RMIT

Geospatial Science

NOTES ON LEAST SQUARES

RMIT
Geospatial Science
School of Mathematical and Geospatial Science
NOTES ON LEAST SQUARES
This edition printed in 2005

CONTENTS

CHAPTER		PAGES
1.	**Introduction**	3
2.	**Least Squares Adjustment of Indirect Observations**	57
3.	**Propagation of Variances**	37
4.	**Approximate Values**	5
5.	**Propagation of Variances Applied to Least Squares Adjustment of Indirect Observations**	**7**
6.	**Least Squares Adjustment of Observations Only**	39
7.	**Linearization Using Taylor's Theorem and the Derivation of Some Common Surveying Observation Equations**	15
8.	**The Standard Error Ellipse**	10
9.	**Least Squares Resection**	14
10.	**Least Squares Bearing Intersection**	12
APPENDICES		
A	**Matrix Algebra**	23
REFERENCES		2

NOTES ON LEAST SQUARES

1. INTRODUCTION

The theory of least squares and its application to adjustment of survey measurements is well known to every geodesist. The invention of the method is generally attributed to Karl Freidrich Gauss (1777-1855) but could equally be credited to Adrien-Marie Legendre (1752-1833).

Gauss used the method of least squares to compute the elements of the orbit of the minor planet *Ceres* and predicted its position in October 1801 from a few observations made in the previous year. He published the technique in 1809 in *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium* (Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections), mentioning that he had used it since 1795, and also developed what we now know as the normal law of error, concluding that: *"... the most probable system of values of the quantities ... will be that in which the sum of the squares of the differences between the actually observed and computed values multiplied by numbers that measure the degree of precision, is a minimum."* (Gauss 1809).

Legendre published an independent development of the technique in *Nouvelles méthodes pour la détermination des orbites des comètes* (New methods for the determination of the orbits of comets), Paris, 1806 and also as the "Méthod des moindres carriés" (Method of Least Squares), published in the *Mémoires de l'Institut national des sciences at arts*, vol. 7, pt. 2, Paris, 1810.

After these initial works, the topic was subjected to rigid analysis and by the beginning of the 20th century was the universal method for the treatment of observations. Merriman (1905) compiled a list of 408 titles, including 72 books, written on the topic prior to 1877 and publication has continued unabated since then. Leahy (1974) has an excellent summary of the

development of least squares and clearly identifies the historical connection with mathematical statistics, which it pre-dates.
The current literature is extensive; the books *Observations and Least Squares* (Mikhail 1976) and *Analysis and Adjustment of Survey Measurements* (Mikhail and Gracie 1981), and lecture notes by Cross (1992), Krakiwsky (1975) and Wells and Krakiwsky (1971) stand out as the simplest modern treatments of the topic.

Following Wells and Krakiwsky (1971, pp.8-9), it is interesting to analyse the following quotation from Gauss' *Theoria Motus* (Gauss, 1809, p.249).

> "If the astronomical observations and other quantities, on which the computation of orbits is based, were absolutely correct, the elements also, whether deduced from three or four observations, would be strictly accurate (so far indeed as the motion is supposed to take place exactly according to the laws of KEPLER), and, therefore, if other observations were used, they might be confirmed, but not corrected. But since all our measurements and observations are nothing more than approximations to the truth, the same must be true of all calculations resting upon them, and the highest aim of all computations made concerning concrete phenomena must be to approximate, as nearly as practicable, to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities. This problem can only be properly undertaken when an approximate knowledge of the orbit has been already attained, which is afterwards to be corrected so as to satisfy all the observations in the most accurate manner possible."

This single paragraph, written almost 200 years ago, embodies the following concepts, which are as relevant today as they were then.

(i) Mathematical models may be incomplete,

(ii) Physical measurements are inconsistent,

(iii) All that can be expected from computations based on inconsistent measurements are estimates of the "truth",

(iv) Redundant measurements will reduce the effect of measurement inconsistencies,

(v) Initial approximations to the final estimates should be used, and finally,

(vi) Initial approximations should be corrected in such a way as to minimise the inconsistencies between measurements (by which Gauss meant his method of least squares).

These notes contain a development of Least Squares processes applicable to surveying and geodesy. Examples and exercises of least squares processes are given using MATLAB, an interactive, matrix-based system for scientific and engineering computation and visualization. The name MATLAB is derived from MATrix LABoratory and is licensed by The MathWorks, Inc.

2. LEAST SQUARES ADJUSTMENT OF INDIRECT OBSERVATIONS

2.1. Introduction

The modern professional surveyor must be competent in all aspects of surveying measurements such as height differences, linear distances, horizontal and vertical angle measurements and combinations thereof which form the fundamental observations used to determine position in space. To obtain these measurements with any degree of confidence the surveyor must be aware of the principles and operation of various pieces of surveying equipment as well as the nature of measurements and the possible effects of errors on these measurements and any derived quantities. The nature of errors in measurements, studied by Gauss and leading to his theory of errors (the normal law of error) is the basis of statistical rules and tests that the surveyor employs to assess the quality of measurements; these rules and tests are covered in basic statistics courses during the undergraduate degree program. In the simple least squares processes and applications which follow it is sufficient to assume that the measurements are affected by small accidental or random errors and the least squares "solutions" provide a means of determining the best estimate of a measured quantity. Least squares solutions also imply that the quantity of interest has been determined from a redundant system of measurements, i.e., there are more measurements than the minimum number required to calculate the quantity.

2.1.1. Definition and classification of measurements

Crandall and Seabloom (1970, pp. 4-5) give a definition of a measurement as:

> A measurement is a comparison between an unknown quantity and a predefined standard, determined by some measuring device and hence, any measured value is an approximation of the exact or true value, not the true value itself. Since the true value of a quantity cannot be measured, any measurement contains by definition, an error.

Direct measurements (or observations) are those that are made directly upon the quantity to be determined. Measurements of a line by direct chaining, or Electronic Distance Measurement (EDM), or measurement of an angle by theodolite or Total Station are examples of direct measurements.

Indirect measurements (or observations) are not made upon the quantity itself but are made on some other quantity or quantities related to it. For example, the coordinates of a point *P* are indirectly determined by measuring bearings and distances to *P* from other points; the latitude of *P* may be determined from altitudes to certain stars; and the height of *P* may be determined by measured height differences from a known point.

2.1.2. Classification of errors of measurement

Since, by definition, every measurement contains an error it is necessary to consider the various kinds of errors that occur in practice. Rainsford (1968, p. 1) provides a derivation of the word error as:

> coming from the Latin *errare* which means to wander and not to sin.

Rainsford divides errors into four classes

(a) blunders or mistakes
(b) constant errors
(c) systematic errors
(d) accidental or random errors

Blunders or mistakes are definite mis-readings, booking errors or other like occurrences. They are usually caused by poor measurement technique and/or a lack of attention to detail by the person making the measurement. They may be eliminated or minimized by correct and careful measurement techniques, and a thorough understanding of the operation of the equipment used for the measurement.

Constant errors are those that do not vary throughout the particular measurement period. They are always of the same sign. Neglecting to standardize a measuring tape introduces a constant error; failure to use the correct prism-offset value introduces constant errors in EDM measurements. A faulty joint between sections of a levelling staff will introduce a constant error into height differences from spirit levelling. Constant errors can be eliminated from measurements by a thorough understanding of the measurement process and the equipment used.

Systematic errors are those errors that follow some fixed law (possibly unknown) dependent on local conditions and/or the equipment being used. For example, if the temperature and pressure (which are indicators of atmospheric conditions) are not measured when using EDM equipment then a systematic error may be introduced, since the modulated electromagnetic beam of the EDM passes through the atmosphere and its time of travel (indirectly measured by phase comparison of emitted and reflected beams) is affected by atmospheric conditions. All EDM measurements must be corrected for atmospheric conditions that depart from "standard conditions".

Accidental or Random errors are the small errors remaining in a measurement after mistakes, constant errors and systematic errors have been eliminated. They are due to the imperfection of the instruments used, the fallibility of the observer and the changing environmental conditions in which the measurements are made, all of which affect the measurement to a lesser or greater degree.

Bearing in mind the aforementioned, it could be said that all careful measurements (where mistakes, constant errors and systematic errors have been eliminated) contain small random errors and from experience, three axioms relating to random errors can be stated.

1. Small errors occur more frequently, or are more probable then large errors.
2. Positive and negative errors of the same magnitude are equally probable
3. Very large errors do not occur.

These axioms are the basic premises on which the theory of errors (the normal law of error) is founded.

2.1.3. Errors, corrections and residuals

A measured quantity has a true value and a most probable value. The most probable value is often called the best estimate and the two terms can be taken as synonymous.

No matter how many times a quantity is measured, its true value will remain unknown and only a best estimate can be obtained from the measurements. In the case of a single measured quantity, the best estimate is the arithmetic mean (or average) of the measurements.

If a quantity has been measured a number of times, the difference between the true (but unknown) value and any measurement is the true error and the difference between the best estimate and any measurement is the apparent error.

These relationships can be established by defining a correction to have the same magnitude as an error but the opposite sign. In surveying, the terms correction and residual are regarded as synonymous, and are universally denoted by the letter *v*.

Suppose an unknown quantity *x* is measured *n* times giving values $x_1, x_2, x_3, \ldots, x_n$. The true value (unknown) of the measured quantity is μ (mu) and is estimated by the arithmetic mean $\overline{x}$ where

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{k=1}^{n} x_k}{n} \qquad (2.1)$$

The arithmetic mean is regarded as the best estimate or most probable value. A correction *v* having the same magnitude as an error but the opposite sign is defined as

$$v_k = \overline{x} - x_k$$

Since these corrections relate to the measurements and arithmetic mean, they could be called apparent corrections and hence according to our definition of corrections and errors, apparent errors $-v$ are defined as

$$-v_k = x_k - \overline{x}$$

In a similar manner, we may define true errors ε (epsilon) as

$$\varepsilon_k = x_k - \mu$$

These relationships may be expressed as

measurement + residual = best estimate
measurement – best estimate = apparent error
measurement – true value = true error

True errors are unknown and are approximated by apparent errors. The closer the best estimate (or most probable value) approaches the true value, the closer the apparent error approaches the true error. The laws defining the nature and behaviour of true errors were derived from practical axioms deduced from the nature of apparent errors and hence any theory of errors may also be regarded as a theory of corrections (or residuals) and the distinction between true errors and apparent errors is ignored for all practical purposes.

The following sections contain simple examples of least squares processes, the mean, the weighted mean, line of best fit (linear regression) and polynomial curve fitting. In each case, Gauss' least squares principle: *"... the most probable system of values of the quantities ... will be that in which the sum of the squares of the differences between the actually observed and computed values multiplied by numbers that measure the degree of precision, is a minimum."* will be employed to determine equations or systems of equations that may be regarded as least squares solutions to the problems. Furthermore, it is assumed that all measurements are free of mistakes, constant errors and systematic errors and "contain" only random errors and that the precision of the measurements is known a priori (Latin *a priori* from what is before). Solutions to some of the examples are provided as MATLAB script files (.m files).

2.2. The Mean

It is well known practice that when a single quantity is measured a number of times the arithmetic mean is taken as the best estimate of the measured quantity. Few people realise that when they adopt this practice they are employing Gauss' least squares principle.

Consider a series of measurements $x_1, x_2, x_3, \ldots, x_n$ of a quantity and denote the best estimate of this quantity as p. According to our general definition of measurements and corrections we may write: measurement + correction (or residual) = best estimate or

$$x_1 + v_1 = p,\ x_2 + v_2 = p,\ x_3 + v_3 = p,\ \cdots,\ x_n + v_n = p$$

These equations can be rearranged as

$$v_1 = p - x_1,\ v_2 = p - x_2,\ v_3 = p - x_3,\ \cdots,\ v_n = p - x_n$$

Now if all the measurements can be regarded as having equal precision we may state the least squares principle as

> The best estimate *p* is that value which makes the sum of the squares of the residuals a minimum.

We may define a least squares function φ (phi) as

$$\varphi = \text{ the sum of the squares of the residuals} = \sum_{k=1}^{n} v_k^2 \tag{2.2}$$

or $$\varphi = \sum_{k=1}^{n} v_k^2 = (p - x_1)^2 + (p - x_2)^2 + \cdots + (p - x_n)^2$$

We say that φ is a function of *p,* the single parameter or variable in this equation. The minimum value of the function (i.e. making the sum of squares of residuals a minimum) can be found by equating the derivative $\frac{d\varphi}{dp}$ to zero, i.e.,

$$\varphi \text{ is a minimum when } \frac{d\varphi}{dp} = 0$$

and $$\frac{d\varphi}{dp} = 2(p - x_1) + 2(p - x_2) + \cdots + 2(p - x_n) = 0$$

Cancelling the 2's and rearranging gives the best estimate *p* as the arithmetic mean.

$$p = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{k=1}^{n} x_k}{n} \tag{2.3}$$

Hence, the arithmetic mean of a series of measurements is the best estimate according to Gauss' least squares principle.

2.3. The Weighted Mean

Before demonstrating that the weighted mean of a set of observations is the result of a least squares process, some discussion of the term weight and its connection with precision is required.

2.3.1. Measures of Precision of a Finite Population

In every least squares process it is assumed that the precision of measurements is known. The precision is a measure of the dispersion (or spread) of a number of measurements from their mean (or average) value. A common statistical measure of precision is the variance σ^2 and the positive square root of the variance is the standard deviation σ. Equations for the variance and standard deviation differ depending on whether the population of measurements is finite or infinite and a population is a potential set of quantities that we want to make inference about based on a sample from that population.

Following Deakin and Kildea (1999), consider a finite population, such as the examination marks m_k of a group of *N* students in a single subject. Since we have complete information about the population, i.e., its size is known, the mean μ, the variance σ^2 and the standard deviation σ of the finite population are

$$\mu = \frac{\sum_{k=1}^{N} m_k}{N} \tag{2.4}$$

$$\sigma^2 = \frac{\sum_{k=1}^{N} (m_k - \mu)^2}{N} \tag{2.5}$$

$$\sigma = \sqrt{\frac{\sum_{k=1}^{N} (m_k - \mu)^2}{N}} \tag{2.6}$$

Note that the variance σ^2 is the average squared difference of a member of the population m_k from the population mean μ. The mean, variance and standard deviation are known as population parameters.

2.3.2. Estimates of Precision of Samples of an Infinite Population

Consider surveying measurements, drawn from infinite populations with the attendant difficulties of estimation since population averages can never be known. In such cases we are usually dealing with small samples of measurements of size *n* and we can only obtain estimates of the unknown population parameters μ, σ^2 and σ. For a sample of *n*

measurements $x_1, x_2, x_3, \ldots, x_n$ from an infinite population, estimates of the mean, variance and standard deviation, denoted by $\overline{x}$, s_x^2 and s_x are

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{2.7}$$

$$s_x^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \overline{x}\right)^2 \tag{2.8}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \overline{x}\right)^2} \tag{2.9}$$

Note the divisor $n-1$ (known as the degrees of freedom) in equations for the estimates of variance and standard deviation. This ensures that s_x^2 is an unbiased estimate of the population variance σ^2, but does not ensure that s_x is an unbiased estimate of the population standard deviation σ; the action of "taking a square-root" changes the property of unbiasedness. This is more an accident of mathematics rather than a cause of faulty estimation but it is not well appreciated in general. Deakin and Kildea (1999, p. 76) show that an unbiased estimator s_x^* of the population standard deviation σ is given by

$$s_x^* = \sqrt{\frac{1}{c_n}\sum_{k=1}^{n}\left(x_k - \overline{x}\right)^2} \tag{2.10}$$

Values of c_n are given in Table 2.1

n	2	3	4	5	10	15	20	30	90
n-1	1	2	3	4	9	14	19	29	89
c_n	0.64	1.57	2.55	3.53	8.51	13.51	18.51	28.50	88.50

Table 2.1 Values of divisor c_n for unbiased estimation of σ

In these notes, it is always assumed that the terms mean, variance and standard deviation refer to estimates of population values.

2.3.3. Relationship between Weights and Estimates of Variance

Another measure of precision, often used in least squares applications is weight *w* and the weight of an observation (or measurement) is defined as being inversely proportional to the variance

$$w_k \propto \frac{1}{s_k^2} \tag{2.11}$$

or

$$w_k = \frac{\sigma_0^2}{s_k^2} \tag{2.12}$$

σ_0^2 is a constant of proportionality known as the reference variance or variance factor. This is the classical definition of weight and if an observation has unit weight $(w_k = 1)$ its variance equals σ_0^2, hence the reference variance is sometimes called the variance of an observation of unit weight; a term often encountered in older surveying texts. In this definition of weight, there is an assumption that the measurements are uncorrelated (a statistical term relating to the dependence between measurements, see section 2.5). In cases where measurements are correlated, weights are not an adequate means of describing relative precisions.

As an example of the connection between weights and standard deviations consider three uncorrelated (i.e., independent) observations of a particular distance, where each observation is the mean of several measurements and standard deviations of each observation have been determined from the measurements

observation 1	136.225 m (st. dev. 0.010 m)
observation 2	136.233 m (st. dev. 0.032 m)
observation 3	136.218 m (st. dev. 0.024 m)

Since the weight is inversely proportional to the variance, the observation with the smallest weight will have the largest variance (standard deviation squared). For convenience, this observation is given unit weight i.e., $w_2 = 1$ and the other observations (with smaller variances) will have higher weight. Hence from (2.12)

$$w_2 = 1 = \frac{\sigma_0^2}{(0.032)^2} \quad \text{and} \quad \sigma_0^2 = (0.032)^2$$

The weights of the three observations are then

$$w_1 = \frac{(0.032)^2}{(0.010)^2} = 10.24$$

$$w_2 = \frac{(0.032)^2}{(0.032)^2} = 1$$

$$w_3 = \frac{(0.032)^2}{(0.024)^2} = 1.78$$

Weights are often assigned to observations using "other information". Say for example, a distance is measured three times and a mean value determined. If two other determinations of the distance are from the means of six and four measurements respectively, the weights of the three observations may simply be assigned the values 3, 6 and 4. This assignment of weights is a very crude reflection of the (likely) relative precisions of the observations since it is known that to double the precision of a mean of a set of measurements, we must quadruple the number of measurements taken (Deakin and Kildea, 1999, p. 76).

2.3.4. Derivation of Equation for the Weighted Mean

Consider a set of measurements of a quantity as $x_1, x_2, x_3, \ldots, x_n$ each having weight $w_1, w_2, w_3, \ldots, w_n$ and denote the best estimate of this quantity as q. According to our general definition of measurements and corrections we may write:

measurement + correction (or residual) = best estimate

or $$x_1 + v_1 = q,\ x_2 + v_2 = q,\ x_3 + v_3 = q,\ \cdots,\ x_n + v_n = q$$

These equations can be rearranged as

$$v_1 = q - x_1,\ v_2 = q - x_2,\ v_3 = q - x_3,\ \cdots,\ v_n = q - x_n$$

Now each measurement has a weight reflecting relative precision and we may state the least squares principle as

> The best estimate q is that value which makes the sum of the squares of the residuals, multiplied by their weights, a minimum.

We may define a least squares function φ (phi) as

$$\varphi = \text{ the sum of the weighted squared residuals} = \sum_{k=1}^{n} w_k v_k^2 \qquad (2.13)$$

or $$\varphi = \sum_{k=1}^{n} w_k v_k^2 = w_1 (q - x_1)^2 + w_2 (q - x_2)^2 + \cdots + w_n (q - x_n)^2$$

We say that φ is a function of *q*, the single parameter or variable in this equation. The minimum value of the function (i.e., making the sum of the weighted squared residuals a minimum) can be found by equating the derivative $\frac{d\varphi}{dq}$ to zero, i.e.,

$$\varphi \text{ is a minimum when } \frac{d\varphi}{dq} = 0$$

and $$\frac{d\varphi}{dq} = 2w_1 (q - x_1) + 2w_2 (q - x_2) + \cdots + 2w_n (q - x_n) = 0$$

Cancelling the 2's and expanding gives

$$w_1 q - w_1 x_1 + w_2 q - w_2 x_2 + \cdots + w_n q - w_n x_n = 0$$

Rearranging gives the weighted arithmetic mean *q*

$$q = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum_{k=1}^{n} w_k x_k}{\sum_{k=1}^{n} w_k} \tag{2.14}$$

Hence, the weighted arithmetic mean of a series of measurements x_k each having weight w_k is the best estimate according to Gauss' least squares principle.

It should be noted that the equation for the weighted mean (2.14) is valid only for measurements that are statistically independent. If observations are dependent, then a measure of the dependence between the measurements, known as covariance, must be taken into account. A more detailed discussion of weights, variances and covariances is given in later sections of these notes.

2.4. Line of Best Fit

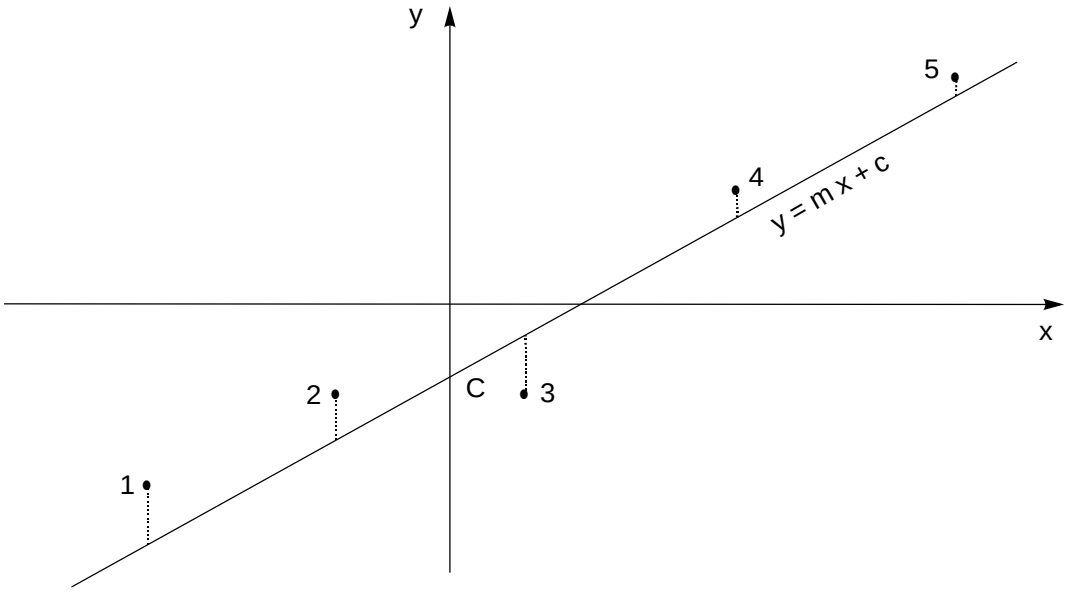


Figure 2.1 Line of Best Fit through data points 1 to 5

The line of best fit shown in the Figure 2.1 has the equation $y = m\,x + c$ where *m* is the slope of the line $\left(m = \tan\theta = \frac{y_2 - y_1}{x_2 - x_1} \right)$ and *c* is the intercept of the line on the *y* axis.

m and *c* are the parameters and the data points are assumed to accord with the mathematical model $y = m\,x + c$. Obviously, only two points are required to define a straight line and so three of the five points in Figure 2.1 are redundant measurements (or observations). In this example the *x,y* coordinate pairs of each data point are considered as indirect measurements of the parameters *m* and *c* of the mathematical model.

To estimate (or compute) values for *m* and *c*, pairs of points in all combinations (ten in all) could be used to obtain average values of the parameters; or perhaps just two points selected as representative could be used to determine *m* and *c*.

A better way is to determine a line such that it passes as close as possible to all the data points. Such a line is known as a Line of Best Fit and is obtained (visually) by minimising the differences between the line and the data points. No account is made of the "sign" of these differences, which can be considered as corrections to the measurements or residuals. The Line of Best Fit could also be defined as the result of a least squares process that determines estimates of the parameters *m* and *c* such that those values will make the sum of the squares of the residuals, multiplied by their weights, a minimum. Two examples will be considered, the first with all measurements considered as having equal precisions, i.e., all weights of equal value, and the second, measurements having different precisions, i.e., unequal weights.

2.4.1. Line of Best Fit (equal weights)

In Figure 2.1 there are five data points whose *x,y* coordinates (scaled from the diagram in mm's) are

Point	*x*	*y*
1	–40.0	–24.0
2	–15.0	–24.0
3	10.0	–12.0
4	38.0	15.0
5	67.0	30.0

Table 2.2 Coordinates of data points (mm's) shown in Figure 2.1

Assume that the data points accord with the mathematical model $y = m\,x + c$ and each measurement has equal precision. Furthermore, assume that the residuals are associated with the *y* values only, which leads to an observation equation of the form

$$y_k + v_k = m\,x_k + c \tag{2.15}$$

By adopting this observation equation we are actually saying that the measurements (the *x,y* coordinates) don't exactly fit the mathematical model, i.e., there are inconsistencies between the model and the actual measurements, and these inconsistencies (in both *x* and *y* measurements) are grouped together as residuals v_k and simply added to the left-hand-side of the mathematical model. This is simply a convenience. We could write an observation equation of the form

$$y_k + v_{y_k} = m\left(x_k + v_{x_k}\right) + c$$

v_{x_k} , v_{y_k} are residuals associated with the *x* and *y* coordinates of the k^{th} point. Observation equations of this form require more complicated least squares solutions and are not considered in this elementary section.

Equations (2.15) can be re-written as residual equations of the form

$$v_k = m\,x_k + c - y_k \tag{2.16}$$

The distinction here between observation equations and residual equations is simply that residual equations have only residuals on the left of the equals sign. Rearranging observation equations into residual equations is an interim step to simplify the function φ = sum of squares of residuals.

Since all observations are of equal precision (equal weights), the least squares function to be minimised is

$$\varphi = \text{ the sum of the squares of the residuals} = \sum_{k=1}^{n} v_k^2$$

or $$\varphi = \sum_{k=1}^{5} v_k^2 = (m\,x_1 + c - y_1)^2 + (m\,x_2 + c - y_2)^2 + ... + (m\,x_5 + c - y_5)^2$$

φ is a function of the *u* = 2 "unknown" parameters *m* and *c* and so to minimise the sum of squares of residuals, the partial derivatives $\frac{\partial \varphi}{\partial m}$ and $\frac{\partial \varphi}{\partial c}$ are equated to zero.

$$\frac{\partial \phi}{\partial m} = 2(m\,x_1 + c - y_1)(x_1) + 2(m\,x_2 + c - y_2)(x_2) + ... + 2(m\,x_5 + c - y_5)(x_5) = 0$$
$$\frac{\partial \phi}{\partial c} = 2(m\,x_1 + c - y_1)(1) + 2(m\,x_2 + c - y_2)(1) + ... + 2(m\,x_5 + c - y_5)(1) = 0$$

Cancelling the 2's, simplifying and re-arranging gives two normal equations of the form

$$\begin{aligned} m\sum_{k=1}^{n} x_k^2 + c\sum_{k=1}^{n} x_k &= \sum_{k=1}^{n} x_k y_k \\ m\sum_{k=1}^{n} x_k + c\,n &= \sum_{k=1}^{n} y_k \end{aligned} \tag{2.17}$$

The normal equations can be expressed in matrix form as

$$\begin{bmatrix} \sum x_k^2 & \sum x_k \\ \sum x_k & n \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} \sum x_k y_k \\ \sum y_k \end{bmatrix} \tag{2.18}$$

or $$\mathbf{N}\,\mathbf{x} \;=\; \mathbf{t} \tag{2.19}$$

Matrix algebra is a powerful mathematical tool that simplifies the theory associated with least squares. The student should become familiar with the terminology and proficient with the algebra. Appendix A contains useful information relating to matrix algebra.

$\mathbf{N} = \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix}$ is the (u,u) normal equation coefficient matrix

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is the $(u,1)$ vector of parameters (or "unknowns"), and

$\mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$ is the $(u,1)$ vector of numeric terms.

The solution of the normal equations for the vector of parameters is

$$\mathbf{x} \;=\; \mathbf{N}^{-1}\,\mathbf{t} \tag{2.20}$$

In this example (two equations in two unknowns) the matrix inverse $\mathbf{N}^{-1}$ is easily obtained (see Appendix A 4.8) and the solution of (2.20) is given as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{(n_{11}\,n_{22} - n_{12}\,n_{21})} \begin{bmatrix} n_{22} & -n_{12} \\ -n_{21} & n_{11} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \tag{2.21}$$

From the data given in Table 2.2, the normal equations are

$$\begin{bmatrix} 7858.00 & 60.00 \\ 60.00 & 5.00 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 3780.00 \\ -15.00 \end{bmatrix}$$

and the solutions for the best estimates of the parameters *m* and *c* are

$$\begin{bmatrix} m \\ c \end{bmatrix} = \frac{1}{(7858.00)(5.00) - (60.00)(60.00)} \begin{bmatrix} 5.00 & -60.00 \\ -60.00 & 7858.00 \end{bmatrix} = \begin{bmatrix} 0.554777 \\ -9.657327 \end{bmatrix}$$

Substitution of the best estimates of the parameters *m* and *c* into the residual equations $v_k = m\,x_k + c - y_k$ gives the residuals (mm's) as

$$
\begin{aligned}
v_1 &= -7.8 \\
v_2 &= \ \ 6.0 \\
v_3 &= \ \ 7.9 \\
v_4 &= -3.6 \\
v_5 &= -2.5
\end{aligned}
$$

2.4.2. Line of Best Fit (unequal weights)

Consider again the Line of best Fit shown in Figure 2.1 but this time the *x,y* coordinate pairs are weighted, i.e., some of the data points are considered to have more precise coordinates than others. Table 2.3 shows the *x,y* coordinates (scaled from the diagram in mm's) with weights.

Point	*x*	*y*	weight *w*
1	–40.0	–24.0	2
2	–15.0	–24.0	5
3	10.0	–12.0	7
4	38.0	15.0	3
5	67.0	30.0	3

Table 2.3 Coordinates (mm) and weights of data points shown in Figure 2.1

Similarly to before, a residual equation of the form given by (2.16) can be written for each observation but this time a weight w_k is associated with each equation and the least squares function to be minimised is

$$\varphi = \text{ the sum of the weighted squared residuals} = \sum_{k=1}^{n} w_k v_k^2$$

or $$\varphi = \sum_{k=1}^{5} w_k v_k^2 \ = \ w_1(m\,x_1 + c - y_1)^2 + w_2(m\,x_2 + c - y_2)^2 + \ldots. + w_5(m\,x_5 + c - y_5)^2$$

φ is a function of the $u = 2$ "unknown" parameters m and c and so to minimise φ the partial derivatives $\frac{\partial \varphi}{\partial m}$ and $\frac{\partial \varphi}{\partial c}$ are equated to zero.

$$\frac{\partial \phi}{\partial m} = 2w_1(m\,x_1 + c - y_1)(x_1) + 2w_2(m\,x_2 + c - y_2)(x_2) + ... + 2w_5(m\,x_5 + c - y_5)(x_5) = 0$$
$$\frac{\partial \phi}{\partial c} = 2w_1(m\,x_1 + c - y_1)(1) + 2w_2(m\,x_2 + c - y_2)(1) + ... + 2w_5(m\,x_5 + c - y_5)(1) = 0$$

Cancelling the 2's simplifying and re-arranging gives two normal equations of the form

$$m\sum_{k=1}^{n} w_k x_k^2 + c\sum_{k=1}^{n} w_k x_k = \sum_{k=1}^{n} w_k x_k y_k$$
$$m\sum_{k=1}^{n} w_k x_k + c\sum_{k=1}^{n} w_k = \sum_{k=1}^{n} w_k y_k$$

The normal equations expressed in matrix form $\mathbf{Nx} = \mathbf{t}$ are

$$\begin{bmatrix} \sum w_k x_k^2 & \sum w_k x_k \\ \sum w_k x_k & \sum w_k \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} \sum w_k x_k y_k \\ \sum w_k y_k \end{bmatrix}$$

Substituting the data in Table 2.3, the normal equations are

$$\begin{bmatrix} 22824.00 & 230.00 \\ 230.00 & 20.00 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 10620.00 \\ -117.00 \end{bmatrix}$$

The solution for the best estimates of the parameters *m* and *c* is found in exactly the same manner as before (see section 2.4.1)

$$m = \quad 0.592968$$
$$c = -12.669131$$

Substitution of *m* and *c* into the residual equations $v_k = m\,x_k + c - y_k$ gives the residuals (mm's) as

$$v_1 = -12.4$$
$$v_2 = \quad 2.4$$
$$v_3 = \quad 5.3$$
$$v_4 = -5.1$$
$$v_5 = -2.9$$

Comparing these residuals with those from the Line of Best Fit (equal weights), shows that the line has been pulled closer to points 2 and 3, i.e.; the points having largest weight.

2.5. Variances, Covariances, Cofactors and Weights

Some of the information in this section has been introduced in previously in section 2.3 The Weighted Mean and is re-stated here in the context of developing general matrix expressions for variances, covariances, cofactors and weights of sets or arrays of measurements.

In surveying applications, we may regard a measurement *x* as a possible value of a continuous random variable drawn from an infinite population. To model these populations, and thus estimate the quality of the measurements, probability density functions have been introduced. In surveying, Normal (Gaussian) probability density functions are the usual model. A probability density function is a non-negative function where the area under the curve is one. For $f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ the values of $f(x)$ are not probabilities. The probability a member of the population lies in the interval *a* to *b* is $\int_a^b f(x)\,dx$. The population mean μ, population variance σ_x^2 and the family of Normal probability density functions are given by Kreyszig (1970) as

$$\mu_x = \int_{-\infty}^{+\infty} x\, f(x)\,dx \tag{2.22}$$

$$\sigma_x^2 = \int_{-\infty}^{+\infty} \left(x - \mu_x\right)^2 f(x)\,dx \tag{2.23}$$

$$f(x;\, \mu_x, \sigma_x) = \frac{1}{\sigma_x \sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \tag{2.24}$$

Since the population is infinite, means and variances are never known, but may be estimated from a sample of size *n*. The sample mean $\bar{x}$ and sample variance s_x^2, are unbiased estimates of the population mean μ_x and population variance σ_x^2

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{2.25}$$

$$s_x^2 = \frac{1}{n-1}\sum_{k=1}^{n} \left(x_k - \bar{x}\right)^2 \tag{2.26}$$

The sample standard deviation s_x is the positive square root of the sample variance and is a measure of the precision (or dispersion) of the measurements about the mean $\bar{x}$.

In least squares applications, an observation may be the mean of a number measurements or a single measurement. In either case, it is assumed to be from an infinite population of measurements having a certain (population) standard deviation and that an estimate this standard deviation is known.

When two or more observations are jointly used in a least squares solution then the interdependence of these observations must be considered. Two measures of this interdependence are covariance and correlation. For two random variables *x* and *y* with a joint probability density function $f(x, y)$ the covariance σ_{xy} is

$$\sigma_{xy} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_x)(y - \mu_y) f(x, y)\, dx\, dy \tag{2.27}$$

and the correlation coefficient ρ is given by

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{2.28}$$

The correlation coefficient ρ will vary between ± 1. If $\rho_{xy} = 0$ random variables *x* and *y* are said to be uncorrelated and, if $\rho_{xy} = \pm 1$, *x* and *y* are linked by a linear relationship (Kreyszig 1970, pp.335-9). Correlation and statistical dependence are not the same, although both concepts are used synonymously. It can be shown that the covariance σ_{xy} is always zero when the random variables are statistically independent (Kreyszig 1970, p.137-9). Unfortunately, the reverse is not true in general. Zero covariance does not necessarily imply statistical independence. Nevertheless, for multivariate Normal probability density functions, zero covariance (no correlation) is a sufficient condition for statistical independence (Mikhail 1976, p.19).

The sample covariance s_{xy} between *n* pairs of values (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) with means $\bar{x}$ and $\bar{y}$ is (Mikhail 1976, p.43)

$$s_{xy} = \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y}) \tag{2.29}$$

Variances and covariances of observations can be conveniently represented using matrices. For *n* observations x_1 , x_2 , x_3 , ..., x_n with variances σ_1^2 , σ_2^2 , σ_3^2 , ..., σ_n^2 and covariances σ_{12} , σ_{13} , ... the variance-covariance matrix $\boldsymbol{\Sigma}$ is defined as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \ldots \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \ldots \sigma_{2n} \\ .... & .... & .... & .... \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \ldots \sigma_{nn}^2 \end{bmatrix} \tag{2.30}$$

Note that the variance-covariance matrix $\boldsymbol{\Sigma}$ is symmetric since in general $\sigma_{kj} = \sigma_{jk}$.

In practical applications of least squares, population variances and covariances are unknown and are replaced by estimates s_1^2 , s_2^2 , … , s_n^2 and s_{12} , s_{13} , … or by other numbers representing relative variances and covariances. These are known as cofactors and the cofactor matrix **Q**, which is symmetric, is defined as

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} & \ldots q_{1n} \\ q_{21} & q_{22} & q_{23} & \ldots q_{2n} \\ .... & .... & .... & .... \\ q_{n1} & q_{n2} & q_{n3} & \ldots q_{nn} \end{bmatrix} \tag{2.31}$$

The relationship between variance-covariance matrices and cofactor matrices is

$$\boldsymbol{\Sigma} = \sigma_0^2 \mathbf{Q} \tag{2.32}$$

σ_0^2 is a scalar quantity known as the variance factor. The variance factor is also known as the reference variance and the variance of an observation of unit weight (see section 2.3 for further discussion on this subject).

The inverse of the cofactor matrix **Q** is the weight matrix **W**.

$$\mathbf{W} = \mathbf{Q}^{-1} \tag{2.33}$$

Note that since **Q** is symmetric, its inverse **W** is also symmetric. In the case of uncorrelated observations, the variance-covariance matrix $\boldsymbol{\Sigma}$ and the cofactor matrix **Q** are both diagonal matrices (see Appendix A) and the weight of an observation *w* is a value that is inversely proportional to the estimate of the variance i.e.,

$$w_{kk} = \sigma_0^2 / q_{kk} \quad \text{or} \quad w_{kk} = \sigma_0^2 / s_{kk}^2 \tag{2.34}$$

For uncorrelated observations, the off-diagonal terms will be zero and the double subscripts may be replaced by single subscripts; equation (2.34) becomes

$$w_k = \sigma_0^2 / s_k^2 \tag{2.35}$$

This is the classical definition of a weight where σ_0^2 is a constant of proportionality.

Note: The concept of weights has been extensively used in classical least squares theory but is limited in its definition to the case of independent (or uncorrelated) observations. (Mikhail 1976, pp.64-65 and Mikhail and Gracie 1981, pp.66-68).

2.6. Matrices and Simple Least Squares Problems

Matrix algebra is a powerful mathematical tool that can be employed to develop standard solutions to least squares problems. The previous examples of the Line of Best Fit will be used to show the development of standard matrix equations that can be used for any least squares solution.

In previous developments, we have used a least squares function φ as meaning either the sum of squares of residuals or the sum of squares of residuals multiplied by weights.

In the Line of Best Fit (equal weights), we used the least squares function

$$\varphi = \text{ the sum of the squares of the residuals} = \sum_{k=1}^{n} v_k^2$$

If the residuals v_k are elements of a (column) vector **v**, the function φ can be written as the matrix product

$$\varphi = \sum_{k=1}^{n} v_k^2 = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \mathbf{v}^T \mathbf{v}$$

In the Line of Best Fit (unequal weights), we used the least squares function

$$\varphi = \text{ the sum of the weighted squared residuals} = \sum_{k=1}^{n} w_k v_k^2$$

If the residuals v_k are elements of a (column) vector **v** and the weights are the diagonal elements of a diagonal weight matrix **W**, the function φ can be written as the matrix product

$$\varphi = \sum_{k=1}^{n} w_k v_k^2 = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & w_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \mathbf{v}^T \mathbf{W} \mathbf{v}$$

Note that in this example the weight matrix **W** represents a set of uncorrelated measurements.

In general, we may write least squares function as a matrix equation

$$\varphi = \mathbf{v}^T \mathbf{W} \mathbf{v} \tag{2.36}$$

Note that replacing **W** with the identity matrix **I** gives the function for the case of equal weights and that for *n* observations, the order of **v** is (*n*,1), the order of **W** is (*n*,*n*) and the function $\varphi = \mathbf{v}^T \mathbf{W} \mathbf{v}$ is a scalar quantity (a single number).

In both examples of the Line of Best Fit an observation equation $y_k + v_k = m x_k + c$ was used that if re-arranged as $v_k - m x_k + c = -y_k$ yields five equations for the coordinate pairs

$$\begin{aligned} v_1 - mx_1 - c &= -y_1 \\ v_2 - mx_2 - c &= -y_2 \\ v_3 - mx_3 - c &= -y_3 \\ v_4 - mx_4 - c &= -y_4 \\ v_5 - mx_5 - c &= -y_5 \end{aligned}$$

Note that these re-arranged observation equations have all the unknown quantities *v*, *m* and *c* on the left of the equals sign and all the known quantities on the right.

These equations can be written in matrix form

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_5 \end{bmatrix} + \begin{bmatrix} -x_1 & -1 \\ -x_2 & -1 \\ \vdots & \vdots \\ -x_5 & -1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} -y_1 \\ -y_2 \\ \vdots \\ -y_5 \end{bmatrix}$$

and written symbolically as

$$\mathbf{v} + \mathbf{Bx} = \mathbf{f} \tag{2.37}$$

where $$\mathbf{f} = \mathbf{d} - \mathbf{l} \tag{2.38}$$

If n is the number of observations (equal to the number of equations) and u is the number of unknown parameters

- **v** is an (n,1) vector of residuals,
- **B** is an (n,u) matrix of coefficients,
- **x** is the (u,1) vector of unknown parameters,
- **f** is the (n,1) vector of numeric terms derived from the observations,
- **d** is an (n,1) vector of constants and
- **l** is the (n,1) vector of observations.

Note that in many least squares problems the vector **d** is zero.

By substituting (2.37) into (2.36), we can obtain an expression for the least squares function

$$\begin{aligned} \varphi &= \mathbf{v}^T \mathbf{W} \mathbf{v} \\ &= \left(\mathbf{f} - \mathbf{Bx}\right)^T \mathbf{W} \left(\mathbf{f} - \mathbf{Bx}\right) \\ &= \left(\mathbf{f}^T - \left(\mathbf{Bx}\right)^T\right) \mathbf{W} \left(\mathbf{f} - \mathbf{Bx}\right) \\ &= \left(\mathbf{f}^T - \mathbf{x}^T \mathbf{B}^T\right) \mathbf{W} \left(\mathbf{f} - \mathbf{Bx}\right) \end{aligned}$$

and multiplication, observing the rule of matrix algebra gives

$$\varphi = \mathbf{f}^T \mathbf{W} \mathbf{f} - \mathbf{f}^T \mathbf{W} \mathbf{B} \mathbf{x} - \mathbf{x}^T \mathbf{B}^T \mathbf{W} \mathbf{f} + \mathbf{x}^T \mathbf{B}^T \mathbf{W} \mathbf{B} \mathbf{x} \tag{2.39}$$

Since φ is a scalar (a number), the four terms on the right-hand-side of (2.39) are also scalars. Furthermore, since the transpose of a scalar is equal to itself, the second and third terms are equal $\left(\mathbf{f}^T \mathbf{W} \mathbf{B} \mathbf{x}\right)^T = \mathbf{x}^T \mathbf{B}^T \mathbf{W} \mathbf{f}$, remembering that **W** is symmetric hence $\mathbf{W} = \mathbf{W}^T$, giving

$$\varphi = \mathbf{f}^T \mathbf{W} \mathbf{f} - 2\mathbf{f}^T \mathbf{W} \mathbf{B} \mathbf{x} + \mathbf{x}^T \left(\mathbf{B}^T \mathbf{W} \mathbf{B}\right) \mathbf{x} \tag{2.40}$$

In equation (2.40) all matrices and vectors are numerical constants except **x**, the vector of unknown parameters, therefore for the least squares function φ to be a minimum, its partial derivative with respect to each element in vector **x** must be equated to zero, i.e., φ will be a minimum when $\frac{\partial \varphi}{\partial \mathbf{x}} = \mathbf{0}^T$. The first term of (2.40) does not contain **x** so its derivative is automatically zero and the second and third terms are bilinear and quadratic forms respectively and their derivatives are given in Appendix A, hence φ will be a minimum when

$$\frac{\partial \varphi}{\partial \mathbf{x}} = -2\mathbf{f}^T\mathbf{W}\mathbf{B} + 2\mathbf{x}^T\left(\mathbf{B}^T\mathbf{W}\mathbf{B}\right) = \mathbf{0}^T$$

Cancelling the 2's, re-arranging and transposing gives a set of normal equations

$$\left(\mathbf{B}^T\mathbf{W}\mathbf{B}\right)\mathbf{x} = \mathbf{B}^T\mathbf{W}\mathbf{f} \tag{2.41}$$

Equation (2.41) is often given in the form

$$\mathbf{N}\mathbf{x} = \mathbf{t} \tag{2.42}$$

where $\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B}$ is a (u,u) coefficient matrix (the normal equation coefficient matrix),

$\mathbf{x}$ is the (u,1) vector of unknown parameters and

$\mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f}$ is a (u,1) vector of numeric terms.

The solution for the vector of parameters **x** is given by

$$\mathbf{x} = \mathbf{N}^{-1}\mathbf{t} \tag{2.43}$$

After solving for the vector **x**, the residuals are obtained from

$$\mathbf{v} = \mathbf{f} - \mathbf{B}\mathbf{x} \tag{2.44}$$

and the vector of "adjusted" or estimated observations $\hat{\mathbf{l}}$ is

$$\hat{\mathbf{l}} = \mathbf{l} + \mathbf{v} \tag{2.45}$$

The "hat" symbol (^) is used to denote quantities that result from a least squares process. Such quantities are often called adjusted quantities or least squares estimates.

These equations are the standard matrix solution for

least squares adjustment of indirect observations.

The name "least squares adjustment of indirect observations", adopted by Mikhail (1976) and Mikhail & Gracie (1981), recognises the fact that each observation is an indirect measurement of the unknown parameters. This is the most common technique employed in surveying and geodesy and is described by various names, such as

parametric least squares
least squares adjustment by observation equations
least squares adjustment by residual equations

The technique of least squares adjustment of indirect observations has the following characteristics

- A mathematical model (equation) links observations, residuals (corrections) and unknown parameters.
- For *n* observations, there is a minimum number n_0 required to determine the *u* unknown parameters. In this case $n_0 = u$ and the number of redundant observations is $r = n - n_0$.
- An equation can be written for each observation, i.e., there are *n* observation equations. These equations can be represented in a standard matrix form; see equation (2.37), representing *n* equations in *u* unknowns and solutions for the unknown parameters, residuals and adjusted observations obtained from equations (2.41) to (2.45).

The popularity of this technique of adjustment is due to its easy adaptability to computer-programmed solutions. As an example, the following MATLAB program *best_fit_line.m* reads a text file containing coordinate pairs (measurements) *x* and *y* and a weight *w* (a measure of precision associated with each coordinate pair) and computes the parameters *m* and *c* of a line of best fit $y = mx + c$.

MATLAB program *best_fit_line*

```
function best_fit_line
%
% BEST_FIT_LINE reads an ASCII textfile containing coordinate pairs (x,y)
%   and weights (w) associated with each pair and computes the parameters
%   m and c of the line of best fit y = mx + c using the least squares
%   principle.  Results are written to a textfile having the same path and
%   name as the data file but with the extension ".out"

%=========================================================================
% Function:  best_fit_line
%
% Author:
%  Rod Deakin,
%  Department of Geospatial Science, RMIT University,
%  GPO Box 2476V, MELBOURNE VIC 3001
%  AUSTRALIA
%  email: rod.deakin@rmit.edu.au
%
% Date:
%  Version 1.0  18 March 2003
%
% Remarks:
%  This function reads numeric data from a textfile containing coordinate
%  pairs (x,y) and weights (w) associated with each pair and computes the
%  parameters m and c of a line of best fit y = mx + c using the least
%  squares principle.  Results are written to a textfile having the same
%  path and name as the data file but with the extension ".out"
%
% Arrays:
%  B       - (n,u) coeff matrix of observation equation v + Bx = f
%  f       - (n,1) vector of numeric terms
%  N       - (u,u) coefficient matrix of Normal equations Nx = t
%  Ninv    - (u,u) inverse of N
%  t       - (u,1) vector of numeric terms of Normal equations Nx = t
%  v       - (n,1) vector of residuals
%  W       - (n,n) weight matrix
%  weight  - (n,1) vector of weights
%  x       - (u,1) vector of solutions
%  x_coord - (n,1) vector of x coordinates
%  y_coord - (n,1) vector of y coordinates
%
%
% Variables
%  n       - number of equations
%  u       - number of unknowns
%
% References:
%  Notes on Least Squares (2003), Department of Geospatial Science, RMIT
%      University, 2003
%
%=========================================================================


%-------------------------------------------------------------------------
% 1. Call the User Interface (UI) to choose the input data file name
% 2. Concatenate strings to give the path and file name of the input file
% 3. Strip off the extension from the file name to give the rootName
% 4. Add extension ".out" to rootName to give the output filename
% 5. Concatenate strings to give the path and file name of the output file
%-------------------------------------------------------------------------
filepath = strcat('c:\temp\','*.dat');
[infilename,inpathname] = uigetfile(filepath);
infilepath = strcat(inpathname,infilename);
rootName   = strtok(infilename,'.');
```

MATLAB program *best_fit_line*

```
outfilename = strcat(rootName,'.out');
outfilepath = strcat(inpathname,outfilename);

%-----------------------------------------------------------
% 1. Load the data into an array whose name is the rootName
% 2. set fileTemp = rootName
% 3. Copy columns of data into individual arrays
%-----------------------------------------------------------
load(infilepath);
fileTemp = eval(rootName);
x_coord = fileTemp(:,1);
y_coord = fileTemp(:,2);
weight  = fileTemp(:,3);

% determine the number of equations
n = length(x_coord);

% set the number of unknowns
u = 2;

% set the elements of the weight matrix W
W = zeros(n,n);
for k = 1:n
  W(k,k) = weight(k);
end

% form the coefficient matrix B of the observation equations
B = zeros(n,u);
for k = 1:n
  B(k,1) = -x_coord(k);
  B(k,2) = -1;
end

% for the vector of numeric terms f
f = zeros(n,1);
for k = 1:n
  f(k,1) = -y_coord(k);
end

% form the normal equation coefficient matrix N
N = B'*W*B;

% form the vector of numeric terms t
t = B'*W*f;

% solve the system Nx = t for the unknown parameters x
Ninv = inv(N);
x = Ninv*t;

% compute residuals
v = f - (B*x);

% open the output file print the data
fidout  = fopen(outfilepath,'wt');

fprintf(fidout,'\n\nLine of Best Fit Least Squares Solution');

fprintf(fidout,'\n\nInput Data');
fprintf(fidout,'\n     x(k)       y(k)     weight w(k)');
for k = 1:n
  fprintf(fidout,'\n%10.4f %10.4f %10.4f',x_coord(k),y_coord(k),weight(k));
end
```

MATLAB program *best_fit_line*

```
fprintf(fidout,'\n\nCoefficient matrix B of observation equations v + Bx = f');
for j = 1:n
  fprintf(fidout,'\n');
  for k = 1:u
    fprintf(fidout,'%10.4f',B(j,k));
  end
end

fprintf(fidout,'\n\nVector of numeric terms f of observation equations v + Bx =
f');
for k = 1:n
  fprintf(fidout,'\n%10.4f',f(k,1));
end

fprintf(fidout,'\n\nCoefficient matrix N of Normal equations Nx = t');
for j = 1:u
  fprintf(fidout,'\n');
  for k = 1:u
    fprintf(fidout,'%12.4f',N(j,k));
  end
end

fprintf(fidout,'\n\nVector of numeric terms t of Normal equations Nx = t');
for k = 1:u
  fprintf(fidout,'\n%10.4f',t(k,1));
end

fprintf(fidout,'\n\nInverse of Normal equation coefficient matrix');
for j = 1:u
  fprintf(fidout,'\n');
  for k = 1:u
    fprintf(fidout,'%16.4e',Ninv(j,k));
  end
end

fprintf(fidout,'\n\nVector of solutions x');
for k = 1:u
  fprintf(fidout,'\n%10.4f',x(k,1));
end

fprintf(fidout,'\n\nVector of residuals v');
for k = 1:n
  fprintf(fidout,'\n%10.4f',v(k,1));
end

fprintf(fidout,'\n\n');

% close the output file
fclose(fidout);
```

MATLAB program *best_fit_line*

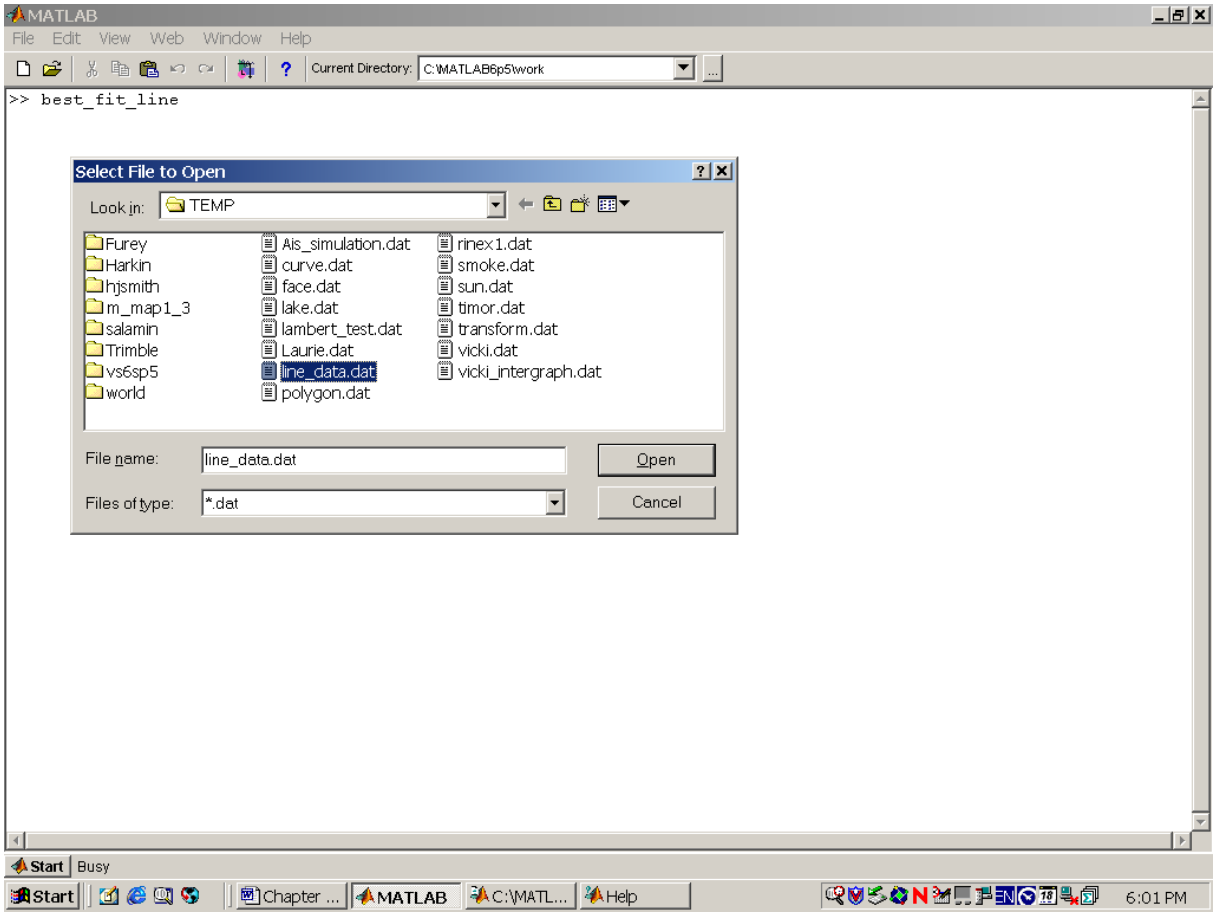
Data file `c:\Temp\line_data.dat`

```
% data file for function "best_fit_line.m"
%   x       y      w
  -40.0  -24.0     2
  -15.0  -24.0     5
   10.0  -12.0     7
   38.0   15.0     3
   67.0   30.0     3
```

MATLAB program *best_fit_line*

Running the program from the MATLAB command window prompt >> opens up a standard Microsoft Windows file selection window in the directory c:\Temp. Select the appropriate data file (in this example: `line_data.dat`) by double clicking with the mouse and the program reads the data file, computes the solutions and writes the output data to the file `c:\Temp\line_data.out`

MATLAB command window



MATLAB program *best_fit_line*

Output file `c:\Temp\line_data.out`

```
Line of Best Fit Least Squares Solution

Input Data
     x(k)       y(k)     weight w(k)
  -40.0000   -24.0000     2.0000
  -15.0000   -24.0000     5.0000
   10.0000   -12.0000     7.0000
   38.0000    15.0000     3.0000
   67.0000    30.0000     3.0000

Coefficient matrix B of observation equations v + Bx = f
   40.0000   -1.0000
   15.0000   -1.0000
  -10.0000   -1.0000
  -38.0000   -1.0000
  -67.0000   -1.0000

Vector of numeric terms f of observation equations v + Bx = f
   24.0000
   24.0000
   12.0000
  -15.0000
  -30.0000

Coefficient matrix N of Normal equations Nx = t
  22824.0000    230.0000
    230.0000     20.0000

Vector of numeric terms t of Normal equations Nx = t
10620.0000
 -117.0000

Inverse of Normal equation coefficient matrix
     4.9556e-005    -5.6990e-004
    -5.6990e-004     5.6554e-002

Vector of solutions x
    0.5930
  -12.6691

Vector of residuals v
  -12.3878
    2.4363
    5.2605
   -5.1363
   -2.9403
```

The data in this example is taken from section **2.4.2 Line of Best Fit (unequal weights)**

MATLAB program *best_fit_line*

By adding the following lines to the program, the Line of Best Fit is shown on a plot together with the data points.

```
%--------------------------------------
% plot data points and line of best fit
%--------------------------------------

%  copy solutions from vector x
m = x(1,1);
c = x(2,1);

% find minimum and maximum x coordinates
xmin = min(x_coord);
xmax = max(x_coord);

% create a vector of x coordinates at intervals of 0.1
% between min and max coordinates
x = xmin:0.1:xmax;

% calculate y coordinates of Line of Best Fit
y = m*x + c;

% Select Figure window and clear figure
figure(1);
clf(1);
hold on;
grid on;
box on;

% plot line of best fit and then the data points with a star (*)
plot(x,y,'k-');
plot(x_coord,y_coord,'k*');

% anotate the plot
title('Least Squares Line of Best Fit')
xlabel('X coordinate');
ylabel('Y coordinate');
```

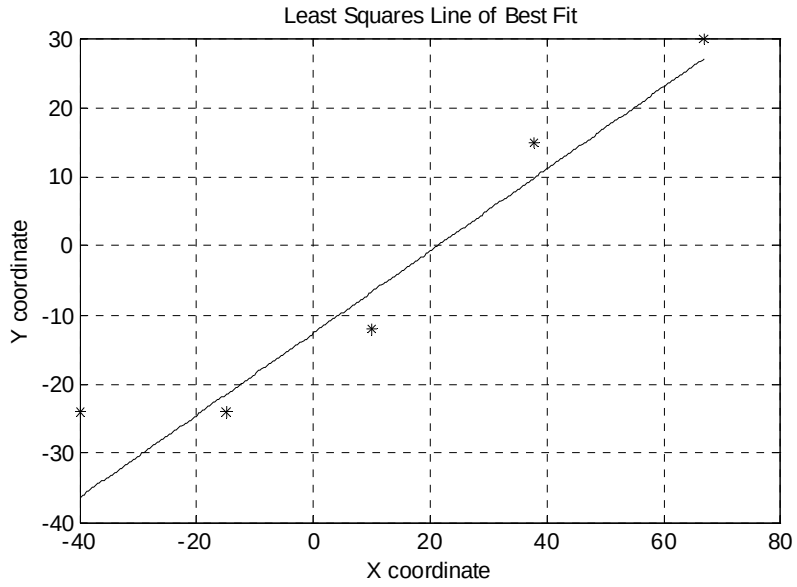


Figure 2.3 Least Squares Line of Best Fit

2.7. Least Squares Curve Fitting

The general matrix solutions for least squares adjustment of indirect observations (see equations (2.37) to (2.45) of section 2.6) can be applied to curve fitting. The following two examples (parabola and ellipse) demonstrate the technique.

2.7.1. Least Squares Best Fit Parabola

Consider the following: A surveyor working on the re-alignment of a rural road is required to fit a parabolic vertical curve such that it is a best fit to the series of natural surface Reduced Levels (RL's) on the proposed new alignment. Figure 2.2 shows a Vertical Section of the proposed alignment with Chainages (*x*-values) and RL's (*y*-values).

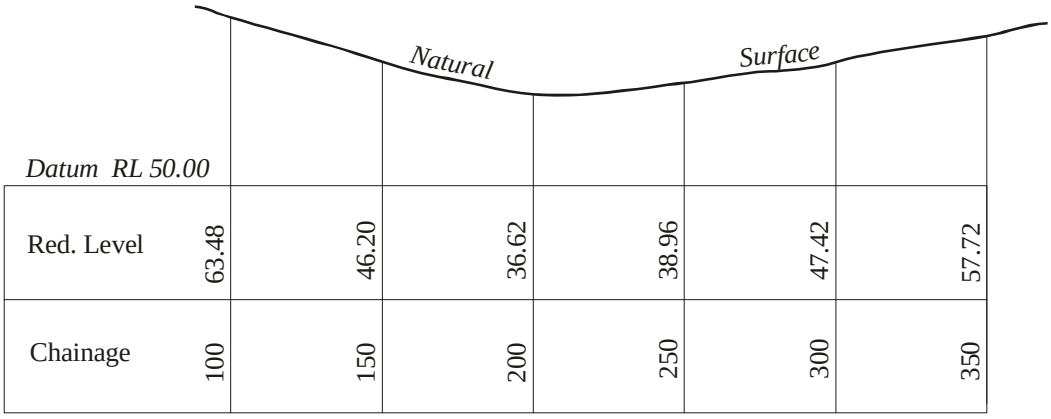


Figure 2.2 Vertical Section of proposed road alignment

The general equation of a parabolic curve is

$$y = ax^2 + bx + c \tag{2.46}$$

This is the mathematical model that we assume our data accords with and to account for the measurement inconsistencies, due to the irregular natural surface and small measurement errors we can add residuals to the left-hand-side of (2.46) to give an observation equation

$$y_k + v_k = ax_k^2 + bx_k + c \tag{2.47}$$

Equation (2.47) can be re-arranged as

$$v_k - ax_k^2 - bx - c = -y_k \tag{2.48}$$

n = 6 equations in u = 3 unknown parameters a, b, c can be written in matrix form $\mathbf{v} + \mathbf{Bx} = \mathbf{f}$ as

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_6 \end{bmatrix} + \begin{bmatrix} -x_1^2 & -x_1 & -1 \\ -x_2^2 & -x_2 & -1 \\ \vdots & \vdots & \vdots \\ -x_6^2 & -x_6 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} -y_1 \\ -y_2 \\ \vdots \\ -y_6 \end{bmatrix}$$

where

$$\mathbf{v}_{(6,1)} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_6 \end{bmatrix}, \qquad \mathbf{B}_{(6,3)} = \begin{bmatrix} -x_1^2 & -x_1 & -1 \\ -x_2^2 & -x_2 & -1 \\ \vdots & \vdots & \vdots \\ -x_6^2 & -x_6 & -1 \end{bmatrix}, \qquad \mathbf{x}_{(3,1)} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \qquad \mathbf{f}_{(6,1)} = \begin{bmatrix} -y_1 \\ -y_2 \\ \vdots \\ -y_6 \end{bmatrix}$$

Considering all the measurements to be of equal precision, i.e., $\mathbf{W} = \mathbf{I}$, the least squares solution for the three parameters in the vector $\mathbf{x}$ is given by the following sequence of operations

- form the normal coefficient matrix: $\mathbf{N} = \mathbf{B}^T\mathbf{WB}$
- form the vector of numeric terms: $\mathbf{t} = \mathbf{B}^T\mathbf{Wf}$
- compute the matrix inverse: $\mathbf{N}^{-1}$
- compute the solutions: $\mathbf{x} = \mathbf{N}^{-1}\mathbf{t}$
- compute the residuals: $\mathbf{v} = \mathbf{f} - \mathbf{Bx}$

This is the identical series of operations to solve for the parameters of the Line of Best Fit, except in this case u = 3. With minor modifications to the MATLAB program *best_fit_line.m* another MATLAB program *best_fit_parabola.m* can be created to determine the parameters *a, b, c* of the best fit parabola. The relevant modifications are shown below.

MATLAB program *best_fit_parabola*

Making the following changes to the MATLAB program *best_fit_line*, a new program *best_fit_parabola* can be created.

Changes to function name and help instructions

```
function best_fit_parabola
%
% BEST_FIT_PARABOLA reads an ASCII textfile containing coordinate pairs (x,y)
%   and weights (w) associated with each pair and computes the parameters
%   a, b and c of a best fit parabola y = a(x*x) + bx + c using the least
%   squares principle.  Results are written to a textfile having the same
%   path and name as the data file but with the extension ".out"
```

Changes to function remarks in documentation section

```
% Remarks:
%  This function reads numeric data from a textfile containing coordinate
%  pairs (x,y) and weights (w) associated with each pair and computes the
%  parameters a, b, and c of a best fit parabola y = a(x*x) + bx + c using
%  the least squares principle.  Results are written to a textfile having
%  the same path and name as the data file but with the extension ".out"
```

Changes to formation of coefficient matrix **B**

```
% form the coefficient matrix B of the observation equations
B = zeros(n,u);
for k = 1:n
  B(k,1) = -(x_coord(k)^2);
  B(k,2) = -x_coord(k);
  B(k,3) = -1;
end
```

Changes to data plotting section

```
%-----------------------------------------
% plot data points and Parabola of best fit
%-----------------------------------------

%  copy solutions from vector x
a = x(1,1);
b = x(2,1);
c = x(3,1);

% find minimum and maximum x coordinates
xmin = min(x_coord);
xmax = max(x_coord);

% create a vector of x coordinates at intervals of 0.1
% between min and max coordinates
x = xmin:0.1:xmax;

% calculate y coordinates of Parabola of Best Fit
y = a*(x.*x) + b*x + c;
```

MATLAB program *best_fit_parabola*

Using the data from Figure 2.2 a data file `c:\Temp\parabola_data.dat` was created

```
% data file for function "best_fit_parabola.m"
%   x       y      w
 100.0    63.48    1
 150.0    46.20    1
 200.0    36.62    1
 250.0    38.96    1
 300.0    47.42    1
 350.0    57.72    1
```

Running the program from the MATLAB command window generated the following output file `c:\Temp\parabola_data.out` and a plot of the Least Squares Parabola of best Fit

```
Parabola of Best Fit Least Squares Solution

Input Data
     x(k)       y(k)     weight w(k)
  100.0000    63.4800     1.0000
  150.0000    46.2000     1.0000
  200.0000    36.6200     1.0000
  250.0000    38.9600     1.0000
  300.0000    47.4200     1.0000
  350.0000    57.7200     1.0000

Coefficient matrix B of observation equations v + Bx = f
    -10000.0000      -100.0000        -1.0000
    -22500.0000      -150.0000        -1.0000
    -40000.0000      -200.0000        -1.0000
    -62500.0000      -250.0000        -1.0000
    -90000.0000      -300.0000        -1.0000
   -122500.0000      -350.0000        -1.0000

Vector of numeric terms f of observation equations v + Bx = f
       -63.4800
       -46.2000
       -36.6200
       -38.9600
       -47.4200
       -57.7200

Coefficient matrix N of Normal equations Nx = t
29218750000.0000   97875000.0000     347500.0000
   97875000.0000     347500.0000       1350.0000
     347500.0000       1350.0000          6.0000

Vector of numeric terms t of Normal equations Nx = t
  16912600.0000
     64770.0000
       290.4000

Inverse of Normal equation coefficient matrix
     4.2857e-009    -1.9286e-006     1.8571e-004
    -1.9286e-006     8.9071e-004    -8.8714e-002
     1.8571e-004    -8.8714e-002     9.3714e+000

Vector of solutions x
       0.001500
      -0.688221
     116.350000
```

MATLAB program *best_fit_parabola*

```
Vector of residuals v
         -0.948
          0.676
          2.103
         -0.889
         -2.498
          1.555
```

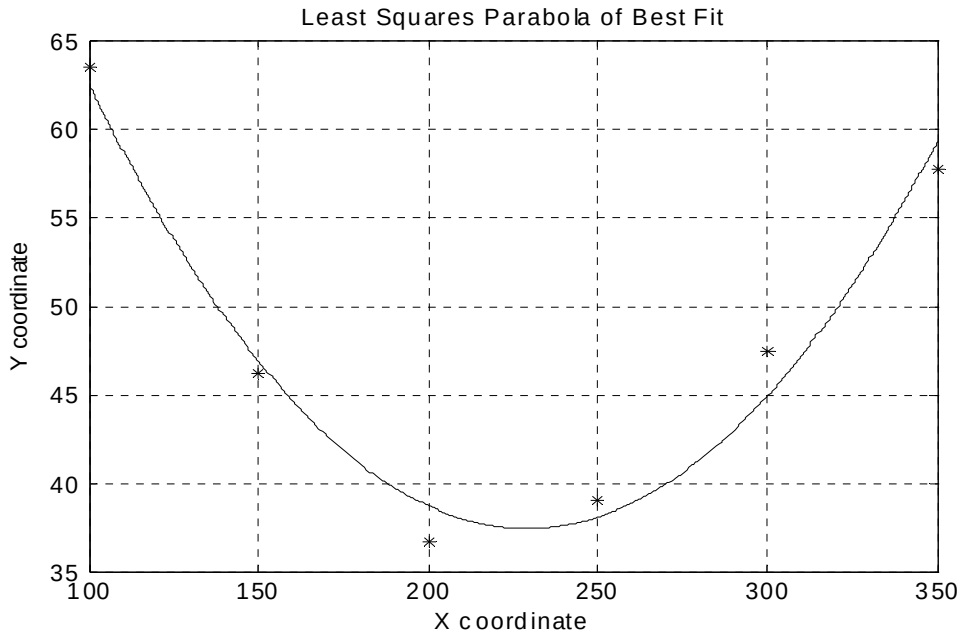


Figure 2.4 Least Squares Parabola of Best Fit

2.7.2. Least Squares Best Fit Ellipse

In November 1994, a survey was undertaken by staff of the Department of Geospatial Science at the Melbourne Cricket Ground (MCG) to determine the dimensions of the playing surface. This survey was to decide which of two sets of dimensions was correct, those of the Melbourne Cricket Club (MCC) or those of the Australian Football League (AFL). The MCC curator Tony Ware and the AFL statistician Col Hutchison both measured the length of the ground (Tony Ware with a 100-metre nylon tape and Col Hutchison with a measuring wheel) and compared their distances with the "true" distance determined by Electronic Distance Measurement (EDM) with a Topcon 3B Total Station. Their measurements were both reasonably close to the EDM distance and it turned out that the "official" AFL dimensions

were incorrect. After this "measure-off", observations (bearings and distances) were made to seventeen points around the edge of the playing surface to determine the Least Squares Ellipse of Best Fit and to see if the major axis of this ellipse was the actual line between the goals at either end. The Total Station was set-up close to the goal-to-goal axis and 20-25 metres from the centre of the ground. An arbitrary *X,Y* coordinate system was used with the origin at the Total Station and the positive *X*-axis in the direction of the Brunton Avenue end of the Great Southern Stand (approximately west). The table of coordinates is given below; point numbers 1 to 6 were not points on the edge of the ground.

Point No.	*X*-coordinate	*Y*-coordinate
7	-54.58	17.11
8	-45.47	36.56
9	-28.40	53.22
10	-2.02	63.72
11	28.12	63.44
12	57.49	52.55
13	80.85	34.20
14	98.08	9.14
15	105.69	-17.30
16	103.83	-46.96
17	88.42	-71.50
18	61.26	-86.84
19	26.47	-91.07
20	-6.59	-81.37
21	-34.55	-59.24
22	-51.51	-29.28
23	-56.30	-2.31

Table 2.4 Arbitrary coordinates of points around the perimeter of the playing surface of the MCG (date of survey November 1994)

To develop an observation equation for the Least Squares Ellipse of Best Fit and to determine the lengths and directions of the axes of the ellipse the following derivation of the *general equation of an ellipse* is necessary.

Figure 2.5 shows an ellipse whose axes are aligned with the *u*-*v* axes. The semi-axes lengths are *a* and *b* $(a > b)$, the centre of the ellipse is at X_0, Y_0 and the ellipse axes are rotated by an

angle β, measured positive anti-clockwise from the *x*-axis. The *x*-*y* axes are parallel to the *X*-*Y* axes and pass through the centre of the ellipse.

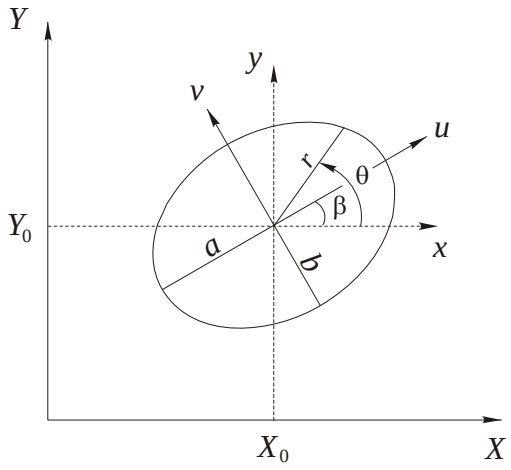


Figure 2.5

The *u*,*v* Cartesian equation of the ellipse is

$$\frac{u^2}{a^2}+\frac{v^2}{b^2}=1 \tag{2.49}$$

Equation (2.49) can be expressed in matrix form as

$$\begin{bmatrix} u & v \end{bmatrix}\begin{bmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}=1 \tag{2.50}$$

The *u*,*v* axes are rotated (positive anti-clockwise) by an angle β from the *x*,*y* axes and the relationship between coordinates is shown in Figure 2.6

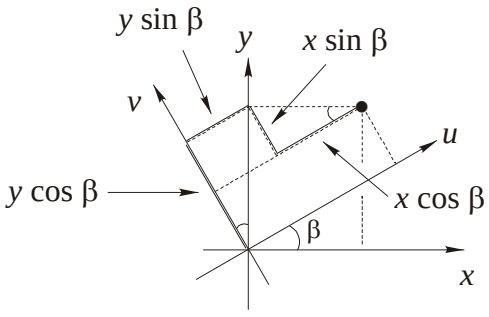


Figure 2.6

Inspection of Figure 2.6 shows

$$\begin{aligned} u &= x\cos\beta + y\sin\beta \\ v &= -x\sin\beta + y\cos\beta \end{aligned} \tag{2.51}$$

Replacing $\cos\beta$ and $\sin\beta$ with the letters c and s the coordinate relationships can be represented as a matrix equation

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \mathrm{c} & \mathrm{s} \\ -\mathrm{s} & \mathrm{c} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.52}$$

Transposing this equation (remembering the reversal rule with the transpose of matrix products) gives

$$\begin{bmatrix} u & v \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \mathrm{c} & -\mathrm{s} \\ \mathrm{s} & \mathrm{c} \end{bmatrix} \tag{2.53}$$

Substituting (2.52) and (2.53) into (2.50) and multiplying the matrices gives

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \mathrm{c} & -\mathrm{s} \\ \mathrm{s} & \mathrm{c} \end{bmatrix} \begin{bmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{bmatrix} \begin{bmatrix} \mathrm{c} & \mathrm{s} \\ -\mathrm{s} & \mathrm{c} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \left(\dfrac{\mathrm{c}^2}{a^2} + \dfrac{\mathrm{s}^2}{b^2}\right) & \left(\dfrac{\mathrm{cs}}{a^2} - \dfrac{\mathrm{cs}}{b^2}\right) \\ \left(\dfrac{\mathrm{cs}}{a^2} - \dfrac{\mathrm{cs}}{b^2}\right) & \left(\dfrac{\mathrm{s}^2}{a^2} + \dfrac{\mathrm{c}^2}{b^2}\right) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$

Replacing the elements of the square matrix with the symbols *A, B* and *H,* noting that the top-right and lower-left elements are the same, this equation may be written in a general form as

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} A & H \\ H & B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1$$

or $$Ax^2 + 2Hxy + By^2 = 1 \tag{2.54}$$

Equation (2.54) is the equation of an ellipse centred at the coordinate origin but with axes rotated from the *x,y* axes. The semi axes lengths *a* and *b,* and the rotation angle β can be determined from (2.54) by the following method.

Letting $x = r\cos\theta$ and $y = r\sin\theta$ in equation (2.54) gives the <u>polar equation of the ellipse</u>

$$A\cos^2\theta + 2H\cos\theta\sin\theta + B\sin^2\theta = \frac{1}{r^2} \tag{2.55}$$

r is the radial distance from the centre of the ellipse and θ is the angle measured positive anti-clockwise from the *x*-axis. Equation (2.55) has maximum and minimum values defining the lengths and directions of the axes of the ellipse. To determine these values from (2.55), consider the following

Let
$$\frac{1}{r^2} = f = A\cos^2\theta + 2H\cos\theta\sin\theta + B\sin^2\theta$$
$$= A\cos^2\theta + H\sin 2\theta + B\sin^2\theta$$

and aim to find the optimal (maximum and minimum) values of *f* and the values of θ when these occur by investigating the first and second derivatives f' and f'' respectively, i.e.,

$$f \text{ is } \left\{\begin{matrix}\max\\ \min\end{matrix}\right\} \text{ when } \left\{\begin{matrix} f' = 0 \text{ and } f'' < 0\\ f' = 0 \text{ and } f'' > 0\end{matrix}\right\}$$

where
$$f' = (B - A)\sin 2\theta + 2H\cos 2\theta$$
$$f'' = 2(B - A)\cos 2\theta - 4H\sin 2\theta \tag{2.56}$$

Now the maximum or minimum value of *f* occurs when $f' = 0$ and from the first member of (2.56) the value of θ is given by

$$\tan 2\theta = \frac{2H}{A - B} \tag{2.57}$$

But this value of θ could relate to either a maximum or a minimum value of *f*. So from the second member of equations (2.56) with a value of 2θ from equation (2.57) this ambiguity can be resolved by determining the sign of the second derivative f'' giving

$$\left\{\begin{matrix} f_{\max}\\ f_{\min}\end{matrix}\right\} \text{ when } \left\{\begin{matrix} f'' < 0\\ f'' > 0\end{matrix}\right\}$$

In the polar equation of the ellipse given by equation (2.55) $f_{\max}$ coincides with $r_{\min}$ and $f_{\min}$ coincides with $r_{\max}$ so the angle β (measured positive anti-clockwise) from the *x*-axis to the major axis of the ellipse (see Figure 2.5) is found from

$$\left\{\begin{matrix} r_{\max}\\ r_{\min}\end{matrix}\right\} \text{ when } \left\{\begin{matrix} f'' > 0\\ f'' < 0\end{matrix}\right\} \text{ and } \left\{\begin{matrix} \beta = \theta\\ \beta = \theta - \frac{1}{2}\pi\end{matrix}\right\} \tag{2.58}$$

These results can be verified by considering the definitions of *A*, *B* and *H* used in the derivation of the polar equation of the ellipse, i.e.,

$$A=\frac{\cos^2\beta}{a^2}+\frac{\sin^2\beta}{b^2},\ B=\frac{\sin^2\beta}{a^2}+\frac{\cos^2\beta}{b^2},\ H=\frac{\cos\beta\sin\beta}{a^2}-\frac{\cos\beta\sin\beta}{b^2}$$

and $$A-B=\left(\frac{1}{a^2}-\frac{1}{b^2}\right)\cos 2\beta,\ 2H=\left(\frac{1}{a^2}-\frac{1}{b^2}\right)\sin 2\beta$$

giving $$\tan 2\beta=\frac{2H}{A-B}$$

Noting that the values of θ coinciding with the maximum or minimum values of the function *f* are found from equation (2.57) then $\tan 2\beta=\frac{2H}{A-B}=\tan 2\theta$ or

$$\tan 2\theta=\tan 2\beta$$

whereupon

$$2\theta=2\beta+n\pi \ \text{ or } \ \theta=\beta+\tfrac{1}{2}n\pi \ \text{ where } n \text{ is an integer}$$

Also, from the second member of equations (2.56)

$$f''=2\left(B-A\right)\cos 2\theta-4H\sin 2\theta$$

Now, for $n=0$

$$\theta=\beta,\ f''\big|_{\theta=\beta}=-2\left(\frac{1}{a^2}-\frac{1}{b^2}\right) \text{ and since } a>b,\ f''\big|_{\theta=\beta}>0$$

So $\theta=\beta$ makes *f* minimum and so *r* is maximum and

$$\begin{aligned} f_{\min} &= A\cos^2\beta+2H\sin\beta\cos\beta+B\sin^2\beta \\ &= \frac{\left(\cos^2\beta+\sin^2\beta\right)^2}{a^2}=\frac{1}{a^2} \end{aligned}$$

So $r_{\max}=a$

When $n=1$

$$\theta=\beta+\tfrac{1}{2}\pi,\ \sin 2\theta=-\sin 2\beta,\ \cos 2\theta=-\cos 2\beta \text{ and so}$$

$$f''\big|_{\theta=\beta+\frac{1}{2}\pi}=2\left(\frac{1}{a^2}-\frac{1}{b^2}\right) \text{ and since } a>b,\ f''\big|_{\theta=\beta+\frac{1}{2}\pi}<0$$

So $\theta=\beta+\tfrac{1}{2}\pi$ makes *f* maximum and so *r* is minimum and

$$f_{\max} = \frac{\left(\sin^2\beta + \cos^2\beta\right)^2}{b^2} = \frac{1}{b^2}$$

So $r_{\min} = b$

When $n = 2$

$$\theta = \beta + \pi,\ \sin 2\theta = \cos 2\beta,\ \cos 2\theta = \cos 2\beta \text{ and } f''\big|_{\theta=\beta+\pi} > 0$$

So $\theta = \beta + \frac{1}{2}\pi$ makes $f_{\min} = \frac{1}{a^2}$ and $r_{\max} = a$

When $n = 3$

$$\theta = \beta + \tfrac{3}{2}\pi,\ \sin 2\theta = -\cos 2\beta,\ \cos 2\theta = -\cos 2\beta \text{ and } f''\big|_{\theta=\beta+\frac{3}{2}\pi} < 0$$

So $\theta = \beta + \frac{3}{2}\pi$ makes $f_{\max} = \frac{1}{b^2}$ and $r_{\min} = b$

All other even values of *n* give the same result as $n = 2$ and all other odd values of *n* give the same result as $n = 1$

Now consider Figure 2.5 and the general Cartesian equation of the ellipse, re-stated again as

$$aX^2 + 2hXY + bY^2 + dX + eY = 1 \qquad (2.59)$$

where the translated *x,y* coordinate system is related to the *X,Y* system by

$$X = x + X_0 \quad \text{and} \quad Y = y + Y_0$$

Substituting these relationships into (2.59) gives

$$a\left(x + X_0\right)^2 + 2h\left(x + X_0\right)\left(y + Y_0\right) + b\left(y + Y_0\right)^2 + d\left(x + X_0\right) + e\left(y + Y_0\right) = 1$$

Expanding and gathering terms gives

$$\begin{aligned} ax^2 + 2hxy + by^2 &+ \left(2aX_0 + 2hY_0 + d\right)x \\ &+ \left(2aY_0 + 2hX_0 + e\right)y \\ &+ aX_0^2 + 2hX_0Y_0 + bY_0^2 + dX_0 + eY_0 = 1 \end{aligned}$$

Inspection of the left-hand-side of this equation reveals three parts:

(i) $ax^2 + 2hxy + by^2$ is the left-hand-side of the equation of an ellipse, similar in form to equation (2.54) ,

(ii) coefficient terms of *x* and *y;* $\left(2aX_0 + 2hY_0 + d\right)$ and $\left(2aY_0 + 2hX_0 + e\right)$,

(iii) a constant term $aX_0^2 + 2hX_0Y_0 + bY_0^2 + dX_0 + eY_0$

Now when the coefficients of *x* and *y* are zero the ellipse will be centred at the origin of the *x*,*y* axes with an equation of the form

$$ax^2 + 2hxy + by^2 = c \tag{2.60}$$

where $$c = 1 - \left(aX_0^2 + 2hX_0Y_0 + bY_0^2 + dX_0 + eY_0\right) \tag{2.61}$$

and $$\begin{aligned} 2aX_0 + 2hY_0 + d &= 0 \\ 2hX_0 + 2bY_0 + e &= 0 \end{aligned} \tag{2.62}$$

Equations (2.62) can be written in matrix form and solved (using the inverse of a 2,2 matrix) to give X_0 and Y_0

$$\begin{bmatrix} d \\ e \end{bmatrix} = \begin{bmatrix} -2a & -2h \\ -2h & -2b \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix}$$

$$\begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} = \frac{1}{2ab - 2h^2} \begin{bmatrix} -b & h \\ h & -a \end{bmatrix} \begin{bmatrix} d \\ e \end{bmatrix}$$

giving $$X_0 = \frac{eh - bd}{2\left(ab - h^2\right)} \quad \text{and} \quad Y_0 = \frac{dh - ae}{2\left(ab - h^2\right)} \tag{2.63}$$

Dividing both sides of (2.60) by *c* gives

$$Ax^2 + 2Hxy + By^2 = 1 \tag{2.64}$$

where $$A = \frac{a}{c}, \quad H = \frac{h}{c}, \quad B = \frac{b}{c}$$

Equation (2.64), identical to equation (2.54), is the equation of an ellipse centred at the *x*,*y* coordinate origin whose axes are rotated from the *x*,*y* axes by an angle β. The rotation angle β and semi-axes lengths *a* and *b* of the ellipse can be determined using the method set out above and equations (2.58), (2.57), (2.56) and (2.55). Thus, we can see from the development that the general Cartesian equation of an ellipse is given by

$$aX^2 + 2hXY + bY^2 + dX + eY = 1 \tag{2.65}$$

Note that the coefficients *a* and *b* in this equation are not the semi-axes lengths of the ellipse.

Returning to the problem of the Least Squares Ellipse of Best Fit for the MCG, the size, shape location and orientation of this ellipse can be determined from a set of observation equations of the form

$$v_k + aX_k^2 + 2hX_kY_k + bY_k^2 + dX_k + eY_k = 1 \qquad (2.66)$$

This observation equation is the general Cartesian equation of an ellipse with the addition of the residual v_k. The addition of v_k to the left-hand-side of (2.65) is simply a convenience and reflects the fact that the measured coordinates X_k, Y_k are inconsistent with the mathematical model. For each of the 17 measured points around the perimeter of the MCG an equation can be written and arranged in the matrix form $\mathbf{v} + \mathbf{Bx} = \mathbf{f}$

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{17} \end{bmatrix} + \begin{bmatrix} X_1^2 & X_1Y_1 & Y_1^2 & X_1 & Y_1 \\ X_2^2 & X_2Y_2 & Y_2^2 & X_2 & Y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{17}^2 & X_{17}Y_{17} & Y_{17}^2 & X_{17} & Y_{17} \end{bmatrix} \begin{bmatrix} a \\ 2h \\ b \\ d \\ e \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The vector **x** contains the parameters *a, h, b, d* and *e* of the general equation of the ellipse and with a weight matrix **W** = **I** (i.e., all observations of equal precision) the solution for **x** is given by the following sequence of operations

- form the normal coefficient matrix: $\mathbf{N} = \mathbf{B}^T\mathbf{WB}$
- form the vector of numeric terms: $\mathbf{t} = \mathbf{B}^T\mathbf{Wf}$
- compute the matrix inverse: $\mathbf{N}^{-1}$
- compute the solutions: $\mathbf{x} = \mathbf{N}^{-1}\mathbf{t}$
- compute the residuals: $\mathbf{v} = \mathbf{f} - \mathbf{Bx}$

This is the identical series of operations to solve for the parameters of the Line of Best Fit, and for the Parabola of Best Fit except in this case the residuals **v** have little practical meaning because they are not connected to quantities such as distances or coordinates. In the case of the Least Squares Ellipse of Best Fit, it is better to compute the offsets *h* (perpendicular distances) from the ellipse to the data points rather than the residuals **v**.

To compute offsets *h* the following preliminary sequence of operations is required:

(i) compute the parameters *a, h, b, d* and *e* using the Least Squares process set out above.

(ii) compute the coordinates of the origin X_0, Y_0 and the constant *c* using equations (2.63) and (2.61).

(iii) compute coefficients *A, H* and *B* of the ellipse given by (2.64) which can then be used to compute the rotation angle β and the semi-axes lengths *a* and *b* from equations (2.55) to (2.58) and.

(iv) compute the *u,v* coordinates of the data points using equations (2.51).

Now, having the *u,v* coordinates, the offsets *h* can be computed. Consider the sectional view of a quadrant of an ellipse in Figure 2.7. The *u,v* axes are in the direction of the major and minor axes respectively (*a* and *b* are the semi-axes lengths) and *P* is a point related to the ellipse by the normal, which makes an angle ϕ with the major axis, and the distance *h = QP* along the normal. The *u,v* coordinates of *P* are the distances *LP* and *MP* respectively. From the geometry of an ellipse, the normal intersects the minor axis at *H* and the distance $QH = \nu$ (where ν is the Greek symbol *nu*) and the distances *DH* and *OH* are νe^2 and $\nu e^2 \sin\phi$ respectively. *e* is the eccentricity of the ellipse and the eccentricity and flattening *f* of an ellipse are related to the semi-axes *a* and *b*.

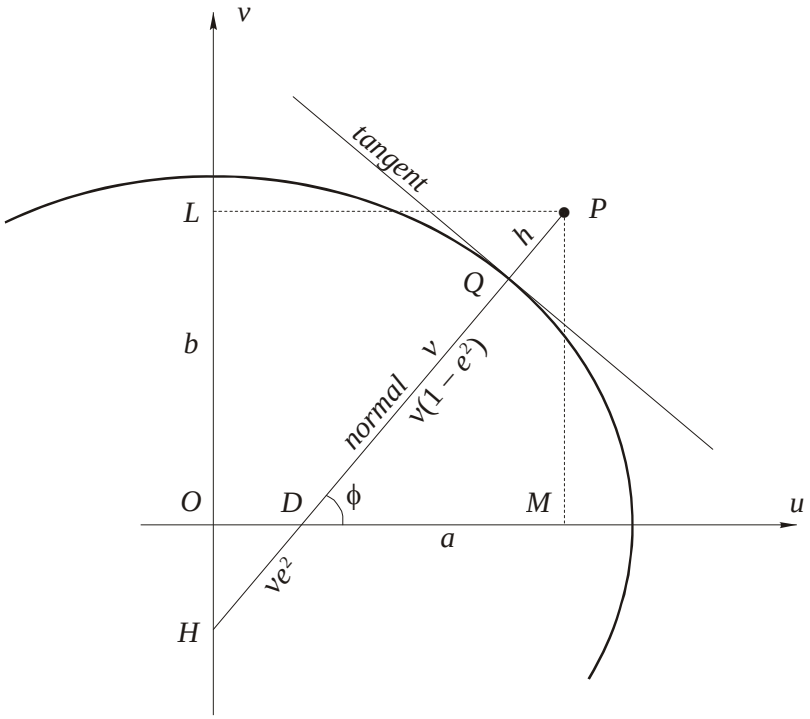


Figure 2.7 *u,v* coordinates of *P* related to an ellipse (*a,b*)

The equations for *f,* e^2 and ν are

$$f = \frac{a-b}{a}$$

$$e^2 = f\left(2-f\right)$$

$$\nu = \frac{a}{\sqrt{1-e^2\sin^2\phi}}$$

Using these relationships, the angle ϕ and perpendicular distance *h* are given by

$$\tan\phi = \frac{v+\nu e^2\sin\phi}{u} \tag{2.67}$$

$$h = \frac{u}{\cos\phi} - \nu \tag{2.68}$$

Inspecting these equations; if the semi-axes *a,b* and the *u,v* coordinates of *P* are known, the perpendicular offset *h* can be determined from (2.68) and (2.67). It should be noted that functions of ϕ appear on both sides of the equals sign of equation (2.67) and ϕ must be solved by iteration.

To determine the Least Square Best Fit Ellipse for the playing surface of the MCG a MATLAB program *best_fit_ellipse.m* operating in the same way as the MATLAB programs *best_fit_line* and *best_fit_parabola* with a data file (in this example: `MCG_ellipse_data.dat`)

Data file `c:\Temp\MCG_ellipse_data.dat`

```
% Data file for MCG Survey, November 1994
% Coordinates of 17 boundary points (edge of concrete)
%
% point      X       Y    weight
    7     -54.58   17.11    1
    8     -45.47   36.56    1
    9     -28.40   53.22    1
   10      -2.02   63.72    1
   11      28.12   63.44    1
   12      57.49   52.55    1
   13      80.85   34.20    1
   14      98.08    9.14    1
   15     105.69  -17.30    1
   16     103.83  -46.96    1
   17      88.42  -71.50    1
   18      61.26  -86.84    1
   19      26.47  -91.07    1
   20      -6.59  -81.37    1
   21     -34.55  -59.24    1
   22     -51.51  -29.28    1
   23     -56.30   -2.31    1
```

gives the following results (contained in an output file having the same name and path as the data file but with the extension .dat) and plot on the screen.

Output file `c:\Temp\MCG_ellipse_data.out`

```
Ellipse of Best Fit Least Squares Solution

Input Data
 point     x(k)         y(k)       weight w(k)
  7     -54.5800      17.1100       1.0000
  8     -45.4700      36.5600       1.0000
  9     -28.4000      53.2200       1.0000
 10      -2.0200      63.7200       1.0000
 11      28.1200      63.4400       1.0000
 12      57.4900      52.5500       1.0000
 13      80.8500      34.2000       1.0000
 14      98.0800       9.1400       1.0000
 15     105.6900     -17.3000       1.0000
 16     103.8300     -46.9600       1.0000
 17      88.4200     -71.5000       1.0000
 18      61.2600     -86.8400       1.0000
 19      26.4700     -91.0700       1.0000
 20      -6.5900     -81.3700       1.0000
 21     -34.5500     -59.2400       1.0000
 22     -51.5100     -29.2800       1.0000
 23     -56.3000      -2.3100       1.0000
```

```
 General Equation of Ellipse with X,Y origin not at centre of ellipse
 aXX + 2hXY + bYY + dX + eY = 1
 a =  1.720717e-004
 h =  2.690541e-005
 b =  1.865607e-004
 d = -7.743828e-005
 e =  3.729881e-005

 Equation of Ellipse with x,y origin at centre of ellipse
 Axx + 2Hxy + Byy = 1
 A =  1.535544e-004
 H =  2.401002e-005
 B =  1.664842e-004

 Ellipse parameters
 semi-major axis a =   86.017
 semi-minor axis b =   73.544

 Bearing of major axis
 beta(degrees) =   -37.465030
 beta(DMS)     =  -37 27 54.11
 Brg(degrees)  =   127.465030
 Brg(DMS)      =  127 27 54.11

 Coordinates of centre of ellipse
 X(centre) =       24.620
 Y(centre) =      -13.547
```

Data and offsets to ellipse of best fit

pt	offset	X	Y	x	y	u	v
7	0.129	-54.580	17.110	-79.200	30.657	-81.511	-23.842
8	0.159	-45.470	36.560	-70.090	50.107	-86.111	-2.863
9	0.162	-28.400	53.220	-53.020	66.767	-82.696	20.744
10	0.164	-2.020	63.720	-26.640	77.267	-68.145	45.124
11	0.060	28.120	63.440	3.500	76.987	-44.051	63.236
12	-0.090	57.490	52.550	32.870	66.097	-14.116	72.457
13	-0.223	80.850	34.200	56.230	47.747	15.588	72.101
14	-0.216	98.080	9.140	73.460	22.687	44.507	62.691
15	-0.123	105.690	-17.300	81.070	-3.753	66.630	46.334
16	0.936	103.830	-46.960	79.210	-33.413	83.195	21.661
17	0.284	88.420	-71.500	63.800	-57.953	85.891	-7.191
18	-1.181	61.260	-86.840	36.640	-73.293	73.664	-35.887
19	-0.224	26.470	-91.070	1.850	-77.523	48.624	-60.407
20	0.627	-6.590	-81.370	-31.210	-67.823	16.483	-72.817
21	0.554	-34.550	-59.240	-59.170	-45.693	-19.171	-72.259
22	-0.438	-51.510	-29.280	-76.130	-15.733	-50.856	-58.796
23	-0.703	-56.300	-2.310	-80.920	11.237	-71.063	-40.303

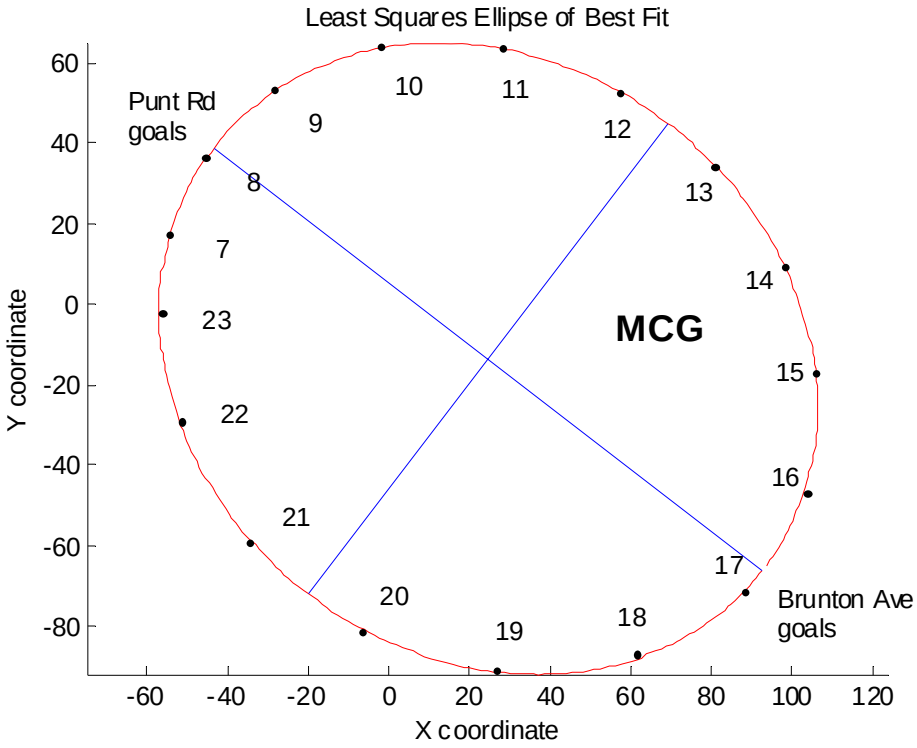


Figure 2.8 Plot of Least Squares Best Fit Ellipse and data points for the MCG

The MATLAB program *best_fit_ellipse.m* calls two other MATLAB functions *ellipse.m* (a function to compute the coordinates of points on an ellipse) and *DMS.m* (a function to convert decimal degree to degrees, minutes and seconds). A copy of these programs is shown below.

MATLAB program *best_fit_ellipse*

```
function best_fit_ellipse
%
% BEST_FIT_ELLIPSE reads an ASCII textfile containing point numbers of
%   coordinate pairs (X,Y) and weights (W) associated with each pair and
%   computes the cordinates of the origin, the lengths of the axes and
%   the rotation angle of the Best Fit Ellipse using the least squares
%   principle.  Results are written to a textfile having the same path
%   and name as the data file but with the extension ".out"

%=========================================================================
% Function:  best_fit_ellipse
%
% Author:
%  Rod Deakin,
%  School of Mathematical and Geospatial Sciences, RMIT University,
%  GPO Box 2476V, MELBOURNE VIC 3001
%  AUSTRALIA
%  email: rod.deakin@rmit.edu.au
%
% Date:
%  Version 1.0  22 March 2003
%  Version 1.1  10 May   2003
%  Version 1.2   9 November 2005
%
% Functions Required:
%    [X,Y] = ellipse(a,b,theta)
%  [D,M,S] = DMS(DecDeg)
%
% Remarks:
%  The general equation of an ellipse is
%    aXX + 2hXY + bYY + dX + eY = 1
%  This function computes the parameters a,h,b,d,e of a Least Squares Best Fit
%  Ellipse given a set of X,Y coordinate pairs and weights (w) associated with
%  each pair.  The centre of the best fit ellipse is at Xo = (eh-db)/(2ab-2hh)
%  and Yo = (dh-ea)/(2ab-2hh).
%  A constant c = 1 - (aXoXo + 2hXoYo + bYoYo + dXo + eYo) is divided into a,
%  h and b giving A = a/c, H = h/c and B = b/c which are the parameters of an
%  ellipse Axx +2Hxy + Byy = 1.  The major axis of this ellipse is rotated
%  from the coordinate axes by an angle beta which can be determined from the
%  polar equation of an ellipse
%  A*cos_squared(theta) + 2H*cos(theta)*sin(theta) + B*sin_squared(theta) =
1/r_squared
%  The maximum and minimum values of this function occur for theta given by
%  tan(2*theta) = 2H/(A-B) and the angle beta is determined by evaluating
%  the sign of the second derivative of the polar equation.  This angle is
%  substituted into the polar equation to determine the length of the semi-major
%  axis length a.  Beta - 90 degrees will give the length of the semi-minor
%  axis length b
%  Note that the semi-axes lengths a,b are not the same as the parameters
%  a and b in the general equation of the ellipse.
%  Results are written to a textfile having the same path and name as the
%  data file but with the extension ".out"
%
% References:
%  Notes on Least Squares (2005), Geospatial Science, RMIT
%      University, 2005
%
```

MATLAB program *best_fit_ellipse*

```
% Arrays:
%  B       -  coeff matrix of observation equation v + Bx = f
%  f       -  vector of numeric terms
%  N       -  coefficient matrix of Normal equations Nx = t
%  Ninv    -  inverse of N
%  p       -  vector of perpendicular distances from ellipse to points
%  point   -  vector of point numbers
%  t       -  vector of numeric terms of Normal equations Nx = t
%  u,v     -  vectors of u,v coords of ellipse
%  W       -  weight matrix
%  weight  -  vector of weights
%  x       -  vector of solutions
%  x,y     -  vectors of x,y coords of ellipse
%  x_coord -  vector of X coordinates
%  y_coord -  vector of Y coordinates
%  xpt,ypt -  vectors of coords for point number locations on plot
%  Xpt,Ypt -  vectors Xpt = xpt + Xc, Ypt = ypt + Yc
%
%
% Variables:
%  A,B,H   - parameters of ellipse Axx + 2Hxy + Byy = 1
%  a,h,b,  - parameters of ellipse aXX + 2hXY + bYY + dX + eY = 1
%  d,e
%  a1,b1   - semi-major and semi-minor axes of ellipse
%  beta    - angle between x-axis and major axis of ellipse (degrees)
%  brg     - bearing of major axis (u-axis) of ellipse (degrees)
%  c       - constant of translated ellipse or cos(x)
%  d2r     - degree to radian conversion factor = 180/pi = 57.29577951...
%  e2      - eccentricity squared
%  flat    - flattening of ellipse
%  f_dd    - second derivative of the function "f" where f is the polar
%            equation of an ellipse
%  lat     - latitude (radians) of point related to an ellipse
%  n       - number of equations
%  new_lat - new latitude in iteration
%  nu      - radius of curvature in prime meridian
%  pion2   - 90 degrees or pi/2
%  u       - number of unknowns
%  s       - sin(x)
%  s1,s2   - sin(lat) and sin_squared(lat)
%  scale   - scale factor to reduce size of numbers in normal equations
%  theta   - angle for which polar equation of ellipse gives max/min
%            values
%  two_theta - 2*theta
%  Xc,Yc   - coords of centre of ellipse X = x + Xc, Y = y + Yc
%  X0,Y0   - scaled coords of centre of ellipse
%
%=========================================================================

%
% Set program constants
d2r   = 180/pi;
pion2 = pi/2;
scale = 100;
```

MATLAB program *best_fit_ellipse*

```
%--------------------------------------------------------------------------
% 1. Call the User Interface (UI) to choose the input data file name
% 2. Concatenate strings to give the path and file name of the input file
% 3. Strip off the extension from the file name to give the rootName
% 4. Add extension ".out" to rootName to give the output filename
% 5. Concatenate strings to give the path and file name of the output file
%--------------------------------------------------------------------------
filepath = strcat('c:\temp\','*.dat');
[infilename,inpathname] = uigetfile(filepath);
infilepath = strcat(inpathname,infilename);
rootName   = strtok(infilename,'.');
outfilename = strcat(rootName,'.out');
outfilepath = strcat(inpathname,outfilename);

%-----------------------------------------------------------
% 1. Load the data into an array whose name is the rootName
% 2. set fileTemp = rootName
% 3. Copy columns of data into individual arrays
%-----------------------------------------------------------
load(infilepath);
fileTemp = eval(rootName);
point   = fileTemp(:,1);
x_coord = fileTemp(:,2);
y_coord = fileTemp(:,3);
weight  = fileTemp(:,4);

% Determine the number of equations and set the number of unknowns
n = length(point);
u = 5;

% Set the elements of the weight matrix W
W = zeros(n,n);
for k = 1:n
  W(k,k) = weight(k);
end

% Form the coefficient matrix B of the observation equations.
% Note that the coordinates are scaled by a factor 1/100 to
% reduce the size of equations.
B = zeros(n,u);
for k = 1:n
  B(k,1) = (x_coord(k)/scale)^2;
  B(k,2) = (x_coord(k)/scale)*(y_coord(k)/scale);
  B(k,3) = (y_coord(k)/scale)^2;
  B(k,4) = x_coord(k)/scale;
  B(k,5) = y_coord(k)/scale;
end

% Form the vector of numeric terms f
f = ones(n,1);

% Form the normal equation coefficient matrix N
% and the vector of numeric terms t
N = B'*W*B;
t = B'*W*f;

% Compute the inverse and solve the system Nx = t
Ninv = inv(N);
x = Ninv*t;
```

MATLAB program *best_fit_ellipse*

```
% Copy the results into the variables a,h,b,d,e
a = x(1,1);
h = x(2,1)/2;
b = x(3,1);
d = x(4,1);
e = x(5,1);

% Compute the coordinates of the centre of the ellipse
X0 = (e*h - b*d)/(2*(a*b - h*h));
Y0 = (d*h - a*e)/(2*(a*b - h*h));
Xc = X0*scale;
Yc = Y0*scale;

% Compute the variables A,H,B and then the lengths
% of the axes and the rotation angle beta
c  = 1-(a*X0*X0 + 2*h*X0*Y0 + b*Y0*Y0 + d*X0 + e*Y0);
A = a/c;
H = h/c;
B = b/c;

% compute the angle theta for max or min
two_theta = atan2(2*H,(A-B));
% compute second derivative
f_dd = 2*(B-A)*cos(two_theta) - 4*H*sin(two_theta);
% test the second derivative to determine max or min
theta     = two_theta/2;
if f_dd < 0
    beta = theta - pion2;
else
    beta = theta;
end
% compute semi-major axis length
c = cos(beta);
s = sin(beta);
a1 = sqrt(1/(A*c*c + 2*H*c*s + B*s*s))*scale;
% compute semi-minor axis length
c = cos(beta+pion2);
s = sin(beta+pion2);
b1 = sqrt(1/(A*c*c + 2*H*c*s + B*s*s))*scale;
% convert beta to degrees
beta = beta*d2r;

% Calculate bearing of major axis noting that the rotation angle
% beta is considered positive anti-clockwise from the X-axis to
% the major axis of the ellipse
brg = 360+90-beta;
if(brg>360)
  brg = brg-360;
end

%-------------------------------------------------------------------------
% Compute perpendicular distances from points to the ellipse of best fit
%-------------------------------------------------------------------------
% Create a set of u,v coordinates by first reducing the X,Y coords
% to x,y coordinates and then rotating these coordinates by the
% rotation angle beta.  The u-axis is the major axis of the ellipse.
x = x_coord-Xc;
y = y_coord-Yc;
for k=1:n
  u(k,1) =  x(k)*cos(beta/d2r) + y(k)*sin(beta/d2r);
  v(k,1) = -x(k)*sin(beta/d2r) + y(k)*cos(beta/d2r);
end
```

MATLAB program *best_fit_ellipse*

```
% Compute the flattening and eccentricity squared for the ellipse
flat = (a1-b1)/a1;
e2   = flat*(2-flat);

% Compute the distance along the normal to the ellipse passing
% through the point
for k=1:n
  % Compute the latitude of the normal to the ellipse through
  % the point by iteration.  nu is the radius of curvature of
  % the prime vertical normal section
  lat     = pi/2;
  new_lat = atan2(v(k,1),u(k,1));
  while(abs(new_lat-lat)>1e-10)
    lat     = new_lat;
    s1      = sin(lat);
    s2      = s1*s1;
    nu      = a1/sqrt(1-e2*s2);
    new_lat = atan2((v(k,1)+nu*e2*s1),u(k,1));
  end
  % p is the distance along the normal from the ellipse to the point
  p(k,1) = (u(k,1)/cos(lat))-nu;
end

%-----------------------------------------------------
% Compute the coordinate locations for a point number
% to be shown on the plot.  These locations used in
% in the plot routines below.
%-----------------------------------------------------
for k=1:n
  theta  = atan2(x(k),y(k));
  if theta<0
    theta = theta + 2*pi;
  end
  r      = sqrt(x(k)^2 + y(k)^2)-10;
  xpt(k) = r*sin(theta);
  ypt(k) = r*cos(theta);
end
Xpt = xpt + Xc;
Ypt = ypt + Yc;


%-----------------------------
% print the data to the screen
%-----------------------------
fprintf('\n Ellipse of Best Fit\n');
fprintf('\n General Equation of Ellipse with X,Y origin not at centre of ellipse');
fprintf('\n aXX + 2hXY + bYY + dX + eY = 1');
fprintf('\n a = %14.6e',a/scale^2);
fprintf('\n h = %14.6e',h/scale^2);
fprintf('\n b = %14.6e',b/scale^2);
fprintf('\n d = %14.6e',d/scale^2);
fprintf('\n e = %14.6e\n',e/scale^2);
fprintf('\n Equation of Ellipse with x,y origin at centre of ellipse');
fprintf('\n Axx + 2Hxy + Byy = 1');
fprintf('\n A = %14.6e',A/scale^2);
fprintf('\n H = %14.6e',H/scale^2);
fprintf('\n B = %14.6e\n',B/scale^2);
fprintf('\n Ellipse parameters');
fprintf('\n semi-major axis a = %8.3f',a1);
fprintf('\n semi-minor axis b = %8.3f\n',b1);
fprintf('\n Bearing of major axis');
fprintf('\n beta(degrees) = %12.6f',beta);
[D,M,S] = DMS(beta);
fprintf('\n beta(DMS)      = %4d %2d %5.2f',D,M,S);
fprintf('\n Brg(degrees)   = %12.6f',brg);
```

MATLAB program *best_fit_ellipse*

```
[D,M,S] = DMS(brg);
fprintf('\n Brg(DMS)       = %4d %2d %5.2f\n',D,M,S);
fprintf('\n Coordinates of centre of ellipse');
fprintf('\n X(centre) = %12.3f',Xc);
fprintf('\n Y(centre) = %12.3f\n',Yc);
fprintf('\n Data and offsets to ellipse of best fit');
fprintf('\n  pt      offset       X          Y          x          y          u
v');
for k=1:n
  fprintf('\n %3d  %10.3f %10.3f %10.3f %10.3f %10.3f %10.3f
%10.3f',point(k),p(k,1),x_coord(k),y_coord(k),x(k),y(k),u(k,1),v(k,1));
end

fprintf('\n\n');

%----------------------------------
% print the data to the output file
%----------------------------------

% Open the output file
fidout  = fopen(outfilepath,'wt');

fprintf(fidout,'\n\nEllipse of Best Fit Least Squares Solution');

fprintf(fidout,'\n\nInput Data');
fprintf(fidout,'\n point     x(k)         y(k)       weight w(k)');
for k = 1:n
  fprintf(fidout,'\n%3d %12.4f %12.4f
%12.4f',point(k),x_coord(k),y_coord(k),weight(k));
end

fprintf(fidout,'\n\n General Equation of Ellipse with X,Y origin not at centre of
ellipse');
fprintf(fidout,'\n aXX + 2hXY + bYY + dX + eY = 1');
fprintf(fidout,'\n a = %14.6e',a/scale^2);
fprintf(fidout,'\n h = %14.6e',h/scale^2);
fprintf(fidout,'\n b = %14.6e',b/scale^2);
fprintf(fidout,'\n d = %14.6e',d/scale^2);
fprintf(fidout,'\n e = %14.6e\n',e/scale^2);
fprintf(fidout,'\n Equation of Ellipse with x,y origin at centre of ellipse');
fprintf(fidout,'\n Axx + 2Hxy + Byy = 1');
fprintf(fidout,'\n A = %14.6e',A/scale^2);
fprintf(fidout,'\n H = %14.6e',H/scale^2);
fprintf(fidout,'\n B = %14.6e\n',B/scale^2);
fprintf(fidout,'\n Ellipse parameters');
fprintf(fidout,'\n semi-major axis a = %8.3f',a1);
fprintf(fidout,'\n semi-minor axis b = %8.3f\n',b1);
fprintf(fidout,'\n Bearing of major axis');
fprintf(fidout,'\n beta(degrees) = %12.6f',beta);
[D,M,S] = DMS(beta);
fprintf(fidout,'\n beta(DMS)      = %4d %2d %5.2f',D,M,S);
fprintf(fidout,'\n Brg(degrees)   = %12.6f',brg);
[D,M,S] = DMS(brg);
fprintf(fidout,'\n Brg(DMS)       = %4d %2d %5.2f\n',D,M,S);
fprintf(fidout,'\n Coordinates of centre of ellipse');
fprintf(fidout,'\n X(centre) = %12.3f',Xc);
fprintf(fidout,'\n Y(centre) = %12.3f\n',Yc);
fprintf(fidout,'\n Data and offsets to ellipse of best fit');
fprintf(fidout,'\n  pt      offset       X          Y          x          y
u          v');
for k=1:n
  fprintf(fidout,'\n %3d  %10.3f %10.3f %10.3f %10.3f %10.3f %10.3f
%10.3f',point(k),p(k,1),x_coord(k),y_coord(k),x(k),y(k),u(k,1),v(k,1));
end
```

MATLAB program *best_fit_ellipse*

```
fprintf(fidout,'\n\n');

% Close the output file
fclose(fidout);

%-------------------------------------------------------------------
% Call function 'ellipse' with parameters a,b,theta and receive back
% X,Y coordinates whose origin is at the centre of the ellipse
%-------------------------------------------------------------------
[X,Y] = ellipse(a1,b1,beta);
X = X + Xc;
Y = Y + Yc;

%-------------------------------------------------------------------
% Set the X,Y coordinates of the major and minor axes of the ellipse
%-------------------------------------------------------------------
aX = [X(180) X(360)];
aY = [Y(180) Y(360)];
bX = [X(90) X(270)];
bY = [Y(90) Y(270)];

%-------------------------------------------------
% plot the ellipse of Best Fit and the data points
%-------------------------------------------------
figure(1);
clf(1);
plot(X,Y,'r-',aX,aY,'b-',bX,bY,'b-');
hold on;
plot(x_coord,y_coord,'k.');
axis equal;
box off;

% plot the point numbers inside the ellipse
point_string=int2str(point);
text(Xpt,Ypt,point_string);

% anotate the plot
title('Least Squares Ellipse of Best Fit')
xlabel('X coordinate');
ylabel('Y coordinate');
```

MATLAB program *ellipse*

```
function [X,Y] = ellipse(a,b,theta)

% ELLIPSE[X Y] = (A,B,THETA)  Function to compute the X,Y coordinates of
%   an ellipse given semi-axes A and B and a rotation angle THETA.
%   The angle THETA is considered to be positive anti-clockwise from the x-axis

% set degree to radian conversion factor
d2r = pi/180;

% Calculate u,v coordinates of ellipse using parametric equations
% u = a*cos(psi)
% v = b*sin(psi)
% where the u-axis is the major axis, the v-axis is the minor axis
% and psi is the auxiliary angle measured positive anti-clockwise
% from the u-axis to a point moving around the auxiliary circle of
% radius a.  The x,y coordinates are computed by rotating the ellipse
% axes by an angle theta, considered as positive anti-clockwise from the x-axis.
% x = u*cos(theta) - v*sin(theta)
% y = u*sin(theta) + v*cos(theta)

for k=1:360
  u    = a*cos(k*d2r);
  v    = b*sin(k*d2r);
  X(k) = u*cos(theta*d2r) - v*sin(theta*d2r);
  Y(k) = u*sin(theta*d2r) + v*cos(theta*d2r);
end
return
```

MATLAB program *DMS*

```
function [D,M,S] = DMS(DecDeg)
% [D,M,S] = DMS(DecDeg)  This function takes an angle in decimal degrees
%   and returns Degrees, Minutes and Seconds

val = abs(DecDeg);
D = fix(val);
M = fix((val-D)*60);
S = (val-D-M/60)*3600;
if(DecDeg<0)
  D = -D;
end
return
```

3. PROPAGATION OF VARIANCES

In least squares problems, where measurements (with associated estimates of variances and covariances) are used to determine the best estimates unknown quantities it is important to be able to determine the precisions of these estimated (or calculated) quantities. To do this requires an understanding of propagation of variances so that certain rules and techniques can be developed.

3.1. Some Statistical Definitions

Students studying Least Squares must become familiar with statistical definitions, terminology and rules. Some of these rules and definitions have been introduced in earlier sections of these notes, e.g., in Chapter 2 the definition and classification of measurements and measurement errors was discussed as well as the rules for computing means and variances for finite and infinite populations. In addition, Chapter 2 contains sections explaining matrix representations of variances and covariances, known as variance-covariance matrices $\boldsymbol{\Sigma}$ and the related cofactor matrices $\mathbf{Q}$ and weight matrices $\mathbf{W}$. The following sections in this chapter repeat some of the rules and definitions already introduced as well as expanding on some concepts previously mentioned.

3.1.1. Experiments, Sets, Sample Spaces, Events and Probability

The term statistical experiment can be used to describe any process by which several chance observations are obtained. All possible outcomes of an experiment comprise a set called the sample space and a set or sample space contains N elements or members. An event is a subset of the sample space containing n elements. Experiments, sets, sample spaces and events are the fundamental "tools" used to determine the probability of certain events where probability is defined as

$$P\left(Event\right) = \frac{n}{N} \tag{3.1}$$

For example, if a card is drawn from a deck of playing cards, what is the probability that it is a heart? In this case, the experiment is the drawing of the card and the possible outcomes of the experiment could be one of 52 different cards, i.e., the sample space is the set of $N = 52$

possible outcomes and the event is the subset containing $n = 13$ hearts. The probability of drawing a heart is

$$P(\text{Heart}) = \frac{n}{N} = \frac{13}{52} = 0.25$$

This definition of probability is a simplification of a more general concept of probability that can be explained in the following manner (see Johnson & Leone, 1964, pp.32-3).

> Suppose observations are made on a series of occasions (often termed trials) and during these trials it is noted whether or not a certain event occurs. The event can be almost any observable phenomenon, for example, that the height of a person walking through a doorway is greater than 1.8 metres, that a family leaving a cinema contains three children, that a defective item is selected from an assembly line, and so on. These trials could be conducted twice a week for a month, three times a day for six months or every hour for every day for 10 years. In the theoretical limit, the number of trials *N* would approach infinity and we could assume, at this point, that we had noted every possible outcome. Therefore, as $N \to \infty$ then *N* becomes the number of elements in the sample space containing all possible outcomes of the trials. Now for each trial we note whether or not a certain event occurs, so that at the end of *N* trials we have noted n_N events. The probability of the event (if it in fact occurs) can then be defined as
>
> $$P(Event) = \lim_{N \to \infty} \left(\frac{n_N}{N} \right)$$
>
> Since n_N and *N* are both non-negative numbers and n_N is not greater than *N* then
>
> $$0 \le \frac{n_N}{N} \le 1$$
>
> Hence
>
> $$0 \le P\{Event\} \le 1$$
>
> If the event occurs at every trial then $n_N = N$ and $n_N / N = 1$ for all *N* and so $P(Event) = 1$. This relationship can be described as: *the probability of a certain (or sure) event is equal to 1.*

If the event never occurs, then $n_N = 0$ and $n_N / N = 0$ for all *N* and so $P(Event) = 0$. This relationship can be described as: *the probability of an impossible event is zero.*

The converse of these two relationships need not hold, i.e., a probability of one need not imply certainty since it is possible that $\lim_{N\to\infty} n_N / N = 1$ without $n_N = 1$ for all values of *N* and a probability of zero need not imply impossibility since it is possible that $\lim_{N\to\infty} n_N / N = 0$ even though $n_N > 0$. Despite these qualifications, it is useful to think of probability as measured on a scale varying from (near) impossibility at 0 to (near) certainty at 1. It should also be noted that this definition of probability (or any other definition) is not directly verifiable in the sense that we cannot actually carry out the infinite series of trials to see whether there really is a unique limiting value for the ratio n_N / N. The justification for this definition of probability is utilitarian, in that the results of applying theory based on this definition prove to be useful and that it fits with intuitive ideas. However, it should be realized that it is based on the *concept* of an infinitely long series of trials rather than an actual *series*, however long it may be.

3.1.2. Random Variables and Probability Distributions of Random Variables

A random variable *X* is a rule or a function, which associates a real number with each point in a sample space. As an example, consider the following experiment where two identical coins are tossed; *h* denotes a head and *t* denotes a tail.

Experiment: Toss two identical coins.

Sample space: $S = \{hh, ht, th, tt\}$.

Random Variable: *X*, the number of heads obtained, may be written as

$$X(hh) = 2$$
$$X(ht) = 1$$
$$X(th) = 1$$
$$X(tt) = 0$$

In this example *X* is the random variable defined by the rule "the number of heads obtained". The possible values (or real numbers) that *X* may take are 0, 1, 2. These possible values are usually denoted by *x* and the notation $X = x$ denotes *x* as a possible real value of the random variable *X*.

Random variables may be discrete or continuous. A discrete random variable assumes each of its possible values with a certain probability. For example, in the experiment above; the tossing of two coins, the sample space $S = \{hh, ht, th, tt\}$ has $N = 4$ elements and the probability the random variable *X* (the number of heads) assumes the possible values 0, 1 and 2 is given by

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Note that the values of *x* exhaust all possible cases and hence the probabilities add to 1

A continuous random variable has a probability of zero of assuming any of its values and consequently, its probability distribution cannot be given in tabular form. The concept of the probability of a continuous random variable assuming a particular value equals zero may seem strange, but the following example illustrates the point. Consider a random variable whose values are the heights of all people over 21 years of age. Between any two values, say 1.75 metres and 1.85 metres, there are an infinite number of heights, one of which is 1.80 metres. The probability of selecting a person at random exactly 1.80 metres tall and not one of the infinitely large set of heights so close to 1.80 metres that you cannot humanly measure the difference is extremely remote, and thus we assign a probability of zero to the event. It follows that probabilities of continuous random variables are defined by specifying an interval within which the random variable lies and it does not matter whether an end-point is included in the interval or not.

$$\begin{aligned} P(a < X \le b) &= P(a < X < b) + P(X = b) \\ &= P(a < X < b) \end{aligned}$$

It is most convenient to represent all the probabilities of a random variable *X* by a formula or function denoted by $f_X(x)$, $g_X(x)$, $h_X(x)$, etc, or by $F_X(x)$, $G_X(x)$, $H_X(x)$, etc.

In this notation the subscript *X* denotes that $f_X(x)$ or $F_X(x)$ is a function of the random variable *X* which takes the numerical values *x* within the function. Such functions are known as probability distribution functions and they are paired; i.e., $f_X(x)$ pairs with $F_X(x)$, $g_X(x)$ pairs with $G_X(x)$, etc. The functions with the lowercase letters are probability density functions and those with uppercase letters are cumulative distribution functions.

For discrete random variables, the probability density function has the properties

1. $f_X(x_k) = P(X = x_k)$

2. $\sum_{k=1}^{\infty} f_X(x_k) = 1$

and the cumulative distribution function has the properties

1. $F_X(x_k) = P(X \le x_k)$

2. $F_X(x) = \sum_{x_k \le x} f_X(x_k)$

As an example consider the probability distribution functions $f_X(x)$ and $F_X(x)$ of the sum of the numbers when a pair of dice is tossed.

Experiment: Toss two identical dice.

Sample space:
$$S = \begin{Bmatrix} 1,1 & 1,2 & 1,3 & 1,4 & 1,5 & 1,6 \\ 2,1 & 2,2 & 2,3 & 2,4 & 2,5 & 2,6 \\ 3,1 & 3,2 & 3,3 & 3,4 & 3,5 & 3,6 \\ 4,1 & 4,2 & 4,3 & 4,4 & 4,5 & 4,6 \\ 5,1 & 5,2 & 5,3 & 5,4 & 5,5 & 5,6 \\ 6,1 & 6,2 & 6,3 & 6,4 & 6,5 & 6,6 \end{Bmatrix}$$

Random Variable: *X*, the total of the two numbers

The probability the random variable *X* assumes the possible values $x = 2, 3, 4, \ldots, 12$ is given in Table 3.1

x	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Table 3.1 Table of probabilities

Note that the values of *x* exhaust all possible cases and hence the probabilities add to 1

The probability density function $f_X(x)$ can be deduced from Table 3.1

$$f_X(x) = \frac{6-|x-7|}{36}, \qquad x = 2,3,4,\ldots,12$$

Probability distributions are often shown in graphical form. For discrete random variables, probability distributions are generally shown in the form of histograms consisting of series of rectangles associated with values of the random variable. The width of each rectangle is one unit and the height is the probability given by the function $f_X(x)$ and the sum of the areas of all the rectangles is 1. Figure 3.1 shows the Probability histogram for the random variable *X*, the sum of the numbers when a pair of dice is tossed.

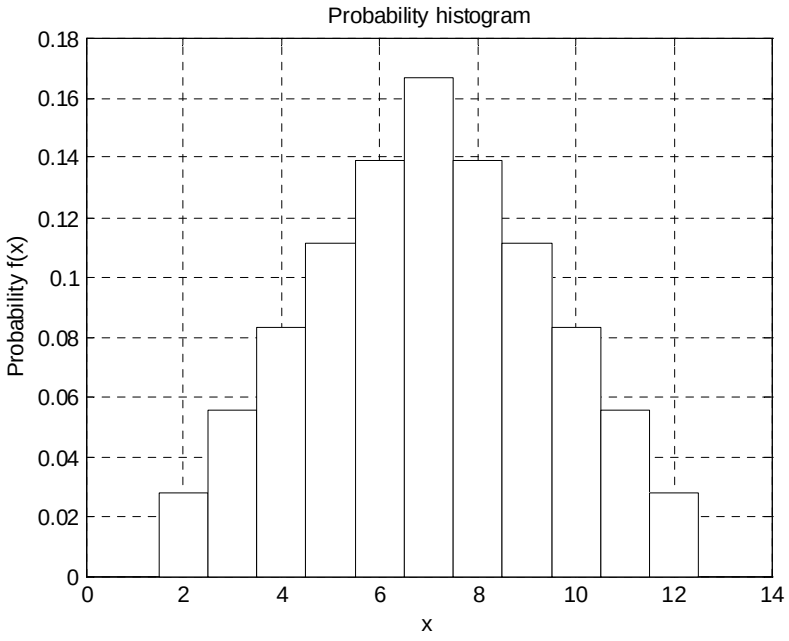


Figure 3.1 Probability histogram

The MATLAB function *dice_pdf.m* was used to create the Probability histogram of Figure 3.1

```
function dice_pdf
% Function DICE_PDF calculates the probability of a random variable
% taking the sum of the values when two dice are tossed and
% plots the probability density function as a histogram.

% create an array of the possible outcomes
x = 2:12;

% calculate the probability
y = (6-abs(x-7))/36;

% Clear Figure 1 and plot the histogram
figure(1);
clf(1);
box on;
grid on;
bar(x,y,1,'-w');

% anotate the plot
title('Probability histogram')
xlabel('x');
ylabel('Probability f(x)');
```

Figure 3.2 shows the cumulative distribution function $F_X(x) = \sum_{x_k \le x} f_X(x_k)$ for the random variable *X*, the sum of the numbers when a pair of dice is tossed.

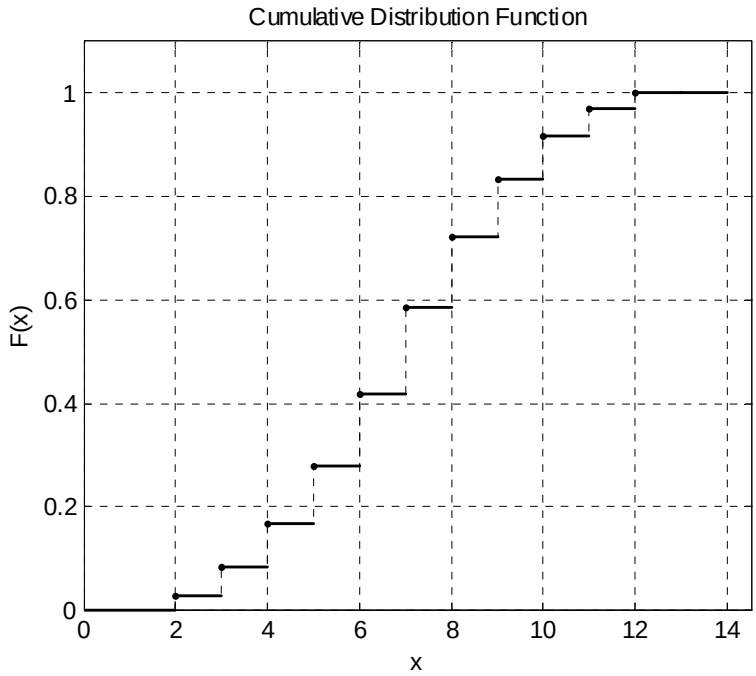


Figure 3.2 Cumulative distribution function. [The dots at the left ends of the line segments indicate the value of $F_X(x)$ at those values of *x*.

For continuous random variables, the probability distribution functions $f_X(x)$ and $F_X(x)$ are curves, which may take various forms depending on the nature of the random variable. Probability density functions $f_X(x)$ that are used in practice to model the behaviour of continuous random variables are always positive and the total area under its curve, bounded by the *x*-axis, is equal to one. These density functions have the following properties

1. $f_X(x) \geq 0$ for any value of *x*

2. $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$

The probability that a random variable *X* lies between any two values $x = a$ and $x = b$ is the area under the density curve between those two values and is found by methods of integral calculus

$$P(a < X < b) = \int_a^b f_X(x)\,dx \tag{3.2}$$

The equations of the density functions $f_X(x)$ are usually complicated and areas under their curves are found from tables. In surveying, the Normal probability density function is the usual model for the behaviour of measurements (regarded as random variables) and the probability density function is (Kreyszig, 1970, p. 107)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{3.3}$$

μ and σ are the mean and standard deviation respectively of the infinite population of *x* and Figure 3.3 shows a plot of the Normal probability density curve for $\mu = 2.0$ and $\sigma = 2.5$.

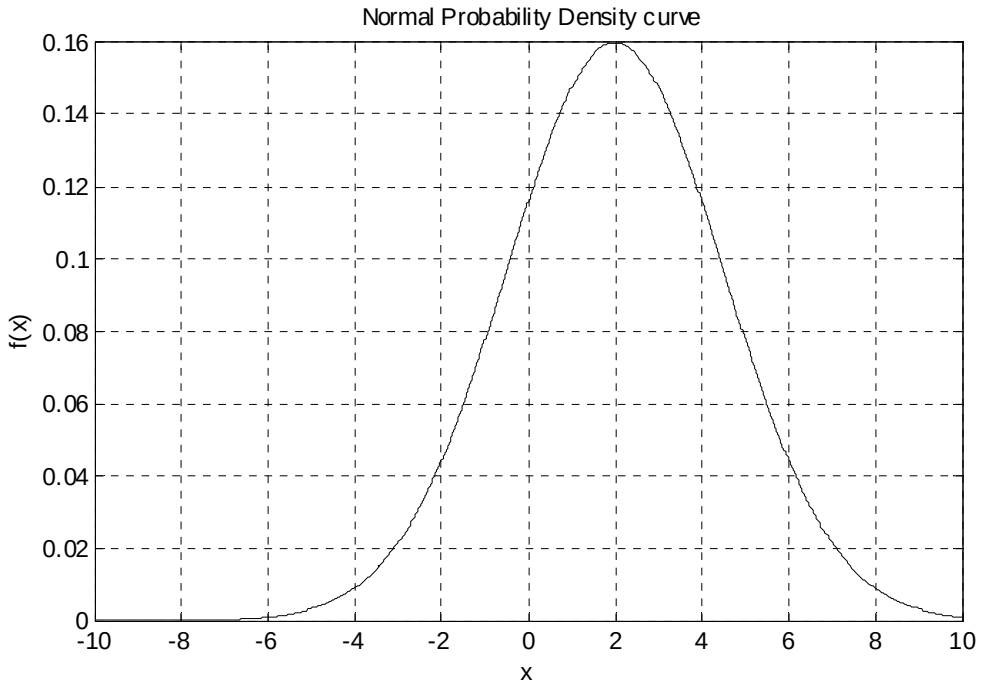


Figure 3.3 Normal probability density function for $\mu = 2.0$ and $\sigma = 2.5$

The MATLAB function *normal_pdf.m* was used to create the Normal probability density curve of Figure 3.3

```
function normal_pdf(mx,sx)
% Function NORMAL_PDF(MX,SX) calculates the probability of a random variable X
% having a NORMAL distribution with mean MX and standard deviation
% SX and plots the probability density function

% create an array of x-values between -10 and 10 at 0.01 intervals
x = -10:0.01:10;

a = (x-mx)./sx;

% calculate the density function
y = 1/(sx*sqrt(2*pi)).*exp(-0.5.*a.^2);

% Clear Figure 1 and plot the probability density curve
figure(1);
clf(1);
box on;
plot(x,y,'k');
grid on;
% anotate the plot
title('Normal Probability Density curve')
xlabel('x');
ylabel('f(x)');
```

For continuous random variables *X*, the cumulative distribution function $F_X(x)$ has the following properties

1. $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(x)\,dx$

2. $\frac{d}{dx} F_X(x) = f_X(x)$

In surveying, the Normal distribution is the usual model for the behaviour of measurements and the cumulative distribution function is (Kreyszig, 1970, p. 108)

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \tag{3.4}$$

The probability that *X* assumes any value in an interval $a < X < b$ is

$$P(a < X < b) = F_X(b) - F_X(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \tag{3.5}$$

Figure 3.4 shows a plot of the Normal cumulative distribution curve for $\mu = 2.0$ and $\sigma = 2.5$.

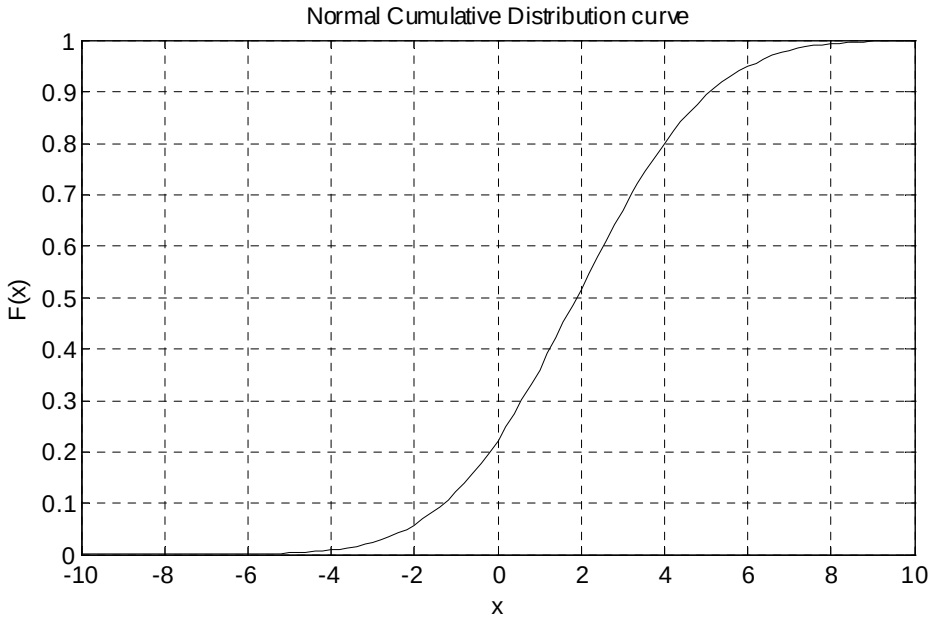


Figure 3.4 Normal cumulative distribution function for $\mu = 2.0$ and $\sigma = 2.5$

3.1.3. Multivariate Probability Density Functions

For multiple random variables, $X_1, X_2, X_3, \ldots$ the term multivariate probability density function $f_{\mathbf{X}}(\mathbf{x})$ is used to define a function whose integral gives the probability of X_1 lying in the range $a_1 < X_1 < b_1$, X_2 lying in the range $a_2 < X_2 < b_2$, X_3 lying in the range $a_3 < X_3 < b_3$, etc. This probability is

$$\begin{aligned} P &= \int_{a_1}^{b_1}\int_{a_2}^{b_2}\int_{a_3}^{b_3} \cdots f_{X_1X_2\cdots}(x_1, x_2, x_3, \cdots)\, dx_1\, dx_2\, dx_3 \cdots \\ &= \int_{\mathbf{a}}^{\mathbf{b}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \end{aligned} \tag{3.6}$$

where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots \end{bmatrix}^T$, $\mathbf{a} = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots \end{bmatrix}^T$ and $\mathbf{b} = \begin{bmatrix} b_1 & b_2 & b_3 & \cdots \end{bmatrix}^T$

Although least squares adjustment theory does not require the random variables (measurements) to have particular probability distributions, the Normal distribution is the usual model assumed to represent measurements and associated errors and corrections (residuals). The Multivariate Normal distribution of random variables has a density function of the following form (Mikhail, 1976, p. 27)

$$f_{\mathbf{X}}(x_1, x_2, \cdots, x_n) = f_{\mathbf{X}}(\mathbf{x}) = \left\{ \frac{1}{(2\pi)^{n/2}\sqrt{|\boldsymbol{\Sigma}|}} \right\} \times \exp\left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_x) \right\} \tag{3.7}$$

with mean vector $\boldsymbol{\mu}_x$ and variance-covariance matrix $\boldsymbol{\Sigma}$. For the case of two random variables *X* and *Y* the Bivariate Normal probability density function has the following form

$$\begin{aligned} f_{XY}(x, y) = \frac{1}{2\pi\sqrt{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2}} \times \exp\Bigg\{ & -\frac{\sigma_x^2\sigma_y^2}{2(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)} \\ & \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\sigma_{xy}\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x^2\sigma_y^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \Bigg\} \end{aligned} \tag{3.8}$$

where μ_x, μ_y are the means, σ_x^2, σ_y^2 are the variances of the random variables *X* and *Y* respectively and σ_{xy} is the covariance.

Figure 3.5 shows a 3-dimensional plot of a Bivariate Normal probability density function with $\mu_x = 0.8, \mu_y = -0.2$, $\sigma_x = 1.5, \sigma_y = 1.2$ and $\sigma_{xy} = -0.5$ over a range of possible values *x*, *y* of the random variables *X* and *Y*.

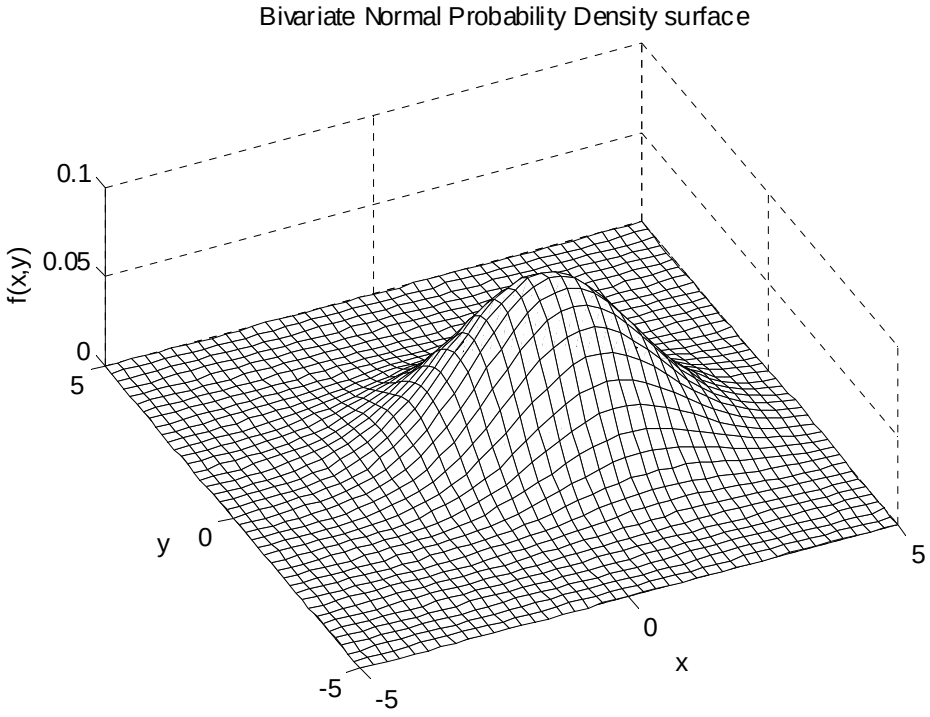


Figure 3.5 Bivariate Normal probability density surface

Planes having values $f_{XY}(x, y) = \text{constant}$ that cut the density surface will create elliptical curves of intersection.

The MATLAB function *bivariate_normal.m* was used to create the Bivariate Normal probability density surface of Figure 3.5. The equation of the Bivariate Normal probability density function $f_{XY}(x, y)$ used in the function is a modified form of (3.8) where the correlation coefficient

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{3.9}$$

is used and equation (3.8) becomes

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \times \exp\left\{ -\frac{1}{2\left(1-\rho_{xy}^2\right)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho_{xy}\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right\} \tag{3.10}$$

```
function bivariate_normal(mx,my,sx,sy,sxy)
% Function BIVARIATE_NORMAL(MX,MY,SX,SY,SXY) calculates the bivariate
% normal density function f(x,y) of two random variables X and Y having
% NORMAL distributions with means MX, MY, standard deviations SX, SY and
% covariance SXY and plots the probability density surface.

% calculate correlation coefficient r (rho)
r  = sxy/(sx*sy);
r2 = r*r;

% create arrays of x and y values between -3 and 3 at 0.2 intervals
[x,y] = meshgrid(-5:0.25:5);

a = (x-mx)./sx;
b = (y-my)./sy;
c = a.^2 - (a.*b).*(2*r) + b.^2;

% calculate the density function z = f(x,y)
z = 1/(2*pi*sx*sy*sqrt(1-r2)).*exp(-1/(2*(1-r2)).*c);

% Clear Figure 1 and plot the probability density surface
figure(1);
clf(1);
box on;
mesh(x,y,z,'EdgeColor','black');
% anotate the plot
title('Bivariate Normal Probability Density surface')
xlabel('x');
ylabel('y');
zlabel('f(x,y)');
```

3.1.4. Expectations

The expectation $E\{X\}$ of a random variable X is defined as the average value μ_X of the variable over all possible values. It is computed by taking the sum of all possible values of $X = x$ multiplied by its corresponding probability. In the case of a discrete random variable the expectation is given by

$$E\{X\} = \mu_X = \sum_{k=1}^{N} x_k P(x_k) \tag{3.11}$$

Equation (3.11) is a general expression from which we can obtain the usual expression for the arithmetic mean

$$\mu = \frac{1}{N}\sum_{k=1}^{N} x_k \tag{3.12}$$

If there are *N* possible values x_k of the random variable *X*, each having equal probability $P(x_k) = 1/N$ (which is a constant), then the expectation computed from (3.11) is identical to the arithmetic mean of the *N* values of x_k from (3.12).

In the case of a continuous random variable the expectation is given by

$$E\{X\} = \mu_X = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx \tag{3.13}$$

This relationship may be extended to a more general form if we consider the expectation of a function $g(X)$ of a random variable *X* whose probability density function is $f_X(x)$. In this case

$$E\{g(X)\} = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx \tag{3.14}$$

Extending (3.14) to the case of two random variables *X* and *Y*

$$E\{g(X,Y)\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\, f_{XY}(x,y)\, dx\, dy$$

Similarly for *n* random variables

$$E\{g(X_1, X_2, \ldots, X_n)\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(x_1, x_2, \ldots, x_n)\, f_X(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n \tag{3.15}$$

Expressing (3.15) in matrix notation gives a general form of the expected value of a multivariate function $g(\mathbf{X})$ as

$$E\{g(\mathbf{X})\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \tag{3.16}$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the multivariate probability density function.

There are some rules that are useful in calculating expectations. They are given here without proof but can be found in many statistical texts, e.g., Walpole, 1974. With *a* and *b* as constants and *X* and *Y* as random variables

$$E\{a\} = a$$

$$E\{aX\} = a\,E\{X\}$$

$$E\{aX + b\} = a\,E\{X\} + b$$

$$E\{g(X) \pm h(X)\} = E\{g(X)\} \pm E\{h(X)\}$$

$$E\{g(X,Y) \pm h(X,Y)\} = E\{g(X,Y)\} \pm E\{h(X,Y)\}$$

3.1.5. Special Mathematical Expectations

The mean of a random variable

$$\mu_X = E\{X\} = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx \tag{3.17}$$

The mean vector $\mathbf{m}_X$ of a multivariate distribution is

$$\mathbf{m}_X = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \mu_{X_3} \\ \vdots \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \\ \vdots \end{bmatrix} = E\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \end{bmatrix} = E\{\mathbf{X}\} \tag{3.18}$$

$\mathbf{m}_X$ can be taken as representing the mean of a multivariate probability density function.

The variance of a random variable

$$\sigma_X^2 = E\left\{(X - \mu_X)^2\right\} = \int_{-\infty}^{+\infty} (x - \mu_x)^2\, f_X(x)\, dx \tag{3.19}$$

The covariance between two random variables *X* and *Y* is

$$\sigma_{XY} = E\{(X - \mu_X)(Y - \mu_Y)\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - \mu_x)(y - \mu_y)\, f_{XY}(x,y)\, dx\, dy \tag{3.20}$$

Equation (3.20) can be expanded to give

$$\begin{aligned}\sigma_{XY} &= E\{(X-\mu_X)(Y-\mu_Y)\}\\ &= E\{XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y\}\\ &= E\{XY\} - E\{X\mu_Y\} - E\{Y\mu_X\} + E\{\mu_X\mu_Y\}\\ &= E\{XY\} - \mu_Y E\{X\} - \mu_X E\{Y\} + \mu_X\mu_Y\\ &= E\{XY\} - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y\\ &= E\{XY\} - \mu_X\mu_Y\end{aligned}$$

If the random variables *X* and *Y* are independent, the expectation of the product is equal to the product of the expectations, i.e., $E\{XY\} = E\{X\}E\{Y\}$. Since the expected values of *X* and *Y* are the means μ_X and μ_Y then $E\{XY\} = \mu_X\mu_Y$ if *X* and *Y* are independent. Substituting this result into the expansion above shows that the covariance σ_{XY} is zero if *X* and *Y* are independent.

For a multivariate function, variances and covariances of the random variables **X** is given by the matrix equation

$$\boldsymbol{\Sigma}_{XX} = E\left\{[\mathbf{X}-\mathbf{m}_X][\mathbf{X}-\mathbf{m}_Y]^T\right\} \tag{3.21}$$

$\boldsymbol{\Sigma}_{XX}$ is a symmetric matrix known as the variance-covariance matrix and its general form can be seen when (3.21) is expanded

$$\boldsymbol{\Sigma}_{XX} = E\left\{\begin{bmatrix} X_1-\mu_{X_1}\\ X_2-\mu_{X_2}\\ \vdots\\ X_n-\mu_{X_n}\end{bmatrix}\begin{bmatrix} X_1-\mu_{X_1} & X_2-\mu_{X_2} & \cdots & X_n-\mu_{X_n}\end{bmatrix}\right\}$$

giving

$$\boldsymbol{\Sigma}_{XX} = \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_n}\\ \sigma_{X_2X_1} & \sigma^2_{X_1} & \cdots & \sigma_{X_2X_n}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \cdots & \sigma^2_{X_1}\end{bmatrix} \tag{3.22}$$

3.2. Law of Propagation of Variances for Linear Functions

Consider two vectors of random variables $\mathbf{x} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix}^T$ that are linearly related by the matrix equation

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \tag{3.23}$$

where **A** is a coefficient matrix and **b** is a vector of constants. Then, using the rules for expectations developed above we may write an expression for the mean $\mathbf{m}_Y$ using (3.18)

$$\begin{aligned} \mathbf{m}_Y &= E\{\mathbf{y}\} \\ &= E\{\mathbf{A}\mathbf{x} + \mathbf{b}\} \\ &= E\{\mathbf{A}\mathbf{x}\} + E\{\mathbf{b}\} \\ &= \mathbf{A}\,E\{\mathbf{x}\} + \mathbf{b} \\ &= \mathbf{A}\mathbf{m}_X + \mathbf{b} \end{aligned}$$

Using (3.21), the variance-covariance matrix $\mathbf{\Sigma}_{yy}$ is given by

$$\begin{aligned} \mathbf{\Sigma}_{yy} &= E\left\{\left(\mathbf{y} - \mathbf{m}_y\right)\left(\mathbf{y} - \mathbf{m}_y\right)^T\right\} \\ &= E\left\{\left(\mathbf{A}\mathbf{x} + \mathbf{b} - \mathbf{A}\mathbf{m}_x - \mathbf{b}\right)\left(\mathbf{A}\mathbf{x} + \mathbf{b} - \mathbf{A}\mathbf{m}_x - \mathbf{b}\right)^T\right\} \\ &= E\left\{\left(\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{m}_x\right)\left(\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{m}_x\right)^T\right\} \\ &= E\left\{\mathbf{A}\left(\mathbf{x} - \mathbf{m}_x\right)\left(\mathbf{A}\left(\mathbf{x} - \mathbf{m}_x\right)\right)^T\right\} \\ &= E\left\{\mathbf{A}\left(\mathbf{x} - \mathbf{m}_x\right)\left(\mathbf{x} - \mathbf{m}_x\right)^T\mathbf{A}^T\right\} \\ &= \mathbf{A}\,E\left\{\left(\mathbf{x} - \mathbf{m}_x\right)\left(\mathbf{x} - \mathbf{m}_x\right)^T\right\}\mathbf{A}^T \\ &= \mathbf{A}\mathbf{\Sigma}_{xx}\mathbf{A}^T \end{aligned}$$

or

If $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ and **y** and **x** are random variables linearly related then

$$\mathbf{\Sigma}_{yy} = \mathbf{A}\mathbf{\Sigma}_{xx}\mathbf{A}^T \tag{3.24}$$

Equation (3.24) is known as the *Law of Propagation of Variances*.

Using the relationships previously established between variance-covariance matrices and cofactor matrices, i.e., $\boldsymbol{\Sigma} = \sigma_0^2 \mathbf{Q}$, the Law of Propagation of Variances may also applies to cofactor matrices

If $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ and $\mathbf{y}$ and $\mathbf{x}$ are random variables linearly related then

$$\mathbf{Q}_{yy} = \mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T \tag{3.25}$$

3.3. Law of Propagation of Variances for Non-Linear Functions

In many practical applications of variance propagation the random variables in $\mathbf{x}$ and $\mathbf{y}$ are nonlinearly related, i.e.,

$$\mathbf{y} = f(\mathbf{x}) \tag{3.26}$$

In such cases, we can expand the function on the right-hand-side of (3.26) using Taylor's theorem.

For a non-linear function of a single variable Taylor's theorem may be expressed in the following form

$$\begin{aligned} f(x) = f(a) &+ \left.\frac{df}{dx}\right|_a (x-a) + \left.\frac{d^2 f}{dx^2}\right|_a \frac{(x-a)^2}{2!} + \left.\frac{d^3 f}{dx^3}\right|_a \frac{(x-a)^3}{3!} + \cdots \\ &+ \left.\frac{d^{n-1} f}{dx^{n-1}}\right|_a \frac{(x-a)^{n-1}}{(n-1)!} + R_n \end{aligned} \tag{3.27}$$

where R_n is the remainder after n terms and $\lim_{n\to\infty} R_n = 0$ for $f(x)$ about $x = a$ and $\left.\frac{df}{dx}\right|_a$, $\left.\frac{d^2 f}{dx^2}\right|_a$ etc are derivatives of the function $f(x)$ evaluated at $x = a$.

For a non-linear function of two random variables, say $\phi = f(x, y)$, the Taylor series expansion of the function ϕ about $x = a$ and $y = b$ is

$$\begin{aligned} \phi = f(a,b) &+ \left.\frac{\partial f}{\partial x}\right|_{a,b} (x-a) + \left.\frac{\partial f}{\partial y}\right|_{a,b} (y-b) \\ &+ \frac{1}{2!}\left\{ \left.\frac{\partial^2 f}{\partial x^2}\right|_{a,b} (x-a)^2 + \left.\frac{\partial^2 f}{\partial y^2}\right|_{a,b} (y-b)^2 + \left.\frac{\partial f}{\partial x}\right|_{a,b} \left.\frac{\partial f}{\partial y}\right|_{a,b} (x-a)(y-b) \right\} + \cdots \end{aligned} \tag{3.28}$$

where $f(a,b)$ is the function ϕ evaluated at $x=a$ and $y=b$, and $\left.\frac{\partial f}{\partial x}\right|_{a,b}$, $\left.\frac{\partial f}{\partial y}\right|_{a,b}$, $\left.\frac{\partial^2 f}{\partial x^2}\right|_{a,b}$ etc are partial derivatives of the function ϕ evaluated at $x=a$ and $y=b$.

Extending to *n* random variables, we may write a Taylor series approximation of the function $f(\mathbf{x})$ as a matrix equation

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0} (\mathbf{x}-\mathbf{x}^0) + \text{ higher order terms} \tag{3.29}$$

where $f(\mathbf{x}^0)$ is the function evaluated at the approximate values $\mathbf{x}^0$ and $\left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0}$ are the partial derivatives evaluated at approximations $\mathbf{x}^0$.

Replacing $f(\mathbf{x})$ in (3.26) by its Taylor series approximation, ignoring higher order terms, gives

$$\mathbf{y} = f(\mathbf{x}) = f(\mathbf{x}^0) + \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0} (\mathbf{x}-\mathbf{x}^0) \tag{3.30}$$

Then, using the rules for expectations

$$\begin{aligned}
\mathbf{m}_y &= E\{\mathbf{y}\} \\
&= E\left\{ f(\mathbf{x}^0) + \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0} (\mathbf{x}-\mathbf{x}^0) \right\} \\
&= E\left\{ f(\mathbf{x}^0) \right\} + E\left\{ \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0} (\mathbf{x}-\mathbf{x}^0) \right\} \\
&= f(\mathbf{x}^0) + \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0} E\left\{ (\mathbf{x}-\mathbf{x}^0) \right\} \\
&= f(\mathbf{x}^0) + \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0} \left(E\{\mathbf{x}\} - E\{\mathbf{x}^0\} \right) \\
&= f(\mathbf{x}^0) + \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0} (\mathbf{m}_x - \mathbf{x}^0)
\end{aligned}$$

And

$$\begin{aligned}\mathbf{y}-\mathbf{m}_y &= \left[f\left(\mathbf{x}^0\right)+\left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0}\left(\mathbf{x}-\mathbf{x}^0\right)\right]-\left[f\left(\mathbf{x}^0\right)+\left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0}\left(\mathbf{m}_x-\mathbf{x}^0\right)\right] \\ &= \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}^0}\left(\mathbf{x}-\mathbf{m}_x\right) \\ &= \mathbf{J}_{yx}\left(\mathbf{x}-\mathbf{m}_x\right)\end{aligned} \tag{3.31}$$

$\mathbf{J}_{yx}$ is the (m,n) Jacobian matrix of partial derivatives, noting that $\mathbf{y}$ and $\mathbf{x}$ are $(m,1)$ and $(n,1)$ vectors respectively

$$\mathbf{J}_{yx}=\begin{bmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 & \cdots & \partial y_1/\partial x_n \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 & \cdots & \partial y_2/\partial x_n \\ \vdots & & & \vdots \\ \partial y_m/\partial x_1 & \partial y_m/\partial x_2 & \cdots & \partial y_m/\partial x_n \end{bmatrix} \tag{3.32}$$

Using (3.21), the variance-covariance matrix $\mathbf{\Sigma}_{yy}$ is given by

$$\begin{aligned}\mathbf{\Sigma}_{yy} &= E\left\{\left(\mathbf{y}-\mathbf{m}_y\right)\left(\mathbf{y}-\mathbf{m}_y\right)^T\right\} \\ &= E\left\{\left(\mathbf{J}_{yx}\left(\mathbf{x}-\mathbf{m}_x\right)\right)\left(\mathbf{J}_{yx}\left(\mathbf{x}-\mathbf{m}_x\right)\right)^T\right\} \\ &= E\left\{\mathbf{J}_{yx}\left(\mathbf{x}-\mathbf{m}_x\right)\left(\mathbf{x}-\mathbf{m}_x\right)^T\mathbf{J}_{yx}^T\right\} \\ &= \mathbf{J}_{yx}\,E\left\{\left(\mathbf{x}-\mathbf{m}_x\right)\left(\mathbf{x}-\mathbf{m}_x\right)^T\right\}\mathbf{J}_{yx}^T \\ &= \mathbf{J}_{yx}\mathbf{\Sigma}_{xx}\mathbf{J}_{yx}^T\end{aligned}$$

Thus, in a similar manner to above, we may express the Law of Propagation of Variances for non-linear functions of random variables as

If $\mathbf{y}=f\left(\mathbf{x}\right)$ and $\mathbf{y}$ and $\mathbf{x}$ are random variables non-linearly related then

$$\mathbf{\Sigma}_{yy}=\mathbf{J}_{yx}\mathbf{\Sigma}_{xx}\mathbf{J}_{yx}^T \tag{3.33}$$

This rule also applies to cofactor matrices

If $\mathbf{y}=f\left(\mathbf{x}\right)$ and $\mathbf{y}$ and $\mathbf{x}$ are random variables non-linearly related then

$$\mathbf{Q}_{yy}=\mathbf{J}_{yx}\mathbf{Q}_{xx}\mathbf{J}_{yx}^T \tag{3.34}$$

3.4. The Special Law of Propagation of Variances

The Law of Propagation of Variances is often expressed as an algebraic equation. For example, if *z* is a function of two random variables *x* and *y*, i.e., $z = f(x, y)$ then the variance of *z* is

$$\sigma_z^2 = \left(\frac{\partial z}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 \sigma_y^2 + 2\frac{\partial z}{\partial x}\frac{\partial z}{\partial y}\sigma_{xy} \tag{3.35}$$

Equation (3.35) can be derived from the general matrix equation (3.33) in the following manner. Let $z = f(x, y)$ be written as $\mathbf{y} = f(\mathbf{x})$ where $\mathbf{y} = [z]$, a (1,1) matrix and $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ is a (2,1) vector. The variance-covariance matrix of the random vector **x** is $\mathbf{\Sigma}_{xx} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$, the Jacobian $\mathbf{J}_{yx} = \begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{bmatrix}$ and the variance-covariance matrix $\mathbf{\Sigma}_{yy}$ which contains the single element σ_z^2 is given by

$$\mathbf{\Sigma}_{yy} = \left[\sigma_z^2\right] = \begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{bmatrix} \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \frac{\partial z}{\partial x} \\ \frac{\partial z}{\partial y} \end{bmatrix}$$

Expanding this equation gives (3.35).

In the case where the random variables in **x** are independent, i.e., their covariances are zero; we have the Special Law of Propagation of Variances. For the case of $z = f(x, y)$ where the random variables *x* and *y* are independent, the *Special Law of Propagation of Variances* is written as

If $z = f(x, y)$ and *x* and *y* are independent random variables then

$$\sigma_z^2 = \left(\frac{\partial z}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 \sigma_y^2 \tag{3.36}$$

3.5. Applications of Propagation of Variances

Propagation of variances, using either the *Law of Propagation of Variances* for linear functions, equations (3.24) or (3.25) and non-linear functions, equations (3.33) or (3.34), or the *Special Law of Propagation of Variances*, equation (3.36) where the variables are independent are important "tools" for assessing the precision of computed quantities arising from measurements. Implicit in every application of variance propagation is an a priori knowledge of the precision of the measurements. For example, if quantities are derived from Total Station EDM distances, then some knowledge of the precision of those distances is assumed; if height differences are computed from a combination of Total Station EDM distances and zenith angles, then the precisions of distances and zenith angles is assumed. These a priori precision estimates may come from information supplied by the equipment manufacturer, statistical analysis of observations, prior knowledge or simply educated guesses. Whatever the source of knowledge, it is assumed that these precisions are known before any variance propagation is made. The following sections set out some useful techniques that are applicable to many surveying operations.

3.5.1. Variances of Total Station Horizontal Distances and Height Differences

Total Stations are modern surveying instruments combining an electronic theodolite (for angle measurement) and an EDM (for distance measurement to a reflecting prism). EDM is an abbreviation of Electronic Distance Measurement. The primary Total Station measurements are horizontal and vertical circle readings α and β respectively (from which angles may be obtained) and slope distances *D*. Total Stations have "on board" computers and may display (at the push of a button) computed quantities such as vertical height differences $\pm V$ and horizontal distances *H*, between the instrument and the prism (the sighting target).

A vertical circle reading β made with a Total Station is a clockwise angle from the zenith (defined by the vertical axis of the Total Station) measured in a vertical plane. This vertical plane is swept out by the telescope rotating about the Total Station's horizontal axis (the trunnion axis). The horizontal distance *H* and the vertical component $\pm V$ of a measured slope distance *S* are

$$H = D\sin\beta$$

$$V = D\cos\beta$$

H and V are functions of D and β which we may write as $H = H(D,\beta)$ and $V = V(D,\beta)$. Treating H, V, D and β as random variables and assuming that D and β are independent the Special Law of Propagation of Variances (3.36) can be used to compute the variances of H and V

$$\sigma_H^2 = \left(\frac{\partial H}{\partial D}\right)^2 \sigma_D^2 + \left(\frac{\partial H}{\partial \beta}\right)^2 \sigma_\beta^2 = \sin^2\beta\left(\sigma_D^2\right) + D^2\cos^2\beta\left(\sigma_\beta^2\right)$$

$$\sigma_V^2 = \left(\frac{\partial V}{\partial D}\right)^2 \sigma_D^2 + \left(\frac{\partial V}{\partial \beta}\right)^2 \sigma_\beta^2 = \cos^2\beta\left(\sigma_D^2\right) + D^2\sin^2\beta\left(\sigma_\beta^2\right)$$

σ_D^2, σ_β^2 are the variances of the slope distance and zenith angle respectively. For any properly calibrated Total Station EDM and prism combination, the standard deviation of a distance can be expressed in the form $\sigma_D = x + y$ ppm (ppm is parts per million) and the standard deviation of a zenith angle as $\sigma_\beta = x$ seconds of arc .

As an example: say D = 60.000 m, $\beta = 85^\circ\,36'00''$ and $\sigma_D = 8$ mm + 5 ppm = 0.0083 m, $\sigma_\beta = 10'' = 4.8481\times10^{-5}$ radians ; the variances in H and V are

$$\begin{aligned}\sigma_H^2 &= (0.9971)^2\,(0.0083)^2 + (60)^2\,(0.0767)^2\left(4.8418\times10^{-5}\right)^2\\ &= 6.85\times10^{-5}\ \text{m}^2\end{aligned}$$

$$\begin{aligned}\sigma_V^2 &= (0.0767)^2\,(0.0083)^2 + (60)^2\,(0.9971)^2\left(4.8418\times10^{-5}\right)^2\\ &= 8.82\times10^{-6}\ \text{m}^2\end{aligned}$$

The standard deviations are of H and V are

$$\sigma_H = \sqrt{6.85\times10^{-5}} = 0.0083\ \text{m}$$

$$\sigma_V = \sqrt{8.82\times10^{-6}} = 0.0030\ \text{m}$$

It is interesting to note that the computed quantities H and V are not independent, even though they have been computed from independent quantities. This can be seen by computing the

variances in the following manner. We may write the computation of the components *H* and *V* from the observations *D* and β as the vector equation

$$\mathbf{y} = f(\mathbf{x})$$

$\mathbf{y} = \begin{bmatrix} H \\ V \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} D \\ \beta \end{bmatrix}$ and **y** and **x** are non-linearly related. Using (3.33) we may write

$$\begin{aligned}
\boldsymbol{\Sigma}_{yy} = \begin{bmatrix} \sigma_H^2 & \sigma_{HV} \\ \sigma_{HV} & \sigma_V^2 \end{bmatrix} &= \mathbf{J}_{yx}\boldsymbol{\Sigma}_{xx}\mathbf{J}_{yx}^T \\
&= \begin{bmatrix} \partial H/\partial D & \partial H/\partial \beta \\ \partial V/\partial D & \partial V/\partial \beta \end{bmatrix} \begin{bmatrix} \sigma_D^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix} \begin{bmatrix} \partial H/\partial D & \partial V/\partial D \\ \partial H/\partial \beta & \partial V/\partial \beta \end{bmatrix} \\
&= \begin{bmatrix} \sin\beta & D\cos\beta \\ \cos\beta & -D\sin\beta \end{bmatrix} \begin{bmatrix} \sigma_D^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix} \begin{bmatrix} \sin\beta & \cos\beta \\ D\cos\beta & -D\sin\beta \end{bmatrix} \\
&= \begin{bmatrix} 0.9971 & 4.6031 \\ 0.0767 & -59.8232 \end{bmatrix} \begin{bmatrix} (0.0083)^2 & 0 \\ 0 & (4.8481\times10^{-5})^2 \end{bmatrix} \begin{bmatrix} 0.9971 & 0.0767 \\ 4.6031 & -59.8232 \end{bmatrix} \\
&= \begin{bmatrix} 6.85\times10^{-5} & 4.62\times10^{-6} \\ 4.62\times10^{-6} & 8.82\times10^{-6} \end{bmatrix}
\end{aligned}$$

The diagonal elements of $\boldsymbol{\Sigma}_{yy}$ are $\sigma_H^2 = 6.85\times10^{-5}$ m^2 and $\sigma_V^2 = 8.82\times10^{-6}$ m^2 which are the same values as computed above and the off-diagonal elements are the covariance $\sigma_{HV} = 4.62\times10^{-6}$ m^2. These elements, which are non-zero, indicate that computed quantities are correlated.

3.5.2. Variances of Height Differences from Spirit Levelling

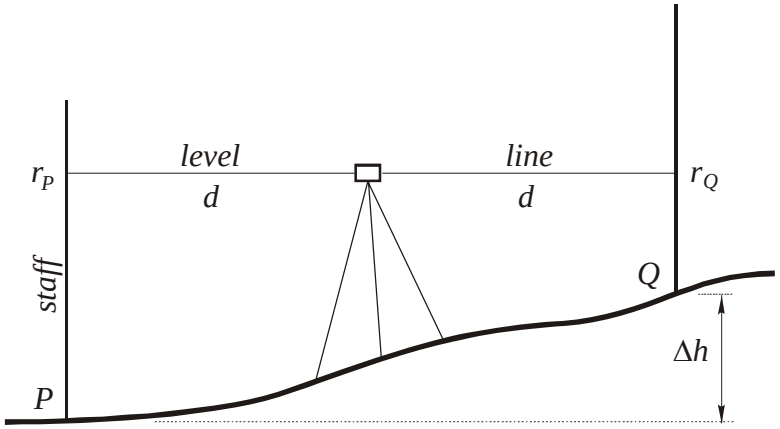


Figure 3.1

Figure 3.1 shows a schematic diagram of a spirit level and a single height difference Δh_{PQ} between positions of the levelling staves at *P* and *Q*. The backsight and foresight staff readings are r_P and r_Q and the length of the sights to the levelling staff are the same and equal to *d*. The height difference is

$$\Delta h_{PQ} = r_P - r_Q$$

Considering the backsight and foresight staff readings to be independent and of equal precision, the Special Law of Propagation of Variances (3.36) gives the variance of a single height difference from spirit levelling as

$$\sigma^2_{\Delta h} = \left(\frac{\partial \Delta h}{\partial r_P}\right)^2 \sigma^2_{r_P} + \left(\frac{\partial \Delta h}{\partial r_Q}\right)^2 \sigma^2_{r_Q} = \sigma^2_{r_P} + \sigma^2_{r_Q} = 2\sigma^2_r \qquad (3.37)$$

σ_r is the standard deviation of a staff reading.

Suppose a flight of levels is run between two points *A* and *B* that are a distance *D* apart and that at every set up of the level, the backsight and foresight distances are the same and equal to *d* and the staff readings are all of equal precision σ_r. There will be $n = D/2d$ set ups and

the height difference ΔH_{AB} will be the sum of the individual height differences of the level run

$$\Delta H_{AB} = \Delta h_1 + \Delta h_2 + \cdots + \Delta h_n$$

Considering each Δh to be independent and of equal precision, the variance of the total height difference ΔH_{AB} is given by the Special Law of Propagation of Variances

$$\sigma^2_{\Delta H_{AB}} = \sigma^2_{\Delta h_1} + \sigma^2_{\Delta h_2} + \cdots + \sigma^2_{\Delta h_n} = n\sigma^2_{\Delta h} \tag{3.38}$$

It is usual practice in spirit levelling to "close the level run" by returning to the start, therefore the mean height difference of a closed level run (or a levelling loop) is

$$\Delta H_{MEAN} = \frac{\Delta H_{AB} + \Delta H_{BA}}{2}$$

Again, considering ΔH_{AB} and ΔH_{BA} to be independent and of equal precision, the Special Law of Propagation of Variances gives, bearing in mind (3.37) and (3.38) the variance of the mean height difference of closed level run as

$$\sigma^2_{\Delta H_{MEAN}} = \frac{1}{4}\sigma^2_{\Delta H_{AB}} + \frac{1}{4}\sigma^2_{\Delta H_{BA}} = \frac{1}{2}\sigma^2_{\Delta H} = \frac{n}{2}\sigma^2_{\Delta h} = n\sigma^2_r \tag{3.39}$$

$n = D/2d$ is the number of set ups in the level run between *A* and *B*, that are distance *D* apart (*d* is the length of the backsight/foresight distance) hence the variance in the mean height difference is proportional *D*. Since weights are inversely proportional to variances, it is common to express precisions of spirit levelled height differences as weights that are defined as being inversely proportional to distances.

$$w_{\Delta H} \propto \frac{1}{\text{distance}} \tag{3.40}$$

3.5.3. Variance of Trigonometric Heights

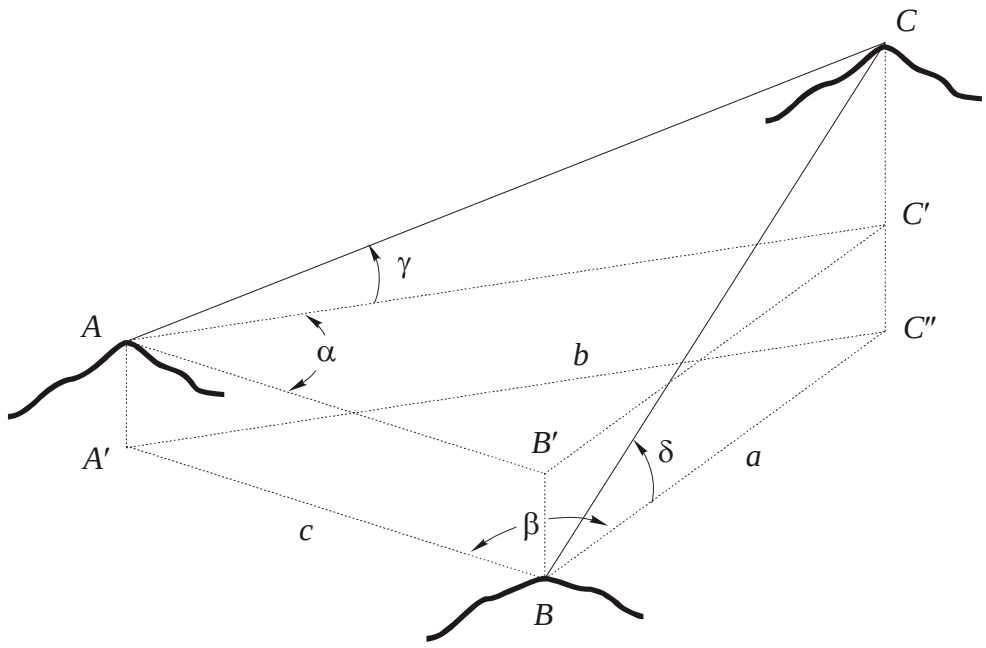


Figure 3.2

Figure 3.2 is a schematic diagram of vertical angle observations γ and δ at *A* and *B* to a distant point *C*. Horizontal angles α and β are measured at *A* and *B* and the horizontal distance *c* between *A* and *B* is measured. $AC'B'$ and $A'C''B$ are horizontal planes. If the heights of *A* and *B* are known then the height of *C* can be computed by plane trigonometry. Heights computed in this manner are known as trigonometric heights.

If the observations, angles $\alpha, \beta, \gamma, \delta$ and distance *c*, have particular precisions then the computed height of *C* will also have a precision that can be determined by propagation of variances. The method of propagation will be demonstrated by the following example.

Example. Referring to Figure 3.2, the following observations and standard deviations are known together with the Reduced Levels of *A* and *B*

$$\begin{array}{lll} \alpha = 75^\circ\,33'25'' & \sigma_\alpha = 10'' & RL_A = 105.450 \text{ m} \\ \beta = 87^\circ\,40'00'' & \sigma_\beta = 20'' & RL_B = 92.330 \text{ m} \\ \gamma = 10^\circ\,42'10'' & \sigma_\gamma = 15'' & \\ \delta = 11^\circ\,28'05'' & \sigma_\delta = 30'' & \\ c = 505.450\,\text{m} & \sigma_c = 0.050 \text{ m} & \end{array}$$

Method of Computation of RL of *C*

(i) Compute sides *a* and *b* using the Sine Rule:

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\left(180 - (\alpha+\beta)\right)} = \frac{c}{\sin(\alpha+\beta)}$$

giving

$$a = \frac{c\sin\alpha}{\sin(\alpha+\beta)} = 1695.816217 \text{ m} \tag{3.41}$$

$$b = \frac{c\sin\beta}{\sin(\alpha+\beta)} = 1749.707936 \text{ m} \tag{3.42}$$

(ii) Compute the mean RL of *C*

$$\begin{array}{ll} RL_C \text{ from } A: & RL_C = RL_A + b\tan\gamma = 436.1486 \text{ m} \\ RL_C \text{ from } B: & RL_C = RL_B + a\tan\delta = 436.3632 \text{ m} \end{array}$$

$$RL_C = \frac{RL_A + RL_B + b\tan\gamma + a\tan\delta}{2} = 436.256 \text{ m} \tag{3.43}$$

Method of Computation of Variance of RL of *C*

(i) Inspection of equations (3.41) and (3.42) shows that *a* and *b* are non-linear functions of the variables α, β and *c* which we may write as

$$\mathbf{y} = f(\mathbf{x})$$

where $\mathbf{y} = \begin{bmatrix} a & b \end{bmatrix}^T$, $\mathbf{x} = \begin{bmatrix} \alpha & \beta & c \end{bmatrix}^T$ and **y** and **x** are non-linearly related. Using (3.33) we may write

$$\boldsymbol{\Sigma}_{yy} = \mathbf{J}_{yx} \boldsymbol{\Sigma}_{xx} \mathbf{J}_{yx}^T \tag{3.44}$$

where $\boldsymbol{\Sigma}_{yy} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ba} & \sigma_b^2 \end{bmatrix}$, $\mathbf{J}_{yx} = \begin{bmatrix} \partial a/\partial\alpha & \partial a/\partial\beta & \partial a/\partial c \\ \partial b/\partial\alpha & \partial b/\partial\beta & \partial b/\partial c \end{bmatrix}$ and $\boldsymbol{\Sigma}_{xx} = \begin{bmatrix} \sigma_\alpha^2 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & \sigma_c^2 \end{bmatrix}$

noting that the observations α, β and c are considered as independent random variables, hence $\boldsymbol{\Sigma}_{xx}$ is diagonal. The elements of the Jacobian $\mathbf{J}_{yx}$ are the partial derivatives of equations (3.41) and (3.42); they are obtained as follows.

(ii) partial derivative $\partial a/\partial\alpha$:

Using the rule for derivatives $\dfrac{d}{dx}\left(\dfrac{u}{v}\right) = \dfrac{v\left(du/dx\right) - u\left(dv/dx\right)}{v^2}$ and the relation

$\dfrac{d}{d\alpha}\sin\left(\alpha+\beta\right) = \cos\alpha$ we may write

$$\frac{\partial a}{\partial\alpha} = \frac{\sin\left(\alpha+\beta\right)c\cos\alpha - c\sin\alpha\cos\left(\alpha+\beta\right)}{\sin^2\left(\alpha+\beta\right)}$$
$$= \frac{c\left[\sin\left(\alpha+\beta\right)\cos\alpha - \cos\left(\alpha+\beta\right)\sin\alpha\right]}{\sin^2\left(\alpha+\beta\right)}$$

Using the trigonometric function $\sin\left(A-B\right) = \sin A\cos B - \cos A\sin B$ where $A = \alpha+\beta$ and $B = \alpha$ then $A - B = \beta$ and we may write

$$\frac{\partial a}{\partial\alpha} = \frac{c\sin\beta}{\sin^2\left(\alpha+\beta\right)}$$

Noting that $b = \dfrac{c\sin\beta}{\sin\left(\alpha+\beta\right)}$ the partial derivative is

$$\frac{\partial a}{\partial\alpha} = \frac{b}{\sin\left(\alpha+\beta\right)} \tag{3.45}$$

(iii) partial derivative $\partial a/\partial \beta$:

Similarly to the above derivation, we may write

$$\frac{\partial a}{\partial \beta} = \frac{\sin(\alpha+\beta)(0) - c\sin\alpha\cos(\alpha+\beta)}{\sin^2(\alpha+\beta)}$$
$$= \frac{-c\sin\alpha\cos(\alpha+\beta)}{\sin^2(\alpha+\beta)}$$

Noting that $a = \frac{c\sin a}{\sin(\alpha+\beta)}$ and that $\frac{\cos(\alpha+\beta)}{\sin(\alpha+\beta)} = \frac{1}{\tan(\alpha+\beta)}$ the partial derivative is

$$\frac{\partial a}{\partial \beta} = \frac{-a}{\tan(\alpha+\beta)} \tag{3.46}$$

(iii) partial derivative $\partial a/\partial c$:

$$\frac{\partial a}{\partial c} = \frac{\sin\alpha}{\sin(\alpha+\beta)} = \frac{a}{c} \tag{3.47}$$

(iv) partial derivatives $\partial b/\partial \alpha$, $\partial b/\partial \beta$, $\partial b/\partial c$

Similarly to the above derivations

$$\frac{\partial b}{\partial \alpha} = \frac{-b}{\tan(\alpha+\beta)} \tag{3.48}$$

$$\frac{\partial b}{\partial \beta} = \frac{a}{\sin(\alpha+\beta)} \tag{3.49}$$

$$\frac{\partial b}{\partial c} = \frac{\sin\beta}{\sin(\alpha+\beta)} = \frac{b}{c} \tag{3.50}$$

(v) Substitute numeric values into the equations for the partial derivatives and set the elements of the Jacobian $\mathbf{J}_{yx}$

$$\mathbf{J}_{yx} = \left[\begin{array}{cc:c} \frac{b}{\sin(\alpha+\beta)} & \frac{-a}{\tan(\alpha+\beta)} & \frac{a}{c} \\ \hdashline \frac{-b}{\tan(\alpha+\beta)} & \frac{a}{\sin(\alpha+\beta)} & \frac{b}{c} \end{array}\right] = \begin{bmatrix} 6061.961236 & 5625.191025 & 3.355062 \\ 5803.955218 & 5875.250353 & 3.461684 \end{bmatrix}$$

(vi) Substitute numeric values into the elements of the variance matrix $\boldsymbol{\Sigma}_{xx}$

$$\boldsymbol{\Sigma}_{xx} = \begin{bmatrix} \sigma_\alpha^2 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & \sigma_c^2 \end{bmatrix} = \begin{bmatrix} \left(4.8481\times10^{-5}\right)^2 & 0 & 0 \\ 0 & \left(9.6963\times10^{-5}\right)^2 & 0 \\ 0 & 0 & \left(0.050\right)^2 \end{bmatrix}$$

(vii) Perform the matrix multiplications in (3.44) to give the variance-covariance matrix of the computed distances *a* and *b*

$$\boldsymbol{\Sigma}_{yy} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ba} & \sigma_b^2 \end{bmatrix} = \begin{bmatrix} 0.412012 & 0.422455 \\ 0.422455 & 0.433671 \end{bmatrix} \tag{3.51}$$

Note that the off-diagonal terms are not zero, indicating that the computed quantities are correlated. The leading-diagonal elements are the variances of *a* and *b*, hence the standard deviations are

$$\sigma_a = \sqrt{0.412012} = 0.642 \text{ m}$$
$$\sigma_b = \sqrt{0.433671} = 0.659 \text{ m}$$

(viii) Inspection of equation (3.43) shows that the mean RL of *C* is non-linear function of the correlated variables *a, b* and the independent variables γ, δ that we may write as

$$\mathbf{y} = f\left(\mathbf{x}\right)$$

$\mathbf{y} = \left[RL_C\right]$, $\mathbf{x} = \begin{bmatrix} a & b & \gamma & \delta \end{bmatrix}^T$ and **y** and **x** are non-linearly related. Using (3.33) we may write

$$\boldsymbol{\Sigma}_{yy} = \mathbf{J}_{yx}\boldsymbol{\Sigma}_{xx}\mathbf{J}_{yx}^T \tag{3.52}$$

where $\boldsymbol{\Sigma}_{yy} = \left[\sigma_{RL_C}^2\right]$, $\mathbf{J}_{yx} = \begin{bmatrix} \partial RL_C/\partial a & \partial RL_C/\partial b & \partial RL_C/\partial\gamma & \partial RL_C/\partial\delta \end{bmatrix}$ and

$$\boldsymbol{\Sigma}_{xx} = \left[\begin{array}{cc:cc} \sigma_a^2 & \sigma_{ab} & 0 & 0 \\ \sigma_{ba} & \sigma_b^2 & 0 & 0 \\ \hdashline 0 & 0 & \sigma_\gamma^2 & 0 \\ 0 & 0 & 0 & \sigma_\delta^2 \end{array}\right]$$

Note that $\boldsymbol{\Sigma}_{yy}$ is a matrix containing a single element only, the variance of the mean RL of *C* and $\boldsymbol{\Sigma}_{xx}$ is a partitioned matrix where the upper-left part is the variance-covariance matrix in equation (3.51) and the lower-right part contains the variances of the vertical angles γ and δ. The Jacobian $\mathbf{J}_{yx}$ is a row-vector containing the partial derivatives of equation (3.43) that are given as follows.

(ix) partial derivatives $\partial RL_C/\partial a$, $\partial RL_C/\partial b$, $\partial RL_C/\partial \gamma$, $\partial RL_C/\partial \delta$

Noting that $\dfrac{d}{dx}\tan x = \sec^2 x = \dfrac{1}{\cos^2 x}$ the partial derivatives are

$$\frac{\partial RL_C}{\partial a} = \frac{\tan\delta}{2}, \quad \frac{\partial RL_C}{\partial b} = \frac{\tan\gamma}{2}, \quad \frac{\partial RL_C}{\partial \gamma} = \frac{b}{2\cos^2\gamma}, \quad \frac{\partial RL_C}{\partial \delta} = \frac{a}{2\cos^2\delta}$$

(x) Substitute numeric values into the equations for the partial derivatives and set the elements of the Jacobian $\mathbf{J}_{yx}$

$$\begin{aligned}\mathbf{J}_{yx} &= \begin{bmatrix} \dfrac{\tan\delta}{2} & \dfrac{\tan\gamma}{2} & \dfrac{b}{2\cos^2\gamma} & \dfrac{a}{2\cos^2\delta} \end{bmatrix} \\ &= \begin{bmatrix} 0.101436 & 0.094501 & 906.105346 & 882.805418 \end{bmatrix}\end{aligned}$$

(xi) Substitute numeric values into the elements of the variance matrix $\boldsymbol{\Sigma}_{xx}$

$$\boldsymbol{\Sigma}_{xx} = \left[\begin{array}{cc:cc} \sigma_a^2 & \sigma_{ab} & 0 & 0 \\ \sigma_{ba} & \sigma_b^2 & 0 & 0 \\ \hdashline 0 & 0 & \sigma_\gamma^2 & 0 \\ 0 & 0 & 0 & \sigma_\delta^2 \end{array}\right] = \left[\begin{array}{cc:cc} 0.412012 & 0.422455 & 0 & 0 \\ 0.422455 & 0.433671 & 0 & 0 \\ \hdashline 0 & 0 & \left(7.2722\times10^{-5}\right)^2 & 0 \\ 0 & 0 & 0 & \left(1.4544\times10^{-4}\right)^2 \end{array}\right]$$

(xii) Perform the matrix multiplications in (3.52) to give the variance-covariance matrix of the computed mean RL of *C.*

$$\boldsymbol{\Sigma}_{yy} = \begin{bmatrix} 0.037040 \end{bmatrix}$$

Hence, the variance of the mean RL of *C* is

$$\sigma^2_{RL_C} = 0.037040 \text{ m}^2$$

and the standard deviation of the mean RL of *C* is

$$\sigma_{RL_C} = \sqrt{0.037040} = 0.192 \text{ m}$$

3.6. Propagation of Variances of Multiple Functions of Random Variables

In some applications of surveying and least squares, we must deal with multiple functions of random variables that may be correlated. To handle these cases, Mikhail (1976, pp.83-87) develops general rules and techniques that are repeated in the following sections. Note that cofactor matrices **Q** replace variance-covariance matrices $\boldsymbol{\Sigma}$ in the following developments.

Consider $\mathbf{x}_{(n,1)}$ and $\mathbf{t}_{(m,1)}$ to be two correlated vectors with cofactor matrices $\mathbf{Q}_{xx}, \mathbf{Q}_{xt}$ and $\mathbf{Q}_{tt}$. $\mathbf{Q}_{xt}$ is an (*n*,*m*) crosscofactor matrix containing estimates of covariances between the elements of the **x** and **t** vectors. Two other vectors $\mathbf{y}_{(q,1)}$ and $\mathbf{z}_{(p,1)}$ are functions of **x** and **t** (**y** is functionally independent of **t** and **z** of **x**).

$$\begin{aligned} \mathbf{y} &= y(\mathbf{x}) \\ \mathbf{z} &= z(\mathbf{t}) \end{aligned} \tag{3.53}$$

The Jacobians of these functions are

$$\begin{aligned} \mathbf{J}_{yx} &= \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \\ \mathbf{J}_{zt} &= \frac{\partial \mathbf{z}}{\partial \mathbf{t}} \end{aligned} \tag{3.54}$$

Letting $\mathbf{r} = \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix}$ and $\mathbf{s} = \begin{bmatrix} \mathbf{x} \\ \mathbf{t} \end{bmatrix}$ equation (3.53) can be written as

$$\mathbf{r} = f(\mathbf{s}) \tag{3.55}$$

The Law of Propagation of Variances (3.34) can be applied to give

$$\mathbf{Q}_{rr} = \mathbf{J}_{rs}\mathbf{Q}_{ss}\mathbf{J}^T_{rs} \tag{3.56}$$

with the Jacobian

$$\mathbf{J}_{rs} = \begin{bmatrix} \mathbf{J}_{yx} & \mathbf{J}_{yt} \\ \mathbf{J}_{zx} & \mathbf{J}_{zt} \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{yx} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt} \end{bmatrix} \tag{3.57}$$

Note that $\mathbf{J}_{yt}$ and $\mathbf{J}_{zx}$ are zero since **y** is independent of **t** and **z** is independent of **x**.

Equation (3.56) can be expanded to give

$$\begin{aligned} \begin{bmatrix} \mathbf{Q}_{yy} & \mathbf{Q}_{yz} \\ \mathbf{Q}_{zy} & \mathbf{Q}_{zz} \end{bmatrix} &= \begin{bmatrix} \mathbf{J}_{yx} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{Q}_{xt} \\ \mathbf{Q}_{tx} & \mathbf{Q}_{tt} \end{bmatrix} \begin{bmatrix} \mathbf{J}_{yx} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt} \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{J}_{yx} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{Q}_{xt} \\ \mathbf{Q}_{tx} & \mathbf{Q}_{tt} \end{bmatrix} \begin{bmatrix} \mathbf{J}_{yx}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt}^T \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{J}_{yx}\mathbf{Q}_{xx}\mathbf{J}_{yx}^T & \mathbf{J}_{yx}\mathbf{Q}_{xt}\mathbf{J}_{zt}^T \\ \mathbf{J}_{zt}\mathbf{Q}_{tx}\mathbf{J}_{yx}^T & \mathbf{J}_{zt}\mathbf{Q}_{tt}\mathbf{J}_{zt}^T \end{bmatrix} \end{aligned}$$

From this expanded equation, we may write the following symbolic equation and four general relationships

$$\text{If } \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} == \begin{bmatrix} y(\mathbf{x}) \\ z(\mathbf{t}) \end{bmatrix} \text{ and } \mathbf{x} \text{ and } \mathbf{t} \text{ are correlated}$$

$$\begin{aligned} \mathbf{Q}_{yy} &= \mathbf{J}_{yx}\mathbf{Q}_{xx}\mathbf{J}_{yx}^T \\ \mathbf{Q}_{yz} &= \mathbf{J}_{yx}\mathbf{Q}_{xt}\mathbf{J}_{zt}^T \\ \mathbf{Q}_{zy} &= \mathbf{J}_{zt}\mathbf{Q}_{tx}\mathbf{J}_{yz}^T \\ \mathbf{Q}_{zz} &= \mathbf{J}_{zt}\mathbf{Q}_{tt}\mathbf{J}_{zt}^T \end{aligned} \tag{3.58}$$

The cofactor matrices where the subscripts are different letters, i.e., $\mathbf{Q}_{xt}$, $\mathbf{Q}_{tx}$, $\mathbf{Q}_{yz}$ and $\mathbf{Q}_{zy}$ are crosscofactor matrices and their form is easily constructed, for example

$$\underset{m,n}{\mathbf{Q}_{xt}} = \begin{bmatrix} q_{x_1,t_1} & q_{x_1,t_2} & \cdots & q_{x_1,t_n} \\ \vdots & & & \\ q_{x_m,t_1} & q_{x_m,t_2} & \cdots & q_{x_m,t_n} \end{bmatrix} \tag{3.59}$$

This directly leads to the fact that crosscofactor matrices are not necessarily square and symmetric, as cofactor matrices are, and that

$$\underset{m,n}{\mathbf{Q}_{xt}} = \mathbf{Q}_{tx}^T$$

3.6.1. Symbolic Multiplication in Propagation Using Matrices

To assist the practitioner in variance propagation (or cofactor propagation) of linear and non-linear functions of random variables a technique known as *symbolic multiplication* can be used. This mnemo-technical rule was originally devised by Tienstra (1966) to obtain covariances of random variables related by systems of linear (or linearized) equations. His rule, developed before the extensive use of matrix algebra, can be employed with matrix equations in the following manner. For example, for linear functions

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a} \qquad \text{(a)}$$

$$\mathbf{z} = \mathbf{B}\mathbf{t} + \mathbf{b} \qquad \text{(b)}$$

The constant vectors **a** and **b** play no part in variance propagation and can be ignored, and the crosscofactor matrix $\mathbf{Q}_{yz}$ is

$$\mathbf{Q}_{yz} = \mathbf{A}\mathbf{Q}_{xt}\mathbf{B}^T \qquad \text{(c)}$$

Equation (c) can be obtained by symbolic multiplication by:

(i) writing the cofactor matrix **Q** on the left-hand-side of the equals sign with subscripts *y* and *z* representing the vectors **y** and **z** on the left-hand-sides of equations (a) and (b) in the order of (a) first and (b) second, then

(ii) writing the coefficient matrix of the random variable in equation (a) on the right-hand-side of the equals sign, then

(iii) multiplying by the cofactor matrix **Q** with the subscripts *x* and *t* representing the random vectors **x** and **t** on the right-hand-sides of equations (a) and (b) in the order of (a) first and (b) second, then

(iv) multiplying by the transpose of the coefficient matrix of the random variable on the right-hand-side of equation (b).

Note the in the case of non-linear functions, the coefficient matrices $\mathbf{J}_{yx}$ and $\mathbf{J}_{zt}$ replace **A** and **B** respectively, and equation (c) becomes identical to the 2nd equation of (3.58).

3.6.2. Further Remarks on Propagation Using Matrices

The rules and techniques of propagation of variances (and cofactors) given in the preceding sections allow for propagation of variances through several transformations. These propagations can be carried out in two ways, (i) substitution and (ii) stepwise. To demonstrate these we consider the following three relations

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{x} + \mathbf{a} \\ \mathbf{z} &= \mathbf{B}\mathbf{y} + \mathbf{b} \\ \mathbf{r} &= \mathbf{C}\mathbf{z} + \mathbf{c} \end{aligned} \tag{3.60}$$

Let the random vector **x** be known, with its cofactor matrix $\mathbf{Q}_{xx}$ and it is desired to obtain the cofactor matrices of **z** and **r**. The vectors **a**, **b** and **c** contain constants and **A**, **B** and **C** are coefficient matrices.

Propagation Through Substitution

We express **z** and **r** in terms of **x** by substitution:

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{x} + \mathbf{a} \\ \mathbf{z} = \mathbf{B}\mathbf{y} + \mathbf{b} &= \mathbf{B}\left(\mathbf{A}\mathbf{x} + \mathbf{a}\right) + \mathbf{b} \\ &= \left(\mathbf{B}\mathbf{A}\right)\mathbf{x} + \left(\mathbf{B}\mathbf{a} + \mathbf{b}\right) \\ \mathbf{r} = \mathbf{C}\mathbf{z} + \mathbf{c} &= \mathbf{C}\left(\mathbf{B}\mathbf{y} + \mathbf{b}\right) + \mathbf{c} \\ &= \left(\mathbf{C}\mathbf{B}\right)\mathbf{y} + \left(\mathbf{C}\mathbf{b} + \mathbf{c}\right) \\ &= \left(\mathbf{C}\mathbf{B}\right)\left(\mathbf{A}\mathbf{x} + \mathbf{a}\right) + \left(\mathbf{C}\mathbf{b} + \mathbf{c}\right) \\ &= \left(\mathbf{C}\mathbf{B}\mathbf{A}\right)\mathbf{x} + \left(\mathbf{C}\mathbf{B}\mathbf{a} + \mathbf{C}\mathbf{b} + \mathbf{c}\right) \end{aligned}$$

Noting that the last terms in the equations for **z** and **r** evaluate to vectors, say **d** and **e** we may rewrite the equations and apply propagation using symbolic multiplication to give

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{x} + \mathbf{a} & \mathbf{Q}_{yy} &= \mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T \\ \mathbf{z} &= \left(\mathbf{B}\mathbf{A}\right)\mathbf{x} + \mathbf{d} & \mathbf{Q}_{zz} &= \left(\mathbf{B}\mathbf{A}\right)\mathbf{Q}_{xx}\left(\mathbf{B}\mathbf{A}\right)^T &&= \mathbf{B}\mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T \\ \mathbf{r} &= \left(\mathbf{C}\mathbf{B}\mathbf{A}\right)\mathbf{x} + \mathbf{e} & \mathbf{Q}_{rr} &= \left(\mathbf{C}\mathbf{B}\mathbf{A}\right)\mathbf{Q}_{xx}\left(\mathbf{C}\mathbf{B}\mathbf{A}\right)^T &&= \mathbf{C}\mathbf{B}\mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \\ && \mathbf{Q}_{yz} &= \mathbf{A}\mathbf{Q}_{xx}\left(\mathbf{B}\mathbf{A}\right)^T &&= \mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T \\ && \mathbf{Q}_{yr} &= \mathbf{A}\mathbf{Q}_{xx}\left(\mathbf{C}\mathbf{B}\mathbf{A}\right)^T &&= \mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \\ && \mathbf{Q}_{zr} &= \left(\mathbf{B}\mathbf{A}\right)\mathbf{Q}_{xx}\left(\mathbf{C}\mathbf{B}\mathbf{A}\right)^T &&= \mathbf{B}\mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \end{aligned} \tag{3.61}$$

Stepwise Propagation

The same result in equations (3.61) can be obtained by applying propagation in steps as follows:

$$\begin{aligned}
&\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a} && \mathbf{Q}_{yy} = \mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T \\
&\mathbf{z} = \mathbf{B}\mathbf{y} + \mathbf{b} && \mathbf{Q}_{zz} = \mathbf{B}\mathbf{Q}_{yy}\mathbf{B}^T = \mathbf{B}\mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T \\
&\mathbf{r} = \mathbf{C}\mathbf{z} + \mathbf{c} && \mathbf{Q}_{rr} = \mathbf{C}\mathbf{Q}_{zz}\mathbf{C}^T = \mathbf{C}\mathbf{B}\mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \\
& && \mathbf{Q}_{yz} = \mathbf{A}\mathbf{Q}_{xy}\mathbf{B}^T \\
& && \mathbf{Q}_{yr} = \mathbf{A}\mathbf{Q}_{xz}\mathbf{C}^T \\
& && \mathbf{Q}_{zr} = \mathbf{B}\mathbf{Q}_{yz}\mathbf{C}^T
\end{aligned} \tag{3.62}$$

The last three (crosscofactor) relations of equations (3.62) do not correspond to those in equations (3.61), particularly because of the absence of the matrices $\mathbf{Q}_{xy}$, $\mathbf{Q}_{xz}$ and $\mathbf{Q}_{yz}$. However, these matrices can be derived if equations (3.60) are supplemented by simple identities $\mathbf{x} = \mathbf{I}\mathbf{x}$ and $\mathbf{y} = \mathbf{I}\mathbf{y}$ giving the following three pairs of equations from which the matrix relationships can be obtained by symbolic multiplication.

$$\begin{aligned}
&\mathbf{x} = \mathbf{I}\mathbf{x} && \mathbf{Q}_{xy} = \mathbf{I}\mathbf{Q}_{xx}\mathbf{A}^T = \mathbf{Q}_{xx}\mathbf{A}^T \\
&\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a} && \\
&\mathbf{x} = \mathbf{I}\mathbf{x} && \mathbf{Q}_{xz} = \mathbf{I}\mathbf{Q}_{xy}\mathbf{B}^T = \mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T \\
&\mathbf{z} = \mathbf{B}\mathbf{y} + \mathbf{b} && \\
&\mathbf{y} = \mathbf{I}\mathbf{y} && \mathbf{Q}_{yz} = \mathbf{I}\mathbf{Q}_{yy}\mathbf{B}^T \\
&\mathbf{z} = \mathbf{B}\mathbf{y} + \mathbf{b} &&
\end{aligned} \tag{3.63}$$

Substituting these relations into the last three relations in (3.62) leads directly to equations (3.61). This demonstrates that propagation through substitution is equivalent to stepwise propagation.

4. APPROXIMATE VALUES

In many least squares problems, the unknown quantities being sought may be quite large numbers and or the coefficients of these quantities may be large numbers. This can lead to numerical problems in the formation of normal equations where large numbers are multiplied and summed. To overcome this problem, approximate values of the unknown quantities may be used and small, unknown corrections to the approximate values become the quantities being sought.

In general, we denote unknown values as x, approximate values as x^0 and small corrections as Δx or δx and

$$x = x^0 + \delta x \tag{4.1}$$

In the case of a vector of unknown quantities $\mathbf{x}$ we have a vector of approximate values $\mathbf{x}^0$ and a vector of small corrections $\delta\mathbf{x}$ and

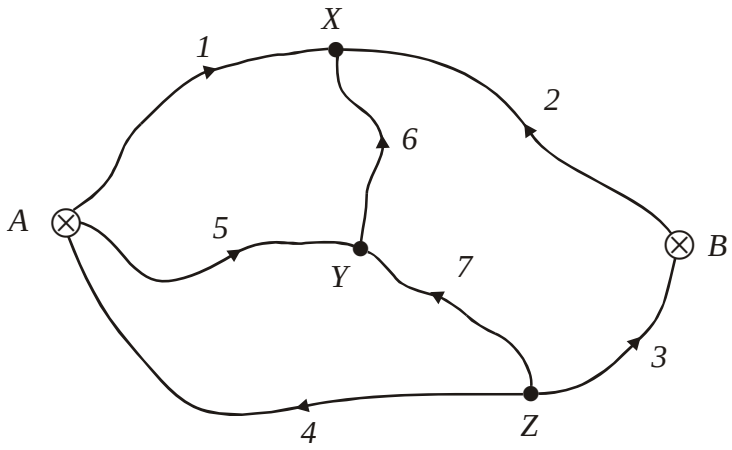
$$\mathbf{x} = \mathbf{x}^0 + \delta\mathbf{x} \tag{4.2}$$

The use of approximate values can best be explained by example and the following sections contain worked examples of some simple least squares problems that demonstrate the use of approximate values.

4.1. LEVEL NET ADJUSTMENT

The diagram below shows a level network of height differences observed between the fixed stations *A* (RL 102.440 m) and *B* (RL 104.565 m) and "floating" stations *X, Y* and *Z* whose Reduced Levels (RL's) are unknown. The arrows on the diagram indicate the direction of rise. The Table of Height differences shows the height difference for each line of the network and the distance (in kilometers) of each level run.

Line	Height Diff	Dist (km)
1	6.345	1.7
2	4.235	2.5
3	3.060	1.0
4	0.920	3.8
5	3.895	1.7
6	2.410	1.2
7	4.820	1.5



The method of Least Squares can be used to determine the best estimates of the RL's of *X*, *Y* and *Z* bearing in mind that the precision of the observed height differences is inversely proportional to the distance of the level run.

The observation equation for the RL's of two points *P* and *Q* connected by an observed spirit levelled height difference ΔH_{PQ} can be written as

$$P + \Delta H_{PQ} + v_{PQ} = Q \tag{4.3}$$

where *P* and *Q* are the RL's of points *P* and *Q* and v_{PQ} is the residual, a small unknown correction to the observed height difference. If the RL's of *P* and *Q* are unknown but have approximate values, say $P = P^0 + \delta P$ and $Q = Q^0 + \delta Q$ we may write a general observation equation for an observed height difference as

$$P^0 + \delta P + \Delta H_{PQ} + v_{PQ} = Q^0 + \delta Q \tag{4.4}$$

Using this general observation equation we may write an equation for each observed height difference

$$\begin{aligned}
A \qquad & +\Delta H_1 + v_1 = X^0 + \delta X \\
B \qquad & +\Delta H_2 + v_2 = X^0 + \delta X \\
Z^0 + \delta Z & +\Delta H_3 + v_3 = B \\
Z^0 + \delta Z & +\Delta H_4 + v_4 = A \\
A \qquad & +\Delta H_5 + v_5 = Y^0 + \delta Y \\
Y^0 + \delta Y & +\Delta H_6 + v_6 = X^0 + \delta X \\
Z^0 + \delta Z & + \Delta H_7 + v_7 = Y^0 + \delta Y
\end{aligned}$$

Rearranging these equations so that all the unknown quantities are on the left-hand-side of the equals sign and all the known quantities are on the right-hand-side gives

$$\begin{aligned}
v_1 - \delta X &= X^0 - A - \Delta H_1 \\
v_2 - \delta X &= X^0 - B - \Delta H_2 \\
v_3 + \delta Z &= B - Z^0 - \Delta H_3 \\
v_4 + \delta Z &= A - Z^0 - \Delta H_4 \\
v_5 - \delta Y &= Y^0 - A - \Delta H_5 \\
v_6 - \delta X + \delta Y &= X^0 - Y^0 - \Delta H_6 \\
v_7 - \delta Y + \delta Z &= Y^0 - Z^0 - \Delta H_7
\end{aligned}$$

The approximate RL's of the unknown points *X*, *Y* and *Z* can be determined from the RL's of *A* and *B* and appropriate height differences

$$\begin{aligned}
X^0 &= A + \Delta H_1 = 108.785 \text{ m} \\
Y^0 &= A + \Delta H_5 = 106.335 \text{ m} \\
Z^0 &= A - \Delta H_4 = 101.520 \text{ m}
\end{aligned}$$

Writing these equations in the standard form $\mathbf{v} + \mathbf{Bx} = \mathbf{f}$ gives

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix} + \begin{bmatrix} -1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \delta X \\ \delta Y \\ \delta Z \end{bmatrix} = \begin{bmatrix} \left(X^0 - A\right) - \Delta H_1 \\ \left(X^0 - B\right) - \Delta H_2 \\ \left(B - Z^0\right) - \Delta H_3 \\ \left(A - Z^0\right) - \Delta H_4 \\ \left(Y^0 - A\right) - \Delta H_5 \\ \left(X^0 - Y^0\right) - \Delta H_6 \\ \left(Y^0 - Z^0\right) - \Delta H_7 \end{bmatrix} = \begin{bmatrix} 0.000 \\ -0.015 \\ -0.015 \\ 0.000 \\ 0.000 \\ 0.040 \\ -0.005 \end{bmatrix}$$

The weight matrix for the adjustment is

$$\begin{aligned}
\mathbf{W} &= \text{diag}\left[\frac{1}{1.7} \quad \frac{1}{2.5} \quad \frac{1}{1} \quad \frac{1}{3.8} \quad \frac{1}{1.7} \quad \frac{1}{1.2} \quad \frac{1}{1.5}\right] \\
&= \text{diag}\left[0.5882 \quad 0.4000 \quad 1.0000 \quad 0.2632 \quad 0.5882 \quad 0.8333 \quad 0.6667\right]
\end{aligned}$$

The least squares solution for the vector of corrections **x** can be obtained from the MATLAB function *least_squares.m* with the following data file `c:\Temp\Level_Net_Data.dat`

```
% Data file for Level Net Adjustment
%
%  dX   dY   dZ    f     weight
   -1   0    0    0.000  0.5882
   -1   0    0   -0.015  0.4000
    0   0    1   -0.015  1.0000
    0   0    1    0.000  0.2632
    0  -1    0    0.000  0.5882
   -1   1    0    0.040  0.8333
    0  -1    1   -0.005  0.6667
```

Running the program from the MATLAB command window created the following output file `c:\Temp|Level_Net_Data.out`

```
Least Squares Adjustment of Indirect Observations

Input Data

Coefficient matrix B of observation equations v + Bx = f
   -1.0000    0.0000    0.0000
   -1.0000    0.0000    0.0000
    0.0000    0.0000    1.0000
    0.0000    0.0000    1.0000
    0.0000   -1.0000    0.0000
   -1.0000    1.0000    0.0000
    0.0000   -1.0000    1.0000

Vector of numeric terms f and weights w of observation equations v + Bx = f
    0.0000      0.5882
   -0.0150      0.4000
   -0.0150      1.0000
    0.0000      0.2632
    0.0000      0.5882
    0.0400      0.8333
   -0.0050      0.6667

Coefficient matrix N of Normal equations Nx = t
      1.8215     -0.8333      0.0000
     -0.8333      2.0882     -0.6667
      0.0000     -0.6667      1.9299

Vector of numeric terms t of Normal equations Nx = t
   -0.0273
    0.0367
   -0.0183

Inverse of Normal equation coefficient matrix
     6.9073e-001     3.0981e-001     1.0703e-001
     3.0981e-001     6.7720e-001     2.3394e-001
     1.0703e-001     2.3394e-001     5.9898e-001

Vector of solutions x
   -0.0095
    0.0121
   -0.0053
```

```
Vector of residuals v
   -0.0095
   -0.0245
   -0.0097
    0.0053
    0.0121
    0.0184
    0.0124
```

The adjusted RL's of X, Y and Z are

$$X = X^0 + \delta X = 108.785 - 0.0095 = 108.776 \text{ m}$$
$$Y = Y^0 + \delta Y = 106.335 + 0.0121 = 106.347 \text{ m}$$
$$Z = Z^0 + \delta Z = 101.520 - 0.0053 = 101.515 \text{ m}$$

The adjusted height differences are

Line	Observed ΔH	Residual v	Adjusted ΔH
1	6.345	-0.0095	6.336
2	4.235	-0.0245	4.211
3	3.060	-0.0097	3.050
4	0.920	0.0053	0.925
5	3.895	0.0121	3.907
6	2.410	0.0184	2.428
7	4.820	0.0124	4.832

5. PROPAGATION OF VARIANCES APPLIED TO LEAST SQUARES ADJUSTMENT OF INDIRECT OBSERVATIONS

A most important outcome of a least squares adjustment is that estimates of the precisions of the quantities sought, the elements of **x**, the unknowns or the parameters, are easily obtained from the matrix equations of the solution. Application of the Law of Propagation of Variances demonstrates that $\mathbf{N}^{-1}$, the inverse of the normal equation coefficient matrix is equal to the cofactor matrix $\mathbf{Q}_{xx}$ that contains estimates of the variances and covariances of the elements of **x**. In addition, estimates of the precisions of the residuals and adjusted observations may be obtained. This most useful outcome enables a statistical analysis of the results of a least squares adjustment and provides the practitioner with a degree of confidence in the results.

5.1. Cofactor matrices for adjustment of indirect observations

The observation equations for adjustment of indirect observations is given by

$$\mathbf{v} + \mathbf{B}\mathbf{x} = \mathbf{f} \tag{5.1}$$

f is an (*n*,1) vector of numeric terms derived from the (*n*,1) vector of observations **l** and the (*n*,1) vector of constants **d** as

$$\mathbf{f} = \mathbf{d} - \mathbf{l} \tag{5.2}$$

Associated with the vector of observations **l** is a variance-covariance matrix Σ_{ll} as well as a cofactor matrix $\mathbf{Q}_{ll}$ and a weight matrix $\mathbf{W}_{ll} = \mathbf{Q}_{ll}^{-1}$. Remember that in most practical applications of least squares, the matrix Σ_{ll} is unknown, but estimated *a priori* by $\mathbf{Q}_{ll}$ that contains estimates of the variances and covariances and $\mathbf{\Sigma}_{ll} = \sigma_0^2 \mathbf{Q}_{ll}$ where σ_0^2 is the reference variance or variance factor.

Note: In the derivations that follow, the subscript *"ll"* is dropped from $\mathbf{Q}_{ll}$ and $\mathbf{W}_{ll}$

If equation (5.2) is written as

$$\mathbf{f} = (-\mathbf{I})\mathbf{l} + \mathbf{d} \tag{5.3}$$

then (5.3) is in a form suitable for employing the Law of Propagation of Variances developed in Chapter 3; i.e., if $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ and **y** and **x** are random variables linearly related and **b** is a vector of constants then $\mathbf{Q}_{yy} = \mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T$. Hence, the cofactor matrix of the numeric terms **f** is

$$\mathbf{Q}_{ff} = (-\mathbf{I})\mathbf{Q}(-\mathbf{I})^T = \mathbf{Q}$$

Thus the cofactor matrix of **f** is also the *a priori* cofactor matrix of the observations **l**.

The solution "steps" in the least squares adjustment of indirect observations are set out Chapter 2 and restated as

$$\begin{aligned}
\mathbf{N} &= \mathbf{B}^T\mathbf{W}\mathbf{B} \\
\mathbf{t} &= \mathbf{B}^T\mathbf{W}\mathbf{f} \\
\mathbf{x} &= \mathbf{N}^{-1}\mathbf{t} \\
\mathbf{v} &= \mathbf{f} - \mathbf{B}\mathbf{x} \\
\hat{\mathbf{l}} &= \mathbf{l} + \mathbf{v}
\end{aligned}$$

To apply the Law of Propagation of Variances, these equations may be re-arranged in the form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ where the terms in parenthesis ( ) constitute the **A** matrix.

$$\mathbf{t} = \left(\mathbf{B}^T\mathbf{W}\right)\mathbf{f} \tag{5.4}$$

$$\mathbf{x} = \left(\mathbf{N}^{-1}\right)\mathbf{t} \tag{5.5}$$

$$\begin{aligned}
\mathbf{v} &= \mathbf{f} - \mathbf{B}\mathbf{x} \\
&= \mathbf{f} - \mathbf{B}\mathbf{N}^{-1}\mathbf{t} \\
&= \mathbf{f} - \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}\mathbf{f} \\
&= \left(\mathbf{I} - \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}\right)\mathbf{f}
\end{aligned} \tag{5.6}$$

$$\begin{aligned}
\hat{\mathbf{l}} &= \mathbf{l} + \mathbf{v} \\
&= \mathbf{l} + \mathbf{f} - \mathbf{B}\mathbf{x} \\
&= \mathbf{d} - \mathbf{B}\mathbf{x} \\
&= (-\mathbf{B})\mathbf{x} + \mathbf{d}
\end{aligned} \tag{5.7}$$

Applying the Law of Propagation of Variances to equations (5.4) to (5.7) gives the following cofactor matrices

$$\mathbf{Q}_{tt} = \left(\mathbf{B}^T\mathbf{W}\right)\mathbf{Q}_{ff}\left(\mathbf{B}^T\mathbf{W}\right)^T = \mathbf{N} \tag{5.8}$$

$$\mathbf{Q}_{xx} = \left(\mathbf{N}^{-1}\right)\mathbf{Q}_{tt}\left(\mathbf{N}^{-1}\right)^T = \mathbf{N}^{-1} \tag{5.9}$$

$$\begin{aligned}\mathbf{Q}_{vv} &= \left(\mathbf{I} - \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}\right)\mathbf{Q}_{ff}\left(\mathbf{I} - \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}\right)^T \\ &= \mathbf{Q} - \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\end{aligned} \tag{5.10}$$

$$\begin{aligned}\mathbf{Q}_{\hat{l}\hat{l}} &= \left(-\mathbf{B}\right)\mathbf{Q}\left(-\mathbf{B}\right)^T \\ &= \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T \\ &= \mathbf{Q} - \mathbf{Q}_{vv}\end{aligned} \tag{5.11}$$

Variance-covariance matrices for **t**, **x**, **v** and $\hat{\mathbf{l}}$ are obtained by multiplying the cofactor matrix by the variance factor σ_0^2.

5.2. Calculation of the quadratic form $\mathbf{v}^T\mathbf{W}\mathbf{v}$

The *a priori* estimate of the variance factor may be computed from

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T\mathbf{W}\mathbf{v}}{r} \tag{5.12}$$

where $\mathbf{v}^T\mathbf{W}\mathbf{v}$ is the quadratic form, and

$r = n - u$ is the degrees of freedom where n is the number of observations and u is the number of unknown parameters. r is also known as the number of redundancies.

A derivation of equation (5.12) is given below. The quadratic form $\mathbf{v}^T\mathbf{W}\mathbf{v}$ may be computed in the following manner.

Remembering, for the method of indirect observations, the following matrix equations

$$\begin{aligned}\mathbf{N} &= \mathbf{B}^T\mathbf{W}\mathbf{B} \\ \mathbf{t} &= \mathbf{B}^T\mathbf{W}\mathbf{f} \\ \mathbf{x} &= \mathbf{N}^{-1}\mathbf{t} \\ \mathbf{v} &= \mathbf{f} - \mathbf{B}\mathbf{x}\end{aligned}$$

then

$$\begin{aligned}\mathbf{v}^T\mathbf{W}\mathbf{v} &= (\mathbf{f}-\mathbf{B}\mathbf{x})^T\mathbf{W}(\mathbf{f}-\mathbf{B}\mathbf{x})\\ &= (\mathbf{f}^T-\mathbf{x}^T\mathbf{B}^T)\mathbf{W}(\mathbf{f}-\mathbf{B}\mathbf{x})\\ &= (\mathbf{f}^T\mathbf{W}-\mathbf{x}^T\mathbf{B}^T\mathbf{W})(\mathbf{f}-\mathbf{B}\mathbf{x})\\ &= \mathbf{f}^T\mathbf{W}\mathbf{f}-\mathbf{f}^T\mathbf{W}\mathbf{B}\mathbf{x}-\mathbf{x}^T\mathbf{B}^T\mathbf{W}\mathbf{f}+\mathbf{x}^T\mathbf{B}^T\mathbf{W}\mathbf{B}\mathbf{x}\\ &= \mathbf{f}^T\mathbf{W}\mathbf{f}-2\mathbf{f}^T\mathbf{W}\mathbf{B}\mathbf{x}+\mathbf{x}^T\mathbf{B}^T\mathbf{W}\mathbf{B}\mathbf{x}\\ &= \mathbf{f}^T\mathbf{W}\mathbf{f}-2\mathbf{t}^T\mathbf{x}+\mathbf{x}^T\mathbf{N}\mathbf{x}\\ &= \mathbf{f}^T\mathbf{W}\mathbf{f}-2\mathbf{x}^T\mathbf{t}+\mathbf{x}^T\mathbf{t}\end{aligned}$$

and

$$\mathbf{v}^T\mathbf{W}\mathbf{v} = \mathbf{f}^T\mathbf{W}\mathbf{f}-\mathbf{x}^T\mathbf{t} \tag{5.13}$$

5.3. Calculation of the Estimate of the Variance Factor $\hat{\sigma}_0^2$

The variance-covariance matrices of residuals $\mathbf{\Sigma}_{vv}$, adjusted observations $\mathbf{\Sigma}_{\hat{l}\hat{l}}$ and computed parameters $\mathbf{\Sigma}_{xx}$ are calculated from the general relationship

$$\mathbf{\Sigma} = \sigma_0^2\mathbf{Q} \tag{5.14}$$

Cofactor matrices $\mathbf{Q}_{vv}$, $\mathbf{Q}_{xx}$ and $\mathbf{Q}_{\hat{l}\hat{l}}$ are computed from equations (5.9) to (5.11) and so it remains to determine an estimate of the variance factor $\hat{\sigma}_0^2$.

The development of a matrix expression for computing $\hat{\sigma}_0^2$ is set out below and follows Mikhail (1976, pp.285-288). Some preliminary relationships will be useful.

1. If **A** is an (*n*,*n*) square matrix, the sum of its diagonal elements is a scalar quantity called the trace of **A** and denoted by $tr(\mathbf{A})$ The following relationships are useful

$$tr(\mathbf{A}+\mathbf{B}) = tr(\mathbf{A})+tr(\mathbf{B}) \quad \text{for } \mathbf{A} \text{ and } \mathbf{B} \text{ of same order} \tag{5.15}$$

$$tr(\mathbf{A}^T) = tr(\mathbf{A}) \tag{5.16}$$

and for the quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ where **A** is symmetric

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = tr(\mathbf{x}\mathbf{x}^T\mathbf{A}) \tag{5.17}$$

2. The variance-covariance matrix $\boldsymbol{\Sigma}_{xx}$ given by equation (3.21) can be expressed in the following manner, remembering that **x** is a vector of random variables and $\mathbf{m}_x$ is a vector of means.

$$\begin{aligned}\Sigma_{xx} &= E\left\{(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^T\right\}\\ &= E\left\{(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}^T-\mathbf{m}_x^T)\right\}\\ &= E\left\{\mathbf{x}\mathbf{x}^T-\mathbf{x}\mathbf{m}_x^T-\mathbf{m}_x\mathbf{x}^T+\mathbf{m}_x\mathbf{m}_x^T\right\}\\ &= E\left\{\mathbf{x}\mathbf{x}^T\right\}-E\left\{\mathbf{x}\mathbf{m}_x^T\right\}-E\left\{\mathbf{m}_x\mathbf{x}^T\right\}+E\left\{\mathbf{m}_x\mathbf{m}_x^T\right\}\\ &= E\left\{\mathbf{x}\mathbf{x}^T\right\}-E\left\{\mathbf{x}\right\}\mathbf{m}_x^T-\mathbf{m}_x E\left\{\mathbf{x}^T\right\}+\mathbf{m}_x\mathbf{m}_x^T\end{aligned}$$

Now from equation (3.18) $\mathbf{m}_x = E\{\mathbf{x}\}$ hence

$$\begin{aligned}\Sigma_{xx} &= E\left\{\mathbf{x}\mathbf{x}^T\right\}-\mathbf{m}_x\mathbf{m}_x^T-\mathbf{m}_x\mathbf{m}_x^T+\mathbf{m}_x\mathbf{m}_x^T\\ &= E\left\{\mathbf{x}\mathbf{x}^T\right\}-\mathbf{m}_x\mathbf{m}_x^T\end{aligned} \qquad (5.18)$$

or $$E\left\{\mathbf{x}\mathbf{x}^T\right\} = \boldsymbol{\Sigma}_{xx}+\mathbf{m}_x\mathbf{m}_x^T \qquad (5.19)$$

3. The expected value of the residuals is zero, i.e.,

$$E\{\mathbf{v}\} = \mathbf{m}_v = \mathbf{0} \qquad (5.20)$$

4. By definition (see Chapter 2) the weight matrix **W**, the cofactor matrix **Q** and the variance-covariance matrix $\boldsymbol{\Sigma}$ are related by

$$\mathbf{W} = \mathbf{Q}^{-1} = \sigma_0^2\boldsymbol{\Sigma}^{-1} \qquad (5.21)$$

Now, for the least squares adjustment of indirect observations the following relationships are recalled

$$\mathbf{v}+\mathbf{B}\mathbf{x} = \mathbf{f}, \quad \mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B}, \quad \mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f}$$
$$\mathbf{Q}_{ff} = \mathbf{Q}, \quad \mathbf{Q}_{tt} = \mathbf{N}, \quad \mathbf{Q}_{xx} = \mathbf{N}^{-1}$$

Bearing in mind equation (5.21), the following relationships may be introduced

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_0^2}\mathbf{W}, \quad \mathbf{M} = \mathbf{B}^T\boldsymbol{\Sigma}^{-1}\mathbf{B}$$

and from these follow

$$\Sigma_{ff} = \mathbf{\Sigma}, \quad \Sigma_{tt} = \sigma_0^4 \mathbf{M}, \quad \mathbf{\Sigma}_{xx} = \mathbf{M}^{-1}$$

In addition, the expectation of the vector **f** is the mean $\mathbf{m}_f$ and so we may write

$$\mathbf{m}_f = E\{\mathbf{f}\} = E\{\mathbf{v} + \mathbf{B}\mathbf{x}\} = E\{\mathbf{v}\} + \mathbf{B}E\{\mathbf{x}\}$$

Now since $E\{\mathbf{x}\} = \mathbf{m}_x$ and $E\{\mathbf{v}\} = \mathbf{0}$

$$\mathbf{m}_f = \mathbf{B}\mathbf{m}_x \tag{5.22}$$

Now the quadratic form

$$\mathbf{v}^T\mathbf{W}\mathbf{v} = \sigma_0^2\left(\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v}\right) \tag{5.23}$$

and from equation (5.13)

$$\begin{aligned}\mathbf{v}^T\mathbf{W}\mathbf{v} &= \mathbf{f}^T\mathbf{W}\mathbf{f} - \mathbf{x}^T\mathbf{t} \\ &= \mathbf{f}^T\mathbf{W}\mathbf{f} - \mathbf{x}^T\mathbf{N}\mathbf{x}\end{aligned}$$

Using the relationships above

$$\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v} = \mathbf{f}^T\mathbf{\Sigma}^{-1}\mathbf{f} - \mathbf{x}^T\mathbf{M}\mathbf{x}$$

Now the expected value of this quadratic form is

$$\begin{aligned}E\left\{\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v}\right\} &= E\left\{\mathbf{f}^T\mathbf{\Sigma}^{-1}\mathbf{f} - \mathbf{x}^T\mathbf{M}\mathbf{x}\right\} \\ &= E\left\{\mathbf{f}^T\mathbf{\Sigma}^{-1}\mathbf{f}\right\} - E\left\{\mathbf{x}^T\mathbf{M}\mathbf{x}\right\}\end{aligned}$$

Recognising that the terms on the right-hand-side are both quadratic forms, equation (5.17) can be used to give

$$\begin{aligned}E\left\{\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v}\right\} &= E\left\{tr\left(\mathbf{f}\mathbf{f}^T\mathbf{\Sigma}^{-1}\right)\right\} - E\left\{tr\left(\mathbf{x}\mathbf{x}^T\mathbf{M}\right)\right\} \\ &= tr\left(E\left\{\mathbf{f}\mathbf{f}^T\mathbf{\Sigma}^{-1}\right\}\right) - tr\left(E\left\{\mathbf{x}\mathbf{x}^T\mathbf{M}\right\}\right) \\ &= tr\left(E\left\{\mathbf{f}\mathbf{f}^T\right\}\mathbf{\Sigma}^{-1}\right) - tr\left(E\left\{\mathbf{x}\mathbf{x}^T\right\}\mathbf{M}\right)\end{aligned}$$

Now using equation (5.19)

$$\begin{aligned}E\left\{\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v}\right\} &= tr\left(\left[\mathbf{\Sigma}_{ff} + \mathbf{m}_f\mathbf{m}_f^T\right]\mathbf{\Sigma}^{-1}\right) - tr\left(\left[\mathbf{\Sigma}_{xx} + \mathbf{m}_x\mathbf{m}_x^T\right]\mathbf{M}\right) \\ &= tr\left(\mathbf{I}_{nn} + \mathbf{m}_f\mathbf{m}_f^T\mathbf{\Sigma}^{-1}\right) - tr\left(\mathbf{I}_{uu} + \mathbf{m}_x\mathbf{m}_x^T\mathbf{M}\right) \\ &= tr\left(\mathbf{I}_{nn} - \mathbf{I}_{uu}\right) - tr\left(\mathbf{m}_f\mathbf{m}_f^T\mathbf{\Sigma}^{-1} + \mathbf{m}_x\mathbf{m}_x^T\mathbf{M}\right) \\ &= \left(n - u\right) - \mathbf{m}_f^T\mathbf{\Sigma}^{-1}\mathbf{m}_f + \mathbf{m}_x^T\mathbf{M}\mathbf{m}_x\end{aligned}$$

From equation (5.22) $\mathbf{m}_f = \mathbf{B}\mathbf{m}_x$ hence using the rule for matrix transpose

$\mathbf{m}_f^T = \left(\mathbf{B}\mathbf{m}_x\right)^T = \mathbf{m}_x^T\mathbf{B}^T$, then

$$\begin{aligned} E\left\{\mathbf{v}^T\boldsymbol{\Sigma}^{-1}\mathbf{v}\right\} &= \left(n-u\right) - \mathbf{m}_x^T\mathbf{B}^T\boldsymbol{\Sigma}^{-1}\mathbf{B}\mathbf{m}_x + \mathbf{m}_x^T\mathbf{M}\mathbf{m}_x \\ &= \left(n-u\right) - \mathbf{m}_x^T\mathbf{M}\mathbf{m}_x + \mathbf{m}_x^T\mathbf{M}\mathbf{m}_x \\ &= \left(n-u\right) \end{aligned}$$

Thus according to equation (5.23) and the expression above

$$\begin{aligned} E\left\{\mathbf{v}^T\mathbf{W}\mathbf{v}\right\} &= \sigma_0^2 E\left\{\mathbf{v}^T\boldsymbol{\Sigma}^{-1}\mathbf{v}\right\} \\ &= \sigma_0^2\left(n-u\right) \end{aligned}$$

from which follows

$$\sigma_0^2 = \frac{E\left\{\mathbf{v}^T\mathbf{W}\mathbf{v}\right\}}{n-u}$$

Consequently, an unbiased estimate of the variance factor $\hat{\sigma}_0^2$ can be computed from

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T\mathbf{W}\mathbf{v}}{n-u} = \frac{\mathbf{v}^T\mathbf{W}\mathbf{v}}{r} \tag{5.24}$$

$r = n-u$ is the number of redundancies in the adjustment and is known as the degrees of freedom

Using equation (5.13) an unbiased estimate of the variance factor $\hat{\sigma}_0^2$ can be computed from

$$\hat{\sigma}_0^2 = \frac{\mathbf{f}^T\mathbf{W}\mathbf{f} - \mathbf{x}^T\mathbf{t}}{r} \tag{5.25}$$

6. LEAST SQUARES ADJUSTMENT OF OBSERVATIONS ONLY

In Chapter 2 the least squares technique of *adjustment of indirect observations* was introduced using the example of fitting a straight line through a series of data points. The "observations" in this example were the *x*,*y* coordinates that were indirect measurements of the unknown parameters *m* and *c,* the slope and intercept of the line on the *y*-axis respectively. Subsequent examples of curve fitting (parabola and ellipse) demonstrated this technique and in Chapter 4 adjustment of indirect observations was applied to a level network. An alternative to this technique, known as *least squares adjustment of observations only*, will be introduced in this chapter using the level network example of Chapter 4.

6.1. Adjustment of a Level Network using Least Squares Adjustment of Observations Only

Figure 6.1 shows a diagram of a level network of height differences observed between the fixed stations *A* (RL 102.440 m) and *B* (RL 104.565 m) and "floating" stations *X, Y* and *Z* whose Reduced Levels (RL's) are unknown. The arrows on the diagram indicate the direction of rise. The Table of Height differences shows the height difference for each line of the network and the distance (in kilometers) of each level run. The height differences can be considered as independent (uncorrelated) and of unequal precision, where the weights of the height differences are defined as being inversely proportional to the distances in kilometres (see Chapter 3, Section 3.5.2)

Line	Height Diff	Dist (km)
1	6.345	1.7
2	4.235	2.5
3	3.060	1.0
4	0.920	3.8
5	3.895	1.7
6	2.410	1.2
7	4.820	1.5

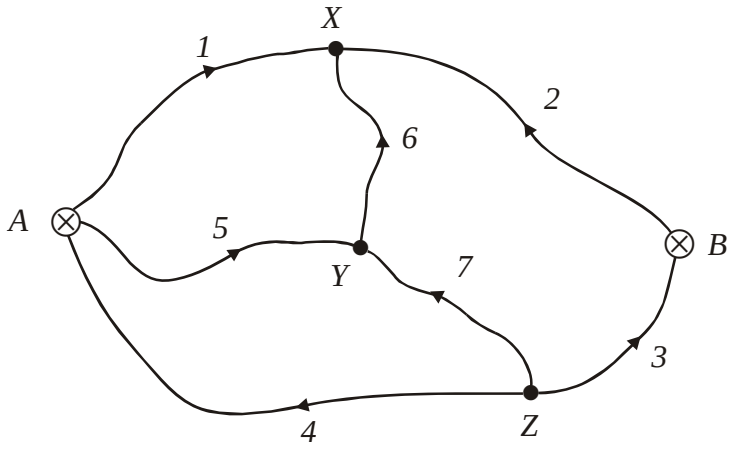


Figure 6.1 Level network

The measured height differences do not accord with the simple principle that they should sum to zero around a "closed loop", i.e., there are misclosures. For example:

in the loop *AXYA* $\quad \Delta H_1 - \Delta H_6 - \Delta H_5 = +0.040$ m

in the loop *XBZYX* $\quad -\Delta H_2 - \Delta H_3 + \Delta H_7 + \Delta H_6 = -0.065$ m

in the loop *AYZA* $\quad \Delta H_5 - \Delta H_7 + \Delta H_4 = -0.005$ m

Hence it is required to determine the adjusted height differences (that will sum to zero) and the RL's of *X, Y* and *Z.*

There are $n = 7$ observations (the measured height differences) and a minimum of $n_0 = 3$ observations are required to fix the RL's of *X, Y* and *Z.* Hence there are $r = n - n_0 = 4$ redundant measurements, which equals the number of independent condition equations. Denoting the observed height differences as l_1, l_2 etc, residuals as v_1, v_2 etc and the RL's of *A* and *B* as *A* and *B,* these condition equations are

$$\begin{aligned} (l_1 + v_1) - (l_6 + v_6) - (l_5 + v_5) &= 0 \\ -(l_2 + v_2) - (l_3 + v_3) + (l_7 + v_7) + (l_6 + v_6) &= 0 \\ (l_5 + v_5) - (l_7 + v_7) + (l_4 + v_4) &= 0 \\ (l_1 + v_1) - (l_2 + v_2) &= B - A \end{aligned} \tag{6.1}$$

The first 3 equations of (6.1) are the loop closure conditions and the last equation is a condition linking the RL's of *A* and *B.*

Since the measurements are of unequal precision, there is an associated weight w_k with each observation and the application of the least squares principle calls for the minimization of the least squares function φ as

$$\varphi = \text{ the sum of the weighted squared residuals} = \sum_{k=1}^{n} w_k v_k^2 \tag{6.2}$$

or $$\varphi = w_1 v_1^2 + w_2 v_2^2 + w_3 v_3^2 + \cdots + w_7 v_7^2$$

Considering equation (6.1) it is clear that separate expressions for residuals cannot be derived and substituted into φ, as was possible in the technique for adjustment of indirect observations (see Chapter 2). Therefore another approach is needed to ensure that φ is a

minimum as well as satisfying equations (6.1). This is accomplished by using a method of function minimization developed by Lagrange[1] and set out in the following manner.

(i) Gather the terms in equations (6.1) together

$$\begin{aligned} v_1 - v_6 - v_5 = 0 - (l_1 - l_6 - l_5) &= f_1 \\ -v_2 - v_3 + v_7 + v_6 = 0 - (-l_2 - l_3 + l_7 + l_6) &= f_2 \\ v_5 - v_7 + v_4 = 0 - (l_5 - l_7 + l_4) &= f_3 \\ v_1 - v_2 = (B - A) - (l_1 - l_2) &= f_4 \end{aligned} \tag{6.3}$$

(ii) Rewrite equations (6.3) in normal form (zero on the right-hand-side)

$$\begin{aligned} v_1 - v_6 - v_5 - f_1 &= 0 \\ -v_2 - v_3 + v_7 + v_6 - f_2 &= 0 \\ v_5 - v_7 + v_4 - f_3 &= 0 \\ v_1 - v_2 - f_4 &= 0 \end{aligned} \tag{6.4}$$

(iii) Now form an augmented function φ' of the form

$$\begin{aligned} \varphi' = {} & w_1 v_1^2 + w_2 v_2^2 + w_3 v_3^2 + \cdots + w_7 v_7^2 \\ & -2k_1 (v_1 - v_6 - v_5 - f_1) - 2k_2 (-v_2 - v_3 + v_7 + v_6 - f_2) \\ & -2k_3 (v_5 - v_7 + v_4 - f_3) - 2k_4 (v_1 - v_2 - f_4) \end{aligned} \tag{6.5}$$

where k_1, k_2, k_3 and k_4 are *Lagrange multipliers* and there are as many multipliers as there are conditions. The introduction of -2 preceding each multiplier is for convenience only. Inspection of equations (6.5), (6.4) and (6.2) show that φ and φ' are equal since the additional terms in φ' equate to zero.

(iv) The unknowns in equation (6.5) are the residuals $v_1, v_2, \ldots, v_7$ and the Lagrange multipliers k_1, k_2, k_3 and k_4, and so for φ' to be a minimum, the partial derivatives of φ' with respect to each of the unknowns must be zero. Setting the

[1] Joseph Louis LAGRANGE (1713-1813), a great French mathematician whose major work was in the calculus of variation, celestial and general mechanics, differential equations and algebra. Lagrange spent 20 years of his life in Prussia and then returned to Paris where his masterpiece, *Mécanique analytique*, published in 1788, formalized much of Newton's work on calculus.

partial derivatives of φ' with respect to the residuals leads to the following equations

$$\begin{aligned}
\frac{\partial\varphi'}{\partial v_1} &= 2w_1v_1 - 2k_1 - 2k_4 = 0 \quad \text{or} \quad v_1 = \frac{1}{w_1}(k_1 + k_4)\\
\frac{\partial\varphi'}{\partial v_2} &= 2w_2v_2 + 2k_2 + 2k_4 = 0 \quad \text{or} \quad v_2 = \frac{1}{w_2}(-k_2 - k_4)\\
\frac{\partial\varphi'}{\partial v_3} &= 2w_3v_3 + 2k_2 = 0 \quad \text{or} \quad v_3 = \frac{1}{w_3}(-k_2)\\
\frac{\partial\varphi'}{\partial v_4} &= 2w_4v_4 - 2k_3 = 0 \quad \text{or} \quad v_4 = \frac{1}{w_4}k_3\\
\frac{\partial\varphi'}{\partial v_5} &= 2w_5v_5 + 2k_1 - 2k_3 = 0 \quad \text{or} \quad v_5 = \frac{1}{w_5}(-k_1 + k_3)\\
\frac{\partial\varphi'}{\partial v_6} &= 2w_6v_6 + 2k_1 - 2k_2 = 0 \quad \text{or} \quad v_6 = \frac{1}{w_6}(-k_1 + k_2)\\
\frac{\partial\varphi'}{\partial v_7} &= 2w_7v_7 - 2k_2 + 2k_3 = 0 \quad \text{or} \quad v_7 = \frac{1}{w_7}(k_2 - k_3)
\end{aligned} \tag{6.6}$$

and when φ' is differentiated with respect to the Lagrange multipliers and equated to zero

$$\begin{aligned}
\frac{\partial\varphi'}{\partial k_1} &= -2(v_1 - v_6 - v_5 - f_1) = 0 \quad \text{or} \quad v_1 - v_6 - v_5 = f_1\\
\frac{\partial\varphi'}{\partial k_2} &= -2(-v_2 - v_3 + v_7 + v_6 - f_2) = 0 \quad \text{or} \quad -v_2 - v_3 + v_7 + v_6 = f_2\\
\frac{\partial\varphi'}{\partial k_3} &= -2(v_5 - v_7 + v_4 - f_3) = 0 \quad \text{or} \quad v_5 - v_7 + v_4 = f_3\\
\frac{\partial\varphi'}{\partial k_4} &= -2(v_1 - v_2 - f_4) = 0 \quad \text{or} \quad v_1 - v_2 = f_4
\end{aligned} \tag{6.7}$$

the original condition equations (6.4) result. This demonstrates that the introduction of Lagrange multipliers ensures that the conditions will be satisfied when φ' is minimized.

(v) Now, substituting equations (6.6) into (6.7) gives four normal equations

$$\begin{aligned}
\left(\frac{1}{w_1}+\frac{1}{w_6}+\frac{1}{w_5}\right)k_1-\frac{1}{w_6}k_2-\frac{1}{w_5}k_3+\frac{1}{w_1}k_4 &= f_1\\
-\frac{1}{w_6}k_1+\left(\frac{1}{w_2}+\frac{1}{w_3}+\frac{1}{w_6}+\frac{1}{w_7}\right)k_2-\frac{1}{w_7}k_3+\frac{1}{w_2}k_4 &= f_2\\
-\frac{1}{w_5}k_1-\frac{1}{w_7}k_2+\left(\frac{1}{w_4}+\frac{1}{w_5}+\frac{1}{w_7}\right)k_3 &= f_3\\
\frac{1}{w_1}k_1+\frac{1}{w_2}k_2+\left(\frac{1}{w_1}+\frac{1}{w_2}\right)k_4 &= f_4
\end{aligned} \qquad (6.8)$$

Equations (6.8) can be solved to give the Lagrange multipliers k_1, k_2, k_3 and k_4, which can be substituted back into equations (6.6) to give the residuals $v_1, v_2, \ldots, v_7$. Note that the coefficient terms $\frac{1}{w_k}$ in equations (6.8) are known as weight reciprocals and in the case of levelling are simply the distances of the level runs in kilometres.

Using the data from Figure 6.1 the weight reciprocals are the distances (in kilometres)

$$\frac{1}{w_k} = \{1.7 \quad 2.5 \quad 1 \quad 3.8 \quad 1.7 \quad 1.2 \quad 1.5\}$$

the numeric terms *f* are given by equations (6.3)

$$\begin{aligned}
f_1 &= -(l_1 - l_6 - l_5) &&= -0.040 \text{ m}\\
f_2 &= -(-l_2 - l_3 + l_7 + l_6) &&= 0.065 \text{ m}\\
f_3 &= -(l_5 - l_7 + l_4) &&= 0.005 \text{ m}\\
f_4 &= (B - A) - (l_1 - l_2) &&= 0.015 \text{ m}
\end{aligned}$$

and the normal equations (in matrix form) are

$$\begin{bmatrix} 4.6 & -1.2 & -1.7 & 1.7 \\ -1.2 & 6.2 & -1.5 & 2.5 \\ -1.7 & -1.5 & 7.0 & 0 \\ 1.7 & 2.5 & 0 & 4.2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} = \begin{bmatrix} -0.040 \\ 0.065 \\ 0.005 \\ 0.015 \end{bmatrix} \qquad (6.9)$$

The solution of equations (6.9) for the Lagrange multipliers gives

$$k_1 = -0.005700, \quad k_2 = 0.009671, \quad k_3 = 0.001402, \quad k_4 = 0.000122$$

Substituting these values (k_1, k_2, k_3 and k_4) together with the weight reciprocals $\frac{1}{w_k}$ into equations (6.6) gives the residuals $v_1, v_2, \ldots, v_7$. The height differences, residuals and the adjusted height differences (observed value + residual) of the level network are shown below.

Line	Observed ΔH	Residual v	Adjusted ΔH
1	6.345	-0.0095	6.336
2	4.235	-0.0245	4.211
3	3.060	-0.0097	3.050
4	0.920	0.0053	0.925
5	3.895	0.0121	3.907
6	2.410	0.0184	2.428
7	4.820	0.0124	4.832

These are identical results to those obtained by least squares adjustment of indirect observations set out in Chapter 4.

6.2. Some Comments on the Two Applications of the Method of Least Squares

1. The method of least squares has been applied in two examples:

   (a) determining the parameters of a "line of best fit" through a number of data points (see Chapter 2) and

   (b) determining the adjusted height differences in a level network.

2. Consider the first example: the *line of best fit.*

   - A mathematical model (equation) was established linking observations, residuals (corrections) and unknown parameters.

- For *n* observations, there is a minimum number n_0 required to determine the *u* unknown parameters. In this case $n_0 = u$ and the number of redundant observations is $r = n - n_0$
- An equation was written for each observation, i.e., there were *n* observation equations. The observation equations were recast as residual equations.
- Since there were *n* equations in *u* unknowns ($n \rangle u$) there is no unique solution and the least squares principle was used to determine the *u* normal equations from which the best estimates of the *u* unknown parameters were calculated.

This technique of least squares "adjustment" is known by various names, some of which are

parametric least squares,
least squares adjustment by observation equations,
least squares adjustment by residual equations, and
least squares adjustment of indirect observations.

The last of these is perhaps the most explicit since each observation is in fact an indirect measurement of the unknown parameters. Least squares adjustment of indirect observations is the name adopted for this technique by Mikhail (1976) and Mikhail & Gracie (1981) and will be used in these notes.

3. Consider the second example: the *level network.*

- A relationship or condition that the observations (and residuals) must satisfy was established. In this case, the condition to be satisfied was that observed height differences (plus some unknown corrections or residuals) should sum to zero around a closed level loop.
- The minimum number of observations n_0 required to fix the heights of *X, Y* and *Z* and satisfy the condition between the fixed points *A* and *B* was determined giving the number of independent condition equations equal to the number of redundant observations $r = n - n_0$.
- There were *r* equations in *n* unknown residuals, and since $r = n - n_0$ was less than *n,* there was no unique solution for the residuals. The least squares principle was

used to determine a set of *r* normal equations, which were solved for *r* Lagrange multipliers which in turn, were used to obtain the *n* residuals.

- The residuals were added to the observations to obtain the adjusted observations which were then used to determine the heights of points *X, Y* and *Z.*

This technique of least squares "adjustment" is known by various names, two of which are

least squares adjustment by condition equations, and
least squares adjustment of observations only.

The second of these is the more explicit since equations involve only observations. No parameters are used. Least squares adjustment of observations only is the name adopted for this technique by Mikhail (1976) and Mikhail & Gracie (1981) and will be used in these notes.

It should be noted that in practice, the method of adjustment of observations only is seldom employed, owing to the difficulty of determining the independent condition equations required as a starting point. This contrasts with the relative ease of the technique of adjustment of indirect observations, where every observation yields an equation of fixed form. Computer solutions of least squares problems almost invariably use the technique of adjustment of indirect observations.

6.2.1. A Note on Independent Condition Equations.

Consider the level network shown in Figure 6.2. The RL of *A* is known and the RL's of *B, C* and *D* are to be determined from the observed height differences. The arrows on the diagram indicate the direction of rise.

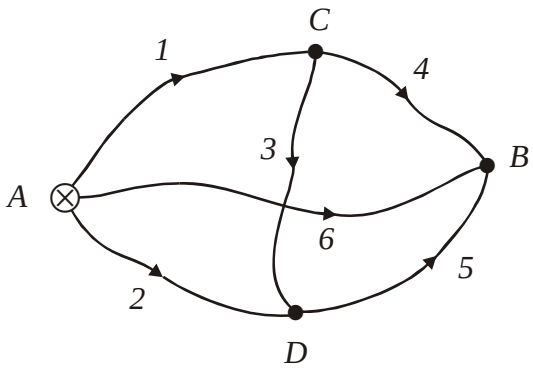


Figure 6.2 Level network

There are $n = 6$ observations with a minimum of $n_0 = 3$ required to fix the RL's of *B, C* and *D* with respect to *A.* Hence there are $r = n - n_0 = 3$ redundant measurements, which equal the number of independent condition equations. Omitting the residuals, these equations are

$$\begin{aligned} l_1 + l_3 - l_2 &= 0 \\ l_4 - l_5 - l_3 &= 0 \\ l_1 + l_4 - l_6 &= 0 \end{aligned} \tag{6.10}$$

Alternatively, here is another set of independent condition equations

$$\begin{aligned} l_1 + l_3 + l_5 - l_6 &= 0 \\ l_1 + l_4 - l_6 &= 0 \\ l_1 + l_4 - l_5 - l_2 &= 0 \end{aligned} \tag{6.11}$$

But, here is a further set of condition equations, which <u>are not</u> independent

$$\begin{aligned} l_1 + l_3 - l_2 &= 0 \\ l_4 - l_5 - l_3 &= 0 \\ l_1 + l_4 - l_5 - l_2 &= 0 \end{aligned} \tag{6.12}$$

where the third equation of (6.12) is obtained by adding the first two.

Care needs to be taken in determining independent equations and it is easy to see that this could become quite difficult as the complexity of the adjustment problem increases.

6.3. Matrix Methods and Least Squares Adjustment of Observations Only

Matrix methods may be used to develop standard equations and solutions for this technique of least squares adjustment.

Consider again the example of the *level net* shown in Figure 6.1. The independent condition equations, (reflecting the fact that height differences around closed level loops should sum to zero and the condition between the known RL's of *A* and *B*), are

$$\begin{aligned} (l_1 + v_1) - (l_6 + v_6) - (l_5 + v_5) &= 0 \\ -(l_2 + v_2) - (l_3 + v_3) + (l_7 + v_7) + (l_6 + v_6) &= 0 \\ (l_5 + v_5) - (l_7 + v_7) + (l_4 + v_4) &= 0 \\ (l_1 + v_1) - (l_2 + v_2) &= B - A \end{aligned} \tag{6.13}$$

These equations could be expressed in matrix form as

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} l_1 + v_1 \\ l_2 + v_2 \\ l_3 + v_3 \\ l_4 + v_4 \\ l_5 + v_5 \\ l_6 + v_6 \\ l_7 + v_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ B - A \end{bmatrix} \tag{6.14}$$

or $$\mathbf{A}\mathbf{l} + \mathbf{A}\mathbf{v} = \mathbf{d} \tag{6.15}$$

which can be written as

$$\mathbf{A}\mathbf{v} = \mathbf{f} \tag{6.16}$$

where $$\mathbf{f} = \mathbf{d} - \mathbf{A}\mathbf{l} \tag{6.17}$$

and

n	is the number of measurements or observations,
n_0	is the minimum number of observations required,
$r = n - n_0$	is the number of redundant observations (equal to the number of condition equations,
v	is an (n,1) vector of residuals,
l	is the (n,1) vector of observations,
A	is an (r,n) matrix of coefficients,
f	is an (r,1) vector of numeric terms derived from the observations,
d	is an (r,1) vector of constants. Note that in many least squares problems the vector **d** is zero.

Now if each observation has an *a priori* estimate of its variance then the (n,n) weight matrix of the observations **W** is known and the least squares function φ is

$$\varphi = \text{ the sum of the weighted squared residuals} = \sum_{k=1}^{n} w_k v_k^2$$

In matrix form, the least squares function is expressed as

$$\varphi = \mathbf{v}^T \mathbf{W} \mathbf{v} \tag{6.18}$$

Now φ is the function to be minimised but with the constraints imposed by the condition equations (6.16). This is achieved by adding an (r,1) vector of *Lagrange multipliers* **k** and forming a new function φ'.

$$\varphi' = \mathbf{v}^T\mathbf{W}\mathbf{v} - 2\mathbf{k}^T(\mathbf{A}\mathbf{v} - \mathbf{f}) \tag{6.19}$$

Note that the second term of (6.19) equals zero, since $\mathbf{A}\mathbf{v} - \mathbf{f} = \mathbf{0}$.

Minimising φ' is achieved by differentiating with respect to the unknowns, **v** and **k** and equating these differentials to zero

$$\frac{\partial\varphi'}{\partial\mathbf{k}} = -2\mathbf{v}^T\mathbf{A}^T + 2\mathbf{f}^T = \mathbf{0}^T \tag{6.20}$$

$$\frac{\partial\varphi'}{\partial\mathbf{v}} = 2\mathbf{v}^T\mathbf{W} - 2\mathbf{k}^T\mathbf{A} = \mathbf{0}^T \tag{6.21}$$

Dividing by two, re-arranging and transposing equations (6.20) and (6.21) gives

$$\mathbf{A}\mathbf{v} = \mathbf{f} \tag{6.22}$$

$$\mathbf{W}\mathbf{v} - \mathbf{A}^T\mathbf{k} = \mathbf{0} \tag{6.23}$$

Note that equations (6.22) are the original condition equations and also that $\mathbf{W} = \mathbf{W}^T$ due to symmetry.

From (6.23), the (n,1) vector of residuals **v** is

$$\mathbf{v} = \mathbf{W}^{-1}\mathbf{A}^T\mathbf{k} = \mathbf{Q}\mathbf{A}^T\mathbf{k} \tag{6.24}$$

which, when substituted into (6.22), gives

$$\mathbf{A}\left(\mathbf{Q}\mathbf{A}^T\mathbf{k}\right) = \left(\mathbf{A}\mathbf{Q}\mathbf{A}^T\right)\mathbf{k} = \mathbf{f} \tag{6.25}$$

The matrix $\mathbf{A}\mathbf{Q}\mathbf{A}^T$ is symmetric and of order (r,r) and equations (6.25) are often termed the *normal equations.* The solution of the (r,1) vector of *Lagrange multipliers* **k** is

$$\mathbf{k} = \left(\mathbf{A}\mathbf{Q}\mathbf{A}^T\right)^{-1}\mathbf{f} \tag{6.26}$$

Now the term $\mathbf{A}\mathbf{Q}\mathbf{A}^T$ in equations (6.25) and (6.26) can be "simplified" if an equivalent set of observations $\mathbf{l}_e$ is considered, i.e.,

$$\mathbf{l}_e = \mathbf{A}\mathbf{l} \tag{6.27}$$

Applying the general law of propagation of variances (cofactors) to (6.27) gives

$$\mathbf{Q}_e = \mathbf{A}\mathbf{Q}_{ll}\mathbf{A}^T = \mathbf{A}\mathbf{Q}\mathbf{A}^T \tag{6.28}$$

and

$$\mathbf{W}_e = \mathbf{Q}_e^{-1} = \left(\mathbf{A}\mathbf{Q}\mathbf{A}^T\right)^{-1} \tag{6.29}$$

Substituting (6.29) into (6.26) gives another expression for **k**

$$\mathbf{k} = \mathbf{Q}_e^{-1}\mathbf{f} = \mathbf{W}_e\mathbf{f} \tag{6.30}$$

After computing **k** from either (6.26) or (6.30) the residuals **v** are computed from (6.24) and the vector of adjusted observations $\hat{\mathbf{l}}$ is given by

$$\hat{\mathbf{l}} = \mathbf{l} + \mathbf{v} \tag{6.31}$$

This is the standard matrix solution for least squares adjustment of observations only.

6.4. Propagation of Variances for Least Squares Adjustment of Observations Only

In this technique of least squares adjustment, the condition equations in matrix form are

$$\mathbf{A}\mathbf{v} = \mathbf{f} \tag{6.32}$$

with

$$\mathbf{f} = \mathbf{d} - \mathbf{A}\mathbf{l} \tag{6.33}$$

Similarly to Chapter 5, equation (6.33) can be expressed in a form similar to equation (3.23) and the general law of propagation of variances applied to give the cofactor matrix of the numeric terms **f**.

$$\mathbf{f} = -\mathbf{A}\mathbf{l} + \mathbf{d}$$

and

$$\mathbf{Q}_{ff} = (-\mathbf{A})\mathbf{Q}(-\mathbf{A})^T = \mathbf{A}\mathbf{Q}\mathbf{A}^T = \mathbf{Q}_e \tag{6.34}$$

Thus the cofactor matrix of ***f*** is also the cofactor matrix of an equivalent set of observations.

The solution "steps" in the least squares adjustment of observations only are set out above and restated as

$$\begin{aligned}\mathbf{Q}_e &= \mathbf{A}\mathbf{Q}\mathbf{A}^T \\ \mathbf{W}_e &= \mathbf{Q}_e^{-1} \\ \mathbf{k} &= \mathbf{W}_e\mathbf{f} \\ \mathbf{v} &= \mathbf{Q}\mathbf{A}^T\mathbf{k} \\ \hat{\mathbf{l}} &= \mathbf{l}+\mathbf{v}\end{aligned}$$

Applying the law of propagation of variances (remembering that cofactor and weight matrices are symmetric) gives the following cofactor matrices

$$\mathbf{Q}_{kk} = (\mathbf{W}_e)\mathbf{Q}_{ff}(\mathbf{W}_e)^T = \mathbf{W}_e \tag{6.35}$$

$$\mathbf{Q}_{vv} = (\mathbf{Q}\mathbf{A}^T)\mathbf{Q}_{kk}(\mathbf{Q}\mathbf{A}^T)^T = \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} \tag{6.36}$$

and

$$\begin{aligned}\hat{\mathbf{l}} &= \mathbf{l}+\mathbf{v} \\ &= \mathbf{l}+\mathbf{Q}\mathbf{A}^T\mathbf{k} \\ &= \mathbf{l}+\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{f} \\ &= \mathbf{l}+\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\left(\mathbf{d}-\mathbf{A}\mathbf{l}\right)\end{aligned}$$

from which follows

$$\hat{\mathbf{l}} = \left(\mathbf{I}-\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\right)\mathbf{l}+\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{d} \tag{6.37}$$

Applying the law of propagation of variances to (6.37) gives

$$\mathbf{Q}_{\hat{l}\hat{l}} = \left(\mathbf{I}-\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\right)\mathbf{Q}\left(\mathbf{I}-\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\right)^T$$

which reduces to

$$\mathbf{Q}_{\hat{l}\hat{l}} = \mathbf{Q}-\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} = \mathbf{Q}-\mathbf{Q}_{vv} \tag{6.38}$$

Variance-covariance matrices for **k**, **v** and $\hat{\mathbf{l}}$ are obtained by multiplying the cofactor matrix by the variance factor σ_0^2 - see equation (2.32).

The *a priori* estimate of the variance factor may be computed from

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T\mathbf{W}\mathbf{v}}{r} \tag{6.39}$$

where

$\mathbf{v}^T\mathbf{W}\mathbf{v}$ is the quadratic form, and

r is the degrees of freedom.

A derivation of equation (6.39) is given in Chapter 5. The quadratic form $\mathbf{v}^T\mathbf{W}\mathbf{v}$ may be computed in the following manner.

Remembering, for the method of observations only, the following matrix equations

$$\begin{aligned}\mathbf{Q}_e &= \mathbf{A}\mathbf{Q}\mathbf{A}^T\\ \mathbf{W}_e &= \mathbf{Q}_e^{-1}\\ \mathbf{k} &= \mathbf{W}_e\mathbf{f}\\ \mathbf{v} &= \mathbf{Q}\mathbf{A}^T\mathbf{k}\end{aligned}$$

then

$$\begin{aligned}\mathbf{v}^T\mathbf{W}\mathbf{v} &= \left(\mathbf{Q}\mathbf{A}^T\mathbf{k}\right)^T\mathbf{W}\left(\mathbf{Q}\mathbf{A}^T\mathbf{k}\right)\\ &= \mathbf{k}^T\mathbf{A}\mathbf{Q}\mathbf{W}\mathbf{Q}\mathbf{A}^T\mathbf{k}\\ &= \mathbf{k}^T\mathbf{A}\mathbf{Q}\mathbf{A}^T\mathbf{k}\\ &= \mathbf{k}^T\mathbf{Q}_e\mathbf{k}\\ &= \mathbf{k}^T\mathbf{W}_e^{-1}\mathbf{k}\\ &= \mathbf{f}^T\mathbf{W}_e\mathbf{W}_e^{-1}\mathbf{k}\end{aligned}$$

and

$$\mathbf{v}^T\mathbf{W}\mathbf{v} = \mathbf{f}^T\mathbf{k} \tag{6.40}$$

6.5. Adjustment of a Single Closed Traverse using the method of Least Squares Adjustment of Observations Only

The basic component of many surveys is a traverse whose bearings have been determined by theodolite or total station observations and distances measured by EDM. If careful observations are made with well maintained equipment, the measurements are usually free of systematic errors and mistakes and the surveyor is left with small random errors which, in the case of a closed traverse, reveal themselves as angular and linear misclosures. If the misclosures are within acceptable limits, it is standard practice to remove the misclosures by adjusting the original observations to make the traverse a mathematically correct figure. In this section, only single closed traverses are considered and such traverses may begin and end at different fixed points or close back on the starting point. Traverse networks, consisting of two or more single traverses with common junction points, are not considered here; such networks are usually adjusted by a method commonly known as *Variation of Coordinates,* based on Least Squares Adjustment of Indirect Observations.

6.5.1. Some single traverse adjustment methods and their deficiencies

A traverse adjustment method should be based on sound mathematical principles related to the measurement techniques with due allowance made for independence (or dependence) of those measurements and also allow for differing measurement precisions.

Bowditch's Rule and the *Transit Rule,* both of which adjust lengths and bearings of traverse lines and *Crandall's method,* which adjusts the lengths only of the traverse lines, are three popular adjustment methods that fail to meet the general guidelines above. Although Crandall's method, which is explained in detail in later sections, does have mathematical rigour if it assumed that the bearings of a traverse close and require no further adjustment.

Bowditch's Rule and the Transit Rule for adjusting single traverses are explained below by applying the rules to adjust a four-sided polygon having an unusually large misclose. The polygon, shown in Figure 6.3, does not reflect the usual misclosures associated with traverses using modern surveying equipment.

Bowditch's Rule

Nathaniel Bowditch (1773-1838) was an American mathematician and astronomer (see citation below). In 1808, in response to a prize offered by a correspondent in *The Analyst*[2], Bowditch put forward a method of adjusting the misclose in a chain and compass survey (bearings measured by magnetic compass and distances measured by surveyor's chain). His method of adjustment was simple and became widely used. It is still used today for the adjustment of a figure prior to the computation of the area, where the area-formula assumes a closed mathematical figure.

Prior to the advent of programmable calculators and computers, Bowditch's Rule was often used to adjust traverses that did not close due to the effects of random errors in the measurement of bearings and distances. Its use was justified entirely by its simplicity and whilst it had theoretical rigour – if the bearings of traverse lines were independent of each other, as they are in compass surveys – it is incompatible with modern traversing techniques. Bowditch's rule cannot take into account different measurement precisions of individual traverse lines nor can it accommodate complicated networks of connecting traverses. Nevertheless, due to its long history of use in the surveying profession, its simplicity and its practical use in the computation of areas of figures that misclose, Bowditch's Rule is still prominent in surveying textbooks and is a useful adjustment technique.

Bowditch, Nathaniel (b. March 26, 1773, Salem, Mass., U.S. – d. March 16, 1838, Boston, Mass., U.S.), self-educated American mathematician and astronomer, author of the best book on navigation of his time, and discoverer of the **Bowditch** curves, which have important applications in astronomy and physics. Between 1795 and 1799 **Bowditch** made four lengthy sea voyages, and in 1802 he was put in command of a merchant vessel. Throughout that period he pursued his interest in mathematics. After investigating the accuracy of *The Practical Navigator,* a work by the Englishman J.H. Moore, he produced a revised edition in 1799. His additions became so numerous that in 1802 he published *The New American Practical Navigator,* based on Moore's book, which was adopted by the U.S. Department of the Navy and went through some 60 editions. **Bowditch** also wrote many scientific papers, one of which, on the motion of a pendulum swinging simultaneously about two axes at right angles, described the so-called **Bowditch** curves (better known as the Lissajous figures, after the man who later studied them in detail). **Bowditch** translated from the French and updated the first four volumes of Pierre-Simon Laplace's monumental work on the gravitation of heavenly bodies, *Traité de mécanique céleste,* more than doubling its size with his own commentaries. The resulting work, *Celestial Mechanics,* was published in four volumes in 1829-39. **Bowditch** refused professorships at several universities. He was president (1804-23) of the Essex Fire and Marine Insurance Company of Salem and worked as an actuary (1823-38) for the Massachusetts Hospital Life Insurance Company of Boston. From 1829 until his death, he was president of the American Academy of Arts and Sciences.

[2] *The Analyst or Mathematical Museum* was a journal of theoretical and applied mathematics. In Vol. I, No. II, 1808, Robert Patterson of Philadelphia posed a question on the adjustment of a traverse and offered a prize of $10 for a solution; the editor Dr Adrian appointed as the judge of submissions. Bowditch's solution was published in Vol. I, No. IV, 1808, pp. 88-93 (Stoughton, H.W., 1974. 'The first method to adjust a traverse based on statistical considerations', *Surveying and Mapping,* June 1974, pp. 145-49).

Bowditch's adjustment can best be explained by considering the case of plotting a figure (using a protractor and scale ruler) given the bearing and distances of the sides.

Consider Figure 6.3, a plot that does not close, of a four-sided figure *ABCD*. The solid lines *AB, BC, CD* and *DE* are the result of marking point *A,* plotting the bearing *AB* and then scaling the distance *AB* to fix *B*. Then, from point *B,* plotting the bearing and distance *BC* to fix *C,* then from *C,* plotting the bearing and distance *CD* to fix *D* and finally from *D,* plotting the bearing and distance *DA*. However, due to plotting errors, the final line does not meet the starting point, but instead finishes at *E*. The distance *EA* is the linear misclose *d,* due to plotting errors, i.e., errors in protracting bearings and scaling distances.

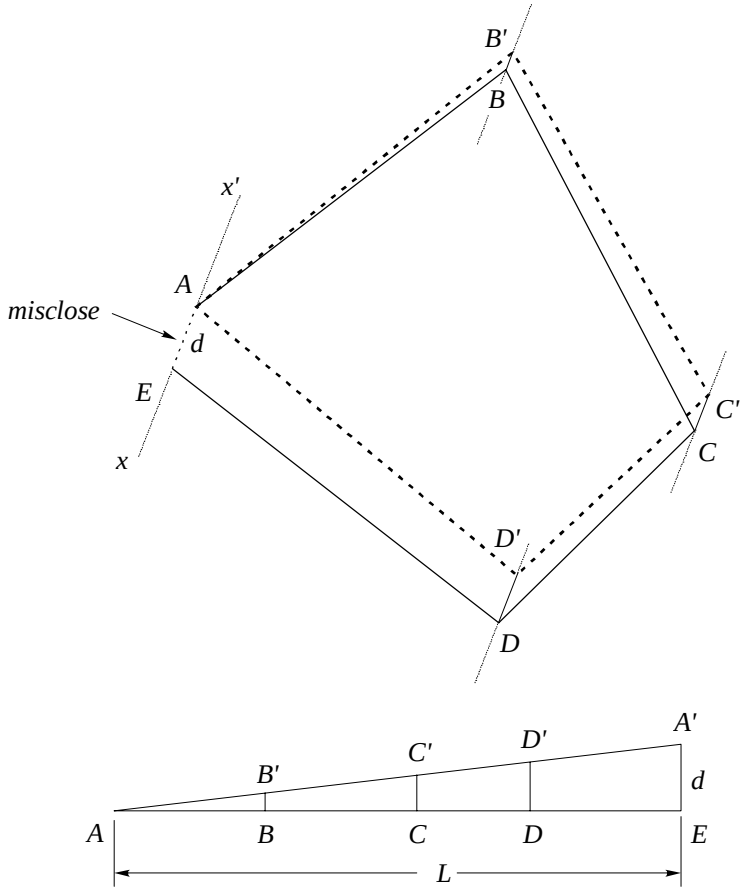


Figure 6.3 Graphical plot of polygon *ABCD* with misclose *d*

To adjust the figure *ABCDE* to remove the misclose *d* the following procedure can be used.

1. Draw lines parallel to the line *xx'* (the misclose bearing) through points *B, C* and *D.*
2. Draw a right-angled triangle *AEA*'. The base of the triangle is *L,* equal to the sum of the lengths of the sides and the height is the linear misclose *d.*
3. Along the base of the triangle, mark in proportion to the total length *L*, the distances *AB, BC* and *CD.* These will be the points *B, C* and *D.*
4. Draw vertical lines from *B, C* and *D* intersecting the hypotenuse of the triangle at *B', C'* and *D'.* These distances are then marked off along the parallel lines of the main figure.
5. The adjusted figure is *AB'C'D'A.*

This adjustment is a graphical demonstration of Bowditch's Rule; i.e., the linear misclose *d* is apportioned to individual sides in the ratio of the length of the side to the total length of all the sides in the direction of the misclose bearing.

Bowditch's Rule as it is normally applied to the adjustment of traverses can be deduced by again considering Figure 6.3. The linear misclose *d* has easting and northing components ΔE_m and ΔN_m, the subscript *m* referring to the misclose. The distances *BB', CC'* and *DD'* each have easting and northing components, say dE_B, dN_B, dE_C, dN_C and dE_D, dN_D, the east misclose $dE_m = dE_B + dE_C + dE_D$ and the north misclose $dN_m = dN_B + dN_C + dN_D$.

Thus, we may express Bowditch's Rule for calculating adjustments dE_k, dN_k to individual easting and northing components $\Delta E_k, \Delta N_k$ of line *k* of a traverse whose total length is *L* as

$$\begin{aligned} dE_k &= dist_k \left(\frac{dE_m}{L} \right) \\ dN_k &= dist_k \left(\frac{dN_m}{L} \right) \end{aligned} \tag{6.41}$$

As an example of a Bowditch adjustment, Table 6.1 shows the bearings and distances of the polygon in Figure 6.3.

The linear misclose, which is quite large, is $d = \sqrt{(-3.173)^2 + (-8.181)^2} = 8.775$ and the length *L,* equal to the sum of the four sides, is $L = 51.53 + 53.86 + 36.31 + 54.71 = 196.41$

The corrections to the easting and northing components of the line *CD* are

$$dE = 36.31 \times \frac{3.173}{196.41} = 0.587$$

$$dN = 36.31 \times \frac{8.181}{196.41} = 1.512$$

Note: (i) Easting and northing misclosures dE_m and dN_m used in equations (6.41) have opposite signs to the misclosures in the tabulation,

(ii) The sums of the corrections are equal and of opposite sign to the misclosures and

(iii) The sums of the adjusted easting and northing components are zero.

			components		*corrections*		*adjusted components*	
Line	*Bearing*	*Dist*	ΔE	ΔN	dN	dN	ΔE	ΔN
AB	52º 31'	51.53	40.891	31.358	0.832	2.146	41.723	33.504
BC	152º 21'	53.86	24.995	-47.709	0.870	2.243	25.865	-45.466
CD	225º 30'	36.31	-25.898	-25.450	0.587	1.512	-25.311	-23.938
DA	307º 55'	54.71	-43.161	33.620	0.884	2.280	-42.277	35.900
		misclose	-3.173	-8.181	3.173	8.181	0.000	0.000

Table 6.1. Bowditch Rule adjustment of polygon *ABCD*

Transit Rule

The Transit Rule has no theoretical basis related to surveying instruments or measuring techniques. Its only justification is its mathematical simplicity, which is no longer a valid argument for the method in this day of pocket computers. The Transit Rule for calculating adjustments dE_k, dN_k to individual easting and northing components $\Delta E_k, \Delta N_k$ of line *k* of a traverse whose east and north misclosures are dE_m and dN_m is

$$dE_k = |\Delta E_k| \left(\frac{dE_m}{\sum_{j=1}^{n} |\Delta E_j|} \right) \qquad dN_k = |\Delta N_k| \left(\frac{dN_m}{\sum_{j=1}^{n} |\Delta N_j|} \right) \tag{6.42}$$

$|\Delta E_k|$ is the absolute value of the east component of the k^{th} traverse leg and $\sum_{j=1}^{n}|\Delta E_j|$ is the sum of the absolute values of the east components of the traverse legs and similarly for $|\Delta N_k|$ and $\sum_{j=1}^{n}|\Delta N_j|$.

As an example of a Transit Rule adjustment, Table 6.2 shows the bearings and distances of the polygon in Figure 6.3. The east and north misclosures are $dE_m = 3.173$ and $dN_m = 8.181$, and the sums of the absolute values of the east and north components of the traverse legs are $\sum_{j=1}^{n}|\Delta E_j| = 134.945$ and $\sum_{j=1}^{n}|\Delta N_j| = 138.137$

The corrections to the easting and northing components of the line *CD* are

$$dE = 25.898 \times \frac{3.173}{134.945} = 0.587$$
$$dN = 25.450 \times \frac{8.181}{138.137} = 1.512$$

Note: (i) Easting and northing misclosures dE_m and dN_m used in equations (6.42) have opposite signs to the misclosures in the tabulation,

(ii) The sums of the corrections are equal and of opposite sign to the misclosures and

(iii) The sums of the adjusted easting and northing components are zero.

			components		*corrections*		*adjusted components*	
Line	*Bearing*	*Dist*	ΔE	ΔN	*dN*	*dN*	ΔE	ΔN
AB	52º 31'	51.53	40.891	31.358	0.961	1.857	41.852	33.215
BC	152º 21'	53.86	24.995	-47.709	0.588	2.826	25.583	-44.883
CD	225º 30'	36.31	-25.898	-25.450	0.609	1.507	-25.289	-23.943
DA	307º 55'	54.71	-43.161	33.620	1.015	1.991	-42.146	35.611
		misclose	-3.173	-8.181	3.173	8.181	0.000	0.000

Table 6.2 Transit Rule adjustment of polygon *ABCD*

6.5.2. Crandall's method. A semi-rigorous single traverse adjustment method

Suppose that the angles of a traverse – either beginning and ending at the same point or between two known points with starting and closing known bearings – have been adjusted so that the traverse has a perfect angular closure and the resulting bearings are considered as correct, or adjusted. We call this a closed traverse. A mathematical closure, using the adjusted bearings and measured distances, will in all probability, reveal a linear misclose, i.e., the sums of the east and north components of the traverse legs will differ from zero (in the case of a traverse beginning and ending at the same point) or certain known values (in the case of a traverse between known points). *Crandall's method,* which employs the least squares principle, can be used to compute corrections to the measured distances to make the traverse close mathematically. The method was first set out in the textbook *Geodesy and Least Squares* by Charles L. Crandall, Professor of Railroad Engineering and Geodesy, Cornell University, Ithaca, New York, U.S.A. and published by John Wiley & Sons, New York, 1906.

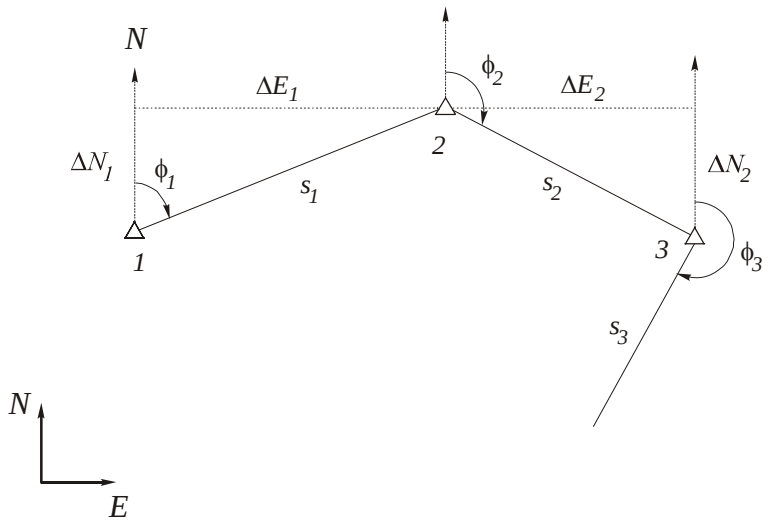


Figure 6.4 Schematic traverse diagram

Figure 6.4 shows a schematic diagram of a traverse of $k = 1, 2, \ldots, n$ legs where ϕ_k, s_k are the adjusted bearing and measured distance respectively of the k^{th} leg. The east and north

components of each traverse leg are $\Delta E_k = s_k \sin\phi_k$ and $\Delta N_k = s_k \cos\phi_k$ respectively. If the adjusted distance of the k^{th} traverse leg is $(s_k + v_k)$ where v_k is the residual (a small unknown correction) then the two conditions that must be fulfilled by the adjusted bearings and adjusted distances in a closed traverse are

$$\begin{aligned}(s_1 + v_1)\sin\phi_1 + (s_2 + v_2)\sin\phi_2 + \cdots + (s_n + v_n)\sin\phi_n &= D_E \\ (s_1 + v_1)\cos\phi_1 + (s_2 + v_2)\cos\phi_2 + \cdots + (s_n + v_n)\cos\phi_n &= D_N\end{aligned} \tag{6.43}$$

where $D_E = E_{END} - E_{START}$ and $D_N = N_{END} - N_{START}$ are the east and north coordinate differences respectively between the terminal points of the traverse. Note that in a traverse beginning and ending at the same point D_E and D_N will both be zero.

Expanding equation (6.43) gives

$$\begin{aligned}v_1\sin\phi_1 + v_2\sin\phi_2 + \cdots + v_n\sin\phi_n + S_E &= D_E \\ v_1\cos\phi_1 + v_2\cos\phi_2 + \cdots + v_n\cos\phi_n + S_N &= D_N\end{aligned} \tag{6.44}$$

where

$$\begin{aligned}S_E &= s_1\sin\phi_1 + s_2\sin\phi_2 + \cdots + s_n\sin\phi_n &&= \sum_{k=1}^{n}\Delta E_k \\ S_N &= s_1\cos\phi_1 + s_2\cos\phi_2 + \cdots + s_n\cos\phi_n &&= \sum_{k=1}^{n}\Delta N_k\end{aligned} \tag{6.45}$$

S_E, S_N are the sums of the east and north components, $\Delta E_k, \Delta N_k$ respectively, of the $k = 1, 2, \ldots, n$ traverse legs.

Equations (6.44) can be expressed in matrix form as

$$\begin{bmatrix}\sin\phi_1 & \sin\phi_2 & \sin\phi_3 & \cdots & \sin\phi_n \\ \cos\phi_1 & \cos\phi_2 & \cos\phi_3 & \cdots & \cos\phi_n\end{bmatrix}\begin{bmatrix}v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n\end{bmatrix} = \begin{bmatrix}D_E - S_E \\ D_N - S_N\end{bmatrix} \tag{6.46}$$

or $$\mathbf{Av} = \mathbf{f}$$

The solution for the vector of residuals **v** is given by equations (6.24) and (6.26) re-stated again as

$$\begin{aligned}\mathbf{v} &= \mathbf{W}^{-1}\mathbf{A}^T\mathbf{k} = \mathbf{Q}\mathbf{A}^T\mathbf{k} \\ \mathbf{k} &= \left(\mathbf{A}\mathbf{Q}\mathbf{A}^T\right)^{-1}\mathbf{f}\end{aligned} \tag{6.47}$$

where **k** is the vector of *Lagrange multipliers,* $\mathbf{Q} = \mathbf{W}^{-1}$ is the cofactor matrix and **W** is the weight matrix, **A** is a coefficient matrix containing sines and cosines of traverse bearings and **f** is a vector containing the negative sums of the east and north components of the traverse legs.

In Crandall's method, weights are considered as inversely proportional to the measured distances and the measured distances are considered to be independent. Hence the weight matrix **W** is diagonal

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & & \vdots \\ \vdots & & & \ddots & \\ 0 & \cdots & & & w_n \end{bmatrix} = \begin{bmatrix} 1/s_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/s_2 & 0 & \cdots & 0 \\ 0 & 0 & 1/s_3 & & \vdots \\ \vdots & & & \ddots & \\ 0 & \cdots & & & 1/s_n \end{bmatrix}$$

and since $\mathbf{Q} = \mathbf{W}^{-1}$ then

$$\mathbf{Q} = \begin{bmatrix} s_1 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & & \vdots \\ \vdots & & & \ddots & \\ 0 & \cdots & & & s_n \end{bmatrix}$$

and

$$\mathbf{Q}\mathbf{A}^T = \begin{bmatrix} s_1 & 0 & 0 & \cdots & 0 \\ 0 & s_2 & 0 & \cdots & 0 \\ 0 & 0 & s_3 & & \vdots \\ \vdots & & & \ddots & \\ 0 & \cdots & & & s_n \end{bmatrix} \begin{bmatrix} \sin\phi_1 & \cos\phi_1 \\ \sin\phi_2 & \cos\phi_2 \\ \sin\phi_3 & \cos\phi_3 \\ \vdots & \vdots \\ \sin\phi_n & \cos\phi_n \end{bmatrix} = \begin{bmatrix} \Delta E_1 & \Delta N_1 \\ \Delta E_2 & \Delta N_2 \\ \Delta E_3 & \Delta N_3 \\ \vdots & \vdots \\ \Delta E_n & \Delta N_n \end{bmatrix} \tag{6.48}$$

$$\mathbf{A}\mathbf{Q}\mathbf{A}^T = \begin{bmatrix} \sum_{k=1}^{n} \sin\phi_k \Delta E_k & \sum_{k=1}^{n} \sin\phi_k \Delta N_k \\ \sum_{k=1}^{n} \cos\phi_k \Delta E_k & \sum_{k=1}^{n} \cos\phi_k \Delta N_k \end{bmatrix}$$

Now, since $\sin\phi_k = \dfrac{\Delta E_k}{s_k}$ and $\cos\phi_k = \dfrac{\Delta N_k}{s_k}$ then $\mathbf{AQA}^T$ can be written as

$$\mathbf{AQA}^T = \begin{bmatrix} \sum_{k=1}^{n} \dfrac{(\Delta E_k)^2}{s_k} & \sum_{k=1}^{n} \dfrac{\Delta E_k \Delta N_k}{s_k} \\ \sum_{k=1}^{n} \dfrac{\Delta E_k \Delta N_k}{s_k} & \sum_{k=1}^{n} \dfrac{(\Delta N_k)^2}{s_k} \end{bmatrix} = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \tag{6.49}$$

and $$\left(\mathbf{AQA}^T\right)^{-1} = \frac{1}{ab - c^2}\begin{bmatrix} b & -c \\ -c & a \end{bmatrix}$$

giving the *Lagrange multipliers* from equations (6.47) as

$$\begin{aligned} k_1 &= \frac{b(D_E - S_E) - c(D_N - S_N)}{ab - c^2} \\ k_2 &= \frac{a(D_N - S_N) - c(D_E - S_E)}{ab - c^2} \end{aligned} \tag{6.50}$$

The residuals **v** (corrections to the measured distances) are given as

$$\mathbf{v} = \mathbf{QA}^T\mathbf{k} = \begin{bmatrix} \Delta E_1 & \Delta N_1 \\ \Delta E_2 & \Delta N_2 \\ \Delta E_3 & \Delta N_3 \\ \vdots & \vdots \\ \Delta E_n & \Delta N_n \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} k_1\Delta E_1 + k_2\Delta N_1 \\ k_1\Delta E_2 + k_2\Delta N_2 \\ k_1\Delta E_3 + k_2\Delta N_3 \\ \vdots \\ k_1\Delta E_n + k_2\Delta N_n \end{bmatrix} \tag{6.51}$$

6.5.3. Example of Crandall's method

Figure 6.5 shows a closed traverse between stations *A, B, C, D* and *E*. The linear misclose (bearing and distance) of the traverse is 222º 57' 31" *0.2340* and the components of the misclose are -0.1594 m east and -0.1712 m north. It is required to adjust the distances using Crandall's method.

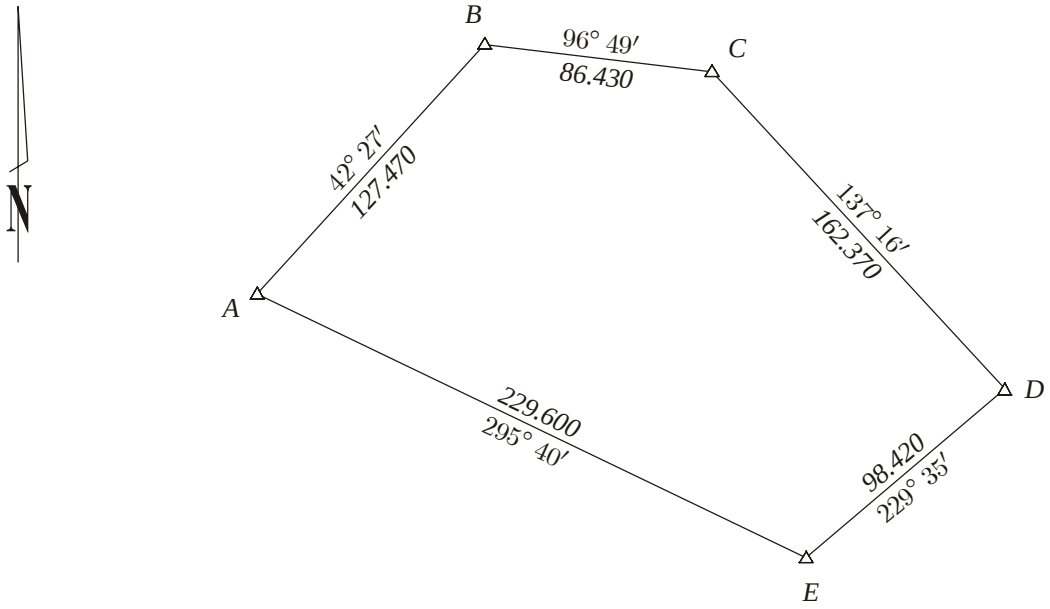


Figure 6.5 Closed traverse between stations *ABCDE*

The adjusted bearings and measured distances and the traverse leg components are shown in Table 6.3 below. S_E and S_N are the summations of east and north components and since this traverse begins and ends at the same point then D_E and D_N will both be zero.

Line	Bearing	Distance	traverse leg components	
k	ϕ_k	s_k	ΔE_k	ΔN_k
1	42º 27′	127.470	86.035437	94.055858
2	96º 49′	86.430	85.819028	-10.258619
3	137º 16′	162.370	110.182189	-119.264002
4	229º 35′	98.420	-74.932042	-63.809760
5	295º 40′	229.600	-206.945175	99.447747
			$S_E = 0.159438$	$S_N = 0.171224$

Table 6.3 Traverse components and sums

Table 6.4 shows the functions of the components for each line and their summations.

Line	$\frac{(\Delta E_k)^2}{s_k}$	$\frac{(\Delta N_k)^2}{s_k}$	$\frac{\Delta E_k \Delta N_k}{s_k}$
1	58.069322	69.400678	63.482677
2	85.212376	1.217624	-10.186101
3	74.768213	87.601787	-80.931014
4	57.049491	41.370509	48.581545
5	186.525721	43.074279	-89.635154
	$a = 461.625124$	$b = 242.664876$	$c =$ -68.688048

Table 6.4 Functions of traverse components

The Lagrange multipliers k_1 and k_2 are computed from equations (6.50) using *a, b, c* from Table 6.4, S_E and S_N from Table 6.3, and since this traverse begins and ends at the same point then D_E and D_N will both be zero.

$$a = \sum_{k=1}^{5} \frac{(\Delta E_k)^2}{s_k} = 461.625124 \qquad D_E - S_E = -0.159438 \qquad k_1 = -4.7018E-04$$

$$b = \sum_{k=1}^{5} \frac{(\Delta N_k)^2}{s_k} = 242.664876 \qquad D_N - S_N = -0.171224 \qquad k_2 = -8.3868E-04$$

$$c = \sum_{k=1}^{5} \frac{\Delta E_k \Delta N_k}{s_k} = -68.688048$$

Table 6.5 shows the original traverse data, the residuals and adjusted traverse distances.

Line	Bearing	Distance	Traverse leg components		Residual	Adjusted Distance
k	ϕ_k	s_k	ΔE_k	ΔN_k	$v_k = k_1 \Delta E_k + k_2 \Delta N_k$	
1	42º 27′	127.470	86.035437	94.055858	-0.119	127.351
2	96º 49′	86.430	85.819028	-10.258619	-0.032	86.398
3	137º 16′	162.370	110.182189	-119.264002	0.048	162.418
4	229º 35′	98.420	-74.932042	-63.809760	0.089	98.509
5	295º 40′	229.600	-206.945175	99.447747	0.014	229.614

Table 6.5 Adjusted traverse distances: Crandall's method

6.5.4. A rigorous single traverse adjustment method

A traverse is a combination of two basic survey measurements, *distances* and *directions*. Ignoring the physical fact that the same measuring equipment is likely to be used on each leg of the traverse, distances and directions are independently determined quantities. *Bearings* ϕ, *angles* α and *coordinates E,N* are derived quantities and in general, cannot be considered as mathematically (or statistically) independent. Restricting the adjustment method to single traverses, means angles at traverse points, derived from directions at those points, can be considered as mathematically independent quantities.

Three conditions, expressing the mathematical relationship between traverse measurements and derived coordinates, may be deduced from Figure 6.6 below, in which P_1 and P_n are "fixed stations" whose east and north coordinates are known and $P_2, P_3, P_4, \ldots, P_{n-1}$ are "floating stations" whose coordinates are to be determined from the traverse angles α and distances *s*. The starting bearing ϕ_0 and the finishing bearing ϕ_n are known.

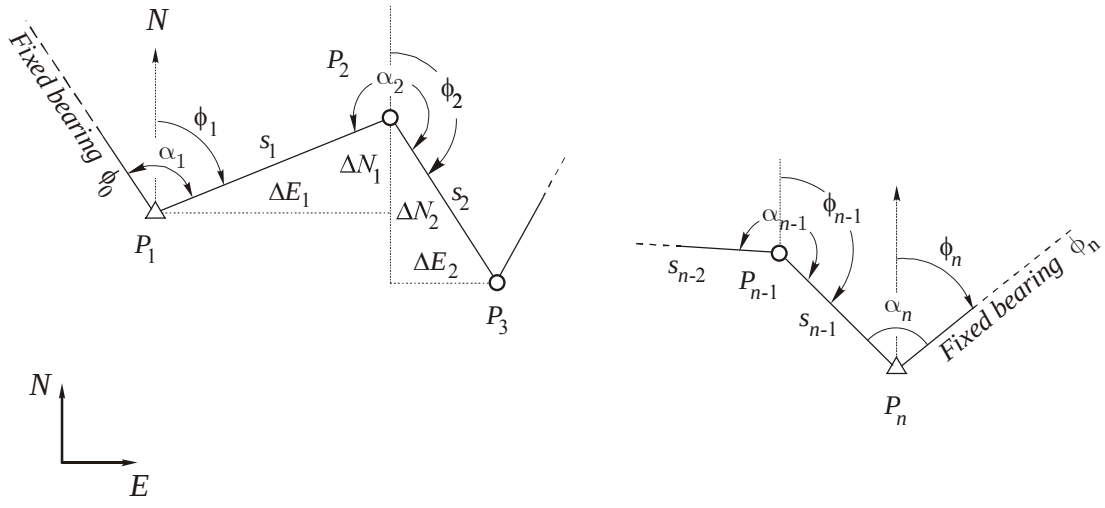


Figure 6.6 A closed traverse between two fixed stations

These three conditions are:

(i) The starting bearing ϕ_0 plus all the measured angles should equal the known finishing bearing ϕ_n,

(ii) The starting east coordinate plus all the east components of the traverse legs should equal the known east coordinate at the end point and

(iii) The starting north coordinate plus all the north components of the traverse legs should equal the known north coordinate at the end point.

These conditions apply to all single traverses whether they start and end at different fixed points or close back on the starting point and can be expressed mathematically as

$$\begin{aligned} \phi_0 + \alpha_1 + \alpha_2 + \alpha_3 + \cdots + a_n &= \phi_n \\ E_1 + \Delta E_1 + \Delta E_2 + \Delta E_3 + \cdots + \Delta E_{n-1} &= E_n \\ N_1 + \Delta N_1 + \Delta N_2 + \Delta N_3 + \cdots + \Delta N_{n-1} &= N_n \end{aligned} \tag{6.52}$$

Equations (6.52) are relationships between adjusted quantities $\alpha_1, \alpha_2, \ldots, \alpha_n$ or functions of adjusted quantities $\Delta E_1, \Delta E_2, \ldots, \Delta E_{n-1}$ and $\Delta N_1, \Delta N_2, \ldots, \Delta N_{n-1}$.

Traverses will generally misclose due to the small random errors in the angles (derived from the measured directions) and the measured distances. To make the traverse mathematically correct, small corrections must be applied to the measurements to give adjusted quantities. These adjusted quantities are:

$$s = s' + v_s$$

$$\alpha = \alpha' + v_\alpha$$

where s and α are adjusted distance and angle respectively, s' and α' are the measured angle and distance, and v_s and v_α are small corrections. Replacing the adjusted quantities with measurements and corrections allows the first member of equations (6.52) to be expressed as

$$\phi_0 + \left(\alpha_1 + v_{\alpha_1}\right) + \left(\alpha_2 + v_{\alpha_2}\right) + \left(\alpha_3 + v_{\alpha_3}\right) + \cdots + \left(\alpha_n + v_{\alpha_n}\right) = \phi_n$$

and summing the measured angles and rearranging gives a simple expression for the summation of corrections to measured angles as

$$\boxed{v_{\alpha_1} + v_{\alpha_2} + v_{\alpha_3} + \cdots + v_{\alpha_n} = f_1} \tag{6.53}$$

where, apart from some multiple of 180º

$$\boxed{f_1 = \phi_n - \left(\phi_0 + \sum_{k=1}^{n} \alpha'_k \right) = \phi_n - \phi'_n} \tag{6.54}$$

f_1 is the angular misclose in the traverse and equation (6.54) simply states that the sum of the corrections to the measured angles is equal to the angular misclose.

The second and third members of equations (6.52) can also be expressed as a summation of corrections by considering the following

$$\Delta E = s \sin\phi \quad \text{and} \quad \Delta N = s \cos\phi$$

where $\Delta E, \Delta N$ are east and north components of a traverse leg and

$$s = s' + v_s \quad \text{and} \quad \phi = \phi' + v_\phi$$

where ϕ' and v_ϕ are "measured" bearing and correction respectively, hence we express the east and north components as

$$\Delta E = \left(s' + v_s\right) \sin\left(\phi' + v_\phi\right)$$
$$\Delta N = \left(s' + v_s\right) \cos\left(\phi' + v_\phi\right)$$

Using the trigonometric expansions for $\sin\left(A + B\right)$ and $\cos\left(A + B\right)$, and the approximations $\sin v_\phi \simeq v_\phi$ and $\cos v_\phi \simeq 1$ since v_ϕ is a <u>small</u> quantity gives

$$\Delta E = \left(s' + v_s\right)\left\{\sin\phi' \cos v_\phi + \sin v_\phi \cos\phi'\right\} = s' \sin\phi' + s' v_\phi \cos\phi' + v_s \sin\phi' + v_s v_\phi \cos\phi'$$
$$\Delta N = \left(s' + v_s\right)\left\{\cos\phi' \cos v_\phi - \sin\phi' \sin v_\phi\right\} = s' \cos\phi' - s' v_\phi \sin\phi' + v_s \cos\phi' - v_s v_\phi \sin\phi'$$

and since v_s and v_ϕ are <u>both small</u> then their product $v_s v_\phi \simeq 0$, hence

$$\Delta E = s' \sin\phi' + v_\phi s' \cos\phi' + v_s \sin\phi'$$
$$\Delta N = s' \cos\phi' - v_\phi s' \sin\phi + v_s \cos\phi'$$

Finally, the east and north components of a traverse leg computed using the measured quantities are $\Delta E' = s' \sin\phi'$ and $\Delta N' = s' \cos\phi'$, and we may write

$$\begin{aligned}\Delta E &= \Delta E' + v_\phi \Delta N' + v_s \sin\phi' \\ \Delta N &= \Delta N' - v_\phi \Delta E' + v_s \cos\phi'\end{aligned} \tag{6.55}$$

Substituting equations (6.55) into the second and third members of equations (6.52) gives

$$\begin{aligned}
E_1 \; &+\left(\Delta E_1{}' + v_{\phi_1}\Delta N_1' + v_{s_1}\sin\phi_1'\right) \\
&+\left(\Delta E_2{}' + v_{\phi_2}\Delta N_2' + v_{s_2}\sin\phi_2'\right) \\
&+\cdots \\
&+\left(\Delta E_{n-1}' + v_{\phi_{n-1}}\Delta N_{n-1}' + v_{s_{n-1}}\sin\phi_{n-1}'\right) = E_n \\
N_1 \; &+\left(\Delta N_1{}' - v_{\phi_1}\Delta E_1' + v_{s_1}\cos\phi_1'\right) \\
&+\left(\Delta N_2' - v_{\phi_2}\Delta E_2' + v_{s_2}\cos\phi_2'\right) \\
&+\cdots \\
&+\left(\Delta N_{n-1}' - v_{\phi_{n-1}}\Delta E_{n-1}' + v_{s_{n-1}}\cos\phi_{n-1}'\right) = N_n
\end{aligned}$$

Letting the misclose in the east and north coordinates be

$$\boxed{\begin{aligned} f_2 &= E_n - \left\{E_1 + \sum_{k=1}^{n-1}\Delta E_k'\right\} = E_n - E_n' \\ f_3 &= N_n - \left\{N_1 + \sum_{k=1}^{n-1}\Delta N_k'\right\} = N_n - N_n' \end{aligned}} \tag{6.56}$$

and recognising that $v_{\phi_1} = v_{\alpha_1}$, $v_{\phi_2} = v_{\alpha_1} + v_{\alpha_2}$, $v_{\phi_3} = v_{\alpha_1} + v_{\alpha_2} + v_{\alpha_3}$ etc, and $v_{\phi_{n-1}} = \sum_{k=1}^{n-1} v_{\alpha_k}$ then we may write

$$\begin{aligned}
&v_{\alpha_1}\Delta N_1' + v_{s_1}\sin\phi_1' + \left(v_{\alpha_1} + v_{\alpha_2}\right)\Delta N_2' + v_{s_2}\sin\phi_2' + \left(v_{\alpha_1} + v_{\alpha_2} + v_{\alpha_3}\right)\Delta N_3' + v_{s_3}\sin\phi_3' + \cdots \\
&\cdots + \left(\sum_{k-1}^{n-1} v_{\alpha_k}\right)\Delta N_{n-1}' + v_{s_{n-1}}\sin\phi_{n-1}' = f_2
\end{aligned}$$

$$\begin{aligned}
&-v_{\alpha_1}\Delta E_1' + v_{s_1}\cos\phi_1' - \left(v_{\alpha_1} + v_{\alpha_2}\right)\Delta E_2' + v_{s_2}\cos\phi_2' - \left(v_{\alpha_1} + v_{\alpha_2} + v_{\alpha_3}\right)\Delta E_3' + v_{s_3}\cos\phi_3' - \cdots \\
&\cdots - \left(\sum_{k-1}^{n-1} v_{\alpha_k}\right)\Delta E_{n-1}' + v_{s_{n-1}}\cos\phi_{n-1}' = f_3
\end{aligned}$$

Gathering together the coefficients of $v_{\alpha_1}, v_{\alpha_2}, v_{\alpha_3},$ etc and rearranging gives

$$\boxed{\begin{aligned}
&\left(\sin\phi_1'\right)v_{s_1} + \left(\sin\phi_2'\right)v_{s_2} + \left(\sin\phi_3'\right)v_{s_3} + \cdots + \left(\sin\phi_{n-1}'\right)v_{s_{n-1}} \\
&+\left(N_n' - N_1\right)v_{\alpha_1} + \left(N_n' - N_2'\right)v_{\alpha_2} + \left(N_n' - N_3'\right)v_{\alpha_3} + \cdots + \left(N_n' - N_{n-1}'\right)v_{\alpha_{n-1}} = f_2
\end{aligned}} \tag{6.57}$$

$$\begin{aligned}&\left(\cos\phi'_1\right)v_{s_1}+\left(\cos\phi'_2\right)v_{s_2}+\left(\cos\phi'_3\right)v_{s_3}+\cdots+\left(\cos\phi'_{n-1}\right)v_{s_{n-1}}\\&-\left(E'_n-E_1\right)v_{\alpha_1}-\left(E'_n-E'_2\right)v_{\alpha_2}-\left(E'_n-E'_3\right)v_{\alpha_3}-\cdots-\left(E'_n-E'_{n-1}\right)v_{\alpha_{n-1}}=f_3\end{aligned} \qquad (6.58)$$

Equations (6.53), (6.57) and (6.58) are the three equations that relate corrections to angles and distances, v_α and v_s respectively to angular and coordinate misclosures f_1, f_2 and f_3 given by equations (6.54) and (6.56). In equation (6.53) the coefficients of corrections to angles are all unity, whilst in equations (6.57) and (6.58) the coefficients of the corrections are sines and cosines of bearings and coordinate differences derived from the measurements. Equations (6.53), (6.57) and (6.58) are applicable to any single closed traverse.

6.5.5. Application of Least Squares Adjustment of Observations Only to Particular Single Closed Traverses

There are three types of single closed traverses.

Type I Traverses that begin and end at different fixed points with fixed orienting bearings at the terminal points. Figure 6.7(a).

Type II Traverses that begin and end at the same point with a single fixed orienting bearing. Figure 6.7(b)

Type III Traverses that begin and end at the same point with a fixed datum bearing. Figure 6.7(b)

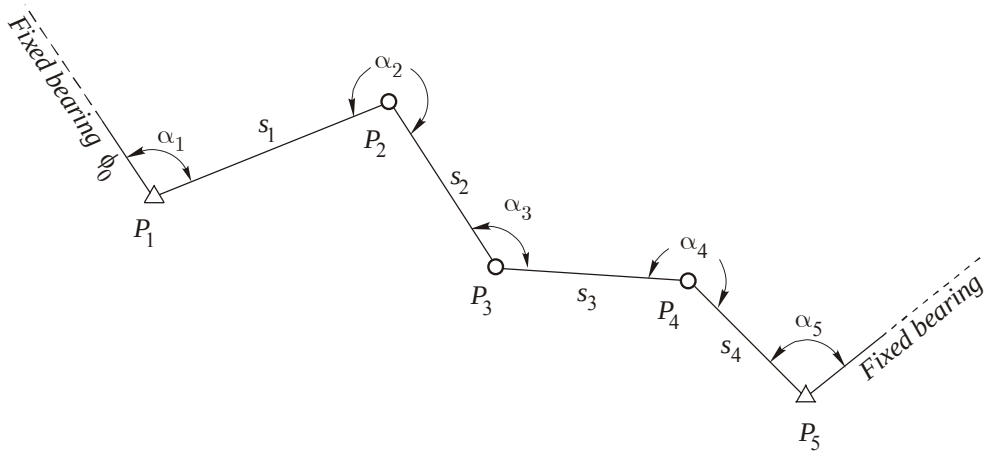


Figure 6.7(a) Type I traverse

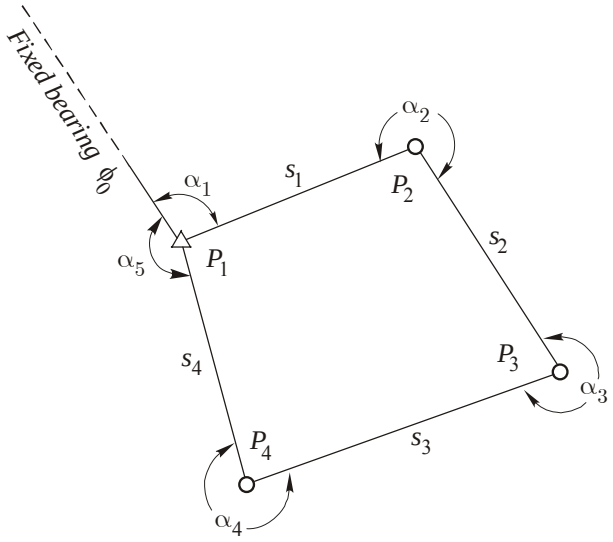


Figure 6.7(b) Type II traverse

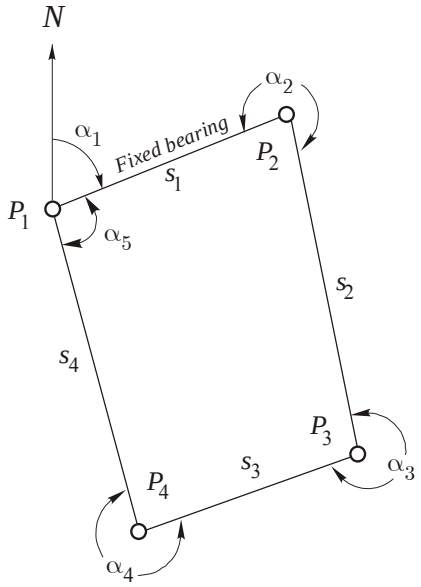


Figure 6.7(c) Type III traverse

Figures 6.7(a), 6.7(b) and 6.7(c) show three types of closed traverses. In each case, the traverse consists of four(4) distances s_1 to s_4 and five(5) angles α_1 to α_5. Traverse points shown with a triangle (Δ) can be regarded as fixed with known coordinates.

In Figures 6.7(a) and 6.7(b) the bearing of the traverse line $P_1 \rightarrow P_2$ is found by adding the observed angle α_1 to the fixed bearing ϕ_0. In both of these traverses five angles must be observed to "close" the traverse.

In Figure 6.7(c) the bearing of the traverse line $P_1 \rightarrow P_2$ is fixed and only four angles need be observed to close the traverse. The angle α_1 at P_1, clockwise from north to P_2, is the bearing of the traverse line $P_1 \rightarrow P_2$. α_1 is used in the adjustment as an observation with a standard deviation of zero.

For any single closed traverse, the method of adjustment is as follows:

(i) Calculate the coordinates of the traverse points by using the observed bearings and distances beginning at point P_1.

(ii) Calculate the angular and coordinate misclosures. In each case, the misclose is the fixed value minus the observed or calculated value. These three values are the elements f_1, f_2 and f_3 in the vector of numeric terms **f**

(iii) Calculate the coefficients of the correction (or residuals) in equations (6.53), (6.57) and (6.58). These coefficients are either zero or unity for equation (6.53), or sines and cosines of observed bearings together with coordinate differences in equations (6.57) and (6.58). These values are the elements of the coefficient matrix **A**

(iv) Assign precisions (estimated standard deviations squared) of the observations. These will be the diagonal elements of the cofactor matrix **Q**

Note: In Type III traverses where the bearing $P_1 \rightarrow P_2$ is fixed, the angle α_1 (which is not observed) is assigned a variance (standard deviation squared) of zero.

(v) Form a set of three(3) normal equations $\left(\mathbf{AQA}^T\right)\mathbf{k} = \mathbf{f}$

(vi) Solve the normal equations for the three(3) Lagrange multipliers k_1, k_2 and k_3 which are the elements of the vector **k** from $\mathbf{k} = \left(\mathbf{AQA}^T\right)^{-1}\mathbf{f}$ and then compute the vector of residuals (corrections) from $\mathbf{v} = \mathbf{QA}^T\mathbf{k}$

(vii) Calculate the adjusted bearings and distances of the traverse by adding the corrections to the observed angles and distances.

6.5.6. Example of Traverse Adjustment using Least Squares Adjustment of Observations Only

Figure 6.8 is a schematic diagram of a traverse run between two fixed stations *A* and *B* and oriented at both ends by angular observations to a third fixed station *C*.

The bearings of traverse lines shown on the diagram, unless otherwise indicated, are called "observed" bearings and have been derived from the measured angles (which have been derived from observed theodolite directions) and the fixed bearing *AC*. The difference between the observed and fixed bearings of the line *BC* represents the angular misclose. The coordinates of the traverse points *D, E* and *F* have been calculated using the observed bearings and distances and the fixed coordinates of *A*. The difference between the observed and fixed coordinates at *B* represents the coordinate misclosures.

In this example estimated standard deviations of measured angles α are $s_\alpha = 5''$ and for measured distances *s* are $s_s = 10\text{mm} + 15\text{ppm}$ where ppm is parts per million.

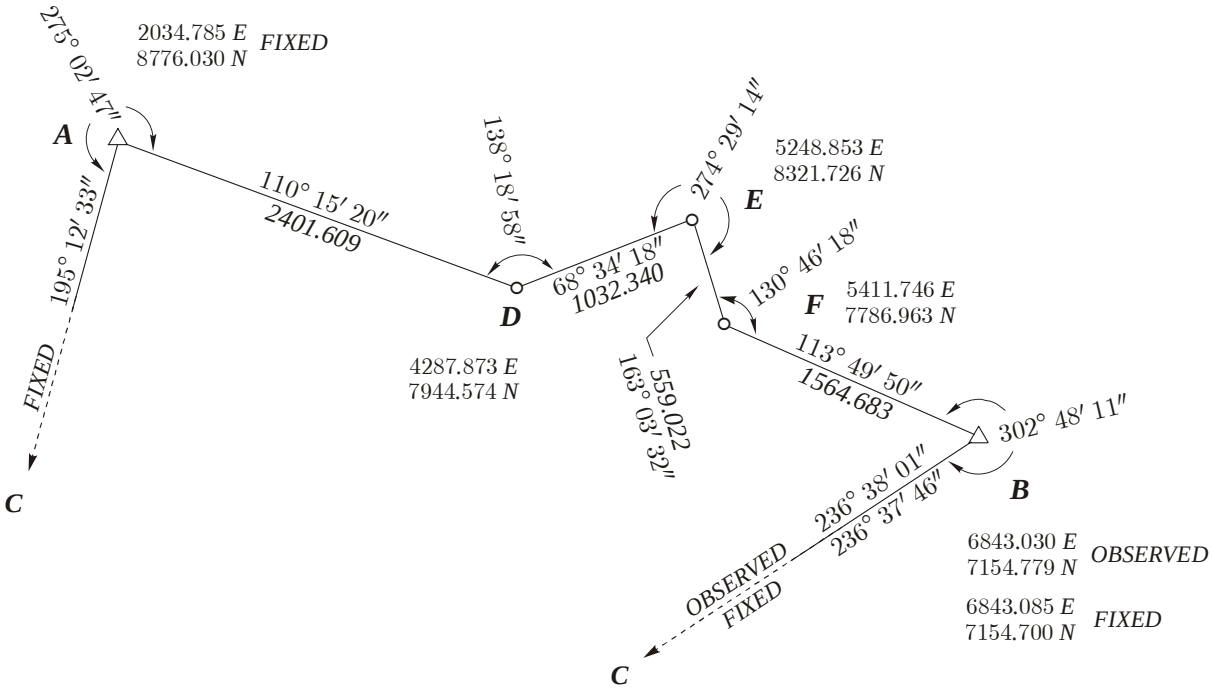


Figure 6.8 Traverse diagram showing field measurements, derived values and fixed values.

Step 1: Calculation of misclosures and formation of vector **f**

From equations (6.54) and (6.56) the angular and coordinate misclosures are the elements f_1, f_2 and f_3 of the vector of numeric terms **f**. These misclosures may be characterised as *misclose = fixed - observed*

angular misclose:
$$\begin{aligned} f_1 &= \phi_n - \phi'_n \\ &= 236^\circ\,37'\,46'' - 236^\circ\,38'\,01'' \\ &= -15'' \end{aligned}$$

east misclose:
$$\begin{aligned} f_2 &= E_n - E'_n \\ &= 6843.085 - 6843.030 \\ &= 0.055 \text{ m} \\ &= 5.5 \text{ cm} \end{aligned}$$

north misclose:
$$\begin{aligned} f_3 &= N_n - N'_n \\ &= 7154.700 - 7154.779 \\ &= -0.079 \text{ m} \\ &= -7.9 \text{ cm} \end{aligned}$$

vector of numeric terms:
$$\mathbf{f} = \begin{bmatrix} -15 \\ 5.50 \\ -7.9 \end{bmatrix} \begin{matrix} \text{sec} \\ \text{cm} \\ \text{cm} \end{matrix}$$

Note that the units of the numeric terms are seconds of arc (sec) and centimetres (cm)

Step 2: Form the coefficient matrix **A** of the equations (6.16) **Av = f**

The first row of **A** contains coefficients of zero or unity from equation (6.53)

$$v_{\alpha_1} + v_{\alpha_2} + v_{\alpha_3} + \cdots + v_{\alpha_n} = f_1$$

The second row of **A** contains the coefficients $\sin\phi'_k$ and $\left(N'_n - N'_k\right)\left(\frac{100}{\rho''}\right)$ from equation (6.57).

$$\begin{aligned} &\left(\sin\phi'_1\right)v_{s_1} + \left(\sin\phi'_2\right)v_{s_2} + \left(\sin\phi'_3\right)v_{s_3} + \cdots + \left(\sin\phi'_{n-1}\right)v_{s_{n-1}} \\ &+\left(N'_n - N_1\right)v_{\alpha_1} + \left(N'_n - N'_2\right)v_{\alpha_2} + \left(N'_n - N'_3\right)v_{\alpha_3} + \cdots + \left(N'_n - N'_{n-1}\right)v_{\alpha_{n-1}} = f_2 \end{aligned}$$

Note that the coefficients of the distance residuals are dimensionless quantities and the coefficients of the angle residuals have the dimensions of sec/cm where $\rho'' = \frac{180}{\pi} \times 3600$ is the number of seconds in one radian.

The third row of **A** contains the coefficients $\cos\phi'_k$ and $-\left(E'_n - E'_k\right)\left(\frac{100}{\rho''}\right)$ from equation (6.58).

$$\left(\cos\phi'_1\right)v_{s_1} + \left(\cos\phi'_2\right)v_{s_2} + \left(\cos\phi'_3\right)v_{s_3} + \cdots + \left(\cos\phi'_{n-1}\right)v_{s_{n-1}}$$
$$-\left(E'_n - E_1\right)v_{\alpha_1} - \left(E'_n - E'_2\right)v_{\alpha_2} - \left(E'_n - E'_3\right)v_{\alpha_3} - \cdots - \left(E'_n - E'_{n-1}\right)v_{\alpha_{n-1}} = f_3$$

Note that the coefficients of the distance residuals are dimensionless quantities and the coefficients of the angle residuals have the dimensions of sec/cm where $\rho'' = \frac{180}{\pi} \times 3600$ is the number of seconds in one radian. The equation $\mathbf{Av} = \mathbf{f}$ is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0.9382 & 0.9309 & 0.2914 & 0.9147 & -0.7860 & -0.3829 & -0.5658 & -0.3065 & 0 \\ -0.3462 & 0.3653 & -0.9566 & -0.4040 & -2.3311 & -1.2388 & -0.7729 & -0.6939 & 0 \\ 4.6\text{ cm} & 2.5\text{ cm} & 1.8\text{ cm} & 3.3\text{ cm} & 5'' & 5'' & 5'' & 5'' & 5'' \end{bmatrix} \begin{bmatrix} v \\ v \\ v \\ v \\ v \\ v \\ v \\ v \\ v \end{bmatrix} \begin{matrix} \uparrow \\ \text{angles} \\ \\ \\ \downarrow \\ \uparrow \\ \\ \text{distances} \\ \\ \downarrow \end{matrix} = \begin{bmatrix} -15 \\ 5.50 \\ -7.90 \end{bmatrix}$$

Note that the numbers below the columns of **A** are the estimates of the standard deviations of the distances or angles associated with the coefficients.

Step 3: Form the normal equations

The normal equations are given by equations (6.25) as $\left(\mathbf{AQA}^T\right)\mathbf{k} = \mathbf{f}$

where $\mathbf{Q} = \mathbf{W}^{-1}$ is the cofactor matrix containing estimates of the variances of the measurements. **Q** and the weight matrix are **W** are square diagonal matrices, i.e., all off-diagonal elements are zero and since weights are inversely proportional to the estimates of the

variances, the diagonal elements of $\mathbf{Q} = \left\{ \frac{1}{s_{\alpha 1}^2} \quad \frac{1}{s_{\alpha 2}^2} \quad \frac{1}{s_{\alpha 3}^2} \quad \frac{1}{s_{\alpha 4}^2} \;\vdots\; \frac{1}{s_{s1}^2} \quad \frac{1}{s_{s2}^2} \quad \frac{1}{s_{s3}^2} \quad \frac{1}{s_{s4}^2} \quad \frac{1}{s_{s5}^2} \right\}$

where the first 4 elements relate to the angles and the remaining 5 elements relate to the distances. Now consider a diagonal matrix that denoted $\sqrt{\mathbf{Q}}$ whose diagonal elements are the square-roots of the elements of $\mathbf{Q}$ and $\mathbf{Q} = \sqrt{\mathbf{Q}}\sqrt{\mathbf{Q}}$ and another matrix $\bar{\mathbf{A}} = \mathbf{A}\sqrt{\mathbf{Q}}$. Each element of $\bar{\mathbf{A}}$ is the original element of $\mathbf{A}$ multiplied by the estimate of the standard deviation associated with the particular element and the normal equations are given by $\left(\bar{\mathbf{A}}\bar{\mathbf{A}}^T\right)\mathbf{k} = \mathbf{f}$ where

$$\bar{\mathbf{A}} = \left[\begin{array}{ccccc:cccc} 0 & 0 & 0 & 0 & 5 & 5 & 5 & 5 & 5 \\ 4.3155 & 2.3272 & 0.5245 & 3.0186 & -1.9145 & -1.9145 & -2.8288 & -1.5324 & 0 \\ -1.5926 & 0.9134 & -1.7219 & -1.3333 & -6.1939 & -6.1939 & -3.8644 & -3.4696 & 0 \end{array}\right]$$

and

$$\bar{\mathbf{A}}\bar{\mathbf{A}}^T = \begin{bmatrix} 125 & -51.0285 & -125.9168 \\ -51.0285 & 62.8875 & 64.2375 \\ -125.9168 & 64.2375 & 209.2995 \end{bmatrix}$$

Step 4: Solve the normal equations for the vector of Lagrange multipliers **k**

From equations (6.47) and with the modification mentioned above

$$\mathbf{k} = \left(\mathbf{AQA}^T\right)^{-1}\mathbf{f} = \left(\bar{\mathbf{A}}\bar{\mathbf{A}}^T\right)^{-1}\mathbf{f} = \begin{bmatrix} 0.0219 & 0.0063 & 0.0112 \\ & 0.0250 & -0.0039 \\ \text{symmetric} & & 0.0127 \end{bmatrix}\begin{bmatrix} -15 \\ 5.5 \\ -7.9 \end{bmatrix}$$

and

$$\mathbf{k} = \begin{bmatrix} -0.3825 \\ 0.0738 \\ -0.2906 \end{bmatrix}$$

Step 5: Calculation of residuals and adjusted traverse dimensions

The residuals are obtained from equation (6.24) $\mathbf{v} = \mathbf{Q}\mathbf{A}^T\mathbf{k}$

Since the cofactor matrix **Q** is diagonal, the individual residuals can be calculated from

$$v_j = s_j^2 \left(a_{1j}k_1 + a_{2j}k_2 + a_{3j}k_3 \right) \tag{6.59}$$

where

$a_{1j},\ a_{2j},\ a_{3j}$ are elements of the coefficient matrix **A**

$k_1,\ k_2,\ k_3$ are the elements of the vector **k**

s_j^2 is the estimate of the variance of the j^{th} measurement

For example, the residual for the second distance $\left(j = 2 \right)$ is

$$\left(2.5\right)^2 \left\{ \left(0\right)\left(-0.3825\right) + \left(0.9309\right)\left(0.0738\right) + \left(0.3653\right)\left(-0.2906\right) \right\} = -0.23 \text{ cm}$$

and the residual for the third measured angle $\left(j = 7 \right)$ is

$$\left(5\right)^2 \left\{ \left(1\right)\left(-0.3825\right) + \left(-0.5658\right)\left(0.0738\right) + \left(-0.7729\right)\left(-0.2906\right) \right\} = -4.99''$$

Exactly the same result can be obtained by using the estimate of the standard deviations s_j and the elements of the matrix $\bar{\mathbf{A}}$

$$v_j = s_j \left(\bar{a}_{1j}k_1 + \bar{a}_{2j}k_2 + \bar{a}_{3j}k_3 \right) \tag{6.60}$$

Both methods give

$$\mathbf{v} = \begin{bmatrix} 3.59 \text{ cm} \\ -0.23 \\ 0.97 \\ 2.01 \text{ cm} \\ 5.92'' \\ -1.27 \\ -4.99 \\ -5.09 \\ -9.56'' \end{bmatrix} \begin{matrix} \uparrow \\ \\ \text{distances} \\ \downarrow \\ \uparrow \\ \\ \text{angles} \\ \\ \downarrow \end{matrix}$$

The residuals for the bearings are the cumulative residuals for the angles up to the particular traverse line. They are

$$\mathbf{v}_\phi = \begin{bmatrix} 5.92'' \\ 4.65 \\ -0.34 \\ -5.43 \\ -14.99'' \end{bmatrix}$$

Applying these residuals (or corrections) to the measured quantities gives the adjusted traverse dimensions as

Line	Bearing	Distance
k	ϕ_k	s_k
1	110º 27′ 25.9″	2401.645
2	68º 34′ 22.6″	1032.338
3	163º 03′ 31.7″	559.032
4	113º 49′ 44.6″	1564.703

Table 6.6 Adjusted traverse distances

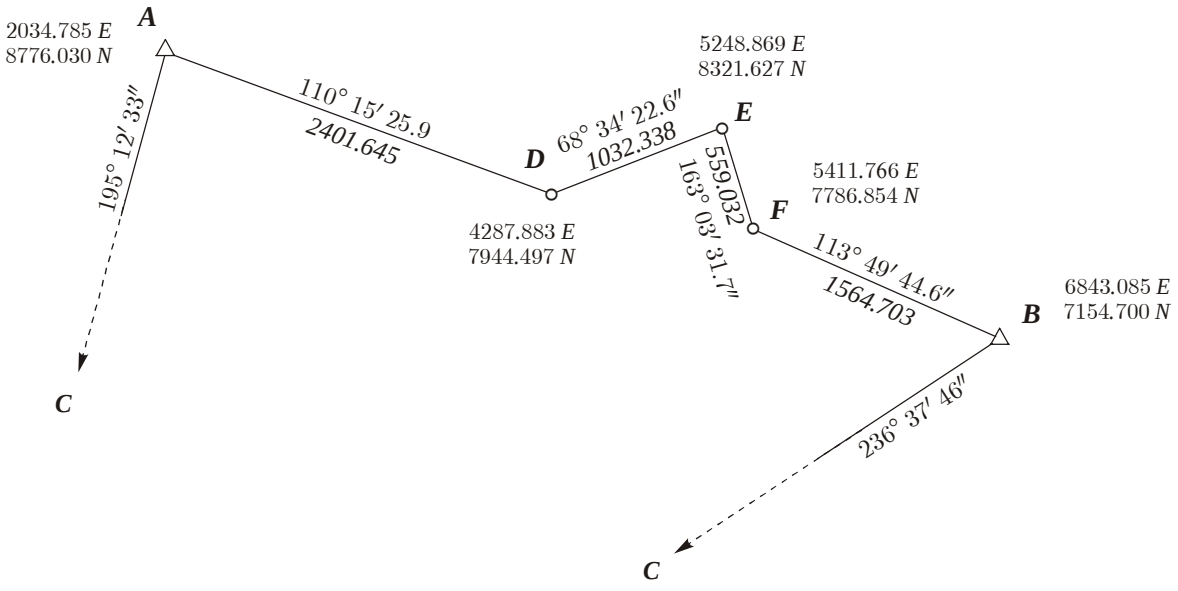


Figure 6.9 Traverse diagram showing adjusted measurements.

7. LINEARIZATION USING TAYLOR'S THEOREM AND THE DERIVATION OF SOME COMMON SURVEYING OBSERVATION EQUATIONS

In many surveying "problems" the solution depends upon selection of a *mathematical model* suitable to the problem, and using this, together with the *observations* (or measurements) obtain a solution.

For example, a surveyor is required to determine the location (the coordinates) of a point. From this "unknown" point, they can see three known points (i.e., points of known coordinates). Understanding geometric principles, the surveyor measures the directions to these three known points with a theodolite, determines the two angles α and β between the three lines and "solves the problem". In surveying parlance, this technique of solution of position is known as a resection; the mathematical model is based on geometric principles and the observations are the directions, from which the necessary angles are obtained for a solution.

Choosing a resection, as an example of a "surveying problem" is appropriate, since it demonstrates the case of determining quantities (the coordinates of the unknown point) from indirect measurements. That is, the surveyor's measurements of directions are indirect measurements of coordinate differences between the unknown point and the known points.

In many surveying problems, the observations exceed the necessary number required for a unique solution. Again, using a resection as an example, consider the case where the surveyor (at an unknown point) measures the directions to four known points. There are now multiple solutions for the resection point, since the four directions give rise to three angles, exceeding the minimum geometric requirements for a unique solution. That is, there is a *redundancy* in the mathematical model. In this case of the resection, and other surveying problems where there are redundant measurements, the method of *least squares* can be employed to obtain the *best estimate* of the "unknowns".

Least squares (as a method of determining best estimates), depends upon the formation of sets of *observation equations* and their solution. The normal techniques of solution of systems of equations require that the sets of observation equations must be *linear,* i.e., "unknowns"

linearly related to measurements. This is not always the case. For example, in a resection, the measurements, directions α_{ik} from the unknown point P_i to known points P_k, are non-linear functions of the coordinate differences (the unknowns).

The observation equations for observed directions in a mathematical model of a resection have the general form

$$\alpha_{ik} + v_{ik} + z_i = \tan^{-1}\left(\frac{E_k - E_i}{N_k - N_i}\right) \tag{7.1}$$

α_{ik} are the observed directions from the resection point P_i to the known points P_k,

v_{ik} are the residuals (small corrections) associated with observed directions,

z_i is an orientation "constant"; the bearing of the Reference Object (RO) for the set of observed directions,

E_k, N_k are the east and north coordinates of the known points, and

E_i, N_i are the east and north coordinates of the resection point.

Clearly, in this case, the measurements α_{ik} are non-linear functions of the unknowns E_i, N_i and any system of equations in the form of (7.1) would be *non-linear* and could not be solved by normal means. Consequently, whenever the equations in a mathematical model are non-linear functions linking the measurements with the unknowns, some method of linearization must be employed to obtain sets of *linear equations.*

The most common method of linearization is by using *Taylor's* theorem to represent the function as a power series consisting of zero order terms, 1st order terms, 2nd order terms and higher order terms. By choosing suitable approximations, second and higher-order terms can be neglected, yielding a linear approximation to the function. This linear approximation of the mathematical model can be used to form sets of linear equations, which can be solved by normal means.

7.1. Taylor's Theorem

This theorem, due to the English mathematician Brook Taylor (1685–1731) enables the value of a real function $f(x)$ near a point $x = a$ to be estimated from the values $f(a)$ and the derivatives of $f(x)$ evaluated at $x = a$. Taylor's theorem also provides an estimate of the error made in a polynomial approximation to a function. The Scottish mathematician Colin Maclaurin (1698–1746) developed a special case of Taylor's theorem, which was named in his honour, where the function $f(x)$ is expanded about the origin $x = a = 0$. The citations below, from the Encyclopaedia Britannica give some historical information about Taylor and Maclaurin.

> **Taylor**, Brook (b. Aug. 18, 1685, Edmonton, Middlesex, Eng.– d. Dec. 29, 1731, London), British mathematician noted for his contributions to the development of calculus.
> In 1708 **Taylor** produced a solution to the problem of the centre of oscillation. The solution went unpublished until 1714, when his claim to priority was disputed by the noted Swiss mathematician Johann Bernoulli. **Taylor**'s *Methodus incrementorum directa et inversa* (1715; "Direct and Indirect Methods of Incrementation") added to higher mathematics a new branch now called the calculus of finite differences. Using this new development, he was the first to express mathematically the movement of a vibrating string on the basis of mechanical principles. *Methodus* also contained the celebrated formula known as **Taylor**'s theorem, the importance of which remained unrecognized until 1772. At that time the French mathematician Joseph-Louis Lagrange realized its importance and proclaimed it the basic principle of differential calculus.
> A gifted artist, **Taylor** set forth in *Linear Perspective* (1715) the basic principles of perspective. This work and his *New Principles of Linear Perspective* contained the first general treatment of the principle of vanishing points. **Taylor** was elected a fellow of the Royal Society of London in 1712 and in that same year sat on the committee for adjudicating Sir Isaac Newton's and Gottfried Wilhelm Leibniz's conflicting claims of priority in the invention of calculus.

> **Maclaurin**, Colin (b. February 1698, Kilmodan, Argyllshire, Scot.–d. June 14, 1746, Edinburgh), Scottish mathematician who developed and extended Sir Isaac Newton's work in calculus, geometry, and gravitation. A child prodigy, he entered the University of Glasgow at age 11. At the age of 19, he was elected professor of mathematics at Marischal College, Aberdeen, and two years later he became a fellow of the Royal Society of London. At this time he became acquainted with Newton. In his most important work, *Geometrica Organica; Sive Descriptio Linearum Curvarum Universalis* (1720; "Organic Geometry, with the Description of the Universal Linear Curves"), **Maclaurin** developed several theorems similar to some in Newton's *Principia,* introduced the method of generating conics (the circle, ellipse, hyperbola, and parabola) that bears his name, and showed that certain types of curves (of the third and fourth degree) can be described by the intersection of two movable angles. On the recommendation of Newton, he was made professor of mathematics at the University of Edinburgh in 1725. In 1740 he shared, with the mathematicians Leonhard Euler and Daniel Bernoulli, the prize offered by the Académie des Sciences for an essay on tides. His *Treatisw of Fluxions* (1742) was written in reply to criticisms by George Berkeley of England that Newton's calculus was based on faulty reasoning. In this essay he showed

that stable figures for a homogeneous rotating fluid mass are the ellipsoids of revolution, later known as **Maclaurin**'s ellipsoids. He also gave in his *Fluxions*, for the first time, the correct theory for distinguishing between maxima and minima in general and pointed out the importance of the distinction in the theory of the multiple points of curves. The **Maclaurin** series, a special case of the Taylor series, was named in his honour. In 1745, when Jacobites (supporters of the Stuart king James II and his descendants) were marching on Edinburgh, **Maclaurin** took a prominent part in preparing trenches and barricades for the city's defense. As soon as the rebel army captured Edinburgh, **Maclaurin** fled to England until it was safe to return. The ordeal of his escape ruined his health, and he died at age 48. **Maclaurin**'s *Account of Sir Isaac Newton's Philosophical Discoveries* was published posthumously, as was his *Treatise of Algebra* (1748). "De Linearum Geometricarum Proprietatibus Generalibus tractatus" ("A Tract on the General Properties of Geometrical Lines"), noted for its elegant geometric demonstrations, was appended to his *Algebra.*

Taylor's theorem may be expressed in the following form

$$f(x) = f(a) + (x-a) f'(a) + \frac{(x-a)^2}{2!} f''(a) + \frac{(x-a)^3}{3!} f'''(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) + R_n \tag{7.2}$$

where R_n is the remainder after n terms and $\lim_{n\to\infty} R_n = 0$ for $f(x)$ about $x = a$

$f'(a),\ f''(a), \cdots$ etc are derivatives of the function $f(x)$ evaluated at $x = a$.

Taylor's theorem can also be expressed as power series

$$f(x) = \sum_{k=0}^{n} f^{(k)}(a) \frac{(x-a)^k}{k!} \tag{7.3}$$

where $f^{(k)}(a)$ denotes the kth derivative of the function $f(x)$ evaluated at $x = a$ and $f^{(0)}(a)$ is the function $f(x)$ evaluated at $x = a$, and $0! = 1$.

Other forms of Taylor's theorem may be obtained by a change of notation, for example: let $x = a + h$, then $f(x) = f(a+h)$ and $x - a = h$. Substitution into equation (7.2) gives

$$f(x) = f(a+h) = f(a) + h f'(a) + \frac{h^2}{2!} f''(a) + \frac{h^3}{3!} f'''(a) + \cdots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(a) + R_n \tag{7.4}$$

This may be a more convenient form of Taylor's theorem for a particular application.

Inspection of equations (7.2), (7.3) and (7.4) show that Taylor's theorem can be used to expand a non-linear function (about a point) into a linear series. Expansions of this form, also called **Taylor's series**, are a convergent power series approximating $f(x)$.

Taylor's series for functions of two variables

Say $\phi = f(x,y)$ then the Taylor series expansion of the function ϕ about $x = a$ and $y = b$ is

$$\phi = f(a,b) + (x-a)\frac{\partial f}{\partial x} + (y-b)\frac{\partial f}{\partial y} + \frac{1}{2!}\left\{(x-a)^2\frac{\partial^2 f}{\partial x^2} + (y-b)^2\frac{\partial^2 f}{\partial y^2} + (x-a)(y-b)\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right\} + \cdots \qquad (7.5)$$

where $f(a,b)$ is the function ϕ evaluated at $x = a$ and $y = b$

$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial^2 f}{\partial x^2}$, etc are partial derivatives of the function ϕ evaluated at $x = a$ and $y = b$.

Taylor's series for functions of three variables

Say $\phi = f(x,y,z)$ then the Taylor series expansion of the function ϕ about $x = a$, $y = b$ and $z = c$

$$\phi = f(a,b,c) + (x-a)\frac{\partial f}{\partial x} + (y-b)\frac{\partial f}{\partial y} + (z-c)\frac{\partial f}{\partial z} + \frac{1}{2!}\left\{(x-a)^2\frac{\partial^2 f}{\partial x^2} + (y-b)^2\frac{\partial^2 f}{\partial y^2} + (z-c)^2\frac{\partial^2 f}{\partial z^2} + (x-a)(y-b)\frac{\partial f}{\partial x}\frac{\partial f}{\partial y} + (x-a)(z-c)\frac{\partial f}{\partial x}\frac{\partial f}{\partial z} + (y-b)(z-c)\frac{\partial f}{\partial y}\frac{\partial f}{\partial z}\right\} + \cdots \qquad (7.6)$$

where $f(a,b,c)$ is the function ϕ evaluated at $x = a$, $y = b$ and $z = c$

$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial^2 f}{\partial x^2}$, etc are partial derivatives evaluated at $x = a$, $y = b$ and $z = c$.

Extensions to four or more variables follow a similar pattern. Equations (7.5) and (7.6) show only terms up to the 2nd order; no remainder terms are shown.

7.2. Linear Approximations to Functions using Taylor's Theorem

In the Taylor expansions of functions shown above, suppose that the variables $x, y, z, \cdots$ etc are expressed as $x = x^0 + \Delta x$, $y = y^0 + \Delta y$, $z = z^0 + \Delta z \cdots$ etc where $x^0, y^0, z^0, \cdots$ etc are approximate values and $\Delta x, \Delta y, \Delta z, \cdots$ etc are small corrections.

The Taylor series expansion of a single variable can be expressed as

$$\begin{aligned}\phi = f(x) &= f(x^0) + (x - x^0)\frac{df}{dx} + (x - x^0)^2 \frac{d^2 f}{dx^2} + (x - x^0)^3 \frac{d^3 f}{dx^3} + \cdots \\ &= f(x^0) + \Delta x \frac{df}{dx} + (\Delta x)^2 \frac{d^2 f}{dx^2} + (\Delta x)^3 \frac{d^3 f}{dx^3} + \cdots \\ &= f(x^0) + \Delta x \frac{df}{dx} + \text{ higher order terms}\end{aligned}$$

where the derivatives $\frac{df}{dx}, \frac{d^2 f}{dx^2}, \frac{d^3 f}{dx^3}, \cdots$ etc are evaluated at the approximation x^0. If the correction Δx is small, then $(\Delta x)^2, (\Delta x)^3, \cdots$ etc will be exceedingly small and the higher order terms may be neglected, giving the following linear approximation

For $\phi = f(x)$ $$\phi = f(x) \simeq f(x^0) + \Delta x \frac{df}{dx} \tag{7.7}$$

Using similar reasoning, linear approximations can be written for functions of two and three variables.

For $\phi = f(x, y)$ $$\phi = f(x, y) \simeq f(x^0, y^0) + \Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y} \tag{7.8}$$

For $\phi = f(x, y, z)$, $$\phi = f(x, y, z) \simeq f(x^0, y^0, z^0) + \Delta x \frac{\partial f}{\partial x} + \Delta y \frac{\partial f}{\partial y} + \Delta z \frac{\partial f}{\partial z} \tag{7.9}$$

Similar linear approximations can be written for functions of four or more variables. In equations (7.7), (7.8) and (7.9) the derivatives are evaluated at the approximations x^0, y^0, z^0.

Generalizing this linear form gives, for $\phi = f\left(x_1, x_2, x_3, \cdots x_n\right)$

$$\phi = f(x_1, x_2, x_3, \cdots x_n) \simeq f\left(x_1^0, x_2^0, x_3^0, \cdots x_n^0\right) + \Delta x_1 \frac{\partial f}{\partial x_1} + \Delta x_2 \frac{\partial f}{\partial x_2} + \Delta x_3 \frac{\partial f}{\partial x_3} + \cdots + \Delta x_n \frac{\partial f}{\partial x_n} \quad (7.10)$$

This equation can be written in matrix form

$$\phi = f\left(\mathbf{x}\right) \simeq f\left(\mathbf{x}^0\right) + \mathbf{j}\Delta\mathbf{x} \quad (7.11)$$

where **x** is a vector of variables, $\mathbf{x}^0$ a vector of approximate values of the variables, **j** is a row vector of partial derivatives and $\Delta\mathbf{x}$ is a column vector of corrections.

Suppose this generalized form, equation (7.11), is extended to the general case of *m* variables $y_1, y_2, y_3, \cdots y_m$ and each variable y_k is a function of a set of variables $x_1, x_2, x_3, \cdots x_n$ i.e.,

$$\begin{aligned} y_1 &= f_1\left(x_1, x_2, x_3, \cdots x_n\right) \\ y_2 &= f_2\left(x_1, x_2, x_3, \cdots x_n\right) \\ &\vdots \\ y_m &= f_m\left(x_1, x_2, x_3, \cdots x_n\right) \end{aligned}$$

Expressing each variable y_k in a linearized form gives

$$\begin{aligned} y_1 &= y_1^0 + \frac{\partial y_1}{\partial x_1}\Delta x_1 + \frac{\partial y_1}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial y_1}{\partial x_n}\Delta x_n \\ y_2 &= y_2^0 + \frac{\partial y_2}{\partial x_1}\Delta x_1 + \frac{\partial y_2}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial y_2}{\partial x_n}\Delta x_n \\ &\vdots \\ y_m &= y_m^0 + \frac{\partial y_m}{\partial x_1}\Delta x_1 + \frac{\partial y_m}{\partial x_2}\Delta x_2 + \cdots + \frac{\partial y_m}{\partial x_n}\Delta x_n \end{aligned} \quad (7.12)$$

Equations (7.12) can be expressed in matrix notation as

$$\mathbf{y} = \mathbf{y}^0 + \mathbf{J}\,\Delta\mathbf{x} \quad (7.13)$$

where

y is an (*m*,1) vector of (unknown) function values, $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix}^T$

$\mathbf{y}^0$ is an (*m*,1) vector of approximate values of the functions, $\mathbf{y}^0 = \begin{bmatrix} y_1^0 & y_2^0 & \cdots & y_m^0 \end{bmatrix}^T$

$\Delta\mathbf{x}$ is an (*n*,1) vector of corrections to the approx. values, $\Delta\mathbf{x} = \begin{bmatrix} \Delta x_1 & \Delta x_2 & \cdots & \Delta x_n \end{bmatrix}^T$

J is the (*m*,*n*) the Jacobian matrix of partial derivatives

$$\mathbf{J} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_3} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_2}{\partial x_3} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & & & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \frac{\partial y_m}{\partial x_3} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

7.3. The Derivation of some Common Surveying Observation Equations

Consider Figure 7.1. P_i is the instrument point and directions α_{ik} and distances s_{ik} have been observed to stations $P_1, P_2, P_3 \cdots P_k$. P_1 is the Reference Object (RO) and the direction $\alpha_{i1} = 0^\circ\,00'00''$. A bearing is assigned to the RO and bearings to all other stations may be obtained by adding the observed directions to the bearing of the RO.

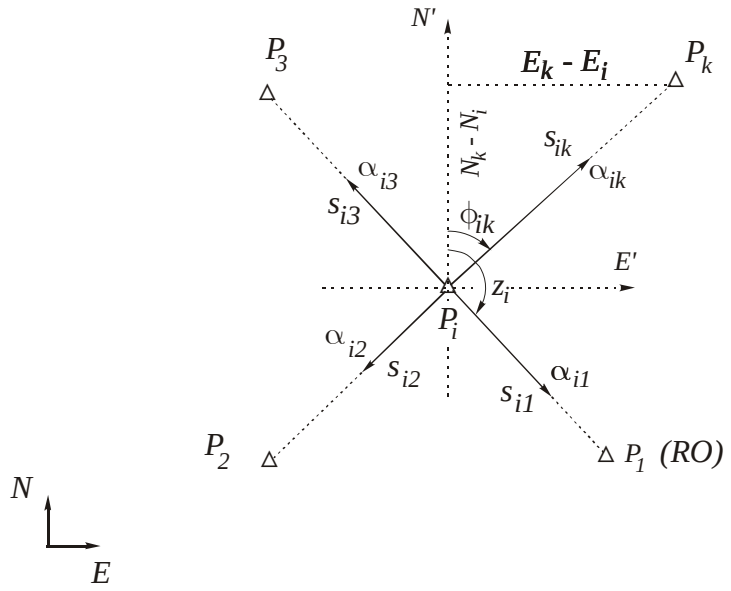


Figure 7.1 Observed directions α and distances *s* from P_i

α_{ik} observed direction P_i to P_k

ϕ_{ik} bearing P_i to P_k

s_{ik}	distance P_i to P_k
z_i	orientation constant for directions at P_i (bearing of the RO)
θ_{ik}	"observed" bearing $\theta_{ik} = \alpha_{ik} + z_i$
E_i, N_i	coordinates of P_i
E_k, N_k	coordinates of P_k

From Figure 7.1 the bearings ϕ_{ik} and distances s_{ik} are non-linear functions of the coordinates of points P_i and P_k

$$\phi_{ik} = \tan^{-1}\left(\frac{E_k - E_i}{N_k - N_i}\right) \tag{7.14}$$

$$s_{ik} = \sqrt{\left(E_k - E_i\right)^2 + \left(N_k - N_i\right)^2} \tag{7.15}$$

With $E = E^0 + \Delta E$ and $N = N^0 + \Delta N$ where E^0, N^0 are approximate values of the coordinates and ΔE, ΔN are small corrections, linear approximations of ϕ_{ik} and s_{ik} can be written as

$$\phi_{ik} = \phi_{ik}^0 + \Delta E_k \frac{\partial \phi_{ik}}{\partial E_k} + \Delta N_k \frac{\partial \phi_{ik}}{\partial N_k} + \Delta E_i \frac{\partial \phi_{ik}}{\partial E_i} + \Delta N_i \frac{\partial \phi_{ik}}{\partial N_i} \tag{7.16}$$

$$s_{ik} = s_{ik}^0 + \Delta E_k \frac{\partial s_{ik}}{\partial E_k} + \Delta N_k \frac{\partial s_{ik}}{\partial N_k} + \Delta E_i \frac{\partial s_{ik}}{\partial E_i} + \Delta N_i \frac{\partial s_{ik}}{\partial N_i} \tag{7.17}$$

where ϕ_{ik}^0 and s_{ik}^0 are approximate bearings and distances respectively, obtained by substituting the approximate coordinates $E_k^0, N_k^0, E_i^0, N_i^0$ into equations (7.14) and (7.15).

The partial derivatives in equations (16) are evaluated in the following manner.

Using the relationships: $\frac{d}{dx}\tan^{-1} u = \frac{1}{1+u^2}\frac{du}{dx}$ and $\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$

The partial derivative $\frac{\partial \phi_{ik}}{\partial E_k}$

$$\frac{\partial \phi_{ik}}{\partial E_k} = \frac{1}{1+\left(\dfrac{E_k - E_i}{N_k - N_i}\right)^2} \frac{\partial}{\partial E_k}\left(\frac{E_k - E_i}{N_k - N_i}\right) = \frac{\left(N_k - N_i\right)^2}{\left(N_k - N_i\right)^2 + \left(E_k - E_i\right)^2} \frac{N_k - N_i}{\left(N_k - N_i\right)^2}$$

giving

$$\frac{\partial \phi_{ik}}{\partial E_k} = \frac{N_k - N_i}{\left(N_k - N_i\right)^2 + \left(E_k - E_i\right)^2} = \frac{N_k - N_i}{s_{ik}^2} = \frac{\cos\phi_{ik}}{s_{ik}} = b_{ik} \tag{7.18}$$

Similarly

$$\frac{\partial \phi_{ik}}{\partial N_k} = \frac{-\left(E_k - E_i\right)}{\left(N_k - N_i\right)^2 + \left(E_k - E_i\right)^2} = \frac{-\left(E_k - E_i\right)}{s_{ik}^2} = \frac{-\sin\phi_{ik}}{s_{ik}} = a_{ik} \tag{7.19}$$

$$\frac{\partial \phi_{ik}}{\partial E_i} = \frac{-\left(N_k - N_i\right)}{\left(N_k - N_i\right)^2 + \left(E_k - E_i\right)^2} = \frac{-\left(N_k - N_i\right)}{s_{ik}^2} = \frac{-\cos\phi_{ik}}{s_{ik}} = -b_{ik} \tag{7.20}$$

$$\frac{\partial \phi_{ik}}{\partial N_i} = \frac{\left(E_k - E_i\right)}{\left(N_k - N_i\right)^2 + \left(E_k - E_i\right)^2} = \frac{\left(E_k - E_i\right)}{s_{ik}^2} = \frac{\sin\phi_{ik}}{s_{ik}} = -a_{ik} \tag{7.21}$$

a_{ik} and b_{ik} are known as *direction coefficients.*

The partial derivatives of equation (7.17) are evaluated in the following manner

The partial derivative $\dfrac{\partial s_{ik}}{\partial E_k}$

$$\frac{\partial s_{ik}}{\partial E_k} = \frac{1}{2}\left[\left(E_k - E_i\right)^2 + \left(N_k - N_i\right)^2\right]^{-\frac{1}{2}} 2\left(E_k - E_i\right) = \frac{E_k - E_i}{s_{ik}} = \sin\phi_{ik} = d_{ik} \tag{7.22}$$

Similarly

$$\frac{\partial s_{ik}}{\partial N_k} = \frac{N_k - N_i}{s_{ik}} = \cos\phi_{ik} = c_{ik} \tag{7.23}$$

$$\frac{\partial s_{ik}}{\partial E_i} = \frac{-\left(E_k - E_i\right)}{s_{ik}} = -\sin\phi_{ik} = -d_{ik} \tag{7.24}$$

$$\frac{\partial s_{ik}}{\partial N_i} = \frac{-\left(N_k - N_i\right)}{s_{ik}} = -\cos\phi_{ik} = -c_{ik} \tag{7.25}$$

c_{ik} and d_{ik} are known as *distance coefficients.*

7.3.1. Observation equation for measured directions

An observation equation, relating observed directions to coordinates P_i and P_k can be written as

$$\alpha_{ik} + v_{ik} + z_i = \phi_{ik} = \tan^{-1}\left(\frac{E_k - E_i}{N_k - N_i}\right) \tag{7.26}$$

where v_{ik} are the residuals (small corrections) associated with observed directions. Using equation (7.16) together with the partial derivatives given in equations (7.18) to (7.21) gives a linear approximation of the observation equation for an observed direction

$$\alpha_{ik} + v_{ik} + z_i = a_{ik}\,\Delta N_k + b_{ik}\,\Delta E_k - a_{ik}\,\Delta N_i - b_{ik}\,\Delta E_i + \phi_{ik}^0 \tag{7.27}$$

where $a_{ik} = \dfrac{-(E_k - E_i)}{s_{ik}^2} = \dfrac{-\sin\phi_{ik}}{s_{ik}}$ and $b_{ik} = \dfrac{N_k - N_i}{s_{ik}^2} = \dfrac{\cos\phi_{ik}}{s_{ik}}$ are the direction coefficients

7.3.2. Observation equation for measured distances

An observation equation, relating observed distances to coordinates P_i and P_k can be written as

$$s_{ik} + v_{ik} = \sqrt{(E_k - E_i)^2 + (N_k - N_i)^2} \tag{7.28}$$

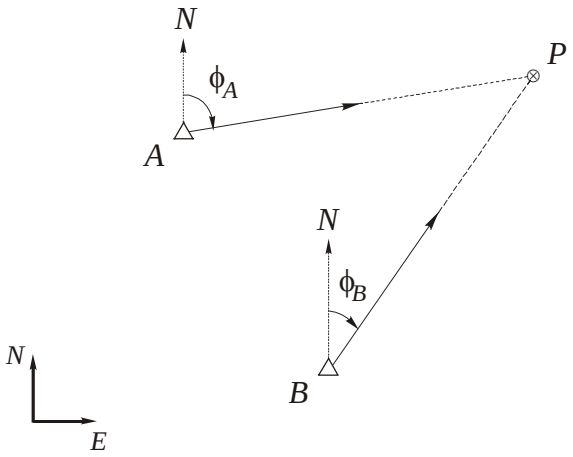
where v_{ik} are the residuals (small corrections) associated with observed distances. Using (7.17) together with the partial derivatives given in equations (7.22) to (7.25) gives a linear approximation of the observation equation for an observed distance

$$s_{ik} + v_{ik} = c_{ik}\,\Delta N_k + d_{ik}\,\Delta E_k - c_{ik}\,\Delta N_i - d_{ik}\,\Delta E_i + s_{ik}^0 \tag{7.29}$$

where $c_{ik} = \dfrac{N_k - N_i}{s_{ik}} = \cos\phi_{ik}$ and $d_{ik} = \dfrac{E_k - E_i}{s_{ik}} = \sin\phi_{ik}$ are the distance coefficients

7.4. An Example of Taylor's Theorem in Practice

Figure 7.2 shows a point *P*, whose coordinates are unknown, intersected by bearings from stations *A* and *B* whose coordinates are known.



Coordinates

A:	12273.910 E	*B:* 12875.270 E
	29612.310 N	28679.600 N

Bearings

$\phi_A = 81^\circ\,01'\,23''$ $\phi_B = 34^\circ\,47'\,52''$

Approximate coords *P:* 13677 E
29834 N

Figure 7.2 Bearing intersection

The information given above can be used to compute the coordinates of *P* by using an iterative technique employing linearized observation equations approximating the bearings ϕ_A and ϕ_B. These observation equations [see equation (7.27)] have been derived using Taylor's theorem.

In general, a bearing is a function of the coordinates of the ends of the line, i.e.,

$$\phi_{ik} = \tan^{-1}\left(\frac{E_k - E_i}{N_k - N_i}\right) = f\left(E_k, N_k, E_i, N_i\right) \tag{7.30}$$

where subscripts *i* and *k* represent instrument and target respectively. In this example (intersection) *A* and *B* are instrument points and are known and *P* is a target point and is unknown hence

$$\phi_{ik} = f\left(E_k, N_k\right)$$

is a non-linear function of the variables E_k and N_k only (the coordinates of *P*). Using equations (7.26) and (7.27) with modifications $\Delta E_i = \Delta N_i = 0$ since the coordinates of the instrument points are known gives

$$\phi_{ik} = a_{ik}\,\Delta N_k + b_{ik}\,\Delta E_k + \phi_{ik}^0 \tag{7.31}$$

$E_k = E_k^0 + \Delta E_k$, $N_k = N_k^0 + \Delta N_k$ and E_k^0, N_k^0, $\Delta E_k, \Delta N_k$ are approximate coordinates and small corrections respectively. $a_{ik} = \dfrac{-(E_k - E_i)}{s_{ik}^2} = \dfrac{-\sin\phi_{ik}}{s_{ik}}$ and $b_{ik} = \dfrac{N_k - N_i}{s_{ik}^2} = \dfrac{\cos\phi_{ik}}{s_{ik}}$ are direction coefficients and ϕ_{ik}^0 is an approximate bearing. Note that ϕ_{ik}^0 and the direction coefficients a_{ik} and b_{ik} are computed using the approximate coordinates of *P*.

Using equation (7.31), two equations for bearings ϕ_A and ϕ_B may be written as

$$\phi_A = a_A \Delta N_P + b_A \Delta E_P + \phi_A^0$$
$$\phi_B = a_B \Delta N_P + b_B \Delta E_P + \phi_B^0$$

These equations can be rearranged and expressed in matrix form as

$$\begin{bmatrix} a_A & b_A \\ a_B & b_B \end{bmatrix} \begin{bmatrix} \Delta N_P \\ \Delta E_P \end{bmatrix} = \begin{bmatrix} \phi_A - \phi_A^0 \\ \phi_B - \phi_B^0 \end{bmatrix}$$

or $$\mathbf{Cx} = \mathbf{u}$$

where **C** is a matrix of direction coefficients, **x** is the vector of corrections to the approximate coordinates of *P* and **u** is a vector of numeric terms (observed bearing – computed bearing).

The solution for the corrections in vector **x** is given by

$$\mathbf{x} = \mathbf{C}^{-1}\mathbf{u}$$

From the information given with Figure 7.2 the computed bearings $\left(\phi_{ik}^0\right)$ and distances $\left(s_{ik}^0\right)$ using the approximate coordinates of *P* are

AP:	81º 01' 17.1"	*BP:*	34º 46' 47.8"
	1420.5 m		*1405.5* m

and the numeric terms in vector **u** are

$$u_A = \phi_A - \phi_A^0 = 81^\circ\,01'23'' - 81^\circ\,01'17.1'' = 5.9''$$

$$u_B = \phi_B - \phi_B^0 = 34^\circ\,47'52'' - 34^\circ\,46'47.8'' = 64.2''$$

With the elements of **x** (the corrections to the approximate coordinates of *P*) in centimetres and the elements of **u** (the differences in observed and computed bearings) in seconds of arc

$$\mathbf{C}\mathbf{x} = \mathbf{u}$$
$$\downarrow \quad \searrow$$
$$\text{cm's} \quad \text{seconds}$$

the elements of the coefficient matrix **C** will be computed in sec/cm (seconds per centimetre) to maintain consistency of units so that

$$\mathbf{x} = \mathbf{C}^{-1}\mathbf{u}$$
$$\swarrow \quad \downarrow \quad \searrow$$
$$\text{cm} \quad \text{cm}/\text{sec} \quad \text{sec}$$

Note that if the units (or dimensions) of the elements of **C** are sec/cm then the units of the elements of the inverse $\mathbf{C}^{-1}$ are cm/sec.

The elements of **C** are the direction coefficients and with distances s_{ik}^0 in centimetres

$$a_{ik} = \frac{-\sin\phi_{ik}^0}{s_{ik}^0} \times \rho'' \qquad \text{and} \qquad b_{ik} = \frac{\cos\phi_{ik}^0}{s_{ik}^0} \times \rho'' \qquad \text{where} \qquad \rho'' = \frac{180}{\pi} \times 3600$$

giving

$$a_A = \frac{-\sin\left(81^\circ\,01'17.1''\right)}{(1420.5)(100)} \times \rho'' = -1.43426 \ \text{sec}/\text{cm}$$
$$b_A = \frac{\cos\left(81^\circ\,01'17.1''\right)}{(1420.5)(100)} \times \rho'' = 0.22662 \ \text{sec}/\text{cm}$$
$$a_B = -0.83714 \ \text{sec}/\text{cm}$$
$$b_B = 1.20538 \ \text{sec}/\text{cm}$$

The matrix equation $\mathbf{C}\mathbf{x} = \mathbf{u}$ is

$$\begin{bmatrix} -1.43426 & 0.22662 \\ -0.83714 & 1.20538 \end{bmatrix} \begin{bmatrix} \Delta N_P \\ \Delta E_P \end{bmatrix} = \begin{bmatrix} 5.9 \\ 64.2 \end{bmatrix}$$

and the solution $\mathbf{x} = \mathbf{C}^{-1}\mathbf{u}$ is

$$\begin{bmatrix} \Delta N_P \\ \Delta E_P \end{bmatrix} = \begin{bmatrix} -0.78316 & 0.14724 \\ -0.54391 & 0.93187 \end{bmatrix} \begin{bmatrix} 5.9 \\ 64.2 \end{bmatrix} = \begin{bmatrix} 4.83 \\ 56.62 \end{bmatrix} \text{cm}$$

giving the "adjusted" coordinates of P as

$$N_P = N_P^0 + \Delta N_P = 29834.400 + 0.048 = 29834.048$$
$$E_P = E_P^0 + \Delta E_P = 13677.000 + 0.566 = 13677.566$$

These are the "new" approximate coordinates for P. A further iteration will show that the corrections to these values are less than 0.5 mm, hence the values above could be regarded as exact.

8. THE STANDARD ERROR ELLIPSE

After a *Least Squares Adjustment* of survey data, intersection, resection or a combination of both, using the method of *Variation of Coordinates,* the cofactor matrix $\mathbf{Q}_{xx} = \mathbf{N}^{-1}$ contains estimates of the variances and covariances of the adjusted quantities. These precision estimates, variances s_E^2, s_N^2 and covariance s_{EN} can be used to define a geometric figure known as the *Standard Error Ellipse,* which is a useful graphical representation of the precision of a position fix. Poor or "weak" fixes are indicated by narrow elongated ellipses and good or "strong" position fixes are indicated by near circular ellipses.

Error ellipses may be computed for points before any observations are made provided the approximate locations of points (fixed and floating) are known. Observations (directions, bearings and distances) may be scaled from maps and diagrams and an approximate set of normal equations formed. The inverse of the coefficient matrix **N** yields all the information required for the computation of the parameters of the error ellipses. In such cases, error ellipses are an important analysis tool for the surveyor in planning survey operations

8.1. The Pedal Curve of the Standard Error Ellipse

Consider a point whose precision estimates, variances s_E^2, s_N^2 and covariance s_{EN} are known. The variance in any other direction *u* may be calculated by considering the projection of *E* and *N* onto the *u*-axis, which is rotated anti-clockwise from the *E*-axis by an angle ϕ

$$\begin{aligned} u &= E\cos\phi + N\sin\phi \\ &= f\left(E, N\right) \end{aligned} \qquad (8.1)$$

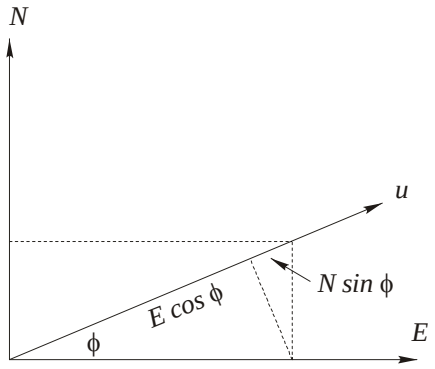


Figure 8.1

Applying the law of propagation of variances to equation (8.1) gives an expression for variance in the *u*-direction

$$s_u^2 = s_E^2 \left(\frac{\partial f}{\partial E} \right)^2 + s_N^2 \left(\frac{\partial f}{\partial N} \right)^2 + 2 s_{EN} \frac{\partial f}{\partial E} \frac{\partial f}{\partial N} \tag{8.2}$$

The partial derivatives $\frac{\partial f}{\partial E} = \cos\phi, \frac{\partial f}{\partial N} = \sin\phi$ are obtained from (8.1) to give an equation for the variance s_u^2 in a direction ϕ (positive anti-clockwise) from the *E*-axis.

$$s_u^2 = s_E^2 \cos^2\phi + s_N^2 \sin^2\phi + 2 s_{EN} \cos\phi \sin\phi \tag{8.3}$$

Equation (8.3) defines the *pedal curve* of the *Standard Error Ellipse*

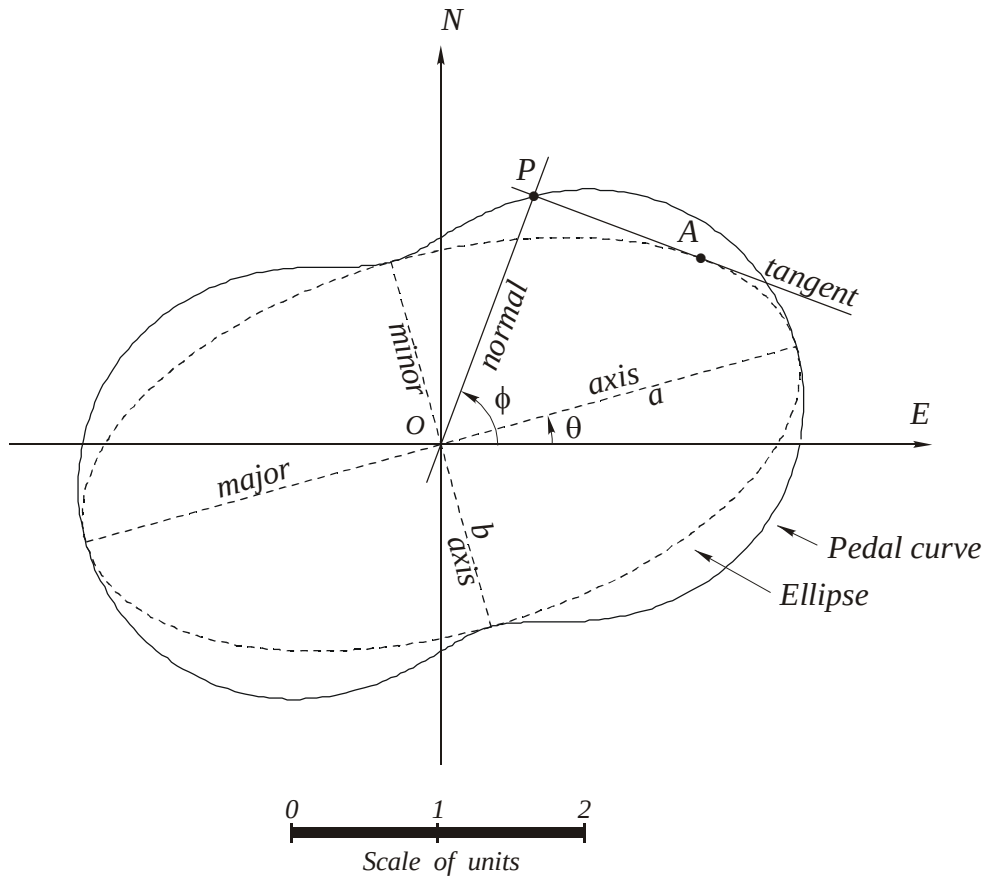


Figure 8.2. The pedal curve of the Standard Error Ellipse

In Figure 8.2, *A* is a point on an ellipse. The tangent to the ellipse at *A* intersects a normal to the tangent passing through *O* at *P*. As *A* moves around the ellipse, the locus of all points *P* is

the pedal curve of the ellipse. The distance $OP = \sqrt{s_u^2}$ for the angle ϕ. The maximum and minimum values of s_u^2 define the directions and lengths of the axes of the ellipse and the following section details the equations linking variances s_E^2, s_N^2 and covariance s_{EN} with the Standard Error Ellipse parameters a, b and θ.

8.2. Parameters of the Standard Error Ellipse

Equation (8.3) has maximum and minimum values defining the lengths and directions of the axes of the error ellipse. To determine these values from (8.3) the trigonometric identities $1-\cos 2\phi = 2\sin^2\phi$, $1+\cos 2\phi = 2\cos^2\phi$, $\sin 2\phi = 2\sin\phi\cos\phi$ can be used to give

$$\begin{aligned} s_u^2 &= \frac{1}{2}s_E^2\left(1+\cos 2\phi\right) + \frac{1}{2}s_N^2\left(1-\cos 2\phi\right) + s_{EN}\sin 2\phi \\ &= \frac{1}{2}\left(s_E^2 + s_N^2\right) + \frac{1}{2}\left(s_E^2 - s_N^2\right)\cos 2\phi + s_{EN}\sin 2\phi \end{aligned}$$

Letting $A = \frac{1}{2}\left(s_E^2 - s_N^2\right)$ and $B = s_{EN}$ this expression has the general form

$$s_u^2 = \frac{1}{2}\left(s_E^2 + s_N^2\right) + A\cos 2\phi + B\sin 2\phi \tag{8.4}$$

Equation (8.4) can be expressed as a trigonometric addition

$$\begin{aligned} s_u^2 &= \frac{1}{2}\left(s_E^2 + s_N^2\right) + R\cos\left(2\phi - \alpha\right) \\ &= \frac{1}{2}\left(s_E^2 + s_N^2\right) + R\cos 2\phi\cos\alpha + R\sin 2\phi\sin\alpha \end{aligned} \tag{8.5}$$

Equating the coefficients of $\cos 2\phi$ and $\sin 2\phi$ in equations (8.4) and (8.5) gives $R\cos\alpha = A$ and $R\sin\alpha = B$ from which we obtain

$$\begin{aligned} R &= \sqrt{A^2 + B^2} \\ &= \sqrt{\tfrac{1}{4}\left(s_E^2 - s_N^2\right)^2 + \left(s_{EN}\right)^2} \\ &= \frac{1}{2}\sqrt{\left(s_E^2 - s_N^2\right)^2 + 4\left(s_{EN}\right)^2} \\ &= \frac{1}{2}W \end{aligned} \tag{8.6}$$

where $$W = \sqrt{\left(s_E^2 - s_N^2\right)^2 + 4\left(s_{EN}\right)^2} \tag{8.7}$$

and the angle α from

$$\tan\alpha = \frac{B}{A} = \frac{2s_{EN}}{s_E^2 - s_N^2} \tag{8.8}$$

Inspection of equation (8.5) shows that $s_u^2 = \frac{1}{2}\left(s_E^2 + s_N^2\right) + R\cos\left(2\phi - \alpha\right)$ will have a maximum value when $\left(2\phi - \alpha\right) = 0$ i.e., $\cos(0) = 1$ and a minimum value when $\left(2\phi - \alpha\right) = \pi$ i.e., $\cos(\pi) = -1$ or

$$\begin{aligned} s_u^2(\max) &= \frac{1}{2}\left(s_E^2 + s_N^2\right) + R = \frac{1}{2}\left(s_E^2 + s_N^2 + W\right) \\ s_u^2(\min) &= \frac{1}{2}\left(s_E^2 + s_N^2\right) - R = \frac{1}{2}\left(s_E^2 + s_N^2 - W\right) \end{aligned} \tag{8.9}$$

Inspection of Figure 8.2 shows that the maximum and minimum values of s_u^2 are in the directions of the major and minor axes of the Standard Error Ellipse and the semi-axes lengths are

$$\begin{aligned} a &= \sqrt{\frac{1}{2}\left(s_E^2 + s_N^2 + W\right)} \\ b &= \sqrt{\frac{1}{2}\left(s_E^2 + s_N^2 - W\right)} \end{aligned} \tag{8.10}$$

The value of ϕ when s_u^2 is a maximum is when $\left(2\phi - \alpha\right) = 0$, i.e., when $\alpha = 2\phi$ thus from equation (8.8), letting $\theta = \phi$ when s_u^2 is a maximum, the angle θ, measured anti-clockwise from the *E*-axis to the major axis of the Standard Error Ellipse, is given by

$$\tan 2\theta = \frac{2s_{EN}}{s_E^2 - s_N^2} \tag{8.11}$$

Note that s_u^2 is a minimum when $2\phi - \alpha = \pi$, i.e., when $\alpha = 2\phi - \pi$ thus from equation (8.8), letting $\theta = \phi$ and recognizing that $\tan\left(x - \pi\right) = \tan x$, then $\tan 2\theta = 2s_{EN} \big/ \left(s_E^2 - s_N^2\right)$ which is the same equation for the angle to the major axis. Hence, it is not possible to distinguish between the angles to the major or minor axes and the ambiguity must be resolved by using equation (8.3).

Alternatively, the parameters of the Standard Error Ellipse can be determined from equation (8.3) by the methods outlined in Chapter 2 (Sction 2.7.2 Least Squares Best Fit Ellipse). Consider equation (8.3) expressed as

$$f = s_E^2 \cos^2\phi + s_N^2 \sin^2\phi + 2s_{EN}\cos\phi\sin\phi \tag{8.12}$$

and the aim is to find the maximum and minimum values of *f* (the maximum and minimum variances) and the values of ϕ when these occur by investigating the first and second derivatives f' and f'' respectively, i.e.,

$$f \text{ is } \left\{\begin{matrix}\max\\ \min\end{matrix}\right\} \text{ when } \left\{\begin{matrix} f' = 0 \text{ and } f'' < 0\\ f' = 0 \text{ and } f'' > 0\end{matrix}\right\}$$

where

$$\begin{aligned} f' &= \left(s_N^2 - s_E^2\right)\sin 2\phi + 2s_{EN}\cos 2\phi \\ f'' &= 2\left(s_N^2 - s_E^2\right)\cos 2\phi - 4s_{EN}\sin 2\phi \end{aligned} \tag{8.13}$$

Now the maximum or minimum value of *f* occurs when $f' = 0$ and from the first member of (8.13) the value of ϕ is given by

$$\tan 2\phi = \frac{2s_{EN}}{s_E^2 - s_N^2} \tag{8.14}$$

But this value of ϕ could relate to either a maximum or a minimum value of *f*. So from the second member of equations (8.13) with a value of 2ϕ from equation (8.14) this ambiguity can be resolved by determining the sign of the second derivative f'' since it is known that

$$\left\{\begin{matrix} f_{\max}\\ f_{\min}\end{matrix}\right\} \text{ when } \left\{\begin{matrix} f'' < 0\\ f'' > 0\end{matrix}\right\}$$

In the equation of the pedal curve of the Standard Error Ellipse given by equation (8.12) $f_{\max}$ coincides with $s_{\max}^2$ and $f_{\min}$ coincides with $s_{\min}^2$ so the angle θ (measured positive anti-clockwise) from the *E*-axis to the major axis of the ellipse (see Figure 8.2) is found from

$$\left\{\begin{matrix} s_{\max}^2\\ s_{\min}^2\end{matrix}\right\} \text{ when } \left\{\begin{matrix} f'' < 0\\ f'' > 0\end{matrix}\right\} \text{ and } \left\{\begin{matrix} \theta = \phi\\ \theta = \phi - \frac{1}{2}\pi\end{matrix}\right\}$$

Substituting $\phi = \theta$ and $\phi = \theta + \frac{1}{2}\pi$ into equation (8.3) will give the max. and min. values of the variance which are the lengths of the semi axes *a* and *b* of the Standard Error Ellipse.

8.3. Example Computation

In Figure 8.2, $s_E^2 = 6.0$, $s_N^2 = 2.0$ and $s_{EN} = 1.2$

The lengths of the semi-axes of the Standard Error Ellipse are

$$\begin{aligned} W &= \sqrt{\left(s_E^2 - s_N^2\right)^2 + 4\left(s_{EN}\right)^2} \\ &= \sqrt{4^2 + 4\left(1.2\right)^2} \\ &= 4.6648 \end{aligned}$$

$$a = \sqrt{\frac{1}{2}\left(s_E^2 + s_N^2 + W\right)} = 2.5164$$

$$b = \sqrt{\frac{1}{2}\left(s_E^2 + s_N^2 - W\right)} = 1.2914$$

The angle between the *E*-axis and the major axis (positive anti-clockwise), noting the quadrant signs to determine the proper quadrant of 2θ

$$\begin{aligned} \tan 2\theta &= \frac{2s_{EN}}{s_E^2 - s_N^2} = \frac{2\left(1.2\right)}{6-2} = \left(\frac{+}{+}\right) \\ 2\theta &= 30^\circ\, 57' 50'' \\ \theta &= 15^\circ\, 28' 55'' \end{aligned}$$

Substituting the values $\phi = \theta = 15^\circ\, 28' 55''$ and $\phi = 90^\circ + \theta = 105^\circ\, 28' 55''$ into equation (8.3) gives $s_u = 2.5164$ and $s_u = 1.2914$ respectively so $\theta = 15^\circ\, 28' 55''$ is the angle (positive anti-clockwise) between the *E*-axis and the major axis. Hence the bearing of the major axis is $90^\circ - \theta = 74^\circ\, 31' 05''$

Alternatively, using the method of evaluating the second derivative we have from equation (8.14)

$$\tan 2\phi = \frac{2s_{EN}}{s_E^2 - s_N^2} = \frac{2(1.2)}{6-2} = \left(\frac{+}{+}\right)$$
$$2\phi = 30^\circ\, 57' 50''$$
$$\phi = 15^\circ\, 28' 55''$$

The second derivative, from the second member of equations (8.13) is

$$\begin{aligned} f'' &= 2\left(s_N^2 - s_E^2\right)\cos 2\phi - 4s_{EN}\sin 2\phi \\ &= 2(2-6)\cos\left(30^\circ\, 57' 50''\right) - 4(1.2)\sin\left(30^\circ\, 57' 50''\right) \\ &= -9.3295 \end{aligned}$$

Now, since $f'' < 0$ then $\theta = \phi = 15^\circ\, 28' 55''$ is the angle (positive anti-clockwise) from the *E*-axis to the major axis of the ellipse. The bearing of the major axis is $90^\circ - \theta = 74^\circ\, 31' 05''$.

The Standard Error Ellipse semi-axes lengths *a* and *b* are obtained from equation (8.3) with $\phi = \theta = 15^\circ\, 28' 55''$ and $\phi = \theta + \frac{1}{2}\pi = 105^\circ\, 28' 55''$ respectively giving

$\phi = 15^\circ\, 28' 55''$ $\qquad s_{\max}^2 = 6.3324$ and $a = \sqrt{s_{\max}^2} = 2.5164$

$\phi = 105^\circ\, 28' 55''$ $\qquad s_{\min}^2 = 1.6677$ and $b = \sqrt{s_{\min}^2} = 1.2914$

8.4. Some Examples of Resections and Error Ellipses

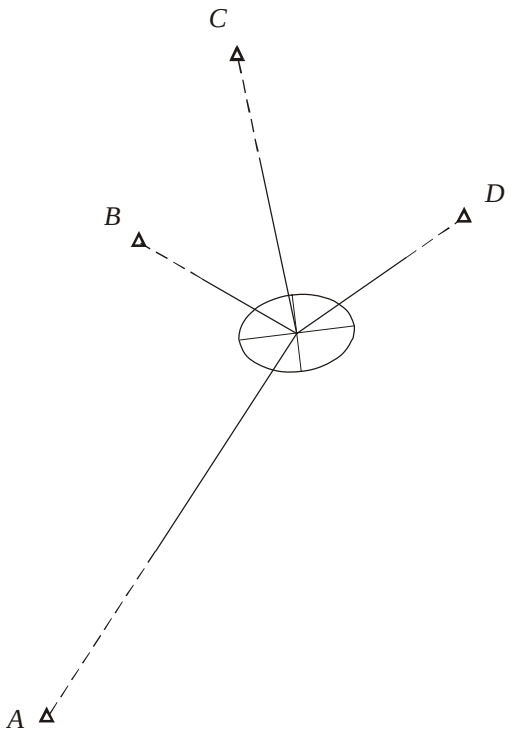


Figure 8.3. Resection 1

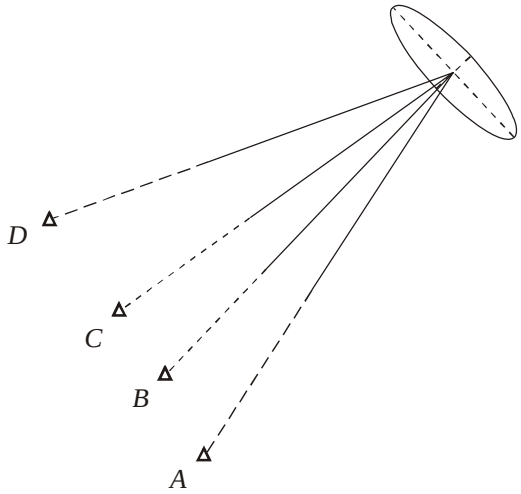
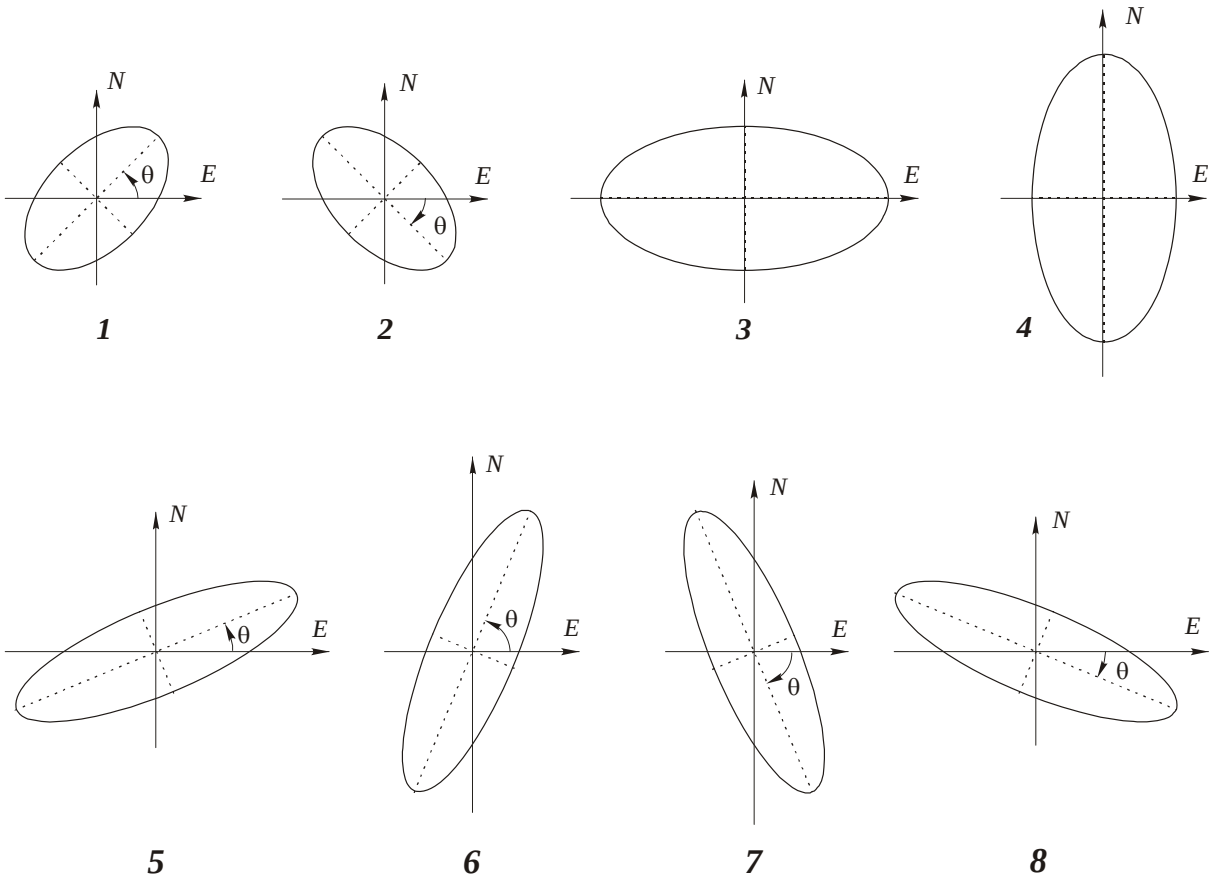


Figure 8.4. Resection 2

In Resection 1 (Figure 8.3) the error ellipse indicates a strong position fix and the observed stations are spread through an arc of approximately 200°.

In Resection 2 (Figure 8.4) the error ellipse indicates a poor position fix. The observed stations lay in a small arc of approximately 35°.

8.5. Some Examples of Orientation and Shape of Ellipses



Ellipse	s_E^2	s_N^2	s_{EN}	ρ_{EN}	a	b	θ
1	4	4	2	0.5	2.45	1.41	45°
2	4	4	-2	-0.5	2.45	1.41	-45°
3	16	4	0	0	4	2	0°
4	4	16	0	0	4	2	90°
5	16	4	6	0.75	4.30	1.23	22° 30′
6	4	16	6	0.75	4.30	1.23	67° 30′
7	4	16	-6	-0.75	4.30	1.23	-67° 30′
8	16	4	-6	-0.75	4.3	1.23	-22° 30′

Note: $\rho_{EN} = \dfrac{s_{EN}}{s_E s_N}$ is the correlation coefficient $-1 \le \rho_{EN} \le 1$

8.6. References

Additional information on error ellipses can be found in the following references.

Cooper, M.A.R., 1982. *Fundamentals of Survey Measurement and Analysis,* Granada, London.

Mikhail, E.M., 1976. *Observations and Least Squares,* IEP–A Dun-Donnelley, New York.

Mikhail, E.M. and Gracie, G., 1981. *Analysis and Adjustment of Survey Measurements,* Van Nostrand Reinhold Company, New York.

Richardus, P., 1966. *Project Surveying,* North-Holland Publishing Company, Amsterdam.

9. LEAST SQUARES RESECTION

A resection is a mathematical technique for determining the plane coordinates of an "unknown" point by observing theodolite (or Total Station) directions to three or more "known" points; known points being those whose coordinates are known. Directions to three known points are the minimum requirement for a solution, which may be obtained from geometric principles. There are many techniques for geometric solution, two of which have been studied in the practical projects associated with this course; (i) the Collins Point method and (ii) the method of Auxiliary angles.

In the case of four or more observed directions to known points, the method of *least squares* (least squares adjustment of indirect observations) may be employed to obtain the best estimates of the coordinates of the resected point. This technique requires the formation of a set of observation equations that yield "normal equations" that are solved for the best estimates of the coordinates of the resected point. Owing to the nature of the observation equation, which is a linearized approximation, the least squares solution process is iterative. That is, approximate values are assumed, corrections computed and approx values updated; with the process repeated until the corrections to approximate vales become negligible. This least squares technique is often called *Variation of Coordinates.*

9.1. The Resection Observation Equation

In the case of four or more observed directions to known points, the method of *least squares* may be employed to obtain the best estimates of the coordinates of the resected point. This technique requires the formation of a set of observation equations; each equation based on a linearized form of the following equation whose elements are shown in Figure 9.1. (see Chapter 7, Section 7.3 for details regarding the linearization process using Taylor's theorem).

$$\alpha_k + v_k + z = \tan^{-1}\left(\frac{E_k - E_P}{N_k - N_P}\right) \tag{9.1}$$

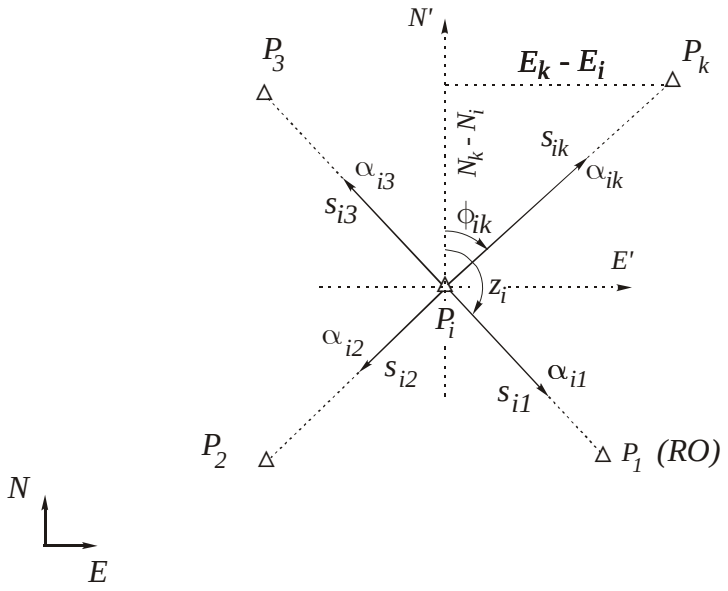


Figure 9.1 Observed directions α and distances *s* from P_i

where

α_k are the observed directions from the resection point *P* to the known points P_k,

v_k are the residuals (small corrections) associated with observed directions,

z is an orientation "constant"; the bearing of the Reference Object (RO) for the set of observed directions,

E_k, N_k are the east and north coordinates of the known points, and

E_P, N_P are the east and north coordinates of the resection point.

Equation (9.1) is a non-linear equation, which has a linear approximation of the form

$$\alpha_k + v_k + z^0 + \Delta z = -a_k \Delta N_P - b_k \Delta E_P + \phi_k^0 \tag{9.2}$$

where a_k, b_k are *direction coefficients* (see Chapter 7, Section 7.3)

$$\begin{aligned} a_k &= \frac{-\left(E_k - E_P^0\right)}{\left(s_k^0\right)^2} = \frac{-\sin\phi_k^0}{s_k^0} \\ b_k &= \frac{N_k - N_P^0}{\left(s_k^0\right)^2} = \frac{\cos\phi_k^0}{s_k^0} \end{aligned} \tag{9.3}$$

$\Delta E,\ \Delta N,\ \Delta z$ are small corrections to approximate coordinates $E_P^0,\ N_P^0$ of the resected point and the approximate orientation constant z^0 such that

$$\begin{aligned} E_P &= E_P^0 + \Delta E \\ N_P &= N_P^0 + \Delta N \\ z &= z^0 + \Delta z \end{aligned} \tag{9.4}$$

$\phi_k^0,\ s_k^0$ are approximate bearings and distances obtained by substituting the approximate coordinates E_P^0, N_P^0 into

$$\phi_k^0 = \tan^{-1}\left(\frac{E_k - E_P^0}{N_k - N_P^0}\right) \tag{9.5}$$

$$s_k^0 = \sqrt{\left(E_k - E_P^0\right)^2 + \left(N_k - N_P^0\right)^2} \tag{9.6}$$

9.2. Formation of the Observation Equations into Standard Matrix Form

The observation equation (9.2) can be re-arranged into a "standard" form

$$v_k + a_k\,\Delta N + b_k\,\Delta E + \Delta z = \phi_k^0 - \left(\alpha_k + z^0\right) \tag{9.7}$$

A set of *n* such equations can be represented in matrix form as

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} a_1 & b_1 & 1 \\ a_2 & b_2 & 1 \\ \vdots & & \vdots \\ a_n & b_n & 1 \end{bmatrix} \begin{bmatrix} \Delta N \\ \Delta E \\ \Delta z \end{bmatrix} = \begin{bmatrix} \phi_1^0 - \left(\alpha_1 + z^0\right) \\ \phi_2^0 - \left(\alpha_2 + z^0\right) \\ \vdots \\ \phi_n^0 - \left(\alpha_n + z^0\right) \end{bmatrix}$$

or $$\mathbf{v} + \mathbf{B}\mathbf{x} = \mathbf{f} \tag{9.8}$$

where
- *n* is the number of observed directions (= to the number of equations)
- *u* is the number of "unknowns" (in the case of a resection *u* = 3)
- **v** is an (*n*,1) vector of residuals
- **B** is an (*n*,*u*) coefficient matrix containing the direction coefficients and ones
- **x** is a (*u*,1) vector of "unknowns" which are corrections to approximate coordinates and orientation constant

f is an (*n*,1) vector of numeric terms which are "computed bearing − observed bearing"

9.3. Formation of the Normal Equations and Solution of the Unknowns

In any least squares adjustment, every measurement (or observation) has an associated precision (a variance) and a measure of connection with every other measurement (covariances). These statistics are contained in a *covariance matrix* $\mathbf{\Sigma}$. The elements of a covariance matrix are population statistics and in practice, the covariance matrix is estimated a priori by a *cofactor matrix* **Q**. Covariance matrices and cofactor matrices are related by $\mathbf{\Sigma} = \sigma_0^2\mathbf{Q}$ where σ_0^2 is the *variance factor*. An estimate of the variance factor $\hat{\sigma}_0^2$ may be computed after the adjustment. In least squares theory it is often useful to express the relative precision of observations in terms of *weights*, where a weight is defined as being inversely proportional to a variance, this leads to the definition of a weight matrix as the inverse of a cofactor matrix, i.e. $\mathbf{W} = \mathbf{Q}^{-1}$. Weight matrices, covariance matrices and cofactor matrices are square and symmetric (see Chapter 2, Section 2.5).

Applying the least squares principle to equation (9.8) with the precisions of the observations estimated by a weight matrix leads to a set of *normal equations* of the form

$$\left(\mathbf{B}^T\mathbf{W}\mathbf{B}\right)\mathbf{x} = \mathbf{B}^T\mathbf{W}\mathbf{f}$$

or
$$\mathbf{N}\mathbf{x} = \mathbf{t} \tag{9.9}$$

where **W** is an (*n*,*n*) weight matrix

$\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B}$ is a (*u*,*u*) coefficient matrix of the set of normal equations

$\mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f}$ is a (*u*,1) vector of numeric terms

The solution of the *n* = 3 unknowns is

$$\mathbf{x} = \begin{bmatrix} \Delta N \\ \Delta E \\ \Delta z \end{bmatrix} = \mathbf{N}^{-1}\mathbf{t} \tag{9.10}$$

9.4. The form of the Coefficient Matrix N and the Vector of Numeric Terms t

For *n* observations, the coefficient matrix **B** and the vector of numeric terms **f** have the following form

$$\mathbf{B} = \begin{bmatrix} a_1 & b_1 & 1 \\ a_2 & b_2 & 1 \\ a_3 & b_3 & 1 \\ \vdots & & \\ a_n & b_n & 1 \end{bmatrix}, \qquad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_n \end{bmatrix} = \begin{bmatrix} \phi_1^0 - \left(\alpha_1 + z^0\right) \\ \phi_2^0 - \left(\alpha_2 + z^0\right) \\ \phi_3^0 - \left(\alpha_3 + z^0\right) \\ \vdots \\ \phi_n^0 - \left(\alpha_n + z^0\right) \end{bmatrix}$$

Each observed direction has an associated variance σ^2 estimated by s^2 where σ is the standard deviation (a population statistic) and *s* is its estimate. The observed directions are assumed independent, hence the covariances between observations are zero and the covariance matrix, cofactor matrix and weight matrix are all diagonal matrices. The cofactor matrix **Q** and the weight matrix **W** have the following form

$$\begin{aligned} \mathbf{Q} &= \begin{bmatrix} s_1^2 & 0 & 0 & \cdots & 0 \\ 0 & s_2^2 & 0 & \cdots & 0 \\ 0 & 0 & s_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_n^2 \end{bmatrix} \\ \mathbf{W} = \mathbf{Q}^{-1} &= \begin{bmatrix} 1/s_1^2 & 0 & 0 & \cdots & 0 \\ 0 & 1/s_2^2 & 0 & \cdots & 0 \\ 0 & 0 & 1/s_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/s_n^2 \end{bmatrix} = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n \end{bmatrix} \end{aligned} \tag{9.11}$$

The normal equation coefficient matrix **N** and vector of numeric terms **t** have the following form

$$\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B} = \begin{bmatrix} \sum_{k=1}^{n} w_k a_k^2 & \sum_{k=1}^{n} w_k a_k b_k & \sum_{k=1}^{n} w_k a_k \\ \sum_{k=1}^{n} w_k a_k b_k & \sum_{k=1}^{n} w_k b_k^2 & \sum_{k=1}^{n} w_k b_k \\ \sum_{k=1}^{n} w_k a_k & \sum_{k=1}^{n} w_k b_k & \sum_{k=1}^{n} w_k \end{bmatrix}, \quad \mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f} = \begin{bmatrix} \sum_{k=1}^{n} w_k a_k f_k \\ \sum_{k=1}^{n} w_k b_k f_k \\ \sum_{k=1}^{n} w_k f_k \end{bmatrix} \tag{9.12}$$

Note that if each element of the coefficient matrix **B** and vector of numeric terms **f** is divided by the appropriate estimate of the standard deviation *augmented matrices* $\bar{\mathbf{B}}$ and $\bar{\mathbf{f}}$ can be formed

$$\bar{\mathbf{B}} = \begin{bmatrix} a_1/s_1 & b_1/s_1 & 1/s_1 \\ a_2/s_2 & b_2/s_2 & 1/s_2 \\ a_3/s_3 & b_3/s_3 & 1/s_3 \\ \vdots & & \\ a_n/s_n & b_n/s_n & 1/s_n \end{bmatrix}, \qquad \bar{\mathbf{f}} = \begin{bmatrix} f_1/s_1 \\ f_2/s_2 \\ f_3/s_3 \\ \vdots \\ f_n/s_n \end{bmatrix} = \begin{bmatrix} \left(\phi_1^0 - \alpha_1 + z^0\right)/s_1 \\ \left(\phi_2^0 - \alpha_2 + z^0\right)/s_2 \\ \left(\phi_3^0 - \alpha_3 + z^0\right)/s_3 \\ \vdots \\ \left(\phi_n^0 - \alpha_n + z^0\right)/s_n \end{bmatrix} \tag{9.13}$$

and the normal equation coefficient matrix **N** and vector of numeric terms **t** are given by

$$\begin{aligned} \mathbf{N} &= \mathbf{B}^T\mathbf{W}\mathbf{B} = \bar{\mathbf{B}}^T\bar{\mathbf{B}} \\ \mathbf{t} &= \mathbf{B}^T\mathbf{W}\mathbf{f} = \bar{\mathbf{B}}^T\bar{\mathbf{f}} \end{aligned} \tag{9.14}$$

9.5. The Iterative Solution and Assessment of Precision

The elements of the solution vector $\mathbf{x} = \begin{bmatrix} \Delta N & \Delta E & \Delta z \end{bmatrix}^T$ are corrections to the approximate coordinates E_P^0, N_P^0 and the approximate orientation constant z^0. These corrections are added to the approximate values to obtain updated values of the approximations and another iteration performed.

When the corrections for the k[th] iteration reach some desired value, say less than 0.5 mm, then the current "approximate" values may be regarded as exact and the solution is complete.

At the end of the iterative process, the residuals **v** are computed from equation (9.8) and an assessment of the "quality" of the observations can be made. Large residuals may indicate poor observations.

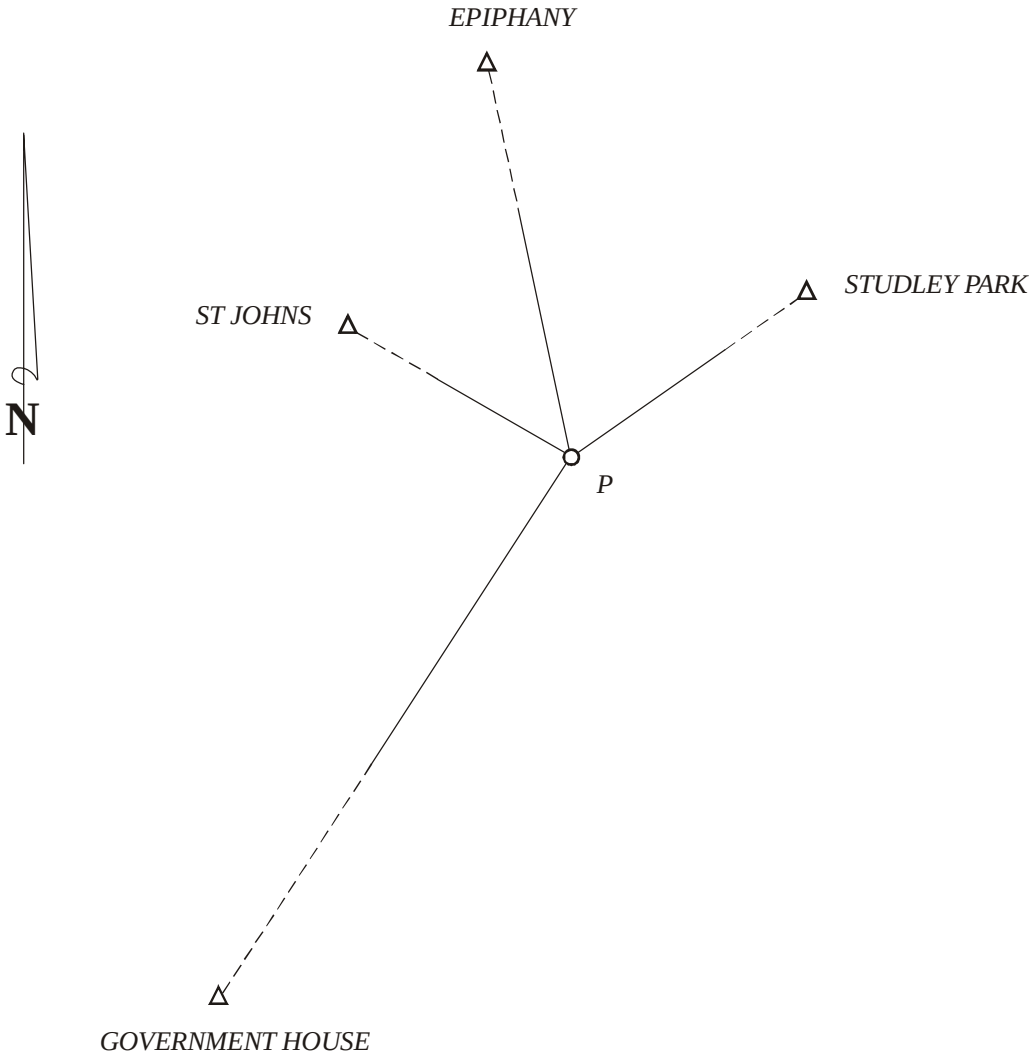
An estimate of the variance factor $\hat{\sigma}_0^2$ can be computed from the residuals after the adjustment process by using the following equation (see Chapter 5, equation (5.25))

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T\mathbf{W}\mathbf{v}}{n-u} = \frac{\mathbf{f}^T\mathbf{W}\mathbf{f} - \mathbf{x}^T\mathbf{t}}{n-u} \tag{9.15}$$

Note that in (9.15) the variance factor can be computed without calculating the residuals.

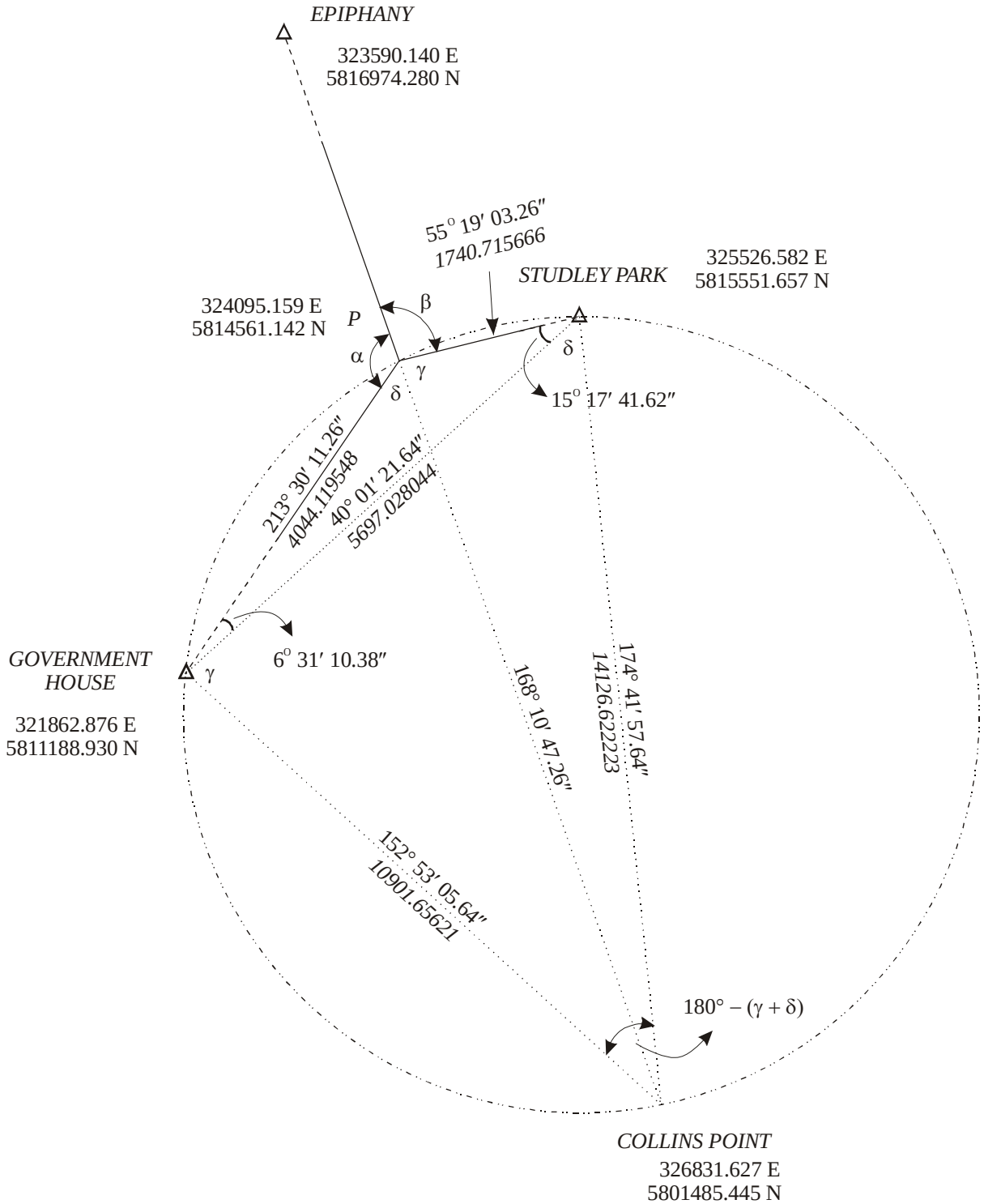
The following pages show an example resection. Page 8 shows a diagram of the resection observations to four known stations. Page 9 shows a geometric solution for the coordinates of *P* using the Collins Point technique and pages 10 to 14 show the calculations required for a least squares solution using all four observations.

9.6. Example Resection: Diagram and Observations



STATION	COORDINATES	OBSERVED DIRECTION
Government House (Flagpole)	321862.876 E 5811188.930 N	0° 00′ 00″
Saint. Johns (Spire)	322731.700 E 5815369.270 N	87° 09′ 09″
Epiphany (Flagpole)	323590.140 E 5816974.280 N	134° 40′ 36″
Studley Park (Beacon)	325526.582 E 5815551.657 N	201° 48′ 52″

9.7. Example Resection: Collins Point Solution



$\alpha = 134° 40′ 36″$ $\quad \delta = 180 - \alpha = 45° 19′ 24″$

$\beta = 67° 08′ 16″$ $\quad \gamma = 180 - \beta = 112° 51′ 44″$

9.8. Example Resection: Least Squares Solution

The data for this resection is shown on the preceding pages. The coordinates of *P* are computed from the Collins Point resection and are rounded to the nearest 0.1 metres.

Observations and computed data

	Coordinates		Observed Directions			Computed Bearing and Distance			
Station	**East**	**North**	Deg	Min	Sec	Deg	Min	Sec	Dist (m)
P	324095.200	5814561.100							
Gov't House	321862.876	5811188.930	0	00	00	213	30	14.18	4044.107
St. Johns	322731.700	5815369.270	87	09	09	300	39	21.44	1585.015
Epiphany	323590.140	5816974.280	134	40	36	348	10	44.62	2465.466
Studley Park	325526.582	5815551.657	201	48	52	55	18	56.40	1740.706

Computation of Direction coefficients and numeric terms for the observation equation

$$v_k + a_k\,\Delta N + b_k\,\Delta E + \Delta z = \phi_k^0 - \left(\alpha_k + z^0\right) \tag{9.16}$$

Note 1: In equation (9.16) the numeric terms on the right-hand-side are computed bearing - "observed" bearing. The observed bearings are obtained by adding the observed directions to the computed bearing of the RO (Government House).

Note 2: The dimensions (or units) of the numeric terms on the right-hand-side of (9.16) are seconds of arc. This means that the elements on the left-hand-side must have consistent dimensions, i.e. v_k (seconds), a_k, b_k (seconds/length), $\Delta E, \Delta N$ (length) and Δz (seconds). If the corrections to approximate coordinates are expressed in centimetres (cm), then the direction coefficients have dimensions of sec/cm and equations (9.3) are

$$\begin{aligned} a_k &= \frac{-\left(E_k - E_P^0\right)}{\left(s_k^0\right)^2} \times \rho'' = \frac{-\sin\phi_k^0}{s_k^0} \times \rho'' \\ b_k &= \frac{N_k - N_P^0}{\left(s_k^0\right)^2} \times \rho'' \qquad = \frac{\cos\phi_k^0}{s_k^0} \times \rho'' \end{aligned} \tag{9.17}$$

where distances and coordinate differences are in cm's and $\rho'' = \frac{180}{\pi} \times 3600$ (seconds in one radian).

	Direction coefficients		Computed Bearing			Observed Bearing			Numeric term
Station	**a** (sec/cm)	**b** (sec/cm)	Deg	Min	Sec	Deg	Min	Sec	comp-obs (sec)
Gov't House	0.281538	-0.425294	213	30	14.18	213	30	14.18	0.00
St. Johns	1.119473	0.663531	300	39	21.44	300	39	23.18	-1.74
Epiphany	0.171384	0.818873	348	10	44.62	348	10	50.18	-5.56
Studley Park	-0.974383	0.674301	55	18	56.40	55	19	06.18	-9.78

Coefficient matrix B, vector of numeric terms f and normal equation coefficient matrix

$\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B}$

$$\mathbf{B} = \begin{array}{c} \begin{matrix} \Delta N & \Delta E & \Delta z \end{matrix} \\ \begin{bmatrix} 0.281538 & -0.425294 & 1 \\ 1.119473 & 0.663531 & 1 \\ 0.171384 & 0.818873 & 1 \\ -0.974383 & 0.674301 & 1 \end{bmatrix} \end{array} \qquad \mathbf{f} = \begin{bmatrix} 0.00 \\ -1.74 \\ -5.56 \\ -9.78 \end{bmatrix}$$

In this example, the observations are assumed to be of equal precision. In such cases the weight matrix **W** can be replaced by the Identity matrix **I**, and the normal equation coefficient matrix $\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B} = \mathbf{B}^T\mathbf{I}\mathbf{B} = \mathbf{B}^T\mathbf{B}$ and the vector of numeric terms $\mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f} = \mathbf{B}^T\mathbf{I}\mathbf{f} = \mathbf{B}^T\mathbf{f}$

$$\mathbf{N} = \mathbf{B}^T\mathbf{B} \begin{bmatrix} 2.311280 & 0.106383 & 0.598012 \\ 0.106383 & 1.746384 & 1.731411 \\ 0.598012 & 1.731411 & 4.000000 \end{bmatrix} \qquad \mathbf{t} = \mathbf{B}^T\mathbf{f} = \begin{bmatrix} 6.630754 \\ -12.302480 \\ -17.081376 \end{bmatrix}$$

Inverse of normal equation coefficient matrix $\mathbf{N}^{-1}$ and solution vector $\mathbf{x} = \mathbf{N}^{-1}\mathbf{t}$

$$\mathbf{N}^{-1} = \begin{bmatrix} 0.454844 & 0.069562 & -0.098111 \\ 0.069562 & 1.013710 & -0.449187 \\ -0.098111 & -0.449187 & 0.459100 \end{bmatrix} \qquad \mathbf{x} = \mathbf{N}^{-1}\mathbf{t} = \begin{bmatrix} 3.8360 \text{ (cm)} \\ -4.3372 \text{ (cm)} \\ -2.9665 \text{ (sec)} \end{bmatrix}$$

Corrections to approximate coordinates and correction to approximate orientation constant

$\Delta E = -4.3372$ cm $= -0.043$ m, $\Delta N = 3.8360$ cm $= 0.038$ m and $\Delta z = -2.9665$ sec $= -3.0$ sec

Adjusted coordinates of *P* and residuals (after one iteration)

$$E_P = E_P^0 + \Delta E = 324095.200 - 0.043 = 324095.157 \text{ m}$$
$$N_P = N_P^0 + \Delta N = 5814561.100 + 0.038 = 5814561.138 \text{ m}$$

Station	**Residual**
Government House	0.04 (sec)
St. Johns	-0.19
Epiphany	0.30
Studley Park	-0.15

Variance factor σ_0^2 and precision of computed coordinates

An estimate of the variance factor $\hat{\sigma}_0^2$ can be computed from equation (9.15) with **W** = **I**, $n = 4$, $u = 3$ and

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T\mathbf{v}}{n-u} = \frac{\mathbf{f}^T\mathbf{f} - \mathbf{x}^T\mathbf{t}}{1} = 129.620678 - 129.465415 = 0.155263 \text{ sec}^2$$

Assuming all the observations are of equal precision and letting **W** = **I** is equivalent to assigning an estimated standard deviation of 1 second to each observation. Inspection of the variance factor shows that if a standard deviation of 0.39 sec ($0.39 = \sqrt{0.155263}$) was used as an estimate of the standard deviation of the observed directions, the estimate of the variance factor computed from the adjustment would have been approximately unity (a variance factor of unity indicates that the estimates of variances are close to the population statistics).

From this adjustment, we may conclude that the standard deviation of the observed directions was approximately 0.4 sec.

A most important "by-product" of a least squares adjustment is the ability to estimate the precision of the computed quantities. Theory shows that this information is contained in the inverse of the normal equations and the covariance matrix of the computed quantities $\mathbf{\Sigma}_{xx}$ is given by

$$\mathbf{\Sigma}_{xx} = \sigma_0^2 \mathbf{Q}_{xx} = \sigma_0^2 \mathbf{N}^{-1} = 0.155263 \begin{bmatrix} 0.454844 & 0.069562 & -0.098111 \\ 0.069562 & 1.013710 & -0.449187 \\ -0.098111 & -0.449187 & 0.459100 \end{bmatrix} = \begin{bmatrix} 0.070 & 0.011 & -0.015 \\ 0.011 & 0.157 & -0.070 \\ -0.015 & -0.070 & 0.071 \end{bmatrix}$$

and the standard deviation of the adjusted coordinates are

$$\sigma_E = \sqrt{0.157} \text{ cm} = 0.004 \text{ m}$$
$$\sigma_N = \sqrt{0.070} \text{ cm} = 0.003 \text{ m}$$

Parameters of Standard Error Ellipse

From the variance matrix $\mathbf{\Sigma}_{xx}$ above we have $\sigma_E^2 = 0.157$, $\sigma_N^2 = 0.070$ and $\sigma_{EN} = 0.011$

Using the formulae from Chapter 8, Section 8.3 (replacing *s* with σ), the lengths of the semi-axes of the Standard Error Ellipse are

$$\begin{aligned} W &= \sqrt{\left(\sigma_E^2 - \sigma_N^2\right)^2 + 4\left(\sigma_{EN}\right)^2} \\ &= \sqrt{\left(0.157 - 0.070\right)^2 + 4\left(0.011\right)^2} \\ &= 0.0897 \end{aligned}$$

$$a = \sqrt{\frac{1}{2}\left(\sigma_E^2 + \sigma_N^2 + W\right)} = 0.3980 \text{ cm}$$

$$b = \sqrt{\frac{1}{2}\left(s_E^2 + s_N^2 - W\right)} = 0.2620 \text{ cm}$$

The angle between the *E*-axis and the major axis (positive anti-clockwise), noting the quadrant signs to determine the proper quadrant of 2θ

$$\tan 2\theta = \frac{2\sigma_{EN}}{\sigma_E^2 - \sigma_N^2} = \frac{2(0.011)}{0.157 - 0.070} = \left(\frac{+}{+}\right)$$
$$2\theta = 14^\circ 11' 28''$$
$$\theta = 7^\circ 05' 44''$$

Substituting the values $\phi = \theta = 7^\circ 05' 44''$ and $\phi = 90^\circ + \theta = 97^\circ 05' 44''$ into equation (8.3) gives $s_u = 0.3980$ cm and $s_u = 0.2620$ cm respectively so $\theta = 7^\circ 05' 44''$ is the angle (positive anti-clockwise) between the *E*-axis and the major axis. Hence the bearing of the major axis is $90^\circ - \theta = 82^\circ 54' 16''$

Figure 9.2 shows a schematic diagram of the example resection and the Standard Error Ellipse

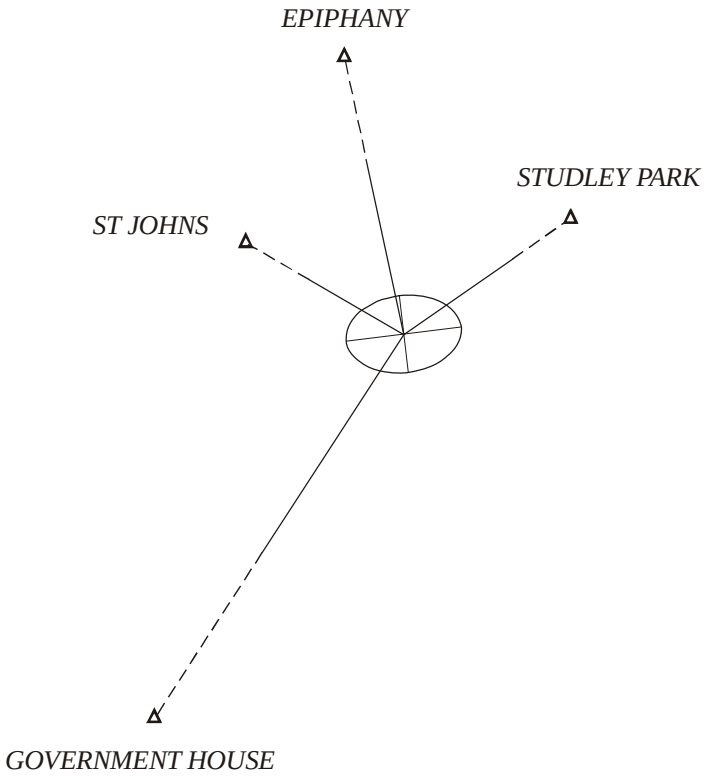


Figure 9.2 Schematic diagram of resection and Standard Error Ellipse

10. LEAST SQUARES BEARING INTERSECTION

Intersection is a mathematical technique for determining the plane coordinates of an "unknown" point by observing bearings (clockwise angles from North) to the unknown point from two or more "known" points; known points being those whose coordinates *E,N* are known. Figure 10.1 shows an unknown point *P* intersected by bearings $\phi_A, \phi_B, \phi_C, \phi_D$ observed from known stations *A, B, C* and *D.* Bearings from two known points are the minimum requirement for a solution, which may be obtained from geometric principles set out below. In the case of three or more observed bearings from known points, the method of *least squares* may be employed to obtain the best estimates of the coordinates of the intersected point *P*. This technique requires the formation of a set of observation equations that yield "normal equations" that are solved for the best estimates of the coordinates of *P*. Owing to the nature of the observation equation, which is a linearized approximation, the least squares solution process is iterative. That is, approximate values are assumed, corrections computed and approx values updated; with the process repeated until the corrections to approximate vales become negligible.

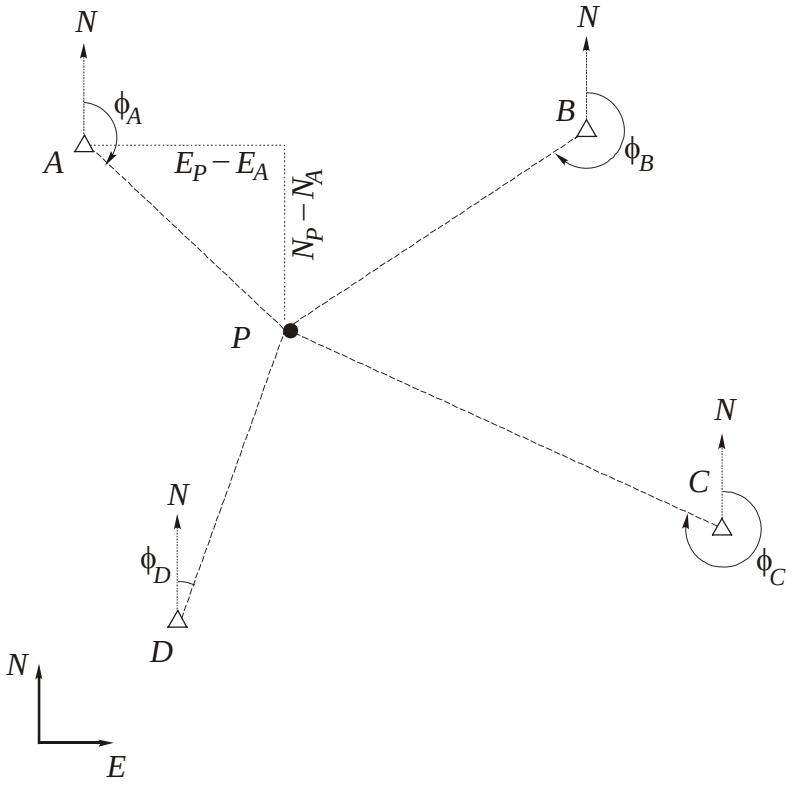


Figure 10.1. Unknown point *P* intersected by bearings from known points *A,B,C* and *D*

10.1. Position from Two Observed Bearings

From Figure 10.1, using the bearings ϕ_A, ϕ_B to the unknown point *P* from known points *A* and *B* with coordinates E_A, N_A and E_B, N_B the following two equations can be obtained

$$\begin{aligned} \tan\phi_A &= \frac{E_P - E_A}{N_P - N_A} \\ \tan\phi_B &= \frac{E_P - E_B}{N_P - N_B} \end{aligned} \tag{10.1}$$

Expanding these equations and re-arranging gives

$$\begin{aligned} E_P &= N_P \tan\phi_A - N_A \tan\phi_A + E_A \\ E_P &= N_P \tan\phi_B - N_B \tan\phi_B + E_B \end{aligned} \tag{10.2}$$

Equating equations (10.2) gives a solution for N_P

$$N_P = \frac{N_A \tan\phi_A - N_B \tan\phi_B + E_B - E_A}{\tan\phi_A - \tan\phi_B} \tag{10.3}$$

Having obtained a solution for N_P from (10.3) then E_P can be obtained from either of equations (10.2).

It should be noted that if *P* lies on the line between *A* and *B* then its position is indeterminate.

10.2. The Least Squares Bearing Intersection Observation Equation

In the case of four or more observed directions to known points, the method of *least squares* may be employed to obtain the best estimates of the coordinates of the resected point. This technique requires the formation of a set of observation equation; each equation based on a linearized form of the following equation whose elements are shown in Figure 10.1.

$$\phi_k + v_k = \tan^{-1}\left(\frac{E_P - E_k}{N_P - N_k}\right) \tag{10.4}$$

where ϕ_k are the observed bearings known points P_k to the unknown point *P*,

v_k are the residuals (small corrections) associated with observed bearings,

E_k , N_k are the east and north coordinates of the known points, and

E_P , N_P are the east and north coordinates of the intersection point *P*.

Equation (10.4) is a non-linear equation, which has a linear approximation of the form

$$\phi_k + v_k = a_k \, \Delta N_P + b_k \, \Delta E_P + \phi_k^0 \tag{10.5}$$

where a_k, b_k are *direction coefficients* (see Chapter 7, Section 7.3)

$$\begin{aligned} a_k &= \frac{-\left(E_P^0 - E_k\right)}{\left(s_k^0\right)^2} = \frac{-\sin\phi_k^0}{s_k^0} \\ b_k &= \frac{N_P^0 - N_k}{\left(s_k^0\right)^2} = \frac{\cos\phi_k^0}{s_k^0} \end{aligned} \tag{10.6}$$

$\Delta E, \Delta N$ are small corrections to approximate coordinates E_P^0, N_P^0 of the intersected point such that

$$\begin{aligned} E_P &= E_P^0 + \Delta E \\ N_P &= N_P^0 + \Delta N \end{aligned} \tag{10.7}$$

ϕ_k^0, s_k^0 are approximate bearings and distances obtained by substituting the approximate coordinates E_P^0, N_P^0 into

$$\phi_k^0 = \tan^{-1}\left(\frac{E_P^0 - E_k}{N_P^0 - N_k}\right) \tag{10.8}$$

$$s_k^0 = \sqrt{\left(E_P^0 - E_k\right)^2 + \left(N_P^0 - N_k\right)^2} \tag{10.9}$$

10.3. Formation of the Observation Equations into Standard Matrix Form

The observation equation (10.5) can be re-arranged into a "standard" form

$$v_k - a_k \, \Delta N - b_k \, \Delta E = \phi_k^0 - \phi_k \tag{10.10}$$

A set of *n* such equations can be represented in matrix form as

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} -a_1 & -b_1 \\ -a_2 & -b_2 \\ \vdots & \\ -a_n & -b_n \end{bmatrix} \begin{bmatrix} \Delta N \\ \Delta E \end{bmatrix} = \begin{bmatrix} \phi_1^0 - \phi_1 \\ \phi_2^0 - \phi_2 \\ \vdots \\ \phi_n^0 - \phi_n \end{bmatrix}$$

or $$\mathbf{v} + \mathbf{Bx} = \mathbf{f} \qquad (10.11)$$

where
- n is the number of observed bearings (= to the number of equations)
- u is the number of "unknowns" (in the case of an intersection $u = 2$)
- $\mathbf{v}$ is an $(n,1)$ vector of residuals
- $\mathbf{B}$ is an (n,u) coefficient matrix containing the direction coefficients
- $\mathbf{x}$ is a $(u,1)$ vector of "unknowns" which are corrections to approximate coordinates
- $\mathbf{f}$ is an $(n,1)$ vector of numeric terms; "computed bearing – observed bearing"

10.4. Formation of the Normal Equations and Solution of the Unknowns

In any least squares adjustment, every measurement (or observation) has an associated precision (a variance) and a measure of connection with every other measurement (covariances). These statistics are contained in a *covariance matrix* $\mathbf{\Sigma}$. The elements of a covariance matrix are population statistics and in practice, the covariance matrix is estimated a priori by a *cofactor matrix* $\mathbf{Q}$. Covariance matrices and cofactor matrices are related by $\mathbf{\Sigma} = \sigma_0^2\mathbf{Q}$ where σ_0^2 is the *variance factor*. An estimate of the variance factor $\hat{\sigma}_0^2$ may be computed after the adjustment. In least squares theory it is often useful to express the relative precision of observations in terms of *weights*, where a weight is defined as being inversely proportional to a variance, this leads to the definition of a weight matrix as the inverse of a cofactor matrix, i.e. $\mathbf{W} = \mathbf{Q}^{-1}$. Weight matrices, covariance matrices and cofactor matrices are square and symmetric (see Chapter 2, Section 2.5).

Applying the least squares principle to equation (10.8) with the precisions of the observations estimated by a weight matrix leads to a set of *normal equations* of the form

$$\left(\mathbf{B}^T\mathbf{WB}\right)\mathbf{x} = \mathbf{B}^T\mathbf{Wf}$$

or $$\mathbf{Nx} = \mathbf{t} \qquad (10.12)$$

where $\mathbf{W}$ is an (n,n) weight matrix

$\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B}$ is a (u,u) coefficient matrix of the set of normal equations

$\mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f}$ is a $(u,1)$ vector of numeric terms

The solution of the $n = 2$ unknowns is

$$\mathbf{x} = \begin{bmatrix} \Delta N \\ \Delta E \end{bmatrix} = \mathbf{N}^{-1}\mathbf{t} \tag{10.13}$$

10.5. The form of the Coefficient Matrix N and the Vector of Numeric Terms t

For n observations, the coefficient matrix $\mathbf{B}$ and the vector of numeric terms $\mathbf{f}$ have the following form

$$\mathbf{B} = \begin{bmatrix} -a_1 & -b_1 \\ -a_2 & -b_2 \\ -a_3 & -b_3 \\ \vdots & \\ -a_n & -b_n \end{bmatrix}, \qquad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_n \end{bmatrix} = \begin{bmatrix} \phi_1^0 - \phi_1 \\ \phi_2^0 - \phi_2 \\ \phi_3^0 - \phi_3 \\ \vdots \\ \phi_n^0 - \phi_n \end{bmatrix}$$

Each observed bearing has an associated variance σ^2 estimated by s^2 where σ is the standard deviation and s is its estimate. The observed bearings are assumed independent, hence the covariances between observations are zero and the covariance matrix, cofactor matrix and weight matrix are all diagonal matrices. The cofactor matrix $\mathbf{Q}$ and the weight matrix $\mathbf{W}$ have the following form

$$\mathbf{Q} = \begin{bmatrix} s_1^2 & 0 & 0 & \cdots & 0 \\ 0 & s_2^2 & 0 & \cdots & 0 \\ 0 & 0 & s_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_n^2 \end{bmatrix}$$

and

$$\mathbf{W} = \mathbf{Q}^{-1} = \begin{bmatrix} 1/s_1^2 & 0 & 0 & \cdots & 0 \\ 0 & 1/s_2^2 & 0 & \cdots & 0 \\ 0 & 0 & 1/s_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/s_n^2 \end{bmatrix} = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n \end{bmatrix} \tag{10.14}$$

The normal equation coefficient matrix **N** and vector of numeric terms **t** have the following form

$$\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B} = \begin{bmatrix} \sum_{k=1}^{n} w_k a_k^2 & \sum_{k=1}^{n} w_k a_k b_k \\ \sum_{k=1}^{n} w_k a_k b_k & \sum_{k=1}^{n} w_k b_k^2 \end{bmatrix}, \quad \mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f} = \begin{bmatrix} \sum_{k=1}^{n} w_k a_k f_k \\ \sum_{k=1}^{n} w_k b_k f_k \end{bmatrix} \tag{10.15}$$

Note that if each element of the coefficient matrix **B** and vector of numeric terms **f** is divided by the appropriate estimate of the standard deviation *augmented matrices* $\overline{\mathbf{B}}$ and $\overline{\mathbf{f}}$ can be formed

$$\overline{\mathbf{B}} = \begin{bmatrix} -a_1/s_1 & -b_1/s_1 \\ -a_2/s_2 & -b_2/s_2 \\ -a_3/s_3 & -b_3/s_3 \\ \vdots & \\ -a_n/s_n & -b_n/s_n \end{bmatrix}, \qquad \overline{\mathbf{f}} = \begin{bmatrix} f_1/s_1 \\ f_2/s_2 \\ f_3/s_3 \\ \vdots \\ f_n/s_n \end{bmatrix} = \begin{bmatrix} \left(\phi_1^0 - \phi_1\right)/s_1 \\ \left(\phi_2^0 - \phi_2\right)/s_2 \\ \left(\phi_3^0 - \phi_3\right)/s_3 \\ \vdots \\ \left(\phi_n^0 - \phi_n\right)/s_n \end{bmatrix} \tag{10.16}$$

and the normal equation coefficient matrix **N** and vector of numeric terms **t** are given by

$$\begin{aligned} \mathbf{N} &= \mathbf{B}^T\mathbf{W}\mathbf{B} = \overline{\mathbf{B}}^T\overline{\mathbf{B}} \\ \mathbf{t} &= \mathbf{B}^T\mathbf{W}\mathbf{f} = \overline{\mathbf{B}}^T\overline{\mathbf{f}} \end{aligned} \tag{10.17}$$

10.6. The Iterative Solution and Assessment of Precision

The elements of the solution vector $\mathbf{x} = \begin{bmatrix} \Delta N & \Delta E \end{bmatrix}^T$ are corrections to the approximate coordinates $E_P^0,\ N_P^0$. These corrections are added to the approximate values to obtain updated values of the approximations and another iteration performed.

When the corrections for kth iteration reach some desired value, say less than 0.5 mm, then the current "approximate" values may be regarded as exact and the solution is complete.

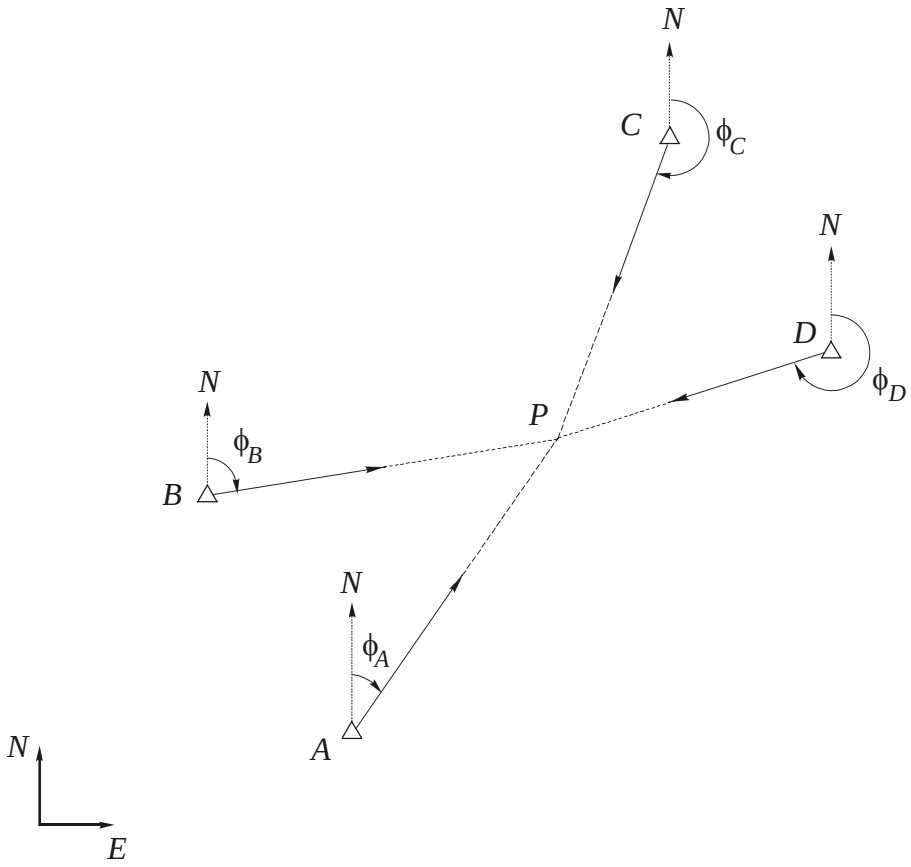
At the end of the iterative process, residuals **v** are computed from equation (10.11) and an assessment of the "quality" of the observations can be made. Large residuals may indicate poor observations.

The variance factor σ_0^2 can be computed from the residuals after the adjustment process by using the following

$$\sigma_0^2 = \frac{\mathbf{v}^T\mathbf{W}\mathbf{v}}{n-u} = \frac{\mathbf{f}^T\mathbf{W}\mathbf{f} - \mathbf{x}^T\mathbf{t}}{n-u} \tag{10.18}$$

Note that in (10.18) the variance factor can be computed without calculating the residuals.

10.7. Example Intersection: Diagram and Observations



STATION	COORDINATES	OBSERVED BEARING
A	12875.270 E 28679.600 N	34° 47′ 52″
B	12273.910 E 29612.310 N	81° 01′ 23″
C	14117.390 E 30999.980 N	200° 40′ 18″
D	14717.690 E 30168.700 N	252° 09′ 35″

10.8. Example Intersection: Least Squares Solution

The data for this intersection is shown on the previous page. The coordinates of *P* are computed from equations (10.3) and (10.2) then rounded down to the nearest 0.1 metres.

Observations and computed data

	Coordinates		Observed Bearing			Computed Bearing and Distance			
Station	**East**	**North**	Deg	Min	Sec	Deg	Min	Sec	Dist
P	13677.500	29834.000							
A	12875.270	28679.600	34	47	52	34	47	48.08	1405.778
B	12273.910	29612.310	81	01	23	81	01	28.43	1420.990
C	14117.390	30999.980	200	40	18	200	40	12.06	1246.199
D	14717.690	30168.700	252	09	35	252	09	48.52	1092.712

Computation of Direction coefficients and numeric terms for the observation equation

$$v_k - a_k\,\Delta N - b_k\,\Delta E = \phi_k^0 - \phi_k \qquad (10.19)$$

Note 1 In equation (10.19) the numeric terms on the right-hand-side are computed bearing minus observed bearing.

Note 2 The dimensions (or units) of the numeric terms on the right-hand-side of (10.19) are seconds of arc. This means that the elements on the left-hand-side must have consistent dimensions, i.e., v_k (seconds), a_k, b_k (seconds/length) and $\Delta E, \Delta N$ (length). If the corrections to approximate coordinates are expressed in centimetres (cm), then the direction coefficients have dimensions of sec/cm and equations (10.6) are

$$a_k = \frac{-\left(E_P^0 - E_k\right)}{\left(s_k^0\right)^2} \times \rho'' = \frac{-\sin\phi_k^0}{s_k^0} \times \rho''$$

$$b_k = \frac{N_P^0 - N_k}{\left(s_k^0\right)^2} \times \rho'' \qquad = \frac{\cos\phi_k^0}{s_k^0} \times \rho''$$

where distances and coordinate differences are in cm's and $\rho'' = \frac{180}{\pi} \times 3600$ (seconds in one radian).

	Direction coefficients		Computed Bearing			Observed Bearing			Numeric term
Station	**a** (sec/cm)	**b** (sec/cm)	Deg	Min	Sec	Deg	Min	Sec	comp-obs (sec)
A	0.837318	-1.204891	34	47	48.08	34	47	52	-3.92
B	1.433784	-0.226459	81	01	28.43	81	01	23	5.43
C	-0.584244	1.548607	200	40	12.06	200	40	18	-5.94
D	-1.796911	0.578189	252	09	48.52	252	09	35	13.52

Coefficient matrix B, vector of numeric terms f, normal equation coefficient matrix

$\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B}$ and numeric terms $\mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f}$

$$\mathbf{B} = \begin{array}{c} \begin{matrix} \Delta N & \Delta E \end{matrix} \\ \begin{bmatrix} 0.837318 & -1.204891 \\ 1.433784 & -0.226459 \\ -0.584244 & 1.548607 \\ -1.796911 & 0.578189 \end{bmatrix} \end{array} \qquad \mathbf{f} = \begin{bmatrix} -3.92 \\ 5.43 \\ -5.94 \\ 13.52 \end{bmatrix}$$

In this example, the observations are assumed to be of equal precision. In such cases the weight matrix **W** can be replaced by the Identity matrix **I**, and the normal equation coefficient matrix $\mathbf{N} = \mathbf{B}^T\mathbf{W}\mathbf{B} = \mathbf{B}^T\mathbf{I}\mathbf{B} = \mathbf{B}^T\mathbf{B}$ and the vector of numeric terms $\mathbf{t} = \mathbf{B}^T\mathbf{W}\mathbf{f} = \mathbf{B}^T\mathbf{I}\mathbf{f} = \mathbf{B}^T\mathbf{f}$

$$\mathbf{N} = \mathbf{B}^T\mathbf{B} \begin{bmatrix} 6.327065 & -3.277288 \\ -3.277288 & 4.235533 \end{bmatrix} \qquad \mathbf{t} = \mathbf{B}^T\mathbf{f} = \begin{bmatrix} -16.323890 \\ 2.114714 \end{bmatrix}$$

Inverse of normal equation coefficient matrix $\mathbf{N}^{-1}$ and solution vector $\mathbf{x} = \mathbf{N}^{-1}\mathbf{t}$

$$\mathbf{N}^{-1} = \begin{bmatrix} 0.263767 & 0.204092 \\ 0.204092 & 0.394016 \end{bmatrix} \qquad \mathbf{x} = \mathbf{N}^{-1}\mathbf{t} = \begin{bmatrix} -3.8741 \text{ (cm)} \\ -2.4983 \text{ (cm)} \end{bmatrix}$$

Corrections to approximate coordinates

$\Delta E = -2.4983$ cm $= -0.025$ m , $\Delta N = -3.8741$ cm $= -0.039$ m

Adjusted coordinates of *P* and residuals (after one iteration)

$$E_P = E_P^0 + \Delta E \quad = 13677.500 - 0.025 = 13677.475 \text{ m}$$
$$N_P = N_P^0 + \Delta N = 29834.000 - 0.039 = 29833.961 \text{ m}$$

Station	**Residual**
A	-3.68 (sec)
B	10.42
C	-4.33
D	8.01

Variance factor σ_0^2 and precision of computed coordinates

An estimate of the variance factor $\hat{\sigma}_0^2$ can be computed from equation (10.18) with **W** = **I**, $n = 4$, $u = 2$ and

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T\mathbf{v}}{n-u} = \frac{\mathbf{f}^T\mathbf{f} - \mathbf{x}^T\mathbf{t}}{4-2} = \frac{262.925300 - 57.957192}{2} = 102.484054 \text{ sec}^2$$

Assuming all the observations are of equal precision and letting **W** = **I** is equivalent to assigning an estimated standard deviation of 1 second to each observation. Inspection of the variance factor shows that if a standard deviation of 10.12 sec ($10.12 = \sqrt{102.484054}$) was used as an estimate of the standard deviation of the observed directions, the estimate of the variance factor computed from the adjustment would have been approximately unity (a variance factor of unity indicates that the estimates of variances are close to the population statistics).

From this adjustment, we may conclude that the standard deviation of the observed directions was approximately 10.1 sec.

A most important "by-product" of a least squares adjustment is the ability to estimate the precision of the computed quantities. Theory shows that this information is contained in the inverse of the normal equations and the covariance matrix of the computed quantities $\mathbf{\Sigma}_{xx}$ is given by

$$\boldsymbol{\Sigma}_{xx} = \sigma_0^2 \mathbf{Q}_{xx} = \sigma_0^2 \mathbf{N}^{-1} = 102.484054 \begin{bmatrix} 0.263767 & 0.204092 \\ 0.204092 & 0.394016 \end{bmatrix} = \begin{bmatrix} 27.0319 & 20.9162 \\ 20.9162 & 40.3804 \end{bmatrix}$$

and the standard deviation of the adjusted coordinates are

$$\sigma_E = \sqrt{40.3804} \text{ cm} = 0.064 \text{ m}$$
$$\sigma_N = \sqrt{27.0319} \text{ cm} = 0.052 \text{ m}$$

Parameters of Standard Error Ellipse

From the variance matrix $\boldsymbol{\Sigma}_{xx}$ above we have

$\sigma_E^2 = 40.3804$, $\sigma_N^2 = 27.0319$ and $\sigma_{EN} = 20.9162$

Using the formulae from Chapter 8, Section 8.3 (replacing *s* with σ), the lengths of the semi-axes of the Standard Error Ellipse are

$$\begin{aligned} W &= \sqrt{\left(\sigma_E^2 - \sigma_N^2\right)^2 + 4\left(\sigma_{EN}\right)^2} \\ &= \sqrt{\left(40.3804 - 27.0319\right)^2 + 4\left(20.9162\right)^2} \\ &= 43.9105 \end{aligned}$$

$$a = \sqrt{\frac{1}{2}\left(\sigma_E^2 + \sigma_N^2 + W\right)} = 7.4607 \text{ cm}$$
$$b = \sqrt{\frac{1}{2}\left(\sigma_E^2 + \sigma_N^2 - W\right)} = 3.4280 \text{ cm}$$

The angle between the *E*-axis and the major axis (positive anti-clockwise), noting the quadrant signs to determine the proper quadrant of 2θ

$$\tan 2\theta = \frac{2\sigma_{EN}}{\sigma_E^2 - \sigma_N^2} = \frac{2\left(20.9162\right)}{40.3804 - 27.0319} = \left(\frac{+}{+}\right)$$
$$2\theta = 72^\circ\, 18' 09''$$
$$\theta = 36^\circ\, 09' 04''$$

Substituting the values $\phi = \theta = 36^\circ\, 09' 04''$ and $\phi = 90^\circ + \theta = 126^\circ\, 09' 04''$ into equation (8.3) gives $s_u = 7.4607$ cm and $s_u = 3.4280$ cm respectively so $\theta = 36^\circ\, 09' 04''$ is the angle

(positive anti-clockwise) between the *E*-axis and the major axis. Hence the bearing of the major axis is $90^\circ - \theta = 53^\circ\, 50' 56''$

Figure 10.2 shows a schematic diagram of the example intersection and the Standard Error Ellipse

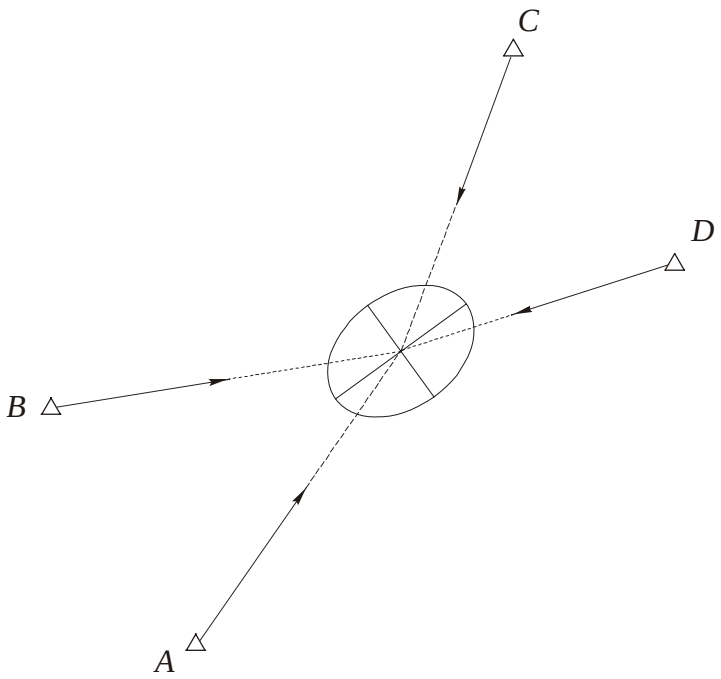


Figure 10.2 Schematic diagram of intersection and Standard Error Ellipse

REFERENCES

Crandall, K.C. and Seabloom, R.W., 1970, *Engineering fundamentals in measurements, probability, statistics and dimensions,* McGraw-Hill, New York.

Cross, P.A. 1992, *Advanced Least Squares Applied to Position Fixing,* Working Paper No. 6, Department of Land Information, University of East London.

Deakin, R.E. and Kildea, D.G., 1999, 'A note on standard deviation and RMS', *The Australian Surveyor,* Vol. 44, No. 1, pp. 74-79, June 1999.

Gauss, K.F. 1809, *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections,* a translation of *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium* by C.H. Davis, Dover, New York, 1963.

Johnson, N.L. and Leone, F.C., 1964, *Statistics and Experimental Design In Engineering and the Physical Sciences,* Vol. I, John Wiley & Sons, Inc., New York

Krakiwsky, E.J. 1975, *A Synthesis of Recent Advances in the Method of Least Squares,* Lecture Notes No. 42, 1992 reprint, Department of Surveying Engineering, University of New Brunswick, Fredericton, Canada

Kreyszig, Erwin, 1970, *Introductory Mathematical Statistics,* John Wiley & Sons, New York.

Leahy, F.J. 1974, 'Two hundred years of adjustment of survey measurements', *Two Centuries of Surveying: Proceedings of the 17th Australian Survey Congress,* Melbourne, 23 Feb. – 4 Mar. 1974, Institution of Surveyors, Australia, pp.19-29.

Merriman, M. 1905, *Method of Least Squares,* 8th edn, John Wiley & Sons, New York.

Mikhail, E.M. 1976, *Observations and Least Squares,* IEP–A Dun-Donnelley, New York.

Mikhail, E.M. and Gracie, G. 1981. *Analysis and Adjustment of Survey Measurements.* Van Nostrand Reinhold, New York, 340 pages.

Rainsford, H.F., 1968, *Survey adjustments and least Squares*, Constable, London.

Tienstra, J.M., 1966, *Theory of adjustment of normally distributed observations*, Argus, Amsterdam

Walpole, R.E., 1974, *Introduction to Statistics*, 2nd edn, Macmillan Publishing Co., Inc. New York.

Wells, D.E. and Krakiwsky, E.J. 1971. *The Method of Least Squares*. Lecture Notes No. 18, Department of Surveying Engineering, University of New Brunswick, May 1971, reprinted September 1992, 180 pages.

APPENDIX A MATRIX ALGEBRA

A 1 INTRODUCTION

Matrix algebra is a powerful mathematical tool, which is extremely useful in modern computational techniques applicable to spatial information science. It is neither new nor difficult, but prior to the development of the electronic computer was thought to be too cumbersome for practical applications. In today's computer age large masses of data are accumulated and matrix algebra is a convenient and concise way of expressing algorithms and computer routines for the manipulation of data.

The advantages of matrix algebra may be set out as:

1. It provides a systematic means of storing and manipulating large arrays of data. Such data may range from numerical coefficients of equations to characters and symbols related to scanned digital images.
2. It provides a means of reducing large and complicated systems of equations to simple expressions, which can be easily visualised and analysed.
3. It provides a concise method of expressing algorithms and of directing computer execution of those algorithms via computer programs.

A 2 DEFINITIONS

A 2.1 Matrix

A matrix is a set of numbers or symbols arranged in a square or rectangular array of m rows and n columns as

$$\mathbf{A}_{m,n} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

A letter or symbol refers to the whole matrix. In many texts and references, matrices are denoted by **boldface** type, ie,

A X P Q W

Matrices may also be indicated by pacing a tilde (~) under a symbol, ie,

$\underset{\sim}{A}$ $\underset{\sim}{X}$ $\underset{\sim}{P}$ $\underset{\sim}{Q}$ $\underset{\sim}{W}$

Example A 1

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & -6 \\ 5 & 1 & 9 \\ 2 & 2 & -8 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 2 & 5 \\ 3 & 5 \\ -4 & 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} -14 & 22 & 3 & -8 \end{bmatrix}$$

A, **C** and **x** are all matrices. Note that the matrix **x** is a *row matrix* or *row vector*. Row matrices or row vectors are usually denoted by lowercase letters.

A 2.2 Matrix element

Individual elements of a matrix are shown by lowercase letters a_{ij} where the subscripts i and j indicate the element lies at the intersection of the i^{th} row and the j^{th} column. The first subscript always refers to the row number and the second to the column number.

$$\mathbf{A}_{m,n} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & & & & & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & \left(a_{ij}\right) & \cdots & a_{in} \\ \vdots & & & & & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix} \leftarrow \text{row } i$$

$$\text{column } j \downarrow$$

Another way of representing a matrix is by typical element, for example

$$\mathbf{A} = \left\{a_{ij}\right\} \quad \begin{matrix} i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, m \end{matrix}$$

A 2.3 Matrix order

A matrix is said to be of order *m* by *n* (or *m, n*) where *m* is the number of rows and *n* is the number of columns. The order of a matrix may be expressed in various ways ie,

$$\mathbf{A}_{m,n} \quad \mathbf{A}_{(m,n)} \quad \mathbf{A}_{m\times n} \quad \mathbf{A}_{(m\times n)} \quad {}_m\mathbf{A}_n$$

Example A 2

$$\mathbf{A}_{3,3} = \begin{bmatrix} 2 & 2 & -6 \\ 5 & 1 & 9 \\ 2 & 2 & -8 \end{bmatrix} \qquad \mathbf{C}_{3,2} = \begin{bmatrix} 2 & 5 \\ 3 & 5 \\ -4 & 1 \end{bmatrix} \qquad \mathbf{x}_{1,4} = \begin{bmatrix} -14 & 22 & 3 & -8 \end{bmatrix}$$

Matrix **A** is of order (3,3), matrix **C** is (3,2) and **x** is (1,4). If a matrix is of order (1,1), it is called a *scalar.*

A 3 TYPES OF MATRICES

A 3.1 Square Matrix

A square matrix is a matrix with an equal number of rows and columns. A square matrix would be indicated by $\mathbf{A}_{m,m}$ and said to be of order *m.* Square matrices have a principal or leading diagonal whose elements are a_{ij} for $i = j$. In matrix **A** below, order (5,5), elements *a, g, m, s* and *y* lie on the leading diagonal.

$$\mathbf{A}_{5,5} = \begin{bmatrix} \underline{a} & b & c & d & e \\ f & \underline{g} & h & i & j \\ k & l & \underline{m} & n & o \\ p & q & r & \underline{s} & t \\ u & v & w & x & \underline{y} \end{bmatrix}$$

Special cases of square matrices are *symmetric* and *skew-symmetric* which are described below.

A 3.2 Column Matrix or Column Vector

A column matrix or column vector is a matrix composed of only one column. Column vectors are usually designated by lowercase letters, for example

$$\mathbf{b}_{m,1} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

A 3.3 Row Matrix or Row Vector

A row matrix or row vector is a matrix composed of only one row. Row vectors are usually designated by lowercase letters, for example

$$\mathbf{b}_{1,n} = \begin{bmatrix} b_1 & b_2 & b_3 & \cdots & b_n \end{bmatrix}$$

A 3.4 Diagonal Matrix

A diagonal matrix is a square matrix with all "off-diagonal" elements equal to zero

$$\mathbf{D}_{m,m} = \begin{bmatrix} d_{11} & 0 & 0 & \cdots & 0 \\ 0 & d_{22} & 0 & \cdots & 0 \\ 0 & 0 & d_{33} & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & d_{mm} \end{bmatrix} \quad \text{where} \quad d_{ij} = 0 \ \text{ for } \ i \neq j$$

A diagonal matrix may have some diagonal elements equal to zero. A diagonal matrix is often shown in the form

$$\mathbf{D} = \text{diag}\{d_1, \ d_2, \ d_3, \ \cdots, \ d_m\}$$

A 3.5 Scalar Matrix

A scalar matrix is a diagonal matrix whose elements are <u>all</u> equal to the same scalar quantity

$$\mathbf{A} = \begin{bmatrix} a & 0 & 0 & \cdots & 0 \\ 0 & a & 0 & \cdots & 0 \\ 0 & 0 & a & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a \end{bmatrix} \quad \text{where} \quad \begin{array}{l} a_{ij} = 0 \ \text{ for } \ i \neq j \\ a_{ij} = a \ \text{ for } \ i = j \end{array}$$

Example A 3

$$\mathbf{W} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad \text{is a (3,3) scalar matrix}$$

A 3.6 Identity or Unit Matrix

An identity or unit matrix is a diagonal matrix whose elements are all equal to 1 (unity). It is always referred to as **I** where

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Note that all the "off-diagonal" elements are zero and all the elements of the leading diagonal are unity.

A 3.7 Null or Zero Matrix

A null or zero matrix is a matrix whose elements are all zero. It is denoted by boldface **0**.

A 3.8 Triangular Matrix

A triangular matrix is a square matrix whose elements above, or below, but not including the leading diagonal, are all zero. Square matrices whose elements above the leading diagonal are zero are known as *lower triangular matrices*.

$$\mathbf{L}_{m,m} = \begin{bmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ l_{31} & l_{32} & l_{33} & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ l_{m1} & l_{m2} & l_{m3} & \cdots & l_{mm} \end{bmatrix} \quad \text{where} \quad l_{ij} = 0 \ \text{ for } \ i < j$$

Square matrices whose elements below the leading diagonal are zero are known as *upper triangular matrices.*

$$\mathbf{U}_{m,m} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1m} \\ 0 & u_{22} & u_{23} & \cdots & u_{2m} \\ 0 & 0 & u_{33} & \cdots & u_{3m} \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & u_{mm} \end{bmatrix} \quad \text{where} \quad u_{ij} = 0 \ \text{ for } \ i > j$$

Example A 4

$$\mathbf{G} = \begin{bmatrix} 2 & 1 & -2 & 4 & 3 \\ 0 & 5 & 4 & -3 & 2 \\ 0 & 0 & 3 & 2 & 1 \\ 0 & 0 & 0 & 5 & 6 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} 8 & 0 & 0 \\ 3 & 4 & 0 \\ 5 & -2 & 6 \end{bmatrix}$$

G and **H** are both triangular matrices. **G** is an upper triangular matrix and **H** is a lower triangular matrix. Triangular matrices of order *n* have $\left(n^2 + n\right)/2$ non-zero elements.

A 3.9 Unit Lower Triangular Matrix

This is a special case of a lower triangular matrix, in which all the elements of the leading diagonal are equal to unity.

$$\mathbf{L}_{m,m} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{21} & 1 & 0 & \cdots & 0 \\ l_{31} & l_{32} & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ l_{m1} & l_{m2} & l_{m3} & \cdots & 1 \end{bmatrix} \quad \text{where} \quad \begin{array}{l} l_{ij} = 0 \ \text{ for } \ i < j \\ l_{ij} = 1 \ \text{ for } i = j \end{array}$$

A 3.10 Unit Upper Triangular Matrix

This is a special case of an upper triangular matrix, in which all the elements of the leading diagonal are equal to unity.

$$\mathbf{U}_{m,m} = \begin{bmatrix} 1 & u_{12} & u_{13} & \cdots & u_{1m} \\ 0 & 1 & u_{23} & \cdots & u_{2m} \\ 0 & 0 & 1 & \cdots & u_{3m} \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad \text{where} \quad \begin{array}{l} u_{ij} = 0 \ \text{ for } \ i > j \\ u_{ij} = 1 \ \text{ for } \ i = j \end{array}$$

A 3.11 Banded Matrix

A banded matrix is any square matrix in which the only non-zero elements occur in a band about the leading diagonal. Thus, if **A** is to be a banded matrix

$$a_{ij} = 0 \quad \text{when} \quad |i - j| > k$$

A typical banded matrix of order 4 is

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix} \quad \text{where} \quad a_{ij} = 0 \quad \text{for} \quad |i-j| > 1$$

A 4 MATRIX OPERATIONS

A 4.1 Equality

Two matrices **A** and **B** are equal if and only if they are the same order and $a_{ij} = b_{ij}$ for all *i* and *j*. Matrices of different order cannot be equated.

A 4.2 Addition

The sum of two matrices **A** and **B**, of the *same order*, is a matrix **C** of that order whose elements are $c_{ij} = a_{ij} + b_{ij}$ for all *i* and *j*. Matrices of different order cannot be added. The following laws of addition hold true for matrix algebra:

commutative law $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

associative law $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + \mathbf{B} + \mathbf{C}$

Example A 5

$$\mathbf{A}_{3,3} = \begin{bmatrix} 8 & 3 & -1 \\ 3 & 4 & 2 \\ 5 & -2 & 6 \end{bmatrix} \quad \mathbf{B}_{3,3} = \begin{bmatrix} 2 & 2 & -5 \\ 5 & 4 & 3 \\ 6 & -7 & 12 \end{bmatrix} \quad \mathbf{A} + \mathbf{B} = \begin{bmatrix} 10 & 5 & -6 \\ 8 & 8 & 5 \\ 11 & -9 & 18 \end{bmatrix}$$

A 4.3 Scalar Multiplication

Multiplication of a matrix **A** by a scalar *k* is another matrix **B** of the same order whose elements are $b_{ij} = k\, a_{ij}$ for all *i* and *j*.

Example A 6

$$\mathbf{A} = \begin{bmatrix} 8 & 3 & -1 \\ 3 & 4 & 2 \\ 5 & -2 & 6 \end{bmatrix} \quad \mathbf{B} = \tfrac{1}{2}\mathbf{A} = \begin{bmatrix} 4.0 & 1.5 & -0.5 \\ 1.5 & 2.0 & 1.0 \\ 2.5 & -1.0 & 3.0 \end{bmatrix}$$

The following laws relating to scalar multiplication hold true

$$k(\mathbf{A}+\mathbf{B}) = k\mathbf{A} + k\mathbf{B}$$
$$(k+q)\mathbf{A} = k\mathbf{A} + q\mathbf{A}$$
$$k(\mathbf{AB}) = (k\mathbf{A})\mathbf{B} = \mathbf{A}(k\mathbf{B})$$
$$k(q\mathbf{A}) = (kq)\mathbf{A}$$

A 4.4 Scalar Product

In vector algebra, it is customary to denote the scalar product of two vectors $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, **i**, **j** and **k** being unit vectors in the direction of the *x*, *y* and *z* axes respectively, as

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1b_1 + a_2b_2 + a_3b_3$$

In matrix algebra, a set of three simultaneous equations represented as

$$\mathbf{Ax} = \mathbf{b} \quad \text{or} \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

means that each element of the column vector **b** is the scalar product of each row of **A** by the column vector **x**, i.e.,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$$

A 4.5 Matrix Multiplication

For three matrices **A**, **B** and **C** with their respective elements a_{ij}, b_{ij} and c_{ij}, then

$$\mathbf{AB} = \mathbf{C} \quad \text{implies} \quad c_{ij} = \sum_k a_{ik}\, b_{kj}$$

which states that the element in row *i* and column *j* of **C** is equal to the scalar product of row *i* of **A** and column *j* of **B**

It is important to note that for matrix multiplication to be defined the number of columns of the first matrix must be equal to the number of rows of the second matrix.

As a quick method of assessing whether a matrix multiplication is defined, write down the matrices to be multiplied with their associated rows and columns, ie,

$$\mathbf{A}_{4,\underset{\uparrow}{2}}\,\mathbf{B}_{\underset{\uparrow}{2},6} \qquad \text{and check that the "inner numbers" are the same.}$$

If they are the same, then the multiplication is defined and the product matrix has an order equal to the "outer numbers".

$$\mathbf{A}_{\underset{\uparrow}{4},2}\,\mathbf{B}_{2,\underset{\uparrow}{6}} = \mathbf{C}_{4,6}$$

Remember, that in all cases, the first number of the matrix order refers to the number of rows and the second number refers to the number of columns.

Example A 7

$$\mathbf{A}_{3,2} = \begin{bmatrix} 1 & 2 \\ 3 & 0 \\ 6 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{B}_{2,4} = \begin{bmatrix} 5 & 1 & 1 & 3 \\ 2 & 3 & 1 & 2 \end{bmatrix}$$

$$\begin{aligned} \mathbf{C}_{3,4} &= \mathbf{A}_{3,2}\,\mathbf{B}_{2,4} \\ &= \begin{bmatrix} 1 & 2 \\ 3 & 9 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} 5 & 7 & 1 & 3 \\ 2 & 3 & 8 & 2 \end{bmatrix} \\ &= \begin{bmatrix} (1\times5+2\times2) & (1\times7+2\times3) & (1\times1+2\times8) & (1\times3+2\times2) \\ (3\times5+9\times2) & (3\times7+9\times3) & (3\times1+9\times8) & (3\times3+9\times2) \\ (6\times5+4\times2) & (6\times7+4\times3) & (6\times1+4\times8) & (6\times3+4\times2) \end{bmatrix} \\ &= \begin{bmatrix} 9 & 13 & 17 & 7 \\ 33 & 48 & 75 & 27 \\ 38 & 54 & 38 & 26 \end{bmatrix} \end{aligned}$$

In forming the product **AB** we say that **B** has been *pre-multiplied* by **A**, or that **A** has been *post-multiplied* by **B**.

The following relationships regarding matrix multiplication hold:

$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$	with **I** = the Identity matrix
$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} = \mathbf{ABC}$	(associative law)
$\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$	(distributive law)
$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$	(distributive law)

In these relationships above, the sequence of the matrices is strictly preserved. Note that in general, the commutative law of algebra does not hold for matrix multiplication even if multiplication is defined in both orders, ie,

$$\mathbf{AB} \neq \mathbf{BA} \quad \text{in general}$$

Example A 8

A is of order (2,3), **B** is of order (3,2), **AB** is of order (2,2) and **BA** is of order (3,3). Even if both matrices are square and of the same order, the results will in general not be the same when the order of multiplication is reversed.

$$\mathbf{A} = \begin{bmatrix} 4 & 6 & 2 \\ 3 & 1 & 8 \\ 5 & 9 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & 7 & 3 \\ 1 & 9 & 5 \\ 6 & 8 & 4 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} 26 & 98 & 50 \\ 55 & 94 & 46 \\ 61 & 172 & 88 \end{bmatrix} \quad \text{and} \quad \mathbf{BA} = \begin{bmatrix} 44 & 46 & 81 \\ 56 & 60 & 109 \\ 68 & 80 & 104 \end{bmatrix}$$

If the product of two matrices **A** and **B** is equal to the null matrix **0** then it does not follow that either **A** or **B** is zero

Example A 9

$$\mathbf{A} = \begin{bmatrix} -5 & 4 & 3 \\ 3 & 6 & 8 \\ 5 & 2 & 4 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} 2 & -6 \\ 7 & -21 \\ -6 & 18 \end{bmatrix}; \quad \mathbf{AB} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

This differs from ordinary algebra where if, for example, $a \times b = 0$ then either a or b or both a and b are zero.

Some particular results involving diagonal matrices are useful. If **A** is a square matrix and **D** is a diagonal matrix of the same order, then

1. **AD** causes each *column* $\mathbf{A}_j$ of **A** to be multiplied by the corresponding element d_{jj} of **D**.

2. **DA** causes each *row* $\mathbf{A}_i$ of **A** to be multiplied by the corresponding element d_{ii} of **D**.

Example A 10

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{33} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{bmatrix}$$

$$\mathbf{AD} = \begin{bmatrix} \alpha a_{11} & \beta a_{12} & \gamma a_{13} \\ \alpha a_{21} & \beta a_{22} & \gamma a_{33} \\ \alpha a_{31} & \beta a_{32} & \gamma a_{33} \end{bmatrix} \quad \text{and} \quad \mathbf{DA} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ \beta a_{21} & \beta a_{22} & \beta a_{33} \\ \gamma a_{31} & \gamma a_{32} & \gamma a_{33} \end{bmatrix}$$

If a diagonal matrix $\mathbf{D}_{n,n}$ has non-negative elements $d_{ii} \geq 0$ then for $p > 0$

$$\mathbf{D}_{n,n}^{p} = \begin{bmatrix} d_{11}^{p} & 0 & \cdots & 0 \\ 0 & d_{22}^{p} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn}^{p} \end{bmatrix}$$

and for $p > 0$ and $q > 0$

$$\mathbf{D}^{p}\mathbf{D}^{q} = \mathbf{D}^{p+q} \qquad \text{and in particular} \qquad \mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}} = \mathbf{D}$$

A 4.6 Matrix Transpose

The transpose of a matrix **A** of order (*m*,*k*) is a (*k*,*m*) matrix formed from **A** by interchanging rows and columns such that row *i* of **A** becomes column *i* of the transposed matrix. The transpose of **A** is denoted by $\mathbf{A}^{T}$. There are various other notations used to indicate the transpose of a matrix, such as: $\mathbf{A}^{t}$, $\mathbf{A}'$, $\mathbf{A}^{*}$ and $\overline{\mathbf{A}}$.

If $\mathbf{B} = \mathbf{A}^{T}$ then $b_{ij} = a_{ji}$ for all *i* and *j*

Example A 11

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & 3 \\ 1 & -2 & 3 \\ 4 & 0 & 6 \end{bmatrix} \qquad \mathbf{A}^{T} = \begin{bmatrix} 3 & 1 & 4 \\ 2 & -2 & 0 \\ 3 & 3 & 6 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 4 & 5 \\ 1 & -2 \\ 7 & 0 \end{bmatrix} \qquad \mathbf{B}^{T} = \begin{bmatrix} 4 & 1 & 7 \\ 5 & -2 & 0 \end{bmatrix}$$

The following relationships hold true

$$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$$

$$(\mathbf{ABC}\ldots)^T = \ldots\mathbf{C}^T\mathbf{B}^T\mathbf{A}^T$$

$$(k\mathbf{A})^T = k\mathbf{A}^T$$

$$\left(\mathbf{A}^T\right)^T = \mathbf{A}$$

If **D** is a diagonal matrix,	then $\mathbf{D}^T = \mathbf{D}$
If **H** is a scalar matrix,	then $\mathbf{H} = \mathbf{H}^T$
If **I** is the Identity matrix,	then $\mathbf{I} = \mathbf{I}^T$
If **x** is a *column* vector,	then $\mathbf{x}^T\mathbf{x}$ is a non-negative scalar that is equal to the *sum of the squares* of the vector components.
If **x** is a *row* vector,	then $\mathbf{x}\mathbf{x}^T$ is a *symmetric matrix* (square) of the same order as the vector **x**.

A 4.7 Bilinear and Quadratic forms

If **x** is an (*m*,1) vector of variables, **y** an (*n*,1) vector of variables and **A** an (*m*,*n*) matrix of constants, the *scalar* function

$$u = \mathbf{x}^T\mathbf{A}\mathbf{y}$$

is known as a *bilinear form.*

If **x** is an (*m*,1) vector of variables and ***B*** an (*m*,*m*) square matrix of constants, the *scalar* function

$$q = \mathbf{x}^T\mathbf{B}\mathbf{x}$$

is known as a *quadratic form.* An example of a quadratic form is the sum of the squares of the weighted residuals; the function $\varphi = \mathbf{v}^T\mathbf{W}\mathbf{v}$, which is minimised in least squares.

A 4.8 Matrix Inverse

Division is not defined in matrix algebra. In place of division, the inverse $\mathbf{A}^{-1}$ of a square matrix **A** is introduced. This inverse, if it exists, has the property

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

This relationship defines the *Cayley Inverse* for square matrices only. A square matrix whose determinant is zero is *singular* and a singular matrix does not have an inverse. A square

matrix whose determinant is non-zero is *non-singular* and does have an inverse. Furthermore, if the inverse exists it is unique. Rectangular matrices have no determinants and so they are taken to be singular but they may have an inverse (such as Moore-Penrose inverses), defined using Generalised Matrix Algebra. These "generalised inverses" are not used in these notes.

Consider the matrix equation $\mathbf{Ax} = \mathbf{b}$. If **A** and **b** are known, then **x** may be determined from $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. **x** is found in a sense, by "dividing" **b** by **A**, but in actual fact **x** is determined by *pre-multiplying* both sides of the original equation by the inverse $\mathbf{A}^{-1}$. For example

$$\mathbf{Ax} = \mathbf{b}$$

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \qquad \text{(pre-multiply both sides by } \mathbf{A}^{-1}\text{)}$$

giving

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \qquad \text{(since } \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \text{ and } \mathbf{Ix} = \mathbf{x}\text{)}$$

The following rules regarding matrix inverses hold:

$$(\mathbf{ABC}\ldots)^{-1} = \ldots\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$$
$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$
$$(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}$$
$$(\alpha\,\mathbf{A})^{-1} = \frac{1}{\alpha}\mathbf{A}^{-1} \qquad (\alpha \text{ is a scalar})$$

Matrix inversion plays an important part in least squares, primarily in the solution of systems of linear equations. If the order of **A** is small, say (2,2) or (3,3), then manual calculation of the inverse is relatively simple. But as the order of **A** increases, computer programs or software products such as Microsoft's *Excel* or *The MathWorks* MATLAB are the appropriate tools to calculate inverses and solve systems of equations.

For a (2,2) matrix the inverse is simple and may be computed from the following relationship

$$\text{If} \qquad \mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \text{then} \quad \mathbf{A}^{-1} = \frac{1}{(A_{11}A_{22} - A_{12}A_{21})}\begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

A 4.9 Matrix Differentiation

1. If every element of a matrix **A** of order (*m*,*n*) is a differentiable function of a (scalar) variable *u*, then the derivative $d\mathbf{A}/du$ is an (*m*,*n*) matrix of derivatives

$$\frac{d\mathbf{A}}{du}_{m,n} = \begin{bmatrix} \frac{da_{11}}{du} & \frac{da_{12}}{du} & \cdots & \frac{da_{1n}}{du} \\ \frac{da_{21}}{du} & \frac{da_{22}}{du} & \cdots & \frac{da_{2n}}{du} \\ \vdots & & \ddots & \vdots \\ \frac{da_{m1}}{du} & \frac{da_{m2}}{du} & \cdots & \frac{da_{mn}}{du} \end{bmatrix}$$

Example A 12

$$\mathbf{A} = \begin{bmatrix} 3u^2 & 2u^3 \\ u^2 & 4u^4 \end{bmatrix} \quad \text{then} \quad \frac{d\mathbf{A}}{du} = \begin{bmatrix} 6u & 6u^2 \\ 2u & 16u^3 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} 2u^2 \\ u^3 \\ 3u^4 \end{bmatrix} \quad \text{then} \quad \frac{d\mathbf{x}}{du} = \begin{bmatrix} 4u \\ 3u^2 \\ 12u^3 \end{bmatrix}$$

2. For the matrix product **C** = **AB** where the elements of the matrices **A** and **B** are differentiable functions of the (scalar) variable *u* then $d\mathbf{C}/du$ is given by

$$\frac{d\mathbf{C}}{du} = \frac{d}{du}(\mathbf{AB}) = \frac{d\mathbf{A}}{du}\mathbf{B} + \mathbf{A}\frac{d\mathbf{B}}{du}$$

Note that the sequence adopted in the product terms must be followed exactly, since for example, the derivative of **AB** is in general not the same as the derivative of **BA**.

3. If a vector $\mathbf{y}_{m,1}$ represents *m* functions of the *n* elements of a variable vector $\mathbf{x}_{n,1}$ then the total differential $d\mathbf{y}$ is given by

$$d\mathbf{y} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} d\mathbf{x}$$

where the (*m*,1) and (*n*,1) vectors $d\mathbf{y}$ and $d\mathbf{x}$ contain differentials

$$d\mathbf{y}_{m,1} = \begin{bmatrix} dy_1 \\ dy_2 \\ \vdots \\ dy_m \end{bmatrix} \quad \text{and} \quad d\mathbf{x}_{n,1} = \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}$$

and the partial derivative $\partial\mathbf{y}/\partial\mathbf{x}$ is an (m,n) matrix known as the *Jacobian Matrix*

$$\frac{\partial\mathbf{y}}{\partial\mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \dots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

4. The derivative of the inverse $\mathbf{A}^{-1}$ is obtained from

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$
$$\frac{d}{dx}\left(\mathbf{A}\mathbf{A}^{-1}\right) = \frac{d\mathbf{I}}{dx} = \mathbf{0}$$
$$\frac{d\mathbf{A}}{dx}\mathbf{A}^{-1} + \mathbf{A}\frac{d\mathbf{A}^{-1}}{dx} = \mathbf{0}$$

hence

$$\frac{d\mathbf{A}^{-1}}{dx} = \mathbf{A}^{-1}\frac{d\mathbf{A}}{dx}\mathbf{A}^{-1}$$

A 4.10 Differentiation of Bilinear and Quadratic forms

For the bilinear form $u = \mathbf{x}^T\mathbf{A}\mathbf{y}$ where $\mathbf{A}$ is independent of both $\mathbf{x}$ and $\mathbf{y}$

$$\frac{\partial u}{\partial\mathbf{x}} = \mathbf{y}^T\mathbf{A}^T \quad \text{and} \quad \frac{\partial u}{\partial\mathbf{y}} = \mathbf{x}^T\mathbf{A}$$

For the quadratic form $q = \mathbf{x}^T\mathbf{B}\mathbf{x}$ where $\mathbf{B}$ is independent of $\mathbf{x}$

$$\frac{\partial q}{\partial\mathbf{x}} = 2\mathbf{x}^T\mathbf{A}$$

These differentials are given without proof, but can be verified in the following manner

$$\text{let } \mathbf{x}_{3,1} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \ \mathbf{y}_{2,1} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \text{ and } \mathbf{A}_{3,2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

then

$$u = \mathbf{x}^T\mathbf{A}\mathbf{y} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= \begin{bmatrix} x_1a_{11} + x_2a_{21} + x_3a_{31} & x_1a_{12} + x_2a_{22} + x_3a_{32} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= y_1x_1a_{11} + y_1x_2a_{21} + y_1x_3a_{31} + y_2x_1a_{12} + y_2x_2a_{22} + y_2x_3a_{32}$$

and

$$\frac{\partial u}{\partial \mathbf{x}} = \begin{bmatrix} y_1a_{11} + y_2a_{12} & y_1a_{21} + y_2a_{22} & y_1a_{31} + y_2a_{32} \end{bmatrix}$$

$$= (\mathbf{A}\mathbf{y})^T$$

$$= \mathbf{y}^T\mathbf{A}^T$$

and

$$\frac{\partial u}{\partial \mathbf{y}} = \begin{bmatrix} x_1a_{11} + x_2a_{21} + x_3a_{31} & x_1a_{12} + x_2a_{22} + x_3a_{32} \end{bmatrix}$$

$$= \mathbf{x}^T\mathbf{A}$$

Using similar methods, the partial differential for the quadratic form *q* can also be verified. More explicit proofs can be found in Mikhail (1976, pp.457-460) and Mikhail & Gracie (1981, pp.322-324).

A 4.11 Matrix Partitioning

A subset of elements from a given matrix **A** is called a *sub-matrix* and *matrix partitioning* allows the matrix to be written in terms of sub-matrices rather than individual elements. Thus, the matrix **A** can be partitioned into sub-matrices as follows

$$\mathbf{A}_{m,n} = \left[\begin{array}{ccc:ccc} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots & a_{3n} \\ \hdashline a_{41} & a_{42} & a_{43} & a_{44} & \cdots & a_{4n} \\ \vdots & & & & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & a_{m4} & \cdots & a_{mn} \end{array}\right]$$

Considering the vertical dotted line only, **A** could be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix}$$

where $\mathbf{A}_1$ is an $(m,3)$ sub-matrix and $\mathbf{A}_2$ is an $m,(n-3)$ sub-matrix. Similarly, considering the horizontal dotted line only

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}$$

where in this case $\mathbf{A}_1$ is a $(2,n)$ sub-matrix and $\mathbf{A}_2$ is an $(m-2),n$ sub-matrix.

Considering both the horizontal and vertical lines

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

where $\mathbf{A}_{11}$ is a $(2,3)$ sub-matrix, $\mathbf{A}_{12}$ is a $(2,(n-3))$ sub-matrix, $\mathbf{A}_{21}$ is an $((m-2),3)$ sub-matrix and $\mathbf{A}_{22}$ is an $((m-2),(n-3))$ sub-matrix.

All matrix operations outlined in the previous sections can be performed on the sub-matrices as if they are normal matrix elements providing necessary precautions are exercised regarding dimensions.

Example A 13

Transposing partitioned matrices

$$\mathbf{A} = \left[\begin{array}{cc:cc} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \hdashline 9 & 10 & 11 & 12 \end{array}\right] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

and

$$\mathbf{A}^T = \begin{bmatrix} \mathbf{A}_{11}^T & \mathbf{A}_{21}^T \\ \mathbf{A}_{12}^T & \mathbf{A}_{22}^T \end{bmatrix} = \begin{bmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{bmatrix}$$

Example A 14

Multiplying partitioned matrices

$$\mathbf{A}_{3,4} = \left[\begin{array}{cc:cc} 1 & 2 & 3 & 4 \\ 2 & -1 & 5 & 2 \\ \hdashline 3 & 2 & 1 & -2 \end{array}\right] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}_{2,2} \quad \text{and} \quad \mathbf{B}_{4,2} = \left[\begin{array}{cc} 1 & 7 \\ 2 & 4 \\ \hdashline 7 & 6 \\ -2 & 5 \end{array}\right] = \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{21} \end{bmatrix}_{2,1}$$

the product is

$$\mathbf{AB} = \mathbf{C} = \begin{bmatrix} (\mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21}) \\ (\mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21}) \end{bmatrix}_{2,1}$$

where

$$\mathbf{A}_{11}\mathbf{B}_{11} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 7 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 0 & 10 \end{bmatrix}; \quad \mathbf{A}_{12}\mathbf{B}_{21} = \begin{bmatrix} 3 & 4 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 3 & 6 \\ -2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 38 \\ 11 & 40 \end{bmatrix}$$

$$\mathbf{A}_{21}\mathbf{B}_{11} = \begin{bmatrix} 3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 7 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 29 \end{bmatrix}; \quad \mathbf{A}_{22}\mathbf{B}_{21} = \begin{bmatrix} 1 & -2 \end{bmatrix}\begin{bmatrix} 3 & 6 \\ -2 & 5 \end{bmatrix} = \begin{bmatrix} 7 & -4 \end{bmatrix}$$

noting that columns of $\mathbf{A}_{xx}$ must equal rows of $\mathbf{B}_{xx}$. The product is

$$\mathbf{AB} = \mathbf{C} = \begin{bmatrix} 6 & 53 \\ 11 & 50 \\ 14 & 25 \end{bmatrix}$$

A 5 SOME SPECIAL MATRICES

A 5.1 Symmetric Matrices

A symmetric matrix is defined to be a matrix that remains invariant when transposed, ie,

$$\mathbf{A}^T = \mathbf{A}$$

Symmetric matrices are <u>always</u> square matrices. For any symmetric matrix **A**, the elements conform to the following

$$a_{ij} = a_{ji}$$

Example A 15

$$\mathbf{A} = \begin{bmatrix} a_{11} & 3 & 5 & 7 \\ 3 & a_{22} & 9 & 11 \\ 5 & 9 & a_{33} & 13 \\ 7 & 11 & 13 & a_{44} \end{bmatrix}$$

For any matrix **A** and for any symmetric matrix **B** the matrices

$$\mathbf{A}\mathbf{A}^T, \quad \mathbf{A}^T\mathbf{A}, \quad \mathbf{A}\mathbf{B}\mathbf{A}^T \quad \text{and} \quad \mathbf{A}^T\mathbf{B}\mathbf{A}$$

are all symmetric matrices.

In least squares, we are often dealing with symmetric matrices. For example, the matrix equation $\mathbf{N}_{u,u} = \mathbf{B}^T_{u,n}\mathbf{W}_{n,n}\mathbf{B}_{n,u}$ often appears. **B** is an (*n*,*u*) matrix of coefficients of the *u* unknowns in *n* equations, **W** is an (*n*,*n*) weight matrix (always symmetric) and **N** is the (*u*,*u*) symmetric coefficient matrix of the set of *u* normal equations.

A 5.2 Skew-symmetric Matrices

In contrast to the above, a skew-symmetric (or anti-symmetric) matrix is defined to be a square matrix that changes sign when transposed, so that

$$\mathbf{A}^T = -\mathbf{A}$$

and the elements conform to the rule

$$a_{ij} = -a_{ji}$$

Note that this definition means that the elements of the leading diagonal can only be zero. An example of a skew-symmetric matrix of order 3 is

$$\mathbf{A} = \begin{bmatrix} 0 & b & c \\ -b & 0 & d \\ -c & -d & 0 \end{bmatrix}$$

Example A 16

Skew-symmetric matrices are found in some surveying and geodesy applications. For instance, a 3D conformal transformation from one orthogonal coordinate system (*x*,*y*,*z*) to another (*X*,*Y*,*Z*) is defined by the matrix equation

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \lambda \mathbf{R}_{\kappa\phi\omega} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

where λ is a scale factor, T_X, T_Y and T_Z translations between the coordinate origins and $\mathbf{R}_{\kappa\phi\omega}$ is a rotation matrix derived by considering successive rotations ω, ϕ and κ about the *x*, *y* and *z* axes respectively

$$\mathbf{R}_{\kappa\phi\omega} = \begin{bmatrix} c_\phi c_\kappa & c_\omega s_\kappa + s_\omega s_\phi c_\kappa & s_\omega s_\kappa - c_\omega s_\phi c_\kappa \\ -c_\phi s_\kappa & c_\omega c_\kappa - s_\omega s_\phi s_\kappa & s_\omega c_\kappa + c_\omega s_\phi s_\kappa \\ s_\phi & -s_\omega c_\phi & c_\omega c_\phi \end{bmatrix}$$

Note that $c_\kappa s_\phi s_\omega = \cos\kappa \sin\phi \sin\omega$, and *x, y, z* and *X, Y, Z* refer to the axes of right-handed orthogonal coordinate systems. Rotations ω, ϕ and κ are considered as positive anti-clockwise according to the "right-hand-grip rule".

In many applications the rotation matrix $\mathbf{R}_{\kappa\phi\omega}$ can be simplified because ω, ϕ and κ are small (often less than 3°). In such cases, the sines of angles are approximately equal to their radian measures, the cosines are approximately 1 and products of sines are approximately zero. This allows the rotation matrix $\mathbf{R}_{\kappa\phi\omega}$ to be approximated by $\mathbf{R}_S$

$$\mathbf{R}_S = \begin{bmatrix} 1 & \kappa & -\phi \\ -\kappa & 1 & \omega \\ \phi & -\omega & 1 \end{bmatrix}$$

This matrix is sometimes referred to as a skew-symmetric matrix. Although the elements of the upper-triangular part are the opposite sign of those in the lower triangular part, it does not conform to the definition above, since the leading diagonal elements are not zero.

Note that $\mathbf{R}_S$ can be expressed as the sum of the identity matrix **I** and a skew-symmetric matrix **S**

$$\mathbf{R}_S = \begin{bmatrix} 1 & \kappa & -\phi \\ -\kappa & 1 & \omega \\ \phi & -\omega & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & \kappa & -\phi \\ -\kappa & 0 & \omega \\ \phi & -\omega & 0 \end{bmatrix} = \mathbf{I} + \mathbf{S}$$

A 5.3 Symmetry and Skew-symmetry

Every square matrix can be uniquely decomposed into the sum of a symmetric and skew-symmetric matrix. Consider the following

$$\begin{aligned} \mathbf{A} &= \mathbf{A} + \tfrac{1}{2}\mathbf{A}^T - \tfrac{1}{2}\mathbf{A}^T \\ &= \tfrac{1}{2}\left(\mathbf{A} + \mathbf{A}^T\right) + \tfrac{1}{2}\left(\mathbf{A} - \mathbf{A}^T\right) \\ &= \mathbf{A}_{Sym} + \mathbf{A}_{Skew} \end{aligned}$$

where

$$\mathbf{A}_{Sym} = \tfrac{1}{2}(\mathbf{A} + \mathbf{A}^T) \text{ is symmetric because}$$
$$\mathbf{A}_{Sym}^T = \tfrac{1}{2}(\mathbf{A} + \mathbf{A}^T)^T = \tfrac{1}{2}(\mathbf{A}^T + \mathbf{A}) = \mathbf{A}_{Sym}$$

and

$$\mathbf{A}_{Skew} = \tfrac{1}{2}(\mathbf{A} - \mathbf{A}^T) \text{ is skew-symmetric because}$$
$$\mathbf{A}_{Skew}^T = \tfrac{1}{2}(\mathbf{A} - \mathbf{A}^T)^T = \tfrac{1}{2}(\mathbf{A}^T - \mathbf{A}) = -\tfrac{1}{2}(\mathbf{A} - \mathbf{A}^T) = -\mathbf{A}_{Skew}$$

A 5.4 Orthogonal Matrix

An orthogonal matrix is a square matrix satisfying the following two conditions:

1. the *norms* of all its rows and columns are equal to unity, and
2. any row is *orthogonal* to every other row in the matrix and similarly for the columns.

These two conditions imply that if **A** is an orthogonal matrix, then

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I} \qquad (\mathbf{I} \text{ is the Identity matrix})$$

and hence an orthogonal matrix has the very useful property that its inverse matrix is the same as its transpose matrix, or

$$\mathbf{A}^{-1} = \mathbf{A}^T \qquad (\text{if } \mathbf{A} \text{ is orthogonal})$$

The terms *norm* and *orthogonal* are applicable to vector algebra. The *norm* of a vector is the magnitude of the vector and is the square root of the product of the vector and its transpose.

Any row (or column) of a matrix has all the characteristics of a vector, and hence the norm of any row (or column) of a matrix is the square root of the product of the row (or column) by its transpose. Two vectors are *orthogonal* if, and only if, their scalar product is zero. Considering rows and columns of the matrix as vectors, then any two matrix rows (or columns) are orthogonal if their scalar product is zero.

Example A 17

Rotation matrices are examples of orthogonal matrices. For example, consider a point *P* with coordinates $P(e,n)$ in the *east-north* coordinate system. If the axes are rotated about the origin by an angle θ (measured clockwise from north), *P* will have coordinates e',n' in the rotated system equal to

$$e' = e\cos\theta - n\sin\theta$$
$$n' = e\sin\theta + n\cos\theta$$

These equations can be written in matrix notation as

$$\begin{bmatrix} e' \\ n' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} e \\ n \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} e' \\ n' \end{bmatrix} = \mathbf{R}_\theta \begin{bmatrix} e \\ n \end{bmatrix}$$

where $\mathbf{R}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is known as the rotation matrix.

The norms of the columns and rows of $\mathbf{R}_\theta$ are unity since $\sin^2\theta + \cos^2\theta = 1$ and the columns and rows are orthogonal since $\sin\theta\cos\theta - \sin\theta\cos\theta = 0$. Hence $\mathbf{R}_\theta$ is orthogonal and its inverse is equal to its transpose. This is useful in defining the transformation from e',n' to e,n coordinates. Pre-multiplying both sides of the original transformation by the inverse $\mathbf{R}_\theta^{-1}$ gives

$$\mathbf{R}_\theta^{-1}\begin{bmatrix} e' \\ n' \end{bmatrix} = \mathbf{R}_\theta^{-1}\mathbf{R}_\theta \begin{bmatrix} e \\ n \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} e \\ n \end{bmatrix} = \mathbf{R}_\theta^{-1}\begin{bmatrix} e' \\ n' \end{bmatrix} \quad \text{since } \mathbf{R}_\theta^{-1}\mathbf{R}_\theta = \mathbf{I}$$

and

$$\begin{bmatrix} e \\ n \end{bmatrix} = \mathbf{R}_\theta^T \begin{bmatrix} e' \\ n' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} e' \\ n' \end{bmatrix} \quad \text{since } \mathbf{R}_\theta^{-1} = \mathbf{R}_\theta^T$$

APPENDIX A REFERENCES

Mikhail, E.M. 1976, *Observations and Least Squares*, IEP–A Dun-Donelley, New York.

Mikhail, E.M. and Gracie, G. 1981, *Analysis and Adjustment of Survey Measurements.* Van Nostrand Reinhold, New York.

Williams, I.P. 1972, *Matrices for Scientists,* Hutchinson &Co Ltd, London.