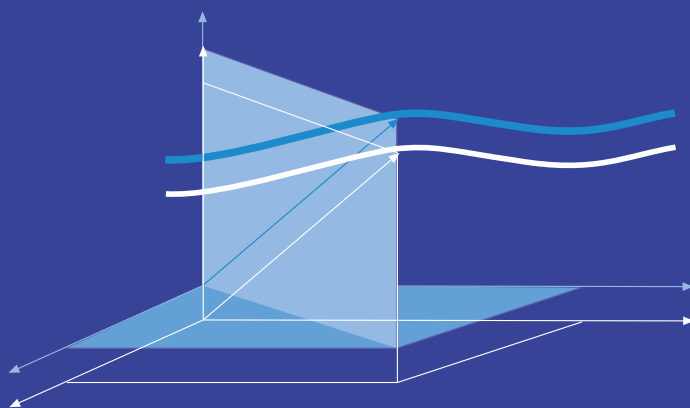


Dynamic data processing

recursive least-squares

P.J.G. Teunissen



Series on Mathematical Geodesy and Positioning

Dynamic data processing
recursive least-squares

Dynamic data processing
recursive least-squares

P.J.G. Teunissen



Delft University of Technology
Department of Mathematical Geodesy and Positioning

VSSD

© VSSD

First edition 2001-2007

Published by:

VSSD

Leeghwaterstraat 42, 2628 CA Delft, The Netherlands

tel. +31 15 27 82124, telefax +31 15 27 87585, e-mail: hlf@vssd.nl

internet: <http://www.vssd.nl/hlf>

URL on this book and related books: <http://www.vssd.nl/hlf/a030.htm>

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photo-copying, recording, or otherwise, without the prior written permission of the publisher.

Ebook: ISBN 978-90-6562-217-4

Gedrukte versie: ISBN 978-90-407-1976-9

NUR 930

Keywords: dynamic, data processing, geodesy

Foreword

This book is based on the lecture notes of the course *Dynamic data processing* as it has been given by the Department of Mathematical Geodesy and Positioning (MGP) of the Delft University of Technology since 1990. The prerequisites are a solid knowledge of adjustment theory and geodetic positioning, together with linear algebra, statistics and calculus. The theory and application of least-squares adjustment are treated in Adjustment theory (Delft University Press, 2000). The material of the present course extends the theory to the recursive estimation of time-varying or dynamic parameters. The time-varying parameters could for instance be geometric parameters such as position, attitude and shape, physical parameters such as temperature and humidity, or instrumental parameters such as clock drifts and biases. The time-varying parameters are said to be determined recursively when the method of determination enables sequential, rather than batch processing of the measurement data. The main goal is therefore to convey the knowledge necessary to be able to process sequentially collected measurement data in an optimal and efficient manner for the purpose of estimating time-varying parameters.

Following the *Introduction*, the basic theory of least-squares estimation is reviewed in *Chapter 1*. This is done for the model of observation equations and for the model of condition equations. In *Chapter 2* the principle of recursive least-squares estimation is introduced. The recursive principle allows one to update the least-squares solution for new observations without the need to store all past observations. Two different forms of the measurement-update equations are given. The results of Chapter 2, which hold true for time-invariant parameters, are generalized in *Chapter 3* to the case of time-varying parameters. The time-varying nature of the parameters is assumed captured by means of polynomial equations of motion. The recursive solution now consists of two types of update equations, the measurement-update equations and the time-update equations. Since there still exist many dynamic systems for which the rather simple polynomial model of Chapter 3 does not apply, a larger class of dynamic models is introduced in *Chapter 4*. These models are formulated using the state-space description of dynamic systems. In order to include randomness in the state-space description of dynamic systems, some of the elementary concepts of the theory of random functions are discussed in *Chapter 5*. This chapter also includes a description of the propagation laws for linear, time-varying systems. The results of Chapter 5 are used in *Chapter 6* to model possible uncertainties associated with the dynamic model. As a result the update equations are obtained for the recursive least-squares filtering and prediction of time-varying parameters.

Many colleagues of the Department of Mathematical Geodesy and Positioning whose assistance made the completion of this book possible are gratefully acknowledged. The typing of the book was done by Mrs. M.P.M. Scholtes, while C.D. de Jong took care of the editing. Various lecturers have taught the book's material over the past years. In particular the feedback and valuable recommendations of the lecturers H.M. de Heus, C.D. de Jong and C.C.J.M. Tiberius are acknowledged.

P.J.G. Teunissen
July, 2001

Contents

Introduction	1
1 Least-squares: a review	5
1.1 The linear A-model	5
1.1.1 Consistency and inconsistency	5
1.1.2 Least-squares estimates	13
1.1.3 A stochastic model for the observations	18
1.1.4 Least-squares estimators	19
1.1.5 Summary	23
1.2 The nonlinear A-model	24
1.2.1 Nonlinear observation equations	24
1.2.2 The linearized A-model	27
1.2.3 Least-squares iteration	33
1.3 The B-model	35
1.3.1 The linear B-model	35
1.3.2 The nonlinear B-model	38
2 Recursive least-squares: the static case	43
2.1 Introduction	43
2.2 Recursive least-squares: the A-form	46
2.3 Recursive least-squares: the B-form	54
2.4 Linearization, iteration and recursion	63
3 Recursive least-squares: the time-varying case	73
3.1 Introduction	73
3.2 Equations of motion: a polynomial model	74
3.3 Prediction, filtering and smoothing	80
3.4 Recursive prediction and filtering: the A-form	87
3.5 Recursive prediction and filtering: the B-form	90
4 State-space models for dynamic systems	103
4.1 Introduction	103
4.2 Equations of motion: kinematics	103
4.3 Equations of motion: dynamics	111
4.4 State vector description of dynamic systems	118
4.5 Linearization of a nonlinear state equation	126
4.6 Linear time-varying state equations	132
4.7 Linear time-invariant state equations	136
4.8 Evaluation of the matrix exponential	138
4.8.1 The Taylor-series method	138
4.8.2 The Jordan canonical form method	142

4.9	Summary	150
5	Random functions	153
5.1	Introduction	153
5.2	The mean and covariance of random functions	153
5.3	Propagation laws for linear systems: the general case	162
	5.3.1 The mean of the output $\underline{x}(t)$	163
	5.3.2 The cross-covariance between the output $\underline{x}(t)$ and input $\underline{z}(t)$	164
	5.3.3 The auto-covariance of the output $\underline{x}(t)$	170
5.4	White noise	183
5.5	The auto-covariance of the output of a linear system: the white noise case	190
5.6	Random polynomial equations of motion	201
	5.6.1 Random constants as input	201
	5.6.2 White noise as input	203
	5.6.3 Exponentially correlated noise as input	204
5.7.	Summary	208
6	Recursive least-squares: the dynamic case	209
6.1	Introduction: filter divergence	209
6.2	The dynamic model of observation equations	213
6.3	Recursive prediction and filtering	214
6.4	State vector augmentation	232
	6.4.1 Exponentially correlated zero-mean velocity	233
	6.4.2 Exponentially correlated zero-mean acceleration	235
6.5	Summary	239
	Literature	240
	Index	241

Introduction

As in other physical sciences, empirical data are used in geodesy to make inferences so as to describe physical reality. Many such problems involve the determination of unknown parameters from a set of redundant measurements. Measurements are said to be redundant when they exceed the minimum necessary for a unique determination of the parameters. There are two main reasons for collecting redundant measurements. First the requirement to be able to check for mistakes or errors. Second the wish to increase the accuracy of the results computed. As a consequence of measurement uncertainty (exact measurements do not exist), the redundant data are usually inconsistent in the sense that each sufficient subset yields results which will differ from the results obtained from another subset. To obtain a unique solution, consistency needs to be restored by applying corrections to the data. This computational process of making the measurement data consistent with the model such that the unknown parameters can be determined uniquely, is referred to as adjustment. Adjustment theory therefore deals with the optimal combination of redundant measurements together with the estimation of unknown parameters. An introductory course on adjustment was presented in *Adjustment theory* (Delft University Press, 2000). This theory is extended in this book to the case of *time-varying* or dynamic parameters with an emphasis on their recursive estimation.

Time-varying parameters occur in many geodetic models. They could be geometric parameters such as position, attitude and shape, physical parameters such as temperature and humidity, or instrumental parameters such as clock drifts and biases. When a body (e.g. satellite, aircraft, car, or ship) is in motion, its position changes as function of time. Being able to track the position of such a moving object is of importance, for instance for navigation and guidance. A moving body may also change its attitude as function of time. Attitude determination is sometimes needed as an aid to navigation and guidance, but it also applies, in case of earth rotation, to the Earth as a whole. Objects that are subject to deformation change their shape as a function of time. On a global scale, for instance, the earth deforms due to various geophysical processes. But the earth's surface may also change its shape on more local or regional scales. Subsidence due to gas extraction is one such example. Apart from time-varying geometric parameters, also physical and instrumental parameters may change as function of time. Atmospheric parameters such as those of the ionosphere and troposphere, change on an hourly, daily and even seasonal basis. Also the performance of instruments often displays a dependence on time. This is the reason why calibrations are carried out, so as to keep the time-varying instrumental parameters in control.

A parameter solution is said to be recursive when the method of determination enables sequential, rather than batch processing of the measurement data. The need for a recursive solution is usually driven by the efficiency with which such solutions can be computed. This holds true in particular for applications in which the time-varying parameters need to be determined instantly or in real-time. We speak of a (near) real-time determination when the time of determination (almost) coincides with the time the parameter takes on the value to be determined. Such applications can typically be found in the area of navigation and guidance. In the case of navigation, for instance, it does not make sense to determine one's position with a too long time delay. In these applications there is therefore a real need to have a computational cycle time of the position determination that is as short as possible. This is feasible when

recursive methods are used. But even in case real-time solutions are not an important issue, the use of recursive methods can still be attractive due to their computational efficiency.

When determining time-varying parameters from sequentially collected measurement data, one can discriminate between three types of estimation problems (see Figure 0.1). When the time at which a parameter estimate is required coincides with the time the last measurements are collected, the problem is referred to as *filtering*. When the time of interest falls within the time span of available measurement data, the problem is referred to as *smoothing*, and when the time of interest occurs after the time the last measurements are collected, the problem is called *prediction*. Thus filtering aims at the determination of current parameter values, while smoothing and prediction aim respectively at the determination of past and future parameter values. The emphasis in this book will be on recursive filtering.

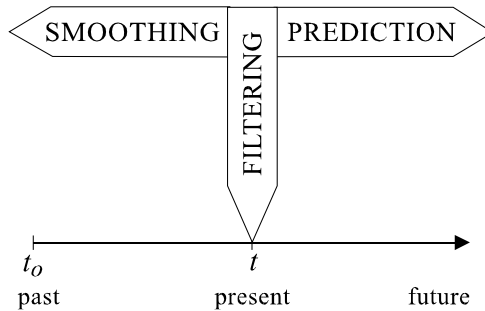


Figure 0.1: Prediction, filtering and smoothing.

The essence of a recursive method is that it enables one to update the parameter estimates for new measurements without the need to store all past measurements. Assume, for example, that one has collected at epoch $t-1$ a redundant set of measurements y_{t-1} which bears a linear relationship with an unknown parameter vector x . The measurements y_{t-1} can then be used to obtain a linear least-squares estimate \hat{x}_{t-1} of the unknown parameter vector x . Now assume that at the next epoch t a new set of measurements y_t becomes available which also bears a linear relationship to the same unknown parameter vector x . Since these additional measurements also contain information about the unknown parameter vector x , they can be used to improve the estimate \hat{x}_{t-1} of x . One approach would be to use both y_{t-1} and y_t and to repeat the least-squares adjustment. As a result one obtains the improved least-squares estimate \hat{x}_t of x . Although this approach is valid, it requires that one saves the past measurements y_{t-1} . In some cases this may be a too heavy computational burden, in particular if there are many past measurements or many epochs that precede the current epoch. Fortunately there is an alternative approach available, the recursive solution. It can be shown (under some mild restrictions) that the same improved least-squares estimate \hat{x}_t of x , can also be computed from \hat{x}_{t-1} and y_t instead of from y_{t-1} and y_t . The solution will then have the recursive structure:

$$\hat{x}_t = \hat{x}_{t-1} + K_t(y_t - A_t \hat{x}_{t-1})$$

in which K_t and A_t are matrices. This recursive equation, which holds true for any epoch t , is referred to as the *measurement-update* equation: the new measurements y_t are used to update the previous parameter estimate \hat{x}_{t-1} so as to obtain the current parameter estimate \hat{x}_t .

Some elements of recursive estimation were already briefly introduced in *Adjustment theory* (Chapter 6, Section 3). However, just as in the above example, this brief introduction only dealt with models in which the parameter vector remained constant in time. In this book we will extend the theory to the case of time-varying parameters. This implies that some additional modeling needs to be done, namely one that describes the time-dependence of the parameter vector. Depending on the application at hand, these equations of motion can be of a kinematic or of a dynamic nature. Kinematics is used to relate position, velocity, acceleration and time without reference to the cause of motion, whereas dynamics also includes an explicit description of the forces responsible for the motion. As a consequence of having incorporated the time-varying nature of the parameter vector into the model, the recursion will now consist of two different update equations, the *time-update* (TU) and the *measurement update* (MU):

$$\hat{x}_{t|t-1} = \Phi_{t,t-1} \hat{x}_{t-1|t-1} \quad (\text{TU}) \quad \text{and} \quad \hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - A_t \hat{x}_{t|t-1}) \quad (\text{MU})$$

with $\Phi_{t,t-1}$ the transition matrix. The time-update uses the filtered estimate $\hat{x}_{t-1|t-1}$ of epoch $t-1$ to predict the parameter vector of the next epoch, x_t , as $\hat{x}_{t|t-1}$. This predicted estimate together with the new measurements y_t are then combined in the measurement update to obtain the filtered estimate of x_t as $\hat{x}_{t|t}$.

1 Least-squares: a review

1.1 The linear A-model

1.1.1 Consistency and inconsistency

Assume that we want to determine n parameters $x_\alpha \in \mathbb{R}$, $\alpha = 1, \dots, n$. An m -number of measurements $y_i \in \mathbb{R}$, $i = 1, \dots, m$, are carried out to determine these parameters. If the measurements bear a known linear relationship with the unknown parameters, we may write the model of observation equations as:

$$(1) \quad y_i = \sum_{\alpha=1}^n a_{i\alpha} x_\alpha, \quad i = 1, \dots, m.$$

In this equation the known scalars $a_{i\alpha}$ model the assumed linear relationships between the measurements y_i and the parameters x_α . By introducing the matrix and vectors:

$$\underset{m \times n}{\mathbf{A}} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad \underset{m \times 1}{\mathbf{y}} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \quad \underset{n \times 1}{\mathbf{x}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

equation (1) can be written in matrix-vector form as:

$$(2) \quad \underset{m \times 1}{\mathbf{y}} = \underset{m \times n}{\mathbf{A}} \underset{n \times 1}{\mathbf{x}}.$$

This is a system of an m -number of linear equations in an n -number of unknown parameters. It is now of interest to know under what conditions a solution to the linear system (2) exists and if a solution exists, whether it is unique or not. It will be clear that a solution to (2) exists if and only if the vector \mathbf{y} can be written as a linear combination of the column vectors of matrix \mathbf{A} . If this is the case the vector \mathbf{y} is an element of the column space or range space of matrix \mathbf{A} . This space is denoted as $R(\mathbf{A})$. Thus a solution to (2) exists if and only if:

$$(3) \quad \mathbf{y} \in R(\mathbf{A}).$$

Systems for which this holds are called consistent systems. A system is said to be inconsistent if and only if:

$$(4) \quad \mathbf{y} \notin R(\mathbf{A}).$$

In this case the vector \mathbf{y} cannot be written as a linear combination of the column vectors of matrix \mathbf{A} and hence no vector \mathbf{x} exists such that (2) holds. The difference between consistency and inconsistency is depicted geometrically in Figure 1.1.

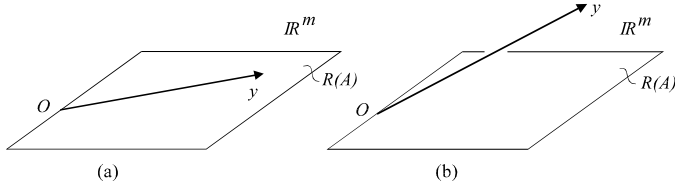


Figure 1.1: $\left\{ \begin{array}{l} (a) \text{ consistency} : y \in R(A) \subset \mathbb{R}^m \\ (b) \text{ inconsistency} : y \notin R(A) \subset \mathbb{R}^m \end{array} \right.$

Since $y \in \mathbb{R}^m$, it follows from (3) that consistency is guaranteed if $R(A) = \mathbb{R}^m$. But $R(A) = \mathbb{R}^m$ only holds if the dimension of $R(A)$ equals the dimension of \mathbb{R}^m . Hence, if $\dim R(A) = m$. It follows therefore, since $\dim R(A)$ equals the rank of matrix A (= number of linear independent columns or rows of A), that consistency is guaranteed if and only if:

$$(5) \quad \text{rank } A = m .$$

In all other cases, $\text{rank } A < m$, the linear system may or may not be consistent. Assuming consistency, the next question one can ask is whether the solution to (2) is unique or not. That is, whether the information content of the measurements collected in the vector y is sufficient for determining the parameter vector x uniquely. The solution is unique only if all the columns of matrix A are linearly independent. Hence, the solution is unique if the rank of matrix A equals the number of unknown parameters:

$$(6) \quad \text{rank } A = n .$$

To clarify this, assume x and $x' \neq x$ to be two different solutions of (2). Then $Ax = Ax'$ or $A(x - x') = 0$ must hold. But this can only be the case if some of the columns of matrix A are linearly dependent, which contradicts the assumption of full column rank (6). In all other cases, $\text{rank } A < n$, there will be more than one solution to a consistent system. In this book we will always assume that (6) holds. The case that $\text{rank } A < n$ is treated elsewhere [Teunissen, 1985a]. With $\text{rank } A = n$ and the fact that the rank of matrix A is always equal to or less than the number of rows or columns of A , it follows that two cases can be distinguished:

$$(7) \quad m = n = \text{rank } A \quad \text{or} \quad m > n = \text{rank } A .$$

In the first case, both (5) and (6) are satisfied, implying that the linear system (2) is consistent and that a unique solution exists. The unique solution, denoted by \hat{x} , is found through an inversion of the matrix A :

$$(8) \quad \left. \begin{array}{l} y = A x \\ \begin{matrix} m \times 1 & m \times n & n \times 1 \end{matrix} \end{array} \right\} \Rightarrow \hat{x} = A^{-1}y .$$

$$m = n = \text{rank } A$$

In the second case, only (6) is satisfied, implying that a unique solution to (2) exists provided that the system is consistent. Consistency in this case is however not guaranteed. But if we assume the system to be consistent, that is $y \in R(A)$, one way to obtain the unique solution is to invert n out of the $m > n$ linear equations:

$$(9) \quad \left. \begin{array}{l} y = \mathbf{A} x \\ m \times 1 \quad m \times n \quad n \times 1 \\ m > n = \text{rank } \mathbf{A} \\ \text{consistent} \end{array} \right\} \Rightarrow \underbrace{\begin{array}{c} n \\ \left(\begin{array}{c} y_1 \\ \dots \\ y_2 \end{array} \right) \\ m-n \\ y \\ m \times 1 \end{array}} = \underbrace{\begin{array}{c} n \\ \left(\begin{array}{c} \mathbf{A}_1 \\ \dots \\ \mathbf{A}_2 \end{array} \right) \\ m-n \\ \mathbf{A} \\ m \times n \end{array}} \Rightarrow \hat{x} = \mathbf{A}_1^{-1} y_1.$$

Since the columns of matrix \mathbf{A} are linearly independent, it is possible to find a matrix \mathbf{A}_1 for which the columns are linearly independent as well, implying that the inverse of the square matrix \mathbf{A}_1 exists. Note that y_2 is not used in computing \hat{x} . This is allowed in the present situation since y_2 is consistent with y_1 and hence does not contain any additional information.

Example 1

Consider the linear system:

$$(10) \quad \underbrace{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_x.$$

In this case we have: $m = 2$, $n = 2$ and $\text{rank } \mathbf{A} = 2$. Thus, the system is consistent since $\text{rank } \mathbf{A} = m = 2$, and the system has a unique solution since $\text{rank } \mathbf{A} = n = 2$. The unique solution of (10) reads: $\hat{x} = (5/7, 3/7)^*$.

Example 2

A particle is moving with constant velocity along a straight line. If we denote the position of the particle as function of time as $u(t)$, we have:

$$u(t) = u(t_0) + \dot{u}(t_0)(t - t_0)$$

with $u(t_0)$ and $\dot{u}(t_0)$ being the initial position and initial velocity respectively of the particle at time t_0 . It is assumed that the initial position and initial velocity of the particle are unknown. The unknown parameters $u(t_0)$ and $\dot{u}(t_0)$ can then be determined from two measurements of position at times t_1 and $t_2 \neq t_1$. This results in the following linear system (see Figure 1.2):

$$(11) \quad \underbrace{\begin{pmatrix} u(t_1) \\ u(t_2) \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 1 & (t_1 - t_0) \\ 1 & (t_2 - t_0) \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}}_x.$$

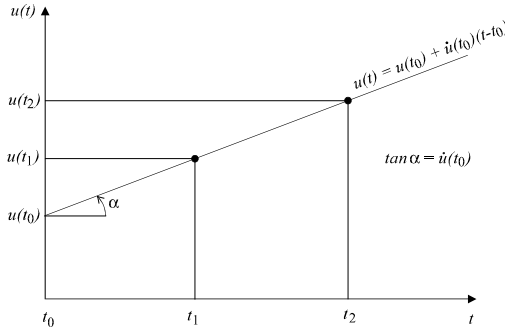


Figure 1.2: Position as function of time.

In this case we have: $m = 2, n = 2$ and $\text{rank } A = 2$. Thus, the system is consistent since $\text{rank } A = m = 2$, and the system has a unique solution since $\text{rank } A = n = 2$. With

$$A^{-1} = \frac{1}{(t_2 - t_1)} \begin{pmatrix} (t_2 - t_0) & -(t_1 - t_0) \\ -1 & 1 \end{pmatrix}$$

the unique solution follows as:

$$\begin{pmatrix} \hat{u}(t_0) \\ \hat{u}(t_0) \end{pmatrix} = \frac{1}{(t_2 - t_1)} \begin{pmatrix} (t_2 - t_0) & -(t_1 - t_0) \\ -1 & 1 \end{pmatrix} \begin{pmatrix} u(t_1) \\ u(t_2) \end{pmatrix}$$

or as:

$$(12) \quad \begin{cases} \hat{u}(t_0) = \frac{t_2 - t_0}{t_2 - t_1} u(t_1) - \frac{t_1 - t_0}{t_2 - t_1} u(t_2) = u(t_1) - \hat{u}(t_0)(t_1 - t_0) \\ \hat{u}(t_0) = \frac{u(t_2) - u(t_1)}{t_2 - t_1} \end{cases}$$

Note that $\text{rank } A = 1$ if $t_2 = t_1$. In this case no unique solution exists.

Example 3

Consider the linear system:

$$(13) \quad \underbrace{\begin{pmatrix} -2 \\ 3 \\ -1 \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 1 & 3 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}}_A \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_x$$

In this case we have: $m = 3, n = 2$ and $\text{rank } A = 2$. Since $m = 3 > \text{rank } A = 2$, consistency of the system is not automatically guaranteed. A closer look at the measurement vector y of (13) shows however that:

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

This shows that y can be written as a linear combination of the column vectors of A . Therefore $y \in R(A)$, showing that the system is consistent. And since $n = \text{rank } A = 2$, its solution is also unique. If we partition (13) as:

$$\begin{pmatrix} -2 \\ 3 \\ \dots \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & -1 \\ \dots & \dots \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

the unique solution follows as:

$$(14) \quad \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}^{-1} \begin{pmatrix} -2 \\ 3 \end{pmatrix} = -\frac{1}{7} \begin{pmatrix} -1 & -3 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The system (13) may of course also be partitioned as:

$$\begin{pmatrix} -2 \\ \dots \\ 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ \dots & \dots \\ 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

The unique solution follows then as:

$$(15) \quad \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ -1 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Example 4

Consider again the situation of a particle moving along a straight line with constant velocity. But now assume that three measurements of position are carried out at the three different times t_1, t_2 and t_3 . The linear system reads then:

$$(16) \quad \underbrace{\begin{pmatrix} u(t_1) \\ u(t_2) \\ u(t_3) \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 1 & (t_1 - t_0) \\ 1 & (t_2 - t_0) \\ 1 & (t_3 - t_0) \end{pmatrix}}_A \underbrace{\begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}}_x.$$

With the times $t_0 = 0$, $t_1 = 1$, $t_2 = 2$ and $t_3 = 3$, and the position measurements $u(t_1 = 1) = 3$, $u(t_2 = 2) = 4$, $u(t_3 = 3) = 5$ system (16) becomes:

$$(17) \quad \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}.$$

In this case we have: $m = 3$, $n = 2$ and $\text{rank } A = 2$. Since $m = 3 > \text{rank } A = 2$, consistency of the system is not automatically guaranteed. But a closer look at the measurement vector of (17) shows that:

$$\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Thus the measurement vector y can be written as a linear combination of the column vectors of matrix A . Therefore $y \in R(A)$, showing that (17) is consistent. And since $n = \text{rank } A = 2$ its solution is also unique. If we partition (17) as:

$$\begin{pmatrix} 3 \\ 4 \\ \dots \\ 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \dots & \dots \\ 1 & 3 \end{pmatrix} \begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}$$

the unique solution follows as:

$$(18) \quad \begin{pmatrix} \hat{u}(t_0) \\ \hat{\dot{u}}(t_0) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

In this case the solution for $u(t_0), \dot{u}(t_0)$ is found from fitting the line $u(t) = u(t_0) + \dot{u}(t_0)(t - t_0)$ through the two points $(t_1 = 1, u(t_1) = 3)$ and $(t_2 = 2, u(t_2) = 4)$. See Figure 1.3a. We may of course partition (17) also as:

$$\begin{pmatrix} 3 \\ \dots \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \dots & \dots \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}$$

the unique solution follows then as:

$$(19) \quad \begin{pmatrix} \hat{u}(t_0) \\ \hat{\dot{u}}(t_0) \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

In this case the solution for $u(t_0), \dot{u}(t_0)$ is found from fitting the line $u(t) = u(t_0) + \dot{u}(t_0)(t - t_0)$ through the two points $(t_2 = 2, u(t_2) = 4)$ and $(t_3 = 3, u(t_3) = 5)$. See Figure 1.3b. Since the linear system (17) is consistent, which means that all three points $(t_1, u(t_1))$, $(t_2, u(t_2))$ and $(t_3, u(t_3))$ lie on the same line (see Figure 1.3c), the two solutions (18) and (19) are of course identical.

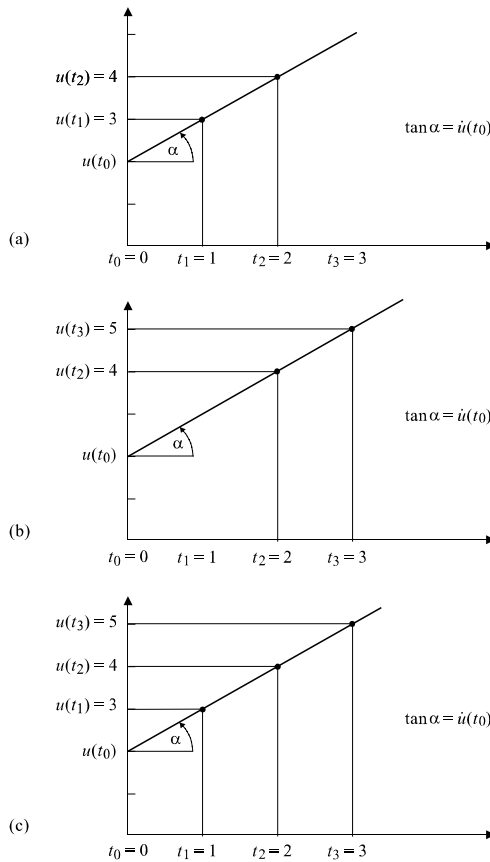


Figure 1.3: Fitting a straight line through consistent measurements.

Example 5

Consider the situation of Example 4. But now assume that the position measurements read:

$$u(t_1=1) = 3, \quad u(t_2=2) = 5, \quad u(t_3=3) = 6.$$

The linear system (16) then becomes:

$$(20) \quad \underbrace{\begin{pmatrix} 3 \\ 5 \\ 6 \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}}_A \underbrace{\begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}}_x.$$

In this case the measurement vector y cannot be written as a linear combination of the column vectors of matrix A . Hence $y \notin R(A)$, showing that the system is inconsistent. This inconsistency can clearly be seen from Figure 1.4a. Figure 1.4a shows clearly that no straight line

$u(t) = u(t_0) + \dot{u}(t_0)(t - t_0)$ exists that passes through all three measurement points. In order to find a solution one could disregard measurement $u(t_3 = 3) = 6$ and solve the system:

$$\begin{pmatrix} u(t_1) = 3 \\ u(t_2) = 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}.$$

The solution of this system reads (see Figure 1.4b):

$$(21) \quad \begin{pmatrix} \dot{u}(t_0) \\ \hat{u}(t_0) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

But instead of disregarding $u(t_3 = 3) = 6$ one could also opt for disregarding measurement $u(t_1 = 1) = 3$ or measurement $u(t_2 = 2) = 5$. In case of disregarding measurement $u(t_1 = 1) = 3$, one has to solve the system:

$$\begin{pmatrix} u(t_2) = 5 \\ u(t_3) = 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}.$$

The solution of this system reads (see Figure 1.4b):

$$(22) \quad \begin{pmatrix} \dot{u}(t_0) \\ \hat{u}(t_0) \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

This solution differs however from solution (21) (see also Figure 1.4b). So, which solution should we accept? The problem with the above approach is the arbitrariness in disregarding measurements. Why should we disregard measurement $u(t_3 = 3) = 6$ and completely rely on the measurements $u(t_1 = 1) = 3$ and $u(t_2 = 2) = 5$? It seems more appropriate to have a solution method which somehow takes all measurements into account. In case of the present example one could for instance think of computing the line $u(t) = u(t_0) + \dot{u}(t_0)(t - t_0)$ such that it fits all three measurement points as closely as possible (see Figure 1.4c). A method that accomplishes this task in a predefined way, is the *method of least-squares*. This method will be introduced in the next section.

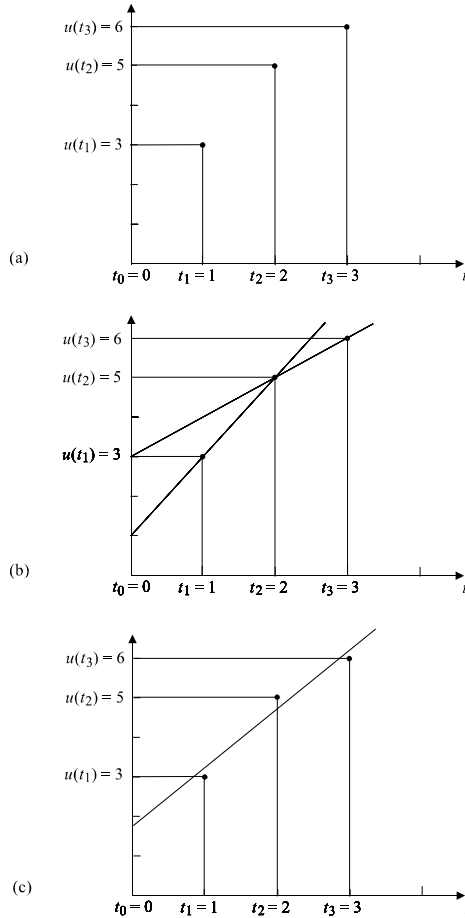


Figure 1.4: Fitting a straight line through inconsistent measurements.

1.1.2 Least-squares estimates

An inconsistent system, that is, a system for which $y \notin R(A)$ holds, can be made consistent by introducing an $m \times 1$ error vector e as (see Figure 1.5):

$$(23) \quad \underset{m \times 1}{y} = \underset{m \times n}{A} \underset{n \times 1}{x} + \underset{m \times 1}{e}, \quad m > n = \text{rank } A.$$

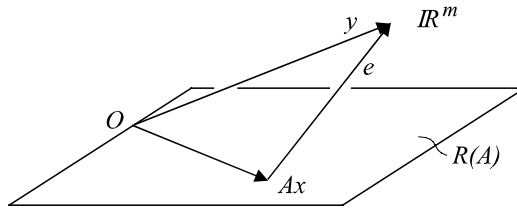


Figure 1.5: The geometry of $y = Ax + e$.

In (23), y and A are given, whereas x and e are unknown. From the geometry of Figure 1.5 it seems intuitively appealing to estimate x as \hat{x} such that $A\hat{x}$ is as close as possible to the given measurement- or observation vector y . In other words, the idea is to find that value of x that minimizes the length of the vector $e = y - Ax$. This idea leads to the following minimization problem:

$$(24) \quad \boxed{\underset{x}{\text{minimize}} (y - Ax)^*(y - Ax)} .$$

From calculus we know that \hat{x} is a solution of (24) if \hat{x} satisfies:

$$(25) \quad \frac{\partial E}{\partial x}(\hat{x}) = 0 \quad \text{and} \quad \frac{\partial^2 E}{\partial x^2}(\hat{x}) \text{ positive-definite}$$

where $E(x)$ is given as:

$$(26) \quad E(x) = (y - Ax)^*(y - Ax) = y^*y - 2y^*Ax + x^*A^*Ax .$$

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

$$(27) \quad \frac{\partial E}{\partial x}(x) = -2A^*y + 2A^*Ax \quad \text{and} \quad \frac{\partial^2 E}{\partial x^2}(x) = 2A^*A .$$

Equating the first equation of (27) to zero shows that \hat{x} satisfies the normal equations:

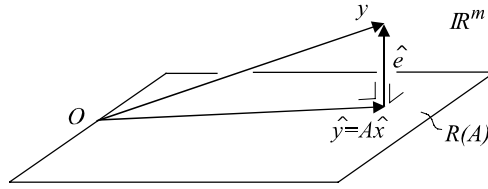
$$(28) \quad A^*A\hat{x} = A^*y .$$

Since $\text{rank } A^*A = \text{rank } A = n$, the system is consistent and has a unique solution. Through an inversion of the normal matrix A^*A the unique solution of (28) is found as:

$$(29) \quad \boxed{\hat{x} = (A^*A)^{-1}A^*y} .$$

That this solution \hat{x} is the minimizer of (26) follows from the fact that the matrix $\partial^2 E / \partial x^2$ of (27) is indeed positive-definite. The vector \hat{x} is known as the least-squares estimate of x , since it produces the smallest possible value of the sum-of-squares function $E(x)$. From the normal equations (28) it follows that $A^*(y - A\hat{x}) = 0$. This shows that the vector $\hat{e} = y - A\hat{x}$, which is the least-squares estimate of e , is orthogonal to the range space of matrix A (see Figure 1.6):

$$(30) \quad \boxed{A^*\hat{e} = 0, \text{ with } \hat{e} = y - A\hat{x}} .$$

Figure 1.6: The geometry of least-squares: $y = A\hat{x} + \hat{e}$.**Example 6**

Consider again the situation of a particle moving with constant velocity along a straight line. We assume that three observations of position are carried out at three different times t_1 , t_2 and t_3 : $u(t_1)$, $u(t_2)$ and $u(t_3)$. We also assume that the time instances are equidistant: $t_1 - t_0 = t_2 - t_1 = t_3 - t_2 = T$. The linear system reads then:

$$\begin{pmatrix} u(t_1) \\ u(t_2) \\ u(t_3) \end{pmatrix} = \begin{pmatrix} 1 & T \\ 1 & 2T \\ 1 & 3T \end{pmatrix} \begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}.$$

If the system is inconsistent, it can be made consistent by introducing an error vector e as:

$$(31) \quad \underbrace{\begin{pmatrix} u(t_1) \\ u(t_2) \\ u(t_3) \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 1 & T \\ 1 & 2T \\ 1 & 3T \end{pmatrix}}_A \underbrace{\begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}}_x + \underbrace{\begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}}_e.$$

The least-squares solution of (31) follows now from minimizing $e^*e = \sum_{i=1}^3 e_i^2$ as function of the parameters $u(t_0)$ and $\dot{u}(t_0)$. With

$$(A^*A)^{-1} = \begin{pmatrix} 3 & 6T \\ 6T & 14T^2 \end{pmatrix}^{-1} = \frac{1}{6T^2} \begin{pmatrix} 14T^2 & -6T \\ -6T & 3 \end{pmatrix}$$

and

$$A^*y = \begin{pmatrix} 1 & 1 & 1 \\ T & 2T & 3T \end{pmatrix} \begin{pmatrix} u(t_1) \\ u(t_2) \\ u(t_3) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^3 u(t_i) \\ T \sum_{i=1}^3 i u(t_i) \end{pmatrix}$$

the least-squares estimate of $x = (u(t_0), \dot{u}(t_0))^*$ follows as:

$$\begin{pmatrix} \hat{u}(t_0) \\ \dot{\hat{u}}(t_0) \end{pmatrix} = \frac{1}{6T^2} \begin{pmatrix} 14T^2 & -6T \\ -6T & 3 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^3 u(t_i) \\ T \sum_{i=1}^3 iu(t_i) \end{pmatrix} = \begin{pmatrix} \frac{1}{3}(4u(t_1) + u(t_2) - 2u(t_3)) \\ \frac{1}{2T}(u(t_3) - u(t_1)) \end{pmatrix}.$$

This may be rearranged to give:

$$(32) \quad \begin{cases} \hat{u}(t_0) = \frac{1}{3} \sum_{i=1}^3 (u(t_i) - \dot{\hat{u}}(t_0)iT) \\ \dot{\hat{u}}(t_0) = \frac{1}{2T} (u(t_3) - u(t_1)) \end{cases}.$$

This result shows that the slope of the straight line, $\dot{u}(t_0)$ (= velocity of particle), is determined from the two outer points $(t_1, u(t_1))$ and $(t_3, u(t_3))$, and that the intercept of the straight line, $u(t_0)$ (= initial position of particle), equals the average of $u(t_i) - \dot{\hat{u}}(t_0)iT$, $i = 1, 2, 3$.

So far we have discussed the unweighted least-squares principle. The least-squares principle can be generalized, however, by introducing a positive-definite $m \times m$ weight matrix W . This is done by replacing (24) by the following minimization problem:

$$(33) \quad \boxed{\underset{x}{\text{minimize}} (y - Ax)^* W (y - Ax)}.$$

The solution of (33) can be derived along lines which are similar to the ones used for solving (24). The solution of (33) reads:

$$(34) \quad \boxed{\hat{x} = (A^*WA)^{-1}A^*Wy}.$$

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

$$(35) \quad \boxed{A^*W\hat{e} = 0, \text{ with } \hat{e} = y - A\hat{x}}.$$

If the inner product of the observation space \mathbb{R}^m is defined as $(a,b) = a^*Wb$, $\forall a,b \in \mathbb{R}^m$, (35) can also be written as $(Ax, \hat{e}) = 0$, $\forall x \in \mathbb{R}^n$. This shows that also in the case of weighted least-squares, the vector \hat{e} can be considered to be orthogonal to the range space of A . A summary of the least-squares algorithm is given in Table 1.1.

<p>Inconsistent linear system</p> $y = Ax + e, \quad y, e \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \quad m > n = \text{rank } A$ <p>Weighted least-squares principle</p> <p>minimize $(y - Ax)^T W (y - Ax)$, $W =$ positive-definite</p> <p>Weighted least-squares estimates</p> <p>parameter vector : $\hat{x} = (A^T W A)^{-1} A^T W y$ observation vector : $\hat{y} = A \hat{x}$ error vector : $\hat{e} = y - \hat{y}$</p>

Table 1.1: Weighted least-squares.

Example 7

The elements of the weight matrix W can be chosen to emphasize (or de-emphasize) the influence of specific observations upon the estimate \hat{x} . In this way different levels of importance may be attached to the different observations. This is of importance if one believes that some observations are more trustworthy than other observations. For instance, some observations may be more trustworthy than others if they are obtained from more accurate measurement instruments. In order to illustrate the influence of the weight matrix, we consider a stationary particle with unknown position $u(t_0)$. We assume that two observations of position are carried out at times t_1 and t_2 . The linear system reads then:

$$(36) \quad \begin{pmatrix} u(t_1) \\ u(t_2) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} u(t_0) + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

A diagonal matrix is taken as weight matrix:

$$(37) \quad W = \begin{pmatrix} w_{11} & 0 \\ 0 & w_{22} \end{pmatrix}.$$

Then

$$\hat{u}(t_0) = (A^T W A)^{-1} A^T W y = \frac{\sum_{i=1}^2 w_{ii} u(t_i)}{\sum_{i=1}^2 w_{ii}}$$

or

$$(38) \quad \hat{u}(t_0) = \frac{u(t_1)}{1 + \frac{w_{22}}{w_{11}}} + \frac{u(t_2)}{1 + \frac{w_{11}}{w_{22}}}.$$

This shows that $\hat{u}(t_0)$ equals the average of $u(t_i)$, $i = 1, 2$, if $w_{11} = w_{22}$ (see Figure 1.7a). In this case both observations have the same influence on $\hat{u}(t_0)$. However, if $w_{22} < w_{11}$ then less weight is attached to the second observation and $\hat{u}(t_0)$ is closer to $u(t_1)$, see Figure 1.7b.

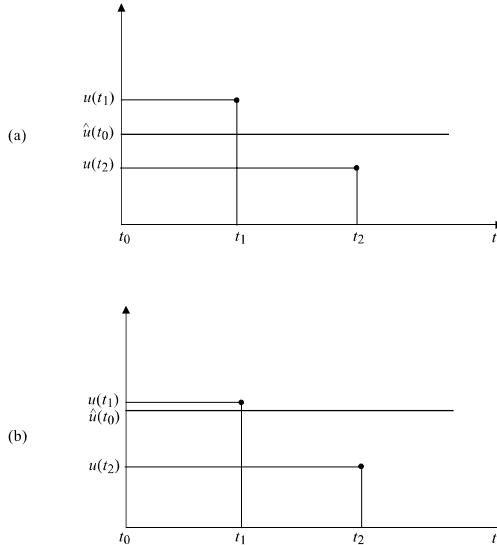


Figure 1.7: Weighted least-squares estimate of $u(t_0)$
 (a) $w_{11} = w_{22}$; (b) $w_{11} > w_{22}$.

1.1.3 A stochastic model for the observations

In the previous section the principle of least-squares was introduced. The least-squares principle enables us, in case of inconsistent systems, to obtain an intuitively appealing estimate \hat{x} of the parameter vector x . But although the least-squares estimate \hat{x} is intuitively appealing, no quality measures as yet can be attached to the estimate. That is, we know how to compute the estimate \hat{x} , but we are not able yet to say how good the estimate really is. Of course, the numerical value of the sum of squares, $\hat{e}^*W\hat{e}$, does indicate something about the quality of \hat{x} . If $\hat{e}^*W\hat{e}$ is small one is inclined to have more confidence in the estimate \hat{x} , than if $\hat{e}^*W\hat{e}$ is large. But how small is small? Besides, $\hat{e}^*W\hat{e}$ is identically zero if the linear system is consistent. Would this then automatically imply that the estimate \hat{x} has good quality? Not really, since the observations may still be subject to measurement errors. In order to obtain quality measures for the results of least-squares estimation, we start by introducing a qualitative description of the input, that is of the observations. This description will be of a probabilistic nature. The introduction of a probabilistic description is motivated by the experimental fact that the variability in the outcome of

measurements, when repeated under similar circumstances, can be described to a sufficient degree by stochastic or random variables. We will therefore assume that the observation vector y , which contains the numerical values of the measurements, constitutes a sample of the random vector of observables \underline{y} (note: the underscore indicates that we are dealing with a random variable). It is furthermore assumed that the vector of observables \underline{y} can be written as the sum of a deterministic functional part Ax and a random residual part \underline{e} :

$$(39) \quad \boxed{\underline{y} = Ax + \underline{e}} .$$

Although a random vector is completely described by its probability density function, we will restrict ourselves for the time being to the first two moments of random variables. That is, we will restrict ourselves to the mean and to the variance matrix. If we assume that \underline{e} models the probabilistic nature of the variability in the measurements, it seems acceptable to assume that this variability is zero *on the average* and therefore that the mean of \underline{e} is zero:

$$(40) \quad E\{\underline{e}\} = \mathbf{0}$$

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

$$(41) \quad D\{\underline{e}\} = Q_y$$

where $D\{\cdot\}$ stands for the dispersion operator. It is defined in terms of $E\{\cdot\}$ as $D\{\cdot\} = E\{(-E\{\cdot\})(-E\{\cdot\})^*\}$. With (40) and (41) we are now in the position to determine the mean and variance matrix of the vector of observables \underline{y} . Application of the law of propagation of means and the law of propagation of variances to (39) gives with (40) and (41):

$$(42) \quad \boxed{E\{\underline{y}\} = Ax \quad ; \quad D\{\underline{y}\} = Q_y} .$$

This will be our model of observation equations for the vector of observables \underline{y} . As the results of the next section show, model (42) enables us to describe the quality of the results of least-squares estimation in terms of the mean and the variance matrix.

1.1.4 Least-squares estimators

Functions of random variables are again random variables. It follows therefore, that if the vector of observables is assumed to be a random vector \underline{y} and substituted for y in the formulae of Table 1.1 in Section 1.1.2 the results are again random variables:

$$(43) \quad \begin{cases} \hat{x} = (A^*WA)^{-1}A^*W\underline{y} \\ \hat{y} = A\hat{x} \\ \hat{e} = \underline{y} - \hat{y} \end{cases} .$$

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

The first moment: the mean

Together with $E\{\underline{y}\} = Ax$, application of the propagation law of means to (43) gives:

$$(44) \quad \begin{cases} E\{\hat{\underline{x}}\} = x \\ E\{\hat{\underline{y}}\} = E\{\underline{y}\} \\ E\{\hat{\underline{e}}\} = E\{\underline{e}\} = \mathbf{0} . \end{cases}$$

This important result shows that under the assumption that (42) holds, the least-squares estimators are unbiased estimators. Note that this property of unbiasedness is independent of the choice for the weight matrix W .

The second moment: the variance matrix and covariance matrix

Together with $D\{\underline{y}\} = Q_y$, application of the propagation law of variances and covariances to (43) gives:

$$(45) \quad \begin{cases} Q_{\hat{\underline{x}}} = (A^*WA)^{-1}A^*WQ_yWA(A^*WA)^{-1} \\ Q_{\hat{\underline{y}}} = AQ_xA^* \\ Q_{\hat{\underline{e}}} = [I - A(A^*WA)^{-1}A^*W]Q_y[I - A(A^*WA)^{-1}A^*W]^* \end{cases}$$

and

$$(46) \quad \begin{cases} Q_{\hat{\underline{x}}\hat{\underline{y}}} = Q_xA^* \\ Q_{\hat{\underline{x}}\hat{\underline{e}}} = (A^*WA)^{-1}A^*WQ_y - Q_xA^* . \\ Q_{\hat{\underline{y}}\hat{\underline{e}}} = AQ_{\hat{\underline{x}}\hat{\underline{e}}} \end{cases}$$

The above variance matrices enable us now to give a complete precision description of any arbitrary linear function of the estimators. Consider for instance the linear function $\hat{\underline{\theta}} = a^*\hat{\underline{x}}$. Application of the propagation law of variances gives then for the precision of $\hat{\underline{\theta}}$: $\sigma_{\hat{\underline{\theta}}}^2 = a^*Q_{\hat{\underline{x}}}a$. The above results enable us to describe the quality of the results of least-squares estimation in terms of the mean and the variance matrix. The introduction of a stochastic model for the vector of observables \underline{y} enables us however also to judge the merits of the least-squares principle itself. Recall that the least-squares principle was introduced on the basis of intuition and not on the basis of probabilistic reasoning. With the mathematical model (42) one could now however try to develop an estimation procedure that produces estimators with certain well-defined probabilistic optimality properties. One such procedure is based on the principle of Best Linear Unbiased Estimation (BLUE), [Teunissen, 2000]. Assume that we are interested in estimating a parameter θ which is a linear function of x :

$$(47) \quad \begin{matrix} \theta & = & a^* & x & . \\ 1 \times 1 & & 1 \times m & m \times 1 & \end{matrix}$$

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

$$(48) \quad \hat{\theta} = l^* \mathbf{y}$$

$\begin{matrix} 1 \times 1 & & 1 \times m & m \times 1 \end{matrix}$

such that it is unbiased:

$$(49) \quad E\{\hat{\theta}\} = \theta$$

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

$$(50) \quad \sigma_{\hat{\theta}}^2 \rightarrow \text{minimum}.$$

The objective is thus to find a vector $l \in \mathbb{R}^m$ such that with (48), the conditions (49) and (50) are satisfied. From Adjustment theory, [Teunissen, 2000] we know that the solution to the above problem is given by:

$$l^* = \mathbf{a}^*(\mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{A})^{-1} \mathbf{A}^* \mathbf{Q}_y^{-1}.$$

If we substitute this into (48) we get:

$$(51) \quad \hat{\theta} = \mathbf{a}^*(\mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{A})^{-1} \mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{y}.$$

This is the best linear unbiased estimator of θ . The important result (51) shows that the best linear unbiased estimator of x is given by:

$$(52) \quad \hat{x} = (\mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{A})^{-1} \mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{y}.$$

A comparison between (43) and (52) shows that the BLUE of x is identical to the weighted least-squares estimator of x if the weight matrix W is taken to be equal to the inverse of the variance matrix of \mathbf{y} :

$$(53) \quad W = \mathbf{Q}_y^{-1}.$$

This is an important result, because it shows that the weighted least-squares estimators are best in the probabilistic sense of having minimal variance if (53) holds. The variances and covariances of these estimators follow if the weight matrix W is replaced in (45) and (46) by \mathbf{Q}_y^{-1} . From now on we will always assume, unless stated otherwise, that the weight matrix W is chosen to be equal to \mathbf{Q}_y^{-1} . Consequently no distinction will be made anymore in this book between weighted least-squares estimators and Best Linear Unbiased Estimators. Instead we will simply speak of least-squares estimators.

Example 8

Consider again the situation of a stationary particle. Assume that it is required to determine its position with a variance of $\sigma^2/10$. The position measurements are uncorrelated and all have a variance equal to σ^2 . How many position measurements are then needed in order to estimate the particle's position with sufficient precision?

In order to answer this question we first introduce our linear A -model:

$$E\left\{ \underbrace{\begin{pmatrix} \underline{u}(t_1) \\ \vdots \\ \underline{u}(t_m) \end{pmatrix}}_{\underline{y}} \right\} = \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{\mathbf{A}} \underbrace{u(t_0)}_x \quad ; \quad \mathbf{Q}_y = \sigma^2 \mathbf{I}_m$$

The variance of the least-squares estimators $\hat{u}(t_0)$ of $u(t_0)$ reads then:

$$\sigma_{\hat{u}(t_0)}^2 = (\mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{A})^{-1} = \sigma^2/m.$$

This result shows that $m = 10$ position measurements are needed to satisfy the requirements.

Example 9

Consider the situation of Example 2. The position observables are assumed to be uncorrelated and to have the same variance σ^2 . An interesting question is now how the times of measurement t_1 and t_2 should be chosen, in order to minimize the variances of the least-squares estimators of initial position $u(t_0)$ and initial velocity $\dot{u}(t_0)$. In order to answer this question we first introduce our linear A-model:

$$E\left\{ \underbrace{\begin{pmatrix} \underline{u}(t_1) \\ \underline{u}(t_2) \end{pmatrix}}_{\underline{y}} \right\} = \underbrace{\begin{pmatrix} 1 & (t_1 - t_0) \\ 1 & (t_2 - t_0) \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}}_x \quad ; \quad \mathbf{Q}_y = \sigma^2 \mathbf{I}_2$$

Note that this model consists of two equations in two unknowns. Hence the redundancy $m-n$ equals zero, and matrix \mathbf{A} is square and invertible (provided that $t_2 \neq t_1$). In this case the least-squares estimator and its variance matrix simply reduce to:

$$\begin{cases} \hat{\mathbf{x}} = (\mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{A})^{-1} \mathbf{A}^* \mathbf{Q}_y^{-1} \underline{y} = (\mathbf{A}^{-1}) (\mathbf{Q}_y^{-1})^{-1} (\mathbf{A}^*)^{-1} \mathbf{A}^* \mathbf{Q}_y^{-1} \underline{y} = \mathbf{A}^{-1} \underline{y} \\ \mathbf{Q}_{\hat{\mathbf{x}}} = (\mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{A})^{-1} = \mathbf{A}^{-1} \mathbf{Q}_y \mathbf{A}^{*-1} \end{cases}$$

With

$$\mathbf{A}^{-1} = \frac{1}{(t_2 - t_1)} \begin{pmatrix} (t_2 - t_0) & -(t_1 - t_0) \\ -1 & 1 \end{pmatrix}$$

this gives for the variance matrix:

$$(54) \quad \underline{Q}_x = \mathbf{A}^{-1} \underline{Q}_y \mathbf{A}^{* -1} = \frac{\sigma^2}{(t_2 - t_1)^2} \begin{pmatrix} \sum_{i=1}^2 (t_i - t_0)^2 & -\sum_{i=1}^2 (t_i - t_0) \\ -\sum_{i=1}^2 (t_i - t_0) & 2 \end{pmatrix}.$$

Let us now first consider the variance of the velocity estimator $\hat{u}(t_0)$. It reads:

$$\sigma_{\hat{u}(t_0)}^2 = 2\sigma^2 / (t_2 - t_1)^2.$$

This result shows that the variance of $\hat{u}(t_0)$ gets smaller, i.e. its precision gets better, if the time interval $t_2 - t_1$ gets larger. Thus one can improve the precision of the velocity estimator by increasing the time interval between the two position measurements. This is also quite understandable if one looks at Figure 1.2 in Section 1.1.1. A straight line can be fitted better through two points that are far apart than through two points that are close together. And in fact it becomes impossible to fit the line uniquely if the two points coincide, just like it is impossible to estimate the velocity if $t_2 = t_1$. Let us now consider the variance of the position estimator $\hat{u}(t_0)$. It reads (see (54)):

$$\sigma_{\hat{u}(t_0)}^2 = \sigma^2 \frac{\sum_{i=1}^2 (t_i - t_0)^2}{(t_2 - t_1)^2} = \sigma^2 \left[1 + \frac{2(t_1 - t_0)}{(t_2 - t_1)} + \frac{2(t_1 - t_0)^2}{(t_2 - t_1)^2} \right].$$

Also this result shows that the variance of the initial position estimator gets smaller if the time interval $t_2 - t_1$ gets larger. Also note that in case $t_1 \neq t_0$, the variance of $\hat{u}(t_0)$ is always larger than the variance of σ^2 of the position observables. The smallest value of $\sigma_{\hat{u}(t_0)}^2$ is obtained for $t_1 = t_0$. This shows that the precision of the initial position estimator is best if the first position measurement is taken at the initial time t_0 .

1.1.5 Summary

In Table 1.2 an overview is given of the main characteristics of least-squares estimation.

The linear A-model	
$E\{\underline{y}\} = \underset{m \times 1}{A} \underset{m \times n \ n \times 1}{x} \quad , \quad D\{\underline{y}\} = \underset{m \times m}{Q_y} \quad , \quad m \geq n = \text{rank } A$	
Least-squares estimators	
$\begin{aligned} \hat{x} &= (A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} y & \hat{e} &= y - \hat{y} \\ \hat{y} &= A \hat{x} & \hat{\theta} &= a^* \hat{x} \end{aligned}$	
Mean	
$\begin{aligned} E\{\hat{x}\} &= x & E\{\hat{e}\} &= E\{e\} = 0 \\ E\{\hat{y}\} &= E\{y\} = Ax & E\{\hat{\theta}\} &= \theta = a^* x \end{aligned}$	
Variances and covariances	
$\begin{aligned} Q_{\hat{x}} &= (A^* Q_y^{-1} A)^{-1} & Q_{\hat{e}} &= Q_y - Q_{\hat{y}} \\ Q_{\hat{y}} &= A Q_{\hat{x}} A^* & \sigma_{\hat{\theta}}^2 &= a^* Q_{\hat{x}} a \\ Q_{\hat{x}\hat{y}} &= Q_{\hat{x}} A^* , \quad Q_{\hat{x}\hat{e}} = 0 , \quad Q_{\hat{y}\hat{e}} = 0 \end{aligned}$	

Table 1.2: Least-squares estimation.

1.2 The nonlinear A-model

1.2.1 Nonlinear observation equations

Up to this point the development of our estimation theory was based on the assumption that the m -vector $E\{\underline{y}\}$ is linearly related to the n -vector of unknown parameters x . In geodetic applications there are however only a few cases where this assumption truly holds. A typical example is levelling. In the majority of applications, however, the m -vector $E\{\underline{y}\}$ is nonlinearly related to the n -vector of unknown parameters x . This implies that instead of the linear A-model (42), we are generally dealing with a nonlinear model of observation equations:

(55)

$$E\{\mathbf{y}\} = \mathbf{A}(x) ; D\{\mathbf{y}\} = \mathbf{Q}_y$$

where $\mathbf{A}(\cdot)$ is a nonlinear vector function from \mathbb{R}^n into \mathbb{R}^m . The following two simple examples will make this clear.

Example 10

Consider the configuration of Figure 1.8a. The x, y coordinates of the three points 1, 2 and 3 are known and the coordinates x_4 and y_4 of point 4 are unknown. The observables consist of the three azimuth variates \underline{a}_{14} , \underline{a}_{24} and \underline{a}_{34} . Since azimuth and coordinates are related as (see Figure 1.8b):

$$\tan a_{ij} = \frac{x_{ij}}{y_{ij}}$$

the model of observation equations for the configuration of Figure 1.8a reads:

$$E\left\{\begin{array}{c} \underline{a}_{14} \\ \underline{a}_{24} \\ \underline{a}_{34} \end{array}\right\} = \left\{\begin{array}{c} \arctan[x_{14}/y_{14}] \\ \arctan[x_{24}/y_{24}] \\ \arctan[x_{34}/y_{34}] \end{array}\right\}.$$

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

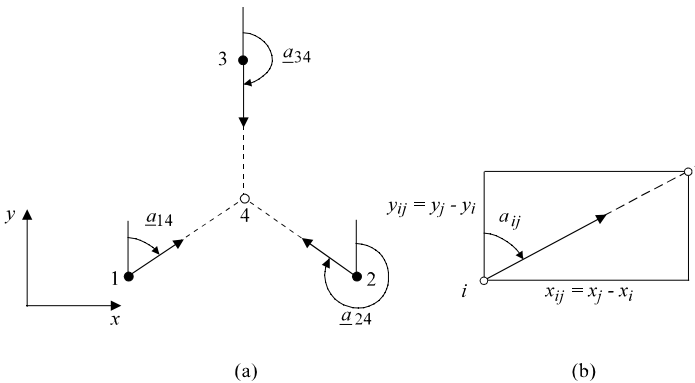


Figure 1.8: Azimuth resection.

Example 11

Consider the situation of Figure 1.9. It shows two cartesian coordinate systems: the x, y -system and the u, v -system. The two systems only differ in their orientation. This means that if the coordinates of a point i are given in the u, v -system, (u_i, v_i) , a rotation by an angle α is needed to obtain the coordinates of the same point i in the x, y -system, (x_i, y_i) :

$$(56) \quad \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix}.$$

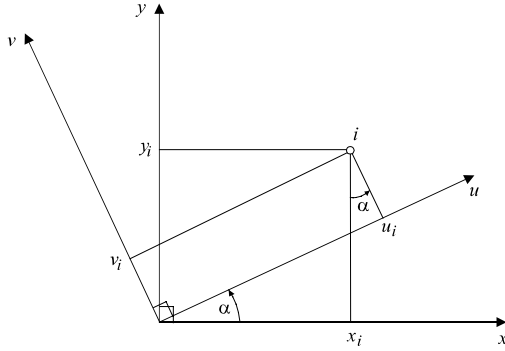


Figure 1.9: A coordinate transformation.

Let us now assume that we have at our disposal the coordinate observables of two points in both coordinate systems: $(\underline{x}_i, \underline{y}_i)$ and $(\underline{u}_i, \underline{v}_i)$, $i = 1, 2$. Using (56), our model reads then:

$$(57) \quad E \begin{pmatrix} \underline{x}_i \\ \underline{y}_i \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} E \begin{pmatrix} \underline{u}_i \\ \underline{v}_i \end{pmatrix}, \quad i = 1, 2.$$

This model is however still not in the form of observation equations. If we consider the orientation angle α and the coordinates of the two points in the u, v -system as the unknown parameters, (57) can be written in terms of observation equations as:

$$(58) \quad E \begin{pmatrix} \underline{x}_1 \\ \underline{y}_1 \\ \underline{x}_2 \\ \underline{y}_2 \\ \underline{u}_1 \\ \underline{v}_1 \\ \underline{u}_2 \\ \underline{v}_2 \end{pmatrix} = \begin{pmatrix} u_1 \cos\alpha & -v_1 \sin\alpha \\ u_1 \sin\alpha & +v_1 \cos\alpha \\ u_2 \cos\alpha & -v_2 \sin\alpha \\ u_2 \sin\alpha & +v_2 \cos\alpha \\ u_1 \\ v_1 \\ u_2 \\ v_2 \end{pmatrix}.$$

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

1.2.2 The linearized A-model

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

Taylor's Theorem

Let $f(x)$ be a function from \mathbb{R}^n into \mathbb{R} . Let $x_0 \in \mathbb{R}^n$ be an approximation to $x \in \mathbb{R}^n$ and define $\Delta x = x - x_0$, and $\theta = x_0 + t(x - x_0)$ with $t \in \mathbb{R}$. Then a scalar $t \in (0,1)$ exists such that:

$$(59) \quad \begin{aligned} f(x) = & f(x^0) + \sum_{\alpha=1}^n \partial_{\alpha} f(x^0) \Delta x_{\alpha} + \frac{1}{2} \sum_{\alpha=1}^n \sum_{\beta=1}^n \partial_{\alpha\beta}^2 f(x^0) \Delta x_{\alpha} \Delta x_{\beta} + \dots \\ & \dots + \frac{1}{(q-1)!} \sum_{\alpha_1=1}^n \dots \sum_{\alpha_{q-1}=1}^n \partial_{\alpha_1 \dots \alpha_{q-1}}^{q-1} f(x^0) \Delta x_{\alpha_1} \dots \Delta x_{\alpha_{q-1}} + R_q(\theta, \Delta x) \end{aligned}$$

with the remainder:

$$(60) \quad R_q(\theta, \Delta x) = \frac{1}{q!} \sum_{\alpha_1=1}^n \dots \sum_{\alpha_q=1}^n \partial_{\alpha_1 \dots \alpha_q}^q f(\theta) \Delta x_{\alpha_1} \dots \Delta x_{\alpha_q}.$$

In (59) and (60), $\partial_{\alpha_1 \dots \alpha_q}^q f(x)$ denotes the q th-order partial derivative of $f(x)$ evaluated at x . For the case $q = 2$, it follows from (59) and (60) that:

$$(61) \quad f(x) = f(x^0) + \sum_{\alpha=1}^n \partial_{\alpha} f(x^0) \Delta x_{\alpha} + \frac{1}{2} \sum_{\alpha=1}^n \sum_{\beta=1}^n \partial_{\alpha\beta}^2 f(\theta) \Delta x_{\alpha} \Delta x_{\beta}.$$

If we introduce the gradient vector and Hessian matrix of $f(x)$ respectively as:

$$\partial_x f(x) = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix} \quad \text{and} \quad \partial_{xx}^2 f(x) = \begin{pmatrix} \partial_{11}^2 f(x) & \dots & \partial_{1n}^2 f(x) \\ \vdots & & \vdots \\ \partial_{n1}^2 f(x) & \dots & \partial_{nn}^2 f(x) \end{pmatrix}$$

then equation (61) may be written in the more compact matrix-vector form as:

$$(62) \quad \boxed{f(x) = f(x^0) + \partial_x f(x^0)^* \Delta x + \frac{1}{2} \Delta x^* \partial_{xx}^2 f(\theta) \Delta x}.$$

This important result shows that a nonlinear function $f(x)$ can be written as a sum of three terms. The first term in this sum is the zero-order term $f(x^0)$. The zero-order term depends on x^0 but is independent of x . The second term in the sum is the first-order term $\partial_x f(x^0)^* \Delta x$. It depends on x^0 and is linearly dependent on x . Finally, the third term in the sum is the second-order remainder $R_2(\theta, \Delta x)$. An important consequence of Taylor's theorem is that the remainder $R_2(\theta, \Delta x)$ can be made arbitrarily small by choosing the approximation x^0 close enough to x . Now assume that

the approximation x^0 is chosen such that the second-order remainder can indeed be neglected. Then, instead of (62) we may write to a sufficient degree of approximation:

$$(63) \quad \boxed{f(x) = f(x^0) + \partial_x f(x^0) \Delta x} .$$

Hence, if x^0 is sufficiently close to x , the nonlinear function $f(x)$ can be approximated to a sufficient degree by the function $f(x^0) + \partial_x f(x^0) \Delta x$ which is linear in x . This function is the linearized version of $f(x)$. A geometric interpretation of this linearization is given in Figure 1.10 for the case $n = 1$.

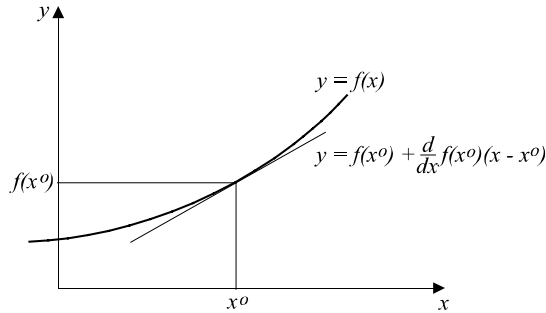


Figure 1.10: The nonlinear curve $y = f(x)$ and its linear tangent $y = f(x^0) + \frac{d}{dx}f(x^0)(x - x^0)$.

Let us now apply the above linearization to our nonlinear observation equations:

$$(64) \quad E\{\underline{y}\} = \mathbf{A}(x) = \begin{pmatrix} a_1(x) \\ \vdots \\ a_m(x) \end{pmatrix} .$$

Each nonlinear observation equation $a_i(x)$, $i = 1, \dots, m$, can now be linearized according to (63). This gives:

$$(65) \quad \begin{matrix} \begin{pmatrix} a_1(x) \\ \vdots \\ a_m(x) \end{pmatrix} \\ m \times 1 \end{matrix} = \begin{matrix} \begin{pmatrix} a_1(x^0) \\ \vdots \\ a_m(x^0) \end{pmatrix} \\ m \times 1 \end{matrix} + \begin{matrix} \begin{pmatrix} \partial_x a_1(x^0) \\ \vdots \\ \partial_x a_m(x^0) \end{pmatrix} \\ m \times n \end{matrix} \Delta x \quad n \times 1 .$$

If we denote the $m \times n$ matrix of (65) as $\partial_x \mathbf{A}(x^0)$, and substitute (65) into (64) we get:

$$E\{\underline{y}\} = \mathbf{A}(x^0) + \partial_x \mathbf{A}(x^0) \Delta x .$$

If we bring the constant m -vector $\mathbf{A}(x^0)$ to the left-hand side of the equation and define $\Delta \underline{y} = \underline{y} - \mathbf{A}(x^0)$, we finally obtain our linearized model of observation equations:

$$(66) \quad \boxed{E\{\Delta \underline{y}\} = \partial_x \mathbf{A}(x^0) \Delta x \quad ; \quad D\{\Delta \underline{y}\} = \mathbf{Q}_y} .$$

This is the linearized A-model. Compare (66) with (55) and (42). Note when comparing (66) with (42) that in the linearized A-model $\Delta \underline{y}$ takes the place of \underline{y} , $\partial_x \mathbf{A}(x^0)$ takes the place of \mathbf{A}

and Δx takes the place of x . Since the linearized A-model is linear in $\Delta x = x - x^0$ our standard formulae of least-squares can be applied again. This gives for the least-squares estimator $\hat{x} = x^0 + \Delta \hat{x}$ of x :

$$(67) \quad \hat{x} = x^0 + [\partial_x \mathbf{A}(x^0)^* \mathbf{Q}_y^{-1} \partial_x \mathbf{A}(x^0)]^{-1} \partial_x \mathbf{A}(x^0)^* \mathbf{Q}_y^{-1} \Delta \mathbf{y}.$$

Application of the propagation law of variances to (67) gives:

$$(68) \quad \mathbf{Q}_{\hat{x}} = [\partial_x \mathbf{A}(x^0)^* \mathbf{Q}_y^{-1} \partial_x \mathbf{A}(x^0)]^{-1}.$$

It will be clear that the above results, (67) and (68), are approximate in the sense that the second-order remainder is neglected. But these approximations are good enough if the second-order remainder can be neglected to a sufficient degree. In this case also the optimality conditions of least-squares (unbiasedness, minimal variance) hold to a sufficient degree. A summary of the linearized least-squares estimators is given in Table 1.3.

The nonlinear A-model	
$E\{\mathbf{y}\} = \mathbf{A}(x) \ ; \ D\{\mathbf{y}\} = \mathbf{Q}_y \ ; \ \mathbf{A}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m$	
The linearized A-model	
$E\{\Delta \mathbf{y}\} = \underbrace{\partial_x \mathbf{A}}_{m \times n} (x^0) \underbrace{\Delta x}_{n \times 1} \ ; \ D\{\Delta \mathbf{y}\} = \underbrace{\mathbf{Q}_y}_{m \times m} \ ; \ m \geq n = \text{rank } \underbrace{\partial_x \mathbf{A}(x^0)}_{m \times m}$	
Least-squares estimators	
$\begin{aligned} \hat{x} &= x^0 + [\partial_x \mathbf{A}(x^0)^* \mathbf{Q}_y^{-1} \partial_x \mathbf{A}(x^0)]^{-1} \partial_x \mathbf{A}(x^0)^* \mathbf{Q}_y^{-1} (\mathbf{y} - \mathbf{A}(x^0)) \\ \hat{\mathbf{y}} &= \mathbf{A}(\hat{x}) \\ \hat{\mathbf{e}} &= \mathbf{y} - \hat{\mathbf{y}} \end{aligned}$	
Variances	
$\begin{aligned} \mathbf{Q}_{\hat{x}} &= [\partial_x \mathbf{A}(x^0)^* \mathbf{Q}_y^{-1} \partial_x \mathbf{A}(x^0)]^{-1} \\ \mathbf{Q}_{\hat{\mathbf{y}}} &= \partial_x \mathbf{A}(x^0) \mathbf{Q}_{\hat{x}} \partial_x \mathbf{A}(x^0)^* \\ \mathbf{Q}_{\hat{\mathbf{e}}} &= \mathbf{Q}_y - \mathbf{Q}_{\hat{\mathbf{y}}} \end{aligned}$	

Table 1.3: Linearized least-squares estimation.

Example 12

Consider the configuration of Figure 1.11a. The x , y coordinates of the three points 1, 2 and 3 are known and the two coordinates x_4 and y_4 of point 4 are unknown. The observables consist of the three distance variates l_{14} , l_{24} and l_{34} . Since distance and coordinates are related as (see Figure 1.11b):

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

the model of observation equations for the configuration of Figure 1.11a reads:

$$(69) \quad E \begin{pmatrix} l_{14} \\ l_{24} \\ l_{34} \end{pmatrix} = \begin{pmatrix} (x_{14}^2 + y_{14}^2)^{\frac{1}{2}} \\ (x_{24}^2 + y_{24}^2)^{\frac{1}{2}} \\ (x_{34}^2 + y_{34}^2)^{\frac{1}{2}} \end{pmatrix}.$$

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

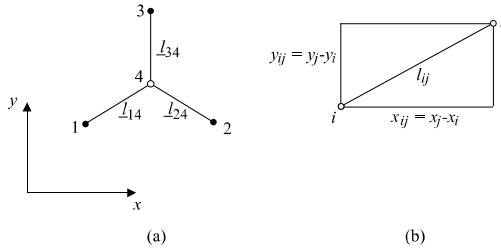


Figure 1.11: Distance resection.

In order to linearize (69) we need approximate values for the unknown coordinates x_4 and y_4 . These approximate values will be denoted as x_4^0 and y_4^0 . With these approximate values a linearization of (69) gives:

$$(70) \quad \underbrace{E \begin{pmatrix} \Delta l_{14} \\ \Delta l_{24} \\ \Delta l_{34} \end{pmatrix}}_{\Delta y} = \underbrace{\begin{pmatrix} (x_4^0 - x_1)/l_{14}^0 & (y_4^0 - y_1)/l_{14}^0 \\ (x_4^0 - x_2)/l_{24}^0 & (y_4^0 - y_2)/l_{24}^0 \\ (x_4^0 - x_3)/l_{34}^0 & (y_4^0 - y_3)/l_{34}^0 \end{pmatrix}}_{\partial_x A(x_0)} \underbrace{\begin{pmatrix} \Delta x_4 \\ \Delta y_4 \end{pmatrix}}_{\Delta x}$$

where:

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

Model (70) is the linearized version of the nonlinear A-model (69).

Example 13

Consider the nonlinear A-model (58) of Example 11. The unknown parameters are α and u_i, v_i for $i = 1, 2$. The approximate values of these parameters will be denoted as α^0 and u_i^0, v_i^0 for $i = 1, 2$. Linearization of (58) gives then:

$$(71) \quad \underbrace{\begin{pmatrix} \Delta x_1 \\ \Delta y_1 \\ \Delta x_2 \\ \Delta y_2 \\ \Delta u_1 \\ \Delta v_1 \\ \Delta u_2 \\ \Delta v_2 \end{pmatrix}}_{\Delta y} = \underbrace{\begin{pmatrix} -u_1^0 \sin \alpha^0 - v_1^0 \cos \alpha^0 & \cos \alpha^0 & -\sin \alpha^0 & 0 & 0 \\ u_1^0 \cos \alpha^0 - v_1^0 \sin \alpha^0 & \sin \alpha^0 & \cos \alpha^0 & 0 & 0 \\ -u_2^0 \sin \alpha^0 - v_2^0 \cos \alpha^0 & 0 & 0 & \cos \alpha^0 & -\sin \alpha^0 \\ u_2^0 \cos \alpha^0 - v_2^0 \sin \alpha^0 & 0 & 0 & \sin \alpha^0 & \cos \alpha^0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\partial_x A(x^0)} \underbrace{\begin{pmatrix} \Delta \alpha \\ \Delta u_1 \\ \Delta v_1 \\ \Delta u_2 \\ \Delta v_2 \end{pmatrix}}_{\Delta x}$$

where:

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

Example 14

Consider the situation of Figure 1.12. A satellite orbiting the earth is assumed to have a circular orbit with unknown radius R . Distance measurements from two known points 1 and 2 on the

earth surface are carried out to the two satellite positions 3 and 4. It is assumed that the earth is a non-rotating body.

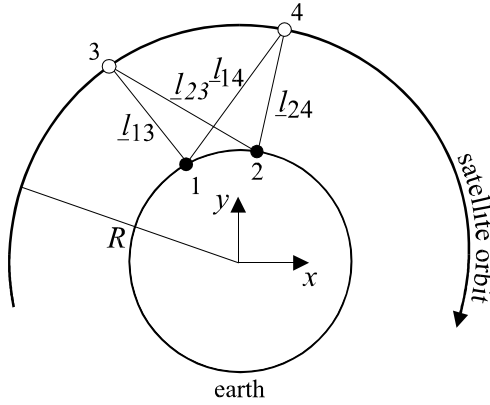


Figure 1.12: Distance measurement to a satellite orbiting the earth.

The distance l_{ij} between two points i and j can be parameterized in terms of cartesian coordinates as:

$$(72) \quad l_{ij} = (x_{ij}^2 + y_{ij}^2)^{\frac{1}{2}}.$$

The circular satellite orbit itself can be parameterized in terms of polar coordinates as:

$$(73) \quad \begin{cases} x_j = R \cos \phi_j \\ y_j = R \sin \phi_j \end{cases}.$$

With (72) and (73) the nonlinear model of observation equations becomes:

$$(74) \quad E \begin{pmatrix} l_{13} \\ l_{14} \\ l_{23} \\ l_{24} \end{pmatrix} = \begin{pmatrix} [(R \cos \phi_3 - x_1)^2 + (R \sin \phi_3 - y_1)^2]^{\frac{1}{2}} \\ [(R \cos \phi_4 - x_1)^2 + (R \sin \phi_4 - y_1)^2]^{\frac{1}{2}} \\ [(R \cos \phi_3 - x_2)^2 + (R \sin \phi_3 - y_2)^2]^{\frac{1}{2}} \\ [(R \cos \phi_4 - x_2)^2 + (R \sin \phi_4 - y_2)^2]^{\frac{1}{2}} \end{pmatrix}.$$

The unknowns in these observation equations are besides the orbital radius R also the coordinates ϕ_3 and ϕ_4 . The approximate values of these parameters are denoted as R^0 , ϕ_3^0 and ϕ_4^0 . Linearization of (74) gives then:

$$\underbrace{E \begin{pmatrix} \Delta l_{13} \\ \Delta l_{14} \\ \Delta l_{23} \\ \Delta l_{24} \end{pmatrix}}_{\Delta y} = \underbrace{\begin{pmatrix} \frac{R^0 - x_1 \cos \phi_3^0 - y_1 \sin \phi_3^0}{l_{13}^0} & \frac{R^0(x_1 \sin \phi_3^0 - y_1 \cos \phi_3^0)}{l_{13}^0} & 0 \\ \frac{R^0 - x_1 \cos \phi_4^0 - y_1 \sin \phi_4^0}{l_{14}^0} & 0 & \frac{R^0(x_1 \sin \phi_4^0 - y_1 \cos \phi_4^0)}{l_{14}^0} \\ \frac{R^0 - x_2 \cos \phi_3^0 - y_2 \sin \phi_3^0}{l_{23}^0} & \frac{R^0(x_2 \sin \phi_3^0 - y_2 \cos \phi_3^0)}{l_{23}^0} & 0 \\ \frac{R^0 - x_2 \cos \phi_4^0 - y_2 \sin \phi_4^0}{l_{24}^0} & 0 & \frac{R^0(x_2 \sin \phi_4^0 - y_2 \cos \phi_4^0)}{l_{24}^0} \end{pmatrix}}_{\partial_x \mathbf{A}(x^0)} \underbrace{\begin{pmatrix} \Delta R \\ \Delta \phi_3 \\ \Delta \phi_4 \end{pmatrix}}_{\Delta x}.$$

(75)

1.2.3 Least-squares iteration

Up to this point it was assumed that the second-order remainder was sufficiently small and that x_0 was a good enough approximation for x . If this is not the case, then \hat{x} as computed by (67) is not the least-squares estimate and hence an unacceptable error is made. In order to repair this situation, we need to improve upon the approximation x^0 . It seems reasonable to expect that the estimate:

$$x^1 = x^0 + [\partial_x \mathbf{A}(x^0)^* \mathbf{Q}_y^{-1} \partial_x \mathbf{A}(x^0)]^{-1} \partial_x \mathbf{A}(x^0)^* \mathbf{Q}_y^{-1} (y - \mathbf{A}(x^0))$$

is a better approximation than x^0 . That is, it seems reasonable to expect that x^1 is closer to the true least-squares estimate than x^0 . In fact one can show that this is indeed the case for most practical applications. But if x^1 is a better approximation than x^0 , a further improvement can be expected if we replace x^0 by x^1 in the linearization of the nonlinear model. The recomputed linearized least-squares estimate reads then:

$$x^2 = x^1 + [\partial_x \mathbf{A}(x^1)^* \mathbf{Q}_y^{-1} \partial_x \mathbf{A}(x^1)]^{-1} \partial_x \mathbf{A}(x^1)^* \mathbf{Q}_y^{-1} (y - \mathbf{A}(x^1)).$$

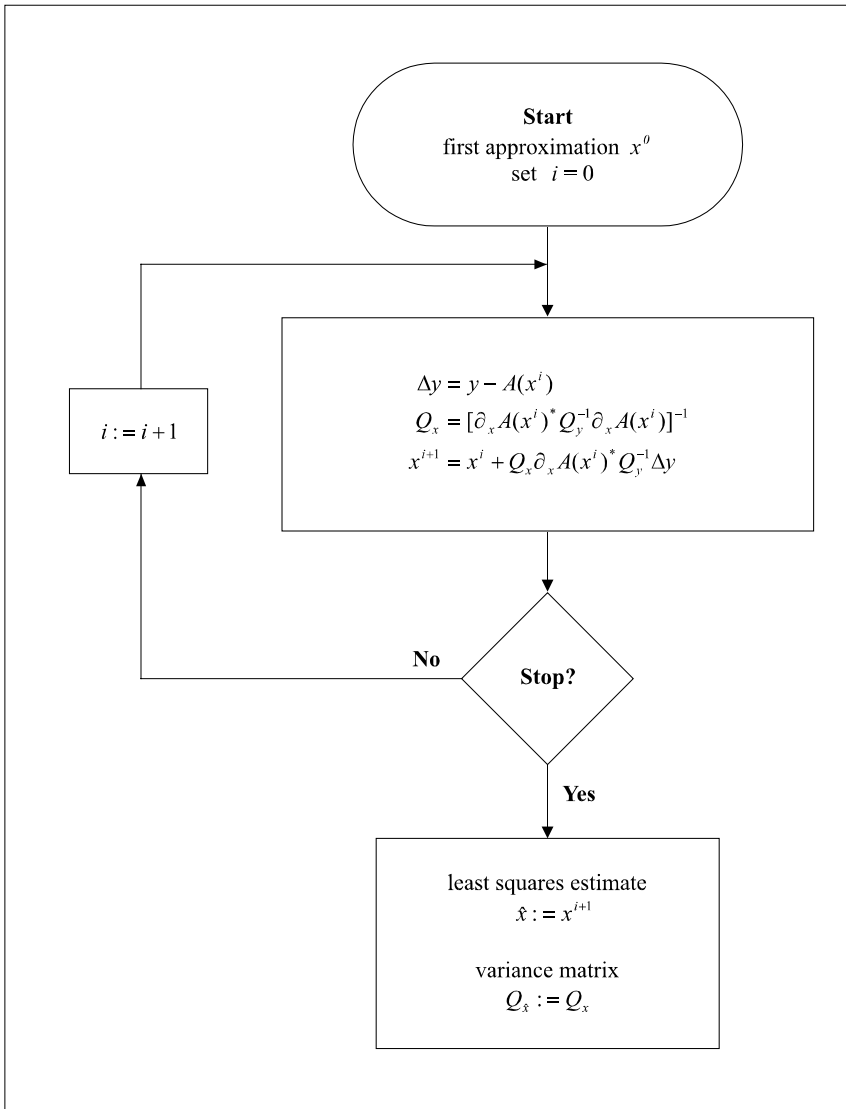


Table 1.4: Least-squares iteration.

By repeating this process a number of times, one can expect that finally the solution converges to the actual least-squares estimate \hat{x} . This is called the least-squares iteration process. The iteration is usually terminated if the difference between successive solutions is negligible. A flow diagram of the least-squares iteration process is shown in Table 1.4. For more details on the numerical properties of the iteration process and on the probabilistic properties of nonlinear least-squares estimators the reader is referred to the theory as developed in [Teunissen, 1985b].

1.3 The B-model

1.3.1 The linear B-model

In the previous sections we considered the model of observation equations. In this section and the next we briefly review the model of condition equations. For more details the reader is again referred to Adjustment theory, [Teunissen, 2000]. As our starting point we take the linear A-model:

$$(76) \quad E\{ \underset{m \times 1}{\underline{y}} \} = \underset{m \times m \times 1}{\mathbf{A}} \underset{m \times 1}{x} \quad ; \quad D\{ \underset{m \times m}{\underline{y}} \} = \underset{m \times m}{\mathbf{Q}_y} \quad ; \quad m \geq n = \text{rank } \mathbf{A} .$$

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

$$(77) \quad \boxed{\mathbf{b} = m - n} .$$

We will now show how one can construct the condition equations, given the linear A-model (76). Each of the column vectors of matrix A is an element of the observation space \mathbb{R}^m . Together the n -number of linearly independent column vectors of A span the range space of A . This range space has dimension n and it is a linear subspace of \mathbb{R}^m : $R(A) \subset \mathbb{R}^m$. Since $\dim R(A) = n$ and $\dim \mathbb{R}^m = m$, exactly $(m - n)$ -number of linearly independent vectors can be found that are orthogonal to $R(A)$. Let us denote these vectors as: $b_i \in \mathbb{R}^m$, $i = 1, \dots, (m - n)$. Then:

$$b_i \perp R(A) \text{ or } \mathbf{A}^* b_i = 0, \quad i = 1, \dots, (m - n) .$$

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

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

that

$$(78) \quad \boxed{\underset{(m - n) \times m \times n}{\mathbf{B}^*} \underset{(m - n) \times n}{\mathbf{A}} = \underset{(m - n) \times n}{\mathbf{0}} \quad ; \quad \text{rank } \mathbf{B} = m - n} .$$

This result may now be used to obtain the model of condition equations from (76). Premultiplication of the linear system of observation equations in (76) by B^* gives together with (78) the following linear model of condition equations:

$$(79) \quad \boxed{\underset{b \times m}{\mathbf{B}^*} E\{ \underset{m \times 1}{\underline{y}} \} = \underset{b \times 1}{\mathbf{0}} \quad ; \quad D\{ \underset{m \times m}{\underline{y}} \} = \underset{m \times m}{\mathbf{Q}_y} \quad ; \quad \text{rank } \mathbf{B} = \mathbf{b} = m - n} .$$

Example 15

Consider the following linear A-model:

$$(80) \quad E \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} x \quad ; \quad D\{y\} = Q_y.$$

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

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

are linearly independent and are both orthogonal to the single column vector of matrix A in (80). Hence the with (80) corresponding linear model of condition equations reads:

$$(81) \quad \underbrace{\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}}_{B^*} \underbrace{E \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{D\{y\}} = Q_y.$$

Example 16

Consider the linear A-model of Example 6 in Section 1.1.2:

$$(82) \quad \underbrace{E \begin{pmatrix} u(t_1) \\ u(t_2) \\ u(t_3) \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 1 & T \\ 1 & 2T \\ 1 & 3T \end{pmatrix}}_A \underbrace{\begin{pmatrix} u(t_0) \\ \dot{u}(t_0) \end{pmatrix}}_x.$$

Since $m = 3$, $n = 2$ and $\text{rank } A = 2 = n$, the redundancy equals $m-n = 1$. Hence, only one condition equations can be formulated. The with (82) corresponding linear B-model reads:

$$(83) \quad \underbrace{(1 \quad -2 \quad 1)}_{\mathbf{B}^*} \quad E \quad \underbrace{\begin{pmatrix} \mathbf{u}(t_1) \\ \mathbf{u}(t_2) \\ \mathbf{u}(t_3) \end{pmatrix}}_{\mathbf{y}} = \mathbf{0} .$$

Verify that $\mathbf{B}^* \mathbf{A} = \mathbf{0}$ holds.

Now that we have the linear B-model (79) at our disposal, how are we going to compute the corresponding least-squares estimators? We know how to compute the least-squares estimators for the linear A-model. The corresponding formulae are however all expressed in terms of the A-matrix. What is needed therefore is to transform these formulae such that they are expressed in terms of the B-matrix. This is possible with the following important matrix identity:

$$(84) \quad \boxed{\mathbf{A}(\mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{A})^{-1} \mathbf{A}^* \mathbf{Q}_y^{-1} = \mathbf{I} - \mathbf{Q}_y \mathbf{B}(\mathbf{B}^* \mathbf{Q}_y \mathbf{B})^{-1} \mathbf{B}^*} .$$

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

$$(85) \quad \mathbf{C} \triangleq (\mathbf{A} : \mathbf{Q}_y \mathbf{B}) \quad \text{and} \quad \bar{\mathbf{C}} \triangleq \begin{pmatrix} (\mathbf{A}^* \mathbf{Q}_y^{-1} \mathbf{A})^{-1} \mathbf{A}^* \mathbf{Q}_y^{-1} \\ \dots\dots\dots \\ (\mathbf{B}^* \mathbf{Q}_y \mathbf{B})^{-1} \mathbf{B}^* \end{pmatrix} .$$

Since both matrices \mathbf{C} and $\bar{\mathbf{C}}$ are of dimension $m \times m$ and since both can be shown to be of full rank, it follows that they are invertible. From (85) it follows with the help of (78) that $\bar{\mathbf{C}} \mathbf{C} = \mathbf{I}_m$. Hence $\bar{\mathbf{C}} = \mathbf{C}^{-1}$ and therefore $\mathbf{C} \bar{\mathbf{C}} = \mathbf{I}_m$. Substitution of (85) into this last expression proves (84). With (84) and the least-squares results of Table 1.2 of Section 1.1.5 we are now in the position to derive the expressions for the least-squares estimators in terms of the matrix \mathbf{B} . The results are summarized in Table 1.5.

<p>The linear B-model</p> <div style="border: 1px solid black; padding: 5px; margin: 10px auto; width: 80%;"> $\mathbf{B}^* \mathbf{E}\{\mathbf{y}\} = \mathbf{0} \quad ; \quad \mathbf{D}\{\mathbf{y}\} = \mathbf{Q}_y \quad ; \quad m \geq \text{rank } \mathbf{B} = b$ $\begin{matrix} b \times m & m \times 1 & b \times 1 & m \times m & m \times m \end{matrix}$ </div> <p style="text-align: center; margin-top: 20px;">Least-squares estimators</p> <div style="border: 1px solid black; padding: 5px; margin: 10px auto; width: 80%;"> $\hat{\mathbf{y}} = [\mathbf{I} - \mathbf{Q}_y \mathbf{B} (\mathbf{B}^* \mathbf{Q}_y \mathbf{B})^{-1} \mathbf{B}^*] \mathbf{y} \quad ; \quad \hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$ </div> <p style="text-align: center; margin-top: 20px;">Variances and covariances</p> <div style="border: 1px solid black; padding: 5px; margin: 10px auto; width: 80%;"> $\mathbf{Q}_{\hat{\mathbf{y}}} = \mathbf{Q}_y - \mathbf{Q}_y \mathbf{B} (\mathbf{B}^* \mathbf{Q}_y \mathbf{B})^{-1} \mathbf{B}^* \mathbf{Q}_y \quad ; \quad \mathbf{Q}_{\hat{\mathbf{e}}} = \mathbf{Q}_y - \mathbf{Q}_{\hat{\mathbf{y}}}$ $\mathbf{Q}_{\hat{\mathbf{y}} \hat{\mathbf{e}}} = \mathbf{0}$ </div>

Table 1.5: Least-squares estimation.

1.3.2 The nonlinear B-model

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

(86)
 $\mathbf{B}^*(\mathbf{E}\{\mathbf{y}\}) = \mathbf{0} \quad ; \quad \mathbf{D}\{\mathbf{y}\} = \mathbf{Q}_y$
 .

Where $\mathbf{B}^*(.)$ is a nonlinear vector function from \mathbb{R}^n into \mathbb{R}^{m-n} . The relationship between the nonlinear B-model and the nonlinear A-model is given by:

(87)
 $\mathbf{B}^*(\mathbf{A}(x)) = \mathbf{0} \quad ; \quad \forall x \in \mathbb{R}^n$
 .

This is the nonlinear generalization of (78). If we take the partial derivative with respect to x of (87) and apply the chain rule, we get:

(88)
 $[\partial_y \mathbf{B}(y^0)]^* [\partial_x \mathbf{A}(x^0)] = \mathbf{0} \quad ; \quad y^0 = \mathbf{A}(x^0)$
 .

This is the linearized version of (87). Compare (88) with (78). With (88) we are now in the position to construct the linearized B-model from the linearized A-model (66). Premultiplication of (66) with the matrix $[\partial_y \mathbf{B}(y^0)]^*$ gives together with (88) the result:

(89)

$$\boxed{[\partial_y \mathbf{B}(y^0)]^* E(\Delta \mathbf{y}) = \mathbf{0} \ ; \ D(\Delta \mathbf{y}) = \mathbf{Q}_y} .$$

This is the linearized B-model. With (89) we are now in the position again to apply our standard least-squares estimation formulae.

Example 17

The cartesian coordinates of three points 1, 2 and 3 are measured. The observables are therefore: $\underline{x}_1, \underline{y}_1, \underline{x}_2, \underline{y}_2, \underline{x}_3$ and \underline{y}_3 . The three points are assumed to lie on a circle with unknown radius R , see Figure 1.13. Since the circle can be parameterized as:

$$\begin{cases} x = R \cos \phi \\ y = R \sin \phi \end{cases}$$

the nonlinear A-model reads:

$$(90) \quad E \left(\begin{array}{c} \underline{x}_1 \\ \underline{y}_1 \\ \underline{x}_2 \\ \underline{y}_2 \\ \underline{x}_3 \\ \underline{y}_3 \end{array} \right) = \begin{pmatrix} R \cos \phi_1 \\ R \sin \phi_1 \\ R \cos \phi_2 \\ R \sin \phi_2 \\ R \cos \phi_3 \\ R \sin \phi_3 \end{pmatrix} .$$

This model consists of six nonlinear observation equations in the four unknown parameters R, ϕ_1, ϕ_2 and ϕ_3 . The approximate values of these parameters are denoted as R^0, ϕ_1^0, ϕ_2^0 and ϕ_3^0 . Linearization of (90) gives:

$$(91) \quad \underbrace{\begin{pmatrix} \Delta x_1 \\ \Delta y_1 \\ \Delta x_2 \\ \Delta y_2 \\ \Delta x_3 \\ \Delta y_3 \end{pmatrix}}_{\Delta y} = \underbrace{\begin{pmatrix} \cos\phi_1^0 & -R^0 \sin\phi_1^0 & 0 & 0 \\ \sin\phi_1^0 & R^0 \cos\phi_1^0 & 0 & 0 \\ \cos\phi_2^0 & 0 & -R^0 \sin\phi_2^0 & 0 \\ \sin\phi_2^0 & 0 & R^0 \cos\phi_2^0 & 0 \\ \cos\phi_3^0 & 0 & 0 & -R^0 \sin\phi_3^0 \\ \sin\phi_3^0 & 0 & 0 & R^0 \cos\phi_3^0 \end{pmatrix}}_{\partial_x A(x^0)} \underbrace{\begin{pmatrix} \Delta R \\ \Delta\phi_1 \\ \Delta\phi_2 \\ \Delta\phi_3 \end{pmatrix}}_{\Delta x}$$

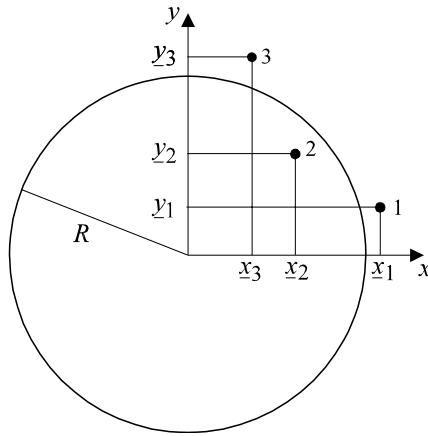


Figure 1.13: Circle with radius R .

Instead of parameterizing the circle, one can alternatively describe the circle implicitly as:

$$x^2 + y^2 = R^2.$$

This description leads to the following nonlinear model of condition equations:

$$(92) \quad \begin{pmatrix} (E\{x_1\}^2 + E\{y_1\}^2) - (E\{x_2\}^2 + E\{y_2\}^2) \\ (E\{x_1\}^2 + E\{y_1\}^2) - (E\{x_3\}^2 + E\{y_3\}^2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The number of independent condition equations equals the redundancy, which is equal to $6 - 4 = 2$. If the approximate values are chosen such that:

$$(x_1^0)^2 + (y_1^0)^2 = (x_2^0)^2 + (y_2^0)^2 = (x_3^0)^2 + (y_3^0)^2$$

linearization of (92) gives:

$$(93) \quad \underbrace{\begin{pmatrix} 2x_1^0 & 2y_1^0 & -2x_2^0 & -2y_2^0 & 0 & 0 \\ 2x_1^0 & 2y_1^0 & 0 & 0 & -2x_3^0 & -2y_3^0 \end{pmatrix}}_{\partial_y B(y^0)^*} E \left\{ \underbrace{\begin{pmatrix} \Delta x_1 \\ \Delta y_1 \\ \Delta x_2 \\ \Delta y_2 \\ \Delta x_3 \\ \Delta y_3 \end{pmatrix}}_{\Delta y} \right\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Verify yourself that $[\partial_y B(y^0)]^* [\partial_x A(x^0)] = 0$ holds for $y^0 = A(x^0)$.
