Network quality control

Network quality control

P.J.G. Teunissen

Delft Institute of Earth Observation and Space systems (DEOS)
Delft University of Technology

VSSD

Series on Mathematical Geodesy and Positioning http://www.vssd.nl/hlf/a030.htm
Website of DEOS: http://www.lr.tudelft.nl/mgp

Adjustment Theory • P.J.G. Teunissen
2003 / 201 p. / ISBN-13 978-90-407-1974-5

Dynamic Data Processing • P.J.G. Teunissen
2001 / 241 + x p. / ISBN-13 978-90-407-1976-9

Testing Theory • P.J.G. Teunissen
2006 / 147+viii p. / ISBN-13 978-90-407-1975-2

Global Positioning System (in Dutch) • G.J. Husti
2000 / 151 pp. / ISBN-13 978-90-71301-1977-6

Hydrography • C.D. de Jong, G. Lachapelle, S. Skone, I.A. Elema
2003 / x+351 pp. / ISBN-13 978-90-407-2359-9 / hardback

Network Quality Control • P.J.G. Teunissen
2001 / viii + 128 p. / ISBN-13 978-90-71301-98-8

First edition 2006

Published by:
VSSD
Leeghwaterstraat 42, 2628 CA Delft, The Netherlands
tel. +31 15 278 2124, telefax +31 15 278 7585, e-mail: hlf@vssd.nl
internet: http://www.vssd.nl/hlf
URL about this book: http://www.vssd.nl/hlf/a034.htm

A collection of digital pictures and an elctronic version can be made available for lecturers who adopt this book. Please send a request by e-mail to hlf@vssd.nl

Printed version:
ISBN-10 90-71301-98-2
ISBN-13 978-90-71301-98-8
Electronic version:
ISBN-10 90-71301-99-0
ISBN-13 978-90-71301-99-5
NUR 930
Keywords: estimation, precision, testing, reliability, free networks, connected networks, constrained networks

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

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.