Adjustment theory

An introduction

Teunissen, P.J.G.

DOI
10.59490/tb.95

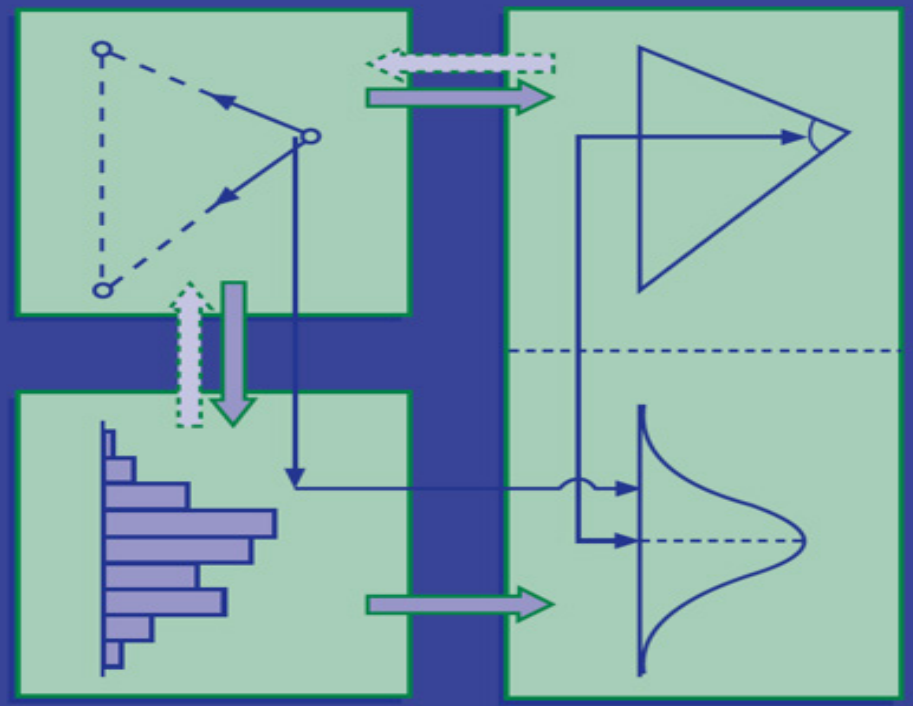
Publication date
2024

Document Version
Final published version

Citation (APA)
Teunissen, P. J. G. (2024). *Adjustment theory: An introduction*. TU Delft OPEN Publishing. https://doi.org/10.59490/tb.95

Adjustment theory: an introduction

Peter J.G. Teunissen



TU Delft OPEN

Adjustment theory

an introduction

Adjustment theory

an introduction

P.J.G.Teunissen

Delft University of Technology
Faculty of Civil Engineering and Geosciences
Department of Mathematical Geodesy and Positioning

Series on Mathematical Geodesy and Positioning

ISBN: 978-94-6366-884-2 (Ebook)
ISBN: 978-94-6366-885-9 (Paperback/softback)
Doi: https://doi.org/10.59490/tb.95

Keywords: adjustment theory, geodesy

Preface 2nd Edition

To promote open access, this new edition of Adjustment Theory is published by TU Delft Open Publishing instead of Delft Academic Press.
Appendix G has been added to this edition. It places the origin of Adjustment Theory in its historical context and provides a brief account of the early methods of adjusting geodetic and astronomical observations.

May, 2024
Peter J.G. Teunissen

Preface

As in other physical sciences, empirical data are used in Geodesy and Earth Observation to make inferences so as to describe the physical reality. Many such problems involve the determination of a number of unknown parameters which bear a linear (or linearized) relationship to the set of data. In order to be able to check for errors or to reduce for the effect these errors have on the final result, the collected data often exceed the minimum necessary for a unique determination (redundant data). As a consequence of (measurement) uncertainty, redundant data are usually inconsistent in the sense that each sufficient subset yields different results 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 data consistent such that the unknown parameters can be determined uniquely, is referred to as *adjustment*. This book presents the material for an introductory course on Adjustment theory. The main goal of this course is to convey the knowledge necessary to perform an optimal adjustment of redundant data. Knowledge of linear algebra, statistics and calculus at the undergraduate level is required. Some prerequisites, repeatedly needed in the book, are summarized in the Appendices A-F.
Following the *Introduction*, the elementary adjustment principles of 'least-squares', unbiased minimum variance' and 'maximum likelihood' are discussed and compared in *Chapter 1*. These results are generalized to the multivariate case in chapters 2 and 3. In *Chapter 2* the model of observation equations is presented, while the model of condition equations is discussed in *Chapter 3*. Observation equations and condition equations are dual to each other in the sense that the first gives a parametric representation of the model, while the second gives an implicit representation. Observables that are either stochastically or functionally related to another set of observables are referred to as y^R-variates. Their adjustment is discussed in *Chapter 4*. In *Chapter 5* we consider mixed model representations and in *Chapter 6* the partitioned models. Models may be partitioned in different ways: partitioning of the observation equations, partitioning of the unknown parameters, partitioning of the condition equations, or all of the above. These different partitionings are discussed and their relevance is shown for various applications. Up tot this point all the material is presented on the basis of the assumption that the data are linearly related. In geodetic applications however, there are only a few cases where this assumption holds true. In the last chapter, *Chapter 7*, we therefore start extending the theory from linear to nonlinear situations. In particular an introduction to the linearization and iteration of nonlinear models is given.

Many colleagues of the Department of Mathematical Geodesy and Positioning whose assistance made the completion of this book possible are greatly acknowledged. C.C.J.M. Tiberius took care of the editing, while the typing was done by Mrs. J. van der Bijl and Mrs. M.P.M. Scholtes, and the drawings by Mr. A. Smit. Various lecturers have taught the book's material over the past years. In particular the feedback and valuable recommendations of the lecturers C.C.J.M. Tiberius, F. Kenselaar, G.J. Husti and H.M. de Heus are acknowledged.

P.J.G. Teunissen
April, 2003

Contents

Introduction

The *calculus of observations* (in Dutch: *waarnemingsrekening*) can be regarded as the part of mathematical statistics that deals with the results of measurement. Mathematical statistics is often characterized by saying that it deals with the making of decisions in uncertain situations. In first instance this does not seem to relate to measurements. After all, we perform measurements in order to obtain sharp, objective and unambiguous quantitative information, information that is both correct and precise, in other words accurate information. Ideally we would like to have 'mathematical certainty'. But mathematical certainty pertains to statements like:

$$\text{if } a=b \text{ and } b=c \text{ and } c=d, \text{ then } a=d.$$

If a, b, c, and d are measured terrain distances however, a surveyor to be sure will first compare d with a before stating that $a=d$. Mathematics is a product of our mind, whereas with measurements one is dealing with humans, instruments, and other materialistic things and circumstances, which not always follow the laws of logic. From experience we know that measurements produce results that are only sharp to a certain level. There is a certain vagueness attached to them and sometimes even, the results are simply wrong and not usable.
In order to eliminate uncertainty, it seems obvious that one seeks to improve the instruments, to educate people better and to try to control the circumstances better. This has been successfully pursued throughout the ages, which becomes clear when one considers the very successful development of geodetic instruments since 1600. However, there are still various reasons why measurements will always remain uncertain to some level:

1. 'Mathematical certainty' relates to abstractions and not to physical, technical or social matters. A mathematical model can often give a useful description of things and relations known from our world of experience, but still there will always remain an essential difference between a model and the real world.
2. The circumstances of measurement can be controlled to some level in the laboratory, but in nature they are uncontrollable and only partly descriptive.
3. When using improved methods and instrumentation, one still, because of the things mentioned under 1 and 2, will have uncertainty in the results of measurement, be it at another level.
4. Measurements are done with a purpose. Dependent on that purpose, some uncertainty is quite acceptable. Since better instruments and methods usually cost more, one will have to ask oneself the question whether the reduction in uncertainty is worth the extra costs.

Summarizing, it is fundamentally and practically impossible to obtain absolute certainty from measurements, while at the same time this is also often not needed on practical grounds. Measurements are executed to obtain quantitative information, information which is accurate enough for a specific purpose and at the same time cheap enough for that purpose. It is because of this intrinsic uncertainty in measurements, that we have to consider the calculus of observations. The calculus of observations deals with:

1. The description and analysis of measurement processes.
2. Computational methods that take care, in some sense, of the uncertainty in measurements.
3. The description of the quality of measurement and of the results derived therefrom.
4. Guidelines for the design of measurement set-ups, so as to reach optimal working procedures.

The calculus of observations is essential for a geodesist, its meaning comparable to what mechanics means for the civil-engineer or mechanical-engineer. In order to shed some more light on the type of problems the calculus of observations deals with, we take some surveying examples as illustration. Some experience in surveying will for instance already lead to questions such as:

1. Repeating a measurement does usually not give the same answer. How to describe this phenomenon?
2. Results of measurement can be corrupted by systematic or gross errors. Can these errors be traced and if so, how?
3. Geometric figures (e.g. a triangle) are, because of 2, usually measured with *redundant* (*overtallig*) observations (e.g. instead of two, all three angles of a triangle are measured). These redundant observations will usually not obey the mathematical rules that hold true for these geometric figures (e.g. the three angular measurements will usually not sum up to π). An *adjustment* (*vereffening*) is therefore needed to obtain consistent results again. What is the best way to perform such an adjustment?
4. How to describe the effect of 1 on the final result? Is it possible that the final result is still corrupted with non-detected gross errors?

Once we are able to answer questions like these, we will be able to determine the required measurement set-up from the given desired quality of the end product. In this book we mainly will deal with points 1 and 3, the elementary aspects of adjustment theory.

Functional and stochastic model

In order to analyze scientific or practical problems and/or make them accessible for a quantitative treatment, one usually tries to rid the problem from its unnecessary frills and describes its essential aspects by means of notions from an applicable mathematical theory. This is referred to as the making of a *mathematical model*. On the one hand the model should give a proper description of the situation, while, on the other hand, the model should not be needlessly complicated and difficult to apply. The determination of a proper model can be considered the 'art' of the discipline. Well-tried and well-tested experience, new experience from experiments, intuition and creativity, they all play a role in formulating the model.
Here we will restrict ourselves to an example from surveying. Assume that we want to determine the relative position of terrain-'points', whereby their differences in height are of no interest to us. If the distances between the 'points' are not too large, we are allowed, as experience tells us, to describe the situation by projecting the 'points' onto a level surface (niveauvlak) and by treating this surface as if it was a plane. For the description of the relative position of the points, we are then allowed to make use of the two-dimensional Euclidean geometry (vlakke

meetkunde), the theory of which goes back some two thousand years. This theory came into being as an abstraction from how our world was seen and experienced. But for us the theory is readily available, thus making it possible to anticipate on its use, for instance by using sharp and clear markings for the terrain-'points' such that they can be treated as if they were mathematically defined points. In this example, the two-dimensional Euclidean geometry acts as our *functional model*. That is, notions and rules from this theory (such as the sine- and cosine rule) are used to describe the relative position of the terrain-'points'.

Although the two-dimensional Euclidean geometry serves its purpose well for many surveying tasks, it is pointed out that the gratuitously use of a 'standard' model can be full of risks. Just like the laws of nature (natuurwetten), a model can only be declared 'valid' on the basis of experiments. We will come back to this important issue in some of our other courses.

When one measures some of the geometric elements of the above mentioned set of points, it means that one assigns a number (the observation) according to certain rules (the measurement procedure) to for instance an angle α (a mathematical notion). From experience we know that a measurement of a certain quantity does not necessarily give the same answer when repeated. In order to describe this phenomenon, one could repeat the measurements a great many times (e.g. 100). And on its turn the experiment itself could also be repeated. If it then turns out that the histograms of each experiment are sufficiently alike, that is, their shapes are sufficiently similar and their locations do not differ too much, one can describe the measurement variability by means of a stochastic variable. In this way notions from mathematical statistics are used in order to describe the inherent variability of measurement processes. This is referred to as the *stochastic model*.

In practice not every measurement will be repeated of course. In that case one will have to rely on the extrapolation of past experience. When measuring, not only the measurement result as a number is of interest, but also additional information such as the type of instrument used, the measurement procedure followed, the observer, the weather, etc. Taken together this information is referred to as the *registration* (*registratie*). The purpose of the registration is to establish a link with past experience or experiments such that a justifiable choice for the stochastic model can be made.

In surveying and geodesy one often may assume as stochastic model that the results of measurement (the observations) of a certain measurement method can be described as being an independent sample drawn (aselecte trekking) from a normal (Gaussian) distribution with a certain standard deviation. The mathematical *expectation* (*verwachting*) of the normally distributed stochastic variable (the observable, in Dutch: de waarnemingsgrootheid), is then assumed to equal the unknown angle etc. From this it follows, if the three angles α, β and γ of a triangle are measured, that the expectations of the three angular stochastic variables obey the following relation of the functional model:

$$\alpha + \beta + \gamma = \pi$$

For the individual observations themselves this is usually not the case. A sketch of the relation between terrain situation, functional model, measurement, experiment and stochastic model is given in figure 0.1. The functional and stochastic model together from the mathematical model.

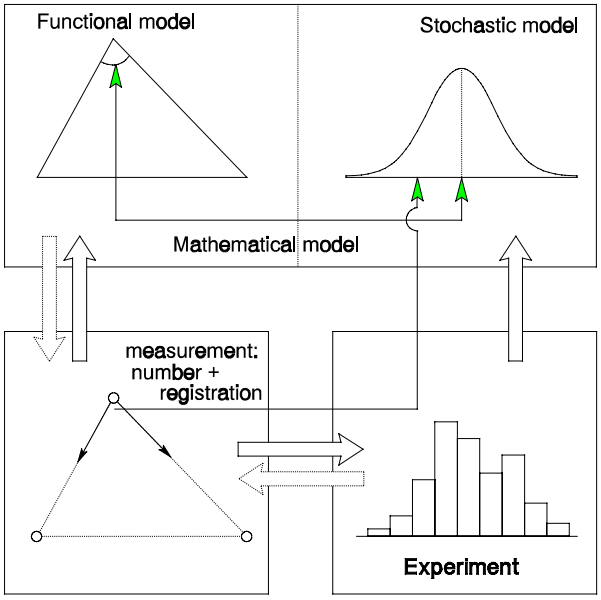


Figure 0.1: Diagram of the fundamental relations in adjustment theory

1 Linear estimation theory: an introduction

1.1 Introduction

Before stating the general problem, let us define a simple example that contains most of the elements that will require our consideration.

Suppose that an unknown distance, which we denote as x, has been measured twice. The two measurements are denoted as y_1 and y_2, respectively. Due to all sorts of influences (e.g. instrumental effects, human effects, etc.) the two measurements will differ in general. That is $y_1 \neq y_2$ or $y_1 - y_2 \neq 0$. One of the two measurements, but probably both, will therefore also differ from the true but unknown distance x. Thus, $y_1 \neq x$ and $y_2 \neq x$. Let us denote the two differences $y_1 - x$ and $y_2 - x$, the so-called measurement errors, as e_1 and e_2 respectively. We can then write:

$$y_i = x + e_i \quad , \quad i = 1,2$$

or in vector notation:

(1)
$$\underset{2\times1}{y} = \underset{2\times1}{a}\ \underset{1\times1}{x} + \underset{2\times1}{e}$$

with:

$$y = (y_1, y_2)^*$$
$$a = (1,1)^*$$
$$e = (e_1, e_2)^*$$

(* denotes the operation of transposition; thus y^* is the transpose of y).

Our problem is now, how to estimate the unknown x from the given measurement or observation vector y. It is to this and related questions that we address our attention. In this introductory chapter we will restrict ourselves to the case of two observations and one unknown. In the next section we will introduce the least-squares principle. Least-squares can be regarded as a deterministic approach to the estimation problem. In the sections following we will introduce additional assumptions regarding the probabilistic nature of the observations and measurement errors. This leads then to the so-called best linear unbiased estimators and the maximum likelihood estimators.

1.2 Least-squares estimation (orthogonal projection)

The equation $\underset{2x1}{y} = \underset{2x1}{a}\ \underset{1x1}{x} + \underset{2x1}{e}$, in which y and a are given, and x and e are unknown, can be visualized as in figure 1.1a.

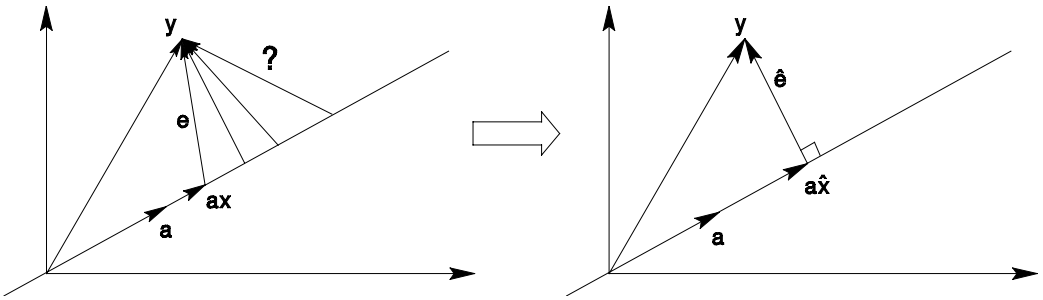


Figure 1.1a: x and e unknown

Figure 1.1b: $\hat{x}$ as estimate of x
$\hat{e}$ as estimate of e

From the geometry of the figure it seems intuitively appealing to estimate x as $\hat{x}$, such that $a\hat{x}$ is as close as possible to the given observation vector y. From figure 1.1b follows that $a\hat{x}$ has minimum distance to y if $\hat{e} = y - a\hat{x}$ is orthogonal to a. That is if:

(2)

$$a^*(y-a\hat{x}) = 0 \qquad \text{"orthogonality"}$$
$$\Downarrow$$
$$(a^*a)\hat{x} = a^*y \qquad \text{"normal equation"}$$
$$\Downarrow$$
$$\hat{x} = (a^*a)^{-1}a^*y \qquad \text{"LSQ-estimate of } x\text{"}$$
$$\Downarrow$$
$$\hat{y} = a\hat{x} = [a(a^*a)^{-1}a^*]y \qquad \text{"LSQ-estimate of observations"}$$
$$\Downarrow$$
$$\hat{e} = y-\hat{y} = [I-a(a^*a)^{-1}a^*]y \qquad \text{"LSQ-estimate of measurement errors"}$$
$$\Downarrow$$
$$\hat{e}^*\hat{e} = y^*[I-a(a^*a)^{-1}a^*]y \qquad \text{"Sum of squares"}$$

The name *"Least-Squares" (LSQ)* is given to the estimate $\hat{x} = (a^*a)^{-1}a^*y$ of x, since $\hat{x}$ minimizes the sum of squares $(y-ax)^*(y-ax)$. We will give two different proofs for this:

***First proof*:** (algebra)

$$\begin{aligned}(y-ax)^*(y-ax) &= [y-a\hat{x}-a(x-\hat{x})]^*[y-a\hat{x}-a(x-\hat{x})]\\ &= [\hat{e}-a(x-\hat{x})]^*[\hat{e}-a(x-\hat{x})]\\ &= \hat{e}^*\hat{e} - \hat{e}^*a(x-\hat{x}) - (x-\hat{x})^*a^*\hat{e} + (x-\hat{x})^*a^*a(x-\hat{x})\\ &= \hat{e}^*\hat{e} + (x-\hat{x})^*a^*a\,(x-\hat{x})\\ &= \text{minimal for } x = \hat{x}.\end{aligned}$$

End of proof.

***Second proof*:** (calculus)

Call $E(x) \doteq (y-ax)^*(y-ax) = y^*y - 2y^*ax + a^*ax^2$.

From calculus we know that $\hat{x}$ is a solution of the minimization problem $\min_x E(x)$ if:

$$\frac{dE}{dx}(\hat{x}) = 0 \text{ and } \frac{d^2E}{dx^2}(\hat{x}) > 0 .$$

With:

$$\frac{dE}{dx}(\hat{x}) = -2a^*y + 2a^*a\hat{x} = 0 \quad \Rightarrow \quad (a^*a)\hat{x} = a^*y$$

and:

$$\frac{d^2E}{dx^2}(\hat{x}) = 2a^*a > 0$$

the proof follows. End of proof.

***Example 1*:** Let us consider the problem defined in section 1.1:

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

With $a = (1,1)^*$ the normal equation $(a^*a)\hat{x} = a^*y$ reads:

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

Hence the least-squares estimate $\hat{x} = (a^*a)^{-1}a^*y$ is:

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

Thus, to estimate x, one adds the measurements and divides by the number of measurements. Hence the least-squares estimate equals in this case the arithmetic average. The least-squares estimates of the observations and observation errors follow from $\hat{y} = a\hat{x}$ and $\hat{e} = y - \hat{y}$ as:

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

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

$$\hat{e}^*\hat{e} = \tfrac{1}{2}(y_1 - y_2)^2.$$

End of example.

The matrices $a(a^*a)^{-1}a^*$ and $I - a(a^*a)^{-1}a^*$ that occur in (2) are known as *orthogonal projectors*. They play a central role in estimation theory. We will denote them as:

(3) $$\boxed{P_a = a(a^*a)^{-1}a^* \quad , \quad P_a^\perp = I - a(a^*a)^{-1}a^*}$$

The matrix P_a projects *onto* a line spanned by a and *along* a line orthogonal to a. To see this, note that we can write $P_a y$ with the help of the cosine rule as (see figure 1.2):

$$P_a y = a(a^*a)^{-1}a^* y = a(a^*a)^{-1}\|a\|\,\|y\|\cos\alpha = \|y\|\cos\alpha\,\frac{a}{\|a\|} \ .$$

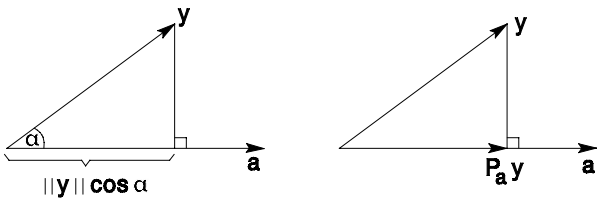


Figure 1.2

In a similar way we can show that $P_a^\perp$ projects *onto* a line orthogonal to a and *along* a line spanned by a (see figure 1.3).

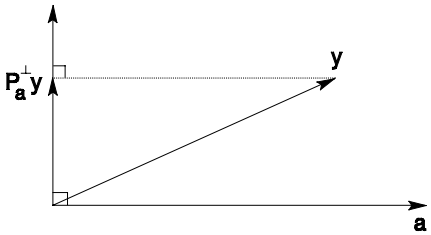


Figure 1.3

If we denote the vector that is orthogonal to a as b, thus $b^*a=0$, the projectors P_a and $P_a^\perp$ can alternatively be represented as:

(4) $$\boxed{P_a = P_b^\perp = I - b(b^*b)^{-1}b^* \quad , \quad P_a^\perp = P_b = b(b^*b)^{-1}b^*}$$

To see this, consult figure 1.4 and note that with the help of the cosine rule we can write:

$$P_a^{\perp} y \;=\; \|y\| \cos\beta \, \frac{b}{\|b\|} \;=\; b\|b\|^{-2} b^{*} y \;=\; b(b^{*}b)^{-1} b^{*} y \;=\; P_b y .$$

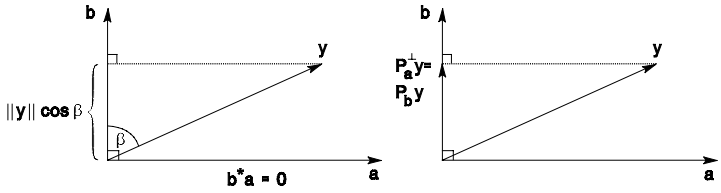


Figure 1.4

With (3) and (4) we have now two different representations of the orthogonal projectors $P_a=P_b^{\perp}$ and $P_a^{\perp}=P_b$. The representations of (3) are used in (2), which followed from solving equation (1) in a least-squares sense. We will now show that the representations of (4) correspond to solving the equation:

(5)
$$\underset{1\times 2}{b^{*}} \underset{2\times 1}{y} \;=\; \underset{1\times 2}{b^{*}} \underset{2\times 1}{e}$$

in a least-squares sense. Equation (5) follows from pre-multiplying equation (1) with b^{*} and noting that $b^{*}a=0$. We know that the least-squares solution of (1) follows from solving the minimization problem:

$$\min_{x} \; \|y - ax\|^2 .$$

We have used the notation $\|.\|^2$ here to denote the square of the length of a vector. The minimization problem $\|y\text{-}ax\|^2$ can also be written as:

$$\min_{z} \; \|y - z\|^2 \quad \text{subject to} \quad z \;=\; ax \; .$$

The equation $z=ax$ is the *parametric representation* of a line through the origin with direction vector a. From linear algebra we know that this line can also be represented in *implicit form* with the help of the normal vector b. Thus we may replace $z=ax$ by $b^{*}z=0$. This gives:

$$\min_{z} \; \|y - z\|^2 \quad \text{subject to} \quad b^{*}z \;=\; 0 \; .$$

With $e=y - z$, this may also be written as:

$$\min_{e} \; \|e\|^2 \quad \text{subject to} \quad b^{*}y \;=\; b^{*}e \; .$$

The least-squares estimate $\hat{e}$ of the unknown e in (5) thus follows as the vector $\hat{e}$ which satisfies (5) and has minimal length. Figure 1.5 shows that $\hat{e}$ is obtained as the orthogonal projection of y onto the vector b. The least-squares estimates of (5) read therefore:

(6)
$$\hat{e} = P_b y = [b(b^*b)^{-1}b^*]y$$
$$\hat{y} = y - \hat{e} = P_b^{\perp} y = [I - b(b^*b)^{-1}b^*]y$$

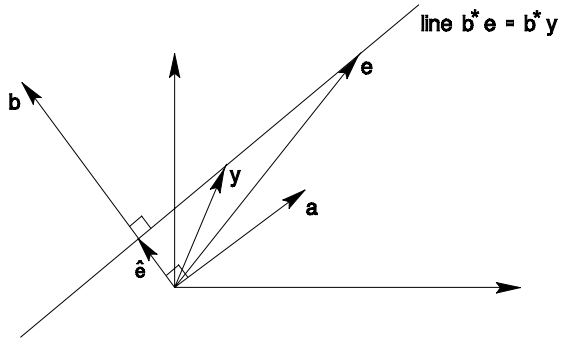


Figure 1.5

***Example 2*:** Let us consider the problem defined in section 1.1:

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

The form corresponding to (5) is:

$$(1 \quad -1)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = (1 \quad -1)\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

This equation expresses the fact that the difference between the two observed distances y_1 and y_2 is unequal to zero in general because of the presence of measurement errors e_1 and e_2. Since the vector b is given in this case as:

$$b = (1 - 1)^*,$$

we have:

$$P_b = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad P_b^{\perp} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The least-squares estimates read therefore:

$$\begin{pmatrix}\hat{e}_1\\ \hat{e}_2\end{pmatrix} = \boldsymbol{P_b}\begin{pmatrix}y_1\\ y_2\end{pmatrix} = \tfrac{1}{2}\begin{pmatrix}y_1-y_2\\ y_2-y_1\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}\hat{y}_1\\ \hat{y}_2\end{pmatrix} = \boldsymbol{P_b^{\perp}}\begin{pmatrix}y_1\\ y_2\end{pmatrix} = \tfrac{1}{2}\begin{pmatrix}y_1+y_2\\ y_1+y_2\end{pmatrix},$$

which is identical to the results obtained earlier.

End of example.

1.3 Weighted least-squares estimation

Up to now, we have regarded the two observations y_1 and y_2 as equally accurate, and solved:

$$\underset{2\times 1}{y} = \underset{2\times 1}{\boldsymbol{a}}\ \underset{1\times 1}{x} + \underset{2\times 1}{\boldsymbol{e}}$$

by looking for the value $\hat{x}$ that minimized:

$$E(x) = (y_1 - \boldsymbol{a}_1 x)^2 + (y_2 - \boldsymbol{a}_2 x)^2 = (y-\boldsymbol{a}x)^* \ I\ (y-\boldsymbol{a}x)\ .$$

The least-squares principle can however be generalized by introducing a *symmetric, positive-definite weight matrix W* so that:

$$E(x) = w_{11}(y_1 - \boldsymbol{a}_1 x)^2 + 2w_{12}(y_1 - \boldsymbol{a}_1 x)(y_2 - \boldsymbol{a}_2 x) + w_{22}(y_2 - \boldsymbol{a}_2 x)^2 = (y-\boldsymbol{a}x)^*\ W\ (y-\boldsymbol{a}x)\ .$$

The elements of the 2×2 matrix W can be chosen to emphasize (or de-emphasize) the influence of specific observations upon the estimate $\hat{x}$. For instance, if $w_{11}>w_{22}$, then more importance is attached to the first observation, and the process of minimizing $E(x)$ tries harder to make $(y_1-a_1x)^2$ small. This seems reasonable if one believes that the first observation is more trustworthy than the second; for instance if observation y_1 is obtained from a more accurate instrument than observation y_2.

The solution $\hat{x}$ of:

$$\min_x E(x) = \min_x (y-\boldsymbol{a}x)^* W (y-\boldsymbol{a}x),$$

is called the *weighted* least-squares (WLSQ) estimate of x. As we know the solution can be obtained by solving:

$$\frac{dE}{dx}(\hat{x}) = 0,$$

and checking whether:

$$\frac{d^2E}{dx^2}(\hat{x}) > 0\ .$$

From:

$$E(x) = y^*Wy - 2y^*Wax + a^*Wax^2$$

follows:

$$\frac{dE}{dx}(\hat{x}) = -2a^*Wy + 2a^*Wa\hat{x} = 0$$

and:

$$\frac{d^2E}{dx^2}(\hat{x}) = 2a^*Wa\ .$$

Since the weight matrix W is assumed to be positive-definite we have $a^*Wa>0$ (recall from linear algebra that a matrix M is said to be *positive-definite* if it is symmetric and $z^*Mz>0 \forall\ z\neq 0$). From the first equation we therefore get:

$$\begin{aligned}
&a^*W(y-a\hat{x}) = 0 \\
&\Downarrow \\
&(a^*Wa)\hat{x} = a^*Wy && \text{"normal equation"} \\
&\Downarrow \\
&\hat{x} = (a^*Wa)^{-1}a^*Wy && \text{"WLSQ-estimate of } x\text{"} \\
&\Downarrow \\
&\hat{y} = a\hat{x} = [a(a^*Wa)^{-1}a^*W]y && \text{"WLSQ-estimate of } y\text{"} \\
&\Downarrow \\
&\hat{e} = y-\hat{y} = [I-a(a^*Wa)^{-1}a^*W]y && \text{"WLSQ-estimate of } e\text{"} \\
&\Downarrow \\
&\hat{e}^*W\hat{e} = y^*[W-Wa(a^*Wa)^{-1}a^*W]y && \text{"weighted sum of squares"}
\end{aligned} \tag{7}$$

Compare this result with (2).

***Example 3*:** Consider again the problem defined in section 1.1:

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

As weight matrix we take:

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

With $a=(1\ 1)^*$ the normal equation $(a^*Wa)\hat{x}=a^*Wy$ reads:

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

Hence, the weighted least-squares estimate $\hat{x}=(a^*Wa)^{-1}a^*Wy$ is:

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

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

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

and the WLSQ-estimate of e, $\hat{e}=y - \hat{y}$, is:

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

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

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

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

End of example.

In section 1.2 we introduced the least-squares principle by using the intuitively appealing geometric principle of orthogonal projection. In this section we introduced the weighted least-squares principle as the problem of minimizing the weighted sum of squares $(y\text{-}ax)^*W(y\text{-}ax)$. An interesting question is now, whether it is also possible to give a geometric interpretation to the *weighted* least-squares principle. This turns out to be the case. First note that the equation:

$$z^*Wz = \text{constant} = c$$

or:

$$w_{11}z_1^2 + 2w_{12}z_1z_2 + w_{22}z_2^2 = c$$

describes an *ellipse*. If $w_{11}=w_{22}$ and $w_{12}=0$, the ellipse is a circle. If $w_{11}\neq w_{22}$ and $w_{12}=0$, the ellipse has its major and minor axis parallel to the z_1- and z_2- axes, respectively. And if $w_{12}\neq 0$, the ellipse may have an arbitrary orientation (see figure 1.6).

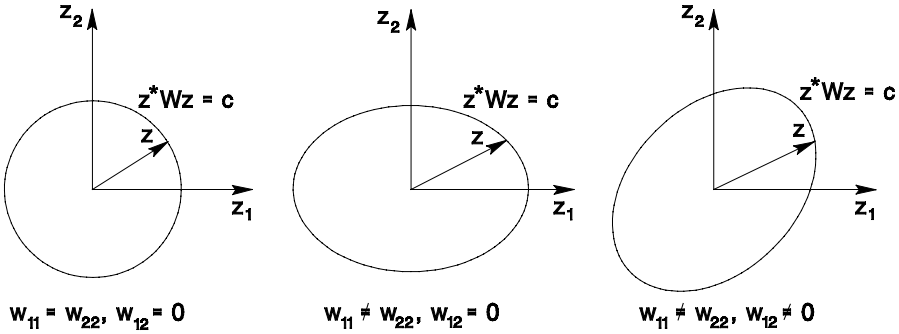


Figure 1.6

If we define a function $F(z)$ as $F(z)=z^*Wz$, then we know from calculus that the *gradient* of $F(z)$ evaluated at z_0 is a vector which is *normal* to the ellipse $F(z)=c$ at z_0. The gradient of $F(z)$ evaluated at z_0 is:

$$\begin{pmatrix} \frac{\partial F}{\partial z_1} \\ \frac{\partial F}{\partial z_2} \end{pmatrix}_{z_o} = \begin{pmatrix} 2w_{11}z_1 + 2w_{12}z_2 \\ 2w_{12}z_1 + 2w_{22}z_2 \end{pmatrix}_{z_o} = 2\,Wz_o \ .$$

This vector is thus normal or orthogonal to the ellipse at z_0 (see figure 1.7).

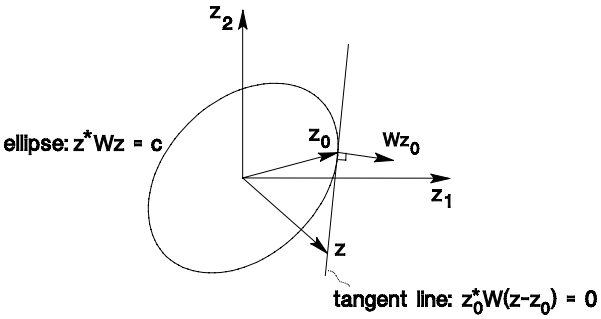


Figure 1.7

From this follows that the tangent line of the ellipse $z^*Wz=c$ at z_0 is given by:

$$z_o^*W(z-z_o) = 0 \ .$$

Comparing this evaluation with the first equation of (7) shows that the WLSQ-estimate of x is given by that $\hat{x}$ for which $y\text{-}a\hat{x}$ is orthogonal to the normal Wa at a of the ellipse $z^*Wz=a^*Wa$. Hence, $y\text{-}a\hat{x}$ is parallel to the tangent line of the ellipse $z^*Wz=a^*Wa$ at a (see figure 1.8). Note that if $W=I$, the ellipse becomes a circle and the unweighted least-squares estimate is obtained through orthogonal projection (see figure 1.9).

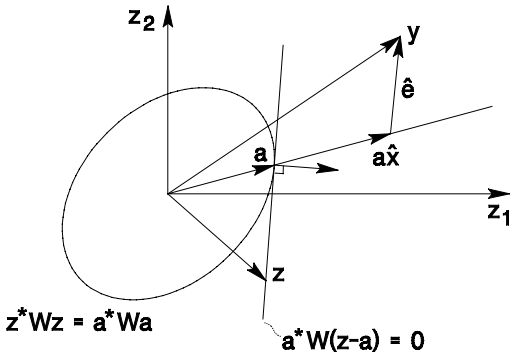


Figure 1.8

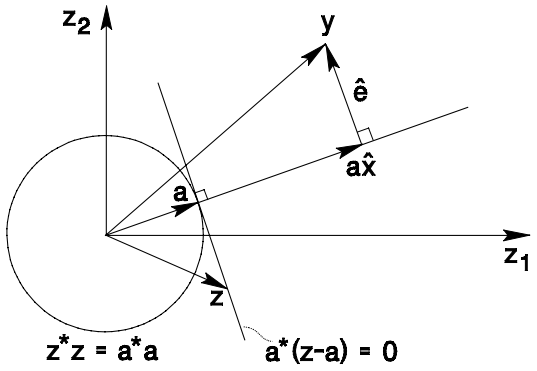


Figure 1.9

The matrices $a(a^*Wa)^{-1}a^*W$ and $I\text{-}a(a^*Wa)^{-1}a^*W$ that occur in (7) are known as *oblique projectors*.

We will denote them as:

(8)
$$\begin{aligned} P_{a,(Wa)^\perp} &= a(a^*Wa)^{-1}a^*W\ , \\ P^{\perp}_{a,(Wa)^\perp} &= P_{(Wa)^\perp,a} = I-a(a^*Wa)^{-1}a^*W \end{aligned}$$

The matrix $P_{a,(Wa)^\perp}$ projects *onto* a line spanned by a and *along* a line orthogonal to Wa.

To see this, note that we can write $P_{a,(Wa)^\perp}y$ with the help of the cosine rule as (see figure 1.10):

$$P_{a,(Wa)^\perp}y = a(a^*Wa)^{-1}a^*Wy = a\frac{\|Wa\|\|y\|\cos\alpha}{\|Wa\|\|a\|\cos\beta} = \|y\|\frac{\cos\alpha}{\cos\beta}\frac{a}{\|a\|}$$

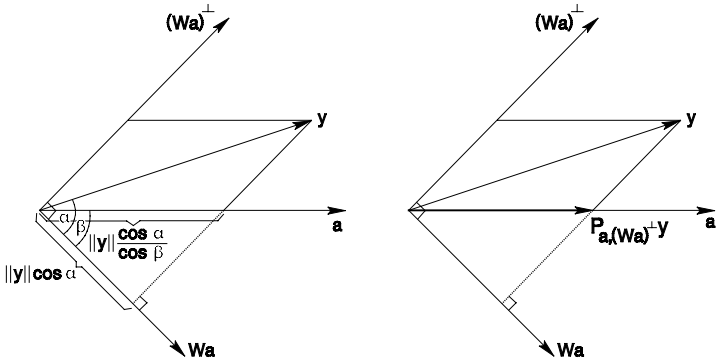


Figure 1.10

In a similar way we can show that $P^\perp_{a,(Wa)^\perp}$ projects *onto* a line orthogonal to Wa and *along* a line spanned by a (see figure 1.11).

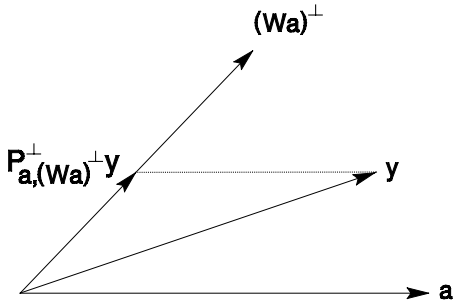


Figure 1.11

Above we have denoted the vector orthogonal to Wa as $(Wa)^\perp$; thus $((Wa)^\perp)^*(Wa)=0$. With the help of the vector b, which is defined in section 1.2 as being orthogonal to a, thus $b^*a=0$, we can write instead of $(Wa)^\perp$ also $W^{-1}b$, because $(W^{-1}b)^*(Wa)=0$ (the inverse W^{-1} of W exists, since we assumed W to be positive-definite). With the help of $W^{-1}b$ we can now represent the two projectors of (8) alternatively as:

(9)
$$\begin{aligned} P_{a,(Wa)^{\perp}} &= P_{a,W^{-1}b} = I - W^{-1}b(b^{*}W^{-1}b)^{-1}b^{*}\,, \\ P_{(Wa)^{\perp},a} &= P_{W^{-1}b,a} = W^{-1}b(b^{*}W^{-1}b)^{-1}b^{*} \end{aligned}$$

To see this, consult figure 1.12 and note that with the help of the cosine rule we can write:

$$P_{W^{-1}b,a}y = \|y\|\frac{\cos\alpha}{\cos\beta}\,\frac{W^{-1}b}{\|W^{-1}b\|} = W^{-1}b\frac{\|y\|\cos\alpha}{\|W^{-1}b\|\cos\beta} = W^{-1}b\frac{b^{*}y}{(W^{-1}b)^{*}b}\,,$$

or:

$$P_{W^{-1}b,a}y = W^{-1}b(b^{*}W^{-1}b)^{-1}b^{*}y\,.$$

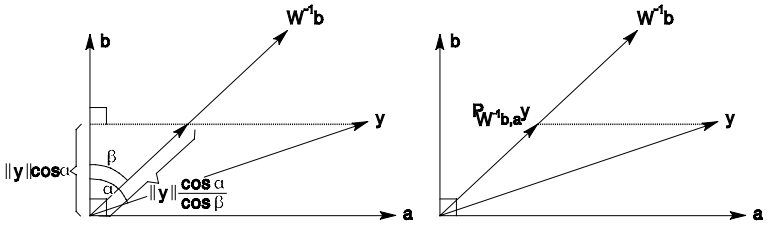


Figure 1.12

Note that if $W=I$, then $W^{-1}=I$ and equations (8) and (9) reduce to equations (3) and (4) respectively. In section 1.2 we showed that for $W=I$ the representations of (9) correspond to solving equation (5) in an unweighted least-squares sense. We will now show that the representations of (9) correspond to solving (5) in a weighted least-squares sense. From the equivalence of the following minimization problems:

$$\min_{x}\ (y-ax)^{*}W(y-ax),$$

$$\min_{z}\ (y-z)^{*}W(y-z) \qquad \text{subject to} \qquad z = ax,$$

$$\min_{z}\ (y-z)^{*}W(y-z) \qquad \text{subject to} \qquad b^{*}z = 0,$$

$$\min_{e}\ (e)^{*}W(e) \qquad \text{subject to} \qquad b^{*}y = b^{*}e,$$

follows that the weighted least-squares estimate of e is given by that $\hat{e}$ on the line $b^{*}y=b^{*}e$ for which $e^{*}We$ is minimal. We know that $e^{*}We$ = constant = c is the equation of an ellipse. For different values of c we get different ellipses and the larger the value of c is, the larger the ellipse (see figure 1.13).

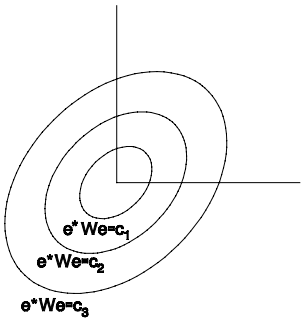


Figure 1.13 $c_3>c_2>c_1$

The solution $\hat{e}$ is therefore given by the *point of tangency* of the line $b^*y=b^*e$ with one of the ellipses $e^*We=c$ (see figure 1.14).

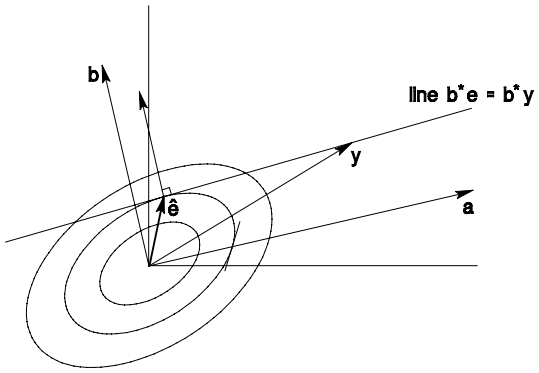


Figure 1.14

At this point of tangency the normal of the ellipse is parallel to the normal of the line $b^*e=b^*y$. But since the normal of the line $b^*e=b^*y$ is given by b, it follows that the normal of the ellipse at $\hat{e}$ is also given by b. From this we can conclude that $\hat{e}$ has to be parallel to $W^{-1}b$. Hence we may write:

$$\hat{e} = W^{-1}b\alpha \ .$$

The unknown scalar α may now be determined from the fact that $\hat{e}$ has to lie on the line $b^*e=b^*y$. Pre-multiplying $\hat{e}=W^{-1}b\alpha$ with b^* gives for α

$$\alpha = (b^*W^{-1}b)^{-1}b^*\hat{e} = (b^*W^{-1}b)^{-1}b^*y \ .$$

Substituting this in the equation $\hat{e}=W^{-1}b\alpha$ shows that the weighted least-squares estimates of equation (5) read:

$$\begin{aligned} \hat{\boldsymbol{e}} &= \boldsymbol{P}_{W^{-1}\boldsymbol{b},\boldsymbol{a}}\, y = [W^{-1}\boldsymbol{b}(\boldsymbol{b}^{*}W^{-1}\boldsymbol{b})^{-1}\boldsymbol{b}^{*}]\, y \\ \hat{y} &= y-\hat{\boldsymbol{e}} = \boldsymbol{P}^{\perp}_{W^{-1}\boldsymbol{b},\boldsymbol{a}}\, y = [I-W^{-1}\boldsymbol{b}(\boldsymbol{b}^{*}W^{-1}\boldsymbol{b})^{-1}\boldsymbol{b}^{*}]\, y \end{aligned} \tag{10}$$

Compare this result with equation (6).

***Example 4*:** Consider again the problem defined in section 1.1:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} x + \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{pmatrix}$$

The form corresponding to (5) is:

$$(1-1)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = (1-1)\begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{pmatrix}.$$

thus the vector b is given as:

$$\boldsymbol{b} = (1-1)^{*}\ .$$

As weight matrix we take:

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

Its inverse reads therefore:

$$W^{-1} = \begin{pmatrix} w_{11}^{-1} & 0 \\ 0 & w_{22}^{-1} \end{pmatrix}.$$

The two projectors $\boldsymbol{P}_{W^{-1}\boldsymbol{b},\boldsymbol{a}}$ and $\boldsymbol{P}^{\perp}_{W^{-1}\boldsymbol{b},\boldsymbol{a}}$ can now be computed as:

$$\boldsymbol{P}_{W^{-1}\boldsymbol{b},\boldsymbol{a}} = W^{-1}\boldsymbol{b}(\boldsymbol{b}^{*}W^{-1}\boldsymbol{b})^{-1}\boldsymbol{b}^{*} = \frac{w_{11}w_{22}}{w_{11}+w_{22}} \begin{pmatrix} w_{11}^{-1} & -w_{11}^{-1} \\ -w_{22}^{-1} & w_{22}^{-1} \end{pmatrix},$$

and

$$\boldsymbol{P}^{\perp}_{W^{-1}\boldsymbol{b},\boldsymbol{a}} = I - W^{-1}\boldsymbol{b}(\boldsymbol{b}^{*}W^{-1}\boldsymbol{b})^{-1}\boldsymbol{b}^{*} = \frac{w_{11}w_{22}}{w_{11}+w_{22}}\begin{pmatrix} w_{22}^{-1} & w_{11}^{-1} \\ w_{22}^{-1} & w_{11}^{-1} \end{pmatrix}.$$

The weighted least-squares estimates $\hat{e}$ and $\hat{y}$ read therefore:

$$\begin{pmatrix} \hat{\boldsymbol{e}}_1 \\ \hat{\boldsymbol{e}}_2 \end{pmatrix} = \boldsymbol{P}_{W^{-1}\boldsymbol{b},\boldsymbol{a}}y = \frac{1}{w_{11}+w_{22}}\begin{pmatrix} w_{22}(y_1-y_2) \\ w_{11}(y_2-y_1) \end{pmatrix},$$

and

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = \boldsymbol{P}^{\perp}_{W^{-1}\boldsymbol{b},\boldsymbol{a}}y = \frac{1}{w_{11}+w_{22}}\begin{pmatrix} w_{11}y_1 + w_{22}y_2 \\ w_{11}y_1 + w_{22}y_2 \end{pmatrix}.$$

These results are identical to the results obtained in example 3. End of example.

1.4 Weighted least-squares is also orthogonal projection

In the previous two sections we have seen that unweighted and weighted least-squares estimation could be interpreted as *orthogonal* and *oblique* projection respectively (see figure 1.15).

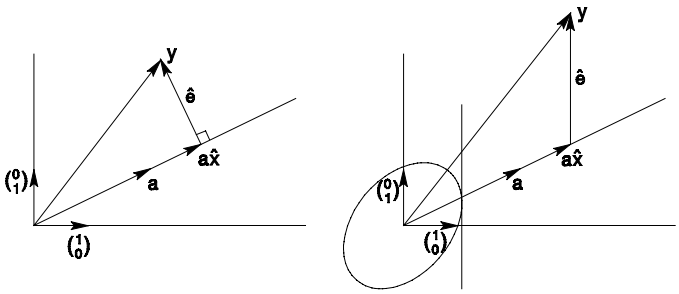


Figure 1.15

It is important to realize however that these geometric interpretations are very much dependent on the choice of base vectors with the help of which the vectors y, a and e are constructed. In all the cases considered so far we have implicitly assumed to be dealing with the orthogonal base vectors $(1,0)^{*}$ and $(0,1)^{*}$. That is, the vectors y, a and e of the equation:

$$y = \boldsymbol{a}x + \boldsymbol{e},$$

were constructed from the scalars y_1, y_2, a_1, a_2, e_1 and e_2 as:

$$
\begin{aligned}
y &= \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\
\boldsymbol{a} &= \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\
\boldsymbol{e} &= \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = e_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + e_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}.
\end{aligned} \tag{11}
$$

Let us now investigate what happens if instead of the orthonormal set $(1,0)^*$ and $(0,1)^*$, a different set, i.e. a non-orthonormal set, of base vectors is chosen. Let us denote these new base vectors as d_1 and d_2. With the help of the base vectors d_1 and d_2, the vectors y, a and e can then be constructed from the scalars y_1, y_2, a_1, a_2, e_1 and e_2 as:

$$
\begin{aligned}
y &= y_1 \boldsymbol{d}_1 + y_2 \boldsymbol{d}_2 \ , \\
\boldsymbol{a} &= a_1 \boldsymbol{d}_1 + a_2 \boldsymbol{d}_2 \ , \\
\boldsymbol{e} &= e_1 \boldsymbol{d}_1 + e_2 \boldsymbol{d}_2 \ .
\end{aligned} \tag{12}
$$

Again the equation $y=ax+e$ holds. This is shown in figure 1.16a. Note that the vectors of (12) are not the same as those of (11). They do however have the same components.

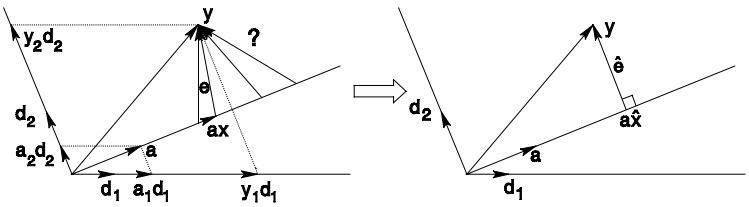


Figure 1.16a Figure 1.16b

Looking at figure 1.16 it seems again, like in section 1.2, intuitively appealing to estimate x as $\hat{x}$ such that $a\hat{x}$ is as close as possible to given vector y. This is the case if the vector $y\text{-}a\hat{x}$ is orthogonal to the vector a; that is, if:

$$
\boldsymbol{a}^*(y - \boldsymbol{a}\hat{x}) = 0. \tag{13}
$$

"orthogonality"

With (12) this gives:

$$
(a_1\boldsymbol{d}_1 + a_2\boldsymbol{d}_2)^*\left[y_1\boldsymbol{d}_1 + y_2\boldsymbol{d}_2 - (a_1\boldsymbol{d}_1 + a_2\boldsymbol{d}_2)\hat{x}\right] = 0,
$$

or

$$
\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^* \begin{pmatrix} \boldsymbol{d}_1^*\boldsymbol{d}_1 & \boldsymbol{d}_1^*\boldsymbol{d}_2 \\ \boldsymbol{d}_2^*\boldsymbol{d}_1 & \boldsymbol{d}_2^*\boldsymbol{d}_2 \end{pmatrix} \left[\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \hat{x} \right] = 0,
$$

or

(14)
$$\boxed{\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^* D^* D \left[\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \hat{x} \right] = 0,}$$

with

$$D = (d_1 \; d_2).$$

The matrix D^*D of (14) is symmetric, since $(D^*D)^*=D^*D$, and it is positive-definite, since $z^*D^*Dz = (Dz)^*(Dz)>0$ for all $z\neq 0$. Hence, the matrix D^*D may be interpreted as a weight matrix. By interpreting the matrix D^*D of (14) as a weight matrix we see that the solution of (13) can in fact be interpreted as a *weighted least-squares* solution, i.e. a solution of the minimization problem:

$$\min_x \left[\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} x \right]^* D^* D \left[\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} x \right].$$

We have thus shown that if in a weighted least-squares problem the weight matrix W can be written as the product of a matrix D^* and its transpose D, then the weighted least-squares estimate $\hat{y}=a\hat{x}$ can be interpreted as being obtained through *orthogonal* projection of $y=y_1d_1+y_2d_2$ onto the line with direction vector $a=a_1d_1+a_2d_2$, where d_1 and d_2 are the column vectors of matrix D (see figure 1.17).

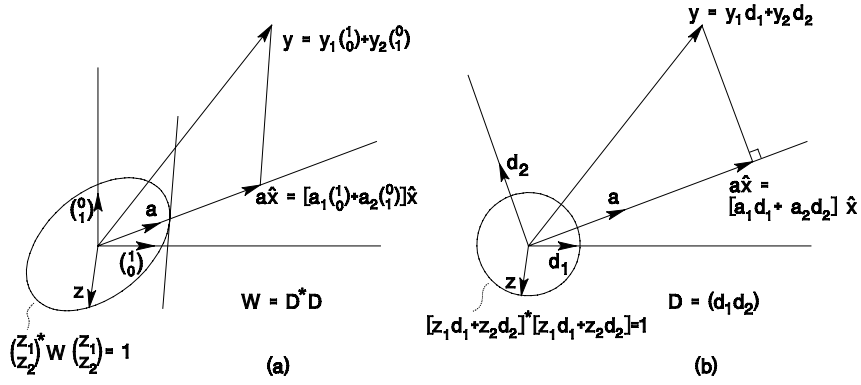


Figure 1.17

The crucial question is now whether *every* weight matrix W, i.e. every symmetric and positive-definite matrix, can be written as a product of a matrix and its transpose, i.e. as $W=D^*D$. If this is the case, then every weighted least-squares problem can be interpreted as a problem of orthogonal projection! The answer is: *yes*, every weight matrix W can be written as $W=D^*D$. For a proof, see the *intermezzo* at the end of this section.

The weight matrix $W=D^*D$ can be interpreted as defining an *inner product* (Strang, 1988). In the unweighted case, $W=I$, the inner product of two vectors u and v with components u_1,u_2 and v_1,v_2 respectively, is given as:

$$(u,v)_I = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^* I \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \left[u_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]^* \left[v_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] .$$

This is the so-called *standard inner product*, where the inner product is represented by the unit matrix I. The square of the length of the vector u reads then:

$$\|u\|_I^2 = u_1^2 + u_2^2 .$$

In the weighted case, $W=D^*D$, the inner product of u and v is given as:

(15)
$$(u,v)_W = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}^* \boldsymbol{D}^*\boldsymbol{D} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \left(u_1\boldsymbol{d}_1 + u_2\boldsymbol{d}_2\right)^* \left(v_1\boldsymbol{d}_1 + v_2\boldsymbol{d}_2\right) .$$

The square of the length of the vector u reads then:

$$\|u\|_W^2 = u_1^2\boldsymbol{d}_1^*\boldsymbol{d}_1 + 2u_1u_2\boldsymbol{d}_1^*\boldsymbol{d}_2 + u_2^2\boldsymbol{d}_2^*\boldsymbol{d}_2 .$$

Note that the two ways in which $(u,v)_W$ is written in (15) correspond with the two figures 1.17a and 1.17b respectively. In figure 1.17a orthonormal base vectors are used and the metric is *visualized* by the ellipse:

$$(z,z)_W = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}^* W \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 1 .$$

In figure 1.17b however, the ellipse is replaced by a circle:

$$(z,z)_W = \left(z_1\boldsymbol{d}_1 + z_2\boldsymbol{d}_2\right)^* I \left(z_1\boldsymbol{d}_1 + z_2\boldsymbol{d}_2\right) = 1 ,$$

and the metric is *visualized* by the non-orthonormal base vectors d_1 and d_2.

Now that we have established the fact that weighted least-squares estimation can be interpreted as orthogonal projection, it must also be possible to interpret the oblique projectors of equation (8) as orthogonal projectors. We will show that the matrix $P_{a,(Wa)^\perp}$ of (8) can be interpreted as the matrix which projects orthogonally onto a line spanned by the vector a, where orthogonality is measured with respect to the inner product defining weight matrix W. From figure 1.18 follows that:

$$\hat{y} = \|y\|_W \cos\alpha \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_W} = \boldsymbol{a}\|\boldsymbol{a}\|_W^{-2} (\boldsymbol{a},y)_W ,$$

or

$$\hat{y}_1 \boldsymbol{d}_1 + \hat{y}_2 \boldsymbol{d}_2 = (a_1 \boldsymbol{d}_1 + a_2 \boldsymbol{d}_2) \left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^* W \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \right]^{-1} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^* W \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

or

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^* W \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \right]^{-1} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^* W \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

and in this expression we recognize indeed the matrix $\boldsymbol{P}^{\perp}_{\boldsymbol{a},(W\boldsymbol{a})}$ of equation (8). From now on we will assume that it is clear from the context which inner-product is taken and therefore also denote the matrix $\boldsymbol{P}^{\perp}_{\boldsymbol{a},(W\boldsymbol{a})}$ as P_a.

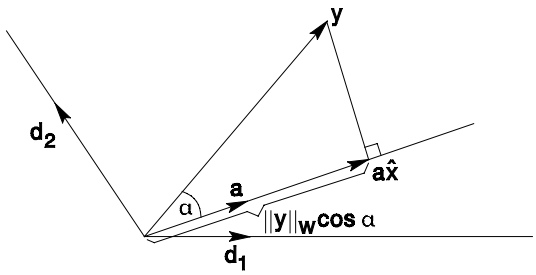


Figure 1.18

***Intermezzo: Cholesky decomposition*:**

If W is an $n \times n$ symmetric positive definite matrix, there is a nonsingular lower triangular matrix L with positive diagonal elements such that $W=LL^*$.

***Proof*:** The proof is by induction. Clearly, the result is true for $n=1$ and the induction hypothesis is that it is true for any $(n\text{-}1)\times(n\text{-}1)$ symmetric positive definite matrix.

Let

$$W = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{a} \\ \boldsymbol{a}^* & b_{nn} \end{pmatrix},$$

where A is a $(n\text{-}1)\times(n\text{-}1)$ matrix. Since W is symmetric positive definite, it will be clear that A is also symmetric positive definite. Thus, by the induction hypothesis, there is a nonsingular lower triangular matrix L_{n-1} with positive diagonal elements such that $A=L_{n-1}L^*_{n-1}$.

Let

$$L = \begin{pmatrix} L_{n-1} & 0 \\ l^* & \beta \end{pmatrix},$$

where l and β are to be obtained by equating LL^* to W:

$$LL^* = \begin{pmatrix} L_{n-1} & 0 \\ l^* & \beta \end{pmatrix}\begin{pmatrix} L_{n-1}^* & l \\ 0 & \beta \end{pmatrix} = \begin{pmatrix} A & a \\ a^* & b_{nn} \end{pmatrix} = W.$$

This implies that l and β must satisfy:

$$L_{n-1}l = a \quad , \quad l^*l + \beta^2 = b_{nn}.$$

Since L_{n-1} is nonsingular, the first equation determines l as $l=L^{-1}{}_{n-1}\, a$. The second equation allows a positive β provided that $b_{nn} - l^*l>0$. In order to proof that $b_{nn} - l^*l>0$ we make use of the fact that W is positive definite, i.e.:

$$x^*Wx = (x_1^* x_2)\begin{pmatrix} A & a \\ a^* & b_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1^* x_2)\begin{pmatrix} L_{n-1}L_{n-1}^* & a \\ a^* & b_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} > 0 \quad \forall x \neq 0$$

Substitution of $x_1 = -(L_{n-1}^*)^{-1}l$ and $x_2=1$ shows that $b_{nn} - l^*l >0$.

End of proof.

A simple example of a Cholesky decomposition is:

$$\begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & \sqrt{3} \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 0 & \sqrt{3} \end{pmatrix}.$$

1.5 The mean and covariance matrix of least-squares estimators

From now on we shall make no distinction between unweighted and weighted least-squares estimation, and just speak of least-squares estimation. From the context will be clear whether $W=I$ or not.

In our discussions so far no assumptions regarding the probabilistic nature of the observations or measurement errors were made. We started by writing the two equations:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}x + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

in vector notation as:

$$y = ax + e,$$

and showed that the least-squares estimates of x, y and e follow from solving the minimization problem:

$$\min_x \ (y - ax)^* W (y - ax)$$

as:

$$\begin{aligned} \hat{x} &= (a^* W a)^{-1} a^* W y, \\ \hat{y} &= a\hat{x}, \\ \hat{e} &= y - a\hat{x}. \end{aligned} \tag{16}$$

Vector $\hat{e}$ contains the least-squares residuals. Since functions of random variables are again random variables, it follows that if the observation vector is assumed to be a random vector $\underline{y}$ (note: the underscore indicates that we are dealing with random variables) and substituted for y in (16), the left-hand sides of (16) also become random variables:

$$\boxed{\begin{aligned} \underline{\hat{x}} &= (a^* W a)^{-1} a^* W \underline{y}, \\ \underline{\hat{y}} &= a\underline{\hat{x}}, \\ \underline{\hat{e}} &= \underline{y} - a\underline{\hat{x}} \end{aligned}} \tag{17}$$

These random variables are called least-squares *estimators*, whereas if $\underline{y}$ is replaced by its sample value y one speaks of least-squares *estimates*.

From your lecture notes in statistics you know how to derive the probability density functions of the estimators of (17) from the probability density function $p_{\underline{y}}(y)$ of $\underline{y}$. You also know that if the probability density function $p_{\underline{y}}(y)$ of $\underline{y}$ is *"normal"* or *"Gaussian"*, then also $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$ are normally distributed. This follows, since $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$ are *linear functions of* $\underline{y}$. You know furthermore how to compute the means and covariance matrices of $\underline{y}$. This is particularly straightforward in case of linear functions (see also appendix E and F).

Application of the *"propagation law of means"* to (17) gives:

$$\begin{aligned} E\{\underline{\hat{x}}\} &= (a^* W a)^{-1} a^* W E\{\underline{y}\} \\ E\{\underline{\hat{y}}\} &= a E\{\underline{\hat{x}}\} = P_a E\{\underline{y}\} \\ E\{\underline{\hat{e}}\} &= E\{\underline{y}\} - a E\{\underline{\hat{x}}\} = (I - P_a) E\{\underline{y}\} = P_a^{\perp} E\{\underline{y}\} \end{aligned} \tag{18}$$

where $E\{.\}$ denotes *mathematical expectation*. The result (18) holds true irrespective the type of probability distribution one assumes for $\underline{y}$. Let us now, without assuming anything about the type of the probability distribution of $\underline{y}$, be a bit more specific about the mean $E\{\underline{y}\}$ of $\underline{y}$. We will assume that the randomness of the observation vector $\underline{y}$ is a consequence of the random nature of the measurement errors. Thus we assume that $\underline{y}$ can be written as the sum of a deterministic, i.e. non-random, component ax and a random component $\underline{e}$:

$$\boxed{\underline{y} = ax + \underline{e}} \tag{19}$$

Furthermore we assume that the mean $E\{\underline{e}\}$ of $\underline{e}$ is zero. This assumption seems acceptable if the measurement errors are zero "on the average".

With (19) the assumption $E\{\underline{e}\}=0$ implies that:

(20)
$$\boxed{E\{\underline{y}\} = \boldsymbol{a}x}$$

Substitution of (20) into (18) gives then:

(21)
$$\boxed{\begin{aligned} E\{\underline{\hat{x}}\} &= x \\ E\{\underline{\hat{y}}\} &= \boldsymbol{a}x = E\{\underline{y}\} \\ E\{\underline{\hat{e}}\} &= E\{\underline{y}\} - \boldsymbol{a}x = 0 \end{aligned}}$$

This important result shows that if (19) with (20) holds, the least-squares estimators $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$ are *unbiased* estimators. Note that this unbiasedness property of least-squares estimators is independent of the choice for the weight matrix W ! Let us now derive the covariance matrices of the estimators $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$. We will denote the covariance matrix of $\underline{y}$ as Q_y, the variance of $\underline{\hat{x}}$ as $\sigma_{\hat{x}}^2$ and the covariance matrices of $\underline{\hat{y}}$ and $\underline{\hat{e}}$ as $Q_{\hat{y}}$ and $Q_{\hat{e}}$ respectively. Application of the *"propagation law of covariances"* to (17) gives:

(22)
$$\boxed{\begin{aligned} \sigma_{\hat{x}}^2 &= (\boldsymbol{a}^*\boldsymbol{W}\boldsymbol{a})^{-1}\boldsymbol{a}^*\boldsymbol{W}\boldsymbol{Q}_y\boldsymbol{W}\boldsymbol{a}(\boldsymbol{a}^*\boldsymbol{W}\boldsymbol{a})^{-1} \\ \boldsymbol{Q}_{\hat{y}} &= \boldsymbol{a}\sigma_{\hat{x}}^2\boldsymbol{a}^* = \boldsymbol{P}_{\boldsymbol{a}}\boldsymbol{Q}_y\boldsymbol{P}_{\boldsymbol{a}}^* \\ \boldsymbol{Q}_{\hat{\boldsymbol{e}}} &= \boldsymbol{Q}_y - \boldsymbol{a}\boldsymbol{Q}_{\hat{x}y} - \boldsymbol{Q}_{yx}\boldsymbol{a}^* + \boldsymbol{a}\sigma_{\hat{x}}^2\boldsymbol{a}^* \\ &\quad \boldsymbol{Q}_y - \boldsymbol{P}_{\boldsymbol{a}}\boldsymbol{Q}_y - \boldsymbol{Q}_y\boldsymbol{P}_{\boldsymbol{a}}^* + \boldsymbol{P}_{\boldsymbol{a}}\boldsymbol{Q}_y\boldsymbol{P}_{\boldsymbol{a}}^* \end{aligned}}$$

As you know from your lecture notes in Statistics, the variance of a random variable is a measure for the spread of the probability density function around the mean value of the random variable. It is a measure of the variability one can expect to see when independent samples are drawn from the probability distribution. It is therefore in general desirable to have estimators with small variances. This is particularly true for our least-squares estimators. Together with the property of unbiasedness, a small variance would namely imply that there is a high probability that a sample from the distribution of $\underline{\hat{x}}$ will be close to the true, but unknown, value x. Since we left the choice for the weight matrix W open up till now, let us investigate what particular weight matrix makes $\sigma_{\hat{x}}^2$ minimal. Using the decomposition $Q_y=D^*D$, it follows from (22) that $\sigma_{\hat{x}}^2$ can be written as:

$$\sigma_{\hat{x}}^2 = \frac{\boldsymbol{a}^*\boldsymbol{W}\boldsymbol{D}^*\boldsymbol{D}\boldsymbol{W}\boldsymbol{a}}{(\boldsymbol{a}^*\boldsymbol{D}^{-1}\boldsymbol{D}\boldsymbol{W}\boldsymbol{a})^2} = \frac{\|\boldsymbol{D}\boldsymbol{W}\boldsymbol{a}\|^2}{\|\boldsymbol{D}\boldsymbol{W}\boldsymbol{a}\|^2\|(\boldsymbol{D}^*)^{-1}\boldsymbol{a}\|^2\cos^2\boldsymbol{\alpha}} = \frac{1}{\|(\boldsymbol{D}^*)^{-1}\boldsymbol{a}\|^2\cos^2\boldsymbol{\alpha}}$$

where α is the angle between the two vectors DWa and $(D^*)^{-1}a$. From this result follows that the variance $\sigma_{\hat{x}}^2$ is minimal when the angle α is zero; that is, when the two vectors DWa and $(D^*)^{-1}a$ are parallel. The variance $\sigma_{\hat{x}}^2$ is therefore minimal when $DWa=(D^*)^{-1}a$ or $D^*DWa=a$ or $Q_yWa=a$. Hence, the variance is minimal when we choose the weight matrix W as:

(23) $$\boxed{W = Q_y^{-1}}$$

With this choice for the weight matrix, the variance $\sigma_{\hat{x}}^2$ becomes:

(24) $$\sigma_{\hat{x}}^2 = (a^* Q_y^{-1} a)^{-1}.$$

1.6 Best Linear Unbiased Estimators (BLUE's)

Up till now we have been dealing with least-squares estimation. The least-squares estimates were derived from the *deterministic* principles of "orthogonality" and "minimum distance". That is, no probabilistic considerations were taken into account in the derivation of least-squares estimates.
Nevertheless, it was established in the previous section that least-squares estimators do have some "optimal properties" in a probabilistic sense. They are *unbiased estimators* independent of the choice for the weight matrix *W*, and their *variance can be minimized* by choosing the weight matrix as the inverse of the covariance matrix of the observations. Moreover, they are *linear* estimators, i.e. linear functions of the observations.

Estimators which are linear, unbiased and have minimum variance are called best linear unbiased estimators or simply BLUE's. "Best" in this case means minimum variance. Thus least-squares estimators are BLUE's in case the weight matrix equals the inverse of the covariance matrix of the observations. In order to find out how large the class of best linear unbiased estimators is (is there only one? or are there more than one?), we will now derive this class. Thus instead of starting from the least-squares principle of "orthogonality" we start from the conditions of linearity, unbiasedness and best. Again we assume that the mean of $\underline{y}$ satisfies:

(25) $$E\{\underline{y}\} = ax.$$

Our problem is to find an estimator $\underline{\hat{x}}$ of the unknown parameter x, where $\underline{\hat{x}}$ is a function of $\underline{y}$. Since we want the estimator $\underline{\hat{x}}$ to be a *linear* function of $\underline{y}$, $\underline{\hat{x}}$ must be of the form:

(26) $$\underline{\hat{x}} = l^* \underline{y} \qquad \text{"L-property"}$$

Furthermore, we want $\underline{\hat{x}}$ to be an *unbiased* estimator of x for all the values x may take. This means that:

(27) $$E\{\underline{\hat{x}}\} = x \ , \ \forall x \qquad \text{"U-property"}$$

And finally, we want to find those estimators $\underline{\hat{x}}$, which in the class of estimators that satisfy (26) and (27), have *minimum variance:*

(28) $\sigma_{\hat{x}}^2$ minimal in the class of LU-estimators "B-property"

It will be clear that the class of linear estimators is infinite. Let us therefore look at the class of linear and unbiased estimators.

From taking the expectation of (26) follows, with (25) and (27) that:

$$E\{\underline{\hat{x}}\} = l^* E\{\underline{y}\} = l^* \boldsymbol{a} x = x \quad \forall x$$

or

(29) $$\boxed{l^* \boldsymbol{a} = 1}$$

Thus for LU-estimators the vector l has to satisfy (29) (see figure 1.19).

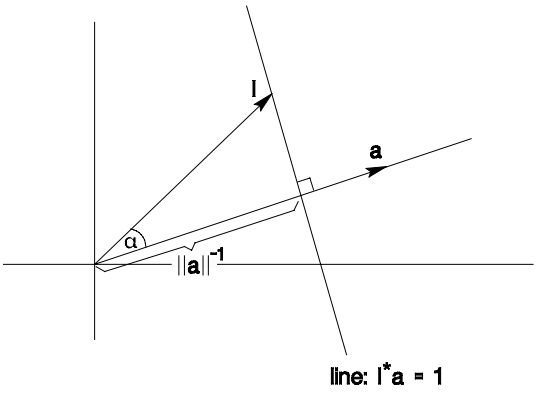


Figure 1.19: $l^* \boldsymbol{a} = \|l\|\|\boldsymbol{a}\|\cos\alpha = 1$

It will be clear that, although the class of permissable vectors l has been narrowed down to the line $l^*a=1$, the class of LU-estimators is still infinite. For instance, every vector l of the form:

$$l = c(\boldsymbol{a}^* c)^{-1} ,$$

where c is arbitrary, is permitted.

Let us now include the "B-property". The variance of the estimator $\underline{\hat{x}}$ follows from applying the "propagation law of variances" to (26). This gives:

(30) $$\sigma_{\hat{x}}^2 = l^* \boldsymbol{Q}_y l$$

The "B-property" implies now that we have to find the vectors l, let us denote them by $\hat{l}$, that minimize $\sigma_{\hat{x}}^2$ of (30) and at the same time satisfy (29). We thus have the minimization problem:

(31) $$\boxed{\min_l \; l^* \boldsymbol{Q}_y l \quad \text{subject to} \quad l^* \boldsymbol{a} = 1}$$

With the decomposition $Q_y=D^*D$, this can also be written as:

$$\min_l \; (\boldsymbol{D} l)^* (\boldsymbol{D} l) \quad \text{subject to} \quad (\boldsymbol{D} l)^* (\boldsymbol{D}^*)^{-1} \boldsymbol{a} = 1$$

or as:

(32) $$\min_{\bar{l}} \bar{l}^* \bar{l} \quad \text{subject to} \quad \bar{l}^* (D^*)^{-1} a = 1 \ ,$$

where:

(33) $$\bar{l} = Dl.$$

Geometrically, the minimization problem (32) boils down to finding those vectors $\hat{\bar{l}}$ whose endpoints lie along the line $\bar{l}^*(D^*)^{-1}a=1$ and who have minimum length (see figure 1.20).

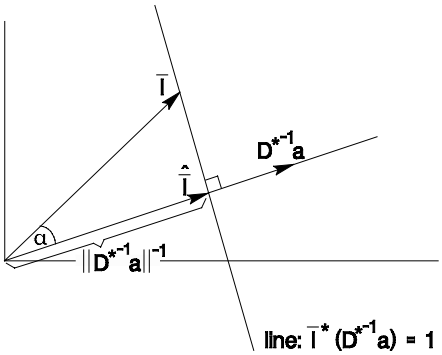


Figure 1.20: $\bar{l}^*(D^*)^{-1}a = \|\bar{l}\| \, \|(D^*)^{-1}a\| \cos \alpha = 1$

From figure 1.20 it is clear that there is only one such vector $\hat{\bar{l}}$. This vector lies along vector $D^{*-1}a$ and has length $\|(D^*)^{-1}a\|^{-1}$. Thus:

(34) $$\hat{\bar{l}} = (D^*)^{-1} a \beta \ ,$$

where the positive scalar β follows from:

$$\hat{\bar{l}}^* \hat{\bar{l}} = \|(D^*)^{-1}a\|^2 \ \beta^2 = \|(D^*)^{-1}a\|^{-2}$$

as:

(35) $$\beta = \|(D^*)^{-1}a\|^{-2}$$

Substitution of (35) into (34) gives with (33):

$$\hat{l} = D^{-1}(D^*)^{-1}a \left[a^* D^{-1}(D^*)^{-1}a\right]^{-1}$$

or:

(36) $$\boxed{\hat{l} = Q_y^{-1} a \ (a^* Q_y^{-1} a)^{-1}}$$

The best linear unbiased estimator $\underline{\hat{x}}$ of x follows therefore with (26) as:

(37) $$\underline{\hat{x}} = (a^* Q_y^{-1} a)^{-1} \ a^* Q_y^{-1} \underline{y}$$

From this result we can draw the important conclusion that the best linear unbiased estimator is *unique* and equal to the least-squares estimator if $W=Q_y^{-1}$!

We will now give an *alternative* derivation of the solution (37). Equation (29) is the equation of a line with a direction vector orthogonal to a (see figure 1.20a). Since the vector b

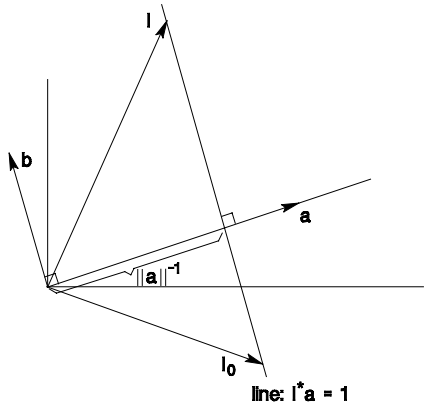


Figure 1.20a: $l^*a=1$ is equivalent to $l=l_o+b\alpha$

was defined as being orthogonal to a, $b^*a=0$, the line $a^*l=1$ can be represented parametrically as:

(37a)
$$\underset{2\times1}{l} = \underset{2\times1}{l_o} + \underset{2\times1}{b}\ \underset{1\times1}{\alpha}\ ,$$

where l_o is a particular solution of (29). Hence, the condition $a^*l=1$ in (31) may be replaced by $l=l_o+b\alpha$:

$$\min_{l}.\ l^*Q_y l \quad \text{subject to} \quad l = l_o + b\alpha$$

This may also be written as:

$$\min_{\alpha}.\ (l_o + b\alpha)^*Q_y(l_o + b\alpha).$$

The solution of this minimization problem is:

(37b)
$$\hat{\alpha} = -(b^*Q_y b)^{-1} b^*Q_y l_o.$$

This is easily verified by using an algebraic proof like the one given in section 1.2. Substitution of (37b) into (37a) gives:

$$\hat{l} = \left[I - b(b^*Q_y b)^{-1} b^*Q_y\right] l_o$$

The best linear unbiased estimator $\underline{\hat{x}}$ of x follows therefore with (26) as:

(37c) $$\underline{\hat{x}} = l_o^* \left[I - Q_y b (b^* Q_y b)^{-1} b^* \right] \underline{y}.$$

From equations (8) and (9) in section 1.3 we know that:

$$I - Q_y b (b^* Q_y b)^{-1} b^* = a (a^* Q_y^{-1} a)^{-1} a^* Q_y^{-1}.$$

Substitution in (37c) gives therefore:

$$\underline{\hat{x}} = l_o^* a (a^* Q_y^{-1} a)^{-1} a^* Q_y^{-1} \underline{y}$$

or:

$$\underline{\hat{x}} = (a^* Q_y^{-1} a)^{-1} a^* Q_y^{-1} \underline{y},$$

since $l_o^* a = 1$. This concludes our alternative derivation of (37).

1.7 The Maximum Likelihood (ML) method

In the previous sections we have seen the two principles of *least-squares* estimation and *best linear unbiased* estimation at work. The least-squares principle is a deterministic principle where no probability assumptions concerning the vector of observations is made. In case $y=ax+e$, the least-squares principle is based on minimizing the quadratic form $(y-ax)^* W(y-ax)$, with the weight matrix W.

In contrast with the least-squares principle, the principle of best linear unbiased estimation does make use of some probability assumptions. In particular it is based on assumptions concerning the mean and covariance matrix of the random vector of observables $\underline{y}$. However, only the mean and covariance matrix of $\underline{y}$ need to be specified, not its complete probability distribution. We will now present a method of estimation that in contrast with the BLUE's method needs the complete probability distribution of $\underline{y}$.

Let $p(y;x)$ denote the probability density function of $\underline{y}$. For a fixed value of x the function $p(y;x)$ is a function of y and for different values of x the function $p(y;x)$ may take different forms (see figure 1.21).

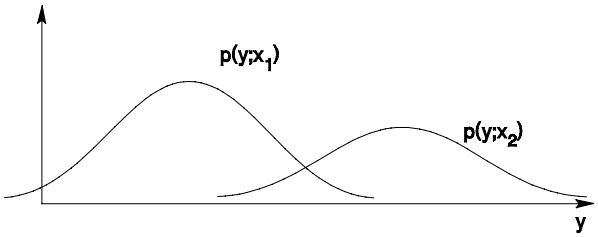


Figure 1.21: Two one-dimensional probability density functions $p(y;x_1)$, $p(y;x_2)$

Our objective is now to determine or estimate the parameter x on the basis of the observation vector y. That is, we wish to determine the correct value of x that produced the observed y. This suggests considering for each possible x how probable the observed y would be if x were the true value. The higher this probability, the more one is attracted to the explanation that the x in question produced y, and the more likely the value of x appears. Therefore, the expression $p(y;x)$ considered for fixed y as a function of x has been called the *likelihood* of x. Note that the likelihood of x is a probability density or a differential probability. The estimation principle is now to maximize the likelihood of x, given the observed value of y:

(38)
$$\max_x \; p(y;x)$$

Thus in the *method of maximum likelihood* we choose as an estimate of x the value which maximizes $p(y;x)$ for the given observed y. Note that in the ML-method one needs to know the complete mathematical expression for the probability density function $p(y;x)$ of $\underline{y}$. Thus it does not suffice, as is the case with the BLUE's method, to specify only the mean and covariance matrix of $\underline{y}$.

As an important example we consider the case that $p(y;x)$ is the probability density function of the normal distribution. Let us therefore assume that $\underline{y}$ is normally distributed with mean $E\{\underline{y}\}=ax$ and covariance matrix Q_y. The probability density function of $\underline{y}$ reads then in case it is a two-dimensional vector:

(39)
$$p(y;x) = (2\pi)^{-1} |Q_y|^{-1/2} \exp[-\tfrac{1}{2}(y-ax)^* Q_y^{-1}(y-ax)].$$

As it is often more convenient to work with $\ln p(y;x)$ then with $p(y;x)$ itself, the required value for x is found by solving:

(40)
$$\max_x \; \ln p(y;x).$$

this is permitted since ln $p(y;x)$ and $p(y;x)$ have their maxima at the same values of x. Taking the logarithm of (39) gives:

$$\ln p(y;x) = -\ln(2\pi) - \tfrac{1}{2}\ln|Q_y| - \tfrac{1}{2}(y-ax)^* Q_y^{-1}(y-ax). \tag{41}$$

Since the first two terms on the right-hand side of (41) are constant, it follows that maximization of ln $p(y;x)$ amounts to maximization of $-1/2(y-ax)^* Q_y^{-1}(y-ax)$. But this is equivalent to minimization of $(y-ax)^* Q_y^{-1}(y-ax)$. Hence we have obtained the important conclusion that maximum likelihood estimators, in case $p(y;x)$ is the normal distribution, are identical to least-squares estimators for which the weight matrix is $W=Q_y^{-1}$!

Let us now exemplify the above discussion geometrically. From (39) follows that the contours of constant probability density are ellipses:

$$p(y;x) = \text{constant} \quad \Leftrightarrow \quad (y-ax)^* Q_y^{-1}(y-ax) = \text{constant} = c.$$

For different values of c we get different ellipses. The *smaller* the probability density, the *larger* the value of c and the larger the ellipses are (see figure 1.22).

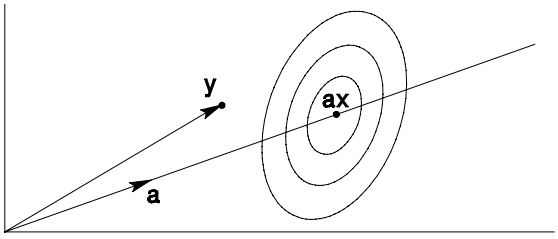


Figure 1.22: Contours of $p(y;x)$ = constant

By varying the values of x the family of contour lines centred at ax translate along the line with direction vector a (see figure 1.23).

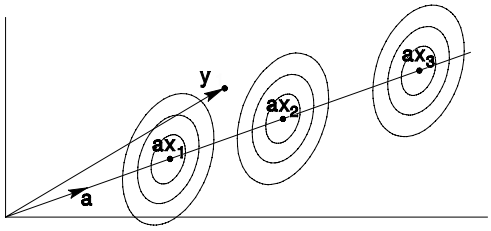


Figure 1.23

From the figure it is clear that the likelihood of x obtains its maximum there where the observation point y is a point of tangency of one of the ellipses (see figure 1.24).

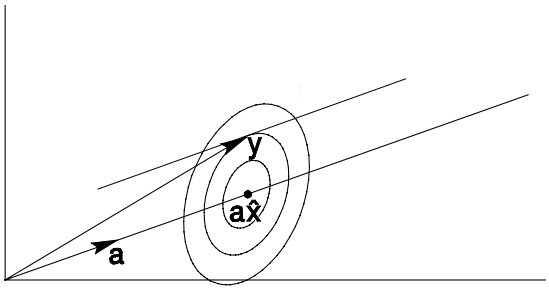


Figure 1.24: The ML-estimate $\hat{x}$

Hence, after translating the family of ellipses along the line with direction vector a such that y becomes a point of tangency, the corresponding centre $a\hat{x}$ of the ellipses gives the maximum likelihood estimate $\hat{x}$. At the point of tangency y, the normal to the ellipse is $Q_y^{-1}(y-a\hat{x})$ (see figure 1.25).

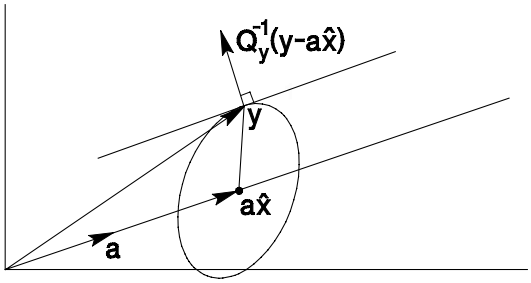


Figure 1.25: The normal $Q_y^{-1}(y-a\hat{x})$

The normal $Q_y^{-1}(y-a\hat{x})$ is orthogonal to the tangent line at y. But this tangent line is parallel to the line with direction vector a. Thus we have $a^* Q_y^{-1}(y-a\hat{x})=0$, which shows once again that in case of the normal distribution, ML-estimators and LSQ-estimators with $W=Q_y^{-1}$ are identical!

1.8 Summary

In this chapter we have discussed three different methods of estimation: the *LSQ*-method, the *BLUE*-method and the *ML*-method. For an overview of the results see table 1. The LSQ-method, being a deterministic method, was applied to the model:

(42)
$$\underset{2\times 1}{y} = \underset{2\times 1}{\boldsymbol{a}}\ \underset{1\times 1}{x} + \underset{2\times 1}{\boldsymbol{e}}\ .$$

In case the measurement errors are considered to be random, $\underline{e}$, the observables also become random, $\underline{y}$, and (42) has to be written as:

(43) $$\underline{y} = \boldsymbol{a}x + \underline{\boldsymbol{e}}.$$

With the assumptions that $E\{\underline{e}\}=0$ and $E\{\underline{e}\underline{e}^*\}=Q_y$, we get with (43) the model:

(44) $$E\{\underline{y}\} = \boldsymbol{a}x \;;\; E\{(\underline{y}-\boldsymbol{a}x)(\underline{y}-\boldsymbol{a}x)^*\} = \boldsymbol{Q}_y \;.$$

These two assumption were the starting point of the BLUE-method. If in addition to (44), it is also assumed that $\underline{y}$ is normally distributed then the model becomes:

(45) $$\underline{y} \sim N(\boldsymbol{a}x, \boldsymbol{Q}_y)$$

It was shown that for this model the ML-estimator is identical to the BLU-estimator and the LSQ-estimator if $W=Q_y^{-1}$.

We have restricted ourselves in this introductory chapter to the case of two observations y_1, y_2 and one unknown x. In the next chapter we will start to generalize the theory to higher dimensions. That is, in the next chapter we will start developing the theory for m observations y_1, y_2,...,y_m and n unknowns x_1, x_2,...,x_n. In that case model (45) generalizes to:

(46) $$\underset{m\times 1}{\underline{y}} \sim N(\underset{m\times n}{\boldsymbol{A}} \; \underset{n\times 1}{x} \; , \underset{m\times m}{\boldsymbol{Q}_y}),$$

where A is an $m\times n$ matrix, i.e. it has m rows and n columns.

Least Squares Estimation (LSQ)	Best Linear Unbiased Estimation (BLUE)	Maximum Likelihood Estimation (ML)
model: $\boxed{y = ax + e}$ *estimation principle*: $\min_x (y-ax)^* W (y-ax)$ *solution:* $\hat{x} = (a^* W a)^{-1} a^* W y$ $\hat{y} = a\hat{x} = P_a y$ $\hat{e} = y - a\hat{x} = P_a^{\perp} y$ $\hat{e}^* W \hat{e} = y^* W P_a^{\perp} y$ $P_a = a(a^* W a)^{-1} a^* W$ $P_a^{\perp} = I - P_a$	*model*: $\boxed{\begin{array}{c} \underline{y} = ax + \underline{e} \\ E\{\underline{e}\} = 0 \;;\; E\{\underline{e}\underline{e}^*\} = Q_y \end{array}}$ or $\boxed{E\{\underline{y}\} = ax \;;\; E\{(\underline{y}-ax)(\underline{y}-ax)^*\} = Q_y}$ *estimation principle*: *L*: $\underline{\hat{x}} = l^* \underline{y}$; *U*: $E\{\underline{\hat{x}}\} = x$; ***B***: $\sigma_{\hat{x}}^2$ min. *solution*: $\underline{\hat{x}} = (a^* Q_y^{-1} a)^{-1} a^* Q_y^{-1} \underline{y}$; $\sigma_{\hat{x}}^2 = (a^* Q_y^{-1} a)^{-1}$ $\underline{\hat{y}} = a\underline{\hat{x}} = P_a \underline{y}$; $Q_{\hat{y}} = P_a Q_y$ $\underline{\hat{e}} = y - a\underline{\hat{x}} = P_a^{\perp} \underline{y}$; $Q_{\hat{e}} = P_a^{\perp} Q_y$ $\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}} = \underline{y}^* Q_y^{-1} P_a^{\perp} \underline{y}$; $P_a = a(a^* Q_y^{-1} a)^{-1} a^* Q_y^{-1}$ $P_a^{\perp} = I - P_a$	*model*: $\boxed{\underline{y} \sim p(y;x)}$ *estimation principle*: $\max_x p(y;x)$ In case $\boxed{\underline{y} \sim N(ax, Q_y)}$ the solution is identical to BLUE and to LSQ if $W = Q_y^{-1}$

Table 1.1

2 The model with observation equations

2.1 Introduction

As an introduction consider the following example: Figure 2.1 shows a *levelling network*. The points 0,1,2,3,4 and 5 are connected with levelled height differences. The total number of levelled height differences is 9. The height of point 0 is assumed to be known in the *NAP-reference system* (NAP= Normaal Amsterdams Peil). The heights of the remaining points are unknown. Our objective is to determine their heights from the measured height differences.

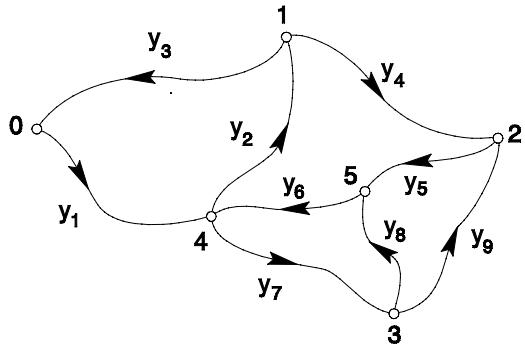


Figure 2.1: Levelling network

One way to determine their heights is to use the configuration of figure 2.2. One starts at the reference point 0 of which the height is known, and then determines the heights of the points 4,5,3,2 and 1 as shown in figure 2.2. In this case only 5 out of the 9 available levelled height differences are used, namely y_1, y_6, y_8, y_9 and y_4.

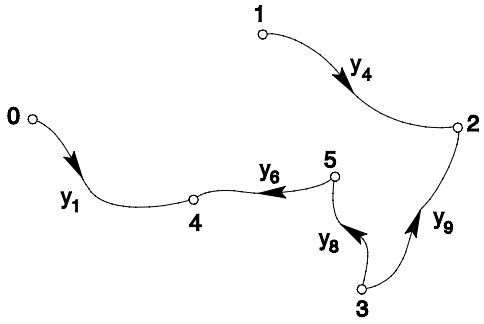


Figure 2.2

An alternative way to determine the heights of the points is to use the configuration of figure 2.3. One starts again at the reference point 0, and then determines the heights of the points 1,2,5,3 and 4 as shown in figure 2.3. In this case again 5 out of the 9 available levelled height differences are used. Note however that the height differences used differ from the ones used in the previous case.

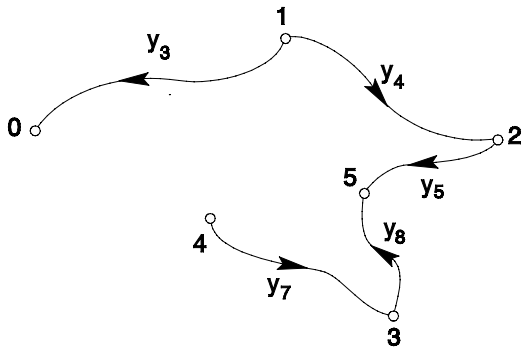


Figure 2.3

It will be clear that the two solutions of figure 2.2 and figure 2.3 are identical in case the height differences were *perfectly* levelled. However, perfect measurements do not exist! Hence, since one relies in the two cases considered on different observations with different measurement errors, the computed heights will also differ. Therefore, if one wants to make use of all the 9 levelled height differences, one has to take care in one way or another of the resulting inconsistencies or discrepancies.

One could also argue that in the case of figure 2.1 too many height differences are measured. This is true as far as the construction of the levelling network is concerned. That is, strictly speaking only 5 out of the 9 available height differences are needed to determine the 5 unknown heights. However, with 5 measured height differences one will *never* be able to detect *blunders or outliers* in the measurements. This is one of the reasons why always more measurements are taken than strictly necessary for determining the geometric figure of the network. In the course of your studies in *Mathematical Geodesy* you will learn to answer questions as: how many measurements are needed, how should they be distributed over the network etc. That is, in the course of your studies you will learn, given the purpose of a network, how to *design* a network with a certain *quality*.

In the above example of figure 2.1 we have 9 measured height differences and 5 unknown heights. As we did in the introductory section of chapter 1, we can write down the 9 equations that relate the measurements to the unknowns. If we denote the unknown heights as x_1, x_2,...,x_5 and the known height of point 0 as x_0, we obtain with the help of figure 2.1:

$$
\begin{aligned}
\underline{y}_1 &= x_4 - x_0 + \underline{e}_1 \\
\underline{y}_2 &= x_1 - x_4 + \underline{e}_2 \\
\underline{y}_3 &= x_0 - x_1 + \underline{e}_3 \\
\underline{y}_4 &= x_2 - x_1 + \underline{e}_4 \\
\underline{y}_5 &= x_5 - x_2 + \underline{e}_5 \\
\underline{y}_6 &= x_4 - x_5 + \underline{e}_6 \\
\underline{y}_7 &= x_3 - x_4 + \underline{e}_7 \\
\underline{y}_8 &= x_5 - x_3 + \underline{e}_8 \\
\underline{y}_9 &= x_2 - x_3 + \underline{e}_9
\end{aligned} \tag{1}
$$

where the $\underline{e}_i$, i=1,...,9, are the unknown random measurements errors.The above 9 equations can be written in matrix notation as:

$$
\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \\ \underline{y}_6 \\ \underline{y}_7 \\ \underline{y}_8 \\ \underline{y}_9 \end{pmatrix} = \begin{pmatrix} & & & 1 & \\ 1 & & & -1 & \\ -1 & & & & \\ -1 & 1 & & & \\ & -1 & & & 1 \\ & & & 1 & -1 \\ & & 1 & -1 & \\ & & -1 & & 1 \\ & 1 & -1 & & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} + \begin{pmatrix} -x_0 \\ 0 \\ x_0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \underline{e}_3 \\ \underline{e}_4 \\ \underline{e}_5 \\ \underline{e}_6 \\ \underline{e}_7 \\ \underline{e}_8 \\ \underline{e}_9 \end{pmatrix} \tag{2}
$$

For simplicity we will assume that the known height x_0 of the reference point 0 is equal to zero. The matrix equation of (2) is then of the form:

$$
\underset{m\times 1}{\underline{y}} = \underset{m\times n}{\boldsymbol{A}} \; \underset{n\times 1}{x} + \underset{m\times 1}{\underline{e}}\,, \tag{3}
$$

with m=9 and n=5. Matrix A thus has 9 rows and 5 columns. If the known height x_0 is unequal to zero, then (2) can still be brought into the form of (3) by bringing the known vector containing x_0 in (2) to the left-hand side.

Equation (3) is the multi dimensional generalization of equation (19) in section 1.5. In this chapter we will generalize the theory of chapter 1 for multi dimensional cases.

2.2 Least-squares estimation

The equations of the linear system:

(4) $$\underset{m\times 1}{\underline{y}} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x} + \underset{m\times 1}{\underline{\boldsymbol{e}}},$$

are known as *observation equations* (waarnemingsvergelijkingen). Let us assume that an observation vector y is given, and that we want to solve for the unknown vectors x and e in the system:

(5) $$\underset{m\times 1}{y} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x} + \underset{m\times 1}{\boldsymbol{e}}.$$

If we denote the column vectors of the matrix A by a_i, i=1,...,n, then A=(a_1, a_2,....,a_n), and (5) may be written as:

(6) $$\underset{m\times 1}{y} = \sum_{i=1}^{n} \underset{m\times 1}{\boldsymbol{a_i}} x_i + \underset{m\times 1}{\boldsymbol{e}}$$

In chapter 1 we solved equation (6) in a least-squares sense for the case n=1. We will now consider the general case that n may be arbitrary, provided that $n \le m$. That is, the restriction is that the number of observations, m, has to be larger than or equal to the number of unknown parameters, n. For m=3 and n=2 we may visualize equation (6) as in figure 2.4.

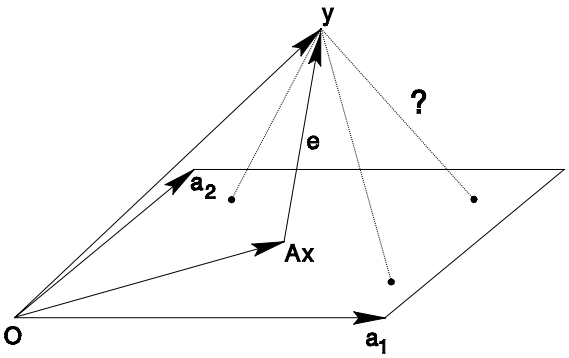


Figure 2.4: x and e unknown in y=Ax+e.

From the geometry of the figure it seems intuitively appealing to estimate x as $\hat{x}$, such that $\sum_{i=1}^{n} a_i\hat{x}_i$ or $A\hat{x}$ is as close as possible to the given observation vector y. This is the case if $\hat{e}=y-\sum_{i=1}^{n} a_i\hat{x}_i$ is *orthogonal* to the "plane" spanned by the column vectors a_i, i=1,...,n, of A.

That is, if:

$$\boldsymbol{a_j}^{*}\left(y-\sum_{i=1}^{n}\boldsymbol{a_i}\hat{x}_i\right)=0\ ,\ j=1,..,n$$

or if:

$$a_1^*(y - \sum_{i=1}^{n} a_i \hat{x}_i) = 0$$

$$a_2^*(y - \sum_{i=1}^{n} a_i \hat{x}_i) = 0$$

$$\vdots \qquad\qquad \vdots$$

$$a_n^*(y - \sum_{i=1}^{n} a_i \hat{x}_i) = 0$$

or if:

$$A^*(y - A\hat{x}) = 0$$

or if:

(7)

$$\underset{n\times n}{A^*A}\ \underset{n\times 1}{\hat{x}} = \underset{n\times m}{A^*}\ \underset{m\times 1}{y}$$

This is a system of n linear equations with n unknowns. If the matrix A^*A is invertible, the solution of (7) is unique and reads:

(8)

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

Compare this result with equation (2) in section 1.2. The $n\times n$ matrix A^*A is invertible if and only if rank $A^*A=n$. And since rank A^*A = rank A (for a proof see the end of this section), it follows that A^*A is invertible if rank $A=n$. We will therefore assume that matrix A has *full rank n*. Verify that the matrix of equation (2) has full rank; its rank is 5. Verify also that if the height of the reference point 0 is unknown, the matrix which follows from including x_o in the vector of unknowns has a *rankdefect* of 1. The reason is of course that "absolute heights" cannot be determined from height differences only. The derivation of the estimate $\hat{x}$ of (8) was based on the principle of orthogonality. The estimate $\hat{x}$ however also minimizes the sum of squares $(y\text{-}Ax)^*(y\text{-}Ax)$. That is, $\hat{x}$ of (8) is the solution of:

(9)

$$\min_x \ (y - Ax)^*(y - Ax)$$

This is easily shown, by using the same algebraic derivation as was used in section 1.2. Along the same lines one can show that the solution of:

(10)

$$\boxed{\min_x \ (y - Ax)^*W(y - Ax)}\ .$$

where W is an $m\times m$ symmetric and positive definite weight matrix, is:

(11)

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

Proof: (algebra)

$$\begin{aligned}(y-Ax)^*W(y-Ax) &= [y-A\hat{x}-A(x-\hat{x})]^*W[y-Ax-A(x-\hat{x})]\\ &= (y-A\hat{x})^*W(y-A\hat{x}) - 2(y-A\hat{x})^*WA(x-\hat{x}) + (x-\hat{x})^*A^*WA(x-\hat{x}).\end{aligned}$$

$$\text{But } A^*W(y-A\hat{x}) = A^*W[(y-A(A^*WA)^{-1}A^*Wy)] = 0, \text{ hence,}$$

$$\begin{aligned}(y-Ax)^*W(y-Ax) &= (y-A\hat{x})^*W(y-A\hat{x})+(x-\hat{x})^*A^*WA(x-\hat{x})\\ &= \text{minimal for } x=\hat{x}\end{aligned}$$

End of proof.

The estimate $\hat{x}$ of (11) is the *weighted least-squares estimate* of x. As a generalization of equation (7) in section 1.3, the weighted least-squares estimators read therefore for the multi dimensional case:

(12)

$$\boxed{\begin{aligned}\underline{\hat{x}} &= (\boldsymbol{A}^*W\boldsymbol{A})^{-1}\boldsymbol{A}^*W\underline{y}\\ \underline{\hat{y}} &= \boldsymbol{A}\underline{\hat{x}} = \boldsymbol{A}(\boldsymbol{A}^*W\boldsymbol{A})^{-1}\boldsymbol{A}^*W\underline{y}\\ \underline{\hat{e}} &= \underline{y}-\underline{\hat{y}} = [I-\boldsymbol{A}(\boldsymbol{A}^*W\boldsymbol{A})^{-1}\boldsymbol{A}^*W]\underline{y}\\ \underline{\hat{e}}^*W\underline{\hat{e}} &= \underline{y}^*[W-W\boldsymbol{A}(\boldsymbol{A}^*W\boldsymbol{A})^{-1}\boldsymbol{A}^*W]\underline{y}\end{aligned}}$$

***Intermezzo*:**

(13)
$$\textbf{rank } \boldsymbol{A} = \textbf{rank } \boldsymbol{A}^*\boldsymbol{A}$$

From theory in Linear Algebra we know that:

$$\text{rank } \boldsymbol{A} = n - \text{dim.} N(\boldsymbol{A}),$$

where $N(A)$ denotes the *nullspace* (kern) of the $m\times n$ matrix A. For A^*A we have:

$$\text{rank } \boldsymbol{A}^*\boldsymbol{A} = n - \text{dim.} N(\boldsymbol{A}^*\boldsymbol{A}).$$

Hence, rank A = rank A^*A is equivalent to:

(14)
$$\text{dim.} N(\boldsymbol{A}) = \text{dim.} N(\boldsymbol{A}^*\boldsymbol{A})$$

The nullspace of A is defined as the set of vectors x that satisfy:

$$\boldsymbol{A}x = \boldsymbol{0}.$$

Since $A^*Ax=0$ if $Ax=0$, it follows that:

(15)
$$N(\boldsymbol{A}) \subset N(\boldsymbol{A}^*\boldsymbol{A})$$

But also:

(16) $$N(\boldsymbol{A}^{*}\boldsymbol{A}) \subset N(\boldsymbol{A})$$

holds. Because if $x \in N(A^*A)$ then $A^*Ax=0$, and this implies that $Ax=0$ or $x \in N(A)$. $A^*Ax=0$ can never imply that $Ax \neq 0$. From (15) and (16) follows that $N(A)=N(A^*A)$ and thus that dim. $N(A)$ = dim $N(A^*A)$ or rank A = rank A^*A.

End of proof.

2.3 The mean and covariance matrix of least-squares estimators

In section 5 of chapter 1 we derived the means and covariance matrices of the least-squares estimators of the model $\underline{y}=ax+\underline{e}$. This was done through application of the propagation laws for means and variances to (17) in section 1.5. In the present section we want to generalize these results. That is, we want to derive the mean and covariance matrix of the estimators of (12). This is again done through application of the propagation laws for means and variances. For completeness sake we first state and prove three laws for the general multi-dimensional case.

Let $\underline{v}$ be a $p\times 1$ dimensional random vector and let $\underline{u}$ be a $q\times 1$ dimensional random vector. Furthermore let $\underline{u}$ and $\underline{v}$ be functionally related by:

(17) $$\underline{v} = F(\underline{u}),$$

where $F(.)$ is a vector function.

The mean of $\underline{v}$, $E\{\underline{v}\}$, is then *defined* as (see also appendices E and F):

(18) $$\boxed{E\{\underline{v}\} = \int F(u)\,\boldsymbol{p}_{\underline{u}}(u)\,du.}$$

where $\boldsymbol{p}_{\underline{u}}(u)$ is the probability density function of $\underline{u}$. Equation (18) shows that the mean of $\underline{v}$ can be computed once the vector function $F(.)$ and probability density function $\boldsymbol{p}_{\underline{u}}(.)$ are known. We will now use (18) to derive the two propagation laws of means and variances for the case of *linear* functions.

The propagation law of means

Let the two random vectors $\underline{u}$ and $\underline{v}$ be related as:

(19) $$\boxed{\underset{p\times 1}{\underline{v}} = \underset{p\times q}{M}\ \underset{q\times 1}{\underline{u}} + \underset{p\times 1}{m}}$$

where M is a constant $p \times q$ matrix and m is a constant $p \times 1$ vector. Using (18) we get for the mean of $\underline{v}$:

$$
\begin{aligned}
E\{\underline{v}\} &= \int (Mu + m)\, p_{\underline{u}}(u)\, du = \int Mu\, p_{\underline{u}}(u)\, du + \int m\, p_{\underline{u}}(u)\, du \\
&= M \int u\, p_{\underline{u}}(u)\, du + m \int p_{\underline{u}}(u)\, du
\end{aligned} \tag{20}
$$

According to (18), $\int u\, p_{\underline{u}}(u)\, du$ equals the mean of $\underline{u}$, $E\{\underline{u}\}$. Furthermore, since the "area" or "volume" of a probability density function equals 1, we have that $\int p_{\underline{u}}(u)\, du = 1$. Hence, (20) reduces to:

$$
\underset{p\times 1}{E\{\underline{v}\}} = \underset{p\times q}{M}\ \underset{q\times 1}{E\{\underline{u}\}} + \underset{p\times 1}{m} \tag{21}
$$

This result shows that in case $\underline{u}$ and $\underline{v}$ are *linearly* related, the mean of $\underline{v}$, $E\{\underline{v}\}$, can be computed once the mean of $\underline{u}$, $E\{\underline{u}\}$, is known! This is generally *not* the case for nonlinear functions.

The propagation law of variances and covariances

The covariance matrix of the random vector $\underline{v}$ is *defined* as:

$$
Q_v = E\{(\underline{v} - E\{\underline{v}\})(\underline{v} - E\{\underline{v}\})^*\}.
$$

If $\underline{u}$ and $\underline{v}$ are related as (19), it follows with (21) that:

$$
\begin{aligned}
Q_v &= E\{(\underline{v} - E\{\underline{v}\})(\underline{v} - E\{\underline{v}\})^*\} = E\{(M\underline{u} + m - ME\{\underline{u}\} - m)(M\underline{u} + m - ME\{\underline{u}\} - m)^*\} \\
&= E\{[M(\underline{u} - E\{\underline{u}\})][M(\underline{u} - E\{\underline{u}\})]^*\} \\
&= E\{M(\underline{u} - E\{\underline{u}\})(\underline{u} - E\{\underline{u}\})^* M^*\}
\end{aligned}
$$

Using (18) this gives:

$$
\begin{aligned}
Q_v &= \int M(u - E\{\underline{u}\})(u - E\{\underline{u}\})^* M^*\, p_{\underline{u}}(u)\, du \\
&= M \int (u - E\{\underline{u}\})(\underline{u} - E\{\underline{u}\})^*\, p_{\underline{u}}(u)\, du\, M^* \\
&= M E\{(\underline{u} - E\{\underline{u}\})(\underline{u} - E\{\underline{u}\})^*\} M^*
\end{aligned} \tag{22}
$$

But $E\{(\underline{u}\text{-}E\{\underline{u}\})(\underline{u}\text{-}E\{\underline{u}\})^*\}$ is by definition the covariance matrix of $\underline{u}$, Q_u. Hence, (22) reduces to:

(23)
$$\underset{p\times p}{Q_v} = \underset{p\times q}{M}\ \underset{q\times q}{Q_u}\ \underset{q\times p}{M^*}$$

This result shows that in case $\underline{u}$ and $\underline{v}$ are *linearly* related, the covariance matrix of $\underline{v}$, Q_v, can be computed once the covariance matrix of $\underline{u}$, Q_u, is known! Thus in case of linear functions one does not need to know the complete probability density function of $\underline{u}$.

If we denote the covariance between $\underline{u}$ and $\underline{v}$ as Q_{uv}, then by *definition:*

$$Q_{uv} = E\{(\underline{u} - E\{\underline{u}\})(\underline{v} - E\{\underline{v}\})^*\}.$$

With a derivation similar to the one given above folllows then that:

(24)
$$\underset{q\times p}{Q_{uv}} = \underset{q\times q}{Q_u}\ \underset{q\times p}{M^*}$$

Prove this yourself!

It will now be clear since the first three estimators of (12) are all *linear* functions of $\underline{y}$, and thus of the form (19), that their means and covariance matrices can be derived from the results of (21), (23) and (24).

The model used for deriving the estimators of (12) was:

(25)
$$\underset{m\times 1}{\underline{y}} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x} + \underset{m\times 1}{\underline{e}}$$

We will now assume that $E\{\underline{e}\}=0$ and $E\{\underline{e}\ \underline{e}^*\}=Q_y$. With (25) this gives:

(26)
$$E\{\underline{y}\} = \boldsymbol{A}x \ ; \ E\{(\underline{y} - \boldsymbol{A}x)(\underline{y} - \boldsymbol{A}x)^*\} = Q_y$$

With the first equation of (26), application of the *propagation law for means*, (21), to (12) gives:

(27)
$$\begin{aligned} E\{\underline{\hat{x}}\} &= x \\ E\{\underline{\hat{y}}\} &= \boldsymbol{A}x \\ E\{\underline{\hat{e}}\} &= \boldsymbol{0} \end{aligned}$$

Compare this result with equation (21) in section 1.5. This important result shows that if (26) holds, the least-squares estimators are *unbiased* estimators. And this property of unbiasedness is independent of the choice for W.

With the second equation of (26), application of the *propagation law for variances*, (23), to (12) gives:

(28)
$$\begin{aligned} Q_{\hat{x}} &= (A^*WA)^{-1}A^*WQ_yWA(A^*WA)^{-1} \\ Q_{\hat{y}} &= AQ_{\hat{x}}A^* \\ Q_{\hat{e}} &= [I - A(A^*WA)^{-1}A^*W]\,Q_y\,[I - WA(A^*WA)^{-1}A^*] \end{aligned}$$

In a similar way, application of the *propagation law of covariances*, (24), to (12) gives:

(29)
$$\begin{aligned} Q_{\hat{x}\hat{y}} &= (A^*WA)^{-1}A^*WQ_yWA(A^*WA)^{-1}A^* = Q_{\hat{x}}A^* \\ Q_{\hat{x}\hat{e}} &= (A^*WA)^{-1}A^*WQ_y - (A^*WA)^{-1}A^*WQ_yWA(A^*WA)^{-1}A^* \\ Q_{\hat{y}\hat{e}} &= A(A^*WA)^{-1}A^*WQ_y - A(A^*WA)^{-1}A^*WQ_yWA(A^*WA)^{-1}A^* \end{aligned}$$

In order to derive the mean of the random variable $\underline{\hat{e}}^*W\underline{\hat{e}}$ of (12), we cannot make use of (21) directly. The random variable $\underline{\hat{e}}^*W\underline{\hat{e}}$ is namely *not* a linear function of $\underline{y}$. It is possible however to write $\underline{\hat{e}}^*W\underline{\hat{e}}$ as a linear combination of products of the elements of $\underline{\hat{e}}$, in which case (21) can be applied again. If we denote the ith-element of $\underline{\hat{e}}$ by $\underline{\hat{e}}_i$, and the (i,j)th-element of the matrix W by $(W)_{ij}$, then:

$$\underline{\hat{e}}^*W\underline{\hat{e}} = \sum_{i=1}^{m}\sum_{j=1}^{m}(W)_{ij}\,\underline{\hat{e}}_i\underline{\hat{e}}_j$$

This shows that $\underline{\hat{e}}^*W\underline{\hat{e}}$ is a *linear* combination of the products $\underline{\hat{e}}_i\underline{\hat{e}}_j$. Hence:

$$E\{\underline{\hat{e}}^*W\underline{\hat{e}}\} = \sum_{i=1}^{m}\sum_{j=1}^{m}(W)_{ij}E\{\underline{\hat{e}}_i\underline{\hat{e}}_j\},$$

where $E\{\underline{\hat{e}}_i\underline{\hat{e}}_j\}$ is the covariance between $\underline{\hat{e}}_i$ and $\underline{\hat{e}}_j$. But the covariance between $\underline{\hat{e}}_i$ and $\underline{\hat{e}}_j$ is the (i,j)th-element of the covariance matrix $Q_{\hat{e}}$. Thus:

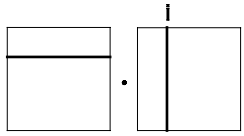
$$E\{\underline{\hat{e}}^*W\underline{\hat{e}}\} = \sum_{i=1}^{m}\sum_{j=1}^{m}(W)_{ij}\,(Q_{\hat{e}})_{ij}$$

But since the covariance matrix $Q_{\hat{e}}$ is symmetric and thus $(Q_{\hat{e}})_{ij}=(Q_{\hat{e}})_{ji}$, we may write:

$$E\{\underline{\hat{e}}^*W\underline{\hat{e}}\} = \sum_{i=1}^{m}\sum_{j=1}^{m}(W)_{ij}(Q_{\hat{e}})_{ji}$$

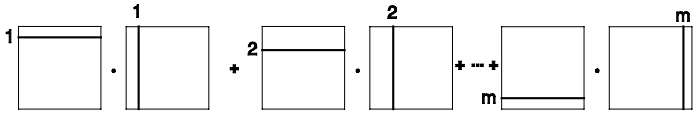
For fixed i, the term $\sum_{j=1}^{m}(W)_{ij}(Q_{\hat{e}})_{ji}$ equals the product of the ith-row of W with the ith-column of $Q_{\hat{e}}$:

$$\sum_{j=i}^{m}(W)_{ij}(Q_{\hat{e}})_{ji} \quad = \quad$$



Thus $E\{\underline{\hat{e}}^* W \underline{\hat{e}}\}$ is the sum of these m products:

$E\{\underline{\hat{e}}^* W \underline{\hat{e}}\}$ =



This shows that $E\{\underline{\hat{e}}^* W \underline{\hat{e}}\}$ equals the *trace* (spoor) of the matrix product $WQ_{\hat{e}}$:

(30) $$E\{\underline{\hat{e}}^* W \underline{\hat{e}}\} = \text{trace } (WQ_{\hat{e}})$$

For a definition of the trace of a square matrix consult your lecture notes on Linear Algebra.

It should be noted that all the results concerning the means, variances and covariances of the least-squares estimators, were derived without making any assumptions on the probability distribution of $\underline{y}$! This is in principle also possible for the variance of $\underline{\hat{e}}^* W \underline{\hat{e}}$. The derivation is however rather complicated and therefore ommitted. Even for the case that $\underline{y}$ is normally distributed, the derivation of the variance of $\underline{\hat{e}}^* W \underline{\hat{e}}$ is complex. We therefore state without proof that if $\underline{y}$ is *normally* distributed with mean and covariance matrix as given in (26), the variance of $\underline{\hat{e}}^* W \underline{\hat{e}}$ reads:

(31) $$\sigma^2_{\underline{\hat{e}}^* W \underline{\hat{e}}} = 2 \; trace \; [[W - WA(A^* WA)^{-1} A^* W] Q_y [W - WA(A^* WA)^{-1} A^* W] Q_y]$$

2.4 Best Linear Unbiased Estimation

We consider again the model:

(32) $$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{A} \; \underset{n\times 1}{x} \quad ; \quad \underset{m\times m}{E\{(\underline{y} - Ax)(\underline{y} - Ax)^*\}} = \underset{m\times m}{Q_y}$$

Let us assume that we are interested in estimating a parameter θ, which is a linear function of x:

(33) $$\underset{1\times 1}{\theta} = \underset{1\times n}{f^*} \; \underset{n\times 1}{x}$$

For instance, in case of the levelling network of figure 2.1 we could be interested in estimating the height difference between the points 1 and 5. Then θ of (33) plays the role of the height difference between the points 1 and 5, and the vector f of (33) takes the form:

$$f = (-1 \; 0 \; 0 \; 0 \; 1)^* .$$

Let $\underline{\hat{\theta}}$ be the estimator of θ. Then according to the criteria of best linear unbiased estimation, we want the estimator $\underline{\hat{\theta}}$ to be a linear function of $\underline{y}$:

(34) $\underset{1\times1}{\hat{\underline{\theta}}} = \underset{1\times m}{l^*}\ \underset{m\times1}{\underline{y}}$ "L-property"

such that

(35) $E\{\hat{\underline{\theta}}\} = \theta,$ "U-property"

and:

(36) $\sigma^2_{\hat{\theta}}$ *is minimal* "B-property"

Our objective is now to find a vector l such that with (34), the conditions (35) and (36) are fulfilled. Substitution of (34) into (35) gives:

$$E\{\hat{\underline{\theta}}\} = l^* E\{\underline{y}\} = \theta,$$

and with (32) and (33) this gives:

$$l^* \boldsymbol{A} x = f^* x$$

Since we want this condition to hold for all x, it follows that:

(37) $$\boxed{\underset{n\times1}{f} = \underset{n\times m}{\boldsymbol{A}^*}\ \underset{m\times1}{l}}$$

Compare this with equation (29) in section 1.6. Equation (37) is a system of n linear equations with m unknowns. The unknowns are the elements of the vector l. Every vector l that satisfies (37) gives with (34) a linear unbiased estimator of θ. If C is an $m\times n$ matrix such that $(A^*C)^{-1}$ exists, then:

(38) $$l = C(\boldsymbol{A}^* C)^{-1} f$$

is a solution of (37). This means that:

(39) $$\hat{\underline{\theta}} = f^*(C^* \boldsymbol{A})^{-1} C^* \underline{y}$$

is an unbiased estimator of θ. That this is indeed the case, follows if we take the expectation of:

(39)

$$E\{\hat{\underline{\theta}}\} = f^*(C^* \boldsymbol{A})^{-1} C^* E\{\underline{y}\} = f^*(C^* \boldsymbol{A})^{-1} C^* \boldsymbol{A} x = f^* x = \theta.$$

The $m\times n$ matrix C of (38) is not unique. One choice for matrix C would be:

$$C = W\boldsymbol{A}$$

Substitution into (39) gives:

$$\underline{\hat{\theta}} = f^*(\boldsymbol{A}^*\boldsymbol{W}\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{W}\underline{y}$$

This shows once again that the weighted least-squares estimator is an unbiased estimator.
From your lecture notes in Linear Algebra (Strang, 1988) you know that the general solution of the *inhomogeneous* system (37) is given by the sum of a *particular solution* of (37) and the solution of the *homogeneous* system:

(40)
$$\underset{n\times 1}{\mathbf{0}} = \underset{n\times m}{\boldsymbol{A}^*}\ \underset{m\times 1}{l}$$

The solution of this homogeneous system is the *nullspace*, $N(A^*)$, of the matrix A^*:

$$N(\boldsymbol{A}^*) = \{l \in \mathbb{R}^m \,|\, \boldsymbol{A}^* l = \mathbf{0}\}$$

The dimension of this space follows from the dimension theorem (Strang, 1988) as:

$$\mathbf{dim}.N(\boldsymbol{A}^*) = m - \mathbf{dim}.\boldsymbol{R}(\boldsymbol{A}^*),$$

where $R(A^*)$ is the *range space* or column space (beeldruimte of kolomruimte) of A^*. Since dim. $R(A^*)$= rank A^*= rank A and since we assumed that rank $A=n$, it follows that:

$$\mathbf{dim}.N(\boldsymbol{A}^*) = m - n.$$

Since the dimension of the solution space of (40) equals m-n, there exist m-n linearly independent vectors b_i, i=1,..., m-n, that form a basis of $N(A^*)$. Each of these vectors satisfy:

(41)
$$\mathbf{0} = \boldsymbol{A}^*\boldsymbol{b}_i\ ,\ i = 1,...,m-n.$$

If we collect the m-n vectors b_i in a matrix B:

$$\underset{m\times(m-n)}{\boldsymbol{B}} = (\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_{m-n}),$$

then matrix B is of order $m\times(m$-$n)$ with full rank, rank $B=m$-n, and satisfies because of (41):

(42)
$$\boxed{\underset{n\times(m-n)}{\mathbf{0}} = \underset{n\times m}{\boldsymbol{A}^*}\ \underset{m\times(m-n)}{\boldsymbol{B}}}$$

This shows that every solution of (40) can be represented as:

$$\underset{m\times 1}{l} = \underset{m\times(m-n)(m-n)\times 1}{\boldsymbol{B}\boldsymbol{\alpha}}$$

Hence, if l_o is a particular solution of (37), the general solution of (37) takes the form:

(43)
$$l = l_o + \boldsymbol{B}\boldsymbol{\alpha}$$

Compare this result with equation (37a) in section 1.6. Equation (43) shows that all vectors l that satisfy (37) lie on a *linear variety or linear manifold* (lineaire variëteit) parallel to the range space of matrix B.

Let us now continue our derivation of the best linear unbiased estimator of (33). With (37) or (43) we have found the class of linear unbiased estimators. We now want to find within this class, the estimator that has minimum variance. Application of the propagation law of variances to (34) gives with (32):

$$\sigma^2_{\hat{\theta}} = l^* Q_y l .$$

Hence, our minimization problem reads:

(44)
$$\boxed{\min_{l}.\ l^* Q_y l \text{ subject to } f = A^* l}$$

Compare this with equation (31) in section 1.6. Since (43) is the solution of (37), (44) may also be written as:

$$\min_{l}.\ l^* Q_y l \quad \text{subject to} \quad l = l_o + B\alpha$$

or as:

(45)
$$\min_{\alpha}.\ (l_o + B\alpha)^* Q_y (l_o + B\alpha) .$$

The solution of this minimization problem is:

(46)
$$\hat{\alpha} = -(B^* Q_y B)^{-1} B^* Q_y l_o .$$

This is easily verified by using an algebraic proof like the one given in section 1.2. Substitution of (46) into (43) gives:

(47)
$$\hat{l} = [I - B(B^* Q_y B)^{-1} B^* Q_y] l_o .$$

The best linear unbiased estimator $\underline{\hat{\theta}}$ of θ follows therefore with (34) as:

(48)
$$\underline{\hat{\theta}} = l_o^* [I - Q_y B (B^* Q_y B)^{-1} B^*] \underline{y} .$$

This solution is expressed in terms of the matrix B, which is related to matrix A of (32) through (42). Since our starting point was the model of (32), it would be nice if we could express the solution (48) in terms of the matrix A. In chapter 1 (see section 1.6) this was possible by making use of the identity:

$$I - Q_y b (b^* Q_y b)^{-1} b^* = a (a^* Q_y^{-1} a)^{-1} a^* Q_y^{-1}$$

Compare with equation (8) and (9) in section 1.3. For the multi-dimensional case, this *suggests* the identity:

(49) $$\boxed{I - Q_y B(B^* Q_y B)^{-1} B^* = A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1}}$$

We will prove this identity, by proving that:

(50) $$\underset{m\times m}{I} = \underset{m\times m}{A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1}} + \underset{m\times m}{Q_y B(B^* Q_y B)^{-1} B^*}$$

Since $Ix=x$, $\forall x$, we have proved (50) once we can show that:

(51) $$[A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} + Q_y B(B^* Q_y B)^{-1} B^*]\, x = x \quad \forall x \in \mathbb{R}^m$$

If R is a square and *invertible* matrix, then (51) is equivalent to:

(52) $$[A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} + Q_y B(B^* Q_y B)^{-1} B^*]\; Rz = Rz \;\; \forall z \in \mathbb{R}^m$$

If we take for R the $m\times m$ partitioned matrix:

(53) $$\underset{m\times m}{R} = (\underset{m\times n}{A} \vdots \underset{m\times(m-n)}{Q_y B}),$$

then:

(54)

$$\begin{aligned}[A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} + Q_y B(B^* Q_y B)^{-1} B^*]\, R &= [A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} + Q_y B(B^* Q_y B)^{-1} B^*](A \vdots Q_y B) \\ &= A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1}(A \vdots Q_y B) + Q_y B(B^* Q_y B)^{-1} B^*(A \vdots Q_y B) \\ &= (A \vdots 0) + (0 \vdots Q_y B) = (A \vdots Q_y B) = R,\end{aligned}$$

where we have made use of the fact that $A^*B=0$, see (42). Equation (54) shows that with the choice (53) for R, the equation (52) holds. What remains to be shown is that matrix R of (53) is invertible. This square matrix is invertible if it has full rank m, and it has full rank only if the solution of the system:

(55) $$\underset{m\times n}{(A} \vdots \underset{m\times(m-n)}{Q_y B}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix}_{m\times 1} = \underset{m\times 1}{0},$$

is the *trivial* solution $(\alpha^* \; \beta^*)^* = 0$.

Since $B^*A=0$, premultiplication of (55) with B^* gives:

(56) $$\underset{(m-n)\times(m-n)}{B^* Q_y B} \quad \underset{(m-n)\times 1}{\beta} = \underset{(m-n)\times 1}{0}.$$

But matrix B has full rank m-n and therefore $\beta=0$ is the only solution of (56). With $\beta=0$ equation (55) reduces to:

(57)
$$\underset{m\times n}{A}\ \underset{n\times 1}{\alpha} = \underset{m\times 1}{0}$$

But matrix A has full rank n and therefore $\alpha=0$ is the only solution of (57). Hence $(\alpha^* \beta^*)^*=0$ is the only solution of (55), which shows that $(A \vdots Q_y B)$ is invertible. This concludes the proof of the identity (49).

Substitution of (49) into (48) gives:

$$\underline{\hat{\theta}} = l_o^* A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} \underline{y}$$

or since l_o is a particular solution of (37):

(58)
$$\underline{\hat{\theta}} = f^*(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} \underline{y}.$$

This important result tells us two things. First of all it tells us that the best linear unbiased estimator of x is given by:

(59)
$$\boxed{\underline{\hat{x}} = (A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} \underline{y}}$$

This follows from (58) if we take for the vector f respectively $(1,0,..,0)^*$, $(0,1,0,...,0)^*$...., $(0,0,...,1)^*$. Compare (59) with equation (37) in section 1.6. Equation (59) shows that the BLUE of x is identical to the weighted least-squares estimator of x if $W=Q_y^{-1}$. The result (58) tells us also that the BLUE of $\theta=f^*x$ is:

(60)
$$\boxed{\underline{\hat{\theta}} = f^* \underline{\hat{x}}}$$

where $\underline{\hat{x}}$ is the BLUE of x.

Now that we know that the best linear unbiased estimators are identical to the least-squares estimators if $W=Q_y^{-1}$, we can use the results of the previous section 2.3 to obtain the covariance matrices of the BLUE's. Substitution of $W=Q_y^{-1}$ into equation (28) gives:

(61)
$$\boxed{\begin{aligned} Q_{\hat{x}} &= (A^* Q_y^{-1} A)^{-1} \\ Q_{\hat{y}} &= A Q_{\hat{x}} A^* \\ Q_{\hat{e}} &= Q_y - Q_{\hat{y}} \end{aligned}}$$

Substitution of $W=Q_y^{-1}$ into equation (29) gives:

(62)
$$\boxed{\begin{aligned} Q_{\hat{x}\hat{y}} &= Q_{\hat{x}} A^* \\ Q_{\hat{x}\hat{e}} &= 0 \\ Q_{\hat{y}\hat{e}} &= 0 \end{aligned}}$$

This shows that in case of best linear unbiased estimation, the estimator $\underline{\hat{e}}$ is *not* correlated with either $\underline{\hat{x}}$ or $\underline{\hat{y}}$.

Substitution of $W=Q_y^{-1}$ into (30) and (31) gives:

$$E\{\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}}\} = \text{trace } (Q_y^{-1} Q_{\hat{e}}) = m - \text{trace } (Q_y^{-1} Q_{\hat{y}})$$

$$\sigma^2_{\hat{e}^* Q_y^{-1} \hat{e}} = 2\,\text{trace } (Q_y^{-1} Q_{\hat{e}}) = 2m - 2\,\text{trace } (Q_y^{-1} Q_{\hat{y}})$$

Since trace $(Q_y^{-1}Q_{\hat{y}})$ = trace $(Q_{\hat{y}}Q_y^{-1})=n$ if rank $A=n$ (for a proof see end of next section) we get:

(63)

$$\boxed{\begin{aligned} E\{\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}}\} &= m - n \\ \sigma^2_{\hat{e}^* Q_y^{-1} \hat{e}} &= 2(m-n) \end{aligned}}$$

To conclude this section let us investigate what happens with our results if we assume that the number of observations, m, equals the number of unknowns, n. If $m=n$, then matrix A is a square matrix. We assumed that rank $A=n$. This implies that if $m=n$, the matrix A is invertible. That is, A^{-1} exists. If A^{-1} exists, the matrix $(A^* Q_y^{-1} A)^1$ can be written as:

(64)
$$(A^* Q_y^{-1} A)^{-1} = A^{-1} Q_y A^{*^{-1}}$$

Substitution into (59) gives then:

$$\underline{\hat{x}} = A^{-1} \underline{y},$$

and from this follows that:

$$\underline{\hat{y}} = A\underline{\hat{x}} = \underline{y} \quad \text{and} \quad \underline{\hat{e}} = \underline{y} - \underline{\hat{y}} = 0.$$

Similarly we find from substituting (64) into (61) that:

$$\begin{aligned} Q_{\hat{x}} &= A^{-1} Q_y A^{*^{-1}} \\ Q_{\hat{y}} &= Q_y \\ Q_{\hat{e}} &= 0 \end{aligned} \quad .$$

These results show that if $m=n$ (there is *no redundancy*), the estimate $\hat{x}$ is simply the unique solution of the system $y=Ax$ and the best linear unbiased estimator of $E\{\underline{y}\}=Ax$, $\underline{\hat{y}}$, is simply $\underline{y}$. Thus if $m=n$ we have just enough observations available to determine the unknown parameter vector x uniquely, but not enough observations to get a better estimator for $E\{\underline{y}\}=Ax$ than $\underline{y}$.

2.5 The orthogonal projector

As a generalization of the theory developed in chapter 1 we will prove in this section that the matrix $A(A^*Q_y^{-1}A)^{-1}A^*Q_y^{-1}$ is an *orthogonal projector* that projects onto the rangespace of A and along its orthogonal complement.

Let the m-dimensional linear vectorspace $\mathbb{R}^m$ have the *inner product:*

(65) $$(u,v)_{Q_y^{-1}} = u^* Q_y^{-1} v \quad \forall u,v \in \mathbb{R}^m .$$

Matrix Q_y^{-1} is thus the *matrix representation* of the inner product (65).

The *range space* of the $m \times n$ matrix A is defined as:

(66) $$R(A) = \{ z \in \mathbb{R}^m \,|\, z = A\alpha \text{ for some } \alpha \in \mathbb{R}^n \}.$$

The range space $R(A)$ is a linear subspace of $\mathbb{R}^m$, $R(A) \subset \mathbb{R}^m$, and its dimension equals the rank of A, which in our case is n:

(67) $$\dim . R(A) = \operatorname{rank} A = n .$$

The *orthogonal complement* of $R(A)$ is defined as the set of vectors that are orthogonal to $R(A)$ (Strang, 1988):

(68) $$R(A)^\perp = \{ z \in \mathbb{R}^m \,|\, A^* Q_y^{-1} z = 0 \}.$$

According to your lecture notes in Linear Algebra, dim. $R(A)$+dim. $R(A)^\perp$= dim $\mathbb{R}^m$. With dim. $R(A)=n$ and dim. $\mathbb{R}^m=m$ this gives:

(69) $$\dim . R(A)^\perp = m - n .$$

In section 2.4 a full rank matrix B of order $m \times (m-n)$ was introduced that satisfied:

(70) $$\underset{n \times m}{A^*} \quad \underset{m \times (m-n)}{B} \quad = \quad \underset{n \times (m-n)}{0} .$$

With this matrix B we can represent the orthogonal complement $R(A)^\perp$ of (68) as:

(71) $$R(A)^\perp = \{ z \in \mathbb{R}^m \,|\, z = Q_y B \beta \text{ for some } \beta \in \mathbb{R}^{m-n} \}.$$

According to theory in Linear Algebra, the space $\mathbb{R}^m$ is the *direct sum* of $R(A)$ and $R(A)^\perp$:

(72) $$\mathbb{R}^m = R(A) \oplus R(A)^\perp$$

This means that $R(A) \cap R(A)^\perp = \{0\}$, i.e. the two subspaces $R(A)$ and $R(A)^\perp$ only have the null vector in common, and that a basis of $R(A)$ together with a basis of $R(A)^\perp$ forms a basis of $\mathbb{R}^m$. Since the $m \times n$ matrix A has full rank n, its column vectors form a basis of $R(A)$. Similarly, since the $m \times (m-n)$ matrix B has full rank $m-n$ and the $m \times m$ matrix Q_y has full rank m, the $m \times (m-n)$ matrix $Q_y B$ has full rank $m-n$ and its column vectors form a basis of $R(A)^\perp$. From this follows that the column vectors of the $m \times m$ matrix:

(73) $$\underset{m\times n}{\boldsymbol{A}} \;\vdots\; \underset{m\times(m-n)}{\boldsymbol{Q}_y\boldsymbol{B}}$$

form a basis of $\mathbb{R}^m$. This implies that the $m\times m$ matrix of (73) is square and invertible (this was already shown in the previous section). If we write equation (50) of the previous section as:

$$(\boldsymbol{A} \vdots \boldsymbol{Q}_y\boldsymbol{B})\begin{pmatrix}(\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\\ (\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{B}^*\end{pmatrix} = I,$$

it follows that the inverse of (73) is given as:

(74) $$(\boldsymbol{A} \vdots \boldsymbol{Q}_y\boldsymbol{B})^{-1} = \begin{pmatrix}(\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\\ (\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{B}^*\end{pmatrix}.$$

According to theory in Linear Algebra, a vector $z \in \mathbb{R}^m$ can be decomposed in only one way along $R(A)$ and $R(A)^\perp$. If:

(75) $$z = u + v \text{ with } u \in \boldsymbol{R}(\boldsymbol{A}) \text{ and } v \in \boldsymbol{R}(\boldsymbol{A})^\perp,$$

then u is called the *orthogonal projection* of z on $R(A)$. The matrix P_A, defined as:

(76) $$\boldsymbol{P}_{\boldsymbol{A}} z = u,$$

is called the *orthogonal projector* on $R(A)$. We now want to find the matrix representation of P_A. Since the column vectors of the matrix of (73) form a basis of $\mathbb{R}^m$, every $z \in \mathbb{R}^m$ can be represented as:

$$z = (\boldsymbol{A} \vdots \boldsymbol{Q}_y\boldsymbol{B})\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \boldsymbol{A}\alpha + \boldsymbol{Q}_y\boldsymbol{B}\beta.$$

In this representation $u=A\alpha$ is the component of z along $R(A)$. With (76) this gives:

$$\boldsymbol{P}_{\boldsymbol{A}}(\boldsymbol{A} \vdots \boldsymbol{Q}_y\boldsymbol{B})\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \boldsymbol{A}\alpha$$

or:

(77) $$\boldsymbol{P}_{\boldsymbol{A}}(\boldsymbol{A} \vdots \boldsymbol{Q}_y\boldsymbol{B})\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = (\boldsymbol{A} \vdots 0)\begin{pmatrix}\alpha\\ \beta\end{pmatrix}.$$

Since (76) holds for all $z \in \mathbb{R}^m$, (77) holds for all α and β:

$$\boldsymbol{P}_{\boldsymbol{A}}(\boldsymbol{A} \vdots \boldsymbol{Q}_y\boldsymbol{B}) = (\boldsymbol{A} \vdots 0)\,.$$

Post-multiplication with $(A \vdots Q_yB)^{-1}$ gives:

$$\boldsymbol{P}_{\boldsymbol{A}} = (\boldsymbol{A} \vdots 0)\,(\boldsymbol{A} \vdots \boldsymbol{Q}_y\boldsymbol{B})^{-1}\,.$$

And substitution of (74) gives then finally:

(78) $$\boxed{\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}}$$

According to theory in Linear Algebra, P_A has the following properties:

(79) $$\begin{cases} \boldsymbol{R}(\boldsymbol{P}_{\boldsymbol{A}}) = \boldsymbol{R}(\boldsymbol{A}) \\ N(\boldsymbol{P}_{\boldsymbol{A}}) = \boldsymbol{R}(\boldsymbol{A})^{\perp} = \boldsymbol{R}(\boldsymbol{Q}_y\boldsymbol{B}) \\ \boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{A}} \ \text{(idempotence)}\,. \end{cases}$$

From Linear Algebra we know that $I\text{-}P_A$ is also an orthogonal projector. $I\text{-}P_A$ projects onto $R(A)^{\perp}$ and along $R(A)$. $I\text{-}P_A$ has the properties:

(80) $$\begin{cases} \boldsymbol{R}(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{A}}) = \boldsymbol{R}(\boldsymbol{A})^{\perp} \\ N(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{A}}) = \boldsymbol{R}(\boldsymbol{A}) \\ (\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{A}})(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{A}}) = \boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{A}} \ \text{(idempotence)}\,. \end{cases}$$

Instead of $I\text{-}P_A$ we will sometimes use the notation $P_A^{\perp}$. It follows from equation (49) that:

(81) $$\boxed{\boldsymbol{P}_{\boldsymbol{A}}^{\perp} = \boldsymbol{Q}_y\boldsymbol{B}(\boldsymbol{B}^{*}\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{B}^{*}}$$

To conclude this section we will prove that:

(82) $$\text{trace}\ (\boldsymbol{Q}_{\hat{y}}\boldsymbol{Q}_y^{-1}) = n$$

Substitution of $Q_{\hat{y}}=A(A^{*}Q_y^{-1}A)^{-1}A^{*}$ (see (61)) gives:

$$\text{trace}\left[(\boldsymbol{A}(\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1})\right] = n$$

or with (78):

(83) $$\text{trace}\ (\boldsymbol{P}_{\boldsymbol{A}}) = n$$

Since *the trace of a matrix is the sum of its eigenvalues* (Strang, 1988 and Lay, 1997), we need the eigenvalues of P_A in order to prove (83). The *eigenvalue problem* for P_A reads:

(84) $$\boldsymbol{P_A} z = \lambda z$$

with $P_A P_A z = \lambda P_A z = \lambda^2 z$ and $P_A P_A z = P_A z = \lambda z$ follows that $\lambda^2 z = \lambda z$ or for $z \neq 0$:

(85) $$\lambda (\lambda - 1) = 0.$$

This shows that the eigenvalues of P_A are 1 or 0. In order to prove (83) we thus need to show that P_A has n-number of eigenvalues equal to 1. This corresponds to showing that for $\lambda=1$ the solution space of (84) has dimension n. For $\lambda=1$ the solution space of (84) is the solution space of $(I-P_A)\ z=0$ and thus the nullspace of $I-P_A$. According to (80) the nullspace of $I-P_A$ is the rangespace of A. Hence, dim. $N(I-P_A)$=dim. $R(A)$= rank $A=n$. This concludes the proof.

2.6 Summary

In the table on the next page an overview is given of the results of Best Linear Unbiased Estimation.

Best Linear Unbiased Estimation

Model

$$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x}\ ;\ \underset{m\times m}{E\{(\underline{y}-\boldsymbol{A}x)(\underline{y}-\boldsymbol{A}x)^*\}} = \underset{m\times m}{Q_y}$$

Estimators

$$\hat{\underline{x}} = (\boldsymbol{A}^* Q_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^* Q_y^{-1}\underline{y}\ ;\quad \hat{\underline{\theta}} = f^*\hat{\underline{x}}$$

$$\hat{\underline{y}} = \boldsymbol{A}\hat{\underline{x}} = \boldsymbol{P}_{\boldsymbol{A}}\underline{y}\quad ;\quad \hat{\underline{e}} = \underline{y}-\hat{\underline{y}} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\underline{y}$$

$$\hat{\underline{e}}^* Q_y^{-1}\hat{\underline{e}} = \underline{y}^* Q_y^{-1}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\underline{y}$$

Mean

$$E\{\hat{\underline{x}}\} = x\quad ;\quad E\{\hat{\underline{\theta}}\} = f^* x = \theta$$

$$E\{\underline{\hat{y}}\} = \boldsymbol{A}x = E\{\underline{y}\}\ ;\ E\{\underline{\hat{e}}\} = 0$$

$$E\{\underline{\hat{e}}^* Q_y^{-1}\underline{\hat{e}}\} = m-n$$

Variances and covariances

$$Q_{\hat{x}} = (\boldsymbol{A}^* Q_y^{-1}\boldsymbol{A})^{-1}\quad ;\quad \sigma_{\hat{\theta}}^2 = f^* Q_{\hat{x}} f$$

$$Q_{\hat{y}} = \boldsymbol{A}Q_{\hat{x}}\boldsymbol{A}^* = \boldsymbol{P}_{\boldsymbol{A}} Q_y \boldsymbol{P}_{\boldsymbol{A}}^*\ ;\ Q_{\hat{e}} = Q_y - Q_{\hat{y}} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp} Q_y \boldsymbol{P}_{\boldsymbol{A}}^{\perp *}$$

$$= \boldsymbol{P}_{\boldsymbol{A}}^{\perp} Q_y = Q_y \boldsymbol{P}_{\boldsymbol{A}}^{\perp *}$$

$$\sigma^2_{\hat{e}^* Q_y^{-1}\hat{e}} = 2(m-n)$$

$$Q_{\hat{x}\hat{y}} = Q_{\hat{x}}\boldsymbol{A}^*\ ;\ Q_{\hat{x}\hat{e}} = 0\ ;\ Q_{\hat{y}\hat{e}} = 0$$

Orthogonal projectors

$$\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^* Q_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^* Q_y^{-1}\quad ;\quad \boldsymbol{R}(\boldsymbol{P}_{\boldsymbol{A}}) = \boldsymbol{R}(\boldsymbol{A})\quad ,\quad N(\boldsymbol{P}_{\boldsymbol{A}}) = \boldsymbol{R}(\boldsymbol{A})^{\perp}$$

$$\boldsymbol{P}_{\boldsymbol{A}}^{\perp} = I - \boldsymbol{A}(\boldsymbol{A}^* Q_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^* Q_y^{-1}\ ;\ \boldsymbol{R}(\boldsymbol{P}_{\boldsymbol{A}}^{\perp}) = \boldsymbol{R}(\boldsymbol{A})^{\perp}\quad ,\ N(\boldsymbol{P}_{\boldsymbol{A}}^{\perp}) = \boldsymbol{R}(\boldsymbol{A})$$

3 The model with condition equations

3.1 Introduction

Figure 3.1 shows the levelling network of section 2.1.

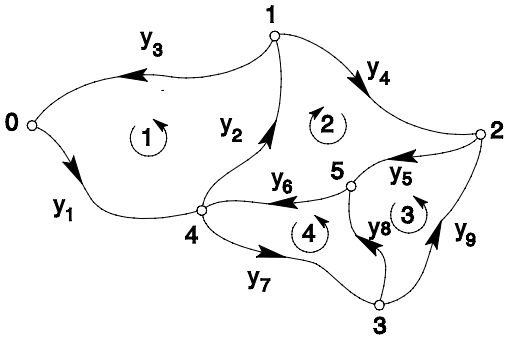


Figure 3.1: Levelling network

In section 2.1 it was shown that the *observation equations* of the levelling network of figure 3.1 take the form:

(1)
$$\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \\ \underline{y}_6 \\ \underline{y}_7 \\ \underline{y}_8 \\ \underline{y}_9 \end{pmatrix} = \begin{pmatrix} & & & & 1 & \\ 1 & & & -1 & & \\ -1 & & & & & \\ -1 & 1 & & & & \\ & -1 & & & & 1 \\ & & & & 1 & -1 \\ & & & 1 & -1 & \\ & & -1 & & & 1 \\ & 1 & -1 & & & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} + \begin{pmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \underline{e}_3 \\ \underline{e}_4 \\ \underline{e}_5 \\ \underline{e}_6 \\ \underline{e}_7 \\ \underline{e}_8 \\ \underline{e}_9 \end{pmatrix}$$

For simplicity it is assumed here that the known height of the reference point 0 is equal to zero. The matrix equation of (1) is of the form:

(2)
$$\underset{m\times 1}{\underline{y}} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x} + \underset{m\times 1}{\underline{e}}$$

with m=9 and n=5. Thus since rank A=n=5, the number of *redundant* observations is m-n=4.

Instead of writing the observed height differences as functions of the unknown heights, as is done in (1), it is also possible to write down *condition equations* (voorwaarde vergelijkingen) which the height differences have to fulfill. In figure 3.1 four loops (kringen) are shown. With perfect measurements the height differences of each loop would sum up to zero. Due however to the presence of measurement errors these sums will differ from zero. That is, misclosures will exist. If we denote these four *misclosures* (tegenspraken) by respectively $\underline{t}_1$, $\underline{t}_2$, $\underline{t}_3$ and $\underline{t}_4$, the four condition equations become (see figure 3.1):

$$(3) \qquad \begin{cases} \underline{y}_1 + \underline{y}_2 + \underline{y}_3 = \underline{t}_1 \\ \underline{y}_2 + \underline{y}_4 + \underline{y}_5 + \underline{y}_6 = \underline{t}_2 \\ \underline{y}_9 + \underline{y}_5 - \underline{y}_8 = \underline{t}_3 \\ \underline{y}_6 + \underline{y}_7 + \underline{y}_8 = \underline{t}_4 \end{cases}$$

These four condition equations can be written in matrix notation as:

$$(4) \qquad \begin{pmatrix} 1 & 1 & 1 & & & & & & \\ & 1 & & 1 & 1 & 1 & & & \\ & & & & 1 & & & -1 & 1 \\ & & & & & 1 & 1 & 1 & \end{pmatrix} \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \\ \underline{y}_6 \\ \underline{y}_7 \\ \underline{y}_8 \\ \underline{y}_9 \end{pmatrix} = \begin{pmatrix} \underline{t}_1 \\ \underline{t}_2 \\ \underline{t}_3 \\ \underline{t}_4 \end{pmatrix}$$

This matrix equation is of the form:

$$(5) \qquad \boxed{\underset{b\times m}{\boldsymbol{B}^{*}} \underset{m\times 1}{\underline{y}} = \underset{b\times 1}{\underline{t}}}$$

with b=4 and m=9. Thus matrix B has m=9 rows and b=4 columns.

Note that in the above example of the levelling network the number of redundant observations, m-n=4, equals the number of condition equations, b=4. This is generally true, because each additional observation on top of the n-number of observations which are needed for the unique determination of x in (2), gives rise to an extra condition equation. Thus:

$$(6) \qquad \boxed{b = m - n}$$

Also note that if the matrix of (1) is pre-multiplied with the matrix of (4), the result equals the zero-matrix. In terms of the matrices A and B this means that:

(7) $$\underset{b\times m}{\boldsymbol{B}^{*}}\ \underset{m\times n}{\boldsymbol{A}} = \underset{b\times n}{0}$$

Also this relation between the two matrices A and B is generally true. This can be seen as follows. Let us start from equation (2). If we pre-multiply equation (2) with a $b\times m$ matrix B^* we get:

(8) $$\underset{b\times m}{\boldsymbol{B}^{*}}\ \underset{m\times 1}{\underline{y}} = \underset{b\times m}{\boldsymbol{B}^{*}}\ \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x} + \underset{b\times m}{\boldsymbol{B}^{*}}\ \underset{m\times 1}{\underline{e}}$$

Our objective is now to investigate whether a full rank matrix B^*, with rank $B^*=b=m-n$, exists such that $B^*A=0$. If such a matrix exists, then the unknown parameter vector x gets eliminated from (8), and (8) reduces to the form of (5) with:

(9) $$\underline{t} = \boldsymbol{B}^{*}\underline{e}$$

If we denote the b-number of columnvectors of matrix B by b_i, $i=1,...,b$, then $B^*A=0$ is equivalent to:

(10) $$\underset{n\times m}{\boldsymbol{A}^{*}}\ \underset{m\times 1}{\boldsymbol{b}_{i}} = \underset{n\times 1}{0}\quad ,\boldsymbol{i}=1,....\boldsymbol{b}$$

The question is thus if and how many linearly independent vectors exist that satisfy (10). From Linear Algebra we know that for an $n\times m$ matrix A^*:

(11) $$\begin{cases} \dim.\boldsymbol{R}(\boldsymbol{A}^{*}) = \operatorname{rank}\boldsymbol{A}^{*} \\ \operatorname{rank}\boldsymbol{A}^{*} = \operatorname{rank}\boldsymbol{A} \\ \dim.\boldsymbol{R}(\boldsymbol{A}^{*}) + \dim.N(\boldsymbol{A}^{*}) = \dim.\mathbb{R}^{m} = m \end{cases}$$

Since rank $A=n$, it follows from (11) that:

(12) $$\dim.N(\boldsymbol{A}^{*}) = m - n\,.$$

Thus the dimension of the nullspace of A^* is $m-n$. This shows that indeed $m-n$ linearly independent vectors b_i exist that satisfy (10). Note that if there are no redundant observations $m=n$ and (10) has only the trivial zero vector as solution. Thus if $m=n$, no matrix B exists for which (7) holds and no condition equations exist.

With (2) and (5) we now have two equivalent representations of the same model. Equation (2) is a parametric representation. The equations of (2) are known as *observation equations*. Equation (5) is an implicit representation. The equations of (5) are known as *condition equations*. In

chapter 2 we developed the theory of linear estimation for (2). In this chapter we will develop the theory for the model with condition equations (5).

3.2 Best Linear Unbiased Estimation

From equations (5) and (9) follows that:

(13) $$\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{\underline{y}} = \underset{b\times 1}{\underline{t}} \quad with\ \underset{b\times 1}{\underline{t}} = \underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{\underline{e}}$$

This shows that the misclosure vector $\underline{t}$ is due to the random measurement error vector $\underline{e}$. We will assume, like we did in chapter 2, that:

(14) $$E\{\underline{e}\} = 0 \quad \text{and} \quad E\{\underline{e}\underline{e}^*\} = Q_y$$

This, together with (13) gives:

(15) $$\boxed{\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0}\ ;\ \underset{m\times m}{E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\}} = \underset{m\times m}{Q_y}}$$

This model will be our starting point for best linear unbiased estimation. Let us assume that we are interested in estimating a parameter φ, which is a linear function of $E\{\underline{y}\}$:

(16) $$\underset{1\times 1}{\varphi} = \underset{1\times m}{\boldsymbol{g}^*}\ \underset{m\times 1}{E\{\underline{y}\}}$$

For instance, in case of the levelling network of figure 3.1 we could be interested in estimating the height difference between the points 1 and 5. Then φ of (16) plays the role of the height difference between the points 1 and 5, and the vector g of (16) takes the form:

$$\boldsymbol{g} = (0\ 0\ 0\ 1\ 1\ 0\ 0\ 0)^*$$

Let $\underline{\hat{\varphi}}$ be the estimator of φ. Then according to the criteria of best linear unbiased estimation, we want the estimator $\underline{\hat{\varphi}}$ to be a linear function of $\underline{y}$:

(17) $$\underset{1\times 1}{\underline{\hat{\varphi}}} = \underset{l\times m}{l^*}\ \underset{m\times 1}{\underline{y}} \qquad \text{"L-property"}$$

such that

(18) $$E\{\underline{\hat{\varphi}}\} = \varphi\ , \qquad \text{"U-property"}$$

and

(19) $$\sigma^2_{\hat{\varphi}} = \text{minimal} \qquad \text{"B-property"}$$

Our objective is now to find a vector l such that with (17), the conditions (18) and (19) are fulfilled.

Substitution of (17) into (18) gives:

(20) $$E\{\hat{\underline{\varphi}}\} = l^{*}E\{\underline{y}\} = \boldsymbol{\varphi}\,.$$

If we subtract (16) from (20) we get:

(21) $$\underset{1\times m}{(l-\boldsymbol{g})^{*}}\;\underset{m\times 1}{E\{\underline{y}\}} = \underset{1\times 1}{\mathbf{0}}\,.$$

This equation states that the vector l-g is orthogonal to the mean $E\{\underline{y}\}$. From the first equation of (15) we learn that the subspace orthogonal to $E\{\underline{y}\}$ is spanned by the b linearly independent column vectors of matrix B. This, together with (21) shows that the vector l-g is an element of the range space of the matrix B:

$$l-\boldsymbol{g} \;\in\; \boldsymbol{R}(\boldsymbol{B})\,.$$

This means that a $b\times 1$ vector α exists such that:

$$l-\boldsymbol{g} = \boldsymbol{B}\boldsymbol{\alpha}$$

or:

(22) $$\boxed{\underset{m\times 1}{l} = \underset{m\times 1}{\boldsymbol{g}} + \underset{m\times \boldsymbol{b}\boldsymbol{b}\times 1}{\boldsymbol{B}\;\boldsymbol{\alpha}}}$$

The meaning of this result is that every vector l that can be represented as (22), gives with (17) an unbiased estimator of (16). To verify this, we substitute (22) into (17):

$$\hat{\underline{\varphi}} = \boldsymbol{g}^{*}\underline{y} + \boldsymbol{\alpha}^{*}\boldsymbol{B}^{*}\underline{y}$$

If we take the expectation we get:

$$E\{\hat{\underline{\varphi}}\} = \boldsymbol{g}^{*}E\{\underline{y}\} + \boldsymbol{\alpha}^{*}\boldsymbol{B}^{*}E\{\underline{y}\}\;,$$

and with (15):

$$E\{\hat{\underline{\varphi}}\} = \boldsymbol{g}^{*}E\{\underline{y}\}\;,$$

which according to (16) is equal to φ.

Let us now introduce condition (19), the condition of minimum variance. Application of the *propagation law of variances* to (17) gives with (15):

$$\sigma^{2}_{\hat{\varphi}} = l^{*}\boldsymbol{Q}_{y}\,l$$

Hence, our minimization problem reads:

(23) $$\boxed{\min_{l}.\ l^{*}Q_{y}l \text{ subject to } l=g+B\alpha}$$

This minimization problem may also be written as:

$$\min_{\alpha}.\ (g+B\alpha)^{*}Q_{y}(g+B\alpha).$$

The solution of this minimization problem is:

(24) $$\hat{\alpha}=-(B^{*}Q_{y}B)^{-1}B^{*}Q_{y}g.$$

This is easily verified by using an algebraic proof like the one given in section 1.2. Substitution of (24) into (22) gives:

(25) $$\hat{l}=[I-B(B^{*}Q_{y}B)^{-1}B^{*}Q_{y}]g.$$

The best linear unbiased estimator $\hat{\underline{\Phi}}$ of Φ follows therefore with (17) as:

(26) $$\boxed{\hat{\underline{\Phi}}=g^{*}[I-Q_{y}B(B^{*}Q_{y}B)^{-1}B^{*}]\underline{y}}$$

If we take for the vector g respectively $(1,0,...,0)^{*}$, $(0,1,0,...,0)^{*}$, ..., $(0,...,0,1)^{*}$, the BLUE of $E\{\underline{y}\}$ follows as:

(27) $$\boxed{\hat{\underline{y}}=[I-Q_{y}B(B^{*}Q_{y}B)^{-1}B^{*}]\underline{y}}$$

This, together with (26) shows that:

(28) $$\boxed{\hat{\underline{\Phi}}=g^{*}\hat{\underline{y}}}$$

If we pre-multiply (27) with B^{*}, we get:

(29) $$B^{*}\hat{\underline{y}}=0\ ,$$

which shows that the estimator $\hat{\underline{y}}$ satisfies the condition equations. Since $\underline{t}=B^{*}\underline{y}$ (see (13)), application of the propagation law of variances gives $Q_{t}=B^{*}Q_{y}B$. This shows that (27) can also be written as:

(30) $$\hat{\underline{y}}=\underline{y}-Q_{y}BQ_{t}^{-1}\underline{t}$$

This equation shows perhaps more clearly than (27), that the estimator $\hat{\underline{y}}$ is obtained from $\underline{y}$ by subtracting a part which depends on the misclosure vector $\underline{t}$ such that $\hat{\underline{y}}$ satisfies (29). Since:

(31) $$I - Q_y B(B^* Q_y B)^{-1} B^* = A(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} \quad ,$$

(see equation (49) in section 2.2), the estimator of (27) is *identical* to the estimator of $E\{\underline{y}\}$ which was derived in the previous chapter (see for instance the summary in section 2.6). This is of course understandable, because we used the *same* principle of estimation, namely BLUE, and the *same* model. Model (15) is namely equivalent to the model in section 2.6. The two models only differ in their representation.

The results of estimation for model (15) are therefore identical to the results of the previous chapter. The results only differ in their appearance. In chapter 2 everything was expressed in terms of the matrix A, whereas in this chapter everything is expressed in terms of the matrix B.

3.3 The case $B^* E\{\underline{y}\} = b^0, b^0 \neq 0$

The model of condition equations (4) was based on the fact that the observed height difference of each levelling loop in figure 3.1 should sum up to zero. Now consider the triangulation network of figure 3.2.

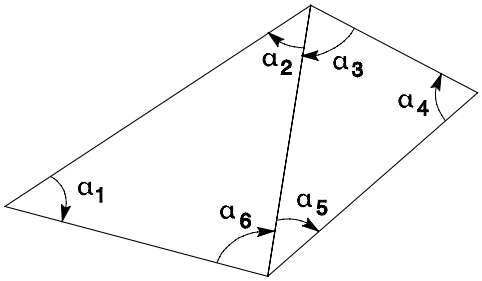


Figure 3.2: Small triangulation network

From the figure it is clear that the six angle observables should fulfill two conditions: in each triangle the expectation of the three observed angles sums up to π. Due however to the presence of measurement errors these condition equations will result in small misclosures. The two condition equations become in matrix notation:

(32) $$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \underline{\alpha}_1 \\ \underline{\alpha}_2 \\ \underline{\alpha}_3 \\ \underline{\alpha}_4 \\ \underline{\alpha}_5 \\ \underline{\alpha}_6 \end{pmatrix} = \begin{pmatrix} \pi \\ \pi \end{pmatrix} + \begin{pmatrix} \underline{t}_1 \\ \underline{t}_2 \end{pmatrix}.$$

Compared to (5) this matrix equation is of the form:

$$\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{\underline{y}} = \underset{b\times 1}{\boldsymbol{b}^0} + \underset{b\times 1}{\underline{t}} \tag{33}$$

with b=2, m=6 and $\boldsymbol{b}^0=(\pi,\pi)^*$.

For a non-zero vector b^0 straightforward application of the estimation formulae of the previous section is no longer possible. At first sight we can rewrite (32) e.g. as:

$$\begin{pmatrix}1&1&0&0&0&1\\0&0&1&1&1&0\end{pmatrix}\begin{pmatrix}\underline{\alpha}_1-\pi\\ \underline{\alpha}_2\\ \underline{\alpha}_3\\ \underline{\alpha}_4-\pi\\ \underline{\alpha}_5\\ \underline{\alpha}_6\end{pmatrix}=\begin{pmatrix}\underline{t}_1\\ \underline{t}_2\end{pmatrix},$$

so that a model of the form of (5) is obtained. However, in general it is difficult to redefine the vector of observables so that it is of the form of (5). Therefore we will adapt the estimation formulae so that the non-zero vector b^0 can be taken into account. From (33) the vector of misclosures and its variance matrix are obtained as:

$$\underline{t}=\boldsymbol{B}^*\underline{y}-\boldsymbol{b}^0;\quad Q_t=\boldsymbol{B}^*Q_y\boldsymbol{B}. \tag{34}$$

Then with (30) the estimators $\underline{\hat{y}}$ and $\underline{\hat{e}}$ become:

$$\begin{cases}\underline{\hat{y}} &= \left[I-Q_y\boldsymbol{B}(\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1}\boldsymbol{B}^*\right]\underline{y} + Q_y\boldsymbol{B}(\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1}\boldsymbol{b}^0\\ \underline{\hat{e}} &= Q_y\boldsymbol{B}(\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1}\boldsymbol{B}^*\underline{y} - Q_y\boldsymbol{B}(\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1}\boldsymbol{b}^0\end{cases} \tag{35}$$

Comparing this result with (27) shows that the non-zero vector b^0 results in an additional term for the least-squares estimators. Since b^0 is a vector of constants the extension of the model of condition equations with a non-zero b^0-vector does not affect the variance matrices Q_r, $Q_{\hat{y}}$ and $Q_{\hat{e}}$.

3.4 Summary

Since the results of estimation for model (15) are identical to the results of the previous chapter, we merely need to represent the results of section 2.6 in terms of the matrix B to obtain a summary of the results for model (15). This is done on the next page, there upon the main results of this and the previous chapter are compared.

Best Linear Unbiased Estimation

Model

$$\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0} \quad ; \quad \underset{m\times m}{E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\}} = \underset{m\times m}{Q_y}$$

Estimators

$$\hat{\underline{y}} = P_{Q_y B}^{\perp}\underline{y} \ ; \quad \hat{\underline{e}} = \underline{y} - \hat{\underline{y}} = Q_y B Q_t^{-1}\underline{t} = P_{Q_y B}\underline{y}$$

$$\hat{\underline{\varphi}} = \boldsymbol{g}^*\hat{\underline{y}}$$

$$\hat{\underline{e}}^* Q_y^{-1}\hat{\underline{e}} = \underline{t}^* Q_t^{-1}\underline{t} = \underline{y}^* Q_y^{-1} P_{Q_y B}\underline{y}$$

Mean

$$E\{\hat{\underline{y}}\} = E\{\underline{y}\} \ ; \ E\{\hat{\underline{e}}\} = 0$$
$$E\{\hat{\underline{\varphi}}\} = \boldsymbol{g}^* E\{\underline{y}\}$$
$$E\{\hat{\underline{e}}^* Q_y^{-1}\hat{\underline{e}}\} = E\{\underline{t}^* Q_t^{-1}\underline{t}\} = \boldsymbol{b}$$

Variances and covariances

$$Q_{\hat{y}} = P_{Q_y B}^{\perp} Q_y P_{Q_y B}^{\perp *} \quad ; \quad Q_{\hat{e}} = Q_y - Q_{\hat{y}} = Q_y B Q_t^{-1} B^* Q_y$$
$$= P_{Q_y B} Q_y P_{Q_y B}^*$$
$$\sigma^2_{\hat{e}^* Q_y^{-1}\hat{e}} = \sigma^2_{t^* Q_t^{-1} t} = 2b$$
$$Q_{\hat{y}\hat{e}} = 0$$

Orthogonal projectors

$$P_{Q_y B}^{\perp} = I - Q_y B(B^* Q_y B)^{-1} B^* \quad ; \quad R(P_{Q_y B}^{\perp}) = R(Q_y B)^{\perp} \quad ; \quad N(P_{Q_y B}^{\perp}) = R(Q_y B)$$
$$P_{Q_y B} = Q_y B(B^* Q_y B)^{-1} B^* \quad ; \quad R(P_{Q_y B}) = R(Q_y B) \quad ; \quad N(P_{Q_y B}) = R(Q_y B)^{\perp}$$

Best Linear Unbiased Estimation

$$\underset{n\times m}{\boldsymbol{A}}{}^{*}\,\underset{n\times b}{\boldsymbol{B}} = 0$$

$$b = m - n$$

$$N(\boldsymbol{A}^{*}) = \boldsymbol{R}(\boldsymbol{B})$$

$$E\{\underline{y}\} = \boldsymbol{A}x \;,\; \operatorname{rank}\boldsymbol{A} = n$$

$$E\{(\underline{y} - \boldsymbol{A}x)(\underline{y} - \boldsymbol{A}x)^{*}\} = \boldsymbol{Q}_y$$

observation equations

$$\boldsymbol{B}^{*}E\{\underline{y}\} = 0 \;,\; \operatorname{rank}\boldsymbol{B} = b$$

$$E\{(\underline{y} - E\{\underline{y}\})(\underline{y} - E\{\underline{y}\})^{*}\} = \boldsymbol{Q}_y$$

condition equations

$$\underline{\hat{x}} = (\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\underline{y}$$

$$\underline{\hat{\theta}} = f^{*}\underline{\hat{x}}$$

$$\underline{\hat{y}} = \boldsymbol{P}_{\boldsymbol{A}}\underline{y}$$

$$\underline{\hat{e}} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\underline{y}$$

$$\underline{\hat{e}}^{*}\boldsymbol{Q}_y^{-1}\underline{\hat{e}} = \underline{y}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\underline{y}$$

$$\underline{\hat{\varphi}} = \boldsymbol{g}^{*}\underline{\hat{y}}$$

$$\underline{\hat{y}} = \boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}}^{\perp}\underline{y}$$

$$\underline{\hat{e}} = \boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}}\underline{y}$$

$$\underline{\hat{e}}^{*}\boldsymbol{Q}_y^{-1}\underline{\hat{e}} = \underline{y}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}}\underline{y}$$

$$\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}$$

$$\boldsymbol{P}_{\boldsymbol{A}}^{\perp} = I - \boldsymbol{A}(\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}$$

$$\boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}}^{\perp} = I - \boldsymbol{Q}_y\boldsymbol{B}(\boldsymbol{B}^{*}\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{B}^{*}$$

$$\boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}} = \boldsymbol{Q}_y\boldsymbol{B}(\boldsymbol{B}^{*}\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{B}^{*}$$

$$\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}}^{\perp}$$

$$\boldsymbol{P}_{\boldsymbol{A}}^{\perp} = \boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}}$$

4 $\underline{y}^R$-Variates

4.1 Introduction

For the case that m=2 and n=1 the model with *observation equations* reads:

$$E\{\underset{2\times1}{\underline{y}}\} = \underset{2\times1}{\boldsymbol{a}}\ \underset{1\times1}{x}\ ;\ \underset{2\times2}{E\{(\underline{y}-\boldsymbol{a}x)(\underline{y}-\boldsymbol{a}x)^*\}} = \underset{2\times2}{Q_y}\ . \tag{1}$$

The corresponding model with *condition equations* reads:

$$\underset{1\times2}{\boldsymbol{b}^*}\ \underset{2\times1}{E\{\underline{y}\}} = \underset{1\times1}{0}\ ;\ \underset{2\times2}{E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\}} = \underset{2\times2}{Q_y}\ . \tag{2}$$

In chapter 1 (see figure 1.8 or figure 1.14 in section 1.3) it was shown that the estimator $\underline{\hat{y}}$ of $E\{\underline{y}\}$ is the oblique projection of $\underline{y}$ onto the line with direction vector a. The direction of projection is parallel to the tangent line of the ellipse $z^*Q_y^{-1}z = a^*Q_y^{-1}a$ at a (see figure 4.1).

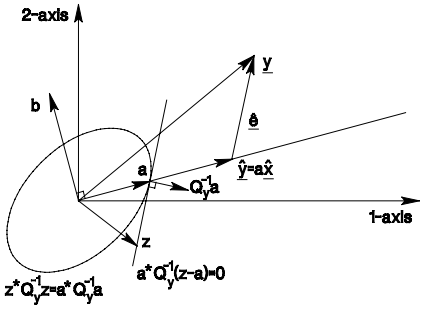


Figure 4.1: Oblique projection of $\underline{y}$

If we rotate the direction vector a so that it becomes parallel to the 1-axis, figure 4.1 transforms into figure 4.2.

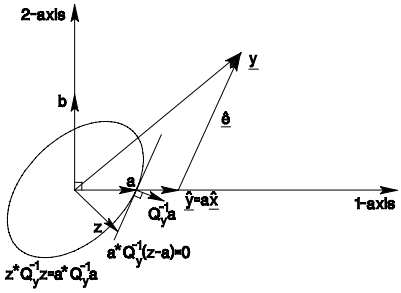


Figure 4.2: Oblique projection of $\underline{y}$ in case a= $(a_1\ 0)^*$

In this case the second component a_2 of the vector a is zero, and the first component b_1 of the vector b is zero. The corresponding two models (1) and (2) then take the form:

$$E\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\right\} = \begin{pmatrix}a_1\\ 0\end{pmatrix}x \;;\; E(\underline{y}-ax)(\underline{y}-ax)^* = \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{pmatrix} \tag{3}$$

and:

$$\begin{pmatrix}0\\ 1\end{pmatrix}^* E\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\right\} = 0 \;\;;\;\; E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\} = \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{pmatrix}. \tag{4}$$

Equation (4) shows that the coefficient of $E\{\underline{y}_1\}$ equals zero. This implies that the observable $\underline{y}_1$ does *not* appear in the condition equation. Observables that do not appear in the condition equations are called *free variates*. Thus $E\{\underline{y}_1\}$ is a free variate. If one writes (4) as:

$$E\{\underline{y}_2\} = 0 \;\;;\;\; E\{(\underline{y}_2 - E\{\underline{y}_2\})(\underline{y}_2 - E\{\underline{y}_2\})^*\} = \sigma_2^2\,, \tag{5}$$

one might be inclined to conclude, since $E\{\underline{y}_1\}$ does not appear in (5), that the estimator $\underline{\hat{y}}_1$ of $E\{\underline{y}_1\}$ equals $\underline{y}_1$ and thus that $\underline{\hat{e}}_1=\underline{y}_1-\underline{\hat{y}}_1$ is zero. This however is *not true*. Figure 4.2 shows namely clearly that both $\underline{\hat{e}}_1$ and $\underline{\hat{e}}_2$ are unequal to zero. The reason for this is that the major and minor axes of the ellipse of figure 4.2 make a non-zero angle with the 1-axis and 2-axis. If the major and minor axes of the ellipse are parallel to the 1-axis and the 2-axis, then figure 4.2 transforms into figure 4.3.

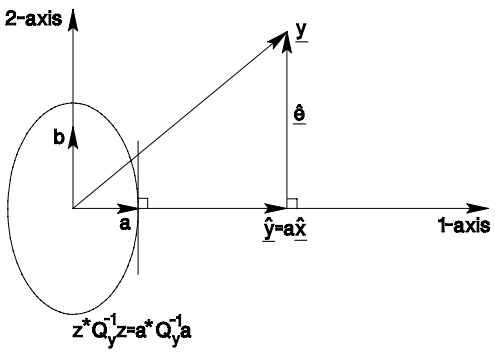


Figure 4.3: The covariance matrix Q_y is diagonal

In this case $\sigma_{12}=0$, Q_y is diagonal and $\underline{\hat{e}}_1=0$. It thus seems that the covariance, σ_{12}, between $\underline{y}_1$ and $\underline{y}_2$ determines to what extent $\underline{\hat{e}}_1=\underline{y}_1 - \underline{\hat{y}}_1$ differs from zero. If $\sigma_{12}=0$ then $\underline{\hat{e}}_1=0$, and if

$\sigma_{12}\neq 0$ then also $\underline{\hat{e}}_1\neq 0$. In order to find out how $\underline{\hat{e}}_1$ depends on σ_{12} consult figure 4.4. Figure 4.4 shows that:

(6) $$\frac{\hat{e}_1}{\hat{e}_2} = \tan\beta \ ,$$

and that β is the angle between the vector $Q_y^{-1}a$ and the 1-axis.

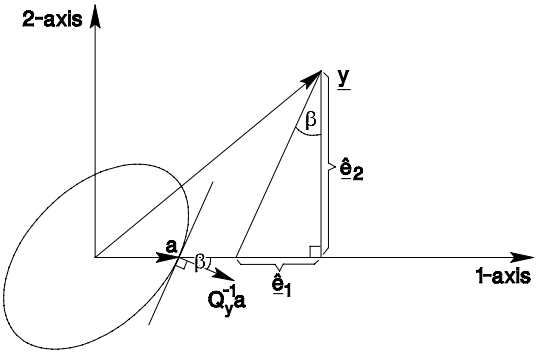


Figure 4.4: $\tan\beta = \frac{\hat{e}_1}{\hat{e}_2}$

Since the inverse of Q_y is:

$$Q_y^{-1} = \left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right)^{-1} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}$$

it follows with $a=(a_1\ 0)^*$ that:

$$Q_y^{-1}a = \left(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\right)^{-1} \begin{pmatrix} \sigma_2^2 a_1 \\ -\sigma_{12}a_1 \end{pmatrix} .$$

From this it follows that:

(7) $$\tan\beta = \frac{\sigma_{12}}{\sigma_2^2} \ ,$$

which with (6) gives that:

(8) $$\boxed{\underline{\hat{e}}_1 = \sigma_{12}\sigma_2^{-2}\underline{\hat{e}}_2.}$$

This result shows how $\underline{\hat{e}}_1=\underline{y}_1 - \underline{\hat{y}}_1$, and thus the estimator $\underline{\hat{y}}_1=\underline{y}_1 - \underline{\hat{e}}_1$ of free variates, can be determined from $\underline{\hat{e}}_2=\underline{y}_2 - \underline{\hat{y}}_2$. In the next section we will generalize (8) to the multi-dimensional case.

Free variates are examples of so-called $\underline{y}^R$-variates. The *definition* of $\underline{y}^R$-variates reads as follows:

> $\underline{y}^R$-variates are observables that are either stochastically or functionally Related to another set of observables $\underline{y}$.

There are three types of $\underline{y}^R$-variates:

1. $\underline{y}^R$-variates that correlate with $\underline{y}$-variates. These are the *free variates* (vrije grootheden);
2. $\underline{y}^R$-variates that are functions of $\underline{y}$-variates. These are the *derived variates* (afgeleide grootheden);
3. $\underline{y}^R$-variates of which the $\underline{y}$-variates are functions. These are the *constituent variates* (samenstellende grootheden).

ad 1. These variates occur for instance when new measurements need to be included in existing geodetic networks. Consider for instance the levelling network of figure 3.1 of Chapter 3. Assume that the five heights have been estimated with the available nine levelled height differences. This gives the five estimators of the heights, $\underline{\hat{x}}_1$, $\underline{\hat{x}}_2$, $\underline{\hat{x}}_3$, $\underline{\hat{x}}_4$, $\underline{\hat{x}}_5$. Now assume that a new height difference has been measured, say between the points 1 and 5. If we denote this height difference by y_{10}, we can write the condition equation as:

(9) $$E\{\underline{y}_{10}\} - E\{\underline{\hat{x}}_5 - \underline{\hat{x}}_1\} = 0 \ .$$

On the basis of this condition equation we can compute $\underline{\hat{y}}_{10}$ and $\underline{\hat{x}}_5$ $\underline{\hat{x}}_1$, the improved estimators for the heights of the points 1 and 5. However, since the estimators $\underline{\hat{x}}_2$, $\underline{\hat{x}}_3$ and $\underline{\hat{x}}_4$ are correlated with $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_5$, also the estimators $\underline{\hat{x}}_2$, $\underline{\hat{x}}_3$ and $\underline{\hat{x}}_4$ can be improved. The variates $\underline{\hat{x}}_2$, $\underline{\hat{x}}_3$ and $\underline{\hat{x}}_4$ are in this case the free variates since they do not appear in condition equation (9).

ad 2. Examples of derived variates in the case of the levelling network are height differences which are not directly measured, but which can be computed as functions of the estimated heights.

ad 3. Consider the levelling network of figure 4.5. Assume that the height difference between the points 1 and 2 has not been measured directly, but that instead it equals the sum of a number of measured height differences.

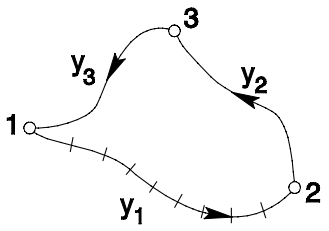


Figure 4.5: Levelling network

One can now include either the individual measured height differences in the condition equation or their sum $\underline{y}_1$. In the last case the individual measured height differences are the constituent variates.

In the following sections it will be shown that formula (8) holds for all three types of $\underline{y}^R$-variates.

4.2 Free variates

In this section we will generalize formula (8) to the multi-dimensional case. We will give two derivations. One based on the model with *condition equations*, and one based on the model with *observation equations*.

We know that the solution of the model with condition equations

(10) $$\boldsymbol{B}^* E\{\underline{y}\} = \mathbf{0} \quad ; \quad E\{(\underline{y} - E\{\underline{y}\})(\underline{y} - E\{\underline{y}\})^*\} = \boldsymbol{Q}_y \ ,$$

reads:

(11) $$\begin{cases} \underline{\hat{y}} = [I - \boldsymbol{Q}_y \boldsymbol{B}(\boldsymbol{B}^* \boldsymbol{Q}_y \boldsymbol{B})^{-1} \boldsymbol{B}^*] \ \underline{y} \\ \underline{\hat{e}} = \underline{y} - \underline{\hat{y}} = \boldsymbol{Q}_y \boldsymbol{B}(\boldsymbol{B}^* \boldsymbol{Q}_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y}. \end{cases}$$

Now let us assume that instead of (10) we have the model:

(12) $$\begin{pmatrix} \boldsymbol{B} \\ 0 \end{pmatrix}^* E\left\{\begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix}\right\} = \mathbf{0}; E\left\{\left[\begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix} - E\left\{\begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix}\right\}\right]\left[\begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix} - E\left\{\begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix}\right\}\right]^*\right\} = \begin{pmatrix} \boldsymbol{Q}_y & \boldsymbol{Q}_{y\boldsymbol{R}} \\ \boldsymbol{Q}_{\boldsymbol{R}y} & \boldsymbol{Q}_{\boldsymbol{R}} \end{pmatrix}.$$

In this case the coefficients of $\underline{y}^R$ are zero; thus $\underline{y}^R$ is a vector of *free variates*. When we compare (12) with (10) we see that B^* of (10) is replaced by (B^* 0) and Q_y of (10) by:

$$\begin{pmatrix} \boldsymbol{Q}_y & \boldsymbol{Q}_{y\boldsymbol{R}} \\ \boldsymbol{Q}_{\boldsymbol{R}y} & \boldsymbol{Q}_{\boldsymbol{R}} \end{pmatrix}.$$

Using the second equation of (11), we therefore get for model (12):

$$\begin{pmatrix} \underline{\hat{e}} \\ \underline{\hat{e}}^{\boldsymbol{R}} \end{pmatrix} = \begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix} - \begin{pmatrix} \underline{\hat{y}} \\ \underline{\hat{y}}^{\boldsymbol{R}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_y & \boldsymbol{Q}_{y\boldsymbol{R}} \\ \boldsymbol{Q}_{\boldsymbol{R}y} & \boldsymbol{Q}_{\boldsymbol{R}} \end{pmatrix} \begin{pmatrix} \boldsymbol{B} \\ 0 \end{pmatrix} (\boldsymbol{B}^* \boldsymbol{Q}_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y}$$

or:

(13) $$\begin{pmatrix} \underline{\hat{e}} \\ \underline{\hat{e}}^{\boldsymbol{R}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_y \boldsymbol{B}(\boldsymbol{B}^* \boldsymbol{Q}_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y} \\ \boldsymbol{Q}_{\boldsymbol{R}y} \boldsymbol{B}(\boldsymbol{B}^* \boldsymbol{Q}_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y} \end{pmatrix}.$$

From the first equation of (13) it follows that:

$$\boldsymbol{Q}_y^{-1} \underline{\hat{e}} = \boldsymbol{B}(\boldsymbol{B}^* \boldsymbol{Q}_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y} \ .$$

If we substitute this into the second equation of (13) we finally get:

(14) $$\boxed{\underline{\hat{e}}^{\boldsymbol{R}} = Q_{\boldsymbol{R}y} Q_y^{-1} \underline{\hat{e}}}$$

This is the multi-dimensional generalization of equation (8) of section 4.1.

We will now derive formula (14) using the model with observation equations. The equivalent of model (12) in terms of observation equations reads:

(15) $$E\{\begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix}\} = \begin{pmatrix} \boldsymbol{A} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} x \\ E\{\underline{y}^{\boldsymbol{R}}\} \end{pmatrix} \; ; \; E\{\begin{bmatrix} \begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix} - E\{\begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix}\} \end{bmatrix} \begin{bmatrix} \begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix} - E\{\begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix}\} \end{bmatrix}^*\} = \begin{pmatrix} Q_y & Q_{y\boldsymbol{R}} \\ Q_{\boldsymbol{R}y} & Q_{\boldsymbol{R}} \end{pmatrix} .$$

Note that:

holds. $$\begin{pmatrix} \boldsymbol{B} \\ 0 \end{pmatrix}^* \begin{pmatrix} \boldsymbol{A} & 0 \\ 0 & I \end{pmatrix} = (0 \; 0)$$

If we denote the inverse of the covariance matrix of the observables by:

(16) $$\begin{pmatrix} Q_y & Q_{y\boldsymbol{R}} \\ Q_{\boldsymbol{R}y} & Q_{\boldsymbol{R}} \end{pmatrix}^{-1} = \begin{pmatrix} G_y & G_{y\boldsymbol{R}} \\ G_{\boldsymbol{R}y} & G_{\boldsymbol{R}} \end{pmatrix} ,$$

the *normal equations* for model (15) take the form:

(17) $$\begin{pmatrix} \boldsymbol{A}^* G_y \boldsymbol{A} & \boldsymbol{A}^* G_{y\boldsymbol{R}} \\ G_{\boldsymbol{R}y} \boldsymbol{A} & G_{\boldsymbol{R}} \end{pmatrix} \begin{pmatrix} \underline{\hat{x}} \\ \underline{\hat{y}}^{\boldsymbol{R}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^* G_y & \boldsymbol{A}^* G_{y\boldsymbol{R}} \\ G_{\boldsymbol{R}y} & G_{\boldsymbol{R}} \end{pmatrix} \begin{pmatrix} \underline{y} \\ \underline{y}^{\boldsymbol{R}} \end{pmatrix} .$$

From the last equation of (17) it follows that:

$$G_{\boldsymbol{R}} \underline{\hat{y}}^{\boldsymbol{R}} = G_{\boldsymbol{R}} \underline{y}^{\boldsymbol{R}} + G_{\boldsymbol{R}y} (\underline{y} - \boldsymbol{A} \underline{\hat{x}})$$

or:

$$\underline{\hat{e}}^{\boldsymbol{R}} = \underline{y}^{\boldsymbol{R}} - \underline{\hat{y}}^{\boldsymbol{R}} = -G_{\boldsymbol{R}}^{-1} G_{\boldsymbol{R}y} (\underline{y} - \boldsymbol{A} \underline{\hat{x}})$$

or:

(18) $$\underline{\hat{e}}^{\boldsymbol{R}} = -G_{\boldsymbol{R}}^{-1} G_{\boldsymbol{R}y} \underline{\hat{e}} \; .$$

This results looks already very similar to (14). However, equation (14) is expressed in terms of covariance matrices, whereas (18) is expressed in terms of weight matrices.
Since:

$$\begin{pmatrix} G_y & G_{y\boldsymbol{R}} \\ G_{\boldsymbol{R}y} & G_{\boldsymbol{R}} \end{pmatrix} \begin{pmatrix} Q_y & Q_{y\boldsymbol{R}} \\ Q_{\boldsymbol{R}y} & Q_{\boldsymbol{R}} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

(see (16)), it follows that $G_{Ry}\ Q_y + G_R\ Q_{Ry} = 0$ and thus that:

(19) $$-G_{\boldsymbol{R}}^{-1} G_{\boldsymbol{R}y} = \boldsymbol{Q}_{\boldsymbol{R}y} \boldsymbol{Q}_y^{-1} \ .$$

This, together with (18) proves (14).

4.3 Derived variates

In this section we will prove that formula (14) also holds for *derived variates*. $\underline{y}^R$-variates are derived variates if they are functions of the observables $\underline{y}$. Let us assume that this functional relationship takes the form:

(20) $$\underset{\boldsymbol{p}\times 1}{\underline{y}^{\boldsymbol{R}}} = \underset{\boldsymbol{p}\times m}{\Lambda}\ \underset{m\times 1}{\underline{y}} \ .$$

Then also:

(21) $$\underline{\hat{y}}^{\boldsymbol{R}} = \Lambda\ \underline{\hat{y}} \ .$$

Subtraction of (21) from (20) gives:

(22) $$\underline{\hat{\boldsymbol{e}}}^{\boldsymbol{R}} = \Lambda\ \underline{\hat{\boldsymbol{e}}} \ .$$

What remains to be shown is now that Λ equals $Q_{Ry}\ Q_y^{-1}$. Application of the *propagation law of covariances* to (20) gives:

$$\boldsymbol{Q}_{\boldsymbol{R}y} = \Lambda\ \boldsymbol{Q}_y$$

or:

$$\Lambda = \boldsymbol{Q}_{\boldsymbol{R}y} \boldsymbol{Q}_y^{-1} \ .$$

This, together with (22) shows that formula (14) also holds for derived variates.

4.4 Constituent variates

In this section we will prove that formula (14) holds for *constituent variates*. Constituent variates are $\underline{y}^R$-variates of which the $\underline{y}$-variates are functions. Let us assume that this functional relationship takes the form:

(23) $$\underset{m\times 1}{\underline{y}} = \underset{m\times \boldsymbol{p}}{\Lambda}\ \underset{\boldsymbol{p}\times 1}{\underline{y}^{\boldsymbol{R}}} \ .$$

Application of the propagation law of variances gives:

(24) $$\boldsymbol{Q}_y = \Lambda\ \boldsymbol{Q}_{\boldsymbol{R}}\ \Lambda^* \ ,$$

and application of the propagation law of covariances gives:

(25) $$Q_{\boldsymbol{R}y} = Q_{\boldsymbol{R}}\Lambda^* \ .$$

As we know $\underline{\hat{e}} = \underline{y} - \underline{\hat{y}}$ follows from solving the model:

(26) $$\boldsymbol{B}^* E\{\underline{y}\} = 0 \ ; \ E\{(\underline{y} - E\{\underline{y}\})(\underline{y} - E\{\underline{y}\})^*\} = Q_y$$

as:

(27) $$\underline{\hat{e}} = Q_y \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y} \ .$$

With (23) we may formulate the model with condition equations also as:

(28) $$\boldsymbol{B}^* \Lambda E\{\underline{y}^{\boldsymbol{R}}\} = 0 \ ; \ E\{(\underline{y}^{\boldsymbol{R}} - E\{\underline{y}^{\boldsymbol{R}}\})(\underline{y} - E\{\underline{y}^{\boldsymbol{R}}\})^*\} = Q_{\boldsymbol{R}} \ .$$

When compared with (26) this means that B of (26) is replaced by $\Lambda^* B$, that $\underline{y}$ is replaced by $\underline{y}^R$ and that Q_y is replaced by Q_R. Instead of (27) we therefore get for model (28):

(29) $$\underline{\hat{e}}^{\boldsymbol{R}} = Q_{\boldsymbol{R}}\Lambda^* \boldsymbol{B}\left(\boldsymbol{B}^* \Lambda Q_{\boldsymbol{R}} \Lambda^* \boldsymbol{B}\right)^{-1} \boldsymbol{B}^* \Lambda \underline{y}^{\boldsymbol{R}} \ .$$

With (23), (24) and (25) this can also be written as:

$$\underline{\hat{e}}^{\boldsymbol{R}} = Q_{\boldsymbol{R}y} \boldsymbol{B}\left(\boldsymbol{B}^* Q_y \boldsymbol{B}\right)^{-1} \boldsymbol{B}^* \underline{y}$$

or with (27) as:

$$\underline{\hat{e}}^{\boldsymbol{R}} = Q_{\boldsymbol{R}y} Q_y^{-1} \underline{\hat{e}} \ ,$$

which is identical to (14).

Let us, as an example of the above, "rederive" the solution of the model with condition equations. The model with condition equations reads:

(30) $$\boldsymbol{B}^* E\{\underline{y}\} = 0 \ ; \ E\{(\underline{y} - E\{\underline{y}\})(\underline{y} - E\{\underline{y}\})^*\} = Q_y \ .$$

If we define the misclosure vector as:

(31) $$\underline{t} = \boldsymbol{B}^* \underline{y} \ ,$$

then (30) can also be written as:

(32) $$I^* \ E\{\underline{t}\} = 0 \ ; \ E\{(\underline{t} - E\{\underline{t}\})(\underline{t} - E\{\underline{t}\})^*\} = Q_t \ .$$

This model is also in the form of condition equations. The coefficient matrix of the condition equations is in this case the unit matrix I. If we solve for model (32) we get for $\underline{\hat{e}}_t = \underline{t} - \underline{\hat{t}}$:

(33) $$\underline{\hat{e}}_t = Q_t I(I^* Q_t I)^{-1} I^* \underline{t} = \underline{t} \ .$$

We can now use our formula (14) in order to derive $\underline{\hat{e}} = \underline{y} - \underline{\hat{y}}$. This is done by interpreting $\underline{y}$ of (31) as an $\underline{y}^R$-variate and $\underline{t}$ of (31) as an $\underline{y}$-variate. Application of formula (14) gives then:

(34) $$\underline{\hat{e}} = Q_{yt} Q_t^{-1} \underline{\hat{e}}_t \ .$$

Application of the propagation law of variances to (31) gives:

(35) $$Q_t = \boldsymbol{B}^* Q_y \boldsymbol{B} \ ,$$

and application of the propagation law of covariances to (31) gives:

(36) $$Q_{yt} = Q_y \boldsymbol{B} \ .$$

Substitution of (33), (35) and (36) into (34) gives with (31) then finally:

$$\underline{\hat{e}} = Q_y \boldsymbol{B} (\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y} \ .$$

Hence, we have obtained our well-known and familiar result again.

5 Mixed model representations

5.1 Introduction

In chapter 2 we developed the theory of linear estimation for the model with *observation equations*:

(1) $$\boxed{E\{\underline{y}\} = \boldsymbol{A}x \quad ; \quad E\{(\underline{y}-\boldsymbol{A}x)(\underline{y}-\boldsymbol{A}x)^*\} = \boldsymbol{Q}_y.}$$

It was shown in chapter 2 that the estimator $\underline{\hat{x}}$ of x follows from solving the system of *normal equations*:

(2) $$(\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\boldsymbol{A})\ \underline{\hat{x}} = \boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\underline{y}$$

as:

(3) $$\underline{\hat{x}} = (\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\underline{y}\ .$$

From this result the estimators $\underline{\hat{y}} = A\underline{\hat{x}}$ and $\hat{e} = y - \hat{y}$ follow as:

(4) $$\underline{\hat{y}} = \boldsymbol{A}(\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\underline{y}$$

and:

(5) $$\underline{\hat{e}} = [I-\boldsymbol{A}(\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{Q}_y^{-1}]\ \underline{y}\ .$$

In chapter 3 we developed the theory of linear estimation for the model with *condition equations*:

(6) $$\boxed{\boldsymbol{B}^*E\{\underline{y}\} = 0 \quad ; \quad E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\} = \boldsymbol{Q}_y.}$$

In this case the estimators $\underline{\hat{y}}$ and $\underline{\hat{e}} = y - \underline{\hat{y}}$ can be represented in terms of the matrix B as:

(7) $$\underline{\hat{y}} = [I-\boldsymbol{Q}_y\boldsymbol{B}(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{B}^*]\ \underline{y}$$

and:

(8) $$\underline{\hat{e}} = \boldsymbol{Q}_y\boldsymbol{B}(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{B}^*\underline{y}\ .$$

In the present chapter we will introduce two additional model representations. The first new model representation we will consider takes the form:

(9) $$\boxed{\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x}\ ;\ \underset{m\times m}{E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\}} = \underset{m\times m}{\boldsymbol{Q}_y}.}$$

In this model representation x is an $n\times 1$ vector of unknown parameters and $\underline{y}$ is an $m\times 1$ vector of observables. Note that (9) is a mixture of (1) and (6). If matrix B of (9) equals the identity matrix, then (9) reduces to the model with observation equations (1). If on the other hand matrix

A of (9) equals the zero matrix, then (9) reduces to the model with condition equations (6). In the next section the solution of (9) will be derived.

The second new model representation that we will consider takes the form:

(10) $$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x}\ ,\ \underset{b\times n}{\boldsymbol{B}^*}\ \underset{n\times 1}{x} = \underset{b\times 1}{0}\ ;\ \underset{m\times m}{E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\}} = \underset{m\times m}{\boldsymbol{Q}_y}.$$

This representation is in the form of observation equations with conditions on the parameter vector x. If B=0 then (10) reduces simply to (1). The solution of (10) will be derived in section 5.3.

5.2 The model representation $B^*E\{\underline{y}\}=Ax$

Our starting point is the representation:

(11) $$\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x}\ ;\ \underset{m\times m}{E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\}} = \underset{m\times m}{\boldsymbol{Q}_y}.$$

We will assume that both matrices A and B have full column rank. Thus:

(12) $$\text{rank } \boldsymbol{A} = n \quad \text{and} \quad \text{rank } \boldsymbol{B} = b\ .$$

Note that matrix A can only have full column rank if $n \leq b$: the number of unknown parameters should not exceed the number of conditions on the observations.

There are different derivations possible for deriving the estimators $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$ for model (11). One possibility would be to transform (11) into a form with *observation equations only*. This can be done as follows. Consider the system:

(13) $$\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x}$$

as an inhomogeneous system of linear equations with the unknown vector $E\{\underline{y}\}$. Vector $E\{\underline{y}\}$ can then be written as the sum of a particular solution of (13) and the homogeneous solution of:

(14) $$\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0}\ .$$

If we denote the $m\times(m-b)$ matrix of which the column vectors are orthogonal to the column vectors of matrix B by $B^{\perp}$, then:

$$\underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times(m-b)}{\boldsymbol{B}^{\perp}} = \underset{b\times(m-b)}{0}\ .$$

From this it follows that the solution of (14) takes the form:

(15) $$\underset{m\times 1}{E\{\underline{y}\}_h} = \underset{m\times(m-b)}{\boldsymbol{B}^{\perp}} \quad \underset{(m-b)\times 1}{\lambda} \quad .$$

This is the homogeneous solution. A particular solution of (13) is:

(16) $$\underset{m\times 1}{E\{\underline{y}\}_p} = \underset{m\times n}{\boldsymbol{B}(\boldsymbol{B}^{*}\boldsymbol{B})^{-1}\boldsymbol{A}} \quad \underset{n\times 1}{x} \qquad \text{(verify !)}$$

Hence, the solution for $E\{\underline{y}\}$ of (13) takes the form of the sum of the particular solution (16) and the homogeneous solution (15):

(17) $$E\{\underline{y}\} = \boldsymbol{B}(\boldsymbol{B}^{*}\boldsymbol{B})^{-1}\boldsymbol{A}x + \boldsymbol{B}^{\perp}\lambda = [\boldsymbol{B}(\boldsymbol{B}^{*}\boldsymbol{B})^{-1}\boldsymbol{A} : \boldsymbol{B}^{\perp}] \begin{pmatrix} x \\ \lambda \end{pmatrix}.$$

Note that this representation is in terms of observation equations only. Since (17) is completely equivalent to the first equation of (11), we can now apply our solution method for observation equations to (17) in order to get the solution of (11).

Another possibility to solve for (11) would be to transform (11) into a form with *condition equations only*. This can be done as follows. Consider again the system (13). If we denote the $b\times(b-n)$ matrix of which the column vectors are orthogonal to the column vectors of matrix A by $A^{\perp}$, then

$$\underset{(b-n)\times b}{\boldsymbol{A}^{\perp *}} \quad \underset{b\times n}{\boldsymbol{A}} = \underset{(b-n)\times n}{0} \quad .$$

From this it follows that pre-multiplication of (13) with $A^{\perp *}$ gives:

(18) $$\underset{(b-n)\times m}{(\boldsymbol{B}\boldsymbol{A}^{\perp})^{*}} \quad \underset{m\times 1}{E\{\underline{y}\}} = \underset{(b-n)\times 1}{0} \quad .$$

Note that this representation is in terms of condition equations only. Since (18) is completely equivalent to the first equation of (11), we can now apply our solution method for condition equations to (18) in order to get the solution of (11).

Instead of following the above two mentioned approaches for solving (11), we will make use of a third approach. This approach makes use of the concept of $\underline{y}^{R}$-variates. Define a $b\times 1$ vector $\underline{t}$ as:

(19) $$\underset{b\times 1}{\underline{t}} = \underset{b\times m}{\boldsymbol{B}^{*}} \quad \underset{m\times 1}{\underline{y}} \quad .$$

Substitution into (11) gives the representation:

(20) $$\underset{b\times 1}{E\{\underline{t}\}} = \underset{b\times n}{\boldsymbol{A}} \quad \underset{n\times 1}{x} \quad ; \quad \underset{b\times b}{E\{(\underline{t}-E\{\underline{t}\})(\underline{t}-E\{\underline{t}\})^{*}\}} = \underset{b\times b}{Q_t} \quad .$$

This representation is in the form of observation equations only. Hence, we know how to compute the estimators $\underline{\hat{x}}$, $\underline{\hat{t}}$ and $\underline{\hat{e}}_t$. The estimator $\underline{\hat{x}}$ follows as:

$$\underline{\hat{x}} = (\boldsymbol{A}^*\boldsymbol{Q}_t^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{Q}_t^{-1}\underline{t} \quad . \tag{21}$$

Application of the propagation law for variances to (19) gives for $\boldsymbol{Q}_t$:

$$\boldsymbol{Q}_t = \boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B} \quad . \tag{22}$$

Substitution of (19) and (22) into (21) gives then for the estimator $\underline{\hat{x}}$ of x:

$$\boxed{\underline{\hat{x}} = [\boldsymbol{A}^*(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{A}]^{-1}\boldsymbol{A}^*(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{B}^*\underline{y}.} \tag{23}$$

The estimators $\underline{\hat{t}}$ and $\underline{\hat{e}}_t$ follow simply as:

$$\begin{cases} \underline{\hat{t}} = \boldsymbol{A}\underline{\hat{x}} \\ \underline{\hat{e}}_t = \underline{t} - \underline{\hat{t}} = \underline{t} - \boldsymbol{A}\underline{\hat{x}} \end{cases} \quad . \tag{24}$$

Now, in order to obtain $\underline{\hat{e}}_y = \underline{y} - \underline{\hat{y}}$, note that $\underline{y}$ of (19) can be interpreted as a constituent $\underline{y}^R$-variate.

Application of formula (14) of chapter 4 gives then:

$$\underline{\hat{e}}_y = \boldsymbol{Q}_{yt}\boldsymbol{Q}_t^{-1}\underline{\hat{e}}_t \quad . \tag{25}$$

The matrix $\boldsymbol{Q}_{yt}$ follows from application of the propagation law of covariances to (19) as:

$$\boldsymbol{Q}_{yt} = \boldsymbol{Q}_y\boldsymbol{B} \quad . \tag{26}$$

The estimators $\underline{\hat{e}}_y$ and $\underline{\hat{y}} = \underline{y} - \underline{\hat{e}}_y$ follow then finally with (22), (23), (24), (25), and (26) as:

$$\boxed{\begin{aligned} \underline{\hat{e}}_y &= \boldsymbol{Q}_y\boldsymbol{B}(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}[\boldsymbol{I} - \boldsymbol{A}[\boldsymbol{A}^*(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{A}]^{-1}\boldsymbol{A}^*(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}]\boldsymbol{B}^*\underline{y} \\ \underline{\hat{y}} &= \underline{y} - \boldsymbol{Q}_y\boldsymbol{B}(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}[\boldsymbol{I} - \boldsymbol{A}[\boldsymbol{A}^*(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}\boldsymbol{A}]^{-1}\boldsymbol{A}^*(\boldsymbol{B}^*\boldsymbol{Q}_y\boldsymbol{B})^{-1}]\boldsymbol{B}^*\underline{y} \end{aligned}} \quad . \tag{27}$$

5.3 The model representation $E\{\underline{y}\}=\boldsymbol{A}x$, $\boldsymbol{B}^*x=0$

Our starting point is the representation:

$$\boxed{\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x}\ ,\ \underset{\boldsymbol{b}\times n}{\boldsymbol{B}^*}\ \underset{n\times 1}{x} = \underset{\boldsymbol{b}\times 1}{0}\ ;\ \underset{m\times m}{E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^*\}} = \underset{m\times m}{\boldsymbol{Q}_y}.} \tag{29}$$

This representation is in the form of observation equations with conditions on the parameter vector x. In order to derive the estimators $\underline{\hat{x}}$, $\underline{\hat{y}}$ and $\underline{\hat{e}}$ we will first transform (29) into a form with observation equations only. This is done by finding the parametric representation for:

(30) $$\underset{b\times n}{\boldsymbol{B}^{*}}\ \underset{n\times 1}{x} = \underset{b\times 1}{0} \ .$$

Let us denote the $n\times(n\text{-}b)$ matrix of which the column vectors are orthogonal to the column vectors of matrix B by $B^{\perp}$. Then:

(31) $$\underset{b\times n}{\boldsymbol{B}^{*}}\ \underset{n\times(n-b)}{\boldsymbol{B}^{\perp}} = \underset{b\times(n-b)}{0} \ .$$

With (31) the parametric representation of the homogeneous system, (30) becomes:

(32) $$\underset{n\times 1}{x} = \underset{n\times(n-b)}{\boldsymbol{B}^{\perp}}\ \underset{(n-b)\times 1}{\lambda} \ .$$

Thus (32) is completely equivalent to (30). It is remarked that we assume that both the matrices A and B are of full column rank; thus rank $A = n$ and rank $B = b$. Note that matrix B can only have full column rank if $b \leq n$: one cannot impose more than n independent conditions on the n unknown parameters.

If we substitute (32) into the first equation of (29) we get:

(33) $$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times(n-b)}{\boldsymbol{A}\boldsymbol{B}^{\perp}}\ \underset{(n-b)\times 1}{\lambda} \quad ; \quad \underset{m\times m}{E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^{*}\}} = \underset{m\times m}{\boldsymbol{Q}_{y}} \ .$$

This representation is completely equivalent to (29) and it is in the form of observation equations only. Thus we know how to compute the estimator $\underline{\hat{\lambda}}$ of λ. This estimator reads:

(34) $$\underline{\hat{\lambda}} = (\boldsymbol{B}^{\perp *}\boldsymbol{A}^{*}\boldsymbol{Q}_{y}^{-1}\boldsymbol{A}\boldsymbol{B}^{\perp})^{-1}\boldsymbol{B}^{\perp *}\boldsymbol{A}^{*}\boldsymbol{Q}_{y}^{-1}\underline{y} \ .$$

Now let us consider the solution of (29) *without* the conditions on the parameter vector. This is our standard model representation with observation equations. As you know the solution of (29) without the conditions on the parameter vector satisfies the normal equations:

(35) $$\boldsymbol{A}^{*}\boldsymbol{Q}_{y}^{-1}\boldsymbol{A}\underline{\hat{x}}_{\boldsymbol{A}} = \boldsymbol{A}^{*}\boldsymbol{Q}_{y}^{-1}\underline{y} \ .$$

We have given $\underline{\hat{x}}_{A}$ of (35) the subscript A to emphasize that $\underline{\hat{x}}_{A}$ is *not* the solution of (29) but of:

(36) $$E\{\underline{y}\} = \boldsymbol{A}x \quad ; \quad E\{(\underline{y}-E\{\underline{y}\})(\underline{y}-E\{\underline{y}\})^{*}\} = \boldsymbol{Q}_{y} \ .$$

The covariance matrix of $\underline{\hat{x}}_{A}$ reads:

(37) $$Q_{\hat{x}_A} = (A^* Q_y^{-1} A)^{-1} \ .$$

With (35) and (37) we can now write (34) as:

(38) $$\underline{\hat{\lambda}} = (B^{\perp *} Q_{\hat{x}_A}^{-1} B^{\perp})^{-1} B^{\perp *} Q_{\hat{x}_A}^{-1} \underline{\hat{x}}_A \ .$$

This result together with (32) gives the estimator of x for model (29):

(39) $$\underline{\hat{x}} = B^{\perp} (B^{\perp *} Q_{\hat{x}_A}^{-1} B^{\perp})^{-1} B^{\perp *} Q_{\hat{x}_A}^{-1} \underline{\hat{x}}_A \ ,$$

If we look at the two formulae (38) and (39) we note that they take a very familiar form. And indeed, (38) and (39) are the solution of the model:

(40) $$E\{\underline{\hat{x}}_A\} = B^{\perp} \lambda \quad ; \quad E\{(\underline{\hat{x}}_A - E\{\underline{\hat{x}}_A\})(\underline{\hat{x}}_A - E\{\underline{\hat{x}}_A\})^*\} = Q_{\hat{x}_A} \ .$$

This model representation is in terms of observation equations. The equivalent of (40) in terms of condition equations follows with (31) as:

(41) $$B^* E\{\underline{\hat{x}}_A\} = 0 \ ; \ E\{(\underline{\hat{x}}_A - E\{\underline{\hat{x}}_A\})(\underline{\hat{x}}_A - E\{\underline{\hat{x}}_A\})^*\} = Q_{\hat{x}_A} \ .$$

As you know the solution of (41) is:

(42) $$\underline{\hat{\hat{x}}}_A = [I - Q_{\hat{x}_A} B (B^* Q_{\hat{x}_A} B)^{-1} B^*] \underline{\hat{x}}_A \ .$$

But since (41) is equivalent to (40) and (39) is the solution to (40), (42) must be identical to (39). This implies that the estimator $\underline{\hat{x}}$ of x for model (29) can be written as:

(43) $$\underline{\hat{x}} = [I - Q_{\hat{x}_A} B (B^* Q_{\hat{x}_A} B)^{-1} B^*] \underline{\hat{x}}_A \ .$$

We thus have now found an expression for the estimator $\underline{\hat{x}}$ of x for model (29), which is in terms of the matrices A and B of (29). It follows then with (35), (37) and (43) that we can compute the estimators for model (29) as:

(44)

$$\begin{aligned}
\underline{\hat{x}}_A &= Q_{\hat{x}_A} A^* Q_y^{-1} \underline{y} \quad ; \quad Q_{\hat{x}_A} = (A^* Q_y^{-1} A)^{-1} \\
\underline{\hat{x}} &= [I - Q_{\hat{x}_A} B (B^* Q_{\hat{x}_A} B)^{-1} B^*] \underline{\hat{x}}_A \\
\underline{\hat{y}} &= A \underline{\hat{x}} \\
\underline{\hat{e}} &= \underline{y} - \underline{\hat{y}}
\end{aligned}$$

This result shows that the solution of (29) can be obtained in *two steps*. In the first step one considers (29) without the conditions on the parameters. That is, one solves for the model with *observation equations* (36). This gives the estimator $\underline{\hat{x}}_A$. Then in a second step one includes the conditions on the parameters. That is, one solves for the model with *condition equations* (41). This then finally gives the estimator $\underline{\hat{x}}$.

It follows from (44) that:

$$\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}} = \|\underline{\hat{e}}\|^2 = \|\underline{y} - A\underline{\hat{x}}_A + AQ_{\hat{x}_A}B(B^* Q_{\hat{x}_A}B)^{-1}B^* \underline{\hat{x}}_A\|^2,$$

or that:

$$\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}} = \|P_A^{\perp} \underline{y} + AQ_{\hat{x}_A}B(B^* Q_{\hat{x}_A}B)^{-1}B^* \underline{\hat{x}}_A\|^2 .$$

Since $P_A^{\perp *} Q_y^{-1} A = 0$, this may also be written as:

$$\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}} = \|P_A^{\perp} \underline{y}\|^2 + \|AQ_{\hat{x}_A}B(B^* Q_{\hat{x}_A}B)^{-1}B^* \underline{\hat{x}}_A\|^2 .$$

This gives finally:

(45)
$$\boxed{\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}} = \underline{y}^* Q_y^{-1} P_A^{\perp} \underline{y} + \underline{\hat{x}}_A^* B(B^* Q_{\hat{x}_A}B)^{-1}B^* \underline{\hat{x}}_A} .$$

Note that the first term on the right hand side of (45) corresponds with the squared norm of the least-squares residual vector of the model $E\{\underline{y}\}=Ax$, whereas the second term on the right hand side of (45) corresponds with the squared norm of the least-squares residual vector of the model $B^* E\{\underline{\hat{x}}_A\} = 0$.

6 Partitioned model representations

6.1 Introduction

In chapters 2 and 3 the theory of linear estimation was developed for the model with *observation equations* and for the model with *condition equations*. Recall from chapter 2 that the estimators of the model:

(1)
$$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x}\ ;\ \underset{m\times m}{\boldsymbol{D}\{\underline{y}\}} = \underset{m\times m}{Q_y}$$ [1)]

are given as:

(2)
$$\begin{cases} \hat{\underline{x}} = (\boldsymbol{A}^*Q_y^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^*Q_y^{-1}\underline{y} \ ; \ Q_{\hat{x}} = (\boldsymbol{A}^*Q_y^{-1}\boldsymbol{A})^{-1} \\ \hat{\underline{y}} = \boldsymbol{A}\hat{\underline{x}} = \boldsymbol{P}_{\boldsymbol{A}}\underline{y} \ ; \ Q_{\hat{y}} = \boldsymbol{P}_{\boldsymbol{A}}Q_y \\ \hat{\underline{e}} = \underline{y}-\hat{\underline{y}} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\underline{y} \ ; \ Q_{\hat{e}} = Q_y - Q_{\hat{y}} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp}Q_y \\ \hat{\underline{e}}^*Q_y^{-1}\hat{\underline{e}} = \underline{y}^*Q_y^{-1}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\underline{y} \ ; \ \sigma^2_{\hat{e}^*Q_y^{-1}\hat{e}} = 2(m-n). \end{cases}$$

And recall from chapter 3 that the estimators of the model:

(3)
$$\underset{\boldsymbol{b}\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{E\{\underline{y}\}} = \underset{\boldsymbol{b}\times 1}{0}\ ;\ \underset{m\times m}{\boldsymbol{D}\{\underline{y}\}} = \underset{m\times m}{Q_y}$$

are given as:

(4)
$$\begin{cases} \hat{\underline{y}} = [I - Q_y\boldsymbol{B}(\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1}\boldsymbol{B}^*]\underline{y} = \boldsymbol{P}_{Q_y\boldsymbol{B}}^{\perp}\underline{y} \ ; \ Q_{\hat{y}} = \boldsymbol{P}_{Q_y\boldsymbol{B}}^{\perp}Q_y \\ \hat{\underline{e}} = \underline{y}-\hat{\underline{y}} = \boldsymbol{P}_{Q_y\boldsymbol{B}}\underline{y} \ ; \ Q_{\hat{e}} = Q_y - Q_{\hat{y}} = \boldsymbol{P}_{Q_y\boldsymbol{B}}Q_y \\ \hat{\underline{e}}^*Q_y^{-1}\hat{\underline{e}} = \underline{y}^*Q_y^{-1}\boldsymbol{P}_{Q_y\boldsymbol{B}}\underline{y} \ ; \ \sigma^2_{\hat{e}^*Q_y^{-1}\hat{e}} = 2\boldsymbol{b} \end{cases}.$$

In this chapter we will partition the matrices A and B of (1) and (3) respectively, *columnwise* and *rowwise*. The following four cases will therefore be considered in this chapter:

$$1^{\circ} \qquad E\{\underline{y}\} = (\boldsymbol{A}_1 \vdots \boldsymbol{A}_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

[1)] The notation "$D\{\underline{y}\}$" will be used instead of the notation "$E\{(\underline{y}\text{-}Ax)(\underline{y}\text{-}Ax)^*\}$". The kernel "$D$" stands for "*Dispersion*".

$$2^{o} \qquad E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix}\} = \begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{pmatrix} x \ ,$$

$$3^{o} \qquad \begin{pmatrix} \boldsymbol{B}_1^{*} \\ \boldsymbol{B}_2^{*} \end{pmatrix} E\{\underline{y}\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} ,$$

$$4^{o} \qquad \left(\boldsymbol{B}_1^{*} \ \boldsymbol{B}_2^{*}\right) E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix}\} = 0 \ .^{2)}$$

In this chapter it will be shown how the above partitioning of A and B can be carried through to the estimators of (2) and (4) respectively.

6.2 The partitioned model $E\{\underline{y}\} = (\boldsymbol{A}_1 \vdots \boldsymbol{A}_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.

Let us consider the partitioned model:

$$E\{\underline{y}\} = (\boldsymbol{A}_1 \vdots \boldsymbol{A}_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ ; \ D\{\underline{y}\} = \boldsymbol{Q}_y \ . \tag{5}$$
$$m\times 1 \qquad m\times n_1 \ m\times n_2 \qquad (n_1+n_2)\times 1 \qquad m\times m \qquad m\times m$$

The corresponding partitioned system of *normal equations* reads:

$$\begin{pmatrix} \boldsymbol{A}_1^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}_1 & \boldsymbol{A}_1^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}_2 \\ \boldsymbol{A}_2^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}_1 & \boldsymbol{A}_2^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}_2 \end{pmatrix} \begin{pmatrix} \hat{\underline{x}}_1 \\ \hat{\underline{x}}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_1^{*}\boldsymbol{Q}_y^{-1}\underline{y} \\ \boldsymbol{A}_2^{*}\boldsymbol{Q}_y^{-1}\underline{y} \end{pmatrix} . \tag{6}$$

Now let us assume that for a particular application we are only interested in the estimator $\hat{\underline{x}}_1$. There are two ways for finding $\hat{\underline{x}}_1$. Either we solve the complete system of normal equations (6) and then look at the first n_1-elements of the estimator $\hat{\underline{x}}$. Or, we eliminate $\hat{\underline{x}}_2$ from the first n_1 equations of (6) and solve for $\hat{\underline{x}}_1$. This second approach is more efficient since it gets round the necessity of solving the complete system of normal equations. The estimator $\hat{\underline{x}}_2$ gets eliminated from the first n_1 equations of (6), if we premultiply (6) with the *square* and *full rank* matrix:

$$\begin{pmatrix} I & -\boldsymbol{A}_1^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}_2(\boldsymbol{A}_2^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}_2)^{-1} \\ 0 & I \end{pmatrix} . \tag{7}$$

Premultiplication of (6) with (7) gives:

[2)] This case will not be dealt with explicitely in this chapter.

$$(8)\quad \begin{pmatrix} A_1^* Q_y^{-1}[I - A_2(A_2^* Q_y^{-1} A_2)^{-1} A_2^* Q_y^{-1}]A_1 & 0 \\ A_2^* Q_y^{-1} A_1 & A_2^* Q_y^{-1} A_2 \end{pmatrix} \begin{pmatrix} \hat{\underline{x}}_1 \\ \hat{\underline{x}}_2 \end{pmatrix} = \begin{pmatrix} A_1^* Q_y^{-1}[I - A_2(A_2^* Q_y^{-1} A_2)^{-1} A_2^* Q_y^{-1}]\underline{y} \\ A_2^* Q_y^{-1} \underline{y} \end{pmatrix}$$

In the first n_1 equations of (8) we recognize the expression for the orthogonal projector:

$$(9)\quad P_{A_2}^{\perp} = I - A_2(A_2^* Q_y^{-1} A_2)^{-1} A_2^* Q_y^{-1} \ .$$

Using the properties:

$$(10)\quad Q_y^{-1} P_{A_2}^{\perp} = P_{A_2}^{\perp *} Q_y^{-1} = P_{A_2}^{\perp *} Q_y^{-1} P_{A_2}^{\perp} \ ,$$

and the abbreviation:

$$(11)\quad \bar{A}_1 = P_{A_2}^{\perp} A_1 \ ,$$

we may therefore write (8) also as:

$$(12)\quad \begin{pmatrix} \bar{A}_1^* Q_y^{-1} \bar{A}_1 & 0 \\ A_2^* Q_y^{-1} A_1 & A_2^* Q_y^{-1} A_2 \end{pmatrix} \begin{pmatrix} \hat{\underline{x}}_1 \\ \hat{\underline{x}}_2 \end{pmatrix} = \begin{pmatrix} \bar{A}_1^* Q_y^{-1} \underline{y} \\ A_2^* Q_y^{-1} \underline{y} \end{pmatrix} .^{3)}$$

The solution for $\hat{\underline{x}}_1$ follows therefore as:

$$(13)\quad \hat{\underline{x}}_1 = (\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1} \ \bar{A}_1^* Q_y^{-1} \underline{y} \ .$$

Application of the propagation law of variances gives:

$$(14)\quad Q_{\hat{x}_1} = (\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1} \ .$$

Once $\hat{\underline{x}}_1$ is known, $\hat{\underline{x}}_2$ can be found from the last n_2 equations of (12). This gives:

$$(15)\quad \hat{\underline{x}}_2 = (A_2^* Q_y^{-1} A_2)^{-1} \ A_2^* Q_y^{-1} (\underline{y} - A_1 \hat{\underline{x}}_1) \ .$$

Application of the propagation law of variances gives with (13) and (14):

$$Q_{\hat{x}_2} = (A_2^* Q_y^{-1} A_2)^{-1} \ A_2^* Q_y^{-1} \ [Q_y - \bar{A}_1 Q_{\hat{x}_1} A_1^* - A_1 Q_{\hat{x}_1} \bar{A}_1^* + A_1 Q_{\hat{x}_1} A_1^*] Q_y^{-1} A_2 (A_2^* Q_y^{-1} A_2)^{-1} \ .$$

Since $A_2^* Q_y^{-1} \bar{A}_1 = 0$, this reduces to:

$$(16)\quad Q_{\hat{x}_2} = (A_2^* Q_y^{-1} A_2)^{-1} + (A_2^* Q_y^{-1} A_2)^{-1} A_2^* Q_y^{-1} A_1 Q_{\hat{x}_1} A_1^* Q_y^{-1} A_2 \ (A_2^* Q_y^{-1} A_2)^{-1} \ .$$

Summarizing, we thus have shown that:

---

[3] The normal matrix of this system is called a *"reduced normal matrix"*. It is said to be reduced for x_2.

(17)
$$\begin{aligned}
\underline{\hat{x}}_1 &= (\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1}\ \bar{A}_1^* Q_y^{-1} \underline{y}\ , \text{ with } \bar{A}_1 = P_{A_2}^{\perp} A_1\ , \\
Q_{\hat{x}_1} &= (\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1}\ , \\
\underline{\hat{x}}_2 &= (A_2^* Q_y^{-1} A_2)^{-1}\ A_2^* Q_y^{-1} (\underline{y} - A_1 \underline{\hat{x}}_1)\ , \\
Q_{\hat{x}_2} &= (A_2^* Q_y^{-1} A_2)^{-1} + (A_2^* Q_y^{-1} A_2)^{-1}\ A_2^* Q_y^{-1} A_1 Q_{\hat{x}_1} A_1^* Q_y^{-1} A_2\ (A_2^* Q_y^{-1} A_2)^{-1}
\end{aligned}$$

Instead of eliminating the estimator $\underline{\hat{x}}_2$ from the first n_1 equations of (6), one may also eliminate the estimator $\underline{\hat{x}}_1$ from the last n_2 equations of (6). This is achieved by premultiplying (6) with the *square* and *full rank* matrix:

(18)
$$\begin{pmatrix} I & 0 \\ -A_2^* Q_y^{-1} A_1\ (A_1^* Q_y^{-1} A_1)^{-1} & I \end{pmatrix}\ .$$

This gives analogous to (12) the equations:

(19)
$$\begin{pmatrix} A_1^* Q_y^{-1} A_1 & A_1^* Q_y^{-1} A_2 \\ 0 & \bar{A}_2^* Q_y^{-1} \bar{A}_2 \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \end{pmatrix} = \begin{pmatrix} A_1^* Q_y^{-1} \underline{y} \\ \bar{A}_2^* Q_y^{-1} \underline{y} \end{pmatrix}\ ,$$

with

(20)
$$\bar{A}_2 = P_{A_1}^{\perp} A_2\ .$$

From (19) the estimators $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_2$ and their variance matrices follow as:

(21)
$$\begin{aligned}
\underline{\hat{x}}_1 &= (A_1^* Q_y^{-1} A_1)^{-1}\ A_1^* Q_y^{-1} (\underline{y} - A_2 \underline{\hat{x}}_2) \\
Q_{\hat{x}_1} &= (A_1^* Q_y^{-1} A_1)^{-1} + (A_1^* Q_y^{-1} A_1)^{-1}\ A_1^* Q_y^{-1} A_2 Q_{\hat{x}_2} A_2^* Q_y^{-1}\ A_1 (A_1^* Q_y^{-1} A_1)^{-1} \\
\underline{\hat{x}}_2 &= (\bar{A}_2^* Q_y^{-1} \bar{A}_2)^{-1}\ \bar{A}_2^* Q_y^{-1} \underline{y}\ , \text{ with } \bar{A}_2 = P_{A_1}^{\perp} A_2 \\
Q_{\hat{x}_2} &= (\bar{A}_2^* Q_y^{-1} \bar{A}_2)^{-1}
\end{aligned}$$

Compare (21) with (17) and note that (21) follows from (17) by interchanging the role of x_1 and x_2. The above results can now be used to investigate what happens to the precision of the estimators when a model with observation equations is enlarged with additional parameters. Let us assume that our original model is given as:

(22)
$$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n_1}{A_1}\ \underset{n_1\times 1}{x_1}\ ;\ \underset{m\times m}{D\{\underline{y}\}} = \underset{m\times m}{Q_y}\ .$$

The corresponding estimator and its variance matrix will be denoted by $\underline{\hat{x}}_1^1$ and $Q_{\hat{x}_1^1}$ respectively. Then:

(23)
$$\begin{cases} \underline{\hat{x}}_1^1 = (A_1^* Q_y^{-1} A_1)^{-1}\ A_1^* Q_y^{-1} \underline{y} \\ Q_{\hat{x}_1^1} = (A_1^* Q_y^{-1} A_1)^{-1} \end{cases}\ .$$

Suppose now that model (22) is enlarged to :

$$(24) \qquad \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n_1}{(\boldsymbol{A}_1} \vdots \underset{m\times n_2}{\boldsymbol{A}_2)} \underset{(n_1+n_2)\times 1}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} \; ; \; \underset{m\times m}{\boldsymbol{D}\{\underline{y}\}} = \underset{m\times m}{\boldsymbol{Q}_y} \; .$$

Then according to (17):

$$(25) \qquad \begin{cases} \hat{\underline{x}}_1 = (\overline{\boldsymbol{A}}_1^* \boldsymbol{Q}_y^{-1} \overline{\boldsymbol{A}}_1)^{-1} \; \overline{\boldsymbol{A}}_1^* \boldsymbol{Q}_y^{-1} \underline{y} \\ \boldsymbol{Q}_{\hat{x}_1} = (\overline{\boldsymbol{A}}_1^* \boldsymbol{Q}_y^{-1} \overline{\boldsymbol{A}}_1)^{-1} \end{cases} .$$

Note the resemblance between (23) and (25). The two estimators $\hat{\underline{x}}_1^{\mathrm{I}}$ and $\hat{\underline{x}}_1$ are identical if and only if $A_1 = \overline{A}_1$, that is if:

$$(26) \qquad \boldsymbol{A}_1 = \boldsymbol{P}_{\boldsymbol{A}_2}^{\perp} \boldsymbol{A}_1 \; .$$

This is the case if the range space of A_1 is *orthogonal* to the range space of A_2, that is if $R(A_1)\perp R(A_2)$ or $A_1^* Q_y^{-1} A_2 = 0$. Note that in this case the partitioned normal matrix of (6) becomes *blockdiagonal*. We may thus conclude that the estimator of x_1 is not affected by an enlargement of the model if and only if $R(A_1)$ is orthogonal to $R(A_2)$. In order to find out what happens when $R(A_1)$ and $R(A_2)$ are not orthogonal to each other, we take instead of the expression $Q_{\hat{x}_1} = (\overline{A}_1^* Q_y^{-1} \overline{A}_1)^{-1}$, the following expression from (21):

$$\boldsymbol{Q}_{\hat{x}_1} = (\boldsymbol{A}_1^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_1)^{-1} + (\boldsymbol{A}_1^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_1)^{-1} \; \boldsymbol{A}_1^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_2} \boldsymbol{A}_2^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_1 \; (\boldsymbol{A}_1^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_1)^{-1} \; .$$

With (23) this can be written as:

$$(27) \qquad \boldsymbol{Q}_{\hat{x}_1} = \boldsymbol{Q}_{\hat{x}_1^{\mathrm{I}}} + \boldsymbol{Q}_{\hat{x}_1^{\mathrm{I}}} \boldsymbol{A}_1^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_2} \boldsymbol{A}_2^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_1 \boldsymbol{Q}_{\hat{x}_1^{\mathrm{I}}} \; .$$

Now suppose we are interested in a particular linear function, $\theta = f^* x_1$, of x_1. With model (22) the estimator of θ reads $\hat{\underline{\theta}}^{\mathrm{I}} = f^* \hat{\underline{x}}_1^{\mathrm{I}}$ and with model (24) we get as estimator of θ, $\hat{\underline{\theta}} = f^* \hat{\underline{x}}_1$. Their variances read:

$$\sigma_{\hat{\theta}^{\mathrm{I}}}^2 = f^* \boldsymbol{Q}_{\hat{x}_1^{\mathrm{I}}} f \text{ and } \sigma_{\hat{\theta}}^2 = f^* \boldsymbol{Q}_{\hat{x}_1} f \; .$$

With (27) this gives:

$$\sigma_{\hat{\theta}}^2 = \sigma_{\hat{\theta}^{\mathrm{I}}}^2 + f^* \boldsymbol{Q}_{\hat{x}_1^{\mathrm{I}}} \boldsymbol{A}_1^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_2} \boldsymbol{A}_2^* \boldsymbol{Q}_y^{-1} \boldsymbol{A}_1 \boldsymbol{Q}_{\hat{x}_1^{\mathrm{I}}} f \; .$$

Since the second term on the right hand side is always non-negative, it follows that:

$$(28) \qquad \sigma_{\hat{\theta}}^2 \geq \sigma_{\hat{\theta}^{\mathrm{I}}}^2 \; .$$

This important result shows that the precision of the estimators generally gets *poorer* if more parameters are added to the linear model.

Let us now investigate the variance matrix of the estimator of the added parameter vector x_2. According to (21), $Q_{\hat{x}_2}$ is the inverse of the matrix $\bar{A}_2^* Q_y^{-1} \bar{A}_2$; see also (19). With $\bar{A}_2 = P_{A_1}^{\perp} A_2$, we may write:

$$\bar{A}_2^* Q_y^{-1} \bar{A}_2 = A_2^* Q_y^{-1} A_2 - A_2^* Q_y^{-1} P_{A_1} A_2 \quad . \tag{29}$$

Since $P_{A_1} A_2 = A_2$ if and only if $R(A_2) \subset R(A_1)$, it follows that $\bar{A}_2^* Q_y^{-1} \bar{A}_2 = 0$ if the column vectors of A_2 are linear dependent of the column vectors of A_1. Thus, $Q_{\hat{x}_2}$ does *not* exist and x_2 is *not* estimable if $R(A_2) \subset R(A_1)$. In this case matrix $(A_1 : A_2)$ also fails to be of full rank. Now assume that x_2 is a scalar. Then $n_2=1$ and A_2 is a vector. With the cosine rule it follows then that:

$$A_2^* Q_y^{-1} P_{A_1} A_2 = \|A_2\| \, \|P_{A_1} A_2\| \cos\alpha \quad , \tag{30}$$

see figure 6.1. But we also have that:

$$A_2^* Q_y^{-1} P_{A_1} A_2 = (P_{A_1} A_2)^* Q_y^{-1} (P_{A_1} A_2) = \|P_{A_1} A_2\|^2 \quad . \tag{31}$$

From (30) and (31) it follows that:

$$A_2^* Q_y^{-1} P_{A_1} A_2 = \|A_2\|^2 \cos^2\alpha = A_2^* Q_y^{-1} A_2 \cos^2\alpha \quad . \tag{32}$$

This together with (29) shows that:

$$\bar{A}_2^* Q_y^{-1} \bar{A}_2 = A_2^* Q_y^{-1} A_2 \sin^2\alpha \quad . \tag{33}$$

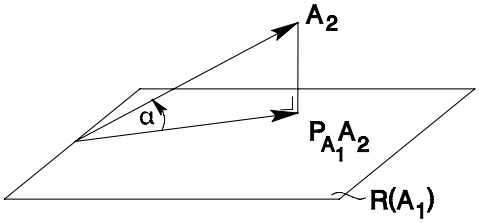


Figure 6.1: $A_2^* Q_y^{-1} P_{A_1} A_2 = \|A_2\| \, \|P_{A_1} A_2\| \cos\alpha$

From this it follows that:

$$\sigma_{\hat{x}_2}^2 = (A_2^* Q_y^{-1} A_2)^{-1} \sin^{-2}\alpha \quad . \tag{34}$$

The angle α measures the *degree of dependence* between A_2 and A_1. If $\alpha = \frac{1}{2}\pi$, then A_2 is orthogonal to $R(A_1)$ and $\sigma_{\hat{x}_2}^2 = (A_2^* Q_y^{-1} A_2)^{-1}$. If $\alpha=0$, then $A_2 \in R(A_1)$ and $\sigma_{\hat{x}_2}^2 = \infty$. If α is small, then $\sigma_{\hat{x}_2}^2$ is large and x_2 is poorly estimable. Since (see(33)):

$$\sin^2\alpha = \frac{\bar{A}_2^* Q_y^{-1} \bar{A}_2}{A_2^* Q_y^{-1} A_2} \quad , \tag{35}$$

the degree of dependence between A_2 and A_1 can be computed in a rather straight forward manner from the ratio of the diagonal elements of the normal matrix *after* and *before* reduction. Compare (19) and (6). The computation of the scalar $\sin^2\alpha$ is usually included in the software in order to diagnose and detect possible singularities or rankdefects in the normal matrix.

So far we have been concerned with the estimators $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_2$. Let us now consider the estimator $\underline{\hat{y}}$. Since $\underline{\hat{y}} = A_1\underline{\hat{x}}_1 + A_2\underline{\hat{x}}_2$, it follows from (17) that:

$$\begin{aligned}
\underline{\hat{y}} &= \boldsymbol{A}_1\underline{\hat{x}}_1 + \boldsymbol{A}_2(\boldsymbol{A}_2^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}_2)^{-1}\,\boldsymbol{A}_2^{*}\boldsymbol{Q}_y^{-1}(\underline{y} - \boldsymbol{A}_1\underline{\hat{x}}_1) \\
&= \boldsymbol{A}_1\underline{\hat{x}}_1 + \boldsymbol{P}_{\boldsymbol{A}_2}(\underline{y} - \boldsymbol{A}_1\underline{\hat{x}}_1) \\
&= \boldsymbol{P}_{\boldsymbol{A}_2}^{+}\boldsymbol{A}_1\underline{\hat{x}}_1 + \boldsymbol{P}_{\boldsymbol{A}_2}\underline{y} \\
&= \bar{\boldsymbol{A}}_1\underline{\hat{x}}_1 + \boldsymbol{P}_{\boldsymbol{A}_2}\underline{y} \\
&= \bar{\boldsymbol{A}}_1\,(\bar{\boldsymbol{A}}_1^{*}\boldsymbol{Q}_y^{-1}\bar{\boldsymbol{A}}_1)^{-1}\,\bar{\boldsymbol{A}}_1^{*}\boldsymbol{Q}_y^{-1}\underline{y} + \boldsymbol{P}_{\boldsymbol{A}_2}\underline{y}
\end{aligned}$$

or:

(36) $$\boxed{\underline{\hat{y}} = \boldsymbol{P}_{\bar{\boldsymbol{A}}_1}\underline{y} + \boldsymbol{P}_{\boldsymbol{A}_2}\underline{y}}$$

Since $\underline{\hat{y}} = P_A\underline{y}$, with $A = (A_1 : A_2)$, we have with (36) the following orthogonal decomposition of the orthogonal projector P_A:

(37) $$\boxed{\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\bar{\boldsymbol{A}}_1} + \boldsymbol{P}_{\boldsymbol{A}_2}}$$

In a completely analogous way it follows from (21) that:

(38) $$\boxed{\underline{\hat{y}} = \boldsymbol{P}_{\boldsymbol{A}_1}\underline{y} + \boldsymbol{P}_{\bar{\boldsymbol{A}}_2}\underline{y}}$$

and:

(39) $$\boxed{\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{A}_1} + \boldsymbol{P}_{\bar{\boldsymbol{A}}_2}}$$

The above orthogonal decomposition is shown in figure 6.2.

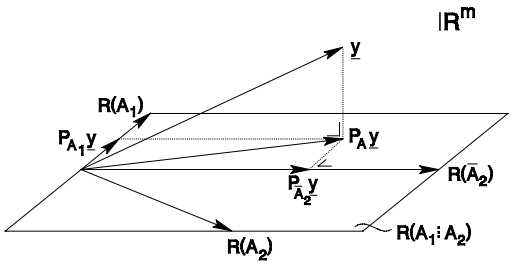


Figure 6.2: $P_A\underline{y} = P_{A_1}\underline{y} + P_{\bar{A}_2}\underline{y}$

Since $\underline{\hat{e}} = (I - P_A)\underline{y} = P_A^{\perp}\underline{y}$, it follows with (39) that:

(40) $$\underline{\hat{e}} = P_A^{\perp}\underline{y} = (I - P_{A_1} - P_{\bar{A}_2})\underline{y} = P_{A_1}^{\perp}\underline{y} - P_{\bar{A}_2}\underline{y} = \underline{\hat{e}}^1 - P_{\bar{A}_2}\underline{y} \ ,$$

where $\underline{\hat{e}}^1$ is the least-squares residual vector of model (22). From (40) it follows that:

(41) $$\underline{\hat{e}}^* Q_y^{-1}\underline{\hat{e}} = \underline{\hat{e}}^{1^*} Q_y^{-1}\underline{\hat{e}}^1 - 2\underline{\hat{e}}_1^* Q_y^{-1} P_{\bar{A}_2}\underline{y} + \underline{y}^* Q_y^{-1} P_{\bar{A}_2}\underline{y} \ .$$

But: $$\underline{\hat{e}}^{1^*} Q_y^{-1} P_{\bar{A}_2}\underline{y} = \underline{y}^* P_{A_1}^{\perp *} Q_y^{-1} P_{\bar{A}_2}\underline{y} = \underline{y}^* Q_y^{-1} P_{A_1}^{\perp} P_{\bar{A}_2}\underline{y} = \underline{y}^* Q_y^{-1} P_{\bar{A}_2}\underline{y} \ .$$

Hence, (41) can be written as:

$$\underline{\hat{e}}^* Q_y^{-1}\underline{\hat{e}} = \underline{\hat{e}}^{1^*} Q_y^{-1}\underline{\hat{e}}^1 - \underline{y}^* Q_y^{-1} P_{\bar{A}_2}\underline{y} \ ,$$

or as:

(42) $$\boxed{\|\underline{\hat{e}}\|^2 = \|\underline{\hat{e}}^1\|^2 - \|P_{\bar{A}_2}\underline{y}\|^2}$$

This important results shows that the length of the residual vector generally gets *smaller* if more paremeters are added to the linear model. In the extreme case that the number of parameters equals the number of observations, that is $m=n_1+n_2$, matrix $A=(A_1 \vdots A_2)$ becomes square and regular, and P_A reduces to the identity matrix I. In this case (see(39)) $I = P_{A_1} + P_{\bar{A}_2}$ or $P_{\bar{A}_2} = P_{A_1}^{\perp}$, and therefore $\underline{\hat{e}}^1 = P_{A_1}^{\perp}\underline{y} = P_{\bar{A}_2}\underline{y}$. With (42) this gives $\|\underline{\hat{e}}\|^2 = 0$. Thus the least-squares residual vector is identically zero if the *redundancy* is zero.

The results of this section will play an important role and are used repeatedly in testing theory.

***Example 1*:**
The model that will be considered first is given as:

$$E\{\underset{m\times 1}{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_m \end{pmatrix}}\} = \underset{m\times 1}{\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}} \underset{1\times 1}{x_1^1} \; ; \; D\{\underline{y}\} = \sigma^2 I_m \; . \tag{43}$$

Since the observation equations are of the form $E\{\underline{y}_i\} = a_i \, x_1^1$, they describe the equation of a straight line through the origin with *slope* x_1^1. This is shown in figure 6.3.

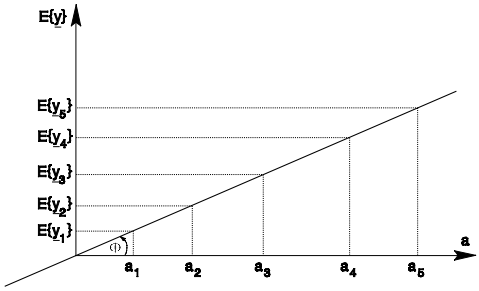


Figure 6.3: The line $E\{\underline{y}\} = a x_1^1$ with slope $x_1^1 = \tan\phi$

The least-squares estimate of x_1^1 follows from the minimization problem:

$$\min_{x_1^1} \; \frac{1}{\sigma^2} \sum_{i=1}^{m} (y_1 - a_i \, x_1^1)^2 \; .$$

Since $|y_i - a_i x_1^1|$ is the *vertical* distance from the point $(a_i, \, y_i)$ to the straight line $E\{\underline{y}\} = a x_1^1$, the least-squares estimate $\hat{x}_1^1$ follows from a minimization of the sum of the squares of these vertical distances. The least-squares estimator $\underline{\hat{x}}_1^1$ reads:

$$\underline{\hat{x}}_1^1 = \sum_{i=1}^{m} a_i \underline{y}_i \; / \; \sum_{i=1}^{m} a_i^2 \; . \tag{44}$$

Its variance is given as:

$$\sigma^2_{\hat{x}_1^1} = (\sum_{i=1}^{m} a_i^2)^{-1} \; \sigma^2 \; . \tag{45}$$

Let us now consider the enlarged linear model:

$$E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_m \end{pmatrix}\} = \begin{pmatrix} a_1 & 1 \\ a_2 & 1 \\ \vdots & \vdots \\ a_m & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \; ; \; D\{\underline{y}\} = \sigma^2 I_m \; . \tag{46}$$

In this case the observation equations $E\{\underline{y}_i\}=a_i x_1+x_2$ describe a straight line with *intercept* x_2 and *slope* x_1. This is shown in figure 6.4.

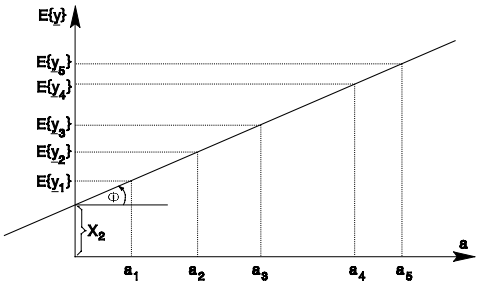


Figure 6.4: The line $E\{\underline{y}\} = a x_1 + x_2$ with slope $x_1 = \tan\phi$ and intercept x_2

With the definitions $A_1 = (a_1, a_2, \ldots, a_m)^*$ and $A_2 = (1 \ldots 1)^*$ it follows that:

$$\begin{cases} P_{A_2}^{\perp} = \begin{pmatrix} 1 & & 0 \\ & 1 & \\ & \ddots & \\ 0 & & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \frac{1}{m} (1 \ldots 1) \; , \\ \\ \bar{A}_1 = P_{A_2}^{\perp} A_1 = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} - \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \frac{1}{m} \sum_{i=1}^{m} a_i \; . \end{cases} \tag{47}$$

In order to simplify the expression for $\bar{A}$ somewhat, we define:

$$\begin{cases} a_c = \frac{1}{m} \sum_{i=1}^{m} a_i & (\text{average value of the } a_i's) \\ \bar{a}_i = a_i - a_c & (\text{centered with respect to } a_c) \end{cases} \tag{48}$$

We now may write $\bar{A}$ as:

$$\bar{A}_1 = (\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_m)^* \; . \tag{49}$$

Using (17) and (49) the least-squares estimator $\underline{\hat{x}}_1$ becomes:

(50) $$\hat{\underline{x}}_1 = \sum_{i=1}^{m} \bar{a}_i \underline{y}_i \; / \; \sum_{i=1}^{m} \bar{a}_i^2 \; .$$

Compare (50) with (44). The variance of $\hat{\underline{x}}_1$ is given as:

(51) $$\sigma^2_{\hat{x}_1} = (\sum_{i=1}^{m} \bar{a}_i^2)^{-1} \; \sigma^2 \; .$$

According to the theory of this section $\sigma^2_{\hat{x}_1} \geq \sigma^2_{\hat{x}_1^1}$, or:

(52) $$(\sum_{i=1}^{m} \bar{a}_i^2)^{-1} \geq (\sum_{i=1}^{m} a_i^2)^{-1}$$

should hold. This is easily verified. From (52) it follows that:

$$\sum_{i=1}^{m} a_i^2 \geq \sum_{i=1}^{m} \bar{a}_i^2$$

or:

$$\sum_{i=1}^{m} (a_i^2 - \bar{a}_i^2) \geq 0$$

or:

$$\sum_{i=1}^{m} (a_i^2 - (a_i - a_c)^2) \geq 0$$

or:

$$\sum_{i=1}^{m} (2a_i a_c - a_c^2) \geq 0$$

or:

$$2ma_c^2 - ma_c^2 \geq 0$$

or:

$$ma_c^2 \geq 0 \; .$$

Hence, inequality (52) holds indeed. Since $A_1 = (a_1, a_2, ..., a_m)^*$ and $A_2 = (1...1)^*$, it follows with (50) from (17) that the least-squares estimator $\hat{\underline{x}}_2$ is given as:

$$\hat{\underline{x}}_2 = \frac{1}{m} \sum_{i=1}^{m} (\underline{y}_i - a_i \hat{\underline{x}}_1) \; ,$$

or as:

(53) $$\hat{\underline{x}}_2 = \frac{1}{m} \sum_{i=1}^{m} \underline{y}_i - a_c \sum_{i=1}^{m} \bar{a}_i \underline{y}_i \; / \; \sum_{i=1}^{m} \bar{a}_i^2 \; .$$

The variance of $\hat{\underline{x}}_2$ follows from (17) as:

$$\sigma^2_{\hat{x}_2} = \frac{\sigma^2}{m} + \frac{1}{m} \sum_{i=1}^{m} a_i \; \sigma^2_{\hat{x}_1} \sum_{i=1}^{m} a_i \frac{1}{m} \; .$$

With (48) and (51) this gives:

$$\sigma_{\hat{x}_2}^2 = \sigma^2 \left(\frac{1}{m} + \frac{a_c^2}{\sum_{i=1}^{m} \bar{a}_i^2} \right) . \tag{54}$$

6.3 The partitioned model $E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix}\} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} x$

Recursive estimation

Consider the partitioned model:

$$E\underset{(m_1+m_2)\times 1}{\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix}\}} = \underset{(m_1+m_2)\times n}{\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}} \underset{n\times 1}{x} \quad ; \; D\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix}\} = \underset{(m_1+m_2)\times(m_1+m_2)}{\begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}} \tag{55}$$

Note that it is assumed that $\underline{y}_1$ and $\underline{y}_2$ are *uncorrelated*. This is usually the case in practical applications. The least-squares estimator of x of model (55) will be denoted as $\underline{\hat{x}}_{(2)}$. It reads:

$$\underline{\hat{x}}_{(2)} = (A_1^* Q_1^{-1} A_1 + A_2^* Q_2^{-1} A_2)^{-1} \; (A_1^* Q_1^{-1} \underline{y}_1 + A_2^* Q_2^{-1} \underline{y}_2) \; . \tag{56}$$

Let us now consider the partial model:

$$\underset{m_1\times 1}{E\{\underline{y}_1\}} = \underset{m_1\times n}{A_1} \underset{n\times 1}{x} \; ; \; \underset{m_1\times m_1}{D\{\underline{y}_1\}} = \underset{m_1\times m_1}{Q_1} \tag{57}$$

Its solution will be denoted as $\underline{\hat{x}}_{(1)}$. It reads:

$$\begin{cases} \underline{\hat{x}}_{(1)} = (A_1^* Q_1^{-1} A_1)^{-1} A_1^* Q_1^{-1} y_1, \\ Q_{\hat{x}_{(1)}} = (A_1^* Q_1^{-1} A_1)^{-1}. \end{cases} \tag{58}$$

Using this result, we may write (56) as:

$$\underline{\hat{x}}_{(2)} = (Q_{\hat{x}_{(1)}}^{-1} + A_2^* Q_2^{-1} A_2)^{-1} \; (Q_{\hat{x}_{(1)}}^{-1} \underline{\hat{x}}_{(1)} + A_2^* Q_2^{-1} \underline{y}_2) \; . \tag{59}$$

But this is also the solution of:

$$E\{\begin{pmatrix} \underline{\hat{x}}_{(1)} \\ \underline{y}_2 \end{pmatrix}\} = \begin{pmatrix} I \\ A_2 \end{pmatrix} x \; ; \; D\{\begin{pmatrix} \underline{\hat{x}}_{(1)} \\ \underline{y}_2 \end{pmatrix}\} = \begin{pmatrix} Q_{\hat{x}_{(1)}} & 0 \\ 0 & Q_2 \end{pmatrix} . \tag{60}$$

Hence we have proven that the solution of the partitioned model (55) can be found in two steps: First one solves for the partial model (57). This gives $\underline{\hat{x}}_{(1)}$ and $Q_{\hat{x}_{(1)}}$. Then in the second step one uses this result together with $\underline{y}_2$ to find $\underline{\hat{x}}_{(2)}$ via model (60). This two step procedure

has an important practical implication. It implies that if new observables, say $\underline{y}_2$, become available one does not need to save the old observables $\underline{y}_1$ to compute $\underline{\hat{x}}_{(2)}$. One can compute $\underline{\hat{x}}_{(2)}$ from the solution $\underline{\hat{x}}_{(1)}$ of the first step and the new observables $\underline{y}_2$. In this way one can recursively determine $\underline{\hat{x}}_{(k)}$ from $\underline{\hat{x}}_{(k-1)}$ and $\underline{y}_k$. This is shown in table 6.1. Expression (59) for $\underline{\hat{x}}_{(2)}$ shows that a matrix of the order n needs to be inverted. One can however also derive an expression for $\underline{\hat{x}}_{(2)}$ in which a matrix of the order m_2 needs to be inverted. This expression is found if we solve (60) via the model of condition equations. Model (60) in terms of condition equations reads:

$$(-\boldsymbol{A}_2 \;\; \boldsymbol{I})\, E\{\begin{pmatrix} \underline{\hat{x}}_{(1)} \\ \underline{y}_2 \end{pmatrix}\} = \boldsymbol{0} \;;\; \boldsymbol{D}\{\begin{pmatrix} \underline{\hat{x}}_{(1)} \\ \underline{y}_2 \end{pmatrix}\} = \begin{pmatrix} \boldsymbol{Q}_{\hat{x}_{(1)}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{pmatrix} \tag{61}$$

Its solution reads:

$$\begin{pmatrix} \underline{\hat{x}}_{(2)} \\ \underline{\hat{y}}_2 \end{pmatrix} = \left[\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} -\boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* \\ \boldsymbol{Q}_2 \end{pmatrix} (\boldsymbol{Q}_2 + \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^*)^{-1} (-\boldsymbol{A}_2 \;\; \boldsymbol{I}) \right] \begin{pmatrix} \underline{\hat{x}}_{(1)} \\ \underline{y}_2 \end{pmatrix} .$$

Hence for $\underline{\hat{x}}_{(2)}$ we get:

$$\underline{\hat{x}}_{(2)} = \underline{\hat{x}}_{(1)} + \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* (\boldsymbol{Q}_2 + \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^*)^{-1} (\underline{y}_2 - \boldsymbol{A}_2 \underline{\hat{x}}_{(1)}) \; . \tag{62}$$

Application of the propagation law of variances gives:

$$\boldsymbol{Q}_{\hat{x}_{(2)}} = \boldsymbol{Q}_{\hat{x}_{(1)}} - \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* (\boldsymbol{Q}_2 + \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^*)^{-1} \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_{(1)}} \; . \tag{63}$$

Both the expressions (59) and (62) give identical results. But expression (62) is more advantageous if m_2 is small compared to n. In particular if m_2=1, only a scalar needs to be inverted in (62), whereas in (59) all three matrices $Q_{\hat{x}_{(1)}}$, Q_2 and $(Q_{\hat{x}_{(1)}}^{-1} + A_2^* Q_2^{-1} A_2)$ need to be inverted.

The equations (59) and (62) are called *measurement update equations*. Expression (62) clearly shows how $\underline{\hat{x}}_{(2)}$ is found from updating $\underline{\hat{x}}_{(1)}$. The correction to $\underline{\hat{x}}_{(1)}$ depends on the difference $\underline{y}_2 - A_2 \underline{\hat{x}}_{(1)}$. Since $A_2 \underline{\hat{x}}_{(1)}$ can be interpreted as the prediction of $E\{\underline{y}_2\}$ based on $\underline{y}_1$, the difference $\underline{y}_2 - A_2 \underline{\hat{x}}_{(1)}$ is called the *predicted residual* of $E\{\underline{y}_2\}$. Note that the predicted residual is *not* equal to the least-squares residual of $E\{\underline{y}_2\}$. This least-squares residual reads namely $\underline{y}_2 - \underline{\hat{y}}_2 = \underline{y}_2 - A_2 \underline{\hat{x}}_{(2)}$.

Expression (62) shows that the correction to $\underline{\hat{x}}_{(1)}$ is small if the predicted residual is small, and that the correction to $\underline{\hat{x}}_{(1)}$ is large if also the predicted residual is large. This is also what one would expect. Also note that the correction to $\underline{\hat{x}}_{(1)}$ is small if the variance of $\underline{\hat{x}}_{(1)}$ is small. This is also understandable, because if the variance of $\underline{\hat{x}}_{(1)}$ is small one has more confidence in $\underline{\hat{x}}_{(1)}$ and therefore would like to give more weight to $\underline{\hat{x}}_{(1)}$ than to $\underline{y}_2$.

Partitioned model

$$E\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \vdots\\ \underline{y}_k\end{pmatrix}\right\} = \begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\\ \vdots\\ \boldsymbol{A}_k\end{pmatrix} x \;;\; \boldsymbol{D}\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \vdots\\ \underline{y}_k\end{pmatrix}\right\} = \begin{pmatrix}\boldsymbol{Q}_1 & & 0\\ & \boldsymbol{Q}_2 & \\ 0 & & \ddots\\ & & & \boldsymbol{Q}_k\end{pmatrix}$$

$$\underline{\hat{x}}_{(k)} = \left(\sum_{i=1}^{k} \boldsymbol{A}_i^{*} \boldsymbol{Q}_i^{-1} \boldsymbol{A}_i\right)^{-1} \left(\sum_{i=1}^{k} \boldsymbol{A}_i^{*} \boldsymbol{Q}_i^{-1} \underline{y}_i\right)$$

$$\boldsymbol{Q}_{\hat{x}_{(k)}} = \left(\sum_{i=1}^{k} \boldsymbol{A}_i^{*} \boldsymbol{Q}_i^{-1} \boldsymbol{A}_i\right)^{-1}$$

Recursive estimation

1° $E\{\underline{y}_1\} = \boldsymbol{A}_1 x \;;\; \boldsymbol{D}\{\underline{y}_1\} = \boldsymbol{Q}_1$

$$\underline{\hat{x}}_{(1)} = (\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1\boldsymbol{Q}_1^{-1}\underline{y}_1$$

$$\boldsymbol{Q}_{\hat{x}_{(1)}} = (\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_1)^{-1}$$

$k°$ $E\left\{\begin{pmatrix}\underline{\hat{x}}_{(k-1)}\\ \underline{y}_k\end{pmatrix}\right\} = \begin{pmatrix}I\\ \boldsymbol{A}_k\end{pmatrix} x \;;\; \boldsymbol{D}\left\{\begin{pmatrix}\underline{\hat{x}}_{(k-1)}\\ \underline{y}_k\end{pmatrix}\right\} = \begin{pmatrix}\boldsymbol{Q}_{\hat{x}_{(k-1)}} & 0\\ 0 & \boldsymbol{Q}_k\end{pmatrix}$

$$\underline{\hat{x}}_{(k)} = (\boldsymbol{Q}^{-1}_{\hat{x}_{(k-1)}} + \boldsymbol{A}_k^{*}\boldsymbol{Q}_k^{-1}\boldsymbol{A}_k)^{-1} (\boldsymbol{Q}^{-1}_{\hat{x}_{(k-1)}}\underline{\hat{x}}_{(k-1)} + \boldsymbol{A}_k^{*}\boldsymbol{Q}_k^{-1}\underline{y}_k)$$

$$\boldsymbol{Q}_{\hat{x}_{(k)}} = (\boldsymbol{Q}^{-1}_{\hat{x}_{(k-1)}} + \boldsymbol{A}_k^{*}\boldsymbol{Q}_k^{-1}\boldsymbol{A}_k)^{-1}$$

$k = 2,3,....$

Table 6.1: Recursive estimation (A-form)

Expression (63) shows how the variance of the estimator gets updated. Because of the minus sign in (63) the precision of the estimator gets better. This is understandable, because by including the additional observable $\underline{y}_2$ more information is available to estimate x. Table 6.2 shows the recursive estimation scheme based on (62) and (63).

Apart from the recursive scheme for the estimator of x, it is also possible to derive a recursive scheme for $\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}}$. Consider again the partitioned model:

$$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{pmatrix} x \;;\; \boldsymbol{D}\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{Q}_1 & 0\\ 0 & \boldsymbol{Q}_2\end{pmatrix} .$$

Then:

$$(\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}})_{(2)} = (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(2)})^* \boldsymbol{Q}_1^{-1} (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(2)}) + (\underline{y}_2 - \boldsymbol{A}_2 \underline{\hat{x}}_{(2)})^* \boldsymbol{Q}_2^{-1} (\underline{y}_2 - \boldsymbol{A}_2 \underline{\hat{x}}_{(2)}) \quad . \tag{64}$$

If we use the abbreviation:

$$\underline{v}_{(2)} = \underline{y}_2 - \boldsymbol{A}_2 \underline{\hat{x}}_{(1)} \;;\; \boldsymbol{Q}_{v_{(2)}} = \boldsymbol{Q}_2 + \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* \quad , \tag{65}$$

we may write the measurements update equation (62) as:

$$\underline{\hat{x}}_{(2)} = \underline{\hat{x}}_{(1)} + \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* \boldsymbol{Q}_{v_{(2)}}^{-1} \underline{v}_{(2)} \quad . \tag{66}$$

Substitution of (66) into the first term on the right-hand side of (64) gives:

(67)

$$(\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(2)})^* \boldsymbol{Q}_1^{-1} (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(2)}) = \left[(\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(1)}) - \boldsymbol{A}_1 \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* \boldsymbol{Q}_{v_{(2)}}^{-1} \underline{v}_{(2)}\right]^* \boldsymbol{Q}_1^{-1} \left[(\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(1)}) - \boldsymbol{A}_1 \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* \boldsymbol{Q}_{v_{(2)}}^{-1} \underline{v}_{(2)}\right]$$

Since $A_1^* Q_1^{-1} (y_1 - A_1 \underline{\hat{x}}_{(1)}) = 0$ and $Q_{\hat{x}_{(1)}}^{-1} = A_1^* Q_1^{-1} A_1$, equation (67) can be written as:

$$(\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(2)})^* \boldsymbol{Q}_1^{-1} (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(2)}) = (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(1)})^* \boldsymbol{Q}_1^{-1} (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(1)}) + \underline{v}_{(2)}^* \boldsymbol{Q}_{v_{(2)}}^{-1} \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* \boldsymbol{Q}_{v_{(2)}}^{-1} \underline{v}_{(2)} \tag{68}$$

Since $\underline{y}_1 - A_1 \underline{\hat{x}}_{(1)}$ is the least-squares residual vector of the partial model:

$$E\{\underline{y}_1\} = \boldsymbol{A}_1 x \;;\; \boldsymbol{D}\{\underline{y}_1\} = \boldsymbol{Q}_1 \quad , \tag{69}$$

we have:

$$(\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}})_{(1)} = (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(1)})^* \boldsymbol{Q}_1^{-1} (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(1)}) \quad . \tag{70}$$

Substitution of (70) into (68) gives:

$$(\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(2)})^* \boldsymbol{Q}_1^{-1} (\underline{y}_1 - \boldsymbol{A}_1 \underline{\hat{x}}_{(2)}) = (\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}})_{(1)} + \underline{v}_{(2)}^* \boldsymbol{Q}_{v_{(2)}}^{-1} \boldsymbol{A}_2 \boldsymbol{Q}_{\hat{x}_{(1)}} \boldsymbol{A}_2^* \boldsymbol{Q}_{v_{(2)}}^{-1} \underline{v}_{(2)} \tag{71}$$

Partitioned model

$$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \vdots\\ \underline{y}_k\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\\ \vdots\\ \boldsymbol{A}_k\end{pmatrix} x \;;\; \boldsymbol{D}\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \vdots\\ \underline{y}_k\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{Q}_1 & & 0\\ & \boldsymbol{Q}_2 & \\ & & \ddots\\ 0 & & \boldsymbol{Q}_k\end{pmatrix}$$

$$\underline{\hat{x}}_{(k)} = (\sum_{i=1}^{k} \boldsymbol{A}_i^{*} \boldsymbol{Q}_i^{-1} \boldsymbol{A}_i)^{-1} (\sum_{i=1}^{k} \boldsymbol{A}_i^{*} \boldsymbol{Q}_i^{-1} \underline{y}_i)$$

$$\boldsymbol{Q}_{\hat{x}_{(k)}} = (\sum_{i=1}^{k} \boldsymbol{A}_i^{*} \boldsymbol{Q}_i^{-1} \boldsymbol{A}_i)^{-1}$$

Recursive estimation

$1^{o}.\quad E\{\underline{y}_1\} = \boldsymbol{A}_1 x \;;\; \boldsymbol{D}\{\underline{y}_1\} = \boldsymbol{Q}_1$

$$\underline{\hat{x}}_{(1)} = (\boldsymbol{A}_1^{*} \boldsymbol{Q}_1^{-1} \boldsymbol{A}_1)^{-1} \boldsymbol{A}_1^{*} \boldsymbol{Q}_1^{-1} \underline{y}_1$$

$$\boldsymbol{Q}_{\hat{x}_{(1)}} = (\boldsymbol{A}_1^{*} \boldsymbol{Q}_1^{-1} \boldsymbol{A}_1)^{-1}$$

$k^{o}.\quad (-\boldsymbol{A}_k \; I)\, E\{\begin{pmatrix}\underline{\hat{x}}_{(k-1)}\\ \underline{y}_k\end{pmatrix}\} = 0 \;;\; \boldsymbol{D}\{\begin{pmatrix}\underline{\hat{x}}_{(k-1)}\\ \underline{y}_k\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{Q}_{\hat{x}_{(k-1)}} & 0\\ 0 & \boldsymbol{Q}_k\end{pmatrix}$

$$\underline{\hat{x}}_{(k)} = \underline{\hat{x}}_{(k-1)} + \boldsymbol{Q}_{\hat{x}_{(k-1)}} \boldsymbol{A}_k^{*} (\boldsymbol{Q}_k + \boldsymbol{A}_k \boldsymbol{Q}_{\hat{x}_{(k-1)}} \boldsymbol{A}_k^{*})^{-1} (\underline{y}_k - \boldsymbol{A}_k \underline{\hat{x}}_{(k-1)})$$

$$\boldsymbol{Q}_{\hat{x}_{(k)}} = \boldsymbol{Q}_{\hat{x}_{(k-1)}} - \boldsymbol{Q}_{\hat{x}_{(k-1)}} \boldsymbol{A}_k^{*} (\boldsymbol{Q}_k + \boldsymbol{A}_k \boldsymbol{Q}_{\hat{x}_{(k-1)}} \boldsymbol{A}_k^{*})^{-1} \boldsymbol{A}_k \boldsymbol{Q}_{\hat{x}_{(k-1)}}$$

$$k = 2,3,....$$

Table 6.2: Recursive estimation (B-form)

Substitution of (66) into the second term on the right-hand side of (64) gives together with (65):

$$(72)\qquad \begin{aligned}(\underline{y}_2-\boldsymbol{A}_2\underline{\hat{x}}_{(2)})^*\boldsymbol{Q}_2^{-1}(\underline{y}_2-\boldsymbol{A}_2\underline{\hat{x}}_{(2)}) &= \underline{\hat{v}}_{(2)}^*\left[\boldsymbol{I}-\boldsymbol{A}_2\boldsymbol{Q}_{\hat{x}_{(1)}}\boldsymbol{A}_2^*\boldsymbol{Q}_{v_{(2)}}^{-1}\right]^*\boldsymbol{Q}_2^{-1}\left[\boldsymbol{I}-\boldsymbol{A}_2\boldsymbol{Q}_{\hat{x}_{(1)}}\boldsymbol{A}_2^*\boldsymbol{Q}_{v_{(2)}}^{-1}\right]\underline{v}_{(2)}\\ &= \underline{v}_{(2)}^*\boldsymbol{Q}_{v_{(2)}}\boldsymbol{Q}_2\boldsymbol{Q}_{v_{(2)}}\underline{v}_{(2)}\end{aligned}$$

Substitution of (71) and (72) into (64) shows finally that:

$$(73)\qquad \boxed{(\underline{\hat{e}}^*\boldsymbol{Q}_y^{-1}\underline{\hat{e}})_{(2)}=(\underline{\hat{e}}^*\boldsymbol{Q}_y^{-1}\underline{\hat{e}})_{(1)}+\underline{v}_{(2)}^*\boldsymbol{Q}_{v_{(2)}}^{-1}\underline{v}_{(2)}.}$$

This equation shows how the norm of the least-squares residual vector should be updated when the observable $\underline{y}_2$ becomes available. The update depends on the predicted residual. Expression (73) can be generalized to more steps. This is shown in table 6.3.

Partitioned model

$$E\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \vdots\\ \underline{y}_k\end{pmatrix}\right\}=\begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\\ \vdots\\ \boldsymbol{A}_k\end{pmatrix}x\ ;\ \boldsymbol{D}\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \vdots\\ \underline{y}_k\end{pmatrix}\right\}=\begin{pmatrix}\boldsymbol{Q}_1 & & & 0\\ & \boldsymbol{Q}_2 & & \\ & & \ddots & \\ 0 & & & \boldsymbol{Q}_k\end{pmatrix}$$

$$(\underline{\hat{e}}^*\boldsymbol{Q}_y^{-1}\underline{\hat{e}})_{(k)}=\sum_{i=1}^{k}(\underline{y}_i-\boldsymbol{A}_i\underline{\hat{x}}_{(k)})^*\boldsymbol{Q}_i^{-1}(\underline{y}_i-\boldsymbol{A}_i\underline{\hat{x}}_{(k)})$$

Recursive update

$$(\underline{\hat{e}}^*\boldsymbol{Q}_y^{-1}\underline{\hat{e}})_{(k)}=(\underline{\hat{e}}^*\boldsymbol{Q}_y^{-1}\underline{\hat{e}})_{(k-1)}+\underline{v}_{(k)}^*\boldsymbol{Q}_{v_{(k)}}^{-1}\underline{v}_{(k)}$$
$$\underline{v}_{(k)}=\underline{y}_k-\boldsymbol{A}_k\underline{\hat{x}}_{(k-1)};\boldsymbol{Q}_{v_{(k)}}=(\boldsymbol{Q}_k+\boldsymbol{A}_k\boldsymbol{Q}_{\underline{\hat{x}}_{(k-1)}}\boldsymbol{A}_k^*)$$

Table 6.3: Recursive update of $(\underline{\hat{e}}^*\boldsymbol{Q}_y^{-1}\underline{\hat{e}})_{(k)}$

So far it was assumed that the observables $\underline{y}_1$ and $\underline{y}_2$ are uncorrelated. Let us now assume that they are *correlated*. Hence, we consider the partitioned model:

$$E\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\right\}=\begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{pmatrix}x\ ;\ \boldsymbol{D}\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\right\}=\begin{pmatrix}Q_1 & Q_{12}\\ Q_{21} & Q_2\end{pmatrix}, \tag{74}$$

With $Q_{12}\neq 0$, $Q_{21}\neq 0$. In this case it is *not* possible to apply the above derived recursive schemes directly. In order to be able to apply the above derived recursive schemes, the observables $\underline{y}_1$ and $\underline{y}_2$ first need to be *de-correlated*. This is possible if we transform model (74) such that the transformed observables are uncorrelated. For instance, if we define new observables as:

$$\begin{pmatrix}\underline{y}_1\\ \underline{\bar{y}}_2\end{pmatrix}=\begin{pmatrix}I & 0\\ -Q_{21}Q_1^{-1} & I\end{pmatrix}\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}, \tag{75}$$

then it follows with (74) that:

$$E\left\{\begin{pmatrix}\underline{y}_1\\ \underline{\bar{y}}_2\end{pmatrix}\right\}=\begin{pmatrix}\boldsymbol{A}_1\\ \bar{\boldsymbol{A}}_2\end{pmatrix}x\ ;\ \boldsymbol{D}\left\{\begin{pmatrix}\underline{y}_1\\ \underline{\bar{y}}_2\end{pmatrix}\right\}=\begin{pmatrix}Q_1 & 0\\ 0 & \bar{Q}_2\end{pmatrix}, \tag{76}$$

with:

$$\bar{\boldsymbol{A}}_2=\boldsymbol{A}_2-Q_{21}Q_1^{-1}\boldsymbol{A}_1\ ;\ \bar{Q}_2=Q_2-Q_{21}Q_1^{-1}Q_{12}\ . \tag{77}$$

Similarly, if we define new observables as:

$$\begin{pmatrix}\underline{\bar{y}}_1\\ \underline{y}_2\end{pmatrix}=\begin{pmatrix}I & -Q_{12}Q_2^{-1}\\ 0 & I\end{pmatrix}\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}, \tag{78}$$

we get with (74) that:

$$E\left\{\begin{pmatrix}\underline{\bar{y}}_1\\ \underline{y}_2\end{pmatrix}\right\}=\begin{pmatrix}\bar{\boldsymbol{A}}_1\\ \boldsymbol{A}_2\end{pmatrix}x\ ;\ \boldsymbol{D}\left\{\begin{pmatrix}\underline{\bar{y}}_1\\ \underline{y}_2\end{pmatrix}\right\}=\begin{pmatrix}\bar{Q}_1 & 0\\ 0 & Q_2\end{pmatrix}, \tag{79}$$

with:

$$\bar{\boldsymbol{A}}_1=\boldsymbol{A}_1-Q_{12}Q_2^{-1}\boldsymbol{A}_2\ ;\ \bar{Q}_1=Q_1-Q_{12}Q_2^{-1}Q_{21} \tag{80}$$

With (76) or (79) it is now possible again to apply the recursive schemes. In this case however the recursive schemes *lose* their practical implication. The transformed observable $\underline{\bar{y}}_2$ of (76) reads namely:

$$\underline{\bar{y}}_2=\underline{y}_2-Q_{21}Q_1^{-1}\underline{y}_1\ .$$

This shows that one still needs to save the old observable $\underline{y}_1$ for computing $\underline{\hat{x}}_{(2)}$ from $\underline{\hat{x}}_{(1)}$ and $\underline{y}_2$. The conclusion reads therefore that the recursive schemes of this section are only of

practical value if the observables are uncorrelated. Fortunately this is the case in most practical applications.

The theory as developed in this section plays an important role in the estimation theory of dynamic systems and in the theory of Kalman filtering.

Example 2

Consider the linear model:

$$E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_k \end{pmatrix}\} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} x \;;\; D\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_k \end{pmatrix}\} = \sigma^2 I_k \;. \tag{81}$$

Its solution reads:

$$\begin{cases} \underline{\hat{x}}_{(k)} = \sum_{i=1}^{k} a_i \underline{y}_i / \sum_{i=1}^{k} a_i^2 \;, \\ \sigma^2_{\hat{x}_{(k)}} = \sigma^2 (\sum_{i=1}^{k} a_i^2)^{-1}. \end{cases} \tag{82}$$

The solution of the partial model:

$$E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_{k-1} \end{pmatrix}\} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{k-1} \end{pmatrix} x \;;\; D\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_{k-1} \end{pmatrix}\} = \sigma^2 I_{k-1} \;, \tag{83}$$

reads:

$$\begin{cases} \underline{\hat{x}}_{(k-1)} = \sum_{i=1}^{k-1} a_i \underline{y}_i / \sum_{i=1}^{k-1} a_i^2 \;, \\ \sigma^2_{\hat{x}_{(k-1)}} = \sigma^2 (\sum_{i=1}^{k-1} a_i^2)^{-1}. \end{cases} \tag{84}$$

According to table 6.2, the solution of (81) can also be written as:

$$\begin{cases} \underline{\hat{x}}_{(k)} = \underline{\hat{x}}_{(k-1)} + \sigma^2_{\hat{x}_{(k-1)}} a_k (\sigma^2 + \sigma^2_{\hat{x}_{(k-1)}} a_k^2)^{-1} (\underline{y}_k - a_k \underline{\hat{x}}_{(k-1)}) \;, \\ \sigma^2_{\hat{x}_{(k)}} = \sigma^2_{\hat{x}_{(k-1)}} - \sigma^2_{\hat{x}_{(k-1)}} a_k (\sigma^2 + \sigma^2_{\hat{x}_{(k-1)}} a_k^2)^{-1} a_k \sigma^2_{\hat{x}_{(k-1).}} \end{cases} \tag{85}$$

We will now show that (85) is indeed identical to (82). We will first consider the first equation of (85). This equation can be written as:

$$\underline{\hat{x}}_{(k)} = \left(1 - \frac{\sigma^2_{\hat{x}_{(k-1)}} a_k^2}{\sigma^2 + \sigma^2_{\hat{x}_{(k-1)}} a_k^2}\right) \underline{\hat{x}}_{(k-1)} + \frac{\sigma^2_{\hat{x}_{(k-1)}} a_k}{\sigma^2 + \sigma^2_{\hat{x}_{(k-1)}} a_k^2} \; y_k$$

or as:

$$\underline{\hat{x}}_{(k)} = \frac{\sigma^2 \underline{\hat{x}}_{(k-1)} + \sigma^2_{\hat{x}_{(k-1)}} a_k \underline{y}_k}{\sigma^2 + \sigma^2_{\hat{x}_{(k-1)}} a_k^2}$$

or with (84) as:

$$\underline{\hat{x}}_{(k)} = \frac{\sigma^2 (\sum_{i=1}^{k-1} a_i \underline{y}_i + a_k \underline{y}_k) / \sum_{i=1}^{k-1} a_i^2}{\sigma^2 (1 + a_k^2 / \sum_{i=1}^{k-1} a_i^2)}$$

or as:

$$\underline{\hat{x}}_{(k)} = \frac{\sigma^2 (\sum_{i=1}^{k-1} a_i \underline{y}_i + a_k \underline{y}_k) / \sum_{i=1}^{k-1} a_i^2}{\sigma^2 (\sum_{i=1}^{k-1} a_i^2 + a_k^2) / \sum_{i=1}^{k-1} a_i^2}$$

which indeed is identical to the first equation of (82). Now consider the second equation of (85). This can be written as:

$$\sigma^2_{\hat{x}_{(k)}} = \left(1 - \frac{\sigma^2_{\hat{x}_{(k-1)}} a_k^2}{\sigma^2 + \sigma^2_{\hat{x}_{(k-1)}} a_k^2}\right) \sigma^2_{\hat{x}_{(k-1)}}$$

or as:

$$\sigma^2_{\hat{x}_{(k)}} = \frac{\sigma^2 \sigma^2_{\hat{x}_{(k-1)}}}{\sigma^2 + \sigma^2_{\hat{x}_{(k-1)}} a_k^2}$$

or with (84) as:

$$\sigma^2_{\hat{x}_{(k)}} = \sigma^2 \frac{\sigma^2 (\sum_{i=1}^{k-1} a_i^2)^{-1}}{\sigma^2 + \sigma^2 (\sum_{i=1}^{k-1} a_i^2)^{-1} a_k^2}$$

or as:

$$\sigma^2_{\hat{x}_{(k)}} = \sigma^2 \frac{1}{\sum_{i=1}^{k-1} a_i^2 + a_k^2} \quad ,$$

which is indeed identical to the second equation of (82).

6.4 The partitioned model $E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\}=\begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{pmatrix}x$

Block estimation I

Consider again the partitioned model:

$$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\}=\begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{pmatrix}x \quad ; \quad \boldsymbol{D}\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\}=\begin{pmatrix}\boldsymbol{Q}_1 & 0\\ 0 & \boldsymbol{Q}_2\end{pmatrix} \tag{86}$$

Its system of normal equations reads:

$$\sum_{i=1}^{2}\boldsymbol{A}_i^{*}\boldsymbol{Q}_i^{-1}\boldsymbol{A}_i\hat{\underline{x}} \;=\; \sum_{i=1}^{2}\boldsymbol{A}_i^{*}\boldsymbol{Q}_i^{-1}\underline{y}_i \;. \tag{87}$$

Now consider the two partial models:

$$\begin{cases} E\{\underline{y}_1\}=\boldsymbol{A}_1x \quad ; \quad \boldsymbol{D}\{\underline{y}_1\}=\boldsymbol{Q}_1\\ E\{\underline{y}_2\}=\boldsymbol{A}_2x \quad ; \quad \boldsymbol{D}\{\underline{y}_2\}=\boldsymbol{Q}_2 \end{cases} \tag{88}$$

Their systems of normal equations read: [4)]

$$\begin{cases} \boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_1\hat{\underline{x}}_{(1)}=\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1\\ \boldsymbol{A}_2^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_2\hat{\underline{x}}_{(2)}=\boldsymbol{A}_2^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2 \end{cases} \tag{89}$$

Comparison of (89) with (87) shows that the normal matrix of model (86) can be found from the *sum* of the two normalmatrices of the two partial models in (88), and that the righthand side of (87) can be found from the *sum* of the two right hand sides of the normal equations in (89).

Let us now generalize model (86) to:

$$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12} & 0\\ 0 & \boldsymbol{A}_{22} & \boldsymbol{A}_{23}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix} \;;\; \boldsymbol{D}\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\}=\begin{pmatrix}\boldsymbol{Q}_1 & 0\\ 0 & \boldsymbol{Q}_2\end{pmatrix} \tag{90}$$

The corresponding partitioned system of normal equations reads:

[4)] Do not confuse $\hat{\underline{x}}_{(2)}$ of (89) with $\hat{\underline{x}}_{(2)}$ of the previous section.

$$(91)\qquad \begin{pmatrix} \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12} & 0 \\ \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \sum_{i=1}^{2}\boldsymbol{A}_{i2}^{*}\boldsymbol{Q}_i^{-1}\boldsymbol{A}_{i2} & \boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23} \\ 0 & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22} & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23} \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1 \\ \sum_{i=1}^{2}\boldsymbol{A}_{i2}^{*}\boldsymbol{Q}_i^{-1}\underline{y}_i \\ \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2 \end{pmatrix}.$$

The estimator $\underline{\hat{x}}_1$ gets eliminated from the second set of normal equations if we premultiply (91) with the square and full rank matrix:

$$(92)\qquad \begin{pmatrix} I & 0 & 0 \\ -\boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11}(\boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11})^{-1} & I & 0 \\ 0 & 0 & I \end{pmatrix}.$$

This gives:

$$(93)\qquad \begin{pmatrix} \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12} & 0 \\ 0 & (\bar{\boldsymbol{A}}_{12}^{*}\boldsymbol{Q}_1^{-1}\bar{\boldsymbol{A}}_{12}+\boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22}) & \boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23} \\ 0 & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22} & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23} \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1 \\ \bar{\boldsymbol{A}}_{12}^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1+\boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2 \\ \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2 \end{pmatrix},$$

where we have used the abbreviation:

$$(94)\qquad \bar{\boldsymbol{A}}_{12}=\boldsymbol{P}_{\boldsymbol{A}_{11}}^{\perp}\boldsymbol{A}_{12}\ . \qquad \text{(see also section 6.2)}$$

The estimator $\underline{\hat{x}}_3$ gets eliminated from the second set of normal equations if we premultiply (93) with the square and full rank matrix:

$$(95)\qquad \begin{pmatrix} I & 0 & 0 \\ 0 & I & -\boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23}(\boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23})^{-1} \\ 0 & 0 & I \end{pmatrix}.$$

This gives:

$$(96)\qquad \begin{pmatrix} \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12} & 0 \\ 0 & \sum_{i=1}^{2}\bar{\boldsymbol{A}}_{i2}^{*}\boldsymbol{Q}_i^{-1}\bar{\boldsymbol{A}}_{i2} & 0 \\ 0 & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22} & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23} \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1 \\ \sum_{i=1}^{2}\bar{\boldsymbol{A}}_{i2}^{*}\boldsymbol{Q}_i^{-1}\underline{y}_i \\ \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2 \end{pmatrix},$$

where we have used the abbreviation:

$$(97)\qquad \bar{\boldsymbol{A}}_{22}=\boldsymbol{P}_{\boldsymbol{A}_{23}}^{\perp}\boldsymbol{A}_{22}\ .$$

The partially reduced system of normal equations (96) can now be solved in the following way. First one uses the second set of normal equations in (96) to solve for $\underline{\hat{x}}_2$. Once $\underline{\hat{x}}_2$ is known, it can be substituted into the first and third set of normal equations in (96) to solve for $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_3$ respectively. Thus the solution of (96) is found as:

(98)
$$\begin{cases} a) & \underline{\hat{x}}_2 = (\sum_{i=1}^{2} \bar{A}_{i2}^{*} Q_i^{-1} \bar{A}_{i2})^{-1} (\sum_{i=1}^{2} \bar{A}_{i2}^{*} Q_i^{-1} \underline{y}_i) \ , \\ b) & \underline{\hat{x}}_1 = (A_{11}^{*} Q_1^{-1} A_{11})^{-1} A_{11}^{*} Q_1^{-1} (\underline{y}_1 - A_{12} \underline{\hat{x}}_2) \ , \\ c) & \underline{\hat{x}}_3 = (A_{23}^{*} Q_2^{-1} A_{23})^{-1} A_{23}^{*} Q_2^{-1} (\underline{y}_2 - A_{22} \underline{\hat{x}}_2) \ . \end{cases}$$

What is the practial implication of the above solution method? Let us assume that two countries, say the Netherlands and Germany, are covered with a large geodetic network. (See figure 6.5.)

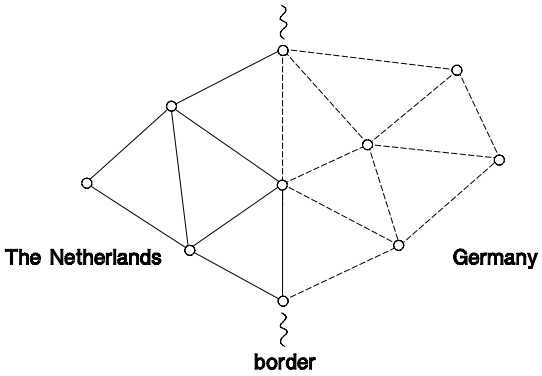


Figure 6.5: Geodetic network covering the Netherlands and Germany

The Dutch part of the network (full lines in figure 6.5) has been measured by the Dutch Triangulation Department (Rijksdriehoeksmeting), and the German part of the network (dashed lines in figure 6.5) has been measured by the German Triangulation Department. We will assume that the vector x_1 contains the unknown coordinates of the network points in the Netherlands, that the vector x_2 contains the unknown coordinates of the network points on the border, and that the vector x_3 contains the unknown coordinates of the network points in Germany.

The partial model of the Dutch part of the network reads then:

(99)
$$E\{\underline{y}_1\} = (A_{11} \ A_{12}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ ; \ D\{\underline{y}_1\} = Q_1 \ .$$

Similarly, the partial model of the German part of the network reads:

(100)
$$E\{\underline{y}_2\} = (A_{22} \ A_{23}) \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \ ; \ D\{\underline{y}_2\} = Q_2 \ .$$

There are now two ways in which a Computing Centre can perform the adjustment of the geodetic network. The *first* method consists of the following: Both the Dutch and the Germans send their partial models, (99) and (100), together with the measured data to the Computing Centre. The Computing Centre merges the two partial models to get (90), the complete model. This model is then solved by the Computing Centre for $\underline{\hat{x}}_1$, $\underline{\hat{x}}_2$ and $\underline{\hat{x}}_3$. The solution $\underline{\hat{x}}_1$, $\underline{\hat{x}}_2$ is then sent by the Computing Centre back to the Dutch, and similarly the solution $\underline{\hat{x}}_2$, $\underline{\hat{x}}_3$ is sent back to the Germans. The *second* method consists of the following: based on their partial model, the Dutch form their partial system of normal equations and reduce it for x_1. This gives:

$$\begin{pmatrix} A_{11}^* Q_1^{-1} A_{11} & A_{11}^* Q_1^{-1} A_{12} \\ 0 & \bar{A}_{12}^* Q_1^{-1} \bar{A}_{12} \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \end{pmatrix} = \begin{pmatrix} A_{11}^* Q_1^{-1} \underline{y}_1 \\ \bar{A}_{12}^* Q_1^{-1} \underline{y}_1 \end{pmatrix} . \tag{101}$$

Similarly, the Germans form, on the basis of their partial model, their partial system of normal equations and reduce it for x_3. This gives:

$$\begin{pmatrix} \bar{A}_{22}^* Q_2^{-1} \bar{A}_{22} & 0 \\ A_{23}^* Q_2^{-1} A_{22} & A_{23}^* Q_2^{-1} A_{23} \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \begin{pmatrix} \bar{A}_{22}^* A_2^{-1} \underline{y} \\ A_{23}^* Q_2^{-1} \underline{y}_2 \end{pmatrix} . \tag{102}$$

The Dutch then sent their reduced normal block $\bar{A}_{12}^* Q_1^{-1} \bar{A}_{12}$ and the reduced right-hand side $\bar{A}_{12}^* Q_1^{-1} \underline{y}_1$ to the Computing Centre. And the Germans do the same for their reduced normal block $A_{22}^* Q_2^{-1} \bar{A}_{22}$ and their reduced right-hand side $\bar{A}_{22}^* Q_2^{-1} \underline{y}_2$. The reduced normal blocks and the reduced right-hand sides are then added by the Computing Centre to form the system:

$$(\sum_{i=1}^{2} \bar{A}_{i2}^* Q_i^{-1} \bar{A}_{i2}) \underline{\hat{x}}_2 = \sum_{i=1}^{2} \bar{A}_{i2}^* Q_i^{-1} \underline{y}_i . \tag{103}$$

This system is solved by the Computing Centre for $\underline{\hat{x}}_2$ (see(98a)). The solution $\underline{\hat{x}}_2$ is then sent back to both the Dutch and the Germans. With $\underline{\hat{x}}_2$ and (101) the Dutch form the system:

$$A_{11}^* Q_1^{-1} A_{11} \underline{\hat{x}}_1 = A_{11}^* Q_1^{-1} (\underline{y}_1 - A_{12} \underline{\hat{x}}_2) . \tag{104}$$

This system is solved by the Dutch for $\underline{\hat{x}}_1$ (see (98b)). In a similar way the Germans use $\underline{\hat{x}}_2$ and (102) to form the system:

$$A_{23}^* Q_2^{-1} A_{23} \underline{\hat{x}}_3 = A_{23}^* Q_2^{-1} (\underline{y}_2 - A_{22} \underline{\hat{x}}_2) . \tag{105}$$

This system is solved by the Germans for $\underline{\hat{x}}_3$ (see (98c)).

When we compare the above two methods we note that:

1°. The computational load for the Computing Centre is less with the second method than with the first method. With the second method some of the computational load stays namely at the two national Triangulation Departments. The order of the system to be solved by the Computing Centre is then equal to the dimension of x_2 (see (103)).

2°. With the second method the Computing Centre never gets to know the original measured data, since $\bar{A}_{i2}^{*}Q_i^{-1}\underline{y}_i$, i=1,2, is provided and not $\underline{y}_i$, i=1,2. Hence, the Computing Centre will never be able to compute the complete network. This may be an advantage in case secrecy of data and coordinates is asked for.

The above developed second method was orginally introduced by the famous German geodesist F.R. Helmert (1843-1917). The method is known as *Helmert's Block method*. The method has been used with great success for the readjustment of the European triangulation network (RETrig: Réseau Européen Trigonometrique).

6.5 The partitioned model $E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\}=\begin{pmatrix}A_1\\ A_2\end{pmatrix}x$

Block estimation II

Consider again the partioned model:

$$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\}=\begin{pmatrix}A_1\\ A_2\end{pmatrix}x \quad ; \quad D\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\}=\begin{pmatrix}Q_1 & 0\\ 0 & Q_2\end{pmatrix} . \tag{106}$$

Its solution reads:

$$\underline{\hat{x}} = (A_1^{*}Q_1^{-1}A_1+A_2^{*}Q_2^{-1}A_2)^{-1}(A_1^{*}Q_1^{-1}\underline{y}_1+A_2^{*}Q_2^{-1}\underline{y}_2) . \tag{107}$$

Now consider the two partial models:

$$\begin{cases} E\{\underline{y}_1\} = A_1x \quad ; \quad D\{\underline{y}_1\}=Q_1 , \\ E\{\underline{y}_2\} = A_2x \quad ; \quad D\{\underline{y}_2\}=Q_2 . \end{cases} \tag{108}$$

Their solutions read:

$$\begin{cases} \underline{\hat{x}}^{(1)} = (A_1^{*}Q_1^{-1}A_1)^{-1}A_1^{*}Q_1^{-1}\underline{y}_1 \quad ; \quad Q_{\hat{x}^{(1)}}=(A_1^{*}Q_1^{-1}A_1)^{-1} , \\ \underline{\hat{x}}^{(2)} = (A_2^{*}Q_2^{-1}A_2)^{-1}A_2^{*}Q_2^{-1}\underline{y}_2 \quad ; \quad Q_{\hat{x}^{(2)}}=(A_2^{*}Q_2^{-1}A_2)^{-1} . \end{cases} \tag{109}$$

Using this results we may write (107) also as:

$$\underline{\hat{x}} = (Q_{\hat{x}^{(1)}}^{-1}+Q_{\hat{x}^{(2)}}^{-1})^{-1}(Q_{\hat{x}^{(1)}}^{-1}\underline{\hat{x}}^{(1)}+Q_{\hat{x}^{(2)}}^{-1}\underline{\hat{x}}^{(2)}) . \tag{110}$$

But this is also the solution of:

$$E\{\begin{pmatrix}\underline{\hat{x}}^{(1)}\\ \underline{\hat{x}}^{(2)}\end{pmatrix}\} = \begin{pmatrix}I\\ I\end{pmatrix}x \quad ; \quad D\{\begin{pmatrix}\underline{\hat{x}}^{(1)}\\ \underline{\hat{x}}^{(2)}\end{pmatrix}\} = \begin{pmatrix}Q_{\hat{x}^{(1)}} & 0\\ 0 & Q_{\hat{x}^{(2)}}\end{pmatrix} . \tag{111}$$

This model is in terms of observation equations. Its equivalent form in terms of condition equations reads:

$$(112) \qquad (I \; -I)\, E\{\begin{pmatrix} \underline{\hat{x}}^{(1)} \\ \underline{\hat{x}}^{(2)} \end{pmatrix}\} \;=\; 0 \;\;;\;\; D\{\begin{pmatrix} \underline{\hat{x}}^{(1)} \\ \underline{\hat{x}}^{(2)} \end{pmatrix}\} \;=\; \begin{pmatrix} Q_{\hat{x}^{(1)}} & 0 \\ 0 & Q_{\hat{x}^{(2)}} \end{pmatrix}$$

And its solution reads:

$$(113) \qquad \begin{pmatrix} \underline{\hat{x}} \\ \underline{\hat{x}} \end{pmatrix} = \left[\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} - \begin{pmatrix} Q_{\hat{x}^{(1)}} \\ -Q_{\hat{x}^{(2)}} \end{pmatrix} (Q_{\hat{x}^{(1)}} + Q_{\hat{x}^{(2)}})^{-1} (I \; -I) \right] \begin{pmatrix} \underline{\hat{x}}^{(1)} \\ \underline{\hat{x}}^{(2)} \end{pmatrix} .$$

Hence we have the following two equivalent expressions for the solution of (112):

$$(114) \qquad \begin{cases} a) \quad \underline{\hat{x}} \;=\; \underline{\hat{x}}^{(1)} - Q_{\hat{x}^{(1)}} (Q_{\hat{x}^{(1)}} + Q_{\hat{x}^{(2)}})^{-1} (\underline{\hat{x}}^{(1)} - \underline{\hat{x}}^{(2)}) , \\ b) \quad Q_{\hat{x}} \;=\; Q_{\hat{x}^{(1)}} - Q_{\hat{x}^{(1)}} (Q_{\hat{x}^{(1)}} + Q_{\hat{x}^{(2)}})^{-1} Q_{\hat{x}^{(1)}} \end{cases}$$

and:

$$(115) \qquad \begin{cases} a) \quad \underline{\hat{x}} \;=\; \underline{\hat{x}}^{(2)} - Q_{\hat{x}^{(2)}} (Q_{\hat{x}^{(1)}} + Q_{\hat{x}^{(2)}})^{-1} (\underline{\hat{x}}^{(2)} - \underline{\hat{x}}^{(1)}) , \\ b) \quad Q_{\hat{x}} \;=\; Q_{\hat{x}^{(2)}} - Q_{\hat{x}^{(2)}} (Q_{\hat{x}^{(1)}} + Q_{\hat{x}^{(2)}})^{-1} Q_{\hat{x}^{(2)}} . \end{cases}$$

Note that (115) follows from (114) by interchanging the role of $\underline{\hat{x}}^{(1)}$ and $\underline{\hat{x}}^{(2)}$. The estimators of (114) and (115) are identical to (110) and therefore also identical to (107). Hence, we have shown how the solution of the partitioned model (106) can be obtained in three separate steps. This is shown in table 6.4.

Partitioned model

$$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{pmatrix} x \;;\; \boldsymbol{D}\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{Q}_1 & 0\\ 0 & \boldsymbol{Q}_2\end{pmatrix}$$

$$\hat{\underline{x}} = (\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_1 + \boldsymbol{A}_2^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_2)^{-1}(\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1 + \boldsymbol{A}_2^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2)$$

$$\boldsymbol{Q}_{\hat{x}} = (\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_1 + \boldsymbol{A}_2^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_2)^{-1} \;.$$

Block estimation

1°.

$$E\{\underline{y}_1\} = \boldsymbol{A}_1 x \;;\; \boldsymbol{D}\{\underline{y}_1\} = \boldsymbol{Q}_1 \;.$$

$$\hat{\underline{x}}^{(1)} = (\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1 \;;\; \boldsymbol{Q}_{\hat{x}^{(1)}} = (\boldsymbol{A}_1^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_1)^{-1} \;.$$

2°.

$$E\{\underline{y}_2\} = \boldsymbol{A}_2 x;\; \boldsymbol{D}\{\underline{y}_2\} = \boldsymbol{Q}_2$$

$$\hat{\underline{x}}^{(2)} = (\boldsymbol{A}_2^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_2)^{-1}\boldsymbol{A}_2^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2 \;;\; \boldsymbol{Q}_{\hat{x}^{(2)}} = (\boldsymbol{A}_2^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_2)^{-1}$$

3°.

$$(I - I)\,E\{\begin{pmatrix}\hat{\underline{x}}^{(1)}\\ \hat{\underline{x}}^{(2)}\end{pmatrix}\} = 0 \;;\; \boldsymbol{D}\{\begin{pmatrix}\hat{\underline{x}}^{(1)}\\ \hat{\underline{x}}^{(2)}\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{Q}_{\hat{x}^{(1)}} & 0\\ 0 & \boldsymbol{Q}_{\hat{x}^{(2)}}\end{pmatrix}$$

$$\begin{cases} \hat{\underline{x}} &= \hat{\underline{x}}^{(1)} - \boldsymbol{Q}_{\hat{x}^{(1)}}(\boldsymbol{Q}_{\hat{x}^{(1)}} + \boldsymbol{Q}_{\hat{x}^{(2)}})^{-1}(\hat{\underline{x}}^{(1)} - \hat{\underline{x}}^{(2)}) \\ \boldsymbol{Q}_{\hat{x}} &= \boldsymbol{Q}_{\hat{x}^{(1)}} - \boldsymbol{Q}_{\hat{x}^{(1)}}(\boldsymbol{Q}_{\hat{x}^{(1)}} + \boldsymbol{Q}_{\hat{x}^{(2)}})^{-1}\boldsymbol{Q}_{\hat{x}^{(1)}} \end{cases}$$

or

$$\begin{cases} \hat{\underline{x}} &= \hat{\underline{x}}^{(2)} - \boldsymbol{Q}_{\hat{x}^{(2)}}(\boldsymbol{Q}_{\hat{x}^{(1)}} + \boldsymbol{Q}_{\hat{x}^{(2)}})^{-1}(\hat{\underline{x}}^{(2)} - \hat{\underline{x}}^{(1)}) \\ \boldsymbol{Q}_{\hat{x}} &= \boldsymbol{Q}_{\hat{x}^{(2)}} - \boldsymbol{Q}_{\hat{x}^{(2)}}(\boldsymbol{Q}_{\hat{x}^{(1)}} + \boldsymbol{Q}_{\hat{x}^{(2)}})^{-1}\boldsymbol{Q}_{\hat{x}^{(2)}} \end{cases}$$

Table 6.4: Block estimation

We will now generalize the above results. Consider therefore the partitioned model:

$$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12} & 0\\ 0 & \boldsymbol{A}_{22} & \boldsymbol{A}_{23}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix};\ D\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{Q}_1 & 0\\ 0 & \boldsymbol{Q}_2\end{pmatrix}. \tag{116}$$

Its solution reads:

$$\begin{pmatrix}\hat{\underline{x}}_1\\ \hat{\underline{x}}_2\\ \hat{\underline{x}}_3\end{pmatrix} = \begin{pmatrix}\boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12} & 0\\ \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \sum_{i=1}^{2}\boldsymbol{A}_{i2}\boldsymbol{Q}_i^{-1}\boldsymbol{A}_{i2} & \boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23}\\ 0 & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22} & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1\\ \sum_{i=1}^{2}\boldsymbol{A}_{i2}^{*}\boldsymbol{Q}_i^{-1}\underline{y}_i\\ \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2\end{pmatrix}. \tag{117}$$

Now consider the two partial models:

$$\begin{cases} E\{\underline{y}_1\} = (\boldsymbol{A}_{11}\ \boldsymbol{A}_{12})\begin{pmatrix}x_1\\ x_2\end{pmatrix};\ D\{\underline{y}_1\} = \boldsymbol{Q}_1, \\ E\{\underline{y}_2\} = (\boldsymbol{A}_{22}\ \boldsymbol{A}_{23})\begin{pmatrix}x_2\\ x_3\end{pmatrix};\ D\{\underline{y}_2\} = \boldsymbol{Q}_2. \end{cases} \tag{118}$$

Their solutions read:

$$\begin{cases} \begin{pmatrix}\hat{\underline{x}}_1^{(1)}\\ \hat{\underline{x}}_2^{(1)}\end{pmatrix} = \begin{pmatrix}\boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12}\\ \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1\\ \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\underline{y}_1\end{pmatrix};\quad \boldsymbol{Q}_{\hat{x}^{(1)}} = \begin{pmatrix}\boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12}\\ \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12}\end{pmatrix}^{-1}, \\ \begin{pmatrix}\hat{\underline{x}}_2^{(2)}\\ \hat{\underline{x}}_3^{(2)}\end{pmatrix} = \begin{pmatrix}\boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22} & \boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23}\\ \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22} & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2\\ \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\underline{y}_2\end{pmatrix};\quad \boldsymbol{Q}_{\hat{x}^{(2)}} = \begin{pmatrix}\boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22} & \boldsymbol{A}_{22}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23}\\ \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{22} & \boldsymbol{A}_{23}^{*}\boldsymbol{Q}_2^{-1}\boldsymbol{A}_{23}\end{pmatrix}^{-1}. \end{cases} \tag{119}$$

If we use the notation:

$$\boldsymbol{Q}_{\hat{x}^{(1)}}^{-1} = \begin{pmatrix}\boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{11}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12}\\ \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}^{*}\boldsymbol{Q}_1^{-1}\boldsymbol{A}_{12}\end{pmatrix} = \begin{pmatrix}(\boldsymbol{Q}_{\hat{x}^{(1)}}^{-1})_{11} & (\boldsymbol{Q}_{\hat{x}^{(1)}}^{-1})_{12}\\ (\boldsymbol{Q}_{\hat{x}^{(1)}}^{-1})_{21} & (\boldsymbol{Q}_{\hat{x}^{(1)}}^{-1})_{22}\end{pmatrix}$$

and a similar notation for $\boldsymbol{Q}_{\hat{x}^{(2)}}$, it follows with (119) that (117) may also be written as:

$$(120)\qquad \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \begin{pmatrix} (Q_{\hat{x}^{(1)}}^{-1})_{11} & (Q_{\hat{x}^{(1)}}^{-1})_{12} & 0 \\ (Q_{\hat{x}^{(1)}}^{-1})_{21} & \sum_{i=1}^{2}(Q_{\hat{x}^{(\mathbf{i})}}^{-1})_{22} & (Q_{\hat{x}^{(2)}}^{-1})_{23} \\ 0 & (Q_{\hat{x}^{(2)}}^{-1})_{32} & (Q_{\hat{x}^{(2)}}^{-1})_{33} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{2}(Q_{\hat{x}^{(1)}}^{-1})_{1i}\underline{\hat{x}}_i^1 \\ \sum_{i=1}^{2}(Q_{\hat{x}^{(1)}}^{-1})_{2i}\underline{\hat{x}}_i^{(1)} + \sum_{i=2}^{3}(Q_{\hat{x}^{(2)}}^{-1})_{2i}\underline{\hat{x}}_i^{(2)} \\ \sum_{i=2}^{3}(Q_{\hat{x}^{(2)}}^{-1})_{3i}\underline{\hat{x}}_i^{(2)} \end{pmatrix}$$

But this is also the solution of:

$$(121)\quad E\{\begin{pmatrix} \underline{\hat{x}}_1^{(1)} \\ \underline{\hat{x}}_2^{(1)} \\ \underline{\hat{x}}_2^{(2)} \\ \underline{\hat{x}}_3^{(2)} \end{pmatrix}\} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad ; \quad D\{\begin{pmatrix} \underline{\hat{x}}_1^{(1)} \\ \underline{\hat{x}}_2^{(1)} \\ \underline{\hat{x}}_2^{(2)} \\ \underline{\hat{x}}_3^{(2)} \end{pmatrix}\} = \begin{pmatrix} (Q_{\hat{x}^{(1)}})_{11} & (Q_{\hat{x}^{(1)}})_{12} & 0 & 0 \\ (Q_{\hat{x}^{(1)}})_{21} & (Q_{\hat{x}^{(1)}})_{22} & 0 & 0 \\ 0 & 0 & (Q_{\hat{x}^{(2)}})_{22} & (Q_{\hat{x}^{(2)}})_{23} \\ 0 & 0 & (Q_{\hat{x}^{(2)}})_{32} & (Q_{\hat{x}^{(2)}})_{33} \end{pmatrix}.$$

This model is in terms of observation equations. Its equivalent form in terms of condition equations reads:

$$(122)\quad (0 \quad I \quad -I \quad 0)E\{\begin{pmatrix} \underline{\hat{x}}_1^{(1)} \\ \underline{\hat{x}}_2^{(1)} \\ \underline{\hat{x}}_2^{(2)} \\ \underline{\hat{x}}_3^{(2)} \end{pmatrix}\} = 0 \quad ; \quad D\{\begin{pmatrix} \underline{\hat{x}}_1^{(1)} \\ \underline{\hat{x}}_2^{(1)} \\ \underline{\hat{x}}_2^{(2)} \\ \underline{\hat{x}}_3^{(2)} \end{pmatrix}\} = \begin{pmatrix} (Q_{\hat{x}^{(1)}})_{11} & (Q_{\hat{x}^{(1)}})_{12} & 0 & 0 \\ (Q_{\hat{x}^{(1)}})_{21} & (Q_{\hat{x}^{(1)}})_{22} & 0 & 0 \\ 0 & 0 & (Q_{\hat{x}^{(2)}})_{22} & (Q_{\hat{x}^{(2)}})_{23} \\ 0 & 0 & (Q_{\hat{x}^{(2)}})_{32} & (Q_{\hat{x}^{(2)}})_{33} \end{pmatrix}$$

Note that in this model, $\underline{\hat{x}}_1^{(1)}$ and $\underline{\hat{x}}_3^{(2)}$ are free $\underline{y}^R$-variates. The solution of (122) reads:

$$(123)\qquad \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \\ \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \left\{ \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix} - \begin{pmatrix} (Q_{\hat{x}^{(1)}})_{12} \\ (Q_{\hat{x}^{(1)}})_{22} \\ -(Q_{\hat{x}^{(2)}})_{22} \\ -(Q_{\hat{x}^{(2)}_{32}} \end{pmatrix} [(Q_{\hat{x}^{(1)}})_{22} + (Q_{\hat{x}^{(2)}})_{22}]^{-1} (0 \quad I \; -I \quad 0) \right\} \begin{pmatrix} \underline{\hat{x}}_1^{(1)} \\ \underline{\hat{x}}_2^{(1)} \\ \underline{\hat{x}}_2^{(2)} \\ \underline{\hat{x}}_3^{(2)} \end{pmatrix}$$

Note that the solution (123) generalizes solution (113). The above results are summarized in table 6.5.

(1) $$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12} & 0\\ 0 & \boldsymbol{A}_{22} & \boldsymbol{A}_{23}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}\ ;\quad D\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}Q_1 & 0\\ 0 & Q_2\end{pmatrix}.$$

Solving the *partial model*

(2) $$E\{\underline{y}_1\} = (\boldsymbol{A}_{11}\,\boldsymbol{A}_{12})\begin{pmatrix}x_1\\ x_2\end{pmatrix}\ ;\quad D\{\underline{y}_1\} = Q_1$$

gives:

(3) $$\begin{pmatrix}\underline{\hat{x}}_1^{(1)}\\ \underline{\hat{x}}_2^{(1)}\end{pmatrix}\ \text{and}\ \begin{pmatrix}(Q_{\hat{x}^{(1)}})_{11} & (Q_{\hat{x}^{(1)}})_{12}\\ (Q_{\hat{x}^{(1)}})_{21} & (Q_{\hat{x}^{(1)}})_{22}\end{pmatrix}.$$

Solving the *partial model*

$$E\{\underline{y}_2\} = (\boldsymbol{A}_{22}\,\boldsymbol{A}_{23})\begin{pmatrix}x_2\\ x_3\end{pmatrix}\ ;\quad D\{\underline{y}_2\} = Q_2$$

gives:

(4) $$\begin{pmatrix}\underline{\hat{x}}_2^{(2)}\\ \underline{\hat{x}}_3^{(2)}\end{pmatrix}\ \text{and}\ \begin{pmatrix}(Q_{\hat{x}^{(2)}})_{22} & (Q_{\hat{x}^{(2)}})_{23}\\ (Q_{\hat{x}^{(2)}})_{32} & (Q_{\hat{x}^{(2)}})_{33}\end{pmatrix}.$$

With (3) and (4) we formulate the model of *condition equations:*

(5) $$(0\quad I\quad -I\quad 0)E\{\begin{pmatrix}\underline{\hat{x}}_1^{(1)}\\ \underline{\hat{x}}_2^{(1)}\\ \underline{\hat{x}}_2^{(2)}\\ \underline{\hat{x}}_3^{(2)}\end{pmatrix}\} = 0\ ;\ D\{\begin{pmatrix}\underline{\hat{x}}_1^{(1)}\\ \underline{\hat{x}}_2^{(1)}\\ \underline{\hat{x}}_2^{(2)}\\ \underline{\hat{x}}_3^{(2)}\end{pmatrix}\} = \begin{pmatrix}(Q_{\hat{x}^{(1)}})_{11} & (Q_{\hat{x}^{(1)}})_{12} & 0 & 0\\ (Q_{\hat{x}^{(1)}})_{21} & (Q_{\hat{x}^{(1)}})_{22} & 0 & 0\\ 0 & 0 & (Q_{\hat{x}^{(2)}})_{22} & (Q_{\hat{x}^{(2)}})_{23}\\ 0 & 0 & (Q_{\hat{x}^{(2)}})_{32} & (Q_{\hat{x}^{(2)}})_{33}\end{pmatrix}$$

Its solution reads:

(6) $$\begin{cases}\underline{\hat{x}}_1 &= \underline{\hat{x}}_1^{(1)} - (Q_{\hat{x}^{(1)}})_{12}[(Q_{\hat{x}^{(1)}})_{22} + (Q_{\hat{x}^{(2)}})_{22}]^{-1}(\underline{\hat{x}}_2^{(1)} - \underline{\hat{x}}_2^{(2)})\\ \underline{\hat{x}}_2 &= \underline{\hat{x}}_2^{(1)} - (Q_{\hat{x}^{(1)}})_{22}[(Q_{\hat{x}^{(1)}})_{22} + (Q_{\hat{x}^{(2)}})_{22}]^{-1}(\underline{\hat{x}}_2^{(1)} - \underline{\hat{x}}_2^{(2)})\\ \underline{\hat{x}}_2 &= \underline{\hat{x}}_2^{(2)} - (Q_{\hat{x}^{(2)}})_{22}[(Q_{\hat{x}^{(1)}})_{22} + (Q_{\hat{x}^{(2)}})_{22}]^{-1}(\underline{\hat{x}}_2^{(2)} - \underline{\hat{x}}_2^{(1)})\\ \underline{\hat{x}}_3 &= \underline{\hat{x}}_3^{(2)} - (Q_{\hat{x}^{(2)}})_{32}[(Q_{\hat{x}^{(1)}})_{22} + (Q_{\hat{x}^{(2)}})_{22}]^{-1}(\underline{\hat{x}}_2^{(2)} - \underline{\hat{x}}_2^{(1)})\end{cases}$$

This is also the solution of (1).

Table 6.5: Block estimation

Example 3

Consider the levelling network of figure 6.6. Point 0 is taken as reference point.

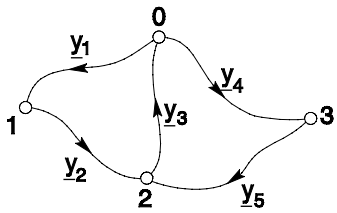


Figure 6.6: Levelling network

Its height is assumed to be known and equal to zero, that is x_0=0. The corresponding model of observation equations reads then:

$$E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \underline{y}_3\\ \underline{y}_4\\ \underline{y}_5\end{pmatrix}\} = \begin{pmatrix}1 & 0 & 0\\ -1 & 1 & 0\\ 0 & -1 & 0\\ 0 & 0 & 1\\ 0 & 1 & -1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}; \quad D\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \underline{y}_3\\ \underline{y}_4\\ \underline{y}_5\end{pmatrix}\} = \sigma^2 I_5 \ . \tag{124}$$

The normal equations read:

$$\begin{pmatrix}2 & -1 & 0\\ -1 & 3 & -1\\ 0 & -1 & 2\end{pmatrix}\begin{pmatrix}\hat{\underline{x}}_1\\ \hat{\underline{x}}_2\\ \hat{\underline{x}}_3\end{pmatrix} = \begin{pmatrix}\underline{y}_1 - \underline{y}_2\\ \underline{y}_2 - \underline{y}_3 + \underline{y}_5\\ \underline{y}_4 - \underline{y}_5\end{pmatrix} . \tag{125}$$

In order to solve (125), we first eliminate $\hat{\underline{x}}_1$ from the second normal equation. This is achieved if we premultiply (125) with the square and regular matrix:

$$\begin{pmatrix}1 & 0 & 0\\ \frac{1}{2} & 1 & 0\\ 0 & 0 & 1\end{pmatrix} .$$

This gives:

$$\begin{pmatrix} 2 & -1 & 0 \\ 0 & 2\frac{1}{2} & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \begin{pmatrix} \underline{y}_1 - \underline{y}_2 \\ \frac{1}{2}\underline{y}_1 + \frac{1}{2}\underline{y}_2 - \underline{y}_3 + \underline{y}_5 \\ \underline{y}_4 - \underline{y}_5 \end{pmatrix}. \tag{126}$$

The estimators $\underline{\hat{x}}_2$ and $\underline{\hat{x}}_3$ follow then from:

$$\begin{pmatrix} \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \begin{pmatrix} 2\frac{1}{2} & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{2}\underline{y}_1 + \frac{1}{2}\underline{y}_2 - \underline{y}_3 + \underline{y}_5 \\ \underline{y}_4 - \underline{y}_5 \end{pmatrix}$$

as

$$\begin{pmatrix} \underline{\hat{x}}_2 \\ \underline{\hat{x}}_3 \end{pmatrix} = \frac{1}{8} \begin{pmatrix} 2\underline{y}_1 + 2\underline{y}_2 - 4\underline{y}_3 + 2\underline{y}_4 + 2\underline{y}_5 \\ \underline{y}_1 + \underline{y}_2 - 2\underline{y}_3 + 5\underline{y}_4 - 3\underline{y}_5 \end{pmatrix}. \tag{127}$$

Substitution of $\underline{\hat{x}}_2$ into the first equation of (126) gives for $\underline{\hat{x}}_1$:

$$\underline{\hat{x}}_1 = \frac{1}{8}(5\underline{y}_1 - 3\underline{y}_2 - 2\underline{y}_3 + \underline{y}_4 + \underline{y}_5). \tag{128}$$

With (127) and (128) we have found the solution of (124). Let us now try to solve (124) using the block estimation scheme of table 6.6. The two partial models read:

$$E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \end{pmatrix}\} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \; ; \; \boldsymbol{D}\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \end{pmatrix}\} = \sigma^2 I_3 \;, \tag{129}$$

and:

$$E\{\begin{pmatrix} \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \; ; \; \boldsymbol{D}\{\begin{pmatrix} \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\} = \sigma^2 I_2 \;. \tag{130}$$

The system of normal equations for partial model (129) reads:

$$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_{1^{(1)}} \\ \underline{\hat{x}}_{2^{(1)}} \end{pmatrix} = \begin{pmatrix} \underline{y}_1 - \underline{y}_2 \\ \underline{y}_2 - \underline{y}_3 \end{pmatrix}. \tag{131}$$

From this it follows that:

$$\begin{pmatrix} \underline{\hat{x}}_{1^{(1)}} \\ \underline{\hat{x}}_{2^{(1)}} \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2\underline{y}_1 - \underline{y}_2 - \underline{y}_3 \\ \underline{y}_1 + \underline{y}_2 - 2\underline{y}_3 \end{pmatrix} \quad ; \quad Q_{\hat{x}^{(1)}} = \frac{1}{3}\sigma^2 \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}. \tag{132}$$

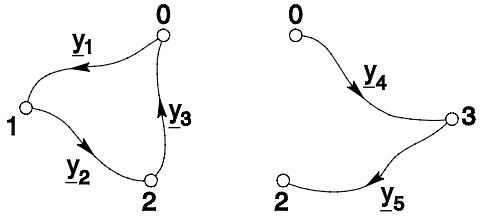
The system of normal equations for partial model (130) reads:

$$\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \hat{\underline{x}}_2^{(2)} \\ \hat{\underline{x}}_3^{(2)} \end{pmatrix} = \begin{pmatrix} \underline{y}_5 \\ \underline{y}_4 - \underline{y}_5 \end{pmatrix}. \tag{133}$$

From this it follows that:

$$\begin{pmatrix} \hat{\underline{x}}_2^{(2)} \\ \hat{\underline{x}}_3^{(2)} \end{pmatrix} = \begin{pmatrix} \underline{y}_4 + \underline{y}_5 \\ \underline{y}_4 \end{pmatrix} \quad ; \quad Q_{\hat{x}^{(2)}} = \sigma^2 \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}. \tag{134}$$

It will be clear that (132) is the solution of the partial levelling network shown in figure 6.7a, and that (134) is the solution of the partial levelling network shown in figure 6.7b.



(a) (b)

Figure 6.7: Two partial levelling networks

Now that the solutions of the two partial networks have been found, we can *connect* the two networks by formulating the model of condition equations:

$$(0 \quad I \quad -I \quad 0)\, E\{\begin{pmatrix} \hat{\underline{x}}_1^{(1)} \\ \hat{\underline{x}}_2^{(1)} \\ \hat{\underline{x}}_2^{(2)} \\ \hat{\underline{x}}_3^{(2)} \end{pmatrix}\} = 0 \quad ; \quad D\{\begin{pmatrix} \hat{\underline{x}}_1^{(1)} \\ \hat{\underline{x}}_2^{(1)} \\ \hat{\underline{x}}_2^{(2)} \\ \hat{\underline{x}}_3^{(2)} \end{pmatrix}\} = \sigma^2 \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}. \tag{135}$$

Note that the number of condition equations in (135) equals one. Thus the redundancy equals one. The redundancy in partial model (130) equals zero and the redundancy of partial model (129) equals one. Hence the total redundancy equals 1+0+1=2. And this is of course equal to the number of linear independent condition equations that can be found in the levelling network of figure 6.6.

The solution of (135) follows as (see also (123)):

$$(136)\qquad \begin{cases} \underline{\hat{x}}_1 &= \underline{\hat{x}}_1^{(1)} - \frac{1}{3}\,(\frac{2}{3}+2)^{-1}(\underline{\hat{x}}_2^{(1)} - \underline{\hat{x}}_2^{(2)}) \;, \\ \underline{\hat{x}}_2 &= \underline{\hat{x}}_2^{(1)} - \frac{2}{3}\,(\frac{2}{3}+2)^{-1}(\underline{\hat{x}}_2^{(1)} - \underline{\hat{x}}_2^{(2)}) \;, \\ \underline{\hat{x}}_2 &= \underline{\hat{x}}_2^{(2)} - 2\,(\frac{2}{3}+2)^{-1}(\underline{\hat{x}}_2^{(2)} - \underline{\hat{x}}_2^{(1)}) \;, \\ \underline{\hat{x}}_3 &= \underline{\hat{x}}_3^{(2)} - 1\,(\frac{2}{3}+2)^{-1}(\underline{\hat{x}}_2^{(2)} - \underline{\hat{x}}_2^{(1)}) \;. \end{cases}$$

We will now verify that solution (136) is indeed identical to (127) and (128). Substitution of $\underline{\hat{x}}_1^{(1)}, \underline{\hat{x}}_2^{(1)}$ and $\underline{\hat{x}}_2^{(2)}$ from (132) and (134) into the first equation of (136) gives:

$$(137)\qquad \begin{aligned} \underline{\hat{x}}_1 &= \tfrac{1}{3}(2\underline{y}_1 - \underline{y}^2 - \underline{y}_3) - \tfrac{1}{3}\tfrac{3}{\mathbf{8}}[(\tfrac{1}{3}\underline{y}_1 + \tfrac{1}{3}\underline{y}_2 - \tfrac{2}{3}\underline{y}_3) - (\underline{y}_{\mathbf{4}} + \underline{y}_5)] \\ &= \tfrac{1}{\mathbf{8}}(5\underline{y}_1 - 3\underline{y}_2 - 2\underline{y}_3 + \underline{y}_{\mathbf{4}} + \underline{y}_5). \end{aligned}$$

And this is indeed identical to (128). The verification of the remaining equations in (136) is left to the reader.

The block estimation scheme of table 6.5 that has been developed in this section, can also be derived from the mixed model representation of section 5.3. To see this, note that (116) is equivalent to:

$$(138)\qquad E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12} & 0 & 0\\ 0 & 0 & \boldsymbol{A}_{22} & \boldsymbol{A}_{23}\end{pmatrix}\begin{pmatrix}x_1^{(1)}\\ x_2^{(1)}\\ x_2^{(2)}\\ x_3^{(2)}\end{pmatrix} \;;\; (0 \quad I \quad -I \quad 0)\begin{pmatrix}x_1^{(1)}\\ x_2^{(1)}\\ x_2^{(2)}\\ x_3^{(3)}\end{pmatrix} = 0 \;.$$

But this representation is of the form:

$$(139)\qquad E\{\underline{y}\} = \boldsymbol{A}x \;;\; \boldsymbol{B}^* x = 0 \;.$$

Show for yourself that the block estimation scheme of table 6.5 follows when the results of section 5.3 are applied to model (138). According to (45) of section 5.3 we have

$$\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}} = \underline{y}^* Q_y^{-1} \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \underline{y} \;+\; \underline{\hat{x}}_{\boldsymbol{A}}^* \boldsymbol{B}(\boldsymbol{B}^* Q_{\hat{x}_{\boldsymbol{A}}} \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{\hat{x}}_{\boldsymbol{A}} \;.$$

If we "translate" this result to model (138) we get with $\underline{\hat{e}}_{1^{(1)}} = \underline{y}_1 - \underline{\hat{y}}_{1^{(1)}} = \underline{y}_1 - \boldsymbol{A}_{11}\underline{\hat{x}}_{1^{(1)}} - \boldsymbol{A}_{12}\underline{\hat{x}}_{2^{(1)}}$ and $\underline{\hat{e}}_{2^{(2)}} = \underline{y}_2 - \underline{\hat{y}}_{2^{(2)}} = \underline{y}_2 - \boldsymbol{A}_{22}\underline{\hat{x}}_{2^{(2)}} - \boldsymbol{A}_{23}\underline{\hat{x}}_{3^{(2)}}$:

(140)
$$\begin{aligned}\hat{\underline{e}}^{*}Q_{y}^{-1}\hat{\underline{e}} &= \hat{\underline{e}}_{1}^{(1)^{*}}Q_{1}^{-1}\hat{\underline{e}}_{1}^{(1)} \\ &+ \hat{\underline{e}}_{2}^{(2)}Q_{2}^{-1}\hat{\underline{e}}_{2}^{(2)} + (\hat{\underline{x}}_{2}^{(1)} - \hat{\underline{x}}_{2}^{(2)})^{*}[(Q_{\hat{x}^{(1)}})_{22} + (Q_{\hat{x}^{(2)}})_{22}]^{-1}(\hat{\underline{x}}_{2}^{(1)} - \hat{\underline{x}}_{2}^{(2)})\end{aligned}$$

This shows that the squared norm of the residuals of model (138) and thus of model (116), can be found from the sum of the squared norm of the residuals of the three models in (118) and (122).

6.6 The partitioned model $\begin{pmatrix} B_1^{*} \\ B_2^{*} \end{pmatrix} E\{\underline{y}\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Estimation in phases

In this section we consider the partitioned model of *condition equations*:

(141)
$$\begin{pmatrix} B_1^{*} \\ B_2^{*} \end{pmatrix} E\{\underline{y}\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad ; \quad D\{\underline{y}\} = Q_y \ .$$

It will be shown that the solution of this model, which will be denoted by $\hat{\underline{y}}_B$, can be found by first solving the partial model:

(142)
$$B_1^{*} E\{\underline{y}\} = 0 \quad ; \quad D\{\underline{y}\} = Q_y \ ,$$

which gives $\hat{\underline{y}}_{B_1}$ and $Q_{\hat{y}_{B_1}}$, and then solving for the model:

(143)
$$B_2^{*} E\{\hat{\underline{y}}_{B_1}\} = 0 \quad ; \quad D\{\underline{y}_{B_1}\} = Q_{\hat{y}_{B_1}} \ .$$

This solution method is known as the *method of estimation in phases* (fase vereffening). This method was originally introduced by the famous Dutch geodesist J.M. Tienstra (1895-1951). Tienstra was appointed professor in Mathematical Geodesy in 1935 at the Delft University of Technology.

With $B=(B_1 \vdots B_2)$, the solution of (141) reads:

(144)
$$\hat{\underline{y}}_{B} = [I - Q_y B(B^{*}Q_y B)^{-1}B^{*}]\underline{y} \ .$$

If we define the vector random variable $\hat{\underline{z}}$ as:

(145)
$$\hat{\underline{z}} = (B^{*}Q_y B)^{-1}B^{*}\underline{y} \ ,$$

then $\hat{\underline{z}} = (\hat{\underline{z}}_1^{*}, \hat{\underline{z}}_2^{*})^{*}$ is the solution of the partitioned system of these "normal equations":

(146)
$$\begin{pmatrix} B_1^{*}Q_y B_1 & B_1^{*}Q_y B_2 \\ B_2^{*}Q_y B_1 & B_2^{*}Q_y B_2 \end{pmatrix} \begin{pmatrix} \hat{\underline{z}}_1 \\ \hat{\underline{z}}_2 \end{pmatrix} = \begin{pmatrix} B_1^{*}\underline{y} \\ B_2^{*}\underline{y} \end{pmatrix} \ .$$

The estimator $\hat{\underline{z}}_1$ can be eliminated from the second set of normal equations if we premultiply (146) with the square and full rank matrix:

$$\begin{pmatrix} I & 0 \\ -\boldsymbol{B}_2^* Q_y \boldsymbol{B}_1 (\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} & I \end{pmatrix} .$$

This gives:

$$\begin{pmatrix} \boldsymbol{B}_1^* Q_y \boldsymbol{B}_1 & \boldsymbol{B}_1^* Q_y \boldsymbol{B}_2 \\ 0 & \boldsymbol{B}_2^* [I - Q_y \boldsymbol{B}_1 (\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^*] Q_y \boldsymbol{B}_2 \end{pmatrix} \begin{pmatrix} \hat{\underline{z}}_1 \\ \hat{\underline{z}}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}_1^* \underline{y} \\ \boldsymbol{B}_2^* [I - Q_y \boldsymbol{B}_1 (\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^*] \underline{y} \end{pmatrix} \tag{147}$$

In (147) we recognize the orthogonal projector:

$$P_{Q_y \boldsymbol{B}_1}^{\perp} = [I - Q_y \boldsymbol{B}_1 (\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^*] .$$

Since:

$$P_{Q_y \boldsymbol{B}_1}^{\perp} Q_y P_{Q_y \boldsymbol{B}_1}^{\perp *} = P_{Q_y \boldsymbol{B}_1}^{\perp} Q_y = Q_y P_{Q_y \boldsymbol{B}_1}^{\perp *} , \tag{148}$$

we may write (147), using the abbreviation:

$$\bar{\boldsymbol{B}}_2 = P_{Q_y \boldsymbol{B}_1}^{\perp *} \boldsymbol{B}_2 , \tag{149}$$

also as:

$$\begin{pmatrix} \boldsymbol{B}_1^* Q_y \boldsymbol{B}_1 & \boldsymbol{B}_1^* Q_y \boldsymbol{B}_2 \\ 0 & \bar{\boldsymbol{B}}_2^* Q_y \bar{\boldsymbol{B}}_2 \end{pmatrix} \begin{pmatrix} \hat{\underline{z}}_1 \\ \hat{\underline{z}}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}_1^* \underline{y} \\ \bar{\boldsymbol{B}}_2^* \underline{y} \end{pmatrix} . \tag{150}$$

The solution of this reduced system of normal equations follows then as:

$$\begin{cases} \hat{\underline{z}}_1 = (\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^* (\underline{y} - Q_y \boldsymbol{B}_2 \hat{\underline{z}}_2) , \\ \hat{\underline{z}}_2 = (\bar{\boldsymbol{B}}_2^* Q_y \bar{\boldsymbol{B}}_2)^{-1} \bar{\boldsymbol{B}}_2^* \underline{y} . \end{cases} \tag{151}$$

With $B=(B_1 \vdots B_2)$ it follows from (144) and (145) that $\hat{\underline{y}}_B$ can be written as:

$$\hat{\underline{y}}_{\boldsymbol{B}} = \underline{y} - Q_y \boldsymbol{B}_1 \hat{\underline{z}}_1 - Q_y \boldsymbol{B}_2 \hat{\underline{z}}_2 .$$

Substitution of (151) gives:

$$\hat{\underline{y}}_{\boldsymbol{B}} = [I - Q_y \boldsymbol{B}_1 (\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^*] (\underline{y} - Q_y \boldsymbol{B}_2 \hat{\underline{z}}_2) ,$$

or:

$$\hat{\underline{y}}_{\boldsymbol{B}} = [I - Q_y \boldsymbol{B}_1 (\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^*] [I - Q_y \boldsymbol{B}_2 (\bar{\boldsymbol{B}}_2^* Q_y \bar{\boldsymbol{B}}_2)^{-1} \bar{\boldsymbol{B}}_2^*] \underline{y}$$

or, with (149):

$$\hat{\underline{y}}_{\boldsymbol{B}} = [P_{Q_y \boldsymbol{B}_1}^{\perp} - P_{Q_y \boldsymbol{B}_1}^{\perp} Q_y \boldsymbol{B}_2 (\boldsymbol{B}_2^* P_{Q_y \boldsymbol{B}_1}^{\perp} Q_y P_{Q_y \boldsymbol{B}_1}^{\perp *} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^* P_{Q_y \boldsymbol{B}_1}^{\perp}] \underline{y}$$

or:

$$\hat{\underline{y}}_{\boldsymbol{B}} = [I - P_{Q_y \boldsymbol{B}_1}^{\perp} Q_y \boldsymbol{B}_2 (\boldsymbol{B}_2^* P_{Q_y \boldsymbol{B}_1}^{\perp} Q_y P_{Q_y \boldsymbol{B}_1}^{\perp *} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^*] P_{Q_y \boldsymbol{B}_1}^{\perp} \underline{y} . \tag{152}$$

Clearly:

(153) $$\underline{\hat{y}}_{\boldsymbol{B}_1} = \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp} \underline{y} \ ,$$

is the solution of (142). The variance matrix of $\underline{\hat{y}}_{\boldsymbol{B}_1}$ reads:

(154) $$\boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} = \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp} \boldsymbol{Q}_y \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp^*} = \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp} \boldsymbol{Q}_y = \boldsymbol{Q}_y \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp^*}.$$

Substitution of (153) and (154) into (152) gives:

(155) $$\boxed{\underline{\hat{y}}_{\boldsymbol{B}} = [I - \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2 (\boldsymbol{B}_2^* \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^*] \underline{\hat{y}}_{\boldsymbol{B}_1}}$$

And this is clearly the solution of (143). Hence we have shown that the solution of the partitioned model of condition equations (141) can indeed be found by first solving for the partial model (142) and then solving for the model (143).

Let us now consider the least-squares residual vector $\underline{\hat{e}}_B = \underline{y} - \underline{\hat{y}}_B$. From (155) it follows that:

(156) $$\underline{\hat{e}}_{\boldsymbol{B}} = \underline{y} - \underline{\hat{y}}_{\boldsymbol{B}_1} + \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2 (\boldsymbol{B}_2^* \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^* \underline{\hat{y}}_{\boldsymbol{B}_1} \ .$$

With:

$$\underline{\hat{e}}_{\boldsymbol{B}_1} = \underline{y} - \underline{\hat{y}}_{\boldsymbol{B}_1} \ ,$$

and:

$$\underline{\hat{e}}_{\boldsymbol{B}_2} = \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2 (\boldsymbol{B}_2^* \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^* \underline{\hat{y}}_{\boldsymbol{B}_1} \ ,$$

this shows that:

(157) $$\boxed{\underline{\hat{e}}_{\boldsymbol{B}} = \underline{\hat{e}}_{\boldsymbol{B}_1} + \underline{\hat{e}}_{\boldsymbol{B}_2}} \ .$$

Hence, the least-squares residual vector of the partitioned model (141) can be found from the sum of the least-squares residual vectors of the two separate models (142) and (143). From (157) it follows that:

(158) $$\underline{\hat{e}}_{\boldsymbol{B}}^* \boldsymbol{Q}_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}} = \underline{\hat{e}}_{\boldsymbol{B}_1}^* \boldsymbol{Q}_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}_1} + 2 \underline{\hat{e}}_{\boldsymbol{B}_1}^* \boldsymbol{Q}_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}_2} + \underline{\hat{e}}_{\boldsymbol{B}_2}^* \boldsymbol{Q}_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}_2}.$$

Since:

$$\begin{aligned} \underline{\hat{e}}_{\boldsymbol{B}_1}^* \boldsymbol{Q}_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}_2} &= (\underline{y} - \underline{\hat{y}}_{\boldsymbol{B}_1})^* \boldsymbol{Q}_y^{-1} \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2 (\boldsymbol{B}_2^* \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^* \underline{\hat{y}}_{\boldsymbol{B}_1} \\ &= \underline{y}^* \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{*} \boldsymbol{Q}_y^{-1} \boldsymbol{Q}_y \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp^*} \boldsymbol{B}_2 (\boldsymbol{B}_2^* \boldsymbol{Q}_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^* \underline{\hat{y}}_{\boldsymbol{B}_1} \\ &= 0 \ , \end{aligned}$$

because:

$$\boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp} \boldsymbol{P}_{Q_y \boldsymbol{B}_1} = 0 \ ,$$

it follows that (158) reduces to:

(159)
$$\boxed{\hat{\underline{e}}_{\boldsymbol{B}}^{*} Q_y^{-1} \hat{\underline{e}}_{\boldsymbol{B}} = \hat{\underline{e}}_{\boldsymbol{B}_1}^{*} Q_y^{-1} \hat{\underline{e}}_{\boldsymbol{B}_1} + \hat{\underline{e}}_{\boldsymbol{B}_2}^{*} Q_y^{-1} \hat{\underline{e}}_{\boldsymbol{B}_2}}$$

From the result it seems that one needs Q_y for computing the scalar $\hat{\underline{e}}_{B_2}^{*} Q_y^{-1} \hat{\underline{e}}_{B_2}$. If this would be the case, it would disrupt the method of estimation in phases, because it would imply that Q_y is also needed in the second phase, that is, when model (143) is solved for. Fortunately one can show that Q_y is *not* required for computing the scalar $\hat{\underline{e}}_{B_2}^{*} Q_y^{-1} \hat{\underline{e}}_{B_2}$. In order to show this we define the vectors of misclosures:

(160)
$$\begin{cases} \underline{t}_{\boldsymbol{B}_1} = \boldsymbol{B}_1^{*} \underline{y} \quad , \; Q_{t_{\boldsymbol{B}_1}} = \boldsymbol{B}_1^{*} Q_y \boldsymbol{B}_1 \; , \\ \underline{t}_{\boldsymbol{B}_2} = \boldsymbol{B}_2^{*} \underline{\hat{y}}_{\boldsymbol{B}_1} \; , \; Q_{t_{\boldsymbol{B}_2}} = \boldsymbol{B}_2^{*} Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2 \; . \end{cases}$$

With (160) we have:

$$\hat{\underline{e}}_{\boldsymbol{B}_1} = Q_y \boldsymbol{B}_1 (\boldsymbol{B}_1^{*} Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^{*} \underline{y} = Q_y \boldsymbol{B}_1 Q_{t_{\boldsymbol{B}_1}}^{-1} \underline{t}_{\boldsymbol{B}_1} \; .$$

From this it follows that:

(161)
$$\hat{\underline{e}}_{\boldsymbol{B}_1}^{*} Q_y^{-1} \hat{\underline{e}}_{\boldsymbol{B}_1} = \underline{t}_{\boldsymbol{B}_1}^{*} Q_{t_{\boldsymbol{B}_1}}^{-1} \underline{t}_{\boldsymbol{B}_1} \; .$$

With (160) we also have:

(162)
$$\hat{\underline{e}}_{\boldsymbol{B}_2} = Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2 (\boldsymbol{B}_2^{*} Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^{*} \underline{\hat{y}}_{\boldsymbol{B}_1} = Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2 Q_{t_{\boldsymbol{B}_2}}^{-1} \underline{t}_{\boldsymbol{B}_2} \; .$$

Since:

$$Q_{\hat{y}_{\boldsymbol{B}_1}} Q_y^{-1} Q_{\hat{y}_{\boldsymbol{B}_1}} = \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp} Q_y Q_y^{-1} Q_y \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp *} = Q_{\hat{y}_{\boldsymbol{B}_1}} \quad \text{(see (154))},$$

it follows from (162) that:

(163)
$$\hat{\underline{e}}_{\boldsymbol{B}_2}^{*} Q_y^{-1} \hat{\underline{e}}_{\boldsymbol{B}_2} = \underline{t}_{\boldsymbol{B}_2}^{*} Q_{t_{\boldsymbol{B}_2}}^{-1} \underline{t}_{\boldsymbol{B}_2} \; .$$

Substitution of (161) and (163) into (159) finally gives:

(164)
$$\boxed{\hat{\underline{e}}_{\boldsymbol{B}}^{*} Q_y^{-1} \hat{\underline{e}}_{\boldsymbol{B}} = \underline{t}_{\boldsymbol{B}_1}^{*} Q_{t_{\boldsymbol{B}_1}}^{-1} \underline{t}_{\boldsymbol{B}_1} + \underline{t}_{\boldsymbol{B}_2}^{*} Q_{t_{\boldsymbol{B}_2}}^{-1} \underline{t}_{\boldsymbol{B}_2}}.$$

A summary of the above results is given in table 6.6.

Partitioned model

$$\begin{pmatrix} \boldsymbol{B}_1^* \\ \boldsymbol{B}_2^* \end{pmatrix} E\{\underline{y}\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \; ; \; D\{\underline{y}\} = Q_y \; .$$

$$\underline{\hat{y}}_{\boldsymbol{B}} = [I - Q_y \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^*]\underline{y}$$

$$\underline{\hat{e}}_{\boldsymbol{B}} = \underline{y} - \underline{\hat{y}}_{\boldsymbol{B}} = Q_y \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y}$$

$$\underline{\hat{e}}_{\boldsymbol{B}}^* Q_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}} = \underline{y}^* \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y} = \underline{t}_{\boldsymbol{B}}^* Q_{t_{\boldsymbol{B}}}^{-1} \underline{t}_{\boldsymbol{B}}, \quad \text{with } \underline{t}_{\boldsymbol{B}} = \boldsymbol{B}^* \underline{y} \; .$$

1.

$$\boldsymbol{B}_1^* E\{\underline{y}\} = 0 \; ; \; D\{\underline{y}\} = Q_y$$

$$\underline{\hat{y}}_{\boldsymbol{B}_1} = [I - Q_y \boldsymbol{B}_1(\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^*]\underline{y}$$

$$\underline{\hat{e}}_{\boldsymbol{B}_1} = \underline{y} - \underline{\hat{y}}_{\boldsymbol{B}_1} = Q_y \boldsymbol{B}_1(\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^* \underline{y}$$

$$\underline{\hat{e}}_{\boldsymbol{B}_1}^* Q_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}_1} = \underline{y}^* \boldsymbol{B}_1(\boldsymbol{B}_1^* Q_y \boldsymbol{B}_1)^{-1} \boldsymbol{B}_1^* \underline{y} = \underline{t}_{\boldsymbol{B}_1}^* Q_{t_{\boldsymbol{B}_1}}^{-1} \underline{t}_{\boldsymbol{B}}, \quad \text{with } \underline{t}_{\boldsymbol{B}_1} = \boldsymbol{B}_1^* \underline{y}$$

2.

$$\boldsymbol{B}_2^* E\{\underline{\hat{y}}_{\boldsymbol{B}_1}\} = 0 \; ; \; D\{\underline{y}_{\boldsymbol{B}_1}\} = Q_{y_{\boldsymbol{B}_1}}$$

$$\underline{\hat{y}}_{\boldsymbol{B}} = [I - Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2(\boldsymbol{B}_2^* Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^*]\underline{\hat{y}}_{\boldsymbol{B}_1}$$

$$\underline{\hat{e}}_{\boldsymbol{B}_2} = \underline{\hat{y}}_{\boldsymbol{B}_1} - \underline{\hat{y}}_{\boldsymbol{B}} = Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2(\boldsymbol{B}_2^* Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^* \underline{\hat{y}}_{\boldsymbol{B}_1}$$

$$\underline{\hat{e}}_{\boldsymbol{B}_2}^* Q_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}_2} = \underline{\hat{y}}_{\boldsymbol{B}_1}^* \boldsymbol{B}_2(\boldsymbol{B}_2^* Q_{\hat{y}_{\boldsymbol{B}_1}} \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^* \underline{\hat{y}}_{\boldsymbol{B}_1} = \underline{t}_{\boldsymbol{B}_2}^* Q_{t_{\boldsymbol{B}_2}}^{-1} \underline{t}_{\boldsymbol{B}_2}, \quad \text{with } \underline{t}_{\boldsymbol{B}_2} = \boldsymbol{B}_2^* \underline{\hat{y}}_{\boldsymbol{B}}$$

$$\underline{\hat{e}}_{\boldsymbol{B}} = \underline{\hat{e}}_{\boldsymbol{B}_1} + \underline{\hat{e}}_{\boldsymbol{B}_2} \; ; \; \underline{\hat{e}}_{\boldsymbol{B}}^* Q_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}} = \underline{\hat{e}}_{\boldsymbol{B}_1}^* Q_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}_1} + \underline{\hat{e}}_{\boldsymbol{B}_2}^* Q_y^{-1} \underline{\hat{e}}_{\boldsymbol{B}_2}$$

$$= \underline{t}_{\boldsymbol{B}_1}^* Q_{t_{\boldsymbol{B}_1}}^{-1} \underline{t}_{\boldsymbol{B}_1} + \underline{t}_{\boldsymbol{B}_2}^* Q_{t_{\boldsymbol{B}_2}}^{*} \underline{t}_{\boldsymbol{B}_2}$$

Table 6.6: Estimation in phases

Example 4

Consider the levelling network of figure 6.8.

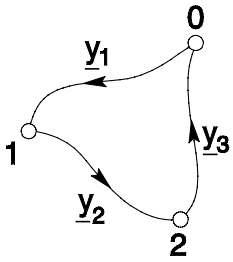


Figure 6.8: A one loop levelling network

The corresponding model of condition equations reads:

$$(1\ \ 1\ \ 1)\,E\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \underline{y}_3\end{pmatrix}\}=0\ ;\ \boldsymbol{D}\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \underline{y}_3\end{pmatrix}\}=\sigma^2 I_3\ . \tag{165}$$

Its solution reads:

$$\begin{pmatrix}\hat{\underline{y}}_1\\ \hat{\underline{y}}_2\\ \hat{\underline{y}}_3\end{pmatrix}=\left[\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}-\sigma^2 I_3\begin{pmatrix}1\\1\\1\end{pmatrix}\frac{1}{3}\sigma^{-2}(1\ \ 1\ \ 1)\right]\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \underline{y}_3\end{pmatrix},$$

or:

$$\begin{pmatrix}\hat{\underline{y}}_1\\ \hat{\underline{y}}_2\\ \hat{\underline{y}}_3\end{pmatrix}=\frac{1}{3}\begin{pmatrix}2\underline{y}_1-\underline{y}_2-\underline{y}_3\\ -\underline{y}_1+2\underline{y}_2-\underline{y}_3\\ -\underline{y}_1-\underline{y}_2+2\underline{y}_3\end{pmatrix}. \tag{166}$$

The variance matrix of (166) reads:

$$Q_{\hat{y}}=\frac{1}{3}\sigma^2\begin{pmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{pmatrix}. \tag{167}$$

Now let us assume that the levelling network of figure 6.8 is enlarged with two additional height difference observables, such that the two loop levelling network of figure 6.9 results.

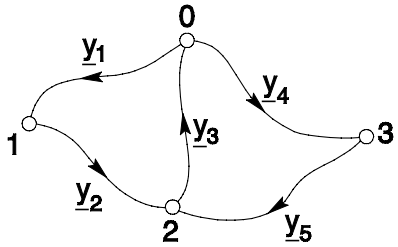


Figure 6.9: A two loop levelling network

The corresponding model of condition equations reads:

$$(168) \qquad \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad ; \quad D\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\} = \sigma^2 I_5 \ .$$

According to table 6.6, this model can be solved in two phases. The *first phase* would then correspond to solving the partial model:

$$(169) \qquad (1 \ 1 \ 1 \ 0 \ 0) E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\} = 0 \quad ; \quad D\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\} = \sigma^2 I_5 \ .$$

Its solution reads:

$$\begin{pmatrix} \underline{\hat{y}}_1 \\ \underline{\hat{y}}_2 \\ \underline{\hat{y}}_3 \\ \underline{\hat{y}}_4 \\ \underline{\hat{y}}_5 \end{pmatrix}_{B_1} = \left[\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} - \sigma^2 I_5 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \frac{1}{3} \sigma^{-2} (1 \ 1 \ 1 \ 0 \ 0) \right] \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix} ,$$

or:

$$(170)\qquad \begin{pmatrix} \underline{\hat{y}}_1 \\ \underline{\hat{y}}_2 \\ \underline{\hat{y}}_3 \\ \underline{\hat{y}}_4 \\ \underline{\hat{y}}_5 \end{pmatrix}_{\boldsymbol{B}_1} = \begin{pmatrix} \frac{1}{3}\begin{pmatrix} 2\underline{y}_1 - \underline{y}_2 - \underline{y}_3 \\ -\underline{y}_1 + 2\underline{y}_2 - \underline{y}_3 \\ -\underline{y}_1 - \underline{y}_2 + 2\underline{y}_3 \end{pmatrix} \\ \begin{pmatrix} \underline{y}_4 \\ \underline{y}_5 \end{pmatrix} \end{pmatrix}.$$

The variance matrix of (170) reads:

$$(171)\qquad Q_{\hat{y}_{\boldsymbol{B}_1}} = \sigma^2 \begin{pmatrix} \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix}.$$

Comparison of (170) and (171) with (166) and (167) shows that the results of the first phase for the first three observables is identical to the solution of (165). Also note that the last two observables do not get a correction in the first phase. Hence, they do not play a role in the first phase. It should be recognized however that this is a consequence of the assumption that $\underline{y}_4$ and $\underline{y}_5$ are not correlated with $\underline{y}_1$, $\underline{y}_2$ and $\underline{y}_3$. This is fortunately the case in most practical applications.

In order to obtain the final solution of (168) we have to solve in the *second phase* the partial model:

$$(172)\qquad (0\ 0\ 1\ 1\ 1)\,E\{\begin{pmatrix} \underline{\hat{y}}_1 \\ \underline{\hat{y}}_2 \\ \underline{\hat{y}}_3 \\ \underline{\hat{y}}_4 \\ \underline{\hat{y}}_5 \end{pmatrix}_{\boldsymbol{B}_1}\} = 0 \quad ; \quad \boldsymbol{D}\{\begin{pmatrix} \underline{\hat{y}}_1 \\ \underline{\hat{y}}_2 \\ \underline{\hat{y}}_3 \\ \underline{\hat{y}}_4 \\ \underline{\hat{y}}_5 \end{pmatrix}_{\boldsymbol{B}_1}\} = Q_{\hat{y}_{\boldsymbol{B}_1}}\ .$$

With (170) and (171) the solution of (172) reads:

$$\begin{pmatrix}\hat{\underline{y}}_1\\ \hat{\underline{y}}_2\\ \hat{\underline{y}}_3\\ \hat{\underline{y}}_4\\ \hat{\underline{y}}_5\end{pmatrix}=\left[\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{pmatrix}-\sigma^2\begin{pmatrix}-\frac{1}{3}\\-\frac{1}{3}\\\frac{2}{3}\\1\\1\end{pmatrix}\frac{3}{8}\sigma^{-2}(0\;0\;1\;1\;1)\right]\begin{pmatrix}\frac{1}{3}\begin{pmatrix}2\underline{y}_1-\underline{y}_2-\underline{y}_3\\-\underline{y}_1+2\underline{y}_2-\underline{y}_3\\-\underline{y}_1-\underline{y}_2+2\underline{y}_3\end{pmatrix}\\\begin{pmatrix}\underline{y}_4\\\underline{y}_5\end{pmatrix}\end{pmatrix}.$$

or:

$$\begin{pmatrix}\hat{\underline{y}}_1\\ \hat{\underline{y}}_2\\ \hat{\underline{y}}_3\\ \hat{\underline{y}}_4\\ \hat{\underline{y}}_5\end{pmatrix}=\frac{1}{8}\begin{pmatrix}5\underline{y}_1-3\underline{y}_2-2\underline{y}_3+\underline{y}_4+\underline{y}_5\\-3\underline{y}_1+5\underline{y}_2-2\underline{y}_3+\underline{y}_4+\underline{y}_5\\-2\underline{y}_1-2\underline{y}_2+4\underline{y}_3-2\underline{y}_4-2\underline{y}_5\\\underline{y}_1+\underline{y}_2-2\underline{y}_3+5\underline{y}_4-3\underline{y}_5\\\underline{y}_1+\underline{y}_2-2\underline{y}_3-3\underline{y}_4+5\underline{y}_5\end{pmatrix}. \tag{173}$$

It will be clear that if point 0 in figure 6.9 is taken as reference point with a height equal to zero, then $\hat{\underline{x}}_1=\hat{\underline{y}}_1$, $\hat{\underline{x}}_2=-\hat{\underline{y}}_3$ and $\hat{\underline{x}}_3=\hat{\underline{y}}_4$. Compare now (173) with (127), (128) and also (137).

There exists a close correspondence between the results of this section and the results of section 6.2. In section 6.2 we considered the partitioned model:

$$\underset{m\times 1}{E\{\underline{y}\}}=\underset{m\times n_1}{(\boldsymbol{A}_1}\vdots\underset{m\times n_2}{\boldsymbol{A}_2)}\underset{(n_1+n_2)\times 1}{\begin{pmatrix}x_1\\x_2\end{pmatrix}}\quad;\quad\underset{m\times m}{\boldsymbol{D}\{\underline{y}\}}=\underset{m\times m}{\boldsymbol{Q}_y}\;, \tag{174}$$

and the partial model:

$$\underset{m\times 1}{E\{\underline{y}\}}=\underset{m\times n_1}{\boldsymbol{A}_1}\;\underset{n_1\times 1}{x_1}\quad;\quad\underset{m\times m}{\boldsymbol{D}\{\underline{y}\}}=\underset{m\times m}{\boldsymbol{Q}_y}\;. \tag{175}$$

In the present section we considered the partitioned model:

$$\begin{pmatrix}\boldsymbol{B}_1^*\\\boldsymbol{B}_2^*\end{pmatrix}E\{\underline{y}\}=\begin{pmatrix}0\\0\end{pmatrix}\quad;\quad\boldsymbol{D}\{\underline{y}\}=\boldsymbol{Q}_y\;, \tag{176}$$

and the partial model:

$$\boldsymbol{B}_1^*E\{\underline{y}\}=0\quad;\quad\boldsymbol{D}\{\underline{y}\}=\boldsymbol{Q}_y\;. \tag{177}$$

Since model (175) follows from model (174) by setting x_2 equal to zero, the *redundancy* of model (175) is higher than that of (174). In fact the redundancy of (174) is m-n_1-n_2, whereas the redundancy of (175) is m-n_1. Since the redundancy of model (175) is higher than that of (174), more condition equations can be formed for model (175) than for model (174). Hence, one may consider (176) to be the equivalent of model (175), and model (177) to be the equivalent of model (174). With $A=(A_1 : A_2)$ and $B=(B_1 : B_2)$ this gives:

$$(178) \qquad \begin{cases} \boldsymbol{B}^{*}\boldsymbol{A}_1 = 0 & \text{for (176) and (175), and} \\ \boldsymbol{B}_1^{*}\boldsymbol{A} = 0 & \text{for (177) and (174).} \end{cases}$$

This implies that the solution of (176) is identical to the solution of (175), and that the solution of (177) is identical to the solution of (174). For the two models (174) and (175) we have the orthogonal decomposition:

$$(179) \qquad \boxed{\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{A}_1} + \boldsymbol{P}_{\bar{\boldsymbol{A}}_2}, \text{ with } \bar{\boldsymbol{A}}_2 = \boldsymbol{P}_{\boldsymbol{A}_1}^{\perp}\boldsymbol{A}_2} \qquad \text{(see (39))}$$

In a similar way we have for the two models (176) and (177) the orthogonal decomposition:

$$(180) \qquad \boxed{\boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}} = \boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}_1} + \boldsymbol{P}_{\boldsymbol{Q}_y\bar{\boldsymbol{B}}_2}, \text{ with } \boldsymbol{Q}_y\bar{\boldsymbol{B}}_2 = \boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}_1}^{\perp}\boldsymbol{Q}_y\boldsymbol{B}_2}$$

With (179) and (180) it is now possible to relate the various orthogonal projectors to each other. Clearly:

$$(181) \qquad \boxed{\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}_1}^{\perp} \text{ and } \boldsymbol{P}_{\boldsymbol{A}_1} = \boldsymbol{P}_{\boldsymbol{Q}_y\boldsymbol{B}}^{\perp}}$$

And with (179) and (180) this shows that:

$$(182) \qquad \boxed{\boldsymbol{P}_{\bar{\boldsymbol{A}}_2} = \boldsymbol{P}_{\boldsymbol{Q}_y\bar{\boldsymbol{B}}_2}}$$

The above results can now be used to show the correspondence between the solutions of the four models (174), (175), (176) and (177). An overview of this correspondence is given in table 6.7.

With the results of table 6.7 we can now also give an alternative proof of the *method of estimation in phases*. Since:

$$\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{A}_1} + \boldsymbol{P}_{\bar{\boldsymbol{A}}_2} \text{ and } \boldsymbol{P}_{\boldsymbol{A}_1}\boldsymbol{P}_{\bar{\boldsymbol{A}}_2} = 0 \ ,$$

it follows that:

$$\boldsymbol{P}_{\boldsymbol{A}_1} = \boldsymbol{P}_{\boldsymbol{A}_1}(\boldsymbol{P}_{\boldsymbol{A}_1} + \boldsymbol{P}_{\bar{\boldsymbol{A}}_2}) = \boldsymbol{P}_{\boldsymbol{A}_1}\boldsymbol{P}_{\boldsymbol{A}} \ .$$

$$\begin{pmatrix} \boldsymbol{B}_1^* \\ \boldsymbol{B}_2^* \end{pmatrix} E\{\underset{\sim}{y}\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \boldsymbol{B}^* \boldsymbol{A}_1 = 0 \qquad E\{\underset{\sim}{y}\} = \boldsymbol{A}_1 x_1$$

$\hat{\underset{\sim}{y}}_{\boldsymbol{B}} = \boldsymbol{P}_{Q_y \boldsymbol{B}}^{\perp} \underset{\sim}{y}$	$\hat{\underset{\sim}{y}}_{\boldsymbol{B}} = \hat{\underset{\sim}{y}}_{\boldsymbol{A}_1}$	$\hat{\underset{\sim}{y}}_{\boldsymbol{A}_1} = \boldsymbol{P}_{\boldsymbol{A}_1} \underset{\sim}{y}$
$\hat{\underset{\sim}{e}}_{\boldsymbol{B}} = \boldsymbol{P}_{Q_y \boldsymbol{B}} \underset{\sim}{y}$	$\hat{\underset{\sim}{e}}_{\boldsymbol{B}} = \hat{\underset{\sim}{e}}_{\boldsymbol{A}_1}$	$\hat{\underset{\sim}{e}}_{\boldsymbol{A}_1} = \boldsymbol{P}_{\boldsymbol{A}_1}^{\perp} \underset{\sim}{y}$
$\|\hat{\underset{\sim}{e}}_{\boldsymbol{B}}\|^2 = \|\boldsymbol{P}_{Q_y \boldsymbol{B}} \underset{\sim}{y}\|^2$	$\|\hat{\underset{\sim}{e}}_{\boldsymbol{B}}\|^2 = \|\hat{\underset{\sim}{e}}_{\boldsymbol{A}_1}\|^2$	$\|\hat{\underset{\sim}{e}}_{\boldsymbol{A}_1}\|^2 = \|\boldsymbol{P}_{\boldsymbol{A}_1}^{\perp} \underset{\sim}{y}\|^2$

$$\boldsymbol{B}_1^* E\{\underset{\sim}{y}\} = 0 \qquad \boldsymbol{B}_1^* \boldsymbol{A} = 0 \qquad E\{\underset{\sim}{y}\} = (\boldsymbol{A}_1 : \boldsymbol{A}_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$\hat{\underset{\sim}{y}}_{\boldsymbol{B}_1} = \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp} \underset{\sim}{y}$	$\hat{\underset{\sim}{y}}_{\boldsymbol{B}_1} = \hat{\underset{\sim}{y}}_{\boldsymbol{A}}$	$\hat{\underset{\sim}{y}}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{A}} \underset{\sim}{y}$
$\hat{\underset{\sim}{e}}_{\boldsymbol{B}_1} = \boldsymbol{P}_{Q_y \boldsymbol{B}_1} \underset{\sim}{y}$	$\hat{\underset{\sim}{e}}_{\boldsymbol{B}_1} = \hat{\underset{\sim}{e}}_{\boldsymbol{A}}$	$\hat{\underset{\sim}{e}}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \underset{\sim}{y}$
$\|\hat{\underset{\sim}{e}}_{\boldsymbol{B}_1}\|^2 = \|\boldsymbol{P}_{Q_y \boldsymbol{B}_1} \underset{\sim}{y}\|^2$	$\|\hat{\underset{\sim}{e}}_{\boldsymbol{B}_1}\|^2 = \|\hat{\underset{\sim}{e}}_{\boldsymbol{A}}\|^2$	$\|\hat{\underset{\sim}{e}}_{\boldsymbol{A}}\|^2 = \|\boldsymbol{P}_{\boldsymbol{A}}^{\perp} \underset{\sim}{y}\|^2$

$$\boldsymbol{P}_{Q_y \boldsymbol{B}} = \boldsymbol{P}_{Q_y \boldsymbol{B}_1} + \boldsymbol{P}_{Q_y \bar{\boldsymbol{B}}_2} \quad ; \quad Q_y \bar{\boldsymbol{B}}_2 = \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp} Q_y \boldsymbol{B}_2 \quad ; \quad \boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{A}_1} + \boldsymbol{P}_{\bar{\boldsymbol{A}}_2} \quad ; \quad \bar{\boldsymbol{A}}_2 = \boldsymbol{P}_{\boldsymbol{A}_1}^{\perp} \boldsymbol{A}_2$$
$$\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{Q_y \boldsymbol{B}_1}^{\perp} \quad ; \quad \boldsymbol{P}_{\boldsymbol{A}_1} = \boldsymbol{P}_{Q_y \boldsymbol{B}}^{\perp} \quad ; \quad \boldsymbol{P}_{\bar{\boldsymbol{A}}_2} = \boldsymbol{P}_{Q_y \bar{\boldsymbol{B}}_2}$$

$\hat{\underset{\sim}{y}}_{\boldsymbol{B}_1} = \hat{\underset{\sim}{y}}_{\boldsymbol{B}} + \boldsymbol{P}_{Q_y \bar{\boldsymbol{B}}_2} \underset{\sim}{y}$	$\hat{\underset{\sim}{y}}_{\boldsymbol{A}} = \hat{\underset{\sim}{y}}_{\boldsymbol{A}_1} + \boldsymbol{P}_{\bar{\boldsymbol{A}}_2} \underset{\sim}{y}$	
$\hat{\underset{\sim}{e}}_{\boldsymbol{B}_1} = \hat{\underset{\sim}{e}}_{\boldsymbol{B}} - \boldsymbol{P}_{Q_y \bar{\boldsymbol{B}}_2} \underset{\sim}{y}$	$\hat{\underset{\sim}{e}}_{\boldsymbol{A}} = \hat{\underset{\sim}{e}}_{\boldsymbol{A}_1} - \boldsymbol{P}_{\bar{\boldsymbol{A}}_2} \underset{\sim}{y}$	
$\|\hat{\underset{\sim}{e}}_{\boldsymbol{B}_1}\|^2 = \|\hat{\underset{\sim}{e}}_{\boldsymbol{B}}\|^2 - \|\boldsymbol{P}_{Q_y \bar{\boldsymbol{B}}_2} \underset{\sim}{y}\|^2$	$\|\hat{\underset{\sim}{e}}_{\boldsymbol{A}}\|^2 = \|\hat{\underset{\sim}{e}}_{\boldsymbol{A}_1}\|^2 - \|\boldsymbol{P}_{\bar{\boldsymbol{A}}_2} \underset{\sim}{y}\|^2$	

Table 6.7: Correspondence between partitioned condition equations and partitioned observation equations

Substitution of $P_{A_1} = P_A - P_{\bar{A}_2}$ in the right-hand side gives:

$$\boldsymbol{P}_{\boldsymbol{A}_1} = (\boldsymbol{P}_{\boldsymbol{A}} - \boldsymbol{P}_{\bar{\boldsymbol{A}}_2})\boldsymbol{P}_{\boldsymbol{A}} = (\boldsymbol{P}_{\bar{\boldsymbol{A}}_2}^{\perp} - \boldsymbol{P}_{\boldsymbol{A}}^{\perp})\boldsymbol{P}_{\boldsymbol{A}} \ .$$

But $P_A^{\perp} P_A = 0$. Hence:

(183) $$\boldsymbol{P}_{\boldsymbol{A}_1} = \boldsymbol{P}_{\bar{\boldsymbol{A}}_2}^{\perp} \boldsymbol{P}_{\boldsymbol{A}} \ .$$

With:

$$\boldsymbol{P}_{\boldsymbol{A}_1} = \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}}^{\perp} \ , \quad \boldsymbol{P}_{\bar{\boldsymbol{A}}_2}^{\perp} = \boldsymbol{P}_{\boldsymbol{Q}_y \bar{\boldsymbol{B}}_2}^{\perp} \ , \quad \boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \ ,$$

equation (183) becomes:

(184) $$\boxed{\boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}}^{\perp} = \boldsymbol{P}_{\boldsymbol{Q}_y \bar{\boldsymbol{B}}_2}^{\perp} \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp}}$$

The geometry of this relation is shown in figure 6.10.

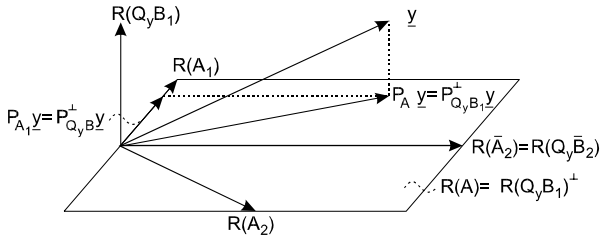


Figure 6.10: Geometry of $P_{Q_y B}^{\perp} = P_{Q_y \bar{B}_2}^{\perp} P_{Q_y B_1}^{\perp}$

Relation (184) already shows the two phases of the method of estimation in phases. Substitution of:

$$\boldsymbol{Q}_y \bar{\boldsymbol{B}}_2 = \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \boldsymbol{Q}_y \boldsymbol{B}_2 \qquad \text{(see (180)),}$$

into:

$$\boldsymbol{P}_{\boldsymbol{Q}_y \bar{\boldsymbol{B}}_2}^{\perp} \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} = [I - \boldsymbol{Q}_y \bar{\boldsymbol{B}}_2 (\bar{\boldsymbol{B}}_2^{*} \boldsymbol{Q}_y \bar{\boldsymbol{B}}_2)^{-1} \bar{\boldsymbol{B}}_2^{*}] \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \ ,$$

using:

$$\boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \boldsymbol{Q}_y \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp *} = \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \boldsymbol{Q}_y = \boldsymbol{Q}_y \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp *} \ ,$$

and:

$$\boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} = \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \ ,$$

gives:

(185) $$\boldsymbol{P}_{\boldsymbol{Q}_y \bar{\boldsymbol{B}}_2}^{\perp} \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} = [I - \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \boldsymbol{Q}_y \boldsymbol{B}_2 (\boldsymbol{B}_2^{*} \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \boldsymbol{Q}_y \boldsymbol{B}_2)^{-1} \boldsymbol{B}_2^{*}] \boldsymbol{P}_{\boldsymbol{Q}_y \boldsymbol{B}_1}^{\perp} \ .$$

Since the variance matrix of $\underset{\sim}{\hat{y}}_{B_1} = P^{\perp}_{Q_y B_1} \underset{\sim}{y}$ is given by:

$$Q_{\hat{y}_{B_1}} = P^{\perp}_{Q_y B_1} Q_y \ ,$$

it follows that (185) may be written as:

$$P^{\perp}_{Q_y \bar{B}_2} P^{\perp}_{Q_y B_1} \quad = \quad [I - Q_{\hat{y}_{B_1}} B_2 (B_2^* Q_{\hat{y}_{B_1}} B_2)^{-1} B_2^*] P^{\perp}_{Q_y B_1} \ .$$

And this shows that:

(186)
$$\boxed{P^{\perp}_{Q_y B} = [I - Q_{\hat{y}_{B_1}} B_2 (B_2^* Q_{\hat{y}_{B_1}} B_2)^{-1} B_2^*][I - Q_y B_1 (B_1^* Q_y B_1)^{-1} B_1^*]}$$

This concludes the proof of the method of estimation in phases.

7 Nonlinear models, linearization, iteration

7.1. The nonlinear A-model

7.1.1 Nonlinear observation equations

Up to this point the theoretical development 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 (1) of section 6.1, we are generally dealing with a nonlinear model of observation equations:

(1)
$$\boxed{E\{\underline{y}\} = \boldsymbol{A}(x) \;;\; \boldsymbol{D}\{\underline{y}\} = \boldsymbol{Q}_y}$$

where $A(.)$ is a nonlinear vector function from $\mathbb{R}^n$ into $\mathbb{R}^m$. The following definition makes clear when a vector function $A(.) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear or nonlinear.

Definition: The vector function $A(.) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is *linear* if and only if:

(2)
$$\begin{array}{ll} \text{a.} & \boldsymbol{A}(\alpha u) = \alpha \boldsymbol{A}(u),\ \forall \alpha \in \mathbb{R},\ u \in \mathbb{R}^n \\ \text{b.} & \boldsymbol{A}(u+v) = \boldsymbol{A}(u) + \boldsymbol{A}(v),\ \forall u,\ v \in \mathbb{R}^n \end{array} .$$

The vector function $A(.)$ is said to be *nonlinear* if it does not satisfy (2).

Example 1

Verify yourself that the vector function:

$$\boldsymbol{A}(x) = \begin{pmatrix} x_1 + 2x_2 \\ 4x_2 \\ 3x_1 + x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 4 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

is linear and from $\mathbb{R}^2$ into $\mathbb{R}^3$.

Example 2

Verify yourself that the vector function:

$$A(x) = \begin{pmatrix} x_1 + (x_2)^2 \\ x_2 \\ (x_1^2) + \tan x_2 \end{pmatrix},$$

is nonlinear and from $\mathbb{R}^2$ into $\mathbb{R}^3$.

Example 3

Consider the levelling network of figure 7.1.

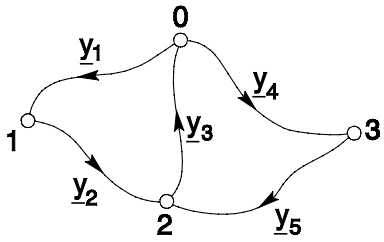


Figure 7.1: Levelling network

Point 0 is taken as reference point. Its height is assumed to be known and equal to zero, that is, x_0=0. The observed height differences are assumed to be uncorrelated and to have equal variances σ^2. The corresponding model of observation equations reads then:

(3)
$$\underbrace{E\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\right\}}_{\underline{y}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{A(x)} ; \quad D\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\right\} = \sigma^2 I_5.$$

Verify yourself that the vector function $A(x)$ of (3) is linear and from $\mathbb{R}^3$ into $\mathbb{R}^5$.

Example 4

Consider the configuration of figure 7.2a. The Cartesian x,y coordinates of the three points 1, 2 and 3 are known and the two Cartesian coordinates x_4 and y_4 of point 4 are unknown. The observables consist of the three distance variates $\underline{l}_{14}$, $\underline{l}_{24}$ and $\underline{l}_{34}$. These observables are assumed to be uncorrelated and to have equal variances σ^2. Since distance and coordinates are related as (see figure 7.2b):

$$l_{ij} = \left[(x_j - x_i)^2 + (y_j - y_i)^2\right]^{1/2} = \left(x_{ij}^2 + y_{ij}^2\right)^{1/2},$$

the model of observation equations for the configuration of figure 7.2a reads:

$$E\left\{\begin{pmatrix} \underline{l}_{14} \\ \underline{l}_{24} \\ \underline{l}_{34} \end{pmatrix}\right\} = \begin{pmatrix} (x_{14}^2 + y_{14}^2)^{1/2} \\ (x_{24}^2 + y_{24}^2)^{1/2} \\ (x_{34}^2 + y_{34}^2)^{1/2} \end{pmatrix} ; \quad D\left\{\begin{pmatrix} \underline{l}_{14} \\ \underline{l}_{24} \\ \underline{l}_{34} \end{pmatrix}\right\} = \sigma^2 I_3. \tag{4}$$

$$\underbrace{\underline{y}} \qquad \underbrace{A(x)}$$

This model consists of 3 nonlinear observation equations in the 2 unknown parameters x_4 and y_4. Hence, the vector function $A(x)$ of (4) is nonlinear and from $\mathbb{R}^2$ into $\mathbb{R}^3$.

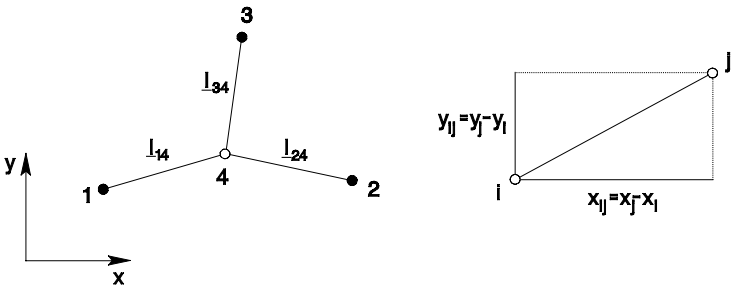
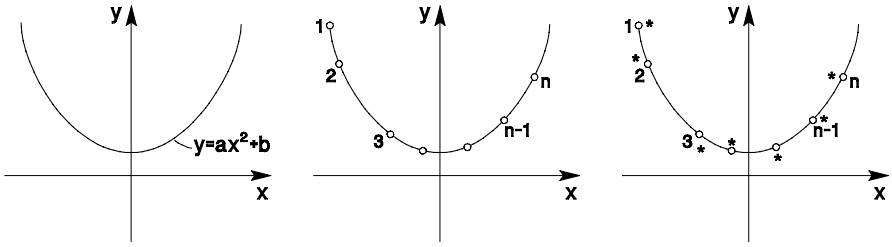


Figure 7.2: Distance resection

Example 5

Consider the parabola of figure 7.3a.



(a) (b) (c)

Figure 7.3
(a) The parabola $y=ax^2+b$
(b) The points (o) on the parabola that need to be digitized.
(c) The sample values ($*$) of the digitized points

We are asked to determine the values of the parameters a and b of the parabola $y=ax^2+b$. In order to solve this problem we decide to measure the x,y coordinates of an n-number of points on the parabola. The coordinates of the points 1,2, .., n (see figure 7.3b) are measured with a digitizer. These coordinate observables $(\underline{x}_i, \underline{y}_i)$, $i=1,...,n$, are assumed to be uncorrelated and to have equal variances σ^2. The observed or sample values of the coordinates are shown in figure 7.3c. The model of observation equations for this problem reads:

(5)
$$E\{\underbrace{\begin{pmatrix}\underline{x}_1\\ \vdots\\ \underline{x}_n\\ \underline{y}_1\\ \vdots\\ \underline{y}_n\end{pmatrix}}_{\underline{y}}\} = \underbrace{\begin{pmatrix}x_1\\ \vdots\\ x_n\\ ax_1^2+b\\ \vdots\\ ax_n^2+b\end{pmatrix}}_{A(x)} \; ; \; D\{\begin{pmatrix}\underline{x}_1\\ \vdots\\ \underline{x}_n\\ \underline{y}_1\\ \vdots\\ \underline{y}_n\end{pmatrix}\} = \sigma^2 I_{2n}.$$

This model consists of $2n$ observation equations in $(n+2)$ unknown parameters, namely $x_1,...,x_n$ and a,b. Note that the first n observation equations are linear, but that the second set of n observation equations are nonlinear. Hence, the vector function $A(x)$, with $x=(x_1,...,x_n,a,b)^*$, of (5) is nonlinear and from $\mathbb{R}^{n+2}$ into $\mathbb{R}^{2n}$.

Example 6

Consider figure 7.4. It shows an equilateral triangle. The position, orientation and size of this triangle are unknown. In order to determine these parameters, the *x, y* coordinates of the three points 1, 2 and 3 are observed.

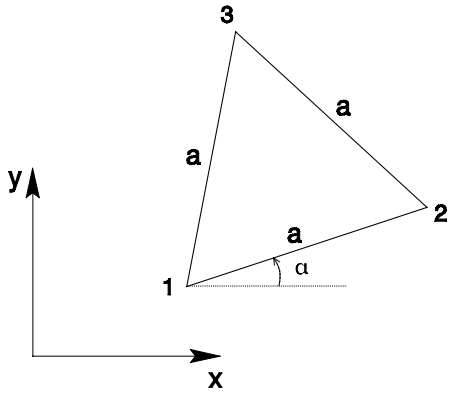


Figure 7.4: An equilateral triangle

The coordinate observables $(\underline{x}_i, \underline{y}_i)$, $i = 1, 2, 3$, are assumed to be uncorrelated and to have equal variances σ^2. The corresponding model of observation equations reads:

$$(6) \qquad \underbrace{E\{\begin{pmatrix}\underline{x}_1\\ \underline{y}_1\\ \underline{x}_2\\ \underline{y}_2\\ \underline{x}_3\\ \underline{y}_3\end{pmatrix}\}}_{\underline{y}} = \underbrace{\begin{pmatrix}x_1\\ y_1\\ x_1 + a\cos\alpha\\ y_1 + a\sin\alpha\\ x_1 + a\cos(\alpha + \frac{1}{3}\pi)\\ y_1 + a\sin(\alpha + \frac{1}{3}\pi)\end{pmatrix}}_{A(x)} ; \quad D\{\begin{pmatrix}\underline{x}_1\\ \underline{y}_1\\ \underline{x}_2\\ \underline{y}_2\\ \underline{x}_3\\ \underline{y}_3\end{pmatrix}\} = \sigma^2 I_6.$$

This model consists of 6 observation equations in 4 unknown parameters, namely x_1, y_1, a and α. The vector function $A(x)$ of (6), with $x=(x_1, y_1, a, \alpha)^*$, is nonlinear and from $\mathbb{R}^4$ into $\mathbb{R}^6$.

7.1.2 Linearization: Taylor's theorem

In the previous section a number of examples were given of nonlinear A-models. We do know how to compute least-squares estimators in case of a linear A-model. But what about a nonlinear A-model? How should we compute the least-squares estimators if the model of observation equations is nonlinear? For the majority of non-linear problems the solution is to approximate the originally nonlinear A-model with a *linear(ized)* one. This approximation is made possible with the theorem of Taylor (see also appendix A). This section is therefore devoted to a discussion of Taylor's theorem.

Taylor's theorem for f: $\mathbb{R} \rightarrow \mathbb{R}$

Let f: $\mathbb{R} \rightarrow \mathbb{R}$ be a function of which all its qth-order derivatives exist and for which $\frac{d^q}{dx^q} f(x)$ is continuous. Then a scalar θ between x and x^0 exists such that:

(7)

$$f(x) = f(x^0) + \frac{d}{dx} f(x^0)\Delta x + \frac{1}{2!}\frac{d^2}{dx^2} f(x^0)\Delta x^2 + \ldots + \frac{1}{(q-1)!}\frac{d^{q-1}}{dx^{q-1}} f(x^0)\Delta x^{q-1} + R_q(\theta, \Delta x),$$

with $\Delta x = x - x^o$ and with the remainder:

(8)

$$R_q(\theta, \Delta x) = \frac{1}{q!}\frac{d^q}{dx^q} f(\theta)\ \Delta x^q$$

It follows from (7) and (8) that for $q = 2$ we have:

(9)

$$f(x) = f(x^0) + \left(\frac{d}{dx} f(x^0) + \frac{1}{2}\frac{d^2}{dx^2} f(\theta)\Delta x \right) \Delta x.$$

Note that the second term within the brackets can be made arbitrarily small with respect to the first term within the brackets, by letting x^0 approach x. If this second term is negligible with respect to the first term, we may decide to ignore it, and approximate (9) as:

(10)

$$f(x) \doteq f(x^0) + \frac{d}{dx} f(x^0)\Delta x$$

In this case we speak of a *linear approximation* of $f(x)$ or a *linearization of* $f(x)$ *at* x^0. A geometric interpretation of (10) is given in figure 7.5.

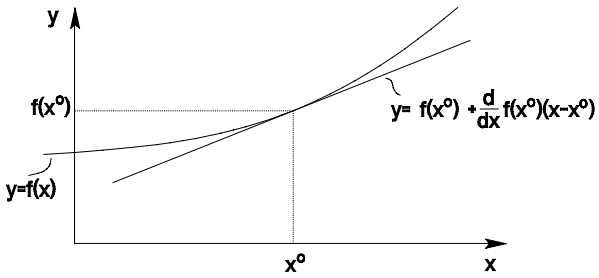


Figure 7.5: The curve $y = f(x)$ and its tangent line $y = f(x^0) + \frac{d}{dx} f(x^0)(x - x^0)$ at the point $(x^0, f(x^0))$

Example 7

The linearization of $f(x) = \tan x$ at x^0 reads:

$$f(x) = \tan x^0 + \frac{1}{\cos^2 x^0}(x - x^0) \quad .$$

Example 8

The linearization of $f(x) = ax^2 + b$ at x^0 reads:

$$f(x) = [a(x^0)^2 + b] + (2ax^0)(x - x^0) \quad .$$

We will now consider functions of more than one variable. A function $f(x_1,...,x_n)$ of the variables $x_1,...,x_n$ will be denoted as $f(x)$ with the n-vector $x=(x_1,...,x_n)^*$. The first-order partial derivatives of $f(x), \frac{\partial}{\partial x_\alpha} f(x), \alpha = 1,...,n,$ will be denoted as $\partial_\alpha f(x)$, $\alpha=1,...,n$. Similarly, we denote the second-order partial derivatives of $f(x)$, $\frac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x)$, as $\partial^2_{\alpha\beta} f(x)$, α, $\beta=1,...,n$. And so on.

Taylor's Theorem for f: $\mathbb{R}^n \to \mathbb{R}$:

Let f: $\mathbb{R}^n \to \mathbb{R}$ be a function of which all its qth-order partial derivatives exist and for which $\partial^q_{\alpha_1 \cdots \alpha_q} f(x)$ is continuous. Let $\Delta x = x - x^0$ and $\theta = x^0 + t(x - x^0)$ with $t \in \mathbb{R}$. Then a scalar $t \in (0,1)$ exists such that:

(11)
$$\boxed{\begin{aligned} f(x) &= f(x^0) + \sum_{\alpha=1}^{n} \partial_\alpha f(x^0)\Delta x_\alpha + \frac{1}{2!}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \partial^2_{\alpha\beta} f(x^0)\Delta x_\alpha \Delta x_\beta + \cdots \\ &\cdots + \frac{1}{(q-1)!}\sum_{\alpha_1=1}^{n}\cdots\sum_{\alpha_{q-1}=1}^{n} \partial^{q-1}_{\alpha_1\cdots\alpha_{q-1}} f(x^0)\Delta x_{\alpha_1}\cdots\Delta x_{\alpha_{q-1}} + \boldsymbol{R}_q(\theta,\Delta x) \end{aligned}}$$

with the remainder:

(12)
$$\boxed{\boldsymbol{R}_q(\theta,\Delta x) = \frac{1}{q!}\sum_{\alpha_1=1}^{n}\cdots\sum_{\alpha_q=1}^{n} \partial^q_{\alpha_1\cdots\alpha_q} f(\theta)\Delta x_{\alpha_1}\cdots\Delta x_{\alpha_q}}$$

See appendix A for a proof of this theorem. Note that (11) and (12) reduce to (7) and (8) for $n = 1$.

For the case $q = 2$, it follows from (11) and (12) that:

(13)
$$f(x) = f(x^0) + \sum_{\alpha=1}^{n} \partial_\alpha f(x^0)\Delta x_\alpha + \frac{1}{2}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \partial^2_{\alpha\beta} f(\theta)\Delta x_\alpha \Delta x_\beta .$$

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

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

then equation (13) may be written in the more compact matrix-vector form as:

$$f(x) = f(x^0) + \partial_x f(x^0)^* \Delta x + \frac{1}{2}\Delta x^* \partial^2_{xx} f(\theta)\Delta x ,$$

or as:

(14)
$$f(x) = f(x^0) + \left(\partial_x f(x^0) + \frac{1}{2}\partial^2_{xx} f(\theta)\Delta x\right)^* \Delta x.$$

Note that this equation reduces to (9) in case $n = 1$. Equation (14) shows that a nonlinear function f: $\mathbb{R}^n \to \mathbb{R}$ can be written as the 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 of (14) 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 x^0 close enough to x. Now assume that x^0 is indeed chosen close enough to x, such that the second-order remainder can be neglected. Then, instead of (14) we may write to a sufficient approximation:

(15) $$\boxed{f(x) \doteq f(x^0) + \partial_x f(x^0)^* \Delta x}$$

Hence, if x^0 is sufficiently close to x, the *nonlinear* function f: $\mathbb{R}^n \to \mathbb{R}$ 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 called the *linearization of $f(x)$ at x^0*. Note that (15) is the n-dimensional generalization of (10).

Example 9

The linearization of $f(x_1, x_2) = (x_1^2 + x_2^2)^{1/2}$ at (x_1^0, x_2^0) reads:

$$\underbrace{f(x_1, x_2)}_{f(x)} = \underbrace{((x_1^0)^2+(x_2^0)^2)^{1/2}}_{f(x^0)} + \underbrace{\left(\frac{x_1^0}{((x_1^0)^2+(x_2^0)^2)^{1/2}}, \frac{x_2^0}{((x_1^0)^2+(x_2^0)^2)^{1/2}}\right)}_{\partial_x f(x^0)^*} \underbrace{\begin{pmatrix}\Delta x_1\\ \Delta x_2\end{pmatrix}}_{\Delta x}$$

Example 10

The linearization of $f(x_1,x_2) = x_1^2 + \sin x_2$ at (x_1^0,x_2^0) reads:

$$\underbrace{f(x_1,x_2)}_{f(x)} = \underbrace{\left((x_1^0)^2+\sin x_2^0\right)}_{f(x^0)} + \underbrace{(2x_1^0, \cos x_2^0)}_{\partial_x f(x^0)^*} \underbrace{\begin{pmatrix}\Delta x_1\\ \Delta x_2\end{pmatrix}}_{\Delta x}$$

7.1.3 The linearized A-model

In the previous section we have seen how Taylor's theorem can be applied to obtain a linear approximation of a nonlinear function. In this section we will see how the results of the previous section can be used to approximate the originally nonlinear A-model:

$$\boxed{E\{\underline{y}\} = \boldsymbol{A}(x);\ \boldsymbol{D}\{\underline{y}\} = \boldsymbol{Q}_y} \tag{16}$$

with its *linearized* version. Consider the nonlinear vector function $A(.)$: $\mathbb{R}^n \rightarrow \mathbb{R}^m$:

$$\boldsymbol{A}(x) = \begin{pmatrix} a_1(x) \\ \vdots \\ a_m(x) \end{pmatrix}. \tag{17}$$

Each function $a_i(x)$, i=1,...,m, of (17) is a function from $\mathbb{R}^n$ into $\mathbb{R}$. Hence, each function $a_i(x)$, i=1,...,m may be linearized according to (15). This gives:

$$\underset{m\times 1}{\boldsymbol{A}(x)} = \underset{m\times 1}{\begin{pmatrix} a_1(x^0) \\ \vdots \\ a_m(x^0) \end{pmatrix}} + \underset{m\times n}{\begin{pmatrix} \partial_x a_1(x^0)^* \\ \vdots \\ \partial_x a_m(x^0)^* \end{pmatrix}} \underset{n\times 1}{\Delta x}. \tag{18}$$

If we denote the $m \times n$ matrix of (18) by $\partial_x A(x^0)$ and substitute (18) into (16) we get:

$$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times 1}{\boldsymbol{A}(x^0)} + \underset{m\times n}{\partial_x \boldsymbol{A}(x^0)} \underset{n\times 1}{\Delta x}\ ;\ \underset{m\times m}{\boldsymbol{D}\{\underline{y}\}} = \underset{m\times m}{\boldsymbol{Q}_y}. \tag{19}$$

If we bring the constant m-vector $A(x^0)$ of (19) to the left hand side of the equation and define $\Delta\underline{y}=\underline{y}$ - $A(x^0)$ we finally obtain our linearized model of observation equations:

$$\boxed{E\{\Delta\underline{y}\} = \partial_x \boldsymbol{A}(x^o)\Delta x\ ;\ \boldsymbol{D}\{\Delta\underline{y}\} = \boldsymbol{Q}_y} \tag{20}$$

This is the *linearized A-model*. Compare (20) with (16) and with (1) of subsection 7.1.1. Note when comparing (20) with (1) that in the linearized A-model $\Delta\underline{y}$ takes the place of $\underline{y}$, $\partial_x A(x^0)$ takes the place of A and Δx takes the place of x. Since the linearized A-model is linear in $\Delta x=x-x^0$, our standard formulae for least-squares estimation can be applied again. This gives for the least-squares estimator $\underline{\hat{x}}=x^0+\Delta\underline{\hat{x}}$

$$\underline{\hat{x}} = x^0 + \left(\partial_x \boldsymbol{A}(x^0)^* \boldsymbol{Q}_y^{-1} \partial_x \boldsymbol{A}(x^0)\right)^{-1}\left[\partial_x \boldsymbol{A}(x^0)^* \boldsymbol{Q}_y^{-1}(\underline{y} - \boldsymbol{A}(x^0))\right]. \tag{21}$$

And application of the propagation law of variances to (21) gives:

$$\boldsymbol{Q}_{\hat{x}} = \left(\partial_x \boldsymbol{A}(x^0)^* \boldsymbol{Q}_y^{-1} \partial_x \boldsymbol{A}(x^0)\right)^{-1}. \tag{22}$$

It will be clear that the above results, (21) and (22), are approximate in the sense that the second-order remainder has been neglected in (20). 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 (e.g. unbiasedness, minimal variance) hold to a sufficient degree. A summary of the *linearized least-squares estimators* is given in table 7.1:

The non linear A-model

$$E\{\underline{y}\} = \boldsymbol{A}(x) \;;\; D\{\underline{y}\} = \boldsymbol{Q}_y \;;\; \boldsymbol{A}(.): \mathbb{R}^n \rightarrow \mathbb{R}^m$$

The linearized A-model

$$\underset{m\times 1}{E\{\Delta \underline{y}\}} = \underset{m\times n}{\partial_x \boldsymbol{A}(x^0)} \underset{n\times 1}{\Delta x} \;;\; D\{\Delta \underline{y}\} = \boldsymbol{Q}_y$$

Least-squares estimators

$$\begin{aligned} \underline{\hat{x}} &= x^0 + \left(\partial_x \boldsymbol{A}(x^0)^* \boldsymbol{Q}_y^{-1} \partial_x \boldsymbol{A}(x^0)\right)^{-1} \left(\partial_x \boldsymbol{A}(x^0)^* \boldsymbol{Q}_y^{-1} (\underline{y} - \boldsymbol{A}(x^0))\right) \\ \underline{\hat{y}} &= \boldsymbol{A}(\hat{x}) \;;\; \underline{\hat{e}} = \underline{y} - \underline{\hat{y}} \end{aligned}$$

Variances

$$\begin{aligned} \boldsymbol{Q}_{\hat{x}} &= \left(\partial_x \boldsymbol{A}(x^0)^* \boldsymbol{Q}_y^{-1} \partial_x \boldsymbol{A}(x^0)\right)^{-1} \\ \boldsymbol{Q}_{\hat{y}} &= \partial_x \boldsymbol{A}(x^0) \boldsymbol{Q}_{\hat{x}} \partial_x \boldsymbol{A}(x^0)^* \\ \boldsymbol{Q}_{\hat{e}} &= \boldsymbol{Q}_y - \boldsymbol{Q}_{\hat{y}} \end{aligned}$$

Table 7.1: Linearized least-squares estimation

Example 11

Consider the *nonlinear* A-model (4) of example 4:

$$E\{\underbrace{\begin{pmatrix} \underline{l}_{14} \\ \underline{l}_{24} \\ \underline{l}_{34} \end{pmatrix}}_{\underline{y}}\} = \underbrace{\begin{pmatrix} (x_{14}^2 + y_{14}^2)^{1/2} \\ (x_{24}^2 + y_{24}^2)^{1/2} \\ (x_{34}^2 + y_{34}^2)^{1/2} \end{pmatrix}}_{\boldsymbol{A}(x)} \quad ; \quad D\{\begin{pmatrix} \underline{l}_{14} \\ \underline{l}_{24} \\ \underline{l}_{34} \end{pmatrix}\} = \sigma^2 I_3. \tag{23}$$

The corresponding *linearized* A-model reads:

$$E\{\underbrace{\begin{pmatrix} \Delta\underline{l}_{14} \\ \Delta\underline{l}_{24} \\ \Delta\underline{l}_{34} \end{pmatrix}}_{\Delta\underline{y}}\} = \underbrace{\begin{pmatrix} x_{14}^0/l_{14}^0 & y_{14}^0/l_{14}^0 \\ x_{24}^0/l_{24}^0 & y_{24}^0/l_{24}^0 \\ x_{34}^0/l_{34}^0 & y_{34}^0/l_{34}^0 \end{pmatrix}}_{\partial_x \boldsymbol{A}(x^0)} \underbrace{\begin{pmatrix} \Delta x_4 \\ \Delta y_4 \end{pmatrix}}_{\Delta x} \quad ; \quad D\{\begin{pmatrix} \Delta\underline{l}_{14} \\ \Delta\underline{l}_{24} \\ \Delta\underline{l}_{34} \end{pmatrix}\} = \sigma^2 I_3. \tag{24}$$

with $x_{i4}^0 = x_4^0 - x_i$, $y_{i4}^0 = y_4^0 - y_i$, $l_{i4}^0 = \left((x_{i4})^2 + (y_{i4})^2\right)^{1/2}$, $i = 1,2,3$.

Up to this point nothing has been said about how one can obtain the approximate values of the parameters, which are needed to perform the linearization. In the majority of geodetic applications one can compute these approximate values of the parameters directly from the observations. In the above problem for instance, the approximate coordinates of point 4, x_4^0 and y_4^0, can be computed from the two observed distances l_{14}, l_{24} and the known coordinates of the two points 1 and 2, x_1, y_1, x_2 and y_2.

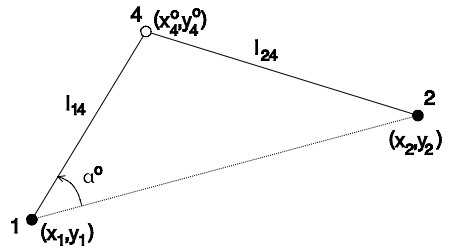


Figure 7.6: Computation of the approximate coordinates x_4^0 and y_4^0

The approximate angle α^o (see figure 7.6) satisfies the cosine-rule:

$$l_{24}^2 = l_{14}^2 + (x_{12}^2 + y_{12}^2) - 2l_{14}(x_{12}^2 + y_{12}^2)^{\frac{1}{2}} \cos\alpha^0.$$

Hence:

$$\alpha^0 = \arccos\left(\frac{l_{14}^2 + (x_{12}^2 + y_{12}^2) - l_{24}^2}{2l_{14}(x_{12}^2 + y_{12}^2)^{\frac{1}{2}}} \right). \tag{25}$$

With this result the approximate coordinates of point 4 can be computed as:

$$\begin{pmatrix} x_4^0 \\ y_4^0 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \frac{l_{14}}{(x_{12}^2 + y_{12}^2)^{\frac{1}{2}}} \begin{pmatrix} \cos\alpha^0 & -\sin\alpha^0 \\ \sin\alpha^0 & \cos\alpha^0 \end{pmatrix} \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix}. \tag{26}$$

These approximate values x_4^0, y_4^0 can now be used to linearize (23) to obtain (24). Of course, there is still an ambiguity in the above computation. It is namely not clear whether the angle α^0 should be counted clockwise or counter clockwise. In expression (26) it has been assumed that α^0 should be counted counter clockwise. In practical applications this ambiguity usually does not occur, since one usually knows whether point 4 lies left or right from the line connecting the two points 1 and 2.

Example 12

Consider the *nonlinear* A-model (6) of example 6:

$$E\{\underbrace{\begin{pmatrix} \underline{x}_1 \\ \underline{y}_1 \\ \underline{x}_2 \\ \underline{y}_2 \\ \underline{x}_3 \\ \underline{y}_3 \end{pmatrix}}_{\underline{y}}\} = \underbrace{\begin{pmatrix} x_1 \\ y_1 \\ x_1 + a\cos\alpha \\ y_1 + a\sin\alpha \\ x_1 + a\cos(\alpha + \frac{1}{3}\pi) \\ y_1 + a\sin(\alpha + \frac{1}{3}\pi) \end{pmatrix}}_{A(x)} ; D\{\begin{pmatrix} \underline{x}_1 \\ \underline{y}_1 \\ \underline{x}_2 \\ \underline{y}_2 \\ \underline{x}_3 \\ \underline{y}_3 \end{pmatrix}\} = \sigma^2 I_6 \tag{27}$$

The corresponding *linearized* A-model reads:

$$E\{\underbrace{\begin{pmatrix}\Delta\underline{x}_1\\ \Delta\underline{y}_1\\ \Delta\underline{x}_2\\ \Delta\underline{y}_2\\ \Delta\underline{x}_3\\ \Delta\underline{y}_3\end{pmatrix}}_{\Delta\underline{y}}\} = \underbrace{\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 1 & 0 & \cos\alpha^0 & -a^0\sin\alpha^0\\ 0 & 1 & \sin\alpha^0 & a^0\cos\alpha^0\\ 1 & 0 & \cos(\alpha^0+\frac{1}{3}\pi) & -a^0\sin(\alpha^0+\frac{1}{3}\pi)\\ 0 & 1 & \sin(\alpha^0+\frac{1}{3}\pi) & a^0\cos(\alpha^0+\frac{1}{3}\pi)\end{pmatrix}}_{\partial_x A(x^0)} \underbrace{\begin{pmatrix}\Delta x_1\\ \Delta y_1\\ \Delta a\\ \Delta\alpha\end{pmatrix}}_{\Delta x} ; \quad D\{\begin{pmatrix}\Delta\underline{x}_1\\ \Delta\underline{y}_1\\ \Delta\underline{x}_2\\ \Delta\underline{y}_2\\ \Delta\underline{x}_3\\ \Delta\underline{y}_3\end{pmatrix}\} = \sigma^2 I_6$$

(28)

The approximate values x_1^0, y_1^0, a^0 and α^0 of the four parameters can be computed from the observed coordinates as follows:

$$\begin{aligned} x_1^0 &= x_1 \;,\; a^0 = \left((x_2-x_1)^2+(y_2-y_1)^2\right)^{1/2} \\ y_1^0 &= y_1 \quad \alpha^0 = \arccos\frac{x_2-x_1}{\left((x_2-x_1)^2+(y_2-y_1)^2\right)^{1/2}} \end{aligned} \quad . \tag{29}$$

7.1.4 Least-squares iteration

Up to this point it was assumed that x^0 was a good enough approximation such that the second-order remainder could be neglected. If this is not the case however, then $\underline{\hat{x}}$ as computed by (21) is not the least-squares estimator and hence an unacceptable error is made. In order to repair this situation, we need to improve upon the approximation x^0. It seems reasonable to expect that the estimate:

$$x^1 = x^0 + (\partial_x A(x^0)^* Q_y^{-1} \partial_x A(x^0))^{-1} \partial_x A(x^0)^* Q_y^{-1} (y - A(x^0))$$

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

$$x^2 = x^1 + (\partial_x A(x^1)^* Q_y^{-1} \partial_x A(x^1))^{-1} \partial_x A(x^1)^* Q_y^{-1} (y - A(x^1)).$$

Now x^2 can be expected to be a better approximation than x^1. But this implies that a further improvement can be expected if we replace x^1 by x^2 in the linearization of the nonlinear A-

model. The recomputed linearized least-squares estimate, based on the approximation x^2 reads then:

$$x^3 = x^2 + (\partial_x \boldsymbol{A}(x^2)^* \boldsymbol{Q}_y^{-1} \partial_x \boldsymbol{A}(x^2))^1 \partial_x \boldsymbol{A}(x^2)^* \boldsymbol{Q}_y^{-1} (y - \boldsymbol{A}(x^2)) \quad .$$

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 process 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 the table on the following page.

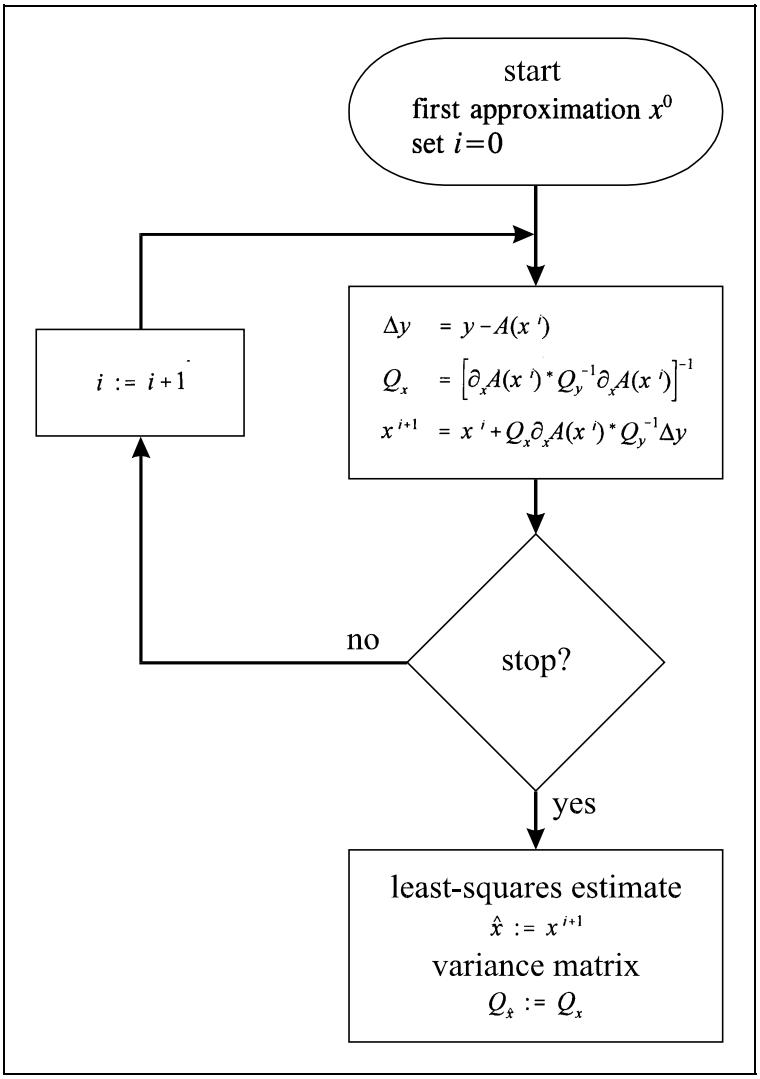


Table 7.2: Least-squares iteration

Example 13

Consider the *nonlinear* A-model:

$$(30)\qquad \underbrace{E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix}\}}_{\underline{y}} = \underbrace{\begin{pmatrix} x^2 \\ x \end{pmatrix}}_{\boldsymbol{A}(x)} \; ; \; \underbrace{\boldsymbol{D}\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix}\} = \sigma^2 I_2}_{Q_y}$$

It consists of two observation equations in one unknown parameter. Hence, $A(.)$: $\mathbb{R} \rightarrow \mathbb{R}^2$. The corresponding *linearized* A-model reads:

$$(31)\qquad \underbrace{E\{\begin{pmatrix} \Delta\underline{y}_1 \\ \Delta\underline{y}_2 \end{pmatrix}\}}_{\Delta\underline{y}} = \underbrace{\begin{pmatrix} 2x^0 \\ 1 \end{pmatrix}\Delta x}_{\partial_x \boldsymbol{A}(x^0)} \; ; \; \underbrace{\boldsymbol{D}\{\begin{pmatrix} \Delta\underline{y}_1 \\ \Delta\underline{y}_2 \end{pmatrix}\} = \sigma^2 I_2}_{Q_y}$$

Let us assume that we are asked to solve model (30). The two observations are given as:

$$(32)\qquad y_1 = 4.1 \; , \; y_2 = 1.9 .$$

In order to find the least-squares estimate $\hat{x}$ of x, we need the linearized model (31), a first approximation x^0, and the iteration scheme:

$$(33)\qquad x^{i+1} = x^i + (\partial_x \boldsymbol{A}(x^i)^* Q_y^{-1} \partial_x \boldsymbol{A}(x^i))^{-1} \partial_x \boldsymbol{A}(x^i)^* Q_y^{-1} (y - \boldsymbol{A}(x^i)), \text{ for } i=0,1,2,\ldots .$$

In our case, expression (33) becomes:

$$(34)\qquad x^{i+1} = x^i + (1 + 4(x^i)^2)^{-1}\left[2x^i(y - (x^i)^2) + (y_2 - x^i)\right]$$

As a first approximation x^0 we take y_2. Thus:

$$(35)\qquad x^0 = y_2 = 1.9 .$$

With (32), (35) and (34) we are now able to compute x^{i+1} for $i = 0,1,2, \ldots$. The results of this least squares iteration process are given in the following table:

iteration step i	solution x^i
0	1.9
1	2.0205959
2	2.0176464
3	2.0176324
4	2.0176344
5	2.0176344

Note that the solution does not change much after iteration step $i = 2$. Hence, depending on the required *numerical* precision, any of these solutions can be taken as the least-squares estimate $\hat{x}$. For instance, if the solution is required to have 5 significant digits, then the least-squares solution can be taken as $\hat{x} = 2.01763$.

7.2 The nonlinear B-model and its linearization

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

(36) $$\boxed{\boldsymbol{B}^*(E\{\underline{y}\}) = 0 \;;\; \boldsymbol{D}\{\underline{y}\} = \boldsymbol{Q}_y}\ ,$$

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

(37) $$\boxed{\boldsymbol{B}^*(\boldsymbol{A}(x)) = 0\ ,\ \forall x \in \mathbb{R}^n}\ .$$

This is the *nonlinear generalization* of (7) of section 3.1. If we take the partial derivative with respect to x of (37) and apply the chain rule (kettingregel), we get:

(38) $$\boxed{\left(\partial_y \boldsymbol{B}(y^0)\right)^* \left(\partial_x \boldsymbol{A}(x^0)\right) = 0, \quad y^0 = \boldsymbol{A}(x^0).}$$

This is the *linearized* version of (37). Compare (38) with (7) of section 3.1. With (38) we are now in the position to construct the linearized B-model from the linearized A-model (20). Pre-multiplication of (20) with the matrix $(\partial_y B(y^o))^*$ gives together with (38) the result:

(39) $$\boxed{\left(\partial_y \boldsymbol{B}(y^0)\right)^* E\{\underline{\Delta y}\} = 0 \;;\; D\{\underline{\Delta y}\} = \boldsymbol{Q}_y.}$$

This is the *linearized B-model.* With (39) we are now in the position again to apply our standard least-squares estimation formulae.

As with the nonlinear A-model, also the solution of the nonlinear B-model needs an *iteration process.* But since the iteration process needed for solving the nonlinear B-model is quite involved, it will not be discussed here.

Example 14

Consider the *nonlinear* A-model (5) of example 5:

(40) $$E\{\underbrace{\begin{pmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \\ \underline{y}_1 \\ \vdots \\ \underline{y}_n \end{pmatrix}}_{\underline{y}}\} = \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ \alpha x_1^2 + \beta \\ \vdots \\ \alpha x_n^2 + \beta \end{pmatrix}}_{\boldsymbol{A}(x)} \;;\; D\{\begin{pmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \\ \underline{y}_1 \\ \vdots \\ \underline{y}_n \end{pmatrix}\} = \sigma^2 I_{2n}.$$

This model consists of $2n$ observation equations in $(n+2)$ unknowns. Hence, the *redundancy* equals $2n-(n+2) = n-2$. Therefore, $(n-2)$ independent condition equations can be formulated. These $(n-2)$ independent nonlinear condition equations are:

$$\frac{E\{\underline{y}_1\} - E\{\underline{y}_j\}}{E\{\underline{y}_2\} - E\{\underline{y}_j\}} - \frac{E\{\underline{x}_1\}^2 - E\{\underline{x}_j\}^2}{E\{\underline{x}_2\}^2 - E\{\underline{x}_j\}^2} = 0 \;,\; j = 3,\ldots,n,$$

or:

(41) $$(E\{\underline{y}_1\} - E\{\underline{y}_j\})(E\{\underline{x}_2\}^2 - E\{\underline{x}_j\}^2) - (E\{\underline{y}_2\} - E\{\underline{y}_j\})(E\{\underline{x}_1\}^2 - E\{\underline{x}_j\}^2) = 0 \quad \text{for} \quad j = 3,\ldots,n.$$

In order to obtain the linearized B-model, we need to linearize (41). Linearization of (41) gives:

$$\underbrace{\begin{pmatrix} -2x_1^0 y_{32}^0 & 2x_2^0 y_{31}^0 & 2x_3^0 y_{12}^0 & & (x_2^{0^2}-x_3^{0^2}) & (x_3^{0^2}-x_1^{0^2}) & (x_1^{0^2}-x_2^{0^2}) & \\ \vdots & \vdots & \ddots & & \vdots & \vdots & \ddots & \\ \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots \\ -2x_1^0 y_{n2}^0 & 2x_2^0 y_{n1}^0 & & 2x_n^0 y_{12}^0 & (x_2^{0^2}-x_n^{0^2}) & (x_n^{0^2}-x_1^{0^2}) & & (x_1^{0^2}-x_2^{0^2}) \end{pmatrix}}_{\partial_y \boldsymbol{B}(\underline{y}^0)^*} \underbrace{E\{\begin{pmatrix} \Delta\underline{x}_1 \\ \Delta\underline{x}_2 \\ \Delta\underline{x}_3 \\ \vdots \\ \Delta\underline{x}_n \\ \Delta\underline{y}_1 \\ \Delta\underline{y}_2 \\ \Delta\underline{y}_3 \\ \vdots \\ \Delta\underline{y}_n \end{pmatrix}\}}_{\Delta\underline{y}} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix},$$

(42)

with $y_{i2}^0 = y_2^0 - y_i^0$, $y_{i1}^0 = y_1^0 - y_i^0$, $i = 3,...,n$. This is the *linearized B-model.* The *linearized A-model* follows from linearizing (40) as:

(43)
$$\underbrace{E\{\begin{pmatrix} \Delta\underline{x}_1 \\ \Delta\underline{x}_2 \\ \Delta\underline{x}_3 \\ \vdots \\ \Delta\underline{x}_n \\ \Delta\underline{y}_1 \\ \Delta\underline{y}_2 \\ \Delta\underline{y}_3 \\ \vdots \\ \Delta\underline{y}_n \end{pmatrix}\}}_{\Delta\underline{y}} = \underbrace{\begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & \ddots & & & & \\ & & & & 1 & & & \\ 2a^0 x_1^0 & & & & & (x_1^0)^2 & 1 & \\ & 2a^0 x_2^0 & & & & (x_2^0)^2 & 1 & \\ & & 2a^0 x_3^0 & & & (x_3^0)^2 & 1 & \\ & & & \ddots & & \vdots & \vdots & \\ & & & & 2a^0 x_n^0 & (x_n^0)^2 & 1 & \end{pmatrix}}_{\partial_x \boldsymbol{A}(x^0)} \underbrace{\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \vdots \\ \Delta x_n \\ \Delta a \\ \Delta b \end{pmatrix}}_{\Delta x}.$$

Verify yourself that $\left(\partial_y B(y^0)\right)^* \left(\partial_x A(x^0)\right) = 0$ indeed holds, if $y_i^0 = a^0 (x_i^0)^2 + b^0$

Example 15

Consider the *nonlinear* A-model (27) of example 12:

$$(44)\qquad \underbrace{E\{\begin{pmatrix}\underline{x}_1\\ \underline{y}_1\\ \underline{x}_2\\ \underline{y}_2\\ \underline{x}_3\\ \underline{y}_3\end{pmatrix}\}}_{\underline{y}} = \underbrace{\begin{pmatrix}x_1\\ y_1\\ x_1 + a\cos\alpha\\ y_1 + a\sin\alpha\\ x_1 + a\cos(\alpha + \frac{1}{3}\pi)\\ y_1 + a\sin(\alpha + \frac{1}{3}\pi)\end{pmatrix}}_{\boldsymbol{A}(x)} ;\quad D\{\begin{pmatrix}\underline{x}_1\\ \underline{y}_1\\ \underline{x}_2\\ \underline{y}_2\\ \underline{x}_3\\ \underline{y}_3\end{pmatrix}\} = \sigma^2 I_6$$

This model consists of 6 observation equations in the four unknown parameters x_1, y_1, a, and α. Hence, the *redundancy* equals 6-4 = 2. Therefore two independent condition equations can be formulated. These two nonlinear condition equations read:

$$(45)\qquad \begin{cases}\left[(E\{\underline{x}_2\} - E\{\underline{x}_1\})^2 + (E\{\underline{y}_2\} - E\{\underline{y}_1\})^2\right] - \left[(E\{\underline{x}_3\} - E\{\underline{x}_1\})^2 + (E\{\underline{y}_3\} - E\{\underline{y}_1\})^2\right] = 0\\ \left[(E\{\underline{x}_2\} - E\{\underline{x}_1\})^2 + (E\{\underline{y}_2\} - E\{\underline{y}_1\})^2\right] - \left[(E\{\underline{x}_3\} - E\{\underline{x}_2\})^2 + (E\{\underline{y}_3\} - E\{\underline{y}_2\})^2\right] = 0\end{cases}$$

Note that these condition equations state that the sides of the equilateral triangle should be equal.

Linearization of (45) gives:

$$(46)\qquad \underbrace{\begin{pmatrix}2x_{23}^0 & 2y_{23}^0 & 2x_{12}^0 & 2y_{12}^0 & 2x_{31}^0 & 2y_{31}^0\\ 2x_{21}^0 & 2y_{21}^0 & 2x_{13}^0 & 2y_{13}^0 & 2x_{32}^0 & 2y_{32}^0\end{pmatrix}}_{\partial_y \boldsymbol{B}^*(y^0)} \underbrace{E\{\begin{pmatrix}\Delta\underline{x}_1\\ \Delta\underline{y}_1\\ \Delta\underline{x}_2\\ \Delta\underline{y}_2\\ \Delta\underline{x}_3\\ \Delta\underline{y}_3\end{pmatrix}\}}_{\Delta\underline{y}} = \begin{pmatrix}0\\ 0\end{pmatrix}.$$

Verify yourself that $\left(\partial_y B(y^0)\right)^*\left(\partial_x A(x^0)\right) = 0$ indeed holds (see (28) of subsection 7.1.3), if the approximate values used in (46) satisfy the nonlinear condition equations.

Appendix A

Taylor's theorem with remainder

In this appendix we review the basic theorem of Taylor for functions of one or more variables. We will start with functions of one variable.

***Definition*:** Let f: $D \subset \mathbb{R} \rightarrow \mathbb{R}$ be a function whose domain D is an open set of $\mathbb{R}$. If all of the qth-order derivatives of $f(x)$ exist at every $x \in D$ and each $\frac{d^q}{dx^q} f(x)$ is a continuous function on D, then $f(x)$ is said to be a *function of class $C^{(q)}$ on D.*

Taylor's Theorem

Let $f(x)$ be a function of class $C^{(q)}$ on $D \subset \mathbb{R}$, $x, x^0 \in D$, and $\Delta x = x - x^0$. Then a scalar θ between x and x_0 exists such that:

$$f(x) = f(x^0) + \frac{d}{dx} f(x^0)\Delta x + \frac{1}{2!}\frac{d^2}{dx^2} f(x^0)\Delta x^2 + \cdots + \frac{1}{(q-1)!}\frac{d^{q-1}}{dx^{q-1}} f(x^0)\Delta x^{q-1} + R_q(\theta,\Delta x) \; . \tag{1}$$

with the remainder:

$$R_q(\theta,\Delta x) = \frac{1}{q!}\frac{d^q}{dx^q} f(\theta)\Delta x^q \; . \tag{2}$$

If we ignore the remainder $R_q(\theta,\Delta x)$, the right-hand side of (1) is a polynomial in Δx of degree q-1. If the remainder is small this polynomial furnishes an approximation to $f(x)$. We speak of a *linear approximation* or a *linearization of $f(x)$ at x^0* if $f(x)$ is approximated as:

$$f(x) \doteq f(x^0) + \frac{d}{dx} f(x^0)(x - x^0) \; . \tag{3}$$

This approximation is acceptable if the second term within the square brackets of:

$$\left(\frac{d}{dx} f(x^0) + \frac{1}{2}\frac{d^2}{dx^2} f(\theta)\Delta x \right) \Delta x \; ,$$

is negligible with respect to the first term. Note that this second term can be made arbitrarily small by letting x^0 approach x. A geometric interpretation of (3) is given in figure A.1.

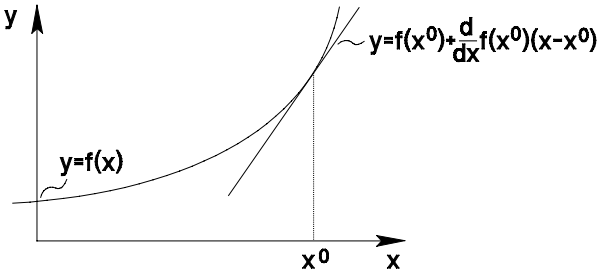


Figure A.1: The curve $y=f(x)$ and its tangentline $y=f(x^0)+\frac{d}{dx}f(x^0)(x-x^0)$ at $f(x^0)$

Examples

1. The linearization of $f(x)=\cos x$ at x^0 reads:

$$f(x) \doteq \cos x^0 - \sin x^0 (x-x^0) \ .$$

2. The linearization of $f(x)=(x^2+a^2)^{1/2}$ at x_o reads:

$$f(x) \doteq \left((x^0)^2+a^2\right)^{1/2} + \frac{x^0}{\left((x^0)^2+a^2\right)^{1/2}} (x-x^0) \ .$$

3. The linearization of $f(x)=x^3$ at x^0 reads:

$$f(x) \doteq (x^0)^3 + 3(x^0)^2 (x-x^0) \ .$$

We will now consider functions of more than one variable. A function $f(x_1, x_2,...,x_n)$ of the variables $x_1, x_2,....,x_n$ will be denoted as $f(x)$ with the n-vector $x=(x_1, x_2,...,x_n)^*$. The first-order partial derivatives of $f(x)$, $\frac{\partial}{\partial x_\alpha} f(x)$, $\alpha=1,...,n$, will be denoted as $\partial_\alpha f(x)$, $\alpha=1,...,n$. Similarly, we denote the second order partial derivatives of $f(x)$, $\frac{\partial^2}{\partial x_\alpha \partial x_\beta} f(x)$, $\alpha,\beta=1,...,n$, as $\partial^2_{\alpha\beta} f(x)$, $\alpha,\beta=1,...,n$. And so on.

***Definition*:** Let $f(x)$: $D\subset\mathbb{R}^n\rightarrow\mathbb{R}$ be a function whose domain D is an open set of $\mathbb{R}^n$.

If all of the qth-order partial derivatives of $f(x)$ exist at every $x=(x_1,x_2,...,x_n)^*\in D$ and each $\partial^q_{\alpha_1..\alpha_q} f(x)$ is a continuous function on D, then $f(x)$ is said to be a *function of class* $C^{(q)}$ on D.

Taylor's Theorem: Let $f(x)$ be a function of class $C^{(q)}$ on $D\subset\mathbb{R}^n$ and x, $x_0\in D$ such that the line segment joining x and x^0 is contained in D. In particular, if D is convex then x and x^0 can be any pair of points in D. Let $\Delta x = x-x^0$ and $\theta=x^0+t(x-x^0)$ with $t\in\mathbb{R}$. Then a scalar $t\in(0,1)$ exists such that:

(4)
$$\begin{aligned} f(x) &= f(x^0) + \sum_{\alpha=1}^{n} \partial_\alpha f(x^0)\,\Delta x_\alpha + \frac{1}{2!}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \partial^2_{\alpha\beta} f(x^0)\,\Delta x_\alpha \Delta x_\beta + \ldots \\ &\ldots + \frac{1}{(q-1)!}\sum_{\alpha_1=1}^{n} \ldots \sum_{\alpha_{q-1}=1}^{n} \partial^{q-1}_{\alpha_1 \ldots \alpha_{q-1}} f(x^0)\Delta x_{\alpha_1} \ldots \Delta x_{\alpha_{q-1}} + \boldsymbol{R}_q(\theta, \Delta x) \end{aligned}$$

with the remainder:

(5)
$$\boldsymbol{R}_q(\theta,\Delta x) = \frac{1}{q!}\sum_{\alpha_1=1}^{n} \ldots \sum_{\alpha_q=1}^{n} \partial^q_{\alpha_1 \ldots \alpha_q} f(\theta)\,\Delta x_{\alpha_1} \ldots \Delta x_{\alpha_q}$$

***Proof*:** Define the functions $\theta(t)$ and $\phi(t)$ by:

$$\theta(t) = x^0 + t(x - x^0) \quad \text{and} \quad \phi(t) = f(\theta(t)) \ .$$

The domain of $\phi(t)$ is $\{t \,|\, x^0 + t(x - x^0) \in D\}$ which is an open set of $\mathbb{R}$ containing the closed interval [0,1]. By repeated application of the *chain rule* we find that:

(6)
$$\begin{cases} \dfrac{d}{dt}\phi(t) &= \displaystyle\sum_{\alpha=1}^{n} \partial_\alpha f(\theta(t))\Delta x_\alpha \\ \dfrac{d^2}{dt^2}\phi(t) &= \displaystyle\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \partial^2_{\alpha\beta} f(\theta(t))\Delta x_\alpha \Delta x_\beta \\ \quad\vdots & \qquad\qquad\qquad\vdots \\ \dfrac{d^q}{dt^q}\phi(t) &= \displaystyle\sum_{\alpha_1=1}^{n} \ldots \sum_{\alpha_q=1}^{n} \partial^q_{\alpha_1 \ldots \alpha_q} f(\theta(t))\Delta x_{\alpha_1} \ldots \Delta x_{\alpha_q}. \end{cases}$$

By Taylor's theorem for functions of one variable there exists a $t \in (0,1)$ such that:

(7)
$$\phi(1) = \phi(0) + \frac{d}{dt}\phi(0) + \ldots + \frac{1}{(q-1)!}\frac{d^{q-1}}{dt^{q-1}}\phi(0) + \frac{1}{q!}\frac{d^q}{dt^q}\phi(t)$$

But $\phi(1) = f(x)$, $\phi(0) = f(x^0)$, and we have, by substitution of (6) into (7), Taylor's formula with remainder.

If we use the vector and matrix notation:

$$(8)\qquad \begin{cases} \Delta x & = (\Delta x_1, \Delta x_2, \cdots, \Delta x_n)^* \\ \partial_x f(x^0) & = (\partial_1 f(x^0), \partial_2 f(x^0), \cdots, \partial_n f(x^0))^* \\ \partial^2_{xx} f(x^0) & = \begin{pmatrix} \partial^2_{11} f(x^0) & \cdots & \partial^2_{1n} f(x^0) \\ \vdots & \ddots & \vdots \\ \partial^2_{n1} f(x^0) & \cdots & \partial^2_{nn} f(x^0) \end{pmatrix} \end{cases}$$

Taylor's formula with third-order remainder becomes:

$$(9)\qquad \boxed{f(x) = f(x^0) + \partial_x f(x^0)^* \Delta x + \frac{1}{2} \Delta x^* \partial^2_{xx} f(x^0) \Delta x + \boldsymbol{R}_3(\boldsymbol{\theta}, \Delta x)}$$

The vector $\partial_x f(x^0)$ is known as the *gradient* of $f(x)$ at x^0. The matrix $\partial^2_{xx} f(x^0)$ is known as the *Hessian matrix* of $f(x)$ at x^0.

***Examples*:**

1. The linearization of $f(x,y) = x\cos y$ at x^0, y^0 reads:

$$f(x,y) \doteq x^0 \cos y^0 + (\cos y^0 - x^0 \sin y^0) \begin{pmatrix} x - x^0 \\ y - y^0 \end{pmatrix}.$$

2. The quadratic approximation of $f(x,y) = x\cos y$ at x^0, y^0 reads:

$$f(x,y) \doteq x^0 \cos y^0 + (\cos y^0 - x^0 \sin y^0) \begin{pmatrix} x - x^0 \\ y - y^0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x - x^0 \\ y - y^0 \end{pmatrix}^* \begin{pmatrix} 0 & -\sin y^0 \\ -\sin y^0 & -x^0 \cos y^0 \end{pmatrix} \begin{pmatrix} x - x^0 \\ y - y^0 \end{pmatrix}.$$

3. The quadratic approximation of $f(x,y) = (x^2 + y^2)^{1/2}$ at $x^0 y^0$ reads:

$$f(x,y) \doteq ((x^0)^2 + (y^0)^2)^{1/2} + ((x^0)^2 + (y^0)^2)^{-1/2} (x^0 \; y^0) \begin{pmatrix} x - x^0 \\ y - y^0 \end{pmatrix}$$

$$+ \frac{1}{2} ((x^0)^2 + (y^0)^2)^{-3/2} \begin{pmatrix} x - x^0 \\ y - y^0 \end{pmatrix}^* \begin{pmatrix} (y^0)^2 & -x^0 y^0 \\ -x^0 y^0 & (x^0)^2 \end{pmatrix} \begin{pmatrix} x - x^0 \\ y - y^0 \end{pmatrix}.$$

It is possible to give an explicit estimate of the remainder in Taylor's formula, in terms of bounds for the qth-order partial derivatives of $f(x)$.

Theorem: Suppose that K is a convex subset of D, such that $x,x^0 \in K$. Then also $\theta = x^0 + t(x - x^0) \in K$ for $t \in [0,1]$. Moreover, suppose that all qth-order partial derivatives of $f(x)$ satisfy:

(10) $$|\partial^q_{\alpha_1 \ldots \alpha_q} f(x)| \leq M \text{ for all } x \in K \ .$$

Then the remainder in Taylor's formula satisfies:

(11) $$\boxed{|R_q(\theta, \Delta x)| \leq M \frac{n^{q/2}}{q!} \|\Delta x\|^q}$$

Proof: From (5) and (10) it follows that:

$$|R_q(\theta, \Delta x)| \ \leq \ M \frac{1}{q!} \sum_{\alpha_1 = 1}^{n} \ldots \sum_{\alpha_q = 1}^{n} |\Delta x_{\alpha_1}| \ \ldots \ |\Delta x_{\alpha_q}|$$

or that:

(12) $$|R_q(\theta, \Delta x)| \ \leq \ M \frac{1}{q!} \left(\sum_{\alpha = 1}^{n} |\Delta x_\alpha| \right)^q \ .$$

For the inner product of two vectors x and y we have:

$$x^* y = \ \|x\| \cdot \|y\| \cdot \cos\alpha \ .$$

From this expression the Cauchy-Schwarz inequality follows:

$$|x^* y| \ \leq \|x\| \cdot \|y\|.$$

This gives for $y=(1,1,\ldots,1)^*$:

(13) $$\sum_{\alpha = 1}^{n} |x_\alpha| \leq n^{1/2} \|x\| \ .$$

If we use this inequality in (12) we obtain the estimate of (11).

Appendix B

Unconstrained optimization: optimality conditions

In this appendix we will formulate necessary and sufficient conditions for finding (local or global) solutions to the minimization problem:

(1) $$\min_x .\ F(x)\ ,\ x\in\mathbb{R}^n\ ,\ F\colon\ \mathbb{R}^n\to\mathbb{R}\ .$$

In the investigation of the minimization problem (1) we distinguish two kinds of solution points: local minimizers and global minimizers.

Definition: The vector $\hat{x}$ is said to be a *local minimum* of $F(x)$ if $F(\hat{x}) \le F(x)\ \forall\ x\in B(\hat{x},\varepsilon)$. The ε-ball, $B(\hat{x},\varepsilon)$, is defined as $B(\hat{x},\varepsilon)=\{x\in\mathbb{R}\,|\,\|x-\hat{x}\|<\varepsilon\}$. The local minimimum is said to be *isolated* if $F(\hat{x})<F(x)\ \forall x\in B(\hat{x},\varepsilon)$.

Definition: The vector $\hat{x}$ is said to be a *global minimum* of $F(x)$ *if* $F(\hat{x}) \le F(x)\forall x\in\ \mathbb{R}^n$. The global minimum is *unique* if $F(\hat{x})<F(x)\ \forall x\in\mathbb{R}^n$.

In the following we will assume that a (local or global) minimum of $F(x)$ exists. The problem of computing the minimum of $F(x)$ can be facilitated by deriving certain properties that must be satisfied by the minimizing vector $\hat{x}$. The following theorem states necessary conditions for $\hat{x}$ to be a minimum of $F(x)$.

Theorem (necessary conditions):
Let $F(x)$ be a class $C^{(3)}$. If $\hat{x}$ is a (local or global) minimum of $F(x)$, then:

(2)
$$\boxed{\begin{aligned} &1.\ \partial_x F(\hat{x}) = 0\\ &2.\ \partial^2_{xx} F(\hat{x}) \ge 0 \end{aligned}}$$

Proof:
First we shall prove (2.1). Consider the vector $x=\hat{x}+td$ where t is a scalar and d an n-vector. Expansion of $F(x)$ in a Taylor series at $\hat{x}$ gives:

(3) $$F(x)=F(\hat{x})+t\partial_x F(\hat{x})^* d+O(t)\ .$$

The order term $O(t)$ indicates the remainder in Taylor's formula. It has the property:

(4) $$\lim_{t\to 0}\frac{O(t)}{t}=0\ .$$

Since $\hat{x}$ is a (local or global) minimum of $F(x)$ by assumption, it follows that for sufficiently small t:

(5) $$F(\hat{x})\ \le\ F(x)=F(\hat{x}+td)\ .$$

This, together with (3) gives after dividing by t:

$$\begin{cases} \partial_x F(\hat{x})^* \boldsymbol{d} + \dfrac{O(t)}{t} \geq 0 \quad \text{for} \quad t>0 \\ \partial_x F(\hat{x})^* \boldsymbol{d} + \dfrac{O(t)}{t} \leq 0 \quad \text{for} \quad t<0 \ . \end{cases}$$

Taking the limit as $t \to 0$ and using (4) shows that (2.1) must be true.

Next we shall prove (2.2). Expanding $F(x)$ in a Taylor series at $\hat{x}$, but now retaining the quadratic terms, gives with (2.1):

$$F(x) = F(\hat{x}) + \frac{1}{2} t^2 \boldsymbol{d}^* \partial^2_{xx} F(\hat{x}) \boldsymbol{d} + O(t^2) \ , \tag{6}$$

where $O(t^2)$ has the property:

$$\lim_{t \to 0} \frac{O(t^2)}{t^2} = 0 \ . \tag{7}$$

Equations (5) and (6) imply that:

$$\frac{1}{2} \boldsymbol{d}^* \partial^2_{xx} F(\hat{x}) \boldsymbol{d} + \frac{O(t^2)}{t^2} \geq 0 \ .$$

Taking the limit as $t \to 0$ and using (7) shows that:

$$\frac{1}{2} \boldsymbol{d}^* \partial^2_{xx} F(\hat{x}) \boldsymbol{d} \geq 0 \ . \tag{8}$$

Since the n-vector d is arbitrary, (8) must hold for all $d \in \mathbb{R}^n$. This means by definition that the Hessian matrix $\partial^2_{xx} F(\hat{x})$ is *positive semi-definite*. Positive semi-definite matrices A are denoted as *"$A \geq 0$"*. This proves (2.2).

Examples

1. The function $F(x,y) = x^2 + xy + y^2$ has a minimum at $(\hat{x},\hat{y})=(0,0)$. According to the theorem, the gradient of $F(x,y)$ must be identical to the zero-vector at this point, and the Hessian matrix of $F(x,y)$ must be positive semi-definite at this point. This is easily verified, since:

$$\begin{pmatrix} \partial_x F(x,y) \\ \partial_y F(x,y) \end{pmatrix} = \begin{pmatrix} 2x + y \\ 2y + x \end{pmatrix} ; \quad \begin{pmatrix} \partial^2_{xx} F(x,y) & \partial^2_{xy} F(x,y) \\ \partial^2_{yx} F(x,y) & \partial^2_{yy} F(x,y) \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} .$$

2. Let the $n \times n$ matrix Q be positive definite. A matrix Q is said to be *positive definite* if:

(9) $$x^* Q x > 0 \quad \forall\, x \in \mathbb{R}^n \setminus \{0\} \ .$$

Since Q is positive definite it follows that the function F: $\mathbb{R}^n \to \mathbb{R}$

(10) $$F(x) = x^* Q x \ ,$$

has a minimum at $\hat{x}=0$. According to the theorem it must hold that $\partial_x F(\hat{x})=0$ and $\partial^2_{xx} F(\hat{x}) \geq 0$. This is easily verified, since it follows from (10) that:

$$\partial_x F(x) = 2Qx \quad \text{and thus} \quad \partial_x F(0) = 2Q0 = 0,$$

and

$$\partial^2_{xx} F(x) = 2Q \quad \text{and thus with (9):} \ \partial^2_{xx} F(\hat{x}=0) \geq 0.$$

The above theorem gives necessary conditions for $\hat{x}$ to be a minimum of $F(x)$. The stated conditions are however *not* sufficient. This is easily illustrated by the following simple example. Consider the function $F_1(x)=x^4$. Clearly it has a (global) minimum $\hat{x}=0$ and $\partial_x F_1(0)=0$ and $\partial^2_{xx} F_1(0)=0$ hold. Consider now the function $F_2(x)=-x^4$. Then once again $\partial_x F_2(0)=0$ and $\partial^2_{xx} F_2(0)=0$ hold. But now, however, $\hat{x}=0$ is a (global) maximum of $F_2(x)$. The following theorem gives sufficient conditions for $\hat{x}$ to be a minimum of $f(x)$.

Theorem (sufficient conditions):
Let $F(x)$ be of class $C^{(3)}$. If:

(11)

1. $\partial_x F(\hat{x}) = 0$
2. $\partial^2_{xx} F(\hat{x}) > 0$

then $\hat{x}$ is a (local or global) minimum of $F(x)$.

Proof:
A Taylor expansion of $F(x)$ at $\hat{x}$ gives with $x=\hat{x}+td$ and (11.1):

(12) $$F(x) = F(\hat{x}) + \left(\frac{1}{2} d^* \partial^2_{xx} F(\hat{x})\, d + \frac{O(t^2)}{t^2} \right) t^2 \ .$$

Since $\partial^2_{xx} F(\hat{x})$ is positive definite by assumption, the first term in the bracket on the right hand side of (12) is strictly positive. Hence, in view of (7), we can conclude that the bracketed term

of (12) is strictly positive for sufficiently small t. Thus from (12) follows that $F(x)>F(\hat{x})$ for all x near $\hat{x}$, that is, for t small.

It should be noted that $\hat{x}$ can be a minimum of $F(x)$ and still violate the sufficient conditions of the above theorem (for example $F(x)=x^4$). Nevertheless the above two theorems facilitate the problem of computing the minimum of $F(x)$. The idea is that one first solves the system of equations $\partial_x F(x)=0$. The solutions of $\partial_x F(x)=0$ are called *stationary or critical points* of $F(x)$. Once the stationary points of $F(x)$ are found, one can check whether $\partial_{xx}^2 F(\hat{x}) > 0$ holds in order to show that the corresponding stationary point must be a local minimum.

If $\partial_{xx}^2 F(\hat{x}) \geq 0$, one first computes the vectors d that satisfy $\partial_{xx}^2 F(\hat{x})d=0$ and then checks whether $F(\hat{x}) \leq F(x)$ for all those x that lie on the ray $x=\hat{x}+td$ that emanates from $\hat{x}$ with direction vector d. After the local minima are found, the global minima follow from comparison of the function values at the local minima.

Example (least-squares)

The function $F:\mathbb{R}^n \rightarrow \mathbb{R}$ is given as:

$$\begin{aligned} F(x) &= (y-\boldsymbol{A}x)^*(y-\boldsymbol{A}x) \\ &= y^*y - 2y^*\boldsymbol{A}x + x^*\boldsymbol{A}^*\boldsymbol{A}x \ . \end{aligned} \tag{13}$$

From this it follows that:

$$\partial_x F(x) = -2\boldsymbol{A}^*y + 2\boldsymbol{A}^*\boldsymbol{A}x \ , \tag{14}$$

and:

$$\partial_{xx}^2 F(x) = 2\boldsymbol{A}^*\boldsymbol{A} \ . \tag{15}$$

If we assume that the $m \times n$ matrix A has full rank n, then A^*A has full rank and $(A^*A)^{-1}$ exists. The solution of $\partial_x F(x)=0$ follows from (14) as:

$$\hat{x} = (\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^*y \ . \tag{16}$$

Thus the function $F(x)$ has only one stationary point. Since $x^*A^*Ax > 0 \ \forall x \in \mathbb{R}^n\backslash\{0\}$ it follows that the Hessian matrix of (15) is positive-definite. Hence, $\hat{x}$ of (16) is the unique global minimum of $F(x)$.

Appendix C

Linearly constrained optimization: optimality conditions

In the following appendix we consider general minimization problems. However, there are special classes of constrained minima which are important and can be studied with the help of the theory of appendix B. These are the *linearly* constrained minima. In problems of this type we restrict our points $y \in \mathbb{R}^m$ to lie in an n-plane in $\mathbb{R}^m$. An n-plane can be considered to be an n-dimensional Euclidean space embedded in $\mathbb{R}^m$. It need not pass through the origin of $\mathbb{R}^m$. If y^0 is a point in the n-plane and $a_1, a_2, ..., a_n$ are n linearly independent vectors in this plane, every point y in the n-plane is expressible in the form:

$$y = y^0 + x_1 a_1 + x_2 a_2 + \ldots + x_n a_n \tag{1}$$

with suitably chosen constants $x_i \in \mathbb{R}, i=1,...,n$. Equation (1) is sketched in figure C.1 for the case $m=3$ and $n=2$.

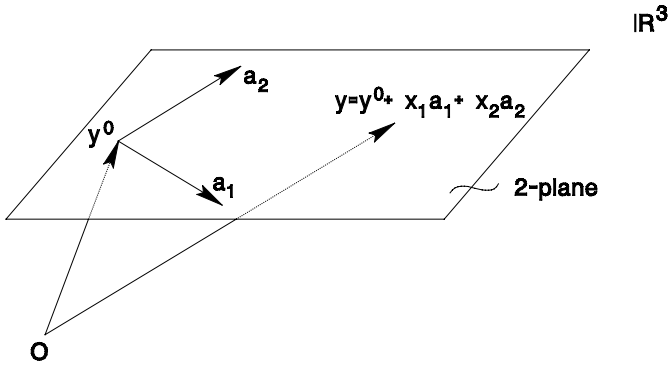


Figure C.1: A 2-plane embedded in $\mathbb{R}^3$

An n-plane is uniquely determined by a point y^0 and a set of n linearly independent vectors $a_1, a_2, ..., a_n$. If $n=1$, we have a line; if $n=2$, we have a two-dimensional plane; and so on. To relate the ideas here presented to those given for more general constrained minima on arbitrary surfaces, it will be convenient to introduce the concept of the tangent space and the normal space to an n-plane. A tangent vector a is a vector parallel to the n-plane. If the n-plane is represented in the *parametric form* (1), a tangent vector a is any vector expressible as a linear combination:

$$a = x_1 a_1 + \ldots + x_n a_n \tag{2}$$

of the vectors $a_1, ..., a_n$. The *tangent space* of the n-plane is the set of all vectors that can be written as (2). It is the n-plane through the origin that is parallel to the set $\{a_1, ..., a_n\}$. Any vector b that is orthogonal to the n-plane and hence to the tangent space will be called a normal vector to the n-plane. The set of all normal vectors to the n-plane is the orthogonal complement of the tangent space and will be called *normal space*. It is of dimension m-n. If $b_1, b_2, ..., b_{m-n}$ form a basis for the normal space, every normal vector b is uniquely expressible as a linear combination:

(3) $$\boldsymbol{b} = \lambda_1 \boldsymbol{b}_1 + \ldots + \lambda_{m-n} \boldsymbol{b}_{m-n}.$$

We may use the normal vectors $b_1,...,b_{m-n}$ to express the n-plane as:

(4) $$\boldsymbol{b}_i^* (y - y^0) = 0 \ , \ i = 1,...,m-n \ .$$

This representation of the n-plane is in contrast to the parametric form (1), known as the *implicit form*. Equation (4) is sketched in figure C.2 for the case m=3 and n=2.

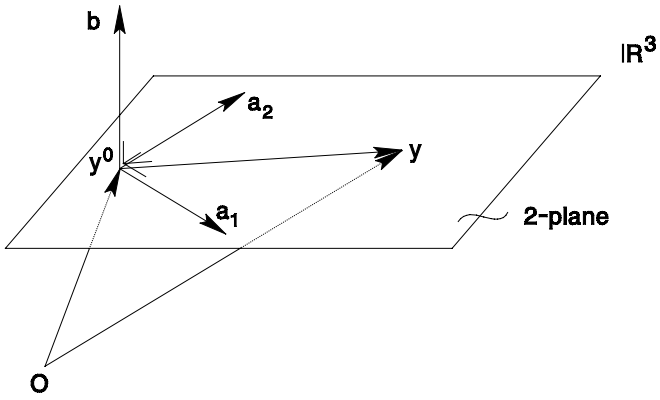


Figure C.2: A 2-plane embedded in $\mathbb{R}^3$ with normal vector b

If we collect the vectors $a_1,...,a_n$ into an $m \times n$ matrix $A=(a_1,a_2,...,a_n)$ and define the n-vector x as $x=(x_1,...,x_n)^*$, we may write (1) in matrix form as:

(5) $$\underset{m\times 1}{y} = \underset{m\times 1}{y^0} + \underset{m\times n}{\boldsymbol{A}} \underset{n\times 1}{x}$$

Similarly, if we collect the vectors $b_1,...,b_{m-n}$ into an $m \times (m-n)$ matrix $B=(b_1,...,b_{m-n})$, we may write (4) in matrix form as:

(6) $$\underset{(m-n)\times m}{\boldsymbol{B}^*} \underset{m\times 1}{(y - y^0)} = \underset{(m-n)\times 1}{0}$$

Observe that the two matrices A and B satisfy the relation:

(7) $$\boxed{\underset{(m-n)\times m}{\boldsymbol{B}^*} \underset{m\times n}{\boldsymbol{A}} = \underset{(m-n)\times n}{0} \ .}$$

We now turn to the characterization of the minimum $\hat{y}$ of a function $F(y)$ on an n-plane. The following theorem gives necessary conditions for $\hat{y}$ to be a solution to the linearly constrained minimization problem:

$$(8)\qquad \begin{cases} \min_{y}.\ F(y),\ y\in\mathbb{R}^m,\ F:\mathbb{R}^m\to\mathbb{R} \\ \text{with } y \text{ constrained to an } n\text{-plane} \end{cases}.$$

Theorem (necessary conditions):
Let $F(y)$ be of class $C^{(3)}$. If $\hat{y}$ is a point on the n-plane that affords a (local or global) minimum of $F(y)$ on the n-plane, then:

$$(9)\qquad \boxed{\begin{aligned} 1.\ \partial_y F(\hat{y})^* A &= 0 \\ 2.\ A^*\partial^2_{yy} F(\hat{y}) A &\geq 0 \end{aligned}}$$

where A is an $m\times n$ matrix of which the column vectors form a basis of the tangent space of the n-plane.

Proof:
The vector $\hat{x}=0$ minimizes:

$$f(x) = F(\hat{y}+Ax)\ .$$

Hence, it follows from the first theorem of appendix B that:

$$\partial_x f(0)^* = \partial_y F(\hat{y})^* A = 0\ ,$$

and:

$$\partial^2_{xx} f(0) = A^*\partial^2_{yy} F(\hat{y}) A \geq 0\ .$$

The following theorem gives sufficient conditions for $\hat{y}$ to be a solution to (8).

Theorem (sufficient conditions):
Let $F(y)$ be of class $C^{(3)}$. If:

$$(10)\qquad \boxed{\begin{aligned} 1.\ \partial_y F(\hat{y})^* A &= 0 \\ 2.\ A^*\partial^2_{yy} F(\hat{y}) A &> 0 \end{aligned}}$$

where A is an $m\times n$ matrix of which the column vectors form a basis of the tangent space of the n-plane, then $\hat{y}$ is a (local or global) minimum of $F(y)$ on this n-plane.

Proof:
The proof follows with $f(x)=F(\hat{y}+Ax)$, $\partial_x f(x)^* = \partial_y F(\hat{y}+Ax)^* A$ and $\partial^2_{xx} f(x) = A^*\partial^2_{yy} F(\hat{y}+Ax) A$ from the second theorem of appendix B.

The above two theorems can be used if the n-plane is given in the parametric form (5). In case the n-plane is defined by the implicit form (6), the results of the above two theorems can be restated as in the following theorem.

Theorem (Lagrange multiplier rule):
Let $\hat{y}$ be a point satisfying the linear constraints:

(11) $$\underset{(m-n)\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{(y-y^0)} = \underset{(m-n)\times 1}{0}$$

If $\hat{y}$ minimizes $F(y)$ subject to these constraints, there exists a multiplier vector $\lambda \in \mathbb{R}^{m-n}$, such that, if we set:

(12) $$L(y) = F(y) + \lambda^* \boldsymbol{B}^* (y - y^0) \ ,$$

then:

(13)
$$\boxed{\begin{aligned} &1.\ \partial_y L(\hat{y}) &&= 0 \\ &2.\ \boldsymbol{a}^* \partial^2_{yy} L(\hat{y})\,\boldsymbol{a} &&\geq 0 \end{aligned}}$$

for all vectors a that satisfy:

(14) $$\underset{(m-n)\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{\boldsymbol{a}} = \underset{(m-n)\times 1}{0} \ .$$

Conversely, if there exists for $\hat{y}$ a function $L(y)$ of form (12) such that:

(15)
$$\boxed{\begin{aligned} &1.\ \partial_y L(\hat{y}) &&= 0 \\ &2.\ \boldsymbol{a}^* \partial^2_{yy} L(\hat{y})\,\boldsymbol{a} &&> 0 \end{aligned}}$$

for solutions $a \neq 0$ of (14), then $\hat{y}$ affords a (local or global) minimum to $F(y)$ subject to the constraints (11).

The result here given is known as the *Lagrange multiplier rule*. The function $L(y)$ of (12) is known as the *Lagrangian* and the vector λ as the *Lagrange* multiplier vector.

Proof:
We first prove (13.1). Condition (9.1) states that the gradient $\partial_y F(\hat{y})$ is orthogonal to the range space of A. This implies that $\partial_y F(\hat{y})$ can be written as a linear combination of the base vectors of the normal space of the n-plane. Hence, there exists a vector $\lambda \in \mathbb{R}^{m-n}$ such that:

$$\partial_y F(\hat{y}) + \boldsymbol{B}\lambda = 0$$

But this implies that the gradient of $L(y)$ must vanish at $\hat{y}$. Condition (13.2) follows simply from (9.2) by noting that $\partial^2_{yy} L(\hat{y}) = \partial^2_{yy} F(\hat{y})$.

The sufficient conditions (15) follow in a similar way from (10).

Example 1 (observation equations)

Let y^0 be a fixed given vector in $\mathbb{R}^m$, and let F: $\mathbb{R}^m \to \mathbb{R}$ be defined as:

$$F(y) = (y^0 - y)^*(y^0 - y) = \|y^0 - y\|^2 \quad . \tag{16}$$

Note that the function $F(y)$ measures the square of the distance between y and the fixed vector y^0. Hence, the $(m\text{-}1)$-dimensional surface $F(y)$=constant is an hypersphere embedded in $\mathbb{R}^m$ (see figure C.3).

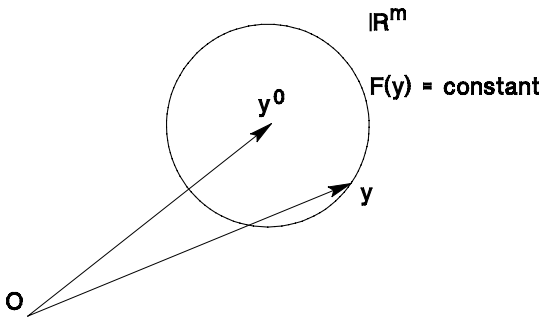


Figure C.3: $F(y)=\|y_o\text{-}y\|^2$= constant

Our objective is to minimize the function $F(y)$ of (16) subject to the linear restrictions:

$$\underset{m\times 1}{y} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{x} \quad , \ \mathrm{rank}\boldsymbol{A} = n \tag{17}$$

Those vectors that satisfy (17) lie on an n-plane through the origin with tangent vectors given by the column vectors of the matrix A (see figure C.4). The problem of minimizing $F(y)$ subject to (17) can now be visualized as the problem of finding the proper radius of the hypersphere of figure C.3 such that the sphere just touches the n-plane of figure C.4. This is shown in figure C.5. The point of contact between sphere and n-plane is the solution $\hat{y}$.

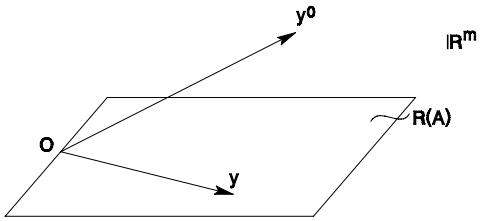


Figure C.4: The n-plane: $y=Ax$, rank $A=n$

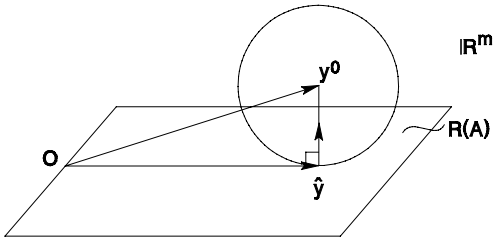


Figure C.5: Hypersphere $\|y^0-y\|^2=$ constant, and n-plane $y=Ax$

The solution $\hat{y}$ can be found using (10). Since $\partial_y F(y)=-2(y^0-y)$, it follows that:

$$\partial_y F(y)^* \boldsymbol{A} = -2(y^0-y)^* \boldsymbol{A}.$$

Substitution of (17) and setting the results equal to zero gives:

$$\boldsymbol{A}^* \boldsymbol{A}x = \boldsymbol{A}^* y^0 \ .$$

This system of equations has the unique solution $\hat{x}=(A^*A)^{-1}A^*y^0$. Hence:

(18) $$\hat{y} = \boldsymbol{A}\hat{x} = \boldsymbol{A}(\boldsymbol{A}^* \boldsymbol{A})^{-1} \boldsymbol{A}^* y^0 \ .$$

Since $\partial^2_{yy} F(y) = 2I_m$ and A^*A is positive definite it follows that:

$$\boldsymbol{A}^* \partial^2_{yy} F(\hat{y}) \boldsymbol{A} = 2\boldsymbol{A}^* \boldsymbol{A} > 0 \ .$$

Hence, $\hat{y}$ is a minimum of $F(y)$ subject to (17).

Example 2 (condition equations)

Let y^0 be a fixed given vector in $\mathbb{R}^m$, and let F: $\mathbb{R}^m \to \mathbb{R}$ be defined as:

(19) $$F(y) = (y^0-y)^*(y^0-y) = \|y^0-y\|^2 \ .$$

Our objective is to minimize the function $F(y)$ of (19) subject to the linear restrictions.

(20) $$\underset{(m-n)\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{y} = \underset{(m-n)\times 1}{0}\ ,\ \operatorname{rank}\boldsymbol{B} = m-n\ .$$

Since these restrictions are in *implicit form*, we have to make use of the Lagrange multiplier rule. We construct the Lagrangian:

(21) $$L(y) = (y^0-y)^*(y^0-y) + 2\lambda^*\boldsymbol{B}^*y\ .$$

The factor 2 in the multiplier vector has merely been introduced for convenience. The stationary points of the Lagrangian have to satisfy:

(22) $$\partial_y L(y) = -2(y^0-y) + 2\boldsymbol{B}\lambda = 0\ .$$

This shows that:

(23) $$\hat{y} = y^0 - \boldsymbol{B}\lambda\ .$$

In order to determine $\hat{y}$, we first have to determine λ. Pre-multiplication of (22) with B^* gives together with (20):

(24) $$\boldsymbol{B}^*\boldsymbol{B}\lambda = \boldsymbol{B}^*y^0\ .$$

Since matrix B has full rank, the inverse of B^*B exists. Hence λ follows from (24) as:

(25) $$\lambda = (\boldsymbol{B}^*\boldsymbol{B})^{-1}\boldsymbol{B}^*y^0.$$

Substitution of this result into (23) gives:

(26) $$\hat{y} = [I - \boldsymbol{B}(\boldsymbol{B}^*\boldsymbol{B})^{-1}\boldsymbol{B}^*]\ y^0\ .$$

Thus the Lagrangian (21) has only one stationary point. Since $\partial^2_{yy}L(y) = 2\ I_{\mathrm{m}}$, it follows that condition (15.2) is satisfied. Hence, $\hat{y}$ of (26) is the solution of our linearly constrained minimization problem.

Example 3 (Best Linear Unbiased Estimation)

In Chapter 2, section 4 the principle of "*Best Linear Unbiased Estimation*" has been discussed. This principle leads to the following linearly constrained minimization problem:

(27) $$\min_{l}.\ l^*\boldsymbol{Q}_y l \quad \text{subject to} \quad \boldsymbol{A}^*l = f\ .$$

This problem may be solved by the Lagrange multiplier rule. We consider the following Lagrangian:

(28) $$L(l) = l^*\boldsymbol{Q}_y l + 2\lambda^*(\boldsymbol{A}^*l - f)\ .$$

The stationary points of the Lagrangian have to satisfy:

(29) $$\partial_l L(l) = 2\boldsymbol{Q}_y l + 2\boldsymbol{A}\lambda = 0\ .$$

This shows that:

(30) $$\hat{l} = -\boldsymbol{Q}_y^{-1}\boldsymbol{A}\lambda \ .$$

In order to determine $\hat{l}$ we first have to determine λ. Pre-multiplication of (29) with $A^{*}Q_y^{-1}$ gives together with the constraints $A^{*}l=f$:

(31) $$\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}\lambda = -f$$

Since A is of full rank and Q_y is positive definite, also $A^{*}Q_y^{-1}A$ is of full rank and positive definite. Hence, the inverse of $A^{*}Q_y^{-1}A$ exists and λ follows from (31) as:

(32) $$\lambda = -(\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}f \ .$$

Substitution into (30) gives:

(33) $$\hat{l} = \boldsymbol{Q}_y^{-1}\boldsymbol{A}(\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A})^{-1}f \ .$$

Since $\partial_{ll}^{2}L(l) = 2Q_y$ and Q_y is positive definite, it follows that $\hat{l}$ of (33) is the unique solution of (27).

Appendix D

Constrained optimization: optimality conditions

In the previous appendix we were concerned with the problem of minimizing a function $F(y)$ with linear equality constraints. In this appendix we seek to minimize a function $F(y)$ with *nonlinear* equality constraints. Hence, we want to solve the constrained minimization problem:

$$(1) \qquad \begin{cases} \min\limits_{y}. \; F(y) \;, \; F\colon \mathbb{R}^m \to \mathbb{R} \\ \text{subject to } \boldsymbol{B}(y) = 0, \; \boldsymbol{B}\colon \mathbb{R}^m \to \mathbb{R}^{m-n} \end{cases}$$

The constraints in (1) are in *implicit form* and they are nonlinear if the vector function $B\colon \mathbb{R}^m \to \mathbb{R}^{m-n}$ is nonlinear. Except in abnormal situations, the constraints of (1) describe an n-dimensional surface (or manifold) embedded in $\mathbb{R}^m$. Examples of $B(y) = 0$ are:

1. $x^2 + y^2 + z^2 - R^2 = 0$,

2. $\begin{cases} x^2 + y^2 + z^2 - R^2 = 0 \\ y - x = 0 \; . \end{cases}$

The first equation describes the surface of a sphere with radius R embedded in $\mathbb{R}^3$. The last two equations together describe a circle embedded in $\mathbb{R}^3$.

By virtue of the implicit function theorem the constraints:

$$(2) \qquad \boldsymbol{B}(y) = 0 \;, \; \boldsymbol{B}\colon \mathbb{R}^m \to \mathbb{R}^{m-n}$$

are expressible locally in *parametric form* as:

$$(3) \qquad y = \boldsymbol{A}(x) \;, \; \boldsymbol{A}\colon \mathbb{R}^n \to \mathbb{R}^m \; .$$

For instance, example 1 is expressible in parametric form as:

1. $\begin{cases} x = R \cos\varphi \cos\lambda \\ y = R \cos\varphi \sin\lambda \\ z = R \sin\varphi \end{cases}$.

And the equations of example 2 as:

2. $\begin{cases} x = \frac{1}{2}\sqrt{2}\, R \cos\varphi \\ y = \frac{1}{2}\sqrt{2}\, R \cos\varphi \\ z = R \sin\varphi \; . \end{cases}$

Since (3) puts the same restrictions on y as (2) we may write the minimization problem (1) also as:

(4)
$$\begin{cases} \min\limits_{y} F(y)\ ,\ F\colon \mathbb{R}^m \to \mathbb{R} \\ \text{subject to}\quad y = \boldsymbol{A}(x)\ ,\ \boldsymbol{A}\colon \mathbb{R}^n \to \mathbb{R}^m\ . \end{cases}$$

Consequently, the problem of minimizing $F(y)$ subject to $B(y) = 0$ is equivalent to the unconstrained problem of minimizing $F(A(x))$. In some cases when it is simple to transform (2) into (3) this is a suitable procedure. However, in many cases it is easier to use the *Lagrange multiplier rule*. The following two theorems, which will be given without proof, give necessary and sufficient conditions for $\hat{y}$ to be a (local or global) solution to (1).

Theorem (necessary conditions):
Suppose that $\hat{y}$ affords a (local or global) minimum to:

$$F(y)\ ,\ F\colon \mathbb{R}^m \to \mathbb{R}\ ,$$

subject to the constraints:

$$\boldsymbol{B}(y) = 0\ ,\ \boldsymbol{B}\colon \mathbb{R}^m \to \mathbb{R}^{m-n}\ .$$

Suppose further that the $m\times(m\text{-}n)$ matrix of partial derivatives $\partial_y B(\hat{y})$ has full rank $m\text{-}n$. Then there exists a unique multiplier vector $\lambda \in \mathbb{R}^{m-n}$ such that, if we set:

$$L(y) = F(y) + \lambda^* \boldsymbol{B}(y)\ ,$$

then:

(5)
$$\boxed{\begin{aligned} &1.\quad \partial_y L(\hat{y}) = \partial_y F(\hat{y}) + \partial_y \boldsymbol{B}(\hat{y})\,\lambda = 0 \\ &2.\quad \boldsymbol{a}^* \partial^2_{yy} L(\hat{y})\ \boldsymbol{a} \geq 0 \end{aligned}}$$

for all vectors a that satisfy:

$$\underset{(m-n)\times m}{\partial_y \boldsymbol{B}(\hat{y})^*}\ \underset{m\times 1}{\boldsymbol{a}} = \underset{(m-n)\times 1}{0}\ .$$

Theorem (sufficient conditions):
Suppose that for a point $\hat{y}$ satisfying $B(y) = 0$ there is a Lagrangian $L(y) = F(y) + \lambda^* B(y)$ such that:

(6)
$$\boxed{\begin{aligned} &1.\quad \partial_y L(\hat{y}) = 0 \\ &2.\quad \boldsymbol{a}^* \partial^2_{yy} L(\hat{y})\ \boldsymbol{a} > 0 \end{aligned}}$$

for all vectors $a \neq 0$ satisfying $\partial_y B(\hat{y})^* a = 0$, then $\hat{y}$ is a (local or global) minimum of $F(y)$ subject to $B(y) = 0$.

Appendix E

Mean and variance of scalar random variables

Mean of a scalar random variable

Let $\underline{x}$ be a continuous scalar random variable with probability density function $p_{\underline{x}}(x)$. The *expectation* or *mean* of $\underline{x}$ is by definition the integral:

(1) $$E\{\underline{x}\} = \int_{-\infty}^{+\infty} x\, p_{\underline{x}}(x)\, dx \quad .$$

This number will also be denoted by m_x.

Mean of a function of a scalar random variable

Given a scalar random variable $\underline{x}$ and a function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. As we see from (1), the mean of $\underline{y}$ is given by:

(2) $$E\{\underline{y}\} = \int_{-\infty}^{+\infty} y\, p_{\underline{y}}(y)\, dy \quad .$$

It appears, therefore, that to determine the mean of $\underline{y}$, we must first find its probability density function $p_{\underline{y}}(y)$. This, however, is not necessary. As the next theorem shows, $E\{\underline{y}\}$ can be expressed directly in terms of the function $f(x)$ and the density $p_{\underline{x}}(x)$ of $\underline{x}$.

***Theorem*:**

(3) $$E\{f(\underline{x})\} = \int_{-\infty}^{+\infty} f(x)\, p_{\underline{x}}(x)\, dx \quad .$$

***Proof*:**
We shall sketch a proof using the curve $f(x)$ of figure E.1. With $y = f(x_1) = f(x_2) = f(x_3)$ as in the figure, we see that:

$$P(y \le \underline{y} \le y + dy) = P(x_1 \le \underline{x} \le x_1 + dx_1) + P(x_2 - dx_2 \le \underline{x} \le x_2) + P(x_3 \le \underline{x} \le x_3 + dx_3)$$

or:

$$p_{\underline{y}}(y)\, dy = p_{\underline{x}}(x_1)\, dx_1 + p_{\underline{x}}(x_2)\, dx_2 + p_{\underline{x}}(x_3)\, dx_3 .$$

Multiplying by y, we obtain:

$$y p_{\underline{y}}(y)\,dy = f(x_1)p_{\underline{x}}(x_1)\,dx_1 + f(x_2)p_{\underline{x}}(x_2)\,dx_2 + f(x_3)p_{\underline{x}}(x_3)\,dx_3 .$$

Thus, to each differential in (2) there corresponds one or more differentials in (3). As dy covers the y-axis, the corresponding dx's are nonoverlapping and they cover the entire x-axis. Hence, the integrals in (2) and (3) are equal.

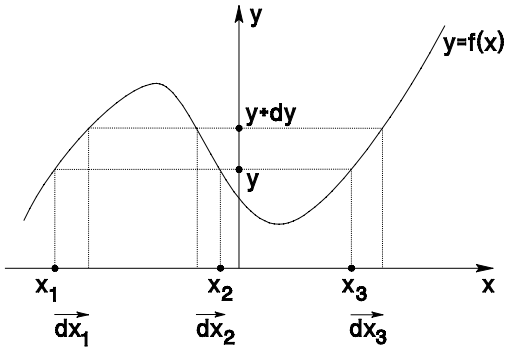


Figure E.1: The curve $y = f(x)$

It appears from (3) that to determine the mean of $\underline{y}$, we must know the probability density function $p_{\underline{x}}(x)$. In general this is true. However, in the special case that the function $f(x)$ is *linear* not the complete density of $\underline{x}$ needs to be known but only the mean m_x of $\underline{x}$.

***Theorem (propagation law of the mean)*:**
Given a scalar random variable $\underline{x}$ and a function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. If the function $f(x)$ is *linear*:

(4) $$f(x) = ax + b ,$$

then:

(5) $$\boxed{m_y = a m_x + b}$$

***Proof*:**
Substitution of (4) into (3) gives:

$$\begin{aligned} m_y &= \int_{-\infty}^{+\infty} (ax+b)p_{\underline{x}}(x)dx \\ &= a\int_{-\infty}^{+\infty} x p_{\underline{x}}(x)dx + b\int_{-\infty}^{+\infty} 1\; p_{\underline{x}}(x)dx \\ &= a m_x + b. \end{aligned}$$

If the function $f(x)$ is nonlinear we need strictly speaking the density $p_{\underline{x}}(x)$ of $\underline{x}$ in order to compute the mean of $\underline{y} = f(\underline{x})$. However, by using Taylor's formula an approximation to the

mean of $\underline{y}$ can be derived that gets around the difficulty of having to know the density $p_x(x)$ of $\underline{x}$.

***Theorem (linearized propagation law of the mean)*:**
Given a scalar random variable $\underline{x}$ and a nonlinear function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. Let x^0 be an approximation to a sample of $\underline{x}$ and define $\Delta\underline{y} = \underline{y}\text{-}f(x^0)$ and $\Delta\underline{x} = \underline{x}\text{-}x^0$. Then a first-order approximation to the mean of $\Delta\underline{y}$ is:

(6)
$$\boxed{m_{\Delta y} \doteq \frac{d}{dx} f(x^0)\, m_{\Delta x}} \quad .$$

***Proof*:**
Application of Taylor's formula gives:

$$\underline{y} = f(x^0) + \frac{d}{dx} f(x^0)(\underline{x}-x^0) + \frac{1}{2}\frac{d^2}{dx^2} f(x^0)(\underline{x}-x^0)^2 + \ldots$$

If we take the expectation we get:

$$E\{\underline{y}\} = f(x^0) + \frac{d}{dx} f(x^0)\, E\{\underline{x}-x^0\} + \frac{1}{2}\frac{d^2}{dx^2} f(x^0)\, E\{(\underline{x}-x^0)^2\} + \ldots$$

This may be written with:

$$m_{\Delta y} = E\{\underline{y}\} - f(x^0) \;,\; m_{\Delta x} = E\{\underline{x}-x^0\} \;,$$

and:

$$\begin{aligned} E\{(\underline{x}-x^0)^2\} &= E\{\left[(\underline{x}-m_x)-(x^0-m_x)\right]^2\} \\ &= E\{(\underline{x}-m_x)^2\} - 2E\{(\underline{x}-m_x)(x^0-m_x)\} + E\{(x^0-m_x)^2\} \\ &= E\{(\underline{x}-m_x)^2\} + m_{\Delta x}^2 \\ &= \sigma_x^2 + m_{\Delta x}^2 \;, \end{aligned}$$

as:

$$m_{\Delta y} = \frac{d}{dx} f(x^0)\, m_{\Delta x} + \frac{1}{2}\frac{d^2}{dx^2} f(x^0)\left(\sigma_x^2 + m_{\Delta x}^2\right) + \ldots$$

If we neglect the second and higher order terms, the result (6) follows.

Examples

1. With $f(x) = \cos x$, we define the random variable $\underline{y} = f(\underline{x})$. Since $\frac{d}{dx} f(x^0) = -\sin x^0$, we find to a first order:
$$E\{\underline{y} - \cos x^0\} \doteq -\sin x^0 \; E\{\underline{x} - x^0\} \; .$$

2. With $f(x) = x^3$, we define the random variable $\underline{y} = f(\underline{x})$. Since $\frac{d}{dx} f(x^0) = 3(x^0)^2$ we find to a first order:
$$E\{\underline{y} - (x^0)^3\} \doteq 3(x^0)^2 \; E\{\underline{x} - x^0\} \; .$$

Variance of a scalar random variable

The *variance* of a scalar random variable is by definition the integral:

(7) $$\boxed{E\{(\underline{x} - E\{\underline{x}\})^2\} = \int_{-\infty}^{+\infty} (x - m_x)^2 p_{\underline{x}}(x)\, dx} \; .$$

This number will also be denoted by σ_x^2. The positive constant σ_x is called the *standard deviation* of $\underline{x}$. From the definition it follows that:

$$\sigma_x^2 = E\{(\underline{x} - m_x)^2\} = E\{\underline{x}^2\} - 2E\{\underline{x}\} m_x + m_x^2 \; ,$$

or

(8) $$\sigma_x^2 = E\{\underline{x}^2\} - m_x^2 \; .$$

Variance of a function of a scalar random variable

***Theorem (propagation law of the variance)*:**
Given a scalar random variable $\underline{x}$ and a function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. If the function $f(x)$ is *linear*:

(9) $$f(x) = ax + b \; ,$$

then:

(10) $$\boxed{\sigma_y^2 = a^2 \sigma_x^2}$$

***Proof*:**
According to definition (7) we have:

$$\sigma_y^2 = E\{(\underline{y} - E\{\underline{y}\})^2\} = E\{(f(\underline{x}) - m_y)^2\} .$$

With (3) this gives:

$$\sigma_y^2 = \int_{-\infty}^{+\infty} (f(x) - m_y)^2 p_{\underline{x}}(x)\, dx .$$

Substitution of (5) and (9) gives:

$$\begin{aligned} \sigma_y^2 &= \int_{-\infty}^{+\infty} \left[(ax + b) - (am_x + b) \right]^2 p_{\underline{x}}(x)\, dx \\ &= a^2 \int_{-\infty}^{+\infty} (x - m_x)^2 p_{\underline{x}}(x)\, dx \\ &= a^2 \sigma_x^2 . \end{aligned}$$

The above result shows that if $f(x)$ is linear, knowledge of σ_x^2 is sufficient for computing the variance of $\underline{y} = f(\underline{x})$. For nonlinear functions $f(x)$ this is generally not true. If the function $f(x)$ is nonlinear one will generally need to know the complete density $p_{\underline{x}}(x)$ of $\underline{x}$. However, by using Taylor's formula an approximation to the variance of $\underline{y}$ can be derived that gets around the difficulty of having to know $p_{\underline{x}}(x)$.

***Theorem (linearized propagation law of the variance)*:**
Given a scalar random variable $\underline{x}$ and a nonlinear function $f(x)$, we form the random variable $\underline{y} = f(\underline{x})$. Let x^0 be an approximation to a sample of $\underline{x}$. Then a first-order approximation to the variance of $\underline{y}$ is:

(11)
$$\boxed{\sigma_y^2 \doteq \left(\frac{d}{dx} f(x^0) \right)^2 \sigma_x^2}$$

***Proof*:**
Substitution of:

$$\begin{cases} f(x) = f(x^0) + \dfrac{d}{dx} f(x^0)(x - x^0) + \ldots \\ m_y = f(x^0) + \dfrac{d}{dx} f(x^0)(m_x - x^0) + \ldots \end{cases}$$

into:

$$\sigma_y^2 = \int_{-\infty}^{+\infty} (f(x) - m_y)^2 p_{\underline{x}}(x)\, dx ,$$

gives after neglecting second and higher order terms:

$$\begin{aligned}\sigma_y^2 &\doteq \int_{-\infty}^{+\infty}\left(\frac{d}{dx}f(x^0)(x-m_x)\right)^2 p_{\underline{x}}(x)dx \\ &\doteq \left(\frac{d}{dx}f(x^0)\right)^2 \int_{-\infty}^{+\infty}(x-m_x)^2 p_{\underline{x}}(x)dx \\ &\doteq \left(\frac{d}{dx}f(x^0)\right)^2 \sigma_x^2\end{aligned}$$

Examples

1. A first order approximation to the variance of $\underline{y} = \cos \underline{x}$ is:

$$\sigma_y^2 \doteq \sin^2 x^0 \ \sigma_x^2 \ .$$

2. A first order approximation to the variance of $\underline{y} = \underline{x}^3$ is:

$$\sigma_y^2 \doteq 9(x^0)^4 \ \sigma_x^2 \ .$$

Appendix F

Mean and variance of vector random variables

Mean of a vector random variable

Let $\underline{x}_i$, $i = 1, 2,\ldots,n$ be n continuous scalar random variables with *joint* probability density function $p_x(x_1, x_2,\ldots,x_n)$. The expectation or mean of the random n-vector $(\underline{x}_1, \underline{x}_2,\ldots,\underline{x}_n)^*$ is by definition the integral:

(1)
$$\boxed{E\left\{\begin{pmatrix}\underline{x}_1\\ \underline{x}_2\\ \vdots\\ \underline{x}_n\end{pmatrix}\right\} = \int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix} p_{\underline{x}}(x_1,x_2,\ldots,x_n)\,dx_1\cdots dx_n}$$

If we use the notation:

$$\underline{x} = (\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)^*,\ x=(x_1,\ldots,x_n)^*,\ p_{\underline{x}}(x_1,\ldots,x_n) = p_{\underline{x}}(x),\ dx_1,\ldots,dx_n = dx \text{ and } \int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty} = \int$$

we may write (1) in the compact form:

(2)
$$\boxed{E\{\underline{x}\} = \int x\, p_{\underline{x}}(x)\,dx}\ .$$

From (1) it follows that:

(3)
$$E\{\underline{x}_i\} = \int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty} x_i\, p_{\underline{x}}(x_1,\ldots,x_i,\ldots,x_n)\,dx_1\cdots dx_n\ .$$

This result seems to contradict definition (1) of appendix E. Note however that (3) reduces to (1) of appendix E, since:

$$\int_{-\infty}^{+\infty} p_{\underline{x}}(x_1,\ldots,x_j,\ldots,x_n)\,dx_j = p_{\underline{x}}(x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n)\ .$$

Hence, (1) may also be written as:

$$E\left\{\begin{pmatrix}\underline{x}_1\\ \underline{x}_2\\ \vdots\\ \underline{x}_n\end{pmatrix}\right\} = \begin{pmatrix}\int_{-\infty}^{+\infty} x_1\, p_{\underline{x}_1}(x_1)\,dx_1\\ \int_{-\infty}^{+\infty} x_2\, p_{\underline{x}_2}(x_2)\,dx_2\\ \vdots\\ \int_{-\infty}^{+\infty} x_n\, p_{\underline{x}_n}(x_n)\,dx_n\end{pmatrix} . \tag{4}$$

Thus in order to compute $E\{\underline{x}_i\}$ one only needs the *marginal* density $p_{\underline{x}_i}(x_i)$.

Mean of a vector function of a random vector

Given a random n-vector $\underline{x}$ and a vector function $F(x)$, F: $\mathbb{R}^n \rightarrow \mathbb{R}^m$, we form the random m-vector $\underline{y} = F(\underline{x})$. As we see from (4) the mean of $\underline{y}_i$ is given by:

$$E\{\underline{y}_i\} = \int_{-\infty}^{+\infty} y_i\, p_{\underline{y}_i}(y_i)\,dy_i . \tag{5}$$

It appears, therefore, that to determine the mean of $\underline{y}_i$, we must first find its *marginal* probability density function $p_{\underline{y}_i}(y_i)$. This, however, is not necessary. As the next theorem shows, $E\{\underline{y}_i\}$ can be expressed directly in terms of the vector function $F(x)$ and the *joint* density of $\underline{x}$.

***Theorem*:**

$$\boxed{E\{F(\underline{x})\} = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} F(x_1,\ldots,x_n)\; p_{\underline{x}}(x_1,\ldots,x_n)\,dx_1\cdots dx_n} \tag{6}$$

or:

$$\boxed{E\{F(\underline{x})\} = \int F(x)\; p_{\underline{x}}(x)\,dx} \tag{7}$$

***Proof*:**
The proof is similar to the proof given in appendix E.

The following theorem is an extremely important one, and it is used frequently.

***Theorem (propagation law of the mean)*:**
Given a random n-vector $\underline{x}$ and a vector function $F(x)$, F: $\mathbb{R}^n \rightarrow \mathbb{R}^m$, we form the random m-vector $\underline{y} = F(\underline{x})$. If the vector function $F(x)$ is linear:

$$\underset{m\times 1}{F(x)} = \underset{m\times n}{A}\;\underset{n\times 1}{x} + \underset{m\times 1}{b} \tag{8}$$

then:

(9)
$$\underset{m\times 1}{m_y} = \underset{m\times n}{\boldsymbol{A}}\ \underset{n\times 1}{m_x} + \underset{m\times 1}{\boldsymbol{b}} \quad .$$

***Proof*:**
If we denote the n column vectors of matrix A by a_i, i=1, .., n, we may write (8) as:

$$F(x) = \sum_{i=1}^{n} a_i\, x_i + b.$$

If we substitute this into (6) we get:

$$E\{F(\underline{x})\} = \sum_{i=1}^{n} a_i \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_i p_{\underline{x}}(x_1,\dots,x_n)\,dx_1 \dots dx_n + b \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p_{\underline{x}}(x_1,\dots,x_n)\,dx_1 \dots dx_n$$

or:

$$\begin{aligned} E\{F(\underline{x})\} &= \sum_{i=1}^{n} a_i \int_{-\infty}^{+\infty} x_i\, p_{\underline{x}_i}(x_i)\,dx_i + b \\ &= \sum_{i=1}^{n} a_i\, m_{x_i} + b \\ &= A\, m_x + b. \end{aligned}$$

Without proof we also give the linearized version of the propagation law of the mean.

***Theorem (linearized propagation law of the mean)*:**
Given a random n-vector $\underline{x}$ and a nonlinear vector function $F(x)$, F: $\mathbb{R}^n \rightarrow \mathbb{R}^m$ we form the random m-vector $\underline{y} = F(\underline{x})$. Let $x^0 \in \mathbb{R}^n$ be an approximation to a sample of $\underline{x}$ and define $\Delta\underline{y} = \underline{y} - F(x^0)$ and $\Delta\underline{x} = \underline{x} - x^0$. Then we have to a first-order:

(10)
$$m_{\Delta y} \doteq \partial_x F(x^0)\, m_{\Delta x} \quad .$$

Example 1
Let the two random variables $\underline{y}_1$ and $\underline{y}_2$ be defined as:

$$\begin{cases} \underline{y}_1 = 1\underline{x}_1 + 3\underline{x}_2 + 5\underline{x}_3 + 2 \\ \underline{y}_2 = 4\underline{x}_1 + 2\underline{x}_2 - 1\underline{x}_3 + 3 \end{cases}$$

Then:

$$E\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\end{pmatrix}\right\} = \begin{pmatrix}1 & 3 & 5\\ 4 & 2 & -1\end{pmatrix} E\left\{\begin{pmatrix}\underline{x}_1\\ \underline{x}_2\\ \underline{x}_3\end{pmatrix}\right\} + \begin{pmatrix}2\\ 3\end{pmatrix}.$$

Example 2

Let the two random variables $\underline{y}_1$ and $\underline{y}_2$ be defined as:

$$\begin{cases}\underline{y}_1 = \sin(\underline{x}_1\underline{x}_2) + \underline{x}_3\\ \underline{y}_2 = \underline{x}_1^2 + \underline{x}_2 + 4.\end{cases}$$

With the approximate values x_1^0 x_2^0 and x_3^0 we have to a first-order:

$$E\left\{\begin{pmatrix}\underline{y}_1 - \sin(x_1^0 x_2^0) - x_3^0\\ \underline{y}_2 - (x_1^0)^2 - x_2^0 - 4\end{pmatrix}\right\} = \begin{pmatrix}x_2^0\cos(x_1^0 x_2^0) & x_1^0\cos(x_1^0 x_2^0) & 1\\ 2x_1^0 & 1 & 0\end{pmatrix} E\left\{\begin{pmatrix}\underline{x}_1 - x_1^0\\ \underline{x}_2 - x_2^0\\ \underline{x}_3 - x_3^0\end{pmatrix}\right\} .$$

Variance matrix of a random vector

Let $\underline{x}$, i=1,2,...,n be n continuous scalar random variables with joint probability density function $p_x(x_1,...,x_n)$. The *variance matrix* of the random n-vector $(\underline{x}_1,\underline{x}_2,...,\underline{x}_n)^*$ is by definition the integral:

$$(11)\quad \boxed{E\left\{\begin{pmatrix}\underline{x}_1 - E\{\underline{x}_1\}\\ \vdots\\ \underline{x}_n - E\{\underline{x}_n\}\end{pmatrix} \cdot \begin{pmatrix}\underline{x}_1 - E\{\underline{x}_1\}\\ \vdots\\ \underline{x}_n - E\{\underline{x}_n\}\end{pmatrix}^*\right\} = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \begin{pmatrix}x_1 - E\{\underline{x}_1\}\\ \vdots\\ x_n - E\{\underline{x}_n\}\end{pmatrix} \cdot \begin{pmatrix}x_1 - E\{\underline{x}_1\}\\ \vdots\\ x_n - E\{\underline{x}_n\}\end{pmatrix}^* p_{\underline{x}}(x_1,...,x_n)\,dx_1...dx_n} .$$

This variance matrix will be denoted by Q_x. Using vector notation we may write (11) in the compact form:

$$(12)\quad \boxed{E\{(\underline{x} - m_x)(\underline{x} - m_x)^*\} = \int (x - m_x)(x - m_x)^*\, p_{\underline{x}}(x)\,dx} .$$

From (11) it follows that:

$$E\{(\underline{x}_i - m_{x_i})(\underline{x}_j - m_{x_j})\} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (x_i - m_{x_i})(x_j - m_{x_j}) p_{\underline{x}}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \ .$$

Integration gives:

$$E\{(\underline{x}_i - m_{x_i})(\underline{x}_j - m_{x_j})\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_i - m_{x_i})(x_j - m_{x_j}) p_{\underline{x}}(x_i, x_j)\, dx_i dx_j \ , \tag{13}$$

in which $p_{\underline{x}}(x_i,x_j)$ is the joint density function of the two random variables $\underline{x}_i$ and $\underline{x}_j$. The scalar (13) is called the *covariance* of the two random variables $\underline{x}_i$ and $\underline{x}_j$. This covariance will be denoted as $\sigma_{x_i x_j}$. Note that the off-diagonal elements of the variance matrix Q_x consist of the covariances between the elements of the random vector $\underline{x}$. If $i=j$ it follows from (13) that:

$$E\{(\underline{x}_i - m_{x_i})^2\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_i - m_{x_i})^2 p_{\underline{x}}(x_i, x_j)\, dx_i dx_j \ .$$

Integration gives:

$$E\{(\underline{x}_i - m_{x_i})^2\} = \int_{-\infty}^{+\infty} (x_i - m_{x_i})^2 p_{\underline{x}_i}(x_i)\, dx_i \ .$$

This is the *variance* $\sigma_{x_i}^2$ of $\underline{x}_i$. Note that the variance of $\sigma_{x_i}^2$ is the ith-diagonal element of the variance matrix Q_x. Hence, the variance matrix Q_x can be written as:

$$\underset{n\times n}{Q_x} = \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_n} \\ \sigma_{x_2x_1} & \sigma_{x_2}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{x_nx_1} & \cdots & \cdots & \sigma_{x_n}^2 \end{pmatrix} . \tag{14}$$

Note that the variance matrix is a *symmetric* matrix.

Variance matrix of a vector function of a random vector

***Theorem (propagation law of variances)*:**
Given a random n-vector $\underline{x}$ and a vector function $F(x)$, F: $\mathbb{R}^n \rightarrow \mathbb{R}^m$, we form the random m-vector $\underline{y} = F(\underline{x})$. If the vector function $F(x)$ is *linear*:

$$\underset{m\times 1}{F(x)} = \underset{m\times n}{A}\ \underset{n\times 1}{x} + \underset{m\times 1}{b} \ , \tag{15}$$

then:

(16)
$$\underset{m\times m}{Q_y} = \underset{m\times n}{A}\ \underset{n\times n}{Q_x}\ \underset{n\times m}{A^*}$$

***Proof*:**
If we denote the n column vectors of matrix A by a_i, i=1,...,n, we may write (15) as:

(17)
$$F(x) = \sum_{i=1}^{n} a_i x_i + b.$$

In a similar way we may write (9) as:

(18)
$$m_y = \sum_{i=1}^{n} a_i m_{x_i} + b.$$

Substitution of (17) and (18) into:

$$Q_y = E\{(\underline{y}-m_y)(\underline{y}-m_y)^*\} \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} (F(x)-m_y)(F(x)-m_y)^*\ p_{\underline{x}}(x_1,\ldots,x_n)\,dx_1\cdots dx_n$$

gives:

$$Q_y = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}(x_i-m_{x_i})(x_j-m_{x_j})\ p_{\underline{x}}(x_1,\ldots,x_n)\ dx_1\cdots dx_n$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* \sigma_{x_i x_j}.$$

This can be written as:

$$Q_y = (a_1,\ldots,a_n)\begin{pmatrix} \sum_{j=1}^{n}\sigma_{x_1x_j}a_j^* \\ \vdots \\ \sum_{j=1}^{n}\sigma_{x_nx_j}a_j^* \end{pmatrix}$$

or as:

$$Q_y = (a_1,\ldots,a_n)\begin{pmatrix} \sigma_{x_1x_1} & \cdots & \sigma_{x_1x_n} \\ \vdots & & \vdots \\ \sigma_{x_nx_1} & \cdots & \sigma_{x_nx_n} \end{pmatrix}\begin{pmatrix} a_1^* \\ \vdots \\ a_n^* \end{pmatrix}$$

$$= AQ_xA^*.$$

Without proof we also give the linearized version of the propagation law of variances.

***Theorem (linearized propagation law of variances)*:**
Given a random n-vector $\underline{x}$ and a nonlinear vector function $F(x)$, F: $\mathbb{R}^n \to \mathbb{R}^m$, we form the random m-vector $\underline{y} = F(\underline{x})$. Let $x^0 \in \mathbb{R}^n$ be an approximation to a sample of $\underline{x}$. Then we have to a first-order:

$$\underset{m\times m}{Q_y} \doteq \underset{m\times n}{\partial_x F(x^0)}\ \underset{n\times n}{Q_x}\ \underset{n\times m}{\partial_x F(x^0)^*} \tag{19}$$

Example 1
Let the two random variables $\underline{y}_1$ and $\underline{y}_2$ be defined as:

$$\begin{cases} \underline{y}_1 = 1\underline{x}_1 + 3\underline{x}_2 + 5\underline{x}_3 + 2 \\ \underline{y}_2 = 4\underline{x}_1 + 2\underline{x}_2 - 1\underline{x}_3 + 3 \end{cases}$$

The variance matrix of $\underline{x} = (\underline{x}_1, \underline{x}_2, \underline{x}_3)^*$ is given as:

$$Q_x = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then:

$$\begin{aligned} Q_y &= \begin{pmatrix} 1 & 3 & 5 \\ 4 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 3 & 2 \\ 5 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 50 & 25 \\ 25 & 41 \end{pmatrix}. \end{aligned}$$

Example 2
Let the two random variables $\underline{y}_1$ and $\underline{y}_2$ be defined as:

$$\begin{cases} \underline{y}_1 = \underline{x}_1 + \underline{x}_2^2 + \underline{x}_1\underline{x}_3 \\ \underline{y}_2 = \underline{x}_1^3 + \sin\underline{x}_2. \end{cases} \tag{20}$$

The variance matrix of $\underline{x} = (\underline{x}_1, \underline{x}_2, \underline{x}_3)^*$ is given as:

$$Q_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

The 2-by-3 matrix of partial derivatives $\partial_x F(x^0)$ follows from (20) as:

$$\underset{2\times 3}{\partial_x F(x^0)} = \begin{pmatrix} \partial_{x_1} F^1(x^0) & \partial_{x_2} F^1(x^0) & \partial_{x_3} F^1(x^0) \\ \partial_{x_1} F^2(x^0) & \partial_{x_2} F^2(x^0) & \partial_{x_3} F^2(x^0) \end{pmatrix} = \begin{pmatrix} 1+x_3^0 & 2x_2^0 & x_1^0 \\ 3(x_1^0)^2 & \cos x_2^0 & 0 \end{pmatrix}. \tag{22}$$

If we take as approximate values $x_1^0=1$, $x_2^0=0$, $x_3^0=0$, (22) becomes:

$$\partial_x F(x^0) = \begin{pmatrix} 1 & 0 & 1 \\ 3 & 1 & 0 \end{pmatrix}. \tag{23}$$

The variance matrix Q_y follows now from (21) and (23) to a first-order as:

$$Q_y \doteq \begin{pmatrix} 1 & 0 & 1 \\ 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\doteq \begin{pmatrix} 3 & 3 \\ 3 & 10 \end{pmatrix}.$$

Appendix G

A BRIEF ACCOUNT ON THE EARLY HISTORY OF ADJUSTING GEODETIC AND ASTRONOMICAL OBSERVATIONS[1]

Prof. dr ir P.J.G. Teunissen, Netherlands Geodetic Commission (KNAW)

I. Introduction

Adjustment theory can be regarded as the part of mathematical geodesy[2] that deals with the optimal combination of redundant observations for the purpose of determining values for parameters of interest. It is essential for a geodesist, its meaning comparable to what mechanics means to a civil engineer or a mechanical engineer. The two main reasons for performing redundant measurements are the wish to increase the accuracy of the results computed and the requirement to be able to check for errors. Due to the intrinsic uncertainty in observations, redundancy generally leads to an inconsistent system of equations. Without additional criteria, such a system is not uniquely solvable.

The problem of solving an inconsistent system of equations has attracted the attention of leading scientists in the middle of the 18th century. Historically, the first methods of combining redundant observations originate from studies in geodesy and astronomy, namely from the problem of determining the size and shape of the Earth, and the problem of finding a mathematical representation of the motions of the moon. Since its discovery almost 200 years ago, least-squares has been the most popular method of adjustment. Although the method of least squares may seem 'natural' for a modern student of adjustment theory, its discovery evolved only slowly from earlier methods of combining redundant observations[3]. In this contribution we sketch the historical line of development of these adjustment methods in the second half of the 18th century.

II. The method of selected points

It is convenient to cast the problem of combining redundant observations in terms of vectors and matrices. Suppose we are given a set of linear equation of the form

$$\underset{m\times 1}{\underbrace{y}} = \underset{m\times n}{\underbrace{A}}\ \underset{n\times 1}{\underbrace{x}}$$

where y is a vector of observations, A is a given matrix of full rank and x is the vector of unknown parameters. This set of linear equations is said to be overdetermined when there are more observations than unknowns, $m > n$. The problem is to combine the m observations so that one can

[1] Based on a presentation given at the occasion of the official opening of the Geodetic-Astronomical Observatory Westerbork, 24 September 1999.

[2] Mathematical geodesy covers the development of theory and its implementation as is needed in order to process, analyse, integrate and validate the various geodetic data. It concerns itself with the calculus of observations (adjustment and estimation theory), with the validation of mathematical models (testing and reliability theory) and with the analysis for spatial and temporal phenomena (interpolation and prediction theory), Founders of the Dutch School of Mathematical Geodesy, internationally known as the 'Delft School', are the professors J.M. Tienstra (1895-1951) and W. Baarda (1917-2005).

[3] Stigler, S.M. (1986): *The History of Statistics*, Belknap, Harvard. Hald, A. (1998): *A History of Mathematical Statistics*, Wiley-Interscience.

solve for the n unknown parameters. If we restrict ourselves to linear combinations of the observations, we can write the general solution in the following form

$$\underset{n\times 1}{\underbrace{x}} = \underset{n\times m}{\underbrace{B}}\ \underset{m\times 1}{\underbrace{y}} \qquad \text{with} \qquad \underset{n\times m}{\underbrace{B}} = \underbrace{(LA)^{-1}}_{n\times n}\ \underset{n\times m}{\underbrace{L}}$$

and where matrix L is a suitably chosen matrix defining the linear combinations. Different choices of L give different linear combinations and therefore different solutions. In modern terminology, matrix B is called a left-inverse of A, since B times A equals the identity matrix. Before 1750 a popular, albeit subjective, method of solving an overdetermined set of linear equations was *the method of selected points*. It consists of choosing n out of m observations (referred to as the selected points) and using their equations to solve for x. If the choice falls on the first n observations, the corresponding L matrix takes the form

$$L = \left[\underset{n\times n}{\underbrace{I}}\ \ \underset{n\times(m-n)}{\underbrace{0}}\right]$$

where I is the identity matrix. For the method of selected points, n residuals (the difference between the observed and adjusted observations) are by definition equal to zero. Many scientists using this method calculated the remaining $m-n$ residuals and studied their sign and size to get an impression of the goodness of fit between observations and the proposed law. The method is subjective because no clear rule is given which observations to select and which to throw out. Selecting another set of n observations leads to a different solution for x. Although the disadvantage of not using all observations was recognized, no simple method existed to tackle this shortcoming. Sometimes all possible combinations of n observations were considered and then averaged to obtain the final result. But since this approach requires handling m over n combinations, it was only practical for problems of low dimensions.

III. The method of averages

Tobias Mayer (1723-1762), professor of mathematics and head of the Göttingen observatory, made numerous observations of the Moon with the purpose of determining the characteristics of the Moon's orbit. In 1750[4] Mayer proposed a new method for adjusting his Moon data, a method which solved the above-mentioned pitfall of the method of selected points. Apart from his adjustment method, Mayer is also known for his other contributions to surveying and navigation. In 1752 he invented the Repeating or Reflecting Circle, an instrument for observing the angle between two celestial bodies. The accuracy of Mayer's instrument was comparable to John Hadley's reflecting octant (1731), but had the advantage that it could be used to measure angles of over 90 degrees[5]. Mayer also contributed to solving the mariner's 'longitude problem'. It was the British Parliament, which in 1714, offered the 'Longitude Prize' to those who could find a 'useful and practical' method for determining longitude at sea. To determine longitude of a ship at sea, the mariner needs to know both his local time and the time at some standard location. Local time was readily determined, but the determination of standard time at sea was more complicated. Mayer's detailed lunar tables (1755) made it possible to translate the instrument readings into longitude positions. The use of Mayer's lunar tables was later superseded

[4] Mayer, T. (1750): Abhandlung ueber die Umwalzung des Monds um seine Axe und die scheinbare Bewegung der Mondsflecten, *Kosmogr. Nachr. Samml. Auf das Jahr 1748*, 1, 52-183.

[5] Forbes, E.G. (1974): *The Birth of Scientific Navigation*, Maritime Monographs and Reports, No. 10, National Maritime Museum, Greenwich, London.

by John Harrison's marine chronometer H-4 (1759). In recognition of their contributions, both men were awarded part of the 'Longitude Prize', with the larger sum going to Harrison[6].

Mayer studied the liberation of the Moon by observing the changing position of the crater Manilius as seen from the Earth. Using spherical geometry, he found a linearized relationship between his observables and some location parameters of Manilius and the Moon's pole. This gave him an inconsistent system of $m = 27$ linear equations in 3 unknown parameters

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & a_{12} & a_{13} \\ \vdots & \vdots & \vdots \\ 1 & a_{m2} & a_{m3} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_3 \end{bmatrix}}_{x}$$

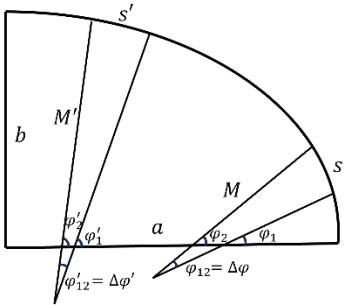
Mayer proposed to divide the 27 equations into 3 groups of 9 each, to sum the equations within each group, and to solve the resulting 3 equations in 3 unknowns. For a general set of m equations in n unknowns, this approach amounts to a separation of the m equations into n groups, followed by a groupwise summation. For example, in case $n = 2$, the corresponding L matrix takes the form

$$L = \begin{bmatrix} e_1 & 0 \\ 0 & e_2 \end{bmatrix}$$

where the e_1 and e_2 are row vectors having only 1's as their entries. Since one may use averages instead of sums, the method became later known as *the method of averages*. Mayer's method of averages soon became popular. It used all observations and it was very simple to apply. However, due to the lack of an objective criterion on how to group the observations, the method was still a subjective one.

IV. The method of least absolute deviations

To determine the Earth's flattening as predicted by Newton's theory of gravitation (Principia[7], 1687), the French Academy of Sciences organized arc-measurement expeditions to Peru, Lapland and the Cape of Good Hope in the period 1735-1754. These expeditions aroused the interest in other countries and in 1750 Pope Benedict XIV commissioned the Jesuit and professor of mathematics Roger Joseph Boscovich (1711-1787) to perform a similar geodetic survey near Rome, the results of which were published in 1755. In a summary of this report, published in 1757[8], Boscovich formulated his new method, now known as *the method of least absolute deviations*, and applied it to the data of the French and Italian arc measurements.



In order to understand the equations used by Boscovich, we first need to introduce some elements from ellipsoidal geodesy. For short meridian arcs, the arc length s (see figure 1) can be written as $s = M(\varphi)\Delta\varphi$, with $M(\varphi)$ the meridian radius of curvature, φ the geodetic latitude of the midpoint of the arc and $\Delta\varphi$ the latitude difference of the two arc endpoints.

Figure 1: Latitude arc measurements along a meridian.

[6] Sobel, D. (1995): *Longitude*. Walker, New York, Dutch translation by E. van Altena (1997): Dava Sobel: *Lengtegraad*, Ambo.
[7] Newton, I. (1687): *Philosophiae Naturalis Principia Mathematica*.
[8] Boscovich, R.J., Maire, C. (1770): *Voyage astronomique et géographique dans l'etat de l'église*. French translation of original 1755 publication. For Boscovich' contributions, see also Sheynin, O.B. (1973): *Arch. History Exact Sci.*, 9: 306-324.

The meridian curvature and its expansion are given as

$$M = \frac{a(1-e^2)}{(1-e^2 sin^2\varphi)^{\frac{3}{2}}} = a(1-e^2)\left\{1 + \frac{3}{2}e^2 sin^2\varphi + \cdots\right\}$$

with $e^2 = (a^2 - b^2)/a^2$ the eccentricity, a and b the half lengths of the major and minor axis and $a(1-e^2)\Delta\varphi$ the length of an arc at the equator. Using only the first two terms in the expansion, the lengths of an $\Delta\varphi$-arc can be written as

$$s = x_1 + sin^2\varphi\, x_2 \quad \text{with} \quad x_1 = a(1-e^2)\Delta\varphi \quad \text{and} \quad x_2 = \tfrac{3}{2}ae^2(1-e^2)\Delta\varphi$$

This is one equation in two unknowns, x_1 and x_2. The arc length s and geodetic latitude φ are determined from astronomical and geodetic measurements, while x_1 and x_2 contain the unknown dimensions of the ellipsoid of revolution. Although a minimum of two arcs is needed to solve for the unknowns, it is preferable to use more than two $\Delta\varphi$-arcs. As a result, one obtains the following system of linear equations

$$\underbrace{\begin{bmatrix} s_1 \\ \vdots \\ s_m \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & sin^2\varphi_1 \\ \vdots & \vdots \\ 1 & sin^2\varphi_m \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x}$$

This is the system of equations which formed the start of Boscovich' analysis. Note that he A-matrix becomes near rank defect, when all arcs are chosen to the same latitude. For an accurate determination of the two unknowns, it is therefore preferable to choose arcs at widely different latitudes. From the data available, Boscovich chose five such arcs ($m = 5$). In his first analyses, Boscovich used the method of selected points. He chose the two arcs with the largest difference in latitude. Not satisfied with the result obtained (the 3 residuals were considered too large), he considers all possible pairs of measured arcs. This gave him 10 selected points to solve, but again he is not satisfied with the results obtained. After having struggled for some time on how to proceed, Boscovich finally formulates his new method of solution in 1757. He states that the parameters x_1 and x_2 should be chosen in such a way that the residuals sum up to zero and have minimum absolute sum. In formula form these two conditions read

$$\sum_{i=1}^{m}(s_i - x_1 - x_2 sin^2\varphi_i) = 0 \qquad \text{and} \qquad \sum_{i=1}^{m}|s_i - x_1 - x_2 sin^2\varphi_i| = min$$

The first condition (although not essential) was motivated by the assumed symmetry in the error distribution, while the second was chosen to get the adjusted values 'as close as possible' to the observed ones. Boscovich gave the graphical algorithm for solving his problem, but no analytical one. The analytical proof of the solution was first given by Laplace in 1793. Using his principle, Boscovich first determined the two parameters x_1 and x_2 and from them the flattening as $f = x_2/3x_1$. Here he only used the first term of the expansion

$$f = \frac{(a-b)}{a} = \frac{1}{2}e^2 + \frac{1}{8}e^4 + \frac{1}{16}e^6 + \cdots$$

The value obtained by Boscovich equals $f = 1/246$, which was smaller than the flattening predicted by Newton. Based on the rotational ellipsoid as an equilibrium figure for a homogeneous, fluid, rotating Earth, Newton obtained the value $f = 1/230$. Boscovich value is however larger than the value known today (International Association of Geodesy (1980): $f = 1/298.257$). Boscovich' method was the first adjustment method that started from the principle of minimizing a function of the

residuals. However, although the method is objective and uses all the observations, it did not reach the same level of popularity as Mayer's method. The method, being nonlinear, was difficult to apply, while the at that time available algorithm could only handle a system of equations with a maximum of two unknowns. In the second half of the 20th century the method gained in popularity due to its property of being resistant (robust) against outliers. Nowadays Boscovich adjustment method is usually referred to as an L_1adjustment, since the L_1-norm of a vector is the sum of absolute values of its entries.

V. The method of least squares

Adrien-Marie Legendre (1752-1833), a professor of mathematics at the École Militaire in Paris, was appointed by the French Academy of Sciences as a member of various committees on astronomical and geodetic projects, among them the committee on the standardization of weights and measures. The committee proposed to define the meter as 10^{-7} times the length of the terrestrial meridian quadrant through Paris at mean sea-level. The arc-measurements took place in the period 1792-1795 and were analysed, among others[9], by both Laplace and Legendre. Legendre's 1805[10] publication on the determination of the orbits of comets, contains a nine-page appendix in which for the first time *the method of least-squares*[11] is described, together with an application of the method to the arc measurements. Legendre used a different equation than Boscovich. Boscovich used the equation

$$s = \left[a(1-e^2) + \frac{3}{2}ae^2(1-e^2)sin^2\varphi\right]\Delta\varphi$$

while Legendre, having the determination of the meter in mind, used a parametrization which differed from the one used by Boscovich. With $\sin^2\varphi = \frac{1}{2}(1-\cos 2\varphi)$, $d = a(1-e^2) + \frac{1}{2}\frac{3}{2}ae^2(1-e^2)$, $\sin\Delta\varphi \approx \Delta\varphi$, and $fd \approx \frac{1}{2}a(1-e^2)e^2$, the above equation can be written as $s = d\Delta\varphi - \frac{3}{2}fd\sin\Delta\varphi\cos 2\varphi$. As a result, we may write for the two endpoints of an arc,

$$\varphi_{12} = s_{12}d^{-1} + \frac{3}{2}fsin\varphi_{12}cos(\varphi_1+\varphi_2).$$

This is one equation in two unknowns, d and f. Note that $d\frac{\pi}{180}$ is the length of a 1-degree arc at 45 degree latitude.

Legendre understood that the observed latitude differences would correlate in case the arcs were connected. He therefore transformed the above equation of differences into an equivalent undifferenced form. This can be achieved by introducing an appropriate additional equation with an additional unknown. As a result, we obtain Legendre's linear system of equations as

---

[9] In 1795 the French government invited other European governments to delegate scientists to Paris for completing and checking the computations for the standardization of weights and measures. The Dutch delegates were professor Jan Hendrik van Swinden and the navy officer Henricus Aeneae. The European scientists met in Paris in 1798 and reported on their findings in 1799. The report on the meter was given by van Swinden, the standard meter, a bar of platinum with rectangular section of 254 millimetres, was placed in the French State Archives, Mètre et Kilogramme des Archives. In 1983 the standard meter was defined as the length travelled by light in vacuum in 1/299.792.458 seconds.

[10] Legendre, A.M. (1805): *Nouvelles méthodes pour la determination des orbits des comètes*. (Appendix: *Sur la méthode des moindre carrés*).

[11] When Carl Friedrich Gauss published his first probabilistic version of the method of least-squares in 1809, he claimed that he had been using the method ('our principle') already since 1795. This claim resulted in a priority dispute between Legendre and Gauss, for a discussion see e.g. Placket, R.L. (1972): The discovery of the method of least-squares. *Biometrica* 59: 239-251.

$$\underbrace{\begin{bmatrix}\varphi_1\\ \vdots\\ \varphi_l\\ \vdots\\ \varphi_m\end{bmatrix}}_{y} = \underbrace{\begin{bmatrix}1 & s_{l1} & \frac{3}{2}sin\,\varphi_{l1}\,cos(\varphi_1+\varphi_l)\\ \vdots & \vdots & \vdots\\ 1 & 0 & 0\\ \vdots & \vdots & \vdots\\ 1 & s_{lm} & \frac{3}{2}sin\,\varphi_{lm}\,cos(\varphi_m+\varphi_l)\end{bmatrix}}_{A} \underbrace{\begin{bmatrix}\varphi_l\\ d^{-1}\\ f\end{bmatrix}}_{x}$$

Since Legendre had four connected arcs ($m = 5$) at his disposal, he had to solve 5 equations in 3 unknowns. In order to solve this overdetermined linear system of equations, Legendre proposed to determine x such that the sum of the squares of the residuals is minimized. In a vector-matrix form $(y - Ax)^T(y - Ax) = min$.

By setting the derivatives of this quadratic form equal to zero, he shows that the solution satisfies the consistent system of linear equations (nowadays referred to as 'the normal equations') $A^T y = A^T Ax$. Note that this result corresponds to the following choice of the L matrix: $L = A^T$.

After solving his set of equations for the three unknowns, φ_3, d and f, Legendre obtained for the flattening the value $f = 1/148$, which he recognizes as being too large. He therefore recomputed his least-squares adjustment, but now with f constrained to the at that time adopted value for the Earth's flattening. As a result, he obtained a value for d, which was now close to the value obtained earlier by Laplace and on which the actual definition of the meter was based. The reason for Legendre having to constrain f lies in the poor resolution of his data. The total arc length of his data covered only about 10 degrees.

Although Legendre did not give a clear motivation for his 'least-squares' criterion, he did realize its potential. His method used all the observations, had an objective criterion and most importantly, resulted – as opposed to Boscovich' method – in a solvable *linear* system of equations. The method met with almost immediate success. Within ten years after Legendre's publication, the method of least-squares became a standard tool in astronomy and geodesy in various European countries. And within twenty years, also the probabilistic foundations of the method were largely completed, the main contributors being Laplace and Gauss.

Literature

1. Calculus

Almering, J.H.J. Revised by H. Bavinck and R.W. Goldbach; *Analyse*
VSSD, 6th ed. (1996)

2. Linear Algebra

Braber, C.A. den, H. van Iperen and M.A. Vlietgever; *Matrixrekening*
VSSD, 2nd ed. (1989)

Lay, D.C.; *Linear algebra and its applications*
Reading, Addison-Wesley, 2nd ed. (1997)

Ortega, J.M.; *Matrix theory, a second course*
New York, Plenum Press (1987)

Strang, G.; *Linear algebra and its applications*
San Diego, Harcourt Brace Jovanovich Publishers, 3rd ed. (1988)

3. Probability Theory

Grimmett, G. and D. Welsh; *Probability; an introduction*
Oxford, Clarendon Press (1986)

Lopuhaä, H.P.; *Lecture notes: Statistiek voor geodeten*
Delft University of Technology, Faculty of Information Technology and Systems (1994)

Papoulis, A.; *Probability, random variables and stochastic processes*
New York, McGraw-Hill Book Company (1991)

Soest, J. van; *Elementaire statistiek*
VSSD, 7th ed. (1997)

4. Estimation Theory

Cooper, M.A.R.; *Control surveys in civil engineering*
London, Collins Publishers (1987)

Koch, K.R.; *Parameter estimation and hypothesis testing in linear models*
Berlin Springer Verlag, 2nd (1999)

Mikhail, E.M., Gracie, G.; *Analysis and adjustment of survey measurements*
New York, van Nostrand Reinhold Company (1981)

Tienstra, J.M.; *Theory of the adjustment of normally distributed observations*
Amsterdam, Argus, (1956)

Index

Adjustment theory:
an introduction

Peter J.G. Teunissen

Adjustment theory can be regarded as the part of mathematical geodesy that deals with the optimal combination of redundant measurements together with the estimation of unknown parameters. It is essential for a geodesist, its meaning comparable to what mechanics means to a civil engineer or a mechanical engineer. Historically, the first methods of combining redundant measurements originate from the study of three problems in geodesy and astronomy, namely to determine the size and shape of the Earth, to explain the long-term inequality in the motions of Jupiter and Saturn, and to find a mathematical representation of the motions of the Moon. Nowadays, the methods of adjustment are used for a much greater variety of geodetic applications, ranging from, for instance, surveying and navigation to remote sensing and global positioning.

The two main reasons for performing redundant measurements are the wish to increase the accuracy of the results computed and the requirement to be able to check for errors. Due to the intrinsic uncertainty in measurements, measurement redundancy generally leads to an inconsistent system of equations. Without additional criteria, such a system of equations is not uniquely solvable. In this introductory course on adjustment theory, methods are developed and presented for solving inconsistent systems of equations. The leading principle is that of least-squares adjustment together with its statistical properties.

The inconsistent systems of equations can come in many different guises. They could be given in parametric form, in implicit form, or as a combination of these two forms. In each case the same principle of least-squares applies. The algorithmic realizations of the solution will differ however. Depending on the application at hand, one could also wish to choose between obtaining the solution in one single step or in a step-wise manner. This leads to the need of formulating the system of equations in partitioned form. Different partitions exist, measurement partitioning, parameter partitioning, or a partitioning of both measurements and parameters. The choice of partitioning also affects the algorithmic realization of the solution. In this introductory text the methodology of adjustment is emphasized, although various samples are given to illustrate the theory. The methods discussed form the basis for solving different adjustment problems in geodesy.

P.J.G. Teunissen
Delft University of Technology
Faculty of Civil Engineering and Geosciences

Dr (Peter) Teunissen is Professor of Geodesy at Delft University of Technology (DUT) and an elected member of the Royal Netherlands Academy of Arts and Sciences. He is research-active in various fields of Geodesy, with current research focused on the development of theory, models, and algorithms for high-accuracy applications of satellite navigation and remote sensing systems. His past DUT positions include Head of the Delft Earth Observation Institute, Education Director of Geomatics Engineering and Vice-Dean of Civil Engineering and Geosciences. His books at TUDelft Open are Adjustment Theory, Testing Theory, Dynamic Data Processing and Network Quality Control.I

TU Delft

ISBN 978-94-6366-884-2
DOI https://doi.org/10.59490/tb.95
textbooks.open.tudelft.nl

Cover image: J.E. Alberda

Series on Mathematical Geodesy and Positioning