Recursive data processing for kinematic GPS surveying

C.C.J.M. Tiberius

NCG Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission

Delft, July 1998

Colophon
Recursive data processing for kinematic GPS surveying
C.C.J.M. Tiberius
Publications on Geodesy 45
ISBN 90 6132 265 0
ISSN 0165 1706

Publications on Geodesy is the continuation of Publications on Geodesy New Series

Published by: NCG Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission, Delft, The Netherlands
Printed by: Meinema Drukkerij, Delft, The Netherlands
Cover: A.B. Smits, TU Delft, Delft, The Netherlands

NCG Nederlandse Commissie voor Geodesie
P.O. Box 5030, 2600 GA Delft, The Netherlands
Tel.: 31 (0)15-278 28 19
Fax: 31 (0)15-278 17 75
E-mail: ncg@geo.tudelft.nl
WWW: www.knaw.nl

The NCG Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW).

Summary

Recursive data processing for kinematic GPS surveying

Through the concept of interferometry and using carrier phase measurements, relative positions can be determined very precisely with the Global Positioning System GPS. Carrier phase measurements are made by phase comparison of the generated and received carrier. Due to this technique, an unknown integer number of cycles (ambiguity) is involved in these measurements. Resolving the ambiguities is the key to **precise GPS positioning**. When the integer (double difference) ambiguities have been successfully resolved, the precision of the **relative** position is at the 1 cm level or better.

The ambiguities are dealt with by the LAMBDA method, the Least-squares AMBiguity Decorrelation Adjustment method, which was introduced in 1993. It features a strict extension of standard **least-squares** to the **integer** domain and by the decorrelating reparametrization of the ambiguities, the integer estimation can be made very fast and efficiently. **Ambiguity resolution** is possible instantaneously, thus based on only one epoch of data, and hence truely On-The-Fly.

In **kinematic GPS surveying**, the roving receiver visits the locations to be surveyed. This set up allows a high productivity in collecting geometric information. It is convenient to obtain the results of the **data processing** already in the field while the survey is run, instead of at the home office one day later. In case of insufficient quality, corrective actions can then be taken immediately in the field. The data processing must therefore be recursive. As mechanization the Square Root Information Filter is applied. The SRIF allows **estimation** and quality control computations to be made in **recursion** and thus possibly in real-time, in close parallel with the gathering of the data. The **quality control** comprises the Detection, Identification and Adaptation of errors in the incoming data (outliers, cycle slips). By the DIA-procedure, the effect of the errors on the estimates for the unknown parameters of interest, the coordinates of the rover, is directly removed.

The data of three GPS measurement campaigns are processed and analysed. The quality, in terms of **precision and reliability**, of the coordinate estimators is considered for various measurement scenarios. Practical results, the positioning performance and the capability of resolving the ambiguities, are given for a **kinematic experiment**. In conclusion is kinematic GPS, as a measurement technique, very well suited for surveying applications.

The **mathematical model** for GPS surveying, or positioning in general, that is currently in use, turns out to be a rough one, and must be further sophisticated. It limits the receiver interdistance to the typical short baseline length ($\sim$10 km). Shortcomings of the **functional** model that show up, when high demands are put on the quality of the positioning results, are explored. Refinements of the functional model concern differential atmospheric delay parameters (ionospheric and tropospheric). In addition, (phase) multipath is still a matter of concern. GPS code and phase observables are extensively analysed; an inventory of refinements to the **stochastic** model is made, and includes elevation dependence, cross-correlation and mutual (channel) correlation.

Contents

Preface

This publication contains my PhD-thesis on kinematic GPS surveying. The research was carried out from 1991 to 1997 at the Faculty of Civil Engineering and Geosciences, Department of Mathematical Geodesy and Positioning, under supervision of professor Peter Teunissen. The major part of the work was under contract (MD508 and MD1078) with the Survey Department of the Rijkswaterstaat, Ministry of Transportation and Public Works. Financial support was provided by the Dr.ir. Cornelis Lely Foundation.

Christian Tiberius
Delft, July 1998.

List of symbols

mathematical geodesy

A : designmatrix, has full rank unless stated otherwise
model of observation equations: $E\{\underline{y}\}=Ax$; $D\{\underline{y}\}=Q_y$

B : matrix with conditions, has full rank unless stated otherwise
model with condition equations: $B^TE\{\underline{y}\}=0$; $D\{\underline{y}\}=Q_y$

C : matrix of alternative hypothesis, specifying a model error (in terms of observables, misclosures or predicted residuals)

D : matrix of alternative hypothesis, specifying a model error (in terms of pseudo-observables)

$E\{.\}$: expectation value of random vector (first moment or mean)

$D\{.\}$: dispersion of random vector (second moment or variance covariance matrix)

I : identity matrix

Q : variance covariance matrix, is symmetric and positive definite unless stated otherwise

$R(A)$: column-space of matrix A

$N(A)$: null-space of matrix A

P_A : orthogonal projector on $R(A)$

S : square root factor of variance covariance matrix $Q_{\hat{x}}=SS^T$
(note that strictly S is not a matrix square root, it would be if $Q_{\hat{x}}=SS$;
$Q_{\hat{x}}$ is the symmetric product of S)

T : transformation matrix

$\underline{T}$: generalized likelihood ratio teststatistic

X : matrix describes impact of model error on x; response matrix

Z : response matrix, describes impact of model error on z

Z^T : transformation matrix (decorrelation)

$\underline{d}$: vector of pseudo-observables (from continuous time dynamic model)

$\underline{e}$: (measurement) noise in observable $\underline{y}$

$\underline{\hat{e}}$: least-squares residuals, $\underline{\hat{e}}=\underline{y}-\underline{\hat{y}}$

k : epoch index, or total number of epochs, or critical value

m : number of observations (from measurements)

n : number of unknowns

q : dimension of model error

r : rank of design matrix A

r : auto-correlation-coefficient

$p_{\underline{y}}(y)$: probability density function of $\underline{y}$

$\underline{t}$: teststatistic for one-dimensional alternative hypothesis ($q=1$), or misclosures
$\underline{v}$: predicted residuals
$\underline{w}$: normalized predicted residuals
$\underline{u}$: normalized misclosures
x : state vector of (time-varying) system; vector of unknowns
$\underline{y}$: vector of observables (from measurements)
z : transformed state vector; $z = S^{-1}x$
or reparametrized (integer) unknowns; $z = Z^T x$

Δ : increment to a parameter or vector
$\Phi_{k,k-1}$: transition matrix, from epoch k-1 to k

α : level of significance
γ : power
λ : non-centrality parameter, reference value indicated by λ_o
or eigenvalue(s) of a square matrix
$\lambda_{\hat{x}}$: bias to noise ratio
ρ : correlation coefficient
σ : standard deviation

∇ : vector of model errors, or Minimal Detectable Bias (one-dimensional, $q=1$)
$\underline{\cdot}$: random vector (or variable)
$\hat{\cdot}$: estimate
$\underline{\hat{\cdot}}$: estimator
$\check{\cdot}$: integer estimate (for x_I)
$|.|$: determinant of a square matrix, or absolute value of a scalar
$\|.\|$: norm of a vector; Euclidean or 2-norm (length)

for n-vector x: $\|x\| = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}$ in standard metric, unless stated otherwise

nl. $\|x\|_{Q_x^{-1}} = \sqrt{x^T Q_x^{-1} x}$

$.^T$: transpose (vector or matrix)
$.^{-1}$: inverse (matrix)
$.^{1/2}$: factor of symmetric positive (semi-) definite matrix (usually Cholesky, thus (upper) triangular with non-negative diagonal elements)
$.^o$: quantity related to null hypothesis
$.^a$: quantity related to alternative hypothesis
$\cdot_o$: approximate value (in linearization)
$\cdot_I$: refers to integer variable or vector (e.g. double difference ambiguities)
$\cdot_R$: refers to real variable or vector (e.g. baseline coordinates)
$\cdot_k$: indicates epoch k for a discrete time quantity e.g. y_k
or interval $[k\text{-}1,k]$ for an originally continuous time quantity e.g. d_k
$\cdot_{l|k}$: quantity concerns epoch l and is determined using data up to epoch k
$\cdot_{l,k}$: quantity concerns interval $[l,k]$

GPS

r : receiver r (subscript)
s : satellite s (superscript)
m : number of satellites

c	: speed of light in vacuum	c =299792458 m/s
f_i	: frequency of carrier in L-band	f_1=1575.42 MHz (L1) f_2=1227.60 MHz (L2)
λ_i	: wavelength	λ_1=0.1903 m (L1) λ_2=0.2442 m (L2)

e : unit direction vector
$x\,i$: 3-dimensional position vector of device i, expressed in Earth Centered Earth Fixed (ECEF) coordinate system
$X\,i$: 3-dimensional position vector of device i, expressed in inertial coordinate system

$p_r^s(t)$: code (pseudo-range) observation on L1 : C1 or P1
$\underline{p}_r^s(t)$: code (pseudo-range) observation on L2 : P2
$\Phi_r^s(t)$: (carrier) phase observation on L1 : L1
$\underline{\Phi}_r^s(t)$: (carrier) phase observation on L2 : L2

$\rho_r^s(t,t-\tau)$: geometric range from satellite to receiver [m], defined from phase center of satellite antenna array to phase center of receiver antenna, when expressed in the ECEF system: $= \|x^s(t-\tau) - x_r(t)\|$

τ : traveltime of signal [s], defined from satellite circuits to receiver circuits
t : GPS system time
$t\,i$: time displayed by device i at time t
$\delta\,it$: clock error of device i; $t = ti - \delta\,it$
$\delta_r t(t)$: receiver clock error [s] at time t
$\delta^s t(t-\tau)$: satellite clock error [s] at time $(t-\tau)$
t_c : time tag (in receiver time) of epoch of computation

$T_r^s(t)$: tropospheric delay [m] for signal from satellite s to receiver r
$I_r^s(t)$: ionospheric (group) delay [m] for L1-signal

$\underline{d}^s(t)$: satellite hardware delay [m] for L1-code
$d^s(t)$: satellite hardware delay [m] for L2-code
$\underline{D}^s(t)$: satellite hardware delay [m] for L1-carrier
$D^s(t)$: satellite hardware delay [m] for L2-carrier

$\underline{d}_r(t)$: receiver hardware delay [m] for L1-code
$d_r(t)$: receiver hardware delay [m] for L2-code
$\underline{D}_r(t)$: receiver hardware delay [m] for L1-carrier
$D_r(t)$: receiver hardware delay [m] for L2-carrier

$\underline{A}_r^s$: phase ambiguity [m] for L1-carrier
A_r^s : phase ambiguity [m] for L2-carrier

Introduction

In this introduction, a sketch of the background of kinematic surveying with GPS is given, followed by a brief outline of the research. The introduction is concluded by a brief description of the contents of this report.

background
The Global Positioning System (GPS) is a worldwide satellite-based positioning system of the US Department of Defense (DoD). The primary purpose of the system is single point positioning with code observations. For a code, or pseudo-range observation, the receiver determines the travel time of the signal transmitted by the satellite. The position is determined in the three-dimensional World Geodetic System 1984 (WGS84). The accuracy lies in the order of ten to hundred meters.

Relative positioning with phase observations yields much higher accuracies. For a phase observation the receiver determines the difference in phase of the carrier received from the satellite and the carrier generated by the receiver itself. The measurement precision lies in the order of one tenth of a radian, which corresponds to three millimeters for the L1-carrier. Like the code, the phase is a range related observation, but the observation is complicated by the presence of an ambiguity.

The concept of relative positioning allows the use of the GPS in surveying. Two, or more, receivers simultaneously make observations to the same satellites. The position of the second receiver is determined, relative to the position of the first one, the reference receiver. The precision of the coordinates lies in the order of centimeters.

In kinematic GPS surveying, the reference receiver is stationary and the second receiver visits the points to be determined. These points are occupied only shortly, i.e. from one instant to a few minutes. Therefore a high productivity in positioning can be realized with this technique. The receiver interdistance is usually limited to 10-15 kilometers. Kinematic GPS surveying can be used for most of the traditional surveying tasks, such as topographic surveying and control surveying, including the determination of photo control points. Kinematic GPS thus aims at precise positioning for survey-applications on a local scale.

research
The purpose of the underlying research is to develop a model and method for the data processing in kinematic GPS surveying. The data processing consists of two parts: estimation and quality control.

Based on the observations, estimates are computed for the unknowns, of which the coordinates are of primary interest. The estimation is carried out in recursion and this

means that the estimation is not held until the survey has been completed, but that 'preliminary' estimates are computed, based on the observations that are already available.

Quality consists of precision and reliability. Precision describes the spread in the estimation results. This description is valid as long as the measurement system is adequately described by the model used in the estimation. When discrepancies between model and reality occur, estimation results and precision description are invalidated. Model errors therefore, need to be detected by a statistical testing procedure, that is executed in combination with the estimation. Subsequently, the estimation should be accounted for the occurrence of the model errors. The nominal performance of the testing procedure is described by reliability.

The procedure for data processing is recursive, and can therefore be carried out in (near) real time, parallel to the gathering of the data. In this way, positioning results can be already available in the field. For a surveyor this is usually not required for the estimation results, but it is on the contrary for the quality control. When quality turns out to be insufficient, corrective actions can be taken immediately. A re-survey, caused by insufficient quality, can then be avoided. The quality of the positioning results is controlled in the field.

synopsis

Chapter 1 is a review of some essential parts of the theory on mathematical geodesy. First the art of modelling is briefly discussed. Once the mathematical model is available, the data processing can be carried out. The principles of both estimation and quality control are reviewed. Chapter 2 covers the data processing. The data processing is carried out in a recursive manner and the mechanization to realize this, the Square Root Information Filter, is treated.

Theory and mechanization are discussed, in principle independent of any application. The application - kinematic GPS surveying - is introduced in chapter 3. It is followed by a general discussion on the modelling of GPS surveying. The mathematical model is set up and developed further to fit in the procedure for recursive data processing. In chapter 4, theory, mechanization and mathematical model come together in the implementation of the data processing for kinematic GPS surveying. The prototype implementation developed is briefly discussed, together with several practical aspects. A separate section deals with the aspects of model validation for kinematic GPS surveying.

The implementation is the basis for the computations and analyses discussed in chapter 5. Three measurement-campaigns are considered, each with a different purpose. The first campaign is used for an analysis of the quality of the measurement system, the second one for a closer analysis of the properties of the (raw) GPS observables, especially in relation to the current mathematical model, and the third one for practical results on positioning and ambiguity resolution. As such, an overview is provided of the geodetic quality of the geometric information obtained with GPS surveying.

Chapter 6 finally provides a summary and gives conclusions and recommendations both for kinematic GPS surveying and further research.

The chapters are written to be self-containing to a high degree. Each chapter is therefore supplied with an introduction, summary, possibly appendices and list of references.

1. Review mathematical geodesy

In this chapter we will briefly review some essential parts of the theory on mathematical geodesy as developed and taught at the 'Delft school' by professors Tienstra, Baarda and Teunissen at present. The topics reviewed here, are those that are directly needed in the data processing for (kinematic) surveying with GPS. This review certainly is not complete and for a thorough treatment of the topics, frequent references will be made to literature.

system of linear equations
We start by discussing the deterministic modelling $y=Ax$. A set of unknowns in vector x are functionally (linearly) related to a set of observations in vector y. In this respect the purpose of estimation is, given a set of observations, to determine the values of the unknowns in x. Two problems may occur when one tries to solve for the unknowns in an arbitrary system of linear equations $y=Ax$. The first problem is that some of the unknowns can not be determined from the given set of observations; the set is not sufficient. This is the estimability problem. One could say that there are 'too few' observations. The second problem is that there may be 'too many' observations for the determination of (certain) unknowns; these are overdetermined. This is the 'problem' of redundancy.

estimation
The first problem is solved by the choice of an S-basis. A number of additional constraints on the unknowns in x is incorporated in the estimation. The second one is solved by using an appropriate estimation criterion: given the redundant set of observations, an estimate for the unknowns is determined that is optimal according to the criterion. The least-squares criterion and the BLUE-estimation principle, together with their relationship, are discussed.

extensions to estimation
In section 4, we will make an excursion to integer estimation. Usually the unknown parameters in vector x are real valued quantities, but in some case (some of) the unknowns are known to be integers. This calls for an extension of the estimation procedure. Up to now we also assumed that a linear relation exists between the observations and unknowns. In practice this is seldom the case. The estimation procedure is extended as to handle non-linear models $y=A(x)$ as well. Keywords are linearization and iteration.

quality
As a result of data processing, the estimates for the unknowns should be accompanied by measures of the quality of the estimator. The quality is described by precision as long as the model used in the estimation indeed holds true. In practice one can never be sure that the model employed, is completely adequate, or in other words, there is a chance that the model is not correct. A statistical testing procedure is used to infer the validity of the model one is using. The presence of redundant data is crucial for statistical testing. After

detection and identification of a misspecification of the model, one has to account for this misspecification in the estimation. The nominal performance of the testing procedure is described by reliability.

1.1 Modelling y=Ax

This thesis deals with the data processing in a dynamic system. We are interested in some quantifiable characteristics of the system. These characteristics constitute the state of the system. As such the state is an abstraction of reality. The state can be mathematically described by a set of parameters, contained in vector x. The adjective dynamic means that the system is time-varying and so will be the parameters: $x(t)$. In this chapter we will consider the state at one instant in time. The time-varying character will be dealt with in chapter 2.

The state of the dynamic system is not known a-priori and we will try to gain knowledge about it, by making measurements on its characteristics. The measurement system thus involves observations in the m-vector y and unknowns in the n-vector x, which are related via the linear model

$$y = Ax \tag{1.1}$$

where A is the mxn design matrix. The unknowns in vector x are deterministic parameters. They are just a medium to express interrelationships between observables of which realizations - observations - are made in real world. Based on the model (1.1), we will compute estimators for the unknowns, and these estimators are linear functions of the observables; this discussion is pursued in section 6. Empirical data y are thus used to make inferences (based on the obtained estimators for the unknowns) about physical reality [Teunissen, 1997].

Model (1.1) is a mathematical description and thus an approximation of the measurement system embedded in reality. In practice, modelling is not a trivial task, see e.g. [Schwarz, 1994]. A thorough knowledge of the measurement system is required. In geodetic measurement systems this knowledge is usually present (to a large extent) and it can be assumed that the model set up, adequately describes reality. Model identification is therefore not of concern; there is no uncertainty in the coefficients contained in matrix A.

1.2 Solution of y=Ax

In respect of the deterministic model $y=Ax$, the purpose of estimation is to obtain values for the unknowns x, based on the observations y, or in other words, to find the solution to the set of linear equations.

The linear operator A maps a vector of the space of unknowns R^n onto a vector in the observation space R^m, figure 2.1.

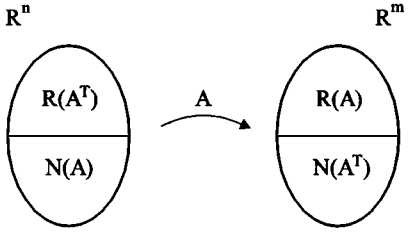


Figure 2.1: Mathematical modelling $y = Ax$, R^n parameter space (left) and R^m observation space (right)

In general, the mxn design matrix A will have rank r, with $r<\min\{n,m\}$. If $r<m$, system (1.1) may be not consistent and it is underdetermined if $r<n$, see [Teunissen, 1984]. The system being not consistent means that the vector y can not be written as a linear combination of the columns of matrix A: $y \notin R(A)$. Underdeterminancy means that, given the observations y, it is not possible to determine the unknowns x uniquely, although there may be redundancy. The redundancy equals $(m\text{-}r)$.

1.2.1 Uniqueness

We will first elaborate on the underdeterminancy. To guarantee a unique solution, the rank of A must be $r=n\leq m$, with preferably $n<m$ for the reason of quality control. At this moment the design matrix is rankdeficient: a set of $(n\text{-}r)$ linearly independent vectors s of dimension n can be found that satisfy $As=0$. The vectors s span the null space of matrix A, $N(A)$. In literature, the system (1.1) is then said to be not observable. Too many unknowns have been introduced in the model. One has to re-set up the model (with less unknowns), bring in deterministic information, or introduce additional observables.

The second strategy implies the choice of an S-basis, see [ibid]: $(n\text{-}r)$ constraints on the unknowns are introduced. The S-basis is constructed in the parameter space R^n by means of the orthogonal complement $S^\perp$ of matrix S; $R(S)\oplus N(A)=R^n$. The constraints read $S^{\perp T}x=0$ and x can be written as a linear combination of the columns of matrix S: $x=Sa$, with a an r-vector. Model (1.1) then becomes $y=ASa$. The n columns of matrix A are linearly dependent and the space $R(A)$ in R^m can be spanned with less, i.e. r, columns: $R(A)=R(AS)$. Choosing an S-basis does not change anything in the observation space R^m, but it does of course, in the parameter space R^n. Different S-bases S_1, S_2 will lead to different estimates for the unknowns: $\hat{x}^{S_1}$, $\hat{x}^{S_2}$. These different estimates are mutually related by means of an S-transformation, see [ibid].

In practice we can discriminate between two types of rankdeficiencies of the design matrix A: the strict deficiency $As=0$, and the near deficiency $As\approx 0$, for a non-zero vector s. No unique estimate can be computed for the first case. In the second case an estimation algorithm may produce results and the quality of these results directly depends on the numerical accuracy of the algorithm. In practice it is better not to risk numerical problems and to avoid this second type of rankdeficiency by one of the strategies mentioned above. With the second strategy of introducing additional constraints, the near rankdeficient system is actually slightly over-constrained. The case $As\approx 0$ can also occur with As not yet sufficiently close to zero to cause numerical problems. A near rank-deficiency anyway means that certain (functions of the) unknowns are hardly estimable. The estimators will have very poor precision.

1.2.2 Consistency

In the sequel we assume that $rank(A)=n\leq m$. In practice, when $n<m$, the vector of (realized) observations will not be consistent with the model (1.1). A system that is inconsistent can be solved routinely by the least-squares algorithm, see p.18 in [ibid]: least-squares generalized inverses. The inconsistency in the observation space R^m is removed by determining the estimate $\hat{y}$ for y, such that $\hat{y}\in R(A)$. This determination is carried out according to the principle of least distance:

$$\min_{\hat{y}} \|y-\hat{y}\|^2_{Q_y^{-1}} \quad \wedge \quad \hat{y} \in R(A) \tag{2.1}$$

with Q_y^{-1} the weight matrix. Equation (2.1) is the (weighted) least-squares estimation principle, specifically classic least-squares based on the L_2-norm [Björck, 1996]. When the inconsistency is significant, one might wonder about the validity of model (1.1). This topic will be further discussed in the section on model validation.

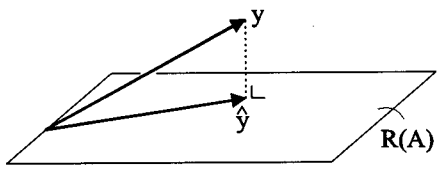


Figure 2.2: Least-squares estimation in the observation space R^m, orthogonal projection of y on $R(A)$

The geometric interpretation of least-squares estimation in the observation space R^m is that the vector $\hat{y}$ is selected such that its end point has shortest distance to the end point of the given vector y.

Estimate $\hat{y}$ is found by orthogonal projection of y on the space spanned by the column vectors of matrix A: $R(A)$.

$$\hat{y} = P_A y \tag{2.2}$$

The orthogonal projector is represented by the *mxm* matrix P_A. The estimates for the unknowns can then be determined from $\hat{y}$ by 'inversion' of the relation (1.1). The least-squares criterion can be rewritten as

$$\min_{x} \|y-Ax\|^2_{Q_y^{-1}} \quad \wedge \quad x \in R^n \tag{2.3}$$

With the design matrix having full rank, $rank(A)=n$, a unique estimate $\hat{x}$ for the unknowns will result.

1.3 BLUE-estimation principle

The parameters contained in vector x are deterministic quantities; an (unknown) parameter is thought to have one and only one value out of R. The observables, contained in vector $\underline{y}$ are random variables. Measurement noise is involved in making observations. The noise is contained in the random vector $\underline{e}$. The description of the state by means of the set of parameters, the observables as random variables and the relation between the observables and the state, constitute the mathematical model of the measurement system:

$$\underline{y} = Ax + \underline{e} \tag{3.1}$$

Note that the noise $\underline{e}$ enters relation (3.1) in an additive fashion. If relation (1.1) adequately models the measurement system in reality, the remaining (small) effects in making the observations, are categorized as noise, and therefore random. Per observable, the effects all together will be zero on the average, so that for the vector of measurement noise holds:

(3.2) $$E\{\underline{e}\} = 0$$

The relations between the observables and the unknowns are expressed in a model of observation equations [Teunissen, 1994a], that define the first and second moment of the vector of observables, the expectation and dispersion:

(3.3) $$E\{\underline{y}\} = Ax \quad ; \quad D\{\underline{y}\} = Q_y$$

A random vector like $\underline{y}$ may have many more moments, but only the first and second are required for the BLUE-estimation principle. The probability density function of $\underline{y}$ will be further discussed in section 6. In the sequel the specification of $E\{.\}$ in (3.3) will be referred to as functional model and $D\{.\}$ as stochastic model. Model (3.3) is also known as the Gauss-Markov model, see [Koch, 1988], or the general univariate linear model, see [Björck, 1996].

Based on the mathematical model an estimator for the state is computed. The Best Linear Unbiased Estimation (BLUE) principle yields, based on a linear functional model as in (3.3), a linear and unbiased estimator $\hat{x}$, that is best in a statistical sense. Under the assumption that the inverse variance covariance matrix of the observables is used as the weight matrix, the least-squares principle, which is a deterministic estimation principle, leads to an identical estimator. If the observables possess a Gaussian distribution, also the Maximum Likelihood principle yields the same estimator.

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

The equivalent model of (3.3) in terms of condition equations reads

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

with matrix B^T an $(m\text{-}n)$xm matrix. It holds that

(3.6) $$B^T A = 0$$

With matrix A of full rank n and matrix B of full rank $(m\text{-}n)$, the column vectors of these two matrices together, span the full observation space R^m.

(3.7) $$R(A : B) = R^m$$

In (3.3) the n unknown parameters in x can be considered as just a medium to express interrelationships between observables. By the model of condition equations (3.5), $(m\text{-}n)$ constraints (relations) are applied to the (first moment) of the vector of observables in the observation space R^m; no unknown parameters are involved. The inconsistency of observations and model directly shows up by the vector of misclosures:

(3.8) $$\underline{t} = B^T \underline{y}$$

When the observations are substituted in (3.8), the misclosures do in general not equal zero. The misclosures will be used for model validation in section 7.

Together estimator $\hat{\underline{x}}$ and the vector of misclosures $\underline{t}$ contain the same information as the vector of observables $\underline{y}$ does. Model (3.3) is premultiplied by the square mxm full rank transformation matrix T

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \begin{pmatrix} (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \\ B^T \end{pmatrix}$$

with T_1 a nxm matrix and T_2 a $(m-n)xm$ matrix. Model $E\{T\underline{y}\} = TAx$ then reads

(3.9) $$E\left\{\begin{pmatrix} \hat{\underline{x}} \\ \underline{t} \end{pmatrix}\right\} = \begin{pmatrix} I_n \\ 0 \end{pmatrix} x \quad ; \quad D\left\{\begin{pmatrix} \hat{\underline{x}} \\ \underline{t} \end{pmatrix}\right\} = \begin{pmatrix} Q_{\hat{x}} & \\ & Q_t \end{pmatrix}$$

where $Q_t = B^T Q_y B$. All information is preserved by the transformation matrix T in (3.9). It is just an invertible linear transformation. Design matrix A is transformed such that its columns $TA = \{e_1, .., e_n\}$. The vectors e_i, with $i = 1, .., m$ constitute the natural basis of R^m. The columns of T_1^T span $R(Q_y^{-1}A)$ and those of T_2^T span $R(B)$. Together the full matrix T^T provides a basis for the observation space R^m.

1.4 Hybrid parameter estimation

In section 2 the real, real in the sense of element from R, parameter estimation problem $y = Ax$ with $x \in R^n$ has been tackled using the least-squares criterion (2.3). The inverse variance covariance matrix Q_y^{-1} is the weight matrix and from (2.3) the weighted least-squares estimate follows: $\hat{x}$. The variance covariance matrix of the estimators reads $Q_{\hat{x}}$. The minimum of the objective function in (2.3) equals the squared norm of the least-squares residuals $\hat{e} = y - \hat{y}$:

(4.1) $$\|\hat{e}\|^2_{Q_y^{-1}} = \|y - A\hat{x}\|^2_{Q_y^{-1}}$$

In this section we will consider the hybrid estimation problem. The n-vector x and the mxn design matrix are partitioned as

$$x = \begin{pmatrix} x_R \\ x_I \end{pmatrix} \quad \text{and} \quad A = (A_R \; A_I)$$

with x_R : n_R-vector, A_R : mxn_R-matrix
x_I : n_I-vector, A_I : mxn_I-matrix

The hybrid estimation problem reads

$$y = (A_R\ A_I)\begin{pmatrix} x_R \\ x_I \end{pmatrix} \quad with \quad x_R \in R^{n_R}\ ;\ x_I \in Z^{n_I} \tag{4.2}$$

1.4.1 Estimation principle

The difference with (1.1) is that now both real and integer parameters are involved. The parameters in vector x_I are constrained to be integers. The least-squares criterion is applied to this hybrid estimation problem [Teunissen, 1995b]:

$$\min_{x_R, x_I} \|y - A_R x_R - A_I x_I\|^2_{Q_y^{-1}} \quad with \quad x_R \in R^{n_R}\ ;\ x_I \in Z^{n_I} \tag{4.3}$$

The objective function can be rewritten in, see [ibid]

$$\|y - A_R x_R - A_I x_I\|^2_{Q_y^{-1}} = \|\hat{e}\|^2_{Q_y^{-1}} + \|\hat{x}_I - x_I\|^2_{Q_{\hat{x}_I}^{-1}} + \|\hat{x}_{R|I} - x_R\|^2_{Q_{\hat{x}_{R|I}}^{-1}} \tag{4.4}$$

The estimation can be carried out in two steps: the unconstrained solution (also float solution) and based on the results of this step, the final solution to (4.3).

First the parameters are treated as being all reals. Problem (4.3) is solved with $x_R \in R^{n_R}$ and $x_I \in R^{n_I}$. The (ordinary) least-squares estimates follow:

$$\begin{pmatrix} \hat{x}_R \\ \hat{x}_I \end{pmatrix}\ ;\ \begin{pmatrix} Q_{\hat{x}_R} & Q_{\hat{x}_R\hat{x}_I} \\ Q_{\hat{x}_I\hat{x}_R} & Q_{\hat{x}_I} \end{pmatrix} \tag{4.5}$$

The first term on the right hand side of (4.4) gives the corresponding minimum of the objective function: $\|\hat{e}\|^2_{Q_y^{-1}}$. The second term equals zero as we consider the estimate $\hat{x}_I$ to be the final estimate: $x_I = \hat{x}_I$. The third term contains the conditional estimate for x_R: $\hat{x}_{R|I}$. It is the estimate for x_R, with the integer parameters conditioned on, or constrained to, x_I. For solution (4.5) we have $x_R = \hat{x}_{R|I} = \hat{x}_R$ and the third term equals zero. The solution $\hat{x}_R, \hat{x}_I$ is also referred to as float solution: the estimates are all floating point numbers. The geometry of the float solution is given in figure 4.1.

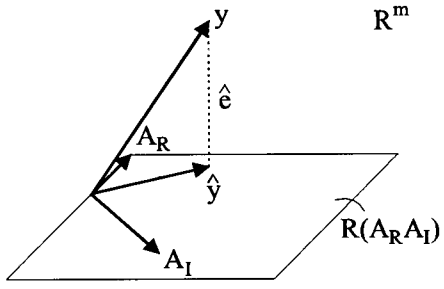


Figure 4.1: Geometry of float solution in R^m

The second step is the integer minimization:

(4.6) $$\min_{x_I} \|\hat{x}_I - x_I\|^2_{Q_{\hat{x}_I}^{-1}} \quad with \quad x_I \in Z^{n_I}$$

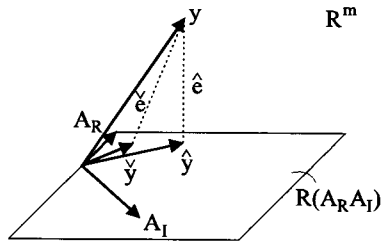
This problem will be dealt with in detail in the next section. The result reads $\check{x}_I$, the integer least-squares estimate for x_I. The minimum of the objective function (4.4) is increased to $\|\hat{e}\|^2_{Q_y^{-1}} + \|\hat{x}_I - \check{x}_I\|^2_{Q_{\hat{x}_I}^{-1}}$. The third term still equals zero with $x_R = \hat{x}_{R|I} = \check{x}_R$, unless additional constraints are put upon the parameters in x_R, see also figures 4.2 and 4.3. Vector $\check{x}_R$ is the estimate for x_R, computed with x_I conditioned on estimate $\check{x}_I$:

(4.7) $$\check{x}_R = \hat{x}_R - Q_{\hat{x}_R\hat{x}_I} Q_{\hat{x}_I}^{-1} (\hat{x}_I - \check{x}_I)$$

The estimates $\check{x}_R$ and $\check{x}_I$ constitute the (final) least-squares solution to (4.3) and is referred to as fixed solution: x_I has been constrained, or fixed, to integer values Z^{n_I} (but not to one particular deterministic value!). The variance covariance matrix of the estimator for x_R with the vector of integer parameters constrained to some deterministic value x_I

(4.8) $$Q_{\hat{x}_{R|I}} = Q_{\hat{x}_R} - Q_{\hat{x}_R\hat{x}_I} Q_{\hat{x}_I}^{-1} Q_{\hat{x}_I\hat{x}_R}$$

is taken as an approximation to $Q_{\check{x}_R}$. The stochastics of integer estimator $\underline{\check{x}}_I$ are neglected, see section 6. In practice the approximation is allowed if sufficient probability mass is located at $\check{x}_I$. The precision description by $Q_{\hat{x}_{R|I}}$ will be (slightly) too optimistic, see [ibid].

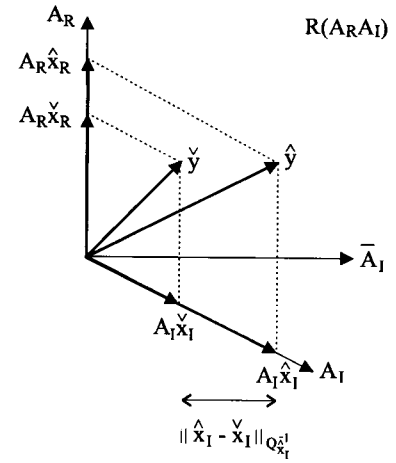


norm of $\hat{e} = \sqrt{\|\hat{e}\|^2_{Q_y^{-1}}}$

norm of $\check{e} = \sqrt{\|\hat{e}\|^2_{Q_y^{-1}} + \|\hat{x}_I - \check{x}_I\|^2_{Q_{\hat{x}_I}^{-1}}}$

Figure 4.2: Hybrid estimation in R^m

The final estimate for y reads $\underline{\check{y}} = A_R\check{x}_R + A_I\check{x}_I$ with $\check{e} = y - \check{y}$. As $(\hat{y} - \check{y}) \in R(A_I)$ it holds that $\|\hat{y} - \check{y}\|_{Q_y^{-1}} = \|\hat{x}_I - \check{x}_I\|_{Q_{\hat{x}_I}^{-1}}$.



$\bar{A}_I$ is A_I but orthogonalized with respect to A_R
in general $\hat{x}_I \notin Z^{n_I}$, but $\check{x}_I \in Z^{n_I}$

Figure 4.3: Hybrid estimation in the sub-space $R(A_R\ A_I)$ of R^m

In hybrid estimation the norm of the least-squares objective function (4.4) is increased as compared with real estimation, by the constraint of part of the parameters being integers. The second term in (4.4) may cause a correction $\|\hat{x}_I - x_I\|^2_{Q_{\hat{x}_I}^{-1}}$ even for the integer estima-

tion case $y=A_I x_I$ with no redundancy $m=n_I$; in this case $\hat{e}=0$. The non-consistency of a hybrid system $y=Ax$ with partly integer parameters, is handled by the least-squares principle. For $y=A_I x_I$, estimate $\check{y}$ is a discrete gridpoint in $R(A_I)$ such that $\check{x}_I \in Z^{n_I}$, see figure 4.4 and compare with figure 2.2. Non-uniqueness of the integer estimate is not considered here. This may occur (in theory) in the case with $y=A_I x_I$ even if the rank of the design matrix A_I equals n_I, see [Teunissen, 1993]. Non-uniqueness of the integer estimate due to a rankdeficient design matrix is discussed in [Teunissen, 1996b].

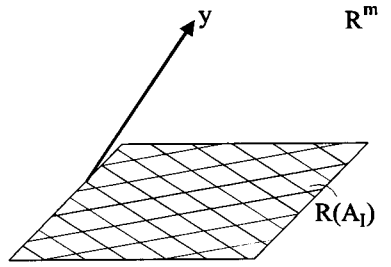


Figure 4.4: Geometric interpretation of integer estimation in R^m, the gridpoints are admissible locations for estimate for y

In the real case, $x \in R$, the value for x can range through the whole R. Computing an estimate is a true determination problem as there is an infinite (and uncountable) number of possible outcomes. In the integer case, $x \in Z$, the value for x is allowed to take on, as seen from R, only discrete values from Z. A-priori, all integers are equally likely. Estimation is now a discrimination problem as based on the observation, one specific value has to be chosen from a set of an infinite (but countable) number of integer values. In practice only a finite number of possibilities has to be considered.

1.4.2 Integer minimization

The second term of the objective function (4.4) needs to be minimized. A vector of integers x_I has to result from

$$\min_{x_I} \|\hat{x}_I - x_I\|^2_{Q_{\hat{x}_I}^{-1}} = \min_{x_I} (\hat{x}_I - x_I)^T Q_{\hat{x}_I}^{-1} (\hat{x}_I - x_I) \qquad \text{with} \quad x_I \in Z^{n_I} \tag{4.9}$$

No direct technique is available for the integer estimation. A search over discrete integer vectors has to be employed instead. For a gridpoint, it holds that all elements of x_I are in Z: $(x_I)_i \in Z$ for all $i=1,..,n_I$. In the n_I-dimensional space, the gridpoint x_I must be found, that has least (minimum) distance to estimate $\hat{x}_I$, in the metric $Q_{\hat{x}_I}^{-1}$. This gridpoint is the integer least-squares estimate $\check{x}_I$ sought for. In practice an appropriate value χ^2 is taken and only gridpoints will be considered, that have a distance equal to or less than χ^2:

$$(\hat{x}_I - x_I)^T Q_{\hat{x}_I}^{-1} (\hat{x}_I - x_I) \le \chi^2 \tag{4.10}$$

The geometric interpretation of (4.10) is the (hyper-) ellipsoid, see figure 4.5, where $n_I=2$. The ellipsoid is centered at estimate $\hat{x}_I$, its shape and orientation are governed by matrix $Q_{\hat{x}_I}$ and its size is controlled by χ^2. The value for χ^2 can be algebraically related to the volume of the ellipsoid and thereby approximately to the number of gridpoints, see [Jonge et al, 1996] and [Teunissen et al, 1996].

The sequential conditional least-squares adjustment interpretation of the search within the ellipsoid for the gridpoint that has the least distance to $\hat{x}_I$, is given in [Teunissen, 1995b]. The variance covariance matrix $Q_{\hat{x}_I}$ needs to be decomposed into a lower triangular L, a diagonal D and an upper triangular L^T: $Q_{\hat{x}_I} = LDL^T$, see [ibid]. The unknown parameters of x_I are successively fixed to integers. Based on value χ^2, the maximum allowable

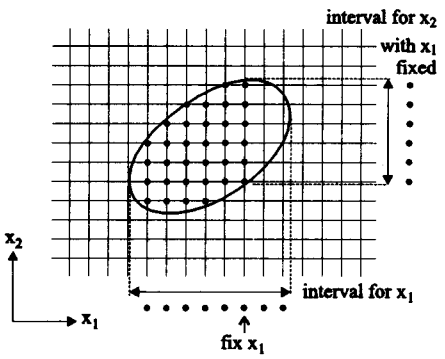


Figure 4.5: Search for gridpoints inside the ellipse, $n_I=2$ and $x_I=(x_1,x_2)^T$

distance, and $\sigma^2_{(\hat{x}_I)_1}$, the interval with valid integers for $(x_I)_1$ is obtained. Then $(x_I)_1$ is preliminary kept fixed to one of these integers. A new, improved estimator is computed for the next unknown parameter $(x_I)_2$, as if a new adjustment was carried out with $(x_I)_1$ known to be that integer; the conditional estimate for $(x_I)_2$ is obtained: $(\hat{x}_I)_{2|1}$.

Distance χ^2 is then reduced by $[(\hat{x}_I)_1 - (x_I)_1]^2 / \sigma^2_{(\hat{x}_I)_1}$ and together with $\sigma^2_{(\hat{x}_I)_{2|1}}$, the variance of the conditional estimator for $(x_I)_2$, the interval for $(x_I)_2$ is computed. It is centered at the conditional estimate $(\hat{x}_I)_{2|1}$. The process goes on with $(x_I)_3$ etc. It continues until all valid integers for all parameters $(x_I)_1$ through $(x_I)_{n_I}$ have been treated, see [Jonge et al, 1996]. The two-dimensional geometric interpretation of the search is given in figure 4.5. The ellipse is scanned through systematically. All gridpoints which are inside, are encountered in this way. Afterwards, the gridpoint with the smallest distance to $\hat{x}_I$ can be selected easily.

The search described here, will provide the integer least-squares estimate $\check{x}_I$. When the variance covariance matrix $Q_{\hat{x}_I}$ is ill conditioned, the ellipsoid in the R^{n_I} will be very out-stretched. The ellipsoid is very elongated; it has some axes which are extremely large, whereas the remaining ones are very short. Correlation between the estimators causes that the ellipsoid has its principal axes not parallel with the gridlines. Poor precision of the estimators in $\hat{x}_I$ and high correlation between them, causes typically, in two dimensions, an ellipse as shown in figure 4.6. If the 2-by-2 matrix is rank-deficient and has rank 1 only, the ellipse collapses even to a single straight line interval [Teunissen, 1996a].

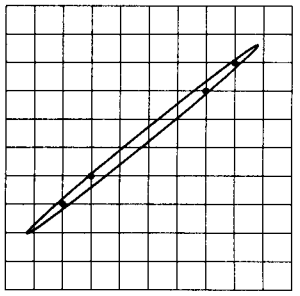


Figure 4.6: Search is inefficient for elongated and rotated ellipse

As in the search, actually intersections of the ellipsoid with the gridlines have to be computed, the search will be highly inefficient for such an elongated ellipsoid. The computation of many intervals is not effective as they do not contain any integer. This phenomena is explained by means of the spectrum of conditional variances in [Teunissen et al, 1994a] and [Teunissen et al, 1994b] for the sequential conditional adjustment of GPS double difference ambiguities. The discontinuity in the spectrum causes that the intervals for the first few parameters are very large, and therefore a lot of integers have to be tried out, whereas the remaining intervals are all very small. Many of them do not contain any integer. A lot of incomplete integer-vectors for x_I are to be encountered. The implication for the search is discussed in detail in [Teunissen, 1994c]. The efficiency of the search can be largely enhanced by a decorrelation of the estimators in $\hat{x}_I$. This decorrelation is realized by means of an integer reparametrization in the next section.

1.4.3 Integer reparametrization

In [Teunissen, 1993] the decorrelating Z-transformation is introduced: the integer unknown parameters in x_I are reparametrized according to

$$z_I = Z^T x_I \tag{4.11}$$

where matrix Z^T is a square $n_I x n_I$ full rank transformation matrix. The estimate and variance covariance matrix are transformed accordingly

$$\hat{z}_I = Z^T \hat{x}_I \qquad Q_{\hat{z}_I} = Z^T Q_{\hat{x}_I} Z \tag{4.12}$$

and the minimization problem (4.9) becomes

$$\min_{z_I} (\hat{z}_I - z_I)^T Q_{\hat{z}_I}^{-1} (\hat{z}_I - z_I) \quad with \quad z_I \in Z^{n_I} \tag{4.13}$$

and will result in $\check{z}_I$. The integer least squares estimate for x_I is then obtained by $\check{x}_I = Z^{-T} \check{z}_I$. The estimate $\check{x}_I$ obtained via $\check{z}_I$ from (4.13), is identical to the one obtained from solving minimization (4.9) directly, thanks to the requirements upon the Z-transformation. In addition to being square and of full rank, this transformation must be integer and volume preserving, see [Teunissen, 1995c].

The purpose of the Z-transformation is to decorrelate the estimators in $\underline{\hat{x}}_I$. The computation of matrix Z^T consists of alternatingly, decorrelation steps (integer Gauss transformation) and interchangings of parameters to allow further decorrelation (re-ordering/permutation). With the $Q_{\hat{x}_I} = LDL^T$ decomposition, see [Teunissen, 1995b], the variance covariance matrix of $\hat{z}_I$ becomes

$$Q_{\hat{z}_I} = Z^T L D L^T Z \tag{4.14}$$

In the hypothetical case that full decorrelation is achieved, $Z = L^{-T}$, matrix (4.14) reduces to

$$Q_{\hat{z}_I} = D \tag{4.15}$$

Then the most simple integer estimation scheme can be used: rounding to the nearest integer, $(\check{z}_I)_j = nint(\hat{z}_I)_j$ for all $j = 1,..,n_I$; no search is needed at all.

In general, full decorrelation is ruled out. As matrix Z^T has to fulfil the integer and volume preserving conditions, one has to be satisfied with an integer approximation of $Z = L^{-T}$. It turns out that in practice, even under the aforementioned conditions, the Z-transformation can largely reduce the correlation between the estimators, and thereby greatly facilitate the search, see e.g. [Teunissen et al, 1994a] and [Teunissen et al, 1994b]. The geometric interpretation of the decorrelating Z-transformation is that the elongated search ellipsoid of the original parameters x_I is transformed into a search ellipsoid for z_I that is much more sphere-like. The intervals for the parameters in z_I in the sequential conditional adjustment, are equalized. The integer estimation for (4.13) can be carried out much more efficiently than it can for (4.9).

1.4.4 LAMBDA-method

The integer minimization of section 1.4.2 and the integer reparametrization of section 1.4.3 are all part of the LAMBDA-method proposed in [Teunissen, 1993]. In summary the method consists of two steps:

1) the decorrelating Z-transformation
2) the actual integer estimation (the search)

Both steps are based on the sequential conditional least-squares adjustment of the unknown integer parameters. Based on the results of the (ordinary) least-squares estimation (4.5), with both x_R and x_I treated as real-valued parameters, the integer minimization (4.9) is carried out using the LAMBDA-method. The result is $\check{x}_I$, the integer least-squares estimate for the integer parameters in x_I.

1.5 Non-linear model y=A(x)

In this section we will review the estimation in case of a non-linear functional model. Direct non-linear estimation is usually not feasible. We will use the Gauss-Newton method [Teunissen, 1990a]. The model is linearized and BLUE estimators for the unknowns of this model are computed; they lead to approximations for the non-linear estimators. For the linearization, approximate values for the unknown parameters are needed and these are improved in an iterative process.

1.5.1 Non-linear functional model

In section 3 the model of observation equations was given as (3.3):

$$E\{y\} = Ax \quad ; \quad D\{y\} = Q_y \tag{5.1}$$

The expectation value of the vector of observables and the unknowns in vector x are linearly related. Based on this model, Best Linear Unbiased Estimators could be computed. They are identical to the weighted least-squares estimators. In practice non-linear models of observation equations are frequently encountered.

$$E\{y\} = A(x) \quad ; \quad D\{y\} = Q_y \tag{5.2}$$

We are, however, still looking for the numerical estimate for x and for the probability density function of the (non-linear) estimator, the latter for the reason of quality control [Teunissen, 1989]. For the BLUE-estimation principle, the first two moments of the estimator $E\{\hat{x}\}$ and $D\{\hat{x}\}$ suffice. With Gaussian distributed observables and a linear estimator, these two moments completely specify the probability density function of the estimator, see also section 6.

In dealing with non-linear functional models, two complications arise. First the (non-linear) estimator must satisfy the least-squares criterion

$$\min_x \|y-A(x)\|^2_{Q_y^{-1}} \tag{5.3}$$

If the map $A(.)$ is non-linear, then generally no direct methods exist for solving (5.3). In other words we will not succeed in finding the non-linear estimator. We will approximate the non-linear model with its linearized version and use an iterative algorithm to compute the estimate. Iterative algorithms to solve (5.3) are discussed in [Teunissen, 1990a]. Secondly, we are not able to propagate the probability density of $\underline{y}$ into the probability density of the non-linear estimator $\underline{\hat{x}}$, [Teunissen, 1989]. The non-linearity also causes that the estimator is biased, i.e. $E\{\underline{\hat{x}}\} \neq x$. In this section, it is reviewed how the non-linear least-squares problem is usually handled.

1.5.2 Linearization

For varying values of x, $A(x)$ traces an n-dimensional surface or manifold embedded in the R^m. If the manifold $A(x)$ is rather smooth, the linearized model can be a sufficient approximation of the non-linear model. With x_o being a good approximate value for the estimate sought for, the Taylor expansion of (5.2) reads

$$E\{\underline{y}\} = A(x_o) + \left.\frac{\partial A(x)}{\partial x}\right|_{x_o} \Delta x + H.O.T. \tag{5.4}$$

and the linearized model becomes

$$E\{\Delta\underline{y}\} = \partial_x A(x_o)\,\Delta x \;;\; D\{\Delta\underline{y}\} = Q_y \tag{5.5}$$

with $\Delta\underline{y} = \underline{y} - A(x_o)$. The higher order terms are neglected, see also [Teunissen, 1994a].

1.5.3 BLUE-estimation

Based on the linearized model (5.5), we will compute, using the BLUE principle, $\Delta\hat{x}$ and $Q_{\Delta\hat{x}}$.

$$\Delta\hat{x} = [\partial_x A(x_o)^T Q_y^{-1} \partial_x A(x_o)]^{-1} \partial_x A(x_o)^T Q_y^{-1}\, \Delta y \tag{5.6}$$

$$Q_{\Delta\hat{x}} = [\partial_x A(x_o)^T Q_y^{-1} \partial_x A(x_o)]^{-1} \tag{5.7}$$

and

$$x_n = x_o + \Delta\hat{x} \tag{5.8}$$

then we obtain

(5.9) $$\hat{x} = x_n \;\; ; \;\; Q_{\hat{x}} = Q_{\Delta\hat{x}}$$

which are approximations to the first two moments of the non-linear estimator [Teunissen, 1989]. The geometric interpretation of this estimation is, that the observation vector y is orthogonally projected (metric Q_y^{-1}) on the tangent space of $A(x)$. The tangent space, plane $\partial_x A(x_o)$, is suspended at $A(x_o)$.

1.5.4 Gauss-Newton method

The approximate value x_o will (slightly) deviate from the (non-linear) estimate sought for. In order to improve the numerical estimate (5.8), iterations have to be made. The geometric interpretation is that the tangent plane has to be put into the right (a better) position. The better the approximate value, the smaller the higher order terms in (5.4), which are neglected in the current estimation. The linearized model will, on a local scale, better approximate the non-linear one. As a result, a better numerical estimate for x is obtained. The Gauss-Newton iteration scheme is employed [Teunissen, 1990a]. The estimate for x (5.8) is taken as the new approximate value in (5.4) and a new estimate (5.8) is computed. This procedure is continued until the convergency test (equation (71) in [ibid]) is passed

(5.10) $$\|x_n - x_o\|_{Q_{\hat{x}}^{-1}} < \epsilon$$

The criterion (5.10) is invariant under a reparametrization of the unknowns x. The estimate is finally set to $\hat{x} = x_n$, (5.9).

The necessary and sufficient conditions for the Gauss-Newton method in order to converge to a (local) minimum of the sum of squares objective function $\|y - A(x)\|^2_{Q_y^{-1}}$, the least-squares estimation criterion (5.3), are given by equation (57) in [ibid]. In practice the curvature of the manifold $A(x)$ must be small and the norm of the residual vector $\hat{e} = y - \hat{y}$ should not be too large. Furthermore, as the Gauss-Newton method belongs to the class of iterative descent methods, the initial guess x_o should be sufficiently close to the solution $\hat{x}$, that has to result from (5.3), see [ibid]. The closer x_o is to the estimate $\hat{x}$, the less iterations will be needed to satisfy the criterion (5.10).

The parameters x are involved in a non-linear way in the vector y-$A(x)$ of which the (squared) norm has to be minimized in (5.3). In [Teunissen, 1989] two types of non-linearity are distinguished: intrinsic non-linearity (curvature of the manifold) and non-linearity of the parameter curves. The non-linearity causes that the estimator obtained (using the procedure in this section), is biased. The non-linearity should be diagnosed in order to assure that the linearized model is a sufficient approximation of the non-linear one. With equation (40) of [ibid], the bias in the estimator can be estimated.

1.6 Estimation with stochastic observations

In the sections 2 and 3, we concentrated on the determination of the values for the unknowns once a set of observations is given. Now we will take into account the stochastics of the observables. The observation y is a sample out of the population $\underline{y}$. The vector of observables $\underline{y}$ contains m random variables. The probability density function of $\underline{y}$

is given by $p_y(y|\mu)$. The kernel μ in this specification, denotes that not all parameters of this function are known yet. In the model of observation equations $E\{\underline{y}\}=Ax$ for instance, the first moment, the expectation, of $\underline{y}$ is unknown in x.

In estimation, a n-vector of values for the unknowns in x is assigned, based on the sample y. The estimator $\underline{\hat{x}}$ for the vector of unknowns, is a linear function of the observables. The functional relation can also be used to propagate the probability density function of $\underline{y}$ into the one for $\underline{\hat{x}}$: $p_{\underline{\hat{x}}}(\hat{x}|\mu)$. Function $p_{\underline{\hat{x}}}(\hat{x}|\mu)$ describes the probability density of each value for $\hat{x}$ and therefore translates into the quality of the estimator under the current modelling. The quality of the estimator may range from infinitely bad to infinitely good. These cases of no knowledge and perfect knowledge concerning the unknowns in vector x, correspond to a probability density function that is infinitely wide and to one that has one infinitely narrow peak respectively. These functions are sketched in figure 6.1 for the case $n=1$. In the last case, the estimator is non-stochastic. It is a deterministic quantity.

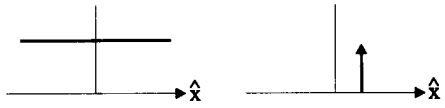


Figure 6.1: Two limiting cases of the probability density function, (left) all values have equal probability density, (right) one (deterministic) value has infinitely large probability density

In this section we will consider the estimation for a model with observation equations with real valued parameters and for a model with integer parameters. In both cases we will elaborate on the probability density function of the estimator. Finally we will discuss the so-called 'fixing' of parameters. A certain estimator is considered to be a deterministic quantity and this implies an approximation to the probability density function. For a treatment in depth on probability theory, the reader is referred to [Koch, 1988] and [Maybeck, 1979].

In the data processing we will carry over the assumption of Gaussian distributed observables. The justification follows [ibid]. With a functional model that is adequate, the measurement noise is typically caused by a number of small effects. When a number of independent random variables are added together, the summed effect can be described very closely by a Gaussian probability density, regardless of the shape of the individual densities (central limit theorem). In other words, the Gaussian distribution will serve as a good approximation in practice. Alternative distributions generally do not yield formulations of the data processing problem that are mathematically tractable.

1.6.1 Real valued parameters

The Gaussian distribution and the least-squares criterion very well fit together. If the observables possess a Gaussian distribution, the estimator obtained with the least-squares criterion, is identical to the Maximum Likelihood estimator. With the probability density function of the observables as $\underline{y}\sim N(Ax,Q_y)$, the probability density function of the estimators is also Gaussian $\underline{\hat{x}}\sim N(x,Q_{\hat{x}})$, as a consequence of the estimator being linearly related to the observables. The estimator is unbiased and has variance covariance matrix $Q_{\hat{x}}$. With the BLUE-estimation principle only the first two moments of $\underline{y}$ are considered, independent of the acual distribution. If $\underline{y}\sim N$, they completely specify the probability density function.

We start considering the estimation for the following model of observation equations

$$E\{\underline{y}\} = Ax \quad ; \quad D\{\underline{y}\} = Q_y \tag{6.1}$$

with $x \in R^n$ and $\underline{y}$ Gaussian, or normally distributed, $\underline{y} \sim N(Ax, Q_y)$

$$p_{\underline{y}}(y|x) = \frac{1}{(2\pi)^{\frac{m}{2}}\sqrt{|Q_y|}} e^{-\frac{1}{2}(y-Ax)^T Q_y^{-1}(y-Ax)} \tag{6.2}$$

The probability density function of the estimator $\hat{\underline{x}}$ reads

$$p_{\hat{\underline{x}}}(\hat{x}|x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|Q_{\hat{x}}|}} e^{-\frac{1}{2}(\hat{x}-x)^T Q_{\hat{x}}^{-1}(\hat{x}-x)} \tag{6.3}$$

with

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

and

$$D\{\hat{\underline{x}}\} = Q_{\hat{x}}$$

An example of function (6.3) is given in figure 6.2. The estimate $\hat{x}$ is a sample out of this population.

Function (6.3) describes the probability density of obtaining value $\hat{x}$, given that x is the expectation value for $\hat{\underline{x}}$, which is usually not known. The following multiple integral gives the probability that estimate $\hat{x}$ lies in the so-called confidence region R:

$$P[\hat{x} \in R] = \int_R \ldots \int p_{\hat{\underline{x}}}(\hat{x}|x)\, dx_1 \ldots dx_n = 1-\alpha \tag{6.4}$$

As $\hat{x}$ can take on values from $<-\infty,\infty>$, regions like (6.4) can be applied to, a (finite) shift over $x-\hat{x}$. In this way one can arrive at a fictitious probability density function of $\underline{x}$, that is centered at $\hat{x}$:

$$p_{\underline{x}}(x|\hat{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|Q_{\hat{x}}|}} e^{-\frac{1}{2}(x-\hat{x})^T Q_{\hat{x}}^{-1}(x-\hat{x})} \tag{6.5}$$

which gives the probability density of x being the expectation value, given the estimate $\hat{x}$. Function (6.5) is used in practice, to set confidence regions for parameters x.

1.6.2 Integer parameters

In this section we start by briefly discussing the discrete probability density function, proposed in [Blewitt, 1989]. Next we point out how the probability density function of the integer estimator should be obtained. For the reason of readibility we will discuss the following, very simple, model of observation equations:

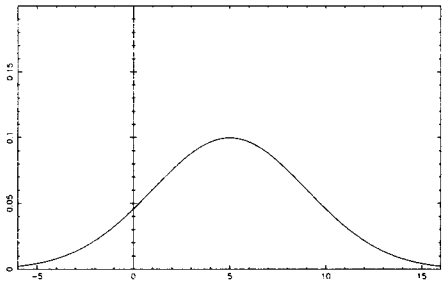


Figure 6.2: Probability density function $p_{\underline{\hat{x}}}(\hat{x}|x)$ for one-dimensional x: $x=5$ and $\sigma=4$

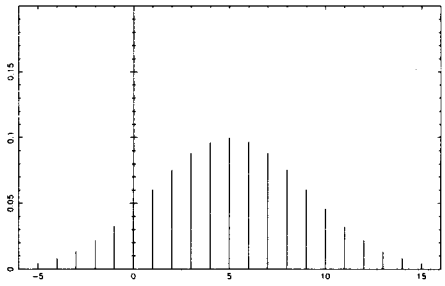


Figure 6.3: Probability density function $p_{\underline{\check{x}}}(\check{x}|x)$ (6.8) for one-dimensional x: $x=5$ and $\sigma=4$

(6.6) $$E\{\underline{y}\} = x \;\; ; \;\; D\{\underline{y}\} = \sigma^2$$

with $x \in Z$ and $\underline{y}$ Gaussian, or normally distributed, $\underline{y} \sim N(x, \sigma^2)$

(6.7) $$p_{\underline{y}}(y|x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-x)^2}{2\sigma^2}}$$

According to [ibid], the probability density function of the integer estimator $\underline{\check{x}} = nint(\underline{y})$, reads

(6.8) $$p_{\underline{\check{x}}}(\check{x}|x) = \frac{1}{\sum_{j=-N}^{N} e^{-\frac{(j-x)^2}{2\sigma^2}}} e^{-\frac{(\check{x}-x)^2}{2\sigma^2}} \sum_{j=-N}^{N} \delta(\check{x}-j)$$

with N some large integer and $\delta(.)$ the Dirac delta function (or impulse function). This function (6.8) describes the probability density of obtaining value $\check{x}$. The function is centered at the integer value x, which is usually not known. An example of function (6.8) is given in figure 6.3. The Dirac delta function was set to $\delta(\tau=0)=1$ and $N=20$ was taken. Function (6.8) is a discrete one, $p_{\underline{\check{x}}}(\check{x}|x)=0$ for all $\check{x} \notin Z$ and anywhere outside the interval $[-N,N]$. The expectation $E\{\underline{\check{x}}\}$ and dispersion $D\{\underline{\check{x}}\}$ of (6.8) can be given; it will possess higher order moments as well.

Probability density function (6.8) can be shown to satisfy the two requirements $\int_{-\infty}^{\infty} p_{\underline{\check{x}}}(\check{x}|x)d\check{x} = 1$ and $p_{\underline{\check{x}}}(\check{x}|x) \geq 0 \;\forall \check{x}$. A generalization of the function to n dimensions is possible. In a way similar as in section 1.6.1, the following probability density function can be found

(6.9) $$p_{\underline{\check{x}}}(x|\check{x}) = \frac{1}{\sum_{j=-N}^{N} e^{-\frac{(j-\check{x})^2}{2\sigma^2}}} e^{-\frac{(x-\check{x})^2}{2\sigma^2}} \sum_{j=-N}^{N} \delta(x-j)$$

which gives the probability density that x is the expectation value, given that the integer estimate is $\check{x}$. Equation (6.9) is similar to formula (A.4) of [ibid].

It can be seen from (6.8) that, apart from the Dirac δ-function and a weighting (the denominator), the value of $p_{\check{x}}$ at some integer $\check{x}=\bar{x}$ just equals the value of the exponential part of the real counterpart $p_{\hat{x}}$, (6.7) (here $p_{\hat{x}}=p_y$), or (6.3) for a one-dimensional x, and this may appear strange. The probability density function of the integer estimator should be obtained by *integration* upon the function of the real-valued estimator. The integer estimation criterion prescribes all real valued samples $\hat{x}$ in the interval $[\bar{x}-\frac{1}{2},\bar{x}+\frac{1}{2}>$ to be mapped onto the integer estimate $\bar{x}$, with $\bar{x}\in Z$, [Teunissen, 1990b]. The corresponding operation of this nearest integer rounding applied to $p_{\hat{x}}$ yields the probability $P[\check{x}=\bar{x}]$, see also figure 6.4,

$$(6.10)\quad p_{\check{x}}(\check{x}=\bar{x}) = \int_{z=\bar{x}-\frac{1}{2}}^{z=\bar{x}+\frac{1}{2}} p_{\hat{x}}(z|x)\,dz \qquad \bar{x}\in Z$$

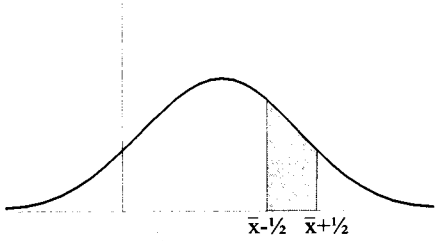


Figure 6.4: Probability density function of integer estimator obtained by integration

The function for the real-valued estimator $\hat{x}$ can be readily obtained as described in the previous section. Integer estimation, however, implies some (non-explicit) non-linear operation (in this example $\check{x}=nint(y)$) and hence the normal distribution of y can not be propagated. The second problem is, provided that (6.8) indeed is the probability density function of the integer estimator, the handling of such a function in further data processing.

We have treated two cases: estimation with real valued parameters and with integer parameters. Also the hybrid case may occur. Some of the parameters are integers and therefore still the same difficulties as described in this section are present. In the next section we will discuss how the problem of the discrete probability density function is tackled in practice.

1.6.3 Fixing of parameters

In practical data processing, integer parameters need special treatment. Estimates for the integer parameters can be obtained by rigorous application of the least-squares principle. This concerns the numerical value for the unknowns and nothing is known yet about the statistical moments of the estimator. The integer estimator $\check{x}$ has a discrete probability density function and this hinders model validation by means of statistical tests currently in use (see section 7) and causes difficulties in further data processing (precision description is troublesome). We will discuss the so-called fixing of parameters. This provides a workable approximation to the stochastics of the estimators for these parameters in further processing.

fixing of real parameters

For model (6.1), least-squares estimation provides the most likely value for x. We assume that the model, used in the estimation, has been validated. The value for the estimate can range through the whole R^n. The probability density function of the estimator is $p_{\hat{x}}(\hat{x}|x)$. If we would like to treat $\hat{x}$ as a deterministic quantity, we have to analyse the variance covariance matrix $Q_{\hat{x}}$. If all elements of $Q_{\hat{x}}$ are sufficiently small, we are allowed to

approximate this matrix by $Q_{\hat{x}}=0$ and continue data processing with $\hat{\underline{x}}$ as a deterministic quantity: $x \doteq \hat{x}$.

The probability density function $p_{\underline{x}}(x|\hat{x})$, see (6.5), is approximated by a probability density function with all probability density mass located at the real-valued estimate $\hat{x}$.

$$p_{\underline{x}}(x) = \delta(x-\hat{x}) \tag{6.11}$$

with the Dirac delta function generalized to n-dimensions. Function $p_{\underline{x}}(x)$ equals zero for all values $x \neq \hat{x}$. By the definition of the Dirac delta function, (6.11) satisfies the two requirements for a probability density function in general.

fixing of integer parameters
For model (6.1) with $x \in Z^n$, integer least-squares estimation provides the most likely integer value for x. Again we assume that the mathematical model has been validated. The values the estimate $\check{x}$ can take on, are restricted to the Z^n, i.e. discrete points in the R^n. This does, however, not yet imply that the estimator is a deterministic quantity. The probability density function of the estimator is $p_{\check{\underline{x}}}(\check{x}|x)$, an n-dimensional generalization of a discrete function like (6.8). Such a function is not convenient for use in further data processing and therefore we would like to treat $\check{\underline{x}}$ as a deterministic quantity. The discrete probability density function, for one dimension depicted in figure 6.3, should sufficiently resemble the single Dirac delta function; there should be one clear peak, as is the case in figure 6.1 on the right. This means that the integer estimate should represent clearly more probability density mass than any other integer vector in Z^n. Means to certify this statement will be treated in appendix A.

The probability density function $p_{\underline{x}}(x|\check{x})$ is then approximated by a function with all probability density mass located at the integer estimate $\check{x}$.

$$p_{\underline{x}}(x) = \delta(x-\check{x}) \tag{6.12}$$

Function $p_{\underline{x}}(x)$ equals zero for all values $x \neq \check{x}$. The data processing is now continued with $\check{\underline{x}}$ as a deterministic quantity $x \doteq \check{x}$. The numerical value of the integer least-squares estimate, which *is* integer, is not changed.

1.7 Model validation

In the previous section, the full probability density function of the observables $\underline{y}$ was specified: $p_{\underline{y}}(y|\mu)$. This assumed modelling of the measurement system embedded in the real world, may not be a completely correct, or adequate description. In that case, when relevant discrepancies between reality and mathematical model occur, the estimates and probability density function of the estimator, computed with the assumed model, the null hypothesis, are invalidated. The optimality property of the estimator $\hat{\underline{x}}$, obtained with the BLUE principle, can be guaranteed only under the null hypothesis H_o.

Parallel to the estimation, one has to check the validity of the null hypothesis. Therefore a testing procedure will be carried out. It is based on the theory of hypothesis testing in linear models and testing is performed with the generalized likelihood ratio test. With a

good testing procedure, it is guaranteed, to some extent, that the mathematical model used in the estimation H_0 is adequate. It is realized that the perfect model simply does not exist. According to [Teunissen, 1997] the question of the model's validity concerns whether the mathematical model, given its intended purposes, is sufficiently consistent with the data observed.

1.7.1 Principle of testing

The purpose of statistical testing is to decide between alternative mathematical models, hypotheses. A hypothesis H is a specification of the probability density function of $\underline{y}$: $p_{\underline{y}}(y|\mu)$ and the hypotheses differ in the parameters μ. Given the observations y, a sample of $\underline{y}$, it must be inferred, which out of two models, the null hypothesis H_0 or another hypothesis, is most likely.

The decision, whether to accept or reject the null hypothesis H_0, is made using a teststatistic. Two possible decisions out of four are incorrect: the probabilities are denoted by α and β in table 7.1. The power, the probability of rejecting H_0 when indeed H_0 is false, is defined as $\gamma=1-\beta$.

reality / decision	H_0 true	H_0 false
accept H_0	OK	β
reject H_0	α	OK

Table 7.1: Four possibilities in testing

Ideally both α and β equal zero, but this can not be achieved. When the probability of false alarm α is specified, power γ is maximized using the Neyman-Pearson testing principle [Teunissen, 1994b]. This yields the teststatistic $\underline{T}$, which is a function of the vector of observables $\underline{T}(\underline{y})$, and the critical region K for the teststatistic, which is among other parameters a function of α. The test reads:

(7.1) if $T \in K \rightarrow$ reject H_0
$T \notin K \rightarrow$ accept H_0 (H_0 not rejected)

1.7.2 Generalized Likelihood Ratio Teststatistic

We will check the validity of H_0, the model used in the estimation, by opposing it against an alternative hypothesis H_a. Both hypotheses are composite ones, i.e. probability density function $p_{\underline{y}}(y|\mu)$ is left partly unspecified. The teststatistic $\underline{T}_q$ is uniformly most powerful invariant (UMPI). It satisfies the Neyman-Pearson principle under the restriction of invariance of the critical region for y under orthogonal transformations [Teunissen, 1994b]. Uniformly means that it is most powerful for all alternative values for the parameters of μ in the set $\Phi\backslash\Phi_o$. The set Φ with values for μ, is divided into Φ_o for the H_0 and the complementary set $\Phi\backslash\Phi_o$ for H_a:

(7.2)
$$H_0: \quad \mu \in \Phi_o$$
$$H_a: \quad \mu \in \Phi\backslash\Phi_o$$

The hypotheses H_0 and H_a are exclusive. The decision with the generalized likelihood ratio then becomes

$$\text{(7.3)} \qquad \text{reject } H_o \text{ if } \quad \frac{\max\limits_{\mu\in\Phi_o} p_y(y|\mu)}{\max\limits_{\mu\in\Phi} p_y(y|\mu)} < a$$

with $a \in <0,1>$.

In this ratio (7.3), the maximum attainable likelihood of the observations y being produced by model H_o and the maximum attainable likelihood of H_a are compared. When the ratio leads to rejection of H_o in favour of H_a, the probability density function $p_y(y|\mu)$ under H_o is decided not to reflect the random character of $\underline{y}$; H_a is found to be a more adequate description.

In the sequel, hypothesis testing is applied to models of observation equations and restricted to misspecifications in the linear functional model. As compared with the null hypothesis, the functional model of the alternative hypothesis is extended with q unknown parameters in vector ∇, with $1 \leq q \leq m\text{-}n$. Based on the observations y, it will be decided that H_o is an adequate model or that additional parameters are needed in the description of $E\{\underline{y}\}$. The stochastic model is assumed to be completely and correctly specified. The observables $\underline{y}$ are normally distributed.

$$\text{(7.4)} \qquad \begin{aligned} H_o: \quad & \underline{y} \sim N(Ax, Q_y) && E\{\underline{y}\} = Ax && ; \quad D\{\underline{y}\} = Q_y \\ H_a: \quad & \underline{y} \sim N(Ax + C_y\nabla, Q_y) && E\{\underline{y}\} = Ax + C_y\nabla && ; \quad D\{\underline{y}\} = Q_y \end{aligned}$$

By the set Φ_o of the null hypothesis H_o, the observation y, in the observation space R^n, is restricted to lie in the n-dimensional space $R(A)$. According to H_a, the observation does not lie in the $R(A)$, but is allowed to lie in the space $R(A : C_y)$. In terms of the model with condition equations we have

$$\text{(7.5)} \qquad \begin{aligned} H_o: \quad & && B^T E\{\underline{y}\} = 0 && ; \quad D\{\underline{y}\} = Q_y \\ H_a: \quad & && B^T E\{\underline{y}\} = C_t\nabla && ; \quad D\{\underline{y}\} = Q_y \end{aligned}$$

with $C_t = B^T C_y$. Specification (7.5) is equivalent to (7.4) for the purpose of testing with the generalized likelihood ratio teststatistic. The hypotheses H_o and H_a of (7.4) are composite hypotheses: the probability density function of $\underline{y}$ is specified with the parameters x left unknown for H_o and also parameters ∇ for H_a.

The generalized likelihood ratio test for testing H_o against H_a, see [ibid], leads to

$$\text{(7.6)} \qquad \text{reject } H_o \text{ if } \quad T_q > \chi^2_\alpha(q,0)$$

with teststatistic

$$\text{(7.7)} \qquad \underline{T}_q = \underline{\hat{e}}^T Q_y^{-1} C_y [C_y^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y]^{-1} C_y^T Q_y^{-1} \underline{\hat{e}}$$

or

$$\underline{T}_q = \underline{t}^T Q_t^{-1} C_t [C_t^T Q_t^{-1} C_t]^{-1} C_t^T Q_t^{-1} \underline{t} \tag{7.8}$$

with

$$\underline{T}_q \sim \chi^2(q,0) \text{ under } H_o \tag{7.9}$$

and

$$\underline{T}_q \sim \chi^2(q,\lambda) \text{ under } H_a, \text{ with } \lambda = \nabla^T C_y^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y \nabla \text{ or } \lambda = \nabla^T C_t^T Q_t^{-1} C_t \nabla \tag{7.10}$$

$$\text{with } Q_{\hat{\nabla}}^{-1} = C_t^T Q_t^{-1} C_t$$

The most powerful property of teststatistic (7.8) implies, we cite [Teunissen, 1991], that - somewhat loosely formulated - the teststatistic has the property of correctly detecting and identifying model errors with the highest possible probability: in other words, when a certain model error is present (and specified in the alternative hypothesis) we have γ% probability of correctly detecting and identifying it by teststatistic (7.8) and no (other) teststatistic can do better.

Two special forms of teststatistic $\underline{T}_q$ are known:

- the Overall Model teststatistic, with $q=m-n$

$$\underline{T}_q = \hat{\underline{e}}^T Q_y^{-1} \hat{\underline{e}} = \|\hat{\underline{e}}\|^2_{Q_y^{-1}} = \underline{t}^T Q_t^{-1} \underline{t} = \|\underline{t}\|^2_{Q_t^{-1}} \tag{7.11}$$

- the so-called $\underline{w}$-teststatistic, with $q=1$

$$\underline{T}_q = \frac{(c_y^T Q_y^{-1} \hat{\underline{e}})^2}{c_y^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y} = \frac{(c_t^T Q_t^{-1} \underline{t})^2}{c_t^T Q_t^{-1} c_t} \tag{7.12}$$

of which the square root is used $\underline{t}^2 = \underline{T}_{q=1}$ and has a $\underline{t} \sim N(0,1)$ distribution under H_o (we use $\underline{t}$ instead of $\underline{w}$).

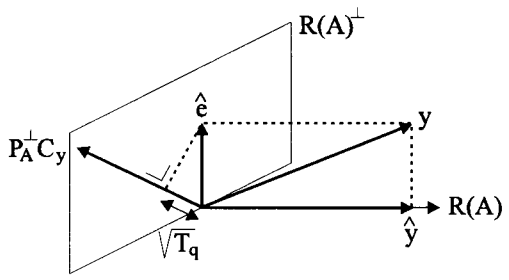


Figure 7.1: Testing in the observation space R^m, $\underline{T}_q = \|P_{P_A^\perp C_y} \hat{\underline{e}}\|^2_{Q_y^{-1}}$

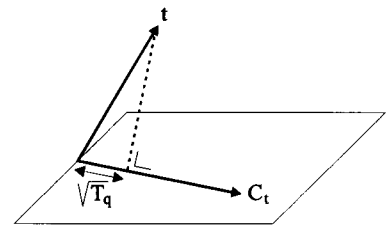


Figure 7.2: Testing in the misclosure space R^{m-n}, $\underline{T}_q = \|P_{C_t} \underline{t}\|^2_{Q_t^{-1}}$

Figure 7.1 clearly shows that hypothesis testing takes place in the observation space R^m. An orthogonal decomposition of y is made: $y=\hat{e}+\hat{y}$ where $\hat{e}\in R(A)^{\perp}$ and $\hat{y}\in R(A)$, with $\hat{e}=P_A^{\perp}y$ and $\hat{y}=P_A y$. The least-squares principle is used to make $y=Ax$ consistent and the deviation of y from model $R(A)$ is considered and judged in testing. If the deviation is significant, this indicates (but does not proof) that model H_o is false.

From figure 7.2 it is clear that the presence of redundant data, $m\text{-}n>0$, is crucial for testing. It is namely the surplus of information which enables us to test whether the data can be considered to be statistically consistent with the assumed model H_o [Teunissen, 1991].

If it is decided that H_a holds true instead of H_o the model used in the estimation, the estimator under the alternative hypothesis has to be computed:

(7.13)
$$\begin{aligned} \underline{\hat{x}}^a &= \underline{\hat{x}}^o - X\underline{\hat{\nabla}} \\ Q_{\hat{x}^a} &= Q_{\hat{x}^o} + XQ_{\hat{\nabla}}X^T \end{aligned}$$

where $\underline{\hat{x}}^o$ and $\underline{\hat{\nabla}}$ are not correlated, thus $Q_{\hat{x}^o\hat{\nabla}}=0$ and with the nxq-matrix X as

(7.14)
$$X = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} C_y$$

and the estimator for the vector $\underline{\hat{\nabla}}$ as

(7.15)
$$\underline{\hat{\nabla}} = (C_y^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y)^{-1} C_y^T Q_y^{-1} \underline{\hat{e}}$$

or

(7.16)
$$\underline{\hat{\nabla}} = (C_t^T Q_t^{-1} C_t)^{-1} C_t^T Q_t^{-1} \underline{t}$$

Note that rejecting H_o and adopting H_a, also yields a new vector of misclosures $\underline{t}^a$ with dimension $m\text{-}n\text{-}q$ (and also a new vector of residuals $\underline{\hat{e}}^a$).

1.7.3 Model validation with hybrid parameters

The generalized likelihood ratio teststatistic $\underline{T}$ introduced in the previous section, is frequently used in geodetic data processing. In some applications, beside real valued parameters, also integer parameters are involved. This has led to an extension of the estimation procedure. The presence of integer parameters has implications also for the testing procedure. The most general hypothesis with both real and integer parameters concerning misspecifications in the functional model reads

(7.17) H_o : $E\{\underline{y}\} = A_R x_R + A_I x_I$ $\quad H_a$: $E\{\underline{y}\} = A_R x_R + A_I x_I + C_y^R \nabla_R + C_y^I \nabla_I$

with $x_R\in R^{n_R}$, $x_I\in Z^{n_I}$, $\nabla_R\in R^{q_R}$ and $\nabla_I\in Z^{q_I}$.

In the case with also integer parameters, the generalized likelihood ratio can be applied as well, but the intricacy is that the distribution of the resulting teststatistic is unknown. The

hypotheses leave the probability density function $p_{\underline{y}}(y|\mu)$ partly unspecified. This function has then to be maximized in the ratio (7.3) and yields an integer estimator with a probability density function that is discrete, as discussed in the previous section. The teststatistic can be derived, but its distribution is unknown. This problem does not occur with the simple likelihood ratio, as $p_{\underline{y}}(y|\mu)$ is then completely specified by both H_o and H_a.

In appendix A we will treat several special cases of this general case. One should keep in mind, the distinction between discriminability and model validation with respect to integerness. For the first, one candidate is to be discriminated from others; a choice between (equivalent) alternatives is to be made (which gridpoint is most likely). Cases 6 and 7 in appendix A concern discriminability. For model validation concerning the integerness of parameters, the parameters being integers is opposed against being reals (is any gridpoint sufficiently likely). This is done by the hypothesis specifications of cases 3 and 4 in appendix A.

1.8 Measures for quality

In the sections 6 and 7, we have discussed the essential aspects of geodetic quality. In section 6, the stochastic characteristics of the observables were propagated, using the null hypothesis, into the stochastic characteristics of the estimators for the unknown parameters. These characteristics are described by the probability density function $p_{\underline{\hat{x}}}(\hat{x})$.

Precise observables do not necessarily yield accurate estimation results. The model used in the estimation, may not imply an adequate modelling of the measurement system and therefore the null hypothesis should be validated. Validation is carried out by hypothesis testing in linear models with the generalized likelihood ratio test.

In order to be able to describe the quality of the results obtained from the data processing, we will consider measures for precision and reliability, sections 1.8.1 and 1.8.2 respectively.

1.8.1 Precision

If we assume that $\underline{y} \sim N$, then also $\underline{\hat{x}} \sim N$, as the estimator is a linear function of the observables. In that case, the probability density function of the estimator $\underline{\hat{x}}$ is specified by the first moment, the expectation $E\{\underline{\hat{x}}\}$, and the second moment, the dispersion $D\{\underline{\hat{x}}\}$. From the population $\underline{y}$, of which in general the expectation $E\{\underline{y}\}$ is not known, one sample y is available and from this, estimate $\hat{x}$ is computed, which then serves as a best estimate for the unknown expectation. The dispersion is given by the variance covariance matrix. Matrix $Q_{\hat{x}}$ is found by propagation of Q_y, using the null-hypothesis. $D\{\underline{\hat{x}}\}$ describes the spread in the estimation results. It represents the quality of the estimator under the working mathematical model, the null hypothesis $\underline{y} \sim N(Ax, Q_y)$.

When design matrix A is rankdeficient, an S-basis has to be chosen, see section 2. It should be realized that estimate $\hat{x}$ will depend on the chosen S-basis, and also the precision description by means of variance covariance matrix $Q_{\hat{x}}$ will. In geodetic network design, the analysis of precision is based on the use of a criterion matrix. The actual variance covariance matrix $Q_{\hat{x}}$ is compared with the criterion matrix and this precision description is S-basis independent [Baarda, 1973].

1.8.2 Reliability

As one can never be 100% sure about the validity of the null hypothesis, beside precision also reliability has to be considered. If an alternative hypothesis holds true instead of the null hypothesis, we have a certain probability of detecting it with the statistical testing procedure of section 7, and decide accordingly. Usually, when the difference between null and alternative hypothesis is significant, the probability will be large as well. The nominal performance of the testing procedure is expressed by reliability.

In section 7, testing was limited to alternative hypotheses, which imply extensions to the functional model as compared with the null hypothesis. Under the assumption that a certain alternative hypothesis holds true instead of the null hypothesis, the size of the model error ∇ can be computed such that it can be found with the testing procedure with a certain probability. Once the level of significance α has been chosen, the power γ follows from the q-dimensional $\chi^2(q,\lambda)$-distribution (7.10). The other way around is that once γ has been fixed (to some γ_o, e.g. 0.80), the non-centrality parameter λ can be computed. This reference value is then denoted by λ_o [Teunissen, 1994b]. The non-centrality parameter is related to the model error by (7.10):

$$\lambda = \nabla^T C_t^T Q_t^{-1} C_t \nabla = \|C_t \nabla\|^2_{Q_t^{-1}} \tag{8.1}$$

With $\lambda=\lambda_o$, equation (8.1) represents the boundaries of a hyper-ellipsoid in the R^q [ibid]. For a one-dimensional model error, $q=1$, the ellipsoid collapses to an interval. The size (not the sign) of the parameter ∇ can be computed from

$$|\nabla| = \sqrt{\frac{\lambda_o}{c_t^T Q_t^{-1} c_t}} \tag{8.2}$$

and describes the nominal performance of the testing procedure in finding a model error of the type specified in the alternative hypothesis. Measure (8.2) is the Minimal Detectable Bias (MDB) related to teststatistic (7.12). The detectable model error in terms of the vector of observations reads

$$\nabla y = C_y \nabla \tag{8.3}$$

and allows to analyse internal reliability; it describes to what extent model validation in the observation space R^n is possible. Testing is based on the deviations of the observation vector y from the model space $R(A)$. Measures for internal reliability do not depend on the S-basis.

In the example below, we consider a two-dimensional misclosure space with two alternative hypotheses ($q=1$). For the testing we may write

$$H_o : E\{\underline{t}\} = 0 \quad \text{and} \quad H_a : E\{\underline{t}\} = c_t \nabla$$

$$H_{\bar{a}} : E\{\underline{t}\} = \bar{c}_t \bar{\nabla}$$

The metric of the R^2-space is given by Q_t^{-1}. The MDBs are denoted by ∇ and $\bar{\nabla}$ respectively.

ellipse: $t^T Q_t^{-1} t = \lambda_o \qquad T_{q=1} = \|P_{c_t} t\|^2_{Q_t^{-1}}$

$\|c_t \nabla\|_{Q_t^{-1}} = \|\bar{c}_t \bar{\nabla}\|_{Q_t^{-1}} = \sqrt{\lambda_o} \qquad \bar{T}_{q=1} = \|P_{\bar{c}_t} t\|^2_{Q_t^{-1}}$

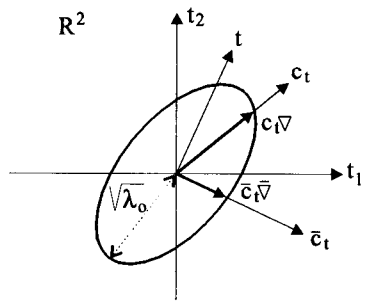


Figure 8.1: Testing in two-dimensional misclosure space

If t originates from a realization of $\underline{y}$ under H_a and the size of the model error equals the MDB ∇, the probability of rejecting H_o in favour of the H_a equals γ_o.

In order to assess the separability between various (one-dimensional) alternative hypotheses, one may compute the angle ϕ between c_t and $\bar{c}_t$.

$$\cos^2\phi = \frac{(\bar{c}_t^T Q_t^{-1} c_t)^2}{\|\bar{c}_t\|^2_{Q_t^{-1}} \|c_t\|^2_{Q_t^{-1}}} \tag{8.4}$$

For $\phi = 0$ or $\phi = \pi$, one can not distinguish between the hypotheses H_a and $H_{\bar{a}}$. For $\phi = \pi/2$ or $\phi = -\pi/2$, there is optimal separability. The correlation between the test-statistics (not misclosures) $\underline{t}$ and $\underline{t}$ (7.12) equals $\rho^2_{\underline{t}\underline{t}} = \cos^2\phi$.

One can go one step further and consider, if a certain alternative hypothesis holds true and this is not detected, the influence on the estimator; one namely has probability $1-\gamma_o$ of non-detection, if the size of the model error equals the MDB. This propagation is external reliability and can be regarded as an analysis on sensitivity of the estimator to unmodelled effects.

With (8.3), the bias in the estimate becomes

$$\nabla\hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \nabla y \tag{8.5}$$

Equation (8.5) shows how a model error ∇y propagates into the estimate $\hat{x}$. The significance of the effect can be judged upon with the bias to noise ratio $\lambda_{\hat{x}}$.

$$\lambda_{\hat{x}} = \nabla\hat{x}^T Q_{\hat{x}}^{-1} \nabla\hat{x} = \|\nabla\hat{x}\|^2_{Q_{\hat{x}}^{-1}} = \|P_A \nabla y\|^2_{Q_y^{-1}} \tag{8.6}$$

The bias in the estimate is compared with the precision of the estimator under H_o. The scalar measure (8.6) provides an upperbound for the bias in an individual element of estimate $\hat{x}$ [Teunissen, 1994b].

$$\left|\frac{\nabla\hat{x}_i}{\sigma_{\hat{x}_i}}\right| \le \sqrt{\lambda_{\hat{x}}} \tag{8.7}$$

The last expression of equation (8.6) shows that $\lambda_{\hat{x}}$ can be written in terms of quantities in the observation space R^n and does therefore not depend on the parametrization in R^n or on the S-basis. It can be shown that a formula similar to (8.7) also holds when a part of the vector x is considered.

1.9 Summary

Some essential theory on data processing - estimation and quality control - has been reviewed. Once the mathematical model, the null hypothesis, has been set up, the estimation is carried out. It is followed by the testing procedure, in which the validity of the model used, is checked. It should be recognized that this validation is of the utmost importance as estimation results, obtained with a model that is invalid, will be invalid as well. Finally, measures for precision and reliability were given. They can be used to describe and analyse the quality of the results.

Appendix 1.A Hypothesis testing: various cases

The most general hypothesis with both real and integer parameters concerning misspecifications in the functional model, is given by the first item in the list below. In this appendix we will treat several special cases of this general case; cases 2 through 7. The list below is certainly not exhaustive as many more variations are possible.

1) $H_o: E\{\underline{y}\} = A_R x_R + A_I x_I \quad H_a: E\{\underline{y}\} = A_R x_R + A_I x_I + C_y^R \nabla_R + C_y^I \nabla_I$

2) $H_o: E\{\underline{y}\} = A_R x_R \quad H_a: E\{\underline{y}\} = A_R x_R + C_y^R \nabla_R$

3) $H_o: E\{\underline{y}\} = A_R x_R + A_I x_I \quad H_a: E\{\underline{y}\} = A_R x_R + A_I x_I$ with $x_I \in R^{n_I}$

4) $H_o: E\{\underline{y}\} = A_I x_I \quad H_a: E\{\underline{y}\} = A_I x_I$ with $x_I \in R^{n_I}$

5) $H_o: E\{\underline{y}\} = A_I x_I \quad H_a: E\{\underline{y}\} = A_I x_I + C_y^I \nabla_I$

6) $H_o: E\{\underline{y}\} = A_I x_I^1 \quad H_a: E\{\underline{y}\} = A_I x_I$

7) $H_o: E\{\underline{y}\} = A_I x_I^1 \quad H_a: E\{\underline{y}\} = A_I x_I^2$

with $x_R \in R^{n_R}$, $x_I \in Z^{n_I}$, $\nabla_R \in R^{q_R}$ and $\nabla_I \in Z^{q_I}$, unless stated otherwise.

1.A.1 General hybrid case

1) $H_o: E\{\underline{y}\} = A_R x_R + A_I x_I \quad H_a: E\{\underline{y}\} = A_R x_R + A_I x_I + C_y^R \nabla_R + C_y^I \nabla_I$

The null hypothesis is the most general case of a hybrid estimation problem. In the alternative hypothesis, the functional model is extended by both real valued and integer parameters. The ultimate goal is to derive the teststatistic and its distribution for this general case. This is not yet achieved and in the sequel we will therefore consider several special cases of the general case.

1.A.2 Only real parameters

2) $H_o: \quad E\{\underline{y}\} = A_R x_R \qquad H_a: \quad E\{\underline{y}\} = A_R x_R \qquad + C_y^R \nabla_R$

This is the default case, treated in section 1.7.2, with only real valued parameters. The H_o is extended by q_R parameters, with $1 \le q_R \le m - n_R$.

For a hybrid estimation problem, case 2 can be used for testing the model used for the computation of the so-called float solution. Once the integer parameters are treated as deterministic quantities, see section 1.6.3, case 2 can also be used for testing the so-called fixed model:

$$H_o: \quad E\{\underline{y} - A_I \check{x}_I\} = A_R x_R$$

to check, within the approximation, the validity of this mathematical model. The fixing of the integer parameters is clearly expressed in the following mixed model, with non-stochastic conditions:

$$H_o: \quad E\{\underline{y}\} = A_R x_R + A_I x_I \qquad x_I \doteq \check{x}_I$$

The values for the Overall Model teststatistic for the float solution T_{float} and for the fixed solution T_{fixed} are related to the value for the teststatistic (A.4) of case 4:

$$T_{fixed} = T_{float} + T(\check{x}_I)$$

The vectors $\hat{e} = y - \hat{y}$ and $(\hat{y} - \check{y})$ are orthogonal, metric Q_y^{-1}, cf. figure 4.2, or $\underline{\hat{e}} = \underline{y} - \underline{\hat{y}}$ and $(\underline{\hat{y}} - \underline{\check{y}})$ do not correlate.

1.A.3 H_o: integer and real parameters vs H_a: only real parameters

In the cases 3 and 4, we have a hybrid and an integer null hypothesis respectively, whereas the alternative hypothesis allows the integer parameters to be real-valued. It will be tested whether the model assumption of the parameters being integers is acceptable, i.e. whether integerness is sufficiently likely as compared to the parameters being reals.

Although the dimension of the parameter space in the alternative hypothesis is not increased as compared with the H_o, H_a implies an extension in the functional model. According to the H_o of case 4, the vector y is allowed to lie only on discrete gridpoints in the $R(A_I)$, whereas under the H_a vector y is allowed to lie everywhere else in the $R(A_I)$. This can be realized by an integer vector x_I together with a real valued increment ∇_R. Thus, an alternative specification of the H_a reads:

$$H_a: \quad E\{\underline{y}\} = A_I x_I + A_I \nabla_R$$

with $\nabla_R \in R^{n_I}$, at least one $(\nabla_R)_i \notin Z$. The alternative hypothesis equals the null hypothesis apart from an extension with n_I real valued parameters. These ∇_R parameters actually contain the fractional parts of the values for the unknowns x_I.

3) $H_o: \quad E\{\underline{y}\} = A_R x_R + A_I x_I \qquad H_a: \quad E\{\underline{y}\} = A_R x_R + A_I x_I \qquad$ with $x_I \in R^{n_I}$

The probability density function $p_{\underline{y}}(y|\mu)$ has to be maximized under both H_o and H_a for the generalized likelihood ratio (7.3). Under H_a the ordinary real valued least-squares estimator results, and under H_o the integer least-squares estimator. The generalized likelihood ratio test can be shown to be

(A.1) reject H_o if $\|\hat{x}_I - \check{x}_I\|^2_{Q_{\hat{x}_I}^{-1}} > \ln a^{-2}$

In the next section on case 4, we will elaborate on this expression and neglect the stochastics of $\underline{\check{x}}_I$ to come to a workable teststatistic. Case 3 is the general case of testing a model with (some) integer parameters against a model with exclusively real valued parameters.

1.A.4 H_o: only integer parameters vs H_a: only real parameters

4) H_o : $E\{\underline{y}\} = A_I x_I$ H_a : $E\{\underline{y}\} = A_I x_I$ with $x_I \in R^{n_I}$

The generalized likelihood ratio test (7.3) becomes

(A.2) reject H_o if $e^{-\frac{1}{2}(\hat{e}^T Q_y^{-1} \hat{e} - \check{e}^T Q_y^{-1} \check{e})} < a$

The exponent is the difference of the integer least-squares norm and ordinary least-squares norm. The test can be rewritten into

(A.3) reject H_o if $\|\hat{x}_I - \check{x}_I\|^2_{Q_{\hat{x}_I}^{-1}} > \ln a^{-2}$

With neglecting the stochastics of the integer estimator $\underline{\check{x}}_I$, the teststatistic becomes

(A.4) $$\underline{T}(\check{x}_I) = (\underline{\hat{x}}_I - \check{x}_I)^T Q_{\hat{x}_I}^{-1} (\underline{\hat{x}}_I - \check{x}_I)$$

which has a $\chi^2(n_p 0)$-distribution under H_o with the assumption $E\{\underline{\hat{x}}_I\} = \check{x}_I$. $\underline{T}(\check{x}_I) \sim \chi^2(n_I, \lambda)$ under H_a with $\lambda = \nabla\check{x}^T Q_{\hat{x}_I}^{-1} \nabla\check{x}$ and $\nabla\check{x} = x^a - \check{x}_I$. If the stochastics of $\underline{\check{x}}_I$ can not be neglected, the distribution of the teststatistic is unknown.

Note that with (A.4), actually H_o with x_I as a deterministic quantity is tested against H_a with $x_I \in R^{n_I}$, as the stochastics of integer estimator $\underline{\check{x}}_I$ are neglected. H_o is a simple hypothesis, H_o : $E\{\underline{y}\} = A_I \check{x}_I$.

The, $n_I = 1$, one-dimensional version of (A.4) reads

(A.5) $$\underline{t}(\check{x}_I) = \frac{\underline{\hat{x}}_I - \check{x}_I}{\sigma_{\hat{x}_I}}$$

which has a $N(0,1)$-distribution under H_o. The critical region for teststatistic (A.5) is two-sided.

In practice the test with statistic (A.4) is carried out as an acceptance test for the parameters in x_I being integers instead of reals. Under the neglect, based on the observati-

ons, it is to be found out whether $x_I = \check{x}_I$ as compared to $x_I \in R^{n_I}$ is sufficiently likely. The geometric interpretation of teststatistic (A.4) is the distance from the real valued estimate $\hat{x}_I$ to the integer estimate $\check{x}_I$, with the distance measured in the metric $Q_{\hat{x}_I}^{-1}$. The distance being less or equal to some constant χ^2 defines the search ellipsoid in the integer minimization, section 4.2.

By the discrete search in the n_I-dimensional ellipsoid, a list of gridpoints ordered after distance to $\hat{x}_I$ is returned. For good discriminination, see section 1.6.3, the integer estimate $\check{x}_I$ should represent clearly more probability density mass than any other gridpoint. In [Teunissen, 1995a] it is proposed to infer this by applying test (A.4) also to the so-called second best candidate $\check{x}_I'$.

$$T(\check{x}_I') = (\hat{x}_I - \check{x}_I')^T Q_{\hat{x}_I}^{-1} (\hat{x}_I - \check{x}_I')$$

The values $T(\check{x}_I')$ and $T(\check{x}_I)$ together bear information on the distance $\|\check{x}_I' - \check{x}_I\|_{Q_{\hat{x}_I}^{-1}}$. It holds

$$\|\hat{x}_I - \check{x}_I'\|_{Q_{\hat{x}_I}^{-1}} - \|\hat{x}_I - \check{x}_I\|_{Q_{\hat{x}_I}^{-1}} \leq \|\check{x}_I' - \check{x}_I\|_{Q_{\hat{x}_I}^{-1}} \leq \|\hat{x}_I - \check{x}_I'\|_{Q_{\hat{x}_I}^{-1}} + \|\hat{x}_I - \check{x}_I\|_{Q_{\hat{x}_I}^{-1}}$$

For good discrimination the difference $T(\check{x}_I') - T(\check{x}_I)$ should be large. The distance $\|\check{x}_I' - \check{x}_I\|_{Q_{\hat{x}_I}^{-1}}$ will play a role in the discrimination, see section A.7.

1.A.5 Only integer parameters

5) H_o : $E\{\underline{y}\} = A_I x_I$ H_a : $E\{\underline{y}\} = A_I x_I + C_y^I \nabla_I$

Case 5 is the equivalent with only integer parameters of case 2 with only real valued parameters. As compared with H_o, H_a is extended with integer parameters.

The estimators both under H_o and H_a have discrete probability density functions and therefore we are not able to derive the distribution of the generalized likelihood ratio teststatistic. The distribution of the least-squares residuals $\underline{\check{e}}$ is unknown. The test reads

(A.6) reject H_o if $\check{e}^{oT} Q_y^{-1} \check{e}^o - \check{e}^{aT} Q_y^{-1} \check{e}^a > \ln a^{-2}$

1.A.6 H_o: simple hypothesis vs H_a: only integer parameters

6) H_o : $E\{\underline{y}\} = A_I x_I^1$ H_a : $E\{\underline{y}\} = A_I x_I$

The H_o, a simple hypothesis, with some deterministic integer vector x_I^1 is tested against a composite hypothesis with only integer parameters (although H_a can be looked upon in the R^m as a set of an infinite number of simple hypotheses). The H_a as compared with H_o is extended in the functional model by integer parameters: H_a : $E\{\underline{y}\} = A_I x_I^1 + A_I \nabla_I$ with $\nabla_I \in Z^{n_I}$ and $\nabla_I \neq 0$. We would like to test the discriminability of one specific gridpoint in the $R(A_I)$, $A_I x_I^1$, with respect to all others.

In practice we would like to test the integer least-squares estimate: $x_I^1 = \check{x}_I$. By definition of the integer least-squares estimate however, both numerator and denominator of the generalized likelihood ratio (7.3) are maximized by $x_I = \check{x}_I$. The ratio equals one by

definition and the test always leads to acceptance of H_o. The expectation of $\underline{y}$ is governed by the integer vector x_I. As $A_I\check{x}_I$ is closest to y, it never makes sense to reject $\check{x}_I$, the H_o, in favour of any other integer vector, the H_a.

Testing the simple hypothesis H_o against H_a with $x_I \in R^{n_I}$ instead of $x_I \in Z^{n_I}$, yields the teststatistic (A.4) discussed under case 4.

1.A.7 H_o and H_a both simple hypotheses

7) $H_o: \quad E\{\underline{y}\} = A_I x_I^1 \qquad H_a: \quad E\{\underline{y}\} = A_I x_I^2$

Case 7 concerns the testing of two simple hypotheses and yields the simple likelihood ratio test. This test is most powerful, see [Teunissen, 1994b]. The alternative hypothesis H_a is not an extension of, but really an alternative to H_o. The values x_I^1 and x_I^2 are constants, deterministic. We will test two integer vectors (gridpoints) against each other. The derivation below of the simple likelihood ratio test parallels [ibid].

The simple likelihood ratio is defined as

(A.7)
reject H_o if $$\frac{p_{\underline{y}}(y|x_I^1)}{p_{\underline{y}}(y|x_I^2)} < a$$

with a some positive constant. The likelihood of sample y being produced by model H_o is directly compared with the likelihood of model H_a. Test (A.7) can be rewritten into

(A.8)
reject H_o if $$2(x_I^2-x_I^1)^T A_I^T Q_y^{-1}(y-A_I x_I^1) > \ln a^{-2} + (x_I^2-x_I^1)^T A_I^T Q_y^{-1} A_I (x_I^2-x_I^1)$$

Vector $\hat{x}_I$ is the least-squares solution to the model of observation equations $E\{\underline{y}\} = A_I x_I$ with $x_I \in R^{n_I}$ and the unit direction vector c is defined as

$$c = \frac{x_I^2-x_I^1}{\|x_I^2-x_I^1\|_{Q_{\hat{x}_I}^{-1}}}$$

Vector c is directed from x_I^1 to x_I^2. The test becomes

reject H_o if $$c^T Q_{\hat{x}_I}^{-1}(\hat{x}_I-x_I^1) > \frac{\ln a^{-1}}{\|x_I^2-x_I^1\|_{Q_{\hat{x}_I}^{-1}}} + \frac{1}{2}\|x_I^2-x_I^1\|_{Q_{\hat{x}_I}^{-1}}$$

or

(A.9)
reject H_o if $$c^T Q_{\hat{x}_I}^{-1}(\hat{x}_I-x_I^1) > K_\alpha$$

Testing with (A.9) takes place in the parameter space R^{n_I}. With H_o: $\underline{\hat{x}}_I \sim N(x_I^1, Q_{\hat{x}_I})$ the teststatistic

(A.10) $$p = c^T Q_{\hat{x}_I}^{-1}(\underline{\hat{x}}_I-x_I^1)$$

is distributed as $N(0,1)$ under H_o and as $N(\|x_I^2 - x_I^1\|_{Q_{\hat{x}_I}^{-1}},1)$ under H_a. The critical region for the teststatistic is single sided. The value of the teststatistic actually is the inner product of the vectors c and $(\hat{x}_I - x_I^1)$. The geometric interpretation is given in figure A.1. The vector $(\hat{x}_I - x_I^1)$ is projected onto the vector c.

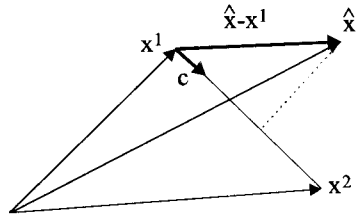


Figure A.1: Simple likelihood ratio test

The teststatistic $\underline{p}$ (A.10) can be related to the difference in distances of case 4:

$$-2\underline{p}\,\|x_I^2 - x_I^1\|_{Q_{\hat{x}_I}^{-1}} + \|x_I^2 - x_I^1\|^2_{Q_{\hat{x}_I}^{-1}} = \underline{T}(x_I^2) - \underline{T}(x_I^1)$$

The test with the expression above reads

reject H_o if $$T(x_I^2) - T(x_I^1) < -2K_\alpha\,\|x_I^2 - x_I^1\|_{Q_{\hat{x}_I}^{-1}} + \|x_I^2 - x_I^1\|^2_{Q_{\hat{x}_I}^{-1}} = \tilde{K}_\alpha$$

In order to not reject the null hypothesis, the difference in distance should be larger than some critical value $\tilde{K}_\alpha$. The real valued estimate, computed from the observations y, should be closer to x_I^1 than to x_I^2. Note that under H_o this teststatistic has a normal distribution (only linear operations were applied to $\underline{p}$):

(A.11) $$\underline{T}(x_I^2) - \underline{T}(x_I^1) \sim N(\|x_I^2 - x_I^1\|^2_{Q_{\hat{x}_I}^{-1}},\ 4\|x_I^2 - x_I^1\|^2_{Q_{\hat{x}_I}^{-1}})$$

The test (A.9) with $\underline{p}$ for the one-dimensional case becomes

(A.12) $$\frac{\hat{x}_I - x_I^1}{\sigma_{\hat{x}_I}} > K_\alpha$$

with $c = \sigma_{\hat{x}_I}$ and where we assumed that $x_I^2 > x_I^1$. We see that for the one-dimensional case, the expression (A.12) of the teststatistic $\underline{p}$ equals (A.5) of case 4. Both are standard normally distributed under H_o. The tests are not identical, as they concern different alternative hypotheses. Moreover the above test has a single sided critical region, whereas (A.5) has a two-sided critical region.

Note that the distance between x_I^2 and x_I^1 can be computed, once the variance covariance matrix $Q_{\hat{x}_I}$ is available; the real valued estimate itself is not needed. In the one-dimensional case, with x_I^2 the so-called second best candidate, we have $\|x_I^2 - x_I^1\|_{Q_{\hat{x}_I}^{-1}} = \sigma_{\hat{x}_I}^{-1}$.

In practice, for judging the discriminability between the integer least squares estimate and the second best integer vector, one constructs the test (A.9) with $x_I^1 = \check{x}_I$ and $x_I^2 = \check{x}_I'$. Both $\check{x}_I$ and $\check{x}_I'$ are not non-stochastic however. A consequence is that, as the integer estimate is the gridpoint that has least distance to the real valued estimate, the teststatistic $\underline{p}$ will *always*

$$p \leq \frac{1}{2}\|\check{x}_I' - \check{x}_I\|_{Q_{\hat{x}_I}^{-1}}$$

and the standard normal distribution of p is violated.

One would like to use test (A.9) to assure that $\check{x}_I$ represents sufficiently more likelihood than any other gridpoint. The discrimination test is needed for fixing the integer parameters, see section 6.3. When the stochastics of the integer estimator are neglected, the integer estimate has to be validated. In theory all (=an infinite number of) gridpoints should be considered. In practice this can be done one by one with the projection teststatistic p for only a limited number of them (the gridpoints in the neighbourhood, or even only the second best one). We should oppose H_o $E\{y\} = A_I\check{x}_I$ against all possible H_a's with $x_I^2 \in Z^{n_I}$ in one time. This was done in case 6.

1.A.8 Concluding remark

Validation of models with integer parameters requires special attention. The major problem is caused by the discrete probability density function of the integer estimator. Various special cases of the general case of testing in hybrid estimation have been considered.

References

[Baarda, 1973] Baarda, W. (1973). S-transformations and criterion matrices. Netherlands Geodetic Commission. Publications on Geodesy. New Series Vol. 5, No. 1. Delft.

[Björck, 1996] Björck, Å. (1996). Numerical methods for least squares problems. SIAM. Philadelphia.

[Blewitt, 1989] Blewitt, G. (1989). Carrier phase ambiguity resolution for the Global Positioning System applied to geodetic baselines up to 2000 km. Journal of Geophysical Research. Vol. 94, No. B8. pp. 10.187-10.203.

[Jonge et al, 1996] Jonge, P.J. de and C.C.J.M. Tiberius (1996). Integer ambiguity estimation with the LAMBDA method. Proceedings IAG Symposium No. 115 'GPS trends in precise terrestrial, airborne and spaceborne applications'. G. Beutler et al. (Eds). Springer Verlag. pp. 280-284.

[Koch, 1988] Koch, K-R (1988). Parameter estimation and hypothesis testing in linear models. Springer-Verlag. New York.

[Maybeck, 1979] Maybeck, P.S. (1979). Stochastic models, estimation and control. Volume 141-1 in Mathematics in science and engineering. Academic Press, Inc. New York.

[Schwarz, 1994] Schwarz, K.P. (1994). The art of modelling; from determinism to chaos. Proceedings of KIS94. Banff, Canada. August 30 - September 2. pp. 3-12.

[Teunissen, 1984] Teunissen, P.J.G. (1984). Generalized inverses, adjustment, the datum problem and S-transformations. Report 84.1. Section Mathematical and Physical Geodesy, Department of Geodetic Engineering, Delft University of Technology.

[Teunissen, 1989] Teunissen, P.J.G. (1989). Nonlinear inversion of geodetic and geophysical data: diagnosing nonlinearity. Developments in four-dimensional geodesy. F.K. Brunner & C. Rizos (Eds). Lecture notes in Earth Sciences, No. 29. Springer-Verlag. pp. 241-264.

[Teunissen, 1990a] Teunissen, P.J.G. (1990). Nonlinear least squares. Manuscripta Geodaetica. Vol. 15. pp. 137-150.

[Teunissen, 1990b] Teunissen, P.J.G. (1990). GPS op afstand bekeken. Een halve eeuw in de goede richting. Lustrumboek Snellius 1985-1990. pp. 215-233. Technische Universiteit Delft.

[Teunissen, 1991] Teunissen, P.J.G. (1991). On the minimal detectable biases of GPS phase ambiguity slips. Proceedings of the First International Symposium 'Real time differential applications of the GPS, DGPS'91'. Stuttgart. Vol. 2. pp. 679-686.

[Teunissen, 1993] Teunissen, P.J.G. (1993). Least-squares estimation of the integer GPS ambiguities. Invited lecture. Section IV Theory and Methodology. IAG General Meeting. Beijing, China. August 1993. 16 p.

[Teunissen, 1994a] Teunissen, P.J.G. (1994). Mathematische geodesie 1; inleiding vereffeningstheorie. Lecture notes Mathematical Geodesy I, course Ge30. Department of Geodetic Engineering, Delft University of Technology.

[Teunissen, 1994b] Teunissen, P.J.G. (1994). Mathematische geodesie 2; inleiding toetsingstheorie. Lecture notes Mathematical Geodesy II, course Ge36. Department of Geodetic Engineering, Delft University of Technology.

[Teunissen, 1994c] Teunissen, P.J.G. (1994). On the GPS double difference ambiguities and their partial search spaces. Proceedings IAG Symposium No. 114 'Geodetic Theory Today'. F. Sansò (Ed). Springer Verlag. pp. 39-48.

[Teunissen et al, 1994a] Teunissen, P.J.G. and C.C.J.M. Tiberius (1994). Integer least-squares estimation of the GPS phase ambiguities. Proceedings of KIS94. Banff, Canada. August 30 - September 2. pp. 221-231.

[Teunissen et al, 1994b] Teunissen, P.J.G., P.J. de Jonge and C.C.J.M. Tiberius (1994). On the spectrum of the GPS DD ambiguities. Proceedings of ION GPS-94. Salt Lake City, USA. September 20-23. pp. 115-124.

[Teunissen, 1995a] Teunissen, P.J.G. (1995). GPS carrier phase ambiguity fixing concepts. Chapter 8 in GPS for Geodesy. A. Kleusberg and P.J.G. Teunissen (Eds). Lecture notes in Earth Sciences, No. 60. Springer Verlag, 1996 (2nd edition, 1998).

[Teunissen, 1995b] Teunissen, P.J.G. (1995). The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation. Journal of Geodesy. Vol. 70. No. 1-2. pp. 65-82.

[Teunissen, 1995c] Teunissen, P.J.G. (1995). The invertible GPS ambiguity transformations. Manuscripta Geodaetica. Vol. 20. pp. 489-497.

[Teunissen, 1996a] Teunissen, P.J.G. (1996). An analytical study of ambiguity decorrelation using dual frequency code and carrier phase. Journal of Geodesy. Vol. 70. No. 8. pp. 515-528.

[Teunissen, 1996b] Teunissen, P.J.G. (1996). Rank defect integer least squares with applications to GPS. Bollettino di Geodesia e Scienze Affini. Vol. 55. No. 3. pp. 225-238.

[Teunissen et al, 1996] Teunissen, P.J.G., P.J. de Jonge and C.C.J.M. Tiberius (1996). The volume of the GPS ambiguity search space and its relevance for integer ambiguity resolution. Proceedings of ION GPS-96. Kansas City, USA. September 17-20. pp. 889-898.

[Teunissen, 1997] Teunissen, P.J.G. (1997). Quality control and GPS. Chapter 7 in GPS for Geodesy. P.J.G. Teunissen and A. Kleusberg (Eds). Lecture notes in Earth Sciences. Springer Verlag, 1998, 2nd edition.

2. Recursive data processing

batch or recursion
Basically, data processing for a model of observation equations $E\{\underline{y}\}=Ax$ can be carried out in batch or in recursion. The recursion stems from a row-wise partitioning of the model. In the first section we will discuss the concepts of both strategies, together with their advantages and disadvantages. The conditions imposed by the kinematic GPS surveying application drive the decision towards the recursive strategy.

implementation
Various implementations of the data processing are possible for the model of observation equations. We can distinguish between two categories of (direct) methods: those that are based on normal equations and those that involve reduction of the designmatrix to a canonical form. The methods with normal equations have already been in use for a long time. In particular the method of normal equations with Cholesky factorization of the normal matrix, is very popular in geodetic data processing. The techniques of the second category have been evolving since the sixties of this century. The reason primarily is numerical accuracy. In appendix A we will treat an alternative from this second category. A QR-factorization of the designmatrix is made and the reduction involves orthogonal transformations. Beside the method of normal equations, also the method of QR-factorization turns out to be very well suited for the Delft approach to data processing with estimation and quality control.

SRIF
Section 2 and 3 deal with the recursive data processing according to the QR-factorization implementation. This implementation is called the Square Root Information Filter and the data processing in the kinematic GPS surveying application will be carried out with this filter. In the section on quality control, the implementation of the DIA-procedure in the SRIF is treated. A review of recursive data processing and the DIA-procedure is given in appendix B.

mechanization aspects
The last section contains a brief review of implementations of recursive data processing and the argumentation for the one chosen, the Square Root Information Filter.

2.1 Recursive estimation

In this section the concepts of estimation both in batch and in recursion are discussed. We elaborate on filtering and introduce the time and measurement update. We conclude with the argumentation for the choice for data processing in recursion.

2.1.1 Concept of filtering

We are concerned with a dynamic system. The state of the system is mathematically described by the vector x. The parameters in x depend on time. Knowledge about the state of the system is gained by making measurements. As a result we have a measurement system with observations on one side, and unknown parameters on the other.

We make a division in time and use an index to denote the epoch a quantity is related to. The system is considered at discrete instants in time from epoch 1 to epoch k. As such we have k vectors, not necessarily of equal dimension, for the state of the system: $x_1, x_2,..,x_k$ and k vectors with observations: $y_1, y_2,..,y_k$.

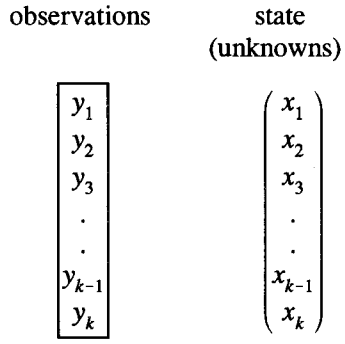


Figure 1.1: Measurement system

The data processing is carried out according to the principles discussed in chapter 1. We distinguish between two strategies: batch and recursive.

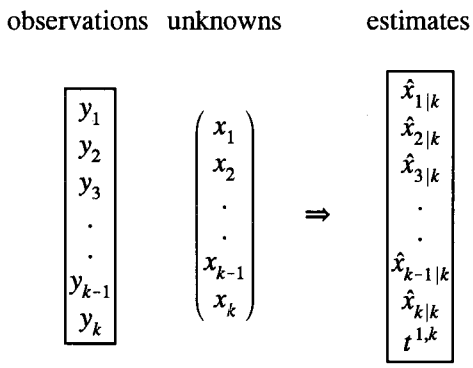


Figure 1.2: Estimation in batch

In batch estimation, all observations are taken together. The full model of observation equations is treated in one go. Therefore estimation can take place only when all observations have been collected and de facto off-line or in post mission.

In the above figure $\hat{x}_{i|k}$ is the estimate for the state at epoch i, with $1 \leq i \leq k$. The index k denotes that all observations from epoch 1 to k are used. $t^{1,k}$ is the vector of misclosures, it stems from the redundancy in the determination of the estimates. The double lined

arrow represents a linear transformation that preserves all information contained in the observations.

Now we restrict ourselves to a special structure of the model of observation equations and discuss the recursive estimation. The recursion stems from a (blocked) row-wise partitioning of the model. We assume that observation y_i is related to x_i and possibly to x_{i-1} (for $i=2,..,k$). The observables $\underline{y}_1,..,\underline{y}_k$ are not time correlated (white noise). Matrix Q_y is a block diagonal matrix. The full model of observation equations reads:

$$(1.1)\quad E\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \cdot \\ \cdot \\ \underline{y}_{k-1} \\ \underline{y}_k \end{pmatrix}\right\} = \begin{pmatrix} A_1 & & & & & \\ A_{21} & A_{22} & & & & \\ & \cdot & \cdot & & & \\ & & \cdot & \cdot & & \\ & & & A_{k-11} & A_{k-12} & \\ & & & & A_{k1} & A_{k2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_k \end{pmatrix} \quad ; \quad D\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \cdot \\ \cdot \\ \underline{y}_{k-1} \\ \underline{y}_k \end{pmatrix}\right\} = \begin{pmatrix} Q_{y_1} & & & & \\ & Q_{y_2} & & & \\ & & \cdot & & \\ & & & Q_{y_{k-1}} & \\ & & & & Q_{y_k} \end{pmatrix}$$

where y_i is a m_i-vector and x_i is a n_i-vector. The matrices A_1 and A_{i2} with $i=2,..,k$ are all assumed to be of full rank. For all epochs i it holds that $m_i \geq n_i$.

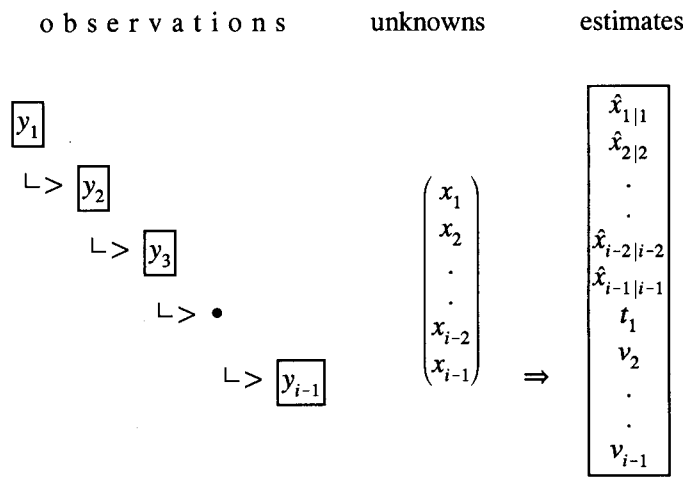


Figure 1.3: Estimation in recursion

The adjective 'recursive' means that the observations y_1 to y_k are treated in a sequential manner, thereby using the same procedure in each step.

In figure 1.3 $\hat{x}_{j|j}$ is the estimate for the state at epoch j, with $1 \leq j \leq i\text{-}1$. v_j is the vector of predicted residuals, it stems from the redundancy in the determination of the estimates at epoch j ($2 \leq j \leq i\text{-}1$). The recursion starts with the computation of estimate $\hat{x}_{1|1}$ from y_1. From this so-called initialization, the misclosures t_1 follow.

At epoch i-1 in the recursive estimation, observations y_1 through y_{i-1} have been processed. The observations are treated sequentially according to some fixed procedure. The so-

called predicted estimates are not considered. Here it is physically impossible to compute estimates $\hat{x}_{i|i-1},..,\hat{x}_{k|i-1}$ without any information about these future states.

In the sequel we will concentrate on filtering: estimates for the state are computed parallel to the data gathering (with possibly a delay). It means going through the data in forward direction, indicated by the arrows in figure 1.3. At epoch i-1, estimate $\hat{x}_{i-1|i-1}$ is of primary interest. The observations of epoch i-1 have been processed and yield the estimate for the state at the current epoch. The double lined arrow again represents a linear transformation that preserves all information contained in the observations (one-to-one mapping).

At epoch i, observation y_i is added.

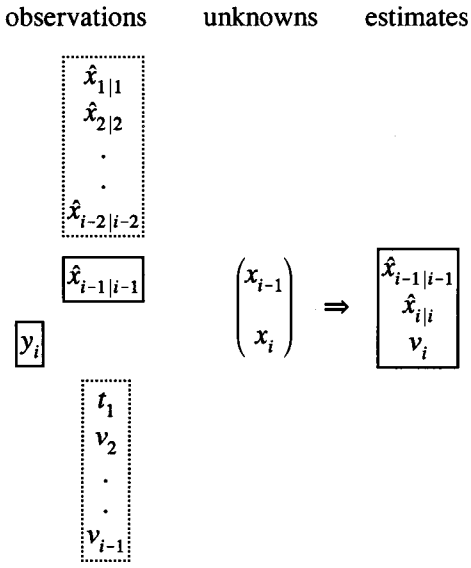


Figure 1.4: Filtering at epoch i

The estimates for the states in the past $\hat{x}_{1|1}$ through $\hat{x}_{i-2|i-2}$, are functionally related to x_1 through x_{i-2} by unit matrices and thus without redundancy. The predicted residuals v_2 through v_{i-1} (and misclosures t_1) in figure 1.4 are not functionally related to any of the unknowns, nor stochastically related to estimate $\hat{x}_{i-1|i-1}$ or the remaining observations. The observations to be treated, y_i through y_k, are related only to the states x_{i-1} through x_k.

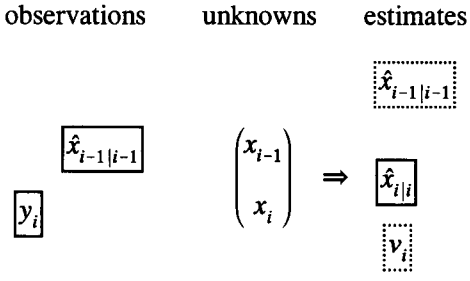


Figure 1.5: Filtering at epoch i

The quantities in the dotted boxes, which are in fact transformed observations, do not contribute to the estimates of these states and are therefore left out of further consideration in filtering, see figure 1.5.

The predicted residuals v_i are the misclosures of the 'local' model in figure 1.5. After filtering at epoch i, estimate $\hat{x}_{i-1|i-1}$ and residual v_i can be left out and observation y_{i+1} can be treated similar to observation y_i, see figure 1.6. The process continues until observation y_k has been treated and estimate $\hat{x}_{k|k}$ has been obtained.

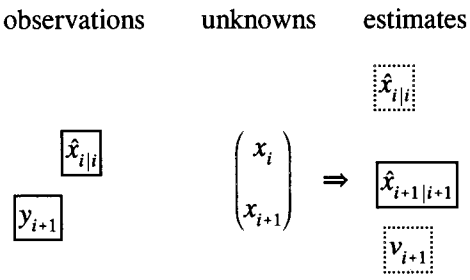


Figure 1.6: Filtering at epoch $i+1$

Filtering starts at epoch 1 and ends at epoch k. We have obtained the estimates $\hat{x}_{1|1},..,$ $\hat{x}_{k|k}$ and predicted residuals $v_2,..,v_k$ (and t_1). They all are (linearly) transformed observations. Smoothing means going through the data in backward direction. Smoothed estimates, are estimates $\hat{x}_{i|k}$ for state x_i with $i=1,..,k$, computed using (indirectly) all observations from y_1 to y_k. Although there is no redundancy in $\hat{x}_{1|1},..,\hat{x}_{k|k}$ determining $x_1,..,x_k$ (actually a one-to-one relation), improved or smoothed estimates $\hat{x}_{1|k},..,\hat{x}_{k|k}$ can be computed, due to the correlation between the filtered state estimates and the predicted residuals. Estimation in batch directly yields these smoothed estimates, figure 1.2.

The predicted residuals become available epochwise in the recursion and represent the redundant data. They are further used in the model validation. It is namely the surplus of information which enables one to test whether the data can be considered to be statistically consistent with the assumed mathematical model, given by equation (1.1). A one-to-one linear mapping exists from the full set of predicted residuals $\{v_2,..,v_k\}$ (with t_1) to the vector of misclosures $t^{1,k}$ obtained with estimation in batch.

2.1.2 Time and measurement update

We will now further detail the model of observation equations (1.1). Observations y_i (with $i \geq 2$) may originate from 'real' measurements, and from knowledge about the (dynamics of the) system. The latter will be referred to as pseudo-observations and denoted by d_i. The dimension of x_i is n_i. In the sequel we assume for the ease of the discussion, that $n_i=n$ for all $i=1,..,k$.

Here, epoch k is the last epoch (with $k \geq 2$). With new observations becoming available (in future), the recursion may of course be continued. In the sequel, the recursive data processing will usually be discussed for this epoch k.

The observation equation for the pseudo-observation at epoch k reads

$$(1.2) \qquad E\{\underline{d}_k\} = x_k - \Phi_{k,k-1}x_{k-1} \;\; ; \;\; D\{\underline{d}_k\} = Q_{d_k}$$

with $\Phi_{k,k-1}$ the nxn transition matrix from epoch k-1 to epoch k. The underlying dynamic models are continuous in time. For a derivation, the reader is referred to [Maybeck, 1979] and [Teunissen, 1995]. Both vector d_k and x_k have dimension n.

The model for the observation at epoch k is

$$(1.3) \qquad E\{\underline{y}_k\} = A_k x_k \;\; ; \;\; D\{\underline{y}_k\} = Q_{y_k}$$

With digital signal processing techniques, sensors provide measurements (samples) at discrete times. Vector y_k has dimension m_k.

With these models (1.2) and (1.3), the observation vector y_k of model (1.1) is split into d_k and y_k. The elements of the designmatrix are detailed according to

$$y_k = \begin{pmatrix} d_k \\ y_k \end{pmatrix} \qquad A_{k1} = \begin{pmatrix} -\Phi_{k,k-1} \\ 0 \end{pmatrix} \qquad A_{k2} = \begin{pmatrix} I_n \\ A_k \end{pmatrix} \qquad Q_{y_k} = \begin{pmatrix} Q_{d_k} & \\ & Q_{y_k} \end{pmatrix}$$

and matrix Q_{y_k} is block diagonal as $\underline{d}_k$ and $\underline{y}_k$ are assumed to be not correlated.

In appendix B it is reviewed how with (1.2) and (1.3), the recursion of figure 1.3 leads to the well known Kalman filter. The recursion of figure 1.3 is then made in two steps. First the pseudo-observation d_k is incorporated. This yields the Time Update, see figure 1.7. In the time update, there are $2n$ unknowns to be determined from $2n$ observations. The redundancy in this step equals zero. $\hat{x}_{k|k-1}$ is called the predicted state estimate.

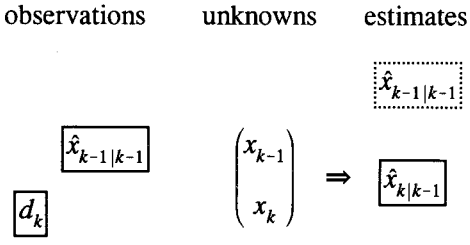


Figure 1.7: Time Update from epoch k-1 to epoch k

In the second step, observation y_k is introduced. This yields the Measurement Update, see figure 1.8. There are n unknowns to be determined from $n+m_k$ observations. The redundancy in the measurement update equals m_k. The vector of predicted residuals v_k is m_k-dimensional.

Parallel to the recursive estimation, model validation can be carried out with the DIA-procedure. This procedure, based on the predicted residuals, allows the detection and identification of, and the subsequent recovery from model errors. The DIA testing procedure is reviewed in appendix B. In the kinematic GPS surveying application, both

estimation and quality control will be carried out in a recursive manner. As a result, the estimators with accompanying quality measures can be available in (near) real time. Estimation and quality control together, constitute the data processing.

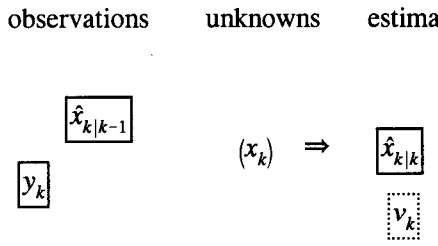


Figure 1.8: Measurement Update at epoch k

2.1.3 Argumentation for recursive data processing

In this section we will consider the arguments for the choice between the recursive and batch data processing strategy. The model of observation equations is assumed to have the particular structure of (1.1).

Both strategies are capable of incorporating observations stemming from measurements as well as those from system knowledge. Although the processing strategy differs, the estimation principle does not. Therefore identical estimators $\hat{\underline{x}}_{k|k}$ for the state at epoch k are obtained. Batch estimation directly provides smoothed estimators $\hat{\underline{x}}_{i|k}$ with $i=1,..,k$ for all states. These estimators are more or at least equally precise as the filtered estimators $\hat{\underline{x}}_{i|i}$, symbolically $Q_{\hat{x}_{i|k}} \leq Q_{\hat{x}_{i|i}}$.

In batch estimation, the complete vector of observations is processed in one go and all unknowns are estimated together. The realization of the data processing may be hindered by practical limitations (computer memory). A partitioning of the problem may be the cure. In a recursion the system per epoch will have relatively small size. Filtered estimates are based on the observations already available at the particular epoch. Collected data can be removed once they are processed (and thus need not be stored). The recursive approach also tends to handle time varying parameters more easily. New parameters can be introduced when needed and estimators for previous states can be removed. In order to start the recursion, the state estimate $\hat{x}_{1|1}$ need to be available. A batch type of solution for the first (few) epoch(s) can be used.

Estimation in batch can provide results only with some delay (post processing). The recursion can run parallel to the data gathering, although also in post-mission the estimation can be carried out in recursion. With filtering, estimates for the state can be available in (near) real time. A surveyor usually does not need to have available the estimation results in real time, but for a navigator this is indispensable. Real time quality control, however, is beneficial to both. Quality can be monitored in the field and if quality requirements are not met, corrective actions can be taken immediately (during the data gathering).

For the kinematic GPS surveying application, the recursive data processing strategy is chosen. Real time quality control thereby constitutes the primary argument.

2.2 Square Root Information Filter: estimation

The general form of the full model of observation equations was given by (1.1). Using the observation equations (1.2) and (1.3), this model was detailed in appendix B (B.1). The variance covariance matrix was assumed to be symmetric and positive definite. Normalization of the observables yields

$$(2.1)\qquad E\left\{\begin{pmatrix} Q_{y_1}^{-1/2}\underline{y}_1 \\ Q_{d_2}^{-1/2}\underline{d}_2 \\ Q_{y_2}^{-1/2}\underline{y}_2 \\ \cdot \\ \cdot \\ Q_{d_k}^{-1/2}\underline{d}_k \\ Q_{y_k}^{-1/2}\underline{y}_k \end{pmatrix}\right\} = \begin{pmatrix} Q_{y_1}^{-1/2}A_1 & & & & \\ -Q_{d_2}^{-1/2}\Phi_{2,1} & Q_{d_2}^{-1/2} & & & \\ & Q_{y_2}^{-1/2}A_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & -Q_{d_k}^{-1/2}\Phi_{k,k-1} & Q_{d_k}^{-1/2} \\ & & & & Q_{y_k}^{-1/2}A_k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix}$$

The designmatrix is of full rank. The number of observations equals $\sum_{i=1}^{k}(m_i)+(k-1)n$, the number of unknowns kn. The observables have unit variance covariance matrix.

Various implementations of the data processing are possible for the model of observation equations. We can distinguish between two categories of direct methods: those that are based on normal equations and those that involve a reduction of the designmatrix to a canonical form [Golub et al, 1989]. In appendix A, estimation via QR-factorization is treated. This method belongs to the second category. The reduction of the designmatrix is realized by orthogonal transformations and the resulting designmatrix is upper-triangular. Employing this method for recursive estimation leads to the SRIF-implementation, or Square Root Information Filter [Bierman, 1977]. The argumentation for this choice is given in section 4.

$$(2.2)\qquad E\left\{\begin{pmatrix} \hat{z}_{1|1} \\ \underline{u}_1 \\ Q_{d_2}^{-1/2}\underline{d}_2 \\ Q_{y_2}^{-1/2}\underline{y}_2 \\ \cdot \\ \cdot \\ Q_{d_k}^{-1/2}\underline{d}_k \\ Q_{y_k}^{-1/2}\underline{y}_k \end{pmatrix}\right\} = \begin{pmatrix} S_{1|1}^{-1} & & & & \\ 0 & & & & \\ -Q_{d_2}^{-1/2}\Phi_{2,1} & Q_{d_2}^{-1/2} & & & \\ & Q_{y_2}^{-1/2}A_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & -Q_{d_k}^{-1/2}\Phi_{k,k-1} & Q_{d_k}^{-1/2} \\ & & & & Q_{y_k}^{-1/2}A_k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix}$$

Before the time and measurement update of the SRIF are discussed in detail, we will show how these steps work out in the model of observation equations as a whole. The underlying sections 2 and 3 closely parallel the treatment of the (standard) Kalman filter in appendix B. We start with model (2.1) and for the initialization an $m_1 x m_1$ orthogonal transformation matrix T is applied to the upper m_1 observation equations. The resulting model is given in (2.2).

The m_1 observables $\underline{y}_1$ are transformed into the n-vector of estimators $\hat{z}_{1|1}$ and the $(m_1\text{-}n)$-vector of misclosures $\underline{u}_1$. The misclosures have all coefficients equal zero. They will be put at the bottom of the model. The designmatrix $Q_{y_1}^{-1/2}A_1$ has been transformed into an upper triangular matrix of which the upper nxn part is $S_{1|1}^{-1}$. The estimator for x_1 follows from $\hat{\underline{x}}_{1|1} = S_{1|1}\hat{z}_{1|1}$ with $Q_{\hat{x}_{1|1}} = S_{1|1}S_{1|1}^T$.

The recursion can start once the first state estimator is available. A sequence of alternatingly time and measurement updates is carried out until all observations have been treated. We will consider the time update from epoch k-1 to epoch k. In equation (2.3) the observations from epoch 1 up to and including epoch k-1 have been treated. The filtered state estimator $\hat{\underline{x}}_{k-1|k-1}$ can be computed from $\hat{z}_{k-1|k-1}$. At epoch k-1 the model reads (2.3).

$$
(2.3) \qquad E\left\{\begin{pmatrix} \underline{z}_1 \\ \underline{z}_2 \\ \cdot \\ \cdot \\ \hat{z}_{k-1|k-1} \\ Q_{d_k}^{-1/2}\underline{d}_k \\ Q_{y_k}^{-1/2}\underline{y}_k \\ \underline{u}_1 \\ \underline{w}_2 \\ \cdot \\ \cdot \\ \underline{w}_{k-1} \end{pmatrix}\right\} = \begin{pmatrix} S_{1,1}^{-1} & S_{1,2}^{-1} & & & & & \\ & S_{2,2}^{-1} & S_{2,3}^{-1} & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & S_{k-1|k-1}^{-1} & & \\ & & & & -Q_{d_k}^{-1/2}\Phi_{k,k-1} & Q_{d_k}^{-1/2} & \\ & & & & & Q_{y_k}^{-1/2}A_k & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix}
$$

The transformed observables $z_1,..,z_{k-2}$ are related to the states $x_1,..,x_{k-1}$. By the upper triangular structure of the designmatrix and together with $\hat{z}_{k-1|k-1}$, smoothed estimators $\hat{\underline{x}}_{1|k-1},..,\hat{\underline{x}}_{k-2|k-1}$ can be determined instead of filtered estimators. $\underline{w}_2,..,\underline{w}_{k-1}$ are the normalized predicted residuals.

The time update from epoch k-1 to k consists of a $2nx2n$ orthogonal transformation matrix T, applied to $\hat{z}_{k-1|k-1}$ and $Q_{d_k}^{-1/2}\underline{d}_k$. After the time update the model reads (2.4). The transformed observables are z_{k-1} and $\hat{z}_{k|k-1}$. From the latter, the predicted state estimator $\hat{\underline{x}}_{k|k-1}$ can be computed.

$$E\left\{\begin{pmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ z_{k-1} \\ \hat{z}_{k|k-1} \\ Q_{y_k}^{-1/2} y_k \\ \underline{u}_1 \\ \underline{w}_2 \\ \cdot \\ \cdot \\ \underline{w}_{k-1} \end{pmatrix}\right\} = \begin{pmatrix} S_{1,1}^{-1} & S_{1,2}^{-1} & & & & & \\ & S_{2,2}^{-1} & S_{2,3}^{-1} & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & S_{k-1,k-1}^{-1} & S_{k-1,k}^{-1} & \\ & & & & & S_{k|k-1}^{-1} & \\ & & & & & Q_{y_k}^{-1/2} A_k & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix} \tag{2.4}$$

Then the measurement update at epoch k can be carried out. An $(n+m_k)$x$(n+m_k)$ orthogonal transformation matrix T is applied to $\hat{z}_{k|k-1}$ and $Q_{y_k}^{-1/2} y_k$.

After the measurement update at epoch k, the model finally becomes (2.5). The resulting observables are $\hat{z}_{k|k}$ and $\underline{w}_k$, the latter has been transferred to the bottom of the model.

$$E\left\{\begin{pmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ z_{k-1} \\ \hat{z}_{k|k} \\ \underline{u}_1 \\ \underline{w}_2 \\ \cdot \\ \cdot \\ \underline{w}_{k-1} \\ \underline{w}_k \end{pmatrix}\right\} = \begin{pmatrix} S_{1,1}^{-1} & S_{1,2}^{-1} & & & & \\ & S_{2,2}^{-1} & S_{2,3}^{-1} & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & S_{k-1,k-1}^{-1} & S_{k-1,k}^{-1} \\ & & & & & S_{k|k}^{-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix} \tag{2.5}$$

The filtered state estimator $\hat{x}_{k|k}$ can now be computed from $\hat{z}_{k|k}$ through upper triangular matrix $S_{k|k}^{-1}$. Model (2.5) is still completely equivalent with the model we started with, equation (2.1). Only orthogonal transformations have been applied.

The lower part of designmatrix (2.5) contains only zeros. This is also the case for the designmatrix of appendix B, (B.5). The upper *knxkn* part has a blocked upper-triangular structure as opposed to the transformed designmatrix (B.5), which contains only k I_n unit

matrices in the upper part. The variance covariance matrix of the observables in (2.5) is a unit matrix, whereas the variance covariance matrix of (B.7) is a block diagonal one.

As model (2.5) is the result of the full transformation upon (2.1), the structure of (A.12) applies. The vector of transformed observables can be split into the 'estimation' part, the vectors z, and the 'redundancy' part, the vectors w (and u). The kn observables of the estimation part, with the blocked upper triangular matrix S^{-1}, uniquely determine (smoothed) estimators for $x_1,..,x_k$. Note that the redundancy part does not play a role in this as these observables are not functionally related to the states $x_1,..,x_k$, nor stochastically related to the transformed state estimators $\underline{z}_1,..,\underline{z}_{k-1}$, $\hat{\underline{z}}_{k|k}$. All redundancy has been transferred to the lower part of the model. The redundancy equals $(\sum_{i=1}^{k} m_i) - n$. The equivalent model of (2.5) in terms of condition equations reads

$$
\begin{pmatrix}
I_{m_1-n} & & & & \\
& I_{m_2} & & & \\
& & \ddots & & \\
& & & I_{m_{k-1}} & \\
& & & & I_{m_k}
\end{pmatrix}
E\left\{
\begin{pmatrix}
\underline{z}_1 \\ \underline{z}_2 \\ \vdots \\ \underline{z}_{k-1} \\ \hat{\underline{z}}_{k|k} \\ \underline{u}_1 \\ \underline{w}_2 \\ \vdots \\ \underline{w}_{k-1} \\ \underline{w}_k
\end{pmatrix}
\right\}
=
\begin{pmatrix}
0 \\ 0 \\ \vdots \\ 0 \\ 0
\end{pmatrix}
\tag{2.6}
$$

The transformed observables $\underline{u}_1$, $\underline{w}_2,..,\underline{w}_k$ are uncorrelated and have all unit variance; they are normalized. From (2.6) we see that together, they constitute a valid set of misclosures to model (2.5) (also linear combinations of them can be taken). A linear one-to-one mapping exists between this set and $\underline{t}_1$, $\underline{v}_2,..,\underline{v}_k$ of (B.6), see also (A.16).

The transformed observables $\underline{u}_1$, $\underline{w}_2,..,\underline{w}_k$ become available epochwise; $\underline{w}_i$ contains the (normalized) misclosures of the local model with $\hat{z}_{i-1|i-1}$, $Q_{d_i}^{-1/2} d_i$ and $Q_{y_i}^{-1/2} y_i$ as observations. In conclusion, $\underline{w}_2,..,\underline{w}_k$ are normalized predicted residuals. These residuals are used for model validation as discussed in section 3.

2.2.1 Time update

The part of (2.3) that is of interest for the time update is

$$
E\left\{
\begin{pmatrix}
\hat{z}_{k-1|k-1} \\ Q_{d_k}^{-1/2} \underline{d}_k
\end{pmatrix}
\right\}
=
\begin{pmatrix}
S_{k-1|k-1}^{-1} & 0 \\
-Q_{d_k}^{-1/2} \Phi_{k,k-1} & Q_{d_k}^{-1/2}
\end{pmatrix}
\begin{pmatrix}
x_{k-1} \\ x_k
\end{pmatrix}
\tag{2.7}
$$

with unit variance covariance matrix. This model has no redundancy. The $2nx2n$ orthogonal transformation matrix T is applied to the observables $\hat{z}_{k-1|k-1}$ and $Q_{d_k}^{-1/2}\underline{d}_k$. The resulting designmatrix is upper triangular:

$$E\{\begin{pmatrix} z_{k-1} \\ \hat{z}_{k|k-1} \end{pmatrix}\} = \begin{pmatrix} S_{k-1,k-1}^{-1} & S_{k-1,k}^{-1} \\ 0 & S_{k|k-1}^{-1} \end{pmatrix} \begin{pmatrix} x_{k-1} \\ x_k \end{pmatrix} \tag{2.8}$$

z_{k-1} is an n-vector of transformed observables, $S_{k-1,k-1}^{-1}$ is an nxn upper triangular matrix and $S_{k-1,k}^{-1}$ is an nxn square full matrix. These quantities are needed for the computation of smoothed estimators, see also [Park et al, 1995]. The notation 'inverse of $S_{k-1,k-1}$' is used as this matrix is part of the upper triangular matrix S^{-1} of the model of observation equations as a whole. The comma is used in the index instead of the vertical line, to denote that the matrix $S_{k-1,k-1}^{-1}$ itself is not the factor of $Q_{\hat{x}_{k-1|k-1}}^{-1}$. The upper n observation equations of (2.8) do not change anymore in further processing. The predicted state estimator follows from the lower n equations

$$\hat{x}_{k|k-1} = S_{k|k-1}\hat{z}_{k|k-1}$$

with the variance covariance matrix as

$$Q_{\hat{x}_{k|k-1}} = S_{k|k-1}S_{k|k-1}^T$$

The observables $\hat{z}_{k|k-1}$ and matrix $S_{k|k-1}^{-1}$ are used in the measurement update.

2.2.2 Measurement update

The part of (2.4) that is of interest for the measurement update is

$$E\{\begin{pmatrix} \hat{z}_{k|k-1} \\ Q_{y_k}^{-1/2} y_k \end{pmatrix}\} = \begin{pmatrix} S_{k|k-1}^{-1} \\ Q_{y_k}^{-1/2} A_k \end{pmatrix} (x_k) \tag{2.9}$$

with unit variance covariance matrix. The redundancy equals m_k. The $(n+m_k)x(n+m_k)$ orthogonal transformation matrix T is applied to the observables $\hat{z}_{k|k-1}$ and $Q_{y_k}^{-1/2}\underline{y}_k$. The resulting designmatrix is upper triangular:

$$E\{\begin{pmatrix} \hat{z}_{k|k} \\ \underline{w}_k \end{pmatrix}\} = \begin{pmatrix} S_{k|k}^{-1} \\ 0 \end{pmatrix} (x_k) \tag{2.10}$$

The filtered state estimator follows from the upper n equations

$$\hat{x}_{k|k} = S_{k|k}\hat{z}_{k|k}$$

with the variance covariance matrix as

$$Q_{\hat{z}_{k|k}} = S_{k|k} S_{k|k}^T$$

The observables $\hat{z}_{k|k}$ and matrix $S_{k|k}^{-1}$ are used in the next time update.

2.2.3 Normalized predicted residuals

In appendix B we have seen that the predicted residuals are the cornerstone of the testing procedure that parallels the recursive estimation. Below (2.6) we concluded that $\underline{w}_2,..,\underline{w}_k$ are the normalized predicted residuals, or normalized innovations [Park et al, 1995]. The normalized predicted residuals can be related to the predicted residuals epochwise, by the square $m_i x m_i$ and full rank transformation matrix W_i:

$$\underline{v}_i = W_i \underline{w}_i$$

for $i=2,..,k$. This relation is used in the next section to rewrite the formula for the generalized likelihood ratio teststatistic (B.18) in terms of normalized predicted residuals. The fact that the normalized predicted residuals have unit variance covariance matrix, will simplify the formulae.

2.3 Square Root Information Filter: quality control

In this section, the quality control in the Square Root Information Filter is discussed. The emphasis lies on validation of the mathematical model. The DIA testing procedure can be executed parallel to the recursive estimation that has been discussed in section 2.2. The DIA-procedure has been reviewed in appendix B, section B.3.

2.3.1 Testing: DIA-procedure

The null hypothesis is given by model (2.1). In the alternative hypothesis, the functional model is extended by adding the q-vector of unknowns ∇. The variance covariance matrix is as in (2.1).

$$E\left\{\begin{pmatrix} Q_{y_1}^{-1/2}\underline{y}_1 \\ Q_{d_2}^{-1/2}\underline{d}_2 \\ Q_{y_2}^{-1/2}\underline{y}_2 \\ \cdot \\ \cdot \\ Q_{d_k}^{-1/2}\underline{d}_k \\ Q_{y_k}^{-1/2}\underline{y}_k \end{pmatrix}\right\} = \begin{pmatrix} Q_{y_1}^{-1/2}A_1 & & & & \\ -Q_{d_2}^{-1/2}\Phi_{2,1} & Q_{d_2}^{-1/2} & & & \\ & Q_{y_2}^{-1/2}A_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & -Q_{d_k}^{-1/2}\Phi_{k,k-1} & Q_{d_k}^{-1/2} \\ & & & & Q_{y_k}^{-1/2}A_k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix} + \begin{pmatrix} Q_{y_1}^{-1/2}C_{y_1} \\ Q_{d_2}^{-1/2}D_2 \\ Q_{y_2}^{-1/2}C_{y_2} \\ \cdot \\ \cdot \\ Q_{d_k}^{-1/2}D_k \\ Q_{y_k}^{-1/2}C_{y_k} \end{pmatrix} \nabla \tag{3.1}$$

In the recursive estimation, the transformations applied to the designmatrix discussed in the previous section, are applied to the model misspecification matrix in (3.1) as well. The alternative hypothesis has been propagated into (3.2) at epoch k. Matrix $Z_{k|k}$ is the so-called response matrix: it determines the effect of the model error ∇ on the estimator $\hat{z}_{k|k}$: $Z_{k|k} = S_{k|k}^{-1} X_{k|k}$ (the matrices Z_i for $\underline{z}_i$). The matrices C_{w_i} are explicitly used in the testing procedure, $C_{v_i} = W_i C_{w_i}$.

$$(3.2) \qquad E\left\{\begin{pmatrix} \underline{z}_1 \\ \underline{z}_2 \\ \cdot \\ \cdot \\ \underline{z}_{k-1} \\ \hat{\underline{z}}_{k|k} \\ \underline{u}_1 \\ \underline{w}_2 \\ \cdot \\ \cdot \\ \underline{w}_{k-1} \\ \underline{w}_k \end{pmatrix}\right\} = \begin{pmatrix} S_{1,1}^{-1} & S_{1,2}^{-1} & & & & \\ & S_{2,2}^{-1} & S_{2,3}^{-1} & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & S_{k-1,k-1}^{-1} & S_{k,k-1}^{-1} \\ & & & & & S_{k|k}^{-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ Z_{k-1} \\ Z_{k|k} \\ C_{u_1} \\ C_{w_2} \\ \cdot \\ \cdot \\ C_{w_{k-1}} \\ C_{w_k} \end{pmatrix} \nabla$$

The null and alternative hypothesis, respectively (2.5) and (3.2), can be reformulated for the purpose of testing with the teststatistic $\underline{T}_q$ in

$$(3.3) \qquad H_o: E\left\{\begin{pmatrix} \underline{u}_1 \\ \underline{w}_2 \\ \cdot \\ \cdot \\ \underline{w}_k \end{pmatrix}\right\} = 0 \quad \textit{and} \quad H_a: E\left\{\begin{pmatrix} \underline{u}_1 \\ \underline{w}_2 \\ \cdot \\ \cdot \\ \underline{w}_k \end{pmatrix}\right\} = \begin{pmatrix} C_{u_1} \\ C_{w_2} \\ \cdot \\ \cdot \\ C_{w_k} \end{pmatrix} \nabla$$

or with the normalized predicted residuals concatenated in vector w

$$(3.4) \qquad H_o: E\{\underline{w}\} = 0 \quad \textit{and} \quad H_a: E\{\underline{w}\} = C_w \nabla$$

where the normalized misclosures $\underline{u}_1$ have been left out, cf. (B.17). These misclosures can however, be treated just like normalized predicted residuals. The general form of the generalized likelihood ratio teststatistic expressed in terms of normalized predicted residuals reads

$$(3.5) \qquad \underline{T}_q = \underline{w}^T C_w \left(C_w^T C_w\right)^{-1} C_w^T \underline{w}$$

and has a central χ^2-distribution with q degrees of freedom under the null hypothesis. The null hypothesis is rejected in favour of the alternative hypothesis if $T_q > K_\alpha$. Under the alternative hypothesis, the non-centrality parameter is given by

$$\lambda = \nabla^T C_w^T C_w \nabla \tag{3.6}$$

The estimator for the model error then reads

$$\hat{\underline{\nabla}} = (C_w^T C_w)^{-1} C_w^T \underline{w} \tag{3.7}$$

detection

In the detection step the overall validity of the null hypothesis is checked. No particular alternative hypothesis is specified. The Global Overall Model (GOM) teststatistic reads

$$\underline{T}^{l,k} = \sum_{i=l}^{k} \underline{w}_i^T \underline{w}_i \tag{3.8}$$

The null hypothesis is rejected in favour of the alternative hypothesis if $T > K_\alpha$; a model error is said to have occurred. The inner product $w^T w$ can be taken epochwise as the variance covariance matrix Q_w is unit. The Overall Model teststatistic is also given on p.71 of [Bierman, 1977] and termed sum of squares of the residuals. For the alternative hypothesis with $l=k$, the Local Overall Model (LOM) teststatistic is obtained.

identification

After an (unspecified) model error has been detected, the identification is performed to trace the model error. We will consider only alternative hypotheses, which are one-dimensional, i.e. ∇ is a scalar ($q=1$). The matrix C_w in the alternative hypothesis reduces to a vector. It can be computed in the recursion.

The Global Slippage (GS) teststatistic for testing against an alternative hypothesis, in which the model misspecification concerns the epochs l to k, reads

$$\underline{t}^{l,k} = \frac{\sum_{i=l}^{k} c_{w_i}^T \underline{w}_i}{\sqrt{\sum_{i=l}^{k} c_{w_i}^T c_{w_i}}} \tag{3.9}$$

The slippage teststatistic $\underline{t}$ has a standard normal distribution, i.e. $\underline{t} \sim N(0,1)$, under the null hypothesis. The null hypothesis is rejected in favour of the alternative hypothesis if $|t| > k_\alpha$; the model error specified by the alternative hypothesis is found to be sufficiently likely and said to have occurred.

The Local Slippage (LS) teststatistic for epoch k is obtained by substitution of $l=k$ in (3.9).

adaptation

After detection and identification of a model error, the estimator for the state under the alternative hypothesis needs to be computed. We will compute it, using the results obtained under the null hypothesis. The n-vector of observables determining the filtered state estimator is denoted by $\hat{z}^o_{k|k}$.

Below, the adaptation is shown for a q-dimensional local ($l=k$) model error with $1 \leq q \leq m_k$. Model (3.2) then reads

$$E\{\begin{pmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ z_{k-1} \\ \hat{z}^o_{k|k} \\ \underline{u}_1 \\ \underline{w}_2 \\ \cdot \\ \cdot \\ \underline{w}_{k-1} \\ \underline{w}_k \end{pmatrix}\} = \begin{pmatrix} S_{1,1}^{-1} & S_{1,2}^{-1} & & & & & & \\ & S_{2,2}^{-1} & S_{2,3}^{-1} & & & & & \\ & & \cdot & & & & & \\ & & & \cdot & & & & \\ & & & & S_{k-1,k-1}^{-1} & S_{k,k-1}^{-1} & & \\ & & & & & S_{k|k}^{o\ -1} & Z_{k|k} & \\ & & & & & & & C_{w_k} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \\ \nabla \end{pmatrix} \tag{3.10}$$

with unit variance covariance matrix. The part of interest for the adaptation is

$$E\{\begin{pmatrix} \hat{z}^o_{k|k} \\ \underline{w}_k \end{pmatrix}\} = \begin{pmatrix} Z_{k|k} & S_{k|k}^{o\ -1} \\ C_{w_k} & 0 \end{pmatrix} \begin{pmatrix} \nabla \\ x_k \end{pmatrix} \tag{3.11}$$

where the order of the unknowns has been reversed. The redundancy equals (m_k-q). In the $(n+m_k)$x$(q+n)$ designmatrix, the nxn block $S_{k|k}^{o\ -1}$ at the right is already upper triangular (factorization-result of the estimation under H_o). The overall upper triangular shape is realized by additional orthogonal transformations.

$$E\{\begin{pmatrix} z_\nabla \\ \hat{z}^a_{k|k} \\ \underline{w}^a_k \end{pmatrix}\} = \begin{pmatrix} S_\nabla^{-1} & S_{\nabla,k}^{-1} \\ & S_{k|k}^{a\ -1} \\ & \end{pmatrix} \begin{pmatrix} \nabla \\ x_k \end{pmatrix} \tag{3.12}$$

z_∇ is a q-vector, $\hat{z}^a_{k|k}$ is an n-vector, $\underline{w}^a_k$ an (m_k-q)-vector, S_∇^{-1} a qxq upper triangular matrix, $S_{\nabla,k}^{-1}$ a (qxn)-matrix and $S_{k|k}^{a\ -1}$ the nxn upper triangular matrix. The estimator for the filtered state under H_a follows from

$$\hat{\underline{x}}^a_{k|k} = S^a_{k|k}\, \hat{z}^a_{k|k}$$

and the variance covariance matrix becomes

$$Q_{\hat{x}^a_{k|k}} = S^a_{k|k}\, S^{a\,T}_{k|k}$$

The first q equations of (3.12) can be left out, if the estimator for ∇ is not needed in further data processing.

The alternative hypothesis H_a has become the new null hypothesis. Yet another alternative hypothesis may be specified and treated similarly. By means of statistical testing, based on the (updated) normalized predicted residuals $\underline{w}^a_k$, one may decide to either accept or reject the (new) null hypothesis (in favour of the specified alternative hypothesis). The adaptation procedure can be repeated again. In this way the adaptation can handle multiple (local) model errors recursively. Finally note that in principle the idea can be carried through to handle also multiple model errors identified by global tests like (3.9). In that case the normalized predicted residuals $\underline{w}_l$ through $\underline{w}_k$ have to be updated.

2.3.2 Measures for quality: precision and reliability

Under the null hypothesis the quality of the estimator is described by the precision. The precision of the filtered state estimator $\hat{\underline{x}}_{k|k}$ is given by the variance covariance matrix $Q_{\hat{x}_{k|k}}$.

The quality of the estimators, if model errors occur, is described by reliability. After taking a reference value λ_o for the non-centrality parameter (3.6), we can compute the Minimal Detectable Bias, that corresponds to teststatistic (3.9)

(3.13)
$$|\nabla^{l,k}| = \sqrt{\frac{\lambda_o}{\sum_{i=l}^{k} c_{w_i}^T c_{w_i}}}$$

The model error in terms of the transformed observations in given by (3.14).

The impact of an undetected model error (3.14) on the filtered estimator for the transformed state reads

(3.15)
$$\nabla\hat{z}_{k|k} = Z_{k|k}\nabla^{l,k}$$

with $X_{k|k} = S_{k|k}Z_{k|k}$, cf. (B.31), and on the state itself

$$\nabla\hat{x}_{k|k} = S_{k|k}\nabla\hat{z}_{k|k}$$

$$\nabla \begin{pmatrix} z_1 \\ \cdot \\ \cdot \\ z_{l-1} \\ z_l \\ \cdot \\ \cdot \\ \hat{z}_{k|k} \\ u_1 \\ \cdot \\ \cdot \\ w_{l-1} \\ w_l \\ \cdot \\ \cdot \\ w_k \end{pmatrix} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ Z_l \\ \cdot \\ \cdot \\ Z_{k|k} \\ 0 \\ \cdot \\ \cdot \\ 0 \\ C_{w_l} \\ \cdot \\ \cdot \\ C_{w_k} \end{pmatrix} \nabla^{l,k} \tag{3.14}$$

The significance of bias (3.15) can be judged upon by means of the following Bias to Noise Ratio (BNR)

$$\lambda_{\hat{x}_{k|k}} = \nabla \hat{z}_{k|k}^T \nabla \hat{z}_{k|k} \tag{3.16}$$

2.4 Mechanization aspects

A filter mechanization is an implementation of the recursive estimation. Two different mechanizations have been given already: the Square Root Information Filter, or SRIF (section 2) and the Kalman, or standard covariance filter (appendix B). Mechanizations differ in the way the computations are made, but for the same problem they all (should) come up with identical estimation results.

This section starts with a very brief overview of the various classes of filter mechanizations. Next, the considerations are given, that lead the decision to the square root information filter for the kinematic GPS surveying application.

2.4.1 Review of mechanizations

A classification of mechanizations can be made after the propagation of the variance covariance matrix $Q_{\hat{x}}$ of the estimator. In information filters $Q_{\hat{x}}^{-1}$ is propagated and in covariance filters $Q_{\hat{x}}$. Beside this division, we will maintain the two categories, mentioned in section 2.2: implementations can be based on the normal equations or on the reduction of the designmatrix to a canonical form. Respectively the full (inverse) variance covariance matrix or a certain factor of this matrix is propagated. For the latter, the use of a square root factor S, e.g. the Cholesky factor, is commonplace.

$$Q_{\hat{x}} = S\, S^T \quad or \quad Q_{\hat{x}}^{-1} = S^{-T} S^{-1}$$

We thus distinguish between two categories: standard mechanizations and square root mechanizations. For the covariance type of filters also the U-D factorization is well known. Instead of propagating a square root factor, the U- and D-factors of the UDU^T-decomposition of $Q_{\hat{x}}$ are propagated. The factorization filters have been developed for the reason of numerical stability.

	full matrix $Q_{\hat{x}}$	square root S
covariance	covariance	square root covariance
information	information	square root information

Table 4.1: Major classes of filter mechanizations

Many modifications and variations within the classes mentioned above do exist. The algorithmic details of these mechanizations will not be discussed; the reader is referred to the textbooks [Bierman, 1977], [Maybeck, 1979], [Anderson et al, 1979], [Minkler et al, 1993] and [Grewal et al, 1993].

In the remaining of this overview, we will point out the duality between the covariance and information filters. The duality concerns the limiting cases of the statistical uncertainty of the state estimator: complete uncertainty and no uncertainty (deterministic state), see also [Morf et al, 1975].

In principle, the information filter can not accomodate problems with perfect (deterministic) knowledge (zero variances), whereas the covariance filter can not accomodate problems involving no a-priori (zero) knowledge (infinite variances). Perfect knowledge about the state implies namely $Q_{\hat{x}}=0 \rightarrow Q_{\hat{x}}^{-1}=\infty$ and zero knowledge corresponds to $Q_{\hat{x}}^{-1}=0 \rightarrow Q_{\hat{x}}=\infty$.

	zero knowledge	perfect knowledge
covariance	$Q_{\hat{x}} = \infty$	$Q_{\hat{x}} = 0$
information	$Q_{\hat{x}}^{-1} = 0$	$Q_{\hat{x}}^{-1} = \infty$

Table 4.2: Limiting cases of knowledge about state

The zero knowledge case may occur at the start up of the recursion. At epoch $k=0$, no knowledge about the system is present. With only one epoch of data, it may still be the case that not all parameters are directly estimable; the (partly) zero knowledge situation may last for some more epochs. Information filters are in principle, able to accomodate this situation, see also [Park et al, 1995]. For a covariance filter, a separate initialization is required to provide an initial state estimator based on the first few epochs and therewith safely start the recursion.

The perfect knowledge situation on the other hand, may arise in the data processing by e.g. external non-stochastic information about the state. The covariance filters can simply incorporate this information. For the information filters, the mathematical model has to be explicitly reformulated or some numerical approximation (to zero variance) must be made.

2.4.2 Argumentation for SRIF

In addition to the initialization and the zero variance case, three more aspects in choosing the mechanization appropriate for a particular application, are the availability of the results, the computational efficiency and the numerical accuracy.

availability of results
All mechanizations should come up with identical estimation results. The standard Kalman filter is the only mechanization in which both the estimate $\hat{x}$ and the variance covariance matrix $Q_{\hat{x}}$ are explicitly available (explaining its numereous use in practice). In the information filters, the 'inverse' information is propagated. The state estimate $\hat{x}$ and matrix $Q_{\hat{x}}$ are available only at extra computational cost. This propagation of 'inverse' information complicates the time update as compared with the covariance filters, but simplifies on the other hand the measurement update.

Testing in recursive data processing (appendix B.3) is based on the predicted residual $\underline{v}$ and its variance covariance matrix Q_v. In general they explicitly follow from the covariance filters at no extra cost. In the U-D covariance factorization filter the predicted residuals are not directly available. In principle they are not computed in the standard information filter. Normalized predicted residuals, however, are directly available in the square root information filter mechanization.

Concerning the availability of estimation results, the standard covariance filter is preferred. The normalized predicted residuals of the square root information filter are very convenient for testing.

computational efficiency
Data processing results should be obtained at reasonable computational cost. The number of arithmetic operations required, should be as small as possible, as this number will directly translate into the amount of CPU-time the data processing algorithm consumes. This will be of particular importance in real-time filter applications.

Two general and indicative statements in literature about computational efficiency are first that the information filters are computationally more demanding than the covariance filters are and secondly that using a factorization mechanization instead of a standard one, increases the computational load; this is the price to pay for increased numerical stability. For comparisons of filter mechanizations on computational efficiency, see [Bierman, 1977] and [Maybeck, 1979].

Three remarks are in order. The statements in literature concern only the estimation, not the quality control computations. Secondly, the importance of the computational efficiency aspect of the filter mechanization is limited by the fact that the peripheral computations, such as administration and I/O, may account for a large part of the total computational burden of the recursive data processing (and thus largely determine the so-called filter-cycle time). These peripherial computations are needed, independent of the filter mechanization. Finally, the actual efficiency will depend very much on the application at hand. A rigorous comparison between the mechanizations, all applied in kinematic GPS surveying, has not been made. Such a comparison will be hard to carry out as per mechanization, a wide variety exists of modifications which may affect the computational efficiency.

In general the conclusion reads that from a computational point of view, the standard covariance filter is to be preferred. This aspect is however, of little weight.

numerical accuracy
For the data processing we set up a mathematical model that is thought to adequately describe the measurement system at hand. For the moment we assume this model to be perfect. Then we have to realize that, the problem we solve with a computer, is an approximation of the problem formulated in the mathematical model: unknowns, observations and model matrices are in general continuous quantities. They are, however, stored in the computer memory using a limited number of bits, as floating numbers. Arithmetic operations among floating numbers introduce additional fractional errors. They are called roundoff errors. Usually roundoff errors accumulate with increasing amounts of calculation. When they play a major role in the computations using a certain mechanization, this mechanization is said to be unstable (filter divergence). The reader is referred to the examples in [Maybeck, 1979].

Numerical inaccuracy may occur in particular in ill-conditioned problems. A 'bad' measurement scenario enters the numerics via the design matrix. The variance covariance matrix $Q_{\hat{x}}$ obtained for the state will be ill-conditioned. This ill-condition shows up in the spectrum of eigenvalues. The eigenvalues typically range from close to zero to very large numbers: certain (linear combinations of) unknowns are determined very precisely, whereas others are nearly unobservable; see also the definition of the condition number [Golub et al, 1989] and [Marel, 1989].

The numerical problems concentrate around the variance covariance matrix $Q_{\hat{x}}$. In kinematic GPS surveying, the occurrence of ill-conditionedness is unavoidable as ambiguous range measurements are made in a configuration that hardly changes during the short observation time span, see also the discussion in [Lichten, 1990]. To avoid numerical problems as much as possible (computed estimation results must be correct at any price), we opt for an information type of filter. In an ill-conditioned system some unknowns can hardly be determined. This was called the zero knowledge case and the information filter was recommended.

Due to roundoff errors in arithmetic operations, the variance covariance matrix (or its inverse) may lose its positive definite property during the recursive data processing. Propagation of a square root instead of the variance covariance matrix itself, guarantees symmetry and positive (semi-)definiteness. These two properties are enforced by computing only an upper or lower triangular part. We cite [Maybeck, 1979]: square root mechanizations exhibit improved numerical precision and stability, particularly in ill-conditioned problems. The square root approach *can* yield twice the effective precision of the conventional mechanization in ill-conditioned problems, see [ibid] and [Lawson et al, 1974]. Note that in practice also other factors play a role in this, see [Golub et al, 1989].

For the aspect of numerical accuracy, the use of the square root information filter is required.

concluding remarks
An ideal mechanization simply does not exist. For the kinematic GPS surveying application, the square root information mechanization is thought to be the best candidate. The two primary considerations are, first that it enables a proper initialization in an ill-conditioned problem (therefore the information type of filter is chosen) and secondly that it is

numerically more stable (factorized Kalman filter approaches possess superior numerical characteristics [Lichten, 1990], and therefore the square root is chosen). An example of geodetic GPS data processing with a square root information filter is given in [ibid].

Finally we give some additional minor considerations concerning the square root information filter. In section 2, it was shown that orthogonal transformations, applied to the full model of observation equations, yield an upper triangular structure for the designmatrix together with an explicit splitting up of the observations into an estimation part and a redundancy part. This straightforward procedure easily allows further data processing, such as the computation of smoothed estimates and the recovery (adaptation) from (multiple) global model errors. Furthermore the normalized predicted residuals have unit variance covariance matrix, which yields some simplification in the quality control computations.

In summary, the Square Root Information Filter (SRIF) turns out to be very well suited for the 'Delft' approach to recursive estimation and quality control in general, and for kinematic GPS surveying in particular.

2.5 Summary

In this chapter, the formulae for the recursive data processing using the Square Root Information Filter have been presented. In each step of the recursion, the designmatrix is reduced to a canonical form. The square root of the inverse variance covariance matrix of the state estimator is directly propagated, and this has positive consequences for the numerical accuracy of the mechanization.

Orthogonal transformations are applied to (a part of) the model of observation equations, thereby preserving the full information content. The observables are split into the estimation part (although the estimator for the state vector is not explicitly computed) and the redundancy part. The redundancy part contains the normalized predicted residuals, which can conveniently be used in the testing procedure. With the DIA-procedure, model validation can be carried out parallel to the recursive estimation.

Appendix 2.A Estimation by means of QR-factorization

In this appendix, estimates for the unknowns in a model of observation equations will be computed by means of the reduction of the designmatrix to a canonical form. In estimation via QR-factorization, the designmatrix is reduced to an upper triangular matrix by orthogonal transformations. This method is opposed to performing a Cholesky factorization upon the normal matrix. For the introduction on estimation via QR-factorization, we will follow the discussion in [Golub et al, 1989], where chapter 5 is primarily concerned with the least-squares solution of an overdetermined system of equations.

2.A.1 Introduction

The mxn design matrix A, with full rank n and $m \geq n$, can be factored as

(A.1) $$A = T^T R$$

in which T^T is an *mxm* orthogonal matrix (the column (and row) vectors are orthogonal unit vectors and thus orthonormal; T^T is therefore better called an orthonormal matrix) and R an *mxn* upper triangular matrix. Note that as opposed to [ibid] where $A=QR$, we use kernel T to denote the orthogonal transformation matrix.

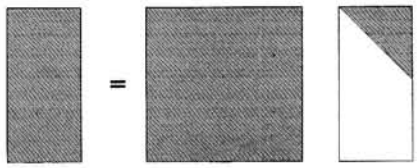


Figure A.1: Factorization of the design matrix

For the orthogonal matrix T^T we have, see section 2.5.1. in [ibid],

$$TT^T = I_m \qquad \text{(A.2)}$$

From this follows that $T^T=T^{-1}$ and $T^TT=I_m$. The matrix T has determinant $|T|=\pm 1$. The eigenvalues of T are $\lambda_i=\pm 1$ for $i=1,..,m$. The orthogonal transformation leaves invariant, the norm of a vector and the angle between two vectors [Park et al, 1995]. Matrix T is perfectly conditioned, the condition number equals 1. The condition of the matrix A can not deteriorate through the application of successive orthogonal transformations: $T_n \ldots T_1 A$.

The model of observation equations under the null hypothesis reads

$$E\{\underline{y}\} = Ax \;\; ; \;\; D\{\underline{y}\} = I_m \qquad \text{(A.3)}$$

We assume for the moment that the observables have a unit variance covariance matrix, i.e. $D\{\underline{y}\}=I_m$ (this can be achieved by normalization, see section A.2). With factorization (A.1) the model can be rewritten into

$$E\{\underline{y}\} = T^TRx \;\; ; \;\; D\{\underline{y}\} = I_m \qquad \text{(A.4)}$$

Premultiplication of the functional model of (A.4) by transformation matrix T, which is square *mxm* and has full rank, yields (A.5). The orthogonal transformation T realizes a conformal mapping of the columns of A onto those of TA [ibid].

$$E\{T\underline{y}\} = TAx = TT^TRx = Rx \;\; ; \;\; D\{T\underline{y}\} = I_m \qquad \text{(A.5)}$$

Note that model (A.5) is equivalent to model (A.3). The BLUE-estimators for x, computed from (A.5) and (A.3) will be identical. Matrix R is upper triangular and has full rank n. The estimates can easily be found by backward substitution upon the first n equations of (A.5). The last $(m\text{-}n)$ transformed observables have expectation equal zero. The estimation via factorization (A.1) will be treated in detail in section A.2.

Matrix T^T can be partitioned as

$$T^T = \left(T_1^T \;\; T_2^T\right) \qquad \text{(A.6)}$$

in which T_1^T has dimensions *mxn* and T_2^T *mx(m-n)*. It is important to note that the column vectors of T_1^T constitute an orthonormal basis for the space $R(A)$ and that the

column vectors of T_2^T are an orthonormal basis for $R(A)^\perp$, the orthogonal (under standard metric) complement of $R(A)$. This is verified by noting that the product $T_2 A$ yields an $(m\text{-}n)$xn zero matrix. The column vectors of T^T together span the full observational space R^m, see also section 5.2.6. in [Golub et al, 1989]. We will come back to these observations in section A.2.

Several methods exist for the construction of the orthonormal matrix T^T. In chapter 5 of [ibid] we find the Householder transformation, the Givens rotation and the Gram-Schmidt orthogonalization, see also chapter 10 of [Lawson et al, 1974]. According to [Golub et al, 1989] the Householder transformation is important because its ability to zero specified components of a matrix. The Householder transformation is useful for zeroing on a grand scale, while the Givens rotation is the tool to zero elements more selectively. The numerical properties of the latter are as favourable as those for the Householder reflection, sections 5.1.5. and 5.1.10. in [ibid]. Least-squares estimation using Givens rotations is also discussed in [Blais, 1985]. The classical Gram-Schmidt orthogonalization has very poor numerical properties and the Modified Gram-Schmidt has no clear advantages (concerning efficiency and numerical precision) over the Householder transformation, sections 5.2.9. and 5.3.6. in [Golub et al, 1989]. In the sequel, although not explicitly stated, we will use the Householder transformation, see also appendix A of chapter 4. It is presented in full detail in [Bierman, 1977] and the references go back via [Golub, 1965] to [Householder, 1958].

2.A.2 Estimation

In this section estimation for a model of observation equations via QR-factorization is treated. Only a theoretical description is given; algorithmic aspects are not dealt with. The estimation consists of normalization of the observables, the orthogonal (Householder) transformation, the actual estimation and the corresponding model with condition equations, which leads to the normalized misclosures. We also make a comparison with the method of normal equations of chapter 1 and treat some aspects of quality control.

normalization
The model of observation equations under the null hypothesis reads

$$E\{\underline{y}\} = Ax \quad ; \quad D\{\underline{y}\} = Q_y \tag{A.7}$$

First the variance covariance matrix of the observables is factorized, e.g. using Cholesky as the matrix is symmetric and positive definite

$$Q_y = Q_y^{1/2} Q_y^{1/2^T}$$

Both sides of the functional model are premultiplied by the inverse of this factor. This transformation matrix is square and has full rank equal m.

$$E\{Q_y^{-1/2}\underline{y}\} = Q_y^{-1/2}Ax \quad ; \quad D\{Q_y^{-1/2}\underline{y}\} = I_m \tag{A.8}$$

The observables $\underline{y}$ have been normalized, the new variance covariance matrix equals the mxm unit matrix.

orthogonal transformation

The estimation is now carried out through an orthogonal transformation. Transformation matrix T is square mxm and has full rank.

(A.9) $$E\{TQ_y^{-1/2}\underline{y}\} = TQ_y^{-1/2}Ax \;\; ; \;\; D\{TQ_y^{-1/2}\underline{y}\} = I_m$$

The purpose of the orthogonal transformation is to reduce the design matrix to an upper triangular one. With the same partition of the orthogonal matrix T as in (A.6)

(A.10) $$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

where T_1 is nxm and T_2 $(m\text{-}n)xm$, model (A.9) becomes

(A.11) $$E\left\{\begin{pmatrix} T_1 Q_y^{-1/2}\underline{y} \\ T_2 Q_y^{-1/2}\underline{y} \end{pmatrix}\right\} = \begin{pmatrix} T_1 Q_y^{-1/2}A \\ T_2 Q_y^{-1/2}A \end{pmatrix} x$$

We define the

n-vector	$\underline{z} = T_1 Q_y^{-1/2}\underline{y}$
$(m\text{-}n)$-vector	$\underline{u} = T_2 Q_y^{-1/2}\underline{y}$
nxn upper triangular matrix	$S^{-1} = T_1 Q_y^{-1/2}A$
$(m\text{-}n)xn$ zero matrix	$0 = T_2 Q_y^{-1/2}A$

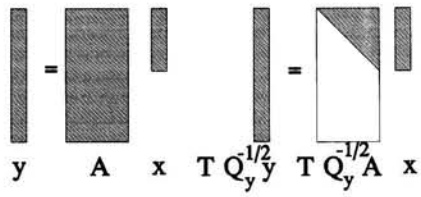


Figure A.2: Model of observation equations, before (left) and after transformation (right)

The model can be rewritten into:

(A.12) $$E\left\{\begin{pmatrix} \underline{z} \\ \underline{u} \end{pmatrix}\right\} = \begin{pmatrix} S^{-1} \\ 0 \end{pmatrix} x \;\; ; \;\; D\left\{\begin{pmatrix} \underline{z} \\ \underline{u} \end{pmatrix}\right\} = I_m$$

estimation

Model (A.12) is still in the form of a model of observation equations. We are, however, already very close to obtaining the BLUE estimator $\hat{\underline{x}}$. The vector of observables is $TQ_y^{-1/2}\underline{y} = \begin{pmatrix} \underline{z} \\ \underline{u} \end{pmatrix}$. The estimator for the vector of observables follows from $\hat{\underline{y}} = P_A \underline{y}$, here

$$TQ_y^{-1/2}\hat{\underline{y}} = P_{(TQ_y^{-1/2}A)}\, TQ_y^{-1/2}\underline{y}$$

The orthogonal projection of vector $TQ_y^{-1/2}y$ onto the range space $R(TQ_y^{-1/2}A)$

$$P_{(TQ_y^{-1/2}A)} = \begin{pmatrix} I_n & \\ & 0 \end{pmatrix}$$

implies maintaining the first n elements and zeroing the last $(m\text{-}n)$ elements. The estimator becomes $TQ_y^{-1/2}\hat{\underline{y}} = \begin{pmatrix} \underline{z} \\ 0 \end{pmatrix}$. As $TQ_y^{-1/2}\hat{\underline{y}} = TQ_y^{-1/2}A\hat{\underline{x}}$, the BLUE estimator for the vector of unknowns follows from the first n equations of (A.12):

(A.13) $$\hat{\underline{x}} = S\,\underline{z}$$

The estimator $\hat{\underline{x}}$ is a linear combination of only the first n transformed observables: $\underline{z}$. The system $z = S^{-1}x$ is namely consistent, $z \in R(S^{-1})$ and yields a unique solution for x. The lower part of (A.12) will be not consistent as in general $u \neq 0$.

The variance covariance matrix becomes

(A.14) $$Q_{\hat{x}} = S\,S^T$$

It can not be stated strictly that matrix S is the (backward) Cholesky factor of $Q_{\hat{x}}$, or that S^{-1} is the Cholesky factor of $Q_{\hat{x}}^{-1}$, as matrix S^{-1} is produced by a sequence of orthogonal transformations; it may have negative diagonal elements. We cite from [Lawson et al, 1974] p. 125: thus if $rank(A)=n$, the matrix S^{-1} of the QR-decomposition is identical with the Cholesky matrix of the normal matrix to within signs of rows.

model with condition equations

The transformed observables $\underline{z}$ of (A.12) will not appear in the condition equations; all redundancy has been concentrated in the lower part: $\underline{u}$. The equivalent model with condition equations reads

(A.15) $$(0\ D)\ E\{\begin{pmatrix} \underline{z} \\ \underline{u} \end{pmatrix}\} = 0 \ ; \ D\{\begin{pmatrix} \underline{z} \\ \underline{u} \end{pmatrix}\} = I_m$$

with D any square $(m\text{-}n)$x$(m\text{-}n)$-matrix of full rank, e.g. $D=I_{m-n}$ is a valid choice.

The last $(m\text{-}n)$ transformed observables $\underline{u}$ have expectation value equal zero and seem therefore to be the misclosures $\underline{t}$. Considering the model with condition equations (A.15) yields the conclusion that they are a set of $(m\text{-}n)$ *normalized* misclosures.

Model $E\{\underline{y}\} = Ax$; $D\{\underline{y}\} = Q_y$ is equivalent to $B^T E\{\underline{y}\} = 0$; $D\{\underline{y}\} = Q_y$ and therefore:

$$E\{TQ_y^{-1/2}\underline{y}\} = TQ_y^{-1/2}A\,x \ ; \ D\{TQ_y^{-1/2}\underline{y}\} = I_m$$

is equivalent to

$$B^T Q_y^{1/2} T^T\, E\{TQ_y^{-1/2}\underline{y}\} = 0 \ ; \ D\{TQ_y^{-1/2}\underline{y}\} = I_m$$

With the definition of the square $(m\text{-}n)$x$(m\text{-}n)$ and full rank matrix V

$$V = B^T Q_y^{1/2} T_2^T$$

and the notion that the columns of T_1^T span $R(Q_y^{-1/2}A)$ and thus $B^T Q_y^{1/2} T_1^T = 0$, the model with condition equations becomes

$$(0\ V)\ E\{\begin{pmatrix} z \\ \underline{u} \end{pmatrix}\} = 0 \quad ; \quad D\{\begin{pmatrix} z \\ \underline{u} \end{pmatrix}\} = I_m$$

Undoing the (invertible) transformations T and $Q_y^{-1/2}$ then yields

$$V\underline{u} = B^T \underline{y} = \underline{t}$$

and the important conclusion

(A.16) $$\underline{t} = V\underline{u}$$

is that there exists a one-to-one linear relation between the misclosures $\underline{t}$ and the normalized misclosures $\underline{u}$. Also the latter can be used in the testing.

comparison
In section 3 of chapter 1, a square and full rank transformation matrix was applied to the observables. The transformed observables were $\hat{\underline{x}}$ and $\underline{t}$. The transformation matrix was based on the normal equations. The estimators become explicitly available.

In this appendix, the orthogonal transformation results in observables $\underline{z}$ and $\underline{u}$. The transformation preserves the unit variance covariance matrix of the normalized observables. The estimator is not explicitly available. We actually determine estimators for z; a reparametrization of the unknowns according to $x = Sz$ has taken place. The estimators $\hat{\underline{z}}$ are fully decorrelated (and normalized).

$$E\{\begin{pmatrix} \hat{\underline{x}} \\ \underline{t} \end{pmatrix}\} = \begin{pmatrix} I \\ 0 \end{pmatrix} x \qquad E\{\begin{pmatrix} \underline{z} \\ \underline{u} \end{pmatrix}\} = \begin{pmatrix} S^{-1} \\ 0 \end{pmatrix} x$$

$$D\{\begin{pmatrix} \hat{\underline{x}} \\ \underline{t} \end{pmatrix}\} = \begin{pmatrix} Q_{\hat{x}} & \\ & Q_t \end{pmatrix} \qquad D\{\begin{pmatrix} \underline{z} \\ \underline{u} \end{pmatrix}\} = \begin{pmatrix} I & \\ & I \end{pmatrix}$$

Estimation with normal equations (left) and QR-factorization (right)

quality control: alternative hypothesis
The alternative hypothesis reads

(A.17) $$E\{\underline{y}\} = Ax + C_y\nabla \quad ; \quad D\{\underline{y}\} = Q_y$$

and normalization yields

(A.18) $$E\{Q_y^{-1/2}\underline{y}\} = Q_y^{-1/2}Ax + Q_y^{-1/2}C_y\nabla \;;\; D\{Q_y^{-1/2}\underline{y}\} = I_m$$

The alternative hypothesis is transformed into

(A.19) $$E\{TQ_y^{-1/2}\underline{y}\} = TQ_y^{-1/2}Ax + TQ_y^{-1/2}C_y\nabla \;;\; D\{TQ_y^{-1/2}\underline{y}\} = I_m$$

Defining the nxq-matrix $C_z = T_1 Q_y^{-1/2} C_y$ and the $(m\text{-}n)$xq-matrix $C_u = T_2 Q_y^{-1/2} C_y$ yields:

(A.20) $$E\{\begin{pmatrix}\underline{z}\\ \underline{u}\end{pmatrix}\} = \begin{pmatrix}S^{-1}\\ 0\end{pmatrix}x + \begin{pmatrix}C_z\\ C_u\end{pmatrix}\nabla \;;\; D\{\begin{pmatrix}\underline{z}\\ \underline{u}\end{pmatrix}\} = I_m$$

The generalized likelihood ratio teststatistic can be expressed in terms of normalized misclosures. Relation (A.16) is used, together with $C_t = VC_u$ and $Q_t = VQ_uV^T$, with $Q_u = I_{m-n}$. The teststatistic becomes

(A.21) $$\underline{T}_q = \underline{u}^T C_u \,(C_u^T C_u)^{-1}\, C_u^T \underline{u}$$

Teststatistic $\underline{T}_q$ has a central χ^2-distribution under the null hypothesis with q degrees of freedom.

quality control: adaptation

We have already computed the estimator for x under the null hypothesis. Based on the results of the testing procedure one may decide to reject the null hypothesis in favour of a specific alternative hypothesis H_a. This alternative hypothesis is declared to be the new working (null) hypothesis and the estimator for x under the alternative hypothesis H_a has to be computed.

We apparently start the estimation procedure again, now using the alternative hypothesis H_a. We do, however, make use of the factorization results obtained in the estimation under H_o. Only a few additional orthogonalization steps will be required. The estimation under H_a can namely be carried out very efficiently once one realizes that switching from the null hypothesis H_o to the alternative hypothesis amounts to only a slight change of the design matrix:

H_o: $(Q_y^{-1/2}A)$ $\qquad mxn$

H_a: $(Q_y^{-1/2}C_y \vdots Q_y^{-1/2}A)$ $\qquad mx(q+n)$

where we put the q-dimensional model error ∇ in top of the vector of unknowns. The alternative hypothesis (A.18) has been transformed into (A.20).

(A.22) $$E\{\begin{pmatrix}\underline{z}\\ \underline{u}\end{pmatrix}\} = \begin{pmatrix}C_z & S^{-1}\\ C_u & 0\end{pmatrix}\begin{pmatrix}\nabla\\ x\end{pmatrix}$$

The $mx(q+n)$-design matrix is not upper triangular yet: the first q columns are filled. The design matrix (A.20) is the one of (A.12) to which q columns are appended. In the QR-factorization of design matrix (A.18) we may use the factorization of design matrix (A.8) as a starting point, see section 12.6 of [Golub et al, 1989]. An additional number of orthogonalization steps, concatenated in the mxm orthonormal matrix H, will be carried out on (A.22).

$$E\left\{\begin{pmatrix} z_\nabla \\ z^a \\ \underline{u}^a \end{pmatrix}\right\} = \begin{pmatrix} S_\nabla^{-1} & S_{\nabla,a}^{-1} \\ & S^{a-1} \end{pmatrix} \begin{pmatrix} \nabla \\ x \end{pmatrix} \tag{A.23}$$

and the estimators $\hat{\nabla}$ and $\hat{x}^a$ can be retrieved in a way analogously to (A.13). Note that the dimension of the vector of misclosures has been reduced from $(m-n)$ for $\underline{u}$ to $(m-q-n)$ for $\underline{u}^a$.

Equation (A.23), which is still equivalent to (A.17), has become the new null hypothesis. If desired, an alternative hypothesis, also encompassing the $\tilde{q}$-dimensional model error $\tilde{\nabla}$, can be formulated

$$E\{Q_y^{-1/2}\underline{y}\} = (Q_y^{-1/2}C_y \vdots Q_y^{-1/2}A)\begin{pmatrix} \nabla \\ x \end{pmatrix} + Q_y^{-1/2}\tilde{C}_y\tilde{\nabla} \quad ; \quad D\{Q_y^{-1/2}\underline{y}\} = I_m \tag{A.24}$$

Application of the transformation HT, yields

$$E\left\{\begin{pmatrix} z_\nabla \\ z^a \\ \underline{u}^a \end{pmatrix}\right\} = \begin{pmatrix} S_\nabla^{-1} & S_{\nabla,a}^{-1} \\ & S^{a-1} \end{pmatrix} \begin{pmatrix} \nabla \\ x \end{pmatrix} + \begin{pmatrix} C_{z_\nabla}^a \\ C_z^a \\ C_u^a \end{pmatrix} \tilde{\nabla} \tag{A.25}$$

where $C_{z_\nabla}^a$ is a $qx\tilde{q}$-matrix, C_z^a is a $nx\tilde{q}$-matrix and C_u^a a $(m-q-n)x\tilde{q}$-matrix. By means of statistical testing, based on the (updated) normalized misclosures $\underline{u}^a$, one can oppose the null hypothesis (A.18) to the alternative hypothesis (A.24). If necessary the procedure described can be repeated again, and hence a recursive estimation procedure in case of multiple model errors becomes feasible.

As in general q, the number of additional model parameters in H_a, is small, it is recommended to carry out the additional orthogonalization H using the Givens rotation [ibid], instead of full orthogonalization with the Householder transformation. One should exploit the fact that although the first q columns in (A.22) are filled, the last n columns have already the upper triangular shape. With Givens rotation the adaptation can be carried out very efficiently.

2.A.3 Concluding remarks

Estimation based on the model of observation equations $E\{\underline{y}\}=Ax$ can be realized basically (1) via normal equations (and factorization of the normal matrix) or (2) via the

factorization of the design matrix A. An example of the second option, QR-factorization with the Householder transformation, has been treated in this appendix.

Beside estimation also the quality control computations can be carried out conveniently. The generalized likelihood ratio teststatistic can be expressed in terms of normalized misclosures and switching from the null hypothesis to some alternative hypothesis is straightforward. We thus described complete data processing based on QR-factorization.

We have seen that in this approach, we are able to compute the estimates $\hat{x}$ and the variance covariance matrix $Q_{\hat{x}}$, directly, i.e. without constructing normal equations. This has positive consequences for the numerical accuracy, [Golub et al, 1989]. The approach is computationally more demanding [ibid], but when both estimation and testing need to be carried out, the Householder transformation may be a viable alternative to the solution via normal equations, as all quantities needed in the testing become automatically available after the transformation, see also the comparison in [Copps et al, 1994]. The Householder orthogonalization directly provides also an orthonormal basis for the orthogonal complement $R(Q_y^{-1/2}A)^{\perp}$.

In [Bierman, 1977] we find that solving the model $E\{\underline{y}\}=Ax$ by means of the Householder transformation is more elegant than the 'brute force' computation of the estimates via the normal equations. The vector of observations is transformed and split into a part that uniquely determines the vector of unknowns x, and another part that expresses the fit of the model to the observations and vice versa, which can be used for model validation. By means of the Householder transformations, we actually do not solve the system of equations explicitly, but only transform it.

Appendix 2.B Review recursive data processing and DIA-procedure

In this appendix we will briefly review the recursive estimation in a time varying system, highlight the predicted residual and discuss the procedure for model validation in the recursion.

2.B.1 Estimation

Model (1.1) of section 2.1 is detailed with observation equations (1.2) and (1.3). In [Teunissen et al, 1988] it was shown that straightforward application of the BLUE-estimation principle to the model of observation equations leads to the well known Kalman filter time- and measurement update. This approach to Kalman filtering is adopted, but instead of treating time- and measurement update explicitly as in [ibid], we treat the full model of observation equations and apply linear transformations to (a part of) the observables, such that the estimators of time- and measurement update become available. After that, we will also give the explicit treatment of the well-known time- and measurement update. They result from considering 'local' models of observation equations.

The model of observation equations is given in (B.1). The designmatrix is of full rank. The number of observations equals $(\sum_{i=1}^{k} m_i)+(k-1)n$, the number of unknowns kn. The

variance covariance matrix is block-diagonal (no time correlation) and assumed to be symmetric positive definite.

$$(B.1)\qquad E\{\begin{pmatrix} \underline{y}_1 \\ \underline{d}_2 \\ \underline{y}_2 \\ . \\ . \\ \underline{d}_k \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} A_1 & & & & & \\ -\Phi_{2,1} & I_n & & & & \\ & A_2 & & & & \\ & & . & & & \\ & & & . & & \\ & & & & -\Phi_{k,k-1} & I_n \\ & & & & & A_k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_{k-1} \\ x_k \end{pmatrix} ;\ D\{\begin{pmatrix} \underline{y}_1 \\ \underline{d}_2 \\ \underline{y}_2 \\ . \\ . \\ \underline{d}_k \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} Q_{y_1} & & & & & & \\ & Q_{d_2} & & & & & \\ & & Q_{y_2} & & & & \\ & & & . & & & \\ & & & & . & & \\ & & & & & Q_{d_k} & \\ & & & & & & Q_{y_k} \end{pmatrix}$$

In the initialization a square $m_1 x m_1$ full rank transformation matrix T is applied to the upper m_1 observation equations, cf. section 1.3. The resulting model is given in (B.2).

The m_1 observables $\underline{y}_1$ are transformed into the n-vector of estimators $\underline{\hat{x}}_{1|1}$ and the $(m_1\text{-}n)$-vector of misclosures $\underline{t}_1$. The misclosures have all coefficients equal zero. They will be put at the bottom of the model. The estimator $\underline{\hat{x}}_{1|1}$ and misclosures $\underline{t}_1$ are not correlated, see again section 3 of chapter 1.

The model of observation equations after the initialization:

$$(B.2)\qquad E\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{t}_1 \\ \underline{d}_2 \\ \underline{y}_2 \\ . \\ . \\ \underline{d}_k \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} I_n & & & & & \\ 0 & & & & & \\ -\Phi_{2,1} & I_n & & & & \\ & A_2 & & & & \\ & & . & & & \\ & & & . & & \\ & & & & -\Phi_{k,k-1} & I_n \\ & & & & & A_k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_{k-1} \\ x_k \end{pmatrix} ; D\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{t}_1 \\ \underline{d}_2 \\ \underline{y}_2 \\ . \\ . \\ \underline{d}_k \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} Q_{\hat{x}_{1|1}} & & & & & & \\ & Q_{t_1} & & & & & \\ & & Q_{d_2} & & & & \\ & & & Q_{y_2} & & & \\ & & & & . & & \\ & & & & & Q_{d_k} & \\ & & & & & & Q_{y_k} \end{pmatrix}$$

The recursion can start once the first state estimator is available. A sequence of so-called time- and measurement updates is carried out until all observables have been treated. We will consider the time update from epoch k-1 to epoch k. In equation (B.3) all observables from epoch 1 to k-1 have been treated. The filtered state estimator is given by $\underline{\hat{x}}_{k-1|k-1}$.

Filtered estimators for previous states are given by $\underline{\hat{x}}_{2|2}$ through $\underline{\hat{x}}_{k-2|k-2}$. The redundancy at each epoch results in the predicted residuals $\underline{v}_2,..,\underline{v}_{k-1}$. In the functional model they have all coefficients equal zero. Similar to the misclosures, they will be used for model validation.

At epoch k-1 the model reads:

$$E\left\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{\hat{x}}_{2|2} \\ \cdot \\ \cdot \\ \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \\ \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \cdot \\ \underline{v}_{k-1} \end{pmatrix}\right\} = \begin{pmatrix} I_n & & & & & & \\ & I_n & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & I_n & & \\ & & & & -\Phi_{k,k-1} & I_n & \\ & & & & & A_k & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix}$$

(B.3)

$$D\left\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{\hat{x}}_{2|2} \\ \cdot \\ \cdot \\ \underline{\hat{x}}_{k-1|k-1} \\ \underline{d}_k \\ \underline{y}_k \\ \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \cdot \\ \underline{v}_{k-1} \end{pmatrix}\right\} = \begin{pmatrix} Q_{\hat{x}_{1|1}} & Q & \cdot & \cdot & Q & & & 0 & Q & \cdot & \cdot & Q \\ Q & Q_{\hat{x}_{2|2}} & \cdot & \cdot & Q & & & & 0 & \cdot & \cdot & Q \\ \cdot & \cdot & \cdot & & \cdot & & & & & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & & & & & \cdot & \cdot \\ Q & Q & \cdot & \cdot & Q_{\hat{x}_{k-1|k-1}} & & & & & & & 0 \\ & & & & & Q_{d_k} & & & & & & \\ & & & & & & Q_{y_k} & & & & & \\ 0 & & & & & & & Q_{t_1} & & & & \\ Q & 0 & & & & & & & Q_{v_2} & & & \\ \cdot & \cdot & \cdot & & & & & & & \cdot & & \\ \cdot & \cdot & & \cdot & & & & & & & \cdot & \\ Q & Q & \cdot & \cdot & 0 & & & & & & & Q_{v_{k-1}} \end{pmatrix}$$

Matrix Q is used to denote any covariance block, i.e. to denote correlation. Note that the filtered estimator $\underline{\hat{x}}_{i|i}$ with i=2,..,k-1, does not correlate with predicted residuals $\underline{v}_j$ for $2 \leq j \leq i$, but does correlate with future predicted residuals $\underline{v}_j$ with $i < j \leq k$-1. The fact that the filtered estimator $\underline{\hat{x}}_{i|i}$ is not correlated with the present predicted residuals $\underline{v}_i$ is explicitly denoted by a zero in (B.3).

The time update from epoch k-1 to k consists of a $2n$x$2n$ square and full rank transformation matrix T, applied to $\underline{\hat{x}}_{k-1|k-1}$ and $\underline{d}_k$. After the time update the model reads (B.4). The transformed observables are $\underline{\hat{x}}_{k-1|k-1}$ and $\underline{\hat{x}}_{k|k-1}$. The latter is the predicted state estimator.

$$E\left\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{\hat{x}}_{2|2} \\ \vdots \\ \underline{\hat{x}}_{k-1|k-1} \\ \underline{\hat{x}}_{k|k-1} \\ \underline{y}_k \\ \underline{t}_1 \\ \underline{v}_2 \\ \vdots \\ \underline{v}_{k-1} \end{pmatrix}\right\} = \begin{pmatrix} I_n & & & & \\ & I_n & & & \\ & & \ddots & & \\ & & & I_n & \\ & & & I_n & \\ & & & A_k & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \\ x_k \end{pmatrix}$$

(B.4)

$$D\left\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{\hat{x}}_{2|2} \\ \vdots \\ \underline{\hat{x}}_{k-1|k-1} \\ \underline{\hat{x}}_{k|k-1} \\ \underline{y}_k \\ \underline{t}_1 \\ \underline{v}_2 \\ \vdots \\ \underline{v}_{k-1} \end{pmatrix}\right\} = \begin{pmatrix} Q_{\hat{x}_{1|1}} & Q & \cdots & Q & Q & 0 & Q & \cdots & Q \\ Q & Q_{\hat{x}_{2|2}} & \cdots & Q & Q & & 0 & \cdots & Q \\ \vdots & \vdots & \ddots & \vdots & \vdots & & & \ddots & \vdots \\ Q & Q & \cdots & Q_{\hat{x}_{k-1|k-1}} & Q & & & & 0 \\ Q & Q & \cdots & Q & Q_{\hat{x}_{k|k-1}} & & & & \\ & & & & & Q_{y_k} & & & \\ 0 & & & & & & Q_{t_1} & & \\ Q & 0 & & & & & & Q_{v_2} & \\ \vdots & \vdots & \ddots & & & & & & \ddots \\ Q & Q & \cdots & 0 & & & & & Q_{v_{k-1}} \end{pmatrix}$$

For the measurement update at epock k an $(n+m_k)$x$(n+m_k)$ square and full rank transformation matrix T is applied to $\underline{\hat{x}}_{k|k-1}$ and $\underline{y}_k$. The resulting observables are $\underline{\hat{x}}_{k|k}$ and $\underline{v}_k$, the latter has been transferred to the bottom of model (B.5). The filtered estimator for the state at epock k is available. Model (B.5) is completely equivalent with model (B.1), only square and full rank (and thus invertible) transformation matrices have been applied.

The lower part of designmatrix (B.5) contains only zeros. The upper *knxkn* part actually is a *kn* unit matrix. The vector of transformed observables can be split into the 'estimation' part, the vectors x, and the 'redundancy' part, the vectors v (and t). The estimation part uniquely determines filtered state estimators. For the computation of smoothed estimators also the redundancy part is needed, see (B.7).

The measurement update at epoch k results in:

$$E\left\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{\hat{x}}_{2|2} \\ \cdot \\ \cdot \\ \underline{\hat{x}}_{k-1|k-1} \\ \underline{\hat{x}}_{k|k} \\ \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \cdot \\ \underline{v}_{k-1} \\ \underline{v}_k \end{pmatrix}\right\} = \begin{pmatrix} I_n & & & & & & \\ & I_n & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & I_n & & \\ & & & & & I_n & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix}$$

(B.5)

$$D\left\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{\hat{x}}_{2|2} \\ \cdot \\ \cdot \\ \underline{\hat{x}}_{k-1|k-1} \\ \underline{\hat{x}}_{k|k} \\ \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \cdot \\ \underline{v}_{k-1} \\ \underline{v}_k \end{pmatrix}\right\} = \begin{pmatrix} Q_{\hat{x}_{1|1}} & Q & \cdot & \cdot & Q & Q & 0 & Q & \cdot & \cdot & Q & Q \\ Q & Q_{\hat{x}_{2|2}} & \cdot & \cdot & Q & Q & & 0 & \cdot & \cdot & Q & Q \\ \cdot & \cdot & \cdot & & \cdot & \cdot & & & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & & & \cdot & \cdot & \cdot \\ Q & Q & \cdot & \cdot & Q_{\hat{x}_{k-1|k-1}} & Q & & & & & 0 & Q \\ Q & Q & \cdot & \cdot & Q & Q_{\hat{x}_{k|k}} & & & & & & 0 \\ 0 & & & & & & Q_{t_1} & & & & & \\ Q & 0 & & & & & & Q_{v_2} & & & & \\ \cdot & \cdot & \cdot & & & & & & \cdot & & & \\ \cdot & \cdot & & \cdot & & & & & & \cdot & & \\ Q & Q & \cdot & \cdot & 0 & & & & & & Q_{v_{k-1}} & \\ Q & Q & \cdot & \cdot & Q & 0 & & & & & & Q_{v_k} \end{pmatrix}$$

The redundancy in (B.5) is concentrated in the lower part of the model. The redundancy equals $(\sum_{i=1}^{k} m_i) - n$. The equivalent model of (B.5) in terms of condition equations is given by (B.6), with the variance covariance matrix as given in (B.5).

The transformed observables $\underline{t}_1$, $\underline{v}_2$,..,$\underline{v}_k$ have a blocked diagonal variance covariance matrix, i.e. they are not time correlated. From (B.6) we see that together they constitute a valid set of misclosures to model (B.5), but also linear combinations of them can be taken. Therefore a one-to-one linear relation will exist between this set of transformed

observables and the vector of misclosures that will result from solving (B.1) in batch. Note that the latter will in general possess a full variance covariance matrix, instead of a block-diagonal one as in (B.5).

(B.6)
$$\begin{pmatrix} I_{m_1-n} & & & & \\ & I_{m_2} & & & \\ & & \cdot & & \\ & & & I_{m_{k-1}} & \\ & & & & I_{m_k} \end{pmatrix} E\left\{ \begin{pmatrix} \hat{\underline{x}}_{1|1} \\ \hat{\underline{x}}_{2|2} \\ \cdot \\ \hat{\underline{x}}_{k-1|k-1} \\ \hat{\underline{x}}_{k|k} \\ \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \underline{v}_{k-1} \\ \underline{v}_k \end{pmatrix} \right\} = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ 0 \\ 0 \end{pmatrix}$$

The transformed observables $\underline{t}_1$, $\underline{v}_2,..,\underline{v}_k$ become available epochwise; v_i contains the misclosures of the 'local' model with $\hat{x}_{i-1|i-1}$, d_i and y_i as observations, see also [Teunissen et al, 1988]. $\underline{v}_2,..,\underline{v}_k$ are the predicted residuals.

Finally we give by (B.7) the equivalent of (B.5) in terms of smoothed estimators. These smoothed estimators and the predicted residuals are not correlated, cf. section 3 of chapter 1. The variance covariance is block-diagonal. Note that the predicted residuals do not change in the smoothing. Once they are computed in the filtering, they are left untouched.

(B.7)
$$E\left\{ \begin{pmatrix} \hat{\underline{x}}_{1|k} \\ \hat{\underline{x}}_{2|k} \\ \cdot \\ \hat{\underline{x}}_{k-1|k} \\ \hat{\underline{x}}_{k|k} \\ \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \underline{v}_{k-1} \\ \underline{v}_k \end{pmatrix} \right\} = \begin{pmatrix} I_n & & & & \\ & I_n & & & \\ & & \cdot & & \\ & & & I_n & \\ & & & & I_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix}$$

$$
(B.7)\qquad D\left\{\begin{pmatrix}\underline{\hat{x}}_{1|k}\\ \underline{\hat{x}}_{2|k}\\ .\\ .\\ \underline{\hat{x}}_{k-1|k}\\ \underline{\hat{x}}_{k|k}\\ \underline{t}_1\\ \underline{v}_2\\ .\\ .\\ \underline{v}_{k-1}\\ \underline{v}_k\end{pmatrix}\right\} = \begin{pmatrix} Q_{\hat{x}_{1|k}} & Q & . & . & Q & Q & & & & & & \\ Q & Q_{\hat{x}_{2|k}} & . & . & Q & Q & & & & & & \\ . & . & . & & . & . & & & & & & \\ . & . & . & . & . & . & & & & & & \\ Q & Q & . & . & Q_{\hat{x}_{k-1|k}} & Q & & & & & & \\ Q & Q & . & . & Q & Q_{\hat{x}_{k|k}} & & & & & & \\ & & & & & & Q_{t_1} & & & & & \\ & & & & & & & Q_{v_2} & & & & \\ & & & & & & & & . & & & \\ & & & & & & & & & . & & \\ & & & & & & & & & & Q_{v_{k-1}} & \\ & & & & & & & & & & & Q_{v_k} \end{pmatrix}
$$

Now we have treated the model of observation equations (B.1) as a whole, we will, in the remaining of this subsection, give an explicit treatment of the well known Kalman filter time- and measurement update equations. They result from considering each time a part only of the full models given so far.

time update

The part of (B.3) that is of interest for the time update is

$$
(B.8)\qquad E\left\{\begin{pmatrix}\underline{\hat{x}}_{k-1|k-1}\\ \underline{d}_k\end{pmatrix}\right\} = \begin{pmatrix} I_n & 0\\ -\Phi_{k,k-1} & I_n\end{pmatrix}\begin{pmatrix}x_{k-1}\\ x_k\end{pmatrix};\ D\left\{\begin{pmatrix}\underline{\hat{x}}_{k-1|k-1}\\ \underline{d}_k\end{pmatrix}\right\} = \begin{pmatrix} Q_{\hat{x}_{k-1|k-1}} & \\ & Q_{d_k}\end{pmatrix}
$$

In this model there is no redundancy. The estimator for x_k reads

$$
\boxed{\underline{\hat{x}}_{k|k-1} = \Phi_{k,k-1}\underline{\hat{x}}_{k-1|k-1} + \underline{d}_k}
$$

The variance covariance matrix becomes

$$
Q_{\hat{x}_{k|k-1}} = \Phi_{k,k-1}\, Q_{\hat{x}_{k-1|k-1}}\, \Phi^T_{k,k-1} + Q_{d_k}
$$

The $2n$x$2n$ transformation matrix T, applied to the observables $\underline{\hat{x}}_{k-1|k-1}$ and $\underline{d}_k$, reads

$$
(B.9)\qquad T = \begin{pmatrix} I_n & 0\\ \Phi_{k,k-1} & I_n\end{pmatrix}
$$

The resulting model is

$$(B.10) \qquad E\{\begin{pmatrix}\hat{\underline{x}}_{k-1|k-1}\\ \hat{\underline{x}}_{k|k-1}\end{pmatrix}\} = \begin{pmatrix}I_n & \\ & I_n\end{pmatrix}\begin{pmatrix}x_{k-1}\\ x_k\end{pmatrix} \; ; \; D\{\begin{pmatrix}\hat{\underline{x}}_{k-1|k-1}\\ \hat{\underline{x}}_{k|k-1}\end{pmatrix}\} = \begin{pmatrix}Q_{\hat{x}_{k-1|k-1}} & Q\\ Q & Q_{\hat{x}_{k|k-1}}\end{pmatrix}$$

measurement update
The part of (B.4) that is of interest for the measurement update is

$$(B.11) \qquad E\{\begin{pmatrix}\hat{\underline{x}}_{k|k-1}\\ \underline{y}_k\end{pmatrix}\} = \begin{pmatrix}I_n\\ A_k\end{pmatrix}(x_k) \; ; \; D\{\begin{pmatrix}\hat{\underline{x}}_{k|k-1}\\ \underline{y}_k\end{pmatrix}\} = \begin{pmatrix}Q_{\hat{x}_{k|k-1}} & \\ & Q_{y_k}\end{pmatrix}$$

The redundancy equals m_k. The estimator for x_k reads

$$\hat{\underline{x}}_{k|k} = \hat{\underline{x}}_{k|k-1} + K_k\,(\underline{y}_k - A_k\hat{\underline{x}}_{k|k-1})$$

with the gain matrix K_k

$$K_k = Q_{\hat{x}_{k|k-1}}A_k^T\,(Q_{y_k} + A_k Q_{\hat{x}_{k|k-1}}A_k^T)^{-1}$$

The variance covariance matrix becomes

$$Q_{\hat{x}_{k|k}} = [I - K_k A_k]\; Q_{\hat{x}_{k|k-1}}$$

The $(n+m_k)$x$(n+m_k)$ transformation matrix T, applied to $\hat{\underline{x}}_{k|k-1}$ and $\underline{y}_k$, reads

$$(B.12) \qquad T = \begin{pmatrix}I_n - K_k A_k & K_k\\ -A_k & I_{m_k}\end{pmatrix}$$

The resulting model is

$$(B.13) \qquad E\{\begin{pmatrix}\hat{\underline{x}}_{k|k}\\ \underline{y}_k\end{pmatrix}\} = \begin{pmatrix}I_n\\ 0\end{pmatrix}(x_k) \; ; \; D\{\begin{pmatrix}\hat{\underline{x}}_{k|k}\\ \underline{y}_k\end{pmatrix}\} = \begin{pmatrix}Q_{\hat{x}_{k|k}} & \\ & Q_{v_k}\end{pmatrix}$$

2.B.2 Predicted residuals

In equation (B.13) the predicted residuals show up:

$$\underline{v}_k = \underline{y}_k - A_k \underline{\hat{x}}_{k|k-1}$$

with variance covariance matrix

$$Q_{v_k} = Q_{y_k} + A_k Q_{\hat{x}_{k|k-1}} A_k^T$$

The predicted residuals $\underline{v}_2,..,\underline{v}_k$ or innovations [Kailath, 1968], are the basis of the statistical testing in dynamic systems [Teunissen et al, 1988].

2.B.3 DIA-procedure

Recursive data processing encompasses recursive estimation and recursive quality control. Recursion estimation can be realized using a filter, as discussed in sections 2.1 and B.1. In quality control, the validity of the mathematical model underlying the estimation is checked. This is important as estimation results obtained with an invalid model, will be invalid as well (e.g. biased). The DIA-testing procedure, proposed in [Teunissen et al, 1988], realizes recursive quality control. The procedure for Detection, Identification and Adaptation of model errors can be executed parallel to the recursive estimation. It is based on the generalized likelihood ratio teststatistic $\underline{T}_q$ of section 1.7.2.

In the detection step, an overall model validation takes place. In the identification step, the most likely model error, the cause of the detection being positive, must be found. Finally we must account for this model error. The recovery takes place in the adaptation step. The estimators, initially computed using the null hypothesis, are corrected.

The null hypothesis is given by model (B.1), and by (B.2) through (B.5) as only one-to-one linear transformations have been applied. It might occur that instead of the null hypothesis, some alternative hypothesis is a (more) adequate description of the measurement system. To be able to guarantee the quality of the estimators (described by the precision), the mathematical model has to be validated. The null hypothesis will be tested against alternative hypotheses by means of the generalized likelihood ratio test. Below we give a general specification of the alternative hypothesis. The misspecification in the mathematical model is restricted to the functional model.

(B.14)
$$E\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{d}_2 \\ \underline{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{d}_k \\ \underline{y}_k \end{pmatrix}\right\} = \begin{pmatrix} A_1 & & & & & & \\ -\Phi_{2,1} & I_n & & & & & \\ & A_2 & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & -\Phi_{k,k-1} & I_n \\ & & & & & & A_k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix} + \begin{pmatrix} C_{y_1} \\ D_2 \\ C_{y_2} \\ \cdot \\ \cdot \\ \cdot \\ D_k \\ C_{y_k} \end{pmatrix} \nabla$$

The variance covariance matrix is as in (B.1). The null hypothesis (B.1) is actually extended by a set of q unknown parameters contained in vector ∇. The parameters are

related to the pseudo-observables $\underline{d}_i$ via D_i and to the observables $\underline{y}_i$ via C_{y_i}. Examples of misspecifications in the functional model are outliers and slips in the observations y_i, see [Salzmann, 1993].

(B.15)

$$
\begin{array}{lll}
 & & X_{1|1} \\
i:=i+1 & \rightarrow & \downarrow \\
 & & X_{i-1|i-1} \\
 & & \downarrow \\
\uparrow & & X_{i|i-1} = \Phi_{i,i-1} X_{i-1|i-1} + D_i \\
 & & \downarrow \\
 & & C_{v_i} = C_{y_i} - A_i X_{i|i-1} \\
\uparrow & & X_{i|i} = X_{i|i-1} + K_i C_{v_i} \\
 & \leftarrow &
\end{array}
$$

In the recursive estimation, the transformations applied to the designmatrix discussed in section B.1, are applied to the model misspecification matrix as well. Once C_{y_1} has been propagated into $\hat{x}_{1|1}$, the recursion for the model misspecification can start with $i=2$, see (B.15). Matrix $X_{i|i}$ is the so-called response matrix: it determines the effect of the model error ∇ on the estimator $\underline{\hat{x}}_{i|i}$. The matrices C_{v_i} will be used explicitly in the testing procedure. Note that $C_{v_i}=0$ for $i<l$, if $C_{y_i}=0$ and $D_j=0$ for all $j<l\leq k$. The predicted residuals have expectation equal zero until the model error starts to occur at epoch l. The model error is a global one for $l<k$. A model error with $l=k$ is a local one.

Model (B.14) after the measurement update at epoch k reads

$$
E\left\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{\hat{x}}_{2|2} \\ \cdot \\ \cdot \\ \underline{\hat{x}}_{k-1|k-1} \\ \underline{\hat{x}}_{k|k} \\ \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \cdot \\ \underline{v}_{k-1} \\ \underline{v}_k \end{pmatrix}\right\} = \begin{pmatrix} I_n & & & & & \\ & I_n & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & I_n & \\ & & & & & I_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \end{pmatrix} + \begin{pmatrix} X_{1|1} \\ X_{2|2} \\ \cdot \\ \cdot \\ X_{k-1|k-1} \\ X_{k|k} \\ C_{t_1} \\ C_{v_2} \\ \cdot \\ \cdot \\ C_{v_{k-1}} \\ C_{v_k} \end{pmatrix} \nabla
$$

The estimators $\underline{\hat{x}}_{1|1},\ldots,\underline{\hat{x}}_{k|k}$ are left out, as they do not contribute to the redundancy. The null and alternative hypothesis can be reformulated for the purpose of testing with the teststatistic $\underline{T}_q$ in

(B.16) $$H_o:\ E\{\begin{pmatrix} \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \cdot \\ \underline{v}_{k-1} \\ \underline{v}_k \end{pmatrix}\} = 0 \quad and \quad H_a:\ E\{\begin{pmatrix} \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \cdot \\ \underline{v}_{k-1} \\ \underline{v}_k \end{pmatrix}\} = \begin{pmatrix} C_{t_1} \\ C_{v_2} \\ \cdot \\ \cdot \\ C_{v_{k-1}} \\ C_{v_k} \end{pmatrix} \nabla$$

or with the predicted residuals concatenated in vector v

(B.17) $$H_o : E\{\underline{v}\} = 0 \quad and \quad H_a : E\{\underline{v}\} = C_v \nabla$$

The misclosures $\underline{t}_1$ have been left out (we assume that the model error occurs after epoch 1, i.e. $C_{y_1}=0$ or $l \geq 2$). This is done for the ease of the discussion and to correspond to the literature. For model validation concerning epoch 1, the misclosures $\underline{t}_1$ should be used.

As discussed below (B.6), a one-to-one mapping exists between $\underline{t}_1$, $\underline{v}_2,..,\underline{v}_k$ and the misclosures that follow from solving (B.1) in batch. The general likelihood ratio teststatistic $\underline{T}_q$ can thus be expressed also in terms of predicted residuals [Teunissen et al, 1988]:

(B.18) $$\underline{T}_q = \underline{v}^T Q_v^{-1} C_v \, (C_v^T Q_v^{-1} C_v)^{-1} \, C_v^T Q_v^{-1} \underline{v}$$

$\underline{T}_q$ has a central χ^2-distribution with q degrees of freedom under the null hypothesis, in which we have assumed that the observables are Gaussian distributed. The null hypothesis is rejected in favour of the alternative hypothesis if $T_q > K_\alpha$. Under the alternative hypothesis, the non-centrality parameter is given by

(B.19) $$\lambda = \nabla^T C_v^T Q_v^{-1} C_v \nabla$$

The estimator for the model error then reads

(B.20) $$\hat{\underline{\nabla}} = (C_v^T Q_v^{-1} C_v)^{-1} C_v^T Q_v^{-1} \underline{v}$$

detection

In the detection step the overall validity of the null hypothesis is checked. No particular alternative hypothesis is specified. The Global Overall Model (GOM) teststatistic reads

(B.21) $$\underline{T}^{l,k} = \sum_{i=l}^{k} \underline{v}_i^T Q_{v_i}^{-1} \underline{v}_i$$

The upper index ^{l,k} denotes that the model error as specified in the alternative hypothesis, concerns the epochs l to k. The degrees of freedom equals $\sum_{i=l}^{k} m_i$. The null hypothesis is

rejected in favour of the alternative hypothesis if $T>K_\alpha$; a model error is said to have occurred.

The predicted residuals are not time correlated; the variance covariance matrix is block diagonal, see also [Teunissen et al, 1989]. Therefore the inner product $v^T Q_v^{-1} v$ in (B.21) can be taken epochwise. For the alternative hypothesis in which $l=k$, the Local Overall Model (LOM) teststatistic for epoch k is obtained.

identification

After an (unspecified) model error has been detected, the identification is performed to trace the model error. We will consider only alternative hypotheses, which are one-dimensional, i.e. ∇ is a scalar ($q=1$). The matrix C_v in the alternative hypothesis (B.17) reduces to a vector. It can be computed recursively according to (B.15).

The Global Slippage (GS) teststatistic for testing against an alternative hypothesis, in which the model misspecification concerns the epochs l to k, reads

(B.22)
$$\underline{t}^{l,k} = \frac{\sum_{i=l}^{k} c_{v_i}^T Q_{v_i}^{-1} \underline{v}_i}{\sqrt{\sum_{i=l}^{k} c_{v_i}^T Q_{v_i}^{-1} c_{v_i}}}$$

The slippage teststatistic $\underline{t}$ actually is the 'square root' of the $\underline{T}_{q=l}$ teststatistic and has a standard normal distribution, i.e. $\underline{t} \sim N(0,1)$, under the null hypothesis. The null hypothesis is rejected in favour of the alternative hypothesis if $|t|>k_\alpha$; the alternative hypothesis (with the model error) is sufficiently more likely than the null hypothesis.

The Local Slippage (LS) teststatistic for epoch k is obtained by substitution of $l=k$ in (B.22).

(B.23)
$$E\left\{\begin{pmatrix} \underline{\hat{x}}_{1|1} \\ \underline{\hat{x}}_{2|2} \\ \cdot \\ \cdot \\ \underline{\hat{x}}_{k-1|k-1} \\ \underline{\hat{x}}^o_{k|k} \\ \underline{t}_1 \\ \underline{v}_2 \\ \cdot \\ \cdot \\ \underline{v}_{k-1} \\ \underline{v}_k \end{pmatrix}\right\} = \begin{pmatrix} I_n & & & & & & \\ & I_n & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & I_n & & \\ & & & & & I_n & X_{k|k} \\ & & & & & & C_{v_k} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \\ x_k \\ \nabla \end{pmatrix}$$

adaptation

After detection and identification of a model error, the estimator for the state under the alternative hypothesis needs to be computed. We will compute it, using the results obtained under the null hypothesis. The filtered state estimator is denoted by $\underline{\hat{x}}^o_{k|k}$.

The adaptation is shown for a q-dimensional local model error with $1 \leq q \leq m_k$. Model (B.5) after the measurement update under H_o, but now for the alternative hypothesis reads (B.23), with the variance covariance matrix as in (B.5). The part of (B.23) that is of interest for the adaptation is

(B.24)
$$E\{\begin{pmatrix}\underline{\hat{x}}^o_{k|k}\\ \underline{v}_k\end{pmatrix}\} = \begin{pmatrix}I_n & X_{k|k}\\ 0 & C_{v_k}\end{pmatrix}\begin{pmatrix}x_k\\ \nabla\end{pmatrix} \; ; \; D\{\begin{pmatrix}\underline{\hat{x}}^o_{k|k}\\ \underline{v}_k\end{pmatrix}\} = \begin{pmatrix}Q_{\hat{x}^o_{k|k}} & \\ & Q_{v_k}\end{pmatrix}$$

The redundancy equals $(m_k\text{-}q)$. The adapted estimator for x_k reads

$$\underline{\hat{x}}^a_{k|k} = \underline{\hat{x}}^o_{k|k} - X_{k|k}\underline{\hat{\nabla}}$$

The variance covariance matrix becomes

$$Q_{\hat{x}^a_{k|k}} = Q_{\hat{x}^o_{k|k}} + X_{k|k}Q_{\hat{\nabla}}X^T_{k|k}$$

The adaptation can be realized for model (B.24) by application of two transformations. First the $(n+m_k)$x$(n+m_k)$ transformation matrix T is applied to $\underline{\hat{x}}^o_{k|k}$ and $\underline{v}_k$ to obtain the estimator for the model error.

(B.25)
$$T = \begin{pmatrix}I_n & 0\\ 0 & \begin{matrix}(C^T_{v_k}Q^{-1}_{v_k}C_{v_k})^{-1}C^T_{v_k}Q^{-1}_{v_k}\\ B^T_{v_k}\end{matrix}\end{pmatrix}$$

The m_k-vector of predicted residuals $\underline{v}_k$ is split into the estimator $\underline{\hat{\nabla}}$ (q-vector) and the new predicted residuals $\underline{v}^a_k$ ($(m_k\text{-}q)$-vector). For the $(m_k\text{-}q)$xm_k-matrix $B^T_{v_k}$ it holds that $B^T_{v_k}C_{v_k}=0$, cf section 1.3.

(B.26)
$$E\{\begin{pmatrix}\underline{\hat{x}}^o_{k|k}\\ \underline{\hat{\nabla}}\\ \underline{v}^a_k\end{pmatrix}\} = \begin{pmatrix}I_n & X_{k|k}\\ & I_q\\ & \end{pmatrix}\begin{pmatrix}x_k\\ \nabla\end{pmatrix} \; ; \; D\{\begin{pmatrix}\underline{\hat{x}}^o_{k|k}\\ \underline{\hat{\nabla}}\\ \underline{v}^a_k\end{pmatrix}\} = \begin{pmatrix}Q_{\hat{x}_{k|k}} & & \\ & Q_{\hat{\nabla}} & \\ & & Q_{v^a_k}\end{pmatrix}$$

Secondly, the actual adaptation is carried out. Matrix T is $(n+m_k)$x$(n+m_k)$.

(B.27)
$$T = \begin{pmatrix} I_n & -X_{k|k} & 0 \\ 0 & I_q & 0 \\ 0 & 0 & I_{(m_k-q)} \end{pmatrix}$$

The resulting model reads

(B.28)
$$E\{\begin{pmatrix} \hat{\underline{x}}^a_{k|k} \\ \hat{\underline{\nabla}} \\ \underline{v}^a_k \end{pmatrix}\} = \begin{pmatrix} I_n & 0 \\ 0 & I_q \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_k \\ \nabla \end{pmatrix} ; \; D\{\begin{pmatrix} \hat{\underline{x}}^a_{k|k} \\ \hat{\underline{\nabla}} \\ \underline{v}^a_k \end{pmatrix}\} = \begin{pmatrix} Q_{\hat{x}^a_{k|k}} & Q & \\ Q & Q_{\hat{\nabla}} & \\ & & Q_{v^a_k} \end{pmatrix}$$

The alternative hypothesis H_a has now become the new null hypothesis. The same procedure for detection, identification and adaptation for model errors can be carried out again if desired (to handle multiple model errors). The testing is then based on the adapted predicted residuals $\underline{v}^a_k$. According to [Teunissen, 1990] we account, step by step, for the most likely model error and test after each step the likelihood of the remaining model errors.

In the case above, where we considered local model errors, the total number of model errors can not exceed m_k. The approach shown, can in principle be followed for the adaptation of global model errors as well, but might be a rather complex matter.

quality measures

Under the null hypothesis the quality of the estimator is described by the precision. The precision of the filtered state estimator $\hat{\underline{x}}_{k|k}$ is given by the variance covariance matrix $Q_{\hat{x}_{k|k}}$.

The purpose of carrying out the testing procedure is to detect model misspecifications if there are any. The quality of the estimators, if model errors occur, is described by reliability. Reliability measures can be divided into measures for internal reliability and measures for external reliability. Internal reliability describes the (nominal) performance of the testing procedure and external reliability describes the impact of undetected model errors on the state estimator.

With (B.19) the Minimal Detectable Bias (MDB) for (B.22) becomes

(B.29)
$$|\nabla^{l,k}| = \sqrt{\frac{\lambda_o}{\sum_{i=l}^{k} c_{v_i}^T Q_{v_i}^{-1} c_{v_i}}}$$

It gives the size the model error has to have, in order to be detected by the testing procedure with probability γ_o. The model error in terms of (transformed) observables is given by (B.30).

Separability between alternative hypotheses can be analysed using the angle ϕ between c_v and $\bar{c}_v$, similar to section 1.8.2.

(B.30)
$$\nabla \begin{pmatrix} \hat{x}_{1|1} \\ \cdot \\ \cdot \\ \hat{x}_{l-1|l-1} \\ \hat{x}_{l|l} \\ \cdot \\ \cdot \\ \hat{x}_{k|k} \\ t_1 \\ \cdot \\ \cdot \\ v_{l-1} \\ v_l \\ \cdot \\ \cdot \\ v_k \end{pmatrix} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ X_{l|l} \\ \cdot \\ \cdot \\ X_{k|k} \\ 0 \\ \cdot \\ \cdot \\ 0 \\ C_{v_l} \\ \cdot \\ \cdot \\ C_{v_k} \end{pmatrix} \nabla^{l,k}$$

The impact of an undetected model error (B.30) on the filtered estimator for the current state x_k reads

(B.31)
$$\nabla \hat{x}_{k|k} = X_{k|k} \nabla^{l,k}$$

and the significance of bias (B.31) can be judged upon by means of the following Bias to Noise Ratio (BNR)

(B.32)
$$\lambda_{\hat{x}_{k|k}} = \nabla \hat{x}_{k|k}^T Q_{\hat{x}_{k|k}}^{-1} \nabla \hat{x}_{k|k}$$

Note that Bias to Noise ratio (B.32) concerns a part of the unknowns (only x_k). The upperbound (cf. section 1.8.2) for the bias in element i of estimate $\hat{x}_{k|k}$ is given by $|\nabla \hat{x}_{k|k\,i} / \sigma_{\hat{x}_{k|k\,i}}| \leq \sqrt{\lambda_{\hat{x}_{k|k}}}$, [Teunissen, 1994].

concluding remarks

In this last section some essential aspects of quality control in a recursive estimation procedure for a dynamic system have been reviewed. The DIA-procedure is a unified procedure for (real time) model validation. It was not intended to give an extensive theoretical treatment nor a thorough discussion of all practical aspects. The reader is referred to the literature. Theoretical aspects of the DIA-procedure are given in [Teunissen et al, 1989] and [Teunissen, 1990]. Reliability aspects of the DIA-procedure are discussed in [Salzmann, 1993]. Navigational examples of recursive data processing can be found in [Salzmann, 1991a] and [Salzmann, 1991b].

The recursive estimation discussed in this appendix is based on the assumption that the observables are not time correlated. The variance covariance matrix is block diagonal, see also model (1.1). A modification however, can be made to the procedure, in order to handle some type of time correlation; state augmentation is treated in [Teunissen, 1995].

References

[Anderson et al, 1979] Anderson B.D.O. and J.B. Moore (1979). Optimal filtering. Information and system sciences series. Prentice-Hall, Inc. Englewood Cliffs.

[Bierman, 1977] Bierman, G.J. (1977). Factorization methods for discrete sequential estimation. Volume 128 in Mathematics in science and engineering. Academic Press, Inc. New York.

[Blais, 1985] Blais, J.A.R. (1985). Least-squares filtering and smoothing using Givens transformations. Manuscripta Geodaetica. Vol 10. pp. 208-212.

[Copps et al, 1994] Copps, E.M. and L.O. Lupash (1994). Extending the generality and robustness of RAIM algorithms. Proceedings of the ION National Technical Meeting. San Diego, California, USA. January 24-26. pp. 31-39.

[Golub, 1965] Golub, G.H. (1965). Numerical methods for solving linear least squares problems. Numerische Mathematik. Vol. 7. Handbook series linear algebra. Springer Verlag. Berlin. pp. 206-216.

[Golub et al, 1989] Golub, G.H. and C.F. van Loan (1989). Matrix computations. The Johns Hopkins University Press. Baltimore, Maryland. Second edition.

[Grewal et al, 1993] Grewal, M.S. and A.P. Andrews (1993). Kalman filtering; theory and practice. Prentice-Hall, Inc. Englewood Cliffs.

[Householder, 1958] Householder, A.S. (1958). Unitary triangularization of a nonsymmetric matrix. Journal of Association for Computing Machinery (ACM). Vol. 5. New York. pp. 339-342.

[Kailath, 1968] Kailath, T. (1968). An innovations approach to least-squares estimation part I: linear filtering in additive white noise. IEEE Transactions on Automatic Control. Vol. AC-13, No. 6, December. pp. 646-655.

[Lawson et al, 1974] Lawson, C.L. and R.J. Hanson (1974). Solving least squares problems. Series in automatic computation. Prentice-Hall, Inc. Englewood Cliffs.

[Lichten, 1990] Lichten, S.M. (1990). Estimation and filtering for high-precision GPS positioning applications. Manuscripta Geodaetica. Vol. 15. pp. 159-176.

[Maybeck, 1979] Maybeck, P.S. (1979). Stochastic models, estimation and control. Volume 141-1 in Mathematics in science and engineering. Academic Press, Inc. New York.

[Marel, 1989] Marel, H. v.d. (1989). Algorithmic aspects of solving large systems of equations. Lecture notes Mathematical aspects of data analysis, ISPRS tutorial. Pisa, Italy. June 1-2.

[Minkler et al, 1993] Minkler, G. and J. Minkler (1993). Theory and application of Kalman filtering. Magellan Book Company, Florida.

[Morf et al, 1975] Morf, M. and T. Kailath (1975). Square-root algorithms for least-squares estimation. IEEE Transactions on Automatic Control. Vol. 20, No. 4. August. pp. 487-497.

[Park et al, 1995] Park, P. and T. Kailath (1995). New square-root algorithms for Kalman filtering. IEEE Transactions on Automatic Control. Vol. 40, No. 5. May. pp. 895-899.

[Salzmann, 1991a] Salzmann, M.A. (1991). MDB: a designtool for integrated navigation systems. Bulletin Géodésique. Vol. 65. pp. 109-115.

[Salzmann, 1991b] Salzmann, M.A. (1991). An assessment of external reliability in the design of integrated navigation systems. Presented at IAG GM1/GM4 Symposium. Vienna, Austria. August 11-24.

[Salzmann, 1993] Salzmann, M.A. (1993). Least squares filtering and testing for geodetic navigation applications. Ph.D. thesis. Department of Geodetic Engineering, Delft University of Technology.

[Teunissen, 1990] Teunissen, P.J.G. (1990). An integrity and quality control procedure for use in multi sensor integration. Proceedings ION GPS-90. Colorado Springs, USA. September 19-21. pp. 513-522.

[Teunissen, 1994] Teunissen, P.J.G. (1994). Mathematische geodesie 2: inleiding toetsingstheorie. Lecture notes Mathematical Geodesy II, course Ge 36. Department of Geodetic Engineering, Delft University of Technology.

[Teunissen, 1995] Teunissen, P.J.G. (1995). Dynamische gegevensverwerking I. Lecture notes Dynamic Data Processing, course Ge37. Department of Geodetic Engineering, Delft University of Technology.

[Teunissen et al, 1988] Teunissen, P.J.G. and M.A. Salzmann (1988). Performance analysis of Kalman filters. Report 88.2. Section Mathematical and Physical Geodesy, Department of Geodetic Engineering, Delft University of Technology.

[Teunissen et al, 1989] Teunissen, P.J.G. and M.A. Salzmann (1989). A recursive slippage test for use in state-space filtering. Manuscripta Geodaetica. Vol. 14. pp. 383-390.

3. GPS surveying: modelling

GPS surveying
We start this chapter by marking the area of GPS surveying. Applications primarily involve the traditional land surveying tasks. GPS surveying is considered to be a part of GPS geodesy and concerns local scale applications. The working area is restricted to ten or a few tens of kilometers (the short baseline limitation). Existing GPS surveying techniques are mentioned and for kinematic GPS surveying, we come to the definition that kinematic is related to the motion of the receiver and not to a particular measurement technique. Kinematic GPS surveying is therefore used to indicate the whole class of precise, high production GPS surveying techniques.

concept of geodetic positioning with GPS
The two fundamental observables for positioning with GPS are the code and the carrier phase. The concepts of both types of measurements are discussed. The unknowns in the observation equations are geometric parameters and bias parameters. Geodetic positioning with code and phase observables implies relative positioning as opposed to single point positioning: two receivers simultaneously make observations. This concept of positioning has been developed from radio interferometry.

Both code and phase are range related observables. In section 3, we will discuss the measurement configuration with satellites and receivers for distance observations. The coordinates of a second receiver are determined with respect to the coordinates of a reference receiver.

mathematical model for GPS surveying
The GPS observation equation is set up and developed further to arrive at a mathematical model for relative positioning over short distances, i.e. the GPS surveying category of applications. The mathematical model consists of a functional and a stochastic model. First the observation equation of a single phase observation, in terms of the expectation and dispersion, is treated. Then also the observation equations for the code observations are given. These observation equations are non-linear in the receiver coordinates, and need to be linearized for the estimation.

In section 5 it is shown that an equivalent formulation of the resulting mathematical model for GPS surveying can be given: the model of double difference observation equations. This implementation has found widespread use in practice. Section 6 briefly treats some related GPS mathematical models. There are the parametrization in terms of ranges instead of (baseline) coordinates, and the model for Differential GPS.

The development of the mathematical model in this chapter will be rather straightforward. No fundamentals on reference systems, signal propagation and receiver technology are

treated. Also the underlying physics of the GPS, the system specifications and model assumptions are not dealt with. The reader is referred to the textbooks on GPS positioning.

3.1 GPS surveying

According to [Alberda, 1990], land surveying, as a part of geodesy in a broad sense, is concerned with a local description of a part of the earth. Concerning topographic surveying, it results in a geometric description of an area that is restricted to about 50 by 50 kilometers. Kinematic GPS surveying falls within this category of land surveying and the above characteristics still apply. They are discussed below.

Traditional topographic surveying results in two-dimensional geometric information (the remaining one-dimensional height system is left out of consideration, as it is related to the gravity potential and therefore not a geometric system). GPS surveying yields Cartesian coordinates in a three-dimensional coordinate system, e.g. WGS84. The coordinates of one receiver relative to the other are determined, or actually the coordinates of the survey marker underneath the tripod with the antenna. It must be realized that the coordinates of this discrete point (idealization) are just a tool in our attempt to describe the topography in the real world, and certainly not a goal in itself.

One main purpose of land surveying is to survey topography. In these normal surveying tasks, GPS receivers replace conventional angle and distance measuring devices like theodolites, EDMs and total stations. Applications of kinematic GPS surveying are discussed at the end of this section.

The working area of surveying concerning topography is typically restricted to 50 by 50 kilometers. This restriction was based on the flat earth approximation. In that case, also most of the geodynamic effects can be left out of consideration. In kinematic GPS surveying, the area (baseline length) is restricted to 10 or a few tens of kilometers, but the restriction has another cause. The relatively simple mathematical models for GPS surveying do not account for differential atmospheric delays. This simplification of equal ionospheric and tropospheric delays from one satellite to both receivers is allowed when the receiver separation is small (for a 10 kilometer separation the differential delays are usually at or below the millimeter level). The errors introduced by this neglect of course directly depend on the actual atmospheric conditions. For larger receiver separations also orbital errors need to be taken into account. In GPS surveying, satellite coordinates are computed from the broadcast ephemeris received from the satellites. The accuracy of these coordinates lies in the order of ten or tens of meters.

Practically speaking, most surveyors' project areas will be limited to a few tens of kilometers. The working area of the mathematical model for GPS surveying, however, can be extended by applying a-priori atmospheric corrections and using precise orbits (instead of the broadcast ones). The orbit and corrective information can be obtained from an Active GPS Reference System, a network of permanently operating GPS receivers.

Now we have sketched the working area of GPS surveying, we will discuss the measurement techniques. Using the GPS for land surveying started with the static survey. We cite from [Remondi, 1988]: 'Static GPS using the carrier phase was developed from ideas which have their origins in VLBI...Carrier phase measurements are recorded at both sites

and the data are sent or carried to the home office for processing. At that time it was typical to collect data for 2-3 hours. In the home office the three-dimensional baseline vector was estimated from these measurements'. From then on three important developments have taken place.

Instead of processing the data afterwards in the home office, the data processing can take place in the field in (near) real time. The data of the reference receiver needs to be transmitted via a telemetry link, to the second receiver. There, the recursive estimation and quality control can be carried out. They have been discussed in chapter 2.

The second development concerns advances in receiver technology and processing software, see also [Blewitt, 1993]. Keywords are digital technology and automated data processing to handle the bulk of (sampled) data. A remarkable sophistication of the algorithms for the GPS data processing has been achieved in the area of ambiguity resolution. In order to obtain precise positioning results using carrier phase observations, the ambiguities have to be resolved. This resolution can be carried out very quickly, i.e. using data of only a little time span. In the limiting case, the ambiguities are resolved instantaneously, or On-The-Fly. This capability largely reduces the site occupation time as compared with the 2-3 hours static survey.

Thirdly, as opposed to static, the GPS receiver can be in motion. We will not be concerned with the forces causing the motion and therefore use the term kinematic, instead of dynamic. With the receiver moving during the survey, surveying productivity can be increased and also the receiver's trajectory can be determined (in a discrete sense: positions at specific instants in time). In this way, kinematic GPS surveying actually includes all GPS measurement techniques. The categories usually identified are, see [Blewitt, 1993] and [Seeber, 1993]:

- semi kinematic or stop-and-go: the roving receiver makes occasional stops at survey marks. Some initialization is needed to determine the ambiguities and after that, the receiver must keep tracking the signals, also during motion.
- true or pure kinematic: the receiver is in permanent motion (includes precise navigation)
- rapid static: the survey marks are quickly surveyed. The ambiguities are resolved using the data of the actual occupation. The receiver can be switched off during transportation

In the sequel we will not further use the above terminology. In summary, we use the adjective 'kinematic' to refer to the motion of the receiver during the survey. Kinematic GPS surveying therefore comprises the whole category of GPS measurement techniques in surveying, that provide precise coordinates in a short time span (possibly instantaneously). Later on in this chapter, we will develop a mathematical model for GPS surveying in general and not for one particular measurement technique.

Applications of kinematic GPS surveying are primarily traditional surveying tasks. Community needs for spatial, or geo information have to be assuaged. The purpose therefore usually is to obtain coordinates for mapping purpose, i.e. to gather geometric information of (new) topography in order to relate it to existing structures. Kinematic GPS surveying can be an efficient provider of geometric geo-information for Geographic Information Systems. [Seeber, 1993] mentions cadastral surveys, control surveys and engineering surveys (stake out, locating points and construction work). Note that surveying, or basically positioning, does not necessarily take place on land but also at sea (hydrography) and in the air. Kinematic GPS surveying can support aerial

photogrammetry by either being an efficient tool to establish ground control, or for the determination of the camera position at exposure, see e.g. [Ackermann, 1997]. In fact, this determination is applicable to (remote) sensing techniques in general (based on e.g. laser or radar).

On the other hand, the measurement technique can be used for navigation as well. In this precise navigation application, coordinates are to be obtained in real time, in order to determine ones position with respect to existing features, enabling the navigator to avoid obstacles and reach the destination. These applications are primarily marine and airborne. The precision requirements are usually less stringent as the size of the vehicle to be positioned may range from several tens to hundreds of meters. Integrity on the contrary is much more stringent as compared with surveying.

The applications of positioning by kinematic GPS mentioned here, are surveying and navigation. The purpose is to determine the coordinates of a (discrete) point (object) that is either stationary or moving.

3.2 Geodetic positioning with GPS: concept

The two fundamental observables for positioning with GPS are the code and the carrier phase. Treating the concepts of both types of measurements is the first step in setting up the mathematical model for positioning.

Precise positioning with GPS has been developed from radio interferometry. This implies relative positioning as opposed to single point positioning: two receivers simultaneously make observations. The unknowns in the observation equations are geometric parameters (the coordinate differences) and possibly bias parameters. As such, GPS surveying can be considered as an example of geometric satellite geodesy, [Seeber, 1993]. In this section we will concentrate on the geometric unknowns; the bias parameters are left out of consideration.

3.2.1 GPS observables

Both the code and phase observable are travel time, and thus range related. The signal, carrier plus modulations, leaves the satellite at time t^s and arrives at the receiver at time t_r, the time of observation. The distance from the satellite at t^s to the receiver at t_r is denoted by l_r^s. In order to clarify the measurement principles, error sources are not considered here, i.e. we assume coherent oscillators, running at nominal frequency and no propagation delays. Satellite and receiver clock are synchronized.

code

For the code observation, the distance from satellite to receiver is determined by measuring the travel time of the signal. This travel time is measured by correlation of the received code modulation with the local generated code modulation. The time shift required to match the codes, is the travel time τ_r^s.

(2.1) $$\tau_r^s = t_r - t^s$$

The code observation is a direct measurement of the difference in time [Hatch et al, 1994]. Multiplication of the travel time with the speed of light c, yields the distance, observed at time t_r.

$$l_r^s = c\tau_r^s \tag{2.2}$$

phase

The phase observation results from a phase comparison. The difference of the phase of the receiver carrier at time of observation t_r and the phase of the satellite carrier at time of signal transmission t^s reads

$$\psi_r^s(t_r) = \phi_r(t_r) - \phi^s(t^s) \tag{2.3}$$

Under the current assumptions and with zero phase at time zero, the phase of a carrier equals the frequency f multiplied by the time t: $\phi(t) = ft$. The phase of the carrier generated by the receiver equals $\phi_r(t_r) = ft_r$ and the carrier transmitted by the satellite $\phi^s(t^s) = f(t_r - \tau_r^s)$. Substitution in (2.3) then yields for the actual phase difference ψ_r^s

$$\lambda\,\psi_r^s(t_r) = c\tau_r^s \tag{2.4}$$

As a consequence of carrying out measurements on pure monotone carriers, one cycle can not be distinguished from another by the receiver. Only the fractional difference in phase is measured: $frac(\psi_r^s(t_r)) \in [0,1>$ cycle. The observed phase difference ϕ_r^s differs by a whole number of cycles N_r^s from the actual phase difference ψ_r^s. At time t_1 when the receiver locks onto the satellite signal, we have

$$\phi_r^s(t_1) = \psi_r^s(t_1) - int(\phi_r^s(t_1)) \tag{2.5}$$

with the integer $N_r^s = -int(\phi_r^s(t_1))$. Once the receiver has locked onto the satellite signal, it keeps track of zero passes in the phase difference and incorporates this in the phase observation ϕ_r^s (accumulated phase). One ambiguity is involved in a timeseries of phase observations of one receiver to one satellite. N_r^s represents the initial number of missing cycles. At any later time t_r we have

$$\phi_r^s(t_r) = \psi_r^s(t_r) + N_r^s \tag{2.6}$$

and with (2.4) and (2.6) we obtain for the phase observation at time t_r

$$l_r^s = c\tau_r^s = \lambda(\phi_r^s - N_r^s) \tag{2.7}$$

or

$$\lambda\phi_r^s = l_r^s + \lambda N_r^s \tag{2.8}$$

The phase observation is ambiguously related to the range l_r^s. The first time derivative of the carrier phase observation is a measure of the change in range (the ambiguity drops out). The difference in time of the phase observation is an integration of the frequency difference: generated frequency minus received frequency (integrated Doppler shift), see [Hatch et al, 1994].

$$\lambda\{\phi_r^s(t_2) - \phi_r^s(t_1)\} = \int_{t_1}^{t_2} \dot{l}_r^s(\sigma)d\sigma = l_r^s(t_2) - l_r^s(t_1) \tag{2.9}$$

3.2.2 Relative positioning

This section is a first attempt to develop the model $y=Ax$ for the measurement system of GPS geodesy. We will establish a relation between observations and unknown parameters, with the emphasis on the geometric parameters.

Relative positioning with GPS implies the reception of signals from the satellites and making observations at two (or more) sites (nearly) simultaneously. This concept has been developed from radio interferometry, [Counselman et al, 1981]. GPS satellites can be used for precise relative positioning, similar to the use of quasars for Very Long Baseline Interferometry, [Seeber, 1993].

In (relative) positioning with GPS, the received signal is compared with a signal generated by the receiver itself (correlation of the code modulation or comparison of the phase). The resulting observations are then brought together for processing; the principle of radio-interferometry with independent clocks [Brouwer, 1985]. A GPS receiver can be considered as an advanced interferometry recorder.

In its design (single point positioning), the basic observable of GPS is the code observation, obtained by correlation of received and generated modulation. At first, developing code correlating receivers was a military business because of the control and knowledge of these code modulations.

In the early eighties, the Jet Propulsion Laboratory developed the SERIES-receiver, Satellite Emission Range Inferred Earth Surveying, [MacDoran, 1979] and [MacDoran et al, 1985]. The received signals were recorded at two sites and physically brought together for the correlation process. The group delay was measured and this can be seen as (differential) codeless pseudo-ranging. As such, the receiver is also a predecessor of the cross correlating capability of nowadays receivers.

The Massachusetts Institute of Technology proposed the MITES-receiver, Miniature Interferometer Terminals for Earth Surveying. The emphasis was on measuring the phase delay of the carriers. We cite [Counselman et al, 1979]: 'The basic quantity is the interferometric phase - the difference between the phases of radio signals from a single satellite received at any two terminals'.

Texas Instruments then developed the first commercial receiver. The TI 4100 could make both code observations (delay of signal; code correlation) and phase observations (phase

comparison). For a full overview of the receiver development, the reader is referred to [Blewitt, 1993], [Seeber, 1993] and [Hofmann-Wellenhof et al, 1997].

We will now discuss both the code and phase observation in relation with the interferometric principle. The distance to the source of the (radio) signal, the satellite, is very large compared with the station separation, the baseline length b, see figure 2.1.

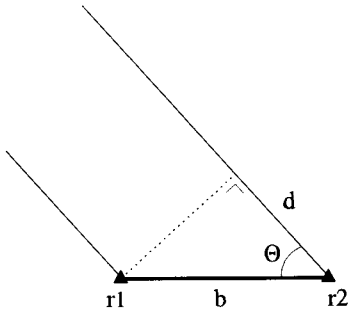


Figure 2.1: Principle of interferometry

By means of interferometry we have an observation that is related to the difference in distances from source to r2 and from source to r1: $d = l_2^s - l_1^s$. With

(2.10) $$d = b \cos\Theta$$

this observation can be related to the baseline length b we are interested in.

With the code observations (2.2) made at both sites r1 and r2, the difference in travel time can be determined:

(2.11) $$d = c(\tau_2^s - \tau_1^s) = c\,\Delta\tau$$

This difference in travel time is also referred to as group delay, [Seeber, 1993].

At both sites phase observations are made simultaneously. The received phase is measured with respect to a local oscillator. We take the difference of two equations (2.7):

(2.12) $$d = \lambda(\phi_2^s - \phi_1^s - N_2^s + N_1^s)$$

The term between brackets is referred to as phase delay, [ibid].

For the phase observation, we will discuss also the interferometric principle in terms of vectors in the three-dimensional space. Vector $\bar{b}$ is the baseline vector (here directed from r2 to r1) and vector $\bar{s}$ is the unit direction vector to the satellite, see figure 2.2.

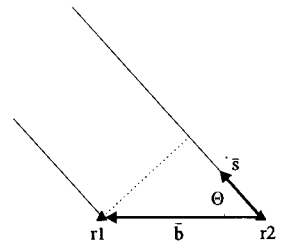


Figure 2.2: Principle of interferometry in terms of vectors

Equation (2.10) now becomes

(2.13) $$d = \|\bar{b}\|\cos\Theta = \bar{b}.\bar{s}$$

The distance d equals the inner product of $\bar{b}$ and $\bar{s}$. Combining (2.12) and (2.13) yields

(2.14) $$\bar{b}.\bar{s} = \lambda(\phi_2^s - \phi_1^s - N_2^s + N_1^s)$$

Apart from the ambiguity term, this equation for the phase observation is given in [Counselman et al, 1981]. The observation is also referred to as fringe phase. In the above equation, the receivers clock term - the difference between the phases of the independent oscillators - is absent because of our assumption of no error sources. Observation (2.14) is actually the single difference and it can be considered as an interferometric observable. The differencing of observations of two sites [Remondi, 1984], has its origin in interferometry.

The baseline vector $\bar{b}$ contains the coordinate differences in three dimensions. The estimation of these coordinates is dealt with in section 4. There we will elaborate on the GPS observation equations for positioning. We will explicitly use the ranges from the satellite to receiver r1 and r2 respectively and compute the position of the satellite, as compared with interferometry on quasars, the satellite is not infinitely far away from earth. The curvature of the wavefront must be taken into account, [Seeber, 1993].

3.3 Measurement configuration

In this section we will discuss different measurement configurations for positioning with GPS. We start with the observation equations without error sources of section 2 and consider the effect of introducing certain bias parameters - clock errors in particular - on the capability of estimating the geometric unknowns (in the three-dimensional Euclidean geometry).

Because of the biases, the observations are not true distance measurements anymore and will therefore be referred to as range-related observations. Both code and phase are range related, but we concentrate on the code observation. We will discuss single point positioning, the purpose of the GPS at first, and relative positioning, developed from interferometry.

single point positioning with distance observations
With three distance observations, like l_r^s of (2.2), the receiver position x_r can be determined in the three-dimensional space. The satellite positions are assumed to be known. The satellites act as (moving) connection or control points. When the satellite coordinates, determined with respect to some S-basis, are treated as stochastic quantities (with their variance covariance matrix propagated), the receiver position is estimated also with respect to this S-basis and a proper variance covariance matrix is obtained. Actually a resection with distances is made.

single point positioning with range-related observations
In practice we have to deal with bias parameters in the observation equations (propagation delays and clock errors). We will use p_r^s for the code observation (the observed distance l_r^s in (2.2)) and P_r^s for the phase observation ($\lambda\phi_r^s$ in (2.8)). The geometric range is denoted by ρ_r^s instead of l_r^s for the distance.

The discussion will be restricted to clock errors. The full observation equations are given in section 4. For a discussion of the error sources affecting the GPS observables, the reader is further referred to the textbooks [Seeber, 1993], [Leick, 1995], [Kleusberg et al, 1995], [Parkinson et al, 1996], [Bauer, 1997], [Hofmann-Wellenhof et al, 1997] and [Strang et al, 1997].

In addition to the three unknown receiver coordinates in x_r, we have to deal with the receiver clock error $\delta_r t$ (multiplied by the speed of light c).

$$E\{p_r^s\} = \rho_r^s + c\,\delta_r t \tag{3.1}$$

The receiver clock error causes that the distance measured in the code observation is not a true distance anymore and therefore observation (3.1) is referred to as pseudo-range observation.

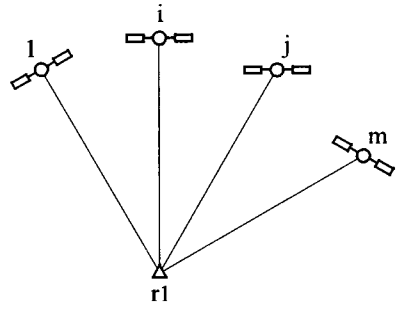


Figure 3.1: Single point positioning

The parameters can be determined with observations to at least 4 satellites. Satellite position and clock error, and atmospheric delays are treated as known quantities in the estimation and are therefore the cause of the accuracy being not better than several tens of meters. The full model of linearized observation equations is given by (3.2).

$$E\left\{\begin{pmatrix}\Delta p_r^1\\ \Delta p_r^i\\ \Delta p_r^j\\ \Delta p_r^m\end{pmatrix}\right\} = \begin{pmatrix}(-e_r^1)^T & 1\\ (-e_r^i)^T & 1\\ (-e_r^j)^T & 1\\ (-e_r^m)^T & 1\end{pmatrix}\begin{pmatrix}\Delta x_r\\ c\,\delta_r t\end{pmatrix} \tag{3.2}$$

In (3.1) the receiver coordinates x_r were contained in the range ρ_r^s, e_r^s is the unit direction vector from receiver to satellite. The linearization will be treated in detail for GPS positioning in section 4.

relative positioning with distance observations

When two receivers make distance observations l_r^s to three satellites, the following free network is obtained, figure 3.2.

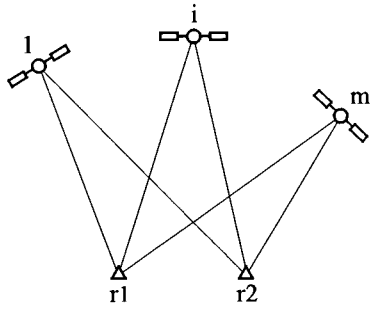


Figure 3.2: Relative positioning with distances

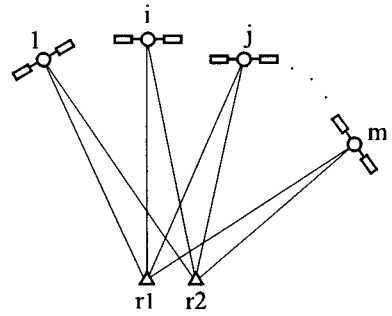


Figure 3.3: Relative positioning; small receiver interdistance

The scale is determined by c, the speed of light. For the translation (3 parameters) and rotation (3 parameters), the network must be connected to six satellite coordinates. They form the (local) S-basis of this network. With sufficient observations, the coordinates of the receivers and remaining satellites can be determined.

In practice, the broadcast satellite coordinates are used, which are nominally given in the WGS84 coordinate system. By this, the above network is linked to the reference frame of satellite coordinates and thus into the WGS84 coordinate system. It should be recognized again that the satellite coordinates are stochastic quantities.

relative positioning with range-related observations
We introduce a satellite related bias for each satellite and a receiver related bias for each receiver (per epoch and per observation type), for example a clock error δt.

(3.3) $$E\{p_r^s\} = \rho_r^s + c\,\delta_r t - c\,\delta^s t$$

Following the derivation in [Teunissen, 1990], we will show that the position of a receiver itself is hardly estimable, as the satellites are far away as compared with the distance between the receivers. We come to the concept of interferometry and estimate only coordinate differences. The effect of many biases on relative positioning is small or even negligible.

Two receivers simultaneously make code observations to m satellites. The linearized model of observation equations reads

(3.4) $$E\left\{\begin{pmatrix} \Delta p_1^1 \\ \vdots \\ \Delta p_1^m \\ \Delta p_2^1 \\ \vdots \\ \Delta p_2^m \end{pmatrix}\right\} = \begin{pmatrix} (-e_1^1)^T & & 1 & -1 & & \\ \vdots & & \vdots & & \ddots & \\ (-e_1^m)^T & & 1 & & & -1 \\ & (-e_2^1)^T & 1 & -1 & & \\ & \vdots & \vdots & & \ddots & \\ & (-e_2^m)^T & 1 & & & -1 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ c\,\delta_1 t \\ c\,\delta_2 t \\ c\,\delta^1 t \\ \vdots \\ c\,\delta^m t \end{pmatrix}$$

The satellite coordinates are assumed to be known. The following square $2m$x$2m$ and full rank transformation matrix is applied to the observables:

$$T = \begin{pmatrix} I_m & \\ -I_m & I_m \end{pmatrix}$$

The resulting model of observation equations becomes

$$E\left\{\begin{pmatrix} \Delta p_1^1 \\ \cdot \\ \cdot \\ \Delta p_1^m \\ \Delta p_{12}^1 \\ \cdot \\ \cdot \\ \Delta p_{12}^m \end{pmatrix}\right\} = \begin{pmatrix} (-e_1^1)^T & & 1 & -1 & & \\ \cdot & & \cdot & & \cdot & \\ \cdot & & \cdot & & & \cdot \\ (-e_1^m)^T & & 1 & & & -1 \\ (e_1^1)^T & (-e_2^1)^T & -1 & 1 & & \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & & \\ (e_1^m)^T & (-e_2^m)^T & -1 & 1 & & \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ c\,\delta_1 t \\ c\,\delta_2 t \\ c\,\delta^1 t \\ \cdot \\ \cdot \\ c\,\delta^m t \end{pmatrix} \tag{3.5}$$

The upper m observables can be left out, as they can be considered to determine only the satellite clock errors. The lower m observables are single differences. They are not related to the satellite clock errors. The geometric unknowns are reparametrized and split into the coordinates of r1 and coordinate differences

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} I_m & \\ I_m & I_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_{12} \end{pmatrix}$$

with the coordinate differences contained in $x_{12}=x_2-x_1$. The model becomes

$$E\left\{\begin{pmatrix} \Delta p_{12}^1 \\ \cdot \\ \cdot \\ \Delta p_{12}^m \end{pmatrix}\right\} = \begin{pmatrix} (e_1^1 - e_2^1)^T & (-e_2^1)^T & -1 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ (e_1^m - e_2^m)^T & (-e_2^m)^T & -1 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_{12} \\ c\,\delta_1 t \\ c\,\delta_2 t \end{pmatrix} \tag{3.6}$$

In GPS surveying, the receiver interdistance is small as compared with the distance to the satellite, figure 3.3. The unit direction vectors are defined as

$$e_1^s = \frac{x_1^s}{\|x_1^s\|} \qquad e_2^s = \frac{x_2^s}{\|x_2^s\|}$$

and $\|x_1^s\| \approx \|x_2^s\| \sim 2.10^7$ m. The difference in unit vectors can then be approximated by

$$(e_1^s - e_2^s) \approx \frac{x_{12}}{\|x_1^s\|}$$

and therefore $\|(e_1^s - e_2^s)\| \approx 10^{-3}$ for a 20 km receiver interdistance, whereas $\|e_2^s\| = 10^0$. In theory, with sufficient observations, all unknown parameters in model (3.6) can be estimated. In practice, x_1, the position of r1, is not estimable, due to numerical singularity. The coefficients for the coordinates in x_1 are about 1000 times smaller than the coefficients for the coordinate differences in x_{12}. A near, not a strict, rankdeficiency occurs. Note that in the estimation clock error $\delta_1 t$ is kept fixed. Clock error $\delta_2 t$ is estimated relative to $\delta_1 t$.

As compared with the coordinate differences, the coordinates of r1 are poorly estimable. Two implications are that first down on earth, an S-basis has to be chosen for the network of coordinate differences: 3 parameters are needed to provide the translation. The near rankdeficiency is treated as a strict one. In GPS surveying, we will estimate only coordinate differences, and keep the coordinates of r1 fixed: $\Delta x_1 = 0$, see also [Bock et al, 1985]. Secondly, in practice all satellite positions are treated as known constants. It may mean a serious overconstraining in a general network. In fact a pseudo least squares adjustment is carried out: $\Delta x^s = 0$. The satellites are moving connection points. The coordinates x^s are stochastic quantities, but in the data processing usually supposed to be non-stochastic: $Q_{x^s} = 0$. This is however, all greatly relaxed in relative positioning with a network as in figure 3.3.

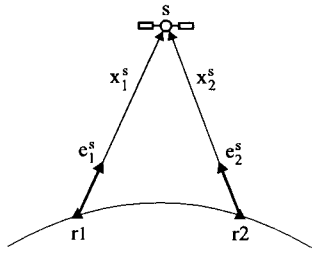


Figure 3.4: Unit direction vectors

In GPS surveying we will estimate the coordinates of the roving receiver in three dimensions, with respect to the coordinates of the reference receiver. As such, actually a small free network is obtained (of one or more baselines). Treating the satellite positions as known quantities provides the orientation of the network (3 parameters). The scale was already provided by the speed of light (1 parameter). The coordinates of the reference receiver (in WGS84) then constitute the S-basis for the free network; the translation (3 parameters). This completes the 7 parameters for the similarity transformation.

To conclude this section, we will consider the connection of the coordinate differences to the coordinates x_1 in more detail. The reparametrization of x_1 and x_2 into x_1 and x_{12} in (3.6) was carried out in order to explicitly show the coordinates of r1 and the coordinate differences. As the coordinates of r1 are (almost) not estimable in GPS surveying, they are kept fixed. The full model of observation equations is (3.7).

From model (3.5) the upper m equations were left out. The first three columns are denoted by A_1, the second three by A_2. The coefficients for the differential clock error are contained in A_3. Vector $\underline{y}$ contains the m observables of (3.6). The observables $\underline{x}_1$ may result from a single point solution. The variance covariance matrix is given by Q_{x_1}.

$$
\begin{aligned}
E\{\Delta \underline{y}\} &= A_1 \Delta x_1 + A_2 \Delta x_{12} + A_3 c\, \delta_{12} t \; ; \quad D\{\Delta \underline{y}\} = Q_y \\
E\{\Delta \underline{x}_1\} &= \Delta x_1 \qquad\qquad\qquad\qquad ; \quad D\{\Delta \underline{x}_1\} = Q_{x_1}
\end{aligned}
\tag{3.7}
$$

For model (3.7) it can be shown that for the limiting case of $A_1 = 0$, the following estimators are obtained

$$(3.8) \qquad \begin{pmatrix} \hat{\underline{x}}_1 \\ \hat{\underline{x}}_{12} \end{pmatrix} = \begin{pmatrix} \underline{x}_1 \\ \hat{\underline{x}}_{12} \end{pmatrix} \; ; \; D\{\begin{pmatrix} \hat{\underline{x}}_1 \\ \hat{\underline{x}}_{12} \end{pmatrix}\} = \begin{pmatrix} Q_{x_1} & \\ & Q_{\hat{x}_{12}} \end{pmatrix}$$

The estimator for the coordinate differences $\hat{\underline{x}}_{12}$ is determined by the observables $\underline{y}$. The coordinates x_1 are left untouched. Estimates $\hat{x}_1$ and $\hat{x}_{12}$ are computed as if the coordinates x_1 were deterministic quantities, $Q_{x_1}=0$. In practice, the reference receiver coordinate observables $\underline{x}_1$ will have a very limited effect on the estimators for the coordinate differences. The coordinates for r2 are obtained from (3.8) by

$$\begin{pmatrix} \hat{\underline{x}}_1 \\ \hat{\underline{x}}_2 \end{pmatrix} = \begin{pmatrix} \underline{x}_1 \\ \underline{x}_1 + \hat{\underline{x}}_{12} \end{pmatrix} \; ; \; D\{\begin{pmatrix} \hat{\underline{x}}_1 \\ \hat{\underline{x}}_2 \end{pmatrix}\} = \begin{pmatrix} Q_{x_1} & Q_{x_1} \\ Q_{x_1} & Q_{x_1} + Q_{\hat{x}_{12}} \end{pmatrix}$$

The connection of the network to existing pointfields on earth and the transformation from WGS84 to another datum are topics beyond the scope of this thesis.

3.4 Mathematical model

In this section a mathematical model will be developed for kinematic GPS surveying. In this application two receivers simultaneously make range-related observations to GPS satellites. One receiver is stationary, the reference receiver, the other may be stationary either moving, the roving receiver. The parameters that are of primary interest are the relative geometric unknowns: the baseline coordinates x_{12}, y_{12} and z_{12}.

First the observation equation per observable is treated. Next the full mathematical model for relative positioning over short distances is given. This model is then analysed on the estimability of the unknown parameters.

3.4.1 Observation equation

First the L1 phase observable will be treated. Then the other types of observations are considered. The modelling of the GPS observables is based on the Rinex convention, a generally accepted standard for GPS observation files, [Gurtner et al, 1990]. The observation equation is non-linear in the geometric unknowns and has to be linearized. The aberration effect is a complicating factor in this. The single observation part concludes with a brief diagnosis on the non-linearity.

3.4.1.1 Functional model phase observable

The full observation equation for the L1-phase observable $\underline{P}_r^s(t)$ follows by extension of (2.8) and reads

$$(4.1) \qquad E\{\underline{P}_r^s(t)\} = \rho_r^s(t,t-\tau) + c\,\delta_r t(t) - c\,\delta^s t(t-\tau) + T_r^s(t) - I_r^s(t) + D^s(t) + D_r(t) + A_r^s$$

where

ρ_r^s : geometric range from satellite to receiver
$\delta_r t$: receiver clock error
$\delta^s t$: satellite clock error
T_r^s : tropospheric delay
I_r^s : ionospheric delay
D^s : satellite hardware delay
D_r : receiver hardware delay
A_r^s : ambiguity

The observation is made by receiver r to satellite s at time t, which is expressed in GPS time. The observation is expressed in [m]. In equation (4.1) t is used as the time tag of the observation. With the definition of the traveltime τ, time t-τ becomes the time at which the signal must leave the satellite circuits to be in time at the receiver for the observation at time t.

Note that the traveltime τ is defined by the geometric range and the various delays and is therefore not an independent parameter:

$$\tau = \frac{1}{c}\left[\rho_r^s(t,t-\tau) + T_r^s(t) - I_r^s(t) + D^s(t) + D_r(t)\right] \tag{4.2}$$

By definition (4.2) the traveltime τ is thus not determined by only the geometric range ρ_r^s.

The geometric range $\rho_r^s(t,t-\tau)$ is a function of the time of observation t: both satellite and receiver are in general in permanent motion. The range is also a function of the traveltime τ: the aberration effect. The satellite will move with respect to the receiver during the the time interval the signal needs to travel from satellite to receiver. Both t and τ are not known.

The measured phase difference ϕ_r^s has been related to the traveltime τ. The carrier phase ambiguity A_r^s in equation (4.1) does not represent an integer number as it now includes initial phase offsets of receiver and satellite. The ambiguity is a constant as long as signal tracking is un-interrupted.

The inaccuracy with the time arguments of ρ_r^s, T_r^s, I_r^s and D^s in equation (4.1) implies only very small approximations ($<10^{-4}$m); e.g. $D^s(t)$ should be $D^s(t-\tau)$ and $x^s(t-\tau)$ should be $x^s(t-\tau+D^s/c)$. For the delays it is a matter of notation: estimates for these quantities are computed, independent of the notation and therefore the problem is not relevant. For the geometric range the effect cancels in relative positioning.

parameter	value [m]	remark
ρ	10^7	
$c\delta_r t$	10^5	$\delta_r t \sim 10^{-3}$ s *)
$c\delta^s t$	10^5	$\delta^s t < 10^{-3}$ s
T	10^0	
I	10^1	
D^s	10^0	several ns
D_r	10^{-1}	some ns
A	10^7	

*) can be several ms when continuously running clock is restored (Rinex conversion)

Table 4.1: Values for parameters in (4.1)

In table 4.1 indications are given for the magnitude of the parameters in observation equation (4.1). The magnitudes in the table represent ordinary to maximum values for these parameters. The minimum elevation is assumed to be 15°.

3.4.1.2 Stochastic model phase observable

In chapter 3 of [Parkinson et al, 1996] the accuracy of the L1-phase observable is stated to be 0.1 radians rms. This translates into $\sigma_P = 3$ mm, which is assumed to hold for normal operational conditions. The phase observable is assumed to possess a Gaussian distribution.

$$D\{\underline{P}_r^s(t)\} = \sigma_P^2$$

We assume here no time correlation (although the correlation length of the receiver noise could be several seconds, as internal filtering may take place in the receiver).

With the above assumption on the stochastics, it must be realized that there are several effects that are not explicitly captured by the current functional model (4.1). There are the propagation effects: multipath, atmospheric diffraction (signal path is not a straight line from satellite to receiver), relativistic delay (signal path is curved due to the earth's gravity field) and higher order ionospheric effects (ionospheric delay is assumed to be inversely proportional to the frequency, cf. (4.3)). Radio frequency interference of the GPS signal is not considered. The non-modelled effects may induce time correlation.

In further data processing we will use the broadcast ephemeris to compute the satellite position. A bias in this position is not accounted for. The geometric range is defined using the (electro-magnetic) phase centers. In order to extract useful geometric information, the range has to be related to some discrete point on the antenna. As the electro-magnetic phase center is difficult to define, the relation will also be problematic and additional biases may enter the estimation process. Interchannel biases of the receiver can be neglected for a nowadays digital receiver, see e.g. [Meehan et al, 1992].

Indications for the magnitudes of the effects above on the observation equation (4.1) are given in table 4.2. Maximum values are given; ordinary values for these effects are of the same order or smaller.

non-modelled effect	value [m]
satellite position	10^{1}
satellite antenna phase center	10^{-2}
multipath at satellite	10^{-2}
diffraction	10^{-2}
higher order ionospheric effects	10^{-2}
relativistic delay	10^{-2}
multipath at receiver	10^{-2}
receiver antenna phase center	10^{-2}

Table 4.2: Maximum values for unmodelled effects

3.4.1.3 Phase and code observables

Beside the phase observation on L1, also L2 phase observations can be made. Furthermore code observations can be made based on the L1 or L2 signal.

The full observation equation for the L2-phase observable $\bar{\underline{P}}_r^s(t)$ reads

$$(4.3) \quad E\{\bar{\underline{P}}_r^s(t)\} = \rho_r^s(t,t-\tau) + c\,\delta_r t(t) - c\,\delta^s t(t-\tau) + T_r^s(t) - \frac{f_1^2}{f_2^2} I_r^s(t) + \bar{D}^s(t) + \bar{D}_r(t) + \bar{A}_r^s$$

For the L1-code observable $\underline{p}_r^s(t)$ we have

$$E\{\underline{p}_r^s(t)\} = \rho_r^s(t,t-\tau) + c\,\delta_r t(t) - c\,\delta^s t(t-\tau) + T_r^s(t) + I_r^s(t) + d^s(t) + d_r(t) \tag{4.4}$$

and for the L2-code observable $\bar{\underline{p}}_r^s(t)$ the observation equation reads

$$E\{\bar{\underline{p}}_r^s(t)\} = \rho_r^s(t,t-\tau) + c\,\delta_r t(t) - c\,\delta^s t(t-\tau) + T_r^s(t) + \frac{f_1^2}{f_2^2} I_r^s(t) + \bar{d}^s(t) + \bar{d}_r(t) \tag{4.5}$$

In the equations above we made the assumption that the geometric range is identical for all observation types. Although we did not denote it, the traveltime is different for different observation types (due to the delays, see (4.2); think of 10^{-7} s effects). This affects the time arguments of ρ_r^s and $\delta^s t$. We are now faced with a contradiction as the range depends on t and τ. The effects involved are $< 10^{-4}$ m and will largely cancel in relative positioning. Furthermore, the electro-magnetic phase centers of the different observation types will not coincide. The geometric ranges will not be identical (mm-cm effects). These effects largely cancel in relative positioning.

The first order ionospheric delay is inversely proportional to the frequency of the carrier. Code and carrier are assumed to be equally (although oppositely) affected.

$$D\left\{\begin{pmatrix} \underline{P}_r^s \\ \bar{\underline{P}}_r^s \\ \underline{p}_r^s \\ \bar{\underline{p}}_r^s \end{pmatrix}\right\} = \begin{pmatrix} \sigma_P^2 & & & \\ & \sigma_{\bar{P}}^2 & & \\ & & \sigma_p^2 & \\ & & & \sigma_{\bar{p}}^2 \end{pmatrix} \tag{4.6}$$

The stochastic model of the observables made by one receiver to one satellite is assumed to be (4.6). The standard deviation for the phase observable will be several millimeters, for the code observable several decimeters.

Variance covariance matrix (4.6) is diagonal. We assume here that the four observables above are not mutually correlated, although in practice there may be for instance (frequency) cross correlation. We also assume here identical stochastics for all channels, no mutual correlation between the channels (satellites) of one receiver and between different receivers.

3.4.1.4 Linearization of observation equation

In this section we will linearize the observation equation (4.1), in order to be able to compute estimates for the geometric unknowns of interest using the BLUE principle, i.e. $\hat{x}_r = x_{r,o} + \Delta\hat{x}_r$. A complication in the linearization is the aberration effect. A straightforward linearization would be possible in case both the satellite and receiver do not move (in one or another coordinate system).

The observation equation (4.1) is linear in the satellite clock error, atmospheric and hardware delays and the ambiguity, but is non-linear in the satellite and receiver coordinates, and in the receiver clock error as it is involved in the geometric range, see (A.1). Time t_r is the time tag of the receiver in the observation file. The time of observation in GPS time reads $t = t_r - \delta_r t(t)$. The satellite position was assumed to be known. In order to compute estimates for the receiver position, equation (4.1) has to be linearized w.r.t. the receiver coordinates. We will thereby forget about the delays in equation (4.2), see also the implementation given in appendix A.

The linearized observation equation reads

$$\Delta P_r^s = \left.\frac{\partial \rho}{\partial x_r}\right|_o \Delta x_r + \left.\frac{\partial \rho}{\partial \delta_r t}\right|_o \Delta\delta_r t + c\Delta\delta_r t - c\delta^s t + T_r^s - I_r^s + D^s + D_r + A_r^s \tag{4.7}$$

and the partial derivatives are evaluated using the approximate values $x_{r,o}$ and $\delta_r t_o$.

We will neglect the second term on the right hand side of (4.7), i.e. the partial derivative of the geometric range w.r.t. the receiver clock error - the coefficient is $-\frac{\partial \rho}{\partial t}$ - see section 2.3 of [Jonge, 1993].

As actually $c\,\delta_r t(t) = c\,\delta_r t(t_r - \delta_r t(t))$, the partial derivative w.r.t. the receiver clock error yields coefficient c and an additional coefficient. The latter is relatively (factor 10^6) small and therefore neglected in this discussion. The clock error is then actually evaluated at t_r instead of t. The clock error appears only with coefficient c in (4.7).

With e_r^s the vector of unit length from x_r at time t to x^s at time $(t-\tau)$ we obtain

$$\Delta P_r^s = -e_r^{sT} \Delta x_r + c\Delta\delta_r t - c\delta^s t + T_r^s - I_r^s + D^s + D_r + A_r^s \tag{4.8}$$

The computed observation is obtained with the non-linear relation (4.1)

$$P_{r,o}^s = \rho_{r,o}^s + c\,\delta_r t_o \tag{4.9}$$

with $x_{r,o}$ and $\delta_r t_o$ as approximate values. The observed minus computed observation is $\Delta P_r^s = P_r^s - P_{r,o}^s$.

Once estimates for the (estimable) unknowns are obtained via (4.8), for example for x_r, they can be used as approximate values in a next iteration. The receiver clock error $\delta_r t$ may not be separately estimable, see section 3.4.4. In that case no improved numerical estimates are obtained.

The difficulties in the estimation process are the computation of the unit vector e_r^s and the computed observation $P_{r,o}^s$ as the aberration effect is present. Both can be computed with the procedure given in appendix A.

3.4.1.5 Non-linearity effect

In the previous section we have performed the linearization of the non-linear observation equation, without questioning its allowance. The linearized equation (4.8) was obtained. The estimates obtained (using the linear approximation) will be biased due to the non-linearity. The non-linearity effect will be briefly diagnosed.

Both code and phase observable are basically range observables. We use equation (2.8) and leave the ambiguity out.

$$P_r^s = l_r^s = \sqrt{(x^s - x_r)^2 + (y^s - y_r)^2 + (z^s - z_r)^2} \tag{4.10}$$

The linearized version, cf. (4.8), becomes

$$\Delta P_r^s = -e_r^{sT} \Delta x_r \tag{4.11}$$

The non-linear relation (4.10) is approximated by a Taylor expansion (4.11), in which the second and higher order terms are neglected. In [Teunissen, 1989] we find that the second order remainder $R_2(x)$ can be used as a first measure of the non-linearity effect. The second order term is the largest term that is neglected in using the linearized observation equation instead of the non-linear one. The measure $R_2(x)$ comprises both the intrinsic non-linearity and the non-linearity due to the parametrization.

The Taylor expansion of $E\{y\}=A(x)$ up to the second order term reads

$$E\{y\} = A(x_o) + \partial_x A(x)\big|_o \Delta x + \tfrac{1}{2}\Delta x^T \partial_{xx}^2 A(x)\big|_o \Delta x \tag{4.12}$$

$\partial_x A(x)\big|_o$ is the 1x3 Jacobian matrix and $\partial_{xx}^2 A(x)\big|_o$ the 3x3 Hessian matrix. For (4.10) the Hessian matrix becomes, see also [ibid],

$$\partial_{x_r x_r}^2 l_r^s = \frac{1}{l_r^{s\,3}} \begin{pmatrix} y_r^{s\,2} + z_r^{s\,2} & -x_r^s y_r^s & -x_r^s z_r^s \\ -x_r^s y_r^s & x_r^{s\,2} + z_r^{s\,2} & -y_r^s z_r^s \\ -x_r^s z_r^s & -y_r^s z_r^s & x_r^{s\,2} + y_r^{s\,2} \end{pmatrix} \tag{4.13}$$

with vector $x_r^s = x^s - x_r = (x_r^s, y_r^s, z_r^s)^T$. The eigenvalues of matrix (4.13) are

$$\begin{aligned} \lambda_{\min} &= 0 \\ \lambda_{\max} &= \frac{1}{l_r^s} \end{aligned} \tag{4.14}$$

and with (12) of [ibid] we obtain the following bounds for the size of the second order remainder

$$0 \le R_2(x) \le \frac{1}{2l_r^s}\|\Delta x\|^2 \tag{4.15}$$

with Δx the 3x1 vector true receiver position minus approximate (used in linearization) receiver position and thus $\|\Delta x\| = \sqrt{(x_r - x_{r,o})^2 + (y_r - y_{r,o})^2 - (z_r - z_{r,o})^2}$. With $l_r^s = 2.10^7$ m we obtain table 4.3.

If the approximate coordinates for receiver r are on the cm-level (0.1-0.01m), possibly after iteration, the second order remainder in the linearization of the observation equation is completely negligible ($< 10^{-9}$ m). The linearized observation equation is a sufficient approximation.

$\|\Delta x\|$ [m]	R_2 [m]
100	$0,25.10^{-3}$
10	$0,25.10^{-5}$
1	$0,25.10^{-7}$
0.1	$0,25.10^{-9}$
0.01	$0,25.10^{-11}$

Table 4.3: Second order remainder

We have seen that for the linearization, when the approximate values for the receiver coordinates are correct up to 100 m, the second order remainder will be below the 1 mm level. In [Lichten, 1990] it is stated that in order to avoid an iteration, receiver locations should be known to within a few hundred meters. In the next section however, we will see that 100 m errors are not allowed in the coordinates of the reference receiver, when these are to be constrained in relative positioning.

In section 5.4 of chapter 1, it was noted that with a non-linear model of observation equations, the estimator is biased: $E\{\hat{x} - x\} \neq 0$. For relative positioning with a short observation time span, the bias in the observables, which propagates into the estimator - see equation (40) of [Teunissen, 1989] - is at or below the 10^{-5} meter level, and thus well below the measurement noise (of the precise phase observables).

3.4.2 Functional model

The model of observation equations is set up for relative positioning over short distances: two receivers (nearly) simultaneously observe m satellites. The effects of many error sources on the *relative* geometric unknowns are significantly less as compared with single point positioning. The development of the model is shown in this section for L1 phase observations.

Both the satellite and the receiver r1 position are not treated as unknowns in relative positioning. The satellite position x^s is computed from the broadcast ephemeris. The receiver r1 position x_1 follows from a reference frame or a separate single point solution. In section 3.3 it was shown that these quantities are hardly estimable. When these parameters are kept fixed in the estimation (introduced as deterministic quantities), the effect of errors therein on the observation equation will be rather limited. The satellite clock error $\delta^s t$ will be estimated and therefore not computed from the clock parameters in the navigation message.

It is assumed that receivers of the same make and type are used together with identical antennas. The baseline vector x_{12} is defined from phase center to phase center and equals in this case the vector between the geometric centers of the antennas (mm-cm translation). The discussion on the definition and interpretation of the baseline is pursued in section 1 of chapter 4.

We will discuss the cancellation of errors in relative positioning. The short distance assumption therein is essential. Three categories of error sources are distinguished: satellite, signal propagation and receiver.

satellite
The signals to r1 and r2 nearly simultaneously ($\sim$ms) leave the satellite. The effects of satellite related error sources on the observations made by r1 and r2 will therefore be largely identical.

Similar to the derivation in section 3.3, one can give an indication for the maximum size of the term concerning the satellite coordinates in the observation equation:

$$(e_1^s - e_2^s)^T \Delta x^s \sim \frac{\|x_{12}\|}{\|x_1^s\|} \|\Delta x^s\| \tag{4.16}$$

In practice (4.16) is interpreted as that a satellite position error dr (true position - broadcast position) has a db effect on the baseline vector in worst case according to $db/b = dr/r$ with b is the baseline length and r the distance to the satellite [Wells et al, 1986]. As such a 20 meter error would have a 1 cm effect on a 10 km baseline, see also [Zieliński, 1989] and [Kuang et al, 1996]. The effect of the satellite position is long term (hours).

The stability of the satellite clock (even with SA on, Selective Availability see chapter 17 in [Parkinson et al, 1996]) is by far sufficient to safely bridge the small time span between departure of the two signals; the clock error $\delta^s t$ can be used in the observation equations of both r1 and r2.

Also the hardware delay is sufficiently stable, and D^s can be used in the observation equations of both r1 and r2.

Antenna phase center effects of the satellite can be neglected as the receiver separation is very small (compared with the distance to the satellite).

signal propagation
For small receiver separation, the signals from s to r1 and to r2 pass through the same part of the atmosphere.

We assume that the atmospheric delays are equal, i.e. $T_1^s = T_2^s$ and $I_1^s = I_2^s$. The remaining effects usually are in the order of several millimeters, but can reach the centimeter level. Under extreme circumstances, the differential ionospheric effect I_{12}^s can reach the 1 cm level for 1 km receiver separation. The temporal variation of differential ionospheric effects can be both short and long term. Differential tropospheric effects tend to be long term.

The diffraction effect largely cancels.

The relativistic delay cancels.

Multipath completely depends on local circumstances and does therefore not cancel in relative positioning. Multipath usually is a short term effect (seconds-minutes). Multipath is not modelled, section 3.4.1.2.

receiver
All m satellites are observed by the same receiver exactly simultaneously. The effects of receiver related error sources are equal for all observations made by one receiver.

The same receiver clock error $\delta_r t$ can be used in the observation equations.

One can also use the same receiver hardware delay D_r.

Antenna phase center effects will cancel in short distance relative positioning, cf. section 1 of chapter 4.

Concerning an error in the receiver position x_1, the magnitude of the effect is roughly the same as for the satellite position error, see also [Bock et al, 1985]. Note that single point positioning with CA-code provides typically 100 m positioning accuracy (SA on).

residual effects
In table 4.4 we give indications for the residual effects in relative positioning. The effects on the observation and the effect on the baseline estimate are assumed to be of the same order of magnitude. The figures given are maximum values; ordinary values are of the same order or smaller. Note that the figures given are related to the 10 km receiver interdistance and to a minimum elevation of 15°.

parameter	value [m]	remark
x^i	10^{-2}	
$c\delta^i t$	$<10^{-4}$	
D^i	$\ll 10^{-4}$	
phase center^i	-	
multipath^i	10^{-2}	for phase
diffraction	10^{-4}	
T	10^{-2}-10^{-3}	
I	10^{-2}-10^{-3}	
relativistic	10^{-5}	
phase center$_r$	-	
multipath$_r$	10^{-2}	for phase
x_r	10^{-2}	

Table 4.4: Residual effects in relative positioning

From table 4.4 we see that the largest residual effects are of the order 10^{-2} m. Taking the baseline length of 10 km into account, yields the 1 ppm positioning accuracy, which is frequently used as an indication in practice.

Apart from multipath, the most important error sources are the atmospheric delays and the satellite position. Modelling and estimation of atmospheric delays and satellite orbits is beyond the scope of this research. The reader is referred to e.g. [Blewitt, 1993].

observation equations
The assumptions and approximations are now realized in the observation equations below (time arguments are omitted). With two satellites i and j and two receivers r1 and r2 we obtain with the non-linear observation equation for the L1 phase observable (4.1):

$$E\{\underline{P}_1^i\} = \rho_1^i + c\,\delta_1 t - c\,\delta^i t + T_1^i - I_1^i + D^i + D_1 + A_1^i \tag{4.17}$$

$$E\{\underline{P}_1^j\} = \rho_1^j + c\,\delta_1 t - c\,\delta^j t + T_1^j - I_1^j + D^j + D_1 + A_1^j \tag{4.18}$$

$$E\{\underline{P}_2^i\} = \rho_2^i + c\delta_2 t - c\delta^i t + T_1^i - I_1^i + D^i + D_2 + A_2^i \tag{4.19}$$

$$E\{\underline{P}_2^j\} = \rho_2^j + c\delta_2 t - c\delta^j t + T_1^j - I_1^j + D^j + D_2 + A_2^j \tag{4.20}$$

The linearized versions read

$$E\{\Delta\underline{P}_1^i\} = -e_1^{iT}\Delta x_1 + c\Delta\delta_1 t - c\delta^i t + T_1^i - I_1^i + D^i + D_1 + A_1^i \tag{4.21}$$

$$E\{\Delta\underline{P}_1^j\} = -e_1^{jT}\Delta x_1 + c\Delta\delta_1 t - c\delta^j t + T_1^j - I_1^j + D^j + D_1 + A_1^j \tag{4.22}$$

$$E\{\Delta\underline{P}_2^i\} = -e_2^{iT}\Delta x_2 + c\Delta\delta_2 t - c\delta^i t + T_1^i - I_1^i + D^i + D_2 + A_2^i \tag{4.23}$$

$$E\{\Delta\underline{P}_2^j\} = -e_2^{jT}\Delta x_2 + c\Delta\delta_2 t - c\delta^j t + T_1^j - I_1^j + D^j + D_2 + A_2^j \tag{4.24}$$

In relative positioning we keep x_1 fixed: $\Delta x_1 = 0$. The coordinates of r2 are estimated with respect to x_1. To emphasize the relative character instead of the absolute, the estimation result can be presented as $\hat{x}_{12} = \hat{x}_2 - x_1$, the estimate for the baseline vector. The observation equations for $\underline{\bar{P}}_r^s$, $\underline{p}_r^s$ and $\underline{\bar{p}}_r^s$ follow in a similar way from (4.3) to (4.5).

mathematical model

Below we will give the full model of observation equations for the case with two receivers observing m satellites for $k=1$ epoch. First the clock errors of both satellite and receiver are redefined. The receiver clock errors (for both r1 and r2) include the hardware delays and the satellite clock errors include the hardware delays and the atmospheric delays, corresponding to the observation type.

$$\begin{array}{ll}
c\delta_r t := c\delta_r t + D_r & c\delta^s t := c\delta^s t - D^s - T_1^s + I_1^s \\
c\bar{\delta}_r t := c\delta_r t + \bar{D}_r & c\bar{\delta}^s t := c\delta^s t - \bar{D}^s - T_1^s + \frac{f_1^2}{f_2^2} I_1^s \\
c\partial_r t := c\delta_r t + d_r & c\partial^s t := c\delta^s t - d^s - T_1^s - I_1^s \\
c\bar{\partial}_r t := c\delta_r t + \bar{d}_r & c\bar{\partial}^s t := c\delta^s t - \bar{d}^s - T_1^s - \frac{f_1^2}{f_2^2} I_1^s \\
\text{for both r1 and r2} & \text{for all } m \text{ satellites}
\end{array} \tag{4.25}$$

The model, which is quite similar to (3.4), then becomes (4.26). The estimability of the parameters in model (4.26) is treated in section 3.4.4.

3.4.3 Stochastic model

The stochastic model was discussed in section 3.4.1. For the moment this model is left unchanged, but two remarks are in order for its application to relative positioning over short distances.

$$(4.26)\qquad E\left\{\begin{pmatrix}\Delta\underline{P}_1^1\\ \cdot\\ \cdot\\ \Delta\underline{P}_1^m\\ \Delta\underline{P}_2^1\\ \cdot\\ \cdot\\ \Delta\underline{P}_2^m\end{pmatrix}\right\} = \begin{pmatrix} & 1 & -1 & & 1 & & & \\ & \cdot & & \cdot & & \cdot & & \\ & \cdot & & & \cdot & & \cdot & \\ & 1 & & & -1 & & 1 & \\ (-e_2^1)^T & & 1 & -1 & & & & 1 \\ \cdot & & \cdot & & \cdot & & & \cdot \\ \cdot & & \cdot & & & \cdot & & \cdot \\ (-e_2^m)^T & & 1 & & -1 & & & 1 \end{pmatrix} \begin{pmatrix}\Delta x_2\\ c\Delta\delta_1 t\\ c\Delta\delta_2 t\\ c\delta^1 t\\ \cdot\\ \cdot\\ c\delta^m t\\ A_1^1\\ \cdot\\ \cdot\\ A_1^m\\ A_2^1\\ \cdot\\ \cdot\\ A_2^m\end{pmatrix}$$

Although the effects of satellite positions and reference receiver position on the estimators will be rather limited, one should realize that they are stochastic quantities. Q_{x^s} and Q_{x_1} should be taken into account in the data processing.

Differential atmospheric delays are not accounted for in the functional model. We cite [Bock et al, 1986]: 'The residual errors of both propagation medium and orbital origin are approximately proportional to the distance between stations, at least for distances of up to several tens of kilometers. Therefore these errors are dependent on baseline distance'. In [ibid], equation (23), it is proposed to account for the baseline length dependence of these unmodelled effects in the dispersion of the observables. These effects may also induce time correlation in the observables with a correlation length of several minutes.

$$(4.27)\qquad D\left\{\begin{pmatrix}\Delta\underline{P}_1^1\\ \cdot\\ \cdot\\ \Delta\underline{P}_1^m\\ \Delta\underline{P}_2^1\\ \cdot\\ \cdot\\ \Delta\underline{P}_2^m\end{pmatrix}\right\} = \begin{pmatrix}\sigma_P^2 & & & & & & \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & \sigma_P^2 & & & \\ & & & & \sigma_P^2 & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & & \sigma_P^2\end{pmatrix}$$

The dispersion of the observables in (4.26) is given by (4.27). The stochastic model is identical for the observables of both receivers r1 and r2.

Equations (4.26) and (4.27) represent the full mathematical model for one epoch of the time varying measurement system. They are of the form $E\{\underline{y}_i\}=A_i x_i$; $D\{\underline{y}_i\}=Q_{y_i}$ (cf. (1.3) in chapter 2). No pseudo-observables are involved yet.

3.4.4 Estimability of parameters

The model of observation equations has been set up in a straightforward manner. A direct implementation of this model is not possible; it turns out to be rankdeficient. The full set of parameters in the vector of unknowns can not be estimated (cf. section 2 of chapter 1). In appendix B the functional model for kinematic GPS surveying is analysed on estimability. Reparametrizations of the unknowns are proposed to remove the rankdeficiencies. The coordinates x_2 are not involved in the S-basis; they are directly estimable, once the considerations of section 3.3 have been taken into account.

In appendix B five measurement scenarios are considered:

C1: code, single frequency
L1: phase, single frequency
L1&L2: phase, dual frequency
C1+L1: phase and code, single frequency
L1&L2+C1&P2: phase and code, dual frequency

Both set ups, with two stationary receivers and with the second receiver in motion, are treated. Below we repeat the conclusions of the analysis.

For all scenarios it holds true that the clock error (per epoch) of the reference receiver can not be (separately) estimated, see also [Blewitt, 1993]: only relative clock behaviour (between receivers) is estimable. Carrier phase ambiguities of the reference receiver are not estimable, neither are those of the second receiver to the first satellite.

If one receiver makes a certain type of observation to a satellite and the other receiver does not, the observations are free observations: they can be left out from the data processing. Only the overlapping part in the data needs to be considered: with the data of one receiver, under current S-bases, one can estimate the satellite clock errors, for each observation type, for the particular epochs.

When both receivers are stationary at least two epochs of data are required ($k\geq 2$), with (single or dual frequency) phase data. As soon as also code data are available, one epoch suffices ($k\geq 1$). In all five scenarios, less satellites ($m<4$) may suffice when more epochs of data are available. The absolute bottom is $m=2$.

When the second receiver is moving, we have 3 geometric coordinate unknowns per epoch instead of 3 for the whole survey. With code observations, at least 4 satellites are required and at least 7 satellites when only phase observations are available. With only phase data, single or dual frequency, at least two epochs of data are needed for this ($k\geq 2$). When code data are available, one epoch suffices ($k\geq 1$). Less than 4 satellites is not possible, even when more epochs of data are available.

To conclude this section we consider mathematical model (4.26), but now with the parameters that constitute the S-basis explicitly left out. Beside the coordinates of r2, the

clock error of r2, the satellite clock errors and (m-1) ambiguities remain. The design matrix is of full rank when at least two epochs of data are considered ($k \geq 2$).

$$(4.28) \qquad E\left\{\begin{pmatrix} \Delta P_1^1 \\ \cdot \\ \cdot \\ \Delta P_1^m \\ \Delta P_2^1 \\ \cdot \\ \cdot \\ \Delta P_2^m \end{pmatrix}\right\} = \begin{pmatrix} & & -1 & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & -1 & \\ (-e_2^1)^T & 1 & -1 & & & & 0 \\ \cdot & \cdot & & \cdot & & & 1 \\ \cdot & \cdot & & & \cdot & & \cdot \\ (-e_2^m)^T & 1 & & & & -1 & 1 \end{pmatrix} \begin{pmatrix} \Delta x_2 \\ c\Delta\delta_2 t \\ c\delta^1 t \\ \cdot \\ \cdot \\ c\delta^m t \\ A_2^2 \\ \cdot \\ A_2^m \end{pmatrix}$$

The ambiguity A_r^s of an undifferenced carrier phase observation, is not an integer number, or better, an integer number of cycles is involved, but it can not be separated from the initial phase offset of the satellite and of the receiver. In the estimable functions $A_2^2 - A_2^1 - A_1^2 + A_1^1$ to $A_2^m - A_2^1 - A_1^m + A_1^1$, see appendix B, the initial phase offsets (of each time two receivers and two satellites) cancel. As such, these parameters represent integer numbers of cycles $\lambda_1 N$, with N integer. They are double difference combinations of ambiguities.

The stochastic model (4.27) is left unchanged. Estimability concerns the unknown parameters, not the observables.

3.5 Model with double differences

For the kinematic GPS surveying application with two receivers, the mathematical model can be reformulated to allow another implementation of the data processing. The model with so-called undifferenced (L1-phase) observables was given by (4.26) and (4.27). The alternative is the model with double difference observables. We will first derive the full mathematical model with double difference observables and then give the considerations in chosing one or the other.

3.5.1 Observation equation

The double difference combination of observations has its origin in interferometry [Counselman et al, 1972]. The double difference is a linear combination of four undifferenced observations, made by two receivers to two satellites, figure 5.1.

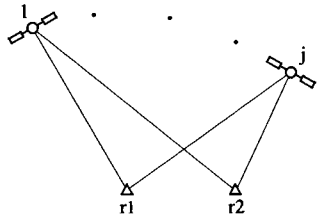


Figure 5.1: Double difference

$$(5.1) \qquad P_{12}^{1j} = P_2^j - P_2^1 - P_1^j + P_1^1$$

With the linearized observation equations for the L1-phase observable in (4.26) we obtain:

$$E\{\Delta\underline{P}_{12}^{1j}\} = (e_2^1)^T\Delta x_2 - (e_2^j)^T\Delta x_2 + A_2^j - A_2^1 - A_1^j + A_1^1 \tag{5.2}$$

or

$$E\{\Delta\underline{P}_{12}^{1j}\} = -(e_2^j - e_2^1)^T\Delta x_2 + A_{12}^{1j} \tag{5.3}$$

Under the assumptions and approximations of section 4, only coordinate unknowns and ambiguities are present in the double difference observation equation. The linear combination $A_{12}^{1j} = A_2^j - A_2^1 - A_1^j + A_1^1$ represents an integer number of cycles, as the initial phase offsets of satellite and receiver oscillators cancel, $A_{12}^{1j} = \lambda_1 N_{12}^{1j}$. When the receivers are stationary, none of the parameters depends on time.

3.5.2 Functional model

Consider model (4.26) for the case with two receivers observing the m satellites 1, 2,..,m. Per epoch we apply the following square $2m$x$2m$ and full rank transformation to the observables:

$$T = \begin{pmatrix} I_m & \\ -I_m & I_m \end{pmatrix}$$

and obtain

$$E\left\{\begin{pmatrix} \Delta\underline{P}_1^1 \\ \vdots \\ \Delta\underline{P}_1^m \\ \Delta\underline{P}_{12}^1 \\ \vdots \\ \Delta\underline{P}_{12}^m \end{pmatrix}\right\} = \begin{pmatrix} & 1 & & -1 & & & 1 & & & \\ & \cdot & & & \cdot & & & \cdot & & \\ & \cdot & & & & \cdot & & & \cdot & \\ & 1 & & & & -1 & & & 1 & \\ (-e_2^1)^T & -1 & 1 & & & & -1 & & & 1 \\ \cdot & \cdot & \cdot & & & & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & & & & & \cdot & \\ (-e_2^m)^T & -1 & 1 & & & & & -1 & & 1 \end{pmatrix} \begin{pmatrix} \Delta x_2 \\ c\Delta\delta_1 t \\ c\Delta\delta_2 t \\ c\delta^1 t \\ \cdot \\ \cdot \\ c\delta^m t \\ A_1^1 \\ \cdot \\ \cdot \\ A_1^m \\ A_2^1 \\ \cdot \\ \cdot \\ A_2^m \end{pmatrix} \tag{5.4}$$

The lower m observables are the single differences. The observables of r2 are differenced with respect to those of r1. In this respect, receiver r1 is the pivot or reference receiver. The upper m undifferenced observables are free observables. Each observable is concerned with its own unknown parameter, namely the satellite clock error $\delta^j t$. These

observables will be left out of further consideration, together with the corresponding unknowns.

With the following reparametrization of the unknown ambiguities, $s=1..m$

$$\begin{pmatrix} A_1^s \\ A_{12}^s \end{pmatrix} = \begin{pmatrix} I_m & \\ -I_m & I_m \end{pmatrix} \begin{pmatrix} A_1^s \\ A_2^s \end{pmatrix}$$

we obtain the model of single difference observation equations:

$$E\left\{\begin{pmatrix} \Delta\underline{P}_{12}^1 \\ \cdot \\ \cdot \\ \Delta\underline{P}_{12}^m \end{pmatrix}\right\} = \begin{pmatrix} (-e_2^1)^T & -1 & 1 & 1 & & \\ \cdot & \cdot & \cdot & & \cdot & \\ \cdot & \cdot & \cdot & & & \cdot \\ (-e_2^m)^T & -1 & 1 & & & 1 \end{pmatrix} \begin{pmatrix} \Delta x_2 \\ c\Delta\delta_1 t \\ c\Delta\delta_2 t \\ A_{12}^1 \\ \cdot \\ \cdot \\ A_{12}^m \end{pmatrix} \tag{5.5}$$

Per epoch, we had $2m$ observables. They have been transformed into m single difference observables, i.e. one per satellite. The clock error of the reference receiver $\delta_1 t$ is not estimable. Only the clock error of r2 (with respect to r1) can be estimated, see appendix B. Besides the designmatrix of (5.5) is rankdeficient by one. The single difference ambiguities interfere with the receiver clock error $\delta_2 t$.

Next, the square mxm and full rank transformation matrix is applied to the single difference observables in (5.5)

$$T = \begin{pmatrix} 1 & & \\ -1 & 1 & \\ \cdot & & \cdot \\ \cdot & & \cdot \\ -1 & & 1 \end{pmatrix}$$

and we obtain

$$E\left\{\begin{pmatrix} \Delta\underline{P}_{12}^1 \\ \Delta\underline{P}_{12}^{12} \\ \cdot \\ \cdot \\ \Delta\underline{P}_{12}^{1m} \end{pmatrix}\right\} = \begin{pmatrix} (-e_2^1)^T & -1 & 1 & 1 & & \\ -(e_2^2 - e_2^1)^T & & & -1 & 1 & \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & \cdot & & \cdot \\ -(e_2^m - e_2^1)^T & & & -1 & & 1 \end{pmatrix} \begin{pmatrix} \Delta x_2 \\ c\Delta\delta_1 t \\ c\Delta\delta_2 t \\ A_{12}^1 \\ \cdot \\ \cdot \\ A_{12}^m \end{pmatrix} \tag{5.6}$$

Satellite 1 is taken as a pivot, or reference satellite. It is a common approach to form double difference combinations using a pivot. There exist other ways, but as long as invertible linear transformations are used (as is the one here), one set of double differences can be transformed into the other, [Teunissen, 1995]. The models will be mathematically equivalent. From the observations made by two receivers to m satellites, $2m$ in total, only (m-1) double difference combinations can be taken that are linearly independent.

We observe in (5.6) that the remaining single difference observable $\underline{P}_{12}^{1}$ is a free observable, i.e. it determines the receiver clock error $\delta_2 t$. This observable, together with the receiver clock error, is left out of further consideration. The single difference ambiguity unknowns are reparametrized according to

$$\begin{pmatrix} A_{12}^{1} \\ A_{12}^{12} \\ . \\ . \\ A_{12}^{1m} \end{pmatrix} = \begin{pmatrix} 1 & & \\ -1 & 1 & \\ . & & . \\ . & & . \\ -1 & & 1 \end{pmatrix} \begin{pmatrix} A_{12}^{1} \\ A_{12}^{2} \\ . \\ . \\ A_{12}^{m} \end{pmatrix}$$

We finally obtain the model of double difference observation equations for L1-phase observables to m satellites for one epoch (k=1). It contains (m-1) double difference observables.

$$(5.7) \qquad E\left\{ \begin{pmatrix} \Delta \underline{P}_{12}^{12} \\ . \\ . \\ \Delta \underline{P}_{12}^{1m} \end{pmatrix} \right\} = \begin{pmatrix} -(e_2^2 - e_2^1)^T & 1 & \\ . & & . \\ . & & . \\ -(e_2^m - e_2^1)^T & & 1 \end{pmatrix} \begin{pmatrix} \Delta x_2 \\ A_{12}^{12} \\ . \\ . \\ A_{12}^{1m} \end{pmatrix}$$

This model is well known and widely used, see also [Seeber, 1993] and [Hofmann-Wellenhof et al, 1997].

If m satellites are continuously tracked for k epochs, we have $k(m\text{-}1)$ double difference observables. The vector of unknowns contains three baseline coordinates (stationary receivers) and (m-1) double difference ambiguities. These parameters are all estimable ($k \geq 2$) and the redundancy equals $k(m\text{-}1)\text{-}m\text{-}2$, compare to appendix B.

3.5.3 Stochastic model

Application of the variance covariance propagation law to diagonal matrix (4.27) yields the variance covariance matrix of the double difference observables. The variance of the (undifferenced) phase observable σ_P^2 has been taken outside the matrix. If the phase observables are not time correlated, as is assumed for (4.27), the double difference observables are not time correlated either.

$$D\left\{\begin{pmatrix} \Delta P_{12}^{12} \\ \cdot \\ \cdot \\ \Delta P_{12}^{1m} \end{pmatrix}\right\} = \sigma_P^2 \begin{pmatrix} 4 & 2 & . & . & 2 \\ 2 & . & . & & . \\ . & . & . & . & . \\ . & & & . & 2 \\ 2 & . & . & 2 & 4 \end{pmatrix} \tag{5.8}$$

The equations (5.7) and (5.8) represent the full mathematical model for one epoch of the time varying measurement system. The model is mathematically equivalent to the mathematical model given by (4.26) and (4.27). No information has been lost in going from (4.26) to (5.7). We have applied square and full rank transformation matrices and left out only free observables.

3.5.4 Argumentation for double differencing

In this argumentation, frequent references are made to the discussion in [Blewitt, 1993]. Under the topic 'information content', we find in [ibid] that both models are equivalent, when a clock error is to be estimated per epoch per observation type, as is the case in section 3.4 with clock errors δit, $\underline{\delta} it$, ∂it and $\underline{\partial} it$. This is the so-called white noise clock estimation. The statement holds true for relative positioning on local and regional scale, when the issues of correlation and data selection are rigorously dealt with.

In the double difference implementation, the problem of estimability of certain parameters, such as ambiguities and clock errors, is avoided (section 3.4.4). The S-basis is implicitly chosen by taking the double difference combinations. This is done prior to the data processing.

Two additional remarks are in order concerning the clock error parameters. In the undifferenced approach estimates for the clock errors are explicitly available. These estimates are usually not needed. Clock errors are time varying parameters and as such their behaviour in time may be modelled (dynamic model; pseudo-observations). In the double difference approach, this modelling is not possible, as the clock error parameters have been explicitly eliminated.

According to [ibid], the double difference implementation is computationally more efficient. As clock parameters have been explicitly removed, the system to solve is smaller: less unknown parameters remain and the number of observations has been reduced by the same number (data compression). Note that matrix Q_y is diagonal in (4.27), whereas (5.8) is a full matrix (per epoch). Double difference observables are not time correlated, as long as the undifferenced observables are not time correlated.

Differencing for the case with two receivers and m satellites is rather trivial. The set of undifferenced observations have to be transformed into a set of linearly independent combinations of undifferenced observations. The transformation matrix must be of full rank. For a fully automatic procedure for generating a set of independent double differences, see [Goad et al, 1988]. [Blewitt, 1989] checks the dependence by Gram-Schmidt orthogonalization.

The above deals with the number of double differences that are to be obtained. The actual set of double difference observations (or ambiguities) is not relevant. It is stated once

more that one (valid) set can be transformed into another (valid) set and back, see [Teunissen, 1995].

3.6 Related mathematical models

In the previous sections a mathematical model was developed for the processing of data obtained with kinematic GPS surveying. To conclude this chapter, we will briefly review two models for processing GPS data that are related to the model developed. The first one concerns a parametrization in terms of (geometric) ranges instead of (baseline) coordinates, the second one is the (basic) model for Differential GPS.

3.6.1 Parametrization in ranges

As discussed in section 4, the observation equations are non-linear in the receiver coordinates. The data, either undifferenced, section 4, or double differenced, section 5, are processed with a linearized version of these equations. In this separate section, we will discuss a stepwise adjustment of the double differenced observables. In the first step a *linear* observation model is used. In the second step, to arrive at estimates for the coordinate unknowns, again a non-linear model is employed.

stepwise adjustment
The linear observation equations for the L1-phase double difference, cf. (5.3), reads

$$E\{\underline{P}_{12}^{1j}\} = \rho_{12}^{1j} + A_{12}^{1j} \tag{6.1}$$

Two unknown parameters are involved: the double differenced geometric range ρ_{12}^{1j} and the double difference ambiguity A_{12}^{1j}. As at least the satellites are in permanent motion, the range ρ_{12}^{1j} depends on time (the receivers are allowed to move as well).

For two receivers, observing m satellites, the model of observation equations for one epoch, cf. (5.7), reads

$$E\left\{\begin{pmatrix} \underline{P}_{12}^{12} \\ \cdot \\ \cdot \\ \underline{P}_{12}^{1m} \end{pmatrix}\right\} = \begin{pmatrix} 1 & & & 1 & & \\ & \cdot & & & \cdot & \\ & & \cdot & & & \cdot \\ & & 1 & & & 1 \end{pmatrix} \begin{pmatrix} \rho_{12}^{12} \\ \cdot \\ \cdot \\ \rho_{12}^{1m} \\ A_{12}^{12} \\ \cdot \\ \cdot \\ A_{12}^{1m} \end{pmatrix} \tag{6.2}$$

This model is simple as it is linear and no coordinates are involved (no satellite ephemeris needed); it is completely expressed in ranges. Under the condition that the model is observable (designmatrix full rank), estimates for the unknown parameters can be computed. With only code observations on one frequency (and no phase observations), the first step is trivial: $\hat{\underline{\rho}}_{12}^{12} = \underline{p}_{12}^{12}, .., \hat{\underline{\rho}}_{12}^{1m} = \underline{p}_{12}^{1m}$.

Next, based on the estimators $\underline{\hat{\rho}}_{12}^{12},..,\underline{\hat{\rho}}_{12}^{1m}$, the second step can be carried out and estimators for the coordinates x_2 are obtained. Only geometric parameters are involved. The non-linear model of observation equations reads

$$E\left\{\begin{pmatrix}\Delta\underline{\hat{\rho}}_{12}^{12}\\ \cdot\\ \cdot\\ \Delta\underline{\hat{\rho}}_{12}^{1m}\end{pmatrix}\right\} = \begin{pmatrix}-(e_2^2-e_2^1)^T\\ \cdot\\ \cdot\\ -(e_2^m-e_2^1)^T\end{pmatrix}(\Delta x_2) \tag{6.3}$$

The receiver-satellite geometry, by means of the unit-direction vectors, is present now in the designmatrix.

In surveying practice, this stepwise approach is of limited value. The approach will not work with only phase observations; code observations on at least one frequency are needed (for the first step). Per epoch, there are (m-1) unknown range parameters, no matter whether the receiver is stationary or moving. For $m>4$, the redundancy in the first step is less than in (5.7). The model is less stringent and this will affect the model validation (detection of outliers and slips in the observations) and the capabilities of resolution of the integer ambiguities.

References concerning the parametrization in terms of ranges are [Melbourne, 1985], [Euler et al, 1991] and [Goad, 1992]. The above model is studied in [Teunissen, 1996] and [Jonkman, 1997] in particular with respect to resolution of the ambiguities. Based on the estimates $\hat{A}_{12}^{12},..,\hat{A}_{12}^{1m}$, integer least-squares estimates can be computed for the double difference ambiguities $N_{12}^{12},..,N_{12}^{1m}$.

The connection between this approach, parametrization in ranges, and the common parametrization in terms of coordinates is systematically elaborated upon in [Teunissen, 1997a]. The two approaches are linked through a particular form of a phased adjustment. The second step (6.3) is a consequence of including the geometric constraints. All (m-1) ranges are then related to the 3 coordinates of the receiving antenna. With m=4 satellites (one epoch), there is no redundancy in this step.

corrections for atmospheric delays

The atmospheric delays play the major role in limiting the baseline length (to 10 km) under the current modelling. Accounting for the ionospheric and tropospheric delay may extend the operational baseline length to several tens of kilometers.

The code and phase observations are corrected a-priori for the ionospheric I_r^s and tropospheric delay T_r^s. The (non-linear) observation equation for the L1-phase, in terms of a single difference, reads

$$E\{\underline{P}_{12}^s - T_{12}^s + I_{12}^s\} = \rho_{12}^s + c\delta_{12}t + A_{12}^s \tag{6.4}$$

Linearization of the model yields again (5.5).

Sample values for the corrections can be provided by an external model or by an Active GPS Reference System (AGRS) [Marel, 1997]; an AGRS is an active infrastructure for geodetic GPS users, which on the hardware side primarily consists of a network of permanent GPS stations. If the corrections are precise and can be considered non-stochastic, the variance covariance matrix of the observables is left unchanged, see also [Odijk, 1997] and [Teunissen, 1997b]. The assumptions in section 4.2 correspond in addition to taking sample values zero.

3.6.2 Differential GPS

The model of observation equations for relative positioning (over short distances, 10 km) with single frequency code observations from one epoch of data, contains $2m$ observations (m satellites are observed by both receivers) and $3+m+1$ unknown parameters, namely the three coordinates x_2, the m satellite clock errors $c\partial^i t$ $(i=1,..,m)$ and the $c\partial_2 t$ differential receiver clock error, see model (4.28) in section 4, but with the carrier phase ambiguities left out. The rankdeficiency concerning the receiver clock error $\partial_1 t$ is discussed in appendix B. All code observations at receiver 1 are related to this clock error and hence all corrections will be. The clock error of the rover $\partial_2 t$ accounts for this; it is part of the S-basis.

$$E\left\{\begin{pmatrix}\Delta p^1_{12}\\ \cdot\\ \cdot\\ \Delta p^m_{12}\end{pmatrix}\right\} = E\left\{\begin{pmatrix}\Delta p^1_2 - \Delta p^1_1\\ \cdot\\ \cdot\\ \Delta p^m_2 - \Delta p^m_1\end{pmatrix}\right\} = \begin{pmatrix}(-e^1_2)^T & 1\\ \cdot & \cdot\\ \cdot & \cdot\\ (-e^m_2)^T & 1\end{pmatrix}\begin{pmatrix}\Delta x_2\\ c\Delta\partial_2 t\end{pmatrix} \tag{6.5}$$

The satellite clock error unknown $\partial^i t$ actually represents the satellite (clock error) and propagation (troposphere and ionosphere) parameters together, cf. (4.25). The effects (on the coordinates) of position errors of receiver 1 and of the satellites will also be largely absorbed by these clock error parameters. The model of single difference observation equations is given by (6.5), cf. (5.5) in section 5.

There are m observations for 4 unknown parameters; the redundancy equals m-4, thus at least $m=4$ satellites are needed.

In the above model, the m undifferenced observables $p^1_1,..,p^m_1$ of receiver 1 have been left out, together with the m unknown satellite clock errors:

$$E\left\{\begin{pmatrix}\Delta p^1_1\\ \cdot\\ \cdot\\ \Delta p^m_1\end{pmatrix}\right\} = \begin{pmatrix}-1 & & \\ & \cdot & \\ & & -1\end{pmatrix}\begin{pmatrix}c\partial^1 t\\ \cdot\\ \cdot\\ c\partial^m t\end{pmatrix} \tag{6.6}$$

We can now see that the m observations of receiver 1, eq. (6.6), are applied as (range) corrections to the m observations of receiver 2, see eq. (6.5). Next, with model (6.5), the position coordinates are estimated.

The conclusion is, that for the above model assumptions, relative positioning with code observations is in principle identical to straightforward Differential GPS. By DGPS is meant, code positioning by means of range-corrections, determined at the reference station and applied at the rover site.

In practice with DGPS, there is a time delay in applying the corrections. It is caused by generating and transmitting the corrections. As such, relative code positioning can be seen as a limiting case of Differential GPS (time delay equal zero).

3.7 Summary and concluding remarks

This chapter has dealt with the mathematical modelling aspects of precise GPS positioning on a local scale. Two receivers simultaneously make observations to m satellites. The receiver separation is usually limited to about 10 kilometers. The most significant error sources are multipath, atmospheric delays and orbital errors.

The coordinates of the second receiver are determined with respect to those of the first one. This determination takes place in the WGS84 coordinate system. The transformation of the geometric information $\hat{x}_2$, into a local datum is not considered. In this chapter, the coordinate vector from r1 to r2, the baseline vector, was defined by the geometric antenna centers of the two receiver antennas. The connection to the survey markers on the ground and other practical aspects of GPS surveying are deferred to chapter 4.

The adjective 'kinematic' refers to increased surveying productivity, as compared with traditional static surveying. As such, a measurement technique is envisaged that is based on rapid static, but with short site occupation times (by quick recovery of the ambiguities) and the possibility (not the necessity!) of keeping lock to the satellites by the roving receiver during transportation to the next site. Although this last aspect enables trajectory determination, the emphasis of surveying remains on the determination of discrete points. In conclusion, kinematic GPS surveying stands for efficient precise coordinate determination on a local scale.

The full mathematical model for the L1-phase observable, for one epoch of the time varying measurement system has been treated in this chapter. Two equivalent implementations were given: in section 4 with undifferenced observations and in section 5 with double difference observations. For this particular application, the latter is thought to be more appropriate. Equations (5.7) and (5.8) are of the form $E\{\underline{y}_i\} = A_i x_i$; $D\{\underline{y}_i\} = Q_{y_i}$, cf. (1.3) in chapter 2.

As a result a mathematical model for kinematic GPS surveying is available and prepared for the recursive estimation, discussed in chapter 2. We were concerned with observables that result from measurements. Dynamic modelling of the time varying parameters in the measurement system is beyond the scope of this research. These pseudo-observations can however, be incorporated in the recursive estimation schemes of chapter 2 very easily (time update). No relation in time is assumed to exist for time varying parameters. In the model with double difference observations for stationary receivers only constant parameters are involved; in the recursive estimation actually a sequential adjustment is made upon these parameters. Further implementation aspects of the data processing, estimation and in particular quality control, are discussed in chapter 4.

Appendix 3.A Unit direction vector e_r^s

The unit direction vector e_r^s from receiver to satellite is needed in the linearization of the observation equation. First a few remarks concerning time-relations and the satellite position are made. Then the procedure for the computation of the unit direction vector e_r^s and the computed observation $\rho_{r,o}^s$ are treated.

time-relations
Time t_r is the receiver time tag in the (Rinex) observation file. The time of observation in GPS time reads $t = t_r - \delta_r t(t)$ and is expressed as $t = t_c + \Delta t - \delta_r t(t)$. Δt is the offset (in receiver time) between the time of observation and the computation time: $\Delta t = t_r - t_c$.

The time arguments of the geometric range can now be rewritten in

$$\begin{aligned} \rho_r^s(t, t-\tau) &= \rho_r^s(t_r - \delta_r t(t)\ ,\ t_r - \delta_r t(t) - \tau) \\ &= \rho_r^s(t_c + \Delta t - \delta_r t(t)\ ,\ t_c + \Delta t - \delta_r t(t) - \tau) \end{aligned} \tag{A.1}$$

The satellite ephemeris will be evaluated at time t_c (then to be interpreted as a GPS time, that is close to $(t-\tau)$). This evaluation needs to be done only once per satellite per observation epoch and can then be used for all receivers. The satellite positions (needed for the observations made by the various receivers) are then computed using a Taylor expansion. The procedure presented here, completely parallels the one in appendix B of [Jonge, 1993].

satellite position
For the definition of the reference system World Geodetic System WGS84 the reader is referred to [Parkinson et al, 1996] and for the current status [Malys et al, 1997].

For a treatment of coordinate systems and orbital motion, see [Seeber, 1993]. The three fundamental laws of Kepler determine the orbit of an object in space (normal orbital theory). The orbit is an ellipse with the earth's center in one of the foci of the ellipse. The perturbed orbit is represented by an extended set of Keplerian parameters. The representation by means of Keplerian elements has been chosen for the reason of graceful degradation in time [Dierendonck et al, 1978].

The satellite position will be computed based on data in the Rinex Navigation file, see [Gurtner et al, 1990]. The broadcast ephemeris are contained in the records 1-5 of the Navigation file, see also the Ephemeris Data Definitions in chapter 4 of [Parkinson et al, 1996] and [USCG, 1995].

For the ECEF coordinate system we take the WGS84: the coordinate vector is denoted by x. The (pseudo)-inertial coordinate system is identical to the WGS84 coordinate system, but the earth rotation has been switched off. The (constant) earth rotation rate $\dot{\Omega}_E$ is defined in the WGS84 reference system. Both systems coincide at 00:00h Sat/Sun midnight (begin of GPS week). The system is pseudo-inertial as the motion of the earth around the sun, together with the precession and nutation are neglected. This system will be used for the computation of the geometric range. Coordinate vectors are denoted by X.

If this pseudo-inertial coordinate system would be a true inertial system and if there would be no perturbing forces, the satellite orbit would be an ellipse and could be described by the six elementary Keplerian elements:

a	: semi major axis of orbital ellipse
e	: eccentricity
Ω	: longitude of ascending node
i	: inclination
M_o	: mean anomaly at reference time
ω	: argument of perigee

With the satellite ephemeris the satellite position (phase center of satellite antenna array) is known as function of time. The satellite position can be computed in WGS84 (coordinates obtained are corrected for polar motion): x^s; see the procedure in chapter 4 of [Parkinson et al, 1996]. We will compute the position in the pseudo-inertial system and also compute the velocity and acceleration of the satellite [Jonge, 1993].

procedure for e_r^s and $\rho_{r,o}^s$

The procedure (A.3) has to be carried out in every iteration step, per receiver-satellite combination. The first part, (A.2) for the satellite position, velocity and acceleration at time t_c, need to be carried out only once per observation epoch (outside the iteration).

We introduced approximate values for the receiver clock error, $\delta_r t_o$, and for the receiver position, $x_{r,o}$, so that we have $x_r = x_{r,o} + \Delta x_r$ and $\delta_r t = \delta_r t_o + \Delta\delta_r t$. The initial approximate values may result from a single point solution. This solution must certainly be carried out for the reference receiver, as the clock error of the reference receiver $\delta_1 t$ is not estimable, section 3.4.4. A good approximate value is necessary, see [Blewitt, 1993]; it is used to compute the satellite position (A.3).

The approximate time of observation (in GPS time) becomes $t_o = t_c + \Delta t - \delta_r t_o$. In the computation scheme (A.2)-(A.3) the $_o$ for approximate value is omitted.

In words, procedure (A.3) consists of the following steps. The receiver position is transformed to the pseudo-inertial system. R denotes the rotation matrix. The rotation angle is $\Omega(t) = \dot{\Omega}_E t$. The satellite position at the appropriate time, is computed using a second order Taylor expansion. With this position the geometric range is computed (in the pseudo-inertial frame). A new traveltime τ results. A new iteration step can be carried out. After termination of the iteration, the satellite position is computed in the ECEF system. The unit vector, expressed in the ECEF (WGS84) system, $e_r^s(t, t-\tau)$ for the design matrix, results. The value for $\rho_{r,o}^s$ is available after the iteration. Note that the word iteration here, applies to the recursive procedure for the geometric range in (A.3).

We will spend a few remarks on procedure (A.3). The geometric range is evaluated in the (pseudo)-inertial system, instead of the ECEF system. The advantage is that the earth rotation is not involved in the expansion of the satellite position. In chapter 4 of [Parkinson et al, 1996] we find: the user shall account for the effects due to earth rotation rate during the time of signal propagation so as to evaluate the path delay in an inertially stable coordinate system.

(A.2) read broadcast ephemeris

$$X^s(t_c)$$
$$\dot{X}^s(t_c)$$
$$\ddot{X}^s(t_c)$$

↓

(A.3) $$X_r(t) = R_{Xx}(-\Omega(t))\; x_r(t)$$

↓

initial guess for τ

↓

→ →

↓

$$X^s(t-\tau) = X^s(t_c) + \dot{X}^s(t_c)(\Delta t - \delta_r t(t) - \tau) + \frac{1}{2}\ddot{X}^s(t_c)(\Delta t - \delta_r t(t) - \tau)^2$$

↑ $$\rho_r^s(t,t-\tau) = \|X^s(t-\tau) - X_r(t)\|$$

$$\tau = \frac{\rho_r^s(t,t-\tau)}{c}$$

↑

↓

new τ ← stop?

↓

$$x^s(t-\tau) = R_{xX}(\Omega(t))\; X^s(t-\tau)$$

↓

$$e_r^s(t,t-\tau) = \frac{x^s(t-\tau) - x_r(t)}{\rho_r^s(t,t-\tau)}$$

The maximum values for the derivatives of the satellite position are

$$\|X^s\| = 2{,}6.10^7 \text{ m}$$
$$\|\dot{X}^s\| = 4{,}0.10^3 \text{ m/s}$$
$$\|\ddot{X}^s\| = 0{,}6 \text{ m/s}^2$$

In the iteration (A.3) we define the (improved) traveltime as geometric range divided by c. This does not comply with (4.2); the atmospheric and hardware delays are forgotten. This can be interpreted as starting each iteration (in computing the estimates for the geometric unknowns) with approximate values equal zero for these unknowns, as is assumed for (4.8) and (4.9). The effect is of the order 10^{-7} s for the traveltime and of the order 10^{-4} m for the geometric range computed using this traveltime (in next iteration). It largely cancels in relative positioning (as the neglected delays will about, but not exactly equal for both receivers).

The procedure can be used to compute $\rho_r^s(t,t-\tau)_o$ for the computed observation as well.

Alternatives to the procedure presented, method 5 of appendix B in [Jonge, 1993], are, iterative computation of the satellite position by repeated use of the standard procedure in [Parkinson et al, 1996], and iteration upon the geometric range with a Taylor expansion of the range itself (derivatives of ρ_r^s have to be computed).

Appendix 3.B Analysis on estimability

The mathematical model developed in section 3.4.2 is rankdeficient. The full set of parameters in the vector of unknowns can not be estimated, cf. section 1.2 of chapter 1. In order to allow a proper implementation of the estimation, we will apply reparametrizations on the unknowns and remove the rankdeficiencies. We consider five measurement scenarios:

C1: code, single frequency
L1: phase, single frequency
L1&L2: phase, dual frequency
C1+L1: phase and code, single frequency
L1&L2+C1&P2: phase and code, dual frequency

The discussion is based on the linearized observation equations, see (4.21) to (4.24) for the L1-phase observable; the Δ-symbols for the observations and unknowns are however omitted in this appendix. We start by analysing the set up with both receivers stationary, thus x_2 consists of three constant parameters. The set up with the rover in motion is briefly treated at the end of this appendix. Pseudo-observables (from a dynamic model) are left out of consideration.

From equation (4.1) it can be seen that hardware delays, clock errors and atmospheric delays are introduced for every epoch, see also (4.25). We assume to have k epochs of data. In the overviews index i runs from $i=1,..,k$ by default and index j from $j=1,..,m$. The part that is underlined in the estimable functions represents (parts of) the chosen S-basis. By P_1^s we mean the vector of L1-phase observations made by receiver r1 to all m satellites.

C1: code, single frequency

observations per epoch: p_1^s, p_2^s	total:	$2mk$
constant unknowns: x_2	total:	3
time-variant unknowns: $c\partial_1 t$, $c\partial_2 t$ and $c\partial^s t$	total:	$(m+2)k$

---

rankdeficiency: k

S-basis: $c\partial_1 t_i \doteq 0$ (k)

estimable functions:	x_2	geometric unknowns (3)
	$c\partial_2 t_i - \underline{c\partial_1 t_i}$	diff. receiver clock error (k)
	$c\partial^s t_i - \underline{c\partial_1 t_i}$	satellite clock error (mk)

redundancy: $k(m-1)-3$ $k=1 \rightarrow m \geq 4$

L1: phase, single frequency

observations per epoch: P_1^s, P_2^s	total:	$2mk$
constant unknowns: x_2, A_1^s and A_2^s	total:	$3+2m$
time-variant unknowns: $c\delta_1 t$, $c\delta_2 t$ and $c\delta^s t$	total:	$(m+2)k$

---

rankdeficiency: $k+m+1$

S-basis: $c\delta_1 t_i \doteq 0$ (k)
$A_1^j \doteq 0$ (m)
$A_2^1 \doteq 0$ (1)

estimable functions:	x_2	geometric unknowns (3)
	$c\delta_2 t_i - \underline{c\delta_1 t_i} + \underline{A_2^1} - \underline{A_1^1}$	diff. receiver clock error (k)
	$c\delta^j t_i - \underline{c\delta_1 t_i} - \underline{A_1^j}$	satellite clock error (mk)
	$A_2^j - \underline{A_2^1} - \underline{A_1^j} + \underline{A_1^1}$	double difference ambiguities (m-1), (with j=2,..,m)

redundancy: $k(m-1)-m-2$ $k=1$ -> not solvable
$k=2$ -> $m \geq 4$

L1&L2: phase, dual frequency

observations per epoch: P_1^s, $\bar{P}_1^s$, P_2^s and $\bar{P}_2^s$	total:	$4mk$
constant unknowns: x_2, A_1^s, $\bar{A}_1^s$, A_2^s and $\bar{A}_2^s$	total:	$3+4m$
time-variant unknowns: $c\delta_1 t$, $c\bar{\delta}_1 t$, $c\delta_2 t$, $c\bar{\delta}_2 t$, $c\delta^s t$ and $c\bar{\delta}^s t$	total:	$2(m+2)k$

---

rankdeficiency: $2(k+m+1)$

S-basis: $c\delta_1 t_i \doteq 0$ (k)
$c\bar{\delta}_1 t_i \doteq 0$ (k)
$A_1^j \doteq 0$ (m)
$\bar{A}_1^j \doteq 0$ (m)
$A_2^1 \doteq 0$ (1)
$\bar{A}_2^1 \doteq 0$ (1)

estimable functions:	x_2	geometric unknowns (3)
	$c\delta_2 t_i - \underline{c\delta_1 t_i} + \underline{A_2^1} - \underline{A_1^1}$	diff. receiver clock error (k)
	$c\bar{\delta}_2 t_i - \underline{c\bar{\delta}_1 t_i} + \underline{\bar{A}_2^1} - \underline{\bar{A}_1^1}$	diff. receiver clock error (k)
	$c\delta^j t_i - \underline{c\delta_1 t_i} - \underline{A_1^j}$	satellite clock error (mk)
	$c\bar{\delta}^j t_i - \underline{c\bar{\delta}_1 t_i} - \underline{\bar{A}_1^j}$	satellite clock error (mk)
	$A_2^j - \underline{A_2^1} - \underline{A_1^j} + \underline{A_1^1}$	double difference ambiguities (m-1), (with j=2,..,m)
	$\bar{A}_2^j - \underline{\bar{A}_2^1} - \underline{\bar{A}_1^j} + \underline{\bar{A}_1^1}$	double difference ambiguities (m-1), (with j=2,..,m)

redundancy: $2k(m-1)-2m-1$ $k=1$ -> not solvable
$k=2$ -> $m \geq 4$ same geometry is involved for L1 and L2; an additional rankdefect shows up in case m=3

C1+L1: phase and code, single frequency

observations per epoch: P_1^s, p_1^s, P_2^s and p_2^s total: $4mk$

constant unknowns: x_2, A_1^s and A_2^s total: $3+2m$

time-variant unknowns: $c\delta_1 t$, $c\partial_1 t$, $c\delta_2 t$, $c\partial_2 t$, $c\delta^s t$ and $c\partial^s t$ total: $2(m+2)k$

---

rankdeficiency: $2k+m+1$

S-basis:

$c\delta_1 t_i = 0$	(k)
$c\partial_1 t_i = 0$	(k)
$A_1^j = 0$	(m)
$A_2^1 = 0$	(1)

estimable functions:

x_2	geometric unknowns (3)
$c\delta_2 t_i - \underline{c\delta_1 t_i} + \underline{A_2^1} - \underline{A_1^1}$	diff. receiver clock error (k)
$c\partial_2 t_i - \underline{c\partial_1 t_i}$	diff. receiver clock error (k)
$c\delta^j t_i - \underline{c\delta_1 t_i} - \underline{A_1^j}$	satellite clock error (mk)
$c\partial^j t_i - \underline{c\partial_1 t_i}$	satellite clock error (mk)
$A_2^j - \underline{A_2^1} - \underline{A_1^j} + \underline{A_1^1}$	double difference ambiguities $(m\text{-}1)$, (with $j=2,..,m$)

redundancy: $2k(m\text{-}1)\text{-}m\text{-}2$ $\quad k=1 \rightarrow m \geq 4$

L1&L2+C1&P2: phase and code, dual frequency

observations per epoch: P_1^s, $\bar{P}_1^s$, p_1^s, $\bar{p}_1^s$, P_2^s, $\bar{P}_2^s$, p_2^s and $\bar{p}_2^s$ total: $8mk$

constant unknowns: x_2, A_1^s, $\bar{A}_1^s$, A_2^s and $\bar{A}_2^s$ total: $3+4m$

time-variant unknowns: $c\delta_1 t$, $c\bar{\delta}_1 t$, $c\partial_1 t$, $c\bar{\partial}_1 t$, $c\delta_2 t$, $c\bar{\delta}_2 t$, $c\partial_2 t$, $c\bar{\partial}_2 t$, $c\delta^s t$, $c\bar{\delta}^s t$, $c\partial^s t$ and $c\bar{\partial}^s t$ total: $4(m+2)k$

---

rankdeficiency: $4k+2m+2$

S-basis:

$c\delta_1 t_i = 0$	(k)
$c\bar{\delta}_1 t_i = 0$	(k)
$c\partial_1 t_i = 0$	(k)
$c\bar{\partial}_1 t_i = 0$	(k)
$A_1^j = 0$	(m)
$\bar{A}_1^j = 0$	(m)
$A_2^1 = 0$	(1)
$\bar{A}_2^1 = 0$	(1)

estimable functions:

x_2	geometric unknowns (3)
$c\delta_2 t_i - \underline{c\delta_1 t_i} + \underline{A_2^1} - \underline{A_1^1}$	diff. receiver clock error (k)
$c\bar{\delta}_2 t_i - \underline{c\bar{\delta}_1 t_i} + \underline{\bar{A}_2^1} - \underline{\bar{A}_1^1}$	diff. receiver clock error (k)
$c\partial_2 t_i - \underline{c\partial_1 t_i}$	diff. receiver clock error (k)
$c\bar{\partial}_2 t_i - \underline{c\bar{\partial}_1 t_i}$	diff. receiver clock error (k)
$c\delta^j t_i - \underline{c\delta_1 t_i} - \underline{A_1^j}$	satellite clock error (mk)
$c\bar{\delta}^j t_i - \underline{c\bar{\delta}_1 t_i} - \underline{\bar{A}_1^j}$	satellite clock error (mk)
$c\partial^j t_i - \underline{c\partial_1 t_i}$	satellite clock error (mk)
$c\bar{\partial}^j t_i - \underline{c\bar{\partial}_1 t_i}$	satellite clock error (mk)
$A_2^j - \underline{A_2^1} - \underline{A_1^j} + \underline{A_1^1}$	double difference ambiguities $(m\text{-}1)$, (with $j=2,..,m$)
$\bar{A}_2^j - \underline{\bar{A}_2^1} - \underline{\bar{A}_1^j} + \underline{\bar{A}_1^1}$	double difference ambiguities $(m\text{-}1)$, (with $j=2,..,m$)

redundancy: $4k(m\text{-}1)\text{-}2m\text{-}1$ $\quad k=1 \rightarrow m \geq 4$

same geometry is involved for L1 and L2; an additional rankdefect shows up in case m=3

We have discussed the estimability of five measurement scenarios in relative GPS surveying. In order to remove the rankdeficiencies in the mathematical models, we have constructed S-bases, thereby aiming at bases that can be given some practical interpretation. Other S-bases are possible as well.

For all scenarios it holds true that the clock error (per epoch) of the reference receiver can not be (separately) estimated, see also [Blewitt, 1993]: only relative clock behaviour (between receivers) is estimable. Carrier phase ambiguities of the reference receiver are not estimable, neither are those of the second receiver to the first satellite.

If one receiver makes a certain type of observation to a satellite and the other receiver does not, the observations are free observations: they can be left out from the data processing. Only the overlapping part in the data needs to be considered: with the data of one receiver, under current S-bases, one can estimate the satellite clock errors, for each observation type, for the particular epochs.

With (single or dual frequency) phase data, at least two epochs of data are required ($k \geq 2$). As soon as also code data are available, one epoch suffices ($k \geq 1$). In all five scenarios less satellites ($m < 4$) may suffice when more epochs of data are available. The absolute bottom is $m = 2$.

With both receivers stationary, we have 3 geometric unknowns for the whole survey. If receiver r2 is moving, we have 3 geometric unknowns per epoch. The redundancy for the five measurement scenarios is given below:

C1: code, single frequency

observations:	$2mk$	
unknowns:	$3k+(m+1)k$	
redundancy:	$k(m-4)$	$k=1 \rightarrow m \geq 4$

L1: phase, single frequency

observations:	$2mk$	
unknowns:	$3k+(m+1)k+(m-1)$	
redundancy:	$k(m-4)-(m-1)$	$k=2 \rightarrow m \geq 7$

L1&L2: phase, dual frequency

observations:	$4mk$	
unknowns:	$3k+2k(m+1)+2(m-1)$	
redundancy:	$2mk-5k-2(m-1)$	$k=2 \rightarrow m \geq 7$

same geometry is involved for L1 and L2; an additional rankdefect shows up in case m=4,5,6

L1+C1: phase and code, single frequency:

observations:	$4mk$	
unknowns:	$3k+2k(m+1)+(m-1)$	
redundancy:	$2mk-5k-(m-1)$	$k=1 \rightarrow m \geq 4$

L1&L2+C1&P2: phase and code, dual frequency

observations:	$8mk$	
unknowns:	$3k+4k(m+1)+2(m-1)$	
redundancy:	$4mk-7k-2(m-1)$	$k=1 \rightarrow m \geq 4$

same geometry is involved for L1 and L2; an additional rankdefect shows up in case m=3

In brief, the rule for a moving receiver r2 reads that with code observations at least 4 satellites are required and at least 7 satellites when only phase observations are available. With only phase data, single or dual frequency, at least two epochs of data are needed ($k \geq 2$). When code data are available, one epoch suffices ($k \geq 1$). Less than 4 satellites is not possible, even when more epochs of data are available.

References

[Ackermann, 1997] Ackermann, F. (1997). GPS for photogrammetry. Bollettino di Geodesia e Scienze Affini. Vol. 56. No. 4. pp. 387-406.

[Alberda, 1990] Alberda, J.E. (1990). Inleiding landmeetkunde. Delftse Uitgevers Maatschappij. Delft.

[Bauer, 1997] Bauer, M. (1997). Vermessung und Ortung mit Satelliten. Wichmann, Heidelberg. 4. Auflage.

[Blewitt, 1989] Blewitt, G. (1989). Carrier phase ambiguity resolution for the Global Positioning System applied to geodetic baselines up to 2000 km. Journal of Geophysical Research. Vol. 94, No. B8 August. pp. 10.187-10.203.

[Blewitt, 1993] Blewitt, G. (1993). Advances in Global Positioning System technology for geodynamics investigations: 1978-1992. Contributions of space geodesy to geodynamics: technology. AGU Geodynamics Series, Vol. 25. D.E. Smith and D.L. Turcotte (Eds). pp. 195-213.

[Bock et al, 1985] Bock, Y. et al (1985). Establishment of three-dimensional geodetic control by interferometry with the Global Positioning System. Journal of Geophysical Research. Vol. 90, No. B9 August. pp. 7689-7703.

[Bock et al, 1986] Bock, Y. et al (1986). Interferometric analysis of GPS phase observations. Manuscripta Geodaetica. Vol. 11. pp. 282-288.

[Brouwer, 1985] Brouwer, F.J.J. (1985). On the principles, assumptions and methods of geodetic very long baseline interferometry. Ph.D. thesis. Department of Geodetic Engineering. Delft University of Technology.

[Counselman et al, 1972] Counselman, C.C., H.F. Hinteregger and I.I. Shapiro (1972). Astronomical applications of differential interferometry. Science. Vol. 178. November. pp. 607-608.

[Counselman et al, 1979] Counselman, C.C. and I.I. Shapiro (1979). Miniature interferometer terminals for earth surveying. Bulletin Géodésique. Vol. 53. pp. 139-163.

[Counselman et al, 1981] Counselman, C.C. and S.A. Gourevitch (1981). Miniature interferometer terminals for earth surveying: ambiguity and multipath with the Global Positioning System. IEEE Transactions on Geoscience and Remote Sensing. Vol. Ge-19, No. 4 October. pp. 244-252.

[Dierendonck et al, 1978] Dierendonck, A.J. v., S.S. Russell, E.R. Kopitzke and M. Birnbaum (1978). The GPS navigation message. Navigation. Vol. 25, No. 2. pp. 147-165.

[Euler et al, 1991] Euler, H-J. and C.C. Goad (1991). On optimal filtering of GPS dual frequency observations without using orbit information. Bulletin Géodésique. Vol. 65. pp. 130-143.

[Goad et al, 1988] Goad C.C. and A. Müller (1988). An automated procedure for generating an optimum set of independent double difference observables using Global Positioning System carrier phase measurements. Manuscripta Geodaetica. Vol. 13, No. 6. pp. 365-369.

[Goad, 1992] Goad, C.C. (1992). Robust techniques for determining GPS phase ambiguities. Sixth International Geodetic Symposium on Satellite Positioning. Columbus, Ohio, USA. March 17-20. pp. 245-254.

[Gurtner et al, 1990] Gurtner, W. and G.L. Mader (1990). The RINEX format: current status, future developments. Proceedings of the 2nd International Symposium on Precise Positioning with the GPS. Ottawa, Canada. September 3-7. pp. 977-992.

[Hatch et al, 1994] Hatch, R.R. and H-J Euler (1994). Comparison of several kinematic techniques. Proceedings of ION GPS-94. Salt Lake City, Utah, USA. September 20-23. pp. 363-370.

[Hofmann-Wellenhof et al, 1997] Hofmann-Wellenhof, B., H. Lichtenegger and J. Collins (1997). GPS Theory and Practice. Springer-Verlag. Wien, New York. Fourth revised edition.

[Jonge, 1993] Jonge, P.J. de (1993). GPSVEC - GPS baseline adjustment software. Internal report. Delft Geodetic Computing Centre. Delft University of Technology. May 6.

[Jonkman, 1997] Jonkman, N.F. (1997). Integer GPS-ambiguity estimation without the receiver-satellite geometry. LGR-series No. 18. Delft Geodetic Computing Centre. February 1998. 95 pp.

[Kleusberg et al, 1995] Kleusberg, A. and P.J.G. Teunissen (Eds) (1995). GPS for Geodesy. Lecture notes in Earth Sciences. No. 60. Springer Verlag, 1996 (2nd edition, 1998).

[Kuang et al, 1996] Kuang, D., B.E. Schutz and M.M. Watkins (1996). On the structure of geometric positioning information in GPS measurements. Journal of Geodesy. Vol. 71. pp. 35-43.

[Leick, 1995] Leick, A. (1995). GPS satellite surveying. 2nd edition. John Wiley, New York.

[Lichten, 1990] Lichten, S.M. (1990). Estimation and filtering for high-precision GPS positioning applications. Manuscripta Geodaetica. Vol. 15. pp. 159-176.

[MacDoran, 1979] MacDoran, P.F. (1979). Satellite emission radio interferometric earth surveying series - GPS geodetic system. Bulletin Géodésique. Vol. 53. pp. 117-138.

[MacDoran et al, 1985] MacDoran, P.F., R.B. Miller, L.A. Buennagel and J.H. Whitcomb (1985). Codeless systems for positioning with NAVSTAR-GPS. Proceedings First International Symposium on Precise Positioning with the Global Positioning System. Rockville, Maryland, USA. April 15-19. pp. 181-190.

[Malys et al, 1997] Malys,.S., J.A. Slater, R.W. Smith, L.E. Kunz and S.C. Kenyon (1997). Status of the World Geodetic System 1984. Proceedings of KIS97. June 3-6. Banff, Canada. pp. 25-34.

[Marel, 1997] Marel, H. v.d. (1997). Active GPS control stations. Chapter 9 in GPS for Geodesy. P.J.G. Teunissen and A. Kleusberg (Eds). Lecture notes in Earth Sciences. Springer Verlag, 1998. 2nd edition.

[Meehan et al, 1992] Meehan, T.K., J.M. Srinivasan, D.J. Spitzmesser, C.E. Dunn, J.Y. Ten, J.B. Thomas, T.N. Munson and C.B. Duncan (1992). The TurboRogue GPS receiver. Proceedings of the Sixth International Geodetic Symposium on Satellite Positioning. Columbus, Ohio, USA. March 17-20. pp. 209-218.

[Melbourne, 1985] Melbourne, W.G. (1985). The case for ranging in GPS-based geodetic systems. Proceedings First International Symposium on Precise Positioning with the Global Positioning System. Rockville, Maryland, USA. April 15-19. pp. 373-386.

[Odijk, 1997] Odijk, D. (1997). Treatment of the ionospheric effect on precise GPS positioning. Internal report. Delft Geodetic Computing Centre (LGR). Delft University of Technology. January.

[Parkinson et al, 1996] Parkinson, B.W. and J.J. Spilker (Eds) (1996). Global Positioning System: theory and applications. Progress in Astronautics and Aeronautics. Vol. 163 and 164. AIAA. Washington.

[Remondi, 1984] Remondi, B.W. (1984). Using the Global Positioning System phase observable for relative geodesy: modeling, processing, and results. Ph.D. thesis. The University of Texas at Austin.

[Remondi, 1988] Remondi, R.W. (1988). Kinematic and pseudo-kinematic GPS. Proceedings of ION GPS-88. Colorado Springs, Colorado, USA. September 19-23. pp. 115-121.

[Seeber, 1993] Seeber, G. (1993). Satellite Geodesy. de Gruyter Berlin, New York.

[Strang et al, 1997] Strang, G. and K. Borre (1997). Linear algebra, geodesy, and GPS. Wellesley-Cambridge Press. MA.

[Teunissen, 1989] Teunissen, P.J.G. (1989). Nonlinear inversion of geodetic and geophysical data: diagnosing nonlinearity. Developments in four-dimensional geodesy. F.K. Brunner and C. Rizos (Eds). Lecture notes in Earth Sciences. No. 29. Springer Verlag. pp. 241-264.

[Teunissen, 1990] Teunissen, P.J.G. (1990). GPS op afstand bekeken. Een halve eeuw in de goede richting. Lustrumboek Snellius 1985-1990. pp. 215-233. Technische Universiteit Delft.

[Teunissen, 1995] Teunissen, P.J.G. (1995). The invertible GPS ambiguity transformations. Manuscripta Geodaetica. Vol. 20. pp. 489-497.

[Teunissen, 1996] Teunissen, P.J.G. (1996). An analytical study of ambiguity decorrelation using dual frequency code and carrier phase. Journal of Geodesy. Vol. 70. No. 8. pp. 515-528.

[Teunissen, 1997a] Teunissen, P.J.G. (1997). GPS double difference statistics: with and without using satellite geometry. Journal of Geodesy. Vol. 71. No. 3. pp. 137-148.

[Teunissen, 1997b] Teunissen, P.J.G. (1997). The geometry free GPS ambiguity search space with a weighted ionosphere. Journal of Geodesy. Vol. 71. No. 6. pp. 370-383.

[USCG, 1995] USCG (1995). Global Positioning System: standard positioning service signal specification. 2nd edition. June 2. Navtech (October 1995).

[Zieliński, 1989] Zieliński, J.B. (1989). GPS baseline error caused by orbital uncertainty. Manuscripta Geodaetica. Vol. 14. pp. 117-124.