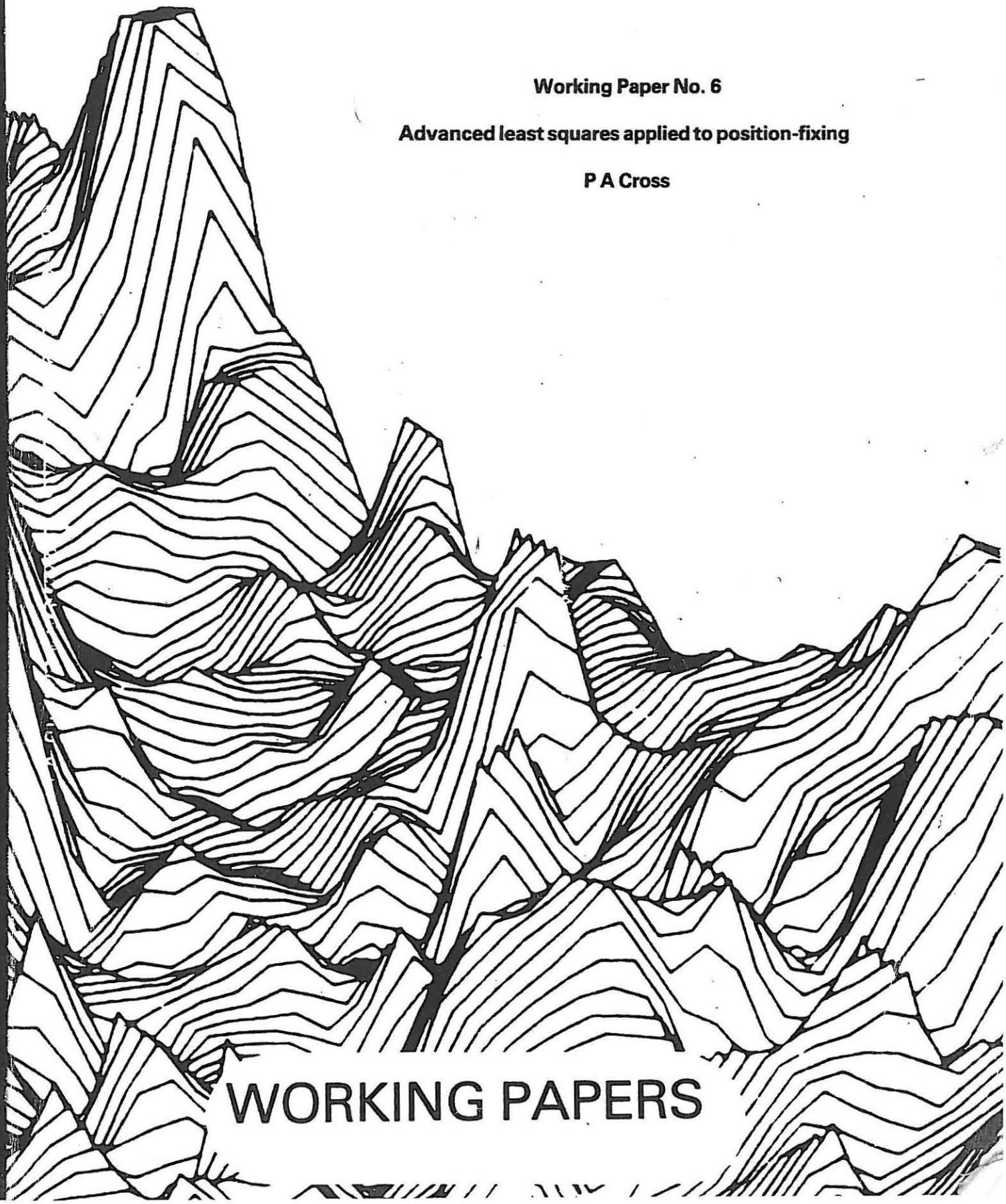
University of East London

School of Surveying

ISSN 0260-9142

University *of*
East London



Working Paper No. 6

Advanced least squares applied to position-fixing

P A Cross

WORKING PAPERS

Working Paper No. 6

Advanced least squares applied to position fixing

P A Cross

University of East London

School of Surveying

January 1994

Note: Change of name:

North East London Polytechnic was renamed Polytechnic of East London - 1989
Polytechnic of East London was renamed University of East London - 1992
Department of Land Surveying and Department of Estate Management were amalgamated and renamed School of Surveying - 1994

First edition April 1983

Reprinted with corrections January 1990

ISBN 0 907382 06 1

ISSN 0260-9142

Abstract

Position-fixing in both the land and marine environment involves two equally important operations: the mathematical combination of the observed data to produce estimates of position, and the study of the errors in the measurements and their propagation through the computational procedure in order to yield the quality of the estimated positions. The usefulness of the results of both operations is largely dependent on the sophistication of the processing methods employed and the current trend is to use increasingly more complicated mathematical procedures.

This Working Paper is concerned with a family of techniques, known generally as least squares, which is now almost universally used for modern position-fixing. The object is to provide a rigorous mathematical background to least squares methods and, at the same time, to identify a number of current, and possible future, applications in the field of position-fixing. Readers are assumed to be conversant with matrix algebra but no prior knowledge of statistics or least squares, or of the general position-fixing methodology, is assumed.

Acknowledgements

Appendix 2 is based on derivations made by Mr A H Fagir.

Table 5.2 is reproduced by permission of Macmillan Publishing Co., Inc., New York. Table 5.3 is reproduced by permission of Chapman and Hall Ltd., London. The author is grateful to the Literary Executor of the late Sir Ronald A Fisher, F.R.S., to Dr Frank Yates, F.R.S. and to Longman Group Ltd., London for permission to reprint Table 5.4, which is taken from Table III of their book Statistical Tables for Biological, Agricultural and Medical Research (6th edition, 1974). Table 5.5 is reproduced by permission of Dover Publications, Inc., New York. Not all of these tables have been reproduced from the original sources and full details of all sources used have been set out in the text (pp 91-94); the cooperation of all authors and publishers involved is acknowledged with thanks.

The author is indebted to both Mr B M Whiting and Mr A H Fagir for reading the original manuscript and for making some useful suggestions. He would also like to thank the Series Editor, Dr A S Walker, for his painstaking editorial work and for his numerous invaluable comments relating to both the content and

the style of this Working Paper. Dr Walker's efforts have substantially increased the quality of this document. It must, however, be emphasised that any mistakes which remain can only be attributed to the author's carelessness or obduracy.

The author would like to thank Mr K Eley for drawing the diagrams and to express particular gratitude to Mrs S A Lacey, who typed a difficult manuscript with dedication, patience and considerable skill.

Contents

Notation

The notation below refers to the general usage of symbols throughout the Working Paper. Where a symbol has only a special localised meaning it is defined in the text and not included in this list.

b	vector of absolute terms in a linearised mathematical model
e	vector of true errors; eastings in 2.5.2
e_i	null vector except for the ith element, which is unity
$\hat{k}$	least squares estimate of the vector of correlatives
ℓ	vector of observed values
$\bar{\ell}$	true values of the vector of observed quantities
$\hat{\ell}$	least squares estimate of $\bar{\ell}$
m	number of parameters
n	number of observations; northings in 2.5.2
r	number of equations in a position-fixing mathematical model
s	scale error or sample standard error (section 5)
v	true values of the vector of residuals
$\hat{v}$	least squares estimate of v
$\bar{x}$	true values of the vector of parameters; mean value (section 5)
x°	approximate values of $\bar{x}$
x	true values of the corrections to x°, $x = \bar{x}-x^\circ$
$\hat{x}$	least squares estimate of x
A,C	design matrices
C_y	covariance matrix of the vector of variates y
E	eastings
$E(y)$	expectation of y
G	Kalman filter gain matrix
H_0	null hypothesis
H_A	alternative hypothesis
I	unit matrix
M	transition matrix (section 8)
N	northings or A^TWA
$P(a \leq y \leq b)$	probability of y being in the range a to b, sometimes written P(a,b)
α	probability of type 1 error (section 5)
β	probability of type 2 error (section 5)

μ	vector of mean values
ν	degrees of freedom
σ	standard error; standard error of population (section 5)
σ_{y_1}	standard error of y_1
$\sigma_{y_1 y_2}$	covariance of y_1 and y_2
σ^2	variance
σ_o^2	unit variance

The following general points apply to the use of matrix algebra.

(i) Upper and lower case letters (usually Roman) are used for matrices and vectors respectively, e.g. Y is a matrix and y is a vector.

(ii) Individual elements are denoted by subscripts, e.g. Y_{ij} and y_i are elements of Y and y respectively.

(iii) Y^T, Y^{-1} and Tr(Y) are the transpose, inverse and trace of Y respectively. The inverse always refers to the Cayley inverse, i.e. $AA^{-1} = I$.

1. Introduction

Position-fixing is concerned with the determination of the coordinates of points on (or possibly above or below) the earth's land and sea surface. It is an activity that is central to the geodetic sciences and one that is required for innumerable purposes. Sometimes single points are fixed, e.g. for navigation and in the provision of photogrammetric control, and at other times groups of points are fixed simultaneously, as in engineering and national control networks. It is convenient to consider position-fixing as having the following steps:

(i) design

(ii) measurement

(iii) mathematical modelling

(iv) estimation

(v) analysis.

The design stage involves decisions upon which measurements to make; often these will be limited by practical considerations but in principle there is a vast choice. Nowadays instrumentation is available to measure a multitude of physical quantities ranging from the traditional angles and distances to frequency shift (e.g. satellite-Doppler), time delay (e.g. from quasars in very long baseline interferometry) and vehicle acceleration (e.g. inertial surveying). The object of the design stage is to select a set of measurements that will yield results of the desired quality with the minimum cost. Once the measurements have been made it is necessary to set up a mathematical model relating the required coordinates and the observations. At this stage the physics of the measurement process and the chosen reference system (and other aspects of geodesy) are taken into account. For instance, if a plane surface mathematical model is used, a distance observed by EDM must be "projected" from the earth to a plane surface via models for both the atmosphere and the reference ellipsoid.

The fourth stage is the statistical estimation of the coordinates and this is of crucial importance if the best possible results are to be derived from the measurements. This point is illustrated by the fact that the massive improvement in the quality of inertial surveying and satellite-Doppler results over the last ten years is entirely due to advances in the adopted processing

methods (the instrumentation has hardly changed). Also, the major errors in many national triangulation networks, e.g. in Great Britain, have been largely eliminated simply by reprocessing the old observations (perhaps along with a few new ones). Fortunately, the current availability of relatively inexpensive computers means that it is possible to apply sophisticated processing methods to almost all position-fixing measurements. The essence of the argument here is that since it is so expensive and time-consuming to make position-fixing measurements, they should be given the best possible treatment.

The final stage, the analysis of the quality of the fix, is, in a sense, the most important. Indeed, it could be argued that coordinates are of no value without some measure of their quality because it would not be known whether or not they were suitable for their intended purpose. The quality of a position-fix is a measure of the probability of its containing errors of a specified size (or the size of an error that can be expected with a specified probability). Since it is usual to class errors in surveying under three headings - random, gross and systematic - it follows that quality must be measured under all three headings. In the cases of random and gross errors the terms precision and reliability are used when assessing the quality of a position-fix.

This working paper is concerned with the mathematical aspects of the foregoing stages (iv) and (v), and to a limited extent stage (i). It is essentially a presentation of a universal methodology for estimating, and analysing the quality of, coordinates (and possibly other quantities relating to the measuring systems and the chosen mathematical model) from any set of measurements. It is important to emphasise at the outset that the method, known to surveyors as least squares, is completely general. In principle it can be applied to all position-fixing problems irrespective of the dimensions of the problem, the number of points being fixed and the number and type of observations. Also,it applies to problems either homogeneous or heterogeneous with respect to the types of observations included. Emphasis throughout this paper will be placed on theory, the main object being to provide readers with a set of "tools" which they can use to solve their own position-fixing problems. Some examples have been included but these are to illustrate either specific detailed points in the text or the breadth of the applicability of a particular "tool". Their simplicity should not obscure the fact that the methodology can be applied to all position-fixing problems and in practice many are more complicated than any of the examples given.

As far as the position-fixing methodology is concerned, this paper is completely self-contained with every result being derived from first principles within the text, or, in two cases, in appendices. No prior knowledge of statistics is assumed and all statistical terms are defined before they are used. Cert n statistical formulae and results, e.g. those pertaining to the sampling distributions, are, however, not derived (although references containing the most relevant proofs are given). Readers are assumed to be proficient in matrix algebra and to be familiar with some standard mathematical results.

To ensure the generality of the approach and the equivalence of the techniques in various sections, some of the terminology may, in places, be slightly different to that with which most practising surveyors are familiar. For instance, the term "adjustment" will not be found in this paper. We prefer to speak of "estimating coordinates" rather than "adjusting observations". This is partly because the work "adjustment" has no place in the language of statistics and partly because it has always been something of a misnomer since observations cannot be changed. Similarly, the work "accuracy" is avoided as it is not well defined for position-fixing problems and the two terms "precision" and "reliability" are preferred.

The title of this paper claims that the subject matter is advanced. Whether or not this is the case is, of course, a matter of opinion. Certainly much of it, especially the statistics, is very elementary, but the word is used because the treatment of least squares is more advanced than that usually presented to first degree (i.e. bachelor's) land and sea surveying students in Great Britain (with the possible exception of those who may specialise in photogrammetry or geodesy).

In order to explain how the various sections of this paper fit together and how they relate to the overall objective, a brief summary of the contents will now be given. Section 2 considers the general form of the position-fixing mathematical model and its linearisation ready for the application of least squares. Then in section 3 the method of least squares is introduced and applied to the linearised general mathematical model. The precision of position-fixing observations and of the results of the least squares computation described in section 3 are discussed in section 4, which includes derivations of all the relevant covariance matrices. Section 5 considers methods for the statistical testing of the results of a least squares computation. This is important for measuring the reliability of a position-

fix and can give useful information on the possibilities of systematic errors, the completeness of the model used in section 2 and the correctness of the covariance matrices assigned in section 3. Section 6 uses the statistical terminology introduced in sections 4 and 5 to justify the least squares method used in section 3.

Section 7 looks at methods for dividing large least squares problems so that they may be handled on modern microcomputers. These methods are especially important nowadays as many modern measuring systems produce vast amounts of data which need to be processed simultaneously, and many applications require large numbers of points to be fixed simultaneously. It should be emphasised here that this paper is not concerned with computer methods as such: only the mathematical aspects of dividing the problem are considered. Similarly, methods for solving systems of linear equations and inverting matrices are not considered.

A number of position-fixing problems, e.g. navigation at sea and inertial surveying, involve parameters (especially position) which vary with time. The general treatment of such problems belongs to the areas of mathematics known as filtering, smoothing and prediction and the principles of these are outlined in section 8, which includes a derivation and discussion of a technique known as Kalman filtering. Section 9 is devoted to the problems of least squares interpolation and collocation. Although these are not usually used to determine positions directly they are included here because (a) they belong to the same "family" of least squares methods as those discussed in the rest of the Working Paper and (b) they have indirect applications in position-fixing, e.g. in the determination and interpolation of coordinate transformation parameters.

Finally, it is proper to make it clear that the intentions of this Working Paper are purely didactic and that it contains nothing new. Its contents have been collected from a large number of sources, most of which are referenced in the text. Some works have, however, had such a great influence on the author that simple references are an insufficient acknowledgement of their importance and the following statements are considered necessary. Wells and Krakiwsky (1971) was used extensively whilst writing sections 2 and 3 and parts of 4, 5, 6 and 7. Pelzer (1979), Sunter (1966) and Gagnon (1976) were used for parts of sections 5, 6 and 7 respectively and section 8 is largely based on Krakiwsky (1976). Section 9 owes a great deal to both Krakiwsky (1976) and

Moritz (1980). Although no specific references are made, Kennedy and Neville (1976) was found to be extremely useful when writing the statistical sections of the text.

2. A general mathematical approach to position-fixing

2.1 The basic mathematical model

The basic data from which all coordinates (and perhaps other parameters) are computed are the observations, so in all position-fixing problems it is logical to consider them first. Hence we begin by defining ℓ as a vector of observations. ℓ may contain such quantities as angles, distances, Doppler counts, time, phase differences, gravity, etc. In some problems we deal with homogeneous data, for example only angles in a resection or only distances in an offshore acoustic fix, but in general problems will be heterogeneous and include a mixture of various types of observations. The true value of the quantities that have been observed are contained in a vector $\bar{\ell}$ where

$$\bar{\ell} = \ell - e \tag{2.1}$$

or

$$\bar{\ell} = \ell + v \tag{2.2}$$

and the vector e contains the true errors of measurement whilst the vector v contains the true residuals. Note that, although errors and residuals are the same but for a change of sign, in practice it is usual to work with residuals and hence to use (2.2) rather than (2.1).

Usually it will be required to estimate a set of parameters (with true values $\bar{x}$) from the observations ℓ. These parameters may well include quantities relating to the observations (e.g. scale errors in EDM, frequency errors in satellite-Doppler, etc.) and to the coordinate system (translation parameters, scale factors, etc.) as well as the coordinates themselves.

There must always be a known mathematical relationship between the true values of the observed quantities and those of the parameters. This relationship constitutes the basic mathematical model and is expressed as a general vector function

$$F(\bar{x}, \bar{\ell}) = 0 \tag{2.3}$$

Note that a vector function simply means a set of r equations

$$f_1(\bar{x}, \bar{\ell}) = 0$$

$$f_2(\bar{x}, \bar{\ell}) = 0$$

$$\vdots \qquad (2.4)$$

$$f_r(\bar{x}, \bar{\ell}) = 0$$

Throughout this paper we will deal with problems where m parameters are to be estimated from n observations with a mathematical model containing r equations. Usually the problems will be redundant, i.e. there will be more observations than are strictly necessary to solve the problem and we will have

$$n \geq r \geq m \qquad (2.5)$$

With r-m being known as the redundancy or degrees of freedom.

Examples of the functions $f_i(\bar{x}, \bar{\ell})$ in (2.4) are as follows:

(a) for an observed bearing, α, from unknown station 1 to unknown station 2 using a plane coordinates model

$$\tan\bar{\alpha} - (\bar{E}_2 - \bar{E}_1)/(\bar{N}_2 - \bar{N}_1) = 0 \qquad (2.6)$$

where $\bar{E}_1$, $\bar{N}_1$, $\bar{E}_2$ and $\bar{N}_2$ would be part of $\bar{x}$ and $\bar{\alpha}$, part of $\bar{\ell}$

(b) for an observed distance, d, with an unknown scale correction, $\bar{s}$, and an unknown index correction, $\bar{i}$, between unknown stations 1 and 2 using a plane coordinate model

$$(\overline{ds} + \bar{i})^2 - \{(\bar{E}_2 - \bar{E}_1)^2 + (\bar{N}_2 - \bar{N}_1)^2\} = 0 \qquad (2.7)$$

where $\bar{E}_1$, $\bar{N}_1$, $\bar{E}_2$, $\bar{N}_2$, $\bar{s}$ and $\bar{i}$ would be part of $\bar{x}$ and $\bar{d}$, part of $\bar{\ell}$

(c) for observed abscissa and ordinate (x_i and y_i) when estimating the gradient ($\bar{m}$) and intercept ($\bar{c}$) of a "best" fitting straight line

$$\bar{y}_i - \bar{m}\bar{x}_i - \bar{c} = 0 \qquad (2.8)$$

where $\bar{m}$ and $\bar{c}$ constitute $\bar{x}$, and $\bar{x}_i$ and $\bar{y}_i$ are part of $\bar{\ell}$

(d) for the estimation of the angles of a triangle ($\bar{\alpha}_1$, $\bar{\alpha}_2$, $\bar{\alpha}_3$) from observed

angles using a plane surface model

$$\bar{\alpha}_1 + \bar{\alpha}_2 + \bar{\alpha}_3 - 180^\circ = 0 \qquad (2.9)$$

where $\bar{\alpha}_1$, $\bar{\alpha}_2$ and $\bar{\alpha}_3$ are part of $\bar{\ell}$ and there are no parameters.

It is important to note the use of the overbar on all the elements of $\bar{x}$ and $\bar{\ell}$ in the above four examples: the basic mathematical model applies to the <u>true</u> values of the observed quantities and parameters.

Some special features of the above examples are also worth noting at this stage.

(i) Examples (a), (b) and (c) are non-linear equations whereas (d) is linear. Note that, although it represents a straight line, (c) is mathematically non-linear due to the product $\overline{mx}_i$.

(ii) Example (b) includes quantities other than coordinates amongst its parameters.

(iii) Examples (a) and (b) could be rewritten in the form

$$f_i(\bar{x}) = \bar{\ell} \qquad (2.10)$$

and example (d) has no parameters and could be rewritten

$$f_i(\bar{\ell}) = 0 \qquad (2.11)$$

but example (c) cannot be simplified and must remain as

$$f_i(\bar{x}, \bar{\ell}) = 0 \qquad (2.12)$$

These three classes of equations are often referred to as "observation equations", "condition equations" and "combined case" respectively.

2.2 Linearisation of the model

For the practical estimation of $\bar{x}$ from (2.3) it is necessary to linearise the basic model (unless, of course, as in example (d) in 2.1, the model is already linear). In mathematics linearisation always necessitates estimating provisional (or approximate) values of the quantities involved. We already have an approximation of $\bar{\ell}$ since we know the observed values ℓ, but we also require provisional values of $\bar{x}$; let these provisional values be x^0 and let $\bar{x}$ be related to x^0 by

$$\bar{x} = x^{o} + x \tag{2.13}$$

where it is now necessary to estimate the small quantities x. Hence, if we substitute (2.13) and (2.2) in (2.3), we can write

$$F(\bar{x}, \bar{\ell}) = F(x^{o} + x, \ell + v) = 0 \tag{2.14}$$

and, applying the Taylor series expansion to first differentials, we obtain

$$F(\bar{x}, \bar{\ell}) = F(x^{o}, \ell) + \frac{\partial F}{\partial \bar{x}} x + \frac{\partial F}{\partial \bar{\ell}} v = 0 \tag{2.15}$$

where the partial derivatives $\frac{\partial F}{\partial \bar{x}}$ and $\frac{\partial F}{\partial \bar{\ell}}$ are evaluated at the points x^{o} and ℓ respectively. It is important to note that, despite the lack of overbars, x and v are the <u>true values</u> of the corrections to the provisional values of the parameters and of the residuals respectively.

Now $F(x^{o}, \ell)$ is a vector which contains the r values of $F(\bar{x}, \bar{\ell})$ computed at the known points x^{o}, ℓ. Let this vector be -b, the negative sign being introduced merely for convenience.

Hence

$$\underset{(r \times 1)}{-b} = \begin{bmatrix} f_1(x^{o}, \ell) \\ f_2(x^{o}, \ell) \\ \vdots \\ f_r(x^{o}, \ell) \end{bmatrix} \tag{2.16}$$

$\frac{\partial F}{\partial \bar{x}}$ is a matrix of order r x m and is denoted by the letter A. The ith row will simply be the partial differentials of $f_i(\bar{x}, \bar{\ell})$ with respect to $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$. Hence

$$\underset{(r \times m)}{A} = \begin{bmatrix} \frac{\partial f_1}{\partial \bar{x}_1} & \frac{\partial f_1}{\partial \bar{x}_2} & \cdots & \frac{\partial f_1}{\partial \bar{x}_m} \\ \frac{\partial f_2}{\partial \bar{x}_1} & \frac{\partial f_2}{\partial \bar{x}_2} & \cdots & \frac{\partial f_2}{\partial \bar{x}_m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_r}{\partial \bar{x}_1} & \frac{\partial f_r}{\partial \bar{x}_2} & \cdots & \frac{\partial f_r}{\partial \bar{x}_m} \end{bmatrix} \tag{2.17}$$

$\frac{\partial F}{\partial \bar{\ell}}$ is a matrix of order r x n and is denoted by the letter C.

The ith row will simply be the partial differentials of $f_i(\bar{x}, \bar{\ell})$ with respect to $\bar{\ell}_1, \bar{\ell}_2, \ldots, \bar{\ell}_n$. Hence

$$\underset{(r \times n)}{C} = \begin{bmatrix} \frac{\partial f_1}{\partial \bar{\ell}_1} & \frac{\partial f_1}{\partial \bar{\ell}_2} & \cdots & \frac{\partial f_1}{\partial \bar{\ell}_n} \\ \frac{\partial f_2}{\partial \bar{\ell}_1} & \frac{\partial f_2}{\partial \bar{\ell}_2} & \cdots & \frac{\partial f_2}{\partial \bar{\ell}_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_r}{\partial \bar{\ell}_1} & \frac{\partial f_r}{\partial \bar{\ell}_2} & \cdots & \frac{\partial f_r}{\partial \bar{\ell}_n} \end{bmatrix} \qquad (2.18)$$

Matrices A and C are often referred to as design matrices and we can write (2.15) as

$$Ax + Cv - b = 0 \qquad (2.19)$$

which is the linearised version of the basic model in (2.3). It is often referred to as the combined case as it is a combination of the following two special cases.

2.3 Special cases

2.3.1 Observation equations

As mentioned in 2.1, if each equation in the basic mathematical model only contains one observed quantity, then (2.3) can be written as

$$F(\bar{x}) - \bar{\ell} = 0 \qquad (2.20)$$

with r = n. Clearly, if we differentiate the ith row of (2.20) with respect to $\bar{\ell}$, we will get zero except for $\frac{\partial f_1}{\partial \bar{\ell}_i} = -1$. Hence

$$\frac{\partial F}{\partial \bar{\ell}} = C = -I \qquad (2.21)$$

and (2.19) becomes

$$Ax - b - v = 0$$

or

$$Ax = b + v \tag{2.22}$$

which is the well known linear model for "observation equations".

2.3.2 Condition equations

If $F(\bar{x}, \bar{\ell})$ is formed without parameters, i.e. in terms of observations only, (2.3) becomes

$$F(\bar{\ell}) = 0 \tag{2.23}$$

with $r = n - m$. At first sight it may seem contradictory to state that the problem has no parameters ($m = 0$) and then to write $r = n - m$. Actually what is meant by m here is the number of parameters that would be used if the same problem were to be solved using observation equations. Clearly, if there are no parameters

$$\frac{\partial F}{\partial \bar{x}} = A = 0 \tag{2.24}$$

and (2.19) becomes

$$Cv - b = 0$$

or

$$Cv = b \tag{2.25}$$

which is the well known model for "condition equations".

2.4 Summary of linear models

The main features of the combined model and its two special cases can be summarised in the following table.

	combined case	observation equations	condition equations
mathematical model	$F(\bar{x}, \bar{\ell}) = 0$	$F(\bar{x}) = \bar{\ell}$	$F(\bar{\ell}) = 0$
no. of equations	r	n	n − m
no. of observations	n	n	n
no. of parameters	m	m	–
linearised model	$Ax + Cv - b = 0$	$Ax = b + v$	$Cv = b$

Interestingly, there is always a choice between the two special cases, i.e. if the combined model does not apply the basic mathematical model can be written down either as a set of observation equations or as a set of condition equations. In practice observation equations are usually easier to form than condition equations. This is because the procedure is more easily automated: one observation will lead to one equation. Hence condition equations are rarely used in practice although they may have some special computational advantages, especially when hand computation is used.

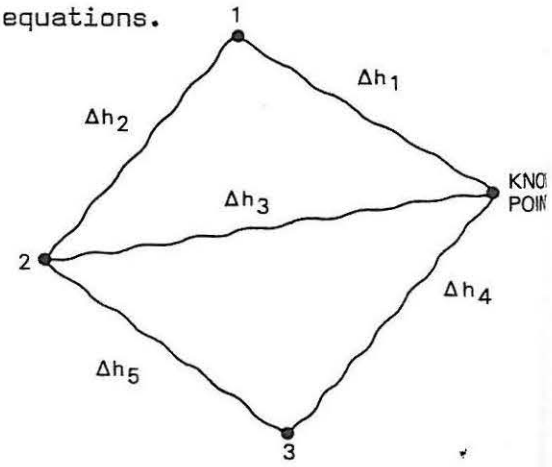
This choice of model is illustrated by the following example of a level network. Note that the linearity of the problem and its geometrical simplici mask the usual difficulties with condition equations.

The observations are

$$\ell = \left[\Delta h_1 \; \Delta h_2 \; \Delta h_3 \; \Delta h_4 \; \Delta h_5\right]^T$$

Let the parameters be the heights of the three unknown stations above the known point:

$$\bar{x} = \left[\bar{x}_1 \; \bar{x}_2 \; \bar{x}_3\right]^T$$



The observation equation model $F(\bar{x}) = \bar{\ell}$ will be the five equations ($n = r = 5, \quad m = 3$)

$$f_1(\bar{x}, \bar{\ell}) = + \bar{x}_1 - \overline{\Delta h}_1 = 0$$

$$f_2(\bar{x}, \bar{\ell}) = - \bar{x}_1 + \bar{x}_2 - \overline{\Delta h}_2 = 0$$

$$f_3(\bar{x}, \bar{\ell}) = - \bar{x}_2 - \overline{\Delta h}_3 = 0$$

$$f_4(\bar{x}, \bar{\ell}) = + \bar{x}_3 - \overline{\Delta h}_4 = 0$$

$$f_5(\bar{x}, \bar{\ell}) = + \bar{x}_2 - \bar{x}_3 - \overline{\Delta h}_5 = 0$$

and the condition equation model $F(\bar{\ell}) = 0$ will be the two equations ($n = 5, \quad r = 2$)

$$f_1(\bar{\ell}) = \overline{\Delta h}_1 + \overline{\Delta h}_2 + \overline{\Delta h}_3 = 0$$

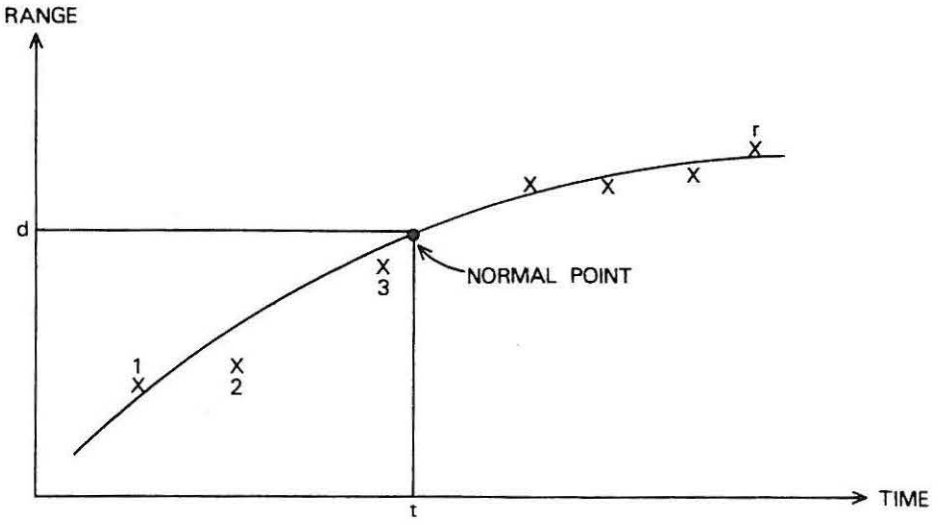
$$f_2(\bar{\ell}) = \overline{\Delta h}_3 + \overline{\Delta h}_4 + \overline{\Delta h}_5 = 0$$

2.5 Examples of position-fixing models

It has been shown that, provided the functional relationship (2.3) between the parameters and observed quantities is known, equations of the type (2.19), (2.22) or (2.25) can be derived for any position-fixing problem simply by differentiation. It makes no difference whether the observations are linear, angular, of frequency or of time (or any other quantity), or whether one, two or three-dimensional coordinates systems are used: the principles are exactly the same. To illustrate these principles three examples have been chosen, one each of the three cases - combined, observation equations and condition equations. Also, Appendix 3 contains numerical worked examples of each of the three cases.

2.5.1 Example of combined case

It is usual when position-fixing using a satellite laser ranging system continuously to track the satellite and observe a very large number of ranges. It would be common to select groups of ranges $d_1, d_2, \ldots, d_r$ measured at times $t_1, t_2, \ldots, t_r$ and reduce them to a single distance d at a specified time t. Such a procedure is often said to produce "normal points" on an orbit.



For this purpose it would be necessary to fit a curve, say a quadratic of the form

$$\bar{d} = \bar{x}_1 + \bar{x}_2\bar{t} + \bar{x}_3\bar{t}^2$$

where $\bar{x}_1$, $\bar{x}_2$ and $\bar{x}_3$ are the true values of the unknown parameters. Hence the basic mathematical model $F(\bar{x}, \bar{\ell})$ is a set of r equations

$$\begin{aligned}
\bar{d}_1 - \bar{x}_1 - \bar{x}_2\bar{t}_1 - \bar{x}_3\bar{t}_1^2 &= 0 \\
\bar{d}_2 - \bar{x}_1 - \bar{x}_2\bar{t}_2 - \bar{x}_3\bar{t}_2^2 &= 0 \\
\vdots \qquad\qquad \vdots \qquad\qquad &\vdots \\
\bar{d}_r - \bar{x}_1 - \bar{x}_2\bar{t}_r - \bar{x}_3\bar{t}_r^2 &= 0
\end{aligned}$$

with m = 3 and n = 2r and the linearised model will be

$$Ax + Cv - b = 0 \qquad (2.19)$$

which must be solved for x and v. Note that in this example the vector of observations ℓ can be arranged as

$$\ell = \left[d_1 \; t_1 \; d_2 \; t_2 \; \cdots \; d_r \; t_r \right]^T$$

and the provisional values, probably determined by fitting a curve to only three points, are

$$x^o = \left[x_1^o \; x_2^o \; x_3^o \right]^T$$

where

$$\bar{x} = x^o + x$$

Then, from (2.17) and (2.18), the design matrices are

$$\underset{(r \times m)}{A} = \frac{\partial F}{\partial \bar{x}} = \begin{bmatrix} -1 & -t_1 & -t_1^2 \\ -1 & -t_2 & -t_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ -1 & -t_r & -t_r^2 \end{bmatrix}$$

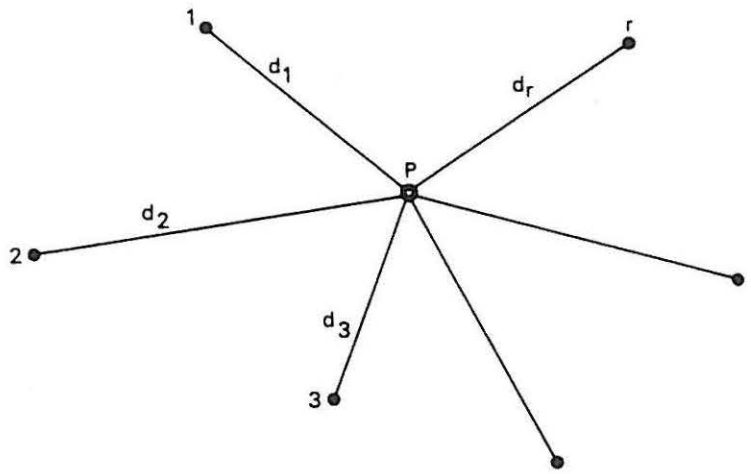
and

$$\underset{(r \times n)}{C} = \frac{\partial F}{\partial \bar{l}} = \begin{bmatrix} 1 & (-x_2^{\circ} - 2x_3^{\circ} t_1) & 0 & 0 & \cdots \\ 0 & 0 & 1 & (-x_2^{\circ} - 2x_3^{\circ} t_2) & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{bmatrix}$$

with the vector of absolute terms given by (from (2.16))

$$\underset{(r \times 1)}{-b} = \begin{bmatrix} d_1 - x_1^{\circ} - x_2^{\circ} t_1 - x_3^{\circ} t_1^2 \\ d_2 - x_1^{\circ} - x_2^{\circ} t_2 - x_3^{\circ} t_2^2 \\ \\ \cdot \\ \cdot \\ \cdot \\ \\ d_r - x_1^{\circ} - x_2^{\circ} t_r - x_3^{\circ} t_r^2 \end{bmatrix}$$

2.5.2 Observation equations example



Consider a point P of unknown coordinates E and N being fixed by distance measurements d_1, d_2, ..., d_r to a number of stations 1, 2, ..., r with known coordinates E_1, N_1, E_2, N_2, etc. If the distances are thought to have a scale error s (due perhaps to the unknown velocity of propagation of the measurement signal), then there are three parameters (m = 3) to be determined and we have

$$\bar{x} = \begin{bmatrix} \bar{E} & \bar{N} & \bar{s} \end{bmatrix}^T$$

Assuming the distances to have been "projected" to the horizontal plane, the basic observation equation mathematical model $F(\bar{x}) - \bar{\ell} = 0$ can be written as the set of r equations

$$\left[(\bar{E} - E_1)^2 + (\bar{N} - N_1)^2\right]^{\frac{1}{2}}/\bar{s} - \bar{d}_1 = 0$$

$$\left[(\bar{E} - E_2)^2 + (\bar{N} - N_2)^2\right]^{\frac{1}{2}}/\bar{s} - \bar{d}_2 = 0$$

$$\vdots$$

$$\left[(\bar{E} - E_r)^2 + (\bar{N} - N_r)^2\right]^{\frac{1}{2}}/\bar{s} - \bar{d}_r = 0$$

If we take provisional values

$$x^o = \left[E^o \quad N^o \quad s^o\right]^T$$

and put

$$d_i{}^o = ((E^o - E_i)^2 + (N^o - N_i)^2)^{\frac{1}{2}}$$

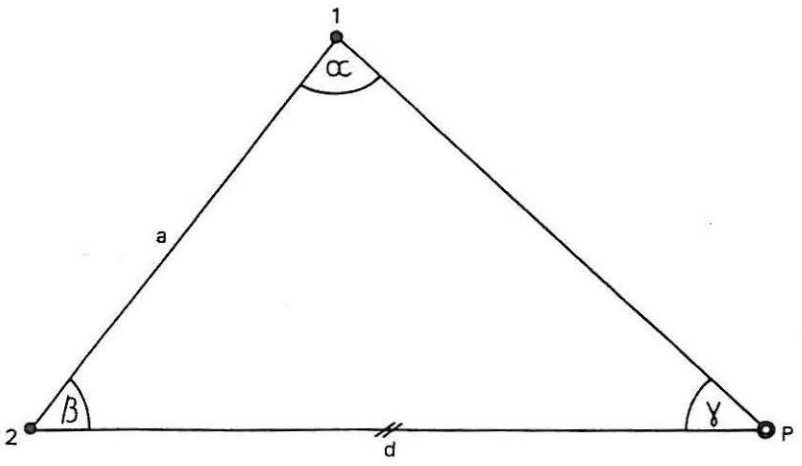
the design matrix will be given by (2.17) as

$$\underset{(n \times m)}{A} = \begin{bmatrix} (E^o - E_1)/(d_1{}^o s^o) & (N^o - N_1)/(d_1{}^o s^o) & -d_1{}^o/(s^o)^2 \\ (E^o - E_2)/(d_2{}^o s^o) & (N^o - N_2)/(d_2{}^o s^o) & -d_2{}^o/(s^o)^2 \\ \vdots & \vdots & \vdots \\ (E^o - E_r)/(d_r{}^o s^o) & (N^o - N_r)/(d_r{}^o s^o) & -d_r{}^o/(s^o)^2 \end{bmatrix}$$

and the vector of absolute terms will be given by (2.16) as

$$\underset{(n \times 1)}{-b} = \begin{bmatrix} d_1{}^o/s^o - d_1 \\ d_2{}^o/s^o - d_2 \\ \vdots \\ d_r{}^o/s^o - d_r \end{bmatrix}$$

2.5.3 Condition equations example



A point P is fixed from two known stations 1 and 2 (with coordinates E_1, N_1 and E_2, N_2 respectively) by measurement of three angles α, β and γ and the distance d. Hence the vector of observations is

$$\ell = \left[\alpha\ \beta\ \gamma\ d\right]^T, \quad n = 4$$

If the problem were solved by the observation equations method there would be two parameters, the easting and northing of P. Therefore m = 2 and the number of condition equations, $F(\bar{\ell}) = 0$, is

$$r = n - m = 2$$

These equations are found by inspection from the diagram and are

$$\bar{\alpha} + \bar{\beta} + \bar{\gamma} - 180^\circ = 0$$

$$\bar{d}/\sin\bar{\alpha} - a/\sin\bar{\gamma} = 0$$

where a is the known distance between the two fixed stations.

The design matrix is given by (2.18) as

$$\underset{(r \times n)}{C} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ -d\cot\alpha\operatorname{cosec}\alpha & 0 & a\cot\gamma\operatorname{cosec}\gamma & \operatorname{cosec}\alpha \end{bmatrix}$$

and the vector of absolute terms is given by (2.16) as

$$\underset{(r \times 1)}{-b} = \begin{bmatrix} \alpha + \beta + \gamma - 180^{\circ} \\ d/\sin\alpha - a/\sin\gamma \end{bmatrix}$$

Note that in practice this type of problem is usually more conveniently solved by the observation equations method because of the difficulty of automating the procedure whereby the condition equations are selected. This difficulty arises because it is necessary to examine the complete set of observations simultaneously in order to extract the condition equations whereas observation equations can be determined consecutively (each observation leads to one observation equation).

3. The combined least squares process

It has been shown that all position-fixing problems can be expressed as a system of linear equations

$$Ax + Cv - b = 0 \qquad (3.1)$$

from which the residuals, v, and parameters, x, must be determined; in other words it is required to derive expressions of the form

$$x = f_1 (A, C, b) \qquad (3.2)$$

$$v = f_2 (A, C, b) \qquad (3.3)$$

which satisfy (3.1).

Clearly once (3.2) and (3.3) have been found we will be able to consider observation equations and condition equations as special cases by putting $C = -I$ and $A = 0$ respectively.

In practice there will usually be more measurements than are strictly necessary to solve the equations and, owing to observational errors, there will be an infinite number of possible solutions to (3.1). This point is best seen by looking at the special case of observation equations:

$$Ax = b + v \qquad (3.4)$$

Here it would be feasible to choose any arbitrary set of values for x, say x^*, and solve for v^* from

$$v^* = Ax^* - b \qquad (3.5)$$

Hence any choice of x^* leads to a set of residuals v^* and unless the observations are perfect it will not be possible to "choose" the true value of x. So, whatever computational process is adopted, it can only produce an estimate of x.

In this paper we will use a computational process known as least squares and we will use the notation $\hat{x}$ and $\hat{v}$ to denote the least squares estimates of x and v respectively.

The least squares estimates are <u>defined</u> as those which minimise a specified quadratic form of the residuals:

$$v^T W v = \text{minimum} \qquad (3.6)$$

where, W is the inverse of C_ℓ, the covariance matrix of the observations, i.e.

$$W = C_{\ell}^{-1} \tag{3.7}$$

The term "covariance matrix" is defined in 4.1 and the problem of assigning values to the variances and covariances in C_{ℓ} is discussed in 4.2. It is worth noting here that the term "least squares" is used because in the special case where all observations are uncorrelated and have the same variance, σ^2, then

$$W = \frac{1}{\sigma^2} I \tag{3.8}$$

and the quadratic form (3.6) can be simplified to

$$\frac{1}{\sigma^2} v^T v = \text{minimum}$$

i.e.
$$\frac{1}{\sigma^2} (v_1^2 + v_2^2 + \ldots v_n^2) = \text{minimum} \tag{3.9}$$

and the "sum of the squares of the residuals" is minimised. A more correct description of the general case of (3.6) would be "minimum quadratic form" but the term "least squares" is universally used and will be retained for this paper.

It ought to be noted here that many texts on statistics differentiate between least squares and weighted least squares, i.e. between $v^T v$ and $v^T W v$. In this paper least squares always means the weighted least squares of statistics texts.

This definition of least squares will now be used to derive explicit expressions of the form (3.2) and (3.3) for $\hat{x}$ and $\hat{v}$. Before doing so, however, it is necessary to outline the basic philosophical approach underlying this, and to mention an equally acceptable alternative. In this paper the least squares process is <u>defined</u> as in (3.6) and (3.7) and then in section 6 its application will be justified by <u>deriving</u>, and then analysing, the statistical properties of the least squares estimates. It will be seen that least squares estimates are, in a certain sense, the "best" estimates. An alternative approach would be to <u>define</u> the statistical properties required of the adopted estimates and then to derive expressions (3.2) and (3.3) to yield estimates with these properties. Clearly, if this approach were adopted, and if we started with the statistical properties of the least squares estimates, we would in fact then <u>derive</u> the least squares process rather than define it as has been done in this paper. The approach adopted is not important, since expressions (3.2) and (3.3) will result in either case.

We now have a straightforward problem of mathematical optimisation, i.e. to find expressions for x and v which make

$$v^T W v = \text{minimum} \qquad (3.6)$$

subject to the constraints

$$Ax + Cv - b = 0 \qquad (3.10)$$

where A, C, b and W are all known.

To solve this problem it is usual to use one of the standard methods of mathematical optimisation, called Lagrange's Method of Undetermined Multipliers. Lagrange showed that if we have a function φ which must be optimised (maximised or minimised), subject to constraints φ_1, φ_2, ..., φ_r, being zero, then the solution is found by optimising a function

$$\Phi = \varphi + p_1\varphi_1 + p_2\varphi_2 + p_3\varphi_3 + \ldots + p_r\varphi_r \qquad (3.11)$$

where p_1, p_2, ..., p_r are unknowns called "undetermined multipliers". A proof of this proposition is given in Krabs (1979, 179) and many other mathematics texts.

For the least squares problem it is convenient to put $p_1 = 2k_1$, $p_2 = 2k_2$, etc., where k_1, k_2, etc. are called "correlatives". If we denote the vector of correlatives by k, then substitution of (3.6) and (3.10) into (3.11) gives

$$\Phi = v^T W v + 2k^T (Ax + Cv - b) \qquad (3.12)$$

which must be minimised. This is achieved by equating the partial derivatives of (3.12) with respect to x and v to zero. Hence

$$\frac{\partial \Phi}{\partial x} = 2A^T\hat{k} = 0 \qquad (3.13)$$

$$\frac{\partial \Phi}{\partial v} = 2W\hat{v} + 2C^T\hat{k} = 0 \qquad (3.14)$$

Notice that as soon as (3.12) has been differentiated and the resulting expressions equated to zero the notation $\hat{x}$, $\hat{v}$ and $\hat{k}$ is introduced to denote the least squares estimates of x, v and k respectively. In principle this could be done in (3.12) but there would be a difficulty with the partial differentiation because $\hat{x}$, $\hat{v}$ and $\hat{k}$ are clearly related so making $\partial\hat{v}/\partial\hat{x}$, $\partial\hat{k}/\partial\hat{x}$ etc. non-zero. There is no problem with (3.12) as it stands because x, the true corrections to the approximate values, is clearly independent of v, the true residuals of the observations.

Hence we need to find values for $\hat{x}$ and $\hat{v}$ which simultaneously satisfy (3.10), (3.13) and (3.14), i.e.

$$\begin{aligned} A\hat{x} + C\hat{v} &= b \\ A^T\hat{k} &= 0 \\ W\hat{v} + C^T\hat{k} &= 0 \end{aligned}$$

which can be rearranged into the "hypermatrix"

$$\begin{bmatrix} W & C^T & 0 \\ C & 0 & A \\ 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{v} \\ \hat{k} \\ \hat{x} \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix} \qquad (3.15)$$

or

$$Py = u \qquad (3.16)$$

which could, in principle, be solved by one of the standard linear algebra techniques for solving simultaneous linear equations to yield

$$y = P^{-1}u$$

Such a solution would, however, involve an unnecessarily large amount of work as P is a matrix of size n + r + m. In practice it is far easier to use (3.15) to derive explicit solutions for $\hat{x}$ and $\hat{v}$. We start by deriving general formulae for the elimination of unknowns from a partitioned set of equations. (3.16) is partitioned as

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \qquad (3.17)$$

which can be multiplied out to give

$$P_{11}\, y_1 + P_{12}\, y_2 = u_1 \qquad (3.18)$$

$$P_{21}\, y_1 + P_{22}\, y_2 = u_2 \qquad (3.19)$$

(3.18) is rearranged to give

$$y_1 = P_{11}^{-1}(u_1 - P_{12}\, y_2) \qquad (3.20)$$

which can be substituted into (3.19) to give

$$P_{21}\, P_{11}^{-1}(u_1 - P_{12}\, y_2) + P_{22}\, y_2 = u_2$$

or

$$(P_{22} - P_{21}\,P_{11}^{-1}\,P_{12})\,y_2 = (u_2 - P_{21}\,P_{11}^{-1}\,u_1) \qquad (3.21)$$

If (3.15) is partitioned as follows

$$\left[\begin{array}{c|cc} W & C^T & 0 \\ \hline C & 0 & A \\ 0 & A^T & 0 \end{array}\right]\left[\begin{array}{c} \hat{v} \\ \hline \hat{k} \\ \hat{x} \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline b \\ 0 \end{array}\right]$$

and (3.21) is applied we obtain

$$\left[\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} - \begin{bmatrix} C \\ 0 \end{bmatrix} W^{-1} \begin{bmatrix} C^T & 0 \end{bmatrix}\right]\begin{bmatrix} \hat{k} \\ \hat{x} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \qquad (3.22)$$

(because $u_1 = 0$), which leads to

$$\left[\begin{array}{c|c} -CW^{-1}C^T & A \\ \hline A^T & 0 \end{array}\right]\left[\begin{array}{c} \hat{k} \\ \hline \hat{x} \end{array}\right] = \left[\begin{array}{c} b \\ \hline 0 \end{array}\right] \qquad (3.23)$$

(3.23) is partitioned as indicated and (3.21) applied to give

$$(0 - A^T(-CW^{-1}C^T)^{-1}A)\hat{x} = 0 - A^T(-CW^{-1}C^T)^{-1}b$$

or

$$\left[A^T(CW^{-1}C^T)^{-1}A\right]\hat{x} = A^T(CW^{-1}C^T)^{-1}b \qquad (3.24)$$

which are the normal equations for the combined least squares problem. If (3.24) is written in the form

$$\hat{x} = \left[A^T(CW^{-1}C^T)^{-1}A\right]^{-1}A^T(CW^{-1}C^T)^{-1}b \qquad (3.25)$$

it can be seen that it is the explicit solution (3.2) for which we were searching. The residuals are obtained via the correlatives. Applying (3.20) to (3.23) gives

$$\hat{k} = -(CW^{-1}C^T)^{-1}(b - A\hat{x}) \qquad (3.26)$$

which can be substituted into (3.14) as follows

$$\hat{v} = -W^{-1}C^T\hat{k} \qquad (3.27)$$

giving

$$\hat{v} = W^{-1}C^T(CW^{-1}C^T)^{-1}(b - A\hat{x}) \qquad (3.28)$$

Substituting (3.25) into (3.28) yields the desired explicit expression of the form (3.3), viz.

$$\hat{v} = W^{-1}C^T(CW^{-1}C^T)^{-1}\{I - A\left[A^T(CW^{-1}C^T)^{-1}A\right]^{-1}A^T(CW^{-1}C^T)^{-1}\}b \qquad (3.29)$$

To solve (3.24) for $\hat{x}$ involves the inversion of $CW^{-1}C^T$, which is of size r, followed by the solution of m simultaneous equations. The correlatives and residuals are obtained by matrix multiplication from (3.26) and (3.27) respectively; no further inversions or solutions are necessary.

3.1 Alternative derivation

Weightman (1982) has given the following alternative derivations for equations (3.25) and (3.28), which have the advantages of being far shorter than the foregoing and of not requiring recourse to correlatives and Lagrange's Method of Undetermined Multipliers. They do, however, involve two substitution (equations (3.32) and (3.33) below) that would be very difficult to find without prior knowledge of the equations being sought.

We begin by rewriting (3.10) as

$$-Cv = Ax - b \tag{3.30}$$

Then, premultiplying and postmultiplying a matrix M, as yet unspecified, by both sides of (3.30), yields

$$(Cv)^T MCv = (Ax - b)^T M(Ax - b) \tag{3.31}$$

Now, putting

$$M = (CW^{-1}C^T)^{-1} \tag{3.32}$$

and introducing a vector q such that

$$v = W^{-1}C^Tq \tag{3.33}$$

we simplify the left hand side of (3.31) to

$$\begin{aligned}(Cv)^T MCv &= q^TCW^{-1}C^T\,(CW^{-1}C^T)^{-1}CW^{-1}C^Tq \\ &= q^TCW^{-1}C^Tq \\ &= q^TCW^{-1}WW^{-1}C^Tq \\ &= v^T Wv \text{ (because of (3.33))}\end{aligned} \tag{3.34}$$

Then substituting (3.34) in (3.31) and expanding gives

$$v^TWv = x^TA^TMAx - b^TMAx - x^TA^TMb + b^TMb \tag{3.35}$$

For least squares we need to minimise v^TWv. So differentiating (3.35) whilst remembering that M is symmetrical gives

$$\frac{\partial}{\partial x}(v^T W v) = 2A^T MA\hat{x} - A^T Mb - A^T Mb = 0 \qquad (3.36)$$

hence

$$\hat{x} = (A^T MA)^{-1} A^T Mb \qquad (3.37)$$

and substituting (3.32) in (3.37) yields

$$\hat{x} = \left[A^T (CW^{-1}C^T)^{-1} A\right]^{-1} A^T (CW^{-1}C^T)^{-1} b \qquad (3.38)$$

which is identical to (3.25).

Now to determine the least squares estimates of the residuals we proceed as follows. Substitute (3.33) in (3.30) to yield

$$-CW^{-1}C^T\hat{q} = A\hat{x} - b$$

$$\hat{q} = (CW^{-1}C^T)^{-1} (b - A\hat{x}) \qquad (3.39)$$

Then substitute (3.39) in (3.33) to give

$$\hat{v} = W^{-1}C^T(CW^{-1}C^T)^{-1}(b - A\hat{x}) \qquad (3.40)$$

which is identical to (3.28).

3.2 Special cases

The results for the combined case can now be simplified to produce explicit expressions (3.2) and (3.3) for the special cases of observation equations and condition equations.

3.2.1 Observation equations

Putting C = -I yields, from (3.24), the normal equations

$$(A^T WA)\hat{x} = A^T Wb$$

with solution, from (3.25),

$$\hat{x} = (A^T WA)^{-1} A^T Wb \qquad (3.41)$$

from (3.26)

$$\hat{k} = -W(b - A\hat{x}) \qquad (3.42)$$

and, from (3.28)

$$\hat{v} = -W^{-1}W(b - A\hat{x}) = A\hat{x} - b \qquad (3.43)$$

The major computational part is the solution of m simultaneous equations.

3.2.2 Condition equations

Putting A = 0 yields, from (3.26), the normal equations

$$(CW^{-1}C^T)\hat{k} = -b$$

with solution

$$\hat{k} = -(CW^{-1}C^T)^{-1}b \qquad (3.44)$$

and, from (3.28),

$$\hat{v} = W^{-1}C^T(CW^{-1}C^T)^{-1}b \qquad (3.45)$$

The major computational part is the solution of (n - m) simultaneous equations.

3.3 Summary of formulae

The formulae for the least squares estimates of x and v for the combined problem and its two special cases can now be summarised as follows

(i) for the parameters

combined case $\hat{x} = [A^T(CW^{-1}C^T)^{-1}A]^{-1}A^T(CW^{-1}C^T)^{-1}b$

observation equations $\hat{x} = (A^TWA)^{-1}A^TWb$

(ii) for the residuals

combined case $\hat{v} = W^{-1}C^T(CW^{-1}C^T)^{-1}(b - A\hat{x})$

observation equations $\hat{v} = A\hat{x} - b$

condition equations $\hat{v} = W^{-1}C^T(CW^{-1}C^T)^{-1}b$

It should be remarked that the above are the algebraic forms of the expression and they do not necessarily indicate the best way to compute $\hat{x}$ or $\hat{v}$ for a practical problem.

3.4 Numerical checks

The accuracy of the numerical solution of the combined case of least squares can be determined by checking the proximity of $A^T(CW^{-1}C^T)^{-1}C\hat{v}$ to a null vector. A proof of this now follows.

From (3.28)

$$A^T(CW^{-1}C^T)^{-1}C\hat{v} = A^T(CW^{-1}C^T)^{-1}CW^{-1}C^T(CW^{-1}C^T)^{-1}(b - A\hat{x})$$

$$= A^T(CW^{-1}C^T)^{-1}b - A^T(CW^{-1}C^T)^{-1}A\hat{x}$$

$$= 0 \text{ (from (3.24))} \qquad (3.46)$$

For the special cases of observation equations the check is

$$A^TW\hat{v} = 0 \text{ (as above with } C = -I) \qquad (3.47)$$

and for the special case of condition equations it is usual to return to the linear model and check

$$C\hat{v} - b = 0 \qquad (3.48)$$

These checks ensure the correct computation and solution of the normal equations and the correct computation of the residuals. It is important to realise that they do not check the proper setting up and linearisation of the basic mathematical model, i.e. they do not check the correctness of (2.3) and (2.19).

3.5 Accuracy of approximate values

If the approximate values used to obtain the numerical values of the elements of the A and C matrices are not close to the final least squares estimates then the equations

$$Ax + Cv - b = 0 \qquad (3.1)$$

will not be a true linearisation of the functional model

$$f(\bar{x}, \bar{l}) = 0 \qquad (2.3)$$

In such cases it is necessary to iterate the least squares process using the least squares estimates from the ith computation as approximate values for the (i + 1)th computation. The iteration is stopped when the vectors of parameters and residuals change by insignificant amounts. In practice such a "convergence" is usually achieved very rapidly and a complete failure to converge only occurs when using nonsensical approximate values. Cross (1981a) includes a discussion of the convergence problem for the computation of individual positions at sea.

This iterative process is very easy to program and is incorporated in most modern software. In cases where it is not desirable, such as for hand

computations or in the methods described in sections 7 and 8, great care must be taken to ensure that approximate values are close to the final least squares estimates. This is best done by deriving them from a preliminary computation using a selection of the observed data. As a general guide Bomford (1980) quotes that so long as coordinates are approximated such that the resulting errors in azimuths and distances are less than 1 minute of arc and 1 in 4000 respectively iteration should not be necessary.

4. The estimation and interpretation of covariance matrices

As explained in section 1, one of the three qualities of a position-fix that needs to be assessed is its precision. This is most conveniently done by the use of covariance matrices (sometimes called variance-covariance matrices). Moreover it is apparent from (3.7) that it is necessary to specify the covariance matrix of the observations, C_{ℓ}, in order to determine the weight matrix, W, needed to compute the least squares estimates of the parameters and residuals. Hence this section deals with the specification of C_{ℓ} and the propagation of random errors through the least squares process to enable the computation of $C_{\hat{x}}$, $C_{\hat{v}}$ and $C_{\hat{\ell}}$, the covariance matrices of the least squares estimates of the parameters, residuals and observed quantities respectively. Finally, methods for abstracting useful information from $C_{\hat{x}}$ are discussed.

In order to keep this paper as far as possible "self-contained" from a statistical point of view, 4.1 contains some basic statistical definitions. It is worth noting that these are the only definitions needed to study the precision of survey measurements and of least squares estimates.

4.1 Some statistical definitions

(i) probability density function (pdf)

The probability density function, $\varphi(y_i)$, of a variate y_i is the function whose integral gives the probability, P(a,b), of y_i lying in the range a to b (see Fig. 4.1), i.e.

$$P(a,b) = \int_a^b \varphi(y_i)\,dy_i \qquad (4.1)$$

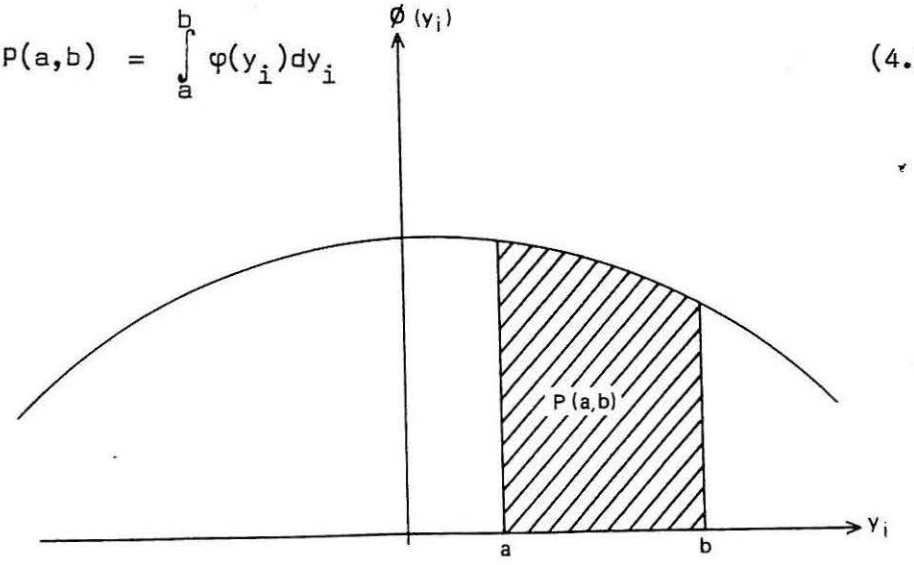


Fig. 4.1

A probability of one is taken to mean that an event is certain to occur, so we may write

$$P(-\infty,\infty) = \int_{-\infty}^{\infty} \varphi(y_1)dy_i = 1 \qquad (4.2)$$

When dealing with more than one variate we use the term "multivariate pdf" to define a function whose integral gives the probability of y_1 lying in the range a_1 to b_1 at the same time as y_2 lies in the range a_2 to b_2, y_3 lies in the range a_3 to b_3, etc., i.e.

$$P = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\int_{a_3}^{b_3} \cdots \varphi_1(y_1)\varphi_2(y_2)\varphi_3(y_3) \cdots dy_1dy_2dy_3 \cdots$$

or

$$P = \int_{a}^{b} \Phi(y)dy \qquad (4.3)$$

where

$$y = \left[y_1\ y_2 \cdots\right]^T$$

$$a = \left[a_1\ a_2 \cdots\right]^T$$

$$b = \left[b_1\ b_2 \cdots\right]^T$$

(ii) <u>expected value and mean</u>

The expected value, or expectation, of a function $f(y_i)$ is the arithmetic average of $f(y_i)$ according to the pdf of y_i, i.e.

$$E\left[f(y_i)\right] = \int_{-\infty}^{\infty} f(y_i)\varphi(y_i)dy_i \qquad (4.4)$$

The mean, μ_i, of a variate, y_i, is the expected value of the variate itself, i.e.

$$y_i = f(y_i)$$

$$\mu_i = E(y_i) = \int_{-\infty}^{\infty} y_i\varphi(y_i)dy_i \qquad (4.5)$$

For a multivariate pdf the expected value is

$$E\left[f(y)\right] = \int_{-\infty}^{\infty} f(y)\Phi(y)dy \qquad (4.6)$$

and the mean is

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} E(y_1) \\ E(y_2) \\ E(y_3) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = E\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = E(y) \qquad (4.7)$$

The following rules apply to expected values (they follow from (4.6)):

$$E(ky) = kE(y) \text{ for non-stochastic } k \qquad (4.8)$$

$$E(y + z + t \ldots) = E(y) + E(z) + E(t) + \ldots \qquad (4.9)$$

(iii) unbiased estimate

An estimate of y, say y*, is said to be unbiased if its expected value is equal to y, i.e.

$$\text{if} \quad E(y^*) = y \qquad (4.10)$$

(iv) covariance matrix

The covariance matrix of y, C_y, is defined as

$$C_y = E\left[(y - \mu)(y - \mu)^T\right] \qquad (4.11)$$

i.e.

$$C_y = E\left(\begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n - \mu_n \end{bmatrix} \begin{bmatrix} y_1 - \mu_1, & y_2 - \mu_2, & \ldots, & y_n - \mu_n \end{bmatrix} \right)$$

or

$$C_y = \begin{bmatrix} \sigma^2_{y_1} & \sigma_{y_1y_2} & \cdot & \cdot & \cdot & \cdot & \cdot & \sigma_{y_1y_n} \\ \sigma_{y_2y_1} & \sigma^2_{y_2} & \cdot & \cdot & \cdot & \cdot & \cdot & \sigma_{y_2y_n} \\ \cdot & \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & & \cdot & & & & \cdot \\ \cdot & \cdot & & & \cdot & & & \cdot \\ \cdot & \cdot & & & & \cdot & & \cdot \\ \cdot & \cdot & & & & & \cdot & \cdot \\ \sigma_{y_ny_1} & \sigma_{y_ny_2} & \cdot & \cdot & \cdot & \cdot & \cdot & \sigma^2_{y_n} \end{bmatrix} \qquad (4.12)$$

where $\sigma^2_{y_i} = E\left[(y_i - \mu_i)^2\right]$ is called the <u>variance</u> of y_i

and $\sigma_{y_i y_j} = E\left[(y_i - \mu_i)(y_j - \mu_j)\right]$ is called the <u>covariance</u> of y_i and

The positive square root of the variance of y_i, σ_{y_i}, is referred to as the <u>standard error</u> (or <u>standard deviation</u>) of y_i. Note that the terms standard error and standard deviation are synonymous.

The <u>coefficient of correlation</u> between y_i and y_j, ρ_{ij}, is defined as

$$\rho_{ij} = \sigma_{y_i y_j} / (\sigma_{y_i} \sigma_{y_j}) \tag{4.13}$$

note that $-1 \leq \rho_{ij} \leq 1$.

(v) <u>propagation of variances and covariances</u>

If the vector of variates z is related to the vector of variates y by the deterministic relationship

$$z = Ry \tag{4.14}$$

the relationship between the variances and covariances of z and y can be derived as follows. Note that the term "deterministic" used above simply means that the relationship between z and y does not depend on chance (R is non-stochastic) although the elements of the vectors z and may themselves depend on chance (z and y are stochastic).

We can write, from (4.7) and (4.8),

$$\mu_z = E(z) = E(Ry) = RE(y) = R\mu_y \tag{4.15}$$

Hence, using (4.11)

$$C_z = E\left[(z - \mu_z)(z - \mu_z)^T\right]$$

Then, from (4.14) and (4.15),

$$\begin{aligned} C_z &= E\left[(Ry - R\mu_y)(Ry - R\mu_y)^T\right] \\ &= RE\left[(y - \mu_y)(y - \mu_y)^T\right]R^T \\ &= R\, C_y\, R^T \end{aligned} \tag{4.16}$$

(4.16) is often referred to as Gauss's propagation of error law for linear equations.

4.2 Covariance matrix of the observations

Before a least squares computation or analysis can take place the covariance matrix of the observations, C_ℓ (often called the "a priori" covariance matrix), has to be estimated. The word "estimated" has been carefully chosen here in order to emphasise that it is necessary for the surveyor carrying out the computation to make a personal assessment of the variances and covariances of the observed quantities. Although there are a number of statistical tools that can be used to aid this assessment, and to check its correctness after the computation of the residuals, the subjective nature of the operation can never be completely removed.

In this section methods for estimating the variances and covariances will be discussed. Their testing for possible subsequent alteration is discussed in section 5.

4.2.1 Estimation of variances

4.2.1.1 Repeated measurements

It can be shown, e.g. by Cooper (1974, 25), that, if a measurement has been repeated n times with results $\ell_1, \ell_2, \ldots, \ell_n$ and mean m_ℓ, then an unbiased estimate of the variance of m_ℓ is given by

$$\sigma^2 = v^T v / \left[n(n-1) \right] \qquad (4.17)$$

where

$$v = \left[v_1 \; v_2 \; \ldots \; v_n \right]^T$$

and

$$v_i = \ell_i - m_\ell$$

Hence in situations where multiple measurements of the same quantity have been made (e.g. rounds of angles, repeated EDM distances), (4.17) can be used to obtain variances directly. It must, however, be pointed out that use of (4.17) (often referred to as the use of internal evidence) has a serious drawback when, as is usually the case, simply repeating the measurement does not involve resampling all the sources of random error. For instance, when a direction is measured with a theodolite, the following are amongst the possible sources of random error:

(i) reading the scale

(ii) bisecting the mark

(iii) centering

(iv) levelling

(v) lateral refraction.

Errors due to sources (iii), (iv) and (v) above would not be reflected in multiple rounds from the same theodolite "set-up" so (4.17) would obviously give too optimistic (variance too small) a measure of the precision of an observed angle. In other words, errors due to (iii), (iv) and (v) are systematic during the repeated measuring process. They will not usually be systematic, however, as far as the position-fix computation is concerned because they are, in general, random from station to station.

4.2.1.2 External comparisons

A better way to estimate the precision of a measuring process is to compare its results with some known values (either true values or values derived from another measuring process that is significantly more precise than the one being used). For instance, with theodolite angle (or direction) measurements it is known that the three angles (or differences from six directions) in a plane triangle should sum to 180°. For a network with n triangles Bomford (1980, 164) quotes Ferrero's formula for the variance of an observed angle, σ_α^2, and observed direction, σ_β^2, as

$$\sigma_\alpha^2 = \Delta^T\Delta/3n \qquad (4.18)$$

$$\sigma_\beta^2 = \Delta^T\Delta/6n \qquad (4.19)$$

where

$\Delta = \left[\Delta_1\ \Delta_2\ \ldots\ \Delta_n\right]^T$ is a vector of triangular misclosures

Ashkenazi et al. (1972) quote an example of the application of (4.19) to the triangular misclosures of the retriangulation of Great Britain. The result is a figure of 0".65 for the standard error of a direction derived from the mean of sixteen rounds. If (4.17) is used with the same data a value of between 0".1 and 0".2 would be derived, illustrating the point made in 4.2.1.1 that the study of repeated measurements will invariably lead to an underestimation of their variances. An important corollary to this is that large numbers of repetitions are usually of little value beyond guarding against gross errors.

For instance, with the foregoing example of direction measurement the major source of error is lateral refraction and no matter how large the theodolite horizontal circle and how skilled the observer (so long as they are both of 1st order geodetic standard) a standard error of 0".65 is to be expected. It may even be argued that two rounds of measurements should be sufficient (the second merely for a check on gross errors). Note that this argument would need to be modified if the repetitions sampled different atmospheric conditions, e.g. some during the day, some at night, etc.

Similarly to (4.18) and (4.19), distance measuring systems are often calibrated on a baseline and an analysis of the calibration readings can lead to a figure for the variance of the equipment, e.g. Ashkenazi and Dodson (1977).

Even these external comparison methods have their drawbacks. For instance, many surveying organisations specify "allowable triangular misclosures" for their field parties and insist that any angles failing to satisfy the criteria be reobserved. If the criteria are too strict, perfectly good angles (in the statistical sense) may be interpreted as blunders and reobserved until, by chance, they cause triangular misclosures to be within the set limits. Hence it is inevitable that (4.18) and (4.19) give small values for the angular variances even though reobservation may cause distortions. Also baseline calibration of distance measuring instruments cannot always gauge how they will perform over much longer distances in different atmospheric environments.

4.2.1.3 Previous performance

In situations where neither external comparisons nor repeated measurements are possible the only way to estimate the variances of observations is to look at evidence of the previous performance of the measuring system, such as may be found in manufacturers' literature or in scientific papers. In fact variances estimated in this manner are likely to be closer to the truth than those estimated from repeated measurements, although repeated measurements may indicate the relative variances of a number of measurements made with the same system.

4.2.1.4 General remark

It is important to remark that when estimating the variances of individual observations we are actually trying to determine statistics (see 5.1) of the infinite populations of errors from which the errors in our measurements have

been drawn. We are not attempting to estimate the actual size of the measurement errors. Hence when a number of measurements have been made using similar instrumentation and observing procedures it would be usual to assign the same variance to all the measurements. In fact very good evidence (such as that which may be yielded by the tests in 5.4.2) would be needed before assigning different variances to any groups of such measurements.

To exemplify this argument consider a modern control network consisting of a number of interlocking traverses measured using similar instrumentation, under similar conditions, by technicians of a similar grade. It has happened in the past that in such situations each traverse has been considered separately and each individual measurement in that traverse assigned a variance related in some way to that traverse's misclosure. Such a procedure is contrary to statistical theory and very dangerous as it can lead to quite ridiculous assigned variances (as found, for instance, by Masson-Smith et al. (1974), when applying a similar argument to gravity loops). The fact that a traverse closes well (perhaps perfectly) does not imply that the population from which its measurement errors are drawn is any different from that relating to a traverse with a large misclosure (so long as the misclosure is not so large as to indicate a gross error when the statistical tests in 5.4.1 are applied). The correct procedure is to compute just one variance from the mean of all the traverse misclosures and to use that to compute the measurement variances. Of course the argument is different if there are good reasons to suspect that different error populations are involved (e.g. if different instrumentation has been used, or if different topographical conditions exist).

A similar argument applies in offshore navigation, where it has happened that position-fixes have been assigned standard errors based on the area of the "cocked-hat" derived from a semi-graphic computation. Clearly what is important is the average size of a "cocked-hat" over a short period of time (during which the measurement process and fix geometry remain unchanged). Unless a "cocked-hat" is large enough to indicate a blunder it should not be interpreted as indicating a fix of a poorer quality than one with an area of zero.

4.2.2 Estimation of covariances

Two approaches to the estimation of covariances are discussed.

4.2.2.1 Simultaneous multiple measurements

If two sets of measurements $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ have been made so that a_i and b_i are simultaneous then it can be shown that the covariance of their means, α and β, can be estimated from

$$\sigma_{\alpha\beta} = v_a^T v_b/(n-1) \qquad (4.20)$$

where

$$v_a = \left[a_1 - \alpha, \quad a_2 - \alpha, \ \ldots\ , \ a_n - \alpha\right]^T$$

and

$$v_b = \left[b_1 - \beta, \quad b_2 - \beta, \ \ldots\ , \ b_n - \beta\right]^T$$

(4.20) is often known as Pearson's formula (after the statistician Karl Pearson).

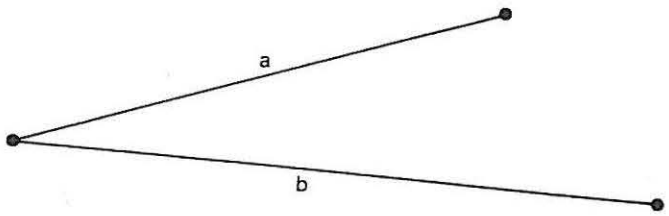


Fig. 4.2

For example, if two adjacent distances a and b in Fig. 4.2 have been measured by EDM over a long period of time as just described, we would expect that if the ith measurement of a is too long then the ith measure of b would also be too long (because the main error source in EDM is usually the inability of the refraction model to describe the actual conditions). Hence both v_{ai} and v_{bi} would be positive giving a positive product in (4.20). Similarly, if both lines were measured too short v_{ai} and v_{bi} would both be negative, again giving a positive product. The result would be a positive value for $\sigma_{\alpha\beta}$ and the measurements would be termed "positively correlated". Conversely, if two adjacent angles were measured in this way, we would probably obtain a negative value for their covariance because when one angle was too large the other would be too small (due to lateral refraction or pointing or reading errors).

It must be pointed out that it is unusual for (4.20) to be used in practice because such simultaneous multiple measurements are rarely made. Moreover (4.20) may not always accurately reflect the degree of correlation. For instance, in the foregoing distance example, if the EDM had a frequency error

additional correlation would exist but would not be taken into account by Pearson's formula because all distances would be in error by the same proportional amount.

4.2.2.2 Error propagation studies

In some situations the quantities used to form the vector $\bar{\ell}$ in the basic mathematical model (2.3) for the position-fix are not the quantities that have been directly (and perhaps independently) observed. In such cases the covariances (and variances) of the derived quantities can be determined by means of error propagation studies using (4.16).

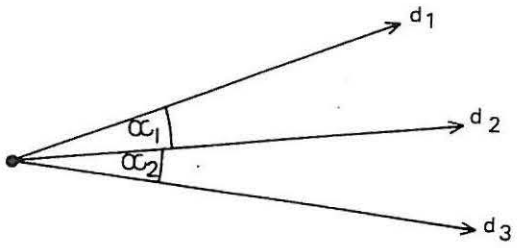


Fig. 4.3

For instance, if directions d_1, d_2 and d_3 in Fig. 4.3 have been independently measured (covariances equal to zero) with variances σ_1^2, σ_2^2 and σ_3^2, and if the mathematical model has been formed in terms of the derived angles α_1 and α_2, the variances of α_1 and α_2 and their covariance can be derived as follows

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = Ad \qquad (4.21)$$

then, using (4.16),

$$C_\alpha = AC_dA^T \qquad (4.22)$$

or

$$\begin{bmatrix} \sigma_{\alpha_1}^2 & \sigma_{\alpha_1\alpha_2} \\ \sigma_{\alpha_2\alpha_1} & \sigma_{\alpha_2}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 & -\sigma_2^2 \\ -\sigma_2^2 & \sigma_2^2 + \sigma_3^2 \end{bmatrix}$$

Another example of the use of this procedure is when combining coordinates derived from satellite-Doppler with other measurements. The Doppler coordinates would themselves have come from a least squares computation and would bring with them a full covariance matrix as in 4.3.1 even though the measured Doppler counts may have been considered uncorrelated.

This particular procedure can be very useful but it must be remembered that it does not remove the need to consider the correlation of the directly observed quantities.

4.2.3 Diagonal covariance matrices

It must be remarked here that it is extremely rare to use a full covariance matrix in practice. It is usually so difficult to estimate covariances that they are ignored even though they are known to exist. In such cases C_ℓ becomes a diagonal matrix and its inversion to obtain W is trivial. If $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ are the variances of the n observations we have

$$W = \begin{bmatrix} \sigma_1^{-2} & & & & \\ & \sigma_2^{-2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \sigma_n^{-2} \end{bmatrix} = \begin{bmatrix} w_1 & & & & \\ & w_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & w_n \end{bmatrix} \qquad (4.23)$$

where $w_1, w_2, \ldots, w_n$ are referred to as the weights of the observations. It is from this special case that we obtain the well known relationship:

$$\text{weight} = \text{reciprocal of standard error squared} \qquad (4.24)$$

It must be emphasised that the term "weight of an observation" only has meaning when observations are uncorrelated. If correlation exists and C_ℓ has off-diagonal elements, the term "weight matrix of a set of observations" must be used. Even if C_ℓ has only a few off-diagonal elements W will be a full matrix (i.e. without zero terms).

4.2.4 Estimation from least squares residuals

After a least squares computation (with an estimated C_ℓ) has taken place it is possible to estimate the precision of the measurements by examining the residuals (the amounts by which the observed values have been "altered" by the

computation process). Obviously high precision (low standard error) observations will have small residuals and vice versa.

It is usual, after computing the least squares residuals from (3.29), (3.43) or (3.45), to compute a quantity σ_o^2, which is known as the unit variance, fro

$$\sigma_o^2 = \hat{v}^T W \hat{v}/(n - m) \tag{4.25}$$

(Note that σ_o is often called the standard error of an observation of unit weight.)

It is shown in Appendix 1 that

$$E(\sigma_o^2) = 1 \tag{4.26}$$

Now, if σ_o^2 is significantly different from unity (see 5.4.2.2 for a statistical test to determine whether or not this is the case), and if there are no gross errors in the measurements (see 5.4.1), it can be concluded that the variances and covariances were, on average, underestimated by a factor of $1/\sigma_o^2$. The reasoning for this conclusion is simple: if we multiply C_ℓ by σ_o^2 we will have a new weight matrix W', where

$$W' = \frac{1}{\sigma_o^2} C_\ell^{-1} \tag{4.27}$$

and if we were to recompute the residuals $\hat{v}$ based on W' they would not change (this can be simply verified by multiplying W in (3.29), (3.43) and (3.45) by any scalar), so the new σ_o^2 would be less than the old one by a factor of σ_o^2, i.e. it would be unity. Hence multiplying by σ_o^2 will "correct" the estimated variances and covariances of the observations.

$$C_\ell \text{ (corrected)} = \sigma_o^2 \, C_\ell \text{ (estimated)} \tag{4.28}$$

It should be noted that in cases where $(n - m)$, the number of degrees of freedom, is small, e.g. in an individual offshore position-fix, σ_o^2 computed from one fix is of little value (this is evident from the statistical test in 5.4.2.2). If the measurement is repeated a large number of times (e.g. on a moving ship or stationary oil rig), however, the mean value of σ_o^2 can be used to correct the original estimate of the measurement variance. This procedure is then essentially equivalent to 4.2.1.1, where variances are computed using repeated measurements of the same quantity.

It is important to realise that the proximity of σ_o^2 to unity can only give

us the factor by which the a priori covariance matrix was on average incorrect. Under certain circumstances, however, it may be known, from testing external to the least squares computation, that some of the elements of C_ℓ are correct and then we can deduce the average errors in the other parts of C_ℓ. Typically we may be mixing angle and distance measurements and know the variances of the angles from, say, Ferrero's formulae (see 4.2.1.2), and we may be able to use σ_o^2 to help estimate the variances of the distances. Details of a practical application of this procedure are given in Ashkenazi et al. (1972).

4.2.5 Importance of weight matrices

After discussion of methods of estimating C_ℓ, it is clearly relevant to consider how important it is to estimate C_ℓ correctly and to investigate the effect on the residuals and parameters of using an incorrect C_ℓ. Certainly if the whole weight matrix is wrong by a constant factor (such as $1/\sigma_o^2$) it will not matter; for example, for the parameters in the combined case (3.25) gives

$$\hat{x} = [A^T(CW^{-1}C^T)^{-1}A]^{-1}A^T(CW^{-1}C^T)^{-1}b \qquad (3.25)$$

If W is multiplied by a scalar, say p, it is obvious that $\hat{x}$ will not change as p will cancel out. What is more important in this case is that the "relative" weighting between the observations should be correct and this is most difficult to achieve when dealing with heterogeneous data (such as mixtures of angles, distances, Doppler fixes, etc.). It is not possible to give general guidelines for the necessary accuracy needed in the estimation of C_ℓ, so it is best to test the sensitivity of any particular problem by computing it with different estimates of C_ℓ and analysing the differences in the results. Cross (1972) discusses the problem for large mixed triangulation networks and it has been generally found that parameters and residuals are not very sensitive to the C_ℓ employed. When dealing with uncorrelated homogeneous data it is usually sufficient to assume that all variances are the same, say σ^2, and C_ℓ becomes a scalar matrix, $\sigma^2 I$.

The argument is considerably different when assessing the quality of a position-fix by means of analysing the covariance matrices of the residuals and parameters. By looking ahead, say to equations (4.67) and (4.68), it can be seen that if C_ℓ is incorrectly estimated by a factor p then the estimate of the quality of the fix will be wrong by the same factor. Hence, although the overall size of C_ℓ is not important for estimating the parameters and

residuals themselves, it is critical for the estimation of their quality. This is why the statistic σ_o^2 (discussed in 4.2.4) is of such special importance in least squares computations.

4.3 A posteriori covariance matrices

It is usual in surveying and geodesy to refer to the covariance matrices of quantities derived from a least squares process as a posteriori covariance matrices. This is to contrast them with the covariance matrix of the observations which is called, as in 4.2, the a priori covariance matrix. Three a posteriori covariance matrices, those of the parameters, residuals and estimated observed quantities, are of especial interest and expressions for them will be derived in this section. The approach will be to consider, for all three matrices, the combined case and then to make the appropriate substitutions (given in 2.3) to obtain the expressions for observation and condition equations. It is relevant to mention at the outset that the terms a priori and a posteriori are not being used here with their strict classical meanings. Cooper (1981) discusses the matter in detail but, as pointed out by Cross (1982), the above meanings of the words are well understood in surveying and geodesy and there would be considerable difficulties and confusion if alternative meanings were adopted.

Before considering the covariance matrices of the parameters, residuals and estimated observed quantities, it is necessary to derive an expression for C_b, the covariance matrix of the absolute term in the combined linearised mathematical model. Rewriting (2.19) we have

$$Ax + Cv - b = 0 \tag{2.19}$$

where

$$b = -F(x^o, \ell) \tag{4.29}$$

$$\ell = \overline{\ell} - v \tag{4.30}$$

$$C = \partial F/\partial \ell \tag{4.31}$$

$$A = \partial F/\partial x \tag{4.32}$$

Substituting (4.30) in (4.29) gives

$$b = -F(x^o, \overline{\ell} - v) \tag{4.33}$$

and expanding to the first differential using Taylor's theorem leads to

$$b = -\left[F(x^o, \bar{\ell}) - \frac{\partial F}{\partial \ell} v\right] \tag{4.34}$$

$$= -F(x^o, \bar{\ell}) + \frac{\partial F}{\partial \ell} v \tag{4.35}$$

Substituting (4.31) in (4.35) gives

$$b = -F(x^o, \bar{\ell}) + Cv \tag{4.36}$$

Applying the propagation of error law, (4.16), to (4.36) and remembering that $F(x^o, \bar{\ell})$ is an errorless constant gives

$$C_b = C\, C_v\, C^T \tag{4.37}$$

but from (4.30) we have $C_v = C_\ell$, giving

$$C_b = C\, C_\ell\, C^T = C W^{-1} C^T \tag{4.38}$$

4.3.1 Covariance matrix of the parameters

The parameters are given by (3.25) as

$$\hat{x} = \left\{\left[A^T (C W^{-1} C^T)^{-1} A\right]^{-1} A^T (C W^{-1} C^T)^{-1}\right\} b \tag{3.25}$$

Defining D as

$$D = C W^{-1} C^T \tag{4.39}$$

and substitutuing (4.39) in (3.25) gives

$$\hat{x} = \left\{(A^T D^{-1} A)^{-1} A^T D^{-1}\right\} b \tag{4.40}$$

Applying (4.16) gives

$$C_{\hat{x}} = \left\{(A^T D^{-1} A)^{-1} A^T D^{-1}\right\} C_b \left\{(A^T D^{-1} A)^{-1} A^T D^{-1}\right\}^T \tag{4.41}$$

We note that, from (4.38) and (4.39),

$$C_b = D \tag{4.42}$$

and so (4.41) becomes

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

which simplifies to

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

i.e., from (4.39),

$$C_{\hat{x}} = \left[A^T (C W^{-1} C^T)^{-1} A\right]^{-1} \tag{4.43}$$

Notice that $C_{\hat{x}}$ is merely the inverse of the left hand side of the normal equations (3.24).

4.3.2 Covariance matrix of the residuals

The covariance matrix of the residuals is derived via the covariance matrix for the correlatives. Rewriting (3.26) using (4.39) gives

$$\hat{k} = - D^{-1}(b - A\hat{x}) \tag{4.44}$$

Substituting (4.40) and rearranging leads to

$$\hat{k} = - D^{-1}(I - A(A^T D^{-1} A)^{-1} A^T D^{-1})b \tag{4.45}$$

Applying (4.16) and (4.42) gives

$$C_{\hat{k}} = D^{-1}\left[1 - A(A^T D^{-1} A)^{-1} A^T D^{-1}\right] D \left[I - D^{-1} A(A^T D^{-1} A)^{-1} A^T\right] D^{-1}$$

$$= D^{-1}\left[D - 2A(A^T D^{-1} A)^{-1} A^T + A(A^T D^{-1} A)^{-1} A^T D^{-1} A(A^T D^{-1} A)^{-1} A^T\right] D^{-1}$$

which simplifies to

$$C_{\hat{k}} = D^{-1}\left[I - A(A^T D^{-1} A)^{-1} A^T D^{-1}\right] \tag{4.46}$$

Now, rewriting (3.27)

$$\hat{v} = - (W^{-1} C^T)\hat{k} \tag{3.27}$$

and applying (4.16) gives

$$C_{\hat{v}} = (W^{-1} C^T) C_{\hat{k}} (C W^{-1}) \tag{4.47}$$

Substituting (4.46) in (4.47)

$$C_{\hat{v}} = W^{-1} C^T D^{-1}\left[I - A(A^T D^{-1} A)^{-1} A^T D^{-1}\right] C W^{-1} \tag{4.48}$$

Finally substituting (4.39) into (4.48) gives the full expression as

$$C_{\hat{v}} = W^{-1}C^T(CW^{-1}C^T)^{-1}\left\{CW^{-1} - A\left[A^T(CW^{-1}C^T)^{-1}A\right]^{-1}A^T(CW^{-1}C^T)^{-1}CW^{-1}\right\} \quad (4.49)$$

4.3.3 Covariance matrix of the estimated observed quantities

The term "estimated observed quantities" is used for $\hat{\ell}$, the least squares estimate of the quantities that were observed, i.e.

$$\hat{\ell} = \ell + \hat{v} \quad (4.50)$$

Using (3.29) and (4.39), we write (4.50) as

$$\hat{\ell} = \ell + \left\{W^{-1}C^TD^{-1}\left[I - A(A^TD^{-1}A)^{-1}A^TD^{-1}\right]\right\}b \quad (4.51)$$

Putting

$$G = W^{-1}C^TD^{-1}\left[I - A(A^TD^{-1}A)^{-1}A^TD^{-1}\right] \quad (4.52)$$

simplifies (4.51) to

$$\hat{\ell} = \ell + Gb \quad (4.53)$$

Now, from (4.36)

$$b = -F(x^o, \bar{\ell}) + Cv$$

$$= -F(x^o, \bar{\ell}) + C(\bar{\ell} - \ell) \quad \text{(from (2.2))}$$

giving

$$b = -F(x^o, \bar{\ell}) + C\bar{\ell} - C\ell \quad (4.54)$$

Substituting (4.54) in (4.53) leads to

$$\hat{\ell} = \ell + G\left[-F(x^o, \bar{\ell}) + C\bar{\ell} - C\ell\right]$$

$$= G\left[-F(x^o, \bar{\ell}) + C\bar{\ell}\right] + (I - GC)\ell \quad (4.55)$$

Applying (4.16) to (4.55), whilst noticing that the first term of (4.55) is a non-stochastic vector and hence has no variances or covariances, we have

$$C_{\hat{\ell}} = (I - GC)C_{\ell}(I - C^TG^T) \quad (4.56)$$

Putting $C_{\ell} = W^{-1}$, from (3.7), and expanding gives

$$C_{\hat{\ell}} = W^{-1} - GCW^{-1} - W^{-1}C^{T}G^{T} + GCW^{-1}C^{T}G^{T} \tag{4.57}$$

We now define the symmetric matrix R by

$$R = A(A^{T}D^{-1}A)^{-1}A^{T}D^{-1} \tag{4.58}$$

Then, from (4.52) and (4.58),

$$G = W^{-1}C^{T}D^{-1}(I - R) \tag{4.59}$$

and (4.57) becomes

$$C_{\hat{\ell}} = W^{-1} - W^{-1}C^{T}D^{-1}(I - R)CW^{-1} - W^{-1}C^{T}(I - R^{T})D^{-1}CW^{-1} + W^{-1}C^{T}D^{-1}(I - R)D(I - R^{T})D^{-1}CW^{-1} \tag{4.60}$$

The fourth term of (4.60) can now be expanded as follows

$$\text{fourth term} = W^{-1}C^{T}D^{-1}(D - RD - DR^{T} + RDR^{T})D^{-1}CW^{-1}$$

$$= W^{-1}C^{T}D^{-1}(I - R - DR^{T}D^{-1} + RDR^{T}D^{-1})CW^{-1} \tag{4.61}$$

but, using (4.58), we may write

$$RDR^{T}D^{-1} = A(A^{T}D^{-1}A)^{-1}A^{T}D^{-1}DD^{-1}A(A^{T}D^{-1}A)^{-1}A^{T}D^{-1} = R \tag{4.62}$$

and

$$DR^{T}D^{-1} = DD^{-1}A(A^{T}D^{-1}A)^{-1}A^{T}D^{-1} = R \tag{4.63}$$

Substituting (4.62) and (4.63) in (4.61) gives

$$\text{fourth term} = W^{-1}C^{T}D^{-1}(I - R)CW^{-1} \tag{4.64}$$

From examination of (4.60) and (4.64) it is evident that the second and fourth terms of (4.60) cancel out and (4.60) becomes

$$C_{\hat{\ell}} = W^{-1} - W^{-1}C^{T}(I - R^{T})D^{-1}CW^{-1} \tag{4.65}$$

which, using (4.39) and (4.58) whilst noticing that R is symmetric, we can write in full as

$$C_{\hat{\ell}} = W^{-1} - W^{-1}C^{T}\left\{I - A\left[A^{T}(CW^{-1}C^{T})^{-1}A\right]^{-1}A^{T}(CW^{-1}C^{T})^{-1}\right\}(CW^{-1}C^{T})^{-1}CW^{-1} \tag{4.66}$$

4.3.4 Special cases

4.3.4.1 Observation equations

Putting C = -I in equations (4.43), (4.49) and (4.66) and simplifying leads to the following expressions for the observation equations special case:

$$C_{\hat{x}} = (A^T W A)^{-1} \tag{4.67}$$

$$C_{\hat{v}} = W^{-1} - A(A^T W A)^{-1} A^T \tag{4.68}$$

and, by transposing the second (symmetric) term in the braces in (4.66),

$$C_{\hat{\ell}} = A(A^T W A)^{-1} A^T = W^{-1} - C_{\hat{v}} \tag{4.69}$$

4.3.4.2 Condition equations

Putting A = 0 in equations (4.49) and (4.66) and simplifying leads to the following expressions for the condition equations special case:

$$C_{\hat{v}} = W^{-1} C^T (C W^{-1} C^T)^{-1} C W^{-1} \tag{4.70}$$

$$C_{\hat{\ell}} = W^{-1} - W^{-1} C^T (C W^{-1} C^T)^{-1} C W^{-1} = W^{-1} - C_{\hat{v}} \tag{4.71}$$

4.3.5 A further note on the use of the unit variance

As was explained in 4.2.4 it is usual, as part of a least squares computation, to determine a value of σ_o^2, the unit variance, from (4.25). If it is decided (as a result of the statistical test described in 5.4.2.2) that the unit variance is significantly different from unity then any a posteriori covariance matrices computed from equations (4.43), (4.49), and (4.66) to (4.71) must be multiplied by it. This is because the weight matrix used in these equations would have been incorrect (on average) by the reciprocal of this factor, as shown in (4.27) and (4.28).

Many authors, such as Krakiwsky (1976), include the unit variance in their expressions for the a posteriori covariance matrices, e.g. (4.67) is written

$$C_{\hat{x}} = \sigma_o^2 (A^T W A)^{-1} \tag{4.72}$$

In this paper the unit variance has been deliberately omitted from all

these expressions to emphasise that it should not be used without first applying the relevant statistical test. To take an extreme example of the possible misuse of (4.72), imagine a traverse which, by chance, closes perfectly (most surveyors are lucky enough to achieve this occasionally). Then a least squares computation could lead to zero residuals and a zero value for σ_o^2. Hence (4.72) would indicate perfect coordinates for all points; this is obviously nonsensical (all that has happened is that the errors around the traverse have summed to zero - they are not all zero).

4.4 Interpretation of a posteriori covariance matrices

This section is primarily concerned with $C_{\hat{x}}$, the covariance matrix of the parameters. Apart from a short section (4.4.5), little attention is paid to $C_{\hat{\ell}}$, the covariance matrix of the estimated observed quantities, because it is seldom computed in practice (except when the method of condition equations is used). $C_{\hat{v}}$, the covariance matrix of the residuals, is discussed in section 5 as its main practical use is in connection with statistical testing.

Basically $C_{\hat{x}}$ is used to assess the precision of a position-fix. A number of different measures of precision can be employed and their computation and interpretation are discussed below.

4.4.1 Positional standard errors

Simply by taking the square roots of the relevant diagonal elements of $C_{\hat{x}}$ we can determine the standard errors of all coordinates. Hence each quoted coordinate can be accompanied by its standard error.

It is important to realise that such standard errors are a function of the chosen coordinate system and are said to be "reference frame dependent". In other words, coordinate standard errors measure the precision of a point's position with respect to the fixed point(s) used when setting up the basic mathematical model (2.3). In situations where a number of points are being fixed simultaneously (e.g. in a triangulation network or an acoustic beacon calibration), great care must be taken with regard to the interpretation of positional standard errors because, in general, the further a point is from the fixed point(s) the larger its positional standard error will be. Hence the proper interpretation of a high standard error is not necessarily that the point is weakly determined but perhaps that it is simply far from a fixed point (i.e. a point where the reference frame is defined).

In some problems, e.g. the fixing of an off-shore oil platform for the purpose of assessing the "ownership" of the oil it recovers, the reference system is of very special importance because of the way the boundaries are defined, and it is likely that knowledge of positional standard errors (so long as they are related to the proper reference system) would be essential. In other situations, such as for most engineering work, the choice of the fixed point is quite arbitrary; yet positional standard errors depend on this point, so they would reflect the choice rather than the required precision. In such cases we search for measures (such as some of those in 4.4.2 and 4.4.3) that are invariant to the choice of a single fixed point.

4.4.2 Error ellipses

In two-dimensional position-fixing work the positional standard errors discussed in 4.4.1 would be the standard errors of northings and eastings or latitudes and longitudes, i.e. in two orthogonal directions. It is often useful to know the standard errors in other directions as well and the error ellipse is a way of describing standard errors in all directions. Consider orthogonal axes x and y rotated by an angle ψ to directions p and q in Fig. 4.4

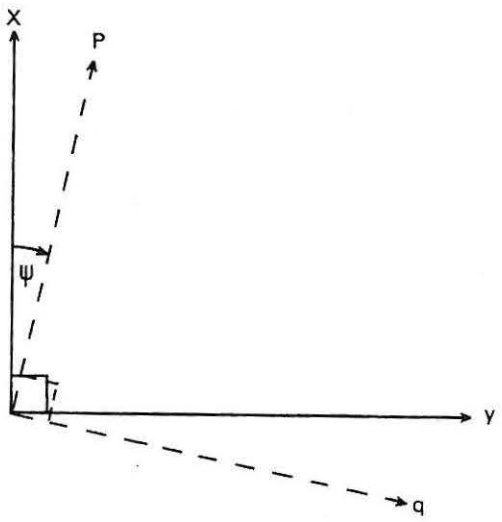


Fig. 4.4

The rotation can be expressed in matrix algebra as

$$\begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad (4.73)$$

Applying (4.16) to (4.73) gives

$$\begin{bmatrix} \sigma_p^2 & \sigma_{pq} \\ \sigma_{qp} & \sigma_q^2 \end{bmatrix} = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}$$

from which we can extract the positional variances of the fix in the directions p and q:

$$\sigma_p^2 = \cos^2\psi\ \sigma_x^2 + \sin^2\psi\ \sigma_y^2 + 2\cos\psi\sin\psi\ \sigma_{xy} \qquad (4.74)$$

$$\sigma_q^2 = \sin^2\psi\ \sigma_x^2 + \cos^2\psi\ \sigma_y^2 - 2\cos\psi\sin\psi\ \sigma_{xy} \qquad (4.75)$$

To determine the direction ψ_m in which σ_p^2 is a maximum (or minimum) we differentiate (4.74) with respect to ψ and equate to zero to yield

$$-2\cos\psi_m\sin\psi_m\ \sigma_x^2 + 2\sin\psi_m\cos\psi_m\ \sigma_y^2 + 2(\cos^2\psi_m - \sin^2\psi_m)\ \sigma_{xy} =$$

giving

$$-\sin 2\psi_m\ \sigma_x^2 + \sin 2\psi_m\ \sigma_y^2 + 2\cos 2\psi_m\ \sigma_{xy} = 0$$

$$\therefore\ \tan 2\psi_m = 2\sigma_{xy}/(\sigma_x^2 - \sigma_y^2) \qquad (4.76)$$

There are two possible solutions to (4.76), which we will call ψ_{max} and ψ_{min} (they will be 90° apart). Substitution of each into (4.74) will yield the maximum and minimum variances σ^2_{max} and σ^2_{min}. Of course until this substitution has been made it will not be known which of the two solutions to (4.76) corresponds to the maximum and which to the minimum variance.

An ellipse drawn with its major axis in the direction ψ_{max} and with semimajor and semiminor axes of length σ_{max} and σ_{min} respectively is called an error ellipse.

It is possible to derive alternative formulae for σ_{max} and σ_{min} which do not involve ψ_{max}. We proceed as follows.

Rearranging (4.74) as

$$\sigma_m^2 = \cos^2\psi(\sigma_x^2 + \tan^2\psi\ \sigma_y^2 + 2\tan\psi\ \sigma_{xy}) \qquad (4.77)$$

and putting

$$\cos^2\psi = 1/\sec^2\psi = 1/(1 + \tan^2\psi)$$

and

$$t = \tan\psi \tag{4.78}$$

leads to

$$\sigma_m^2 = \left[1/(1 + t^2)\right]\sigma_x^2 + \left[t^2/(1 + t^2)\right]\sigma_y^2 + \left[2t/(1 + t^2)\right]\sigma_{xy} \tag{4.79}$$

Also let

$$a = \sigma_x^2 - \sigma_y^2 \tag{4.80}$$

and

$$b = \sigma_{xy} \tag{4.81}$$

Then, for $\sigma_m = \sigma_{max}$ or σ_{min}, we have

$$\tan 2\psi_m = 2t/(1 - t^2) \quad \text{(using (4.78))} \tag{4.82}$$

$$= 2\sigma_{xy}/(\sigma_x^2 - \sigma_y^2) \quad \text{(from (4.76))} \tag{4.83}$$

$$= 2b/a \quad \text{(from (4.80) and (4.81))} \tag{4.84}$$

Equating and cross-multiplying the right hand sides of (4.82) and (4.84) gives

$$bt^2 + at - b = 0 \tag{4.85}$$

with solution (taking positive square root)

$$t = \left[-a + (a^2 + 4b^2)^{\frac{1}{2}}\right]/2b \tag{4.86}$$

putting

$$c = (a^2 + 4b^2)^{\frac{1}{2}} \tag{4.87}$$

simiplifies (4.86) to

$$t = (c - a)/2b \tag{4.88}$$

In preparation for substitution into (4.79) we can derive, from (4.88)

$$1/(1 + t^2) = 2b^2/(c(c - a)) \tag{4.89}$$

$$t^2/(1 + t^2) = (c - a)/2c \tag{4.90}$$

$$t/(1 + t^2) = b/c \tag{4.91}$$

Substituting (4.89), (4.90) and (4.91) into (4.79) gives

$$\sigma_{max}^2 = \left[2b^2/(c(c-a))\right]\sigma_x^2 + \left[(c-a)/2c\right]\sigma_y^2 + \left[2b/c\right]\sigma_{xy} \tag{4.92}$$

which, after using (4.81), becomes

$$\sigma_{max}^2 = \left[1/(2c(c-a))\right]\left[4b^2\sigma_x^2 + (c-a)^2\sigma_y^2 + 4b^2(c-a)\right] \tag{4.93}$$

but, from (4.80), we have

$$\sigma_x^2 = a + \sigma_y^2 \tag{4.94}$$

Substituting (4.94) into (4.93) and rearranging yields

$$\sigma_{max}^2 = \left[1/(2c(c-a))\right]\left[(4b^2 + c^2 + a^2 - 2ac)\sigma_y^2 + 4b^2c\right] \tag{4.95}$$

$$= \left[1/(2c(c-a))\right]\left[(2c^2 - 2ac)\sigma_y^2 + 4b^2c\right] \text{ (using (4.87))} \tag{4.96}$$

$$= \sigma_y^2 + 2b^2/(c-a) \tag{4.97}$$

but, from (4.87), we have

$$b^2 = (c^2 - a^2)/4 = (c-a)(c+a)/4 \tag{4.98}$$

and substitution of (4.98) in (4.97) gives

$$\sigma_{max}^2 = \sigma_y^2 + (c+a)/2$$

$$= \tfrac{1}{2}(\sigma_x^2 + \sigma_y^2 + c) \text{ (using (4.80))} \tag{4.99}$$

Then substituting (4.80) and (4.81) into (4.87) and substituting the result into (4.99) yields

$$\sigma_{max}^2 = \tfrac{1}{2}\left\{\sigma_x^2 + \sigma_y^2 + \left[(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2\right]^{\frac{1}{2}}\right\} \tag{4.100}$$

Similarly taking the negative square root in (4.86) yields

$$\sigma_{min}^2 = \tfrac{1}{2}\left\{\sigma_x^2 + \sigma_y^2 - \left[(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2\right]^{\frac{1}{2}}\right\} \tag{4.101}$$

These alternative formulae are important from a theoretical point of view because they establish a link with the eigenvalues of the covariance matrix $C_{\hat{x}}$ (see (4.104)), but in practice they are not very useful. This is because (4.76) has two solutions, ψ_{max} and ψ_{min}, so (4.74) must be used anyway to distinguish between them. Hence σ_{max} and σ_{min} would have already been determined.

It should be noted that the error ellipse does not quite describe the standard error in all directions, although it is a good approximation to it. The exact value of σ_p, as given by (4.74), is the pedal curve of this ellipse and is shown by the dotted line in Fig. 4.5. It is the locus of the point S where ST is perpendicular to OS and tangential to the error ellipse. It is the error ellipse rather than its pedal curve that is generally most useful in practice so for most applications it is unusual to see this pedal curve drawn.

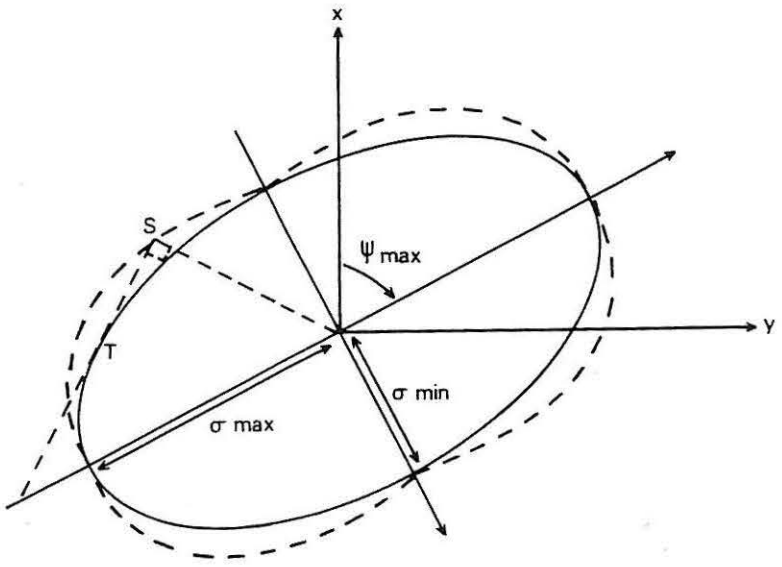


Fig. 4.5

It will be explained in 5.2.2 that if the observational errors belong to a normal distribution then the resulting two-dimensional positional errors belong to a bivariate normal distribution and it follows that there is a 39.4% probability that the least squares estimate of a station's position lies within an error ellipse centred at its true position. Of course in

practice true positions are not known so error ellipses have to be drawn with their centres at the least squares estimates of the stations' positions. It is then sometimes said that a station's true position has a 39.4% probability of lying within its error ellipse but this statement is incorrect and its use should be avoided. It is better simply to refer to the error ellipse as the 39.4% confidence region.

For a more detailed account of the properties of the error ellipse and its pedal curve readers are referred to Bomford (1980, 719-724) and Richardus (1966). It is, however, interesting to note here that the sizes of the semi-major and semiminor axes of an error ellipse are given by the square roots of the two eigenvalues of the matrix

$$\begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}.$$

This is simply demonstrated as follows. Let λ be an eigenvalue of the above matrix; then

$$\left| \begin{bmatrix} \sigma_x^2 - \lambda & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 - \lambda \end{bmatrix} \right| = 0 \qquad (4.102)$$

i.e. $$(\sigma_x^2 - \lambda)(\sigma_y^2 - \lambda) - \sigma_{xy}^2 = 0$$

or

$$\lambda^2 - (\sigma_x^2 + \sigma_y^2)\lambda + \sigma_x^2 \sigma_y^2 - \sigma_{xy}^2 = 0 \qquad (4.103)$$

with solutions

$$\lambda = \tfrac{1}{2}\left\{(\sigma_x^2 + \sigma_y^2) \pm \left[(\sigma_x^2 + \sigma_y^2)^2 - 4(\sigma_x^2 \sigma_y^2 - \sigma_{xy}^2)\right]^{\frac{1}{2}}\right\}.$$

which rearranges to

$$\lambda = \tfrac{1}{2}\left\{(\sigma_x^2 + \sigma_y^2) \pm \left[(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2\right]^{\frac{1}{2}}\right\} \qquad (4.104)$$

which is identical to (4.100) and (4.101).

Also it can be demonstrated that the eigenvectors corresponding to the two eigenvalues contain the direction of the major and minor axes respectively. More generally we often consider that the n eigenvalues of any n x n covariance matrix describe an n-dimensional "hyperellipsoid" and the n

eigenvectors describe, in n-dimensional space, the directions of each of the axes of this hyperellipsoid. Although this may seem at first sight to be somewhat theoretical and impossible to conceive it does have some practical uses, for instance it affords a means by which large networks can be compared (e.g. largest eigenvalue or volume of hyperellipsoid).

Error ellipses as described in the foregoing, with x and y in (4.73) referring to the coordinates of a survey station, are usually called absolute error ellipses. It should be noted that the remarks in 4.4.1 regarding the importance of the reference system to positional standard errors also apply to absolute error ellipses, i.e. they increase in size with distance from the fixed point(s). Ashkenazi and Cross (1972) demonstrate this with a practical example.

For large networks the patterns of the absolute error ellipses contain useful information. Error ellipses with minor axes pointing towards the fixed point indicate orientation weakness and suggest that more azimuth control is needed. Conversely, if all major axes are pointing towards the fixed point, more scale control is required.

In practice we are often more interested in the relative position of two points. In this case if we interpret x and y above as differences in position between the points i and j we can write

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ x_j \\ y_j \end{bmatrix} \qquad (4.105)$$

and applying (4.16) to (4.105) gives

$$\begin{aligned} \sigma_x^2 &= \sigma_{x_i}^2 + \sigma_{x_j}^2 - 2\sigma_{x_i x_j} \\ \sigma_y^2 &= \sigma_{y_i}^2 + \sigma_{y_j}^2 - 2\sigma_{y_i y_j} \\ \sigma_{xy} &= \sigma_{x_i y_i} - \sigma_{x_i y_j} - \sigma_{x_j y_i} + \sigma_{x_j y_j} \end{aligned} \qquad (4.106)$$

which can be used in (4.74) and (4.76) to compute the major and minor axes and orientation of the relative error ellipse between points i and j. Note

that all the quantities on the right hand side of the equality in (4.106) are found in the covariance matrix $C_{\hat{x}}$.

Relative error ellipses are invariant to the choice of a single fixed station and are generally more useful than absolute error ellipses for reasons explained in 4.4.1.

4.4.3 Standard errors of derived quantities

The quality of the coordinates estimated by a least squares computation is often best expressed by the precision of quantities derived from them. This is especially true when this precision relates closely to the intended use of the coordinates. For instance, if two points on either side of a hill have been fixed in order to build a tunnel between them it would be necessary to use these coordinates to compute setting-out directions at each end of the tunnel. In fact the coordinates are not really of prime interest: they were only determined in order to establish the directions. Hence the standard errors of the derived directions could be of special importance and would be the best way to measure the precision of the fixed positions. Expressions for the standard errors of such derived quantities can be obtained as follows.

Let $\hat{q}$ be a vector of k quantities derived from a set of estimated parameters $\hat{x}$ via the vector function

$$\hat{q} = F(x^o + \hat{x}) \tag{4.107}$$

(Note from 2.2 that the final values of the parameters are given by x^o, the provisional values, plus $\hat{x}$, their small changes estimated by the least squares process.)

Expanding (4.107) by Taylor's expansion gives

$$\hat{q} = F(x^o) + \frac{\partial F}{\partial x^o}\hat{x} \tag{4.108}$$

or

$$\hat{q} = F(x^o) + B\hat{x} \tag{4.109}$$

where B is a k x m matrix containing the first differentials of the functional relationship between $\hat{q}$ and x^o. Matrix B is of a similar form to matrix A in (2.19). Applying (4.16) to (4.109) whilst remembering that $F(x^o)$ is non-stochastic leads to

$$C_{\hat{q}} = BC_{\hat{x}}B^T \tag{4.110}$$

As an example, imagine that it is required to find the standard error of a plane direction α between two stations i and j. The relevant elements of $\hat{x}$ would be

$$\hat{x} = (dE_i \; dN_i \; dE_j \; dN_j)^T \tag{4.111}$$

$\hat{q}$ would have just one element α; $F(x^o + \hat{x})$ would be

$$\hat{q} = F(x^o + \hat{x}) = \alpha = \tan^{-1} (E_j - E_i)/(N_j - N_i) \tag{4.112}$$

B would have a single row whose elements would be given by differentiating (4.112), viz.

$$B_{11} = \frac{\partial \alpha}{\partial E_i} = (N_i - N_j)/d^2$$

$$B_{12} = \frac{\partial \alpha}{\partial N_i} = -(E_i - E_j)/d^2$$

$$B_{13} = \frac{\partial \alpha}{\partial E_j} = -(N_i - N_j)/d^2$$

$$B_{14} = \frac{\partial \alpha}{\partial N_j} = (E_i - E_j)/d^2$$

where

$$d^2 = (E_i - E_j)^2 + (N_i - N_j)^2$$

Then, with the required elements of $C_{\hat{x}}$ determined as in 4.3.1, the variance of α (which would be the sole element of $C_{\hat{q}}$) would be given by (4.110). Of course, in general, several quantities may be involved and $C_{\hat{q}}$ would be a full matrix containing both the variances and the covariances of the selected derived quantities. It is worth remarking, however, that in practice the covariances of the derived quantities may not be very useful so it may be efficient, from a computational point of view, to compute the variances only, following the procedure in the above example (i.e. considering each derived quantity independently).

Standard errors of certain relative derived quantities, such as angles, directions, distances and height differences, are invariant to the choice of a single fixed station and are consequently especially useful measures of precision when the coordinate system is arbitrarily chosen. Ashkenazi and Cross (1972, 1976) discuss this point and give examples.

4.4.4 Single number measures of precision

There are some situations in which it is desirable to use a single number to express the precision of the position of single (or groups of) points. For instance, in a comparison of alternative observation schemes for a given position-fixing problem, the precision of each scheme must be "measured" by a single number in order to say which is the more precise. Although it is unrealistic (especially for large networks) to use one number to describe completely the precision, the following are mentioned as possible candidates.

(i) One possibility is a quantity related to the trace of $C_{\hat{x}}$, e.g. $[Tr(C_{\hat{x}})]/m$, the average parameter variance. Care must be taken to include only elements corresponding to positional unknowns in $\hat{x}$. Other parameters may have different units and so invalidate the summation made in computing the trace.

(ii) Secondly, the average size of error ellipses may be informative. Note that if absolute error ellipses are used (and if size is defined as $\sigma^2_{max} + \sigma^2_{min}$) this measure is virtually the same as (i).

(iii) Thirdly, some surveyors use the spectral condition number of the normal equations (the ratio of the maximum to minimum eigenvalues). This can be interpreted as a measure of a network's precision because it reflects the numerical ill-conditioning of these equations. It has the disadvantage of not having a real physical meaning, other than as described in 4.4.2, and of being difficult to compute.

(iv) Finally, the average standard error of a derived quantity may be helpful. A commonly used, and most effective, single number measure can be obtained by averaging the standard error of a number of similar (usually arbitrarily selected) derived quantities. The choice of quantities would be a function of the purpose of the network.

4.4.5 Covariance matrix of the estimated observed quantities

Apart from its limited use in statistical testing (see (5.34) in 5.4.1.3), the covariance matrix of the estimated observed quantities, $C_{\hat{\ell}}$, is seldom of interest. This is because usually the main objective of a least squares computation is to determine a set of parameters and so it is naturally $C_{\hat{x}}$, their covariance matrix, that is required.

An exception is the special case of condition equations, where there are no parameters and we are only concerned with estimating the quantities that have been measured. Usually $C_{\hat{\ell}}$ would be used either to obtain the standard errors of the estimated observed quantities (by simply taking the square roots of the diagonal elements) or else to aid the computation of the covariance matrix of some derived quantities. For instance, if a triangulation network was computed by the method of condition equations, the result would be estimated angles, distances, azimuths, etc. These would later be used to compute coordinates and we would be interested in the covariance matrix of these coordinates so that the techniques of 4.4.1 to 4.4.4 could be used.

The process for computing the required covariance matrix is identical to that described in 4.4.3. If $\hat{y}$ is a vector of derived coordinates

$$C_{\hat{y}} = BC_{\hat{\ell}}B^T \qquad (4.113)$$

where B is a Jacobian containing the first differentials of the functional relationships used to compute the coordinates from the estimated observed quantities.

4.4.6 Non-positional standard errors

In many position-fix computations the vector of parameters will consist only of coordinates (or small changes to coordinates in non-linear problems). In some problems, however, such as example (b) in 2.1, one or more of the parameters may relate to the observations themselves (e.g. frequency errors or scale factors) or to the coordinate system (e.g. translation parameters). In such cases it would be usual to analyse the variances of these parameters separately from those of the coordinates.

Usually only the relevant diagonal elements of $C_{\hat{x}}$ would be of interest and these would be used to compute the standard errors of all the non-positional parameters. Hence we would have a measure of their precision which could be quoted alongside their values.

There are two main uses of such a measure of precision. The first is to help decide on the significance of a parameter (in conjunction with the statistical test in 5.4.3.2). Basically, if a parameter is very small compared with its standard error there would be considerable doubt as to whether or not it should have appeared in the model in the first place and it may be worthwhile to recast the model without it (hence increasing the degrees of freedom which, in general,

is a desirable action from the point of view of the statistical testing discussed in section 5).

The second use is in the measurement of the quality of a network with respect to systematic errors. For example, if a scale unknown were included in the model then its standard error would indicate how well the scale of the distance measurements to which it refers was determined, i.e. it would be a measure of the probability of the existence of scale errors in the estimated coordinates. Of course such an interpretation assumes the "fixed scale" of the network is correct (fixed scale would be provided either by fixed points or by distance measurements without scale unknowns).

4.4.7 A note of caution

Sections 4.4.1 to 4.4.5 have derived and described a number of extremely valuable methods of assessing the precision of position-fixes. It is, however, necessary to issue two warnings on their use.

(i) The correctness of all three a posteriori covariance matrices depends directly on the correctness of the estimation of C_ℓ (see 4.2). If the precision of the measurements has not been properly estimated, the a posteriori covariance matrices will be of little value.

(ii) The a posteriori covariance matrices only measure the precision of position-fixes. This is not sufficient to describe fully the quality of a fix; it is essential to quote also some measures of reliability (e.g. as in 5.4.1.3) and to have some indication of whether or not systematic errors may be present. For example, Ashkenazi and Cross (1972) found that the precision of block VI of the European triangulatio network was such that the average a posteriori standard error of a derived distance was 2.5 ppm but Weightman (1975) found an overall 3.1 ppm systematic error when comparing the triangulation with satellite-Doppler results, hence rendering the earlier estimate of limited value.

4.5 The design of position-fixes

It is evident from (4.43), (4.49), (4.66) and their special cases (4.67) to (4.71) that we do not need the actual measured quantities (vectors ℓ or b) to compute the a posteriori covariance matrices. So long as a set of measurements has been planned (with a covariance matrix C_ℓ) and the

approximate geometry is known (matrices A and C), the precision of a fix can be determined. Hence the a posteriori covariance matrices can be used to design position-fixes.

It is now generally accepted by geodesists that it is useful to consider four separate geodetic design problems as suggested by Grafarend (1974). Most practical design problems will then be a combination of one or more of these four. They can be summarised as follows.

Zero order design is the selection of a suitable reference system. This problem has received little attention and will not be considered here. It is obviously not relevant to position-fixing problems where the coordinate system is not open to choice, e.g. single point positioning offshore.

First order design is the design of a network configuration, i.e. the choice of the positions of the points in the network and of which quantities to measure. There is usually little choice of positions of the points as these are largely dictated by the topography and the requirements of the survey. In some cases, however, there may be a choice: for example, in an off-shore single point-fix there may be a number of possible shore stations that could be used. Whether or not there is a choice of measurements will depend on the situation. For instance, in an engineering network the surveyor can usually choose between angle measurements and distance measurements or a combination of both, whereas at sea a hydrographer may have only one position-fixing system on board.

Second order design is the choice of observation weights, i.e. with what precision should the measurements be made (and hence what instruments should be used)? Again, whether or not there is a choice of instrumentation will depend on the situation. It should be pointed out here that the second order design problem is sometimes used to solve the first order design problem in the following iterative way:

(i) postulate a large number of possible observations

(ii) solve the second order design problem

(iii) discard observations with low weights and, if required, go back to (ii).

Hence in practice the second order design problem is more important than its definition would suggest.

Third order design is the improvement of existing positions by the inclusion of additional observations. This is especially useful in permanently

maintained networks such as national triangulation networks or shore control stations for off-shore position-fixing. As new instrumentation becomes available it is desirable to make additional measurements to improve the quality of the existing positions and the question of which measurements to make then arises.

Using the observation equation model we can write down

$$C_{\hat{x}} = (A^T W A)^{-1} \qquad (4.67)$$

and then say that, given $C_{\hat{x}}$, the required precision, the first and second order design problems are the solutions of (4.67) for A and W respectively.

Usually there is a large (possibly infinite) number of possible solutions to a particular geodetic design problem and we are mainly interested in the solution that can be implemented with the least cost. This solution is called the optimal design and the two approaches currently used to determine it will now be summarised.

4.5.1 Computer simulation

Computer simulation is now common practice in many geodetic and engineering organisations and is also ideally suited to most off-shore position-fixing problems. The starting point is the precision criteria, which are usually expressed in terms of the quantities discussed in 4.4, e.g. all error ellipses should be less than a specified size or all standard errors of derived distances less than a specified size. Then the process is as illustrated in the flow diagram in Fig. 4.6

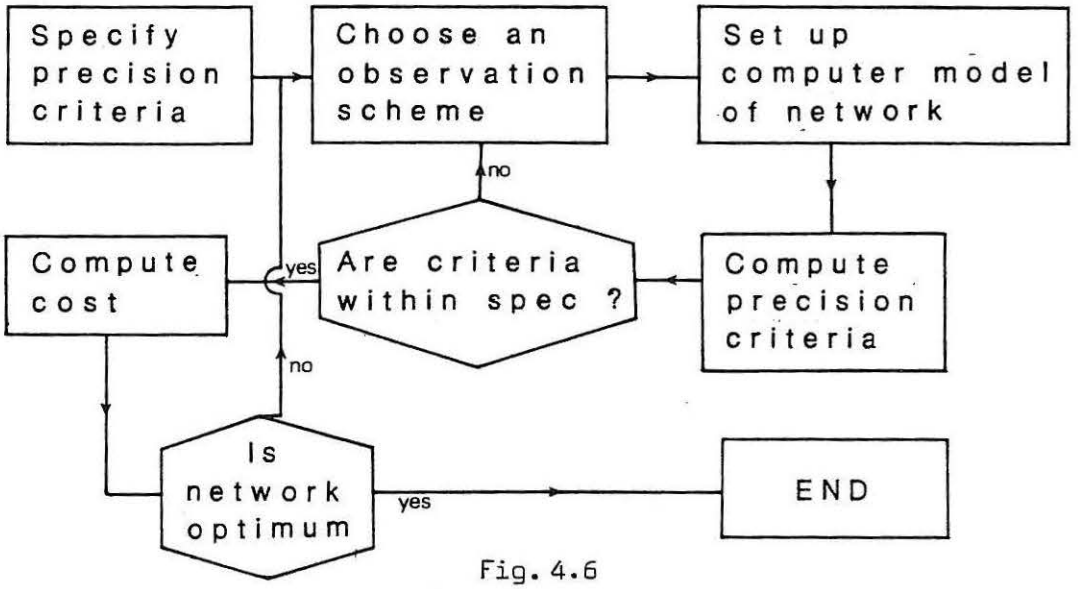


Fig. 4.6

The phrase "set up computer model" in Fig. 4.5 refers to the computation of the a posteriori covariance matrix of the parameters, $C_{\hat{x}}$, and the phrase "choose an observation scheme" includes selections relevant to the first, second and third order design problems, i.e. the network geometry may be changed by adding and deleting points or by changing the measured quantities, the precision of the planned measurements may be changed by altering C_{ℓ}, and additional observations may be included. There are many examples of the application of the computer simulation method to practical problems, for instance Chrzanowski (1981) on the design of tunnel networks, Cross and Whiting (1981) on the design of level networks and Nickerson et al (1978) on the use of computer simulation with interactive graphics.

4.5.2 Analytical methods

Analytical methods involve direct mathematical solutions to optimal design problems and are currently at a very early stage of development. The only problem to which satisfactory solutions have so far been reached is the second order design problem applied to the special case of observation equations, i.e. the solution for W in (4.67)

$$(A^T W A)^{-1} = C_{\hat{x}} \tag{4.67}$$

in the form

$$W = f(A, C_{\hat{x}}) \tag{4.114}$$

where $C_{\hat{x}}$ contains the required variances and covariances of the parameters and is called the "criterion matrix". A full discussion of the various expressions for (4.114) is beyond the scope of this paper as it would need to include advanced mathematical techniques such as generalised matrix algebra and operations research. Cross (1981b) is a review of possible solutions to (4.114) which includes many references to more detailed publications.

It is probably true to say that analytical design techniques are still at the research stage and almost all practical geodetic design problems are solved by the computer simulation method (or else purely by use of experience).

5. Statistical testing

It is usual both before and after a least squares computation to want to ask a number of questions regarding the significance of certain aspects of the observed and computed data. Two examples are as follows.

(i) Do all measurements come from the same normally distributed population? If not, one or more may contain gross errors and need to be rejected or perhaps may have to be assigned different a priori variances from the rest.

(ii) Are two measurements or estimates of the same quantity significantly different? If so, the quantity itself may have changed (e.g. a distance or position may change owing to land movements).

The answers to such questions are investigated by means of a process known as statistical testing. In this section a review of the statistical tests most useful for position-fixing will be given along with some examples of their application. Unlike least squares, which is the primary subject of this paper, statistical testing will not be given a full mathematical treatment (the subject is too large for this anyway). The approach will be to produce a practitioner's guide to the topic with emphasis on methodology rather than theory. Before describing the statistical tests themselves it is necessary to introduce the terminology of statistical testing and to describe the four probability density functions which form the basis of the tests to be described.

5.1 Terminology of statistical testing

Statistics, in particular statistical testing, has its own jargon and it is important to define the following terms carefully before proceeding. The definitions are taken from Wells and Krakiwsky (1971) but are similar to those in virtually any book on mathematical statistics.

Statistic: a quantitative item of information (e.g. a mean or standard error) deduced from the application of statistical methods.

Variate: a quantity (also known as a random variable) which may assume any one of the values of a specified set with a specified probability, i.e. with specified pdf (probability density functions were explained in 4.1).

Population: a collection of all objects having in common a particular measurable variate. We will deal here with infinite populations and use Greek letters (e.g. σ for standard error) to describe population statistics.

Individual: a single member of a population.

Sample: a group of individuals drawn from a population. We will use Roman letters (e.g. s for standard error) to describe sample statistics.

To give an example of the use of these words, let us imagine that we have measured a distance n times under similar circumstances. The length of the line would be the variate and the n distances would be a sample of all the possible infinite number of measurements which constitute the population. Any one distance measurement would be an individual and the mean of the n measures, d, is an example of a sample statistic whilst (assuming no systematic errors) μ, the true distance, would be the mean of the population and would be a population statistic.

In statistical testing we use the term null hypothesis, H_0, to describe the hypothesis we wish to test. For example, if we wanted to know whether the mean of the sample of distances above was significantly different from the true distance (perhaps obtained by some more precise measurement technique) we would write

$$H_0: d = \mu$$

Tested against this null hypothesis is an alternative hypothesis, H_A, which might be, in the above example

$$H_A: d \neq \mu$$

Note that if the null hypothesis involves an equality then the alternative hypothesis will be an inequality and we term the test a two-sided test because in the above example we would reject the null hypothesis both if $d > \mu$ and if $d < \mu$. If, however, we write

$$H_0: d \leq \mu$$

$$H_A: d > \mu$$

i.e. we do not mind by how much d is smaller than μ, we would term this a one-sided test. The terms two-tailed and one-tailed are also used.

When a statistical test is carried out there are two kinds of error which may be made.

(i) The null hyposthesis could be rejected when it ought to have been accepted. Such an error is called a type 1 error and the probability of making such an error is called the level of significance of the test and is usually denoted by α or more commonly by $100\alpha\%$, where α lies between 0 and 1. Alternatively the term level of confidence, given by $(1 - \alpha)100\%$, can be used.

(ii) The null hypothesis could be accepted when it ought to have been rejected. Such an error is called a type 2 error and the probability of it occurring is denoted by β, or $100\beta\%$. $(1 - \beta)$ is usually referred to as the power of the test.

The situation is summarised in Table 5.1. Note that there is an inverse

DECISION	ACTUAL SITUATION	
	Null hypothesis true	Null hypothesis false
Accept null hypothesis	Correct	Type 2 error
Reject null hypothesis	Type 1 error	Correct

Table 5.1

relationship between α and β: if we reduce α then β must be increased and vice versa. For instance, in the case of the rejection of observations with gross errors, if we apply a test which ensures that all observations are accepted we would never reject a "good" observation but there is a high chance of accepting a "bad" observation (type 2 error). Conversely, if we rejected all our measurements we would never make a type 2 error. The choice of values for α and β is subjective, i.e. it needs to be made personally by the surveyor carrying out the tests.

Finally, it is mentioned that the tests described in this paper come under the general heading of parametric tests because they are all based on statistics computed from the samples. Tests not based on these are called non-parametric tests, e.g. sign tests and rank tests. Such tests are not discussed here.

5.2 The normal distribution

5.2.1 Single variate

As far as surveying is concerned (and indeed most other branches of science and technology) the normal distribution is the most important of all the many probability distributions. Its pdf is defined by

$$\varphi(y) = (1/(\sigma\sqrt{2\pi})) \exp\left\{-(y-\mu)^2/2\sigma^2\right\} \qquad (5.1)$$

where y is a single variate with standard error σ. Fig. 5.1 shows the general form of the curve; note that it is symmetrical about $y = \mu$ with points of

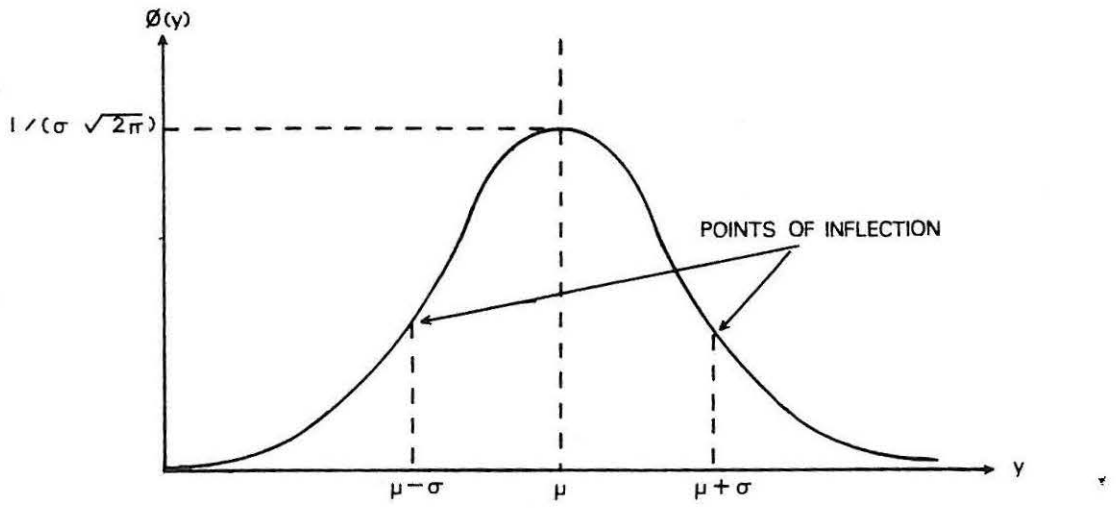


Fig. 5.1

inflection at $y = \mu \pm \sigma$. We find it convenient in practice to use a standard normal distribution which relates to a variate z, where

$$z = (y - \mu)/\sigma \qquad (5.2)$$

and has a pdf

$$\varphi(z) = (1/\sqrt{2\pi}) \exp\left\{-z^2/2\right\} \qquad (5.3)$$

i.e. it has a mean of zero and standard error of unity. From the definition of the pdf (equation (4.1)) the area under the curve is the probability. Hence by integrating (5.3) as follows

$$P(a \leq z \leq b) = \int_a^b (1/\sqrt{2\pi}) \exp\left\{-z^2/2\right\} dz \qquad (5.4)$$

we can determine the probability of z lying in the range a to b. The integration is relatively easy to perform: $\exp(-z^2/2)$ is expanded using the exponential expression

$$\exp(x) = x + x/1! + x^2/2! + x^3/3! + \ldots \text{ etc.} \qquad (5.5)$$

and we then only have to integrate a polynomial. In fact the area under the standard normal curve is usually tabulated for $a = -\infty$ and b typically 0 to +3, for example Table 5.2. Use of such tables is very straightforward. For instance, to find the probability of z lying between -1 and +1 we look up the values of the area for b equals -1 and +1 and subtract them, i.e.

$$\begin{aligned} P(-1 \leq z \leq 1) &= P(-\infty\ z \leq 1) - P(-\infty\ z \leq -1) \\ &= 0.8413 - (1 - 0.8413) \\ &= 0.6826 \end{aligned} \qquad (5.6)$$

Notice that for our real variate y, with mean μ and standard error σ, then, from (5.2)

$$y = z\sigma + \mu \qquad (5.7)$$

Putting $z = -1$ and $z = +1$ into (5.7) gives

$$y = \mu + \sigma \text{ and } y = \mu - \sigma$$

Hence, from (5.5)

$$P(\mu - \sigma \leq y \leq \mu + \sigma) = 0.6826 \qquad (5.8)$$

That is to say that the probability of a variate being within one standard error of its mean is 0.6826 or 68.26%. This is an important result with implications in the interpretation of many of the results in section 4. For instance, if the a posteriori standard error of a coordinate is, say, 10 cm we can state that, so long as a normal population is assumed, there is a 68% (approximately) chance that the true error is less than 10 cm. Roughly

two-thirds of all variates lie within one standard error of their mean. Table 5.2 can also be used to derive the following useful figures:

range of variate	$\mu \pm \sigma$	$\mu \pm 2\sigma$	$\mu \pm 2.5\sigma$	$\mu \pm 3\sigma$
probability (%)	68.26	95.44	98.76	99.74

5.2.2 Multivariate

If we wish to consider several variaties simultaneously it is necessary to use the multivariate normal distribution with the pdf

$$\Phi(x) = \left[\left(\left| C_x^{-1} \right| \right)^{\frac{1}{2}} / (2\pi)^{n/2} \right] \exp \left\{ -(x - \mu)^T C_x^{-1} (x - \mu)/2 \right\} \qquad (5.9)$$

where $x = [x_1, x_2, \ldots, x_n]^T$ is a vector of normally distributed variates with mean vector $\mu = [\mu_1, \mu_2, \ldots, \mu_n]^T$ and covariance matrix C_x. The symbol $\left| C_x^{-1} \right|$ denotes the determinant of the inverse of C_x. To determine the probability of a number of events, e.g. $a_1 \leq x_1 \leq b_1$, $a_2 \leq x_2 \leq b_2$ etc. occurring simultaneously, we need to integrate $\Phi(x)$ as in equation (4.3). Of special interest is the bivariate case where x_1 and x_2 are the eastings and northings (or latitude and longitude) and we would like to find the probability of the true error in x_1 being within one standard error at the same time as the true error in x_2 being within one standard error, i.e. the probability of the least squares estimate of a point lying within an error ellipse centred at the unknown true position of the point (as described in 4.4.2). It can be shown, e.g. Mikhail and Gracie (1981, 230), that this probability is 0.394 or 39.4%. Similarly the following figures are given:

c	1.000	2.000	2.447	3.000
P	0.394	0.865	0.950	0.989

where P is the probability of a point lying within an error ellipse drawn with semi-major and semi-minor axes of $c\sigma_{max}$ and $c\sigma_{min}$. It is common practice to draw error ellipses 2.447 times their "standard" size and then to be 95% sure that the true errors in the positions of the points are smaller than those described by the drawn ellipses.

5.2.3 Importance of the normal distribution

It will become evident in section 6 that observational errors do not need to be normally distributed in order to apply least squares. We do, however,

need to have a normal distribution (or some other known pdf) in order to make statements, such as in 5.2.1 and 5.2.2, regarding the probability of certain events occurring. Also, it will be seen that the normal distribution is crucial to the theory of the statistical tests to be described in 5.4.

It turns out that there are actually good theoretical reasons for using the normal distribution pdf even when we know nothing of the real pdf of our observations or even when we suspect the pdf not to be normal. This is because of a remarkable statistical theorem known as the central limit theorem. The theorem itself will not be explained here but is presented formally in virtually any text on mathematical statistics (although for a proof a more advanced work such as Cramer (1946) would be required). In this paper we merely mention one of the many implications of the theorem, that if we have a variate x that is the sum of a number of independent variates x_1, x_2, ..., x_n then as n increases the pdf of x approaches a normal distribution, irrespective of the pdfs of x_1, x_2, ..., x_n. Now the errors in almost all surveying measurements are in fact the sum of a number of small errors, so it follows from the central limit theorem that, irrespective of the pdfs of the small errors (and so long as there are more than about four of them), a normal pdf will closely describe the total error of a survey measurement. Hence special attention is paid to the normal distribution in almost all books on the treatment of surveying measurements.

5.3 Sampling distributions

In order to carry out certain statistical tests we need to know something of the pdfs of various functions of quantities that are themselves normally distributed. These pdfs, sometimes called sampling distributions, are both lengthy and difficult to derive, so here we will merely name, and explain the relationship between, the three most important ones. Before doing so it should be emphasised that, in general, functions of normally distributed variates will not be normally distributed; for instance, the random variable x^2 is not normally distributed even though x is.

5.3.1 Chi-square distribution

If x_1, x_2, ..., x_ν are ν independent random variables, each normally distributed with zero mean and variance unity, then the function χ^2_ν, where

$$\chi^2_\nu = x_1^2 + x_2^2 + \ldots + x_\nu^2 \qquad (5.10)$$

is said to have a chi-square distribution with ν degrees of freedom. Notice that χ^2_ν will be a random variable. The shape of a typical chi-square pdf is given in Fig. 5.2. It should be noted that as ν increases so the chi-square pdf approaches a normal distribution. Table 5.3 gives what are known as the percentiles of the χ^2_ν distribution. Essentially a percentile is the value of χ^2_ν that, for a given number of degrees of freedom, will be exceeded with a specified probability. For instance, with three degrees of freedom the probability of χ^2_ν being less than 6.25 is 0.90 or 90%. This is illustrated in Fig. 5.2 and can be confirmed by looking up three degrees of freedom in Table 5.3. We would say that 6.25 was the 90% percentile of a chi-square distribution with three degrees of freedom.

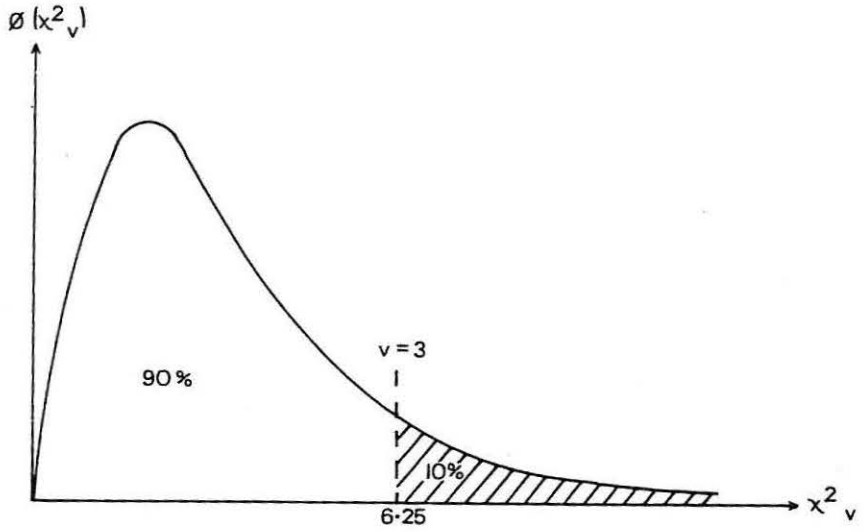


Fig. 5.2

5.3.2 t distribution

If we have two normally distributed random variables, x_1 and x_2, each with mean zero and variance unity, then the quantity t given by

$$t_\nu = x_1/(x_2^2/\nu)^{\frac{1}{2}} \qquad (5.11)$$

is said to have a t distribution with ν degrees of freedom. The distribution is often called the Student's distribution after the statistician W.S. Gosset (at the time a chemist at Guinness's brewery in Dublin) who used the pseudonym "Student". The t pdf is exemplified in Fig. 5.3, which shows the case for four degrees of freedom ($\nu = 4$). The distribution looks somewhat similar

to the normal and is identical to it when $\nu = \infty$. Table 5.4 gives some percentiles for the t distribution: as an example, when $\nu = 4$ there is a 0.9 (90%) probability that t lies between −2.132 and +2.132; this case is also shown in Fig. 5.3.

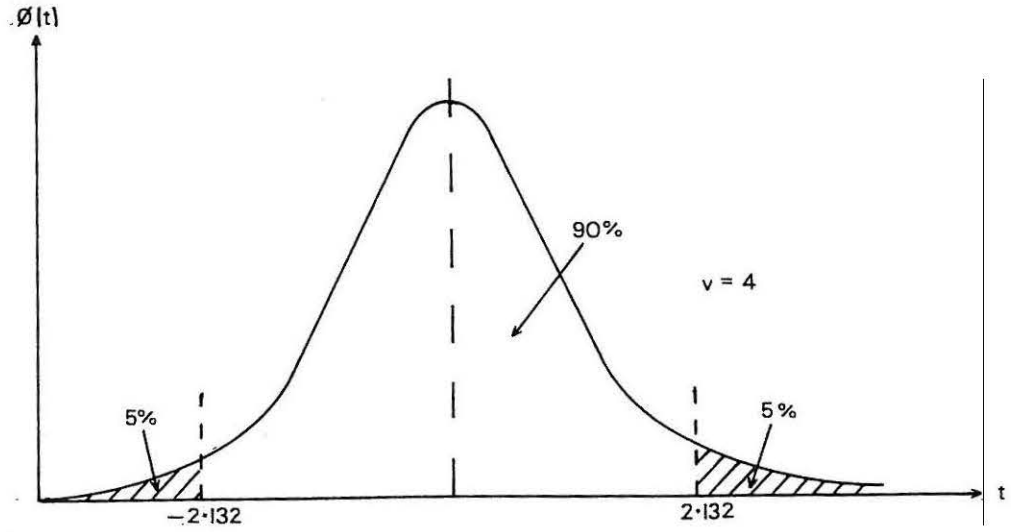


Fig. 5.3

5.3.3 F distribution

If we have two variables x^2 and y^2, both with chi-square distributions, i.e.

$$x^2 = x_1^2 + x_2^2 + \ldots + x_{\nu_1}^2$$
$$y^2 = y_1^2 + y_2^2 + \ldots + y_{\nu_2}^2 \qquad (5.12)$$

with $x_1, x_2, \ldots, x_{\nu_1}$ and $y_1, y_2, \ldots, y_{\nu_2}$ normally distributed with zero mean and unit variance and with ν_1 and ν_2 degrees of freedom respectively, the variable $F_{\nu_1\nu_2}$ given by

$$F_{\nu_1\nu_2} = (x^2/\nu_1)/(y^2/\nu_2) \qquad (5.13)$$

is said to have an F distribution (named after the statistician R.A. Fisher). The F pdf is shown in Fig. 5.4 and some percentiles are given in Table 5.5. As an example, in the case where $\nu_1 = 10$ and $\nu_2 = 20$ there is a 0.95 (95%) probability that F is less than 2.35 (note that F is obviously always positive). It is worth noting that when $\nu_1 = 1$ the F statistic is the

square of the t statistic, and it can be shown that as ν_2 increases the F distribution approaches the χ^2 distribution.

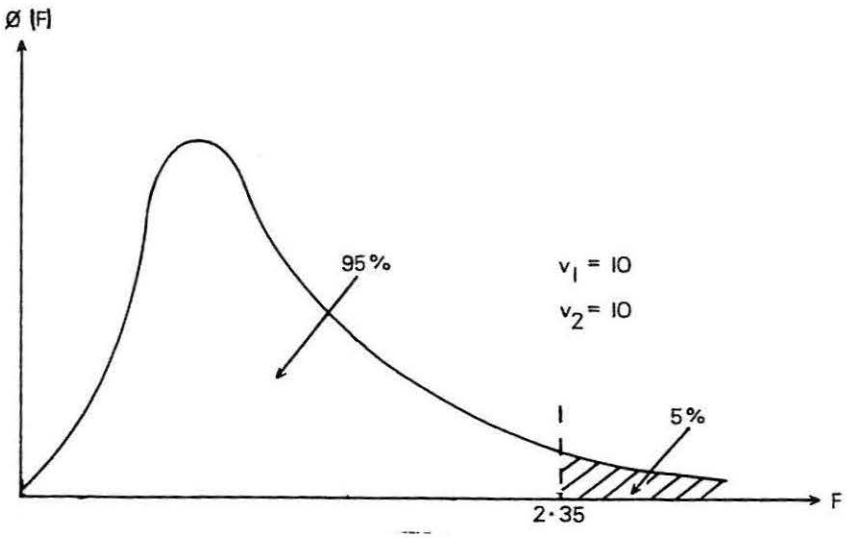


Fig. 5.4

5.4 Statistical tests

In 5.1 the basic terminology of statistical tests was given and in 5.2 and 5.3 the distributions upon which all the tests in this paper are based were described. Now these results will be used in order to explain how to carry out some statistical tests that particularly relate to the computation of position-fixes by least squares.

5.4.1 Identification of outliers

The word "outlier" has no strict statistical definition and can only be defined in an intuitive manner; for instance, Hawkins (1980) defines an outlier as "an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism". The problem of identifying outliers is very important in surveying, especially when data are automatically captured and processed, and never seen in their raw form by the surveyor. Since there is no opportunity in such cases for him to pick out and investigate obvious gross errors, automatic statistical methods of detecting outliers are essential.

In position-fixing the problem can arise in two situations, firstly, as

considered in 5.4.1.1, when we have multiple measurements of a single quantity such as a distance or an angle, or when a position has been computed many times such as in multiple pass Doppler computations, and, secondly, as considered in 5.4.1.2, when we have a number of single measurements (or means) that are "fitted together" during a least squares computation. It must be emphasised that the statistical tests that will be described to deal with these two situations can only "identify" the outliers; they cannot in themselves distinguish between their two possible causes:

(i) there has been a gross error in the measurement

(ii) the basic mathematical model is incorrect.

In practice, (i) above is more common and it would be usual to reject any outlying measurements and recompute the position-fix without them. Before rejection, however, the measurement should be investigated, as the cause of the gross error may be apparent and it may be possible to correct it, e.g. two numbers may have been transposed when keying measured data into a computer. The danger with the automatic rejection of all outliers is that cause (ii) above may be overlooked and important information rejected. For instance, consider a single point position-fix by n measured distances to known stations where n - 1 lines are short and the nth is a long line. If the measuring system has an unmodelled scale error the nth line will have a larger residual than all the rest and may be rejected when the proper solution would be to recast the model with a scale unknown (as in 2.5.2). Model errors are exceptionally difficult to detect and great care must be taken to ensure that any information that may aid their detection is not rejected. Some of the great advances of science, e.g. the discovery of Neptune from the perturbations of Uranus, are due to observations not quite fitting their supposed model.

5.4.1.1 From repeated measurements

If we have a large number of repeated measurements of the same quantity, $x_1, x_2, \ldots, x_n$, it is common practice to compute their mean and standard deviation, $\bar{x}$ and s respectively, and to compute for the ith measurement the statistic

$$r_i = (x_i - \bar{x})/s \qquad (5.14)$$

If the observational errors are normally distributed r_i will have a standard

normal distribution (mean zero and standard error unity) and the probability of r_i being greater than a certain size can be obtained from Table 5.2. If we wish to ascertain whether or not any of the measurements are outliers we set up a test as follows:

H_0: x_i comes from a normal distribution with a mean of μ and standard error σ ($\bar{x}$ and s are estimates of μ and σ)

H_A: x_i comes from a normal distribution with a mean of (μ + gross error) and standard error σ.

We choose a level of significance α (say 0.01, i.e. we are only prepared to identify "inlying" measurements as outliers in one case out of a hundred) and look up the relevant percentile for the normal distribution in Table 5.2 (2.58 for α = 0.01). Then (for this example) we identify as an outlier any measurement for which r_i is greater than 2.58 in modulus. Notice that this is a two-tailed test as we wish to detect both positive and negative outliers. The chances of failing to detect an outlier (type 2 error) are discussed in 5.4.1.2.

The foregoing procedure strictly only applies when n is very large (say greater than 30 as a practical guideline). For smaller values of n a different procedure has to be adopted as the sample mean and standard deviation may not represent very well their equivalents for the population. There are many alternative methods but limited detail will be given here as we are mainly concerned with least squares and 5.4.1.2 is a more relevant problem.

One approach is to use the so-called Chauvenet's criterion which simply posits that the level of confidence should be chosen to be equal to $1/(2n)$, so for say 8 observations, α would be 0.0625 and we would identify as outliers all measurements with r_i greater than 1.82 (from Table 5.2). Whereas if n = 20 then α = 0.025 and the criterion would be 2.24. This method is simple and effective but its mathematical basis is rather doubtful as on average it will identify as an outlier half a good observation (an inlier) per sample regardless of the size of the sample. It should not be used successively to search for more than one outlier because of the manner in which α is related to n.

More rigorous approaches to the problem have mainly been based on the t distribution; for instance, it can be shown that the statistic

$$t = r(n-2)^{\frac{1}{2}}/(n-1-r^2)^{\frac{1}{2}} \qquad (5.15)$$

(where r is defined as in (5.14))has a t distribution with n - 1 degrees of freedom. Hence the t tables (Table 5.4) could be used, just as the normal distribution tables were used in the earlier part of this section, to decide whether or not an observation is an outlier. The procedure is simply to compute the value of t and compare it with a percentile from Table 5.4. Alternatively, it is possible merely to compute r as in (5.14) but to use critical percentiles based on the size of sample and the t distribution. Logan (1955) gives full details of this procedure and includes simple, easy to use tables (for α = 0.05, 0.01 and 0.001). This technique is mathematically identical to the t statistic above but easier to apply in practice.

The identification of outliers is now a very advanced branch of mathematical statistics and readers who wish to persue the topic in depth are recommended to read Hawkins (1980).

5.4.1.2 From the results of a least squares computation

We need to have some method of analysing the results of a least squares computation to determine whether or not any of the observations are outliers. A method for doing this will now be given; note that although the discussion will be limited to the special case of observation equations, the method can be extended to handle the combined case (and of course the special case of condition equations). Also note that although this section is written assuming that the cause of the outlying measurement is a gross error (blunder), the technique is equally applicable to detecting model errors as discussed in 5.4.1.

Let ℓ_i be the ith observation of a vector of n observations ℓ, used in a least squares computation. If we suspect that ℓ_i contains a gross error Δ_i whilst all other observations have only random, normally distributed errors ϵ_i, we have to set up a test as follows

$$H_0 : \ell_i = \bar{\ell}_i + \epsilon_i \qquad (5.16)$$

$$H_A : \ell_i = \bar{\ell}_i + \epsilon_i + \Delta_i \qquad (5.17)$$

Note that if an observation contains a gross error we expect it to belong to a normally distributed population, but about a mean of $\bar{\ell}_i + \Delta_i$ rather than $\bar{\ell}_i$. Clearly in practice we would want to apply (5.16) and (5.17) to every observation in turn.

We now introduce a test statistic $\hat{w}_i$, given by

$$\hat{w}_i = \hat{d}_i / \sigma_{\hat{d}_i} \qquad (5.18)$$

where

$$\hat{d}_i = \ell_i - \hat{\ell}_i^c$$

and

$\hat{\ell}_i^c$ = the ith observed quantity computed from the parameters derived from a least squares computation of all the observations except ℓ_i

$\sigma_{\hat{d}_i}$ = standard error of $\hat{d}_i$.

It has been shown by Baarda (1968) that if H_0 is true $\hat{w}_i$ will be normally distributed with a mean of zero and variance of unity; otherwise, under H_A, the normal distribution will have mean δ_i where

$$\delta_i = \Delta_i / \sigma_{\hat{d}_i} \qquad (5.19)$$

The test is carried out by specifying a level of confidence (e.g. 95%) ($\alpha = 0.05$). We can see from Table 5.2 that $\hat{w}_i$ should be less than 1.96 in 95% of cases; hence if $\hat{w}_i > 1.96$ we reject the observation with a 5% chance of making a type 1 error. If we specify the required power of the test as say 90% ($\beta = 0.10$) we can determine the "upper bound", δ_i^u on δ_i (and hence Δ_i), i.e. we can determine the maximum size of gross error that will be accepted (type 2 error) when it should have been rejected one time in ten. This is done by examining Fig. 5.5. The value of δ_i^u is given as

$$\delta_i^u = a + b \qquad (5.20)$$

and a and b are given from Table 5.2 as 1.96 and 1.28 respectively; hence

$$\delta_i^u = 3.24$$

and, from (5.19),

$$\Delta_i^u = 3.24\, \sigma_{\hat{d}_i} \qquad (5.21)$$

Therefore, the use of the above values for α and β and rejecting all observations with a value of $\hat{w}_i$ greater than 1.96 will cause rejection of "correct" observations in five percent of the observations (type 1 error) whilst one tenth of gross errors greater than or equal to $3.24\, \sigma_{\hat{d}_i}$ will remain

undetected (type 2 error).

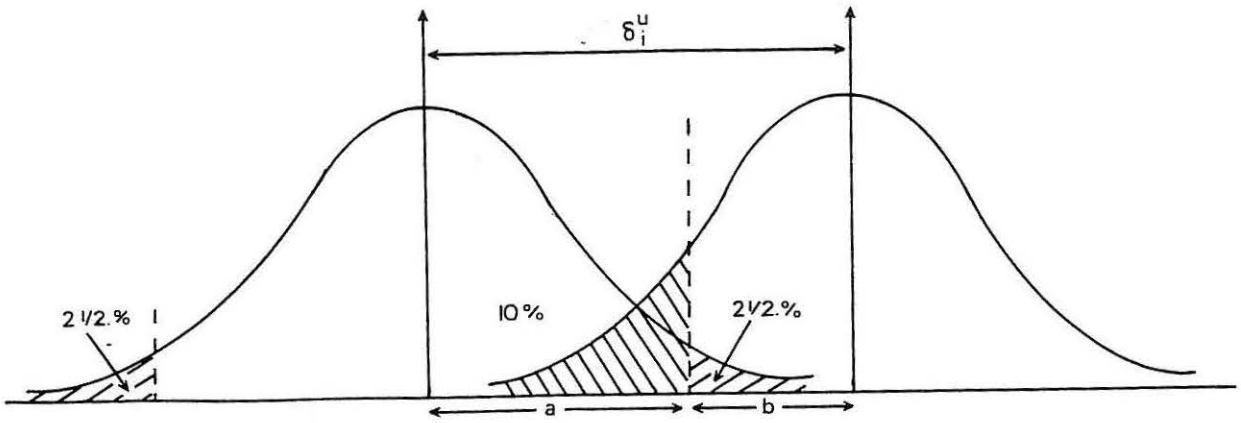


Fig. 5.5

In practice it would be very tedious to have to compute $\hat{w}_i$ and σ_{d_i} as indicated by (5.18); hence we would prefer to have a more straightforward means of computing it. It is shown in Appendix 2 that

$$\hat{w}_i = - e_i^T W \hat{v} / (e_i^T W C_{\hat{v}} W e_i)^{\frac{1}{2}} \tag{5.22}$$

and

$$\sigma_{\hat{d}_i} = 1/(e_i^T W C_{\hat{v}} W e_i)^{\frac{1}{2}} \tag{5.23}$$

where e_i is a null vector but for the ith element which is unity, i.e.

$$e_i = [0, 0, \ldots, 1, \ldots, 0, 0]^T \tag{5.24}$$

In the case of W being diagonal (which is usual in practice) with the standard error of the ith observation being σ_i, it is shown in Appendix 2 that (5.22) and (5.23) simplify to

$$\hat{w}_i = \hat{v}_i / \sigma_{\hat{v}_i} \tag{5.25}$$

and

$$\sigma_{\hat{d}_i} = \sigma_i^2 / \sigma_{\hat{v}_i} \tag{5.26}$$

i.e. $\hat{w}_i$ is the ratio of a least squares residual to its standard error. Note that $\hat{v}_i$ is given by (3.32) and $\sigma_{\hat{v}_i}^2$ is the ith diagonal element of $C_{\hat{v}}$ given

by (4.68). It is really quite remarkable that (5.18) and (5.25) should be equivalent and that there should be such a simple test for the rejection of outliers using the results of a least squares computation.

If it is suspected that a number of observations simultaneously contain gross errors then we can use, according to Kok et al (1980), the same one-dimensional w test described above except that the vector e_i in (5.22) becomes e where e is a null vector but for the elements which correspond to the observations to be tested simultaneously, which will be unity. Such a test is, however, rarely carried out as it would be unusual to suspect a particular group of observations of containing gross errors and to test every possible group would be quite impractical. The test is usually used as described, i.e. for each individual observation in turn, in which case the process is sometimes referred to as"data snooping".

The foregoing testing procedure, based on the normal distribution, is commonly known as the B-method of testing, after W.W. Baarda, who first introduced it. It is strictly only correct when each value of $\sigma_{\hat{v}_i}$ truly reflects the population from which each residual has been drawn. This will only be the case when we are sure that we are using the correct W, either because we have external evidence or because we have a very large sample of observations. In general this is not the case and we should adopt a slightly different testing procedure. Pope (1976) has made a detailed theoretical study of the statistics of residuals and shown that, so long as the unit variance is unity, τ, computed from (5.25), has the following distribution

$$\tau_\nu = (\sqrt{\nu}\, t_{\nu-1})/(\nu - 1 + t^2_{\nu-1})^{\frac{1}{2}} \qquad (5.27)$$

where $t_{\nu-1}$ is a t distribution with $\nu-1$ degrees of freedom and ν is the number of degrees of freedom in the least squares computation. This is known as the tau distribution and some of its percentiles are given in Table 5.6. The testing procedure is as follows.

(i) Carry out the least squares computation in the usual way and compute, for each observation, $\hat{v}_i$ and $\sigma_{\hat{v}_i}$.

(ii) Compute, for each observation, the tau statistic from

$$\tau = \hat{v}_i/(\sigma_o\, \sigma_{\hat{v}_i}) \qquad (5.28)$$

where σ_o is the standard error of an observation of unit weight.

(iii) Select a value for α, the level of significance, and compare τ with the relevant critical value in Table 5.6. If τ is the greater then the observation is liable for rejection.

The above process is based on the assumption that only one blunder exists in the observations. In practice it is quite possible for a number of gross errors to be introduced, e.g. a gross centering error during the measurement of a triangulation network will cause gross errors in all observations at or to that point. Very little work has been done on this problem but interested readers are referred to the already mentioned Kok et al (1980) and to Stefanovic (1978) for details of possible strategies.

5.4.1.3 Reliability of a position-fix

The reliability of a position-fix is a measure of the ease with which gross errors may be detected. Imagine a single point being fixed by measured distances to two known points. There would be no redundant measurements and (3.43) and (4.68) would give both $\hat{v}_i$ and $\sigma_{\hat{v}_i}$ as zero for both measurements. Hence $\hat{w}_i$ would be indeterminate and no check for gross errors would be possible. This of course is reasonable, as any two distance measurements would produce a position and there is obviously no way of detecting a gross error. We would consider that such a position-fix was unreliable, although of course it may be very precise with small error ellipses etc.

It is useful to have some quantitative measure of the reliability of a position-fix. Pelzer (1979) has introduced the quantity τ_i defined by

$$\tau_i^2 = 1/(e_i^T C_\ell e_i e_i^T W C_{\hat{v}} W e_i) \qquad (5.29)$$

or more simply in the case of a diagonal C_ℓ (and hence diagonal W)

$$\begin{aligned} \tau_i^2 &= 1/(\sigma_i^2 e_i^T W C_{\hat{v}} W e_i) \\ &= \sigma_{\hat{d}_i}^2 / \sigma_i^2 \text{ (using (5.23))} \end{aligned} \qquad (5.30)$$

Now, rearranging (5.19), we have

$$\begin{aligned} \Delta_i^u &= \delta_i^u \sigma_{\hat{d}_i} \\ &= \delta_i^u \sigma_i \sigma_{\hat{d}_i} / \sigma_i \end{aligned} \qquad (5.31)$$

and taking the square root of (5.30) and substituting in (5.31) gives

$$\Delta_i^U = \delta_i^U \sigma_i \tau_i \tag{5.32}$$

where δ_i^U depends only on the chosen probabilities, α and β, of type 1 and type 2 errors. Hence once τ_i has been computed for an observation we can simply compute from (5.32) the maximum undetected gross error in that observation (with probability β). For instance, using the figures for α and β in 5.4.1.2 we could say that there was a 10% chance of a gross error greater than 3.24 $\sigma_i \tau_i$ remaining undetected. It is extremely important to emphasise that τ_i does not depend on $\hat{v}$ (i.e. on the measurements themselves) so a position-fix can be analysed for reliability in advance of its actual observation. Hence both precision and reliability can be considered when designing position-fixing procedures.

Note that, when W is diagonal, substitution of (5.26) in (5.30) yields

$$\tau_i = \sigma_i / \sigma_{\hat{v}_i} \tag{5.33}$$

i.e. the ratio of the standard error of an observation to the standard error of its least squares residual (given by (4.68)). Ashkenazi (1980) has suggested an alternative measure, ρ_i, given by

$$\rho_i = \hat{\sigma}_i / \sigma_i \tag{5.34}$$

where $\hat{\sigma}_i$ is the standard error of the least squares estimate of the ith observed quantity and is given by the square root of the ith diagonal element of $C_{\hat{\ell}}$ (see (4.69)). Note that (5.33) and (5.34) are closely related because, when W is diagonal, (4.69) simplifies to

$$\sigma_{\hat{v}}^2 = \sigma_i^2 - \hat{\sigma}_i^2 \tag{5.35}$$

The disadvantage of ρ_i compared with τ_i is that it is not possible to make the same kind of simple statement regarding the largest undetected gross error.

Baarda (1968) has described the foregoing concept of reliability as "internal reliability" and he further defines "external reliability" as the effect of an undetected gross error on the parameters and on quantities computed from the parameters. It could be argued that external reliability is more important than internal reliability as undetected gross errors are of no consequence if they do not significantly effect the parameters. External reliability is determined as follows.

Consider a quantity ψ, with least squares estimate $\hat{\psi}$, computed from the parameters and let $\Delta\hat{\psi}_i$ be the effect on $\hat{\psi}$ of a gross error of size Δ_i^u, as given by (5.32), in the ith observation. Then it is shown in Appendix 2 that, for uncorrelated observations,

$$\Delta\hat{\psi}_i \leq \delta_i^u \gamma_i \sigma_{\hat{\psi}} \tag{5.36}$$

where

$$\gamma_i = \hat{\sigma}_i/\sigma_{\hat{v}_i} \tag{5.37}$$

Hence if a gross error equal to the boundary value is made in the ith observation the resulting error in $\hat{\psi}$ must be less than γ_i multiplied by a constant multiplied by the standard error of $\hat{\psi}$. Note that $\sigma_{\hat{\psi}}$ is evaluated according to 4.4.3. Therefore γ_i is a measure of the external reliability: it is a factor which multiplies the standard error of the desired function of the parameters to give the <u>maximum</u> effect of Δ_i^u on that function. Notice that γ_i depends on the selected observation and not on ψ, i.e. it is the same for all functions of the parameters. Also it can be computed without the values of the observations and can hence be used as a criterion for the design of position-fixes.

In the case of uncorrelated observations the relationship between τ_i and γ_i can be derived as follows.

Combining (5.33) and (5.37) yields

$$\tau_i^2 - \gamma_i^2 = (\sigma_i^2 - \hat{\sigma}_i^2)/\sigma_{\hat{v}_i}^2 \tag{5.38}$$

$$= 1 \text{ (from (5.35))} \tag{5.39}$$

i.e.
$$\gamma_i^2 = \tau_i^2 - 1 \tag{5.40}$$

Hence the computation of τ_i automatically leads to γ_i. It follows from (5.40) that if an observation has high internal reliability it must also have high external reliability, and conversely low internal reliability reflects low external reliability.

5.4.2 Variance ratio tests

5.4.2.1 The testing procedure

There are a number of situations in which it is required to compare the variances (or standard errors) of two samples to discover whether or not they come from populations with the same variances. For instance, one variance may

have been computed from angles measured by observer A and another from those measured by observer B and it may be required (perhaps for the purposes of weighting) to find out if the two observers are performing with the same precision. Further examples will be given after the test has been described.

Let s_1 and s_2 be the variances of samples 1 and 2 computed with ν_1 and ν_2 degrees of freedom. If σ_1 and σ_2 are the (unknown) variances of the two populations from which the samples have been drawn we set up the two-sided test as follows:

$$H_0: \sigma_1 = \sigma_2 \tag{5.41}$$

$$H_A: \sigma_1 \neq \sigma_2 \tag{5.42}$$

It can be shown, e.g. in Mood and Graybill (1963), that F, given by

$$F = s_1^2/s_2^2 \tag{5.43}$$

where $s_1 > s_2$, has the F distribution given in (5.13) with ν_1 and ν_2 degrees of freedom. Hence the testing procedure is to select a level of significance, α, and look up the percentile for ν_1 and ν_2 in the appropriate F distribution table. Note that since we are setting up a two-sided test it is necessary to double the probabilities in the F distribution table, i.e. use the table for $\alpha/2$. If the alternative hypothesis was

$$H_A: \sigma_1 > \sigma_2 \tag{5.44}$$

we would have a one-sided test and would use the table for α.

Once we have looked up the appropriate value of the percentile, we simply compare it with our statistic computed by (5.43). If the statistic is the larger we reject the null hypothesis with a probability of α of making a type 1 error; otherwise we accept the null hypothesis.

As an example say that s_1 is estimated to be 6.0 with 40 degrees of freedom, s_2 is estimated to be 3.2 with 10 degrees of freedom, and we wish to make the two-sided test (5.41) and (5.42) with $\alpha = 0.02$ (2%). We look up the necessary percentile in Table 5.5 (at the 1% level) with $\nu_1 = 40$ and $\nu_2 = 10$ to obtain 4.17. We then compute the F statistic from (5.43) as

$$F = (6.0/3.2)^2 = 3.52$$

and since F is less than 4.17 we would accept the null hypothesis (i.e. that

the two samples come from populations with the same variance) at the 2% level.

5.4.2.2 Examples of variance ratio tests

(i) Unit variance

Probably the most important application of the variance ratio test is the testing of the proximity to unity of the unit variance computed from

$$s_o^2 = (\hat{v}^T W \hat{v})/(n - m) \qquad (5.45)$$

after a least squares position-fix computation. Notice that the notation in (5.45) has been deliberately changed from (4.25) with s_o replacing σ_o. This is to conform with the notation of this section (Greek and Roman letters for population and sample statistics respectively). The test is set up as follows (note s_o is our estimate of σ_o)

$$H_0: \sigma_o^2 = 1 \qquad (5.46)$$

$$H_A: \sigma_o^2 \neq 1 \qquad (5.47)$$

and F is computed as in either (a) or (b) below.

(a) If $s_o > 1$

$$F = s_o^2/1 = s_o^2 \qquad (5.48)$$

$$\nu_1 = n - m, \ \nu_2 = \infty$$

(b) If $s_o < 1$

$$F = 1/s_o^2 \qquad (5.49)$$

$$\nu_1 = \infty, \ \nu_2 = n - m$$

As an example suppose $(n - m)$ is 6 and $s_o^2 = 2.46$, and we wish to set a level of confidence of 90% ($\alpha = 0.10$). The 0.05 (i.e. $\alpha/2$) F table (Table 5.5) gives, for $\nu_1 = 6$ and $\nu_2 = \infty$, a percentile of 2.10. Hence computing F from (5.48) as 2.46 would lead to the rejection of the null hypothesis and the conclusion that our value for the unit variance was significantly different from unity (with a 10% chance of making a type I error).

It was remarked in 4.2.4 that in some position-fixing problems (e.g. on a moving vessel) the unit variance is best estimated by averaging a large number of separately determined values. In such cases the above test is carried out exactly as described but with $\Sigma(n - m) - 1$ degrees of freedom rather than $(n - m)$.

It is interesting to note that when ν_2, or ν_1, is ∞, as is the case for unit variance testing, then the F statistic, or its reciprical, has a χ^2 distribution. Hence it is possible to use the χ^2 tables instead of the F tables. For an example of how to do this see Mikhail and Gracie (1981, 219-220).

It should be mentioned that if the null hypothesis is rejected, i.e. if it is decided that σ_o is significantly different from unity, there are two possible interpretations (assuming that there are no gross errors). The first, discussed in detail in 4.2.4, is that the assigned a priori covariance matrix needs to be multiplied by s_o^2, and the second is that the model used for the least squares computation is incorrect or incomplete. For instance, if a scale error exists in a distance measuring system, and it is not modelled, unexpectedly large residuals would arise, mainly in the corresponding distance equations (note that this would only occur if there was some other scale input into the system, e.g. via more than one fixed point). It is usually very difficult to choose between the two interpretations. Obviously if we were absolutely sure of our a priori variances we could deduce an incomplete model and vice versa but such confidence rarely exists in practice.

(ii) Comparison of instrument or observer performance

To find out whether or not there is any significant difference between the performance of various observers or instruments the F test is applied exactly as in the example in 5.4.2.1. If we merely wish to determine whether or not their performance is different we use a two-sided test but if we wish to test whether or not one instrument or observer is better than another we use a one-sided test. The sample variances may come either from repeated measurements or from the results of a least squares computation (via the covariance matrix $C_{\hat{x}}$).

(iii) Test of observing conditions

If we have observed under certain conditions and later repeated the measurements (or made similar measurements) under different conditions, we can use the F test to determine whether or not the change in conditions has significantly affected the precision of the measurements. The procedure is exactly as in the example in 5.4.2.1.

For example, when carrying out the retriangulation of Great Britain the Ordnance Survey made most of their angle measurements in East Anglia on towers because of the very flat ground and it is obviously relevent (e.g. for assigning C_ℓ in a least squares compuation) to question whether or not using towers significantly degraded the precision of angle measurement. The tables of triangle misclosures in Ordnance Survey (1967) give the data needed to answer this question by means of the F test.

5.4.3 Comparison of means

5.4.3.1 Between two samples

Suppose we measure a distance many times with one instrument and then repeat the process with another and hence determine two mean distances, $\bar{x}_1$ and $\bar{x}_2$, from the first and second instruments respectively. We may wish to know whether or not the difference between the two means is significant, i.e. whether or not one of the instruments produces a significantly different measurement to the other. If so we would conclude that there was some kind of systematic error in one (or perhaps both) of the instruments. Alternatively we may have carried out the two sets of measurements at different epochs with well calibrated instruments and a significant difference could now be interpreted as a change in the distance (crustal movement). We can set up a test for the above as follows: let μ_1 and μ_2 be the population means estimated by $\bar{x}_1$ and $\bar{x}_2$; then we have the two-sided test

$$H_0: \mu_1 = \mu_2 \ (\text{or } \mu_1 \geq \mu_2) \qquad (5.50)$$

$$H_A: \mu_1 \neq \mu_2 \ (\text{or } \mu_1 < \mu_2) \qquad (5.51)$$

where the bracketed alternatives relate to a one-sided test.

If s_1 and s_2 are the two sample standard errors and n_1 and n_2 the number of

measurements in each sample it can be shown, e.g. in Mood and Graybill (1963), that the statistic

$$t = (\bar{x}_1 - \bar{x}_2)/(s^2(1/n_1 + 1/n_2))^{\frac{1}{2}} \qquad (5.52)$$

will have a t distribution with ν degrees of freedom where

$$\nu = n_1 + n_2 - 2$$

and

$$s^2 = (s_1^2(n_1 - 1) + s_2^2(n_2 - 1))/\nu$$

The procedure then is to select α, compute t from (5.52) and look up the relevant percentile in Table 5.4. Then, if t is greater than the percentile we reject the null hypothesis and otherwise we accept it.

For example

$$\bar{x}_1 = 146.214 \quad , \quad s_1 = 0.011 \quad , \quad n_1 = 8$$

and

$$\bar{x}_2 = 146.206 \quad , \quad s_2 = 0.006 \quad , \quad n_2 = 12$$

then $\nu = 18$ and, if we choose to perform a two-sided test with $\alpha = 0.10$, we look up the percentile in Table 5.4 and obtain a value of 1.734. t computed from (5.47) is 2.109 so we reject the null hypothesis at the 10% level of significance (10% chance of a type 1 error) and conclude that the difference between $\bar{x}_1$ and $\bar{x}_2$ is significant.

5.4.3.2 Between one sample and a known value

Sometimes we may wish to compare a mean not with another mean but with a known value, for instance when calibrating a distance measuring device over a known distance or testing a navigation system by repeatedly fixing a known stationary point. In these circumstances we proceed exactly as in 5.4.3.1 but the test statistic (5.47) becomes

$$t = (\bar{x} - \mu)n^{\frac{1}{2}}/s \qquad (5.53)$$

where $\bar{x}$ is the mean from n measurements and s its standard error. μ is the known value of the quantity (i.e. the population mean) and the t tables are used with $\nu = n - 1$.

Notice that (5.52) can be written in the form

$$t = (\bar{x} - \mu)/s_{\bar{x}} \qquad (5.54)$$

where $s_{\bar{x}} = s/n^{\frac{1}{2}}$ is the standard error of the mean $\bar{x}$. In this form it can be used to test the significance of certain results from a least squares computation. For instance, when we have included non-positional unknowns in the parameters of a least squares computation we may wish to question whether or not the quantities they are modelling actually exist (i.e. whether or not there really was any need to model them in the first place). We can proceed as follows.

Let p be the least squares estimate of a parameter with true value ρ. We have

$$H_0: \rho = 0$$

$$H_A: \rho \neq 0$$

then, if s_p is the standard error of p (obtained by taking the square root of the relevant diagonal element of $C_{\hat{x}}$), (5.54) becomes

$$t = (p - 0)/s_p = p / s_p \qquad (5.55)$$

and the test is carried out exactly as in 5.4.3.1 with the number of degrees of freedom given by $\nu - 1$, where ν is the number of observation equations containing the parameter.

5.4.4 Goodness of fit

Throughout this discussion of statistical testing we have assumed that the underlying pdf of our observational errors is normal. Although (as discussed in 5.2.3) there are good reasons for doing this we may nevertheless wish to test this assumption. For example, we may suspect that some non-random influence is interfering in some way with our measurement process. Hence we need a procedure for testing the "distribution" of our data.

Although the goodness of fit test to be described can in fact be used to test the goodness of fit to any distribution, this section is directed specifically at the normal distribution since, in practical position-fixing, we are really only interested in this distribution. Hence we formulate the test as follows.

H_0: $x_1, x_2, \ldots, x_n$ are normally distributed with mean $\bar{x}$ and variance s^2

H_A: $x_1, x_2, \ldots, x_n$ are not normally distributed with mean $\bar{x}$ and variance s^2.

where $x_1, x_2, \ldots, x_n$ are the n variates in whose pdf we are interested and $\bar{x}$ and s^2 are computed in the usual way from these n variates.

The procedure is essentially to split the data into a number, say p, of equal classes and to draw a histogram. If O_i is the number of variates in the ith class and E_i is the number expected in a normal distribution (computed from Table 5.2) it can be shown, e.g. Mood and Graybill (1963), that the statistic

$$\chi^2 = \sum_{i=1}^{p} ((O_i - E_i)^2/E_i) \qquad (5.56)$$

has the χ^2 distribution with p-3 degrees of freedom. Strictly speaking the number of degrees of freedom is p-1-q where q is the number of statistics drawn from the sample. In the above case q = 2: $\bar{x}$ and s^2. If the mean and standard error were known the null hypothesis would be

H_0: $x_1, x_2, \ldots, x_n$ are normally distributed with mean μ and variance σ^2

so we would have q = 0 and there would be p - 1 degrees of freedom.

First we select a level of significance α and look up the relevant percentile in the χ^2 tables (Table 5.3). We then simply compare our value of χ^2 from (5.56) with this percentile and if the percentile is greater we accept the null hypothesis; otherwise we reject it.

There may be a problem with the choice of class-width when dividing the sample and with the cut-off at the two ends of the pdf. The latter arises because E_i becomes very small as the size of the variate increases. If readers are particularly concerned with this they can consult Mann and Wald (1942).

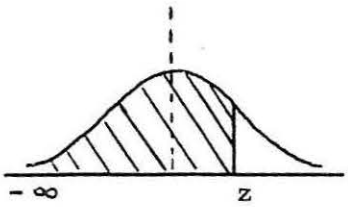
The χ^2 goodness of fit test is most commonly applied to sets of least squares residuals. For instance it is standard practice when computing a geodetic satellite-Doppler fix to check the normality of the two hundred or so residuals from each satellite pass. If it is decided that the residuals are not normally distributed it would be usual to reject the complete pass from the position-fix computation. Similarly, the residuals from a large triangulation network may be checked in this way. Although it is unlikely that the network would be rejected if the residuals failed the test it may uncover some non-random aspect of the observing procedure or possibly a modelling error.

5.5 Concluding remarks

All the foregoing statistical tests are based on a level of confidence $(1 - \alpha)$

which must be subjectively assigned by the individuals carrying them out. Generally tests are performed at the 95% and/or 99% level of confidence (α = 0.05 or 0.01) and values of the statistic larger than the relevant percentile are termed "significant" or "highly significant" respectively. It is fairly common to consider that if the null hypothesis is accepted at the significant level (95%) or rejected at the highly significant level (99%) then there is no need to question the data further, whereas in between (rejection at 95% but not at 99%) it would be usual to look for some other evidence upon which to base a decision.

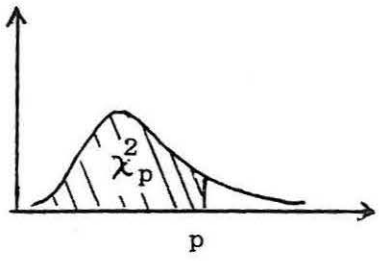
It cannot be emphasised too strongly that we can never be absolutely certain of anything as a result of a statistical test. The level of confidence cannot be set too close to 100% as the probability of type 2 errors would become unacceptable. Hence we should view statistical testing simply as a means (albeit a very valuable one) of procuring evidence, which, along with other evidence, can help the process of making decisions based on the observed data (and on information computed from them, e.g. positions).



z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.	.864[illegible]	.866[illegible]	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.884[illegible]	.[illegible]869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9648	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9874	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.995[illegible]	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.	.9987	.9990	.9993	.9995	.9997	.9998	.9998	.9999	.9999	1.0000

Table 5.2 Area under the normal curve

Source: Mikhail and Gracie (1981, 326) where it is reproduced from Introduction to Probability and Statistics by B W Lindgren and G W McElrath (Macmillan, New York, 1969)

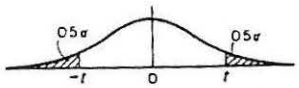


Degrees of freedom (ν)	1	5	90	95	99	99·9
1	$0{\cdot}0^3157$	0·00393	2·71	3·84	6·63	10·83
2	0·0201	0·103	4·61	5·99	9·21	13·81
3	0·115	0·352	6·25	7·81	11·34	16·27
4	0·297	0·711	7·78	9·49	13·28	18·47
5	0·554	1·15	9·24	11·07	15·09	20·52
6	0·872	1·64	10·64	12·59	16·81	22·46
7	1·24	2·17	12·02	14·07	18·48	24·32
8	1·65	2·73	13·36	15·51	20·09	26·12
9	2·09	3·33	14·68	16·92	21·67	27·88
10	2·56	3·94	15·99	18·31	23·21	29·59
11	3·05	4·57	17·28	19·68	24·73	31·26
12	3·57	5·23	18·55	21·03	26·22	32·91
14	4·66	6·57	21·06	23·68	29·14	36·12
16	5·81	7·96	23·54	26·30	32·00	39·25
18	7·01	9·39	25·99	28·87	34·81	42·31
20	8·26	10·85	28·41	31·41	37·57	45·31
22	9·54	12·34	30·81	33·92	40·29	48·27
24	10·86	13·85	33·20	36·42	42·98	51·18
26	12·20	15·38	35·56	38·89	45·64	54·05
28	13·56	16·93	37·92	41·34	48·28	56·89
30	14·95	18·49	40·26	43·77	50·89	59·70

Table 5.3 Percentiles of the χ^2 distribution

Source: Wetherill G B 1982 Elementary Statistical Methods 3rd edition Chapman and Hall London 356 pp page 327

Reproduced by permission of the publishers

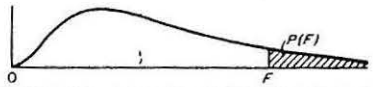
Distribution of t 

Degrees of freedom ν	*Probability* α			
	0.10	*0.05*	*0.01*	*0.001*
1	6.314	12.706	63.657	636.619
2	2.920	4.303	9.925	31.598
3	2.353	3.182	5.841	12.941
4	2.132	2.776	4.604	8.610
5	2.015	2.571	4.032	6.859
6	1.943	2.447	3.707	5.959
7	1.895	2.365	3.499	5.405
8	1.860	2.306	3.355	5.041
9	1.833	2.262	3.250	4.781
10	1.812	2.228	3.169	4.587
11	1.796	2.201	3.106	4.437
12	1.782	2.179	3.055	4.318
13	1.771	2.160	3.012	4.221
14	1.761	2.145	2.977	4.140
15	1.753	2.131	2.947	4.073
16	1.746	2.120	2.921	4.015
17	1.740	2.110	2.898	3.965
18	1.734	2.101	2.878	3.922
19	1.729	2.093	2.861	3.883
20	1.725	2.086	2.845	3.850
21	1.721	2.080	2.831	3.819
22	1.717	2.074	2.819	3.792
23	1.714	2.069	2.807	3.767
24	1.711	2.064	2.797	3.745
25	1.708	2.060	2.787	3.725
26	1.706	2.056	2.779	3.707
27	1.703	2.052	2.771	3.690
28	1.701	2.048	2.763	3.674
29	1.699	2.045	2.756	3.659
30	1.697	2.042	2.750	3.646
40	1.684	2.021	2.704	3.551
60	1.671	2.000	2.660	3.460
120	1.658	1.980	2.617	3.373
∞	1.645	1.960	2.576	3.291

Table 5.4 Percentiles of the t distribution

Source: Kennedy and Neville (1976, 457) where it is taken from Table III, page 46 of Statistical Tables for Biological, Agricultural and Medical Research by Fisher and Yates (6th edition, Longman, London: previously published by Oliver and Boyd, Edinburgh)

5% level of significance



ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.09	251.14	252.20	253.25	254.32
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Table 5.5 Percentiles of the F distribution

Source: Kennedy and Neville (1976, 460–461) where it [illegible] ed from Statistics Manual by E L Crow, F A Davies and M W Maxfield (Dover, New York, 1960)

Reproduced by permission of the publishers

1% level of significance

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052.4	4,999.5	5,403.3	5,624.6	5,763.7	5,859.0	5,928.3	5,981.6	6,022.5	6,055.8	6,106.3	6,157.3	6,208.7	6,234.6	6,260.7	6,286.8	6,313.0	6,339.4	6,366.0
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.74	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.82	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Table 5.5 (continued)

NT=NO. OF OBSERVATIONS
NU=DEGREES OF FREEDOM

REJECTION LEVEL=0.01 (PROBABILITY OF A TYPE-I ERROR)

NT \ NU	4	6	8	10	12	14	16	18	20	30	40	50	60	70	80	90	100	110	130	150
8	1.98	2.32																		
10	1.98	2.33	2.54																	
12	1.98	2.34	2.56	2.70																
14	1.99	2.34	2.57	2.71	2.82															
16	1.99	2.35	2.58	2.73	2.83	2.91														
18	1.99	2.35	2.59	2.74	2.85	2.93	2.99													
20	1.99	2.36	2.59	2.75	2.86	2.94	3.01	3.06												
30	1.99	2.37	2.62	2.79	2.91	3.00	3.06	3.12	3.16											
40	1.99	2.38	2.64	2.81	2.94	3.03	3.10	3.16	3.21	3.35										
50	1.99	2.39	2.65	2.83	2.96	3.06	3.13	3.19	3.24	3.39	3.47									
60	1.99	2.39	2.66	2.84	2.98	3.08	3.16	3.22	3.27	3.43	3.51	3.56								
70	2.00	2.39	2.67	2.85	2.99	3.09	3.17	3.24	3.29	3.45	3.54	3.59	3.62							
80	2.00	2.40	2.67	2.86	3.00	3.11	3.19	3.26	3.31	3.48	3.56	3.62	3.65	3.68						
90	2.00	2.40	2.68	2.87	3.01	3.12	3.20	3.27	3.33	3.50	3.58	3.64	3.68	3.70	3.72					
100	2.00	2.40	2.68	2.88	3.02	3.13	3.22	3.28	3.34	3.51	3.60	3.66	3.70	3.72	3.74	3.76				
150	2.00	2.41	2.70	2.90	3.05	3.17	3.26	3.33	3.39	3.58	3.68	3.74	3.78	3.81	3.83	3.85	3.86	3.87	3.89	
200	2.00	2.41	2.71	2.92	3.08	3.20	3.29	3.37	3.43	3.62	3.73	3.79	3.83	3.86	3.89	3.91	3.92	3.93	3.95	3.96
250	2.00	2.42	2.72	2.93	3.09	3.22	3.31	3.39	3.46	3.66	3.77	3.83	3.88	3.91	3.93	3.95	3.97	3.98	4.00	4.01
300	2.00	2.42	2.72	2.94	3.11	3.23	3.33	3.41	3.48	3.69	3.80	3.86	3.91	3.94	3.97	3.99	4.00	4.02	4.04	4.05
350	2.00	2.42	2.73	2.95	3.12	3.24	3.35	3.43	3.50	3.71	3.82	3.89	3.94	3.97	4.00	4.02	4.03	4.05	4.07	4.08
400	2.00	2.42	2.73	2.96	3.12	3.25	3.36	3.44	3.51	3.73	3.84	3.91	3.96	4.00	4.02	4.04	4.06	4.07	4.10	4.11

Table 5.6 Some critical values of tau

6. Justification of least squares

In section 3 the least squares solution was defined as the one which minimised a specified quadratic form, viz.

$$v^T W v = \text{minimum} \tag{3.6}$$

where

$$W = C_\ell^{-1} \tag{3.7}$$

It is quite natural to ask the question: why use (3.6) and (3.7)? Why not use some other function for example rather than least squares ((3.6) when $W = I$)

$$v_1^2 + v_2^2 + \ldots + v_n^2 = \text{minimum} \tag{6.1}$$

why not have least cubes

$$v_1^3 + v_2^3 + \ldots + v_n^3 = \text{minimum} \tag{6.2}$$

or least product

$$v_1 v_2 v_3 \ \ldots \ v_n = \text{minimum} \tag{6.3}$$

etc?

To justify the use of (3.6) and (3.7) we will now derive and analyse the statistical properties of estimates computed by the least squares method and it will be seen that, from a number of different statistical points of view, the least squares estimates can be described as the "best estimates". To avoid lengthy algebra, discussion will be limited to the special case of observation equations but all results can be extended to the more general combined least squares problem.

Each of the following sections 6.1 to 6.4 considers a particular statistical property of $\hat{x}$, the least squares estimate of the parameters from an observation equations model. 6.5 includes a summary of all properties and of the history of least squares, and some practical points are made in 6.6.

6.1 Unbiased estimate

To prove that $\hat{x}$ is unbiased it is necessary to show, from (4.10), that

$$E(\hat{x}) = x \tag{6.4}$$

We start with (2.22), the basic linearised mathematical model for the case of observation equations,

$$Ax = b + v \tag{2.22}$$

or

$$b = Ax - v \tag{6.5}$$

$$= Ax + e \tag{6.6}$$

where e is the vector of true errors (from (2.1) and (2.2)). Taking expectations in (6.6) gives

$$E(b) = AE(x) + E(e) \tag{6.7}$$

but obviously the expected value of the true errors is zero, $E(e) = 0$; hence we have

$$E(b) = AE(x) \tag{6.8}$$

$$= Ax \text{ (for true values)} \tag{6.9}$$

Now, from (3.41),

$$\hat{x} = (A^T W A)^{-1} A^T W b \tag{3.41}$$

$$\therefore \quad E(\hat{x}) = (A^T W A)^{-1} A^T W E(b) \tag{6.10}$$

Substituting (6.9) in (6.10) gives

$$E(\hat{x}) = (A^T W A)^{-1} A^T W A x \tag{6.11}$$

$$= x \tag{6.12}$$

Which is the same as (6.4). Hence we can say that the least squares estimate is <u>unbiased</u>, i.e. on average the least squares solution is equal to the true solution. It is worth noting that the above proof does not rely in any way on the contents of W, and so the least squares estimate is unbiased irrespective of the choice of W.

6.2 Minimum trace of the covariance matrix of the parameters

It is reasonable to say that the best estimate of a quantity is the one with the minimum variance because, given the choice of two estimates, one with a large variance and the other with a small variance, we would obviously choose the one with the smaller variance. Hence we would like to have some estimate

of x, say x', for which the covariance matrix $C_{x'}$ was smaller than the covariance matrix for any other estimate. Since the words large and small do not have a single meaning where matrices are concerned, we need to define some property of $C_{x'}$ that we require to be minimum. In this section we will minimise the trace of $C_{x'}$, i.e. we will seek an estimate for which the sum (and hence the mean) of the variances of the parameters is a minimum. In 6.3 a different property of $C_{x'}$ will be considered.

The argument here will be restricted to linear unbiased estimates, i.e.

$$x' = Qb \tag{6.13}$$

If x' is unbiased

$$E(x') = x \tag{6.14}$$

and, from (6.13),

$$E(x') = QE(b) \tag{6.15}$$

Substituting (6.9) in (6.15) gives

$$E(x') = QAx \tag{6.16}$$

and combining (6.14) and (6.16) leads to

$$QA = I \tag{6.17}$$

Now applying (4.16) to (6.13) gives

$$C_{x'} = QC_bQ^T \tag{6.18}$$

Hence the problem is to find a linear transformation, Q, which satisfies (6.17) whilst minimising the trace of QC_bQ^T, i.e. we require

$$Tr(QC_bQ^T) = \text{minimum} \tag{6.19}$$

Subject to

$$QA - I = 0 \tag{6.20}$$

We use Lagrange's method of undetermined multipliers given in (3.12) and write the complete Lagrangian function using the trace notation

$$\Phi = Tr(QC_bQ^T) + Tr\left[2(QA - I)K\right]$$

or

$$\Phi = Tr(QC_bQ^T) + 2\,Tr(QAK) - 2\,Tr\,(IK) \tag{6.21}$$

where K is a matrix of correlatives. To minimise we differentiate with respect to Q and equate to zero, i.e.

$$\frac{\partial\Phi}{\partial Q} = 0 \tag{6.22}$$

which leads to

$$2QC_b + 2\,K^TA^T - 0 = 0 \tag{6.23}$$

giving

$$QC_b = -K^TA^T \tag{6.24}$$

and

$$Q = -K^TA^TC_b^{-1} \tag{6.25}$$

Postmultiplying (6.25) by A gives

$$QA = -K^TA^TC_b^{-1}A \tag{6.26}$$

and substituting (6.26) into (6.17) gives

$$I = -K^TA^TC_b^{-1}A \tag{6.27}$$

which re-arranges to

$$K^T = -(A^TC_b^{-1}A)^{-1} \tag{6.28}$$

Substituting (6.28) into (6.25) gives

$$Q = (A^TC_b^{-1}A)^{-1}A^TC_b^{-1} \tag{6.29}$$

Hence, from (6.13), the estimate of x which leads to a covariance matrix of parameters with a minimum trace is

$$x' = (A^TC_b^{-1}A)^{-1}A^TC_b^{-1}b \tag{6.30}$$

The least squares estimate from (3.41) is

$$\hat{x} = (A^T W A)^{-1} A^T W b \qquad (3.41)$$

and it is obvious from (6.30) and (3.41) that the least squares estimate is the minimum trace estimate so long as we put $W = C_b^{-1}$. The basic definition of least squares actually puts $W = C_\ell^{-1}$, but it is clear from (4.38) that for the special case of observation equations $C_b = C_\ell$. Hence it has been proved that the least squares process, as defined by (3.6) and (3.7), yields an estimate with a covariance matrix that has a smaller trace (i.e. smaller sum of variances) than any other linear unbiased estimate.

6.3 Minimum variance of derived quantities

In 4.4.3 the computation of the variances and covariances of a set of quantities $\hat{q}$ derived from the least squares estimates of the parameters was shown and it was explained that such variances and covariances were extremely useful in practice as they are closely related to the probable use of the parameters. Here we consider a single quantity ψ whose least squares estimate is given by

$$\hat{\psi} = s^T \hat{x} \qquad (6.31)$$

where

$$s^T = \left[s_1, s_2, \ldots, s_m \right]$$

and it will be shown that the variance of ψ, σ^2, is less than the variance, σ'^2, of ψ' computed from any other linear unbiased estimate of x, x'. What follows is an extension of the proof given in Sunter (1966) for the special case of $W = I$.

Let $$x' = x + \delta x \qquad (6.32)$$

then

$$\psi' = s^T x' = s^T x + s^T \delta x \qquad (6.33)$$

Also, if x' is a linear estimate of x we can write it in the form of a linear transformation of b

$$x' = \left[((A^T W A)^{-1} A^T + U) W \right] b \qquad (6.34)$$

i.e. it is the least squares transformation of b

$$\hat{\underset{\sim}{x}} = \left[(A^T W A)^{-1} A^T W\right] b \tag{6.35}$$

plus an additional amount. Combining (6.34) and (6.35) gives

$$x' = \hat{\underset{\sim}{x}} + UWb \tag{6.36}$$

which when substituted into (6.32) gives

$$\delta x = UWb \tag{6.37}$$

Substituting (6.37) in (6.33) results in

$$\psi' = s^T \hat{\underset{\sim}{x}} + s^T UWb \tag{6.38}$$

and putting

$$t^T = s^T U \tag{6.39}$$

gives

$$\psi' = s^T \hat{\underset{\sim}{x}} + t^T Wb \tag{6.40}$$

Taking expectations yields

$$E(\psi') = s^T E(\hat{\underset{\sim}{x}}) + t^T W E(b) \tag{6.41}$$

Substituting (6.12) and (6.9) in (6.41) gives

$$E(\psi') = s^T x + t^T WAx \tag{6.42}$$

but since ψ' is an unbiased estimate of ψ (because x' is defined as an unbiased estimate of x) we have

$$E(\psi') = \psi \tag{6.43}$$

which, on using (6.31) with true values, becomes

$$E(\psi') = s^T x \tag{6.44}$$

Then it is evident from (6.42) and (6.44) that

$$t^T WAx = 0 \tag{6.45}$$

Also if (6.45) is to be true for all x we can deduce that

$$t^T W A = 0 \tag{6.46}$$

and

$$A^T W t = 0 \tag{6.47}$$

We now substitute (6.35) into (6.40) to yield

$$\psi' = \left[s^T (A^T W A)^{-1} A^T W + t^T W \right] b \tag{6.48}$$

and applying (4.16) to (6.48) gives

$$\sigma'^2 = \left[s^T (A^T W A)^{-1} A^T W + t^T W \right] C_b \left[s^T (A^T W A)^{-1} A^T W + t^T W \right]^T \tag{6.49}$$

Since, for observation equations, $C_b = C_\ell = W^{-1}$ (from (4.38)), multiplying out (6.49) gives

$$\begin{aligned} \sigma'^2 = {} & s^T (A^T W A)^{-1} A^T W W^{-1} W A (A^T W A)^{-1} s + t^T W W^{-1} W A (A^T W A)^{-1} s \\ & + s^T (A^T W A)^{-1} A^T W W^{-1} W t + t^T W W^{-1} W t \end{aligned} \tag{6.50}$$

Using (6.46) and (6.47) we see that the middle two terms are zero giving

$$\sigma'^2 = s^T (A^T W A)^{-1} s + t^T W t \tag{6.51}$$

which, from (4.67), reduces to

$$\sigma'^2 = s^T C_{\hat{x}} s + t^T W t \tag{6.52}$$

Now applying (4.16) to (6.31) gives

$$\hat{\sigma}^2 = s^T C_{\hat{x}} s \tag{6.53}$$

so from (6.52) and (6.53) we have

$$\sigma'^2 = \hat{\sigma}^2 + t^T W t \tag{6.54}$$

Now assuming W to be positive-definitive (which it will always be, because it is obtained by inverting the positive-definite matrix C_ℓ), $t^T W t$ must be a positive number, irrespective of t, and it follows that σ'^2 must be greater than $\hat{\sigma}^2$.

Hence if we compute, say, a distance from the least squares estimates of some

coordinates, then that distance will have a smaller variance than a similar distance computed from any other linear unbiased estimate of the coordinates. The general theorem that has been proved here is known as the Gauss-Markov theorem and it represents the single most important justification for adopting the least squares procedure.

6.4 Maximum likelihood

An estimate is said to be the maximum likelihood estimate when it maximises the value of the probability density function of the observational errors. If we assume that the true errors (and hence the true residuals) are from a multivariate normal distribution, their pdf can be written as in (5.9) as

$$\Phi = \text{constant} \times \exp\left\{-\tfrac{1}{2}\, v^T C_\ell^{-1} v\right\} \qquad (6.55)$$

Clearly (6.55) will be maximised if $v^T C_\ell^{-1} v$ is minimised. Hence, so long as the observational errors have a normal distribution, we can say that the least squares estimate is the maximum likelihood estimate.

The exact meaning of the term maximum likelihood is complex and its full explanation is beyond the scope of this paper: interested readers are recommended to study Mood and Graybill (1963, chapter 8) for a detailed discussion of the matter or Thompson (1969, chapter 10) for a simpler, but less rigorous, discussion. Here we merely mention that the term is almost equivalent to "most probable" but it is applied to estimated parameters rather than to the observed quantities.

6.5 Summary and historical background

It has been shown in 6.1 to 6.3 that, no matter what the pdf of the observational errors, in particular irrespective of whether or not they are normally distributed, the least squares estimate has the following two properties:

(i) it is unbiased

(ii) it has a minimum variance (in the sense that $\mathrm{Tr}(C_{\hat{x}})$ is a minimum and that the variance of a derived quantity is a minimum).

As a result of these properties, the least squares estimate is often referred to as the BLUE or Best (because of minimum variance) Linear Unbiased Estimate.

Also it has been shown, in 6.4, that if the observational errors are normally

distributed than the least squares estimate has the additional property of being the maximum likelihood estimate.

There seems to have been some confusion within the surveying profession as to the proper statistical justification of the least squares procedure, especially with regard to whether or not observations need to be normally distributed. For instance Rainsford (1957) states,

"Once the normal law of error has been accepted, all other results, such as the principle of least squares ... follow logically from it".

And Gale (1965) states,

"The least squares principle is derived from the normal distribution function ...".

Although neither of these statements is wrong they are over-restrictive and have been interpreted by many as implying that a normal pdf must exist before least squares should be used. To help explain how this confusion arose, and because it is interesting for its own sake, a brief history of the development of least squares will now be given.

The first published description of the use of least squares was by Legendre (1806), who used the method for the orbit determination of comets. Three years later Gauss (1809) wrote that "our principle (least squares with a diagonal weight matrix) which we have made use of since 1795 has lately been published by Legendre" and a bitter personal feud started. It is now generally accepted that Gauss was the first to use least squares. The first theoretical analysis of the method was by Laplace (1812), who essentially showed that least squares estimates were what we should call maximum likelihood estimates. Actually Laplace used the term "most probable" and justified least squares so long as the observations were independent (diagonal W) and normally distributed. His proof basically involved maximising the probability (i.e. area under a normal pdf) of estimates lying within certain limits.

A different approach was taken by Gauss (1821), who was the first to show the minimum variance property of the least squares solution (although his proof was restricted to independent observations). Gauss was therefore the first to justify the use of least squares without recourse to the normal distribution. His work was largely ignored during the nineteenth century when Todhunter (1865), Merriman (1877) and others gave the impression that Laplace was wholly responsible for giving the method a statistical basis.

In the early part of the twentieth century Markov, e.g. in Markov (1912), wrote extensively on Gauss's ideas and drew attention to the importance of his work. Although Markov added nothing new in this area we today refer to the minimum variance property (in 6.3) as the Gauss-Markov theorem. Aitken (1934), using matrix algebra notation, extended the Gauss-Markov theorem to the case where W is a full matrix (i.e. correlated observations).

6.6 The practical argument

We have seen that there is a theoretical justification for the application of the least squares process to virtually any set of measurements (the only assumptions implicit in the foregoing were that the pdf should be continuous and W should be positive-definite, both of which will always be true in practice). There are also a number of more practical reasons why surveyors use least squares and it must be said that it is for these reasons that most least squares computations are carried out, i.e. surveyors would probably use least squares even if it could not be justified from a statistical point of view.

(i) The method is extremely easy to apply because it yields a linear set of normal equations (unlike, say, "least cubes").

(ii) It is unique, i.e. there is only one solution to a given problem. Some of the so-called semi-rigorous methods (e.g. semi-graphic resection) yield a large number of solutions depending on the subjective choice of the surveyor carrying out the computation.

(iii) It is, generally speaking, "unobjectionable": it is very difficult to form an argument against least squares and in favour of some other procedure.

(iv) The method leads to an easy quantitative assessment of quality, e.g. via the covariance matrix $C_{\hat{x}}$.

(v) It is a general method that can be applied to any problem.

To counter these apparently overwhelming arguments, the following warnings should be given.

(i) Least squares does not give the true solution. On average (because of the unbiased property) the least squares solution is the true solution, but for any individual problem there may be some other technique that

will give (perhaps by chance) an answer closer to the truth. Of course, we are very unlikely to know this.

(ii) If either C_ℓ or the basic model (2.3) do not truly reflect the physical situation, none of the derived properties of the least squares estimate will be valid. There are often considerable practical problems with determining both of these.

(iii) Although the normal distribution is not required for the application of least squares, it is essential for the application of most of the statistical tests described in section 5.

7. Sequential and step-by-step methods

There are many practical problems for which it is convenient to divide a least squares computation into parts. Two classes of such problems can be identified as follows:

(i) situations in which not all of the observed data is available at any particular time and it is required to estimate all of the parameters with a view to updating these estimates as new data becomes available

(ii) situations in which it is required to split the parameters into groups and to estimate either only a limited number of parameters or different groups of parameters at different times.

Methods for the two classes of problem are referred to as "sequential" and "step-by-step" respectively. Note that we require that the results from both non-simultaneous approaches to be the same as would be obtained from a full simultaneous least squares computation (as described in section 3). The reason for dividing the problem is purely one of computational convenience (usually to reduce both the storage and execution time required for solution by a computer).

In this section details of one method for each of the two classes of problems will be given. The approach adopted is to begin by quoting a number of examples of problems in which the situation occurs and then to give a full mathematical derivation, followed by the working formulae for the relevant method. In the case of sequential methods there really is only one method, called sequential least squares, and this will be given. There are, however, several alternative step-by-step methods and the one chosen for inclusion here is known as the Helmert-Wolf method. Derivations will be restricted to the special case of observation equations. Gagnon (1976) describes other non-simultaneous methods which fall into the classes identified above.

7.1 Sequential least squares

7.1.1 Examples of applications

(i) Real time data acquisition There are a number of problems in which data is acquired in real time and a continuous estimate of a number of stationary parameters is required. Note that if new parameters are being "generated" by the new data we must use a step-by-step method and if the parameters are changing with time we need to use a filtering technique, e.g. Kalman filtering as discussed in section 8. An example of a problem suitable for the

sequential least squares approach is the fixing of a stationary oil rig by continuous interrogation of a number of different navigation systems.

(ii) Coordinate updates A national mapping organization, or any other body concerned with the provision of coordinates, is, in general, continually undertaking new measurements. Rather than carrying out a completely new least squares computation involving all the original data, these organizations may prefer sequential least squares, which allows an update of the coordinates (and the estimates of their precision) by computing small "corrections" from the new data.

(iii) Design of position-fixes The computer simulation method of designing position-fixes (described in 4.5.1) involves continuously altering a proposed position-fixing configuration to search for the one which fulfils the design criteria most economically. After each alteration a new covariance matrix $C_{\hat{x}}$ is required. The use of the sequential least squares method allows this to be determined by a correction to the previous $C_{\hat{x}}$ and without a completely new computation and inversion.

7.1.2 Derivation of the method

Suppose we have a vector of observations, ℓ_1, from which we wish to estimate some parameters, x. We set up the linearised model in the form

$$A_1 x = b_1 + v_1 \tag{7.1}$$

and determine the least squares estimate of x, which we will call $\hat{x}_1$, from (3.41) by

$$\hat{x}_1 = (A_1^T W_1 A_1)^{-1} A_1^T W_1 b_1 \tag{7.2}$$

where $W_1 = C_{b_1}^{-1}$. We find it convenient to write (7.2) as

$$\hat{x}_1 = N_1^{-1} u_1 \tag{7.3}$$

where

$$N_1 = A_1^T W_1 A_1 \tag{7.4}$$

and

$$u_1 = A_1^T W_1 b_1 \tag{7.5}$$

Also we have, from (4.67),

$$C_{\hat{x}_1} = N_1^{-1} \tag{7.6}$$

Now suppose that we have some additional observations ℓ_2 and we wish to obtain an estimate of x, which we will call $\hat{x}_2$, from <u>both</u> ℓ_1 and ℓ_2. We proceed as follows.

Let the linearised model for the second set of observations be

$$A_2 x = b_2 + v_2 \tag{7.7}$$

with a new weight matrix $W_2 = C_{b_2}^{-1}$.

Then combining (7.1) and (7.7) we obtain the complete model

$$Ax_2 = b + v \tag{7.8}$$

where

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

also we will have

$$W = \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix}$$

and for the correlatives

$$k = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$$

Now, from (3.15), remembering that we are dealing with the special case of observation equations, we write the least squares solution to (7.8) as

$$\begin{bmatrix} W & -I & 0 \\ -I & 0 & A \\ 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{v} \\ \hat{k} \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix} \tag{7.9}$$

Then, partitioning (7.9) further we obtain

$$\left[\begin{array}{cc|cc|c} W_1 & 0 & -I & 0 & 0 \\ 0 & W_2 & 0 & -I & 0 \\ \hline -I & 0 & 0 & 0 & A_1 \\ 0 & -I & 0 & 0 & A_2 \\ \hline 0 & 0 & A_1^T & A_2^T & 0 \end{array}\right] \left[\begin{array}{c} \hat{v}_1 \\ \hat{v}_2 \\ \hline \hat{k}_1 \\ \hat{k}_2 \\ \hline \hat{x}_2 \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ \hline b_1 \\ b_2 \\ \hline 0 \end{array}\right] \qquad (7.10)$$

which can be rearranged to

$$\begin{bmatrix} W_1 & 0 & -I & 0 & 0 \\ 0 & W_2 & 0 & 0 & -I \\ -I & 0 & 0 & A_1 & 0 \\ 0 & 0 & A_1^T & 0 & A_2^T \\ 0 & -I & 0 & A_2 & 0 \end{bmatrix} \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \hat{k}_1 \\ \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ b_1 \\ 0 \\ b_2 \end{bmatrix} \qquad (7.11)$$

Now applying (3.21) to eliminate $\hat{v}_1$ in (7.11) gives

$$\begin{bmatrix} W_2 & 0 & 0 & -I \\ 0 & -W_1^{-1} & A_1 & 0 \\ 0 & A_1^T & 0 & A_2^T \\ -I & 0 & A_2 & 0 \end{bmatrix} \begin{bmatrix} \hat{v}_2 \\ \hat{k}_1 \\ \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ b_1 \\ 0 \\ b_2 \end{bmatrix} \qquad (7.12)$$

Similarly eliminating $\hat{v}_2$ in (7.12) gives

$$\begin{bmatrix} -W_1^{-1} & A_1 & 0 \\ A_1^T & 0 & A_2^T \\ 0 & A_2 & -W_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{k}_1 \\ \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ 0 \\ b_2 \end{bmatrix} \qquad (7.13)$$

and eliminating $\hat{k}_1$ in (7.13), and using (7.4) and (7.5), yields

$$\begin{bmatrix} N_1 & A_2^T \\ A_2 & -W_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} u_1 \\ b_2 \end{bmatrix} \tag{7.14}$$

We now apply (3.21) directly to (7.14) to give

$$-(W_2^{-1} + A_2 N_1^{-1} A_2^T)\hat{k}_2 = b_2 - A_2 N_1^{-1} u_1 \tag{7.15}$$

Substituting (7.3) in (7.15) and rearranging yields

$$\hat{k}_2 = (W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1}(A_2 \hat{x}_1 - b_2) \tag{7.16}$$

From (7.14) we can write

$$N_1 \hat{x}_2 + A_2^T \hat{k}_2 = u_1 \tag{7.17}$$

hence

$$\hat{x}_2 = N_1^{-1}(u_1 - A_2^T \hat{k}_2) \tag{7.18}$$

and substituting (7.3) and (7.16) in (7.18) gives

$$\hat{x}_2 = \hat{x}_1 - N_1^{-1} A_2^T (W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1}(A_2 \hat{x}_1 + b_2) \tag{7.19}$$

which is the sequential least squares expression for the parameters. To determine the residuals for the additional observations we return to (7.10) from which we can write

$$W_2 \hat{v}_2 - \hat{k}_2 = 0 \tag{7.20}$$

giving

$$\hat{v}_2 = W_2^{-1} \hat{k}_2 \tag{7.21}$$

and substituting (7.16) in (7.21) gives

$$\hat{v}_2 = W_2^{-1}(W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1}(A_2 \hat{x}_1 - b_2) \tag{7.22}$$

To determine the covariance matrix $C_{\hat{x}_2}$ we rearrange (7.19) as follows

$$\hat{x}_2 = \left[I - N_1^{-1} A_2^T (W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1} A_2\right] \hat{x}_1 + N_1^{-1} A_2^T (W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1} b_2 \tag{7.23}$$

which can be written

$$\hat{x}_2 = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ b_2 \end{bmatrix} \tag{7.24}$$

where

$$Q_1 = I - N_1^{-1} A_2^T (W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1} A_2 \tag{7.25}$$

and

$$Q_2 = N_1^{-1} A_2^T (W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1} \tag{7.26}$$

Now applying the propagation of error law, (4.16), to (7.24) gives

$$C_{\hat{x}_2} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} C_{\hat{x}_1} & 0 \\ 0 & C_{b_2} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} \tag{7.27}$$

Putting $W_2^{-1} = C_{b_2}$ and $N_1^{-1} = C_{\hat{x}_1}$ in (7.27) and multiplying out gives

$$C_{\hat{x}_2} = Q_1 N_1^{-1} Q_1^T + Q_2 W_2^{-1} Q_2^T \tag{7.28}$$

Then substituting (7.25) and (7.26) in (7.28) with

$$P = (W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1} \tag{7.29}$$

gives

$$\begin{aligned} C_{\hat{x}_2} = {} & (I - N_1^{-1} A_2^T P A_2) N_1^{-1} (I - A_2^T P A_2 N_1^{-1}) \\ & + N_1^{-1} A_2^T P W_2^{-1} P A_2 N_1^{-1} \end{aligned} \tag{7.30}$$

which multiplies out to give

$$\begin{aligned} C_{\hat{x}_2} = {} & N_1^{-1} - N_1^{-1} A_2^T P A_2 N_1^{-1} - N_1^{-1} A_2^T P A_2 N_1^{-1} \\ & + N_1^{-1} A_2^T P A_2 N_1^{-1} A_2^T P A_2 N_1^{-1} + N_1^{-1} A_2^T P W_2^{-1} P A_2 N_1^{-1} \end{aligned} \tag{7.31}$$

The last two terms of (7.31) combine to give

$$N_1^{-1} A_2^T P (A_2 N_1^{-1} A_2^T + W_2^{-1}) P A_2 N_1^{-1} \tag{7.32}$$

But from (7.29) we see that (7.32) is equal to

$$N_1^{-1} A_2^T P A_2 N_1^{-1} \tag{7.33}$$

Hence the sum of the last three terms of (7.31) is zero and we can write, using (7.6) and (7.29)

$$C_{\hat{x}_2} = C_{\hat{x}_1} - N_1^{-1} A_2^T (w_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1} A_2 N_1^{-1} \tag{7.34}$$

Note that, using (7.6), we can write

$$N_2^{-1} = C_{\hat{x}_2} \tag{7.35}$$

7.1.3 Summary of the sequential procedure

We start with a set of measurements ℓ_1 and obtain our first least squares estimate of x ($\hat{x}_1$) and its covariance matrix ($C_{\hat{x}_1}$) in the usual way using (7.3) and (7.6); also we can estimate the residuals ($\hat{v}_1$) from (3.43). If a new set of measurements ℓ_2 becomes available we can obtain a new estimate of x and its covariance matrix from both sets of measurements by using

$$\hat{x}_2 = \hat{x}_1 + \delta\hat{x} \tag{7.36}$$

$$C_{\hat{x}_2} = C_{\hat{x}_1} + \Delta C_{\hat{x}} \tag{7.37}$$

where $\delta\hat{x}$ and $\Delta C_{\hat{x}}$ are given in (7.19) and (7.34). The residuals for the new observations are given by (7.22), or if required the residuals for the complete set of observations ℓ_1 and ℓ_2 can be obtained from

$$\hat{v} = A\hat{x}_2 - b \tag{7.38}$$

Should a third set of measurements become available we can obtain $\hat{x}_3$ etc. from $\hat{x}_2$ etc. by using exactly the same set of formulae but putting the subscripts 3 and 2 instead of 2 and 1 and so on for fourth, fifth ... sets of measurements.

The general form of the equations, written in a way that indicates how the computations would be carried out in practice, is

$$R_i = N_{i-1}^{-1} A_i^T \tag{7.39}$$

$$S_i = (w_i^{-1} + A_i R_i)^{-1} \tag{7.40}$$

$$\hat{k}_i = S_i (A_i \hat{x}_{i-1} - b_i) \tag{7.41}$$

$$\hat{x}_i = \hat{x}_{i-1} - R_i \hat{k}_i \tag{7.42}$$

$$\hat{v}_i = W_i^{-1} \hat{k}_i \tag{7.43}$$

$$C_{\hat{x}_i} = N_i^{-1} = C_{\hat{x}_{i-1}} - R_i S_i R_i^T \tag{7.44}$$

Notice that each computational cycle includes only one inversion, to get S_i in (7.40). No inversion is required to compute R_i in (7.39) because N_{i-1}^{-1} is given by (7.44) from the previous cycle. The size of the matrix to be inverted in (7.40) is the number of new observations. Hence it can be seen that the method will be at its most efficient when only a few (compared with the number of parameters) measurements are added during each cycle. If the number of new measurements approaches the number of parameters it is more efficient to start afresh and carry out a new simultaneous least squares computation in the usual way. Mikhail (1976) and Forster (1980) include detailed discussions of the efficiency of the application of the sequential least squares method and several numerical examples of its use.

It should be noted that there may be situations in which it is required to remove rather than add observations, e.g. when designing position-fixes or after discovering a gross error amongst the latest observations to be added. In such cases the procedure and formulae are exactly the same except for a change of the signs in (7.40), (7.41), (7.42) and (7.44). Finally, it is remarked that a similar procedure can be derived for the combined case of least squares.

7.2 The Helmert-Wolf method

It was explained at the beginning of this section that the Helmert-Wolf method can be used for problems where it is convenient to split the parameters into two or more groups. 7.2.1 gives three examples of such problems whilst 7.2.2 is concerned with the Helmert-Wolf method itself. Further reference to the examples is made in 7.2.2 and 7.2.3.

7.2.1 Example of applications

(i) Calibration of acoustic networks Suppose that there are a number of acoustic beacons, A, B, C, ..., etc. as in Fig. 7.1 placed on the seabed and that it is required to find their positions so that they may be used at a later date to position ships in the area. A ship may be sailed through the area, as

shown in Fig. 7.1, and at points 1, 2, 3, ..., etc. it may measure distances to whichever beacons are within range and the problem would be to compute the

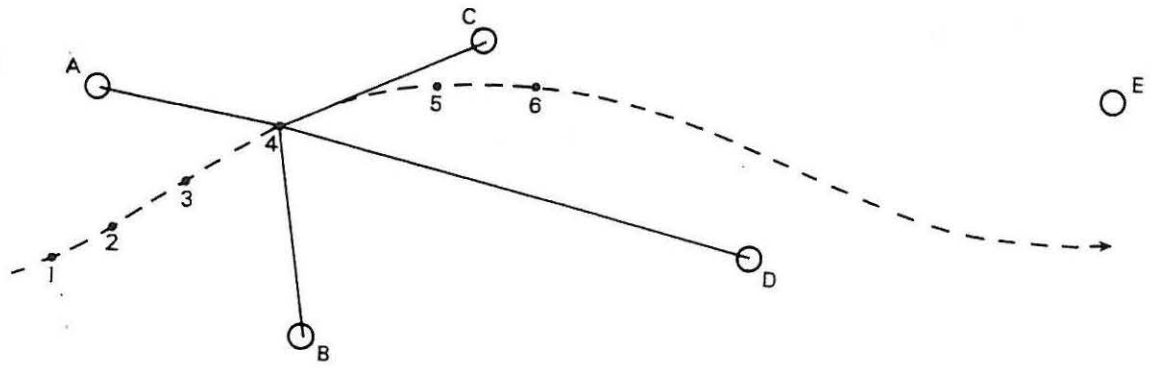


Fig. 7.1

coordinates of points A, B, C, ..., etc. The coordinates of points 1, 2, 3, ..., etc. are unlikely to be needed (although as will be seen later they can be determined if required).

Obviously one solution would be to write down one equation for each distance and solve the problem simultaneously as if it were a geodetic network. The difficulty with this is that there may be hundreds (or even thousands) of ship positions and the least squares computation would involve solving normal equations of very large dimensions, possibly larger than could be handled conveniently by the available computer. It would clearly be advantageous to be able to divide the parameters into two groups - ship positions and beacon positions - and to be able to limit the solution to the beacon positions only.

Note that the problem as described here is often referred to as relative calibration, as one of the beacons would have to be assigned arbitrary coordinates and the result would consist of relative beacon positions. Absolute calibration (i.e. determination of positions in a particular reference system) could be performed by the same method if at some of the ship positions, 1, 2, 3, ..., etc. in Fig. 7.1, measurements were made to points (e.g. shore stations or satellites) whose positions were known in the required reference system.

(ii) Satellite-Doppler positioning When fixing the position of a stationary point using the satellite-Doppler method it is common practice to observe a large number (say fifty) satellite passes. For each pass the observation equations will contain a number of unknowns - three for the position of the

point and a variable number for the pass (e.g. for orbit errors, refraction, frequency offset and frequency drift). The number in the latter group depends on the mathematical model used for the solution. For example, if we had ten pass unknowns (which would be different for every pass), then, for fifty passes, we would have a system of 50 x 10 + 3 = 503 equations to solve if all the data were treated simultaneously.

There are obvious advantages in dividing the parameters into two groups, pass unknowns and position unknowns, as we often do not need to know the former.

(iii) Computation of national and continental networks

In modern geodesy it is fashionable to compute simultaneously very large triangulation networks, e.g. the European network has recently been completed, the North American is currently (1982) being computed and an African network is at the early planning stage. There are several reasons why such networks may not conveniently be handled in one simultaneous computation; some are listed below.

(a) The set of equations may be too large to handle conventionally.

(b) Various countries may be reluctant to make their observations generally available.

(c) Relative weights between areas of triangulation may be difficult to assess.

(d) There may be problems over the publication of coordinates, i.e. a country may not wish others to know coordinates of its stations.

It will be seen that splitting the network into areas (countries or groups of countries in the case of continental networks) and introducing two kinds of position unknowns, those for points within the area and those for points on the boundaries (called internal and junction unknowns respectively), can overcome all of the above difficulties. A central bureau is established to compute the junction point coordinates and then each area can compute its own coordinates independently.

7.2.2 Derivation of the method

Suppose that we have r sets of independent observation equations and that in each set we divide the parameters into two groups which we call local and common parameters. The common parameters, y, will appear in every set whereas

the local parameters, x_i for the ith set of observations, appear only in that one set. We could then partition the ith set of observation equations in the form

$$\left[\begin{array}{c|c} A_i & B_i \end{array}\right] \left[\begin{array}{c} x_i \\ \hline y \end{array}\right] = \left[b_i \right] + \left[v_i \right] \qquad (7.45)$$

i.e. $$A_i x_i + B_i y = b_i + v_i \qquad (7.46)$$

To explain this further we return to the three examples in 7.2.1.

(i) Each set of observations is the group of distance measurements from one ship position. The local parameters are the position of the ship and the common parameters are the positions of the beacons.

(ii) Each set of observations is a single satellite pass, the pass unknowns are the local parameters and the coordinates of the observer's position are the common parameters.

(iii) Each set of observations is an area of triangulation, the interior point positions are the local parameters and the junction point positions are the common parameters.

It is important to note the following two points.

(i) The unknowns in each set of observations must be carefully arranged as in (7.45), i.e. the common parameters must appear after the local parameters in the complete vector of parameters.

(ii) The vector y contains all the common parameters, even though any one set of measurements may only be concerned with a limited number of them. This point is made merely to ensure the rigour of the derivation to follow. In practice it is only necessary to form sets of observation equations using the common parameters with which those observations are concerned. The missing parameters are merely filled with zeroes at the appropriate stage. This will become clear in 7.2.3.

Now if we let W_i $(= C_{b_i}^{-1})$ be the weight matrix for the ith set of observations and combine together all r sets of measurements we will have a block diagonal global weight matrix (because the sets are independent of each other).

$$W = \begin{bmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_r \end{bmatrix}$$

and a global vector of residuals

$$v = \begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix}^T$$

making the least squares requirement

$$v^T W v = v_1^T W_1 v_1 + v_2^T W_2 v_2 + \cdots + v_r^T W_r v_r \quad \text{is a minimum}$$

i.e.
$$\sum_{i=1}^{r} v_i^T W_i v_i = \text{minimum} \qquad (7.47)$$

Substituting (7.46) in (7.47) gives

$$\sum_{i=1}^{r} (A_i x_i + B_i y - b_i)^T W_i (A_i x_i + B_i y - b_i) = \text{minimum}$$

and expanding gives

$$\sum_{i=1}^{r} (x_i^T A_i^T W_i A_i x_i + x_i^T A_i^T W_i B_i y - x_i^T A_i^T W_i b_i$$

$$+ y^T B_i^T W_i A_i x_i + y^T B_i^T W_i B_i y - y^T B_i^T W_i b_i$$

$$- b_i^T W_i A_i x_i - b_i^T W_i B_i y + b_i^T W b_i) = \text{minimum} \qquad (7.48)$$

For a minimum the differentials of (7.48) with respect to x_i and y must be zero. Differentiating first with respect to x_i, and again using $\hat{x}_i$ and $\hat{y}$ to denote the least squares estimates of x_i and y gives

$$\sum_{i=1}^{r} (2A_i^T W_i A_i \hat{x}_i + A_i^T W_i B_i \hat{y} - A_i^T W_i b_i$$

$$+ A_i^T W_i B_i \hat{y} - A_i^T W_i b_i) = 0$$

which simplifies to

$$\sum_{i=1}^{r} (A_i^T W_i A_i \hat{x}_i + A_i^T W_i B_i \hat{y} - A_i^T W_i b_i) = 0 \qquad (7.49)$$

Similarly differentiating with respect to y and simplifying leads to

$$\sum_{i=1}^{r} (B_i^T W_i A_i \hat{x}_i + B_i^T W_i B_i \hat{y} - B_i^T W_i b_i) = 0 \qquad (7.50)$$

(7.49) and (7.50) can be combined to give

$$\sum_{i=1}^{r} \left\{ \begin{bmatrix} A_i^T W_i A_i & A_i^T W_i B_i \\ B_i^T W_i A_i & B_i^T W_i B_i \end{bmatrix} \begin{bmatrix} \hat{x}_i \\ \hat{y} \end{bmatrix} - \begin{bmatrix} A_i^T W_i b_i \\ B_i^T W_i b_i \end{bmatrix} \right\} = 0 \qquad (7.51)$$

Then applying (3.21) to (7.51) gives

$$\left\{ \sum_{i=1}^{r} (B_i^T W_i B_i - B_i^T W_i A_i (A_i^T W_i A_i)^{-1} A_i^T W_i B_i) \right\} \hat{y}$$

$$= \sum_{i=1}^{r} (B_i^T W_i b_i - B_i^T W_i A_i (A_i^T W_i A_i)^{-1} A_i^T W_i b_i) \qquad (7.52)$$

Hence we have derived a set of equations which is in terms of $\hat{y}$ only. We can write (7.52) in the form

$$(N_1 + N_2 + \ldots + N_r)\hat{y} = d_1 + d_2 + d_3 + \ldots + d_r \qquad (7.53)$$

or

$$N\hat{y} = d \qquad (7.54)$$

where

$$N_i = B_i^T W_i B_i - B_i^T W_i A_i (A_i^T W_i A_i)^{-1} A_i^T W_i B_i$$

$$d_i = B_i^T W_i b_i - B_i^T W_i A_i (A_i^T W_i A_i)^{-1} A_i^T W_i b_i$$

The set of equations

$$N_i \hat{y} = d_i \qquad (7.55)$$

is sometimes referred to as the "reduced normal equations" for the ith observation set. This is because they are obtained simply by independently

forming the normal equations for the ith observation set ((7.51) without the summation sign) and eliminating the local parameters, $\hat{x}_i$, according to the procedure of (3.21). Hence the procedure is to compute the reduced normal equations for each set of measurements and then simply to add them together, as in (7.53), to form a set of equations, (7.54), that can be solved for $\hat{y}$, the common parameters. If the values of any of the local parameters are required we return to the normal equations for the relevant set of observations and obtain $\hat{x}_i$ using (3.20), i.e.

$$\hat{x}_i = (A_i^T W_i A_i)^{-1}(A_i^T W_i b_i - A_i^T W_i B_i \hat{y}) \qquad (7.56)$$

Note that it is at the summation stage, (7.53), that we need to ensure that all matrices N_i have the same dimensions. If some of the common parameters do not appear in any particular set of measurements, the N_i for that set will have to be filled with zeroes in the appropriate places. This is not problematical, although great care must be taken when carrying out, say, a continental adjustment to ensure that all countries adopt the same code numbers for the common parameters.

To obtain the least squares estimates of the residuals we return to each of the i sets of measurements, solve for the local parameters using (7.56), and use (7.46) in the following form

$$\hat{v}_i = A_i \hat{x}_i + B_i \hat{y} - b_i \qquad (7.57)$$

In some problems the local parameters and residuals may not be required but it may be desired to compute the variance factor from (4.25), which requires computation of the global quadratic form $\hat{v}^T W \hat{v}$. It is shown in Gagnon (1976,45) that this can be determined without explicitly computing the residuals by using

$$\hat{v}^T W \hat{v} = \hat{y}^T d + \sum_{i=1}^{r} (b_i^T W_i b_i - b_i^T W_i A_i (A_i^T W_i A_i)^{-1} A_i^T W_i b_i) \qquad (7.58)$$

where d is given in (7.54).

The covariance matrix of the common parameters, $C_{\hat{y}}$, is obtained in the usual way

$$C_{\hat{y}} = N^{-1} \qquad (7.59)$$

where N is defined in (7.54). To determine the covariance matrix of the local

parameters of a particular set of observations we return to the normal equations for that set ((7.51) without the summation sign) and write them in the form

$$\begin{bmatrix} N_x & N_{xy} \\ N_{xy}^T & N_y \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \tag{7.60}$$

where

$$N_x = A_i^T W_i A_i$$

$$N_{xy} = A_i^T W_i B_i$$

$$N_y = B_i^T W_i B_i$$

$$p_1 = A_i^T W_i b_i$$

$$p_2 = B_i^T W_i b_i$$

$$\hat{x} = \hat{x}_i$$

The suffix i will be dropped from now on to increase clarity.

Now applying (3.20) to (7.60) gives

$$\hat{x} = N_x^{-1}(p_1 - N_{xy}\hat{y})$$

which expands to

$$\hat{x} = N_x^{-1}p_1 - N_x^{-1}N_{xy}\hat{y}$$

and substituting for p_1 gives

$$\hat{x} = N_x^{-1}A^T W b - N_x^{-1}N_{xy}\hat{y} \tag{7.61}$$

then, applying (4.16) to (7.61) and assuming b and $\hat{y}$ to be independent (which is not true but is a reasonable simplification)

$$C_{\hat{x}} = N_x^{-1}A^T W C_b W A N_x^{-1} + N_x^{-1}N_{xy}C_{\hat{y}}N_{xy}^T N_x^{-1}$$

but $W = C_b^{-1}$ and $N_x = A^T W A$

$$\therefore \quad C_{\hat{x}} = N_x^{-1} + N_x^{-1}N_{xy}C_{\hat{y}}N_{xy}^T N_x^{-1} \tag{7.62}$$

which is the expression for the covariance matrix of a set of local parameters. Writing it in full for the ith set of local parameters we have the approximate formula

$$C_{\hat{x}} = (A_i^T W_i A_i)^{-1}\left[I + A_i^T W_i B_i C_{\hat{y}} B_i^T W_i A_i (A_i^T W_i A_i)^{-1}\right] \qquad (7.63)$$

7.2.3 Summary of the Helmert-Wolf procedure

The Helmert-Wolf method can be summarised in the following steps.

(i) For each set of observations (measurements at one ship position, one satellite pass and one area of triangulation in the three examples in 7.2.1) the normal equations are independently formed, taking care to put the common parameters (beacon, receiver and junction point positions in the three examples) below the local parameters (ship position, pass unknowns and interior point coordinates) in the vector of parameters.

(ii) Each set of normal equations is independently reduced using the strategy of (3.21) to form reduced normal equations in terms of the common parameters only.

(iii) These reduced normal equations are simply added together and solved to give the least squares estimates of the common parameters.

(iv) If the least squares estimates of the local parameters are required it is necessary to return to the independent sets of normal equations (stage (i)) and use (3.20). If the residuals are required they can be obtained from (7.57) but if an estimate of only the unit variance is required (7.58) can be used.

(v) The covariance matrix of the common parameters is obtained from (7.59) and, if required, the approximate covariance matrices of the local parameters are obtained from (7.63).

The following points are worth noting.

(i) The procedure is exactly equivalent to a simultaneous least squares solution (7.2.2 proves this).

(ii) It has special advantages when the local parameters are not required, e.g. in the first example of 7.2.1 where a very small minicomputer could easily cope with what is equivalent to the simultaneous solution of many hundred simultaneous equations.

(iii) It has special advantages, in terms of the organization of the work, for the computation of continental networks; see Gagnon (1976).

(iv) If required the process can be split into a number of levels, i.e. the common parameters could themselves be divided. This is rather complex and only really applicable to exceptionally large problems. Isner (1978) and Dillinger (1978) describe the application of this strategy to the computation of the North American horizontal network.

7.3 A note on linearisation

It is important to point out that both sequential and step-by-step methods are usually only worthwhile when there are no significant linearisation errors, i.e. when the approximate values for the parameters, x^o, are very close to the final estimates. Obviously many of the advantages of the method would be lost if it were necessary to iterate.

7.4 Terminology

It should be realised that other authors use the phrase "step-by-step" to refer to different problems. Vaníček and Krakiwsky (1986) call Kalman Filtering and Sequential Least Squares "step-by-step" methods while for Gagnon (1976) anything that is not a simultaneous solution is a "step-by-step" solution.

8. Filtering, smoothing and prediction

Filtering, smoothing and prediction techniques are used for problems in which the parameters being estimated by the least squares process vary with time. One obvious example is navigation (at sea, on land or in space), where we try to estimate the position of a vehicle as it moves. The methods are, however, valid for any system in which the parameters have temporal variations even though there may be no movement. For instance, a stationary inertial surveying system has time varying system errors (gyro drift especially) which often appear as parameters in a least squares computation. Before any mathematical details are given the basic terminology will be explained by means of an example of navigation at sea.

Suppose we have a vessel moving so that at times t_1, t_2, ..., t_i it is at positions 1, 2, ..., i as in Fig. 8.1.

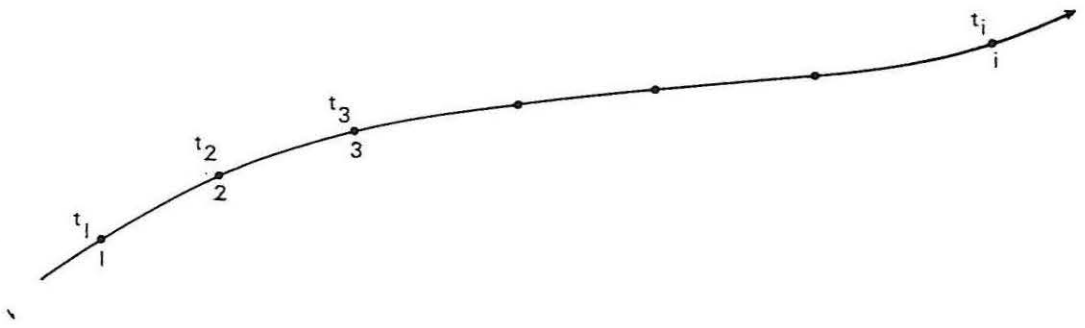


Fig.8.1

Let us say that t_i is the present and t_j is the time at which we want to estimate the position of the ship using all the information available up to t_i (the present). We can make the following definitions:

if $t_i = t_j$ we are <u>filtering</u>

$t_i > t_j$ we are <u>smoothing</u>

$t_i < t_j$ we are <u>predicting</u>.

To make these definitions clearer, imagine a ship at sea carrying out a seismic survey. It would be common to travel in straight lines, taking seismic measurements at specified linear intervals, whilst continuously interrogating some kind of shore-based navigation system. The process of

computing the ship's position at any instant in order to plot the real-time position on a chart would be filtering. The computation of the time at which the ship is expected to be at the correct position to make a seismic measurement would be prediction and the subsequent, possibly office-based, estimation of where the ship actually was when the measurements were taken would be smoothing.

There is a number of different mathematical strategies for filtering, smoothing and prediction. In this section a widely used method, called Kalman filtering, will be derived and discussed. It is important to note that despite its name the set of equations known as the Kalman filter can be used for all three activities, i.e. filtering, smoothing and prediction. Before the Kalman filter is introduced more definitions are required.

Returning to Fig. 8.1, suppose that $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_i$ are the true values of the parameters at points 1, 2, ..., i and that $\ell_1, \ell_2, \ldots, \ell_i$ are the corresponding vectors of measured quantities. For each point a functional model can be written as follows

$$\begin{aligned} F_1(\bar{x}_1, \bar{\ell}_1) &= 0 \quad \text{at point 1} \\ F_2(\bar{x}_2, \bar{\ell}_2) &= 0 \quad \text{at point 2} \\ &\text{etc.} \end{aligned} \tag{8.1}$$

which may be linearised (as in 2.2) to

$$\begin{aligned} A_1 x_1 + C_1 v_1 - b_1 &= 0 \\ A_2 x_2 + C_2 v_2 - b_2 &= 0 \\ &\text{etc.} \end{aligned} \tag{8.2}$$

Note that F_1, F_2, F_3, etc. may be entirely different, for instance point 1 may represent a satellite-Doppler fix, point 2 may be a shore-base navigation system fix, point 3 may be an acoustic fix, etc. In practice, however, it is more likely that most points will be fixed by the same system with only occasional measurements made with other systems. Also note that in this example x may contain a number of parameters besides positions: typically, velocity and heading may be included as well as scale and refraction unknowns. Some unknowns, e.g. scale and refraction, may best be considered as constant

over short periods of time, i.e. not all of the parameters have to be time varying.

(8.1) and (8.2) are usually referred to as <u>primary models</u>; they relate the parameters to the measurements and have associated covariance matrices which are used to determine the weight matrices W_1, W_2, ..., W_i in the usual way.

We further suppose that we have some kind of mathematical model which relates the parameters at point 2 to the parameters at point 1, i.e. we have some additional source of information as to how x varies with time. Such a model is known as the <u>secondary model</u> and can be written as

$$F_{12}(\bar{x}_1, \bar{x}_2, t_1, t_2) = 0 \tag{8.3}$$

or, more generally,

$$F_{i-1,i}(\bar{x}_{i-1}, \bar{x}_i, t_{i-1}, t_i) = 0 \tag{8.4}$$

The secondary model (8.3) can be linearised to

$$x_2 = Mx_1 + y \tag{8.5}$$

where y represents the unknown errors in the secondary model, i.e. the inability of (8.3) to predict correctly the temporal changes in the parameters. Associated with (8.5) is a covariance matrix C_y which reflects the precision of the secondary model. The inverse of C_y is the weight matrix of the secondary model

$$W_M = C_y^{-1} \tag{8.6}$$

For most practical problems, and indeed in the classical approach to filtering theory, it is convenient to consider the vector y as being given by

$$y = Tg \tag{8.7}$$

where g is the vector of the quantities which cause the secondary model to be incorrect and T is a coefficient matrix chosen so that the product Tg represents the effect of these quantities on the parameters. Notice that, in general this transformation matrix T will not be a square matrix as the number of error sources in the secondary model is not necessarily equal to the number of parameters. Usually g will be assumed to be a vector of random quantities with a zero mean and covariance matrix C_g. Then, by use of (4.16),

the weight matrix of the secondary model becomes

$$W_M = C_y^{-1} = (TC_gT^T)^{-1} \qquad (8.8)$$

In practice the secondary model will be determined from the known historical, or theoretically expected, changes in the parameters and the vector g will be some physical effect which is known to exist but is not modelled. For instance when navigating at sea the secondary model may describe straight line motion (constant velocity) and all its errors would be due to vehicle accelerations; hence g would contain these unknown accelerations (which could reasonably be described in practice as being random).

In the jargon of filtering techniques the vector of parameters x is called the state vector and the matrix M is known as the transition matrix. Vector g is known as the driving noise or more generally as the forcing function. Notice that M, T and W_M will all usually be functions of time.

8.1 The Kalman filter equations

The filter which we now call the Kalman filter was first derived in Kalman (1960) as a method for use in electrical control systems. It is now a standard method within the general mathematical subject area known as sequential smoothing and prediction, and most books on this topic, e.g. Morrison (1969), devote considerable space to the technique. Surveyors generally find it difficult to read publications such as these as the notation is very different to that to which they are accustomed. Also, the starting point for the "classical" derivations is usually the maximum likelihood requirement (see 6.4) which may be less familiar than least squares.

Krakiwsky (1975) has, however, shown that the Kalman filter can be derived from the standard least squares requirement and it is this derivation that will be given here. Note that in the case of observations with a normal pdf the least squares estimate is the same as the maximum likelihood estimate. Krakiwsky (1975) gives the derivation for the combined case of least squares but here only the special case of observation equations will be considered. This is simply to restrict the length (and apparent complexity) of the formulae - the procedure is identical for both cases. Readers who require the formulae for the combined case are advised to consult Krakiwsky (1975).

Consider the two times t_1 and t_2. At t_1 we have the primary model

$$A_1 x_1 = b_1 + v_1 \tag{8.9}$$

and at time t_2 another primary model

$$A_2 x_2 = b_2 + v_2 \tag{8.10}$$

Also we have the secondary model

$$x_2 = M x_1 + y \tag{8.11}$$

The filtering, smoothing and prediction problems are to find least squares estimates for x_2, x_1 and x_3 respectively. Hence we wish to minimise

$$v_1^T W_1 v_1 + v_2^T W_2 v_2 + y^T W_M y$$

subject to constraints (8.9), (8.10) and (8.11). Following Lagrange's method of undetermined multipliers described by (3.11) we write down the function

$$\Phi = v_1^T W_1 v_1 + v_2^T W_2 v_2 + y^T W_M y + 2k_1^T(A_1 x_1 - b_1 - v_1)$$
$$+ 2k_2^T(A_2 x_2 - b_2 - v_2) + 2k_3^T(x_2 - M x_1 - y) \tag{8.12}$$

To minimise Φ we differentiate with respect to all five variables

(i) $$\partial\Phi/\partial\hat{v}_1 = 2\hat{v}_1^T W_1 - 2\hat{k}_1^T = 0$$
$$\therefore W_1\hat{v}_1 - \hat{k}_1 = 0 \tag{8.13}$$

(ii) $$\partial\Phi/\partial\hat{v}_2 = 2\hat{v}_2^T W_2 - 2\hat{k}_2^T = 0$$
$$\therefore W_2\hat{v}_2 - \hat{k}_2 = 0 \tag{8.14}$$

(iii) $$\partial\Phi/\partial\hat{y} = 2\hat{y}^T W_M - 2\hat{k}_3^T = 0$$
$$\therefore W_M\hat{y} - \hat{k}_3 = 0 \tag{8.15}$$

(iv) $$\partial\Phi/\partial\hat{x}_1 = 2\hat{k}_1^T A_1 - 2\hat{k}_3^T M = 0$$
$$\therefore A_1^T\hat{k}_1 - M^T\hat{k}_3 = 0 \tag{8.16}$$

(v) $$\partial\Phi/\partial\hat{x}_2 = 2\hat{k}_2^T A_2 + 2\hat{k}_3^T = 0$$
$$\therefore A_2^T\hat{k}_2 + \hat{k}_3 = 0 \tag{8.17}$$

We now combine equations (8.12) to (8.17) with (8.9) to (8.11) to form the normal equations, which we write in the partitioned matrix form

$$\begin{bmatrix} W_1 & 0 & 0 & -I & 0 & 0 & 0 & 0 \\ 0 & W_2 & 0 & 0 & -I & 0 & 0 & 0 \\ 0 & 0 & W_M & 0 & 0 & -I & 0 & 0 \\ -I & 0 & 0 & 0 & 0 & 0 & A_1 & 0 \\ 0 & -I & 0 & 0 & 0 & 0 & 0 & A_2 \\ 0 & 0 & -I & 0 & 0 & 0 & -M & I \\ 0 & 0 & 0 & A_1^T & 0 & -M^T & 0 & 0 \\ 0 & 0 & 0 & 0 & A_2^T & I & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \hat{y} \\ \hat{k}_1 \\ \hat{k}_2 \\ \hat{k}_3 \\ \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ b_1 \\ b_2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (8.18)$$

In theory (8.18) could be solved as a set of simultaneous equations to give solutions for $\hat{x}_1$ and $\hat{x}_2$. Such a procedure would of course be quite impractical owing to the size of the left-hand-side matrix so we search for explicit expressions for $\hat{x}_1$ and $\hat{x}_2$. First we introduce $\hat{x}'_1$ as the estimate of $\hat{x}_1$ using only the information available up to time t_1,

i.e. $$\hat{x}'_1 = (A_1^T W_1 A_1)^{-1} A_1^T W_1 b_1 = N_1^{-1} u_1 \qquad (8.19)$$

with covariance matrix

$$C_{\hat{x}'_1} = (A_1^T W_1 A_1)^{-1} = N_1^{-1} \qquad (8.20)$$

where

$$N_1 = A_1^T W_1 A_1 \qquad (8.21)$$

and

$$u_1 = A_1^T W_1 b_1 \qquad (8.22)$$

Now we eliminate $\hat{v}_1$, $\hat{v}_2$ and $\hat{y}$ from (8.18) by successively using (3.21) to give

$$\begin{bmatrix} -W_1^{-1} & 0 & 0 & A_1 & 0 \\ 0 & -W_2^{-1} & 0 & 0 & A_2 \\ 0 & 0 & -W_M^{-1} & -M & I \\ A_1^T & 0 & -M^T & 0 & 0 \\ 0 & A_2^T & I & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{k}_1 \\ \hat{k}_2 \\ \hat{k}_3 \\ \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (8.23)$$

(8.23) is rearranged to prepare for the elimination of $\hat{k}_1$ and $\hat{x}_1$

$$\begin{bmatrix} -W_1^{-1} & A_1 & 0 & 0 & 0 \\ A_1^T & 0 & -M^T & 0 & 0 \\ 0 & -M & -W_M^{-1} & I & 0 \\ 0 & 0 & I & 0 & A_2^T \\ 0 & 0 & 0 & A_2 & -W_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{k}_1 \\ \hat{x}_1 \\ \hat{k}_3 \\ \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ 0 \\ 0 \\ 0 \\ b_2 \end{bmatrix} \quad (8.24)$$

Applying (3.21) to (8.24) to eliminate $\hat{k}_1$ gives

$$\begin{bmatrix} N_1 & -M^T & 0 & 0 \\ -M & -W_M^{-1} & I & 0 \\ 0 & I & 0 & A_2^T \\ 0 & 0 & A_2 & -W_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{k}_3 \\ \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} u_1 \\ 0 \\ 0 \\ b_2 \end{bmatrix} \quad (8.25)$$

Applying (3.21) to (8.25) to eliminate $\hat{x}_1$ gives

$$\begin{bmatrix} -(MN_1^{-1}M^T + W_M^{-1}) & I & 0 \\ I & 0 & A_2^T \\ 0 & A_2 & -W_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{k}_3 \\ \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} -MN_1^{-1}u_1 \\ 0 \\ b_2 \end{bmatrix} \quad (8.26)$$

We now introduce $\hat{\underset{\sim}{x}}'_2$, the "predicted" state vector computed using the secondary model, and $\hat{\underset{\sim}{x}}'_1$. $\hat{\underset{\sim}{x}}'_2$ is defined by

$$\hat{\underset{\sim}{x}}'_2 = M\hat{\underset{\sim}{x}}'_1 \quad (8.27)$$

and substituting in (8.19) gives

$$\hat{\underset{\sim}{x}}'_2 = MN_1^{-1}u_1 \quad (8.28)$$

Also applying (4.16) to (8.27) whilst remembering that implicit in (8.27) is a vector y with covariance matrix C_y (cf. (8.5)) we obtain

$$C_{\hat{\underset{\sim}{x}}'_2} = MC_{\hat{\underset{\sim}{x}}'_1}M^T + C_y \quad (8.29)$$

and substituting (8.6) and (8.20) in (8.29) gives

$$C_{\hat{\underset{\sim}{x}}'_2} = MN_1^{-1}M^T + W_M^{-1} \quad (8.30)$$

Now we introduce

$$N_2'^{-1} = C_{\hat{x}_2'} = MN_1^{-1}M^T + W_M^{-1} \tag{8.31}$$

Substituting (8.31) and (8.28) into (8.26) gives

$$\begin{bmatrix} -N_2'^{-1} & I & 0 \\ I & 0 & A_2^T \\ 0 & A_2 & -W_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{k}_3 \\ \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} -\hat{x}_2' \\ 0 \\ b_2 \end{bmatrix} \tag{8.32}$$

Applying (3.21) to (8.32) to eliminate $\hat{k}_3$ gives

$$\begin{bmatrix} N_2' & A_2^T \\ A_2 & -W_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{x}_2 \\ \hat{k}_2 \end{bmatrix} = \begin{bmatrix} N_2'\hat{x}_2' \\ b_2 \end{bmatrix} \tag{8.33}$$

Finally, applying (3.21) to (8.33) gives

$$(-W_2^{-1} - A_2N_2'^{-1}A_2^T)\hat{k}_2 = b_2 - A_2N_2'^{-1}N_2'\hat{x}_2'$$

i.e.
$$\hat{k}_2 = -(W_2^{-1} + A_2N_2'^{-1}A_2^T)^{-1}(b_2 - A_2\hat{x}_2') \tag{8.34}$$

Applying (3.20) to (8.33) gives

$$\hat{x}_2 = N_2'^{-1}(N_2'\hat{x}_2' - A_2^T\hat{k}_2)$$

i.e.
$$\hat{x}_2 = \hat{x}_2' - N_2'^{-1}A_2^T\hat{k}_2 \tag{8.35}$$

Substituting (8.34) into (8.35) gives

$$\hat{x}_2 = \hat{x}_2' + N_2'^{-1}A_2^T(W_2^{-1} + A_2N_2'^{-1}A_2^T)^{-1}(b_2 - A_2\hat{x}_2') \tag{8.36}$$

which is usually written as

$$\hat{x}_2 = \hat{x}_2' + G(b_2 - A_2\hat{x}_2') \tag{8.37}$$

where G, given by

$$G = N_2'^{-1}A_2^T(W_2^{-1} + A_2N_2'^{-1}A_2^T)^{-1} \tag{8.38}$$

is called the <u>gain matrix</u>.

The covariance matrix of $\hat{x}_2$ is derived as follows. (8.37) is rearranged as

$$\hat{x}_2 = (I - GA_2)\hat{x}_2' + Gb_2 \tag{8.39}$$

then applying (4.16) to (8.39), whilst noting that $\hat{x}'_2$ and b_2 are independent, gives

$$C_{\hat{x}_2} = (I - GA_2)C_{\hat{x}'_2}(I - GA_2)^T + GC_{b_2}G^T \tag{8.40}$$

Putting $W_2 = C_{b_2}^{-1}$ and substituting (8.31) in (8.40) gives

$$C_{\hat{x}_2} = (I - GA_2)N_2'^{-1}(I - A_2^T G^T) + GW_2^{-1}G^T \tag{8.41}$$

which expands and rearranges to

$$C_{\hat{x}_2} = (N_2'^{-1} - GA_2N_2'^{-1}) - (N_2'^{-1}A_2^T - GA_2N_2'^{-1}A_2^T - GW_2^{-1})G^T \tag{8.42}$$

The second bracketed term can be shown to be zero as follows. From (8.38)

$$N_2'^{-1}A_2^T = G(W_2^{-1} + A_2N_2'^{-1}A_2^T) \tag{8.43}$$

hence

$$N_2'^{-1}A_2^T - GW_2^{-1} - GA_2N_2'^{-1}A_2^T = 0 \tag{8.44}$$

Substituting (8.44) in (8.42) gives

$$C_{\hat{x}_2} = N_2'^{-1} - GA_2N_2'^{-1} = (I - GA_2)N_2'^{-1} \tag{8.45}$$

For smoothing we need to compute $\hat{x}_1$, the least squares estimate of x_1 using all data collected up to and including point 2. To do this we proceed as follows.

From (8.32) we write

$$\hat{k}_3 + A_2^T\hat{k}_2 = 0$$

giving

$$\hat{k}_3 = -A_2^T\hat{k}_2 \tag{8.46}$$

and from (8.25) we write

$$N_1\hat{x}_1 - M^T\hat{k}_3 = u_1$$

giving

$$\hat{x}_1 = N_1^{-1}u_1 + N_1^{-1}M^T\hat{k}_3 \tag{8.47}$$

Substituting (8.19) and (8.47) gives

$$\hat{x}_1 = \hat{x}'_1 - N_1^{-1} M^T A_2^T \hat{k}_2 \tag{8.48}$$

where $\hat{k}_2$ is obtained from (8.34).

8.2 Summary of formulae and procedure

A summary of the working formulae and procedure will now be given. To emphasise the recursive nature of the Kalman filter the suffices 1 and 2 will be changed to i-1 and i respectively and the possible temporal variation of the transition matrix will be highlighted by using the notation $M_{i-1,i}$ for M. Similarly, W_M will also vary with time and it is now denoted by $W_{M_{i-1,i}}$. The procedure is:

(i) set i = 1

(ii) compute the starting estimate of the state vector from (8.19)

$$\hat{x}'_i = N_i^{-1} u_i \tag{8.49}$$

(iii) increment i

$$i = i + 1 \tag{8.50}$$

(iv) predict the state vector from (8.27)

$$\hat{x}'_i = M_{i-1,i} \hat{x}'_{i-1} \tag{8.51}$$

(v) predict the covariance matrix from (8.31)

$$C_{\hat{x}'_i} = N'^{-1}_i = M_{i-1,i} N_{i-1}^{-1} M^T_{i-1,i} + W^{-1}_{M_{i-1,i}} \tag{8.52}$$

(vi) compute the gain matrix from (8.38)

$$G = N'^{-1}_i A_i^T (W_i^{-1} + A_i N'^{-1}_i A_i^T)^{-1} \tag{8.53}$$

(vii) compute the state vector from (8.37)

$$\hat{x}_i = \hat{x}'_i + G(b_i - A_i \hat{x}'_i) \tag{8.54}$$

(viii) compute the covariance matrix of the state vector from (8.45)

$$C_{\hat{x}_i} = (I - GA_i) N'^{-1}_i = N_i^{-1} \tag{8.55}$$

(ix) prepare for the next iteration

$$\hat{x}'_i = \hat{x}_i \tag{8.56}$$

(x) return to step (iii).

The above steps are for filtering. For smoothing we use (8.48)

$$\hat{x}_{i-1} = \hat{x}'_{i-1} - N_{i-1}^{-1} M_{i-1,i}^T A_i^T \hat{k}_i \tag{8.57}$$

where, from (8.34),

$$\hat{k}_i = -(W_i^{-1} + A_i N_i^{-1} A_i^T)^{-1}(b_i - A_i \hat{x}'_i) \tag{8.58}$$

Note that smoothing is carried out after the measurements have been completed. If there are n state vectors (max i = n) we would obtain $\hat{x}_{n-1}$, then $\hat{x}_{n-2}$, etc., until we reached $\hat{x}_1$.

Prediction is carried out at any stage simply by using the secondary model whilst assuming no model errors, i.e. from (8.27)

$$\hat{x}'_{i+1} = M_{i,i+1} \hat{x}'_i \tag{8.59}$$

with covariance matrix, from (8.31),

$$C_{\hat{x}'_{i+1}} = M_{i,i+1} C_{\hat{x}_i} M_{i,i+1}^T + W_{M_{i,i+1}}^{-1} \tag{8.60}$$

So far nothing has been said of the computation of the residuals. These can be simply obtained from (8.10)

$$\hat{v}_i = A_i \hat{x}_i - b_i \tag{8.61}$$

and can be tested in the usual way (see 5.4.1) for possible rejection. Also, the model errors can be examined by using (8.11)

$$\hat{y}_{i-1,i} = \hat{x}_i - M_{i-1,i} \hat{x}_{i-1} \tag{8.62}$$

8.3 Numerical aspects

A quick glance at the formulae in 8.2 may give the impression that the Kalman filter involves a great deal of computational work. In particular, the reader will notice a large number of matrix inversion symbols. It is important to realise that for any one cycle of the filter the only inversion occurs in the computation of the gain matrix when it is necessary to invert

$$W_i^{-1} + A_i N'^{-1}_i A_i^T$$

This matrix will have a size equal to the number of observations in the new primary model, i.e. the number of measurements made at the ith ship position

in the example shown in Fig. 8.1.

The matrices W_i^{-1} and $W_{M_{i-1,i}}^{-1}$ are not obtained by inversions: they are only symbols for the covariance matrices of the primary and secondary models. Also, the matrix N_i^{-1} is found from $N_i'^{-1}$ without carrying out a new inversion. Hence the efficiency of the Kalman filter is largely related to the number of measurements added at each new time epoch - the smaller this number the more efficient the Kalman filter. It is worth mentioning here that there is a mathematically equivalent (i.e. giving identical answers) method called Bayes filtering. This involves the inversion (during each cycle) of a matrix whose size is equal to the number of parameters in the state vector. Hence if the number of observations at each epoch is larger than the number of parameters in the state vector the Bayes filter is more efficient than the Kalman filter, and vice versa. Formulae for the Bayes filter can be found in Krakiwsky (1975).

8.4 Examples of applications

8.4.1 Navigation

Suppose we have a ship travelling in a steady fashion through an area of acoustic beacons as indicated in Fig. 8.2. At points 1, 2, ..., i the ship fixes its position by means of distance measurements to a number of beacons and carries out i independent least squares computations.

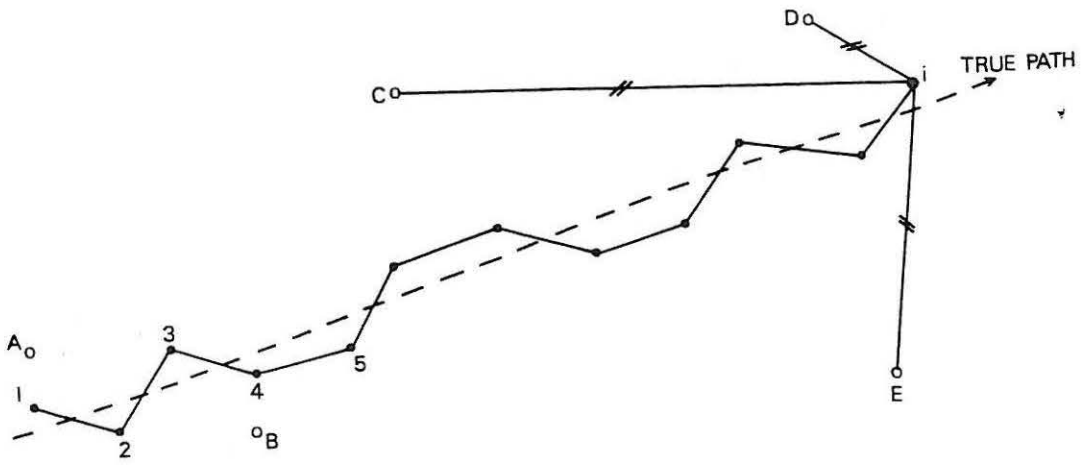


Fig. 8.2

If the results of these computations are plotted and joined up we obtain the solid line in Fig. 8.2. Clearly this line does not represent the true path of the ship: it is uneven simply because of the errors in the position-fixes. A line rather like the dotted line in Fig. 8.2 is the true path of the ship and we need a method which can somehow combine all the individual position-fixes to estimate this in real-time.

The Kalman filter is the ideal method for this because it can cater for both the position-fix measurements (primary model) and a secondary model which in some way assumes that the final path is smooth. As a somewhat over-simplified example the filter could be applied to the foregoing problem as follows.

Let a position-fix be carried out every Δt seconds with the results of the ith position fix being E_i^o and N_i^o with covariance matrix W_i^{-1}. Note that the superscript ^{o} is used to denote an observed value; strictly speaking, the observations are the distance measurements from i to points C, D, E etc. but for this simple example we are assuming that these have been processed to produce "observed" positions with an associated covariance matrix computed in the usual way. The state vector for this problem will contain four elements: two for the ship's position and one each for the east and north components of the ship's velocity, i.e.

$$x_i = \left[E_i \; N_i \; V_{E_i} \; V_{N_i}\right]^T \tag{8.63}$$

The primary model

$$Ax_i = b_i + v_i \tag{8.64}$$

will be

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} E \\ N \\ V_E \\ V_N \end{bmatrix}_i = \begin{bmatrix} E^o \\ N^o \end{bmatrix}_i + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_i \tag{8.65}$$

If we assume the ship to have a constant velocity then the transition matrix $M_{i-1,i}$ will be given by

$$M_{i-1,i} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8.66}$$

In this situation the driving noise (the quantity that will cause the secondary model to be in error) will contain the two components, a_E and a_N, of the ship's acceleration and its effect on the state vector y is given by

$$y = Tg \tag{8.7}$$

$$\therefore\ y = \begin{bmatrix} \Delta t^2/2 & 0 \\ 0 & \Delta t^2/2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \begin{bmatrix} a_E \\ a_N \end{bmatrix} \tag{8.67}$$

Notice that when (8.66) and (8.67) are combined as in (8.5) they represent simple applications of the classical equations relating distance, velocity and acceleration, viz.

$$x_i = Mx_{i-1} + Tg \tag{8.68}$$

$$\begin{bmatrix} E \\ N \\ V_E \\ V_N \end{bmatrix}_i = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} E \\ N \\ V_E \\ V_N \end{bmatrix}_{i-1} + \begin{bmatrix} \Delta t^2/2 & 0 \\ 0 & \Delta t^2/2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \begin{bmatrix} a_E \\ a_N \end{bmatrix} \tag{8.69}$$

which multiply out to

$$\begin{aligned} E_i &= E_{i-1} + V_{E_{i-1}} \Delta t + \tfrac{1}{2} a_E \Delta t^2 \\ N_i &= N_{i-1} + V_{N_{i-1}} \Delta t + \tfrac{1}{2} a_N \Delta t^2 \\ V_{E_i} &= V_{E_{i-1}} + a_E \Delta t \\ V_{N_i} &= V_{N_{i-1}} + a_N \Delta t \end{aligned} \tag{8.70}$$

The ship's accelerations (driving force in this example) would usually be considered random (note that if this were not so, very large and practically impossible velocities would occur). In practice these accelerations are not known (nor are they needed for the Kalman filter) but an estimate of their covariance matrix C_g is essential. It would be usual to consider the accelerations independent and their standard deviations (this term is preferred here to the synonymous standard error because the motion of the ship is not itself an error) equal. The size of these standard deviations would be a function of a number of physical quantities such as the ship's power, the weather, the state of the sea and the type of motion (e.g. straight line or turning). To compute W_M from C_g we use (8.8). For instance if we denote the standard deviation of the ship's accelerations by σ we have

$$W_M^{-1} = \begin{bmatrix} \Delta t^2/2 & 0 \\ 0 & \Delta t^2/2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} \Delta t^2/2 & 0 & \Delta t & 0 \\ 0 & \Delta t^2/2 & 0 & \Delta t \end{bmatrix} \qquad (8.71)$$

Hence in (8.65), (8.66) and (8.71) together with the covariance matrix of the primary model W_i we have all the matrices needed to use the Kalman Filter equations listed in 8.2.

It should be emphasised that the foregoing is not suggested as a practical procedure for applying the Kalman filter to the described navigation problem. It is merely set out as a simple example of how to assign values to the various matrices used by the filter. In practice the problem is likely to be more complex with a need to combine data from a variety of sources, such as velocity and heading sensors, and to solve for a variety of parameters, such as bias errors in the measuring systems.

Finally, it is remarked that the performance of the filter is highly dependent on the weight matrices W_i and W_M. If we assign large weights to the secondary model (i.e. W_M is large compared with W_i) we will obtain a very smooth real-time path for the ship but it will be slow to react to sharp changes in direction (e.g. when a ship turns through 180° when running seismic lines). Conversely, if W_i is large compared with W_M we may obtain an uneven path, but it will react quickly to sharp changes. These problems need to be solved by experience.

8.4.2 Inertial surveying

Inertial surveying systems make regular measurements of velocity, typically about fifty per second, in three orthogonal directions (usually north, east and "down"). For some applications it is required to have a real-time estimate of the position of a system as it moves over the earth's surface. In such cases we would have a state vector consisting of a number of time varying quantities such as position, velocity, heading errors, gyro drift rates and accelerometer drift rates. The quantities to be included in the state vector should be a function of the particular inertial system, the required accuracy and the capabilities of the on-board computer. As an example, Gregerson (1975) states that the Litton system used by the Geodetic Survey of Canada includes seventeen elements in its state vector.

It is not possible, without going into details of the theory of inertial surveying, to give a full account of how the Kalman filter is applied. An indication of its application, however, can be seen by looking at a simplified system.

Adams (1979) describes a two-dimensional system with a seven parameter state vector

$$x = \left[\alpha,\ \beta,\ \gamma,\ d\varphi,\ d\lambda,\ d\dot{\varphi},\ d\dot{\lambda}\right]^T \tag{8.72}$$

where α, β and γ are the orientation errors of the three axes, $d\varphi$ and $d\lambda$ are the errors in the positions derived by velocity integration and $d\dot{\varphi}$ and $d\dot{\lambda}$ are the north and east velocity errors. The filter will be used to evaluate the state vector at every velocity update with the design matrix for the primary model being given by

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{8.73}$$

because the measured quantities at each update are the velocity errors. The primary model weight matrix, W, will be a function of the precision of the velocity measurements and will not vary with time

$$W_i^{-1} = W^{-1} = \begin{bmatrix} \sigma_{\dot{\varphi}}^2 & 0 \\ 0 & \sigma_{\dot{\lambda}}^2 \end{bmatrix} \tag{8.74}$$

The transition matrix describes how the elements of x vary with time. This is rather complicated in inertial surveying owing to the cyclic nature of the error propagation (see, for example, Cross and Webb (1980)). Hence the matrix M will not be quoted here; suffice it to say that it is a square matrix of size seven whose elements are mainly harmonic functions with periods related to the Schuler period (84.4 minutes) and the rate of rotation of the earth. The matrix and its associated covariance matrix W_M^{-1} is given in full in Adams (1979).

8.5 Concluding remarks

The Kalman filter is a fully rigorous least squares method for the filtering, smoothing and prediction of time varying quantities. As such it can be described as an optimal filter. Actually the term optimal filter strictly applies to filters which lead to maximum likelihood solutions, but it has been shown in 6.4 that for a normally distributed pdf the least squares estimate is identical to the maximum likelihood estimate.

Of course the advantage of a filter over a normal least squares computation is in the organization of the data so that it can be handled by the types of mini- and microcomputer commonly dedicated to position-fixing systems. To carry out an ordinary least squares computation using all the available data up to any particular time would involve massive sets of equations to solve and matrices to invert. It must, however, be emphasised that if such a solution were to be made the solution for the latest data point would be identical to the Kalman filter solution (the solution for the previous points would be the same as the smoothed solution).

Finally, it is remarked that, as pointed out by Krakiwsky (1975), the sequential least squares method of 7.1 is actually a special case of the Kalman filter, i.e. sequential least squares is the Kalman filter applied to parameters that do not change with time. The equivalence of the formulae can be seen by dropping the subscripts from the state vector and putting $M = W_M^{-1} = 0$. Then, for example, equations (7.19) and (8.36) become identical.

9. Least squares collocation

The strict mathematical definition of collocation is given by Moritz (1980) as "the determination of a function by fitting an analytical approximation to a given number of linear functionals". The technique is mainly used in surveying and geodesy to determine the values of quantities at points other than those at which measurements have been made (or at which information is known). In its simplest form least squares collocation is exactly equivalent to a technique known as least squares prediction (or, more commonly, as least squares interpolation).

The application of least squares collocation to position-fixing is not as directly obvious as the applications of the other least squares methods described in this Working Paper. It does, however, have some important "indirect" uses, for instance in the prediction of spatially varying quantities such as gravity, height, deviation of the vertical and geoid-ellipsoid separation (all of which are needed in classical geodesy to reduce measured quantities to a reference surface ready for geometric computations). Also it is a very powerful tool for coordinate transformations, especially in cases where measurement errors have resulted in variable transformation parameters. The major geodetic uses of collocation are, however, in the general field of physical geodesy where it can be used for many aspects (some of which have already been mentioned) of the computation of the anomalous gravity field. The technique is included here for the foregoing "indirect" applications and because it belongs to the general family of least squares methods to which this Working Paper is devoted.

An essential feature of the method is that quantities which are by nature deterministic are described in a statistical manner, particularly by the use of covariance matrices. This is in contrast to the techniques in the rest of this Working Paper, which use statistics only for measured quantities (or for functions of them). For instance if we were to use least squares collocation to predict the unknown height of a point surrounded by a number of points of known height we would need to establish a function (known as a covariance function) from which it would be possible to compute the covariance of the heights at any two points. This function would be in terms of quantities such as position and distance between the points. Hence all the elements of a height covariance matrix could be computed. This matrix would then describe the variation of height in the area in a statistical manner: in flat areas the heights of neighbouring points would be highly correlated leading to large

covariances whereas in highly undulating areas we would expect small covariances (i.e. the height of a point would be largely unrelated to the heights of neighbouring points). This procedure is in contrast to the better known process of fitting a mathematical surface to the known points and using it to predict the heights of other points.

Hence before discussing the technique of least squares collocation in detail it is necessary to consider the computation of covariance matrices for quantities that are, in principle, non-stochastic.

9.1 Covariance matrices

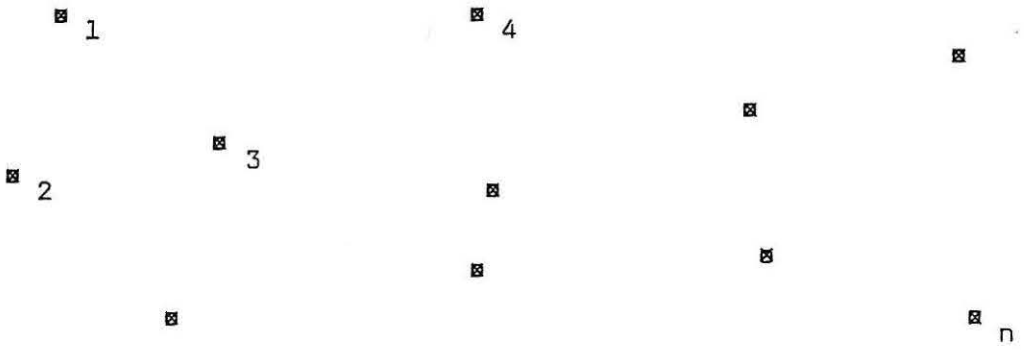


Fig. 9.1

Consider n points (Fig. 9.1) at which we know the value of a quantity (e.g. height, gravity anomaly, deviation of the vertical or a coordinate transformation parameter) u, i.e. we know u_1, u_2, ..., u_n. One method of determining a covariance function would be to proceed as follows.

First we might assume that the correlation of the quantity between any two points i and j is a function only of the distance, d_{ij}, between them. Then using all n_1 pairs of points separated by a distance of up to r_1 metres we compute their covariance from

$$c_1 = \frac{1}{n_1} \Sigma u_i u_j \qquad (9.1)$$

The process is then repeated using all n_2 pairs of points separated by a distance greater than r_1 and less than r_2 metres etc. Generally we can write, for the n_k pairs of points separated by a distance greater than r_{k-1} and less than r_k metres,

$$c_k = \frac{1}{n_k} \Sigma u_i u_j \qquad (9.2)$$

We can now plot, e.g. Fig. 9.2, the covariance histogram and draw a curve to represent the covariance function. The covariance matrix for the n points

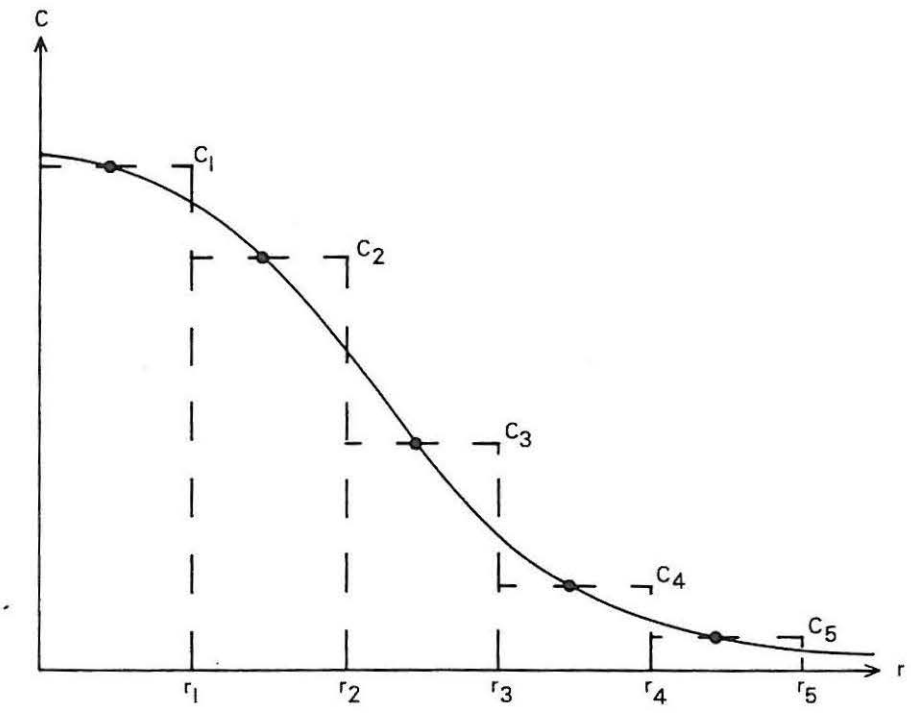


Fig. 9.2

is now written as

$$C_u = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \qquad (9.3)$$

where any individual element c_{ij}, corresponding to the covariance between points i and j, is determined simply by computing the distance r_{ij} between them and reading off the value of c from the curve in Fig. 9.2. Alternatively a mathematical function, e.g.

$$c_{ij} = a \exp(-b\, r_{ij}) \qquad (9.4)$$

where a and b are constants, could be fitted to the data and subsequently used

to compute each element of C_u. Once a covariance function has been computed from a set of data it would be common to use the same function for other similar problems, especially those that do not involve enough data to enable an internal computation of the covariance matrix.

For some problems it is possible to compute theoretically the form of the covariance function. This is particularly appropriate in physical geodesy where much work has been done on the "spatial covariance function of the anomalous potential". For instance Moritz (1980,84) quotes the general form of this and Tscherning and Rapp (1974) quote explicit forms of covariance functions for a variety of geodetic quantities. Furthermore, for some problems we use idealised forms of covariance matrices based on the manner in which we would like a quantity to behave (rather than perhaps how it actually does), e.g. Grafarend (1974) and the use of the Taylor-Karman covariance structure for geodetic networks.

We now extend the concept of the covariance matrix to cover the situation shown in Fig. 9.3 where we have a quantity, u, known at points 1,2,3, etc.

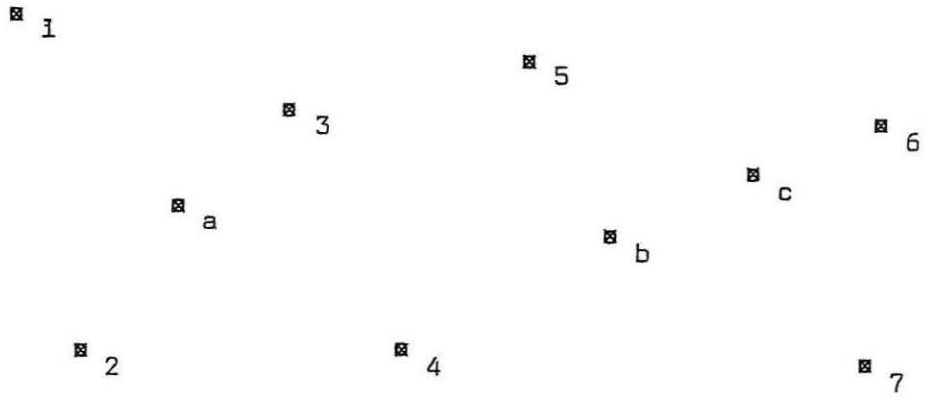


Fig. 9.3

but unknown at points a,b,c. In this case we find it convenient to partition the complete vector u into two parts u_1 and u_2

$$u = [u_1 \mid u_2]^T \qquad (9.5)$$

where u_1 contains the values of the quantity at points 1,2,3, etc. (called data points) and u_2 contains the values at points a,b,c, etc. (called computation points). Then the covariance matrix of u is correspondingly partitioned

$$C_u = \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array}\right] \qquad (9.6)$$

where the generally non-square matrices C_{12} and C_{21} (note that $C_{12} = C_{21}^T$) are often termed the cross-covariance matrices between the data and computation points. All the elements of C_u are determined from formulae such as (9.4) in the manner already described.

Finally it is emphasised that the covariance matrices discussed in this section do <u>not</u> relate to observational errors. If there are observations at the data points then these observations will of course have a covariance matrix related to the precision of the measurements, but this has nothing to do with C_u, which describes how a particular quantity is spatially (or possibly temporally) varying.

9.2 Least squares prediction

As a preliminary to least squares collocation we will consider the simple case of least squares prediction. Referring to 9.1 and Fig. 9.3, let u_1 be a vector of known quantities at points 1,2,3, etc. and let u_2 be the unknown values of the quantities at a,b,c, etc. Again it is emphasised that we are not here concerned with measurement errors, i.e. u_1 is perfectly known, but it is required to <u>estimate</u> u_2. Any linear estimates of u_2, say u_2^*, must be of the form

$$u_2^* = Qu_1 \qquad (9.7)$$

where Q is a linear transformation to be determined.

Let e* be the true error of the estimate u_2^*, then

$$e^* = u_2^* - u_2 \qquad (9.8)$$

and substituting (9.7) in (9.8) we have

$$e^* = Qu_1 - u_2 \qquad (9.9)$$

which can be rewritten as

$$e^* + [Q \mid -I] \left[\begin{array}{c} u_1 \\ \hline u_2 \end{array}\right] \qquad (9.10)$$

Then applying (4.16) to (9.10) and using (9.6) we obtain the covariance matrix of e^*

$$C_{e^*} = [Q \mid -I] \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array}\right] \left[\begin{array}{c} Q^T \\ \hline -I \end{array}\right] \tag{9.11}$$

which multiplies out to

$$C_{e^*} = QC_{11}Q^T - QC_{12} - C_{21}Q^T + C_{22} \tag{9.12}$$

and reorders to

$$C_{e^*} = C_{22} + QC_{11}Q^T - QC_{12} - C_{21}Q^T \tag{9.13}$$

Now $C_{21}C_{11}^{-1}C_{12}$ is subtracted and added to (9.13) to yield

$$C_{e^*} = C_{22} - C_{21}C_{11}^{-1}C_{12} + QC_{11}Q^T - QC_{12} - C_{21}Q^T + C_{21}C_{11}^{-1}C_{12} \tag{9.14}$$

Since $C_{11}C_{11}^{-1} = I$ (9.14) can be written in the expanded form

$$C_{e^*} = C_{22} - C_{21}C_{11}^{-1}C_{12} + QC_{11}Q^T - QC_{11}C_{11}^{-1}C_{12} - C_{21}C_{11}C_{11}^{-1}Q^T + C_{21}C_{11}^{-1}C_{11}C_{11}^{-1}C_{12} \tag{9.15}$$

which, after putting $C_{12} = C_{21}^T$, becomes

$$C_{e^*} = C_{22} - C_{21}C_{11}^{-1}C_{21}^T + (Q - C_{21}C_{11}^{-1})C_{11}(Q - C_{21}C_{11}^{-1})^T \tag{9.16}$$

We can write (9.16) as the sum of two matrices viz.

$$C_{e^*} = F + G \tag{9.17}$$

where

$$F = C_{22} - C_{21}C_{11}^{-1}C_{21}^T \tag{9.18}$$

and

$$G = (Q - C_{21}C_{11}^{-1})C_{11}(Q - C_{21}C_{11}^{-1})^T \tag{9.19}$$

Hence we have derived an expression for the precision of any linear estimate, u_2^*, of u_2. We now pose the question: what choice of Q in (9.7) will result in the best estimate of u_2? Examination of (9.18) shows that F is independent of Q; hence we are only concerned with the effect of Q on G in (9.19). In fact any choice of Q will yield a matrix G with non-negative diagonal elements.

This is because the ith diagonal element of G in (9.19) is given by the quadratic form

$$G_{ii} = g\, C_{11}\, g^T \tag{9.20}$$

where g is the ith row of $Q - C_{21}C_{11}^{-1}$ and C_{11} is positive-definite. Hence any choice of Q will make the variances of the error in each element of u_2^* equal to or larger than the diagonal elements of F. The minimum variance estimate will therefore be obtained when G is a null matrix, i.e.

$$Q - C_{21}C_{11}^{-1} = 0 \tag{9.21}$$

or

$$Q = C_{21}C_{11}^{-1} \tag{9.22}$$

Substituting (9.22) in (9.7) gives the best (in the sense of minimum variance) linear estimate of u_2 as

$$\hat{u}_2 = C_{21}C_{11}^{-1}u_1 \tag{9.23}$$

which, because of its minimum variance property, is also termed the least squares estimate. (9.23) is often written in the following manner for the prediction of u at any particular computation point p:

$$\hat{u}_p = \begin{bmatrix} c_{p1} & c_{p2} & \cdots & c_{pn} \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}^{-1} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{9.24}$$

where c_{pi} represents the vector of covariances between point p and the ith data point and all other symbols are as previously defined, i.e. $u_1, u_2, \ldots, u_n$ are the values of the quantity at the data points and the square matrix to be inverted is the covariance matrix of the quantities at the data points.

Finally, it is worth noting that the evaluation of $\hat{u}_2$ from (9.23) does not, in practice, involve the inversion of C_{11}. This is because the product $C_{11}^{-1}u_1$ can be replaced by a vector y where y is the solution to the square set of linear equations $C_{11}y = u_1$.

9.3 The least squares collocation equations

In its simplest form we can consider least squares collocation as a direct extension of least squares prediction to the case where the quantities being determined at the computation points are not generally the same as those being measured (or known) at the data points. For instance in physical geodesy, from which the technique evolved and in which it finds its greatest application, the quantity to be estimated is generally the anomalous potential and the measured (or known) quantities are usually gravity anomalies and/or deviations of the vertical. Furthermore, the general collocation model is also able to take into account measurement errors at the data points and the possible requirement to compute certain parameters during the prediction process. For instance we may wish to recover the parameters of the normal gravity field in the aforementioned physical geodesy example, or in the application of the method to coordinate transformations we may wish to determine the parameters of a specified transformation. Here we will treat the problem in its most general form and later the least squares prediction formula will be shown to be a special case of collocation.

Consider a set of data points at which we have made n observations. Let there also be q computation points and m parameters to be recovered. As usual we will denote the true values of the modelled observed quantities and the parameters by the vectors $\bar{\ell}$ and $\bar{x}$ respectively. We can write down n, generally non-linear, observation equations of the form

$$F(\bar{x}) - \bar{\ell} = 0 \qquad (2.20)$$

Also we rewrite (2.1) in the form

$$\ell = \bar{\ell} + e \qquad (9.25)$$

where e is the total "error" in the observations (i.e. the difference between the observed and modelled quantities). In collocation this total error is considered to be the sum of two independent errors usually called the signal and noise, and denoted by the symbols s_1 and n respectively. Hence (9.25) is written as

$$\ell = \bar{\ell} + s_1 + n \qquad (9.26)$$

After linearising (2.20) using (9.26) we obtain

$$Ax - b + s_1 + n = 0 \qquad (9.27)$$

An explanation of the terms signal and noise is now necessary. The noise is simply the random measurement error in ℓ; it is basically the residual vector, v, in all the previous sections of this Working Paper, with a change of sign. The signal can be most usefully interpreted as the inability of the chosen model, when simplified to

$$\begin{aligned} Ax &= b + v \\ &= b - n \end{aligned} \qquad (9.28)$$

to describe completely the physical situation. This may be for one of two, often indistinguishable, reasons.

(i) The model $F(\bar{x}) = \bar{\ell}$ may be incomplete and/or inaccurate, i.e. there may be some parameters that have not been considered and/or those that have been considered may not have been correctly related to the observed quantities.

(ii) The measurements may have systematic errors that have not been modelled, either by the inclusion of suitable parameters or by the selection of the correct stochastic model (i.e. the weight matrix).

In other words there may be some signal in the system that is not being modelled by (9.28) and the measurement noise covariance matrix, which we now denote by C_n. The meanings of the terms may be further clarified by an example.

Consider the problem of the geodetic datum transformation from an established local (astrogeodetic) coordinate system, such as OSGB36 (the British national mapping coordinate system) to the satellite-Doppler coordinate system. The process would be to observe Doppler coordinates (X_i^D, Y_i^D, Z_i^D) at, say, p points (i.e. i = 1,2, ..., p) with known coordinates (X_i^L, Y_i^L, Z_i^L) on the local system. We now select a transformation model, say translation only, and write down the collocation observation equations as follows

$$Ax - b + s_1 + n = 0 \qquad (9.27)$$

where

$$
\underset{(3p \times 3)}{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} X_1^D - X_1^L \\ Y_1^D - Y_1^L \\ Z_1^D - Z_1^L \\ X_2^D - X_2^L \\ Y_2^D - Y_2^L \\ \cdot \\ \cdot \\ \cdot \\ Z_P^D - Z_P^L \end{bmatrix}
$$

and the translation parameters are

$$
x = \begin{bmatrix} \Delta X & \Delta Y & \Delta Z \end{bmatrix}^T
$$

In this case n would represent the noise (random observational error) of the Doppler measurements and s_1 would be the signal due to the following two causes, of which the second is likely to be the larger in practice:

(i) oversimplification of the basic model - we should also, in practice, be including parameters for rotations and possibly a scale difference between the two coordinate systems

(ii) distortions in the local coordinate system due to a variety of causes such as lack of scale control, computation in blocks, etc.

The problem of collocation is now to estimate simultaneously the following:

(i) the parameters x

(ii) the signal s_1 and noise n at the data points

(iii) the signal s_2 at the computation points, i.e. in the foregoing example points at which it is required to transform from local to Doppler coordinates but at which there are no Doppler measurements.

It is an essential prerequisite to the application of least squares collocation that we know the covariance matrices of both the signal and the noise. C_n, the covariance matrix for the noise, is obtained in the usual way (it is equivalent to C_ℓ in 4.2) and C_s, the signal covariance matrix for both the data and

computation points, by a study (such as that outlined in 9.1) of the variation of the signal where it is known or can be estimated. C_n will often be diagonal but C_s will invariably be a full matrix as the whole point of differentiating between the signal and the noise is that the signal is highly spatially (or possibly temporally) correlated and has completely different statistical properties to the noise. It should be noted that the statistical expectations of both the signal and the noise are zero, i.e.

$$E(s) = 0 \tag{9.29}$$

$$E(n) = 0 \tag{9.30}$$

The derivation of the collocation equations now proceeds as follows. Let s be a vector containing the signal at both the data and computation points, i.e.

$$s = \left[s_1 \mid s_2 \right]^T \tag{9.31}$$

Then (9.27) can be rewritten as

$$Ax - b + Bs + n = 0 \tag{9.32}$$

with

$$B = \left[I \mid 0 \right] \tag{9.33}$$

Note that if we have q computation points then B will have dimensions $(n + q) \times n$ with I, a unit matrix, being $n \times n$ and 0, a null matrix, being $q \times n$.

We now wish to estimate x, s and n in (9.32) using the method of least squares, i.e. minimising

$$s^T C_s^{-1} s + n^T C_n^{-1} n$$

Hence, using Lagrange's method of undetermined multipliers as in (3.12), we have

$$\Phi = s^T C_s^{-1} s + n^T C_n^{-1} n + 2k^T (Ax - b + Bs + n) \tag{9.34}$$

which is minimised by differentiating as follows

$$\frac{\partial \Phi}{\partial x} = 2A^T \hat{k} = 0 \tag{9.35}$$

$$\frac{\partial \Phi}{\partial s} = 2C_s^{-1} \hat{s} + 2B^T \hat{k} = 0 \tag{9.36}$$

$$\frac{\partial \Phi}{\partial n} = 2C_n^{-1}\hat{n} + 2\hat{k} = 0 \tag{9.37}$$

Also the least squares estimates must satisfy (9.32), viz.

$$A\hat{x} - b + B\hat{s} + \hat{n} = 0 \tag{9.38}$$

Now after dividing (9.35) to (9.37) by 2 and combining them with (9.38) we obtain the following least squares hypermatrix

$$\begin{bmatrix} C_n^{-1} & 0 & I & 0 \\ 0 & C_s^{-1} & B^T & 0 \\ I & B & 0 & A \\ 0 & 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{n} \\ \hat{s} \\ \hat{k} \\ \hat{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ b \\ 0 \end{bmatrix} \tag{9.39}$$

which, after applying (3.21) to eliminate $\hat{n}$, reduces to

$$\begin{bmatrix} C_s^{-1} & B^T & 0 \\ B & -C_n & A \\ 0 & A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{s} \\ \hat{k} \\ \hat{x} \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix} \tag{9.40}$$

A further application of (3.21), to eliminate $\hat{s}$, leads to

$$\begin{bmatrix} -(C_n + BC_sB^T) & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{k} \\ \hat{x} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{9.41}$$

Now, using (9.33) whilst noting (9.31), the expression $-(C_n + BC_sB^T)$ in (9.41) can be simplified as follows

$$-(C_n + BC_sB^T) = -C_n - \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} C_{s_1} & C_{s_1 s_2} \\ C_{s_2 s_1} & C_{s_2} \end{bmatrix} \begin{bmatrix} I \\ 0 \end{bmatrix} \tag{9.42}$$

$$= -(C_n + C_{s_1}) \tag{9.43}$$

where C_{s_1} and C_{s_2} are the covariance matrices of the signal at the data and computation points respectively and $C_{s_1 s_2}$ and $C_{s_2 s_1}$ are their cross covariance matrices. Substituting (9.43) in (9.41) gives

$$\begin{bmatrix} -(C_n + C_{s_1}) & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{k} \\ \hat{x} \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{9.44}$$

Application of (3.21) to (9.44) leads to

$$\hat{x} = (A^T(C_n + C_{s_1})^{-1}A)^{-1}A^T(C_n + C_{s_1})^{-1}b \tag{9.45}$$

which is the least squares collocation expression for the parameters. Similarly, application of (3.20) to (9.44) produces

$$\hat{k} = -(C_n + C_{s_1})^{-1}(b - A\hat{x}) \tag{9.46}$$

Then substituting (9.46) in (9.36) and rearranging leads to

$$\hat{s} = C_s B^T(C_n + C_{s_1})^{-1}(b - A\hat{x}) \tag{9.47}$$

which is the least squares collocation expression for the signal at both the data and computation points. The noise at the data points is obtained by substituting (9.46) in (9.37) to yield

$$\hat{n} = C_n(C_n + C_{s_1})^{-1}(b - A\hat{x}) \tag{9.48}$$

Hence we have derived expressions for the least squares estimates of the parameters, the signal at both the computation and data points, and the noise at the data points. Next we must derive the corresponding covariance matrices so that we can measure their precision. First we need an expression for the covariance matrix of the vector b.

From (2.16) in the special case of observation equations

$$b = F(x^\circ) - \ell \tag{9.49}$$

and substituting (9.26) leads to

$$b = F(x^\circ) - \bar{\ell} - n - s_1 \tag{9.50}$$

Then applying (4.16) to (9.50), whilst noting that $F(x^\circ)$ and $\bar{\ell}$ are not stochastic and that we have already assumed n and s_1 to be independent, we have

$$C_b = C_n + C_{s_1} \tag{9.51}$$

Now we substitute (9.51) in (9.45) to yield

$$\hat{x} = \left[(A^T C_b^{-1} A)^{-1} A^T C_b^{-1}\right] b \qquad (9.52)$$

Application of (4.16) to (9.52) then produces the following expression for the covariance matrix of the parameters:

$$C_{\hat{x}} = \left[(A^T C_b^{-1} A)^{-1} A^T C_b^{-1}\right] C_b \left[(A^T C_b^{-1} A)^{-1} A^T C_b^{-1}\right]^T \qquad (9.53)$$

which simplifies to

$$C_{\hat{x}} = (A^T C_b^{-1} A)^{-1} \qquad (9.54)$$

$$= \left[A^T (C_n + C_{s_1})^{-1} A\right]^{-1} \quad \text{(from (9.51))} \qquad (9.55)$$

To obtain the covariance matrix of the least squares estimate of the signal, at both the computation and data points, we substitute (9.51) and (9.52) in (9.47) to yield

$$\hat{s} = \left\{C_s B^T C_b^{-1} [I - A(A^T C_b^{-1} A)^{-1} A^T C_b^{-1}]\right\} b \qquad (9.56)$$

Then substituting (9.54) and applying (4.16) to (9.56) leads to

$$C_{\hat{s}} = \left\{C_s B^T C_b^{-1} [I - A C_{\hat{x}} A^T C_b^{-1}]\right\} C_b \left\{C_s B^T C_b^{-1} [I - A C_{\hat{x}} A^T C_b^{-1}]\right\}^T \qquad (9.57)$$

which simplifies to

$$C_{\hat{s}} = C_s B^T C_b^{-1} B C_s - C_s B^T C_b^{-1} A C_{\hat{x}} A^T C_b^{-1} B C_s \qquad (9.58)$$

Then substituting (9.51) and (9.54) in (9.58) leads to the following expression for the signal covariance matrix

$$C_{\hat{s}} = C_s B^T (C_n + C_{s_1})^{-1} B C_s - C_s B^T (C_n + C_{s_1})^{-1} A [A^T (C_n + C_{s_1})^{-1} A]^{-1} A^T (C_n + C_{s_1})^{-1} B C_s \qquad (9.59)$$

A similar treatment of (9.48) leads to the covariance matrix for the least squares estimate of the noise at the data points

$$C_{\hat{n}} = C_n (C_n + C_{s_1})^{-1} C_n - C_n (C_n + C_{s_1})^{-1} A [A^T (C_n + C_{s_1})^{-1} A]^{-1} A^T (C_n + C_{s_1})^{-1} C_n \qquad (9.60)$$

Equations (9.45), (9.47), (9.48), (9.55), (9.59) and (9.60) are the working formulae for least squares collocation.

9.3.1 Special cases

The following three special cases can be identified.

(i) Collocation without parameters

Many practical problems to which collocation is suited do not require the recovery of parameters, i.e. the system is completely modelled by the signal and the noise. For instance when using collocation to determine geoid-ellipsoid separation from gravity anomalies the only unknowns are the signal (separations) and noise (gravity anomaly measurement errors). In such circumstances A = 0 and (9.47), (9.59), (9.48) and (9.60) simplify to

$$\hat{s} = C_s B^T (C_n + C_{s_1})^{-1} b \qquad (9.61)$$

with

$$C_{\hat{s}} = C_s B^T (C_n + C_{s_1})^{-1} B C_s \qquad (9.62)$$

and

$$\hat{n} = C_n (C_n + C_{s_1})^{-1} b \qquad (9.63)$$

with

$$C_{\hat{n}} = C_n (C_n + C_{s_1})^{-1} C_n \qquad (9.64)$$

(ii) Collocation without parameters and without noise

It is not unusual to consider that the observed quantities are without error, i.e. $C_n = 0$. This may be either because the particular model is extremely insensitive to observational errors, e.g. in certain circumstances small random errors in deviation of the vertical cause insignificant errors in geoid-ellipsoid separation, or because the statistics of the observations are unknown. The least squares collocation equations for the signal (now the only unknown), (9.47) and (9.59), then simplify to

$$\hat{s} = C_s B^T C_{s_1}^{-1} b \qquad (9.65)$$

with

$$C_{\hat{s}} = C_s B^T C_{s_1}^{-1} B C_s \qquad (9.66)$$

(iii) Least squares prediction

If we further simplify the problem to the situation where the observed (without noise) quantities and the signal are the same (e.g. both gravity anomalies or both transformation distortions) then, using the notation of 9.2, we have

$$s = \left[u_1 \mid u_2 \right]^T \tag{9.67}$$

$$b = u_1 \tag{9.68}$$

$$C_{s_1} = C_{u_1} \tag{9.69}$$

and

$$C_s = \left[\begin{array}{c|c} C_{s_1} & C_{s_1 s_2} \\ \hline C_{s_2 s_1} & C_{s_2} \end{array}\right] = \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array}\right] \tag{9.70}$$

Then, using (9.33) and (9.70), we have

$$C_s B^T = \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{21} & C_{22} \end{array}\right] \left[\begin{array}{c} I \\ - \\ 0 \end{array}\right] \tag{9.71}$$

$$= \left[\begin{array}{c} C_{11} \\ \hline C_{21} \end{array}\right] = \left[C_{11} \mid C_{21} \right]^T \tag{9.72}$$

Substituting (9.72) in (9.65) yields

$$\hat{s} = \left[C_{11} \mid C_{21} \right]^T C_{11}^{-1} b \tag{9.73}$$

$$= \left[\begin{array}{c} b \\ \hline C_{21} C_{11}^{-1} b \end{array}\right] \tag{9.74}$$

Finally, substituting (9.67) and (9.68) in the left and right hand sides respectively of (9.74), we have

$$\left[\begin{array}{c} \hat{u}_1 \\ \hline \hat{u}_2 \end{array}\right] = \left[\begin{array}{c} u_1 \\ \hline C_{21} C_{11}^{-1} u_1 \end{array}\right] \tag{9.75}$$

i.e. $$\hat{u}_1 = u_1 \tag{9.76}$$

and

$$\hat{u}_2 = C_{21}C_{11}^{-1}u_1 \qquad (9.77)$$

with (9.77) being identical to (9.23). Hence we have shown least squares prediction to be a special case of least squares collocation.

9.4 Closing remarks

Least squares collocation is a large and highly advanced mathematical subject. This is exemplified by Moritz (1980), the modern standard work in physical geodesy, which devotes about 250 pages to it. The major development areas are in the determination of suitable covariance matrices for its multitudinous applications and in numerical and organizational techniques, for example step-by-step collocation (often called "stepwise" collocation), to increase its efficiency. There is currently much discussion on the proper role and interpretation of the signal covariance matrices for both collocation and simple least squares prediction. This is largely because surveyors and geodesists are used to using covariance matrices to describe observational errors and hence unknown quantities sampled from infinite populations. The use of covariance matrices to describe the variation of a finite population (e.g. gravity anomalies or height) is, however, not unknown in other sciences, for example Moritz (1978) points out that the world's human population is finite but we commonly take small samples and make inferences based on these.

Finally it is reiterated that this Working Paper is not an attempt at a full, in depth treatment of collocation theory. The objective has been to present the topic as one of a family of advanced least squares techniques, to point to some of its applications and to interpret some of the related terminology.

Appendix 1 Proof that $E(\sigma_o^2) = 1$

We define

$$\sigma_o^2 = \hat{v}^T W \hat{v}/(n-m) \tag{A1.1}$$

and wish to prove that

$$E(\sigma_o^2) = 1 \tag{A1.2}$$

We begin with a set of observation equations

$$A\hat{x} = b + \hat{v} \tag{A1.3}$$

and introduce Δ as the true observational errors and $\delta\hat{x}$ as the true errors of the least squares estimates of the parameters. Hence

$$A(\hat{x} + \delta\hat{x}) = b + \Delta \tag{A1.4}$$

Premultiply (A1.4) by $\hat{v}^T W$ to obtain

$$\hat{v}^T WA\,(\hat{x} + \delta\hat{x}) = \hat{v}^T Wb + \hat{v}^T W\Delta \tag{A1.5}$$

which rearranges to

$$\hat{v}^T W(A\hat{x} - b) + \hat{v}^T WA\delta x = \hat{v}^T W\Delta \tag{A1.6}$$

but from (3.47), $v^T WA = 0$, and substituting (A1.3) in (A1.6) gives

$$\hat{v}^T W\hat{v} = \hat{v}^T W\Delta = \Delta^T W\hat{v} \tag{A1.7}$$

Now premultiplying (A1.4) by $\Delta^T W$ gives

$$\Delta^T WA\,(\hat{x} + \delta\hat{x}) = \Delta^T Wb + \Delta^T W\Delta \tag{A1.8}$$

which rearranges to

$$\Delta^T W(A\hat{x} - b) + \Delta^T WA\delta\hat{x} = \Delta^T W\Delta \tag{A1.9}$$

Substituting (A1.3) and (A1.7) in (A1.9) gives

$$\hat{v}^T W\hat{v} + \Delta^T WA\delta\hat{x} = \Delta^T W\Delta \tag{A1.10}$$

Taking expectations gives

$$E(\hat{v}^T W\hat{v}) = E(\Delta^T W\Delta) - E(\Delta^T WA\delta\hat{x}) \tag{A1.11}$$

but

$$E(\Delta^T W \Delta) = n \text{ (for n observations)}$$

and it will be shown that

$$E(\Delta^T W A \delta\hat{x}) = m \qquad \text{(A1.13)}$$

Substituting (A1.12) and (A1.13) in (A1.11) gives

$$E(\hat{v}^T W \hat{v}/(n-m)) = 1$$

To prove (A1.13) we premultiply (A1.4) by $A^T W$ to give

$$A^T W A(\hat{x} + \delta\hat{x}) = A^T W b + A^T W \Delta \qquad \text{(A1.14)}$$

but $A^T W A\hat{x} = A^T W b$ (normal equations) so (A1.14) becomes

$$A^T W A \delta\hat{x} = A^T W \Delta$$

i.e.

$$\delta\hat{x} = (A^T W A)^{-1} A^T W \Delta \qquad \text{(A1.15)}$$

Premultiplying (A1.15) by $\Delta^T W A$ gives

$$\Delta^T W A \delta\hat{x} = \Delta^T W A (A^T W A)^{-1} A^T W \Delta$$

$$= Tr\left[A^T W \Delta\Delta^T W A \ (A^T W A)^{-1}\right] \qquad \text{(A1.16)}$$

now $E(W\Delta\Delta^T) = I$, so taking expectations in (A1.16) gives

$$E(\Delta^T W A \delta\hat{x}) = Tr\left[A^T W A \ (A^T W A)^{-1}\right]$$

$$= Tr\ (I)$$

$$= m \text{ because there are m parameters}$$

Appendix 2 Derivation of formulae for w testing and external reliability

The purpose of this appendix is to derive the formulae

3) $$\hat{w}_i = e_i^T W \hat{v} / (e_i^T W C_{\hat{v}} W e_i)^{\frac{1}{2}} \quad (5.22)$$

$$\sigma_{\hat{d}_i} = (e_i^T W C_{\hat{v}} W e_i)^{-\frac{1}{2}} \quad (5.23)$$

$$\hat{w}_i = \hat{v}_i / \sigma_{\hat{v}_i} \quad (5.25)$$

and

.4) $$\Delta\hat{\psi}_i \leq \delta_i^{\cup} \gamma_i \sigma_\psi \quad (5.36)$$

used in 5.4.1.2 and 5.4.1.3 for blunder detection and the measurement of reliability. We begin with the linearised model for the special case of observation equations, which is given in 2.3.1 as

L5) $$Ax = b + v \quad (A2.1)$$

where

$$b = \ell - F(x^o) \quad (A2.2)$$

and

$$A = \partial F / \partial x \quad (A2.3)$$

16) The least squares estimates of x and v are given in 3.1.1 as

$$\hat{x} = (A^T W A)^{-1} A^T W b \quad (A2.4)$$

and

$$\hat{v} = A\hat{x} - b \quad (A2.5)$$

$$= \left[A(A^T W A)^{-1} A^T W - I \right] b \quad (A2.6)$$

with covariance matrices given in 4.3 as

$$C_{\hat{x}} = (A^T W A)^{-1} \quad (A2.7)$$

and

$$C_{\hat{v}} = W^{-1} - A(A^T W A)^{-1} A^T \quad (A2.8)$$

Formulae (A2.2) to (A2.8) represent the real computations that would be carried out for a particular data set (irrespective of whether or not it contained a gross error). If, however, the ith observation included a gross error Δ_i then, although (A2.1) and (A2.2) were used for the computations, they would no longer describe the physical relationship between the parameters and the observations. The correct linearised model would be

$$Ax = b - e_i\Delta_i + v \qquad (A2.9)$$

where e_i is a null vector but for the ith element which is unity, i.e.

$$e_i = \left[0\ 0\ \ldots\ 1\ \ldots\ 0\ 0\right]^T \qquad (A2.10)$$

Now in order to detect gross errors Baarda (1968) has introduced a test statistic $\hat{w}_i$ given by

$$\hat{w}_i = \hat{d}_i/\sigma_{\hat{d}_i} \qquad (A2.11)$$

where

$$\hat{d}_i = \ell_i - \hat{\ell}_i^{\,c} \qquad (A2.12)$$

ℓ_i is the ith observed value

$\hat{\ell}_i^{\,c}$ is the ith observed quantity computed from the parameters derived from a least squares computation using all observations except ℓ_i

$\sigma_{\hat{d}_i}$ is the standard error of $\hat{d}_i$

It is clear that, since least squares estimates have been shown (in 6.1) to be unbiased

$$E(\hat{\ell}_i^{\,c}) = \bar{\ell}_i \qquad (A2.13)$$

where $\bar{\ell}_i$ is the true value of the ith observed quantity.

Taking expectations in (A2.12) gives

$$E(\hat{d}_i) = E(\ell_i) - E(\hat{\ell}_i^{\,c}) \qquad (A2.14)$$

but, since ℓ_i contains a gross error Δ_i, we can write

$$E(\ell_i) = \ell_i + \Delta_i \qquad (A2.15)$$

Substituting (A2.13) and (A2.15) in (A2.14) gives

$$E(\hat{d}_i) = \bar{\ell}_i + \Delta_i - \bar{\ell}_i = \Delta_i \qquad (A2.16)$$

and rewriting (A2.16) gives

$$\Delta_i = E(\hat{d}_i) = \bar{d}_i \qquad (A2.17)$$

which can be substituted into (A2.9) to yield

$$Ax + e_i\bar{d}_i - b - v = 0 \qquad \text{(A2.18)}$$

We now apply Lagrange's method to determine the least squares estimates of x, $\bar{d}_i$ and v which we write as $\hat{x}$, $\hat{d}_i$ and $\hat{v}$ respectively. Hence we have

$$\Phi = v^T W v + 2k^T(Ax + e_i\bar{d}_i - b - v) \qquad \text{(A2.19)}$$

with

$$\partial\Phi/\partial\hat{v} = 2\hat{v}^T W - 2\hat{k}^T = 0 \quad \therefore W\hat{v} - \hat{k} = 0 \qquad \text{(A2.20)}$$

$$\partial\Phi/\partial x = 2\hat{k}^T A = 0 \quad \therefore A^T\hat{k} = 0 \qquad \text{(A2.21)}$$

$$\partial\Phi/\partial d_i = 2\hat{k}^T e_i = 0 \quad \therefore e_i{}^T\hat{k} = 0 \qquad \text{(A2.22)}$$

Hence we write the least squares hypermatrix for the simultaneous solution of (A2.18), (A2.20), (A2.21) and (A2.22) as

$$\begin{bmatrix} W & -I & 0 & 0 \\ -I & 0 & A & e_i \\ 0 & A^T & 0 & 0 \\ 0 & e_i{}^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{v} \\ \hat{k} \\ \hat{x} \\ \hat{d}_i \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ 0 \\ 0 \end{bmatrix} \qquad \text{(A2.23)}$$

Application of the block elimination process, (3.21), to eliminate $\hat{v}$ from (A2.23) gives

$$\begin{bmatrix} -W^{-1} & A & e_i \\ A^T & 0 & 0 \\ e_i{}^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{k} \\ \hat{x} \\ \hat{d}_i \end{bmatrix} = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} \qquad \text{(A2.24)}$$

A similar process eliminates $\hat{k}$ from (A2.24) to give

$$\begin{bmatrix} A^TWA & A^TWe_i \\ e_i{}^TWA & e_i{}^TWe_i \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{d}_i \end{bmatrix} = \begin{bmatrix} A^TWb \\ e_i{}^TWb \end{bmatrix} \qquad \text{(A2.25)}$$

Finally the elimination of $\hat{x}$ from (A2.25) yields

$$\hat{d}_i = \left[e_i{}^TWe_i - e_i{}^TWA(A^TWA)^{-1}A^TWe_i\right]^{-1}\left[e_i{}^TW - e_i{}^TWA(A^TWA)^{-1}A^TW\right]b \qquad \text{(A2.26)}$$

We now put

$$p = e_i{}^TW \qquad \text{(A2.27)}$$

and

$$q = A(A^TWA)^{-1}A^TW \qquad \text{(A2.28)}$$

Then substitution of (A2.27) and (A2.28) in (A2.26) gives

$$\hat{d}_i = \left[pe_i - pqe_i\right]^{-1}\left[p - pq\right]b \qquad \text{(A2.29)}$$

Application of the propagation of error law, (4.16), to (A2.29) then gives

$$\sigma_{\hat{d}_i}^2 = \left[pe_i - pqe_i\right]^{-1}\left[p - pq\right]C_b\left[p^T - q^Tp^T\right]\left[pe_i - pqe_i\right]^{-1} \qquad \text{(A2.30)}$$

Then putting $C_b = W^{-1}$ and expanding (A2.30) gives

$$\sigma_{\hat{d}_i}^2 = \left[pe_i - pqe_i\right]^{-1}\left[pW^{-1}p^T - pqW^{-1}p^T - pW^{-1}q^Tp^T + pqW^{-1}q^Tp^T\right]\left[pe_i - pqe_i\right]^{-1} \qquad \text{(A2.31)}$$

The middle term of (A2.31) is now considered using (A2.27) and (A2.28)

$$\begin{aligned}\text{middle term} = {} & e_i^TWW^{-1}We_i \\ & - e_i^TWA(A^TWA)^{-1}A^TWW^{-1}We_i \\ & - e_i^TWW^{-1}WA(A^TWA)^{-1}A^TWe_i \\ & + e_i^TWA(A^TWA)^{-1}A^TWW^{-1}WA(A^TWA)^{-1}A^TWe_i \end{aligned} \qquad \text{(A2.32)}$$

which reduces to

$$\text{middle term} = e_i^TWe_i - e_i^TWA(A^TWA)^{-1}A^TWe_i \qquad \text{(A2.33)}$$

$$= pe_i - pqe_i \qquad \text{(A2.34)}$$

since the third and fourth terms of (A2.32) cancel out.

Now substituting (A2.34) into (A2.31) gives

$$\begin{aligned}\sigma_{\hat{d}_i}^2 &= \left[pe_i - pqe_i\right]^{-1} \\ &= \left[e_i^TWe_i - e_i^TWA(A^TWA)^{-1}A^TWe_i\right]^{-1}\end{aligned} \qquad \text{(A2.35)}$$

Also substituting (A2.8) in (A2.35) yields

$$\sigma_{\hat{d}_i}^2 = \left[e_i^TWC_{\hat{v}}We_i\right]^{-1} \qquad \text{(A2.36)}$$

We now substitute (A2.35) into (A2.26) to give

$$\hat{d}_i = \sigma_{\hat{d}_i}^2 \left[e_i^T W - e_i^T W A (A^T W A)^{-1} A^T W \right] b \qquad (A2.37)$$

and substituting (A2.6) in (A2.37) gives

$$\hat{d}_i = - \sigma_{\hat{d}_i}^2 e_i^T W \hat{v} \qquad (2.38)$$

and, noting that for a matrix containing a single element the inverse is simply the reciprocal, we can substitute (A2.36) into (A2.38) to give

$$\hat{d}_i = e_i^T W \hat{v} / (e_i^T W C_{\hat{v}} W e_i) \qquad (A2.39)$$

and

$$\sigma_{\hat{d}_i} = (e_i^T W C_{\hat{v}} W e_i)^{\frac{1}{2}} \qquad (A2.40)$$

which is (5.23).

Then the $\hat{w}$ statistic of (A2.11) is given by

$$\hat{w}_i = - e_i^T W \hat{v} / (e_i^T W C_{\hat{v}} W e_i)^{\frac{1}{2}} \qquad (A2.41)$$

which is (5.22) (note that the minus sign can be ignored as we are only interested in the magnitude of $\hat{w}_i$). Hence we have proved (5.22) and (5.23).

In the special case of W being a diagonal matrix (whose ith diagonal element is $1/\sigma_i^2$, where σ_i^2 is the variance of the ith observation) then $e_i^T W$ becomes a null vector but for the ith element which will be $1/\sigma_i^2$, i.e.

$$e_i^T W = \left[0 \; 0 \; \ldots \; 1/\sigma_i^2 \; \ldots \; 0 \; 0 \right]^T \qquad (A2.42)$$

In this case the numerator of (A2.41) becomes

$$- e_i^T W \hat{v} = \hat{v}_i / \sigma_i^2$$

and the denominator becomes

$$(e_i^T W C_v W e_i)^{\frac{1}{2}} = \left[(1/\sigma_i^2) \, \sigma_{\hat{v}_i}^2 \, (1/\sigma_i^2) \right]^{\frac{1}{2}} = \sigma_{\hat{v}_i} / \sigma_i^2 \qquad (A2.43)$$

Then (A2.41) becomes

$$\hat{w}_i = \hat{v}_i / \sigma_{\hat{v}_i} \qquad (A2.44)$$

where $\sigma_{\hat{v}_i}$ is the standard error of the ith residual $\hat{v}_i$; hence (5.25) is proved.

To derive (5.36) we proceed as follows.

Let $\hat{\psi}$ be computed from the least squares estimates of the parameters by the linear transformation

$$\hat{\psi} = s^T \hat{x} \tag{A2.45}$$

(A2.4) and (A2.7) are combined to yield

$$\hat{x} = C_{\hat{x}} A^T W b \tag{A2.46}$$

and if the ith observation includes a gross error Δ_i its effect on the parameters will be $\Delta\hat{x}_i$ where

$$\Delta\hat{x}_i = C_{\hat{x}} A^T W \Delta b_i \tag{A2.47}$$

and

$$\Delta b_i = \left[0\ 0\ 0\ \ldots\ \Delta_i\ \ldots\ 0\ 0\right]^T \tag{A2.48}$$

Similarly, using (A2.45), we see that its effect on $\hat{\psi}$, $\Delta\hat{\psi}_i$, will be

$$\Delta\hat{\psi}_i = s^T C_{\hat{x}} A^T W \Delta b_i \tag{A2.49}$$

If W is diagonal (i.e. if we assume the observations to be uncorrelated), (A2.49) can be written, using (A2.48), as

$$\Delta\hat{\psi}_i = s^T C_{\hat{x}} a_i w_i \Delta_i \tag{A2.50}$$

where a_i is the ith column of A and w_i is the ith diagonal element of W. Note that

$$w_i = 1/\sigma_i^2 \tag{A2.51}$$

Now consider the product $s^T C_{\hat{x}} a_i$ in (A2.50),

$$s^T C_{\hat{x}} a_i = s^T E \Delta E^T a_i \tag{A2.52}$$

where E is a matrix of the eigenvectors of $C_{\hat{x}}$ and D is a diagonal matrix whose elements are the positive eigenvalues of $C_{\hat{x}}$. Note this decomposition is always possible with symmetrical positive-definite matrices (see Fox (1964)).

Putting

$$Z = E D^{\frac{1}{2}} \tag{A2.53}$$

in (A2.52) gives

$$s^T C_{\hat{x}} a_i = s^T Z Z^T a_i \tag{A2.54}$$

Furthermore if we define

$$p = Z^T s \tag{A2.55}$$

and

$$q = Z^T a_i \tag{A2.56}$$

then (A2.54) simplifies to

$$s^T C_{\hat{x}} a_i = p^T q \tag{A2.57}$$

Now for any vectors p and q

$$p^T q / \left[(p^T p)^{\frac{1}{2}} (q^T q)^{\frac{1}{2}} \right] \leq 1 \tag{A2.58}$$

because the left hand side of (A2.58) is the cosine of the angle between the two vectors (as defined in n-dimensional space).

Rearranging (A2.58) we have

$$p^T q \leq (p^T p)^{\frac{1}{2}} (q^T q)^{\frac{1}{2}} \tag{A2.59}$$

Then substituting (A2.55), (A2.56) and (A2.57) in (A2.59) yields

$$s^T C_{\hat{x}} a_i \leq (s^T Z Z^T s)^{\frac{1}{2}} (a_i^T Z Z^T a_i)^{\frac{1}{2}} \tag{A2.60}$$

Now, noting from (A2.52) and (A2.53) that

$$Z Z^T = C_{\hat{x}} \tag{A2.61}$$

we rewrite (A2.60) as

$$s^T C_{\hat{x}} a_i \leq (s^T C_{\hat{x}} s)^{\frac{1}{2}} (a_i^T C_{\hat{x}} a_i)^{\frac{1}{2}} \tag{A2.62}$$

Then postmultiplying both sides of (A2.62) by $w_i \Delta_i$ yields

$$s^T C_{\hat{x}} a_i w_i \Delta_i \leq (s^T C_{\hat{x}} s)^{\frac{1}{2}} (a_i^T C_{\hat{x}} a_i)^{\frac{1}{2}} w_i \Delta_i \tag{A2.63}$$

Now we can substitute (A2.50) and (A2.51) in (A2.63), whilst noticing that

$$\sigma_{\hat{\psi}}^2 = s^T C_{\hat{x}} s \tag{A2.64}$$

and

$$\hat{\sigma}_i^2 = a_i^T C_{\hat{x}} a_i \tag{A2.65}$$

and rearrange to yield

$$\Delta \hat{\psi}_i \leq (\hat{\sigma}_i \Delta_i / \sigma_i^2) \sigma_{\hat{\psi}} \tag{A2.66}$$

Now if the size of the blunder Δ_i is equal to the boundary value as defined in 5.4.1.2, i.e.

$$\Delta_i = \Delta_i^u \tag{A2.67}$$

we can write, from (5.32) and (5.33)

$$\Delta_i = \delta_i^u \sigma_i^2 / \sigma_{\hat{v}_i} \tag{A2.68}$$

Then substituting (A2.68) in (A2.66) gives

$$\Delta\hat{\psi} \leq \delta_i^u \frac{\hat{\sigma}_i}{\hat{\sigma}_{v_i}} \sigma_{\hat{\psi}} \tag{A2.69}$$

and, from (5.37), the bracketed term in (A2.69) is equal to γ_i, so we can write

$$\Delta\hat{\psi} \leq \delta_i^u \gamma_i \sigma_{\hat{\psi}} \tag{A2.70}$$

which is (5.36), which was required to be proved.

Appendix 3 Numerical examples

A3.1 Combined case

Photographic observations to a satellite have produced the following five observations of time and altitude with standard errors of 0.001 sec and 2".0 respectively.

Time			Altitude		
18 hrs	04 min	10.1152 sec	45°	17'	40".1
18	04	15.2370	47	51	25.6
18	04	20.2220	50	20	58.2
18	04	24.9580	52	42	59.6
18	04	29.7820	55	07	43.9

It is required to determine

(i) the least squares estimates of the slope and intercept of the straight line in which the satellite is assumed to be moving during the period of observation

(ii) the standard errors of these estimates

(iii) the least squares estimate of the altitude of the satellite at 18 hrs 04 min 22.0 sec and its standard error.

The basic mathematical model for the ith observed time and altitude is written as

$$f_i(\bar{x}, \bar{\ell}) = \bar{x}_1 \bar{t}_i + \bar{x}_2 - \bar{\alpha}_i = 0 \qquad \text{(A3.1)}$$

where

$$\bar{x} = [\bar{x}_1 \quad \bar{x}_2]^T, \quad \bar{x}_1 = \text{true value of the slope}, \quad \bar{x}_2 = \text{true value of the intercept}$$

$$\bar{\ell} = [\bar{t}_1 \quad \bar{\alpha}_1 \quad \bar{t}_2 \quad \bar{\alpha}_2 \quad \ldots \quad \bar{t}_5 \quad \bar{\alpha}_5]^T$$

with $\bar{t}_i$ and $\bar{\alpha}_i$ being the true values of the ith time and altitude respectively.

We note immediately that because (A3.1) cannot be simplified to

$$f_i(\bar{x}) = \bar{\ell}_i$$

or

$$f_i(\bar{\ell}) = 0$$

the problem requires the use of the combined least squares method. Also the problem is non-linear and hence approximate values of $\bar{x}$ are required. For convenience we reduce the number of digits in the problem by subtracting 18 hrs 4 min 10 sec from the times and 45° from the altitudes and then express the observations in units of seconds of time and arc respectively, viz.

$$\begin{aligned} t_1 &= 0.1152 & \alpha_1 &= 1060.1 \\ t_2 &= 5.2370 & \alpha_2 &= 10285.6 \\ t_3 &= 10.2220 & \alpha_3 &= 19258.2 \\ t_4 &= 14.9580 & \alpha_4 &= 27779.6 \\ t_5 &= 19.7820 & \alpha_5 &= 36463.9 \end{aligned}$$

Approximate values are computed as follows

from f_5, putting $\bar{x}_2 = 0$, $x_1^o = 1843$

from f_1, putting $\bar{x}_1 = 1843$, $x_2^o = 850$

Then, using (2.17), (2.18), (2.16) and (3.7) respectively, we obtain matrices A, C, b and W as follows:

$$A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ t_3 & 1 \\ t_4 & 1 \\ t_5 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} x_1^o & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x_1^o & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_1^o & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_1^o & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_1^o & -1 \end{bmatrix}$$

$$b = \begin{bmatrix} \alpha_1 - x_1^o t_1 - x_2^o \\ \alpha_2 - x_1^o t_2 - x_2^o \\ \cdot \\ \cdot \\ \cdot \\ \alpha_5 - x_1^o t_5 - x_2^o \end{bmatrix} \qquad W^{-1} = \begin{bmatrix} \sigma_{t_1}^2 & & & & & & & & & 0 \\ & \sigma_{\alpha_1}^2 & & & & & & & & \\ & & \sigma_{t_2}^2 & & & & & & & \\ & & & \sigma_{\alpha_2}^2 & & & & & & \\ & & & & \cdot & & & & & \\ & & & & & \cdot & & & & \\ & & & & & & \cdot & & & \\ & & & & & & & \cdot & & \\ & & & & & & & & \sigma_{t_5}^2 & \\ 0 & & & & & & & & & \sigma_{\alpha_5}^2 \end{bmatrix}$$

Substituting the observations and their standard errors along with the approximate values of the parameters then leads to

$$A = \begin{bmatrix} 0.1152 & 1 \\ 5.2370 & 1 \\ 10.2220 & 1 \\ 14.9580 & 1 \\ 19.7820 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} 1843 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1843 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1843 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1843 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1843 & -1 \end{bmatrix}$$

$$b = \begin{bmatrix} -2.21 \\ -216.19 \\ -430.95 \\ -637.99 \\ -844.33 \end{bmatrix} \qquad W^{-1} = \begin{bmatrix} 10^{-6} & & & & & & & & & 0 \\ & 4 & & & & & & & & \\ & & 10^{-6} & & & & & & & \\ & & & 4 & & & & & & \\ & & & & 10^{-6} & & & & & \\ & & & & & 4 & & & & \\ & & & & & & 10^{-6} & & & \\ & & & & & & & 4 & & \\ & & & & & & & & 10^{-6} & \\ & & & & 0 & & & & & 4 \end{bmatrix}$$

Using (3.25) we obtain the least squares estimate of x as follows

$$A^T(CW^{-1}C^T)^{-1}A = \begin{bmatrix} 100.9914 & 6.8023 \\ 6.8023 & 0.6760 \end{bmatrix}$$

$$A^T(CW^{-1}C^T)^{-1}b = \begin{bmatrix} -4296.9727 \\ -288.1940 \end{bmatrix}$$

$$\left[A^T(CW^{-1}C^T)^{-1}A\right]^{-1} = \begin{bmatrix} 0.030729 & -0.309211 \\ -0.309211 & 4.590747 \end{bmatrix}$$

$$\hat{x} = \left[A^T(CW^{-1}C^T)^{-1}A\right]^{-1} A^T(CW^{-1}C^T)^{-1}b = \begin{bmatrix} -42.929 \\ 5.645 \end{bmatrix}$$

Hence the least squares estimate of the slope

$$= x_1^o + \hat{x}_1 = 1843 - 42.929 = 1800.07 \text{ "/sec}$$

and the least squares estimate of the intercept

$$= x_2^o + \hat{x}_2 = 850 + 5.645 = 855\text{"}64$$

In practice we would now take a new x^o equal to the above estimates and repeat the computation until $\hat{x}$ was insignificantly different from a null vector.

We use (4.43) to give the standard errors of the least squares estimates of the slope and intercept respectively as follows

$$\sigma_{\hat{x}_1} = (0.030729)^{\frac{1}{2}} = 0.18 \text{ "/sec}$$

$$\sigma_{\hat{x}_2} = (4.590747)^{\frac{1}{2}} = 2\text{"}.14$$

The least squares estimate of the altitude of the satellite at 18 hrs 04 min 22.0 sec and its standard error are obtained as follows.

When time = 18 hrs 04 min 22.0 sec , t = 12

$\therefore$ azimuth = $45^\circ + \{12 \times 1800.07 + 855.64\}$"

= $51^\circ\ 14'\ 16\text{"}.49$

In general $\hat{\alpha} = 45^\circ + t\hat{x}_1 + \hat{x}_2$

$= \alpha_o + a\hat{x}$

where $a = [t \quad 1]$

$\therefore \quad \sigma^2_{\hat{\alpha}} = a\, C_{\hat{x}}\, a^T$

$$= [12 \quad 1] \begin{bmatrix} 0.030729 & -0.309211 \\ -0.309211 & 4.590747 \end{bmatrix} \begin{bmatrix} 12 \\ 1 \end{bmatrix}$$

$\therefore \quad \sigma_{\hat{\alpha}} = 1\text{"}.26$

It should be noted that this example could actually be solved in a simpler way without use of the combined least squares model. The procedure adopted here has been chosen to exemplify the general approach to combined least squares problems rather than to show an efficient procedure for curve fitting problems.

A3.2 Special case of observation equations

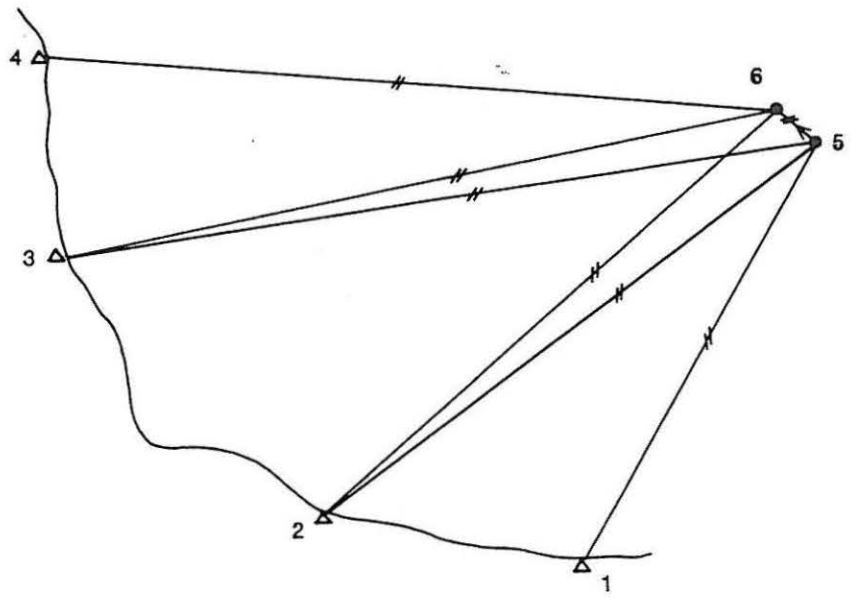


Fig. A3.1

Two nearby offshore platforms are to be fixed from the following measurements, where the numbers in parentheses are their standard errors.

Distances

1	– 5	87921.2 m	(5 m)
2	– 5	114948.5	(5 m)
3	– 5	147802.4	(5 m)
2	– 6	114737.8	(5 m)
3	– 6	146667.7	(5 m)
4	– 6	146156.3	(5 m)
5	– 6	1981.81	(0.02 m)

Azimuth

5 – 6 316° 18' 05".7 (3")

Position

Eastings of 5: 255086.5 m (3 m)
Northings of 5: 964173.1 (3 m)

Given that stations 1, 2, 3 and 4 have known coordinates of

	E	N
1	216498.72 m	885174.98 m
2	163304.56	894962.77
3	108791.23	943117.05
4	109007.10	986075.53

carry out the following.

(i) Compute the least squares estimates of the coordinates of stations 5 and 6.

(ii) Compute the least squares residuals and test for any gross errors in the observations.

(iii) Estimate the precision of the position-fix by computing the following:

(a) the absolute error ellipses at stations 5 and 6

(b) the relative error ellipse between stations 5 and 6

(c) the standard errors of all azimuths and distances computed using the least squares estimates of the coordinates of stations 5 and 6

(d) the standard error of the angle subtended at station 5 between stations 6 and 2 computed using the least squares estimates of the coordinates of these stations.

(iv) Assess the internal and external reliability of each observation by determining its τ and γ factors. Compute the probability of accepting each observation with a gross error of four times its standard error if rejection is carried out using the w statistic with a 5% significance level. Also compute the effects of undetected blunders of four times their standard errors in each observation on the least squares estimate of the derived azimuth between stations 5 and 6.

This problem contains three different types of observation: distance, azimuth and position. The first stage is to derive the general form of each observation equation.

Distance between stations i and j

Functional model $f(\bar{x}) = \bar{\ell}$:

$$\left\{(\bar{E}_i - \bar{E}_j)^2 + (\bar{N}_i - \bar{N}_j)^2\right\}^{\frac{1}{2}} = \bar{d}_{ij}$$

$\therefore$ linearised model $Ax = b + v$:

$$\frac{\partial f}{\partial E_i} dE_i + \frac{\partial f}{\partial N_i} dN_i + \frac{\partial f}{\partial E_j} dE_j + \frac{\partial f}{\partial N_j} dN_j = (d_{ij} - d^o_{ij}) + v$$

where

$$d^o_{ij} = \left\{(E^o_i - E^o_j)^2 + (N^o_i - N^o_j)^2\right\}^{\frac{1}{2}}$$

Differentials are evaluated (at $\bar{x} = x^o$) as follows:

$$f = \left\{(E_i - E_j)^2 + (N_i - N_j)^2\right\}^{\frac{1}{2}}$$

$$\frac{\partial f}{\partial E_i} = \tfrac{1}{2}\left\{(E^o_i - E^o_j)^2 + (N^o_i - N^o_j)^2\right\}^{-\frac{1}{2}} 2(E^o_i - E^o_j)\ 1$$

$$= (E^o_i - E^o_j)/d^o_{ij}$$

Similarly

$$\frac{\partial f}{\partial N_i} = (N^o_i - N^o_j)/d^o_{ij}$$

$$\frac{\partial f}{\partial E_j} = -(E^o_i - E^o_j)/d^o_{ij}$$

$$\frac{\partial f}{\partial N_j} = -(N^o_i - N^o_j)/d^o_{ij}$$

<u>Azimuth from station i to station j</u>

Functional model $f(\bar{x}) = \bar{\ell}$:

$$\tan^{-1}\left\{(E_j - E_i)/(N_j - N_i)\right\} = \alpha_{ij}$$

$\therefore$ linearised model $Ax = b + v$:

$$\frac{\partial f}{\partial E_i} dE_i + \frac{\partial f}{\partial N_i} dN_i + \frac{\partial f}{\partial E_j} dE_j + \frac{\partial f}{\partial N_j} dN_j = (\alpha_{ij} - \alpha^o_{ij}) + v$$

where

$$\alpha^o_{ij} = \tan^{-1}\left\{(E^o_j - E^o_i)/(N^o_j - N^o_i)\right\}$$

Differentials are evaluated (at $\bar{x} = x^o$) as follows:

$$f = \tan^{-1}\left\{(E_j - E_i)/(N_j - N_i)\right\}$$

Note that if $y = \tan^{-1}x$, $\partial y/\partial x = 1/(1 + x^2)$; hence

$$\frac{\partial f}{\partial E_i} = \left\{1/[1 + (E_j - E_i)^2/(N_j - N_i)^2]\right\}[-1/(N_j - N_i)]$$

$$= -(N_j - N_i)/d_{ij}^{o2}$$

$$\frac{\partial f}{\partial N_i} = \left\{1/[1 + (E_j - E_i)^2/(N_j - N_i)^2]\right\}[-1\ (E_j - E_i)(N_j - N_i)^{-2} - 1]$$

$$= (E_j - E_i)/d_{ij}^{o2}$$

Similarly

$$\frac{\partial f}{\partial E_j} = (N_j - N_i)/d_{ij}^{o2}$$

$$\frac{\partial f}{\partial N_j} = -(E_j - E_i)/d_{ij}^{o2}$$

Note that in practice the above four differentials are usually multiplied by 206264"8 to convert the dimensions of the linearised model from radians to seconds.

Position of station i

Observed eastings and northings will lead to two equations.

(i) For eastings functional model is

$$\bar{E}_i = \bar{\ell}_i$$

where $\bar{\ell}_i$ is the true value of the observed eastings, this is "linearised" (actually it is, of course, already linear) to

$$dE_i = \ell_i - \ell_i^o$$

where ℓ_i and ℓ_i^o are the observed and approximate eastings.

(ii) Similarly for northings

$$dN_i = \ell_i - \ell_i^o$$

Now returning to the problem in hand we have ten observations and four parameters; using the usual notation we can write

$$\ell = \begin{bmatrix} d_{15} & d_{25} & d_{35} & d_{26} & d_{36} & d_{46} & d_{56} & \alpha_{56} & E_5 & N_5 \end{bmatrix}^T$$

$$x = \begin{bmatrix} dE_5 & dN_5 & dE_6 & dN_6 \end{bmatrix}^T$$

and we will have ten observation equations of the form

$$f_i(\bar{x}) = \bar{\ell}_i$$

-1]

Then using the following for approximate values of $\bar{x}$, i.e. x^o

	E	N
5	255086.5 m	964173.1 m
6	253717.3	965605.9

and using the general forms of the necessary differentials already developed we can write down directly the observation equations (2.22) as follows

$$\begin{bmatrix} 0.439 & 0.899 & 0 & 0 \\ 0.798 & 0.602 & 0 & 0 \\ 0.990 & 0.142 & 0 & 0 \\ 0 & 0 & 0.788 & 0.616 \\ 0 & 0 & 0.988 & 0.153 \\ 0 & 0 & 0.990 & -0.140 \\ 0.691 & -0.723 & -0.691 & 0.723 \\ -75.248 & -71.904 & 75.248 & -71.904 \\ 1.0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 \end{bmatrix} \begin{bmatrix} dE_5 \\ dN_5 \\ dE_6 \\ dN_6 \end{bmatrix} = \begin{bmatrix} 2.38 \\ -3.64 \\ -0.39 \\ -0.67 \\ 7.16 \\ 5.53 \\ -0.01 \\ 4.69 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \\ v_{10} \end{bmatrix}$$

ed"

with a ten by ten weight matrix whose diagonal elements are
0.04, 0.04, 0.04, 0.04, 0.04, 0.04, 2500, 0.11, 0.11, 0.11.

Then, using (3.41), we write the normal equations as

$$\hat{x} = N^{-1}d$$

where

$$N = \begin{bmatrix} 182.257 & -64.750 & -182.238 & 64.754 \\ -64.750 & 188.136 & 64.754 & -188.120 \\ -182.238 & 64.754 & 182.249 & -64.752 \\ 64.754 & -188.120 & -64.752 & 188.122 \end{bmatrix}$$

and

$$d = \begin{bmatrix} -62.661 & -13.033 & 63.052 & 13.025 \end{bmatrix}^T$$

with solution

$$\hat{x} = \begin{bmatrix} 1.47 & -0.56 & 1.51 & -0.54 \end{bmatrix}^T$$

Hence the least squares estimates of the coordinates of stations 5 and 6 are

$$E_5 = 255086.5 + 1.47 = 255087.97$$
$$N_5 = 964173.1 - 0.56 = 964172.54$$
$$E_6 = 253717.3 + 1.51 = 253718.81$$
$$N_6 = 965605.9 - 0.54 = 965605.36$$

In practice we would now take a new x^o equal to the above estimates and repeat the computation until $\hat{x}$ was insignificantly different from a null vector.

The least squares residuals are computed from (3.43)

$$\hat{v} = A\hat{x} - b = \begin{bmatrix} -2.23 \\ 4.47 \\ 1.76 \\ 1.52 \\ -5.75 \\ -3.96 \\ 0.00 \\ 0.02 \\ 1.47 \\ -0.56 \end{bmatrix}$$

and the unit variance is given, from (4.25), as

$$\sigma_o^2 = \hat{v}^T W \hat{v}/(n-m) = 0.573$$

To check if this is significantly different from unity we compute, following 5.4.2.2, the F statistic

$$F_{\infty,6} = 1/\sigma_o^2 = 1.745 \text{ (note that there are 6 degrees of freedom)}$$

Then using a level of significance of 0.10 we find that, from Table 5.5, the critical value of F is 2.10 so we can accept the hypothesis that the unit variance is not significantly different from unity (note that Table 5.5 is entered with $\alpha = 0.05$ as we are carrying out a two-tailed test).

To test for blunders we compute the covariance matrix of the residuals, $C_{\hat{v}}$ from (4.68)

$$C_{\hat{v}} = W^{-1} - AN^{-1}A^T$$

and extract the square roots of the diagonal elements to obtain

$$\sigma_{\hat{v}_1}, \sigma_{\hat{v}_2} \cdots \sigma_{\hat{v}_{10}}.$$

These are listed below along with the w statistic computed from (5.25)

observation		residual ($\hat{v}$)	$\sigma_{\hat{v}}$	w
distance	1 - 5	-2.23 m	4.50 m	0.50
	2 - 5	4.47	4.63	0.96
	3 - 5	1.76	4.63	0.38
	2 - 6	1.52	4.63	0.33
	3 - 6	-5.75	4.63	1.24
	4 - 6	-3.96	4.55	0.87
	5 - 6	0.00	0.00	1.54*
azimuth	5 - 6	0".02	0".02	0.97
eastings	5	1.47 m	2.30 m	0.64
northings	5	-0.56	1.70	0.33

*Note that the computations have been carried out to more digits than have been displayed.

To test for blunders in the observations we first choose a level of significance, say 0.01 (99% confidence level), and obtain a critical value of w from the normal distribution tables (Table 5.2) of 2.57. Since all values of the w statistic are less than 2.57 we can accept, with a 0.01 probability of making a type I error, the null hypothesis that none of the observations contains a gross error. Alternatively we could compute the tau statistics from (5.28). The largest is

$$\tau_7 = 1.54/(0.573)^{\frac{1}{2}} = 2.03$$

which is less than the critical value of 2.33 given in Table 5.6 for 10 observations and 6 degrees of freedom. Hence we reach the same conclusion and do not reject any observations as blunders at the 0.01 level of significance.

The precision analysis now proceeds as follows. We first compute the covariance matrix of the parameters using (4.67), i.e. by inverting the normal equations matrix

$$C_{\hat{x}} = N^{-1} = \begin{bmatrix} 3.76330 & -1.29788 & 3.76307 & -1.29799 \\ -1.29788 & 6.14226 & -1.29789 & 6.14222 \\ 3.76307 & -1.29789 & 3.76345 & -1.29779 \\ -1.29799 & 6.14222 & -1.29779 & 6.14278 \end{bmatrix} \quad \text{(A3.2)}$$

Absolute error ellipses

The absolute error ellipse at station 5 is obtained as follows. From (4.76)

$$\tan 2\psi_m = 2(-1.298)/(6.142-3.763) = -1.091$$

$$\therefore \quad \psi_m = 66^\circ.3 \text{ or } 156^\circ.3$$

Then using (4.74) and (4.75)

$$\sigma_p^2 = \cos^2 66.3\ 6.142 + \sin^2 66.3\ 3.763 + 2 \cos 66.3 \sin 66.3\ (-1.298)$$

$$\therefore \quad \sigma_p = 1.787$$

and

$$\sigma_q^2 = \cos^2 156.3\ 6.142 + \sin^2 156.3\ 3.763 - 2 \cos 156.3 \sin 156.3\ (-1.2$$

$$\therefore \quad \sigma_q^2 = 2.591$$

Hence σ_q is the major axis (it is larger than σ_p) and the error ellipse is as follows

$$\sigma_{max} = 2.59\,m, \quad \sigma_{min} = 1.78\,m, \quad \psi_{max} = 156^\circ$$

Similarly the error ellipse for station 6 is

$$\sigma_{max} = 2.59\,m, \quad \sigma_{min} = 1.78\,m, \quad \psi_{max} = 156^\circ$$

Relative error ellipse between stations 5 and 6

Using (4.106) we have

$$\sigma_x^2 = 6.14278 + 6.14226 - 2(6.14222) = 0.00060$$

$$\sigma_y^2 = 3.76345 + 3.76330 - 2(3.76307) = 0.00061$$

$$\sigma_{xy} = (-1.29788) - (-1.29789) - (-1.29799) + (-1.29779) = 0.00021$$

Then applying (4.76), (4.74) and (4.75) gives the relative error ellipse as

$$\sigma_{max} = 0.029, \quad \sigma_{min} = 0.020, \quad \psi_{max} = 46^\circ$$

Standard errors of derived quantities

For the distance d between stations 5 and 6, for example, the vector $\hat{q}$ in (4.107) becomes a single element, say $\hat{d}$, where

$$\hat{d} = [(\hat{E}_6 - \hat{E}_5)^2 + (\hat{N}_6 - \hat{N}_5)^2]^{\frac{1}{2}}$$

and the matrix B in (4.109) is the vector b, where

.298)

$$b = \left[\frac{\partial \hat{d}}{\partial E_5}, \frac{\partial \hat{d}}{\partial N_5}, \frac{\partial \hat{d}}{\partial E_6}, \frac{\partial \hat{d}}{\partial N_6}\right]$$

$$= \left[\frac{\Delta\hat{E}}{\hat{d}}, \frac{\Delta\hat{N}}{\hat{d}}, -\frac{\Delta\hat{E}}{d}, -\frac{\Delta\hat{N}}{d}\right]$$

where

(-1.2

$$\Delta\hat{E} = \hat{E}_5 - \hat{E}_6 \text{ and } \Delta\hat{N} = \hat{N}_5 - \hat{N}_6$$

Substituting the least squares solution for $\Delta\hat{E}$, $\Delta\hat{N}$ and $\hat{d}$ we obtain

$$b = \left[0.691 \quad -0.723 \quad -0.691 \quad 0.723\right]$$

and using (4.110) with $C_{\hat{x}}$ as given in (A3.2) we obtain

$$\sigma^2_{\hat{d}} = 0.00040 = (0.020)^2$$

Hence the standard error of the least squares estimate of the distance 5 to 6 (as derived from the least squares estimates of the coordinates of stations 5 and 6) is 0.020 m.

Similarly the following standard errors of distances and azimuths have been computed after linearising the relationships

160

$$\hat{\alpha} = \tan^{-1}[(\hat{E}_j - \hat{E}_i)/(\hat{N}_j - \hat{N}_i)]$$

161

and

121

$$\beta(\text{angle 6-5-2}) = \tan^{-1}[(\hat{E}_6 - \hat{E}_5)/(\hat{N}_6 - \hat{N}_5)] - \tan^{-1}[(E_2 - \hat{E}_5)/(N_2 - \hat{N}_5)]$$

The results are tabulated below.

Stations from	Stations to	Standard error of distance	Standard error of azimuth
1	5	2.16 m	5".4
2	5	1.84	4.6
3	5	1.86	3.6
4	5	2.05	3.3
1	6	2.18	5.3
2	6	1.85	4.6
3	6	1.85	3.6
4	6	2.04	3.4
5	6		

Furthermore the standard error of the derived angle at station 5 between stations 6 and 2 is 5".5.

Reliability

Consider the first observation, i.e. the distance between stations 1 and 5. Then using (5.33) with i = 1 we have

$$\tau_1 = \sigma_1 / \sigma_{\hat{v}_1}$$

where $\sigma_1 = 5.0$ m (the given standard error)

and $\sigma_{\hat{v}_1} = 4.50$ m (already computed and listed earlier with the residuals)

then $\tau_1 = 5.0/4.50 = 1.11$

Now to consider the probability of making a type II error and accepting a gross error of four times its standard error (i.e. 20 m) we proceed as follows. Using (5.32)

$$20 = \delta_i^U \, 5.0 \, 1.11$$

giving

$$\delta_i^U = 3.60$$

But from (5.20) with $\alpha = 5\%$ and a = 1.96 we have

$$3.60 = 1.96 + b$$

giving

$$b = 1.64$$

Then from Table 5.2 we see that a value of b of 1.64 corresponds to a probability of 0.9495 (approximately 0.95). Hence we are 95% sure of rejecting the first observation if it has a blunder greater than or equal to four times its standard error. Conversely there is only a 5% chance that an accepted observation has a gross error greater than or equal to four times its standard error.

The effect of an undetected blunder in the first observation on the derived azimuth between stations 1 and 5 is investigated as follows. First we compute γ_1 from (5.40)

$$\gamma_1^2 = \tau_1^2 - 1$$

$$= (1.11)^2 - 1$$

giving

$$\gamma_1 = 0.48$$

Then, using (5.36) and the already computed standard error of the azimuth from 1 to 5,

$$\Delta\hat{\psi} = 3.6\ 0.48\ 5''.4$$

$$= 9''.3$$

Hence the maximum effect on the derived azimuth 1 to 5 of a blunder in the first observation of four times its standard error is 9".3.

Similar computations with all other observations lead to the following table, in which the observations are listed in the same order as they were given in the problem. Note that P is the probability of rejecting an observation with a blunder greater than or equal to four times its standard error and $\Delta\hat{\psi}$ is the maximum effect of such a blunder (if it was undetected) on the azimuth between stations 1 and 5.

Observation	τ factor	γ factor	P	$\Delta\hat{\psi}$
1	1.11	0.48	0.95	9".3
2	1.08	0.40	0.96	8.0
3	1.08	0.40	0.96	8.0
4	1.08	0.40	0.96	8.0
5	1.08	0.40	0.96	8.0
6	1.10	0.45	0.95	8.8
7	308.90	308.90	0.03	20.0
8	148.87	148.87	0.03	21.7
9	1.31	0.85	0.86	14.0
10	1.77	1.47	0.62	17.9

From the above we can draw the obvious conclusion that measurements 7 and 8 have extremely poor (effectively non-existent) internal reliability and rather poor external reliability, i.e. large blunders in these observations may go undetected and seriously affect the resulting least squares estimates.

A3.3 Special case of condition equations

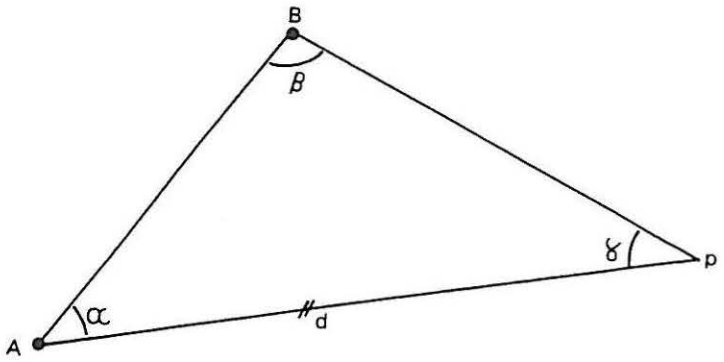


Fig. A3.2

The following observations are made in the above triangle:

angle	α	40°	18' 16"	(5")
	β	106	54 21	(5")
	γ	32	47 40	(5")
distance	d	625.64 m		(0.05 m)

Given the coordinates of A and B as

	E	N
A	10417.62	55061.78
B	10645.28	55333.09

find the least squares estimates of the coordinates of P and their standard errors.

Note that this problem is, in practice, more easily treated using the special case of observation equations and the reader is invited to do this and check that the solution is identical to what follows. Here, in order to show how the condition equations method can be applied to two-dimensional position-fixing problems, it will be considered as an example of the special case of condition equations.

The vector of observations ℓ is given by

$$\ell = \left[\alpha\ \beta\ \gamma\ d\right]^T$$

Since there are 4 observations and 2 parameters (eastings and northings of P) we see from the table in 2.4 that there will be (4–2) = 2 condition equations; putting the known distance AB as a, we can write these:

$$\bar{\alpha} + \bar{\beta} + \bar{\gamma} - 180^\circ = 0$$

$$\bar{d}/\sin\bar{\beta} - a/\sin\bar{\gamma} = 0$$

From (2.18) we write the matrix C as

$$C = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & -\mu\, d \cot\beta \operatorname{cosec}\beta & \mu\, a \cot\gamma \operatorname{cosec}\gamma & \operatorname{cosec}\beta \end{bmatrix}$$

where $\mu = 1/206265''$ and converts the units to agree with W^{-1}; C is evaluated as

$$C = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0.0009635 & 0.0049203 & 1.0452 \end{bmatrix}$$

with a = 354.173

The vector b is given by (2.16) as

$$b = -\begin{bmatrix} \alpha + \beta + \gamma - 180 \\ d/\sin\beta - a/\sin\gamma \end{bmatrix}$$

which is evaluated as

$$b = \begin{bmatrix} -17 \\ 0.007 \end{bmatrix}$$

Also from (3.7) we have

$$W^{-1} = \begin{bmatrix} 25 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 0.0025 \end{bmatrix}$$

From 3.2.2 we write the normal equations as

$$(CW^{-1}C^T)\hat{k} = -b$$

with

$$(CW^{-1}C^T) = \begin{bmatrix} 75 & 0.1471 \\ 0.1471 & 0.003359 \end{bmatrix}$$

and

$$-b = \begin{bmatrix} 17 \\ -0.007 \end{bmatrix}$$

which leads to a solution for $\hat{k}$

$$\hat{k} = -\begin{bmatrix} 0.2525 \\ -13.14 \end{bmatrix}$$

Then using (3.44)

$$\hat{v} = \begin{bmatrix} -6.3 \\ -6.0 \\ -4.7 \\ 0.034 \end{bmatrix}$$

and using (4.50)

$$\hat{\ell} = \ell + \hat{v}$$

we obtain

$$\begin{aligned} \hat{\alpha} &= 40^\circ\ 18'\ 09''.7 \\ \hat{\beta} &= 106\ \ 54\ \ 15.0 \\ \hat{\gamma} &= 32\ \ 47\ \ 35.3 \\ \hat{d} &= 625.674\ \text{m} \end{aligned}$$

which are the least squares estimates of the observed quantities.

Now if we let θ be the bearing of the line AB computed from the known coordinates of A and B we have

$$\theta = 40^\circ\ 00'\ 1\overset{''}{.}73$$

and the least squares estimates of the eastings and northings of P are

$$\hat{y} = \begin{bmatrix} \hat{E}_p \\ \hat{N}_p \end{bmatrix} = \begin{bmatrix} E_A + \hat{d}\sin(\theta + \hat{\alpha}) \\ N_A + \hat{d}\cos(\theta + \hat{\alpha}) \end{bmatrix} = \begin{bmatrix} 11034.35 \\ 55167.17 \end{bmatrix}$$

To compute the standard errors of $\hat{E}_p$ and $\hat{N}_p$ we must first compute $C_{\hat{\ell}}$ using (4.71). This involves some lengthy matrix manipulations using the matrices C and W^{-1} with the result

$$C_{\hat{\ell}} = \begin{bmatrix} 15.881 & -8.734 & -7.154 & 0.0416 \\ -8.734 & 16.462 & -7.735 & 0.0212 \\ -7.154 & -7.735 & 14.882 & -0.0629 \\ 0.0416 & 0.0212 & -0.0629 & 0.000281 \end{bmatrix}$$

Then in order to apply (4.113) to obtain the standard errors of the coordinates $\hat{E}_p$ and $\hat{N}_p$ we need B from $\partial\hat{y}/\partial\hat{\ell}$:

$$B = \begin{bmatrix} \frac{\partial\hat{E}_p}{\partial\alpha} & \frac{\partial\hat{E}_p}{\partial\beta} & \frac{\partial\hat{E}_p}{\partial\gamma} & \frac{\partial\hat{E}_p}{\partial d} \\ \frac{\partial\hat{N}_p}{\partial\alpha} & \frac{\partial\hat{N}_p}{\partial\beta} & \frac{\partial\hat{N}_p}{\partial\gamma} & \frac{\partial\hat{N}_p}{\partial d} \end{bmatrix}$$

Hence

$$B = \begin{bmatrix} \hat{d}\cos(\theta + \hat{\alpha})\,\mu & 0 & 0 & \sin(\theta + \hat{\alpha}) \\ -\hat{d}\sin(\theta + \hat{\alpha})\,\mu & 0 & 0 & \cos(\theta + \hat{\alpha}) \end{bmatrix}$$

which can be evaluated as

$$B = \begin{bmatrix} 0.000511 & 0 & 0 & 0.9857 \\ 0.002990 & 0 & 0 & 0.1684 \end{bmatrix}$$

Then using (4.113) whilst remembering that

$$\hat{y} = \begin{bmatrix} \hat{E}_p & \hat{N}_p \end{bmatrix}^T$$

we have

$$C_{\hat{y}} = \begin{bmatrix} 0.000320 & -0.000097 \\ -0.000097 & 0.000108 \end{bmatrix}$$

and the required standard errors are

$$\sigma_{\hat{E}_p} = (0.000320)^{\frac{1}{2}} = 0.018\,\text{m}$$

$$\sigma_{\hat{N}_p} = (0.000108)^{\frac{1}{2}} = 0.010\,\text{m}$$

Notice that error ellipses and all the usual precision and reliability information could be computed if required.

A3.4 Sequential least squares

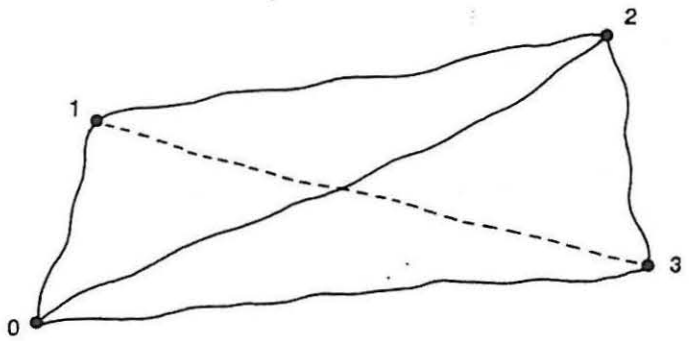


Fig. A3.3

The following five height differences have been observed between the four stations in Fig. A3.3. Note that the number in brackets after each observation is the approximate length of the level route in kilometres.

Stations from	to	Observed height difference	
0	1	61.478 m	(10)
1	2	16.994	(15)
2	3	-25.051	(9)
3	0	-53.437	(18)
0	2	78.465	(20)

Given that the height of station 0 is 214.880 metres above datum and that the standard error of an observed height difference is $0.005\ (d)^{\frac{1}{2}}$ m, where d is the length of the level route in kilometres, compute the least squares estimates of the heights of points 1, 2 and 3 and their standard errors.

Now if, at a later date, the height difference from 1 to 3 is observed to be -8.070 m (length 22) determine new sets of least squares estimates and standard errors.

Let $\bar{\ell}$ be the true values of the observed height differences and $\bar{x} = \left[\bar{x}_1\ \bar{x}_2\ \bar{x}_3\right]^T$ be the true values of the heights of stations 1, 2 and 3. Then we can write down the observation equations, $F(\bar{x}) = \bar{\ell}$ as

$$
\begin{aligned}
-214.880 + \bar{x}_1 &= \bar{\ell}_1 \\
-\bar{x}_1 + \bar{x}_2 &= \bar{\ell}_2 \\
-\bar{x}_2 + \bar{x}_3 &= \bar{\ell}_3 \\
-214.880 + \bar{x}_3 &= \bar{\ell}_4 \\
-214.880 + \bar{x}_2 &= \bar{\ell}_5
\end{aligned}
$$

Then putting

$$
x^\circ = \begin{bmatrix} x_1^\circ \\ x_2^\circ \\ x_3^\circ \end{bmatrix} = \begin{bmatrix} 276.358 \\ 293.345 \\ 268.317 \end{bmatrix}
$$

and applying (2.17) and (2.16) we obtain (2.22)

where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 0.000 \\ 0.007 \\ -0.023 \\ 0.000 \\ 0.000 \end{bmatrix}$$

with, from (3.7),

$$W = \begin{bmatrix} 4000 & 0 & 0 & 0 & 0 \\ 0 & 2667 & 0 & 0 & 0 \\ 0 & 0 & 4444 & 0 & 0 \\ 0 & 0 & 0 & 2222 & 0 \\ 0 & 0 & 0 & 0 & 2000 \end{bmatrix}$$

Then, following 3.3 we have the normal equations

$$\begin{bmatrix} 6667 & -2667 & 0 \\ -2667 & 9111 & -4444 \\ 0 & -4444 & 6667 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix} = \begin{bmatrix} -18.667 \\ 120.889 \\ -102.222 \end{bmatrix}$$

which using (4.67) leads to

$$\hat{x} = \begin{bmatrix} 0.0008 \\ 0.0089 \\ -0.0094 \end{bmatrix}$$

and

$$C_{\hat{x}} = \begin{bmatrix} 0.000181 & 0.000079 & 0.000052 \\ 0.000079 & 0.000197 & 0.000131 \\ 0.000052 & 0.000131 & 0.000237 \end{bmatrix}$$

The least squares estimates of the heights are then given by $x^0 + \hat{x}$ and their standard errors by the square roots of the diagonal elements of $C_{\hat{x}}$, i.e.

station	height	standard error
1	276.359 m	0.013 m
2	293.354	0.014
3	268.308	0.015

Now to implement the sequential least squares process we proceed as follows. First we use the previous results, and following the notation of 7.2.2 we put

$$\hat{\underline{x}}_1 = \hat{\underline{x}} \text{ and } C_{\hat{\underline{x}}_1} = N_1^{-1} = C_{\hat{\underline{x}}}$$

Then we have the additional (sixth) observation equation as

$$f_6(\bar{x}) = \bar{x}_3 - \bar{x}_1 = \bar{\ell}_6$$

giving

$$A_2 = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}, \qquad b_2 = \begin{bmatrix} -0.029 \end{bmatrix}$$

$$W_2^{-1} = \begin{bmatrix} 0.00055 \end{bmatrix}$$

Hence we can compute

$$A_2 N_1^{-1} A_2^T = 0.000314$$

$$W_2^{-1} + A_2 N_1^{-1} A_2^T = 0.000864$$

$$(W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1} = 1157$$

$$A_2 \hat{\underline{x}}_1 = -0.01015$$

$$A_2 \hat{\underline{x}}_1 - b_2 = 0.01885$$

$$N_1^{-1} A_2^T = \begin{bmatrix} -0.000129 \\ 0.000052 \\ 0.000185 \end{bmatrix}$$

$$N_1^{-1} A_2^T \left\{ (W_2^{-1} + A_2 N_1^{-1} A_2^T)^{-1} (A_2 \hat{\underline{x}}_1 - b_2) \right\} = \begin{bmatrix} -0.00281 \\ 0.00113 \\ 0.00404 \end{bmatrix}$$

Hence using (7.19) the new value of $\hat{\underline{x}}$ is

$$\hat{\underline{x}}_2 = \begin{bmatrix} 0.00076 \\ 0.00891 \\ -0.00939 \end{bmatrix} - \begin{bmatrix} -0.00281 \\ 0.00113 \\ 0.00404 \end{bmatrix} = \begin{bmatrix} 0.00357 \\ 0.00778 \\ -0.01343 \end{bmatrix}$$

To obtain the new covariance matrix we compute

$$N_1^{-1} A_2^T (W_2^{-1} - A_2 N^{-1} A_2^T)^{-1} A_2 N_1^{-1}$$

$$= \begin{bmatrix} -0.000129 \\ 0.000052 \\ 0.000185 \end{bmatrix} \begin{bmatrix} 1157 \end{bmatrix} \begin{bmatrix} -0.000129 & 0.000052 & 0.000185 \end{bmatrix}$$

$$= \begin{bmatrix} 0.000019 & -0.000008 & -0.000027 \\ -0.000008 & 0.000003 & 0.000011 \\ -0.000027 & 0.000011 & 0.000040 \end{bmatrix}$$

which, following (7.34), is substracted from $C_{\hat{x}_1}$ to produce

$$C_{\hat{x}_2} = \begin{bmatrix} 0.000162 & 0.000087 & 0.000080 \\ 0.000087 & 0.000194 & 0.000120 \\ 0.000080 & 0.000120 & 0.000198 \end{bmatrix}$$

Hence we have the following least squares estimates from the combination of all six observations:

station	height	standard error
1	276.361 m	0.012 m
2	293.353	0.013
3	268.303	0.014

The reader can check in straightforward fashion that the above solution is identical to one derived from a simultaneous computation with all six observations.

A3.5 Step by step least squares (Helmert-Wolf method)

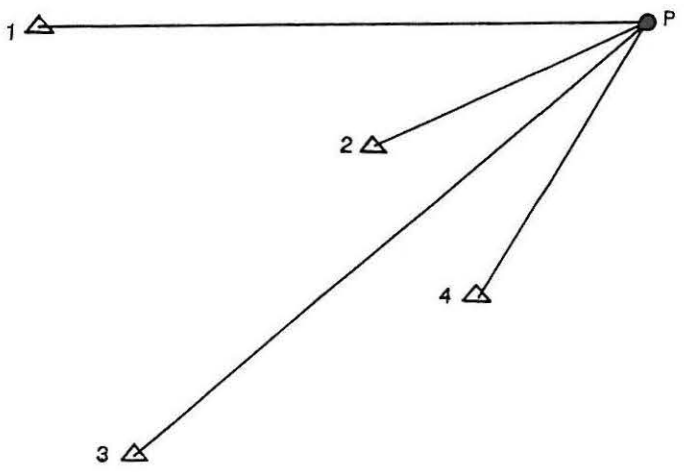


Fig. A3.4

A point P is fixed by the following distance measurements from known points 1, 2, 3 and 4.

1	– P	8622.45 m	with standard error	0.5 m
2	– P	3069.72		0.2
3	– P	6725.24		0.2
4	– P	12138.96		0.5

The known coordinates are

Station	Eastings	Northings
1	24433.11 m	71200.85 m
2	30375.90	70319.63
3	30813.54	65515.77
4	23146.72	64817.22

Some time later a new set of measurements are taken from stations 1, 2, 3 only with the following results:

1	– P	8624.18 m	with standard error	0.5 m
2	– P	3070.80		0.2
3	– P	6726.92		0.2

Assuming each set of measurements is subject to different unknown systematic scale errors s_1 and s_2 respectively, use the Helmert–Wolf step by step method to determine the least squares estimates of the coordinates of P without explicitly determining s_1 and s_2. Determine the standard errors of these estimates.

(Note that these two different scale errors may arise in practice because of different meteorological conditions for the two sets of measurements.)

First we write down the general functional model for an observed distance between stations P and i at the j th epoch

$$\frac{1}{(1-10^{-6}\bar{s}_j)}\left[(\bar{E}_p - E_i)^2 + (\bar{N}_p - N_i)^2\right]^{\frac{1}{2}} - \bar{O}_i = 0$$

where

$\bar{s}_j$ is the true value of the scale error at the j th epoch (in ppm)
$\bar{O}_i$ is the true value of the distance between i and P
$\bar{E}_p$, $\bar{N}_p$ are the true coordinates of P
E_i, N_i are the known coordinates of the i th station.

The differentials required to form the A matrix can then be approximately evaluated as

$$\frac{\partial f}{\partial s_j} \simeq 10^{-6} C_i$$

$$\frac{\partial f}{\partial E_p} \simeq (E^o - E_i)/C_i$$

$$\frac{\partial f}{\partial N_p} \simeq (N^o - N_i)/C_i$$

where E^o and N^o are the approximate values of $\bar{E}_p$ and $\bar{N}_p$ and

$$C_i = \left[(E^o - E_i)^2 + (N^o - N_i)^2\right]^{\frac{1}{2}}$$

Using the above, and taking approximate values of the coordinates of P as 33028.77 E and 71865.58 N, we proceed as follows.

Epoch 1

Observation equations are written as (2.22) with

$$A = \left[\begin{array}{c|cc} 0.00862 & 0.9970 & 0.0771 \\ 0.00307 & 0.8640 & 0.5035 \\ 0.00673 & 0.3294 & 0.9442 \\ 0.01214 & 0.8141 & 0.5807 \end{array}\right]$$

and

$$b = \begin{bmatrix} 1.13 \\ -0.73 \\ 0.11 \\ 0.83 \end{bmatrix}$$

with $x = \left[s_1 \,\middle|\, dE_p \;\; dN_p\right]^T$, which are partitioned according to (7.45). Also we have from (3.7)

$$W = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

which leads to the normal equations, partitioned according to (7.51),

$$\left[\begin{array}{c|cc} 0.002255 & 0.1956 & 0.2282 \\ \hline 0.1956 & 28.002 & 20.850 \\ 0.2282 & 20.850 & 29.998 \end{array}\right] \left[\begin{array}{c} \hat{s}_1 \\ \hline d\hat{E}_p \\ d\hat{N}_p \end{array}\right] = \left[\begin{array}{c} 0.042 \\ \hline -7.653 \\ -4.316 \end{array}\right]$$

These are reduced, following (7.52), to

$$\begin{bmatrix} 11.036 & 1.056 \\ 1.056 & 6.905 \end{bmatrix} \begin{bmatrix} d\hat{E}_p \\ d\hat{N}_p \end{bmatrix} = \begin{bmatrix} -11.296 \\ -8.563 \end{bmatrix} \qquad \text{(A3.3)}$$

Epoch 2

Observation equations are written as (2.22) with

$$A = \left[\begin{array}{c|cc} 0.00862 & 0.9970 & 0.0771 \\ 0.00307 & 0.8640 & 0.5035 \\ 0.00673 & 0.3294 & 0.9442 \end{array}\right] \quad b = \begin{bmatrix} 2.86 \\ 0.35 \\ 1.79 \end{bmatrix}$$

with $x = \left[\begin{array}{c|cc} s_2 & dE_p & dN_p \end{array}\right]^T$. They are partitioned according to (7.45). Also we have, from (3.7),

$$W = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 25 \end{bmatrix}$$

which leads to normal equations, partitioned according to (7.51), of

$$\left[\begin{array}{c|cc} 0.001665 & 0.1561 & 0.2002 \\ \hline 0.1561 & 25.351 & 18.959 \\ 0.2002 & 18.959 & 28.649 \end{array}\right] \left[\begin{array}{c} \hat{s}_2 \\ \hline d\hat{E}_p \\ d\hat{N}_p \end{array}\right] = \left[\begin{array}{c} 0.4266 \\ \hline 33.706 \\ 47.541 \end{array}\right]$$

These are reduced, following (7.52), to

$$\begin{bmatrix} 10.718 & 0.192 \\ 0.192 & 4.580 \end{bmatrix} \begin{bmatrix} d\hat{E}_p \\ d\hat{N}_p \end{bmatrix} = \begin{bmatrix} -6.285 \\ -3.767 \end{bmatrix} \qquad \text{(A3.4)}$$

Combination of epochs

Summing (A3.3) and (A3.4), as in (7.53), yields

$$\begin{bmatrix} 21.753 & 1.248 \\ 1.248 & 11.485 \end{bmatrix} \begin{bmatrix} d\hat{E}_p \\ d\hat{N}_p \end{bmatrix} = \begin{bmatrix} -17.581 \\ -12.330 \end{bmatrix} \qquad \text{(A3.5)}$$

with solution $d\hat{E}_p = -0.751$ and $d\hat{N}_p = -0.992$. Hence the least squares estimates of the coordinates of P are

$$\hat{E}_p = 33028.77 - 0.75 = 33028.02$$
$$\hat{N}_p = 71864.59 - 0.99 = 71864.60$$

The covariance matrix of these estimates is obtained from (7.59) by inverting the left-hand-side matrix of (A3.5)

$$C_{\hat{y}} = \begin{bmatrix} 0.0463 & -0.0050 \\ -0.0050 & 0.0876 \end{bmatrix}$$

Hence

$$\sigma_{\hat{E}_p} = (0.0463)^{\frac{1}{2}} = 0.22\,\text{m}$$
$$\sigma_{\hat{N}_p} = (0.0876)^{\frac{1}{2}} = 0.30\,\text{m}$$

The scale factors can be obtained by use of (7.56), viz.

$$\hat{s}_1 = (1/0.002255)(0.042 - [0.1956 \quad 0.2282] \begin{bmatrix} -0.751 \\ -0.992 \end{bmatrix})$$

$$= 184\ \text{ppm}$$

Similarly

$$\hat{s}_2 = 444\ \text{ppm}$$

Readers are invited to check that the above results are exactly equivalent to a full simultaneous least squares computation by carrying out the latter, i.e. by making one single computation with seven observation equations in terms of four parameters.

A3.6 The Kalman filter

This example follows exactly the navigation problem outlined in 8.4.1 and uses identical notation. Note that the Kalman filter is recursive and takes some time to settle down in practice. Here we enter the process after the i-1 th recursion and simply illustrate the numerical procedure by carrying out one complete set of computations for filtering, smoothing and prediction. The numbers have been arbitrarily chosen and are not supposed to represent any particular navigation system. Also more digits than are really meaningful have been quoted to enable the reader to check his own computations more easily.

The following information is given (linear and time units are metres and seconds respectively throughout).

Standard deviation of ship's (assumed) random acceleration (σ) = 0.0002

Fix interval (Δt) = 60

Covariance matrix of each position-fix (W_i^{-1}):

$$W_i^{-1} = \begin{bmatrix} 91.6 & 42.7 \\ 42.7 & 91.6 \end{bmatrix}$$

Then, using (8.71) we obtain

$$W_M^{-1} = \begin{bmatrix} 0.1296 & 0 & 0.00432 & 0 \\ 0 & 0.1296 & 0 & 0.00432 \\ 0.00432 & 0 & 0.000144 & 0 \\ 0 & 0.00432 & 0 & 0.000144 \end{bmatrix}$$

Also we will take the following as the values of the state vector and its covariance matrix resulting from the i-1 th recursion

$$\hat{x}'_{i-1} = \begin{bmatrix} 15969.933 \\ 25030.638 \\ 2.92214 \\ 2.00528 \end{bmatrix}$$

and

$$N_{i-1}^{-1} = C_{\hat{x}'_{i-1}} = \begin{bmatrix} 29.020576 & 11.740694 & 0.092973 & 0.029312 \\ 11.740694 & 20.661862 & 0.029312 & 0.072305 \\ 0.092973 & 0.029312 & 0.000655 & 0.000111 \\ 0.029312 & 0.072305 & 0.000111 & 0.000576 \end{bmatrix}$$

The results of the position-fix at the i th point (i.e. the observations as far as the Kalman filter is concerned) are

$$E_i^o = 16145.292$$
$$N_i^o = 25158.442$$

Filtering

Using (8.51) with $M_{i-1,i}$ given by (8.66) we predict the state vector as

$$\underset{\sim}{x}_i = \begin{bmatrix} 1 & 0 & 60 & 0 \\ 0 & 1 & 0 & 60 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 15969.933 \\ 25030.638 \\ 2.92214 \\ 2.00528 \end{bmatrix} = \begin{bmatrix} 16145.262 \\ 25150.955 \\ 2.92214 \\ 2.00528 \end{bmatrix}$$

From (8.52) we predict the covariance matrix

$$C_{\underset{\sim}{x}_i} = N_i'^{-1} = \begin{bmatrix} 42.66415 & 15.623718 & 0.136580 & 0.035686 \\ 15.623718 & 31.540946 & 0.035686 & 0.111173 \\ 0.136580 & 0.035686 & 0.000799 & 0.000111 \\ 0.035686 & 0.111173 & 0.000111 & 0.000720 \end{bmatrix}$$

We next compute the gain matrix using (8.53) with A_i given by (8.65)

$$G = \begin{bmatrix} 0.336512 & -0.043163 \\ -0.043163 & 0.367242 \\ 0.001170 & -0.000351 \\ -0.000351 & 0.001419 \end{bmatrix}$$

The least squares estimate of the state vector is then computed from (8.54)

$$\hat{\underset{\sim}{x}}_i = \begin{bmatrix} 16144.949 \\ 25153.703 \\ 2.91955 \\ 2.01590 \end{bmatrix} \begin{matrix} \text{eastings} \\ \text{northings} \\ \text{east velocity} \\ \text{north velocity} \end{matrix}$$

with a covariance matrix given by (8.55)

$$C_{\hat{\underset{\sim}{x}}_i} = \begin{bmatrix} 28.981496 & 11.727531 & 0.092159 & 0.028476 \\ 11.727531 & 20.632154 & 0.028476 & 0.071886 \\ 0.092159 & 0.028476 & 0.000652 & 0.000108 \\ 0.028476 & 0.071886 & 0.000108 & 0.000575 \end{bmatrix}$$

which can be seen to be virtually identical to the covariance matrix for the filtered state vector at point i-1. This is to be expected because once the filter has stabilised the precision of the state vector will only change very slowly.

Prediction

To predict the state vector at point i+1 we use (8.59) to yield

$$\hat{x}'_{i+1} = \begin{bmatrix} 16320.122 \\ 25274.657 \\ 2.91955 \\ 2.01589 \end{bmatrix}$$

the covariance matrix of which, if required, could be computed from (8.60).

Smoothing

Although in practice it would not be usual to do so at this stage it is possible to compute the least squares estimate of the state vector at point i-1 using all data up to and including point i.

First we compute $\hat{k}_i$ from (8.58)

$$\hat{k}_i = \begin{bmatrix} 0.047942 \\ -0.110881 \end{bmatrix}$$

then we obtain $\hat{x}_{i-1}$ from (8.57)

$$\hat{x}_{i-1} = \begin{bmatrix} 15970.097 \\ 25028.513 \\ 2.92453 \\ 1.99515 \end{bmatrix}$$

Exercise

The reader is invited to verify that if, at point i+1, we have observations of

$$E^o_{i+1} = 16324.026$$
$$N^o_{i+1} = 25276.678$$

then the filtered, predicted and smoothed results at points i+1, i+2 and i respectively would be

$$\begin{bmatrix} 16321.346 \\ 25275.229 \\ 2.92338 \\ 2.01738 \end{bmatrix}, \quad \begin{bmatrix} 16496.748 \\ 25396.272 \\ 2.92338 \\ 2.01738 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 16143.952 \\ 25153.220 \\ 2.91584 \\ 2.01444 \end{bmatrix}$$

References

ADAMS J R 1979 Description of the local level inertial survey system and its simulation Technical Report No. 59 Department of Surveying Engineering University of New Brunswick 120 pp

ASHKENAZI V 1970 Adjustment of control networks for precise engineering surveys Chartered Surveyor 102 314-320

ASHKENAZI V 1980 Least squares adjustment: signal or just noise Royal Institution of Chartered Surveyors London Ordinary general meeting October 14 pp

ASHKENAZI V CROSS P A 1972 Strength analysis of block VI of the European triangulation network Bulletin Geodesique 103 5-24

ASHKENAZI V CROSS P A 1976 Strength of long lines in terrestrial geodetic control networks Bulletin Geodesique 50 (2) 169-180

ASHKENAZI V CROSS P A DAVIES M J K PROCTOR D W 1972 The readjustment of the retriangulation of Great Britain and its relationship to the European terrestrial and satellite networks Ordnance Survey Professional Papers New Series No. 24 51 pp

ASHKENAZI V DODSON H 1977 The Nottingham multi-pillar baseline Civil Engineering Surveyor June 1-8

BAARDA W W 1968 A testing procedure for use in geodetic networks Netherlands Geodetic Commission Publications on Geodesy No. 5 97 pp

BOMFORD G 1980 Geodesy 4th edition Clarendon Press Oxford 855 pp

CHRZANOWSKI A 1981 Optimization of breakthrough accuracy in tunnelling surveys Canadian Surveyor 35 (1) 5-16

COOPER M A R 1974 Fundamentals of Survey Measurement and Analysis Crosby Lockwood Staples London 107 pp

COOPER M A R 1981 A priori and a posteriori analyses, or game and guess Survey Review 201 137-140

CRAMER H 1946 Mathematical Methods of Statistics Princeton University Press Princeton

CROSS P A 1972 The effect of errors in weights Survey Review 165 319-325

CROSS P A 1981a The computation of position at sea Hydrographic Journal 20 7-16

CROSS P A 1981b Computer aided design of geodetic networks Proceedings of VI International Symposium on Geodetic Computations Munich 12 pp

CROSS P A 1982 A priori and a posteriori analyses letter to Survey Review 204 303-304

CROSS P A WHITING B M 1981 On the design of vertical control networks using iterative methods Proceedings of VI International Symposium on Geodetic Computations Munich 9 pp

CROSS P A WEBB J P 1980 Instrumentation and methods for inertial surveying Chartered Land Surveyor/Chartered Minerals Surveyor 2 (2) 4-27

DILLINGER W H 1978 Helmert block higher level system Preceedings of the Second International Symposium on Problems Related to the Redefinition of North American Geodetic Networks U.S. Government Printing Office Washington D.C. 417-426

FORSTER P D 1980 A study of aspects of survey adjustment and sequential design M.Sc. thesis Department of Geodesy and Surveying University of Oxford 76 pp

FOX L 1964 An Introduction to Numerical Linear Algebra Clarendon Press Oxford 328 pp

GAGNON P 1976 Step-by-step adjustment procedures for large horizontal, geodetic networks Technical Report No. 38 Department of Surveying Engineering University of New Brunswick 159 pp

GALE L A 1965 Theory of adjustments by least squares Canadian Surveyor 19 (1) 42-63

GAUSS K F 1809 Theoria Motus Corporum Coelestium Hamburg

GAUSS K F 1821 Theoria Combinationis Observationum Erroribus Minimis Obnoxiae Published in 1873 in Werke (4) Gottingen

GRAFAREND E W 1974 Optimization of geodetic networks Bulletino di Geodesia a Science Affini 33 (4) 351-406

GREGERSON L F 1975 Inertial instrumentation at the Geodetic Survey of Canada Proceedings of the Commonwealth Survey Officers' Conference paper C4 14 pp

HAWKINS D M 1980 Identification of Outliers Chapman and Hall London 188 pp

ISNER J F 1978 Helmert block initial level system Proceedings of the Second International Symposium on Problems Related to the Redefinition of North American Geodetic Networks U.S. Government Printing Office Washington D.C. 405-416

KALMAN R E 1960 A new approach to linear filtering and prediction problems ASME Journal of Basic Engineering 82 35-45

KENNEDY J B NEVILLE A M 1976 Basic Statistical Methods for Engineers and Scientists Harper International New York 490 pp

KOK J J EHRNSPERGER W RIETVELD H 1980 The 1979 adjustment of the united European levelling network and its analysis of precision and reliability Proceedings of the Second International Sumposium on Problems Related to the Redefinition of American Vertical Geodetic Networks Canadian Institute of Surveyors Ottawa 455-483

KRABS W 1979 Optimization and Approximation Wiley New York 220 pp

KRAKIWSKY E J 1976 A synthesis of recent advances in the method of least squares Technical Report No. 42 Department of Surveying Engineering University of New Brunswick 125 pp

LAPLACE P S 1812 Theorie Analytique de Probabilities Paris

LEGENDRE A M Nouvelles Methodes pour la Determination des Orbites des Cometes - Appendice sur la Methode des Moindres Carres Paris

LOGAN W R 1955 The rejection of outlying observations Empire Survey Review 97 133-137

MANN H B WALD A 1942 On the choice of the number of class intervals in the application of the chi-square test Annals of Mathematical Statistics 13 306-317

MARKOV A A 1912 Wahrscheinlichkeitsrechnung 2nd edition Leipzig and Berlin

MASSON-SMITH D HOWELL P M ABERTHNEY-CLARK A B D E 1974 The national gravity reference network 1973 Ordnance Survey Professional Papers New Series No. 26 22 pp

MERRIMAN M 1877 A list of writings relating to the method of least squares with historical and critical notes Transactions of the Connecticut Academy of Arts and Sciences 4 151-232

MIKHAIL E M 1976 Observations and Least Squares Dun Donnelley New York 497 pp

MIKHAIL E M GRACIE G 1981 Analysis and Adjustment of Survey Measurements Van Nostrand Reinhold New York 340 pp

MOOD A M GRAYBILL F A 1963 Introduction to the Theory of Statistics McGraw-Hill New York 443 pp

MORITZ H 1972 Advanced least squares methods Ohio State University Report No. 175 Reports of the Department of Geodetic Science Ohio State University 129 pp

MORITZ H 1980 Advanced Physical Geodesy Abacus Press 500 pp

MORRISON N 1969 Introduction to Sequential Smoothing and Prediction McGraw-Hill New York 645 pp

NICKERSON B J KRAKIWSKY E J THOMPSON D B KRAKIWSKY M L CRAWFORD J J 1978 Design of survey networks using interactive computer graphics Research Report Department of Surveying Engineering University of New Brunswick 12 pp

ORDNANCE SURVEY 1967 The History of the Retriangulation of Great Britain 1935-62 HMSO London 395 pp

PELZER H 1979 Criteria for the reliability of geodetic networks Halmos F and Somogyi J (eds) Akademiai Kiado Budapest 553-562

POPE A J 1976 The statistics of residuals and the detection of outliers National Oceanic and Atmospheric Administration Technical Reports USA Department of Commerce Washington 133 pp

RAINSFORD H F 1957 Survey Adjustments and Least Squares Constable London 326 pp

RICHARDUS P 1966 Project Surveying North Holland Amsterdam 467 pp

STEFANOVIC P 1978 Blunders and least squares ITC Journal 1978-1 122-157

SUNTER A B 1966 Statistical properties of least squares estimates Canadian Surveyor 20 36-44

TSCHERNING C C RAPP R H 1974 Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree variance models Report No. 208 Reports of the Department of Geodetic Science Ohio State University

THOMPSON E H 1969 Introduction to the Algebra of Matrices with Some Applications Adam Hilger London 229 pp

TODHUNTER I 1865 A History of the Mathematical Theory of Probability Macmillan

VANIČEK P KRAKIWSKY E J 1986 Geodesy: The Concepts 2nd edition North Holland Amsterdam 697 pp

WEIGHTMAN J A 1975 Doppler ties to European datum and the European Geoid Proceedings of the Commonwealth Survey Officers' Conference Paper B4 32 pp

WEIGHTMAN J A 1982 Personal communication

WELLS D E KRAKIWSKY E J 1971 The method of least squares Lecture Notes No. 18 Department of Surveying Engineering University of New Brunswick 180 pp

The School of Surveying

Head: Dr. Allan Brimicombe BA(Hons) MPhil PhD FRGS FGS

The School aims to provide a comprehensive educational service in the field which is becoming known as geomatics, i.e. measurement science, mapping, land surveying, geographical information systems and their applications. In particular there is a recognition of the important interrelationships between all aspects of surveying and its use of information technology systems in the collection, storage, analysis and dissemination of data which has some form of location.

The School has existed in various forms since 1949, when training for the professional examinations of the Royal Institution of Chartered Surveyors began at the South West Essex Technical College later to become the Waltham Forest Technical College. 1969 saw the start of a degree in Land Surveying shortly before the North East London Polytechnic (NELP) was formed in 1970 with its Department of Land Surveying under the leadership of John Hollwey. The portfolio of courses has increased since then and the School now offers Higher Certificate and Diploma in Land Surveying and Geographic Information Systems, B.Sc honours degrees in Surveying and Mapping, Geographic Information Systems (GIS) and variants specialising in the application of GIS to Business and Property. Recently a programme of postgraduate courses leading to MSc has been introduced to provide a well integrated development path for individuals. To support its work, the School has well-appointed laboratories with a range of equipment and software for hands-on experience.

Further details may be obtained from:

The School of Surveying
University of East London
Longbridge Road
Dagenham
Essex RM8 2AS
England

Tel: +44 (0) 208 223 2355
Fax: +44 (0) 208 223 2918
Email: Land-Surveying@uel.ac.uk
URL: http://www.uel.ac.uk/faculties/surv

Working Papers published by the School of Surveying of the University of East London include lecture notes, instructions on the use of departmental equipment, documentation of computer programs, etc., primarily for sale to undergraduates as essential material for courses run in the School. Several are in preparation and the titles of those published to date are given below.
Copies may be obtained from:

School of Surveying
University of East London
Longbridge Road
Dagenham
Essex RM8 2AS.

1. R W Dixon-Gough
   Proofing: the requirements of the mapmaker and some current techniques
   Second Edition July 1982 56pp ISBN 0 907382 05 3 ISSN 0260-9142

2. P A Cross, J R Hollwey, L G Small
   Geodetic appreciation
   October 1980 185pp ISBN 0 907382 010 ISSN 0260-9142

3. R K Bullard, P J Lakin
   First steps in remote sensing
   July 1981 61pp ISBN 0 907382 02 9 ISSN 0260-9142

4. P A Cross
   The Geodetic suite
   Second Edition June 1984 114pp ISBN 0 907382 09 6 ISSN 0260-9142

5. G M Sears
   First steps in photogrammetry
   January 1982 158pp ISBN 0 907382 04 5 ISSN 0260-9142

6. P A Cross
   Advanced least squares applied to position-fixing
   April 1983 205pp ISBN 0 907382 06 1 ISSN 0260-9142

7. G E D Cole, R W Dixon-Gough, P E Sorrell
   First steps in cartography
   May 1983 132pp ISBN 0 907382 07 X ISSN 0260-9142

8. R K Bullard
   Abandoned land in Thurrock: an application of remote sensing
   July 1983 53pp ISBN 0 907382 08 8 ISSN 0260-9142

9. R W Dixon-Gough
   An introduction to photography for map makers
   February 1987 186pp ISBN 0 907382 10X ISSN 0260-9142

ISBN 0 907382 010
ISSN 0260-9142