\hat{x} & \text{with } Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1} \\ \lambda_{\hat{\theta}} & = & \nabla\hat{\theta}^T Q_{\hat{\theta}}^{-1} \nabla\hat{\theta} & \text{with } Q_{\hat{\theta}} = F^T Q_{\hat{x}} F \\ \lambda_{\hat{x}_1} & = & \nabla\hat{x}_1^T Q_{\hat{x}_1}^{-1} \nabla\hat{x}_1 & \text{with } Q_{\hat{x}_1} = (\bar{A}_1^T Q_y^{-1} \bar{A}_1)^{-1} \end{array} \tag{3.69}$$

Note that the BNR's are dimensionless and that they measure the squared lengths of the bias-vectors in the metric defined by the appropriate variance matrix. Also note that the BNR's are scalars. Thus $\lambda_{\hat{x}}$ is a scalar, whereas $\nabla\hat{x}$ is a vector. This implies that with the BNR's, one only needs to evaluate a scalar per alternative hypothesis, whereas otherwise a complete vector would have to be evaluated for each alternative hypothesis.

From a computational point of view, there are also some shortcuts that can be used when computing the BNR's. For instance, when the complete vector $\nabla\hat{x}$ is considered, it can be shown that

$$\lambda_{\hat{x}} = \nabla\hat{x}^T Q_{\hat{x}}^{-1} \nabla\hat{x} = \nabla^T C^T Q_y^{-1} C\nabla - \lambda(\alpha_q, q, \gamma) \tag{3.70}$$

The second expression on the right hand side may sometimes be easier to compute than the first expression on the right hand side, in particular when Q_y is diagonal. For the BNR of a subset of the parameter vector, one can show that

$$\lambda_{\hat{x}_1} = \nabla\hat{x}_1^T Q_{\hat{x}_1}^{-1}\nabla\hat{x}_1 = \lambda_{\hat{x}} - \nabla^T C^T Q_y^{-1} A_2 (A_2^T Q_y^{-1} A_2)^{-1} A_2^T Q_y^{-1} C\nabla \tag{3.71}$$

An important feature of the BNR's is that they can be used to formulate *upperbounds* on the external reliability of functions of the parameters. For instance, if we consider the function $\hat{\theta} = f^T\hat{x}$ having the *variance* $\sigma_{\hat{\theta}}^2$ and the *bias* $\nabla\hat{\theta} = f^T\nabla\hat{x}$, then

$$\nabla\hat{\theta} \le \sigma_{\hat{\theta}}\sqrt{\lambda_{\hat{x}}} \tag{3.72}$$

This shows, that the potential bias in $\hat{\theta}$ due to an undetected model error of the size of the MDB, will never be larger than the standard deviation times the square root of the BNR.

3.5.4 Connection of two height systems: an example

Let us assume that we need to connect two levelling networks of which the heights are defined in two different height-systems (height-datums). The height of point i in the first system is denoted as h_i and the height of the same point in the second system as H_i. As observables we have available the heights $\underline{h}_i$ and $\underline{H}_i$ of the same n points in both height-systems, $i = 1,\ldots,n$.

We assume that all heights are uncorrelated, that all heights of the first system have the same variance σ_h^2 and that all heights of the second system have the same variance σ_H^2 (*note*: these assumptions are of course not really realistic, but simply serve to make our example analytically tractable). If we assume that the two height systems differ in scale and have a constant offset, then the two different heights of point i are related as

$$h_i = \lambda H_i + t$$

with the scale factor λ and the translation t.

Based on the above assumptions, the model of observation equations for connecting the two networks, becomes

$$E\{\begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_n \\ \underline{H}_1 \\ \vdots \\ \underline{H}_n \end{bmatrix}\} = \begin{bmatrix} \lambda H_1 + t \\ \vdots \\ \lambda H_n + t \\ H_1 \\ \vdots \\ H_n \end{bmatrix} \quad , \quad \begin{bmatrix} Q_h & Q_{hH} \\ Q_{Hh} & Q_H \end{bmatrix} = \begin{bmatrix} \sigma_h^2 I_n & 0 \\ 0 & \sigma_H^2 I_n \end{bmatrix} \tag{3.73}$$

Note that the observation equations are nonlinear due to the presence of the scale factor λ. We will now consider four different cases. First we will assume that the difference in scale and the constant offset are absent. Then we will assume that there is only a constant offset, followed by the case that there is only a difference in scale. Finally, we will assume that there is both a constant offset and a difference in scale. For all four cases we will discuss the MDB's that correspond to the use of data snooping.

Data snooping applied to the above connection model, implies that we are testing for errors in the individual height coordinates. It will be clear that with the above model, one will never be able to discriminate whether a blunder occurred in the h-coordinate or in the H- coordinate of a point i. That is, one will never be able to pinpoint a blunder in a height coordinate to one of the two height systems. Hence, it suffices to restrict our attention to the $\underline{w}_i$-test statistics of one of the two sets of heights. We choose to restrict our attention to the h-coordinates. Since the complete variance matrix of the observables is diagonal, we can make use of the simplified expression (3.64) for the MDB's. For the h_i coordinates, it reads

$$| \nabla_i | = \sigma_h \sqrt{\frac{17.075}{1 - \sigma^2_{\hat{h}_i} / \sigma^2_{h_i}}} \tag{3.74}$$

Note that we assumed $\lambda(\alpha_1, 1, \gamma) = 17.075$. This value is based on the values $\alpha_1 = 0.001$ and $\gamma = 0.80$.

Scale and translation absent: When both scale and translation are absent, the model of observation equations becomes linear and the least- squares solution of (3.73) amounts to taking a simple weighted average of the data. The variance of the least-squares solution $\underline{\hat{h}}_i$ reads then

$$\sigma^2_{\hat{h}_i} = \frac{\sigma^2_h \sigma^2_H}{\sigma^2_h + \sigma^2_H} \tag{3.75}$$

Substitution into (3.74) gives for the MDB

$$| \nabla_i | = \sqrt{17.075(\sigma^2_h + \sigma^2_H)} \tag{3.76}$$

This shows that the MDB will be about 5.8 times the standard deviation of the height coordinate, if the two networks are equally precise. Hence, a blunder of this size in the h_i coordinate can be found with a probability of 80% with the w_i-test.

Only scale is absent: In this case the model of observation equations is again linear. Note however that the redundancy has decreased by one. In the previous case the redundancy equalled n, whereas now, due to the additional unknown translation, it equals $n-1$. Due to the decrease in redundancy, the model will have less 'strenght' and one can therefore expect that the results of the adjustment will also be somewhat less precise. And indeed, the variance of the least-squares solution $\underline{\hat{h}}_i$ reads now

$$\sigma^2_{\hat{h}_i} = \frac{\sigma^2_h \sigma^2_H}{\sigma^2_h + \sigma^2_H}(1 + \frac{\sigma^2_h}{\sigma^2_H}\frac{1}{n}) \tag{3.77}$$

Note that (3.75) and (3.77) will differ less, the larger n is. Hence, the difference between the two variances will become smaller when the number of points in the two overlapping networks increases. The difference will also be small when $\sigma^2_H >> \sigma^2_h$, that is, when the second network is considerably less precise than the first network. But in that case we also would have $\sigma^2_{\hat{h}_i} \approx \sigma^2_h$, thus showing that no significant improvement in precision has taken

place. This is of course understandable, because if the second network is considerably less precise than the first network, it will also not contribute much to the adjustment.

Substitution of (3.77) into (3.74) gives for the MDB

$$| \nabla_i | = \sqrt{\frac{17.075(\sigma_h^2 + \sigma_H^2)}{1 - \frac{1}{n}}} \tag{3.78}$$

Note that the MDB goes to infinity when $n = 1$. This is due to the fact that there is no redundancy when $n = 1$. In that case no model errors at all can be found by the statistical tests and the solution is said to be infinitely unreliable. Also note that the above MDB is larger than the one of (3.76). This is due to the decrease in redundancy of one. That is, the outcomes of the statistical tests will now be somewhat less reliable than they were in the previous case. But again, this difference can be made small by increasing the number of points n.

Only translation is absent: When only the translation is absent, we again have a model with a redundancy of $n-1$. Note however, that the observation equations are now non-linear. Hence, first a linearization needs to be carried out. The linearized model reads

$$E\{\begin{bmatrix} \underline{\Delta h}_1 \\ \vdots \\ \underline{\Delta h}_n \\ \underline{\Delta H}_1 \\ \vdots \\ \underline{\Delta H}_n \end{bmatrix}\} = \begin{bmatrix} \lambda^o & & & H_1^o \\ & \ddots & & \vdots \\ & & \lambda^o & H_n^o \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta H_1 \\ \vdots \\ \Delta H_n \\ \Delta\lambda \end{bmatrix} \tag{3.79}$$

where λ^o is the approximate scale factor and H_i^o, $i = 1, \ldots, n$, are the approximate heights of the second network. Since geodetic networks often only differ slightly in scale, one may take as approximate value $\lambda^o = 1$. We will do so here also.

Note that the design matrix of the above model depends on the approximate heights H_i^o. Hence, also the precision of the least- squares estimators will depend on them. The variance of the least-squares solution $\underline{\Delta\hat{h}}_i$ reads now

$$\sigma^2_{\hat{h}_i} = \frac{\sigma_h^2 \sigma_H^2}{\sigma_h^2 + \sigma_H^2}(1 + \frac{\sigma_h^2}{\sigma_H^2} \frac{(H_i^o)^2}{\sum_{j=1}^n (H_j^o)^2}) \tag{3.80}$$

Compare this result with that of (3.77). Substitution of (3.80) into (3.74) gives for the MDB

$$| \nabla_i | = \sqrt{\frac{17.075(\sigma_h^2 + \sigma_H^2)}{1 - \frac{(H_i^o)^2}{\sum_{j=1}^n (H_j^o)^2}}} \tag{3.81}$$

Compare this result with that of (3.78). In the previous case, the MDB was, apart from the a priori precision, only dependent on the number of points n. In the present case however, the MDB has also become dependent on the heights of the points themselves. Thus in the above expression for the MDB, we see four effects at work:

1. A priori precision: $\sigma_h^2 + \sigma_H^2$
2. Number of points: $\sum_{j=1}^{n}(.)^2$
3. Height of the point tested: H_i^o
4. Height distribution of network: $\sum_{j=1}^{n}(H_j^o)^2$

As before the MDB gets smaller when the a priori precision improves and/or when the number of points increases. But now the MDB also gets smaller when the height of the point being tested is closer to zero and/or when the heights of the remaining points are further away from zero.

Both scale and translation present: As in the previous case, the observation equations are again nonlinear. The linearized model reads

$$E\{\begin{bmatrix} \underline{\Delta h}_1 \\ \vdots \\ \underline{\Delta h}_n \\ \underline{\Delta H}_1 \\ \vdots \\ \underline{\Delta H}_n \end{bmatrix}\} = \begin{bmatrix} \lambda^o & & & 1 & H_1^o \\ & \ddots & & \vdots & \vdots \\ & & \lambda^o & 1 & H_n^o \\ 1 & & & 0 & 0 \\ & \ddots & & \vdots & \vdots \\ & & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta H_1 \\ \vdots \\ \Delta H_n \\ \Delta t \\ \Delta\lambda \end{bmatrix} \tag{3.82}$$

Due to the additional unknown, the translation t, the redundancy now equals $n-2$. Thus again we can expect a somewhat less precise and less reliable result. The variance of the least-squares solution $\underline{\Delta\hat{h}}_i$ reads now

$$\sigma_{\hat{h}_i}^2 = \frac{\sigma_h^2\sigma_H^2}{\sigma_h^2+\sigma_H^2}[1+\frac{\sigma_h^2}{\sigma_H^2}(\frac{1}{n}+\frac{(\bar{H}_i^o)^2}{\sum_{j=1}^{n}(\bar{H}_j^o)^2})] \tag{3.83}$$

where $\bar{H}_i^o = H_i^o - \frac{1}{n}\sum_{j=1}^{n}H_j^o$, i.e. the difference in height of point i with respect to the average height of the network. Compare this result with that of (3.80). Substitution of (3.83) into (3.74) gives for the MDB

$$|\nabla_i| = \sqrt{\frac{17.075(\sigma_h^2+\sigma_H^2)}{1-\frac{1}{n}-\frac{(\bar{H}_i^o)^2}{\sum_{j=1}^{n}(\bar{H}_j^o)^2}}} \tag{3.84}$$

Compare this result with that of (3.81). We see almost the same four effects present in the expression for the MDB. The only difference is that the heights are now referenced to the average height of the network, instead of to zero, as it was the case when the translation was absent. Thus the point having the smallest MDB is the one which height is closest to the average of the network. If it coincides with the average, then $\bar{H}_i^o = 0$ and the above expression reduces to that of (3.78).

3.6 Quality control: precision and reliability

We are now in a position to summarize the two main diagnostics by which the quality of estimation and testing can be characterized. They are precision and reliability. In order to discuss them properly, we follow the general steps involved when performing the adjustment and the testing.

Formulate H_o and H_a: In order to be able to perform the adjustment, one first must have a working hypothesis, the null hypothesis H_o, available. It reads

$$H_o: E\{\underline{y}\} = Ax \ , \ D\{\underline{y}\} = Q_y \tag{3.85}$$

This is the model one beliefs to be true and on which one would like to base the adjustment. But of course, relying on this model without checking its validity would be dangerous, since errors in it could completely ruin the results of the adjustment. The purpose of testing is therefore to check whether the null hypothesis is likely to be true or not. This is done by opposing the above model to one or more alternative hypotheses H_a. In these lecture notes we have restricted ourselves to alternative hypotheses, which differ from the null hypothesis in their functional model only. The alternative hypotheses considered are therefore of the form

$$H_a: E\{\underline{y}\} = Ax + C\nabla \ , \ D\{\underline{y}\} = Q_y \tag{3.86}$$

The two type of hypotheses thus differ in their mean of $\underline{y}$ only, $E\{\underline{y} \mid H_a\} = E\{\underline{y} \mid H_o\} + C\nabla$. The vector $\nabla y = C\nabla$, with matrix C known and vector ∇ unknown, describes the assumed model error.

The quality under H_o and H_a: Based on our working hypothesis H_o, the least-squares estimator of x reads

$$\underline{\hat{x}} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} \tag{3.87}$$

It will be clear that the quality of $\underline{\hat{x}}$, as expressed by its expectation and its dispersion, depends on whether H_o is true or H_a is true. In the first case we have

$$H_o \text{ true} \quad \left\{ \begin{array}{lll} \text{mean:} & E\{\underline{\hat{x}}\} & = \ x \\ \text{variance:} & D\{\underline{\hat{x}}\} & = \ Q_{\hat{x}} \end{array} \right. \tag{3.88}$$

with $Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1}$. In the second case however, we have

$$H_a \text{ true} \quad \left\{ \begin{array}{lll} \text{mean:} & E\{\underline{\hat{x}}\} & = \ x + \nabla\hat{x} \\ \text{variance:} & D\{\underline{\hat{x}}\} & = \ Q_{\hat{x}} \end{array} \right. \tag{3.89}$$

with $\nabla\hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C\nabla$. This shows that the least-squares estimator $\underline{\hat{x}}$ has the same variance matrix under H_a as it has under H_o. Thus the precision of the estimator is not affected by the model error $C\nabla$. Its expectation or its mean however, does get affected. Under H_o it is an unbiased estimator, but under H_a it becomes biased, due to the presence of the model error $C\nabla$.

Since one can never be completely sure whether the null hypothesis is true or whether one of the alternative hypotheses is true, the quality of the estimator is made up of the two components $\nabla\hat{x}$ and $Q_{\hat{x}}$. The variance matrix, which describes the precision, is known and in the last section of the previous chapter we discussed how one could evaluate it. The bias $\nabla\hat{x}$ however, is unknown, since it depends on the unknown model error $C\nabla$.

Testing: In order to minimize the risk that a bias like $\nabla\hat{x}$ indeed occurs, one needs to check the validity of the null hypothesis. This is the whole purpose of testing. Although there is only one null hypothesis, there are in practice often many more than one alternative hypotheses. The class of alternative hypotheses considered depends very much on the application at hand and on ones experience. Through the alternative hypotheses one specifies the model errors that one beliefs are likely to occur. The testing procedure to be applied consists then of detection, identification and adaptation. In the detection step the overall validity of the null hypothesis is checked. Ones this test leads to a rejection of H_o, one needs to identify the most likely alternative hypothesis. In the final step, the adaptation step, one then corrects for the identified misspecification in the null hypothesis.

Reliability: Although the statistical testing of H_o minimizes the risk that a bias like $\nabla\hat{x}$ occurs, one should realize that the outcomes of the statistical tests are not exact and thus still prone to errors (type I and type II error). It depends on the 'strength' of the model, how much confidence one will have in these outcomes. A measure of this confidence is provided for by the concept of reliability. When the w- test statistics are used, the *internal* reliability is described by the set of MDB's, one for each alternative hypothesis. The MDB is given as

$$|\nabla| = \sqrt{\frac{\lambda(\alpha_1, 1, \gamma)}{c^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c}} \tag{3.90}$$

It is the size of the model error that can be found with a probability γ, when using the w-test. The internal reliability improves when the MDB's get smaller and gets worse when the MDB's get larger. Note however, that it only makes sense to consider these MDB's when the statistical tests are actually carried out. Hence, the solution is said to be infinitely unreliable by definition, if no statistical testing has taken place. Also note, that the MDB, apart from being dependent on the chosen values for α_1 and γ, is governed by the c- vector, the design matrix A and the variance matrix Q_y. The vector c can not be changed at will, since it depends on the particular model error one is considering. Hence, this leaves us, much like in the case of precision, with the two matrices A and Q_y for improving the internal reliability.

The *external* reliability describes how model errors of the size of the MDB's propagate into the results of the adjustment,

$$\nabla\hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} c\,|\nabla| \tag{3.91}$$

This vector thus describes the bias in $\underline{\hat{x}}$, when a model error of the size of the MDB has occurred.

Precision and reliability: Once one has evaluated the precision of the least-squares solution and found it adequate enough for the application at hand, one can evaluate the significance of the bias vector $\nabla\hat{x}$, through the Bias-to-Noise Ratio (BNR)

$$\lambda_{\hat{x}} = \nabla\hat{x}^T Q_{\hat{x}}^{-1} \nabla\hat{x} \tag{3.92}$$

The dimensionless BNR measures the bias relative to the precision. For an arbitrary function $\underline{\hat{\theta}} = f^T \underline{\hat{x}}$, the BNR can be used to obtain the upperbound

$$\frac{\nabla\hat{\theta}}{\sigma_{\hat{\theta}}} \leq \sqrt{\lambda_{\hat{x}}} \tag{3.93}$$

Thus if for the particular application at hand, the BNR is considered to be small enough, the bias in any function of $\underline{\hat{x}}$ is guaranteed to be sufficiently insignificant as well.

Chapter 4

Adjustment and validation of networks

4.1 Introduction

In this chapter we will discuss the various computational steps involved for determining the geometry of a geodetic network. Let us first briefly review the general steps involved. They are the design, the adjustment and the testing.

Design: Before one can start, one needs to specify the *functional* model and the *stochastic* model. In the functional model, one formulates the assumed relation between the observables and the unknown parameters. These observation equations depend on the type of observables used (e.g. angles, distances, baselines, etc.) and on the choice of parameterization (e.g. Cartesian coordinates or geographic coordinates). The observation equations may be linear or nonlinear. In the nonlinear case, they first need to be linearized. For the linearization, approximate values for the parameters are needed. When the approximate geometry of the network is known, it is often possible to obtain the approximate coordinates from a map. The approximate coordinates can also be computed from a sufficient set of observations. The design matrix A is known, once the functional model is specified.

With the stochastic model, one specifies the assumed distributional properties of the observables. In geodesy it often suffices to assume the observables to be normally distributed. In addition one has to specify the second moment, the variance matrix, of the distribution. The variance matrix of the observables describes their precision. The specification of this variance matrix Q_y depends on the measurement equipment and on the measurement procedures used.

The two matrices A and Q_y are known, once the functional and stochastic model are known. These two matrices can be used before the actual adjustment and testing is carried out, to infer the expected quality in terms of *precision* and *reliability*, of the network. In order to evaluate the precision of the least-squares solution $\underline{\hat{x}}$, its variance matrix $Q_{\hat{x}}$ is used. This matrix quantifies how random errors in the observables propagate into the least-squares solution. In order to evaluate the *reliability* of the least-squares solution, the minimal detectable bias vector $\nabla\hat{x}$ is used. It quantifies how a potential error $c\nabla$ in the functional model propagates into the least-squares solution. The size of the potential error is coupled to the power of the corresponding test statistic. Both $Q_{\hat{x}}$ and $\nabla\hat{x}$ depend on A and Q_y. Hence, one can improve the precision and reliability, by changing A and/or Q_y.

Adjustment: Once one is satisfied with the design of the network, the actual adjustment can be carried out. It needs, apart from the design matrix A and the variance matrix Q_y, of course also the actual observations y. The adjustment is based on the principle of least-squares and it produces a solution for the unknown parameters. This solution is obtained by solving the *normal equations*, for which usually the Cholesky-decomposition is used. In case of nonlinear observation equations, the least-squares solution is usually iterated

using the normal equations based on the linearized observation equations. The number of iterations will be small, when good approximate values are used.

On the basis of the assumption that the model (functional and stochastic) has been specified correctly, the linear(ized) least-squares estimators are known to be unbiased and of minimal variance. These properties will fail to hold however, when a misspecified model has been used in the computations. Hence, before the least-squares solution $\hat{x}$ can be accepted, one has to test the validity of the model.

Testing: The validity of the model (the null hypothesis) is tested by opposing it to various alternative models (the alternative hypotheses). For each alternative hypothesis, one has an appropriate test statistic. But all test statistics are functions of the least- squares residual vector $\underline{\hat{e}}$. The testing procedure consists of the following three steps: detection, identification and adaptation. In the *detection* step, the null hypothesis is opposed to the most relaxed alternative hypothesis. The purpose of the detection step is to infer whether one has any reason to belief that the null hypothesis is indeed wrong. When the detection step leads to a rejection of the null hypothesis, the next step is the *identification* of the most likely model error. For identification one needs to specify the alternative hypothesis. This choice depends on the type of model error one expects to be present. Hence, it very much depends on the application at hand. It is standard practice however, to have *data snooping* included in the identification step. In case of data snooping each of the individual observations is screened for potential blunders. Once certain model errors have been identified as sufficiently likely, the last step consists of an *adaptation* of the data and/or model. Depending on the situation, two approaches are possible in principle. One can either decide to remeasure some of the observables, or, one can decide to include additional parameters in the model, such that the model errors are accounted for. The first approach is possible in case the model errors are due to clear measurement errors, e.g. blunders in individual observations. The second approach can be used for more complicated situations. In this case the identified alternative hypothesis will become the new null hypothesis. One should be aware however, that in this case, due to the change in model, the precision and reliability of the solution will change as well.

The above considerations for the design, the adjustment and the testing, are valid for any geodetic project where measurements are used to determine unknown parameters. When computing geodetic networks however, some additional aspects need to be considered as well. The construction of a geodetic network implies that the geometry of the configuration of a set of points is determined. The set of points usually consists of: (1) newly established points, of which the coordinates still need to be determined, and (2) already existing points, the so-called control points, of which the coordinates are known. By means of a network adjustment the relative geometry of the new points is determined and integrated into the geometry of the existing control points. The determination of the geometry is usually divided into two parts, the so-called free network adjustment and the connection adjustment.

The free network: In the free network adjustment, the known coordinates of the control

points do not take part in the determination of the geometry of the point field. This adjustment step is thus *free* from the influence of the existing control points. The idea is that a good geodetic network should be sufficiently precise and reliable in itself, without the need of external control.

Since the coordinates of the control points do not take part in the adjustment, one is confronted with the fundamental non-uniqueness in the relation between geodetic observables and coordinates. In a levelling network for instance, absolute heights can not be determined if only height differences are measured. That is, for computing heights, additional information is needed on the absolute height of the network. Similarly, one can not obtain the position, the orientation and the scale of a triangulation network if only angles are measured. The additional information which is needed to be able to compute coordinates from the geodetic observables, is provided for in the form of so-called *minimal constraints*. These minimal constraints are not unique. There is a whole set from which the constraints can be chosen. It is important though, that the constraints are minimal. That is, they should not only be necessary to eliminate the lack of information in the observables, but they should also be sufficient.

After the design of the free network, its coordinates are computed by means of a least-squares adjustment and its validity is checked by means of the statistical testing of the observations. The coordinates depend on the chosen minimal constraints, but the statistical tests do not.

The connected network: Once the geometry of the free network has been determined to ones satisfaction, it needs to be integrated into the existing geometry of the control points. The data used for this connection, are the results of the free network adjustment together with the coordinates of the control points. One can descriminate between the so-called *constrained connection* and the *unconstrained connection*. In most applications, it is not very practical to see the coordinates of the control points change everytime a free network is connected to them. This would happen however, when an ordinary least-squares adjustment is carried out. In that case all observations, including the coordinates of the control points, would get corrections due to the least-squares adjustment. In order to circumvent this, no ordinary adjustment, but a constrained adjustment is carried out. This implies that the connection is carried out, with the explicit constraints that the coordinates of the existing control points remain fixed.

For the statistical testing of the observations however, a constrained adjustment would not be realistic. Although practice dictates that the coordinates of the control points remain fixed, these coordinates are of course still samples from random variables. Hence, for the statistical testing of the observations, the variance matrix of the control points should not be set to zero, but should be included in the adjustment as well. Thus, the constrained connection is carried out for the final computation of the coordinates, but the unconstrained connection is used for the statistical testing of the control points.

This chapter is organized as follows. First we will discuss the free network case and then we will show how they can be connected to existing control. For the free networks, we present the observation equations, discuss their invariance and show how one can choose

appropriate sets of minimal constraints. As an example we discuss the adjustment of a small free GPS network and the testing of a small free levelling network.

4.2 The free network

In this section we consider networks for which the observational data are insufficient to determine either the (horizontal and/or vertical) position, orientation or scale of the network (*note*: here and in the remaining part of the lecture notes, we will disregard so called configuration defects. That is, we assume that sufficient observational data are used to determine the configuration of the network).

As we have seen in the earlier sections on adjustment theory, a consequence of the non-uniqueness is that the design matrix A of the model of observation equations $E\{\underline{y}\} = Ax$, $D\{\underline{y}\} = Q_y$, will have a rank defect. Therefore no unbiased linear estimator $\hat{x} = Ly$ of x exists, since this would require $E\{\underline{\hat{x}}\} = LE\{\underline{y}\} = Ax$ for all x, or $AL = I$, which is impossible since the rank of a product of two matrices can not exceed the rank of either factor. But although x is not unbiased estimable, we have seen that functions of x exist that are unbiased estimable. In particular we have shown that the *minimally constrained* solution $\underline{\hat{x}}_b$ is an unbiased estimator of $S_b x$, with S_b being the S-transformation defined by the null space of the design matrix, $N(A) = R(G)$, and the minimal constraints $B^T x = 0$.

In this section we will show how minimally constrained solutions for geodetic networks can be constructed. We start off by presenting the nonlinear observation equations of some geodetic observables and their linearized versions. Then we discuss the invariance properties of these geodetic observables. Their invariance can usually be related to properties of coordinate transformations. Once these properties of invariance are understood, one will be able to identify the null space of the design matrix A. As a consequence the matrix G, of which the columns span $N(A)$, can be constructed. Understanding the invariance, also allows one to specify the minimal constraints $B^T x = 0$. They are needed to be able to compute a particular least-squares solution of the rank defect model $E\{\underline{y}\} = Ax$, $D\{\underline{y}\} = Q_y$. Such a solution is one of the many free network solutions that can be computed. By means of the S-transformation, one is able to transform one particular free network solution into another. This section will be concluded with an adjustment example of a free GPS network and a testing example of a free levelling network.

4.2.1 Some typical observation equations

In this section we will present some typical examples of geodetic observation equations. Also their linearized versions are given. We will parametrize the observation equations in terms of Cartesian coordinates (in one dimension the height h; in two dimensions: x and y; and in three dimensions: x, y and z).

Height difference: Probably the simplest of all geodetic observation equations is the one that corresponds with observed height differences. Let h_{ij} denote the height difference between two points i and j, and let the heights of these two points be denoted as respectively h_i and h_j. The observation equation for an observed height difference reads then

$$E\{\underline{h}_{ij}\} = h_j - h_i \tag{4.1}$$

Since this equation is already linear, no linearization is needed.

Azimuth: If we consider a two dimensional network and assume that all network points are located in the two-dimensional Euclidean plane, the nonlinear observation equation for an azimuth a_{ij} between the two points i and j reads

$$E\{\underline{a}_{ij}\} = \arctan\frac{x_j - x_i}{y_j - y_i} \tag{4.2}$$

Linearization gives

$$E\{\Delta\underline{a}_{ij}\} = \frac{y_j^o - y_i^o}{(l_{ij}^o)^2}\Delta x_j - \frac{y_j^o - y_i^o}{(l_{ij}^o)^2}\Delta x_i - \frac{x_j^o - x_i^o}{(l_{ij}^o)^2}\Delta y_j + \frac{x_j^o - x_i^o}{(l_{ij}^o)^2}\Delta y_i \tag{4.3}$$

with

$$\Delta\underline{a}_{ij} = \underline{a}_{ij} - \arctan\frac{x_j^o - x_i^o}{y_j^o - y_i^o}$$

and

$$\Delta x_j = x_j - x_j^o\ ,\ \Delta x_i = x_i - x_i^o\ ,\ \Delta y_j = y_j - y_j^o\ ,\ \Delta y_i = y_i - y_i^o$$

and where $(.)^o$ denotes the approximate value.

Distance: For a planar network, the nonlinear observation equation for a distance reads

$$E\{\underline{l}_{ij}\} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{4.4}$$

Linearization gives

$$E\{\Delta\underline{l}_{ij}\} = \frac{x_j^o - x_i^o}{(l_{ij}^o)}\Delta x_j - \frac{x_j^o - x_i^o}{(l_{ij}^o)}\Delta x_i + \frac{y_j^o - y_i^o}{(l_{ij}^o)}\Delta y_j - \frac{y_j^o - y_i^o}{(l_{ij}^o)}\Delta y_i \tag{4.5}$$

with

$$\Delta\underline{l}_{ij} = \underline{l}_{ij} - \sqrt{(x_j^o - x_i^o)^2 + (y_j^o - y_i^o)^2}$$

Angle: An angle α_{ijk} between three points i, j and k, is the difference between their azimuths a_{jk} and a_{ji}. Thus we have

$$E\{\underline{\alpha}_{ijk}\} = E\{\underline{a}_{jk} - \underline{a}_{ji}\} = \arctan\frac{x_k - x_j}{y_k - y_j} - \arctan\frac{x_i - x_j}{y_i - y_j} \tag{4.6}$$

Its linearized version follows then from the linearized versions of the two azimuths.

Distance ratio: The distance ratio v_{ijk} between three points i, j and k, is the ratio of their distances l_{jk} and l_{ji}. Thus we have

$$E\{\underline{v}_{ijk}\} = E\{\frac{\underline{l}_{jk}}{\underline{l}_{ji}}\} = \frac{\sqrt{(x_k - x_j)^2 + (y_k - y_j)^2}}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \tag{4.7}$$

Linearization gives

$$E\{\Delta\underline{v}_{ijk}\} = \frac{l^o_{ji}\Delta l_{jk} - l^o_{jk}\Delta l_{ji}}{(l^o_{ji})^2}$$

This can be further expressed in terms of the coordinate increments by making use of (4.5).

Direction: A direction r_{ij} is an "azimuth with an unknown orientation". Its observation equation reads thus

$$E\{\underline{r}_{ij}\} = a_{ij} + o_j \tag{4.8}$$

where o_j is the orientation unknown. Apart form the additional orientation unknown, the linearized observation equation for a direction is the same as that for an azimuth. Note that the difference of two directions having the same orientation, produces an angle.

Pseudo-distance: A pseudo-distance s_{ij} (not to be confused with the pseudo-distance of GPS) is a "distance with an unknown scale". Its observation equation reads therefore

$$E\{\underline{s}_{ij}\} = \lambda_i l_{ij} \tag{4.9}$$

where λ_i is the scale unknown. Linearization gives

$$E\{\Delta\underline{s}_{ij}\} = \lambda^o_i \Delta l_{ij} + l^o_{ij}\Delta\lambda_i$$

This can be further expressed in terms of the coordinate increments by making use of (4.5). Note that the ratio of two pseudo-distances having the same scale factor, produces a distance ratio.

The three dimensional case: So far we assumed all network points to lie in the two dimensional Euclidean plane. For a three dimensional network though, we have to take the third dimension into account as well. For a network of a sufficiently large extent, also the change in the direction of the plumbline will have to be taken into account. This implies that for direction measurements like the azimuth a_{ij} and the zenith angle z_{ij}, the astronomical latitude Φ_i and astronomical longitude Λ_i will enter the observation equations as well. The nonlinear observation equations for respectively the azimuth, the zenith angle and the distance between two points i and j, are given as

$$\begin{aligned}
E\{\underline{a}_{ij}\} &= \arctan \frac{-\sin\Lambda_i(x_j-x_i)+\cos\Lambda_i(y_j-y_i)}{-\sin\Phi_i\cos\Lambda_i(x_j-x_i)-\sin\Phi_i\sin\Lambda_i(y_j-y_i)+\cos\Phi_i(z_j-z_i)} \\
E\{\underline{z}_{ij}\} &= \arccos \frac{\cos\Phi_i\cos\Lambda_i(x_j-x_i)+\cos\Phi_i\sin\Lambda_i(y_j-y_i)+\sin\Phi_i(z_j-z_i)}{l_{ij}} \\
E\{\underline{l}_{ij}\} &= \sqrt{(x_j-x_i)^2+(y_j-y_i)^2+(z_j-z_i)^2}
\end{aligned} \tag{4.10}$$

The unknowns in these observation equations are now apart from the Cartesian coordinates, also the astronomical latitude and longitude. The linearization of the above observation equations is left as an exercise to the reader.

GPS baseline: As our last example we consider the three dimensional baseline. When expressed in Cartesian coordinates, the observation equations for its three components read

$$\begin{aligned} E\{\underline{x}_{ij}\} &= x_j - x_i \\ E\{\underline{y}_{ij}\} &= y_j - y_i \\ E\{\underline{z}_{ij}\} &= z_j - z_i \end{aligned} \tag{4.11}$$

As with the height differences no linearization is needed, since the equations are already linear in the parameters.

Instead of using Cartesian coordinates, one may of course use other type of coordinates as well. In three dimensions, one often also makes use of the geographic coordinates ϕ, λ and h, where h now refers to the height above the reference ellipsoid. The Cartesian coordinates and geographic coordinates are related as

$$\begin{aligned} x &= (N+h)\cos\phi\cos\lambda \\ y &= (N+h)\cos\phi\sin\lambda \\ z &= (N(1-e^2)+h)\sin\phi \end{aligned} \tag{4.12}$$

where N is the radius of curvature in the prime vertical

$$N = \frac{a}{\sqrt{1-e^2\sin^2\phi}}$$

and $e^2 = (a^2-b^2)/a^2$, with a and b being the lenghts of the major and minor axes of the ellipsoid of revolution.

4.2.2 On the invariance of geodetic observables

It will be clear that most geodetic observables (if not all!) do not contain enough information to determine the coordinates of a geodetic network uniquely. That is, the geodetic observables are invariant against certain coordinate transformations. In this section we will investigate the invariance of some geodetic observables and show how the G-matrix, of which the columns span the null space of the design matrix A, can be constructed. We also give examples of the minimal constraints that can be chosen.

The one dimensional case: As a one dimensional example we consider the case of levelling. Let h_i denote the height of point i in the first coordinate system and let H_i denote the height of the same point in the second coordinate system. The coordinate transformation between the two coordinate systems reads then

$$h_i = H_i + t \tag{4.13}$$

where t is a translation or a shift in height, which is constant for all points. It will be clear that this coordinate transformation leaves observed height differences invariant. That is, for the observation equation of a height difference we have

$$E\{\underline{h}_{ij}\} = h_j - h_i = H_j - H_i \tag{4.14}$$

This shows that the height differences are invariant for the translation t. This immediately implies that the design matrix A of a levelling network that is built up from height differences only, will have a null space which is spanned by the column of the matrix

$$G = \underset{n\times 1}{\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}} \tag{4.15}$$

where n is the number of points in the network.

Since matrix G has only one column, only one constraint is needed to eliminate the nonuniqueness in the relation between height differences and heights. Possible choices for the matrix B of the minimal constraints $B^T x = 0$ are

$$\begin{aligned} B^T &= (1,0,\ldots,0) \\ B^T &= (0,0,\ldots,1) \\ B^T &= (1,0,\ldots,1) \\ B^T &= (1,1,\ldots,1) \end{aligned} \tag{4.16}$$

In the first case, the height of the first point is fixed and in the second case, the height of the last point is fixed. With the third choice, the sum of the heights of the first and last point is fixed. The fourth case, corresponds to a fixing of the sum of heights of all points. When the two matrices G and B are known, one can also construct the S-transformation, $S_b = I - G(B^T G)^{-1} B^T$, that corresponds to the chosen minimal constraints. When the height of the first point is fixed, it reads

$$S_b = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ -1 & 1 & & \\ \vdots & & \ddots & \\ -1 & & & 1 \end{bmatrix}$$

The two dimensional case: In two dimensions we are dealing with observables like angles and distances. Their observation equations are nonlinear. The construction of the G-matrix is now a bit more involved, since a linearization is involved as well. Let us first show the general principle of constructing the G-matrix. Let us assume that the nonlinear observation equations are invariant for the coordinate transformation

$$x = T(u,p) \tag{4.17}$$

where the vector x contains the coordinates of the first coordinate system, the vector u contains the coordinates of the same points in the second coordinate system and where vector p contains the transformation parameters. Invariance of the nonlinear observation equations $A(x)$ for the transformation (4.17), implies that $A(x) = A(u)$ and thus that

$$A(T(u,p)) = A(u) \quad \text{for all } p \tag{4.18}$$

Since the observation equations are insensitive to changes in p, it follows if we take the partial derivatives of (4.18) with respect to p, that

$$\underbrace{\partial_x A(x^o)}_{A}\ \underbrace{\partial_p T(u^o, p^o)}_{G} = 0 \tag{4.19}$$

Thus once we know the transformation which leaves the observation equations invariant, we only need to take its partial derivative with respect to the parameters in order to obtain the G-matrix.

The transformation which plays an important role in case of geodetic observables, is the *Similarity transformation*. For the two dimensional case, it reads (see figure 4.1)

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \lambda \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} u_i \\ v_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{4.20}$$

with

λ	:	scale
α	:	rotation
t_x, t_y	:	translation

and where x_i, y_i are the coordinates of the first coordinate system and u_i, v_i, the coordinates of the second coordinate system. Linearization of the similarity transformation, assuming

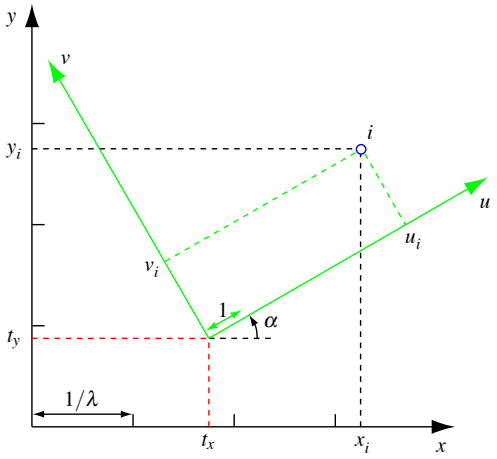


Figure 4.1 The two dimensional similarity transformation.

that $\lambda^o = 1$, $\alpha^o = 0$, $t_x^o = 0$ and $t_y^o = 0$, gives

$$\begin{bmatrix} \Delta x_i \\ \Delta y_i \end{bmatrix} = \begin{bmatrix} \Delta u_i \\ \Delta v_i \end{bmatrix} + \begin{bmatrix} 1 & 0 & x_i^o & -y_i^o \\ 0 & 1 & y_i^o & x_i^o \end{bmatrix} \begin{bmatrix} \Delta t_x \\ \Delta t_y \\ \Delta\lambda \\ \Delta\alpha \end{bmatrix} \tag{4.21}$$

Angles and distance ratios: It will be clear that angles fail to contain information about scale, orientation and translation. This also holds true for distance ratios. This can be seen as follows. Substitution of the coordinate transformation (4.20) into the observation equation of a distance ratio gives

$$E\{\frac{l_{jk}}{l_{ji}}\} = \frac{\sqrt{x_{jk}^2 + y_{jk}^2}}{\sqrt{x_{ji}^2 + y_{ji}^2}} = \frac{\sqrt{u_{jk}^2 + v_{jk}^2}}{\sqrt{u_{ji}^2 + v_{ji}^2}}$$

thus showing that the transformation parameters λ, α, t_x and t_y are absent. These parameters will also be absent when one considers the observation equation of an angle. As a consequence, the design matrix A of a geodetic network, that has been built up from distance ratios and/or angles only, will have a null space which is spanned by the columns of the matrix

$$G = \underset{2n \times 4}{\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & x_i^o & -y_i^o \\ 0 & 1 & y_i^o & x_i^o \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}} \tag{4.22}$$

where n is the number of points in the network.

Since matrix G has four independent columns, four constraints are needed to take care of the nonuniqueness. The simplest way to fix the degrees of freedom of scale, orientation and translation, would be to fix the coordinates of two points. The corresponding B-matrix reads then

$$B^T = \underset{4 \times 2n}{\begin{bmatrix} 0 & 0 & \dots & I_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & I_2 & \dots & 0 & 0 \end{bmatrix}} \tag{4.23}$$

Azimuth: When we substitute the coordinate transformation (4.20) into the observation equation of an azimuth, we get

$$E\{\underline{a}_{ij}\} = \arctan \frac{x_{ij}}{y_{ij}} = \arctan \frac{\cos\alpha\, u_{ij} - \sin\alpha\, v_{ij}}{\sin\alpha\, u_{ij} + \cos\alpha\, v_{ij}}$$

This shows that scale, λ, and the two translations, t_x, t_y, get eliminated, but that the angle of rotation α does not get eliminated. Hence, with azimuth observables one cannot determine the scale and position of the network, but only its orientation. As a consequence, the design matrix A of a geodetic network, that has been built up from azimuths only, will have a null space which is spanned by the columns of the matrix

$$G = \underset{2n \times 3}{\begin{bmatrix} \vdots & \vdots & \vdots \\ 1 & 0 & x_i^o \\ 0 & 1 & y_i^o \\ \vdots & \vdots & \vdots \end{bmatrix}} \tag{4.24}$$

Now three constraints are needed. One could for instance fix the two coordinates of the first point. This takes care of the two translational degrees of freedom. The scale can be fixed by constraining the distance between the first and second point. The corresponding B-matrix reads then

$$B^T = \underset{3 \times 2n}{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & \dots \\ -x_{12}^o & -y_{12}^o & x_{12}^o & y_{12}^o & 0 & \dots \end{bmatrix}} \tag{4.25}$$

Distance: For distance observables one can expect that the rotation angle and the translations get eliminated. And indeed, we have

$$E\{l_{ij}\} = \sqrt{x_{ij}^2 + y_{ij}^2} = \lambda \sqrt{u_{ij}^2 + v_{ij}^2}$$

Thus in this case the G-matrix is given as

$$G = \underset{2n \times 3}{\begin{bmatrix} \vdots & \vdots & \vdots \\ 1 & 0 & -y_i^o \\ 0 & 1 & x_i^o \\ \vdots & \vdots & \vdots \end{bmatrix}} \tag{4.26}$$

In this case an admissible set of minimal constraints would be: the fixing of the two coordinates of one point and the fixing of the orientation between the first and the second point.

The three dimensional case: In three dimensions, the nonlinear similarity transformation is given as

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \lambda \; R(\alpha)R(\beta)R(\gamma) \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{4.27}$$

with

λ	:	scale
α, β, γ	:	rotation
t_x, t_y, t_z	:	translation

and the three rotation matrices

$$R(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix}, \; R(\beta) = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix}, \; R(\gamma) = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Linearization, using the approximate values $\lambda^o = 1$, $\alpha^o = \beta^o = \gamma^o = 0$ and $t_x^o = t_y^o = t_z^o = 0$, gives

$$\begin{bmatrix} \Delta x_i \\ \Delta y_i \\ \Delta z_i \end{bmatrix} = \begin{bmatrix} \Delta u_i \\ \Delta v_i \\ \Delta w_i \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 & 0 & x_i^o & 0 & -z_i^o & y_i^o \\ 0 & 1 & 0 & y_i^o & z_i^o & 0 & -x_i^o \\ 0 & 0 & 1 & z_i^o & -y_i^o & x_i^o & 0 \end{bmatrix}}_{\partial_u T(u^o, p^o)} \begin{bmatrix} \Delta t_x \\ \Delta t_y \\ \Delta t_z \\ \Delta\lambda \\ \Delta\alpha \\ \Delta\beta \\ \Delta\gamma \end{bmatrix} \tag{4.28}$$

As in the two dimensional case, the G-matrix is again built up from some or all of the columns of matrix $\partial_u T(u^o, p^o)$. In case of angles and distance ratios, all the columns are used. For other observables though, only some of the columns are used.

Table 4.1 Entrie(s) of G-column(s) for different observation types.

dim. netw.	dim$N(A)$= dim$R(G)$	actual observation types(s)	entries of G-column(s)		
			translation(s)	rotation(s)	scale
1	1	height differences	1		
2	2	distances and azimuts	$\begin{matrix}1 & 0\\ 0 & 1\end{matrix}$		
	3	distances	$\begin{matrix}1 & 0\\ 0 & 1\end{matrix}$	$\begin{matrix}y_i^o\\ -x_i^o\end{matrix}$	
	4	angles and/or distance ratios	$\begin{matrix}1 & 0\\ 0 & 1\end{matrix}$	$\begin{matrix}y_i^o\\ -x_i^o\end{matrix}$	$\begin{matrix}x_i^o\\ y_i^o\end{matrix}$
3	3	distances, azimuts astronomical latitudes and longitudes	$\begin{matrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{matrix}$		
	6	distances	$\begin{matrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{matrix}$	$\begin{matrix}0 & -z_i^o & y_i^o\\ z_i^o & 0 & -x_i^o\\ -y_i^o & x_i^o & 0\end{matrix}$	
	7	angles and/or distance ratios	$\begin{matrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{matrix}$	$\begin{matrix}0 & -z_i^o & y_i^o\\ z_i^o & 0 & -x_i^o\\ -y_i^o & x_i^o & 0\end{matrix}$	$\begin{matrix}x_i^o\\ y_i^o\\ z_i^o\end{matrix}$

GPS baseline: Since the baseline observable is only invariant for translations and not for rotations and scale changes, the G-matrix of a geodetic network built up from baselines only, is given as

$$G = \underset{3n\times 3}{\begin{bmatrix} \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix}} \tag{4.29}$$

Note that this is the three dimensional analogue of the levelling-case.

To conclude this section, we have given in table 4.1 the various entries of the G-matrix, when different type of geodetic observables are used.

4.2.3 Adjustment of free GPS network: an example

Let us assume that we have three GPS baselines available, between the three points 1, 2 and 3. The baseline between the two points i and j will be denoted as b_{ij}. Thus, when the baseline is expressed in Cartesian coordinates, we have

$$b_{ij} = (x_{ij}, y_{ij}, z_{ij})^T$$

It is our goal to determine the Cartesian coordinates of the three points 1, 2 and 3. The position vector of point i will be denoted as p_i. Thus

$$p_i = (x_i, y_i, z_i)^T$$

We will assume that the three baselines have been determined independently. Thus no correlation is assumed to exist between the baselines. We also assume that the three baselines have been determined with the same precision. Thus we assume that all three baselines have the same variance matrix, which will be denoted as Q.

Based on the above assumptions, we can formulate the model of observation equations as

$$\underbrace{E\{\begin{bmatrix} \underline{b}_{12} \\ \underline{b}_{23} \\ \underline{b}_{31} \end{bmatrix}\}}_{\underline{y}} = \underbrace{\begin{bmatrix} -I_3 & I_3 & 0 \\ 0 & -I_3 & I_3 \\ -I_3 & 0 & I_3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}}_{x} \quad , \quad \underbrace{D\{\begin{bmatrix} \underline{b}_{12} \\ \underline{b}_{23} \\ \underline{b}_{31} \end{bmatrix}\}}_{\underline{y}} = \underbrace{\begin{bmatrix} Q & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & Q \end{bmatrix}}_{Q_y} \tag{4.30}$$

This is a system of 9 observation equations in 9 unknowns. Note however, that the design matrix is not of full rank. It has a rank defect of 3. Thus the *redundancy* of the model equals $9 - 9 + 3 = 3$. Thus if one would formulate the model in terms of conditions equations, one would have 3 independent condition equations. These three equations are given as $E\{\underline{b}_{12} + \underline{b}_{23} + \underline{b}_{31}\} = 0$.

Due to the rank defect of the model of observation equations, we need to specify minimal constraints in order to be able to compute a particular least-squares solution. We know that the range space of the matrix B of the minimal constraints $B^T x = 0$, needs to satisfy $R^n = R(B) \oplus R(A^T) = R(B) \oplus N(A)^\perp$. Thus $R(B)$ needs to be complementary to $N(A)^\perp = R(G)^\perp$. Since the null space of A is spanned by the columns of the matrix

$$G = \begin{bmatrix} I_3 \\ I_3 \\ I_3 \end{bmatrix} \tag{4.31}$$

it follows that $R(B)$ needs to be complementary to

$$G^\perp = \begin{bmatrix} I_3 & 0 \\ -I_3 & I_3 \\ 0 & -I_3 \end{bmatrix}$$

This condition is satisfied by the following three choices for matrix B,

$$B_{(1)} = \begin{bmatrix} I_3 \\ 0 \\ 0 \end{bmatrix} , B_{(3)} = \begin{bmatrix} 0 \\ 0 \\ I_3 \end{bmatrix} , B_{(1+2+3)} = \begin{bmatrix} I_3 \\ I_3 \\ I_3 \end{bmatrix} \tag{4.32}$$

There are of course many more matrices that satisfy the above condition, but we will confine our discussion to the above three.

Point 1 as fixed (datum) point: The choice of matrix $B_{(1)}$ for the minimal constraints, corresponds to a fixing of the coordinates of point 1. The corresponding least- squares solution will be denoted as $\underline{\hat{x}}_{(1)}$. It reads

$$\underline{\hat{x}}_{(1)} = B^{\perp}_{(1)}[(B^{\perp}_{(1)})^T A^T Q_y^{-1} A B^{\perp}_{(1)}]^{-1}(B^{\perp}_{(1)})^T A^T Q_y^{-1}\underline{y}$$

and it has as variance matrix

$$Q_{\hat{x}_{(1)}} = B^{\perp}_{(1)}[(B^{\perp}_{(1)})^T A^T Q_y^{-1} A B^{\perp}_{(1)}]^{-1}(B^{\perp}_{(1)})^T$$

Since

$$A^T Q_y^{-1} A = \begin{bmatrix} 2Q^{-1} & -Q^{-1} & -Q^{-1} \\ -Q^{-1} & 2Q^{-1} & -Q^{-1} \\ -Q^{-1} & -Q^{-1} & 2Q^{-1} \end{bmatrix} \text{ and } B^{\perp}_{(1)} = \begin{bmatrix} 0 & 0 \\ I_3 & 0 \\ 0 & I_3 \end{bmatrix}$$

it follows that

$$[(B^{\perp}_{(1)})^T A^T Q_y^{-1} A B^{\perp}_{(1)}]^{-1} = \begin{bmatrix} 2Q^{-1} & -Q^{-1} \\ -Q^{-1} & 2Q^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{2}{3}Q & \frac{1}{3}Q \\ \frac{1}{3}Q & \frac{2}{3}Q \end{bmatrix}$$

Hence, the variance matrix of the minimally constrained least-squares solution, keeping point 1 fixed, reads

$$Q_{\hat{x}_{(1)}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{2}{3}Q & \frac{1}{3}Q \\ 0 & \frac{1}{3}Q & \frac{2}{3}Q \end{bmatrix} \tag{4.33}$$

This matrix describes the precision of the coordinates of the free GPS network, when the minimal constraints correspond to a fixing of point 1.

Point 3 as fixed (datum) point: Instead of choosing point 1 as fixed point, one may of course also choose another point, say point 3. In that case the variance matrix of the minimally constrained least-squares solution, reads

$$Q_{\hat{x}_{(3)}} = \begin{bmatrix} \frac{2}{3}Q & \frac{1}{3}Q & 0 \\ \frac{1}{3}Q & \frac{2}{3}Q & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.34}$$

This matrix describes the precision of the coordinates of the free GPS network, when the minimal constraints correspond to a fixing of point 3. Note that the two variance matrices of (4.33) and (4.34) differ greatly. It is important to recognize however, that these two variance matrices contain identical information. One can transform the one into the other by means of an S-transformation. The transformation, that transforms any arbitrary least-squares solution to $\hat{x}_{(1)}$, reads

$$\begin{aligned} S_{(1)} &= I_9 - G(B^T_{(1)}G)^{-1}B^T_{(1)} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ -I_3 & I_3 & 0 \\ -I_3 & 0 & I_3 \end{bmatrix} \end{aligned} \tag{4.35}$$

Hence, the variance matrix $Q_{\hat{x}_{(1)}}$ can be obtained from the variance matrix $Q_{\hat{x}_{(3)}}$, by means of the transformation (verify yourself) $Q_{\hat{x}_{(1)}} = S_{(1)} Q_{\hat{x}_{(3)}} S_{(1)}^T$.

Fixing the sum of the coordinates: Instead of fixing the three coordinates of one of the points of the network, one may also decide to fix of all points of the network, their sum of x-coordinates, their sum of y-coordinates and their sum of z-coordinates. Also these three constraints are admissible. The S-transformation that corresponds with this set of minimal constraints is given as

$$\begin{aligned} S_{(1+2+3)} &= I_9 - G(B_{(1+2+3)}^T G)^{-1} B_{(1+2+3)}^T \\ &= \begin{bmatrix} \frac{2}{3}I_3 & -\frac{1}{3}I_3 & -\frac{1}{3}I_3 \\ -\frac{1}{3}I_3 & \frac{2}{3}I_3 & -\frac{1}{3}I_3 \\ -\frac{1}{3}I_3 & -\frac{1}{3}I_3 & \frac{2}{3}I_3 \end{bmatrix} \end{aligned} \tag{4.36}$$

With this S-transformation, we can thus obtain the variance matrix of $\hat{x}_{(1+2+3)}$ from $Q_{\hat{x}_{(3)}}$ as

$$\begin{aligned} Q_{\hat{x}_{(1+2+3)}} &= S_{(1+2+3)} Q_{\hat{x}_3} S_{(1+2+3)}^T \\ &= \begin{bmatrix} \frac{2}{9}Q & -\frac{1}{9}Q & -\frac{1}{9}Q \\ -\frac{1}{9}Q & \frac{2}{9}Q & -\frac{1}{9}Q \\ -\frac{1}{9}Q & -\frac{1}{9}Q & \frac{2}{9}Q \end{bmatrix} \end{aligned} \tag{4.37}$$

The entries of this variance matrix again differ greatly from the entries of $Q_{\hat{x}_{(1)}}$ and $Q_{\hat{x}_{(3)}}$. The three minimally constrained solutions $\hat{x}_{(1)}, \hat{x}_{(3)}$ and $\hat{x}_{(1+2+3)}$ are however completely equivalent. All three produce the same least-squares solution for the measurements. This thus also holds for the variance matrix $Q_{\hat{y}}$. Verify yourself that indeed $Q_{\hat{y}} = A Q_{\hat{x}_{(1)}} A^T = A Q_{\hat{x}_{(3)}} A^T = A Q_{\hat{x}_{(1+2+3)}} A^T$.

Evaluation of free network precision: When evaluating the precision of a free network, one has to make sure that the evaluation is not affected by the choice of minimal constraints. These constraints do not contain information which is essential for the precision-evaluation. They are merely a tool to be able to compute coordinates. Thus when the precision evaluation is based on, say $Q_{\hat{x}_{(1)}}$, one should use a procedure which gives results that are identical to the results that one would obtain, when the precision evaluation is based on, say $Q_{\hat{x}_{(3)}}$. Hence, when one wants to compare $Q_{\hat{x}_{(1)}}$ with a criterium matrix, one has to make sure that the criterium matrix is defined with respect to the same set of minimal constraints. This can be accomplished by transforming the criterium matrix with the appropriate S-transformation. Thus when C_x denotes the criterium matrix, one should compare $Q_{\hat{x}_{(1)}}$ with $S_{(1)} C_x S_{(1)}^T$, and $Q_{\hat{x}_{(3)}}$ with $S_{(3)} C_x S_{(3)}^T$. Only then will the two evaluations give identical results.

In case one decides to base the evaluation on the generalized eigenvalue problem, the appropriate formulation is thus

$$| \, Q_{\hat{x}_{(1)}} - \lambda S_{(1)} C_x S_{(1)}^T \, | = 0 \tag{4.38}$$

and not $| Q_{\hat{x}_{(1)}} - \lambda C_x | = 0$. And if it is $Q_{\hat{x}_{(3)}}$, that needs to be evaluated, the appropriate formulation is

$$| Q_{\hat{x}_{(3)}} - \lambda S_{(3)} C_x S_{(3)}^T | = 0 \tag{4.39}$$

and not $| Q_{\hat{x}_{(3)}} - \lambda C_x | = 0$. Both the eigenvalue problems (4.38) and (4.39), will give identical results for the eigenvalues.

4.2.4 Testing of free levelling network: an example

The levelling network consists of the four points 1, 2, 3 and 4, and the following five observables

$$\underline{h}_{12} \, , \, \underline{h}_{23} \, , \, \underline{h}_{31} \, , \, \underline{h}_{34} \, , \, \underline{h}_{42}$$

Thus the network consists of two levelling loops, loop $1-2-3$ and loop $2-3-4$. The observables are assumed to be uncorrelated, all having the same variance σ^2. We know that we need to introduce one minimal constraint in order to eliminate the rank deficiency. As minimal constraint we choose to fix the height of the first point:

$$h_1 = 0$$

The model of observation equations, with the minimal constraint included, reads then

$$E\{ \begin{bmatrix} \underline{h}_{12} \\ \underline{h}_{23} \\ \underline{h}_{31} \\ \underline{h}_{34} \\ \underline{h}_{42} \end{bmatrix} \} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} h_2^1 \\ h_3^1 \\ h_4^1 \end{bmatrix} \, , \, Q_y = \sigma^2 I_5 \tag{4.40}$$

Note that due to the elimination of h_1, the design matrix is indeed of full rank. The remaining heights are given the upperindex ^{1} to show that they are defined with respect to the fixing of the height of the first point. There are 5 observations and 3 unknowns. The redundancy is therefore equal to 2. The two condition equations can be identified as $E\{\underline{h}_{12} + \underline{h}_{23} + \underline{h}_{31}\} = 0$ and $E\{\underline{h}_{23} + \underline{h}_{34} + \underline{h}_{42}\} = 0$.

Detection: In order to test the above model, we will start with the overall model test. The general expression for the corresponding test statistic reads

$$\underline{T} = \hat{\underline{e}}^T Q_y^{-1} \hat{\underline{e}}$$

When applied to the above model, it becomes (verify yourself)

$$\underline{T} = \frac{1}{8\sigma^2} \begin{bmatrix} \underline{h}_{12} + \underline{h}_{23} + \underline{h}_{31} \\ \underline{h}_{23} + \underline{h}_{34} + \underline{h}_{42} \end{bmatrix}^T \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} \underline{h}_{12} + \underline{h}_{23} + \underline{h}_{31} \\ \underline{h}_{23} + \underline{h}_{34} + \underline{h}_{42} \end{bmatrix} \tag{4.41}$$

Since the redundancy equals 2, this test statistic has a central Chi- squared distribution, with 2 degrees of freedom under the null hypothesis. Note that the overall model test statistic is a function of the misclosure of the levelling loop $1-2-3$ and of the levelling loop $2-3-4$.

Identification: In case identification of model errors is needed, we need to decide what type of model errors are likely to occur. It is standard practice to search for blunders in the individual observations. This is the data snooping approach. Since the variance matrix Q_y is diagonal, the w_i-test statistics for data snooping have the simple form

$$\underline{w}_i = \frac{\hat{\underline{e}}_i}{\sigma_{\hat{e}_i}} \qquad i = 1, \ldots, 5$$

They have a standard normal distribution under the null hypothesis. When the w_i-test statistic is worked out for the above model, we get (verify yourself)

$$\begin{array}{lcl} \underline{w}_1 & = & \frac{1}{2\sigma\sqrt{6}}(3\underline{h}_{12} + 2\underline{h}_{23} + 3\underline{h}_{31} - \underline{h}_{34} - \underline{h}_{41}) \\ \underline{w}_2 & = & \frac{1}{\sigma\sqrt{8}}(\underline{h}_{12} + 2\underline{h}_{23} + \underline{h}_{31} + \underline{h}_{34} + \underline{h}_{41}) \\ \underline{w}_3 & = & \underline{w}_1 \\ \underline{w}_4 & = & \frac{1}{2\sigma\sqrt{6}}(-\underline{h}_{12} + 2\underline{h}_{23} - \underline{h}_{31} + 3\underline{h}_{34} + 3\underline{h}_{41}) \\ \underline{w}_5 & = & \underline{w}_4 \end{array} \tag{4.42}$$

Note that some of the test statistics are identical. This is understandable if one considers the geometry of the levelling network. A blunder in $\underline{h}_{12}$ can not be discriminated from a blunder in $\underline{h}_{31}$ $(\underline{w}_3 = \underline{w}_1)$. Also, a blunder in $\underline{h}_{34}$ can not be discriminated from a blunder in $\underline{h}_{42}$ $(\underline{w}_5 = \underline{w}_4)$.

If in addition to potential blunders in the data, one also suspects that, say all three observed height differences of the levelling loop $1-2-3$ are erroneous by a constant amount, then also an additional identification test needs to be performed. In this case one has to make use of the general expression for the w-test statistic. It reads

$$\underline{w} = \frac{c^T Q_y^{-1} \hat{e}}{\sqrt{c^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c}}$$

The appropriate c-vector for testing whether a constant shift in the first three observations occurred or not, reads

$$c = (1, 1, 1, 0, 0)^T$$

When the expression of the w-test statistic is worked out for the above model, we get (verify yourself)

$$\underline{w} = \frac{\underline{h}_{12} + \underline{h}_{23} + \underline{h}_{31}}{\sqrt{3}\sigma} \tag{4.43}$$

Hence, the appropriate test statistic equals the misclosure of the levelling loop $1-2-3$, divided by its standard deviation.

Adaptation: In case data snooping lead to the conclusion that, say the second observation was erroneous, one can decide to remeasure this observation and after remeasurement, again apply the whole testing procedure. For the case that all first three observations

are off by a constant amount, one can of course also opt for remeasurement. But instead, one may also decide to adapt the model. In that case the new model becomes

$$E\{\begin{bmatrix} \underline{h}_{12} \\ \underline{h}_{23} \\ \underline{h}_{31} \\ \underline{h}_{34} \\ \underline{h}_{42} \end{bmatrix}\} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} h_2^1 \\ h_3^1 \\ h_4^1 \\ \nabla \end{bmatrix}, \; Q_y = \sigma^2 I_5 \tag{4.44}$$

One should be aware however, that this change also results in a change for precision and reliability. Both will become poorer and it depends on the particular application at hand whether one is willing to accept this or not.

4.3 The connected network

4.3.1 The observation equations

The purpose of the connection is to integrate the geometry of the free network with that of the network of existing control points and to express the result in the coordinate system of the control points. Since the coordinate system used for the free network adjustment, may differ from the one used for the control points, one could be dealing with two different coordinate systems. In order to show this difference, we will make use of the following notation. The Cartesian coordinates of the free network are denoted as (x, y) in case of two dimensions and as (x, y, z) in case of three dimensions. The coordinates of the network of control points are denoted as (u, v) or (u, v, w). The coordinates of all free network points are collected in the vector p and the coordinates of all control network points are collected in the vector q. For the two dimensional case, we thus have

$$\begin{aligned} p &= (\ldots, x_i, y_i, \ldots)^T \\ q &= (\ldots, u_i, v_i, \ldots)^T \end{aligned}$$

The free network and the control network will usually overlap. Hence, three type of points can be discriminated. The points that are part of the free network, but not of the control network. The points that are part of both the free network and the control network. And the points that are part of the control network, but not of the free network. The two sets of coordinates of the free network will be denoted as

$$\begin{aligned} p_1 &= (x_1, y_1, \ldots, x_{n_1}, y_{n_1})^T \\ p_2 &= (x_{n_1+1}, y_{n_1+1}, \ldots, x_{n_1+n_2}, y_{n_1+n_2})^T \end{aligned}$$

where p_1 contains the coordinates of the n_1 points of the free network that are not part of the overlap, and where p_2 contains the coordinates of the n_2 points of the free network that are part of the overlap. Thus the total number of points of the free network is assumed to be equal to $n_1 + n_2$. As a result of the free network adjustment, we thus have available

$$E\{\begin{bmatrix} \underline{p}_1 \\ \underline{p}_2 \end{bmatrix}\} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \; D\{\begin{bmatrix} \underline{p}_1 \\ \underline{p}_2 \end{bmatrix}\} = \begin{bmatrix} Q_{p_1} & Q_{p_1 p_2} \\ Q_{p_2 p_1} & Q_{p_2} \end{bmatrix} \tag{4.45}$$

The two sets of coordinates of the control network will be denoted as

$$\begin{aligned} q_2 &= (u_1, v_1, \ldots, u_{n_2}, v_{n_2})^T \\ q_3 &= (u_{n_2+1}, v_{n_2+1}, \ldots, x_{n_2+n_3}, y_{n_2+n_3})^T \end{aligned}$$

where q_2 contains the coordinates of the n_2 points of the control network that are part of the overlap, and where q_3 contains the coordinates of the n_3 points of the control network that are not part of the overlap. Thus the total number of points of the control network is assumed to be equal to $n_2 + n_3$. Of the control network, we assume to have available

$$E\{\begin{bmatrix} \underline{q}_2 \\ \underline{q}_3 \end{bmatrix}\} = \begin{bmatrix} q_2 \\ q_3 \end{bmatrix} \;,\; D\{\begin{bmatrix} \underline{q}_2 \\ \underline{q}_3 \end{bmatrix}\} = \begin{bmatrix} Q_{q_2} & Q_{q_2 q_3} \\ Q_{q_3 q_2} & Q_{q_3} \end{bmatrix} \tag{4.46}$$

In order to be able to combine (4.45) with (4.46), we still need to consider the coordinate transformation between the two coordinate systems. Although the type of coordinate transformation that is needed, depends on the particular application at hand, any coordinate transformation will be of the general form $p = F(q,t)$, where t denotes the vector of transformation parameters. When applied to the two sets p_1 and p_2, we thus have

$$\begin{aligned} p_1 &= F_1(q_1, t) \\ p_2 &= F_2(q_2, t) \end{aligned} \tag{4.47}$$

With (4.45), (4.46) and (4.47), we are now in the position to formulate the model of observation equations for the connection adjustment. It reads

$$E\{\begin{bmatrix} \underline{p}_1 \\ \underline{p}_2 \\ \underline{q}_2 \\ \underline{q}_3 \end{bmatrix}\} = \begin{bmatrix} F_1(q_1,t) \\ F_2(q_2,t) \\ q_2 \\ q_3 \end{bmatrix} \;,\; D\{\begin{bmatrix} \underline{p}_1 \\ \underline{p}_2 \\ \underline{q}_2 \\ \underline{q}_3 \end{bmatrix}\} = \begin{bmatrix} Q_{p_1} & Q_{p_1 p_2} & 0 & 0 \\ Q_{p_2 p_1} & Q_{p_2} & 0 & 0 \\ 0 & 0 & Q_{q_2} & Q_{q_2 q_3} \\ 0 & 0 & Q_{q_3 q_2} & Q_{q_3} \end{bmatrix} \tag{4.48}$$

Note that the observation equations are linear in q_3, but possibly nonlinear in q_1, q_2 and t. Whether the observation equations are nonlinear or not, depends on the type of coordinate transformation used. In case of nonlinear observation equations, we still need to apply a linearization in order to obtain linear(ized) observation equations. Linearization of (4.48) gives

$$E\{\begin{bmatrix} \Delta\underline{p}_1 \\ \Delta\underline{p}_2 \\ \Delta\underline{q}_2 \\ \Delta\underline{q}_3 \end{bmatrix}\} = \begin{bmatrix} T_1 & 0 & 0 & A_1 \\ 0 & T_2 & 0 & A_2 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix} \begin{bmatrix} \Delta q_1 \\ \Delta q_2 \\ \Delta q_3 \\ \Delta t \end{bmatrix} \tag{4.49}$$

where

$$\begin{aligned} T_1 &= \partial_{q_1} F_1(q_1^o, t^o) & A_1 &= \partial_t F_1(q_1^o, t^o) \\ T_2 &= \partial_{q_2} F_2(q_2^o, t^o) & A_2 &= \partial_t F_2(q_2^o, t^o) \end{aligned} \tag{4.50}$$

The structure of these four matrices of course also depends on the type of coordinate transformation used. As an example, we will show how they look like when the three dimensional similarity transformation is used.

The three dimensional similarity transformation between two sets of Cartesian coordinates reads

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \lambda\, R(\alpha)R(\beta)R(\gamma) \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{4.51}$$

with the scale λ, the translation vector $(t_x, t_y, t_z)^T$ and the three rotation matrices

$$R(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix}, R(\beta) = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix}, R(\gamma) = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

After linearization, the four matrices of (4.50) read

$$\begin{aligned} {}^{T_1}_{3n_1\times 3n_1} &= \begin{bmatrix} \lambda^o R^o & & \\ & \ddots & \\ & & \lambda^o R^o \end{bmatrix}, \quad {}^{T_2}_{3n_2\times 3n_2} = \begin{bmatrix} \lambda^o R^o & & \\ & \ddots & \\ & & \lambda^o R^o \end{bmatrix} \\ {}^{A_1}_{3n_1\times 7} &= \begin{bmatrix} \lambda^o R^o W_1^o & I_3 \\ \vdots & \vdots \\ \lambda^o R^o W_{3n_1}^o & I_3 \end{bmatrix}, \quad {}^{A_2}_{3n_2\times 7} = \begin{bmatrix} \lambda^o R^o W_{3n_1+1}^o & I_3 \\ \vdots & \vdots \\ \lambda^o R^o W_{3n_1+3n_2}^o & I_3 \end{bmatrix} \end{aligned} \tag{4.52}$$

with $R^o = R(\alpha^o)R(\beta^o)R(\gamma^o)$, $\Delta t = (\Delta\ln\lambda, \Delta\alpha, \Delta\beta, \Delta\gamma, \Delta t_x, \Delta t_y, \Delta t_z)^T$ and

$$W_i^o = \begin{bmatrix} u_i^o & -w_i^o\cos\beta^o\sin\gamma^o - v_i^o\sin\beta^o & -w_i^o\cos\gamma^o & v_i^o \\ v_i^o & w_i^o\cos\beta^o\cos\gamma^o + u_i^o\sin\beta^o & -w_i^o\sin\gamma^o & -u_i^o \\ w_i^o & -v_i^o\cos\beta^o\cos\gamma^o + u_i^o\cos\beta^o\sin\gamma^o & v_i^o\sin\gamma^o + u_i^o\cos\gamma^o & 0 \end{bmatrix}$$

In case of the two dimensional similarity transformation, the matrices of (4.52) will have the same structure, be it that now $R^o = R(\alpha^o)$, $\Delta t = (\Delta\ln\lambda, \Delta\alpha, \Delta t_x, \Delta t_y)^T$ and

$$W_i^o = \begin{bmatrix} u_i^o & -v_i^o \\ v_i^o & u_i^o \end{bmatrix}$$

We conclude this section with some remarks.

Remark 1: In the above linearization, no particular assumptions were made about the values of the approximate transformation parameters. In some applications it may happen however, that the two coordinate systems differ only slightly in scale and orientation. In that case one can choose as approximate values $\lambda^o = 1$, $\alpha^o = \beta^o = \gamma^o = 0$. Note that this results in a considerable simplification of the above matrices. We then have $\lambda^o R^o = I_3$ and

$$W_i^o = \begin{bmatrix} u_i^o & 0 & -w_i^o & v_i^o \\ v_i^o & w_i^o & 0 & -u_i^o \\ w_i^o & -v_i^o & u_i^o & 0 \end{bmatrix}$$

Remark 2: When the similarity transformation is used in the connection model (4.48) with all its transformation parameters included, one should recognize that the scale, the orientation and the location of the free network is abandoned in favour of the scale, the orientation and the location of the control network. Thus only the shape of the free network will then contribute to the determination of the geometry of the connected network.

Remark 3: Above it was assumed that all transformation parameters of the similarity transformation are unknown. In some applications it may happen however, that all or some of these transformation parameters are known. For instance, it may happen that one knowns (from an earlier adjustment) the relative orientation and scale between the WGS84 coordinate system and the National coordinate system. This would imply that if a free GPS network is expressed in the WGS84 system and the coordinates of the control network in the National coordinate system, that the scale and three orientation parameters are not needed as unknowns in the model of observation equations. If they are known with a sufficient precision, they could be treated as constants. The only remaining transformation parameters would then be the three translation parameters.

Remark 4: Above it was assumed that the coordinates of the control network are of the Cartesian type. This need not be the case of course. The model is easily adapted however, for the case that one uses other than Cartesian coordinates. One only needs to substitute in the above nonlinear model of observation equations, the relation that exists between Cartesian coordinates and the type of coordinates that one needs to use.

Remark 5: In the model (4.48), it was assumed that no correlation exists between the coordinates of the free network and the coordinates of the control network. For most practical applications this assumption is realistic, since the measurement processes that produced the coordinates of the two networks can usually be assumed to be independent.

Remark 6: In the model (4.48) it was also assumed that the variance matrix of the coordinates of the control network, Q_q, is available. In practice this may not be the case however. Fortunately, for the constrained connection adjustment it is not needed. In that case the adjustment is performed with $Q_q = 0$. For the statistical testing of the model (4.48), it is needed however. In that case one will have to work with a substitute 'variance matrix', that then hopefully will give a sufficiently good approximation to the actual variance matrix Q_q.

Remark 7: In the connection model (4.48) we included the coordinates of the nonoverlapping points of the control network, q_3. In the remaining part of these lecture notes, they will be disregarded however. These coordinates do not contribute to the redundancy of the model and therefore also not to the results of statistical testing, and they remain unchanged when a constrained connection adjustment with $Q_q = 0$ is applied.

4.3.2 The unconstrained connection for testing

In this section we will discuss ways of testing the validity of the connection model (4.48). It will be clear that the coordinates of the nonoverlapping points of the two networks do not contribute to the redundancy of the model and therefore also not to the results of the statistical tests. Hence, we can restrict our attention to the linear(ized) model

$$E\{\begin{bmatrix} \Delta\underline{p}_2 \\ \Delta\underline{q}_2 \end{bmatrix}\} = \begin{bmatrix} T_2 & A_2 \\ I & 0 \end{bmatrix}\begin{bmatrix} \Delta q_2 \\ \Delta t \end{bmatrix}, \; D\{\begin{bmatrix} \Delta\underline{p}_2 \\ \Delta\underline{q}_2 \end{bmatrix}\} = \begin{bmatrix} Q_{p_2} & 0 \\ 0 & Q_{q_2} \end{bmatrix} \quad (4.53)$$

The redundancy of this model equals $(2n_2 - n_t)$ in two dimensions and $(3n_2 - n_t)$ in three dimensions, where n_t is the dimension of the vector of transformation parameters.

Thus with the full similarity transformation in two dimensions, we have a redundancy of $(2n_2 - 4)$. This shows that in the overlap of the two networks, a minimum of two points is needed.

We can reduce the number of unknown parameters in this model, if we eliminate Δq_2 by making use of the coordinate differences $\Delta \underline{d} = \Delta \underline{p}_2 - T_2 \Delta \underline{q}_2$. This gives

$$H_o: \; E\{\Delta \underline{d}\} = A_2 \Delta t \;\; , \;\; D\{\Delta \underline{d}\} = Q_{p_2} + T_2 Q_{q_2} T_2^T \tag{4.54}$$

This model will be our *null hypothesis* H_o. As *alternative hypothesis* H_a, we will consider the model

$$H_a: \; E\{\Delta \underline{d}\} = \begin{bmatrix} A_2 & C \end{bmatrix} \begin{bmatrix} \Delta t \\ \nabla \end{bmatrix} \;\; , \;\; D\{\Delta \underline{d}\} = Q_{p_2} + T_2 Q_{q_2} T_2^T \tag{4.55}$$

where ∇ is an $r \times 1$ vector of assumed model errors and C is a matrix that specifies how the model errors are related to the observations. From our chapter on statistical testing, we know that the above null hypothesis can be tested against the above alternative hypothesis, using the test statistic

$$\underline{T}_r = \underline{\hat{e}}_d^T Q_d^{-1} C (C^T Q_d^{-1} Q_{\hat{e}_d} Q_d^{-1} C)^{-1} C^T Q_d^{-1} \underline{\hat{e}}_d \tag{4.56}$$

where $\underline{\hat{e}}_d$ is the least-squares residual vector of $\Delta \underline{d}$ and $Q_{\hat{e}_d}$ its variance matrix. Assuming normally distributed observables, one will decide to reject H_o on the basis of H_a, when

$$T_r > \chi^2_\alpha(r, 0)$$

with $\chi^2_\alpha(r,0)$ the critical value of the central Chi- square distribution, having r degrees of freedom.

Usually one will start with the most relaxed alternative hypothesis. This is the hypothesis for which the matrix (A_2, C) is square and regular. The corresponding test statistic reads

$$\underline{T} = \underline{\hat{e}}_d^T Q_d^{-1} \underline{\hat{e}}_d$$

Under the null hypothesis H_o, it has a central Chi-square distribution with the degrees of freedom being equal to the redundancy of the model under H_o. The test corresponding to the most relaxed alternative hypothesis, is usually referred to as the *overall model* test and its purpose is to *detect* unspecified model errors.

When the overall model test leads to a rejection of the null hypothesis, there is a high likelihood that the model under H_o has been specified incorrectly. Potential model errors are: (1) the assumed relation between the two coordinate systems is wrong; (2) the shape of the free network differs significantly from the shape of the control network, due to, for instance, an error in one or more of the coordinates; (3) the assumptions about the stochastic model are incorrect, due to, for instance, a too optimistically specified variance matrix of the coordinates.

In the following we will restrict ourselves to model errors in the functional model. They can be specified by means of the matrix C. We will now discuss some examples.

Errors in the coordinates: Since the coordinates form the observations in the connection model, it seems reasonable to first check the coordinates on possible errors. First note that one will never be able to find errors in the coordinates of the points in the two *non*overlapping parts of the two networks. These coordinates do not contribute to the redundancy of the model and will therefore not be part of the test statistics. Fortunately, these errors will also have no influence on the adjusted coordinates of the other points after the connection.

Secondly note, that one will also not be able to discriminate whether an error has occurred in the coordinates of the free network or in the coordinates of the control network. This is due to the fact that the observations in the model (4.55) are formed from coordinate *differences*.

In order to check for a potential error in the coordinates, the simplest alternative hypothesis is the one for which it is assumed that only one coordinate of one of the points is wrong. In this case, ∇ is a scalar $(r = 1)$ and matrix C becomes a vector, which will be denoted by the lowercase character c. Hence, instead of using (4.56) one can in this case also use its square-root, being the w-test statistic

$$\underline{w} = \frac{c^T Q_d^{-1} \underline{\hat{e}}_d}{\sqrt{c^T Q_d^{-1} Q_{\hat{e}} Q_d^{-1} c}}$$

For the three dimensional case, the c-vector then takes one of the following three forms

$$\begin{aligned} c_{x_i} &= (0\,0\,0 \ldots 1\,0\,0 \ldots 0\,0\,0)^T \\ c_{y_i} &= (0\,0\,0 \ldots 0\,1\,0 \ldots 0\,0\,0)^T \\ c_{z_i} &= (0\,0\,0 \ldots 0\,0\,1 \ldots 0\,0\,0)^T \end{aligned}$$

where i refers to the point being tested. In this way one can systematically check all coordinates of all overlapping points on potential errors.

Point identification errors: An error in only one coordinate of a point may occur due to typing errors. But in other cases of course, it is more likely, when an error occurs, that it will not be confined to a single coordinate only, but instead will effect all coordinates of a point. Such a type of error occurs when one erroneously beliefs that the free network coordinates and the control network coordinates indeed refer to the same point. These type of errors are called *point identification errors*.

In order to test for these type of errors, the C-matrix takes the form

$$C_i = \begin{bmatrix} 0 & 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 & 1 & \ldots & 0 & 0 & 0 \end{bmatrix}^T$$

Height or eccentricity error: In case of GPS, it not seldom occurs that an error is made in the measured GPS antenna height. The c-vector then takes the form

$$c_i = \begin{bmatrix} 0 & \ldots & 0 & h_{x_i} & h_{y_i} & h_{z_i} & 0 & \ldots & 0 \end{bmatrix}^T$$

where the vector $(h_{x_i}, h_{y_i}, h_{z_i})^T$ specifies the direction of the potential antenna height offset of point i in the approximate coordinate system. In case of an eccentricity error, the C-matrix has two columns,

$$C_i = \begin{bmatrix} 0 & \dots & 0 & e_{x_i} & e_{y_i} & e_{z_i} & 0 & \dots & 0 \\ 0 & \dots & 0 & f_{x_i} & f_{y_i} & f_{z_i} & 0 & \dots & 0 \end{bmatrix}^T$$

where the two vectors $(e_{x_i}, e_{y_i}, e_{z_i})^T$ and $(f_{x_i}, f_{y_i}, f_{z_i})^T$ span the plane in which the eccentricity error of point i is supposed to have taken place.

Misspecified coordinate transformation: Instead of having errors in the data, it may also happen that one has misspecified the coordinate transformation between the two coordinate systems. As an example consider a free GPS network that needs to be connected to a control network. Let us assume that the coordinates of the GPS network are expressed in the system of WGS84. Furthermore it is assumed that the coordinate system of the control network, has a scale and an orientation which is identical to the WGS84 system. The model under the null hypothesis reads then

$$H_o: \; E\{ \begin{bmatrix} \vdots \\ \Delta\underline{x}_i - \Delta\underline{u}_i \\ \Delta\underline{y}_i - \Delta\underline{v}_i \\ \Delta\underline{z}_i - \Delta\underline{w}_i \\ \vdots \end{bmatrix} \} = \begin{bmatrix} \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \Delta t_x \\ \Delta t_y \\ \Delta t_z \end{bmatrix}$$

Thus only the translation vector is included in the coordinate transformation. Now assume that the null hypothesis gets rejected by the overall model test and that one suspects that the two coordinate systems, apart from differing in their location, differ also slightly in scale. In that case one will have to oppose the above null hypothesis to the following alternative hypothesis

$$H_a: \; E\{ \begin{bmatrix} \vdots \\ \Delta\underline{x}_i - \Delta\underline{u}_i \\ \Delta\underline{y}_i - \Delta\underline{v}_i \\ \Delta\underline{z}_i - \Delta\underline{w}_i \\ \vdots \end{bmatrix} \} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & u_i^o \\ 0 & 1 & 0 & v_i^o \\ 0 & 0 & 1 & w_i^o \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \Delta t_x \\ \Delta t_y \\ \Delta t_z \\ \Delta \ln\lambda \end{bmatrix}$$

The model is thus enlarged with one additional column, being the column which models the scale difference between the two coordinate systems. This column vector is thus the appropriate choice for the c-vector

$$c = (\dots, u_i^o, v_i^o, w_i^o, \dots)^T$$

Note that λ^o has been taken equal to one, since the difference in scale between the two coordinate systems, although potentially significant, was still assumed to be small.

4.4 The constrained connection for coordinate computation

In the previous section the *validation* of the connection model was considered. In the present section we will consider the *estimation* problem. For validation, we could restrict our attention to the coordinates of the overlapping points. For estimation however, we need to take the coordinates of the nonoverlapping points of the free network into account as well. After all, the purpose of the connection is to obtain the coordinates of these newly established points in the coordinate system of the control network. Thus instead of the model (4.54), we will now work with the enlarged model

$$E\{\begin{bmatrix} \Delta\underline{p}_1 \\ \Delta\underline{d} \end{bmatrix}\} = \begin{bmatrix} T_1 & A_1 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} \Delta q_1 \\ \Delta t \end{bmatrix} \;,\; D\{\begin{bmatrix} \Delta\underline{p}_1 \\ \Delta\underline{d} \end{bmatrix}\} = \begin{bmatrix} Q_{p_1} & Q_{p_1p_2} \\ Q_{p_2p_1} & Q_d \end{bmatrix} \quad (4.57)$$

Remember that the coordinate differences of the overlapping points are contained in $\Delta\underline{d} = \Delta\underline{p}_2 - T_2\Delta\underline{q}_2$, which has as variance matrix, $Q_d = Q_{p_2} + T_2 Q_{q_2} T_2^T$.

The above model can be solved using the standard least-squares algorithm. As a result one would obtain the *unconstrained* least- squares solution. In practice however, circumstances often dictate that the coordinates of the control network remain unchanged. This can be accomplished by applying the standard least-squares algorithm to the above model, with the additional *constraints* that

$$Q_{q_2} := 0 \;\text{ and }\; \Delta q_2 := 0 \quad (4.58)$$

Setting Q_{q_2} to zero, implies that no least-squares correction is given to $\Delta\underline{q}_2$. This together with the setting of Δq_2 to zero, implies that the coordinates of the control points in the overlap remain fixed to their original values (*note:* the original values are used as approximate values in the linearization). In order to discriminate between the unconstrained and the constrained least-squares solution, we will use the superscript c for the constrained case. Thus $\Delta\underline{d}$ and $\Delta\underline{d}^c$ have the same variance matrix, namely Q_d, but their sample values differ. They are given respectively as

$$\Delta d = \Delta p_2 - T_2\Delta q_2 \;\text{ and }\; \Delta d^c = \Delta p_2$$

Instead of solving the above model in one step, we will solve it in three steps. This will help us in getting a somewhat better insight in the various aspects of the connection adjustment.

The first step: We already remarked that the coordinates of the nonoverlapping points of the free network do not contribute to the redundancy of the above model. Likewise, these coordinates also do not contribute to the determination of the transformation parameters. Only the coordinates of the overlapping points contribute to the determination of the transformation parameters. We therefore start by first considering

$$E\{\Delta\underline{d}\} = A_2\Delta t \;\;,\;\; D\{\Delta\underline{d}\} = Q_d \quad (4.59)$$

From it the *unconstrained* least-squares estimator of the vector of transformation parameters follows as

$$\Delta\underline{\hat{t}} = (A_2^T Q_d^{-1} A_2)^{-1} A_2^T Q_d^{-1} \Delta\underline{d} \;\;,\;\; Q_{\hat{t}} = (A_2^T Q_d^{-1} A_2)^{-1} \quad (4.60)$$

The *constrained* least-squares estimator however, reads

$$\begin{cases} \Delta\hat{\underline{t}}^c &= (A_2^T Q_{p_2}^{-1} A_2)^{-1} A_2^T Q_{p_2}^{-1} \Delta\underline{d}^c \\ Q_{\hat{t}^c} &= (A_2^T Q_{p_2}^{-1} A_2)^{-1} A_2^T Q_{p_2}^{-1} Q_d Q_{p_2}^{-1} A_2 (A_2^T Q_{p_2}^{-1} A_2)^{-1} \end{cases} \tag{4.61}$$

Note that in the application of the error propagation law, the randomness of $\Delta\underline{q}_2$ has been taken into account. After all, although we constrain the solution to the coordinates of the control network, these coordinates are still of a stochastic nature. We know that a least-squares solution that takes the stochasticity of all variables into account will give estimators that are of minimal variance. In the above constrained least-squares solution this is not the case. Thus

$$Q_{\hat{t}} < Q_{\hat{t}^c}$$

showing that the transformation parameters of the unconstrained solution are of a better precision than the ones of the constrained solution. This is thus the price one pays, when one is forced to keep the coordinates of the control network unchanged.

The unconstrained solution for the coordinate differences and their least-squares residuals are given as

$$\begin{array}{lll} \Delta\hat{\underline{d}} = A_2 \Delta\hat{\underline{t}} & , & Q_{\hat{d}} = A_2 Q_{\hat{t}} A_2^T \\ \hat{\underline{e}}_d = \Delta\underline{d} - \Delta\hat{\underline{d}} & , & Q_{\hat{e}_d} = Q_d - A_2 Q_{\hat{t}} A_2^T \end{array} \tag{4.62}$$

As we have seen in the previous section, these least-squares residuals form the basis for the statistical testing of the validity of the connection model.

The second step: In this second step we will determine the solution for the coordinates of the free network in the nonoverlapping part. We know that $\Delta\underline{p}_1$ does not contribute to the redundancy of the model. If in addition, $\Delta\underline{p}_1$ would not correlate with $\Delta\underline{p}_2$ and thus not with $\Delta\underline{d}$, then $\Delta\underline{p}_1 = \Delta\hat{\underline{p}}_2$ would hold true. That is, in that case the coordinates of the free network in the nonoverlapping part would not change as well. But since $Q_{p_1 p_2} \neq 0$, it follows that these coordinates do change. In fact, their residual vector can be computed from $\hat{\underline{e}}_d$ as

$$\hat{\underline{e}}_{p_2} = Q_{p_1 p_2} Q_d^{-1} \hat{\underline{e}}_d$$

Hence, the *unconstrained* least-squares solution reads

$$\begin{array}{lll} \Delta\hat{\underline{p}}_1 &=& \Delta\underline{p}_1 - Q_{p_1 p_2} Q_d^{-1} (\Delta\underline{d} - \Delta\hat{\underline{d}}) \\ Q_{\hat{p}_1} &=& Q_{p_1} - Q_{p_1 p_2} Q_d^{-1} (Q_d - Q_{\hat{d}}) Q_d^{-1} Q_{p_2 p_1} \end{array} \tag{4.63}$$

Of course if the connection model fails to have any redundancy at all, then $\hat{\underline{e}}_d = 0$ and $\Delta\hat{\underline{p}}_1 = \Delta\underline{p}_1$. This would for instance happen when in the two dimensional case, the connection model is based on the full similarity transformation and the number of points in the overlap equals only two.

The *constrained* least-squares solution reads

$$\Delta\hat{\underline{p}}_1^c = \Delta\underline{p}_1 - Q_{p_1 p_2} Q_{p_2}^{-1} (\Delta\underline{d}^c - A_2 \Delta\hat{\underline{t}}^c) \tag{4.64}$$

and again we have $Q_{\hat{p}_1^c} > Q_{\hat{p}_1}$.

The above two solutions are still expressed in the coordinate system of the free network. Hence, in order to express them in the coordinate system of the control network, we still need a third step.

The third step: After the unconstrained adjustment of the connection model (4.57), we have

$$\Delta\underline{\hat{p}}_1 = T_1\Delta\underline{\hat{q}}_1 + A_1\Delta\underline{\hat{t}}$$

Since matrix T_1 is invertible, it follows after substitution of (4.60) and (4.63), that

$$\Delta\underline{\hat{q}}_1 = T_1^{-1}\left[\begin{array}{cc} I & -Q_{p_1p_2}Q_d^{-1}\end{array}\right]\left[\left[\begin{array}{c}\Delta\underline{p}_1\\ \Delta\underline{d}\end{array}\right] - \left[\begin{array}{c}A_1\\ A_2\end{array}\right]\Delta\underline{\hat{t}}\right] \tag{4.65}$$

This is the final *unconstrained* least-squares solution of the coordinates of the nonoverlapping points of the free network, expressed in the coordinate system of the control network. This is thus the solution one gets for the coordinates of the nonoverlapping points, when the model (4.57) is solved using the standard least- squares algorithm.

In case of a constrained adjustment of the connection model (4.57) we have

$$\Delta\underline{\hat{p}}_1^c = T_1\Delta\underline{\hat{q}}_1^c + A_1\Delta\underline{\hat{t}}^c$$

From substitution of (4.61) and (4.64), follows then

$$\Delta\underline{\hat{q}}_1^c = T_1^{-1}\left[\begin{array}{cc} I & -Q_{p_1p_2}Q_{p_2}^{-1}\end{array}\right]\left[\left[\begin{array}{c}\Delta\underline{p}_1\\ \Delta\underline{d}^c\end{array}\right] - \left[\begin{array}{c}A_1\\ A_2\end{array}\right]\Delta\underline{\hat{t}}^c\right] \tag{4.66}$$

This is the final *constrained* least-squares solution of the coordinates of the nonoverlapping points of the free network, expressed in the coordinate system of the control network. Hence, this is the solution of the newly established points, when one requires that the coordinates of the control network remain unchanged.

The above expression is the most general expression one can get. Let us now consider three special cases.

No correlation: In the special case that the coordinates of $\Delta\underline{p}_1$ do not correlate with those of $\Delta\underline{p}_2$, we have $Q_{p_1p_2} = 0$. The above expression simplifies then to

$$\Delta\underline{\hat{q}}_1^c = T_1^{-1}(\Delta\underline{p}_1 - A_1\Delta\underline{\hat{t}}^c)$$

In this case the transformation $A_1\Delta\underline{\hat{t}}^c$ is directly applied to $\Delta\underline{p}_1$, without taking the vector of least-squares residuals $\underline{\hat{e}}_d^c = \Delta\underline{d}^c - \Delta\underline{\hat{d}}^c$ into account.

No redundancy: The above simplified expression is also obtained in case there is no redundancy in the connection model. An absence of redundancy implies that the points in the overlap have as many coordinates as there are transformation parameters. Hence, there is no need for an adjustment to determine the transformation parameters. They are uniquely determined from the coordinates of the points in the overlap and the least-squares residual vector $\underline{\hat{e}}_d^c$ is identically zero.

No coordinate transformation: In case the two coordinate systems of the free network and the control network coincide, the transformation parameters will be absent from the model and T_1 will be equal to the identity matrix. In that case the above constrained solution simplifies to

$$\Delta\hat{\underline{q}}_1^c = \Delta\underline{p}_1 - Q_{p_1p_2}Q_{p_2}^{-1}\Delta\underline{d}^c \tag{4.67}$$

The correction to the coordinates of $\Delta\underline{p}_1$ is now only based on the difference between the coordinates of the points in the overlap.

We already remarked that the precision of the constrained solution is less than that of the unconstrained solution, $Q_{\hat{q}_1^c} > Q_{\hat{q}_1}$. This need not be dramatic. It is simply a consequence of the fact that by constraining the solution, one is in fact using a less than optimal estimator. But what is important to recognize though, is that the constrained least- squares solution can be *poorer* than the solution one had before the connection was carried out. This is easily shown for the above solution (4.67). Application of the error propagation law to (4.67) gives

$$Q_{\hat{q}_1^c} = Q_{p_1} - Q_{p_1p_2}Q_{p_2}^{-1}(Q_{p_2} - Q_{q_2})Q_{p_2}^{-1}Q_{p_2p_1}$$

This shows that

$$Q_{\hat{q}_1^c} > Q_{p_1} \text{ when } Q_{p_2} < Q_{q_2} \tag{4.68}$$

But this means, if the coordinates of the control network are of a poorer precision than the coordinates of the free network ($Q_{p_2} < Q_{q_2}$), that the adjusted coordinates, when constraining is applied, will have a precision which is poorer than before the adjustment was carried out ($Q_{\hat{q}_1^c} > Q_{p_1}$). This is the reason, why in surveying one works from the 'large to the small'. That is, in the densification process, the free networks are connected to 'higher order' control networks, i.e. networks that are of a better precision than the free networks.

4.4.1 A levelling example

To conclude this chapter, we will give an example of testing a connection model. The adjustment is therefore of the unconstrained type. The constrained adjustment is rather straightforward and is left to the reader. The connection model is one where two levelling networks need to be connected. The model is assumed to read

$$E\{\begin{bmatrix} \underline{h}_1 \\ \vdots \\ \underline{h}_n \\ \underline{H}_1 \\ \vdots \\ \underline{H}_n \end{bmatrix}\} = \begin{bmatrix} 1 & & & 1 \\ & \ddots & & \vdots \\ & & 1 & 1 \\ 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} H_1 \\ \vdots \\ H_n \\ t \end{bmatrix}, \begin{bmatrix} Q_h & Q_{hH} \\ Q_{Hh} & Q_H \end{bmatrix} = \begin{bmatrix} \sigma_h^2 I_n & 0 \\ 0 & \sigma_H^2 I_n \end{bmatrix} \tag{4.69}$$

The two overlapping levelling networks are thus assumed to differ in height only. The redundancy of the model equals $(n-1)$. First we will give the least-squares estimators of H_i and t, and then we will discuss the detection and identification step.

Least-squares estimators: After forming the system of normal equations of the above model and solving it for the translation parameter, we get

$$\begin{cases} \hat{\underline{t}} &= \bar{\underline{h}} - \bar{\underline{H}} \\ \sigma^2_{\hat{t}} &= \frac{1}{n}(\sigma^2_h + \sigma^2_H) \end{cases} \tag{4.70}$$

where

$$\bar{\underline{h}} = \frac{1}{n}\sum_{i=1}^{n} \underline{h}_i \ , \ \bar{\underline{H}} = \frac{1}{n}\sum_{i=1}^{n} \underline{H}_i$$

are the average heights of the two networks. In this particular case, the least-squares estimator of the translation parameter thus equals the difference of the height averages of the two networks.

When we solve the system of normal equations for the height coordinate H_i of point i, we get

$$\begin{cases} \hat{\underline{H}}_i &= \underline{H}_i - \frac{\sigma^2_H}{\sigma^2_h+\sigma^2_H}[(\underline{H}_i - \bar{\underline{H}}) - (\underline{h}_i - \bar{\underline{h}})] \\ \sigma^2_{\hat{H}_i} &= \sigma^2_H[1 - \frac{\sigma^2_H}{\sigma^2_h+\sigma^2_H}(1-\frac{1}{n})] \end{cases} \tag{4.71}$$

This shows that the least-squares residual of $\underline{H}_i$ is given as

$$\hat{\underline{e}}_{H_i} = \frac{\sigma^2_H}{\sigma^2_h+\sigma^2_H}[(\underline{H}_i - \bar{\underline{H}}) - (\underline{h}_i - \bar{\underline{h}})]$$

It is constructed from a difference of two differences. The two differences are the height of point i with respect to the average height in the first height system and the height of point i with respect to the average height in the second height system. Note that the residual gets smaller, when σ^2_H gets smaller. This is understandable. The more precise the $\underline{H}_i$ coordinate is, the more confidence one has in it and the less one is willing to give it a large correction. In the limiting case $\sigma^2_H = 0$, we have $\hat{\underline{H}}_i = \underline{H}_i$. In that case no correction is applied at all. If we consider the other limiting case $\sigma^2_h = 0$, then of course the $\underline{h}_i$ coordinate would not get a correction, $\hat{\underline{h}}_i = \underline{h}_i$. In this case, we also would have

$$\begin{aligned} \hat{\underline{H}}_i &= \underline{h}_i + \bar{\underline{H}} - \bar{\underline{h}} \\ &= \underline{h}_i - \hat{\underline{t}} \end{aligned}$$

thus showing that $\hat{\underline{H}}_i$ is now obtained from simply applying the translation estimator to $\underline{h}_i$.

Detection: In order to verify whether it is justified to rely our results of adjustment on the above model (4.69), we need to test it. In order to be able to test it, the least-squares

residuals should at least not be identical to zero, which would happen when there is no redundancy. Since the redundancy equals $(n-1)$, we thus at least need more than one point in the two overlapping networks.

To test the above model, we first consider its overall validity. The general form of the appropriate test statistic for detection reads $\underline{T}_{redundancy} = \underline{\hat{e}}^T Q_y^{-1} \underline{\hat{e}}$. When this expression is worked out for the above model we get

$$\underline{T}_{n-1} = \frac{\sum_{i=1}^{n} (\underline{H}_i - \underline{h}_i)^2 - n(\bar{\underline{H}} - \bar{\underline{h}})^2}{\sigma_h^2 + \sigma_H^2} \tag{4.72}$$

Note that the term $n(\bar{\underline{H}} - \bar{\underline{h}})^2$ would be absent in case the translation parameter t would be absent from the model. In that case the test statistic would simply equal the sum of the squared height differences divided by the sum of the two variances. The redundancy would then also have increased by one to n.

With the above test statistic, the model is invalidated or rejected when

$$T_{n-1} \; > \; \chi_\alpha^2(n-1,0) \tag{4.73}$$

When rejection occurs, there are of course many type of misspecifications in the model that could have caused it. The misspecifications could be in the functional and/or the stochastic model. For instance, when the stochastic model is too optimistic, the variances are too small and the value of T_{n-1} would be unrealistically large. As a result, one could have an unjustified rejection of the model. This is therefore the reason, why the variance σ_H^2 is not set to zero, when testing the connection model. Thus the unconstrained connection is used for testing, while the constrained connection, with $\sigma_H^2 = 0$, is used for the actual coordinate computation.

Instead of having a too optimistic stochastic model, one could of course also have used a too pessimistic stochastic model. In that case the variances are too large and T_{n-1} would be unrealistically small. If one suspects this to be the case, then the above one-sided test should be replaced by its two-sided counterpart. Since the expectation of $\underline{T}_{n-1}/(n-1)$ equals one under the null hypothesis, one then tests whether this ratio is significantly larger or smaller than one. We will however assume that the stochastic model is correct and restrict our attention to misspecifications in the functional model.

Identification: If we assume that the rejection of the model could have been caused by an error in the coordinate set of the H-system, the c-vector of the w-test statistic will have the form

$$c = (0^T, c_H^T)^T \;\text{ with }\; c_H = (c_{H_1}, \ldots, c_{H_n})^T \tag{4.74}$$

When worked out, the corresponding w-test statistic reads then

$$\underline{w}_H = \frac{\sum_{i=1}^{n} c_{H_i} (\underline{H}_i - \underline{h}_i) - (\bar{\underline{H}} - \bar{\underline{h}}) \sum_{i=1}^{n} c_{H_i}}{\sqrt{(\sigma_h^2 + \sigma_H^2)(\sum_{i=1}^{n} c_{H_i}^2 - \frac{1}{n}(\sum_{i=1}^{2} c_{H_i})^2)}} \tag{4.75}$$

and the test would be

$$| \, w_H \, | > N_{\frac{1}{2}\alpha_1}(0,1) \tag{4.76}$$

where the level of significance α_1 is coupled to that of α by $\lambda(\alpha_1, 1, \gamma) = \lambda(\alpha, n-1, \gamma)$, so as to ensure that both tests (4.73) and (4.76) would have an equal probability γ of correct rejection.

If we want to test for an error in the coordinate set of the h-system instead of in the H-system, then one would need to choose instead of (4.74), the c-vector $c = (c_h^T, 0^T)^T$. But for $c_h = c_H$, the two test statistics $\underline{w}_h$ and $\underline{w}_H$ would become identical apart from a change in sign (verify yourself). This simply implies that one is not able to identify whether the model error originated from the h-system or from the H-system. This is much like the situation where an error in a sum or difference can not be unambiguously assigned to either components of the sum or difference. This is also the reason why it is of importance to have a well-tested free network before the connection is carried out. Because if then during the testing of the connection model, rejection occurs, one can confidently assume that the error is present in the coordinate set of the control network, in our case the coordinate set of the H-system.

Data snooping: If we want to screen the individual $\underline{H}_i$ coordinates for the presence of blunders, the c_H-vector takes the form

$$c_{H_i} = (0, \ldots, 0, 1, 0, \ldots, 0)^T$$

and the above $\underline{w}_H$ test statistic reduces to

$$\underline{w}_{H_i} = \frac{(\underline{H}_i - \underline{h}_i) - (\bar{\underline{H}} - \bar{\underline{h}})}{\sqrt{(\sigma_h^2 + \sigma_H^2)(1 - \frac{1}{n})}} \tag{4.77}$$

The size of the blunder that can be found with a probability $\gamma = 0.80$ using this test statistic with a level of significance $\alpha_1 = 0.001$, is given by the MDB,

$$| \nabla_i | = \sqrt{\frac{17.075(\sigma_h^2 + \sigma_H^2)}{1 - \frac{1}{n}}} \tag{4.78}$$

This is the MDB which we already met earlier in the section on Reliability.

Testing for scale: If one suspects that the model was rejected due to the erroneous assumption that scale is absent, the appropriate c- vector for testing for the presence of scale, is

$$c = (c_h^T, 0^T)^T \text{ with } c_h = (H_1^o, \ldots, H_n^o)^T$$

The size of the scale factor that then can be found with probability $\gamma = 0.80$, is given by the MDB

$$| \nabla_\lambda | = \sqrt{\frac{17.075(\sigma_h^2 + \sigma_H^2)}{\sum_{i=1}^n (H_i^o - \bar{H}_i^0)^2}} \tag{4.79}$$

This shows that the presence of scale become better identifiable, when $\sum_{i=1}^n (H_i^o - \bar{H}_i^0)^2$ is large, that is, when the network would have large height differences with respect to the average height.

4.5 Summary

As a conclusion to this chapter, we have summarized the main steps involved in four flow diagrams. In figure 4.2, the adjustment and testing steps are shown for both the free network and the connected network (*note:* the constrained connection is also referred to as a pseudo least-squares connection). The model used for the connection is usually

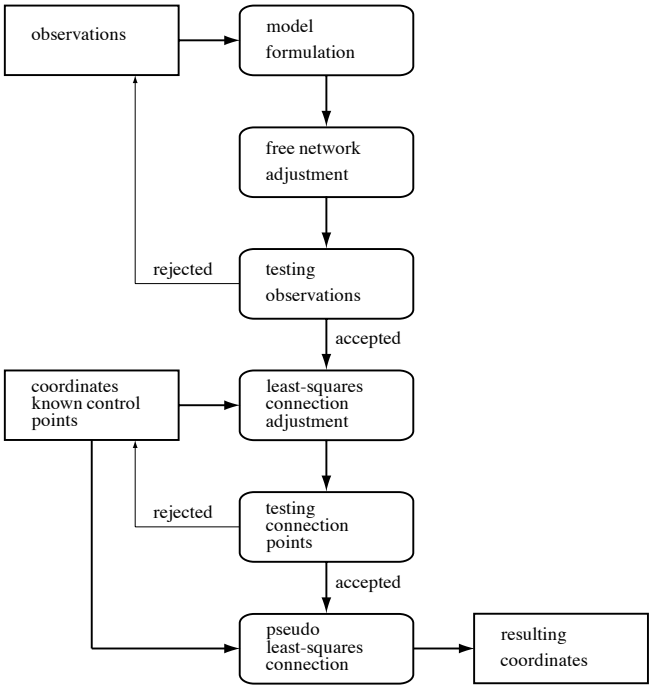


Figure 4.2 Flow diagram of free and connected network phase.

based on a coordinate transformation. After all, the coordinates of the free network may not be defined in the same coordinate system as those of the control network. In case of connecting a GPS-based free network to existing control, one is dealing with WGS84 coordinates and with the coordinates of the National Reference System (NatRS). The model for the connection can be formulated in two ways. Either one formulates it so that the results are expressed in the WGS84 coordinate system, or one formulates it so that the results are expressed in the NatRS coordinate system. The first formulation is shown in figure 4.3 and the second formulation is shown in figure 4.4. With the first formulation one of course will need an additional (back) transformation step to get the final results expressed in the NatRS coordinate system.

The various steps involved in the transformation between the two systems are shown in figure 4.5.

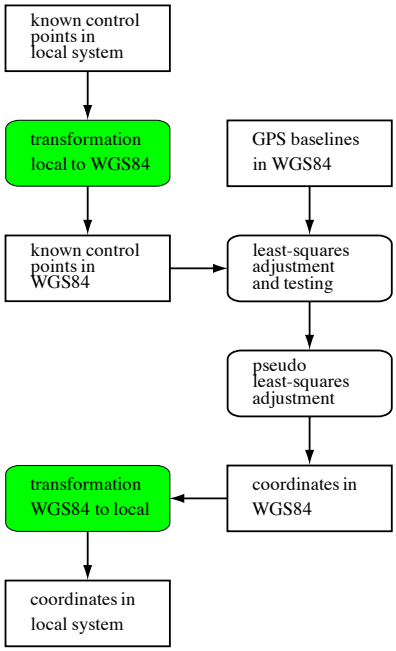


Figure 4.3 Connection adjustment and testing in WGS84 system.

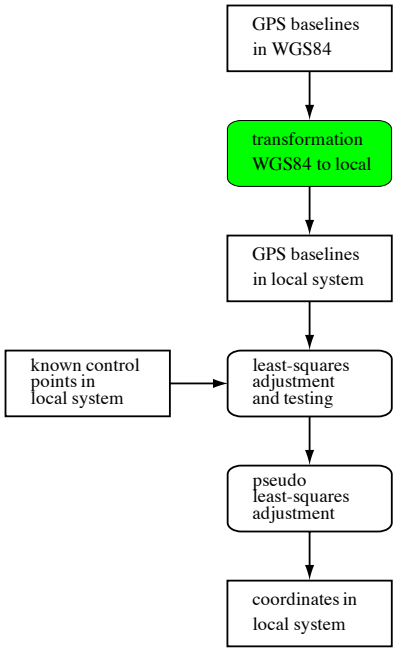


Figure 4.4 Connection adjustment and testing in NatRS or local system.

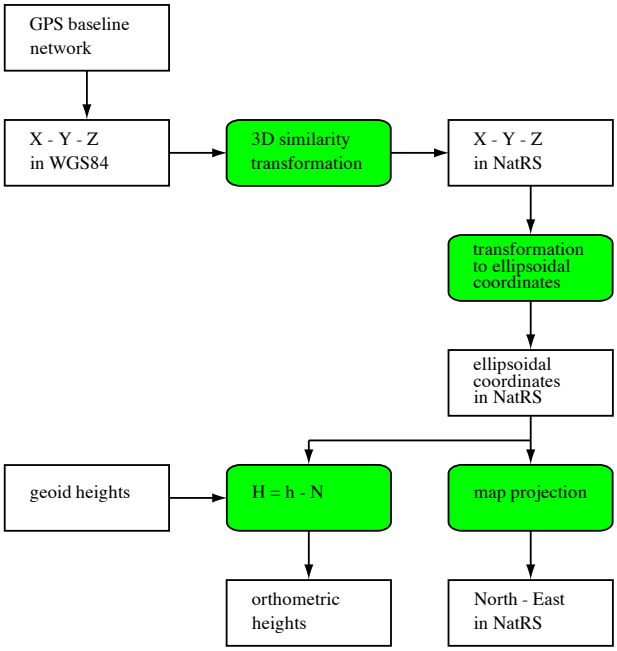


Figure 4.5 Datum transformation WGS84 - NatRS.

Appendix A

Appendix

In this appendix, we consider the first two moments of a random variable, the mean and the variance. We also show how they propagate when functions of the random variable are taken.

First we consider the mean of a scalar random variable and give the propagation law of the mean when the function is either linear or nonlinear. In the latter case, use is made of a linearization. Then we do the same for the variance of a scalar random variable. After the scalar case has been treated, we generalize to the vectorial case.

A.1 Mean and variance of scalar random variables

Mean of a scalar random variable: Let $\underline{x}$ be a continuous scalar random variable with probability density function $p_{\underline{x}}(x)$. The *expectation* or *mean* of x is by definition the integral

$$E\{\underline{x}\} = \int_{-\infty}^{+\infty} x p_{\underline{x}}(x) dx \tag{A.1}$$

This number will also be denoted by m_x.

Mean of a function of a scalar random variable: Given a scalar random variable $\underline{x}$ and a function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. As we see from (A.1), the mean of $\underline{y}$ is given by

$$E\{\underline{y}\} = \int_{-\infty}^{+\infty} y p_{\underline{y}}(y) dy \tag{A.2}$$

It appears therefore, that to determine the mean of $\underline{y}$, we must first find its probability density function $p_{\underline{y}}(y)$. This, however, is not necessary. As the next theorem shows, $E\{\underline{y}\}$ can be expressed directly in terms of the function $f(x)$ and the density $p_{\underline{x}}(x)$ of $\underline{x}$.

Theorem:

$$E\{f(\underline{x})\} = \int_{-\infty}^{+\infty} f(x) p_{\underline{x}}(x) dx \tag{A.3}$$

Proof: We shall sketch a proof using the curve $f(x)$ of figure A.1. With $y = f(x_1) = f(x_2) = f(x_3)$ as in the figure, we see that

$$\begin{aligned} &P(y \le \underline{y} \le y + dy) = \\ &= P(x_1 \le \underline{x} \le x_1 + dx_1) + P(x_2 - dx_2 \le \underline{x} \le x_2) + P(x_3 \le \underline{x} \le x_3 + dx_3) \end{aligned}$$

or

$$p_{\underline{y}}(y) dy = p_{\underline{x}}(x_1) dx_1 + p_{\underline{x}}(x_2) dx_2 + p_{\underline{x}}(x_3) dx_3$$

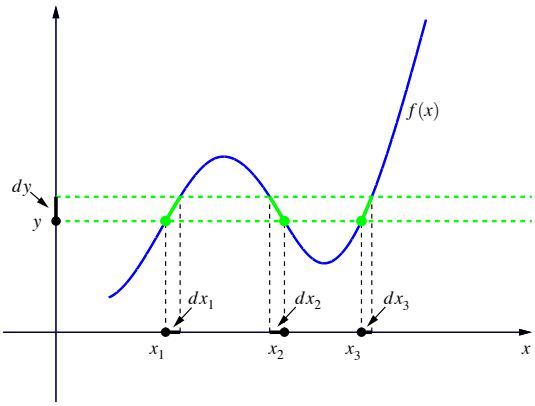


Figure A.1 The curve $y = f(x)$.

Multiplying by y, we obtain

$$y p_{\underline{y}}(y)dy = f(x_1)p_{\underline{x}}(x_1)dx_1 + f(x_2)p_{\underline{x}}(x_2)dx_2 + f(x_3)p_{\underline{x}}(x_3)dx_3$$

Thus, to each differential in (A.2) there corresponds one or more differentials in (A.3). As dy covers the y-axis, the corresponding dx's are nonoverlapping and they cover the entire x-axis. Hence, the integrals in (A.2) and (A.3) are equal.

It appears from (A.3) that to determine the mean of $\underline{y}$, we must know the probability density function $p_{\underline{x}}(x)$. In general this is true. However, in the special case that the function $f(x)$ is *linear* not the complete density of x needs to be known but only the mean m_x of $\underline{x}$.

Theorem (propagation law of the mean): Given a scalar random variable x and a function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. If the function $f(x)$ is linear,

$$f(x) = ax + b \tag{A.4}$$

then

$$m_y = am_x + b \tag{A.5}$$

Proof: Substitution of (A.4) into (A.3) gives

$$\begin{aligned} m_y &= \int_{-\infty}^{+\infty}(ax+b)p_{\underline{x}}(x)dx \\ &= a\int_{-\infty}^{+\infty} x\, p_{\underline{x}}(x)dx + b\int_{-\infty}^{+\infty} 1\, p_{\underline{x}}(x)dx \\ &= am_x + b \end{aligned}$$

If the function $f(x)$ is nonlinear we need strictly speaking the density $p_{\underline{x}}(x)$ of $\underline{x}$ in order to compute the mean of $\underline{y} = f(\underline{x})$. However, by using Taylor's formula an approximation to the mean of $\underline{y}$ can be derived that gets round the difficulty of having to know the density $p_{\underline{x}}(x)$ of $\underline{x}$.

Theorem (linearized propagation law of the mean): Given a scalar random variable $\underline{x}$ and a nonlinear function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. Let x^0 be an

approximation to a sample of $\underline{x}$ and define $\Delta\underline{y} = \underline{y} - f(x^0)$ and $\Delta\underline{x} = \underline{x} - x^0$. Then a first-order approximation to the mean of $\Delta\underline{y}$ is

$$m_{\Delta y} \doteq \frac{d}{dx} f(x^0) m_{\Delta x} \tag{A.6}$$

Proof: Application of Taylor's formula gives

$$\underline{y} = f(x^0) + \frac{d}{dx} f(x^0)(\underline{x} - x^0) + \frac{1}{2}\frac{d^2}{dx^2} f(x^0)(\underline{x} - x^0)^2 + \dots$$

If we take the expectation we get

$$E\{\underline{y}\} = f(x^0) + \frac{d}{dx} f(x^0) E\{(\underline{x} - x^0)\} + \frac{1}{2}\frac{d^2}{dx^2} f(x^0) E\{(\underline{x} - x^0)^2\} + \dots$$

This may be written with

$$m_{\Delta y} = E\{\underline{y}\} - f(x^0) \quad , \quad m_{\Delta x} = E\{(\underline{x} - x^0)\}$$

and

$$\begin{aligned} E\{(\underline{x} - x^0)^2\} &= E\{\left[(\underline{x} - m_x) - x^0 - m_x)\right]^2\} \\ &= E\{(\underline{x} - m_x)^2\} - 2E\{(\underline{x} - m_x)(x^0 - m_x)\} + e\{(x^0 - m_x)^2\} \\ &= E\{(\underline{x} - m_x)^2\} + m_{\Delta x}^2 \\ &= \sigma_x^2 + m_{\Delta x}^2 \end{aligned}$$

as

$$m_{\Delta y} = \frac{d}{dx} f(x^0) m_{\Delta x} + \frac{1}{2}\frac{d^2}{dx^2} f(x^0) \left(\sigma_x^2 + m_{\Delta x}^2\right) + \dots$$

If we neglect the second and higher order terms, the result (A.6) follows.

Examples

1. With $f(x) = \cos x$, we define the random variable $\underline{y} = f(\underline{x})$. Since $\frac{d}{dx} f(x^0) = -\sin x^0$, we find to a first order
$$E\{\underline{y} - \cos x^0\} \doteq -\sin x^0 E\{\underline{x} - x^0\}$$
2. With $f(x) = x^3$, we define the random variable $\underline{y} = f(\underline{x})$. Since $\frac{d}{dx} f(x^0) = 3(x^0)^2$ we find to a first order
$$E\{\underline{y} - (x^0)^3\} \doteq 3(x^0)^2 E\{\underline{x} - x^0\}$$

□

Variance of a scalar random variable: The *variance* of a scalar random variable is by definition the integral

$$E\{(\underline{x} - E\{\underline{x}\})^2\} = \int_{-\infty}^{+\infty} (x - m_x)^2 p_{\underline{x}}(x) dx \tag{A.7}$$

This number will also be denoted by σ_x^2. The positive constant σ_x is called the *standard deviation* of $\underline{x}$. From the definition it follows that

$$\sigma_x^2 = E\{(\underline{x} - m_x)^2\} = E\{\underline{x}^2\} - 2E\{\underline{x}\}m_x + m_x^2$$

or

$$\sigma_x^2 = E\{\underline{x}^2\} - m_x^2 \tag{A.8}$$

Theorem (propagation law of the variance): Given a scalar random variable $\underline{x}$ and a function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. If the function $f(x)$ is *linear*,

$$f(x) = ax + b \tag{A.9}$$

then

$$\sigma_y^2 = a^2\sigma_x^2 \tag{A.10}$$

Proof: According to definition (A.7) we have

$$\sigma_y^2 = E\{(\underline{y} - E\{\underline{y}\})^2\} = E\{(f(\underline{x}) - m_y)^2\}$$

With (A.3) this gives

$$\sigma_y^2 = \int_{-\infty}^{+\infty} (f(x) - m_y)^2 p_{\underline{x}}(x)dx$$

Substitution of (A.5) and (A.9) gives

$$\begin{array}{rcl} \sigma_y^2 & = & \int_{-\infty}^{+\infty} [(ax+b) - (am_x + b)]^2 \, p_{\underline{x}}(x)dx \\ & = & a^2 \int_{-\infty}^{+\infty} (x - m_x)^2 p_{\underline{x}}(x)dx \\ & = & a^2\sigma_x^2 \end{array}$$

The above result shows that if $f(x)$ is linear, knowledge of σ_x^2 is sufficient for computing the variance of $\underline{y} = f(\underline{x})$. For nonlinear functions $f(x)$ this is generally not true. If the function $f(x)$ is nonlinear one will generally need to know the complete density $p_{\underline{x}}(x)$ of $\underline{x}$. However, by using Taylor's formula an approximation to the variance of $\underline{y}$ can be derived that gets round the difficulty of having to know $p_{\underline{x}}(x)$.

Theorem (Linearized propagation law of the variance): Given a scalar random variable $\underline{x}$ and a nonlinear function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. Let x^0 be an approximation to a sample of $\underline{x}$. Then a first-order approximation to the variance of $\underline{y}$ is

$$\sigma_y^2 \doteq \left(\frac{d}{dx} f(x^0)\right)^2 \sigma_x^2 \tag{A.11}$$

Proof: Substitution of

$$\left\{ \begin{array}{rcl} f(x) & = & f(x^0) + \frac{d}{dx} f(x^0)(x - x^0) + \ldots \\ m_x & = & f(x^0) + \frac{d}{dx} f(x^0)(m_x - x^0) + \ldots \end{array} \right.$$

into

$$\sigma_y^2 = \int_{-\infty}^{+\infty} (f(x) - m_y)^2 \, p_{\underline{x}}(x) dx$$

gives after neglecting second and higher order terms

$$\begin{aligned} \sigma_y^2 &= \int_{-\infty}^{+\infty} \left(\tfrac{d}{dx} f(x^0)(x - m_x) \right)^2 p_{\underline{x}}(x) dx \\ &= \left(\tfrac{d}{dx} f(x^0) \right)^2 \int_{-\infty}^{+\infty} (x - m_x)^2 p_{\underline{x}}(x) dx \\ &= \left(\tfrac{d}{dx} f(x^0) \right)^2 \sigma_x^2 \end{aligned}$$

Examples

1. A first order approximation to the variance of $\underline{y} = \cos \underline{x}$ is

   $$\sigma_y^2 \doteq \sin^2 x^0 \sigma_x^2$$

2. A first order approximation to the variance of $\underline{y} = \underline{x}^3$ is

   $$\sigma_y^2 \doteq 9 (x^0)^4 \sigma_x^2$$

□

A.2 Mean and variance of vector random variables

Mean of a vector random variable: Let $\underline{x}_i, i = 1, 2, \ldots, n$ be n continuous scalar random variables with joint probability density function $p_{\underline{x}}(x_1, x_2, \ldots, x_n)$. The expectation or mean of the random n-vector $(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)^T$ is by definition the integral

$$E\{ \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix} \} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} p_{\underline{x}}(x_1, x_2, \ldots, x_n) dx_1 \ldots dx_n \tag{A.12}$$

If we use the notation

$$\begin{aligned} &\underline{x} = (\ \underline{x}_1 \quad \underline{x}_2 \quad \cdots \quad \underline{x}_n\ )^T \quad , \quad x = (\ x_1 \quad x_2 \quad \cdots \quad x_n\ )^T \\ &p_{\underline{x}}(x) = p_{\underline{x}}(x_1, x_2, \ldots, x_n) \quad , \quad dx = dx_1 \ldots dx_n \\ &\text{and } \int = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \end{aligned}$$

we may write (A.12) in the compact form:

$$E\{\underline{x}\} = \int x p_{\underline{x}}(x) dx \tag{A.13}$$

From (A.12) follows that

$$E\{\underline{x}_i\} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i p_{\underline{x}}(x_1, x_2, \ldots, x_n) dx_1 \ldots dx_n \tag{A.14}$$

This result seems to contradict definition (A.1). Note however that (A.14) reduces to (A.1), since

$$\int_{-\infty}^{+\infty} p_{\underline{x}}(x_1,\ldots,x_j,\ldots,x_n)dx_j = p_{\underline{x}}(x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n)$$

Hence, (A.12) may also be written as

$$E\{\begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix}\} = \begin{bmatrix} \int_{-\infty}^{+\infty} x_1 p_{\underline{x}}(x_1)dx_1 \\ \int_{-\infty}^{+\infty} x_2 p_{\underline{x}}(x_2)dx_2 \\ \vdots \\ \int_{-\infty}^{+\infty} x_n p_{\underline{x}}(x_n)dx_n \end{bmatrix} \tag{A.15}$$

Thus in order to compute $E\{\underline{x}_i\}$ one only needs the *marginal* density $p_{\underline{x}_i}(x_i)$.

Mean of a vectorfunction of a random vector: Given a random n-vector $\underline{x}$ and a vectorfunction $F(x), F: R^n \to R^m$ we form the random m-vector $\underline{y} = F(\underline{x})$. As we see from (A.15) the mean of $\underline{y}_i$ is given by

$$E\{\underline{y}_i\} = \int_{-\infty}^{+\infty} y_i p_{\underline{y}_i}(y_i)dy_i \tag{A.16}$$

It appears, therefore, that to determine the mean of y_i, we must first find its *marginal* probability density function $p_{\underline{y}_i}(y_i)$. This, however, is not necessary. As the next theorem shows, $E\{\underline{y}_i\}$ can be expressed directly in terms of the vectorfunction $F(x)$ and the *joint* density of $\underline{x}$.

Theorem:

$$E\{F(\underline{x})\} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} F(x_1,\ldots,x_n) p_{\underline{x}}(x_1,\ldots,x_n)dx_1 \ldots dx_n \tag{A.17}$$

or

$$E\{F(\underline{x})\} = \int F(x) p_{\underline{x}}(x)dx \tag{A.18}$$

Proof: The proof is similar to the proof given in the earlier section of this appendix.

The following theorem is an extremely important one, and it is used frequently in these lecture notes.

Theorem (propagation law of the mean): Given a random n-vector $\underline{x}$ and a vectorfunction $F(x), F: R^n \to R^m$, we form the random m-vector $\underline{y} = F(\underline{x})$. If the vectorfunction $F(x)$ is *linear*,

$$\underset{m\times 1}{F(x)} = \underset{m\times n}{A}\,\underset{n\times 1}{x} + \underset{m\times 1}{b} \tag{A.19}$$

then

$$\underset{m\times 1}{m_y} = \underset{m\times n}{A}\,\underset{n\times 1}{m_x} + \underset{m\times 1}{b} \tag{A.20}$$

Proof: If we denote the n columnvectors of matrix A by $a_i, i = 1,\ldots,n$, we may write (A.19) as

$$F(x) = \sum_{i=1}^{n} a_i x_i + b$$

If we substitute this into (A.17) we get

$$\begin{aligned} E\{F(\underline{x})\} &= \sum_{i=1}^{n} a_i \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i p_{\underline{x}}(x_1,\ldots,x_n) dx_1 \ldots dx_n \\ &\quad + b \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_{\underline{x}}(x_1,\ldots,x_n) dx_1 \ldots dx_n \end{aligned}$$

or

$$\begin{aligned} E\{F(\underline{x})\} &= \textstyle\sum_{i=1}^{n} a_i \int_{-\infty}^{+\infty} x_i p_{\underline{x}_i}(x_i) dx_i + b \\ &= \textstyle\sum_{i=1}^{n} a_i m_{x_i} + b \\ &= A m_x + b \end{aligned}$$

Without proof we also give the linearized version of the propagation law of the mean.

Theorem (linearized propagation law of the mean): Given a random n-vector $\underline{x}$ and a nonlinear vectorfunction $F(x), F : R^n \rightarrow R^m$ we form the random m-vector $\underline{y} = F(\underline{x})$. Let $x^0 \in R^n$ be an approximation to a sample of $\underline{x}$ and define $\Delta\underline{y} = \underline{y} - F(x^0)$ and $\Delta\underline{x} = \underline{x} - x^0$. Then we have to a first-order

$$m_{\Delta y} \doteq \partial_x F(x^0) m_{\Delta x} \tag{A.21}$$

Examples

1. Let the two random variables $\underline{y}_1$ and $\underline{y}_2$ be defined as

$$\begin{cases} \underline{y}_1 &= 1\,\underline{x}_1 + 3\,\underline{x}_2 + 5\,\underline{x}_3 + 2 \\ \underline{y}_2 &= 4\,\underline{x}_1 + 2\,\underline{x}_2 - 1\,\underline{x}_3 + 3 \end{cases}$$

   Then

$$E\{\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix}\} = \begin{bmatrix} 1 & 3 & 5 \\ 4 & 2 & -1 \end{bmatrix} E\{\begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \underline{x}_3 \end{bmatrix}\} + \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

2. Let the two random variables $\underline{y}_1$ and $\underline{y}_2$ be defined as

$$\begin{cases} \underline{y}_1 &= \sin(\underline{x}_1\,\underline{x}_2) + \underline{x}_3 \\ \underline{y}_2 &= \underline{x}_1^2 + \underline{x}_2 + 4 \end{cases}$$

   With the approximate values x_1^0, x_2^0 and x_3^0 we have to a first-order

$$E\{\begin{bmatrix} \underline{y}_1 - \sin(x_1^0\,x_2^0) - x_3^0 \\ \underline{y}_2 - (x_1^0)^2 - x_2^0 - 4 \end{bmatrix}\} = \begin{bmatrix} x_2^0\cos(x_1^0\,x_2^0) & x_1^0\cos(x_1^0\,x_2^0) & 1 \\ 2x_1^0 & 1 & 0 \end{bmatrix} E\{\begin{bmatrix} \underline{x}_1 - x_1^0 \\ \underline{x}_2 - x_2^0 \\ \underline{x}_3 - x_3^0 \end{bmatrix}\}$$

□

Variancematrix of a random vector: Let $\underline{x}_i, i = 1,2,\ldots,n$ be n continuous scalar random variables with joint probability density function $p_{\underline{x}}(x_1,\ldots,x_n)$. The *variancematrix* of the random n-vector $(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)^T$ is by definition the integral

$$E\{\begin{bmatrix} \underline{x}_1 - E\{\underline{x}_1\} \\ \vdots \\ \underline{x}_n - E\{\underline{x}_n\} \end{bmatrix} \begin{bmatrix} \underline{x}_1 - E\{\underline{x}_1\} \\ \vdots \\ \underline{x}_n - E\{\underline{x}_n\} \end{bmatrix}^T\} = \tag{A.22}$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \begin{bmatrix} x_1 - E\{\underline{x}_1\} \\ \vdots \\ x_n - E\{\underline{x}_n\} \end{bmatrix} \begin{bmatrix} x_1 - E\{\underline{x}_1\} \\ \vdots \\ x_n - E\{\underline{x}_n\} \end{bmatrix}^T p_{\underline{x}}(x_1,\ldots,x_n)dx_1\ldots dx_n$$

This variancematrix will also be denoted by Q_x. Using vector notation we may write (A.22) in the compact form

$$E\{(\underline{x}-m_x)(\underline{x}-m_x)^T\} = \int (x-m_x)(x-m_x)^T p_{\underline{x}}(x)dx \tag{A.23}$$

From (A.22) it follows that

$$E\{(\underline{x}_i - m_{x_i})(\underline{x}_j - m_{x_j})\} =$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (x_i - m_{x_i})(x_j - m_{x_j}) p_{\underline{x}}(x_1,\ldots,x_n)dx_1\ldots dx_n$$

Integration gives

$$E\{(\underline{x}_i - m_{x_i})(\underline{x}_j - m_{x_j})\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_i - m_{x_i})(x_j - m_{x_j}) p_{\underline{x}}(x_i,x_j)dx_i dx_j \tag{A.24}$$

in which $p_{\underline{x}}(x_i,x_j)dx_i dx_j$ is the joint density function of the two random variables $\underline{x}_i$ and $\underline{x}_j$. The scalar (A.24) is called the *covariance* of the two random variables $\underline{x}_i$ and $\underline{x}_j$. This covariance will also be denoted as $\sigma_{x_i x_j}$. Note that the off-diagonal elements of the variancematrix Q_x consist of the covariances between the elements of the random vector $\underline{x}$.

If $i = j$ it follows from (A.24) that

$$E\{(\underline{x}_i - m_{x_i})^2\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_i - m_{x_i})^2 p_{\underline{x}}(x_i,x_j)dx_i dx_j$$

Integration gives

$$E\{(\underline{x}_i - m_{x_i})^2\} = \int_{-\infty}^{+\infty} (x_i - m_{x_i})^2 p_{\underline{x}_i}(x_i)dx_i$$

This is the *variance* of $\underline{x}_i$. Note that the variance $\sigma^2_{x_i}$ is the ith-diagonal element of the variancematrix Q_x. Hence, the variancematrix Q_x can be written as

$$\underset{n\times n}{Q_x} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_n} \\ \sigma_{x_2x_1} & \sigma^2_{x_2} & & \\ \vdots & & \ddots & \\ \sigma_{x_nx_1} & & & \sigma^2_{x_n} \end{bmatrix} \tag{A.25}$$

Note that the variancematrix is a *symmetric* matrix.

Theorem (propagation law of variances): Given a random n-vector $\underline{x}$ and a vectorfunction $F(x), F : R^n \to R^m$, we form the random m-vector $\underline{y} = F(\underline{x})$. If the vectorfunction $F(x)$ is *linear*,

$$\underset{m\times 1}{F(x)} = \underset{m\times n\, n\times 1}{A\,x} + \underset{m\times 1}{b} \tag{A.26}$$

then

$$\underset{m\times m}{Q_y} = \underset{m\times n}{A}\ \underset{n\times n}{Q_x}\ \underset{n\times n}{A^T} \tag{A.27}$$

Proof: If we denote the n columnvectors of matrix A by $a_i, i = 1,\ldots,n$, we may write (A.26) as

$$F(x) = \sum_{i=1}^{n} a_i x_i + b \tag{A.28}$$

In a similar way we may write (A.20) as

$$m_y = \sum_{i=1}^{n} a_i m_{x_i} + b \tag{A.29}$$

Substitution of (A.28) and (A.29) into

$$\begin{aligned} Q_y &= E\{(\underline{y}-m_y)(\underline{y}-m_y)^T\} \\ &= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}(F(x)-m_y)(F(x)-m_y)^T p_{\underline{x}}(x_1,\ldots,x_n)dx_1\ldots dx_n \end{aligned}$$

and using

$$\sigma_{x_i x_j} = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}(x_i-m_{x_i})(x_i-m_{x_i})^T p_{\underline{x}}(x_1,\ldots,x_n)dx_1\ldots dx_n$$

gives

$$Q_y = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^T \sigma_{x_i x_j}$$

This can be written as

$$Q_y = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} \sum_{j=1}^{n}\sigma_{x_1 x_j} a_j^T \\ \vdots \\ \sum_{j=1}^{n}\sigma_{x_n x_j} a_j^T \end{bmatrix}$$

or as

$$\begin{aligned} Q_y &= \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} \sigma_{x_1x_1} & \cdots & \sigma_{x_1x_n} \\ \vdots & & \vdots \\ \sigma_{x_nx_1} & \cdots & \sigma_{x_nx_n} \end{bmatrix} \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \\ &= AQ_xA^T \end{aligned}$$

Without proof we also give the linearized version of the propagation law of variances.

Theorem (linearized propagation law of variances): Given a random n-vector $\underline{x}$ and a nonlinear vectorfunction $F(x), F : R^n \rightarrow R^m$, we form the random m-vector $\underline{y} = F(\underline{x})$. Let $x^0 \in R^n$ be an approximation to a sample of $\underline{x}$. Then we have to a first order

$$\underset{m\times m}{Q_y} \doteq \underset{m\times n}{\partial_x F(x^0)} \underset{n\times n}{Q_x} \underset{n\times m}{\partial_x F(x^0)^T} \tag{A.30}$$

Examples

1. Let the two random variables $\underline{y}_1$ and $\underline{y}_2$ be defined as

$$\left\{ \begin{array}{lcl} \underline{y}_1 & = & 1\,\underline{x}_1 + 3\,\underline{x}_2 + 5\,\underline{x}_3 + 2 \\ \underline{y}_2 & = & 4\,\underline{x}_1 + 2\,\underline{x}_2 - 1\,\underline{x}_3 + 3 \end{array} \right.$$

The variancematrix of $\underline{x} = (\underline{x}_1, \underline{x}_2, \underline{x}_3)^T$ is given as

$$Q_x = \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{array} \right]$$

Then

$$\begin{aligned} Q_y &= \left[\begin{array}{ccc} 1 & 3 & 5 \\ 4 & 2 & -1 \end{array} \right] \left[\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{array} \right] \left[\begin{array}{cc} 1 & 4 \\ 3 & 2 \\ 5 & -1 \end{array} \right] \\ &= \left[\begin{array}{cc} 50 & 25 \\ 25 & 41 \end{array} \right] \end{aligned}$$

2. Let the two random variables $\underline{y}_1$ and $\underline{y}_2$ be defined as

$$\left\{ \begin{array}{lcl} \underline{y}_1 & = & \underline{x}_1 + \underline{x}_2^2 + \underline{x}_1 \underline{x}_3 \\ \underline{y}_2 & = & \underline{x}_1^3 + \sin \underline{x}_2 \end{array} \right] \tag{A.31}$$

The variancematrix of $\underline{x} = (\underline{x}_1, \underline{x}_2, \underline{x}_3)^T$ is given as

$$Q_x = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array} \right] \tag{A.32}$$

The 2×3 matrix of partial derivatives $\partial_x F(x^0)$ follows from (A.31) as

$$\begin{aligned} \underset{2\times 3}{\partial_x F(x^0)} &= \left[\begin{array}{ccc} \partial_{x_1} F^1(x^0) & \partial_{x_2} F^1(x^0) & \partial_{x_3} F^1(x^0) \\ \\ \partial_{x_1} F^2(x^0) & \partial_{x_2} F^2(x^0) & \partial_{x_3} F^2(x^0) \end{array} \right] \\ &= \left[\begin{array}{ccc} 1 + x_3^0 & 2x_2^0 & x_1^0 \\ 3(x_1^0)^2 & \cos x_2^0 & 0 \end{array} \right] \end{aligned} \tag{A.33}$$

If we take as approximate values $x_1^0 = 1, x_2^0 = 0, x_3^0 = 0$, (A.33) becomes

$$\partial_x F(x^0) = \begin{bmatrix} 1 & 0 & 1 \\ 3 & 1 & 0 \end{bmatrix} \tag{A.34}$$

The variancematrix Q_y follows now from (A.32) and (A.34) to a first-order as

$$\begin{aligned} Q_y &\doteq \begin{bmatrix} 1 & 0 & 1 \\ 3 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &\doteq \begin{bmatrix} 3 & 3 \\ 3 & 10 \end{bmatrix} \end{aligned}$$

□

Appendix B

References

In order to aid a further study in the field of adjustment, testing and computation of geodetic networks, the list of references given below gives an introductory overview of the existing international literature. With a few exceptions, the list concentrates on books, lectures notes and reports.

Adjustment and testing

[1] Baarda, W. (1967): *Statistical Concepts in Geodesy*, Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 2, No. 4, Delft.

[2] Baarda, W. (1968): *A Testing Procedure for use in Geodetic Networks*, Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 2. No. 5, Delft.

[3] Grafarend, E., B. Schaffrin (1993): *Ausgeleichungsrechnung in Linearen Modellen*, B.I. Wissenschaftsverlag.

[4] Graybill, F.A. (1976): *Theory and Application of the Linear Model*, Duxbury Press.

[5] Koch, K.-R. (1987): *Parameterschätzung und Hypothesestests in linearen Modellen*, 2nd edition, Dümmler Verlag, Bonn.

[6] Meissl, P. (1982): *Least Squares Adjustment: A Modern Approach*, Mitteilungen der Geodätischen Institute der Technische Universität Graz. Folge 43.

[7] Mikhail, E. M. (1976): *Observations and Least Squares*, University Press of America.

[8] Rao, C.R. (1973): *Linear Statistical Inference and its Applications*, 2nd edition, Wiley series in probability and mathematical statistics.

[9] Teunissen, P.J.G. (2001): *Adjustment Theory: an introduction*, 2nd edition, Series on Mathematical Geodesy and Positioning, Delft University Press, ISBN 90-407-1974-8.

[10] Teunissen, P.J.G. (2000): *Testing Theory: an introduction*, Series on Mathematical Geodesy and Positioning, Delft University Press, ISBN 90-407-1975-6.

The first two references concentrate on testing theory. Much of the quality control procedures which have become standard practice nowadays, can be traced back to these two original works. The references [3], [4], [5] and [8] are reference books, whereas the references [6], [7], [9] and [10] are more of the lecture notes type. Reliability theory is only delt with in the first two and in the last reference.

Surveying and geodetic networks

[1] Alberda, J.E. (1974): Planning and Optimization of Networks: Some General Considerations. *Boll. Geod. Sc. Aff.*, 33, pp. 209-240.

[2] Alberda, J.E. (1981): *Inleiding Landmeetkunde*, 3rd edition, Delft University Press.

[3] Baarda, W. (1973): *S-transformations and Criterion Matrices*, Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 5, No.1, Delft.

[4] Borre, K. (1990): *Landmaeling*, Aalborg.

[5] Delft Geodetic Computing Centre (Eds.) (1982): *Forty Years of Thought*. Anniversary volume (Vol. 1 and 2) on the occasion of prof. Baarda's 65th birthday, Delft.

[6] Grafarend, E., Sanso (Eds.) (1985): *Optimization and Design of Geodetic Networks*, Springer Verlag.

[7] Grafarend, E., B. Schaffrin (1974): Unbiased free net adjustment. *Survey review*, 22, pp. 200-218.

[8] Heck, B. (1987): *Rechenverfahren und Auswertemodelle der Landesvermessung*, Herbert Wichmann Verlag.

[9] Kahmen, H., W. Faig (1988): *Surveying*, Walter de Gruyter.

[10] Mierlo van, J. (1979): *Free Network Adjustment and S- transformations*, DGK B, No. 252, pp. 41-54.

[11] Moffitt, F.H., H. Bouchard (1982): *Surveying*, 7th edition, Harper and Row.

[12] Richardus, P. (1984): *Project Surveying*, Balkema.

[13] Teunissen, P.J.G. (1984): *Generalized Inverses, The Datum problem and S-transformations*, Mathematical and Physical Geodesy Report 84.1, Delft, pp. 44.

The references [2], [4], [8], [9], [11], and [12], are all textbooks on surveying. The references [2], [4], [9] and [11], are of an introductory nature. Reference [8] concentrates on functional models, in particular spatial modelling and reference [12] includes aspects of quality control. Optimization and design aspects of geodetic networks are treated in their whole range of variety in the references [1], [5] and [6]. Free networks are treated in particular in [3], [7], [10] and [13]. The concept of S-transformations was introduced in [3] and its relation to generalized inverses is included in [10] and [13].

GPS Surveying

[1] Husti, G.J. (2000): *Global Positioning System* (in Dutch), Series on Mathematical Geodesy and Positioning, Delft University Press, ISBN 90-407-1977-2.

[2] Hofmann-Wellenhof, B., H. Lichtenegger, J. Collins (2001): *Global Positioning System: Theory and Practice*, 6th edition, Springer Verlag.

[3] Leick, A. (2005): *GPS Satellite Surveying*, 3rd edition, John Wiley and Sons.

[4] Seeber, G. (2003): *Satellite Geodesy*, 2nd edition, Walter de Gruyter.

[5] Teunissen, P.J.G. and A. Kleusberg (1998): *GPS for Geodesy*, 2nd edition, Springer Verlag.

Network quality control

Peter J.G. Teunissen

The aim of computing a geodetic network is to determine the geometry of the configuration of a set of points from spatial observations (e.g. GPS baselines and/or terrestrial measurements). The configuration of points usually consists of newly established points, of which the coordinates still need to be determined, and already existing points, the so-called control points, of which the coordinates are known.

Network quality control deals with the qualitative aspects of network design, network adjustment, network validation and network connection. By means of a network adjustment the relative geometry of the new points is determined and integrated into the geometry of the existing control points. Prior to the network adjustment, the geometry of the network is designed on the basis of precision and reliability criteria.

The adjustment and validation of the overall geometry can be divided in two phases, the free network phase and the connected network phase. In the free network phase, the known coordinates of the control points do not take part in the adjustment and validation. The possible use of a free network phase is based on the idea that a good geodetic network should be sufficiently precise and reliable in itself, without the need of external control. Moreover, it allows one to validate the quality of the external control.

In the connected network phase, the geometry of the free network is integrated into the geometry of the control points. Adjustment and validation in this second phase differs from the free network phase. The adjustment in the second phase is a constrained connection adjustment, since it is often not practical to see the coordinates of the control points change every time a free network is connected to them. For the validation of the connected network however, the unconstrained connected adjustment is used as input. This allows one to take the intrinsic uncertainty of the coordinates of the control points in the connection phase into account.

The goal of this introductory text on network quality control is to convey the necessary knowledge for designing, adjusting and testing geodetic networks. For the purpose of network design, the precision and reliability theory is worked out in detail. This includes the minimal detectable biases and the bias-to-noise ratios. For the purpose of the network adjustment, the principles of unconstrained-, constrained-, and minimally constrained least-squares estimation, are treated. For the network testing, the principles of hypothesis testing are presented and worked out for the different network cases. For the free network phase this includes the overall model test, the w-test, and the data snooping procedure. For the connected network phase, it includes the T-test, with an emphasis on the detection and identification of errors in the control points.

TU Delft

P.J.G. Teunissen
Delft University of Technology,
Faculty of Civil Engineering and Geosciences

Dr (Peter) Teunissen is Professor of Geodesy at Delft University of Technology (DUT) and an elected member of the Royal Netherlands Academy of Arts and Sciences. He is research-active in various fields of Geodesy, with current research focused on the development of theory , models, and algorithms for high-accuracy applications of satellite navigation and remote sensing systems. His past DUT positions include Head of the Delft Earth Observation Institute, Education Director of Geomatics Engineering and Vice-Dean of Civil Engineering and Geosciences. His books at TUDelft Open are Adjustment Theory, Testing Theory, Dynamic Data Processing and Network Quality Control.

ISBN 978-94-6366-950-4
DOI https://doi.org/10.59490/tb.100
textbooks.open.tudelft.nl

Cover image: A. Smits