ROYAL INSTITUTE
OF TECHNOLOGY

Theory of Errors and Least Squares Adjustment

Huaan Fan

ISBN 91-7170-200-8

Royal Institute of Technology (KTH)
Division of Geodesy and Geoinformatics
100 44 Stockholm
Sweden

August 2010

Theory of Errors and Least Squares Adjustment

Huaan Fan, PhD

Royal Institute of Technology (KTH)
Division of Geodesy and Geoinformatics
Teknikringen 72, 100 44 Stockholm, Sweden

Phone: +46 8 7907340
Fax: +46 8 7907343
E-mail: hfan@kth.se
URL: http://www.infra.kth.se/impgg

With 22 illustrations and 49 numerical examples

ISBN 91-7170-200-8

Geodesy Report No. 2015
TRITA-GEOFOTO 1997:21

Printed by **Universitetsservice AB**, Drottning Kristinas väg 53 B, Stockholm
October 1997, August 1998, August 1999, May 2000, August 2001, August 2003, October 2005, August 2006, August 2008, August 2009

Printed by **AJ E-print AB**, Oxtorgsgatan 8, 111 57 Stockholm. August 2010.

To the memory of my father

Fuguan Fan (1928-1996)

who loved and inspired a geodesist

Preface

Theory of errors has been a well defined subject within geodesy, surveying and photogrammetry. Presently there are several masterpieces on this subject, such as the famous works by Bjerhammar (1973), Mikhail (1976) and Koch (1980), just to mention a few of them. However, the author still feels the need for a middle-level textbook in this field. A textbook that balances between practical applications and pure mathematical treatments; A textbook that integrates classical adjustment methods with modern developments in geodesy, surveying and photogrammetry.

To meet the above need, an attempt was started in 1995 that has resulted in this compendium. Naturally, it is neither the author's ambition nor within his capability to challenge those great works mentioned above. The compendium will primarily be used for courses given at the Royal Institute of Technology (KTH).

The compendium consists of seven chapters. Chapter 1 deals with basic concepts in theory of errors, such as standard error and error propagation, error ellipse and error ellipsoid, linear equation systems as well as some elementary concepts of statistical analysis. Chapter 2 is devoted to the classical condition adjustment method, including condition adjustment in groups and condition adjustment with unknowns. The method of adjustment by elements is treated in Chapter 3, where adjustment by elements with constraints and sequential adjustment have also been described. Chapters 4-7 deal with diverse topics based on recent developments in theory of errors. These topics include generalized matrix inverses and their applications in least squares adjustment; a posteriori estimation of variance-covariance components; detection of gross and systematic errors; and finally prediction and filtering in linear dynamic systems.

The essential prerequisite for the compendium is a familiarity with linear algebra, mathematical statistics and basic concepts in surveying. In other words, it is assumed that students have already acquired background knowledge in mathematics and surveying. Therefore, efforts have been made to limit discussions on pure mathematical or surveying subjects.

In order to keep mathematical derivations brief and elegant, matrix notations have been used exclusively throughout the compendium. Several old concepts in theory of errors (e.g. the classical Gauss-Doolittle's table for solving normal equations) have, in the author's opinion, become out of date and thus been excluded.

To help readers better understand the theoretical concepts, a number of numerical examples (mostly originating from geodesy and surveying) are provided. For the sake of simplicity, most numerical examples are constructed so that only ideally simple numbers are involved and that they can be solved without needing to use calculators and computers. For those examples with more realistic data, the presented numerical results are obtained based on the author's own programming on an Intel Pentium PC using Lahey Fortran 77 Compiler (version 5.1). Minor decimal differences might occur if the same examples are calculated using other hardware and software.

The author wishes to acknowledge all the help and encouragement from his colleagues and students. Special thanks go to TeknL *George Stoimenov* for proof-reading the manuscript and Mr *Hossein Nahavandchi* for helping draw some of the figures. Of particular benefit have been my students in the classes MK-92, TL-93, TL-94 and TL-95, who have worked through the raw materials as the compendium evolved. To all of them, I express my sincere thanks !

Stockholm, August 1997. *Huaan Fan*

Contents

Chapter 1

Fundamentals of Theory of Errors

Science and engineering often involves measurements of different types . In geodesy and surveying, geometrical quantities (such as angles, distances, heights, etc.) or physical quantities (e.g. gravity) are directly measured, producing large amounts of data which need to be processed. To some extent, a surveying project may be considered as a data production process, from data collection, data processing, to final presentation (graphically and/or digitally).

Due to human limitations, imperfect instruments, unfavourable physical environment and improper measurement routines, which together define the *measurement condition*, all measurement results most likely contain errors. One can discover the existence of measurement errors in different ways. If we repeat the same measurement several times, we will normally get different results due to measurement errors. Another way to discover errors is to check whether the obtained measurement results satisfy some geometrical or physical relations which may exist. For example, one may check whether the sum of three measured angles of a plane triangle is equal to the theoretical value, 180 degrees.

Normally, one may distinguish three types of errors: *systematic errors, gross errors* and *random errors*.

Systematic errors are errors which follow certain physical or mathematical rules and often affect surveying results systematically. The causes of this kind of errors can be the instruments used, physical environment in which measurements are made, human factors and measurement routines. To avoid or reduce systematic errors, one may *(a)* calibrate carefully instruments before field work starts; *(b)* design and use suitable measurement routines and procedures which can reduce or eliminate possible systematic errors; *(c)* if possible, correct measurement results afterwards. One example of systematic errors is the constant error of -5 *cm* for a tape. This constant error will cause a systematic error to all distance measurements by this tape. Another example is the tropospheric effect on GPS satellite signal transmission. To reduce the tropospheric effect on GPS measurements, one may measure both $L1$ and $L2$ frequencies of GPS signals so that the tropospheric effects can be reduced through dual frequency combinations.

Gross errors are errors due to human mistakes, malfunctioning instruments or wrong measurement methods. Gross errors do not follow certain rules and normally cannot be treated by statistical methods. In principle, gross errors are not permitted and should be avoided by surveyors' carefulness and control routines. For example, it can happen that a surveyor might write $50^0\ 32'\ 50.9''$ in his field observation protocol when the actual reading on the theodolite is $50^0\ 32'\ 5.9''$. If the surveyor is highly concentrated during the measurement, he or she may be able to avoid this kind of blunders. On the other hand, if he or she measures the direction by both right circle and left circle, or measure the same direction by more than one complete rounds, the mistake can easily be discovered. Gross errors are also called *blunders* or *outliers*.

Random errors or *stochastic errors* are errors which behave randomly and affect the measurements in a non-systematic way. Random errors can be caused by human factors, instrument errors, physical environment and measurement routines. They can be reduced if the total measurement condition has been improved. The primary study object of theory of errors is just random errors. Probability theory

and mathematical statistics is the science which specializes in studies of random (or stochastic) events, variables and functions. It will serve as the theoretical base for our treatment of the random measurement errors. In Chapter 6, we will briefly discuss how to detect gross errors and systematic errors.

Based on analysis of large amounts of available observation data (e.g. thousands of triangular misclosures in geodetic triangulation networks), it has been found that random errors, though non-systematic, show certain statistical characteristics. If a set of errors ε_1, ε_2, $\cdots$, ε_n have occurred under (roughly) the same measurement condition, then the following statistical characteristics have been discovered:

- The arithmetic mean of ε_i approaches zero when the number n of observations approaches infinity:

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \varepsilon_i}{n} = 0 \tag{1.1}$$

- Positive errors and negative errors with same magnitude occur roughly at equal frequency;
- Errors of smaller magnitude occur more often than errors of larger magnitude;
- Under specific measurement condition, the absolute magnitude of errors is within some limit.

To reduce measurement errors and their effects on the final surveying results, one need to improve the overall measurement condition. As errors are impossible to avoid completely, it is natural to do redundant measurements, both to discover the existence of errors and to increase accuracy and reliability of the final results. When measurement errors are present and redundant measurements are made, there will exist inconsistency or ″contradiction″ among measurements, also called *misclosure*. One of the tasks of geodetic and photogrammetric computations is to get rid of misclosures among measurements in an optimal way according to some estimation criteria (such as the least squares principle). These criteria should naturally be based on the property of the measurement errors. Traditionally, the work or process to eliminate misclosures among measurements and obtain the best results out of available measurement data is called *adjustment*.

Another important task of theory of errors is to assess the quality of observations and results derived from observations. The quality of observations concern three related but different aspects: *precision*, *accuracy* and *reliability*. *Precision* describes the degree of repeatability of the observations and is an internal measure of random errors and their effects on the quality of observation. *Accuracy* is a measure of agreement between the observed value and the correct value. It is influenced not only by random errors but also, more significantly, by systematic or other non-random errors. *Reliability* concerns the capability of observations against gross and systematic errors. Observations or surveying methods with high reliability can easily detect gross and systematic errors, and are said to be *robust*. In general, redundant observations can improve the precision, accuracy and reliability of the observations as well as the derived results.

Theory of errors and least squares adjustment is an important subject within the geomatics programme offered at KTH. This is due to the fact that surveying and mapping (or production of spatial data) often requires mathematical processing of measurement data. Furthermore, the general methodology of spatial data processing is essentially the same as that for data processing in other science and engineering fields, even though data collection procedures and data types can be different. Theory of errors is related to and comparable with what is called *estimation theory* used in automatic control and signal processing. Therefore, studying theory of errors can be helpful for solving general data processing problems in other scientific and engineering fields.

The theoretical foundation of theory of errors is *probability theory and mathematical statistics*. As numerical computations are frequently involved, *numerical methods* in general and *matrix algebra* in particular are useful tools. Other useful mathematical tools include *linear programming*, *optimization theory* and *functional analysis* (theory of vector spaces).

1.1 Standard Errors and Weights

Standard errors and weights are probably the most elementary and fundamental concepts in theory of errors. They are used as *"index numbers"* to describe the accuracy of the measured and derived quantities.

Let $\widetilde{\ell}$ and ℓ denote the true value and the measured value of a quantity, respectively. The *true error* ε of measurement ℓ is defined as the measured value minus the true value :

$$\varepsilon = \ell - \widetilde{\ell} \tag{1.2}$$

ε is also called the *absolute (true) error* of ℓ, while the *relative error* for ℓ can be defined by:

$$\gamma = \left|\frac{\varepsilon}{\ell}\right| = \left|\frac{\ell - \widetilde{\ell}}{\ell}\right| \tag{1.3}$$

γ is a more suitable index for describing the error of ℓ in the case when the error ε tends to increase with the absolute magnitude of ℓ. A typical example is the error of distance measurements in geodesy and surveying. Though two distances, at 1000 and 10 meters respectively, might have equal true errors of 1 cm, the quality of the first measurement would be much better than that of the second one, as their relative errors are different: 1 : 100000 for the first distance and 1 : 1000 for the second one.

Both absolute error ε and relative error γ defined above are describing an individual measurement error. Most measurement errors are random errors that behave in a random way. Therefore, in practice it is very difficult or impossible to describe each individual error which occurs under a specific measurement condition (i.e. under a specific physical environment, at a specific time, with a specific surveyor, using a specific instrument, etc). However, for simplicity we often prefer to use one or several simple index numbers to judge the quality of the obtained measurements (how good or how bad the measurements are). An old and well known approach is to define and use statistical error indices.

1.1.1 Standard Errors

The statistical way of thinking does not try to describe each individual error separately, rather try to judge the collective property of the same group of errors. It assumes that it is the total measurement condition (i.e. measurement instruments used, physical environment of field survey, professional skills of surveyors, etc.) that determines the quality of the obtained observations. In other words, different observations with true errors different in size and sign will still be regarded as having the same quality or accuracy, *if* they are made under same or similar measurement condition.

Standard error is an error index defined in mathematical statistics and used in theory of errors to describe the *average* observation errors coming from the same (or similar) measurement condition. Standard error is also called *standard deviation*, or *mean-square-root error*. Below, we will define standard errors in an intuitive way, while a more rigorous definition based on probability theory will be given in Subsection 1.1.3.

Assume that a quantity with true value $\widetilde{\ell}$ has been measured n times independently under the same measurement condition, i.e. with same accuracy. Denoting the $i-th$ observed value by ℓ_i ($i = 1, 2, 3, \cdots n$) and its (true) error by ε_i , i.e. $\varepsilon_i = \ell_i - \widetilde{\ell}$, the *theoretical standard error* (σ) of this group of observations is then defined by:

$$\sigma^2 = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n} \tag{1.4}$$

When the number of measurements is limited to a finite number n, one gets only an estimate of σ:

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n} \tag{1.5}$$

In probability theory, σ^2 is called the *variance* of the random variable ε (Cf Subsection 1.1.3).

Example 1.1

Let ℓ_1, ℓ_2 and ℓ_3 denote the observed values of three internal angles in a plane triangle. The true error of the sum of these three internal angles can be calculated from the observed angles:

$$w = \ell_1 + \ell_2 + \ell_3 - 180^0 \tag{1.6}$$

w defined above is called the *misclosure of the triangle*. Assume that in a geodetic triangulation network, 30 triangles have been observed with the triangular misclosures (w_i , $i = 1, 2, \cdots, 30$) as listed in Table 1.1 :

Table 1.1: List of Triangular Misclosures

i	w_i ($''$)	i	w_i ($''$)	i	w_i ($''$)
1	+1.5	11	-2.0	21	-1.1
2	+1.0	12	-0.7	22	-0.4
3	+0.8	13	-0.8	23	-1.0
4	-1.1	14	-1.2	24	-0.5
5	+0.6	15	+0.8	25	+0.2
6	+1.1	16	-0.3	26	+0.3
7	+0.2	17	+0.6	27	+1.8
8	-0.3	18	+0.8	28	+0.6
9	-0.5	19	-0.3	29	-1.1
10	+0.6	20	-0.9	30	-1.3

If all triangles are observed independently of each other, the standard error of the misclosures can be estimated using (1.5):

$$\widehat{\sigma}_w{}^2 = \frac{1}{30}\sum_{i=1}^{30} w_i^2 = \frac{25.86}{30} \quad (''^2) \tag{1.7}$$

or:

$$\widehat{\sigma}_w = 0.93''$$

If all three angles in each triangle are uncorrelated and of equal accuracy (i.e. having equal standard error), the standard error of each angle observation can be estimated as:

$$\widehat{\sigma} = \frac{\widehat{\sigma}_w}{\sqrt{3}} = 0.54''$$

The last formula can be obtained by applying error propagation law on Eq.(1.6) (Cf Section 1.2).

In practice, the observed angles of a triangle and angles of different triangles are most likely correlated with each other. Therefore, Eq. (1.7) is not a rigorous formula for estimating the standard error of angle measurements. Nevertheless, it still provides us with a measure on the quality of angle measurements in geodetic triangulation.

1.1.2 Weights and Unit-Weight Standard Error

Sometimes, what we are most interested in is not the absolute magnitude of observation errors, rather the relative accuracy of observations. According to the relative accuracy of different observations, we assign each observation a positive number, called *weight*. The smaller an observation error is, the more accurate the observation will be and consequently the bigger weight the observation should have. Therefore, in theory of errors the weight p_i of an observation ℓ_i is defined to be *inversely proportional to the variance* σ_i^2 (standard error squared) of ℓ_i:

$$p_i = \frac{c_0}{\sigma_i^2} \tag{1.8}$$

where c_0 is an arbitrary positive number.

Based on practical experiences, the following empirical weights are used in geodesy and surveying:

- *levelling:* $p_i = c_0$ divided by the length of the levelling line
- *distance measurement:* $p_i = c_0$ divided by the distance or distance squared
- *direction (angle) measurement*: p_i = number of whole rounds measured divided by c_0

It should be noted that outside the field of theory of errors, the concept of weights may be based on factors other than standard errors. For example, when calculating an "average" price index of a stock market, the total market value of each share (or its daily turn-out) may be used as the basis to define the weight.

Assume that there is an observation with standard error σ_0 and weight $p_0 = 1$. Eq.(1.8) then gives: $c_0 = {\sigma_0}^2$, *i.e.* the arbitrary constant c_0 is equal to the variance of the observation with unit weight $p = 1$. Therefore, σ_0 is called the standard error of unit weight or *unit-weight standard error* (in Swedish: *grundmedelfel*) and $c_0 = {\sigma_0}^2$ is called *variance factor*. Then (1.8) can be written as:

$$p_i = \frac{\sigma_0^2}{\sigma_i^2} \tag{1.9}$$

It is not difficult to find out the geodetic meaning of c_0 in the above empirical weights:

- *levelling:* c_0 = length of the levelling line with weight 1
- *distance measurement:* c_0 = distance or squared distance of weight 1
- *direction (angle) measurement:* c_0 = number of whole rounds by which the direction (angle) of unit weight is measured

Naturally, observations of different weights will have different standard errors. For observation ℓ_i with true error ε_i and weight p_i $(i = 1, 2, 3, \cdots, n)$, one can calculate the *theoretical* unit-weight standard error as:

$$\sigma_0^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} p_i \varepsilon_i \varepsilon_i \tag{1.10}$$

For a finite n , we obtain instead an *estimated* unit-weight standard error:

$$\widehat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^{n} p_i \varepsilon_i \varepsilon_i \tag{1.11}$$

From (1.9), we can get the theoretical variance σ_i^2 and the estimated variance $\widehat{\sigma}_i^2$ for observation ℓ_i of weight p_i:

$$\sigma_i^2 = \frac{\sigma_0^2}{p_i} \ , \quad \widehat{\sigma}_i^2 = \frac{\widehat{\sigma}_0^2}{p_i} \tag{1.12}$$

In practice the true value $\widetilde{\ell}_i$ $(i = 1, 2, 3, \cdots, n)$ of observation ℓ_i is often unknown and thus formulas (1.4) and (1.5) or (1.10) and (1.11) cannot be applied directly. A practical solution is to find first an optimal estimate $\widehat{\ell}_i$ for ℓ_i based on some theoretical criteria (e.g. the least squares principle) and then obtain an estimate for the unit-weight standard error as:

$$\widehat{\sigma}_0^2 = \frac{1}{f} \sum_{i=1}^{n} p_i \widehat{\varepsilon}_i \widehat{\varepsilon}_i \tag{1.13}$$

where :

p_i = weight of ℓ_i ;

$\widehat{\varepsilon}_i = \ell_i - \widehat{\ell}_i$ (= estimated error of ℓ_i);

f = number of redundant observations (statistical degrees of freedom)

The estimated unit-weight standard error given by (1.13) is solely dependent on the estimate $\widehat{\ell}_i$. Only good estimate $\widehat{\ell}_i$ can lead to meaningful estimate $\widehat{\sigma}_0$.

1.1.3 Variance-Covariance Matrix and Cofactor Matrix

For a random variable ε, its probability *distribution function* $F(x)$ and probability *frequency function* $f(\varepsilon)$ are associated by equation:

$$P(\varepsilon \leq x) = F(x) = \int_{-\infty}^{x} f(\varepsilon)d\varepsilon \qquad (-\infty < x < +\infty) \tag{1.14}$$

where $P(\varepsilon \leq x)$ denotes the probability that $\varepsilon \leq x$. $f(\varepsilon)$ is also known as the *density function* of ε.

Eq. (1.14) indicates that geometrically $F(x)$ is equal to the area between the horizontal ε-axis and curve $f(\varepsilon)$ within the interval $(-\infty, x)$. The frequency function characterizes the changing rate of the probability. This can be seen more clearly from the following relation which is obtained by differentiating both sides of Eq.(1.14) with respect to x :

$$\frac{\partial F(x)}{\partial x} = f(x) \tag{1.15}$$

Mathematically, $f(x)$ or $F(x)$ is able to describe the complete analytical properties of random variable ε. Practically one can use several simple index numbers, called *characteristic values*, to describe the most important (but not necessarily complete) properties of the random variable. *Mathematical expectation* and *variance* are two such characteristic numbers. The mathematical expectation of random variable ε , denoted as $E(\varepsilon)$, is defined as the average value of ε weighted against the probability over the whole definition interval:

$$E(\varepsilon) = \int_{-\infty}^{+\infty} \varepsilon \, f(\varepsilon)d\varepsilon \tag{1.16}$$

The variance of ε, denoted as $var(\varepsilon)$ or σ^2, is defined as:

$$var(\varepsilon) = \sigma^2 = E\left\{[\varepsilon - E(\varepsilon)]^2\right\} = \int_{-\infty}^{+\infty} [\varepsilon - E(\varepsilon)]^2 \cdot f(\varepsilon) \cdot d\varepsilon \tag{1.17}$$

If ε denotes a geodetic measurement error, $E(\varepsilon)$ may be regarded as the theoretically true value of ε. When there is no systematical error or gross error in our measurements, it is naturally to believe that the expectation of measurement errors should be equal to zero, *i.e.* $E(\varepsilon) = 0$. On the other hand, σ^2 represents the "average" deviation of ε from the theoretical mean value $E(\varepsilon)$. This explains why σ is also called the *standard deviation* of ε.

Using the mathematical expectation operator $E(\cdot)$, the *covariance* between two random variables ε_1 and ε_2 can be defined:

$$\sigma_{12} = E\left\{[\varepsilon_1 - E(\varepsilon_1)][\varepsilon_2 - E(\varepsilon_2)]\right\} \tag{1.18}$$

If the variances of ε_1 and ε_2 are σ_1^2, σ_2^2, respectively, and the covariance between them is σ_{12}, the correlation coefficient between them is defined as :

$$\rho_{12} = E\left[\frac{\varepsilon_1 - E(\varepsilon_1)}{\sigma_1}\frac{\varepsilon_2 - E(\varepsilon_2)}{\sigma_2}\right] = \frac{\sigma_{12}}{\sigma_1 \cdot \sigma_2} \tag{1.19}$$

It can be easily shown :

$$-1 \leq \rho_{12} \leq +1 \tag{1.20}$$

Now let us look at a random vector $\underset{n\cdot 1}{\varepsilon}$ of n random variables ε_i $(i = 1,\ 2,\ \cdots,\ n)$:

$$\underset{n\cdot 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \tag{1.21}$$

Assume that ε_i has zero expectation and that its variance and covariance with ε_j are as follows:

$$\left.\begin{array}{ll} E(\varepsilon_i) = 0 & \\ E\left\{[\varepsilon_i - E(\varepsilon_i)]^2\right\} = E(\varepsilon_i^2) = \sigma_i^2 & (i,\ j = 1, 2, 3, \cdots, n) \\ E\left\{[\varepsilon_i - E(\varepsilon_i)]\left[\varepsilon_j - E(\varepsilon_j)\right]\right\} = E\left(\varepsilon_i \cdot \varepsilon_j\right) = \sigma_{ij} & \end{array}\right\} \tag{1.22}$$

The *variance-covariance matrix* $C_{\varepsilon\varepsilon}$ of random vector ε is then defined as:

$$\underset{n\cdot n}{C_{\varepsilon\varepsilon}} = E\left\{[\varepsilon - E(\varepsilon)][\varepsilon - E(\varepsilon)]^\top\right\} = E\left\{\varepsilon\varepsilon^\top\right\} = E\left(\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} [\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_n]\right)$$

$$= \begin{bmatrix} E\left(\varepsilon_1^2\right) & E\left(\varepsilon_1\varepsilon_2\right) & \cdots & E\left(\varepsilon_1\varepsilon_n\right) \\ E\left(\varepsilon_2\varepsilon_1\right) & E\left(\varepsilon_2^2\right) & \cdots & E\left(\varepsilon_2\varepsilon_n\right) \\ \cdots & \cdots & \cdots & \cdots \\ E\left(\varepsilon_n\varepsilon_1\right) & E\left(\varepsilon_n\varepsilon_2\right) & \cdots & E\left(\varepsilon_n^2\right) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \tag{1.23}$$

If there is a matrix $\underset{n\cdot n}{Q_{\varepsilon\varepsilon}}$ which satisfies:

$$\underset{n\cdot n}{C_{\varepsilon\varepsilon}} = \sigma_0^2\ \underset{n\cdot n}{Q_{\varepsilon\varepsilon}} \tag{1.24}$$

$\underset{n\cdot n}{Q_{\varepsilon\varepsilon}}$ is called the *cofactor matrix* of random vector ε:

$$\underset{n\cdot n}{Q_{\varepsilon\varepsilon}} = \frac{1}{\sigma_0^2} \cdot\ \underset{n\cdot n}{C_{\varepsilon\varepsilon}} \tag{1.25}$$

The inverse matrix of $\underset{n\cdot n}{Q_{\varepsilon\varepsilon}}$, denoted as $\underset{n\cdot n}{P_{\varepsilon\varepsilon}}$, is called the *weight matrix* of ε :

$$P_{\varepsilon\varepsilon} = \left(Q_{\varepsilon\varepsilon}\right)^{-1} \tag{1.26}$$

When vector ε consists of only *uncorrelated* components $\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_n$, the weight matrix $P_{\varepsilon\varepsilon}$ will be a diagonal matrix and the $i-th$ diagonal element of $P_{\varepsilon\varepsilon}$ $(i = 1,\ 2,\ 3, \cdots,\ n)$ denotes the weight of the corresponding element ε_i. However, if $\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_n$ are correlated, $P_{\varepsilon\varepsilon}$will be a non-diagonal matrix whose diagonal elements are not the weights of the correspponding elements. The weight of each ε_i can be calculated from the definition formulas (1.9). See **Example 1.5** at the end of the next section.

1.2 Error Propagation

When direct measurements contain measurement errors, a function of these measurements will most likely contain errors too. This kind of transferring of errors from directly measured quantities to their functions is called *error propagation*. For instance, two measured directions ℓ_1 , ℓ_2 with errors of $3''$ and $2''$, respectively, will lead to an error of $-1''$ in the derived angle between ℓ_1 and ℓ_2. As it is neither meaningful nor practical to discuss individual measurement errors, we will concentrate on the propagation of the variances and covariances, i.e. determining the variance of a given function (or the variance-covariance matrix of several given functions) of one or more random variables whose variance-covariance matrix is known.

1.2.1 Error Propagation in Linear Functions

Assume that x_1, x_2, $\cdots$, x_n are n observations with random errors ε_1, ε_2, $\cdots$, ε_n, respectively. Let $\widetilde{x}_i$ denote the true value (non-random) of x_i. We have then:

$$\varepsilon_i = x_i - \widetilde{x}_i \qquad (1 \le i \le n) \tag{1.27}$$

Using matrix notations, the above equation can be written as:

$$\underset{n\cdot 1}{\varepsilon} = \underset{n\cdot 1}{x} - \underset{n\cdot 1}{\widetilde{x}} \tag{1.28}$$

where:

$$\underset{n\cdot 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{n\cdot 1}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}, \quad \underset{n\cdot 1}{\widetilde{x}} = \begin{bmatrix} \widetilde{x}_1 \\ \widetilde{x}_2 \\ \cdots \\ \widetilde{x}_n \end{bmatrix}$$

Assume that ε has zero expectation and the following variance-covariance matrix :

$$E(\varepsilon) = E\left\{ \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \right\} = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \cdots \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} = \underset{n\cdot 1}{0} \tag{1.29}$$

$$\underset{n\cdot n}{C_{\varepsilon\varepsilon}} = E\left\{ [\varepsilon - E(\varepsilon)] [\varepsilon - E(\varepsilon)]^\top \right\} = E\left\{ \varepsilon\varepsilon^\top \right\} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \tag{1.30}$$

Then vector x will have the same variance-covariance matrix as ε , since $\widetilde{x}$ is non-random:

$$\left.\begin{aligned} & E\{x\} = E\{\widetilde{x} + \varepsilon\} = E\{\widetilde{x}\} = \widetilde{x} \\ & C_{xx} = E\left\{ [x - E(x)][x - E(x)]^\top \right\} = E[\varepsilon\varepsilon^\top] = \underset{n\cdot n}{C_{\varepsilon\varepsilon}} \end{aligned}\right\} \tag{1.31}$$

Let $\widetilde{y}$ stand for a linear function of $\widetilde{x}_i$:

$$\widetilde{y} = a_1\widetilde{x}_1 + a_2\widetilde{x}_2 + \cdots + a_n\widetilde{x}_n = a \cdot \widetilde{x} \tag{1.32}$$

where a_1, a_2, $\cdots$, a_n are given constants and a is defined as:

$$\underset{1\cdot n}{a} = (\, a_1, a_2, \cdots, a_n)$$

The derived value of $\widetilde{y}$ calculated from observations x_1, x_2, $\cdots$, x_n is:

$$y = a_1x_1 + a_2x_2 + \cdots + a_nx_n = a \cdot x \tag{1.33}$$

and its true error ε_y is :

$$\varepsilon_y = y - \widetilde{y} = ax - a\widetilde{x} = a \cdot \varepsilon \tag{1.34}$$

By definition, the variance of y is:

$$\sigma_y^2 = E\left[\varepsilon_y^2\right] = E\left[a\varepsilon \cdot (a\varepsilon)^\top\right] = a \cdot E\left[\varepsilon\varepsilon^\top\right] \cdot a^\top = a\ C_{xx} a^\top =$$

$$= (\ a_1, a_2, \cdots, a_n) \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{bmatrix} \tag{1.35}$$

Eq. (1.35) expresses the variance of a single linear function of n random observations in terms of the variance-covariance matrix of the observation vector x. It is called the *error propagation law for a single linear function of n random variables* .

If ε_i and ε_j are independent of each other, i.e. $\sigma_{ij} = 0$ $(i \neq j;\ 1 \leq i, j \leq n)$, a special case of Eq. (1.35) is obtained:

$$\sigma_y^2 = a_1^2 \cdot \sigma_1^2 + a_2^2 \cdot \sigma_2^2 + \cdots + a_n^2 \cdot \sigma_n^2 \tag{1.36}$$

which can be called the *error propagation law for a single linear function of n independent random variables.*

Example 1.2

Given: $y = 2\ x_1 - 3\ x_2 - x_3$ and $\sigma_1 = 1\ mm$, $\sigma_2 = 2\ mm$, $\sigma_3 = 3\ mm$; ε_1, ε_2, ε_3 are independent of each other.
Sought: σ_y =?
Solution. Here we have $a_1 = +2$, $a_2 = -3$, $a_3 = -1$ and by Eq. (1.36) we obtain:

$$\sigma_y^2 = (+2)^2 1^2 + (-3)^2 2^2 + (-1)^2 3^2 = 4 + 36 + 9 = 49\ mm^2\ , \quad \text{or} \quad \sigma_y = \sqrt{49} = 7\ mm \qquad \blacksquare$$

Now let us generalize the single linear function y in Eq. (1.32) to a vector y containing m linear functions of x :

$$\underset{m\cdot 1}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \underset{m\cdot n}{A}\ \underset{n\cdot 1}{x} \tag{1.37}$$

where:

$$\underset{m\cdot n}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1.38}$$

The variance-covariance matrix C_{yy} of the random vector y can be found:

$$\underset{m\cdot m}{C_{yy}} = E\left\{[y - E(y)][y - E(y)]^\top\right\} = A\ E\left\{[x - E(x)][x - E(x)]^\top\right\} A^\top = A\ C_{xx} A^\top \tag{1.39}$$

Eq. (1.39) is the *general error propagation law in matrix form*, or the *propagation law of variance-covariance matrices for a random vector with m linear functions of n random variables.*

If we have another random vector z :

$$\underset{p\cdot 1}{z} = \begin{bmatrix} z_1 \\ z_2 \\ \cdots \\ z_p \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{p1} & b_{p2} & \cdots & b_{pn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \underset{p\cdot n}{B}\ \underset{n\cdot 1}{x} \tag{1.40}$$

where:

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{p1} & b_{p2} & \cdots & b_{pn} \end{bmatrix} \tag{1.41}$$

The covariance matrix between y and z can be obtained as:

$$\underset{m\cdot p}{C_{yz}} = E\left\{[y - E(y)][z - E(z)]^\top\right\} = A\ E\left\{[x - E(x)][x - E(x)]^\top\right\} B^\top = AC_{xx}B^\top \tag{1.42}$$

which is also called the *covariance matrix propagation law for two random vectors of linear functions of random variables.*

1.2.2 Error Propagation in Non-Linear Functions

Assume that n observations x_1, x_2, $\cdots$, x_n have true errors ε_1, ε_2, $\cdots$, ε_n, respectively, and that vector $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)$ has the variance-covariance matrix:

$$\underset{n\cdot n}{C_{xx}} = \underset{n\cdot n}{C_{\varepsilon\varepsilon}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \tag{1.43}$$

Let y denote an arbitrary function of x_1, x_2, $\cdots$, x_n :

$$y = f(x_1, x_2, \cdots, x_n) \tag{1.44}$$

Let x_i^0 denote an approximate value of x_i $(1 \le i \le n)$ and y^0 denote the approximate value of y derived from x_i^0 :

$$y^0 = f(x_1^0, x_2^0, \cdots, x_n^0) \tag{1.45}$$

In addition, let δx_i denote a small variation in x_i with respect to x_i^0 and δy the resulted change of y:

$$\left.\begin{array}{ll} x_i = x_i^0 + \delta x_i & (1 \le i \le n) \\ y = y^0 + \delta y & \end{array}\right\} \tag{1.46}$$

Then Eq. (1.44) can be expanded into a Taylor series around the approximate values x_i^0 $(1 \le i \le n)$:

$$y = f(x_1^0 + \delta x_1,\ x_2^0 + \delta x_2,\ \cdots,\ x_n^0 + \delta x_n) =$$

$$= f(x_1^0, x_2^0, \cdots, x_n^0) + \frac{1}{1!}\left[\frac{\partial f}{\partial x_1}\delta x_1 + \frac{\partial f}{\partial x_2}\delta x_2 + \cdots + \frac{\partial f}{\partial x_n}\delta x_n\right]$$

$$+\frac{1}{2!}\left[\frac{\partial^2 f}{\partial x_1^2}\ \delta x_1^2 + \frac{\partial^2 f}{\partial x_1 \partial x_2}\ \delta x_1\ \delta x_2 + \cdots + \frac{\partial^2 f}{\partial x_1 \partial x_n}\ \delta x_1\ \delta x_n\right] + \cdots$$

where the partial derivatives are calculated at the approximate values x_i^0. If all x_i^0 are good enough, the last equation above may be approximated by (after ignoring all non-linear terms of δx_i) :

$$\delta y \approx \frac{\partial f}{\partial x_1}\ \delta x_1 + \frac{\partial f}{\partial x_2}\ \delta x_2 + \cdots + \frac{\partial f}{\partial x_n}\ \delta x_n$$

The above equation expresses the linear relationship between the small variations δx_i in x_i $(i = 1, 2, 3, \cdots, n)$ and the corresponding small variation $in y$. Replacing the small variations δy and δx_i by the corresponding errors $-\varepsilon_y$ and $-\varepsilon_i$, we arrive at a linear relationship between the errors in the variables x_i and the resulted error in function y:

$$\varepsilon_y \approx \frac{\partial f}{\partial x_1}\varepsilon_1 + \frac{\partial f}{\partial x_2}\varepsilon_2 + \cdots + \frac{\partial f}{\partial x_n}\varepsilon_n = a_1\varepsilon_1 + a_2\varepsilon_2 + \cdots + a_n\varepsilon_n = a \cdot \varepsilon \tag{1.47}$$

where

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} , \quad a = (a_1,\ a_2,\ \cdots,\ a_n) , \quad a_i = \frac{\partial f}{\partial x_i} , \quad i = 1, 2, \cdots, n \tag{1.48}$$

Eq. (1.47) expresses the relation between the true errors of y and x_i, corresponding to Eq. (1.33) for a linear function. Therefore the variance σ_y^2 of y follows directly from Eq. (1.35) :

$$\sigma_y^2 = E\left\{[y - E(y)]^2\right\} = a\ C_{xx} a^\top \tag{1.49}$$

If x_1, x_2, $\cdots$, x_n are uncorrelated with each other, *i.e.* $\sigma_{ij} = 0$ for $i \neq j$, Eq. (1.49) reduces to:

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_1^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_2^2 + \cdots + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_n^2 \tag{1.50}$$

Eqs. (1.49) and (1.50) express the *error propagation law for a non-linear function of n random variables.*

Example 1.3

A slope distance has been measured to $\ell_1 = 103.132$ metres with standard error $\sigma_1 = \pm 3\ mm$. The corresponding vertical angle has been measured to $\ell_2 = 45^0$ with standard error $\sigma_2 = \pm 6''$ (Cf Figure 1.1). What is the standard error (σ_d) of horizontal distance $d = \ell_1 \cos \ell_2$?

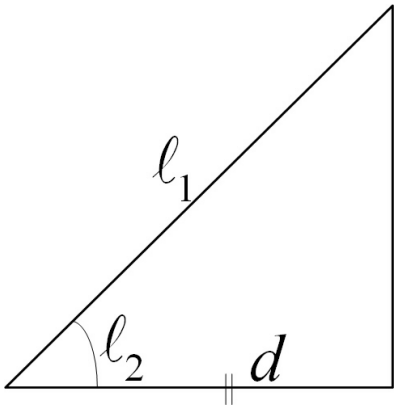

Figure 1.1: Determination of a horizontal distance

This is an error propagation problem for a non-linear function $d = f(\ell_1,\ \ell_2)$:

$\partial f/\partial \ell_1 = \cos \ell_2 = \cos 45^0 = 1/\sqrt{2}$

$\partial f/\partial \ell_2 = -\ell_1 \sin \ell_2 = -103.132\ \sin 45^0 = 103132/\sqrt{2}\ (mm)$

$\sigma_d{}^2 = (\partial f/\partial \ell_1)^2\ \sigma_1^2 + (\partial f/\partial \ell_2)^2\ \sigma_2^2 = (1/\sqrt{2})^2 (3\ mm)^2 + (103132/\sqrt{2})^2 (6''/206265'')^2 = 9\ (mm^2)$

$\sigma_d = \pm 3\ mm.$

Note that constant $206265''$ in the denominator above is used to convert angular quantity σ_2 in arc second ($''$) to radian. This conversion is often necessary when angles and other types of quantities (*e.g.* distances) are involved in the same formula. ∎

The error propagation law can be easily extended to the case with a vector y of m non-linear functions :

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2, \cdots, x_n) \\ f_2(x_1, x_2, \cdots, x_n) \\ \cdots \\ f_m(x_1, x_2, \cdots, x_n) \end{bmatrix} \tag{1.51}$$

Following the same procedure used to derive Eq. (1.47), we can obtain an approximate relation between the error vector ε_y of y and error vector ε of $x = (x_1, x_2, \cdots, x_n)^\top$:

$$\varepsilon_y = \begin{bmatrix} \varepsilon_{y_1} \\ \varepsilon_{y_2} \\ \cdots \\ \varepsilon_{y_m} \end{bmatrix} = \underset{m\cdot n}{A} \; \underset{n\cdot 1}{\varepsilon} \tag{1.52}$$

where:

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{m\cdot n}{A} = \begin{bmatrix} a_{12} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad a_{ij} = \frac{\partial f_i}{\partial x_j} . \quad (i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n) \tag{1.53}$$

The variance matrix of y follows from Eqs. (1.37) and (1.39):

$$C_{yy} = E\left\{[y - E(y)][y - E(y)]^\top\right\} = A\, C_{xx} A^\top \tag{1.54}$$

Comparing Eqs. (1.35) and (1.39) with Eqs. (1.49) and (1.54), respectively, one can see that the error propagation law takes the same forms for both linear and non-linear functions. The only difference is that one needs to calculate the derivatives of the non-linear functions before applying the error propagation law.

Example 1.4

To determine the coordinates (x, y) of an unknown point P (see Fig. 1.2), two angles ℓ_1 and ℓ_2 have been observed independently from two fixed points $A(x_a, y_a)$ and B with equal standard errors $\sigma_1 = \sigma_2 = \sigma = \pm 2''$. The observed angles are $\ell_1 = \ell_2 = 60^0$. The length of the fixed baseline $\overline{AB}$ is $s_0 = 750.000\ metres$, while the fixed azimuth from A to B is $A_0 = 120^0$. We now want to find out the variance-covariance matrix of the coordinates (x, y) at point P.

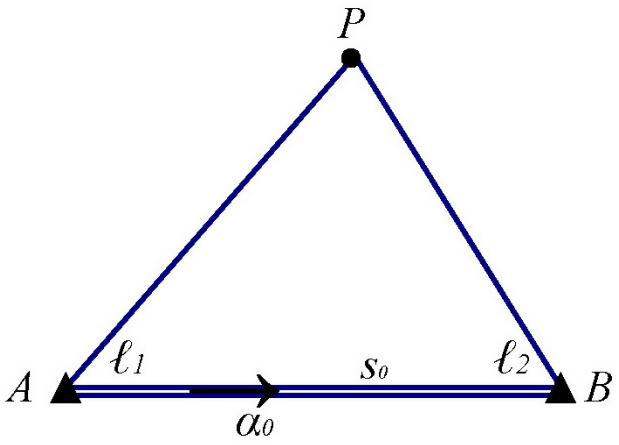

Figure 1.2: Single point positioning by angle measurements

The coordinates of P can be found:

$$x = x_a + s_0 \frac{\sin \ell_2 \ \cos(A_0 - \ell_1)}{\sin(\ell_1 + \ell_2)}$$

$$y = y_a + s_0 \frac{\sin \ell_2 \ \sin(A_0 - \ell_1)}{\sin(\ell_1 + \ell_2)}$$

Let ε_x and ε_y denote the true errors of x and y, and let ε_1, ε_2 denote the true errors of ℓ_1, ℓ_2. From Eq. (1.52), we have:

$$\begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} = A \cdot \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} , \quad \underset{2\cdot 2}{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

where :

$$a_{11} = \frac{\partial x}{\partial \ell_1} = -\frac{s_0}{\rho''} \frac{\sin \ell_2 \cos(A_0 + \ell_2)}{\sin^2(\ell_1 + \ell_2)}$$

$$a_{12} = \frac{\partial x}{\partial \ell_2} = +\frac{s_0}{\rho''} \frac{\sin \ell_1 \cos(A_0 - \ell_1)}{\sin^2(\ell_1 + \ell_2)}$$

$$a_{21} = \frac{\partial y}{\partial \ell_1} = -\frac{s_0}{\rho''} \frac{\sin \ell_2 \sin(A_0 + \ell_2)}{\sin^2(\ell_1 + \ell_2)}$$

$$a_{22} = \frac{\partial y}{\partial \ell_2} = +\frac{s_0}{\rho''} \frac{\sin \ell_1 \sin(A_0 - \ell_1)}{\sin^2(\ell_1 + \ell_2)}$$

where $\rho'' \approx 206\ 265''$ is used to convert ε_1, ε_2 from radian to arc seconds. For $\ell_1 = \ell_2 = 60^0$ and $A_0 = 120^0$, we have :

$$A = \frac{s_0}{0.75 \cdot \rho''} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{4} \\ 0 & \frac{3}{4} \end{bmatrix}$$

The variance-covariance matrix of ε_1, ε_2 has a diagonal structure:

$$C_{\varepsilon\varepsilon} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (\sigma^2 = 4\ ''^2)$$

From the error propagation law (1.54), we can obtain the variance matrix of the coordinates of point P :

$$C_p = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} = A\, C_{\varepsilon\varepsilon}\, A^\top = \sigma^2\, AA^\top = \left(\frac{\sigma\ s_0}{0.75 \cdot \rho''} \right)^2 \frac{3}{16} \begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix} \tag{1.55}$$

For $s_0 = 750\ m$ and $\sigma = \pm 2''$, the above equation gives :

$$C_p = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 88.14 & 30.53 \\ 30.53 & 52.88 \end{bmatrix} \quad (mm^2)$$

or:

$$\sigma_x = \sqrt{88.14} = \pm 9.4\ mm\ , \quad \sigma_y = \sqrt{52.88} = \pm 7.3\ mm. \quad \blacksquare$$

1.2.3 Propagation of Weights

Consider again a linear function y of n random variables $x_i\ (1 \le i \le n)$:

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = a \cdot x \tag{1.56}$$

Let σ_y denote the standard error of y and σ_i the standard error of x_i. For a specific unit-weight standard error σ_0, the weights of y and x_i will be according to Eq. (1.9):

$$p_y = \frac{\sigma_0^2}{\sigma_y^2} , \quad p_i = \frac{\sigma_0^2}{\sigma_i^2} \tag{1.57}$$

which implies:

$$\sigma_y^2 = \frac{\sigma_0^2}{p_y}\ , \quad \sigma_i^2 = \frac{\sigma_0^2}{p_i} \tag{1.58}$$

If x_1, x_2, $\cdots$, x_n are uncorrelated with each other, σ_y can be calculated from Eq. (1.36). Inserting Eq. (1.58) into Eq. (1.36) gives :

$$\frac{\sigma_0^2}{p_y} = a_1^2\frac{\sigma_0^2}{p_1} + a_2^2\frac{\sigma_0^2}{p_2} + \cdots + a_n^2\frac{\sigma_0^2}{p_n}$$

or:

$$\frac{1}{p_y} = \frac{a_1^2}{p_1} + \frac{a_2^2}{p_2} + \cdots + \frac{a_n^2}{p_n} \tag{1.59}$$

Eq. (1.59) expresses the weight of a linear function of n random variables by the weights of the random variables and is called the *reciprocal weight propagation law for a linear function of n independent random variables.*

1.2.4 Propagation of Cofactor Matrices

Let x be an n-dimensional random vector with variance-covariance matrix C_{xx}. Assume that for a unit-weight standard error σ_0 , the cofactor matrix and weight matrix of x are Q_{xx} and P_{xx}, respectively, i.e.:

$$C_{xx} = \sigma_0^2 \cdot Q_{xx} = \sigma_0^2 \cdot P_{xx}^{-1} \tag{1.60}$$

Let y be another vector consisting of linear functions of x :

$$\underset{m\cdot 1}{y} = \underset{m\cdot n}{A}\ \underset{n\cdot 1}{x} \tag{1.61}$$

where A is a given matrix:

$$\underset{m\cdot n}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Let C_{yy}, P_{yy} and Q_{yy} denote the variance matrix, the weight matrix and cofactor matrix of y, respectively:

$$C_{yy} = \sigma_0^2 \cdot Q_{yy} = \sigma_0^2 \cdot P_{yy}^{-1} \tag{1.62}$$

From the error propagation law Eq. (1.39), the following relation holds:

$$C_{yy} = A\ C_{xx}A^\top = A\left[\sigma_0^2 \cdot Q_{xx}\right]A^\top$$

Comparing the last two equations leads to:

$$Q_{yy} = P_{yy}^{-1} = AQ_{xx}A^\top \tag{1.63}$$

Eq. (1.63) expresses the cofactor matrix (or inverse weight matrix) of a vector of linear functions of random variables by the cofactor matrix (or inverse weight matrix) of the random variables. It is called the *propagation law of the cofactor matrix for linear functions.* As for non-linear functions, one should first linearize the non-linear functions and then apply Eq. (1.63), just as in the case of error propagation for non-linear functions.

It can be easily proved that if the n elements of vector x are uncorrelated with each other, Eq. (1.63) reduces to Eq. (1.59) for each element of y . That is to say, the simple reciprocal weight propagation law Eq. (1.59) is a special case of the general propagation law in Eq. (1.63).

Example 1.5

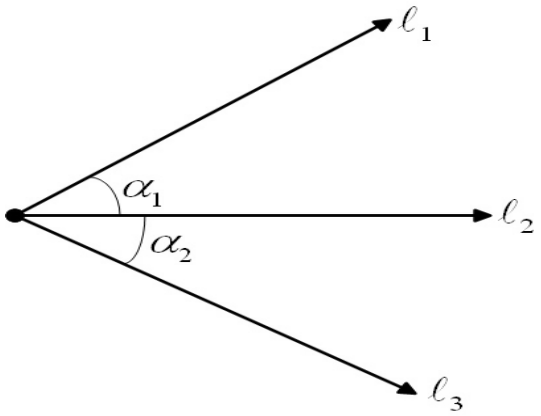

Figure 1.3: Three directions observed from a theodolite station

Let $L = (\ell_1, \ell_2, \ell_3)^\top$ denote 3 uncorrelated directions (clockwise) observed at a theodolite station with equal standard errors $\sigma_1 = \sigma_2 = \sigma_3 = \pm 1.41''$.

Let $\alpha = (\alpha_1, \alpha_2)^\top$ denote the two internal angles formed by the three directions, i.e.

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} -1 & +1 & 0 \\ 0 & -1 & +1 \end{bmatrix} \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{bmatrix} = A \cdot L$$

where :

$$A = \begin{bmatrix} -1 & +1 & 0 \\ 0 & -1 & +1 \end{bmatrix}, \quad L = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{bmatrix}$$

The variance matrix of L is simply a diagonal matrix:

$$C_{LL} = \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \sigma_3^2 \end{bmatrix} \approx \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix}$$

If we choose the unit-weight standard error as $\sigma_0 = \pm 2''$, then from Eq. (1.60) we can obtain the cofactor matrix Q_{LL} and the weight matrix P_{LL} of vector L :

$$Q_{LL} = \frac{C_{LL}}{\sigma_0^2} = \begin{bmatrix} \frac{1}{2} & & \\ & \frac{1}{2} & \\ & & \frac{1}{2} \end{bmatrix}, \quad P_{LL} = (Q_{LL})^{-1} = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix}$$

That is to say, all three directions have weight 2 with respect to the chosen σ_0^2. Applying the error propagation law Eq. (1.39), we can get the variance matrix $C_{\alpha\alpha}$ of α :

$$C_{\alpha\alpha} = AC_{LL}A^\top = \begin{bmatrix} -1 & +1 & 0 \\ 0 & -1 & +1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ +1 & -1 \\ 0 & +1 \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$$

The cofactor matrix $Q_{\alpha\alpha}$ of vector α is obtained either from relation Eq. (1.60):

$$Q_{\alpha\alpha} = \frac{C_{\alpha\alpha}}{\sigma_0^2} = \frac{1}{2} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

or the cofactor propagation law Eq. (1.63):

$$Q_{aa} = A\, Q_{LL} A^\top = \begin{bmatrix} -1 & +1 & 0 \\ 0 & -1 & +1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & & \\ & \frac{1}{2} & \\ & & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -1 & 0 \\ +1 & -1 \\ 0 & +1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \tag{1.64}$$

Finally, the weight matrix P_{aa} of vector α is obtained from the inverse of $Q_{\alpha\alpha}$:

$$P_{\alpha\alpha} = Q_{\alpha\alpha}^{-1} = \frac{2}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{1.65}$$

It should be pointed out that *when random elements in a random vector (such as α) are correlated, the diagonal elements of the weight matrix of this random vector* ***does not*** *represent the weights of the corresponding elements. The weight of each element should be obtained from the* ***reciprocal*** *of the corresponding element in the cofactor matrix of the random vector.* In the case of the above angle vector α , the first diagonal element of the calculated weight matrix $P_{\alpha\alpha}$ is (see Eq.1.65):

$$\frac{2}{3} \cdot 2 = \frac{4}{3}$$

The correct weight p_{α_1} of angle α_1 should be calculated using the cofactor matrix $Q_{\alpha\alpha}$ in (1.64):

$$p_{\alpha_1} = \frac{1}{\frac{2}{2}} = 1$$

The above result can be verified by applying the weight propagation law (1.59) on the relation $\alpha_1 = \ell_2 - \ell_1$:

$$\frac{1}{p_{\alpha_1}} = \frac{(-1)^2}{p_{\ell_1}} + \frac{(+1)^2}{p_{\ell_2}} = \frac{1}{2} + \frac{1}{2} = 1 \ , \quad p_{\alpha_1} = 1$$

where p_{ℓ_1} and p_{ℓ_2} denote the weights of ℓ_1 and ℓ_2, respectively. ∎

1.3 Error Ellipse and Error Ellipsoid

Assume that the planar coordinates (x, y) of a point P have the following variance-covariance matrix :

$$C = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} = \sigma_0^2 Q = \sigma_0^2 \cdot \begin{bmatrix} Q_{xx} & Q_{xy} \\ Q_{yx} & Q_{yy} \end{bmatrix} \tag{1.66}$$

where σ_0^2 denotes the variance factor (unit-weight standard error sqaured) and Q denotes the cofactor matrix of (x, y). In geodesy and surveying, one often calculates the total point error of P as:

$$\sigma_p = \sqrt{\sigma_x^2 + \sigma_y^2} = \sigma_0 \sqrt{Q_{xx} + Q_{yy}} \tag{1.67}$$

In some engineering projects (such as digging underground tunnels or building bridges), it is very important to obtain high accuracy in a specific direction (e.g. in the direction vertical to the tunnel). However, neither σ_x and σ_y nor σ_p can provide point accuracy in a specific direction. In this section we are going to study the standard error of an unknown point at any arbitrary direction, based on the variance-covariance matrix of the coordinates of the point.

1.3.1 Point Errors

Let $(\varepsilon_x,\ \varepsilon_y)$ denote the two error components of planar coordinates (x, y) of point P. The projection of this error vector at an arbitrary direction of azimuth α will be:

$$\varepsilon_\alpha = \cos\alpha\ \varepsilon_x + \sin\alpha\ \varepsilon_y = [\cos\alpha,\ \sin\alpha] \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} \tag{1.68}$$

Applying the error propagation law Eq. (1.39) on the above linear function ε_α , we can obtain the variance of ε_α at the direction α :

$$\sigma_\alpha^2 = [\cos\alpha,\ \sin\alpha]\ \ \sigma_0^2 \begin{bmatrix} Q_{xx} & Q_{xy} \\ Q_{yx} & Q_{yy} \end{bmatrix} \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} = \sigma_0^2 \left\{\cos^2\alpha \cdot Q_{xx} + \sin^2\alpha \cdot Q_{yy} + \sin 2\alpha \cdot Q_{xy}\right\} \tag{1.69}$$

Eq.(1.69) shows that variance σ_α^2 changes with the azimuth α. By letting the derivative of σ_α^2 with respect to α be zero, we can find the directions (α_0) of maximum and minimum variances, respectively:

$$\tan 2\alpha_0 = \frac{2Q_{xy}}{Q_{xx} - Q_{yy}} = \frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2} \tag{1.70}$$

As $\tan 2\alpha_0 = \tan 2(\alpha_0 + 90^0)$, the above equation has two solutions which correspond to the minimum variance direction and the maximum variance direction, respectively. These two directions (at azimuths α_0 and $\alpha_0 + 90^0$) are perpendicular to each other. Substitution of α_0 solved from Eq. (1.70) into Eq.(1.69) gives:

$$\sigma_{\alpha_0}^2 = \frac{1}{2}\sigma_0^2 \left\{Q_{xx} + Q_{yy} + 2(\cot^2 2\alpha_0 + 1)Q_{xy} \sin 2\alpha_0\right\} \tag{1.71}$$

From the above equation, we can deduct that:

- when $Q_{xy} > 0$ or $\sigma_{xy} > 0$:
  - the maximum variance direction (α_e) is in the 1st and 3rd quadrants, *i.e.* $0 < \alpha_e < 90^0$ or $180^0 < \alpha_e < 270^0$;
  - the minimum variance direction (α_f) is in the 2nd and 4th quadrants, *i.e.* $90 < \alpha_f < 180^0$ or $270^0 < \alpha_f < 360^0$;
- when $Q_{xy} < 0$ or $\sigma_{xy} < 0$:

- the maximum variance direction (α_e) is in the 2nd and 4th quadrants, *i.e.* $90 < \alpha_e < 180^0$ or $270^0 < \alpha_e < 360^0$;
- the minimum variance direction (α_f) is in the 1st and 3rd quadrants, *i.e.* $0 < \alpha_f < 90^0$ or $180^0 < \alpha_f < 270^0$.

The maximum variance E^2 when $\alpha = \alpha_e$ and the minimum variance F^2 when $\alpha = \alpha_f$ can be found :

$$E^2 = \frac{1}{2}\sigma_0^2 \left\{ Q_{xx} + Q_{yy} + \sqrt{(Q_{xx} - Q_{yy})^2 + 4Q_{xy}^2} \right\} \tag{1.72}$$

$$F^2 = \frac{1}{2}\sigma_0^2 \left\{ Q_{xx} + Q_{yy} - \sqrt{(Q_{xx} - Q_{yy})^2 + 4Q_{xy}^2} \right\} \tag{1.73}$$

or:

$$E^2 = \frac{1}{2} \left\{ \sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2} \right\} \tag{1.74}$$

$$F^2 = \frac{1}{2} \left\{ \sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2} \right\} \tag{1.75}$$

One can also express σ_α^2 for any arbitrary α by E and F :

$$\sigma_\alpha^2 = E^2 \ \cos^2\psi + F^2 \ \sin^2\psi \tag{1.76}$$

where $\psi = \alpha - \alpha_e$. In practical computations, σ_0^2 used above can be replaced by the estimated variance factor $\widehat{\sigma}_0^2$.

One can compare point error σ_p defined in Eq.(1.67) with σ_α defined above :

$$\sigma_p{}^2 = \sigma_x^2 + \sigma_y^2 = \sigma_0^2 \cdot (Q_{xx} + Q_{yy}) = E^2 + F^2 \tag{1.77}$$

$$\sigma_p{}^2 - \sigma_\alpha^2 = (E^2 + F^2) - (E^2 \ \cos^2\psi + F^2 \ \sin^2\psi) = \sin^2\psi \ E^2 + \cos^2\psi \ F^2 \geq 0 \tag{1.78}$$

Eq.(1.78) indicates that the traditional point error σ_p gives too pessimistic point error estimates.

1.3.2 Error Curve and Error Ellipse

Given the variance-covariance matrix $\sigma_0^2 \, Q$ of an unknown point P, one can calculate by Eq. (1.69) the standard error σ_α of P at an arbitrary azimuth α. If we plot all points with polar coordinates (α, σ_α) for $0 \leq a \leq 360^0$, we will get a closed curve, called the *error curve* of point P. As shown in Figure 1.4, error curve is symmetrical with respect to the axes of maximum and minimum variances. The distance from the curve centre O to a point on the curve in the direction α is the standard error of the point in that direction.

The error curve is not convenient for practical use, as neither Eq. (1.69) nor (1.76) represents a simple mathematical function. However, being very close to an ellipse, the error curve can be approximated by an ellipse whose semi-major axis is equal to E along the direction of azimuth α_e, and whose semi-minor axis is equal to F along the direction of azimuth α_f. Such an ellipse is called the *error ellipse* of point P. Using the directions of E and F as coordinate axes, the error ellipse will follow the following analytical equation :

$$\frac{\varepsilon_x{}^2}{E^2} + \frac{\varepsilon_y{}^2}{F^2} = 1 \tag{1.79}$$

Geometrically, one can obtain σ_α on the error ellipse. At any direction OA of azimuth α (see Figure 1.4), one can find a tangent line to the error ellipse at point B which intersects OA at point C, perpendicularly.

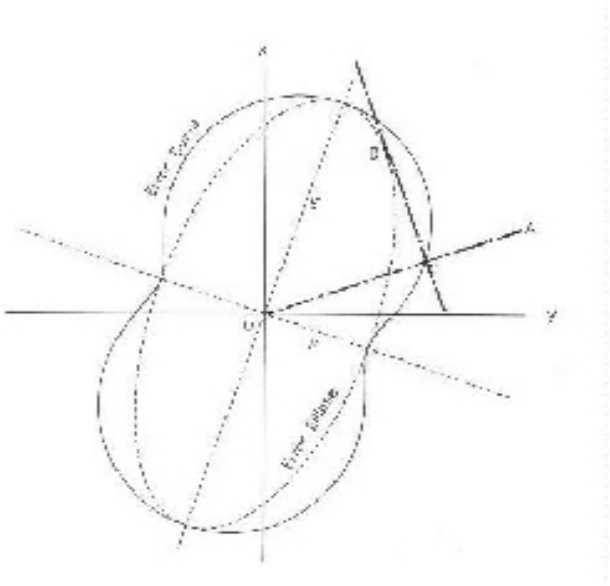

Figure 1.4: Error curve and error ellipse

Point C is on the error curve of azimuth α and thus the length of OC is equal to σ_α . The related proof is left to interested readers.

It can be also proved that the maximum and minimum variances E^2 and F^2 correspond to the maximum and minimum eigenvalues of the variance-covariance matrix C, respectively, and that directions α_e and α_f represent the directions of the corresponding eigenvectors. That is to say, E^2 and F^2 are the solutions of the following equation:

$$|C - \lambda\ I| = 0 \tag{1.80}$$

where λ denotes the eigenvalue of matrix C, I denotes the unit matrix and $|\cdot|$ denotes the matrix determinant. Similarly, α_e and α_f are the directions of eigenvectors v satisfying:

$$(C - \lambda I) \cdot v = 0 \tag{1.81}$$

Example 1.6

We consider the error ellipse of the unknown point P as illustrated in Fig. 1.2 of **Example 1.4**. The variance-covariance matrix of point $P's$ coordinates has already been found :

$$C_p = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 88.14 & 30.53 \\ 30.53 & 52.88 \end{bmatrix} \ (mm^2)$$

From Eqs. (1.74) and (1.75), we can calculate the semi-major and semi-minor axis, respectively, of the error ellipse :

$$E^2 = \tfrac{1}{2}\left\{\sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}\right\} = 105.68 \ \ mm^2\ , \quad E = 10.3\ mm$$

$$F^2 = \tfrac{1}{2}\left\{\sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}\right\} = \ 35.28 \ \ mm^2\ , \quad F = 5.9\ mm.$$

From Eq. (1.70), we have:

$$\tan 2\alpha_0 = \tfrac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2} = 1.732074\ , \quad \alpha_0 = 30^0 \text{ (or } 210^0)$$

As $\sigma_{xy} > 0$, the azimuths of E and F are as follows: $\alpha_e = 30^0$ (or 210^0) and $\alpha_f = 120^0$ (or 300^0).

From Eq.(1.55) of **Example 1.4**, we can see that the variance-covariance matrix C_p and accordingly E and F are proportional to the length (s_0) of the fixed baseline as well as the standard error (σ) of angle measurements. Another factor which also affects the size and orientation of point $P's$ error ellipse is the geometry of the network, namely the two angles ℓ_1 and ℓ_2. To see the effect of the geometry, we have repeated the computations of C_p in **Example 1.4** for different values of ℓ_1 and ℓ_2 and then repeated the computation of error ellipse parameters given by Eqs.(1.74) and (1.75). As in **Example 1.4**, $\ell_1 = \ell_2$, $\sigma_1 = \sigma_2 = \sigma = \pm 2''$, $s_0 = 750\ m$ and $A_0 = 120^0$ are assumed. The results of these computations are listed in Table (1.2).

Table 1.2: Error Ellipse Parameters of Point P

$\ell_1 = \ell_2\ (^0)$	$E\ (mm)$	$F\ (mm)$	$\alpha_e\ (^0)$	$\alpha_f\ (^0)$
15	10.3	2.8	120	30
20	8.0	2.9	120	30
25	6.7	3.1	120	30
30	5.9	3.4	120	30
35	5.5	3.8	120	30
40	5.2	4.4	120	30
45	5.1	5.1	120 or 30	30 or 120
50	6.2	5.2	30	120
55	7.8	5.5	30	120
60	10.3	5.9	30	120
65	14.4	6.7	30	120
75	38.4	10.3	30	120

From Table (1.2), one can notice that:

- For different sizes of angles, the semi-major and semi-minor axis are always parallel with or vertical to the fixed baseline AB (which has azimuth $A_0 = 120^0$). This is due to the assumptions that $\ell_1 = \ell_2$ and that $\sigma_1 = \sigma_2$. It can be proved that this is true independent of the azimuth A_0 of the baseline AB.
- When $\ell_1 = \ell_2 < 45^0$, the major axis E of the error ellipse is parallel with the fixed baseline while the minor axis F is vertical to the baseline. When $\ell_1 = \ell_2 > 45^0$, E is vertical to the baseline but F is parallel with the baseline.
- E decreases and F increases as ℓ_1 and ℓ_2 increase until they reach 45^0. At $\ell_1 = \ell_2 = 45^0$, E and F are equal to each other and the error ellipse becomes a circle.

The absolute magnitude of the above E, F may change if a different angle accuracy σ is assumed and/or the length of the fixed basline AB, which defines the scale of the network, is changed. Generally speaking, the orientation of the error ellipse axes is affected by the network geometry, and the relative accuracies of different observations.

A circular error ellipse implies that the point position has the same accuracy at any direction. This is a good property for geodetic networks of general purposes (such as the national triangulation networks). In this case, the networks are said to be homogeneous in accuracy. However, in some special situations one may require high accuracy at a specific direction (e.g. along the direction vertical to a tunnel). In the latter cases, we need to design our geodetic networks so that the minor axis F of the network points are roughly at the specific direction on which high accuracy is required, and that F is smaller than the prescribed error limit.

Example 1.6 has also shown that error ellipse parameters can be used in accuracy analysis of geodetic and photogrammetric networks. This can be made, both during the network planning stage and after field work is completed.

1.3.3 Probability inside Error Ellipses

Assume that the coordinates (x, y) of point P has a variance matrix C as defined in Eq. (1.66). If there are no systematic or gross errors in the determination of (x, y), the expectation of the errors $\varepsilon = (\varepsilon_x, \varepsilon_y)^\top$ in (x, y) will be zero. When measurement errors are (or are assumed to be) normally distributed, ε will then have the joint frequency function of the *bivariate normal distribution* :

$$f(\varepsilon_x, \varepsilon_y) = \frac{1}{2\pi\,|C|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\varepsilon^\top C^{-1}\varepsilon\right\} =$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{\varepsilon_x}{\sigma_x}\right)^2 - 2\rho\frac{\varepsilon_x}{\sigma_x}\frac{\varepsilon_y}{\sigma_y} + \left(\frac{\varepsilon_y}{\sigma_y}\right)^2\right]\right\} \tag{1.82}$$

where $|C|$ denotes the determinant of C and ρ is the *correlation coefficient* between ε_x and ε_y :

$$|C| = \sigma_x^2\sigma_y^2 - \sigma_{xy}^2 = \sigma_x^2\sigma_y^2(1-\rho^2), \quad \rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \tag{1.83}$$

If we let $f(\varepsilon_x, \varepsilon_y)$ be a constant, we will get the following equation (k^2 is an arbitrary positive constant):

$$\left(\frac{\varepsilon_x}{\sigma_x}\right)^2 - 2\rho\frac{\varepsilon_x}{\sigma_x}\frac{\varepsilon_y}{\sigma_y} + \left(\frac{\varepsilon_y}{\sigma_y}\right)^2 = k^2 \tag{1.84}$$

which geometrically represents an ellipse. For different choices of k, there will be different ellipses. Inside each of such ellipses, the point error $(\varepsilon_x, \varepsilon_y)$ will have different probabilities. Therefore, ellipses defined by Eq. (1.84) are also called the *error ellipses* of point P. This definition of error ellipse is in no contradiction with the definition given in Subsection 1.3.2, since after rotating ε_x- and ε_y-axis by angle α_0 solved from (1.70), Eq.(1.84) will be transformed to:

$$\frac{\varepsilon_x^2}{E^2} + \frac{\varepsilon_y^2}{F^2} = k^2 \tag{1.85}$$

which shows that the error ellipse derived from the error curve is a special case of Eq. (1.85), for $k = 1$.

The probability that point P falls inside the ellipse defined by (1.85) for given k is:

$$p = P\left\{\frac{\varepsilon_x^2}{E^2} + \frac{\varepsilon_y^2}{F^2} \le k^2\right\} = 1 - \exp^{-\frac{1}{2}k^2} \tag{1.86}$$

Numerical values of p for $k = 0, 1, 2, 3, 4$ are listed in Table (1.3) below.

Table 1.3: Probability inside error ellipses

k	p
0	0.0000
1	0.3935
2	0.8647
3	0.9889
4	0.9997

1.3.4 Error Ellipsoid

Let X, $\widetilde{X}$ and ε denote the observed, true three-dimensional coordinates of point P and its error, respectively:

$$\underset{3\times1}{\varepsilon} = \underset{3\times1}{X} - \underset{3\times1}{\widetilde{X}} \tag{1.87}$$

where:

$$\underset{3\times1}{\varepsilon} = \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{bmatrix}, \quad \underset{3\times1}{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \underset{3\times1}{\widetilde{X}} = \begin{bmatrix} \widetilde{x} \\ \widetilde{y} \\ \widetilde{z} \end{bmatrix}$$

Assume that ε (and also X) has the following variance-covariance matrix:

$$\underset{3\times3}{C_{xx}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \tag{1.88}$$

Let v denote an arbitrary direction, defined by three unit vector components e_1, e_2, e_3 :

$$\underset{3\times1}{v} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \tag{1.89}$$

Vector v can also be expressed by the horizontal angle h and vertical angle θ of vector v :

$$\underset{3\times1}{v} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} \cos\theta\ \cos h \\ \cos\theta\ \sin h \\ \sin\theta \end{bmatrix} \tag{1.90}$$

As v is a unit vector, the following condition must be held:

$$v^\top v = e_1^2 + e_2^2 + e_3^2 = 1 \tag{1.91}$$

The positional error of point P along the direction of v is the projection of ε onto v:

$$\varepsilon_v = v^\top \cdot \varepsilon = e_1 \cdot \varepsilon_x + e_2 \cdot \varepsilon_y + e_3 \cdot \varepsilon_z$$

By the error propagation law, the variance of ε_v follows:

$$\sigma_v^2 = v^\top C_{xx}\, v \tag{1.92}$$

To find the direction v at which σ_v^2 attains a maximum (or minimum) value, we form the conditional Lagange function:

$$f = v^\top C_{xx}\, v - \lambda\left(v^\top v - 1\right)$$

where λ is the Lagrange multiplier to be determined. Letting $\dfrac{df}{dv} = 0$, we arrive at a homogeneous equation system:

$$\left(\underset{3\times3}{C_{xx}} - \lambda \cdot \underset{3\times3}{I}\right) \underset{3\times1}{v} = 0 \tag{1.93}$$

Eq.(1.93) shows that the Lagrange multiplier λ and the unknown unit vector v are the *eigenvalue* and the corresponding *eigenvector* of the variance-covariance matrix C_{xx}. Inserting Eq.(1.93) into Eq.(1.92), we obtain the maximum (or minimum) variance of the position of point P :

$$\sigma_v^2 = v^\top \cdot \lambda v = \lambda \cdot v^\top v = \lambda \tag{1.94}$$

which means that the eigenvalue λ is nothing else but the maximum (or minimum) variance of point P. In order to obtain a non-trivial solution for v from Eq.(1.93), the coefficient matrix must have zero determinant, i.e.

$$| C_{xx} - \lambda \cdot I | = 0 \tag{1.95}$$

The above equation represents a third-order equation for λ. Consequently, there are three solutions $\lambda_1,\ \lambda_2,\ \lambda_3$ for λ, corresponding to three eigenvectors $v_1,\ v_2,\ v_3$:

$$\left.\begin{array}{l} (C_{xx} - \lambda_1 I)\ v_1 = 0 \\ (C_{xx} - \lambda_2 I)\ v_2 = 0 \\ (C_{xx} - \lambda_3 I)\ v_3 = 0 \end{array}\right\} \tag{1.96}$$

Considering that C_{xx} is a symmetrical matrix, the above three equations can be rewritten as:

$$C_{xx} = \left[\begin{array}{ccc} v_1 & v_2 & v_3 \end{array}\right] \left[\begin{array}{ccc} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{array}\right] \left[\begin{array}{c} v_1^\top \\ v_2^\top \\ v_3^\top \end{array}\right] = \underset{3\times3}{V} \cdot \underset{3\times3}{\Lambda} \cdot \underset{3\times3}{V^\top} \tag{1.97}$$

where V denotes the eigenvector matrix and Λ denotes the diangonal matrix of eigenvalues:

$$\underset{3\times3}{V} = \left[\begin{array}{ccc} v_1 & v_2 & v_3 \end{array}\right] = \left[\begin{array}{ccc} v_{11} & v_{12} & v_{13}. \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{array}\right],\quad \underset{3\times3}{\Lambda} = \left[\begin{array}{ccc} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{array}\right] \tag{1.98}$$

As C_{xx} is a symmetrical matrix, the eigenvector matrix V is an orthogonal matrix, i.e.

$$V^{-1} = V^\top, \qquad V V^\top = V^\top V = I \tag{1.99}$$

For a given variance-covariance matrix C_{xx}, its eigenvalues $\lambda_1,\ \lambda_2,\ \lambda_3$ and the corresponding eigenvectors $v_1,\ v_2,\ v_3$ can be numerically computed using, e.g., the *Jacobi's algorithm* (See e.g. Dahlquist, et.al., 1974). Eigenvalues and eigenvectors can also be easily obtained using the internal function $EIG(\cdot)$ of commercial programm package MATLAB.

Inserting Eq.(1.96) into Eq.(1.92), we obtain the maximum (or minimum) variances of the position of point P in three directions defined by the eigenvectors:

$$\left(\sigma_v^2\right)_i = v_i^\top \cdot \lambda_i v_i = \lambda_i \cdot v_i^\top v_i = \lambda_i \qquad (i = 1, 2, 3) \tag{1.100}$$

The position variance of point P at any arbitrary direction will be between the maximum and the minimum variances, and can be found from Eq.(1.92).

Similar to error ellipse in the two-dimensional case, the three eigenvalues of C_{xx} along the directions of the three orthogonal eigenvectors of C_{xx} construct an *error ellipsoid* for the position of point P in three-dimensional space:

$$\frac{x_1^2}{E^2} + \frac{x_2^2}{F^2} + \frac{x_3^2}{G^2} = 1 \tag{1.101}$$

where $E,\ F,\ G$ denote the three semi-axes of the ellispoid:

$$\left.\begin{array}{l} E = \sqrt{\lambda_1} \\ F = \sqrt{\lambda_2} \\ G = \sqrt{\lambda_3} \end{array}\right\} \tag{1.102}$$

Example 1.7

The three-dimensional coordinates (x, y, z) of a ground point has been determined through GPS observations with the following variance-covariance matrix:

$$\underset{3\times3}{C_{xx}} = \begin{bmatrix} 4.136\,531 & 2.319\,432 & 0.914\,233 \\ 2.319\,432 & 3.303\,108 & 0.614\,679 \\ 0.914\,233 & 0.614\,679 & 1.207\,910 \end{bmatrix} \quad (cm^2)$$

Matrix C_{xx} can be decomposed using eigevalues and eigenvectors (see Eq.(1.97)):

$$\underset{3\times3}{C_{xx}} = \underset{3\times3}{V} \cdot \underset{3\times3}{\Lambda} \cdot \underset{3\times3}{V^{\top}}$$

where:

$$\underset{3\times3}{V} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} 0.753\,100 & -0.574\,611 & -0.320\,408 \\ 0.623\,493 & 0.778\,795 & 0.068\,813 \\ 0.209\,992 & -0.251\,595 & 0.944\,777 \end{bmatrix}$$

$$\underset{3\times3}{\Lambda} = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} = \begin{bmatrix} 6.311\,713 & & \\ & 1.393\,206 & \\ & & 0.942\,630 \end{bmatrix}$$

The three semi-axes of the error ellipsoid of point P can then be obtained :

$$\begin{cases} E = \sqrt{\lambda_1} = 2.51\ cm \\ F = \sqrt{\lambda_2} = 1.18\ cm \\ G = \sqrt{\lambda_3} = 0.97\ cm \end{cases}$$

The directions of the axes of the error ellipsoid are obtained from the corresponding eigenvectors:

$$\begin{cases} h_1 = \arctan\left(\frac{v_{21}}{v_{11}}\right) = 39.621^0\ , \quad \theta_1 = \arcsin\left(v_{31}\right) = 12.122^0 \\ h_2 = \arctan\left(\frac{v_{22}}{v_{12}}\right) = 126.421^0\ , \quad \theta_2 = \arcsin\left(v_{32}\right) = -14.572^0 \\ h_3 = \arctan\left(\frac{v_{23}}{v_{13}}\right) = 167.879^0\ , \quad \theta_3 = \arcsin\left(v_{33}\right) = 70.870^0 \end{cases}$$

1.4 Statistical Analysis

Probability and mathematical statistics concerns studies of random variables, or stochastic variables. Most measurement errors in geodesy and surveying are random errors and thus can be analysed mathematically using the theory of probability and statistics. In this section, we discuss only several topics in statistical analysis which are directly used in geodesy, photogrammetry and surveying.

1.4.1 Probability Distributions

As most statistical methods are based on the probability distributions of random variables, we summarize below several commonly used distributions. For each distribution, an example of derived statistics from theory of errors is also given.

Normal Distribution

Normal distribution is of fundamental importance in theory of errors, because of the famous *central limiting theorem*, which states that *a sum of n independent random variables with equal expectation and variance will have a distribution that converges toward the normal distribution as $n \to \infty$.*

The density function $f(\varepsilon)$ (or frequency function) and distribution function $F(x)$ of a normally distributed random variable ε are as follows:

$$f(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon-\mu)^2\right\} \qquad (-\infty < \varepsilon < +\infty) \tag{1.103}$$

$$F(x) = \int_{-\infty}^{x} f(\varepsilon)d\varepsilon = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon-\mu)^2\right\} \cdot d\varepsilon \quad (-\infty < x < +\infty) \tag{1.104}$$

where μ, σ are two known constants. The expectation and variance of ε can be found:

$$E(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \varepsilon \cdot \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon-\mu)^2\right\} \cdot d\varepsilon = \mu \tag{1.105}$$

$$Var(\varepsilon) = E\left\{[\varepsilon - E(\varepsilon)]^2\right\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} [\varepsilon - E(\varepsilon)]^2 \cdot \exp\left\{-\frac{1}{2\sigma^2}(\varepsilon-\mu)^2\right\} \cdot d\varepsilon = \sigma^2 \tag{1.106}$$

i.e. the two constants μ and σ^2 are the expectation and variance of ε, respectively. Such a random variable is denoted often by $\varepsilon \sim N(\mu, \sigma^2)$. The density function $f(\varepsilon)$ given in Eq.(1.103) is plotted in Fig. 1.5 below.

Let τ be another random variable defined as:

$$\tau = \frac{\varepsilon - \mu}{\sigma} \tag{1.107}$$

where $\varepsilon \sim N(\mu, \sigma^2)$. Then it can be shown that τ is normally distributed with zero expectation and unit variance:

$$E(\tau) = 0\ , \quad Var(\tau) = 1 \tag{1.108}$$

The frequency function and distribution function of τ are:

$$f(\tau) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\tau^2\right\} \qquad (-\infty < \tau < +\infty) \tag{1.109}$$

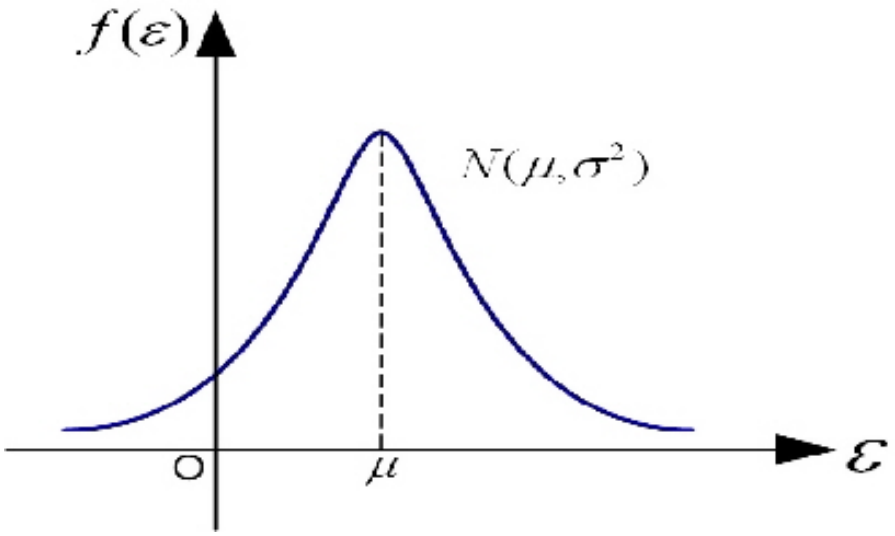

Figure 1.5: Density function of the normal distribution $N(\mu, \sigma^2)$

$$F(x) = \int_{-\infty}^{x} f(\tau)d\tau = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}\tau^2\right\} \cdot d\tau \quad (-\infty < x < +\infty) \tag{1.110}$$

τ is said to have the *standard normal distribution*, denoted as $\tau \sim N(0, 1)$. Numerical values of the distribution function $F(x)$ of the standard normal distribution τ are tabulated in statistical tables which can be found in most statistical books. Values for $x = 0, 1, 2, 3$ are listed in Table (1.4) below.

Table 1.4: Standard Normal Distribution Values

x	$F(x)$
0	0.50000
1	0.84134
2	0.97725
3	0.99865
4	0.99997

Let $x_i \sim N(\mu, \sigma^2)$ $(i = 1, 2, \cdots, n)$ be n independent normally distributed variables with equal expectation μ and equal variance σ^2. A sample mean $\overline{x}$ and an estimated sample variance $\widehat{\sigma}^2$ can be calculated :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ , \quad \widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{1.111}$$

One can show that $\overline{x}$ is also normally distributed:

$$\bar{x} \sim N(\mu, \ \frac{\sigma^2}{n}) \tag{1.112}$$

which leads to a new random variable of standard normal distribution:

$$u = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \tag{1.113}$$

The new statistics u constructed above can be used to test the sample mean $\overline{x}$ against the theoretical mean μ (expectation), when the standard error σ is known (Cf Subsection 1.4.3).

Let ε denote a random vector of n normally distributed variables:

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \tag{1.114}$$

with expectation and variance-covariance matrix as follows:

$$E(\varepsilon) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_n \end{bmatrix} = \underset{n\cdot 1}{\mu} , \quad \underset{n\cdot n}{C} = E\left\{ [\varepsilon - E(\varepsilon)] [\varepsilon - E(\varepsilon)]^\top \right\} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \tag{1.115}$$

then the joint frequency function of n dimensional normally distributed random vector ε is:

$$f(\varepsilon) = \frac{1}{(2\pi)^{\frac{n}{2}} |C|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\varepsilon - \mu)^\top C^{-1} (\varepsilon - \mu) \right\} \tag{1.116}$$

where $|C|$ stands for the determinant of the variance-covariance matrix C.

Binomial Distribution

Assume that a random experiment has two possible outcomes with probability p and $q = 1 - p$, respectively. The event that the outcome of probability p occurs k times ($0 \le k \le n$) when the experiment is repeated n times follows the binomial distribution. k will be a *discrete* random variable with the following frequency function:

$$f(k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} = \frac{n!}{(n-k)!\, k!} p^k q^{n-k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots 3\cdot 2\cdot 1} p^k q^{n-k} \quad (k = 0, 1, 2, \cdots, n) \tag{1.117}$$

The distribution function of k is:

$$F(x) = \sum_{k=0}^{x} f(k) = \sum_{k=0}^{x} \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} \tag{1.118}$$

with expectation and variance of x as follows:

$$E(x) = \sum_{k=0}^{n} k f(k) = np , \quad Var(x) = \sum_{k=0}^{n} \left\{ (k - np)^2 f(k) \right\} = npq \tag{1.119}$$

The number s_+ of positive errors (or negative errors) in a set of random measurement errors will have the binomial distribution with $p = q = \frac{1}{2} = 0.5$. See Subsection 6.1 in Chapter 6 for more details.

Uniform Distribution

The frequency function and distribution function of a uniform distribution are as follows (a, b are two constants):

$$f(\varepsilon) = \begin{cases} \frac{1}{b-a} & \text{for } a \le \varepsilon \le b \\ 0 & \text{otherwise} \end{cases} \tag{1.120}$$

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x > b \end{cases} \tag{1.121}$$

Uniform distribution has the following expectation and variance:

$$E(\varepsilon) = \int_a^b \varepsilon \frac{1}{b-a} d\varepsilon = \frac{a+b}{2} \tag{1.122}$$

$$Var(\varepsilon) = E\left\{[\varepsilon - E(\varepsilon)]^2\right\} = \frac{(b-a)^2}{12} \tag{1.123}$$

Rounding errors in numerical computations are random variables of uniform distribution with $a = -0.5$ and $b = +0.5$. The expectation and variance of rounding errors are:

$$E(\varepsilon) = 0, \quad Var(\varepsilon) = \frac{1}{12} \tag{1.124}$$

χ^2-Distribution

If $\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_n$ are independent of each other and all have standard normal distribution, i.e. $\varepsilon_i \sim N(0,1)$, then the following sum :

$$\chi^2 = \varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2 \tag{1.125}$$

has χ^2-distribution with n degrees of freedom, denoted as $\varepsilon \sim \chi^2(n)$. $\chi^2(n)$ has the following frequency function:

$$f(\chi^2) = \frac{1}{\frac{n}{2}\Gamma(\frac{n}{2})} \left(\chi^2\right)^{\frac{n}{2}-1} \exp\left\{-\frac{1}{2}\chi^2\right\}, \quad (\chi^2 > 0) \tag{1.126}$$

where the Γ-function is defined as:

$$\Gamma(n) = \int_0^{+\infty} x^{n-1} e^{-x} dx \quad (n > 0) \tag{1.127}$$

The distribution function of $\chi^2(n)$ is :

$$F(x) = \int_{-\infty}^{x} f(x) dx, \tag{1.128}$$

while expectation and variance of χ^2 can be derived :

$$E\left(\chi^2\right) = n\ , \quad Var\left(\chi^2\right) = E\left\{\left[\chi^2 - E(\chi^2)\right]^2\right\} = 2n \tag{1.129}$$

Let $x_i \sim N(\mu, \sigma^2)$ $(i = 1, 2, \cdots, n)$ be n independent normally distributed variables with equal expectation μ and equal variance σ^2. And let $\overline{x}$ and $\widehat{\sigma}^2$ be the sample mean and sample variance defined by (1.111). From the above data, one can construct a statistics of χ^2-distribution :

$$\frac{(n-1)\widehat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{\overline{x} - x_i}{\sigma}\right)^2 \sim \chi^2(n-1) \tag{1.130}$$

Considering Eq. (1.129), the expectation and variance of $\widehat{\sigma}^2$ can be found:

$$E\left(\widehat{\sigma}^2\right) = E\left\{\frac{\sigma^2}{n-1}\chi^2(n-1)\right\} = \sigma^2 \tag{1.131}$$

$$Var\left(\widehat{\sigma}^2\right) = E\left\{\left(\widehat{\sigma}^2 - \sigma^2\right)^2\right\} = \frac{2\sigma^4}{n-1} \tag{1.132}$$

The last two equations above indicate that the sample variance $\widehat{\sigma}^2$ is an unbiased estimate of the theoretical variance σ^2 and that the variance of this estimate can be large if n (size of the measurement sample) is too small.

t-Distribution

If $X \sim N(0,1)$ and $Y \sim \chi^2(n)$ are independent of each other, then

$$t = \frac{X}{\sqrt{Y/n}} \tag{1.133}$$

is a random variable of t-distribution with n degrees of freedom, denoted as $t \sim t(n)$, which approaches $N(0,\ 1)$ when $n \to \infty$ .The frequency function of $t(n)$ is:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \qquad (-\infty < t < +\infty) \tag{1.134}$$

and the expectation and variance are :

$$E(t) = 0 \ \ (n > 1), \qquad Var(t) = \frac{n}{n-2} \ \ (n > 2) \tag{1.135}$$

Let $x_i \sim N(\mu, \sigma^2)$ $(i = 1, 2, \cdots, n)$ be n independent normally distributed variables with equal expectation μ and equal variance σ^2 and let $\overline{x}$, $\widehat{\sigma}^2$ are calculated by Eq. (1.111). Taking into account Eqs. (1.112) and (1.130), a new statistics of $t-$distribution can be constructed:

$$t' = \frac{\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)\widehat{\sigma}^2}{\sigma^2} \cdot \frac{1}{n-1}}} = \frac{\overline{x}-\mu}{\widehat{\sigma}/\sqrt{n}} \sim t(n-1) \tag{1.136}$$

The above statistics t' can be used to test the sample mean $\overline{x}$ against the theoretical mean μ (expectation), when the standard error σ is unknown and thus an estimated standard error $\widehat{\sigma}$ is used (Cf Subsection 1.4.3).

F-Distribution

If X, Y are two independent χ^2-distributed random variables :

$$X \sim \chi^2(n)\ , \quad Y \sim \chi^2(m)$$

then the following derived random variable:

$$Z = \frac{X/n}{Y/m} \tag{1.137}$$

has F-distribution with $(n,\ m)$ degrees of freedom, designated as $Z \sim F(n, m)$. The frequency function, expectation and variance of Z are as follows :

$$f(z) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right) \left(\frac{n}{m}z\right)^{\frac{n}{2}-1} \left(1 + \frac{n}{m}z\right)^{-\frac{n+m}{2}} \qquad (z \geq 0) \tag{1.138}$$

$$E(z) = \frac{n}{m-2} \qquad (m > 2) \tag{1.139}$$

$$Var(z) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)} \qquad (m > 4) \tag{1.140}$$

Let $x_i \sim N(\mu_1, \sigma_1^2)$ $(i = 1, 2, \cdots,\ n)$ be a set of n independent normally distributed variables, and let $y_i \sim N(\mu_2, \sigma_2^2)$ $(i = 1, 2, \cdots, m)$ be another set of m independent normally distributed variables . The two sets are independent of each other, and for each set, a sample mean and an estimated sample variance can be calculated as in Eq. (1.111):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ , \quad \widehat{\sigma}_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{1.141}$$

$$\overline{y} = \frac{1}{m}\sum_{i=1}^{m} x_i \ , \quad \widehat{\sigma}_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \overline{y})^2 \tag{1.142}$$

Considering Eq. (1.130), we have:

$$\frac{(n-1)\widehat{\sigma}_1^2}{\sigma_1^2} \sim \chi^2(n-1) \ , \quad \frac{(m-1)\widehat{\sigma}_2^2}{\sigma_2^2} \sim \chi^2(m-1) \tag{1.143}$$

Then a new statistics of F-distribution can be constructed :

$$F' = \frac{\frac{(n-1)\widehat{\sigma}_1^2}{\sigma_1^2} \cdot \frac{1}{n-1}}{\frac{(m-1)\widehat{\sigma}_2^2}{\sigma_2^2} \cdot \frac{1}{m-1}} = \frac{\sigma_2^2 \cdot \widehat{\sigma}_1^2}{\sigma_1^2 \cdot \widehat{\sigma}_2^2} \sim F(n-1, m-1) \tag{1.144}$$

F' constructed above can be used to test if two samples have equal variance (Cf Subsection 1.4.3).

1.4.2 Confidence Intervals and Error Tolerances

For each given interval $[a, b]$ $(-\infty < a, b < +\infty)$, there is a uniquely defined probability α, called *risk level,* that the random measurement error ε will fall outside the interval. *Confidence interval* of ε at risk level α is defined as an interval $[a, b]$ containing the expectation μ of ε such that the probability that ε falls **outside** this interval is α. The probability that ε falls **inside** the confidence interval is equal to $1 - \alpha$, called the *confidence level.* It is clear that the size of the confidence interval is directly related to the risk level α. α is often expressed in % and $[a, b]$ is usually symmetrical with respect to μ, such as $[\mu - k\sigma, \mu + k\sigma]$ for positive k.

If $[\mu - k\sigma, \mu + k\sigma]$ is a confidence interval of ε at risk level α_k, we then have:

$$P\left(\mu - k\sigma \le \varepsilon \le \mu + k\sigma\right) = 1 - \alpha_k = p_k \tag{1.145}$$

If ε is normally distributed, i.e. $\varepsilon \sim N\left(\mu, \sigma^2\right)$, the new variable τ defined below will have standard normal distribution:

$$\tau = \frac{\varepsilon - \mu}{\sigma} \sim N(0, 1) \tag{1.146}$$

The probability (p_k) that ε falls inside the confidence interval $\mu - k\sigma \le \varepsilon \le \mu + k\sigma$ is the same as the probability that $\tau \sim N(0, 1)$ falls inside the interval $[-k, +k]$. Considering the fact that $f(\tau)$ is symmetrical around $\tau = 0$, which implies that the area under $f(\tau)$ over a symmetrical interval $[-k, +k]$ $(k > 0)$ is two times the area over interval $[0, +k]$, we obtain:

$$p_k = P\left(-k \le \tau \le +k\right) = 2\left[P\left(\tau \le +k\right) - P\left(\tau \le 0\right)\right] = 2\left[P\left(\tau \le +k\right) - \frac{1}{2}\right] = 2 \cdot P\left(\tau \le +k\right) - 1 \tag{1.147}$$

For $k = 1, 2, 3$, we have:

$$p_1 = P\left(-1 \le \tau \le +1\right) = 2 \cdot 0.84134 - 1 = 68.27\%$$

$$p_2 = P\left(-2 \le \tau \le +2\right) = 2 \cdot 0.97725 - 1 = 95.45\%$$

$$p_3 = P\left(-3 \le \tau \le +3\right) = 2 \cdot 0.99865 - 1 = 99.73\%$$

When p_k has been found for each confidence interval $[\mu - k\sigma \le \varepsilon \le \mu + k\sigma]$ or $[-k \le \tau \le +k]$, the corresponding risk level α_k follows directly:

$$\alpha_k = 1 - p_k \tag{1.148}$$

The confidence interval $[\mu - k\sigma \le \varepsilon \le \mu + k\sigma]$ can also be written as $\mu \pm k\sigma$. In Table (1.5) below, we have listed several confidence intervals and their risk levels for $k = 1, 2, 3$.

Table 1.5: Confidence Intervals of Normal Distributions

k	*interval for* ε	*interval for* τ	*confidence level* ($p_k = 1 - \alpha_k$)	*risk level* (α_k)
1	$\mu \pm 1\sigma$	± 1	68.27 %	31.73 %
2	$\mu \pm 2\sigma$	± 2	95.45 %	4.55 %
3	$\mu \pm 3\sigma$	± 3	99.73 %	0.27 %
4	$\mu \pm 4\sigma$	± 4	99.99 %	0.01 %

In geodesy and photogrammetry, one often sets 2 or 3 times the theoretical or estimated standard error σ as the accepted error tolerance. According to Table (1.5), the risk that the measurement error with the specified standard error exceeds these two limits will have probability less than 5% and 0.3%, respectively.

Example 1.8

Assume that a height difference has been measured two times, once forward (h_1) and once backward (h_2) with same accuracy. The standard error of the average ($\bar{h}$) of h_1 and h_2 is required to be less than 1 mm. Now we want to know how big the difference $d = h_1 - h_2$ is tolerated.

Let σ denote the standard error of h_1 and h_2 and let $\bar{\sigma}$ ($= \pm 1$ mm) and σ_d denote the standard errors of $\bar{h}$ and d, respectively. Then we have:

$$\bar{h} = (h_1 + h_2)/2$$
$$\bar{\sigma}^2 = (\tfrac{1}{2})^2\sigma^2 + (\tfrac{1}{2})^2\sigma^2 = \tfrac{1}{2}\sigma^2 \ , \ \ i.e. \ \ \sigma = \sqrt{2}\bar{\sigma}$$
$$d = h_1 - h_2$$
$$\sigma_d^2 = (+1)^2\sigma^2 + (-1)^2\sigma^2 = 2\ \sigma^2 \ , \ \ i.e. \ \ \sigma_d = \sqrt{2}\sigma = 2\bar{\sigma} = \pm 2\ mm$$

If we choose 0.3% risk level, the confidence interval of d will be $\mu \pm 3\sigma_d$ *or* $\pm 3\sigma_d$ for $\mu = 0$. Therefore, the maximum tolerated difference d_{max} should be:

$$d_{\max} \leq |3\sigma_d| = 6\ mm$$

If we choose 4.55% risk level instead, then we have:

$$d_{\max} \leq |2\sigma_d| = 4\ mm. \quad \blacksquare$$

1.4.3 Hypothesis Tests

Hypothesis tests in statistical analysis are designed to deduct on the theoretical distribution based on the sample data (measurements). Hypothesis tests try to compare statistics computed from the sample data with the critical value of the assumed distribution at certain *risk level* α in order to decide whether the sample comes from the assumed distribution. The key to hypothesis tests is to construct from the sample data (measurements) a statistics whose distribution is known under the null hypothesis . In Subsection 1.4.1, we have presented several statistics defined in Eqs. (1.113), (1.130), (1.136) and (1.144). Based on these statistics, the following three types of hypothesis tests are widely used in geodesy and photogrammetry.

Testing the expectation when the variance is known (two-way u-test)

Assume that x_1, x_2, $\cdots$, x_n are n independent measurements of the same normal distribution $N(\mu, \sigma^2)$. Their standard error σ is assumed to be known. From the measurement data, we can calculate the sample mean:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.149}$$

Now we want to test if the measurements-derived sample mean is statistically identical to the theoretical mean μ (expectation). If $x_i \sim N(\mu, \sigma^2)$ for $1 \leq i \leq n$, it follows:

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{and} \quad \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

At risk level α , we then have :

$$P\left\{\left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| \leq x_{\frac{1}{2}\alpha}\right\} = 1-\alpha \quad or: P\left\{|\bar{x}-\mu| \leq c_1\right\} = 1-\alpha \tag{1.150}$$

where $x_{\frac{1}{2}\alpha}$ is the critical value of the standard normal distribution for risk level $\frac{1}{2}\alpha$ and c_1 is :

$$c_1 = x_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}} \tag{1.151}$$

To test whether the sample mean $\overline{x}$ of the measurements is statistically identical to μ, we look at the two possible hypotheses:

Null hypothesis H_0 :	$\overline{x} = \mu$
Alternative hypothesis H_1 :	$\overline{x} \neq \mu$

After choosing risk level α, one can find the critical value $x_{\frac{1}{2}\alpha}$ from the standard normal distribution table and then calculate c_1. If $|\bar{x}-\mu| < c_1$, we can accept H_0 at risk level α. Otherwise, we reject H_0.

Testing the expectation when the variance is unknown (two-way $t-$test)

Assume that we have the same task as above, namely testing whether the sample mean of the measurements is equal to the theoretical expectation μ, except that the standard error σ is unknown. For this purpose, we consider the statistics of t-distribution given in Eq.(1.136):

$$\frac{\overline{x}-\mu}{\widehat{\sigma}/\sqrt{n}} \sim t(n-1) \tag{1.152}$$

where $\bar{x}$ and $\widehat{\sigma}$ are calculated by Eq. (1.111). Choosing a risk level α, we have:

$$P\left\{\frac{\overline{x}-\mu}{\widehat{\sigma}/\sqrt{n}} \leq t_{\frac{1}{2}\alpha}(n-1)\right\} = 1-\alpha \quad or: \; P\left\{|\bar{x}-\mu| \leq c_2\right\} = 1-\alpha \tag{1.153}$$

where $t_{\frac{1}{2}a}(n-1)$ is the critical value of t-distribution with $n-1$ degrees of freedom at risk level $\frac{1}{2}\alpha$ and c_2 is:

$$c_2 = t_{\frac{1}{2}\alpha}(n-1)\cdot\frac{\widehat{\sigma}}{\sqrt{n}} \tag{1.154}$$

For a specific risk level α, we can find the corresponding critical value $t_{\frac{1}{2}\alpha}$. If $|\bar{x}-\mu| < c_2$, we accept the null hypothesis H_0. Otherwise, the alternative hypothesis H_1 is accepted.

Example 1.9

A distance x has been measured 10 times with the following results (unit: mm):

i	1	2	3	4	5	6	7	8	9	10
x_i	1322	1324	1318	1319	1318	1323	1325	1328	1326	1317

Assume that all measurements are normally distributed and that we want to test whether this distance is statistically equal to 1323 mm at risk level $\alpha = 5\%$ (i.e. 95% confidence level). Our hypotheses are:

$$H_0 : \overline{x} = \mu = 1323\ mm; \quad H_1 : \overline{x} \neq \mu = 1323\ mm$$

The sample mean and sample standard error are $(n = 10)$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 1322\ mm; \quad \widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 14.67\ mm^2 \ \ or : \widehat{\sigma} = 3.83\ mm$$

At risk level $\alpha = 5\%$, we have :

$$t_{\frac{1}{2}\alpha}(n-1) = 2.26\ , \ \ c_2 = t_{\frac{1}{2}\alpha}(n-1)\frac{\widehat{\sigma}}{\sqrt{n}} = 2.26\ \cdot \frac{3.83}{\sqrt{10}} = 2.74\ , \ \ \bar{x} - \mu = 1322 - 1323 = -1\ mm$$

As $|\bar{x} - \mu| < c_2$, we accept that the distance is statistically equal to 1323 mm at 5% risk level. ∎

Testing whether two distributions have equal variance (two-way $F-$test)

Assume that we have obtained two independent sets of normally distributed measurements:

$$x_i \sim N(\mu_1, \sigma_1^2)\ , \ \ i = 1, 2, \cdots, n \quad \text{and} \quad y_i \sim N(\mu_2, \sigma_2^2)\ , \quad i = 1, 2, \cdots, m \tag{1.155}$$

and that we want to test the following two hypothesis:

$$H_0 :\ \sigma_1^2 = \sigma_2^2\ ; \quad H_1 : \sigma_1^2 \neq \sigma_2^2 \tag{1.156}$$

For this purpose, we calculate first the sample means $(\overline{x},\ \overline{y})$ and sample variances $(\widehat{\sigma}_1^2;\ \widehat{\sigma}_2^2)$ for each set of measurements using Eqs. (1.141) and (1.142). From Eq. (1.144), we obtain then a statistics of $F-$distribution :

$$\frac{\sigma_2^2 \cdot \widehat{\sigma}_1^2}{\sigma_1^2 \cdot \widehat{\sigma}_2^2} \sim F(n-1, m-1) \tag{1.157}$$

For risk level α, we then have :

$$P\left\{F_{1-\frac{1}{2}\alpha}(n-1, m-1) < \frac{\sigma_2^2 \cdot \widehat{\sigma}_1^2}{\sigma_1^2 \cdot \widehat{\sigma}_2^2} < F_{\frac{1}{2}\alpha}(n-1, m-1)\right\} = 1 - \alpha \tag{1.158}$$

where $F_{1-\frac{1}{2}\alpha}(n-1, m-1)$ and $F_{\frac{1}{2}\alpha}(n-1, m-1)$ denote the critical values of F-distribution with $(n-1, m-1)$ degree of freedom at risk level $1 - \frac{1}{2}\alpha$ and $\frac{1}{2}\alpha$, respectively. Under H_0, the above equation becomes

$$P\left\{F_{1-\frac{1}{2}\alpha}(n-1, m-1) < \frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2} < F_{\frac{1}{2}\alpha}(n-1, m-1)\right\} = 1 - \alpha \tag{1.159}$$

One can always name the two samples x_i and y_i so that $\widehat{\sigma}_1^2$ is always larger than $\widehat{\sigma}_2^2$, i.e. $\frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2}$ is always larger than 1. Meanwhile, $F_{1-\frac{1}{2}\alpha}(n-1, m-1)$ is always smaller than 1 for small α. Therefore, the above equation reduces to:

$$P\left\{\frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2} < F_{\frac{1}{2}\alpha}(n-1, m-1)\right\} = 1 - \alpha \tag{1.160}$$

If the computed $\widehat{\sigma}_1^2/\widehat{\sigma}_2^2$ is smaller than the critical value $F_{\frac{1}{2}\alpha}(n-1, m-1)$, we accept H_0. Otherwise, we reject H_0, implying that σ_1^2 is significantly different from σ_2^2.

Example 1.10

A distance x has been measured by two EDM instruments with the number of measurements and sample standard errors as follows :

instrument 1:	$n = 41$,	$\widehat{\sigma}_1 = 4.2\ mm$
instrument 2:	$m = 61$,	$\widehat{\sigma}_2 = 3.0\ mm$

We want to test if these two instruments have the same accuracy at risk level $\alpha = 5\%$. Our two hypotheses to be tested are :

$$H_0 :\ \sigma_1^2 = \sigma_2^2\ ; \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

For $\alpha = 5\%$, we can find:

$$F_{\frac{1}{2}\alpha}(n-1, m-1) = 1.74\ , \quad \frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2} = (4.2/3.0)^2 = 1.96 > F_{\frac{1}{2}\alpha}(n-1, m-1)$$

We reject H_0 at 5% risk level. In other words, the two instruments statistically can not be regarded as having the same accuracy . ∎

1.4.4 Variance Analysis

A statistical experiment (e.g. geodetic measurements) may be influenced by one or several factors. For instance, the instrument used can be one such factor. Another factor can be surveyor who operates the instrument. According to the influencing factors, measurements can be divided into groups. Each group in turn may contain a series of measurements. Variance analysis tries to deduct whether the different groups of measurements follow the same distribution, by analyzing the internal variations within each group as well as the external variations between the groups. For simplicity, we present below only the one-dimensional variance analysis, i.e. with only one influencing factor involved.

Assume that for each outcome F_i $(1 \leq i \leq m)$ of the influencing factor, a group of n_i measurements ℓ_{ij} $(1 \leq j \leq n_i)$ have been made and all measurements within this group have normal distribution $N(\mu_i, \sigma^2)$.

Group	*Measurements*				*Distribution*
F_1	ℓ_{11}	ℓ_{12}	$\cdots$	ℓ_{1,n_1}	$N\left(\mu_1, \sigma^2\right)$
F_2	ℓ_{21}	ℓ_{22}	$\cdots$	ℓ_{2,n_2}	$N\left(\mu_2, \sigma^2\right)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots\cdots\cdots$
F_i	ℓ_{i1}	ℓ_{i2}	$\cdots$	ℓ_{i,n_i}	$N\left(\mu_i, \sigma^2\right)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots\cdots\cdots$
F_m	ℓ_{m1}	ℓ_{m2}	$\cdots$	ℓ_{m,n_m}	$N\left(\mu_m, \sigma^2\right)$

Our task is to test the following hypothesis:

$$H_0 \ : \quad \mu_1 = \mu_2 = \cdots = \mu_m \tag{1.161}$$

First let us look at the sum of squares of the internal variation of all measurements with respect to the corresponding individual expectation $\mu_1,\ \mu_2,\ \cdots,\ \mu_m$, respectively, :

$$S_I^2(\mu_1,\ \mu_2,\ \cdots,\ \mu_m) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\ell_{ij} - \mu_i)^2 \tag{1.162}$$

By minimizing S_I^2, one can derive an estimate of μ_i :

$$\widehat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\ell_{ij} \qquad (i = 1, 2, 3, \cdots, m) \tag{1.163}$$

from which an estimate of the variance for group i can be obtained :

$$\widehat{\sigma}_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(\ell_{ij} - \widehat{\mu}_i)^2 \qquad (i = 1, 2, 3, \cdots, m) \tag{1.164}$$

Using the estimated sample mean $\widehat{\mu}_1,\ \widehat{\mu}_2,\ \cdots,\ \widehat{\mu}_m$, the sum of squares of the internal variation, S_I, becomes now :

$$S_I^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij} - \widehat{\mu}_i\right)^2 = \sum_{i=1}^{m}\left\{(n_i - 1)\cdot\widehat{\sigma}_i^2\right\} \tag{1.165}$$

As the total number of measurements is equal to :

$$n = \sum_{i=1}^{m} n_i = n_1 + n_2 + \cdots + n_m \tag{1.166}$$

and we need to estimate m parameters $(\mu_1, \mu_2, \cdots, \mu_m)$, S_I^2/σ^2 thus has χ^2-distribution with $n - m$ degrees of freedom:

$$\frac{S_I^2}{\sigma^2} \ \sim\ \chi^2(n-m) \tag{1.167}$$

from which an unbiased estimate of the variance σ^2 is obtained :

$$\widehat{\sigma}_I^2 = \frac{S_I^2}{n-m} = \frac{1}{n-m}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij} - \widehat{\mu}_i\right)^2 = \frac{1}{n-m}\sum_{i=1}^{m}\left\{(n_i - 1)\cdot\widehat{\sigma}_i^2\right\} \tag{1.168}$$

$\widehat{\sigma}_I^2$ is called the *internal variance* in variance analysis.

Now let us assume that under hypothesis H_0 defined in Eq.(1.161), all expectations μ_i are equal to μ. With respect to this overall expectation, one can define a total sum of squares of variation:

$$S_T^2(\mu) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij} - \mu\right) \tag{1.169}$$

By minimizing S_T^2, one can obtain an overall estimate of μ :

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\ell_{ij} = \frac{1}{n}\sum_{i=1}^{m}\left(n_i\cdot\widehat{\mu}_i\right) \tag{1.170}$$

where μ_i is given by (1.163).

The variation of each sample mean $\widehat{\mu}_i$ with respect to the overall mean $\widehat{\mu}$ can be expressed as :

$$S_E^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\widehat{\mu}_i - \widehat{\mu}\right)^2 = \sum_{i=1}^{m} n_i\left(\widehat{\mu}_i - \widehat{\mu}\right)^2 \tag{1.171}$$

which leads to a χ^2-distribution with $m - 1$ degrees of freedom:

$$\frac{S_E^2}{\sigma^2} \ \sim\ \chi^2(m-1) \tag{1.172}$$

This gives us another unbiased variance estimate, called the *external variance* in variance analysis :

$$\widehat{\sigma}_E^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left\{n_i\left(\widehat{\mu}_i - \widehat{\mu}\right)^2\right\} = \frac{S_E^2}{m-1} \tag{1.173}$$

Inserting $\widehat{\mu}$ into Eq.(1.169), we have:

$$S_T^2 = S_T^2(\widehat{\mu}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij} - \widehat{\mu}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\left(\ell_{ij} - \widehat{\mu}_i\right) + \left(\widehat{\mu}_i - \widehat{\mu}\right)\right]^2$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij} - \widehat{\mu}_i\right)^2 + \sum_{i=1}^{m} n_i\left(\widehat{\mu}_i - \widehat{\mu}\right)^2 + 2\sum_{i=1}^{m}\left[\left(\widehat{\mu}_i - \widehat{\mu}\right)\sum_{j=1}^{n_i}\left(\ell_{ij} - \widehat{\mu}_i\right)\right]$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij}-\widehat{\mu}_i\right)^2 + \sum_{i=1}^{m} n_i\left(\widehat{\mu}_i-\widehat{\mu}\right)^2$$

Considering Eqs.(1.165) and (1.171), the last equation above lead to :

$$S_T^2 = S_I^2 + S_E^2 \tag{1.174}$$

and

$$\frac{S_T^2}{\sigma^2} = \frac{S_I^2}{\sigma^2} + \frac{S_E^2}{\sigma^2} \;\sim\; \chi^2(n-1) \tag{1.175}$$

From S_T^2 , an overall variance estimate can be obtained :

$$\widehat{\sigma}_T^2 = \frac{S_T^2}{n-1} = \frac{S_I^2+S_E^2}{n-1} = \frac{(n-m)\cdot\widehat{\sigma}_I^2 + (m-1)\cdot\widehat{\sigma}_E^2}{(n-m)+(m-1)} \tag{1.176}$$

where $n-m,\ m-1,\ n-1$ denote the degrees of freedom for $S_I^2,\ S_E^2$ and S_T^2, respectively. The three types of variances are summarized in Table 1.6 below:

Table 1.6: Variance Analysis Table

Type of variances	*Degrees of freedom*	*Sum of squares*	*Variances*
Internal	$n-m$	$S_I^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij}-\widehat{\mu}_i\right)^2$	$\widehat{\sigma}_I^2 = \frac{S_I^2}{n-m}$
External	$m-1$	$S_E^2 = \sum_{i=1}^{m} n_i\left(\widehat{\mu}_i-\widehat{\mu}\right)^2$	$\widehat{\sigma}_E^2 = \frac{S_E^2}{m-1}$
Total	$n-1$	$S_T^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij}-\widehat{\mu}\right)^2$	$\widehat{\sigma}_T^2 = \frac{S_T^2}{n-1}$

From the ratio of the external and internal variances, we obtain the following $F-$distribution :

$$F_0 = \frac{S_E^2/(m-1)}{S_I^2/(n-m)} = \frac{\widehat{\sigma}_E^2}{\widehat{\sigma}_I^2} \sim\; F(m-1, n-m) \tag{1.177}$$

Let $F_\alpha(m-1, n-m)$ denotes the critical value of F-distribution with $(m-1, n-m)$ degrees of freedom at risk level α. If

$$F_0 > F_\alpha(m-1, n-m)$$

we then reject the null hypothesis H_0, i.e. there are significant differences among various groups of measurements.

In numerical computations, it might be favourable to use the following formulas for evaluating $S_E^2,\ S_I^2$:

$$S_E^2 = \sum_{i=1}^{m} n_i\left(\widehat{\mu}_i-\widehat{\mu}\right)^2 = \sum_{i=1}^{m}\left(n_i\widehat{\mu}_i^2\right) - n\cdot\widehat{\mu}^2 \tag{1.178}$$

$$S_T^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij}-\widehat{\mu}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij}^2\right) - n\cdot\widehat{\mu}^2 \tag{1.179}$$

$$S_I^2 = S_T^2 - S_E^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij}^2\right) - \sum_{i=1}^{m}\left(n_i\widehat{\mu}_i^2\right) \tag{1.180}$$

Example 1.11

To investigate whether the time of observation during a day may influence the results of angle measurements, a geodesist used one theodolite to meansure an angle at four different time periods (i.e. $m=4$). The obtained results are listed below and they are assumed to have the same accuracy.

i	*Observation Time*	n_i	*Measurements* α_{ij} $(j = 1, 2, \cdots, n_i)$
1	T_1	3	$51^0\ 23'\ 46''$, $51^0\ 23'\ 48''$, $51^0\ 23'\ 49''$
2	T_2	2	$51^0\ 23'\ 45''$, $51^0\ 23'\ 44''$
3	T_3	3	$51^0\ 23'\ 46''$, $51^0\ 23'\ 47''$, $51^0\ 23'\ 48''$
4	T_4	4	$51^0\ 23'\ 43''$, $51^0\ 23'\ 45''$, $51^0\ 23'\ 44''$, $51^0\ 23'\ 46''$

In order to work with simpler numbers, we reduce the original angle measurements by subtracting a constant $51^0\ 23'\ 46''$:

$$\ell_{ij} = \alpha_{ij} - 51^0\ 23'\ 46''$$

The reduced observations ℓ_{ij} are listed in the following table:

i	*Observation Time*	n_i	ℓ_{ij} *in* $''$ $(j = 1, 2, \cdots, n_i)$	$\sum_{j=1}^{n_i} \ell_{ij}$	$\widehat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \ell_{ij}$	$n_i \cdot \widehat{\mu}_i^2$
1	T_1	3	0, 2, 3	+5	$+\frac{5}{3}$	$\frac{25}{3}$
2	T_2	2	-1, -2	-3	$-\frac{3}{2}$	$\frac{9}{2}$
3	T_3	3	0, 1, 2	+3	+1	3
4	T_4	4	-3, -1, -2, 0	-6	$-\frac{3}{2}$	9
	Sum $(\sum_{i=1}^{m})$	$12 = n$		-1		$\frac{149}{6}$

From the computed results in the last table above, we can calculated the overall mean $\widehat{\mu}$, the external variation S_E^2 and the external variance $\widehat{\sigma}_E^2$:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} \ell_{ij} = -\frac{1}{12}\ ('')$$

$$S_E^2 = \sum_{i=1}^{m}\left(n_i\widehat{\mu}_i^2\right) - n\cdot\widehat{\mu}^2 = \frac{149}{6} - 12\times\left(-\frac{1}{12}\right)^2 = \frac{99}{4} = \frac{297}{12}\ (''^2)$$

$$\widehat{\sigma}_E^2 = \frac{S_E^2}{m-1} = \frac{1}{4-1}\frac{99}{4} = \frac{33}{4}\ (''^2)$$

To calculate S_T^2 using Eq.(1.179), we calculate the squares of ℓ_{ij} in the following table:

i	*Observation Time*	n_i	ℓ_{ij}^2 *in* $''$ $(j = 1, 2, \cdots, n_i)$	$\sum_{j=1}^{n_i} \ell_{ij}^2$
1	T_1	3	0, 4, 9	+13
2	T_2	2	1, 4	+5
3	T_3	3	0, 1, 4	+5
4	T_4	4	9, 1, 4, 0	+14
	Sum $(\sum_{i=1}^{m})$	$12 = n$		+37

Then S_T^2 and S_I^2 are obtained from Eqs.(1.179) and (1.180), respectively :

$$S_I^2 = S_T^2 - S_E^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\ell_{ij}^2\right) - \sum_{i=1}^{m}\left(n_i\widehat{\mu}_i^2\right) = 37 - \frac{149}{6} = \frac{73}{6}\ (''^2)$$

The internal variance is given by Eq.(1.168) :

$$\widehat{\sigma}_I^2 = \frac{S_I^2}{n-m} = \frac{73/6}{12-4} = \frac{73}{48}\ (''^2)$$

The $F-$test statistics in Eq.(1.177) is :

$$F_0 = \frac{\widehat{\sigma}_E^2}{\widehat{\sigma}_I^2} = \frac{\frac{33}{4}}{\frac{73}{48}} = \frac{396}{73} \approx 5.42 > 4.07 = F_\alpha(m-1, n-m) = F_{5\%}(3, 8)$$

Based on the above $F-$test, we conclude that at risk level $\alpha = 5\%$, there are significant differences between angle measurements made at different times[1]. ∎

[1] This example is intended only to show how variance analysis can be used in geodesy and surveying. In practical applications, the size of the sample data, i.e. n and n_i, should be sufficiently large to guarantee reliable statistical inferences.

1.4.5 Regression Analysis

In science and engineering, we often meet the following problem: two or more quantities may be statistically correlated with each other. But there is no exact functional relation among these quantities, or we do not know any such relation. What we want is to estimate some approximate, quantitative functional relation among these quantities based on a series of measurements of these quantities. This kind of problems are called *regression analysis*. If the proposed quantitative relation is a linear function of several unknown parameters (also known as *regression coefficients*), the problem is called *linear regression*. *Non-linear regression* is more complicated and used less often in practice. If the number of quantities analysed is only two, we call the problem *single-variable regression analysis*. If there are three or more quantities involved, we then deal with the *multi-variable regression analysis*.

Linear Regression - Line-Fitting

Assume that two quantities, x and y, have been measured n times with observation data: $x_1,\, x_2, \cdots,\, x_n$ and $y_1, y_2, \cdots, y_n$. We now want to find out whether the following relation holds:

$$y = \alpha + \beta \cdot x \tag{1.181}$$

where $\alpha,\ \beta$ are two unknown regression coefficients to be determined. As Eq.(1.181) represents a line geometrically, the above equation is called the *theoretical regression line*. Since the correlation between x and y can be caused by many complicated factors and meanwhile the measurements $(x_i,\ y_i,\ i = 1, 2, \cdots, n)$ contain errors, Eq.(1.181) need to be modified for the measured data :

$$y_i = \alpha + \beta \cdot x_i + \varepsilon_i \tag{1.182}$$

where ε_i denotes the residual or regression error and all ε_i $(1 \le i \le n)$ are assumed to be independent of each other. To find the optimal estimate of α, β that best fit the measurement data (x_i, y_i), we impose the following condition:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta \cdot x_i)^2 = \text{minimum} \tag{1.183}$$

Regression analysis under condition (1.183) is called a least squares line-fitting. The optimal estimates $\widehat{\alpha}$ and $\widehat{\beta}$ of α and β can be derived by letting the derivatives of the sum of residuals given in Eq.(1.183) be equal to zero:

$$\frac{\partial}{\partial \alpha} \left\{ \sum_{i=1}^{n} (y_i - \alpha - \beta \cdot x_i)^2 \right\}_{\alpha=\widehat{\alpha},\ \beta=\widehat{\beta}} = \sum_{i=1}^{n} 2(y_i - \widehat{\alpha} - \widehat{\beta} \cdot x_i)\,(-1) = 0$$

$$\frac{\partial}{\partial \beta} \left\{ \sum_{i=1}^{n} (y_i - \alpha - \beta \cdot x_i)^2 \right\}_{\alpha=\widehat{\alpha},\ \beta=\widehat{\beta}} = \sum_{i=1}^{n} 2(y_i - \widehat{\alpha} - \widehat{\beta} \cdot x_i)\,(-x_i) = 0$$

which leads to the linear equation system for $\widehat{\alpha}$ and $\widehat{\beta}$:

$$\begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{1.184}$$

where :

$$\left.\begin{aligned} q_{11} &= n \\ q_{12} &= \textstyle\sum_{i=1}^{n} x_i \\ q_{22} &= \textstyle\sum_{i=1}^{n} x_i^2 \\ w_1 &= \textstyle\sum_{i=1}^{n} y_i \\ w_2 &= \textstyle\sum_{i=1}^{n} x_i y_i \end{aligned}\right\} \tag{1.185}$$

From Eq.(1.184), $\widehat{\alpha}$ and $\widehat{\beta}$ can be solved :

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = \frac{1}{q_{11}q_{22} - q_{12}^2} \begin{bmatrix} q_{22} & -q_{12} \\ -q_{12} & q_{11} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{q_{11}q_{22} - q_{12}^2} \begin{bmatrix} q_{22}w_1 - q_{12}w_2 \\ -q_{12}w_1 + q_{11}w_2 \end{bmatrix} \tag{1.186}$$

The least squares residuals $\widehat{\varepsilon}_i$, which is a measure of how good the data point (x_i, y_i) fits the computed regression line $y = \widehat{\alpha} + \widehat{\beta}x$, can be calculated by :

$$\widehat{\varepsilon}_i = y_i - \widehat{\alpha} - \widehat{\beta} \cdot x_i \tag{1.187}$$

If the theoretical residual ε_i has a priori distribution: $\varepsilon_i \sim N(0, \sigma^2)$, then an unbiased estimate of σ^2 can be obtained from the least squares residuals $\widehat{\varepsilon}_i$ given in Eq.(1.188):

$$\widehat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left(y_i - \widehat{\alpha} - \widehat{\beta} \cdot x_i \right)^2 \tag{1.188}$$

$\widehat{\sigma}^2$ is also a measure of how good the overall fitting of all data points to the computed regression line $y = \widehat{\alpha} + \widehat{\beta}x$. The standard errors of the estimated $\widehat{\alpha}$ and $\widehat{\beta}$ can then be calculated as [2]:

$$\sigma_{\widehat{\alpha}} = \widehat{\sigma} \cdot \sqrt{\frac{q_{22}}{q_{11}q_{22} - q_{12}^2}}, \quad \sigma_{\widehat{\beta}} = \widehat{\sigma} \cdot \sqrt{\frac{q_{11}}{q_{11}q_{22} - q_{12}^2}} \tag{1.189}$$

For $\varepsilon_i \sim N(0, \sigma^2)$, the following three statistics have $t-$ and χ^2 distribution, respectively :

$$\frac{\widehat{\alpha}}{\sigma_{\widehat{\alpha}}} \sim t(n-2), \quad \frac{\widehat{\beta}}{\sigma_{\widehat{\beta}}} \sim t(n-2) \tag{1.190}$$

$$(n-2)\frac{\widehat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{y_i - \widehat{\alpha} - \widehat{\beta} \cdot x_i}{\sigma} \right)^2 \sim \chi^2(n-2) \tag{1.191}$$

If σ^2 is known, the overall significance of the linear regression can be tested using the statistics given in Eq.(1.191).

When $\widehat{\alpha}$ and $\widehat{\beta}$ have been estimated in the above linear regression and found to be significant statistically, they may be used to predict quantity y_p for a given value x_p of quantity x :

$$\widehat{y}_p = \widehat{\alpha} + \widehat{\beta} \cdot x_p \tag{1.192}$$

Non-Linear Regression

Generally non-linear regression is not as easy as the linear regression. Non-linear fitting under condition (1.183) often leads to non-linear equations for the unknown regression coefficients which need to be solved. However, sometimes the non-linear regression models can be reduced to linear ones, as the following two examples.

Assume that for two measurement series $x_i,\ y_i$ $(i = 1, 2, \cdots n)$, we have the following non-linear regression model :

$$y_i = \frac{x_i}{\alpha \cdot x_i + \beta} \tag{1.193}$$

where $\alpha,\ \beta$ are the two regression coefficients to be determined. Now we introduce two new quantities $X_i,\ Y_i$ such that :

$$Y_i = \frac{1}{y_i}, \quad X_i = \frac{1}{x_i} \tag{1.194}$$

Eq.(1.193) then reduces to :

$$Y_i = \alpha + \beta \cdot X_i \tag{1.195}$$

[2] The proof of Eqs.(1.188) and (1.189) is given in Subsection 3.1.1 of Chapter 3.

which is a linear regression function of the regression coefficients.

Let us look at another non-linear regression model for two measurement series $x_i,\ y_i\ (i = 1, 2, \cdots n)$:

$$y_i = \alpha \cdot e^{-\frac{x_i^2}{\beta^2}} \tag{1.196}$$

where $\alpha,\ \beta$ are two regression coefficients to be determined. Taking the natural logarithm on both sides of Eq.(1.196), we obtain :

$$\ln y_i = \ln \alpha - \frac{1}{\beta^2} \cdot x_i^2 \tag{1.197}$$

After introducing:

$$X_i = x_i^2, \quad Y_i = \ln y_i, \quad a = \ln \alpha, \quad b = -\frac{1}{\beta^2} \tag{1.198}$$

we obtain a linear regression model with two new regression coefficients a and b :

$$Y_i = a + b \cdot X_i \tag{1.199}$$

When the least squares estimates $\widehat{a}, \widehat{b}$ of the new regression coefficients a, b have been computed using Eq.(1.186), the original regression coefficients can be obtained from Eq.(1.198) :

$$\widehat{\alpha} = e^{\widehat{a}}, \quad \widehat{\beta} = \sqrt{-\frac{1}{\widehat{b}}} \tag{1.200}$$

Whether one should choose a linear or non-linear regression model, depends mainly on the correlation characteristics between the quantities concerned. One may plot the discrete points with planar Cartesian coordinates (x_i, y_i) for $i = 1, 2, \cdots, n$ and then compare with the curves of simple, known functions (lines, polynomials, trigonometric functions, etc.). Based on the comparison, a realistic regression function may be chosen.

Example 1.12

To evaluate the accuracy of an EDM instrument, $n = 10$ baselines are measured with this instrument and compared with the correct lengths. Based on the comparisons, a standard error σ_i $(1 \le i \le 10)$ has been computed for each baseline. The approximate length s_i in km and the corresponding standard error σ_i in cm are listed below:

i	s_i (km)	σ_i (cm)
1	0.5	2.9
2	1.2	3.1
3	1.9	3.2
4	3.0	3.4
5	3.7	3.5
6	4.4	3.7
7	4.9	3.8
8	5.1	4.1
9	5.7	4.2
10	6.0	4.4

We now want to fit the above set of measurement data to the following regression formula:

$$\sigma_i^2 = a^2 + b^2 \cdot s_i^2$$

where a and b are two regression coefficients to be determined. a represents a constant error in the EDM instrument tested, while b represents the scale error. Using the following replacements:

$$y_i = \sigma_i^2, \quad x_i = s_i^2, \quad \alpha = a^2, \quad \beta = b^2$$

the proposed regression formula above reduces to a simple linear line fitting :

$$y_i = \alpha + \beta \cdot x_i + \varepsilon_i$$

The reduced data values (x_i, y_i) are listed in the table below and also plotted in Figure 1.6.

i	$x_i = s_i^2 \ \ (km^2)$	$y_i = \sigma_i^2 \ (cm^2)$	x_i^2	$x_i \cdot y_i$
1	0.25	8.41	0.0625	2.1025
2	1.44	9.61	2.0736	13.8384
3	3.61	10.24	13.0321	36.9664
4	9.00	11.56	81.0000	104.0400
5	13.69	12.25	187.4161	167.7025
6	19.36	13.69	374.8096	265.0384
7	24.01	14.44	576.4801	346.7044
8	26.01	16.81	676.5201	437.2281
9	32.49	17.64	1055.6001	573.1236
10	36.00	19.36	1296.000	696.9600
$\sum$	165.86	134.01	4262.9942	2643.7043

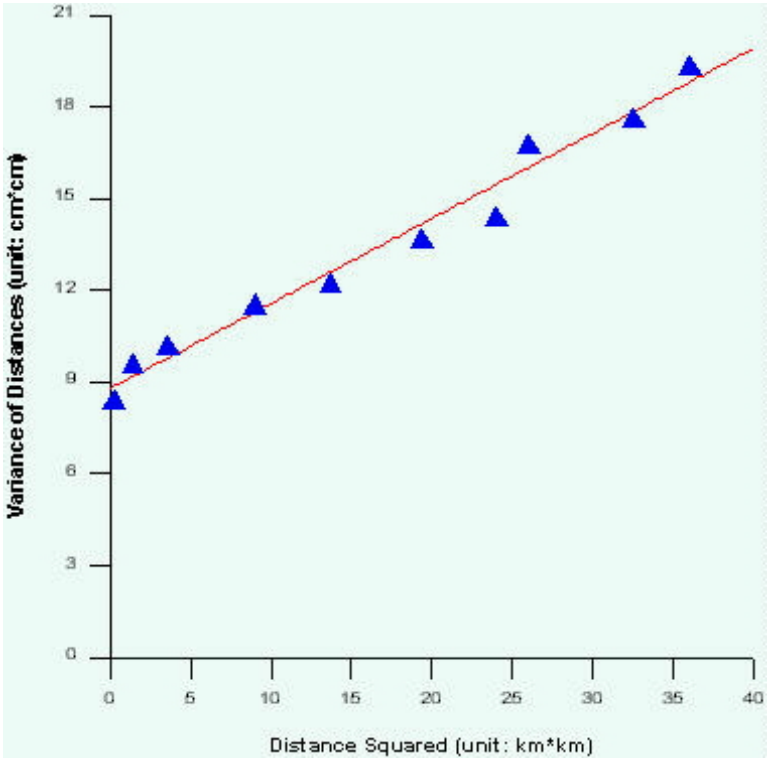

Figure 1.6: Least Squares Line Fitting

From Eqs.(1.184) and (1.185) as well as the computed results in the last table above, we obtain the linear equation for the optimal regression coefficients:

$$\begin{bmatrix} 10 & 165.86 \\ 165.86 & 4262.9942 \end{bmatrix} \begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = \begin{bmatrix} 134.01 \\ 2643.7043 \end{bmatrix}$$

which gives:

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = \frac{1}{15120.403} \begin{bmatrix} 4262.9942 & -165.86 \\ -165.86 & 10 \end{bmatrix} \begin{bmatrix} 134.01 \\ 2643.7043 \end{bmatrix} = \begin{bmatrix} 8.783 \\ 0.278 \end{bmatrix}$$

The a posteriori estimate of the variance factor is :

$$\widehat{\sigma}^2 = \frac{1}{10-2}\sum_{i=1}^{n}\left(y_i - \widehat{\alpha} - \widehat{\beta}\cdot x_i\right)^2 = 0.3782 \quad \rightarrow \quad \widehat{\sigma} = \pm 0.615\ (cm^2)$$

The standard errors of $\widehat{\alpha}$ and $\widehat{\beta}$ are :

$$\sigma_{\widehat{\alpha}} = \widehat{\sigma}\cdot\sqrt{\frac{4262.9942}{15120.403}} = 0.326\ (cm^2), \quad \sigma_{\widehat{\beta}} = \widehat{\sigma}\cdot\sqrt{\frac{10}{15120.403}} = 0.015\ \ (cm/km)^2$$

The original regression coefficients a, b are then obtained:

$$\widehat{a} = \sqrt{\widehat{\alpha}} = \sqrt{8.783} = 2.96\ cm, \quad \widehat{b} = \sqrt{\widehat{\beta}} = \sqrt{0.278} = 0.53\ (cm/km) = 5.3\ ppm$$

where *ppm* stands for "part per million". The standard errors of $\widehat{a}$ and $\widehat{b}$ can be calculated from those of $\widehat{\alpha}$ and $\widehat{\beta}$ with the help of error propagation law:

$$\sigma_{\widehat{a}} = \frac{1}{2\widehat{\alpha}}\sigma_{\widehat{\alpha}} = 0.054\ cm, \quad \sigma_{\widehat{b}} = \frac{1}{2\widehat{\beta}}\sigma_{\widehat{\beta}} = 0.14\ ppm$$

A comparison of the magnitude of the estimated regression coefficients $\widehat{a}$, $\widehat{b}$ with their respective standard errors shows that a quite good regreesion result has been acieved: the EDM instrument has a constant error at about 3 *cm* and a scale error at about 5 *ppm*. ∎

1.5 GUM-Terminology: Uncertainty in Measurement

Traditionally, accuracy and (standard) errors are concepts used to describe the quality of measurements or quality of quanlities derived from the measurements. This is the normal practice in the field of geodesy, surveying, photogrammetry, electrical engineering, as well as many other science and engineering fields.

In earlier 1980's, a *Joint Committee for Guides in Metrology* (JCGM) was formed by seven international organizations:

- Bureau International des Poids et Measures (BIPM)
- International Electrotechnical Commission (IEC)
- International Federation of Clinical Chemistry (IFCC)
- International Organization of Standardization (ISO)
- International Union of Pure and Applied Chemistry (IUPAC)
- International Union of Pure and Applied Physics (IUPAP)
- International Organization of Legal Metrology (OIML)

The task of JCGM is to develop recommendations on how to evaluate and express the uncertainty of measurement in a way which is: *universal* (applicable to all kinds of measurements in different fields), *internally consistent* and *transferable* (possible to propagate uncertainty of one set of measurements to another related set). After several years of work, JCGM published in 1995 its recommendation "*Guide to the expression of uncertainty in measurement* (GUM)", with a revised version in 2008.

However, it should be pointed out that the terminology of GUM is neither completely new nor completely different from traditional geodetic terminonogy. In Table 1.7 below, we can compare the corresponding terminologies:

Table 1.7: GUM terminology versus traditional terminology

Traditional Terminology	*GUM Terminology*
error	uncertainty
standard error (σ)	standard uncertainty
$k\ \sigma$	expanded standard uncertainty ($k\sigma$)
k : factor of the confidence interval	k : coverage factor
the law of propagation of errors	the law of propagation of uncertainty
linearization coefficient	sensitivity coefficient

According to GUM, the uncertainty (error in traditional terminology) in the input measurements can be evaluated either by statistical method based on repeated measurements (the so called Type A) or by other means (Type B). The covergage factor (k) is often taken as 2 (or 1.96) corresponding to a confidence level of about 95%. Moreever, the expanded standard uncertainty ($k\sigma$) is more suitable for situations where high degree safety is desired (such as navigation by GPS or clinical tests involving drugs and patients).

In Sweden, the National Land Survey (Lantmäteriverket) and SP Technical Reserach Institute (Sveriges Provnings- och Forskningsinstitut) have started to use GUM terminology in practical work. Students are advised to get familiar with GUM terminology by consulting the original GUM document which can be downloaded at *http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf* .

1.6 Optimal Estimation

In many scientific and engineering problems, we need to estimate a vector of m unknown parameters:

$$\underset{m\cdot 1}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} \tag{1.201}$$

from a vector of n observations:

$$\underset{n\cdot 1}{\ell} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \cdots \\ \ell_n \end{bmatrix} \tag{1.202}$$

One can construct many different estimates to x from ℓ, depending on the used estimation criteria. Normally, good estimates should have the following properties:

- *Unbiasedness:* Let $\widehat{x}$ be an estimate of x. $\widehat{x}$ is called an *unbiased estimate* of x iff [3]:

$$E(\widehat{x}) = E(x) \tag{1.203}$$

- *Uniformness:* An estimate $\widehat{x}$ is said to be an *uniform estimate* of x if $\widehat{x}$ approaches the true value x when the number of measurements goes to infinity:

$$\lim_{n\to\infty} P\{|\widehat{x} - x| < \delta\} = 1 \tag{1.204}$$

  where δ is an arbitrarily small positive number.

- *Effectiveness:* A good estimate should not only be unbiased, *i.e.* varying around the true value without systematic deviation, but also have as small as possible variations around the true value. An unbiased estimate $\hat{x}$ of x is said to be an *effective estimate* if it has the smallest error variance :

$$E(\widehat{x}) = E(x)\ , \quad Var(\widehat{x} - x) = E\left[(\widehat{x} - x)(\widehat{x} - x)^\top\right] = minimum \tag{1.205}$$

An unbiased estimate with smallest error variance is also called a *Best Unbiased Estimate* (BUE). A subgroup of all BUEs is the group of estimates based on linear combinations of ℓ_i :

$$\underset{m\cdot 1}{\widehat{x}} = \underset{m\cdot n}{A}\ \underset{n\cdot 1}{\ell} \tag{1.206}$$

where A is a matrix with constant elements to be specified. If the above estimate satisfies also Eq. (1.205), then $\widehat{x}$ is called a *Best Linear Unbiased Estimate* (BLUE).

In the following subsections, we are going to introduce three commonly used estimation methods for constructing optimal estimates from given observations.

1.6.1 Least Squares Method

Least squares method is a classical and intuitive method which defines the optimal estimate $\widehat{x}$ by minimizing the sum of the weighted observation residuals squared:

$$\sum_{i=1}^{n} p_i\, \widehat{\varepsilon}_i^2 = \text{minimum} \tag{1.207}$$

[3] *iff* means: if and only if.

where p_i denotes the weight of ℓ_i , $\widehat{\varepsilon}_i$ denotes the estimated error of ℓ_i (also called the *residual* of ℓ_i) for $i = 1, 2, \cdots, n$. The above condition applies when the observations ℓ_i are uncorrelated and the weight matrix P is thus a diagonal matrix. When ℓ_i are correlated, the corresponding condition of the least squares method becomes:

$$\widehat{\varepsilon}^{\top} P \widehat{\varepsilon} = \text{ minimum}, \quad \text{or} \quad \widehat{\varepsilon}^{\top} C^{-1} \widehat{\varepsilon} = \text{minimum} \tag{1.208}$$

where $\widehat{\varepsilon}$ is the residual vector containing all residuals $\widehat{\varepsilon}_i$ which are derived together with the optimal estimate $\widehat{x}$, and C is the variance-covariance matrix of ℓ .

In Chapter 2 and Chapter 3, the least squares method will be used to derive two main adjustment methods used in geodesy and surveying.

Example 1.13 [4]

Assume that a quantity x has been directly measured n times with results: $\ell_1, \ell_2, \cdots, \ell_n$. Let p_i denote the weight of ℓ_i. The *weighted mean* of the n direct measurements is obtained as:

$$\widehat{x} = \frac{p_1\ell_1 + p_2\ell_2 + \cdots + p_n\ell_n}{p_1 + p_2 + \cdots + p_n} = \frac{\sum_{i=1}^{n}(p_i\ell_i)}{\sum_{i=1}^{n}(p_i)} \tag{1.209}$$

The residual of ℓ_i with respect to $\widehat{x}$ is :

$$\widehat{\varepsilon}_i = \ell_i - \widehat{x} \tag{1.210}$$

Let us now assume that x' is another estimate of x and that $\varepsilon'_i = \ell_i - x'$ is the corresponding residual of ℓ_i. We have then :

$$\sum_{i=1}^{n} p_i \varepsilon_i'^2 = \sum_{i=1}^{n} \left\{ p_i \cdot [\ell_i - x']^2 \right\} = \sum_{i=1}^{n} \left\{ p_i \cdot [(\ell_i - \widehat{x}) + (\widehat{x} - x')]^2 \right\}$$

$$= \sum_{i=1}^{n} \left\{ p_i \cdot \left[(\ell_i - \widehat{x})^2 + 2\,(\ell_i - \widehat{x})\,(\widehat{x} - x') + (\widehat{x} - x')^2 \right] \right\}$$

$$= \sum_{i=1}^{n} p_i\, \widehat{\varepsilon}_i^2 + 2\,(\widehat{x} - x') \sum_{i=1}^{n} \{p_i \cdot (\ell_i - \widehat{x})\} + (\widehat{x} - x')^2 \sum_{i=1}^{n} p_i$$

Considering Eq. (1.209), one can show that the second term on the right side of the above equation vanishes:

$$\sum_{i=1}^{n} \{p_i \cdot (\ell_i - \widehat{x})\} = \sum_{i=1}^{n} p_i\ell_i - \widehat{x} \cdot \sum_{i=1}^{n} p_i = 0$$

As $p_i > 0$ and $(\widehat{x} - x')^2 \geq 0$, we obtain finally :

$$\sum_{i=1}^{n} p_i \varepsilon_i'^2 = \sum_{i=1}^{n} p_i\, \widehat{\varepsilon}_i^2 + (\widehat{x} - x')^2 \sum_{i=1}^{n} p_i \;\geq \sum_{i=1}^{n} p_i\, \widehat{\varepsilon}_i^2 \tag{1.211}$$

where the identity holds iff $x' = \widehat{x}$. This shows that the weighted mean $\widehat{x}$ leads to smallest sum of weighted residuals squared. In other words, *the weighted mean $\widehat{x}$ is a least squares estimate* .

For equal weights $p_i = p$ (a constant), we then have the simple mean or direct mean :

$$\widehat{x} = \frac{\ell_1 + \ell_2 + \cdots + \ell_n}{n} = \frac{\sum_{i=1}^{n}\ell_i}{n}. \quad \blacksquare \tag{1.212}$$

[4] The weighted mean of n direct measurements is also treated in **Subsection 3.1.2** using the method of adjustment by elements.

1.6.2 Maximum Likelihood Method

Let $\underset{n\cdot 1}{\ell}$ be an observation vector and $\underset{m\cdot 1}{x}$ the parameter vector to be estimated. Let $f(\ell/x)$ denote the conditional frequency function of the observations ℓ under x. Generally $f(\ell/x)$ is a function of both ℓ and x. But for a specific realization of the observations ℓ, $f(\ell/x)$ is a function of x. If there is an estimate $\widehat{x}$ of x such that $f(\ell/\widehat{x})$ attains maximum, then random vector ℓ will have the maximum probability to occur. Now that ℓ has already occurred (i.e. measurements have been done), the above estimate $\widehat{x}$ should have maximum probability to occur. Such an estimate $\widehat{x}$ of the unknown parameter x based on observations ℓ is called a *maximum likelihood estimate*. The method to obtain an optimal estimate $\widehat{x}$ under condition:

$$f(\ell/x)\mid_{x=\widehat{x}}= \text{ maximum} \tag{1.213}$$

is called the *maximum likelihood method*.

When parameters in x are non-random, the conditional frequency function $f(\ell/x)$ can be simply written as $f(\ell, x)$, still a function of x.

To find the maximum likelihood estimate $\widehat{x}$, one can take the derivative of the conditional frequency function with respect to each parameter in x and let the derivatives vanish:

$$\frac{\partial f(\ell/x)}{\partial x_i}=0, \quad i=1,2,3,\cdots,m \tag{1.214}$$

Eq. (1.214) are known as the likelihood equations. As function $\ln\left(f(\ell/x)\right)$ tends to increase or decrease as $f(\ell/x)$ increase or decrease, the likelihood equations (1.214) can be rewritten as:

$$\frac{\partial \ln\left(f(\ell/x)\right)}{\partial x_i}=0, \quad i=1,2,3,\cdots,m \tag{1.215}$$

Example 1.14

A quantity has been independently observed n times with observations ℓ_i $(i=1,2,\cdots,n)$ and all observations ℓ_i have the same normal distribution, $\ell_i \sim N(\mu,\ \sigma^2)$, with frequency function:

$$f(\ell_i)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\ell_i-\mu)^2}{2\sigma^2}\right\}$$

We want to use the maximum likelihood method to estimate the expectation parameter μ and the variance σ^2 based on the n observations ℓ_i. Our unknown parameter vector is:

$$\underset{2\cdot 1}{x}=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$$

As ℓ_i are independent of each other, the joint frequency function of $\ell_1,\ell_2,\cdots,\ell_n$ is:

$$f(\ell_1,\ell_2,\cdots,\ell_n)=f(\ell_1)\cdot f(\ell_2)\cdots f(\ell_n)=\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ell_i-\mu)^2\right\}$$

Now that x is non-random, the conditional frequency function $f(\ell/x)$ is identical to the above joint frequency function:

$$f(\ell/x)=\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ell_i-\mu)^2\right\} \tag{1.216}$$

The derivatives in Eq. (1.215) can be easily obtained:

$$\frac{\partial \ln\left(f(\ell/x)\right)}{\partial x_1}=\frac{\partial \ln\left(f(\ell/x)\right)}{\partial \mu}=-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[2(\ell_i-\mu)\cdot(-1)\right]=\frac{1}{\sigma^2}\left(\sum_{i=1}^{n}\ell_i-n\mu\right)$$

$$\frac{\partial \ln\left(f(\ell/x)\right)}{\partial x_2} = \frac{\partial \ln\left(f(\ell/x)\right)}{\partial \sigma^2} = -\frac{1}{2\sigma^4}\left(n\sigma^2 - \sum_{i=1}^{n}\left[\ell_i - \mu\right]^2\right)$$

To find the maximum likelihood estimates $\widehat{\mu},\ \widehat{\sigma}^2$, we let the above two derivatives vanish:

$$\sum_{i=1}^{n}\ell_i - n\cdot\widehat{\mu} = 0 \quad \text{and} \quad n\,\widehat{\sigma}^2 - \sum_{i=1}^{n}\left(\ell_i - \widehat{\mu}\right)^2 = 0$$

which gives:

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ell_i \quad \text{and} \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\ell_i - \widehat{\mu}\right)^2 \tag{1.217}$$

The result for $\widehat{\mu}$ shows that the simple mean of n direct measurements is actually a maximum likelihood estimate. Considering Eq. (1.131), we can also obtain:

$$E(\widehat{\mu}) = \frac{1}{n}\sum_{i=1}^{n}E(\ell_i) = \mu\ ,\ E(\widehat{\sigma}^2) = E\left\{\frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left[\ell_i - \widehat{\mu}\right]^2\right\} = \frac{n-1}{n}\ \sigma^2 \tag{1.218}$$

The above equations indicate that $\widehat{\mu}$ is unbiased while $\widehat{\sigma}^2$ is biased. The unbiased estimate $\widehat{\sigma}^2$ given by Eq.(1.111) should be preferred. ∎

1.6.3 Minimum Error Variance Estimation

Assume that both the observations ℓ and the parameters x are random variables. The minimum error variance method says that the optimal estimate $\widehat{x}$ has the smallest error variance:

$$E[(\widehat{x} - x)(\widehat{x} - x)^\top] = \text{minimum} \tag{1.219}$$

It can be proved that the optimal estimate $\widehat{x}$ is equal to the conditional expectation of x under ℓ :

$$\widehat{x} = E(x/\ell) \tag{1.220}$$

If we restrict the optimal estimate $\widehat{x}$ to the set of linear functions of ℓ, the above optimal estimate becomes:

$$\widehat{x} = E(x) + C_{x\ell}C_{\ell\ell}^{-1}\cdot\left[\ell - E(\ell)\right] \tag{1.221}$$

where $C_{x\ell}$ denotes the covariance matrix between x and ℓ, while $C_{\ell\ell}$ denotes the variance-covariance matrix of ℓ. We can easily prove Eq. (1.221) as follows.

Let x' denote a linear estimate of x constructed from observations ℓ :

$$x' = \alpha + \beta\ell \tag{1.222}$$

where $\alpha,\ \beta$ are non-random matrices to be determined. Applying the unbiasedness condition Eq. (1.203) on x', we have:

$$E(x') = E(x) = \alpha + \beta E(\ell)\ , \quad \text{or} \quad \alpha = E(x) - \beta E(\ell)$$

x' can then be rewritten as:

$$x' = E(x) + \beta\left[\ell - E(\ell)\right] \tag{1.223}$$

and the estimation error of x' is:

$$\varepsilon' = x' - x = E(x) - \beta E(\ell) + \left[-1,\ \beta\right]\begin{bmatrix} x \\ \ell \end{bmatrix}$$

Applying error propagation law Eq. (1.39) on ε', we obtain the variance-covariance matrix of ε' :

$$C_{\varepsilon'\varepsilon'} = \left[-1,\ \beta\right]\begin{bmatrix} C_{xx} & C_{x\ell} \\ C_{\ell x} & C_{\ell\ell} \end{bmatrix}\begin{bmatrix} -1 \\ \beta^\top \end{bmatrix} = \beta C_{\ell\ell}\beta^\top - \beta C_{\ell x} - C_{x\ell}\beta^\top + C_{xx} \tag{1.224}$$

By letting the derivative $\partial C_{\varepsilon'\varepsilon'}/\partial\beta$ be zero,

$$\frac{\partial C_{\varepsilon'\varepsilon'}}{\partial \beta^{\top}} = 2\beta C_{\ell\ell} - 2C_{x\ell} = 0$$

we obtain the optimal coefficient matrix β :

$$\widehat{\beta} = C_{x\ell} \cdot C_{\ell\ell}^{-1} \tag{1.225}$$

Substitution of $\widehat{\beta}$ into Eq. (1.223) gives the optimal estimate $\widehat{x}$ as given by Eq. (1.221). The variance-covariance matrix of the error in the optimal estimate $\widehat{x}$:

$$\widehat{\varepsilon} = \widehat{x} - x$$

can be derived by inserting $\widehat{\beta}$ into Eq. (1.224):

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = C_{xx} - C_{x\ell} \cdot C_{\ell\ell}^{-1} \cdot C_{\ell x} \tag{1.226}$$

1.7.2 Special Identities in Matrix Algebra

Let $\underset{r\cdot r}{A}$, $\underset{s\cdot s}{D}$ be two non-singular matrices and $\underset{r\cdot s}{B}$, $\underset{s\cdot r}{C}$ be two arbitrary matrices with dimensions indicated by their subscripts. Then the following two inverse formulas hold:

$$AB(D+CAB)^{-1} = (A^{-1}+BD^{-1}C)^{-1}BD^{-1} \tag{1.267}$$

$$(D+CAB)^{-1} = D^{-1} - D^{-1}C(A^{-1}+BD^{-1}C)^{-1}BD^{-1} \tag{1.268}$$

Proof. Let $G = D + CAB$ and we have:

$$(D+CAB)G^{-1} = DG^{-1} + CABG^{-1} = I$$

From the last equation above, one can solve for G^{-1}:

$$G^{-1} = D^{-1} - D^{-1}CABG^{-1} \tag{1.269}$$

which leads to:

$$BG^{-1} = BD^{-1} - BD^{-1}CABG^{-1}$$

$$BD^{-1} = BG^{-1} + BD^{-1}CABG^{-1} = (A^{-1}+BD^{-1}C)ABG^{-1}$$

$$ABG^{-1} = (A^{-1}+BD^{-1}C)^{-1}BD^{-1} \tag{1.270}$$

Inserting $G = D + CAB$ back to (1.270) gives (1.267). (1.268) follows after substitution of (1.270) into (1.269).

Now let us look at a non-singular square matrix $\underset{n\cdot n}{N}$ which can be partitioned into four submatrices:

$$\underset{n\cdot n}{N} = \begin{bmatrix} \underset{r\cdot r}{N_{11}} & \underset{r\cdot s}{N_{12}} \\ \underset{s\cdot r}{N_{21}} & \underset{s\cdot s}{N_{22}} \end{bmatrix} \qquad (n = r+s) \tag{1.271}$$

where both N_{11} and N_{22} are non-singular. Then the inverse matrix of N can be written as :

$$N^{-1} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \begin{bmatrix} N_{11}^{-1} + N_{11}^{-1}N_{12}\overline{N}_{22}^{-1}N_{21}N_{11}^{-1}\,, & -N_{11}^{-1}N_{12}\overline{N}_{22}^{-1} \\ -\overline{N}_{22}^{-1}N_{21}N_{11}^{-1}\,, & \overline{N}_{22}^{-1} \end{bmatrix} \tag{1.272}$$

or equivalently:

$$N^{-1} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \overline{N}_{11}^{-1}\,, & -\overline{N}_{11}^{-1}N_{12}N_{22}^{-1} \\ -N_{22}^{-1}N_{21}\overline{N}_{11}^{-1}\,, & N_{22}^{-1} + N_{22}^{-1}N_{21}\overline{N}_{11}^{-1}N_{12}N_{22}^{-1} \end{bmatrix} \tag{1.273}$$

where:

$$\left.\begin{aligned} \overline{N}_{11} &= N_{11} - N_{12}N_{22}^{-1}N_{21} \\ \overline{N}_{22} &= N_{22} - N_{21}N_{11}^{-1}N_{12} \end{aligned}\right\} \tag{1.274}$$

Proof The inverse matrix of N can also be partitioned :

$$\underset{n\cdot n}{Q} = \begin{bmatrix} \underset{r\cdot r}{Q_{11}} & \underset{r\cdot s}{Q_{12}} \\ \underset{s\cdot r}{Q_{21}} & \underset{s\cdot s}{Q_{22}} \end{bmatrix} \qquad (n = r+s)$$

Since $N\,Q = I$,

$$\begin{bmatrix} \underset{r\cdot r}{N_{11}} & \underset{r\cdot s}{N_{12}} \\ \underset{s\cdot r}{N_{21}} & \underset{s\cdot s}{N_{22}} \end{bmatrix} \cdot \begin{bmatrix} \underset{r\cdot r}{Q_{11}} & \underset{r\cdot s}{Q_{12}} \\ \underset{s\cdot r}{Q_{21}} & \underset{s\cdot s}{Q_{22}} \end{bmatrix} = \begin{bmatrix} \underset{r\cdot r}{I} & \underset{r\cdot s}{0} \\ \underset{s\cdot r}{0} & \underset{s\cdot s}{I} \end{bmatrix}$$

we then have :

$$\left\{\begin{array}{ll} N_{11}Q_{11} + N_{12}Q_{21} = I & (a) \\ N_{11}Q_{12} + N_{12}Q_{22} = 0 & (b) \\ N_{21}Q_{11} + N_{22}Q_{21} = 0 & (c) \\ N_{21}Q_{12} + N_{22}Q_{22} = I & (d) \end{array}\right.$$

From Eqs. (a) and (c), one can solve for Q_{11} and Q_{21} :

$$\left.\begin{array}{l} Q_{11} = N_{11}^{-1} + N_{11}^{-1}N_{12}\bar{N}_{22}^{-1}N_{21}N_{11}^{-1} \\ Q_{21} = -\bar{N}_{22}^{-1}N_{21}N_{11}^{-1} \end{array}\right\}$$

From Eqs. (b) and (d), one can solve for Q_{12} and Q_{22} :

$$\left.\begin{array}{l} Q_{12} = -N_{11}^{-1}N_{12}\bar{N}_{22}^{-1} \\ Q_{22} = \bar{N}_{22}^{-1} \end{array}\right\}$$

which proves (1.272). Eq.(1.273) can be proved in a similar way using the relation $Q\ N = I$.

As the inverse matrix of N is unique, the corresponding partitioned inverse matrices in (1.272) and (1.273) should be identical. This means:

$$Q_{11} = \bar{N}_{11}^{-1} = N_{11}^{-1} + N_{11}^{-1}N_{12}\bar{N}_{22}^{-1}N_{21}N_{11}^{-1}$$

$$Q_{22} = \bar{N}_{22}^{-1} = N_{22}^{-1} + N_{22}^{-1}N_{21}\bar{N}_{11}^{-1}N_{12}N_{22}^{-1}$$

$$Q_{12} = -N_{11}^{-1}N_{12}\bar{N}_{22}^{-1} = -\bar{N}_{11}^{-1}N_{12}N_{22}^{-1}$$

or more explicitly:

$$\left.\begin{array}{l} (N_{11} - N_{12}N_{22}^{-1}N_{21})^{-1} = N_{11}^{-1} + N_{11}^{-1}N_{12}\bar{N}_{22}^{-1}N_{21}N_{11}^{-1} \\ (N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1} = N_{22}^{-1} + N_{22}^{-1}N_{21}\bar{N}_{11}^{-1}N_{12}N_{22}^{-1} \end{array}\right\} \tag{1.275}$$

and

$$N_{11}^{-1}N_{12}(N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1} = (N_{11} - N_{12}N_{22}^{-1}N_{21})^{-1}N_{12}N_{22}^{-1} \tag{1.276}$$

Eqs. (1.275) and (1.276) represent the same identities as (1.267) and (1.268), respectively.

1.7.3 Trace of Square Matrices

The trace of a square matrix A:

$$\underset{n\cdot n}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

is defined as the sum of all diagonal elements of A :

$$tr(A) = \sum_{i=1}^{n} (a_{ii}) \tag{1.277}$$

Obviously, the trace of an $n \times n$ unit matrix $\underset{n\times n}{I}$ is equal to n :

$$tr\left(\underset{n\times n}{I}\right) = n \tag{1.278}$$

A number a can be regarded as an 1×1 square matrix and therefore, the trace of a number is equal to the number itself:

$$tr(a) = a \tag{1.279}$$

The following important properties can be easily proved:

1. If A, B are two square matrices of same dimension,

$$tr\left(A + B\right) = tr\left(A\right) + tr\left(B\right) \tag{1.280}$$

2. If λ is a constant and A is a square matrix:

$$tr\left(\lambda A\right) = \lambda tr\left(A\right) \tag{1.281}$$

3. A square matrix and its transpose have equal trace:

$$tr\left(A\right) = tr\left(A^{\top}\right) \tag{1.282}$$

4. If A and B are two compatible matrices such that both AB and BA are square matrices,

$$tr(AB) = tr(BA) \tag{1.283}$$

5. If A and B are two compatible matrices such that both $AB^{\top}$ and $A^{\top}B$ are square matrices,

$$tr\left(AB^{\top}\right) = tr\left(A^{\top}B\right) \tag{1.284}$$

6. If matrix A has dimension $n \times m$ and X has dimension $m \times n$,

$$\frac{\partial\left[tr\left(AX\right)\right]}{\partial X} = A^{\top} \tag{1.285}$$

Chapter 2

Condition Adjustment

All observations contain measurement errors. To increase the accuracy and reliability of the final results, we almost always make redundant measurements, *i.e.* taking more observations than absolutely needed. With both measurement errors and redundancy present, there will exist ″contradiction″ among measurements, called *misclosure*. A simple example is the determination of the horizontal position of an unknown point P from two fixed points A and B through angle measurements (Cf Fig. 2.1). Geometrically, it is enough with two angle measurements (e.g. $\ell_1,\ \ell_2$) to determine point P. However, in order to be able to detect possible errors and improve the accuracy of the derived position of point P, we often measure all three angles ℓ_1, ℓ_2, ℓ_3.

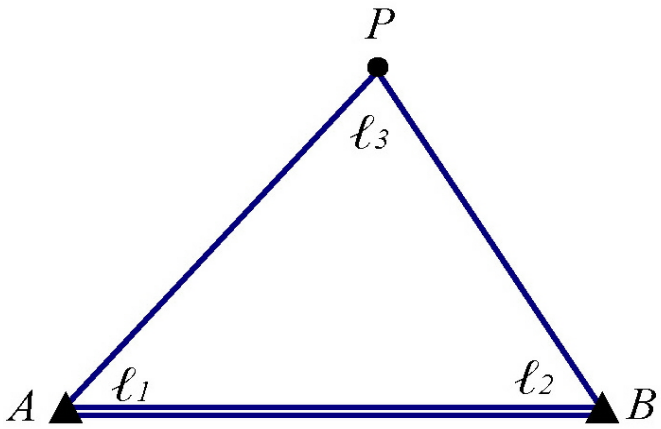

Figure 2.1: A simple triangle

Theoretically, the sum of the three angles should be equal to 180^0 :

$$\widetilde{\ell}_1 + \widetilde{\ell}_2 + \widetilde{\ell}_3 - 180^0 = 0 \tag{2.1}$$

where $\widetilde{\ell}_i$ denotes the true value of ℓ_i $(i = 1, 2, 3)$. Denoting the error of ℓ_i by ε_i and writing $\widetilde{\ell}_i$ as $\ell_i - \varepsilon_i$, Eq.(2.1) becomes:

$$\varepsilon_1 + \varepsilon_2 + \varepsilon_3 = w \tag{2.2}$$

where:

$$w = (\ell_1 + \ell_2 + \ell_3) - 180^0 \tag{2.3}$$

In least squares adjustment, Eq. (2.1) and Eq. (2.22) are called the *condition equation* for the (true) observations and for the (true) observation errors, respectively, while w is called the *misclosure*.

Due to measurement errors, the sum of the observed values of these three angles will not be equal to 180^0, *i.e.* $w \neq 0$. As Eq. (2.2) holds for the true errors ε_i , any meaningful estimate $\widehat{\varepsilon}_i$ of ε_i should naturally satisfy condition Eq. (2.2) too:

$$\widehat{\varepsilon}_1 + \widehat{\varepsilon}_2 + \widehat{\varepsilon}_3 = w \tag{2.4}$$

In this example, we have $n = 3$ observations, $m = 2$ necessary observations needed to determine point P and consequently we have $t = n - m = 1$ condition equation to be satisfied. As t is always smaller than n, there are infinite number of solutions to Eq. (2.2). In geodesy and surveying, we choose the optimal solution to Eq. (2.2) according to the least squares principle described in Subsection 1.5.1, *i.e.* choosing the estimate $\widehat{\varepsilon}_i$ such that:

$$\sum_{i=1}^{n} (p_i \widehat{\varepsilon}_i \widehat{\varepsilon}_i) = minimum \tag{2.5}$$

where p_i denotes the weight of ℓ_i $(i = 1, 2, 3)$ and $\widehat{\varepsilon}_i$ denotes the estimated error of ℓ_i, or the estimated *residual* of ℓ_i.

In mathematical language this is a conditional minimization problem, namely finding a solution $\widehat{\varepsilon}_i$ such that

- Eq.(2.4) is satisfied and
- $\sum_{i=1}^{n} (p_i \widehat{\varepsilon}_i \widehat{\varepsilon}_i)$ is minimized.

The above example for a simple surveying problem demonstrated the basic ideas of condition adjustment using the least squares principle. Normally, geodetic and photogrammetric networks are much more complicated, likely with many observations and consequently many condition equations. In this chapter, we first give a general description on the classical condition adjustment and then present several modified versions of condition adjustment.

2.1 Condition Adjustment with Linear Condition Equations

2.1.1 Basic Formulas

Before we derive the basic formulas of condition adjustment, we need to be clear with our observations and their a priori statistical properties. Assume that we have n observations: ℓ_i ($i = 1, 2, \cdots, n$) and that their true values, true errors and weights are denoted by $\widetilde{\ell}_i$, ε_i and p_i, respectively. Using matrix notations, we then have:

$$\underset{n\cdot 1}{\widetilde{L}} = \underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} \tag{2.6}$$

where:

$$\underset{n\cdot 1}{\widetilde{L}} = \begin{bmatrix} \widetilde{\ell}_1 \\ \widetilde{\ell}_2 \\ \cdots \\ \widetilde{\ell}_n \end{bmatrix}, \quad \underset{n\cdot 1}{L} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \cdots \\ \ell_n \end{bmatrix}, \quad \underset{n\cdot 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \tag{2.7}$$

Furthermore, we assume that ε has zero expectation :

$$E[\varepsilon] = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \cdots \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} = \underset{n\cdot 1}{0} \tag{2.8}$$

and a variance-covariance matrix as:

$$\underset{n\cdot n}{C_{\varepsilon\varepsilon}} = E\left\{[\varepsilon - E(\varepsilon)][\varepsilon - E(\varepsilon)]^\top\right\} = E(\varepsilon\varepsilon^\top) = \sigma_0^2\ P^{-1} \tag{2.9}$$

where σ_0^2 is the a priori variance factor (unit-weight standard error squared) and P is the weight matrix of L:

$$\underset{n\cdot n}{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} = P^\top \tag{2.10}$$

If $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are uncorrelated with each other, P and also $C_{\varepsilon\varepsilon}$ will be diagonal matrices:

$$\underset{n\cdot n}{P} = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \cdots & \\ & & & p_n \end{bmatrix} \tag{2.11}$$

Since L and ε differ from each other only by a non-random vector $\widetilde{L}$, L and ε have the same variance-covariance matrix:

$$C_{LL} = C_{\varepsilon\varepsilon} = \sigma_0^2\ P^{-1} \tag{2.12}$$

Now let us consider the case that there exist t conditions among the true values of the n observations:

$$\left.\begin{array}{l} b_{11}\widetilde{\ell}_1 + b_{12}\widetilde{\ell}_2 + \cdots + b_{1n}\widetilde{\ell}_n + c_1 = 0 \\ b_{21}\widetilde{\ell}_1 + b_{22}\widetilde{\ell}_2 + \cdots + b_{2n}\widetilde{\ell}_n + c_2 = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ b_{t1}\widetilde{\ell}_1 + b_{t2}\widetilde{\ell}_2 + \cdots + b_{tn}\widetilde{\ell}_n + c_t = 0 \end{array}\right\} \tag{2.13}$$

where b_{ij} and c_i ($1 \le i \le t$; $1 \le j \le n$) are known constants. Using matrix notations, the above t equations can be written as one matrix equation:

$$\underset{t\cdot n}{B}\ \underset{n\cdot 1}{\widetilde{L}} + \underset{t\cdot 1}{c} = \underset{t\cdot 1}{0} \tag{2.14}$$

where:

$$\underset{t\cdot n}{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{t1} & b_{t2} & \cdots & b_{tn} \end{bmatrix}, \quad \underset{t\cdot 1}{c} = \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_t \end{bmatrix} \tag{2.15}$$

Substitution of Eq.(2.6) into Eq.(2.14) leads to :

$$\underset{t\cdot n}{B} \cdot \underset{n\cdot 1}{\varepsilon} = \underset{t\cdot 1}{W} \tag{2.16}$$

where W denotes the misclosure vector :

$$\underset{t\cdot 1}{W} = BL + c = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_t \end{bmatrix} \tag{2.17}$$

Eq.(2.14) is called the condition equation for the true observations $\widetilde{L}$, while Eq.(2.16) is called the condition equation for the residuals ε.

The basic task of condition adjustment by the least squares method is to find an estimate $\widehat{\varepsilon}$ for the true error ε which has a priori statistical properties as given by Eq.(2.8) and Eq.(2.9), such that the following two conditions are satisfied:

$$\begin{gathered} \widehat{\varepsilon}^{\top} P \widehat{\varepsilon} = minimum \\ B\widehat{\varepsilon} = W \end{gathered} \tag{2.18}$$

To solve the above conditional minimization problem, one may form the *Lagrangian function*:

$$F = \varepsilon^{\top} P \varepsilon - 2\lambda^{\top} (B\varepsilon - W)$$

where λ is the so called *Lagrange multiplier*. Letting $dF/d\varepsilon = 0$ will lead to the least squares estimate $\widehat{\varepsilon}$:

$$\frac{dF}{d\varepsilon} \mid_{\varepsilon=\widehat{\varepsilon}} = \left(2\varepsilon^{\top} P - 2\ \lambda^{\top} B\right) \mid_{\varepsilon=\widehat{\varepsilon}} = 2\widehat{\varepsilon}^{\top} P - 2\lambda^{\top} B = 0$$

or:

$$\widehat{\varepsilon} = P^{-1} B^{\top} \lambda \tag{2.19}$$

Inserting Eq.(2.19) into Eq.(2.18) leads to the so called *normal equation* of condition adjustment :

$$(BP^{-1}B^{\top})\lambda = W \tag{2.20}$$

from which λ can be solved :

$$\lambda = (BP^{-1}B^{\top})^{-1} W \tag{2.21}$$

Putting λ back into Eq.(2.19) will give the least squares estimate $\widehat{\varepsilon}$:

$$\widehat{\varepsilon} = P^{-1} B^{\top} (BP^{-1}B^{\top})^{-1} W \tag{2.22}$$

The adjusted (corrected) value of observations L will be:

$$\widehat{L} = L - \widehat{\varepsilon} = L - P^{-1} B^{\top} (BP^{-1}B^{\top})^{-1} W \tag{2.23}$$

Example 2.1

We use condition adjustment to adjust the three angles shown in Fig. 2.1. assuming that the observed values of the angles and their weights are as follows:

$$\underset{3\cdot 1}{L} = \begin{bmatrix} 60^0\ 00'\ 03'' \\ 60^0\ 00'\ 03'' \\ 59^0\ 59'\ 51'' \end{bmatrix}, \quad \underset{3\cdot 3}{P} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = \underset{3\cdot 3}{I}$$

where I denotes the unit matrix of dimension 3×3. The triangular misclosure is given by Eq. (2.3):

$$w = (\ell_1 + \ell_2 + \ell_3) - 180^0 = -3''$$

Our condition equation is [see Eq. (2.2)] :

$$\underset{1\cdot 3}{B} \cdot \underset{3\cdot 1}{\varepsilon} = \underset{1\cdot 1}{W}, \quad B = (1 \;\; 1 \;\; 1), \quad W = (w) = (-3)$$

Using the basic formulas of condition adjustment, Eqs. (2.21), (2.19) and (2.23), , we have:

$$BP^{-1}B^\top = BB^\top = 3$$
$$\lambda = (BP^{-1}B^\top)^{-1}W = 1/3 \; w = -1''$$
$$\widehat{\varepsilon} = \begin{bmatrix} \widehat{\varepsilon}_1 \\ \widehat{\varepsilon}_2 \\ \widehat{\varepsilon}_3 \end{bmatrix} = B^\top\lambda = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} (-1'') = \begin{bmatrix} -1'' \\ -1'' \\ -1'' \end{bmatrix}$$
$$\widehat{L} = \begin{bmatrix} \widehat{\ell}_1 \\ \widehat{\ell}_2 \\ \widehat{\ell}_3 \end{bmatrix} = L - \widehat{\varepsilon} = \begin{bmatrix} 60^0\; 00'\; 04'' \\ 60^0\; 00'\; 04'' \\ 59^0\; 59'\; 52'' \end{bmatrix}$$

As verification, we calculate the sum of the three adjusted angles:

$$\widehat{\ell}_1 + \widehat{\ell}_2 + \widehat{\ell}_3 = 60^0\; 00'\; 04'' + 60^0\; 00'\; 04'' + 59^0\; 59'\; 52'' = 180^0\; 00'\; 00''$$

which shows that there exists no longer misclosure among the angles after adjustment. ∎

2.1.2 A Posteriori Unit-Weight Standard Error

The unit-weight standard error σ_0 in (2.12) represents the error information we had before the adjustment and therefore, it is called the *a priori unit-weight standard error*. Eqs. (2.22) and (2.23) show that the results of least squares condition adjustment depends only on P, but not on σ_0. This implies that least squares adjustment requires knowledge on the *relative accuracy* of the observations, not the absolute accuracy as represented by σ_0.

Nevertheless, from the adjustment results $(\widehat{\varepsilon})$ one can construct an estimate to σ_0^2 . As this estimate is obtained after condition adjustment, we call it an *a posteriori* estimate. We are going to prove that the following estimate:

$$\widehat{\sigma}_0^2 = \frac{\widehat{\varepsilon}^\top P \widehat{\varepsilon}}{t} \tag{2.24}$$

is an *unbiased estimate* to the variance factor σ_0^2 , where t is the number of over-determinations, or number of independent condition equations in (2.14). When the weight matrix P is diagonal [Cf (2.11)], (2.24) can be written explicitly as:

$$\widehat{\sigma}_0^2 = \frac{1}{t}\sum_{i=1}^{n}(p_i\widehat{\varepsilon}_i\widehat{\varepsilon}_i) \tag{2.25}$$

To prove the general estimate in (2.24) is an unbiased estimate of σ_0^2, we look at the expectation of $\widehat{\sigma}_0^2$:

$$E\left\{\widehat{\sigma}_0^2\right\} = \frac{1}{t}E\left\{\widehat{\varepsilon}^\top P\widehat{\varepsilon}\right\} = \frac{1}{t}E\left\{W^\top(BP^{-1}B^\top)^{-1}BP^{-1}\cdot P\cdot P^{-1}B^\top(BP^{-1}B^\top)^{-1}W\right\} =$$
$$= \frac{1}{t}E[W^\top(BP^{-1}B^\top)^{-1}W]$$

As the trace[1] of a single number is equal to the number itself, the above result can be rewritten as:

$$E\left\{\widehat{\sigma}_0^2\right\} = tr\left[E\left\{\widehat{\sigma}_0^2\right\}\right] = E\left\{tr\left[\widehat{\sigma}_0^2\right]\right\} = E\left\{\frac{1}{t}tr[W^\top(BP^{-1}B^\top)^{-1}W]\right\} =$$

[1] For definition and properties of matrix trace, see Subsection 1.6.3.

$$= \frac{1}{t} E\left\{tr[(BP^{-1}B^\top)^{-1}W\ W^\top]\right\} = \frac{1}{t} tr\left\{(BP^{-1}B^\top)^{-1}E(WW^\top)\right\}$$

where we have assumed that the order of two operators, $E(\cdot)$ and $tr(\cdot)$, can be exchanged and that identity (1.283) is utilized. The expectation and variance matrix of W can be obtained from (2.16):

$$E[W] = E[B\varepsilon] = B\ E[\varepsilon] = 0 \tag{2.26}$$

$$C_{WW} = E\left[WW^\top\right] = E\left[B\varepsilon \cdot \varepsilon^\top B^\top\right] = \sigma_0^2\ BP^{-1}B^\top \tag{2.27}$$

Thus $E(\widehat{\sigma}_0^2)$ can be further written as:

$$E(\widehat{\sigma}_0^2) = \frac{1}{t} tr\left\{(BP^{-1}B^\top)^{-1} \cdot \sigma_0^2\ BP^{-1}B^\top\right\} = \sigma_0^2 \cdot \frac{1}{t} tr\left(\underset{t\cdot t}{I}\right) = \sigma_0^2$$

which proves that $\widehat{\sigma}_0^2$ is an unbiased estimate of σ_0^2 . Furthermore, it can be also proved that $\widehat{\sigma}_0^2$ is an optimal estimate of σ_0^2 among all estimates derived from a quadratic form of $\widehat{\varepsilon}$ (or W) in the sense that it has the minimum variance. Thus $\widehat{\sigma}_0^2$ in (2.24) can be called a *Best Quadratic Unbiased Estimate* (BQUE) of σ_0^2. See Chapter 5 for more details.

2.1.3 Variance-Covariance Matrices of Adjusted Results

Due to measurement errors in the observations, all results after adjustment also contain errors. Therefore, in condition adjustment one should obtain not only estimates $\widehat{\varepsilon}$ and $\widehat{L}$ but also their variance-covariance matrices. The variance-covariance matrices of the observation L and the misclosure W are already given in Eqs. (2.12) and (2.27), respectively. Considering relation (2.22), the variance-covariance matrix of $\widehat{\varepsilon}$ can be obtained from the error propagation law in (1.39) :

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = P^{-1}B^\top(BP^{-1}B^\top)^{-1}\ C_{WW}\ \left[P^{-1}B^\top(BP^{-1}B^\top)^{-1}\right]^\top = \sigma_0^2 P^{-1}B^\top(BP^{-1}B^\top)^{-1}BP^{-1} \tag{2.28}$$

Inserting Eqs. (2.6) and (2.16) into (2.23), we can rewrite the adjusted observations $\widehat{L}$ as :

$$\widehat{L} = L - \widehat{\varepsilon} = \left(\tilde{L} + \varepsilon\right) - P^{-1}B^\top(BP^{-1}B^\top)^{-1}W = \left[I - P^{-1}B^\top(BP^{-1}B^\top)^{-1}B\right]\varepsilon + \tilde{L} \tag{2.29}$$

Since $\tilde{L}$ does not contain errors, $\widehat{L}$ should have the same variance matrix as the first term on the right side of the above equation:

$$C_{\widehat{L}\widehat{L}} = \left[I - P^{-1}B^\top(BP^{-1}B^\top)^{-1}B\right] \cdot \sigma_0^2 P^{-1} \cdot \left[I - P^{-1}B^\top(BP^{-1}B^\top)^{-1}B\right]^\top =$$

$$= \sigma_0^2\left[P^{-1} - P^{-1}B^\top(BP^{-1}B^\top)^{-1}BP^{-1}\right] \tag{2.30}$$

From (2.22) and (2.29), one can also derive the covariance matrix between $\widehat{\varepsilon}$ and $\widehat{L}$ [see Eq.(1.42)] [2] :

$$C_{\widehat{\varepsilon}\widehat{L}} = P^{-1}B^\top(BP^{-1}B^\top)^{-1}B \cdot \sigma_0^2 P^{-1} \cdot \left[I - P^{-1}B^\top(BP^{-1}B^\top)^{-1}B\right]^\top = 0 \tag{2.31}$$

Comparing Eq. (2.12) with (2.28) and (2.30), one can get an interesting relation:

$$C_{LL} = C_{\widehat{L}\widehat{L}} + C_{\widehat{\varepsilon}\widehat{\varepsilon}} \tag{2.32}$$

If we replace the theoretical variance factor σ_0^2 in Eqs. (2.12), (2.27), (2.28) and (2.30) by the a posteriori estimate $\widehat{\sigma}_0^2$, we then arrive at estimates of the corresponding variance-covariance matrices :

$$\left.\begin{aligned} C_{LL} &= \widehat{\sigma}_0^2\ P^{-1} \\ C_{WW} &= \widehat{\sigma}_0^2\ BP^{-1}B^\top \\ C_{\widehat{\varepsilon}\widehat{\varepsilon}} &= \widehat{\sigma}_0^2\ P^{-1}B^\top(BP^{-1}B^\top)^{-1}BP^{-1} \\ C_{\widehat{L}\widehat{L}} &= \widehat{\sigma}_0^2\ \left[P^{-1} - P^{-1}B^\top(BP^{-1}B^\top)^{-1}BP^{-1}\right] \end{aligned}\right\} \tag{2.33}$$

[2] This interesting result, that the covariance between $\widehat{L}$ and $\widehat{\varepsilon}$ is zero, can be explained by the *projection theorem in vector spaces* (see e.g. Luenberger, 1968, pp.50).

Considering the relationship between the variance-covariance matrix and the cofactor matrix of a random vector, the above equations can be further rewritten as:

$$\left.\begin{aligned} C_{LL} &= \widehat{\sigma}_0^2\, Q_{LL} \\ C_{WW} &= \widehat{\sigma}_0^2\, Q_{WW} \\ C_{\widehat{\varepsilon}\widehat{\varepsilon}} &= \widehat{\sigma}_0^2\, Q_{\widehat{\varepsilon}\widehat{\varepsilon}} \\ C_{\widehat{L}\widehat{L}} &= \widehat{\sigma}_0^2\, Q_{\widehat{L}\widehat{L}} \end{aligned}\right\} \tag{2.34}$$

Comparing equations in (2.33) with equations in (2.34), the cofactor matrices are found :

$$\left.\begin{aligned} Q_{LL} &= P^{-1} \\ Q_{WW} &= BP^{-1}B^{\top} \\ Q_{\widehat{\varepsilon}\widehat{\varepsilon}} &= P^{-1}B^{\top}(BP^{-1}B^{\top})^{-1}BP^{-1} \\ Q_{\widehat{L}\widehat{L}} &= P^{-1} - P^{-1}B^{\top}(BP^{-1}B^{\top})^{-1}BP^{-1} \end{aligned}\right\} \tag{2.35}$$

Now let us consider a random vector y of linear functions of the adjusted observation $\widehat{L}$:

$$y = G\,\widehat{L} \tag{2.36}$$

where G denotes a given matrix. Considering (2.33) and (2.35), one can obtain the variance-covariance matrix and cofactor matrix of y by applying the error propagation laws for variance-covariance matrices and cofactor matrices :

$$\left.\begin{aligned} C_{yy} &= G\,C_{\widehat{L}\widehat{L}}\,G^{\top} = \widehat{\sigma}_0^2\,G\left[P^{-1} - P^{-1}B^{\top}(BP^{-1}B^{\top})^{-1}BP^{-1}\right]G^{\top} \\ Q_{yy} &= G\,Q_{\widehat{L}\widehat{L}}\,G^{\top} = G\left[P^{-1} - P^{-1}B^{\top}(BP^{-1}B^{\top})^{-1}BP^{-1}\right]G^{\top} \end{aligned}\right\} \tag{2.37}$$

The above results on the error estimates of the adjusted quantities indicate that cofactor matrices of the adjusted quantities are uniquely defined by matrix P (describing the relative accuracy of observations L) and matrix B (characterizing the geometrical configuration of the networks), more or less independent of the exact values of L and the adjusted residuals. This provides us a possibility to study the error aspects of a geodetic network before actual measurements are made in the field. The absolute error estimates, namely the variance-covariance matrices, naturally depend on not only P and B but also the actual measurement data L and their accuracy.

Each diagonal element in the variance-covariance matrices derived above denotes the variance of one specific element. The square root of these diagonal elements will be the standard errors of the respective elements, while non-diagonal elements represent the covariances between two elements.

Example 2.2

In a levelling network with 4 unknown benchmarks (see Fig. 2.2), five height differences ℓ_i $(1 \le i \le 5)$ have been measured (s_i denotes the length of the levelling line for ℓ_i) :

$$\begin{aligned} \ell_1 &= +1.002\ metres, & s_1 &\approx 1\ km \\ \ell_2 &= +2.004\ metres, & s_2 &\approx 1\ km \\ \ell_3 &= -2.001\ metres, & s_3 &\approx 2\ km \\ \ell_4 &= -1.002\ metres, & s_4 &\approx 2\ km \\ \ell_5 &= +3.012\ metres, & s_5 &\approx 1\ km \end{aligned}$$

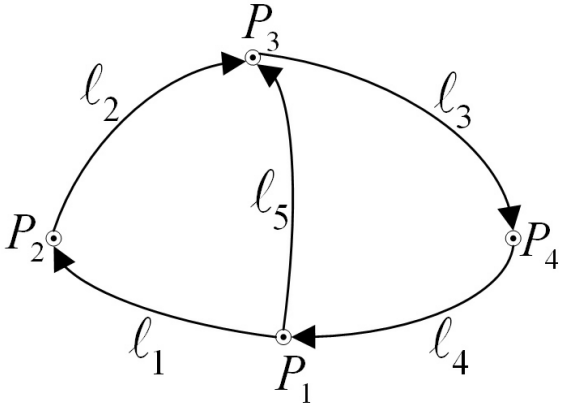

Figure 2.2: A simple levelling network

The condition equations are:

$$\begin{aligned} \widetilde{\ell}_1 + \widetilde{\ell}_2 - \widetilde{\ell}_5 &= 0 \\ \widetilde{\ell}_3 + \widetilde{\ell}_4 + \widetilde{\ell}_5 &= 0 \end{aligned}$$

or:

$$\underset{2\cdot 5}{B} \cdot \underset{5\cdot 1}{\varepsilon} = \underset{2\cdot 1}{W}, \quad B = \begin{bmatrix} 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}, \quad W = \begin{bmatrix} -6\,^{mm} \\ +9\,^{mm} \end{bmatrix}$$

Assume that all height differences ℓ_i are uncorrelated with each other and that the weight of ℓ_i is inversely proportional to levelling length s_i :

$$p_i = \frac{2\,^{km}}{s_i\,^{km}} \tag{2.38}$$

Then our weight matrix P and cofactor matrix Q for the observations are:

$$\underset{5\cdot 5}{P} = \begin{bmatrix} 2 & & & & \\ & 2 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 2 \end{bmatrix}, \quad Q = P^{-1} = \frac{1}{2} \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 2 & & \\ & & & 2 & \\ & & & & 1 \end{bmatrix}$$

The estimated residuals ($\widehat{\varepsilon}$) are given by Eq.(2.22) and the adjusted height differences ($\widehat{L}$) are given by Eq.(2.23):

$$BP^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 1 & +1 \end{bmatrix} \tfrac{1}{2} \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 2 & & \\ & & & 2 & \\ & & & & 1 \end{bmatrix} = \tfrac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 2 & 2 & 1 \end{bmatrix}$$

$$BP^{-1}B^{\top} = \tfrac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & +1 \end{bmatrix} = \tfrac{1}{2} \begin{bmatrix} 3 & -1 \\ -1 & 5 \end{bmatrix}$$

$$(BP^{-1}B^\top)^{-1} = \tfrac{1}{7}\begin{bmatrix} 5 & 1 \\ 1 & 3 \end{bmatrix}$$

$$P^{-1}B^\top(BP^{-1}B^\top)^{-1} = \tfrac{1}{2}\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 2 \\ 0 & 2 \\ -1 & 1 \end{bmatrix} \tfrac{1}{7}\begin{bmatrix} 5 & 1 \\ 1 & 3 \end{bmatrix} = \tfrac{1}{14}\begin{bmatrix} 5 & 1 \\ 5 & 1 \\ 2 & 6 \\ 2 & 6 \\ -4 & 2 \end{bmatrix}$$

$$\widehat{\varepsilon} = P^{-1}B^\top(BP^{-1}B^\top)^{-1}W = \tfrac{1}{14}\begin{bmatrix} 5 & 1 \\ 5 & 1 \\ 2 & 6 \\ 2 & 6 \\ -4 & 2 \end{bmatrix} \begin{bmatrix} -6\ ^{mm} \\ +9\ ^{mm} \end{bmatrix} = \begin{bmatrix} -1.5\ ^{mm} \\ -1.5\ ^{mm} \\ +3.0\ ^{mm} \\ +3.0\ ^{mm} \\ +3.0\ ^{mm} \end{bmatrix} = \begin{bmatrix} -0.0015\ ^{m} \\ -0.0015\ ^{m} \\ +0.0030\ ^{m} \\ +0.0030\ ^{m} \\ +0.0030\ ^{m} \end{bmatrix}$$

$$\widehat{L} = L - \widehat{\varepsilon} = \begin{bmatrix} +1.002 \\ +2.004 \\ -2.001 \\ -1.002 \\ +3.012 \end{bmatrix} - \begin{bmatrix} -0.0015 \\ -0.0015 \\ +0.0030 \\ +0.0030 \\ +0.0030 \end{bmatrix} = \begin{bmatrix} +1.0035 \\ +2.0055 \\ -2.0040 \\ -1.0050 \\ +3.0090 \end{bmatrix} \ (m)$$

The unit-weight standard error is estimated using (2.25):

$$\widehat{\sigma}_0^2 = \frac{\widehat{\varepsilon}^\top P\widehat{\varepsilon}}{t} = \frac{1}{2}\left\{2\cdot(-1.5)^2 + 2\cdot(-1.5)^2 + 1\cdot(+3.0)^2 + 1\cdot(+3.0)^2 + 2\cdot(+3.0)^2\right\} = 22.50\ mm^2$$

$$\widehat{\sigma}_0 = \sqrt{22.50} = \pm 4.74\ \ mm$$

Eq.(2.38) implies that a levelling observation at length of $2\ ^{km}$ has weight 1 and thus has standard error equal to $\widehat{\sigma}_0$. Consequently, one may estimate the standard error for levelling lines at length of $1\ ^{km}$ as follows:

$$\widehat{\sigma}_{1\ km} = \frac{\widehat{\sigma}_0}{\sqrt{2}} = \sqrt{11.25} \approx 3.35\ mm/\sqrt{km}$$

Accordingly, the standard error for a levelling measurement at length of s (in km) will be:

$$\widehat{\sigma}_s = \widehat{\sigma}_{1\ km}\sqrt{s} = 3.35\sqrt{s}\ \ (mm)$$

The cofactor matrices and variance matrices of L, $\widehat{\varepsilon}$ and $\widehat{L}$ are obtained using Eqs. (2.35) and (2.33), respectively:

$$Q_{LL} = Q = P^{-1} = \tfrac{1}{2}\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 2 & & \\ & & & 2 & \\ & & & & 1 \end{bmatrix}$$

$$Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = P^{-1}B^\top(BP^{-1}B^\top)^{-1}BP^{-1} = \tfrac{1}{28}\begin{bmatrix} 5 & 5 & 2 & 2 & -4 \\ 5 & 5 & 2 & 2 & -4 \\ 2 & 2 & 12 & 12 & 4 \\ 2 & 2 & 12 & 12 & 4 \\ -4 & -4 & 4 & 4 & 6 \end{bmatrix}$$

$$Q_{\widehat{L}\widehat{L}} = P^{-1} - P^{-1}B^\top(BP^{-1}B^\top)^{-1}BP^{-1} = Q_{LL} - Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = \tfrac{1}{28}\begin{bmatrix} 9 & -5 & -2 & -2 & 4 \\ -5 & 9 & -2 & -2 & 4 \\ -2 & -2 & 16 & -12 & -4 \\ -2 & -2 & -12 & 16 & -4 \\ 4 & 4 & -4 & -4 & 8 \end{bmatrix}$$

$$C_{LL} = \widehat{\sigma}_0^2\, Q_{LL} = \tfrac{22.50}{2}\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 2 & & \\ & & & 2 & \\ & & & & 1 \end{bmatrix}$$

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = \widehat{\sigma}_0^2\, Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = \frac{22.50}{28} \begin{bmatrix} 5 & 5 & 2 & 2 & -4 \\ 5 & 5 & 2 & 2 & -4 \\ 2 & 2 & 12 & 12 & 4 \\ 2 & 2 & 12 & 12 & 4 \\ -4 & -4 & 4 & 4 & 6 \end{bmatrix}$$

$$C_{\widehat{L}\widehat{L}} = \widehat{\sigma}_0^2\, Q_{\widehat{L}\widehat{L}} = \frac{22.50}{28} \begin{bmatrix} 9 & -5 & -2 & -2 & 4 \\ -5 & 9 & -2 & -2 & 4 \\ -2 & -2 & 16 & -12 & -4 \\ -2 & -2 & -12 & 16 & -4 \\ 4 & 4 & -4 & -4 & 8 \end{bmatrix}$$

From C_{LL} and $C_{\widehat{L}\widehat{L}}$, one can further calculate the standard errors of each height difference before and after the adjustment, respectively:

Before adjustment	*After adjustment*
$\widehat{\sigma}_1 = \sqrt{\frac{22.50}{2} \cdot 1} = \pm 3.35\ mm$	$\widehat{\sigma}_1 = \sqrt{\frac{22.50}{28} \cdot 9} = \pm 2.69\ mm$
$\widehat{\sigma}_2 = \sqrt{\frac{22.50}{2} \cdot 1} = \pm 3.35\ mm$	$\widehat{\sigma}_2 = \sqrt{\frac{22.50}{28} \cdot 9} = \pm 2.69\ mm$
$\widehat{\sigma}_3 = \sqrt{\frac{22.50}{2} \cdot 2} = \pm 4.74\ mm$	$\widehat{\sigma}_3 = \sqrt{\frac{22.50}{28} \cdot 16} = \pm 3.58\ mm$
$\widehat{\sigma}_4 = \sqrt{\frac{22.50}{2} \cdot 2} = \pm 4.74\ mm$	$\widehat{\sigma}_4 = \sqrt{\frac{22.50}{28} \cdot 16} = \pm 3.58\ mm$
$\widehat{\sigma}_5 = \sqrt{\frac{22.50}{2} \cdot 1} = \pm 3.35\ mm$	$\widehat{\sigma}_5 = \sqrt{\frac{22.50}{28} \cdot 8} = \pm 2.53\ mm$

The above table shows that the standard errors of height differences have been reduced after adjustment by about $20\% - 25\%$, implying that the condition adjustment has improved the accuracy of height differences. It should be pointed out, however, that in this simple example, the number of observations ($n = 5$) and the number of redundancy ($t = 2$) are too small so that the estimated $\widehat{\sigma}$ may be not reliable. One should have larger data set and satisfactory redundancy in order to obtain meaningful and reliable error estimates.

The least squares estimate of the height difference h_{24} between P_2 and P_4 can be obtained from the adjusted observations computed above:

$$h_{24} = \widehat{\ell}_2 + \widehat{\ell}_3 = +2.0055 - 2.0040 = 0.0015\ m$$

The a posteriori variance of h_{24} is obtained from (2.37):

$$\widehat{\sigma}_{h_{24}}^2 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \end{bmatrix} C_{\widehat{L}\widehat{L}} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \frac{67.5}{4} = 16.875\ \ mm^2$$

i.e.

$$\widehat{\sigma}_{h_{24}} = \pm 4.11\ mm\ . \quad \blacksquare$$

2.2 Condition Equations

In the pervious section, we have presented the basic formulas of condition adjustment with given condition equations and given a priori statistical information. Now we will discuss the condition equations in more details. We first discuss the number of condition equations in geodetic networks and then the correlation problem among condition equations. After that, non-linear condition equations will be treated in detail.

2.2.1 Number of Condition Equations

The number of condition equations (t) is equal to the number of over-determinations (redundancy), which in turn is equal to the number of observations (n) minus the number of necessary observations (m) :

$$t = n - m \tag{2.39}$$

Therefore, the key to find the number of condition equations is to find out the number of necessary observations (m). Loosely speaking, necessary observations in geodetic networks are the minimum number of observations needed to define the geometry of the network and determine the relative positions among network points. In general, m depends on the character of the network concerned. Below several relevant concepts are first introduced, before we describe how to calculate m.

Number of Necessary Initial Data (m_0)

Most traditional geodetic observations (e.g. height differences, angles, distances) do not describe the absolute position of points, rather the relative geometrical relation among different points within the network. *Necessary initial data* are the minimum number of reference data needed to define the absolute position of the network with respect to some specified reference system. These initial data together are also known as the *coordinate datum* or *reference datum* of the network.

In a levelling network, the necessary initial data is the given height of one fixed point. The number of necessary initial data is thus $m_0 = 1$.

In station adjustment of directions observed by theodolite at one station, the necessary initial data is one direction, from which all other directions can be determined after direction measurements. This means $m_0 = 1$ for station adjustment.

In two-dimensional (2D) triangulation networks (where only angles/directions are measured), one needs *one fixed point* to define the absolute position of the whole network, *one known azimuth* to define the *orientation* of the network and *one fixed baseline* to define the *scale* of the network. In this case, the number of necessary initial data is $m_0 = 4$. Two fixed points will automatically provide the above 4 initial data.

In 2D triangulation-trilateration networks (where distances or both angles/directions and distances are measured), the number of necessary initial data is $m_0 = 3$, *one fixed point* for defining the absolute position of the network and *one known azimuth* for orientation. Traverses or traverse networks can be regarded as a special type of triangulation- trilateration networks and therefore, their necessary initial data are also one fixed point and one fixed azimuth.

In three-dimensional (3D) networks with only angular measurements, the necessary initial data are: *(a)* the coordinates (x_0, y_0, z_0) of one fixed point; *(b)* three orientation parameters and *(c)* one scale factor of the network. The number of necessary initial data is thus $m_0 = 7$. If distances are also measured, the scale factor is not needed and we have then $m_0 = 6$.

In similar way, one can deduct the necessary initial data of other geodetic or photogrammetric networks after analysing their geometrical and physical characters.

The complete set of necessary initial data should include not only enough number of fixed data but also all necessary types. For instance, a triangulation network with four given quantities consisting of one fixed point and two fixed baselines will still have incomplete initial data, as the necessary initial data should include also one fixed azimuth.

Datum Defect and Rank Defect (d)

A network which has no initial data or incomplete initial data is said to have *datum defect*, i.e. the coordinate datum is not sufficiently defined. The difference between m_0 and the number of existing initial data is called the *rank defect* of the network, denoted by d . For the above-mentioned triangulation network with one fixed point and two fixed baselines, the rank defect is $d = 1$.

For networks with datum defect d, one may regard d quantities of the network as known to artifitially establish a complete set of necessary initial data and then carry out the condition adjustment. Datum defect will influence the calculation of the number of necessary observations for a network. A network with datum defect can always be adjusted using the method of condition adjustment, *if* there exist redundant measurements. This is because condition adjustment primarily tries to eliminate the misclosures existing among the observations according to the least squares principle. Condition adjustment does not directly involve the coordinate reference system to which the coordinates of the network are defined.

Chapter 3 treats another adjustment method, known as *adjustment by elements*. By this method, each observation is expressed as a function of m independent parameters. These parameters are often chosen as the coordinates of the unknown network points. Therefore, the adjustment model is directly related to the coordinate reference system and thus to the initial data of the network. When the network has datum defect, adjustment by elements cannot proceed in the normal way. Further discussions on this problem and the solution will be given in **Section 3.2** and **Section 4.3**.

Extra Constraints (e)

If in a network several quantities, besides those necessary initial data, are fixed (regarded as given and free of errors), then this network is said to have extra constraints. Extra constraints can be fixed baselines, extra fixed azimuths or angles among unknown points. These fixed quantities are assumed to be free of errors and thus not to be adjusted. It should be pointed out that extra constraints defined here should not include fixed baseline and azimuth between fixed points, or those forming the necessary initial data.

Necessary Observations (m)

In a geodetic network with sufficient initial data, the number (m) of necessary observations can be calculated by a simple formula given below. If the network has rank defect d, one first assumes d suitable quantities of the network as fixed to establish a sufficient datum and then calculates m by the following simple formula:

$$m = n_p \cdot n_d - e \tag{2.40}$$

where:

n_p = number of unknown points in the network (after artificial initial data are assumed in case of datum defect);

e = number of extra constraints among unknown points;

n_d = dimension of the network:

$$n_d = \left\{ \begin{array}{ll} 1 & \text{for levelling networks} \\ 1 & \text{for station adjustment} \\ 2 & \text{for 2D-triangulation/trilateration networks} \\ 3 & \text{for 3D-networks} \end{array} \right\} \tag{2.41}$$

When the number of observations (n) and number of necessary observations (m) have been found out, the number of condition equations (t) can be easily obtained from (2.39). As examples, we can calculate the number of condition equations for networks in the previous two examples of this chapter:

Example 2.1: $m_0 = 4$; $d = 0$; $n_p = 1$; $n_d = 2$; $n_p \cdot n_d = 2$; $e = 0$; $m = 2$; $n = 3$; $t = 1$.

Example 2.2: $m_0 = 1$; $d = 1$; $n_p = 3$; $n_d = 1$; $n_p \cdot n_d = 3$; $e = 0$; $m = 3$; $n = 5$; $t = 2$.

For the network of **Example 2.2**, there is no fixed point and one of the four unknown benchmarks is thus assumed to be fixed, which leads to $n_p = 3$.

2.2.2 Linear Correlation of Condition Equations

For a network with n observations and m necessary observations, there exist $t = n - m$ independent conditions among observations. If one has found more than t conditions, some or all of them will be *linearly correlated* with each other, i.e. some condition equations can be obtained from linear combinations of other condition equations. The levelling network illustrated in **Example 2.1** has $t = n - m = 2$ independent condition equations. But one can find totally three condition equations:

$$\left. \begin{array}{l} \widetilde{\ell}_1 + \widetilde{\ell}_2 - \widetilde{\ell}_5 = 0 \\ \widetilde{\ell}_3 + \widetilde{\ell}_4 + \widetilde{\ell}_5 = 0 \\ \widetilde{\ell}_1 + \widetilde{\ell}_2 + \widetilde{\ell}_3 + \widetilde{\ell}_4 = 0 \end{array} \right\}$$

It is easy to see that these three equations are linearly correlated, as adding the first two equations together will lead to the third one.

At this kind of situations when more than t condition equations have been chosen, the system of condition equations is linearly dependent or correlated and the resulted normal equation matrix $BP^{-1}B^\top$ will be singular, i.e.:

$$\left| BP^{-1}B^\top \right| = 0$$

In this case, the traditional inverse matrix of $BP^{-1}B^\top$ is not defined. If one still wants to carry out the condition adjustment using all condition equations, it is then necessary to use generalized matrix inverses for matrix B (see Chapter 4). On the other hand, if one wants to avoid generalized matrix inverses, one should choose only t linearly independent condition equations.

Even if we choose only $t = n - m$ condition equations, there is still a risk that some of them are linearly correlated. As there are totally $t = n - m$ linearly independent condition equations, choosing t linearly correlated equations implies that we have missed some of totally t linearly independent condition equations. Condition adjustment using t linearly correlated condition equations will lead to adjusted results which satisfy only some but not all of the t linearly independent condition equations of the adjustment problem.

How to avoid linearly correlated condition equations depends on the type of the network concerned. One can always start with finding out the number of independent condition equations. For levelling networks, condition equations based on non-overlapping loops should normally be linearly independent (see **Example 2.3** below).

Example 2.3

Fig. 2.3 illustrates a levelling network with one fixed benchmark P_0 with given height H_0 and 3 unknown benchmarks P_i $(1 \leq i \leq 3)$. 6 height differences ℓ_i $(1 \leq i \leq 6)$ have been measured.

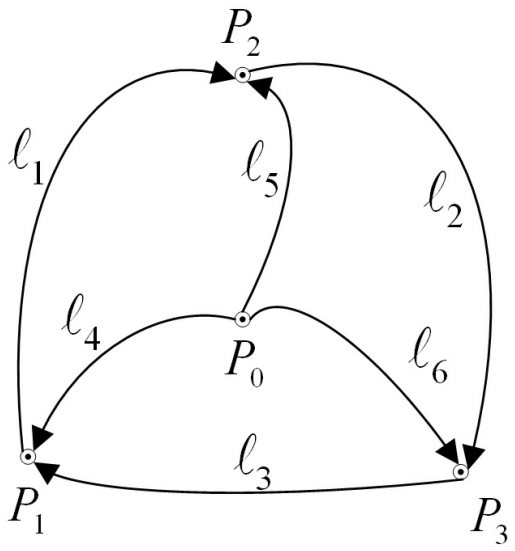

Figure 2.3: Correlated condition equations in a levelling network

For this levelling network, we have:

$$m_0 = 1;\ d = 0;\ n_p = 3;\ n_d = 1;\ n_p \cdot n_d = 3;\ e = 0;\ m = 3;\ n = 6;\ t = n - m = 3$$

It can be found that there exist totally 7 different condition equations for this network. One can thus find:

$$\left(\begin{array}{c} 7 \\ 3 \end{array} \right) = \frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3} = 35$$

combinations or groups, each of which contains three condition equations. The following three groups are examples of linearly independent condition groups:

$$Group\ 1: \quad \begin{cases} \widetilde{\ell}_1 - \widetilde{\ell}_5 + \widetilde{\ell}_4 = 0 \\ \widetilde{\ell}_2 - \widetilde{\ell}_6 + \widetilde{\ell}_5 = 0 \\ \widetilde{\ell}_1 + \widetilde{\ell}_2 + \widetilde{\ell}_3 = 0 \end{cases}$$

$$Group\ 2: \quad \begin{cases} \widetilde{\ell}_1 - \widetilde{\ell}_5 + \widetilde{\ell}_4 = 0 \\ \widetilde{\ell}_2 - \widetilde{\ell}_6 + \widetilde{\ell}_5 = 0 \\ \widetilde{\ell}_3 - \widetilde{\ell}_4 + \widetilde{\ell}_6 = 0 \end{cases}$$

$$Group\ 3: \quad \begin{cases} \widetilde{\ell}_1 - \widetilde{\ell}_5 + \widetilde{\ell}_4 = 0 \\ \widetilde{\ell}_2 - \widetilde{\ell}_6 + \widetilde{\ell}_5 = 0 \\ \widetilde{\ell}_2 + \widetilde{\ell}_3 - \widetilde{\ell}_4 + \widetilde{\ell}_5 = 0 \end{cases}$$

The following three groups are examples of linearly correlated condition groups:

$$Group\ 4: \quad \begin{cases} \widetilde{\ell}_1 - \widetilde{\ell}_5 + \widetilde{\ell}_4 = 0 \\ \widetilde{\ell}_2 - \widetilde{\ell}_6 + \widetilde{\ell}_5 = 0 \\ \widetilde{\ell}_1 + \widetilde{\ell}_2 - \widetilde{\ell}_6 + \widetilde{\ell}_4 = 0 \end{cases}$$

$$\textit{Group } 5: \quad \begin{cases} \widetilde{\ell}_2 - \widetilde{\ell}_6 + \widetilde{\ell}_5 = 0 \\ \widetilde{\ell}_3 - \widetilde{\ell}_4 + \widetilde{\ell}_6 = 0 \\ \widetilde{\ell}_2 + \widetilde{\ell}_3 - \widetilde{\ell}_4 + \widetilde{\ell}_5 = 0 \end{cases}$$

$$\textit{Group } 6: \quad \begin{cases} \widetilde{\ell}_3 - \widetilde{\ell}_4 + \widetilde{\ell}_6 = 0 \\ \widetilde{\ell}_1 - \widetilde{\ell}_5 + \widetilde{\ell}_4 = 0 \\ \widetilde{\ell}_1 - \widetilde{\ell}_5 + \widetilde{\ell}_6 + \widetilde{\ell}_3 = 0 \end{cases}$$

∎

2.2.3 Linearization of Non-Linear Condition Equations

In presenting the general formulas of condition adjustment using the least squares principle, we have assumed that the $t = n - m$ condition equations are linear. However, not all condition equations in geodetic and photogrammetric networks are linear equations. Rigorously speaking, the general formulas given in Section 2.1 are no longer valid for non-linear condition equations. Nevertheless, as measurement errors are relatively very small, these non-linear conditions can be approximated by some linear equations of the measurement errors through the so-called *linearization*. To linearize non-linear condition equations, we use once again the Taylor series expansion as we did when discussing error propagation for non-linear functions (see **Section 1.2**).

Assume that among the true values $\widetilde{\ell}_j$ of observations ℓ_j $(1 \le j \le n)$, there exist t non-linear condition equations, represented by t non-linear functions :

$$\left.\begin{array}{l} f_1(\widetilde{\ell}_1,\ \widetilde{\ell}_2,\ \cdots,\ \widetilde{\ell}_n) = 0 \\ f_2(\widetilde{\ell}_1,\ \widetilde{\ell}_2,\ \cdots,\ \widetilde{\ell}_n) = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ f_i(\widetilde{\ell}_1,\ \widetilde{\ell}_2,\ \cdots,\ \widetilde{\ell}_n) = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ f_t(\widetilde{\ell}_1,\ \widetilde{\ell}_2,\ \cdots,\ \widetilde{\ell}_n) = 0 \end{array}\right\} \tag{2.42}$$

Let ε_j denote the true error of ℓ_j . Then we have: $\widetilde{\ell}_j = \ell_j - \varepsilon_j$ $(1 \le j \le n)$. Now we can expand the $i-$th function in (2.42), $f_i(\widetilde{\ell}_1,\ \widetilde{\ell}_2,\ \cdots,\ \widetilde{\ell}_n)$, into a Taylor series around the measured values ℓ_1 , $\ell_2, \cdots,\ \ell_n$:

$$0 = f_i(\ell_1 - \varepsilon_1,\ \ell_2 - \varepsilon_2, \cdots,\ \ell_n - \varepsilon_n) =$$

$$= f_i(\ell_1,\ \ell_2, \cdots,\ \ell_n) + \frac{1}{1!}\left[\frac{\partial f_i}{\partial \ell_1}(-\varepsilon_1) + \frac{\partial f_i}{\partial \ell_2}(-\varepsilon_2) + \cdots + \frac{\partial f_i}{\partial \ell_n}(-\varepsilon_n)\right] +$$

$$+ \frac{1}{2!}\left[\frac{\partial^2 f_i}{\partial \ell_1^2}(-\varepsilon_1)^2 + \frac{\partial^2 f_i}{\partial \ell_1 \partial \ell_2}(-\varepsilon_1)(-\varepsilon_2) + \cdots + \frac{\partial^2 f_i}{\partial \ell_1 \partial \ell_n}(-\varepsilon_n)^2\right] + \cdots$$

If we neglect all terms of ε_j of order 2 or higher, the above equation will become a linear equation for ε_j :

$$f_i(\ell_1,\ \ell_2, \cdots,\ \ell_n) + \frac{\partial f_i}{\partial \ell_1}(-\varepsilon_1) + \frac{\partial f_i}{\partial \ell_2}(-\varepsilon_2) + \cdots + \frac{\partial f_i}{\partial \ell_n}(-\varepsilon_n) \approx 0 \tag{2.43}$$

or:

$$b_{i1}\varepsilon_1 + b_{i2}\varepsilon_2 + \cdots + b_{ij}\varepsilon_j + b_{i,j+1}\varepsilon_{j+1} + \cdots + b_{in}\varepsilon_n = w_i \tag{2.44}$$

where:

$$b_{ij} = \frac{\partial f_i}{\partial \ell_j}\ , \quad w_i = f_i(\ell_1,\ \ell_2, \cdots,\ \ell_n) \qquad (1 \le i \le t;\ 1 \le j \le n) \tag{2.45}$$

Similarly, one can linearize all equations in (2.42) so that (2.42) reduces to the linearized condition equations in matrix notations:

$$\underset{t\cdot n}{B} \cdot \underset{n\cdot 1}{\varepsilon} = \underset{t\cdot 1}{W} \tag{2.46}$$

Chapter 3

Adjustment by Elements

In a geodetic or photogrammetric network requiring m necessary observations, if we have observed n ($> m$) quantities there must exist t ($= n - m$) conditions among the observed quantities or the observation error ε (residuals). Condition adjustment adjusts the n observations in such a way that the estimated residuals satisfy these t conditions under least squares principle.

An alternative but equivalent method for adjusting the above n observations is the so called *adjustment by elements*. The basic idea of adjustment by elements is to express each observation as a function of m independent *unknown parameters* (corresponding to m necessary observations) and estimate these parameters as well as the residuals under least squares principle.

As an example, we look at the single point positioning network treated at the beginning of **Chapter 2** (see in Fig. 2.1 and **??**).

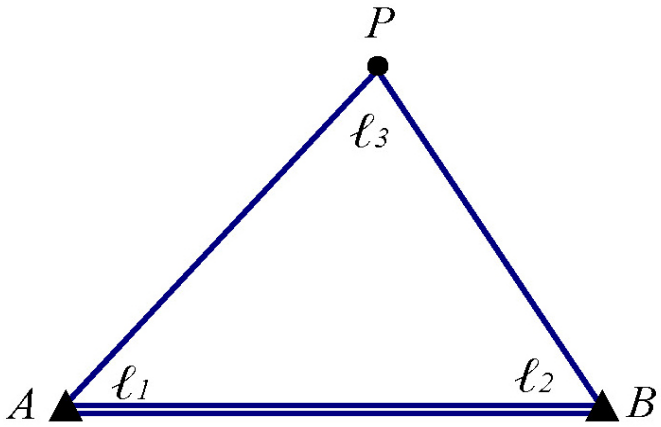

Figure 3.1: A simple triangle

If we choose the internal angles at points A and B as our unknown parameters (x_1, x_2), then all three observations can be written as linear functions of x_1 and x_2 :

$$\begin{cases} \ell_1 = x_1 \\ \ell_2 = x_2 \\ \ell_3 = 180^0 - x_1 - x_2 \end{cases}$$

or in matrix notation:

$$\begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 180^0 \end{bmatrix} \tag{3.1}$$

In adjustment by elements, (3.1) is called *observation equations*. For this simple problem, we have $n = 3$ equations and $m = 2$ unknowns. If all observations were free of errors, the above three equations would be consistent with each other, i.e., one could find values for the two unknowns which would satisfy all three equations at the same time. However, in practice ℓ_1, ℓ_2 and ℓ_3 contain always measurement errors and therefore, the three equations in (3.1) generally do not hold at the same time. In other words, (3.1) is an inconsistent equation system without solutions. To make it consistent, one need to add the true observation error ε to the left side of (3.1):

$$\begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{bmatrix} - \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 180^0 \end{bmatrix} \tag{3.2}$$

or in matrix notation:

$$\underset{3\cdot1}{L} - \underset{3\cdot1}{\varepsilon} = \underset{3\cdot2}{A}\ \underset{2\cdot1}{X} \tag{3.3}$$

where:

$$\underset{3\cdot1}{L} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 180^0 \end{bmatrix}, \quad \underset{3\cdot1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}, \quad \underset{3\cdot2}{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad \underset{2\cdot1}{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{3.4}$$

There are infinite number of solutions to ε and X in (3.2). The method of adjustment by elements chooses such solutions that

(i) Eq.(3.3) is satisfied and

(ii) $\sum_{i=1}^{3}(p_i\varepsilon_i\varepsilon_i) = \text{minimum}$

where p_i denotes the weight of observation ℓ_i $(i = 1, 2, 3)$.

In this chapter, we are going to give a general presentation of adjustment by elements. After describing the simple adjustment by elements in Section 3.1, several modified adjustment methods originating from the simple method will be described. Numerical examples are given after the theoretical derivations to demonstrate practical applications.

3.1 Adjustment by Elements in Linear Models

In this section, we first derive the basic formulas of adjustment by elements with linear observation equations. The treatment of non-linear observation equations will be discussed in Section 3.2.

3.1.1 Basic Formulas

Assume that our geodetic or photogrammetric problem under consideration requires m necessary observations and that we have made n observations ($n > m$). Corresponding to the m necessary observations, we have chosen m independent unknown parameters: $x_1, x_2, \cdots, x_m$ so that the true values $\widetilde{\ell}_i$ $(1 \le i \le n)$ of the observations ℓ_i can be written as linear functions of x_j $(1 \le j \le m)$:

$$\left.\begin{array}{l} \widetilde{\ell}_1 = \ell_1 - \varepsilon_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + c_1 \\ \widetilde{\ell}_2 = \ell_2 - \varepsilon_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + c_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \widetilde{\ell}_n = \ell_n - \varepsilon_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m + c_n \end{array}\right\} \tag{3.5}$$

where a_{ij}, c_i are known constants specific for the problem concerned, and ε_i is the residual of ℓ_i. (3.5) can be rewritten in matrix notation:

$$\underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} = \underset{n\cdot m}{A}\;\underset{m\cdot 1}{X} \tag{3.6}$$

where:

$$\underset{n\cdot 1}{L} = L' - c\,, \quad L' = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \cdots \\ \ell_n \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{bmatrix}, \quad \underset{n\cdot 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \tag{3.7}$$

$$\underset{n\cdot m}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad \underset{m\cdot 1}{X} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} \tag{3.8}$$

Note that in Chapter 2 vector L denoted the direct observations, while in this chapter L denotes the *reduced* observations which differ from the direct observations L' by a constant vector c.

We also assume that ε has the following *a priori* statistical properties, zero expectation:

$$E\left[\varepsilon\right] = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \cdots \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} = \underset{n\cdot 1}{0} \tag{3.9}$$

and a variance-covariance matrix as:

$$\underset{n\cdot n}{C_{\varepsilon\varepsilon}} = E\left\{[\varepsilon - E(\varepsilon)][\varepsilon - E(\varepsilon)]^\top\right\} = E(\varepsilon\varepsilon^\top) = \sigma_0^2\; P^{-1} \tag{3.10}$$

where σ_0^2 is the a priori variance factor (unit-weight standard error squared) and P the corresponding weight matrix. As L', the vector of the direct observations, differs from the "reduced" observation vector L only by a constant vector c, L' and L will have the same variance-covariance matrix as ε :

$$E(L) = E(AX + \varepsilon) = AX \tag{3.11}$$

$$C_{LL} = C_{L'L'} = C_{\varepsilon\varepsilon} = \sigma_0^2 P^{-1} \tag{3.12}$$

There are infinite number of solutions to X in (3.6) and the least squares solution $\widehat{X}$ is obtained under the conditions :

$$L - \widehat{\varepsilon} = A\widehat{X} \tag{3.13}$$

$$\widehat{\varepsilon}^{\top} P \widehat{\varepsilon} = \text{ minimum} \tag{3.14}$$

where $\widehat{\varepsilon}$ is the estimated residual corresponding to estimate $\widehat{X}$. If P is a diagonal matrix and the weight of ℓ_i is denoted by p_i $(1 \le i \le n)$, (3.14) is equivalent to:

$$\sum_{i=1}^{n} \left(p_i \, \widehat{\varepsilon}_i^2 \right) = \text{ minimum} \tag{3.15}$$

To derive the least squares estimate $\widehat{X}$ and $\widehat{\varepsilon}$, we form the Lagrange function F and let its derivative (with respect to X) be zero:

$$F(X) = \varepsilon^{\top} P \varepsilon = (L - AX)^{\top} P (L - AX)$$

$$\frac{\partial F}{\partial X} \mid_{x=\widehat{X},\, \varepsilon=\widehat{\varepsilon}} = 2\, \widehat{\varepsilon}^{\top} \, P(-A) = 0$$

or:

$$A^{\top} P \widehat{\varepsilon} = 0 \tag{3.16}$$

The least squares estimate $\widehat{\varepsilon}$ should also satisfy the observation equations in (3.6), just as the true error ε. Inserting (3.13) into (3.16) leads to the *normal equation* of adjustment by elements:

$$A^{\top} P A \widehat{X} = A^{\top} P L \tag{3.17}$$

from which $\widehat{X}$ can be solved directly :

$$\widehat{X} = (A^{\top} P A)^{-1} A^{\top} P L \tag{3.18}$$

Considering (3.12), the variance-covariance matrix of $\widehat{X}$ follows:

$$C_{\widehat{X}\widehat{X}} = (A^{\top} P A)^{-1} A^{\top} P \cdot \sigma_0^2 P^{-1} \cdot \left[(A^{\top} P A)^{-1} A^{\top} P \right]^{\top} = \sigma_0^2 \, (A^{\top} P A)^{-1} \tag{3.19}$$

After $\widehat{X}$ has been found, one can easily obtain the least squares estimate of ε :

$$\widehat{\varepsilon} = L - A\widehat{X} = \left[I - A(A^{\top} P A)^{-1} A^{\top} P \right] L \tag{3.20}$$

as well as its variance :

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = \left[I - A(A^{\top} P A)^{-1} A^{\top} P \right] \; \sigma_0^2 P^{-1} \; \left[I - A(A^{\top} P A)^{-1} A^{\top} P \right]^{\top} = \sigma_0^2 \left[P^{-1} - A(A^{\top} P A)^{-1} A^{\top} \right] \tag{3.21}$$

The adjusted value $(\widehat{L})$ of L and its variance-covariance matrix are as follows:

$$\widehat{L} = L - \widehat{\varepsilon} = A\widehat{X} = A(A^{\top} P A)^{-1} A^{\top} P L \tag{3.22}$$

$$C_{\widehat{L}\widehat{L}} = A(A^{\top} P A)^{-1} A^{\top} P \; \sigma_0^2 P^{-1} \; \left[A(A^{\top} P A)^{-1} A^{\top} P \right]^{\top} = \sigma_0^2 \, A(A^{\top} P A)^{-1} A^{\top} \tag{3.23}$$

From the first equation in (3.7), we can get the adjusted value of the direct observations L' and its variance-covariance matrix :

$$\widehat{L}' = \widehat{L} + c \tag{3.24}$$

$$C_{\widehat{L}'\widehat{L}'} = C_{\widehat{L}\widehat{L}} = \sigma_0^2 \, A(A^{\top} P A)^{-1} A^{\top} \tag{3.25}$$

Similarly as in condition adjustment, an *a posteriori variance factor* $\widehat{\sigma}_0{}^2$ can be estimated from the least squares residual $\widehat{\varepsilon}$:

$$\widehat{\sigma}_0^2 = \frac{\widehat{\varepsilon}^{\top} P \widehat{\varepsilon}}{n - m} \tag{3.26}$$

To prove that (3.26) is an unbiased estimate of σ_0^2, we first look at the expectation of $LL^\top$:

$$E\left[LL^\top\right] = E[(\varepsilon + AX)(\varepsilon + AX)^\top] =$$
$$E[\varepsilon\varepsilon^\top + \varepsilon(AX)^\top + AX\ \varepsilon^\top + (AX)(AX)^\top] = \sigma_0^2 P^{-1} + A\ XX^\top A^\top \tag{3.27}$$

From (3.20) and (3.26), we can get:

$$\begin{aligned}
E[\widehat{\varepsilon}^\top P\widehat{\varepsilon}] &= E\left\{L^\top[P - PA(A^\top PA)^{-1}A^\top P]L\right\} \\
&= tr\left[E\left\{L^\top[P - PA(A^\top PA)^{-1}A^\top P]L\right\}\right] \\
&= tr\left[\left\{P - PA(A^\top PA)^{-1}A^\top P\right\}E(LL^\top)\right] \\
&= tr\left\{\left[P - PA(A^\top PA)^{-1}A^\top P\right]\left[\sigma_0^2 P^{-1} + AXX^\top A^\top\right]\right\} \\
&= \sigma_0^2\left\{tr[\underset{n\cdot n}{I}] - tr[PA(A^\top PA)^{-1}A^\top]\right\} \\
&= \sigma_0^2\left\{tr[\underset{n\cdot n}{I}] - tr(\underset{m\cdot m}{I})\right\} = \sigma_0^2(n-m)
\end{aligned}$$

which implies:

$$E\left[\widehat{\sigma}_0^2\right] = E\left[\frac{\widehat{\varepsilon}^\top P\widehat{\varepsilon}}{n-m}\right] = \sigma_0^2 \tag{3.28}$$

That is to say, $\widehat{\sigma}_0^2$ is an unbiased estimate of σ_0^2 . One can further prove that $\widehat{\sigma}_0^2$ is an BQUE of σ_0^2 (Cf Chapter 5).

Using the estimated variance factor $\widehat{\sigma}_0^2$, (3.19), (3.21) and (3.23) become :

$$C_{\widehat{X}\widehat{X}} = \widehat{\sigma}_0^2\ (A^\top PA)^{-1} \tag{3.29}$$

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = \widehat{\sigma}_0^2\left[P^{-1} - A(A^\top PA)^{-1}A^\top\right] \tag{3.30}$$
$$C_{\widehat{L}\widehat{L}} = \widehat{\sigma}_0^2\ A(A^\top PA)^{-1}A^\top \tag{3.31}$$

If we make use of the relation between variance matrix (C) and the cofactor matrix (Q), the last three equations can be written as:

$$C_{\widehat{X}\widehat{X}} = \widehat{\sigma}_0^2\ Q_{\widehat{X}\widehat{X}}\ , \quad C_{\widehat{\varepsilon}\widehat{\varepsilon}} = \widehat{\sigma}_0^2\ Q_{\widehat{\varepsilon}\widehat{\varepsilon}}\ , \quad C_{\widehat{L}\widehat{L}} = \widehat{\sigma}_0^2\ Q_{\widehat{L}\widehat{L}} \tag{3.32}$$

where the cofactor matrices are:

$$Q_{\widehat{X}\widehat{X}} = (A^\top PA)^{-1} \tag{3.33}$$
$$Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = P^{-1} - A(A^\top PA)^{-1}A^\top \tag{3.34}$$
$$Q_{\widehat{L}\widehat{L}} = A(A^\top PA)^{-1}A^\top = P^{-1} - Q_{\widehat{\varepsilon}\widehat{\varepsilon}} \tag{3.35}$$

Finally, let us consider a random vector of linear functions of X :

$$y = G\ X \tag{3.36}$$

where G denotes a known matrix. The least squares estimate of y can be obtained as:

$$\widehat{y} = G\ \widehat{X} \tag{3.37}$$

Using error propagation law (1.3.8), the variance-covariance matrix and cofactor matrix of y can be obtained:

$$C_{yy} = GC_{\widehat{X}\widehat{X}}G^\top = \widehat{\sigma}_0^2\ G(A^\top PA)^{-1}G^\top = \widehat{\sigma}_0^2\ Q_{yy} \tag{3.38}$$

$$Q_{yy} = GQ_{\widehat{X}\widehat{X}}G^\top = G(A^\top PA)^{-1}G^\top \tag{3.39}$$

So far, we have presented the basic formulas of adjustment by elements. Below, we give two examples to demonstrate how the theoretical formulas can be applied for practical problems.

Example 3.1

We use the method of adjustment by elements to adjust the three measured angles in **Example 2.1**, as illustrated in Fig. 2.1. The observations and their weights are as follows:

$$\underset{3\cdot1}{L'} = \begin{bmatrix} 60^0\ 00'\ 03'' \\ 60^0\ 00'\ 03'' \\ 59^0\ 59'\ 51'' \end{bmatrix}$$

$$\underset{3\cdot3}{P} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = \underset{3\cdot3}{I}$$

The number of necessary observations is $m = 2$ and we choose the angles at points A and B as unknown parameters x_1 and x_2, respectively. The observation equations of three angles are :

$$\underset{3\cdot1}{L} - \underset{3\cdot1}{\varepsilon} = \underset{3\cdot2}{A}\ \underset{2\cdot1}{X}$$

where:

$$\underset{3\cdot1}{L} = L' - c = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 180^0 \end{bmatrix} = \begin{bmatrix} 60^0\ 00'\ 03'' \\ 60^0\ 00'\ 03'' \\ -120^0\ 00'\ 09'' \end{bmatrix}$$

$$\underset{3\cdot1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

$$\underset{3\cdot2}{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad \underset{2\cdot1}{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The least squares estimates $\widehat{X},\ \widehat{L},\ \widehat{L}'$ can be calculated using (3.18), (3.22) and (3.24), respectively:

$$A^\top PA = A^\top A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad (A^\top PA)^{-1} = \tfrac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\widehat{X} = (A^\top A)^{-1}A^\top L = \tfrac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} L$$

$$= \tfrac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \end{bmatrix}\begin{bmatrix} 60^0\ 00'\ 03'' \\ 60^0\ 00'\ 03'' \\ -120^0\ 00'\ 09'' \end{bmatrix} = \begin{bmatrix} 60^0\ 00'\ 04'' \\ 60^0\ 00'\ 04'' \end{bmatrix}$$

$$\widehat{L} = A\widehat{X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}\begin{bmatrix} 60^0\ 00'\ 04'' \\ 60^0\ 00'\ 04'' \end{bmatrix} = \begin{bmatrix} 60^0\ 00'\ 04'' \\ 60^0\ 00'\ 04'' \\ -120^0\ 00'\ 08'' \end{bmatrix}$$

$$\widehat{L}' = \widehat{L} + c = \begin{bmatrix} 60^0\ 00'\ 04'' \\ 60^0\ 00'\ 04'' \\ -120^0\ 00'\ 08'' \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 180^0 \end{bmatrix} = \begin{bmatrix} 60^0\ 00'\ 04'' \\ 60^0\ 00'\ 04'' \\ 59^0\ 59'\ 52'' \end{bmatrix}$$

The above result for $\widehat{L}'$ is identical to the result obtained in **Example 2.1**, as it should be. In general, *starting from same observation data and same weight matrix, condition adjustment and adjustment by elements should give identical adjustment results.* ∎

Example 3.2

In this example, we use the method of adjustment by elements to adjust the levelling network in **Example 2.2** as illustrated by Fig. 2.2. We assume that benchmark P_4 is fixed with given height $H_4 = 10.000$ metres. From **Example 2.2**, we have the direct observations and their weight matrix:

$$\underset{5\cdot 1}{L'} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \\ \ell_5 \end{bmatrix} = \begin{bmatrix} +1.002 \\ +2.004 \\ -2.001 \\ -1.002 \\ +3.012 \end{bmatrix} (m), \quad \underset{5\cdot 5}{P} = \begin{bmatrix} 2 & & & & \\ & 2 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 2 \end{bmatrix}$$

Choosing the heights of benchmarks P_1, P_2 and P_3 as our unknown parameters x_1, x_2 and x_3, respectively, the observation equations become :

$$\begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \\ \ell_5 \end{bmatrix} - \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ H_4 \\ -H_4 \\ 0 \end{bmatrix} \quad \Leftrightarrow \quad L' - \varepsilon = AX + c$$

or:

$$L - \varepsilon = AX$$

with :

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad L = L' - c = \begin{bmatrix} +1.002 \\ +2.004 \\ -12.001 \\ +8.998 \\ +3.012 \end{bmatrix} (m)$$

Then we have:

$$A^\top P = \begin{bmatrix} -1 & 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & & & & \\ & 2 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 2 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 & 1 & -2 \\ 2 & -2 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 & 2 \end{bmatrix}$$

$$A^\top PA = \begin{bmatrix} -2 & 0 & 0 & 1 & -2 \\ 2 & -2 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 5 \end{bmatrix}$$

$$(A^\top PA)^{-1} = \tfrac{1}{28} \begin{bmatrix} 16 & 14 & 12 \\ 14 & 21 & 14 \\ 12 & 14 & 16 \end{bmatrix}$$

$$(A^\top PA)^{-1} A^\top P = \tfrac{1}{14} \begin{bmatrix} -2 & -2 & -6 & 8 & -4 \\ 7 & -7 & -7 & 7 & 0 \\ 2 & 2 & -8 & 6 & 4 \end{bmatrix}$$

The least squares estimate $\widehat{X}$ is given by Eq.(3.18):

$$\widehat{X} = (A^\top PA)^{-1}A^\top PL = \tfrac{1}{14}\begin{bmatrix} -2 & -2 & -6 & 8 & -4 \\ 7 & -7 & -7 & 7 & 0 \\ 2 & 2 & -8 & 6 & 4 \end{bmatrix}\begin{bmatrix} +1.002 \\ +2.004 \\ -12.001 \\ +8.998 \\ +3.012 \end{bmatrix} = \begin{bmatrix} 8.9950 \\ 9.9985 \\ 12.0040 \end{bmatrix} \ (m)$$

The adjusted values of the reduced observations L and the direct observations L' follow from (3.22) and Eq.(3.24), respectively :

$$\widehat{L} = A\widehat{X} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 8.9950 \\ 9.9985 \\ 12.0040 \end{bmatrix} = \begin{bmatrix} 1.0035 \\ 2.0055 \\ -12.0040 \\ 8.9950 \\ 3.0090 \end{bmatrix} \ (m)$$

$$\widehat{L}' = \widehat{L} + c = \begin{bmatrix} 1.0035 \\ 2.0055 \\ -12.0040 \\ 8.9950 \\ 3.0090 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 10.0000 \\ -10.000 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.0035 \\ 2.0055 \\ -2.0040 \\ -1.0050 \\ 3.0090 \end{bmatrix} \ (m)$$

The estimated residual is obtained by (3.20):

$$\widehat{\varepsilon} = L - A\widehat{X} = L - \widehat{L} = \begin{bmatrix} +1.002 \\ +2.004 \\ -12.001 \\ +8.998 \\ +3.012 \end{bmatrix} - \begin{bmatrix} 1.0035 \\ 2.0055 \\ -12.0040 \\ 8.9950 \\ 3.0090 \end{bmatrix} = \begin{bmatrix} -1.5 \\ -1.5 \\ +3.0 \\ +3.0 \\ +3.0 \end{bmatrix} \ (mm)$$

The *a posteriori* variance factor $\widehat{\sigma}_0^2$ is estimated by Eq.(3.26):

$$\widehat{\sigma}_0^2 = \tfrac{\widehat{\varepsilon}^\top P\widehat{\varepsilon}}{n-m} = \tfrac{1}{5-3}\sum_1^5 (p_i\widehat{\varepsilon}_i\widehat{\varepsilon}_i) = 22.50 \ mm^2$$

$$\widehat{\sigma}_0 = \sqrt{22.50} \approx \pm 4.74 \ mm$$

The cofactor matrices of $\widehat{X}$, $\widehat{L}$ and $\widehat{\varepsilon}$ are given by Eqs.(3.33), (3.34) and (3.35), respectively:

$$Q_{\widehat{X}\widehat{X}} = (A^\top PA)^{-1} = \tfrac{1}{28}\begin{bmatrix} 16 & 14 & 12 \\ 14 & 21 & 14 \\ 12 & 14 & 16 \end{bmatrix}$$

$$Q_{\widehat{L}\widehat{L}} = A(A^\top PA)^{-1}A^\top = \tfrac{1}{28}\begin{bmatrix} 9 & -5 & -2 & -2 & 4 \\ -5 & 9 & -2 & -2 & 4 \\ -2 & -2 & 16 & -12 & -4 \\ -2 & -2 & -12 & 16 & -4 \\ 4 & 4 & -4 & -4 & 8 \end{bmatrix}$$

$$Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = P^{-1} - Q_{\widehat{L}\widehat{L}} = \tfrac{1}{28}\begin{bmatrix} 5 & 5 & 2 & 2 & -4 \\ 5 & 5 & 2 & 2 & -4 \\ 2 & 2 & 12 & 12 & 4 \\ 2 & 2 & 12 & 12 & 4 \\ -4 & -4 & 4 & 4 & 6 \end{bmatrix}$$

The variance-covariance matrices of the above three estimates are:

$$C_{\widehat{X}\widehat{X}} = \widehat{\sigma}_0^2\, Q_{\widehat{X}\widehat{X}} = \tfrac{22.50}{28}\begin{bmatrix} 16 & 14 & 12 \\ 14 & 21 & 14 \\ 12 & 14 & 16 \end{bmatrix}$$

$$C_{\widehat{L}\widehat{L}} = \widehat{\sigma}_0^2\, Q_{\widehat{L}\widehat{L}} = \tfrac{22.50}{28}\begin{bmatrix} 9 & -5 & -2 & -2 & 4 \\ -5 & 9 & -2 & -2 & 4 \\ -2 & -2 & 16 & -12 & -4 \\ -2 & -2 & -12 & 16 & -4 \\ 4 & 4 & -4 & -4 & 8 \end{bmatrix}$$

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = \widehat{\sigma}_0^2\, Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = \frac{22.50}{28} \begin{bmatrix} 5 & 5 & 2 & 2 & -4 \\ 5 & 5 & 2 & 2 & -4 \\ 2 & 2 & 12 & 12 & 4 \\ 2 & 2 & 12 & 12 & 4 \\ -4 & -4 & 4 & 4 & 6 \end{bmatrix}$$

From $C_{\widehat{X}\widehat{X}}$ calculated above, one can further calculate the standard errors of the estimated heights of three benchmarks:

$$\sigma_{\widehat{x}_1} = \widehat{\sigma}_0\sqrt{16/28} \approx \pm 3.59 \;\; mm$$
$$\sigma_{\widehat{x}_2} = \widehat{\sigma}_0\sqrt{21/28} \approx \pm 4.11 \;\; mm$$
$$\sigma_{\widehat{x}_3} = \widehat{\sigma}_0\sqrt{16/28} \approx \pm 3.59 \;\; mm$$

For height difference between P_2 and P_4, we can obtain the least squares estimate:

$$\widehat{h}_{24} = H_4 - \widehat{X}_2 = 10.000 - 9.9985 = 1.5 \; mm$$

and its standard error:

$$\sigma_{\widehat{h}_{24}} = \sigma_{\widehat{x}_2} = \pm 4.11 \; mm. \quad \blacksquare$$

3.1.2 Special Applications: Direct Adjustment and Linear Regression

In this subsection, we are going to look at two special problems using the method of adjustment by elements. The first problem concerns adjustment of n direct measurements which has already been dealt with in **Example 1.12** of Section 1.5, while the second problem is the problem of linear regression with one independent variable, also described in Subsection 1.4.5.

Adjustment of Direct Measurements

Assume that a quantity x has been directly measured n times with results:

$$\underset{n\cdot 1}{L} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \cdots \\ \ell_n \end{bmatrix} \tag{3.40}$$

All observations are assumed to be independent of each other and the weight of the observation ℓ_i $(1 \le i \le n)$ is denoted by p_i. The error vector ε of the observation vector L is assumed to have zero expectation and the following variance-covariance matrix:

$$C_{\varepsilon\varepsilon} = \sigma_0^2 \cdot P^{-1} \tag{3.41}$$

where P denotes the diagonal weight matrix of L or ε :

$$\underset{n\cdot n}{P} = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \cdots & \\ & & & p_n \end{bmatrix} \tag{3.42}$$

Choosing the quantity x as our only unknown parameter to be estimated, we have the following observation equation:

$$\underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} = \underset{n\cdot m}{A} \cdot \underset{m\cdot 1}{X} \quad (m = 1) \tag{3.43}$$

where:

$$\underset{n\cdot 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{n\cdot m}{A} = \underset{n\cdot 1}{A} = \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix}, \quad \underset{m\cdot 1}{X} = \underset{1\cdot 1}{X} = [\, x \,] \tag{3.44}$$

For the above A, L, one can calculate:

$$A^\top PA = \sum_{i=1}^{n} p_i, \quad A^\top PL = \sum_{i=1}^{n} p_i \ell_i$$

The least squares estimate of the unknown parameter can be obtained from Eq.(3.18) :

$$\widehat{X} = \widehat{x} = \left(A^\top PA\right)^{-1} A^\top PL = \frac{\sum_{i=1}^{n} p_i \ell_i}{\sum_{i=1}^{n} p_i} = \frac{p_1\ell_1 + p_2\ell_2 + \cdots + p_n\ell_n}{p_1 + p_2 + \cdots + p_n} \tag{3.45}$$

which is identical to the weighted mean $\widehat{x}$ given in (1.209) of **Example 1.12**, where it has been directly proved that $\widehat{x}$ is a least squares estimate. The least squares residuals $\widehat{\varepsilon}$ then follows:

$$\widehat{\varepsilon} = L - AX = L - A \cdot \widehat{x} = \begin{bmatrix} \ell_1 - \widehat{x} \\ \ell_2 - \widehat{x} \\ \cdots \\ \ell_n - \widehat{x} \end{bmatrix} \tag{3.46}$$

and according to Eq.(3.26), the *a posteriori* estimate of the variance factor σ_0^2 becomes now :

$$\widehat{\sigma}_0^2 = \frac{\widehat{\varepsilon}^\top P \widehat{\varepsilon}}{n - m} = \frac{1}{n-1} \sum_{i=1}^{n} p_i \widehat{\varepsilon}_i^2 = \frac{1}{n-1} \sum_{i=1}^{n} p_i \left(\ell_i - \widehat{x}\right)^2 \tag{3.47}$$

The variance of the estimated $\widehat{x}$ is given by (3.29) :

$$\sigma_{\widehat{x}}^2 = \widehat{\sigma}_0^2 \cdot \left(A^\top P A\right)^{-1} = \widehat{\sigma}_0^2 \cdot \frac{1}{\sum_{i=1}^n p_i} \tag{3.48}$$

which implies that for the variance factor $\widehat{\sigma}_0^2$, the weight of $\widehat{x}$ is:

$$p_x = \sum_{i=1}^n p_i \tag{3.49}$$

When all direct measurements have equal weights $p_i = 1$, $\widehat{x}$ becomes the simple mean:

$$\widehat{x} = \frac{\sum_{i=1}^n \ell_i}{n} = \frac{\ell_1 + \ell_2 + \cdots + \ell_n}{n} \tag{3.50}$$

Eqs. (3.47), (3.48) and (3.49) now reduce to :

$$\widehat{\sigma}_0^2 = \frac{1}{n-1} \sum (\ell_i - \widehat{x})^2, \quad \sigma_{\widehat{x}}^2 = \widehat{\sigma}_0^2 \cdot \frac{1}{n}, \quad p_x = n \tag{3.51}$$

Linear Regression

Assume that we have a set of n pairs of measurements (x_i, y_i) and that we want to estimate the two regression coefficients α, β in the following linear model:

$$y_i = \alpha + \beta \cdot x_i + \varepsilon_i \qquad (1 \le i \le n) \tag{3.52}$$

where ε_i denotes the measurement error (residual) of y_i. The above regression model can be rewritten as observation equations in matrix notation :

$$\underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} = \underset{n\cdot m}{A} \cdot \underset{m\cdot 1}{X} \qquad (m = 2) \tag{3.53}$$

where :

$$\underset{n\cdot 1}{L} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \quad \underset{n\cdot 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{n\cdot m}{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix}, \quad \underset{m\cdot 1}{X} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{3.54}$$

Let ε_i $(1 \le i \le n)$ be ununcorrelated with each other and have zero expectation and variance σ_i^2. For a chosen variance factor σ_0^2, one can obtain the weight matrix for vector ε (or L):

$$\underset{n\cdot n}{P} = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \cdots & \\ & & & p_n \end{bmatrix}, \qquad \text{with } \; p_i = \frac{\sigma_0^2}{\sigma_i^2} \tag{3.55}$$

If all ε_i $(1 \le i \le n)$ have equal variances $\sigma_i^2 = \sigma_0^2$, then we have a unit weight matrix $P = I$ and the adjustment problem using observation equations (3.53) is exactly identical to the linear regression problem with Eqs.(1.182) and (1.183).

Using the general weight matrix P as defined in Eq.(3.55) and the least squares principle:

$$\varepsilon^\top P \varepsilon = \sum_{i=1}^n p_i \varepsilon_i^2 = \sum_{i=1}^n \left\{ p_i \left(y_i - \alpha - \beta x_i\right)^2 \right\} = \min$$

the linear regression by (3.52) can be performed as adjustment by elements using observation equations (3.53). The coefficient matrix and the constant vector of the normal equation can be calculated :

$$A^\top PA = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \cdots & \\ & & & p_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (p_i) & \sum_{i=1}^n (p_i x_i) \\ \sum_{i=1}^n (p_i x_i) & \sum_{i=1}^n \left(p_i x_i^2\right) \end{bmatrix}$$

$$A^\top PL = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \cdots & \\ & & & p_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (p_i y_i) \\ \sum_{i=1}^n (p_i x_i y_i) \end{bmatrix}$$

From Eq.(3.18), we obtain the least squares estimates of X :

$$\widehat{X} = \begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = (A^\top PA)^{-1} A^\top PL$$

$$= \frac{1}{n\sum_{i=1}^n (p_i x_i^2) - \left[\sum_{i=1}^n (p_i x_i)\right]^2} \begin{bmatrix} \sum_{i=1}^n \left(p_i x_i^2\right) & -\sum_{i=1}^n (p_i x_i) \\ -\sum_{i=1}^n (p_i x_i) & \sum_{i=1}^n (p_i) \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n (p_i y_i) \\ \sum_{i=1}^n (p_i x_i y_i) \end{bmatrix} \tag{3.56}$$

The least squares residuals $\widehat{\varepsilon}$ is:

$$\widehat{\varepsilon} = L - A\widehat{X} = \begin{bmatrix} y_1 - \widehat{\alpha} - \widehat{\beta} \cdot x_1 \\ y_2 - \widehat{\alpha} - \widehat{\beta} \cdot x_2 \\ \cdots\cdots\cdots\cdots \\ y_n - \widehat{\alpha} - \widehat{\beta} \cdot x_n \end{bmatrix} \tag{3.57}$$

The *a posteriori* estimate of the variance factor σ^2 follows from Eq.(3.26):

$$\widehat{\sigma}^2 = \frac{\widehat{\varepsilon}^\top P \widehat{\varepsilon}}{n-m} = \frac{1}{n-2} \sum_{i=1}^n p_i \widehat{\varepsilon}_i^2 = \frac{1}{n-2} \sum_{i=1}^n \left\{ p_i \left(y_i - \widehat{\alpha} - \widehat{\beta} \cdot x_i \right)^2 \right\} \tag{3.58}$$

The variance-covariance matrix of the estimated regression coefficients $\widehat{\alpha}$ and $\widehat{\beta}$ is :

$$C_{\widehat{X}\widehat{X}} = \begin{bmatrix} \sigma_{\widehat{\alpha}}^2 & \sigma_{\widehat{\alpha}\widehat{\beta}} \\ \sigma_{\widehat{\alpha}\widehat{\beta}} & \sigma_{\widehat{\beta}}^2 \end{bmatrix} = \widehat{\sigma}_0^2 \cdot \left(A^\top PA\right)^{-1} =$$

$$= \frac{\widehat{\sigma}_0^2}{n\sum_{i=1}^n (p_i x_i^2) - \left[\sum_{i=1}^n (p_i x_i)\right]^2} \begin{bmatrix} \sum_{i=1}^n \left(p_i x_i^2\right) & -\sum_{i=1}^n (p_i x_i) \\ -\sum_{i=1}^n (p_i x_i) & \sum_{i=1}^n (p_i) \end{bmatrix} \tag{3.59}$$

The above adjustment results in Eqs. (3.56), (3.57), (3.58) and (3.59) are exactly identical as the regression results for $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\varepsilon}$, $\widehat{\sigma}^2$, $\sigma_{\widehat{\alpha}}^2$ and $\sigma_{\widehat{\beta}}^2$, given in Eqs.(1.186), (1.187), (1.188) and (1.189), respectively. However, there is one conceptual difference in treating Eq.(3.52) as linear regression model or as observation equation. In linear regression, the measurement pairs $(x_i,\ y_i,\ 1 \le i \le n)$ are assumed to be *statistically* correlated and there is no *functional* relation between the two variables, even though we try to estimate some ″functional relation″ that best fits the measurement data.

On the other hand, adjustment by elements assumes that the functional relations (between observations and unknown parameters) exist but are contaminated by the measurement errors ε. The question for adjustment by elements is how to obtain the best estimates based on the existing functional model, the measured values and the statistical characters of ε.

3.2 Observation Equations

This section treats the establishment of observation equations in more details. We first discuss the selection of unknown parameters and datum problem in adjustment by elements and then describe linearization of non-linear observation equations. Finally, the observation equations for some common types of geodetic observations (height differences, distances, angles, directions) will be presented.

3.2.1 Selection of Unknown Parameters and Datum Problem

Generally speaking, the number of unknown parameters (m) should be equal to the number of necessary observations. These m parameters can be m independent direct observations or some other quantities not directly observed. In **Example 3.1**, we chose two angles which are directly observed as the parameters, while in **Example 3.2**, the heights of unknown benchmarks were chosen as unknown parameters. As shown in Subsection 3.2.3, it is convenient to choose the coordinates (1-D, 2-D, or 3-D) of unknown points as parameters for geodetic networks.

The chosen parameters should also be independent, i.e. no parameter can be expressed as a function of other parameters. In a geodetic network with m necessary observations, if we have chosen more than m quantities as unknown parameters, then these m parameters will not be independent and consequently there must exist functional relations among them. If this happens, the general method and basic formulas presented in Section 3.1 will no longer apply and one should use the method of adjustment by unknowns with constraints to be described in the next section.

For networks with datum defect (i.e. the initial data is not available or insufficient), if we still choose the coordinates of all unknown points as unknown parameters, we will have troubles in the adjustment due to the datum defect problem. Mathematically, the datum defect problem will result in a singular normal equation matrix. The number of rank defect of this matrix is equal to the number of datum defect of the network.

In this case, least squares adjustment is still possible with the help of generalized matrix inverses (see Chapter 4). However, the unknown parameters (absolute coordinates) cannot be estimated meaningfully, as the coordinate estimates will solely depend on the choice of generalized matrix inverse. From geodetic point of view, each choice of the generalized inverses will *implicitly* but *unavoidably* assume a specific coordinate reference datum. Further discussions can be found in Chapter 4.

It should be reminded that even with datum defect, condition adjustment is almost always possible, as condition adjustment only adjusts the relationship among the observations (describing the internal geometry of the networks) without involving the absolute datum of the network.

3.2.2 Linearization of Non-linear Observation Equations

Assume that we have n observations, ℓ_1 , ℓ_2, $\cdots$, ℓ_n, and that the true value and true error of ℓ_i are $\widetilde{\ell}_i$ and ε_i such that $\widetilde{\ell}_i = \ell_i - \varepsilon_i$ $(1 \le i \le n)$. Assume also that each $\widetilde{\ell}_i$ is a non-linear functions of m unknown parameters x_j $(1 \le j \le m)$:

$$\left.\begin{aligned}
\widetilde{\ell}_1 &= f_1(x_1, x_2, \cdots, x_m) \\
\widetilde{\ell}_2 &= f_2(x_1, x_2, \cdots, x_m) \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\widetilde{\ell}_i &= f_i(x_1, x_2, \cdots, x_m) \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\widetilde{\ell}_n &= f_n(x_1, x_2, \cdots, x_m)
\end{aligned}\right\} \tag{3.60}$$

Let x_j^0 and δx_j denote an approximate value of x_j and its corresponding correction such that $x_j = x_j^0 + \delta x_j$ $(1 \le j \le m)$:

$$\underset{m\cdot 1}{X} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} = \underset{m\cdot 1}{X}^0 + \underset{m\cdot 1}{\delta X}\ , \quad \underset{m\cdot 1}{X}^0 = \begin{bmatrix} x_1^0 \\ x_2^0 \\ \cdots \\ x_m^0 \end{bmatrix}, \quad \underset{m\cdot 1}{\delta X} = \begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \cdots \\ \delta x_m \end{bmatrix} \tag{3.61}$$

The $i-th$ function $f_i(x_1,\ x_2,\ \cdots,\ x_m)$ in (3.60) can be expanded into a Taylor series around the approximate parameters x_j^0 $(1 \le j \le m)$:

$$\widetilde{\ell}_i = \ell_i - \varepsilon_i = f(x_1^0 + \delta x_1, x_2^0 + \delta x_2, \cdots, x_m^0 + \delta x_m)$$

$$= f(x_1^0,\ x_2^0,\ \cdots,\ x_m^0) + \tfrac{1}{1!}\left[\tfrac{\partial f}{\partial x_1}\delta x_1 + \tfrac{\partial f}{\partial x_2}\delta x_2 + \cdots + \tfrac{\partial f}{\partial x_m}\delta x_m\right] +$$

$$+ \tfrac{1}{2!}\left[\tfrac{\partial^2 f}{\partial x_1^2}\delta x_1^2 + \tfrac{\partial^2 f}{\partial x_1 \partial x_2}\delta x_1 \delta x_2 + \cdots + \tfrac{\partial^2 f}{\partial x_1 \partial x_m}\delta x_1 \delta x_m\right] + \ \cdots$$

Neglecting all terms of δx_j of order 2 or higher, the above equation becomes:

$$\ell_i - \varepsilon_i = a_{i1}\delta x_1 + a_{i2}\delta x_2 + \cdots + a_{im}\delta x_m + c_i \tag{3.62}$$

where:

$$a_{ij} = \frac{\partial f_i}{\partial x_j}\ , \quad c_i = f_i(x_1^0,\ x_2^0,\ \cdots,\ x_m^0)\ , \quad i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, m \tag{3.63}$$

Eq.(3.62) is the linearized observation equation of ℓ_i. Similarly, one can linearize all other non-linear equations so that the non-linear observation equation system in (3.60) reduces to a linear system:

$$\underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} = \underset{n\cdot m}{A}\ \underset{m\cdot 1}{\delta X} \tag{3.64}$$

where:

$$\underset{n\cdot 1}{L} = L' - c, \quad L' = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \cdots \\ \ell_n \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{bmatrix}, \quad c_i = f_i(x_1^0,\ x_2^0,\ \cdots,\ x_m^0) \tag{3.65}$$

$$\underset{n\cdot m}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad a_{ij} = \frac{\partial f_i}{\partial x_j} \quad (1 \le i \le n;\ 1 \le j \le m) \tag{3.66}$$

The least squares solution of δX can then be obtained from (3.18):

$$\delta\widehat{X} = (A^\top P A)^{-1} A^\top P\ L \tag{3.67}$$

where P is the weight matrix of L' (or L or ε). The least squares estimate of the unknown parameters X then follows:

$$\widehat{X} = X^0 + \delta\widehat{X} \tag{3.68}$$

Since $\widehat{X}$ and $\delta\widehat{X}$ differ from each other only by a non-stochastic vector X^0, $\widehat{X}$ and $\delta\widehat{X}$ should have the same variance-covariance matrix [see Eq.(3.19)]:

$$C_{\widehat{X}\widehat{X}} = C_{\delta\widehat{X}\delta\widehat{X}} = \sigma_0^2\ (A^\top P A)^{-1} \tag{3.69}$$

We can obtain the least squares estimates of ε, L' and L from Eqs.(3.20), (3.22) and (3.24), respectively:

$$\widehat{\varepsilon} = L - A\delta\widehat{X} = \left[I - A(A^\top P A)^{-1} A^\top P\right]\ L \tag{3.70}$$

$$\widehat{L} = L - \widehat{\varepsilon} = A\ \delta\widehat{X} = A(A^\top P A)^{-1} A^\top P L \tag{3.71}$$

$$\widehat{L}' = \ \widehat{L} + c \tag{3.72}$$

and their variance-covariance matrices from Eqs.(3.21) and (3.25):

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = \sigma_0^2 \left[P^{-1} - A(A^\top PA)^{-1}A^\top\right] \tag{3.73}$$

$$C_{\widehat{L}\widehat{L}} = C_{\widehat{L}'\widehat{L}'} = \sigma_0^2\ A(A^\top PA)^{-1}A^\top \tag{3.74}$$

Finally the a posteriori estimate of the variance factor σ_0^2 based on the least squares residual $\widehat{\varepsilon}$ is calculated by [see (3.26]:

$$\widehat{\sigma}_0^2 = \frac{\widehat{\varepsilon}^\top P\widehat{\varepsilon}}{n-m} \tag{3.75}$$

The above derivations show that the key to adjustment with non-linear observation equations is the linearization of the non-linear observation equations. After linearization, the adjustment follows the same procedures and formulas as in the case of linear observation equations discussed in Section 3.1. The linearization of several types of traditional geodetic observations will be given in the next subsection.

3.2.3 Observation Equations of Traditional Geodetic Measurements

In traditional geodetic networks, the most important quantities are the coordinates of unknown points. Therefore, the coordinates of unknown points are normally selected as unknown parameters in adjustment by elements. These coordinates can be 1-dimensional (*e.g.* height), 2-dimensional (*e.g.* horizontal coordinates) or 3-dimensional ($3-D$ coordinates x, y, z).

Geodetic Levelling

In Fig. 3.1, ℓ_{ij} denotes the levelled height difference from benchmark P_i to benchmark P_j :

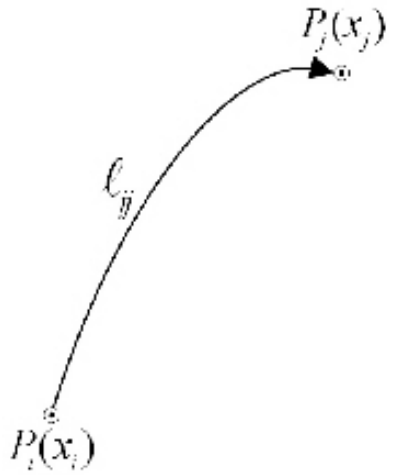

Figure 3.2: Observation equation of a levelled height difference

Let ε_{ij} denote the corresponding error (residual) of ℓ_{ij} and x_i, x_j denote the height of P_1 and P_2, respectively. Then the observation equation of ℓ_{ij} is :

$$\ell_{ij} - \varepsilon_{ij} = x_j\ - x_i \tag{3.76}$$

If benchmark P_i (or P_j) is fixed with known height, then its height will not be regarded as unknown parameter. Accordingly, x_i (or x_j) in (3.76) should be replaced by its known height.

Distance Measurements

Let ℓ_{ij} denote an observed distance between unknown point P_i and unknown point P_j and let ε_{ij} denote its residual (see Fig. 3.2).

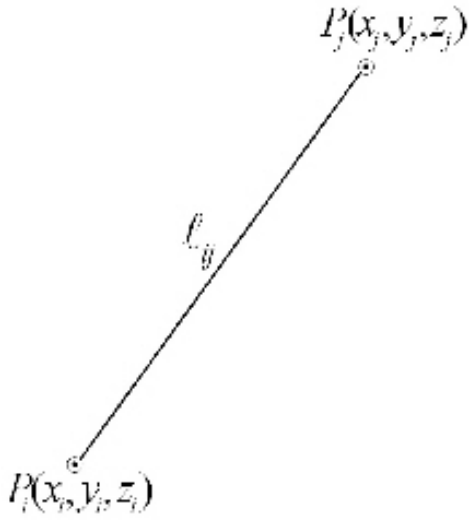

Figure 3.3: Observation equation of a 3D-distance measurement

If we choose the 3-D Cartesian coordinates of unknown points as parameters, the observation equation of ℓ_{ij} will be:

$$\ell_{ij} - \varepsilon_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2} \tag{3.77}$$

where x_i, y_i, z_i and x_j, y_j, z_j stand for the 3-D Cartesian coordinates of points P_i and P_j , respectively.

Denoting the approximate coordinates by x_i^0, y_i^0, z_i^0 and x_j^0, y_j^0, z_j^0 and their corrections by δx_i, $\delta y_i, \delta z_i$ and δx_j , δy_j, δz_j , respectively, the non-linear equation (3.77) can be linearized by (3.62):

$$(\ell_{ij} - s_{ij}^0) - \varepsilon_{ij} = a \cdot \delta x_i + b \cdot \delta y_i + c \cdot \delta z_i + d \cdot \delta x_j \ + e \cdot \delta y_j + f \cdot \delta z_j \tag{3.78}$$

with:

$$\left.\begin{aligned} s_{ij}^0 &= \sqrt{(x_j^0 - x_i^0)^2 + (y_j^0 - y_i^0)^2 + (z_j^0 - z_i^0)^2} \\ a &= \frac{\partial \ell_{ij}}{\partial x_i} = -\frac{x_j^0 - x_i^0}{s_{ij}^0} \\ b &= \frac{\partial \ell_{ij}}{\partial y_i} = -\frac{y_j^0 - y_i^0}{s_{ij}^0} \\ c &= \frac{\partial \ell_{ij}}{\partial z_i} = -\frac{z_j^0 - z_i^0}{s_{ij}^0} \\ d &= \frac{\partial \ell_{ij}}{\partial x_j} = \frac{x_j^0 - x_i^0}{s_{ij}^0} = -a \\ e &= \frac{\partial \ell_{ij}}{\partial y_j} = \frac{y_j^0 - y_i^0}{s_{ij}^0} = -b \\ f &= \frac{\partial \ell_{ij}}{\partial z_j} = \frac{z_j^0 - z_i^0}{s_{ij}^0} = -c \end{aligned}\right\} \tag{3.79}$$

where s_{ij}^0 denotes the approximate distance from point P_i to P_j computed from the approximate coordinates x_i^0, y_i^0, z_i^0 and x_j^0, y_j^0, z_j^0.

If point P_i (or P_j) is fixed, then terms in (3.78) which involve δx_i, δy_i and δz_i (or δx_j, δy_j, δz_j) should be excluded, as a fixed point will not obtain coordinate corrections.

If the network is two-dimensional (e.g. in plane surveying), all terms in (3.77) and (3.78) involving the z-coordinate component will disappear. Eq.(3.78) now becomes :

$$(\ell_{ij} - s_{ij}^0) - \varepsilon_{ij} = a \cdot \delta x_i + b \cdot \delta y_i + c \cdot \delta x_j \ + d \cdot \delta y_j \tag{3.80}$$

with:

$$\left.\begin{aligned} s_{ij}^0 &= \sqrt{(x_j^0 - x_i^0)^2 + (y_j^0 - y_i^0)^2} \\ a &= \frac{\partial \ell_{ij}}{\partial x_i} = -\frac{x_j^0 - x_i^0}{s_{ij}^0} \\ b &= \frac{\partial \ell_{ij}}{\partial y_i} = -\frac{y_j^0 - y_i^0}{s_{ij}^0} \\ c &= \frac{\partial \ell_{ij}}{\partial x_j} = \frac{x_j^0 - x_i^0}{s_{ij}^0} = -a \\ d &= \frac{\partial \ell_{ij}}{\partial y_j} = \frac{y_j^0 - y_i^0}{s_{ij}^0} = -b \end{aligned}\right\} \tag{3.81}$$

Horizontal Direction Measurements in $2-D$ Networks

Let ℓ_{ij} denote a direction observed with theodolite at point P_i toward point P_j, and let ε_{ij} be its residual (see Fig. 3.3).

If we choose the $2-D$ coordinates of unknown points as parameters, the observation equation of ℓ_{ij} is :

$$\ell_{ij} - \varepsilon_{ij} = \arctan\left(\frac{y_j - y_i}{x_j - x_i}\right) - \chi_i \tag{3.82}$$

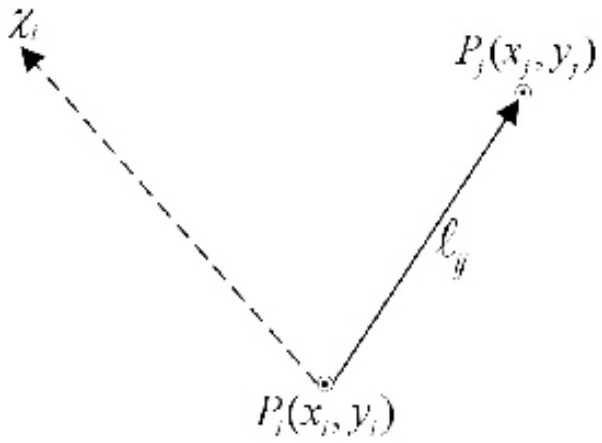

Figure 3.4: Observation equation of a horizontal direction

where χ_i denotes the *orientation angle* at theodolite station P_i (i.e. azimuth corresponding to the zero reading on the horizontal theodolite circle), and x_i, y_i and x_j , y_j stand for the 2-D coordinates of points P_i and P_j . Theoretically, χ_i is a part of the unknown parameters to be determined. In practice, they are no longer useful after the adjustment.

Denoting the approximate coordinates by x_i^0, y_i^0 and x_j^0, y_j^0 and their corrections by δx_i, δy_i and δx_j , δy_j , respectively, the non-linear equation (3.82) can be linearized:

$$(\ell_{ij} - \alpha_{ij}^0) - \varepsilon_{ij} = a \cdot \delta x_i + b \cdot \delta y_i + c \cdot \delta x_j \ + d \cdot \delta y_j \ - \chi_i \tag{3.83}$$

with:

$$\left.\begin{aligned} s_{ij}^0 &= \sqrt{(x_j^0 - x_i^0)^2 + (y_j^0 - y_i^0)^2} \\ \alpha_{ij}^0 &= \arctan\left(\tfrac{y_j^0 - y_i^0}{x_j^0 - x_i^0}\right) \\ a &= \tfrac{\partial \ell_{ij}}{\partial x_i} = \tfrac{y_j^0 - y_i^0}{\left(s_{ij}^0\right)^2} = \tfrac{\sin \alpha_{ij}^0}{s_{ij}^0} \\ b &= \tfrac{\partial \ell_{ij}}{\partial y_i} = -\tfrac{x_j^0 - x_i^0}{\left(s_{ij}^0\right)^2} = -\tfrac{\cos \alpha_{ij}^0}{s_{ij}^0} \\ c &= \tfrac{\partial \ell_{ij}}{\partial x_j} = -a \\ d &= \tfrac{\partial \ell_{ij}}{\partial y_j} = -b \end{aligned}\right\} \tag{3.84}$$

where s_{ij}^0 and α_{ij}^0 denote the approximate distance and azimuth, respectively, from point P_i to P_j computed from the approximate coordinates x_i^0, y^0 and x_j^0, y_j^0.

If point P_i (or P_j) is fixed, then terms in (3.83) which involve δx_i and δy_i (or δx_j , δy_j) should be excluded.

Horizontal Angle Measurements in $2-D$ Networks

Let ℓ_{ijk} denote a horizontal angle, observed with theodolite at station P_i and from object P_j to object P_k (clockwise), and let ε_{ijk} denote the residual of ℓ_{ijk}.

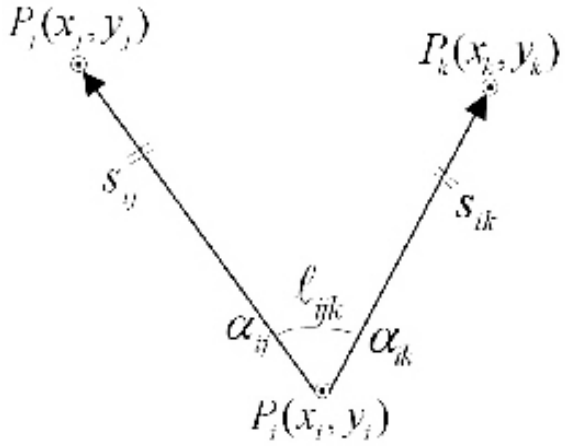

Figure 3.5: Observation equation of a horizontal angle

If the $2-D$ coordinates of unknown points are chosen as parameters, the observation equation of ℓ_{ijk} will be:

$$\ell_{ijk} - \varepsilon_{ijk} = \arctan\left(\frac{y_k - y_i}{x_k - x_i}\right) - \arctan\left(\frac{y_j - y_i}{x_j - x_i}\right) \tag{3.85}$$

where x, y denote the coordinates of points specified by their subscript (i, j or k).

A comparison of (3.85) with (3.82) shows that the observation equation of a horizontal angle (ℓ_{ijk}) can be regarded as the difference of two directions, $\ell_{ik} - \ell_{ij}$, where the orientation angle χ_i for theodolite station P_i disappears after differencing. From (3.83), we have the linearized version of (3.85) :

$$(\ell_{ijk} - \beta^0_{ijk}) - \varepsilon_{ijk} = a \cdot \delta x_i + b \cdot \delta y_i + c \cdot \delta x_j \ + d \cdot \delta y_j \ + e \cdot \delta x_k + f \cdot \delta y_k \tag{3.86}$$

where the coefficients are computed by :

$$\left.\begin{aligned}
a &= \frac{\partial \ell_{ijk}}{\partial x_i} = \frac{y^0_k - y^0_i}{\left(s^0_{ik}\right)^2} - \frac{y^0_j - y^0_i}{\left(s^0_{ij}\right)^2} = \frac{\sin\alpha^0_{ik}}{s^0_{ik}} - \frac{\sin\alpha^0_{ij}}{s^0_{ij}} \\
b &= \frac{\partial \ell_{ijk}}{\partial y_i} = -\frac{x^0_k - x^0_i}{\left(s^0_{ik}\right)^2} + \frac{x^0_j - x^0_i}{\left(s^0_{ij}\right)^2} = -\frac{\cos\alpha^0_{ik}}{s^0_{ik}} + \frac{\cos\alpha^0_{ij}}{s^0_{ij}} \\
c &= \frac{\partial \ell_{ijk}}{\partial x_j} = \frac{y^0_j - y^0_i}{\left(s^0_{ij}\right)^2} = \frac{\sin\alpha^0_{ij}}{s^0_{ij}} \\
d &= \frac{\partial \ell_{ijk}}{\partial y_j} = -\frac{x^0_j - x^0_i}{\left(s^0_{ij}\right)^2} = -\frac{\cos\alpha^0_{ij}}{s^0_{ij}} \\
e &= \frac{\partial \ell_{ijk}}{\partial x_k} = -\frac{y^0_k - y^0_i}{\left(s^0_{ik}\right)^2} = -\frac{\sin\alpha^0_{ik}}{s^0_{ik}} \\
f &= \frac{\partial \ell_{ijk}}{\partial y_k} = \frac{x^0_k - x^0_i}{\left(s^0_{ik}\right)^2} = \frac{\cos\alpha^0_{ik}}{s^0_{ik}}
\end{aligned}\right\} \tag{3.87}$$

with the approximate angle β^0_{ijk}, azimuths α^0_{ij}, α^0_{ik} and distances s^0_{ij}, s^0_{ik} computed from the approximate

coordinates :

$$\left.\begin{aligned}
\beta^0_{ijk} &= \alpha^0_{ik} - \alpha^0_{ij} \\
\alpha^0_{ik} &= \arctan\left(\tfrac{y^0_k - y^0_i}{x^0_k - x^0_i}\right) \\
\alpha^0_{ij} &= \arctan\left(\tfrac{y^0_j - y^0_i}{x^0_j - x^0_i}\right) \\
s^0_{ij} &= \sqrt{(x^0_j - x^0_i)^2 + (y^0_j - y^0_i)^2} \\
s^0_{ik} &= \sqrt{(x^0_k - x^0_i)^2 + (y^0_k - y^0_i)^2}
\end{aligned}\right\} \qquad (3.88)$$

If one or two of the three points (P_i, P_j , P_k) are fixed, the corresponding terms in (3.86) should be excluded accordingly.

Example 3.3

We use the method of adjustment by elements to adjust the geodetic network in **Example 2.4**, as illustrated in Fig 2.4 and the figure below.

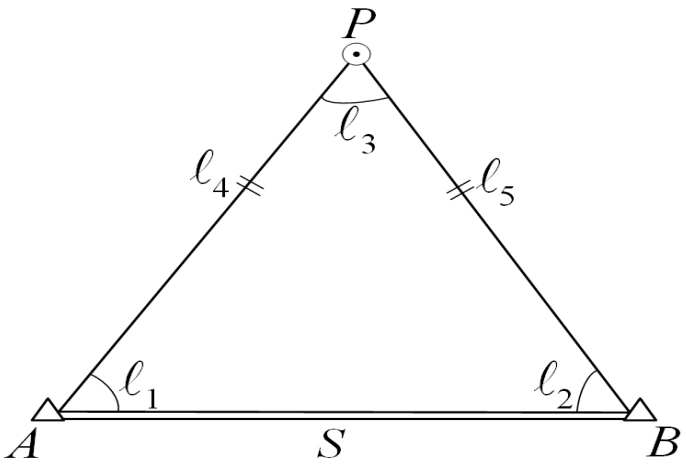

Figure 3.6: Single point positioning by angle and distance measurements

The observations and their a priori standard errors are as follows :

$$L' = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \\ \ell_5 \end{bmatrix} = \begin{bmatrix} 60^0\ 00'\ 05'' \\ 60^0\ 00'\ 05'' \\ 59^0\ 59'\ 58'' \\ 100.008\ m \\ 99.997\ m \end{bmatrix}, \quad \sigma = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \end{bmatrix} = \begin{bmatrix} \pm 6'' \\ \pm 6'' \\ \pm 6'' \\ \pm 3\ mm \\ \pm 3\ mm \end{bmatrix}$$

The given coordinates of the two fixed points A and B are assumed to be :

$$\begin{aligned} x_A &= 6\ 500\ 000.000\ \ m\,, \quad x_B = 6\ 500\ 060.000\ \ m \\ y_A &= 1\ 500\ 000.000\ \ m\,, \quad y_B = 1\ 500\ 080.000\ \ m \end{aligned}$$

For this network, we have $n = 5$ observations and $m = 2$ necessary observations. We choose the coordinates $(x_P,\ y_P)$ of the unknown point P as our parameters. In order to obtain the linearized observation equations, we first compute the approximate coordinates of P and also the approximate distances and azimuths between network points [see (3.88), (3.87)].

The distance and azimuth from P_A to P_B are calculated from the fixed coordinates:

$$S = s_{AB} = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} = \sqrt{60.000^2 + 80.000^2} = 100.000\ (m)$$

$$\alpha_{AB} = \arctan\left(\frac{y_B - y_A}{x_B - x_A}\right) = \arctan\left(\frac{80.000}{60.000}\right) = 53.1301024^0 = 53^0\ 7'\ 48.37''$$

$$\alpha_{BA} = \alpha_{AB} + 180^0 = 233.1301024^0 = 233^0\ 7'\ 48.37''$$

The approximate coordinates of network points to be determined can be obtained from e.g. old determinations, topographical maps or unadjusted observations. Here we use ℓ_1 and ℓ_4, together with s_{AB} and α_{AB}, to calculate the approximate azimuth α^0_{AP} and approximate coordinates $x^0_P,\ y^0_P$ of P:

$$\alpha^0_{AP} = \alpha_{AB} - \ell_1 = 353^0\ 7'\ 43.35''$$

$$x^0_P = x_A + \ell_4 \cos\alpha^0_{AP} = x_A + 99.2897\ m = 6\ 500\ 099.2897\ m$$
$$y^0_p = y_A + \ell_4 \sin\alpha^0_{AP} = y_A - 11.9649\ m = 1\ 499\ 988.0351\ m$$

The approximate distances and azimuths based on x^0_P, y^0_P and the fixed coordinates are then be computed and listed below.

From	*To*	*Distance* s^0 (m)	*Azimuth* α^0 $(^0\ '\ '')$	*Remark*
A	B	100.0000	$53^0\ 07'\ 48.37''$	fixed
A	P	100.0080	$353^0\ 07'\ 43.35''$	approximate
B	A	100.0000	$233^0\ 07'\ 48.37''$	fixed
B	P	100.0061	$293^0\ 08'\ 00.18''$	approximate
P	A	100.0080	$173^0\ 07'\ 43.35''$	fixed
P	B	100.0061	$113^0\ 08'\ 00.18''$	approximate

Denoting the corrections to x^0_p and y^0_P by δx_P and δy_P, respectively, the observation equations of the three angle measurements can be obtained according to Eqs.(3.86), (3.88) and (3.87):

$$\ell_1 - \varepsilon_1 = \left(\alpha_{AB} - \alpha^0_{AP}\right) + \frac{\rho''}{1000}\frac{\sin\alpha^0_{AP}}{s^0_{AP}}\delta x_p - \frac{\rho''}{1000}\frac{\cos\alpha^0_{AP}}{s^0_{AP}}\delta y_p$$

$$\ell_2 - \varepsilon_2 = \left(\alpha^0_{BP} - \alpha_{BA}\right) - \frac{\rho''}{1000}\frac{\sin\alpha^0_{BP}}{s^0_{BP}}\delta x_p + \frac{\rho\text{”}}{1000}\frac{\cos\alpha^0_{BP}}{s^0_{BP}}\delta y_p$$

$$\ell_3 - \varepsilon_3 = \left(\alpha^0_{PA} - \alpha^0_{PB}\right) + \frac{\rho''}{1000}\left(\frac{\sin\alpha^0_{PA}}{s^0_{PA}} - \frac{\sin^0_{PB}}{s^0_{PB}}\right)\delta x_p - \frac{\rho''}{1000}\left(\frac{\cos\alpha^0_{PA}}{s^0_{PA}} - \frac{\cos\alpha^0_{PB}}{s^0_{PB}}\right)\delta y_p$$

and the two distance measurements by Eqs.(3.80) and (3.81):

$$\ell_4 - \varepsilon_4 = s^0_{AP} + \cos\alpha^0_{AP}\delta x_p + \sin\alpha^0_{AP}\delta y_p$$
$$\ell_5 - \varepsilon_5 = s^0_{PB} + \cos\alpha^0_{PB}\delta x_p + \sin\alpha^0_{PB}\delta y_p$$

In the first three equations, factor $\frac{\rho''}{1000}$ ($\rho'' \approx 206265''$) is introduced so that ε_1, ε_2 and ε_3 will be in arc second ($''$), while δx_p and δy_p will be in millimetre (mm). In the last two equations, when δx_p, δy_p and $(\ell_4 - s^0_{AP})$, $(\ell_5 - s^0_{PB})$ are in mm, ε_4 and ε_5 will also be in mm.

Inserting the approximate distances and azimuths into the five observation equations above, we will finally arrive at the linearized observation equations:

$$\left.\begin{aligned} 0\,'' - \varepsilon_1 &= -0.246\,754\,45\,\delta x_p - 2.047\,670\,68\,\delta y_p \\ -8.81\,'' - \varepsilon_2 &= +1.896\,681\,98\,\delta x_p + 0.810\,309\,87\,\delta y_p \\ 14.83\,'' - \varepsilon_3 &= -1.649\,927\,53\,\delta x_p + 1.237\,360\,81\,\delta y_p \\ 0\,^{mm} - \varepsilon_4 &= +0.992\,817\,41\,\delta x_p - 0.119\,639\,41\,\delta y_p \\ -9.12\,^{mm} - \varepsilon_5 &= +0.392\,872\,97\,\delta x_p - 0.919\,592\,75\,\delta y_p \end{aligned}\right\}$$

or in matrix notation,

$$\underset{5\cdot1}{L} - \underset{5\cdot1}{\varepsilon} = \underset{5\cdot2}{A}\ \underset{2\cdot1}{X}$$

where:

$$L = \begin{bmatrix} 0 \\ -8.81 \\ +14.83 \\ 0 \\ -9.12 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}, \quad A = \begin{bmatrix} -0.246\,754\,45 & -2.047\,670\,68 \\ +1.896\,681\,98 & +0.810\,309\,87 \\ -1.649\,927\,53 & +1.237\,360\,81 \\ +0.992\,817\,41 & -0.119\,639\,41 \\ +0.392\,872\,97 & -0.919\,592\,75 \end{bmatrix}, \quad \delta X = \begin{bmatrix} \delta x_P \\ \delta y_P \end{bmatrix}$$

Notice that the 1st and 4th elements in vector L are exactly equal to zero. This is due to : (a) in general, vector L for non-linear observation equations actually represents the differences between the observed values and the approximate values derived from the approximate coordinates; (b) ℓ_1 and ℓ_4 have been used to derive approximate coordinates of point P so that there is no inconsistency between the observed values and the computed values for these two observations.

Choosing the a priori unit-weight standard error as $\sigma_0 = \pm 3\ mm$, the diagonal weight matrix of ε and L can be obtained just as in **Example 2.4** :

$$\underset{5\cdot5}{P} = \begin{bmatrix} \frac{1}{4} & & & & \\ & \frac{1}{4} & & & \\ & & \frac{1}{4} & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$$

With linearized observation equations and the above weight matrix, the adjustment can be carried out following the standard procedures as described in Section 3.1:

$$A^\top PA = \begin{bmatrix} +2.735\,173 & -0.479\,909 \\ -0.479\,909 & +2.455\,119 \end{bmatrix}, \quad \left|A^\top PA\right| = 6.484\,863$$

$$\left(A^\top PA\right)^{-1} = \tfrac{1}{6.484\,863} \begin{bmatrix} +2.455\,119 & +0.479\,909 \\ +0.479\,909 & +2.735\,173 \end{bmatrix} = \begin{bmatrix} +0.378\,592 & +0.074\,005 \\ +0.074\,005 & +0.421\,778 \end{bmatrix}$$

$$A^\top PL = \begin{bmatrix} -13.890\,199 \\ +11.199\,127 \end{bmatrix}$$

$$\delta\widehat{X} = (A^\top PA)^{-1}A^\top PL = \begin{bmatrix} +0.378\,592 & +0.074\,005 \\ +0.074\,005 & +0.421\,778 \end{bmatrix} \begin{bmatrix} -13.890\,199 \\ +11.199\,127 \end{bmatrix} = \begin{bmatrix} -4.43 \\ +3.70 \end{bmatrix}\ (mm)$$

$$\widehat{\varepsilon} = L - A\,\delta\widehat{X} = \begin{bmatrix} +6.45\,'' \\ -3.40\,'' \\ +2.95\,'' \\ +4.82\,^{mm} \\ -3.98\,^{mm} \end{bmatrix}$$

The finally adjusted coordinates of point P become:

$$\widehat{X} = X^0 + \delta\widehat{X} = \begin{bmatrix} 6\,500\,099.2897 \\ 1\,499\,988.0351 \end{bmatrix} + \begin{bmatrix} -0.00443 \\ +0.00370 \end{bmatrix} = \begin{bmatrix} 6500\,099.2853 \\ 1499\,988.0388 \end{bmatrix}\ (m)$$

and the adjusted (direct) observations will be :

$$\widehat{L}' = L' - \widehat{\varepsilon} = \begin{bmatrix} 60^0\ 00'\ 05'' \\ 60^0\ 00'\ 03'' \\ 59^0\ 59'\ 58'' \\ 100.008\ m \\ 99.997\ m \end{bmatrix} - \begin{bmatrix} +6.45\,'' \\ -3.40\,'' \\ +2.95\,'' \\ +4.82\,^{mm} \\ -3.98\,^{mm} \end{bmatrix} = \begin{bmatrix} 59^0\ 59'\ 58.55'' \\ 60^0\ 00'\ 06.40'' \\ 59^0\ 59'\ 55.05'' \\ 100.0032\ m \\ 100.0010\ m \end{bmatrix}$$

The a posteriori estimate of the variance factor σ_0^2 is obtained by Eq.(3.75) :

$$\widehat{\sigma}_0^2 = \frac{\widehat{\varepsilon}^\top P \widehat{\varepsilon}}{n-m} = \frac{1}{5-2}\sum_{i=1}^{n}\{p_i \widehat{\varepsilon}_i^2\} \approx 18.1885\ \ mm^2$$
$$\widehat{\sigma}_0 \approx \pm 4.26\ mm$$

The variance matrix of $\widehat{X}$ (and $\delta\widehat{X}$) follows from Eq.(3.69):

$$C_{\widehat{X}\widehat{X}} = \widehat{\sigma}_0^2 (A^\top P A)^{-1} = \widehat{\sigma}_0^2 \begin{bmatrix} +0.378\,592 & +0.074\,005 \\ +0.074\,005 & +0.421\,778 \end{bmatrix}$$
$$\sigma_{\widehat{x}_P} = \widehat{\sigma}_0 \sqrt{0.378572} \approx \pm 2.62\ mm;$$
$$\sigma_{\widehat{y}_P} = \widehat{\sigma}_0 \sqrt{0.421778} \approx \pm 2.77\ mm.$$
$$\sigma_P = \sqrt{\sigma_{\widehat{x}_P}^2 + \sigma_{\widehat{y}_P}^2} = \widehat{\sigma}_0 \sqrt{0.378572 + 0.421778} \approx \pm 3.82\ mm$$

The variance matrix of the adjusted observations is given by (3.74):

$$C_{\widehat{L}\widehat{L}} = \widehat{\sigma}_0^2\ A(A^\top P A)^{-1} A^\top = \widehat{\sigma}_0^2 \begin{bmatrix} +1.866\,333 & -1.179\,235 & -0.687\,098 & -0.137\,684 & +0.714\,774 \\ -1.179\,235 & +1.866\,364 & -0.687\,129 & +0.714\,765 & -0.137\,698 \\ -0.687\,098 & -0.687\,129 & +1.374\,227 & -0.577\,081 & -0.577\,076 \\ -0.137\,684 & +0.714\,765 & -0.577\,081 & +0.361\,630 & +0.123\,031 \\ +0.714\,774 & -0.137\,698 & -0.577\,076 & +0.123\,031 & +0.361\,639 \end{bmatrix}$$

which implies the following standard errors for each adjusted observation $\widehat{\ell}_i$ $(1 \le i \le 5)$:

$$\widehat{\sigma}_1 = \widehat{\sigma}_0 \sqrt{1.866333} \approx \pm 5.83''$$
$$\widehat{\sigma}_2 = \widehat{\sigma}_0 \sqrt{1.866364} \approx \pm 5.83''$$
$$\widehat{\sigma}_3 = \widehat{\sigma}_0 \sqrt{1.374227} \approx \pm 5.00''$$
$$\widehat{\sigma}_4 = \widehat{\sigma}_0 \sqrt{0.361630} \approx \pm 2.56\ mm$$
$$\widehat{\sigma}_5 = \widehat{\sigma}_0 \sqrt{0.361639} \approx \pm 2.56\ mm$$

A comparison of the above results with those of **Example 2.4** verifies that the condition adjustment and adjustment by elements lead to practically identical results for the same geodetic network, except minor decimal differences (e.g. for the variance matrix of $\widehat{L}$).

3.2.4 Observation Equations of 3D Measurements in A Local Area

Using modern totalstations, one can do 3D positioning in a local area where the $z-$coordinate is the vertical height component and $x-$ and $y-$axes are inside the local horizontal plan. The 3D measurements consist of the (slope) distance between two stations, the horizontal angle among three stations and the vertical angle from one station to another. The observation equations of the distance and horizontal angle have already been presented in Eqs. (3.78) and (3.83), respectively. Below, we will discuss only the observation equation of the vertical angle.

Assume that $P_i,\ P_j$ are two points whose 3D coordinates are denoted by $(x_i,\ y_i,\ z_i)$ and $(x_j,\ y_j,\ z_j)$, respectively. The observation equation of the measured vertical angle ℓ_{ij} from P_i toward P_j can be written as:

$$\ell_{ij} - \varepsilon_{ij} = arctan\left(\frac{z_j - z_i}{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}}\right) \tag{3.89}$$

Using the approximate coordinates $(x_i^0,\ y_i^0,\ z_i^0)$, $(x_j^0,\ y_j^0,\ z_j^0)$ of P_i and P_j, the above equation can be linearized as follows.

$$(\ell_{ij} - \ell_{ij}^0) - \varepsilon_{ij} = a\ \delta x_i + b\ \delta y_i + c\ \delta z_i + d\ \delta x_j + e\ \delta y_j + f\ \delta z_j \tag{3.90}$$

where ℓ_{ij}^0 denotes the computed value of the vertical angle ℓ_{ij} using the approximate coordinates:

$$\ell_{ij}^0 = arctan\left(\frac{z_j^0 - z_i^0}{\sqrt{(x_j^0 - x_i^0)^2 + \left(y_j^0 - y_i^0\right)^2}}\right) \tag{3.91}$$

The six coefficients (partial derivatives) on the right side of Eq.(3.90) are as follows:

$$a = \frac{1}{s_{ij}^{02}}\ \frac{(x_j^0 - x_i^0)\ (z_j^0 - z_i^0)}{\sqrt{(x_j^0 - x_i^0)^2 + \left(y_j^0 - y_i^0\right)^2}} \tag{3.92}$$

$$b = \frac{1}{s_{ij}^{02}}\ \frac{(y_j^0 - y_i^0)\ (z_j^0 - z_i^0)}{\sqrt{(x_j^0 - x_i^0)^2 + \left(y_j^0 - y_i^0\right)^2}} \tag{3.93}$$

$$c = \frac{1}{s_{ij}^{02}}\ \sqrt{(x_j^0 - x_i^0)^2 + \left(y_j^0 - y_i^0\right)^2} \tag{3.94}$$

$$d = -a\ ,\quad e = -b\ ,\quad f = -c$$

$$s_{ij}^0 = \sqrt{(x_j^0 - x_i^0)^2 + \left(y_j^0 - y_i^0\right)^2 + (z_j^0 - z_i^0)^2} \tag{3.95}$$

3.2.5 Observation Equations of GNSS Measurements

Today, more and more geodetic survey tasks are carried out using Global Navigation Satellite Systems (GNSS). In this sub-section, we describe the observation equations for several typical types of GNSS measurements.

The Code Pseudorange Measurement

The simpliest method for positioning and navigation using GNSS is to determine the time interval Δt_{ik} in which the radio signal travels from satellite k to the receiver i (at ground station P_i). If the satellite k has known coordinates $(x_k,\ y_k,\ z_k)$ and the receiver i has unknown coordinates $(x_i,\ y_i,\ z_i)$, the distance between them will be:

$$s_{ik} = c_0 \cdot \Delta t_{ik} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2 + (z_k - z_i)^2} \tag{3.96}$$

where c_0 denotes the speed of light.

There are many error sources which influence the measured value ℓ_{ik} of s_{ik}, such as ionospheric and tropospheric effects, and clock errors in the satellite and the receiver with respect to the correct system time. If we only model the combined clock error (δt_{ik}) of the satellite k and receiver i , the observation equation of the code pseudorange measurement ℓ_{ik} will read as:

$$\ell_{ik} - \varepsilon_{ik} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2 + (z_k - z_i)^2} + c_0\ \delta t_{ik} \tag{3.97}$$

Using the approximate coordinates $(x_i^0,\ y_i^0,\ z_i^0)$ of the receiver P_i , the above observation equation can be linearized:

$$(\ell_{ik} - s_{ik}^0) - \varepsilon_{ik} = a\ \delta x_i + b\ \delta y_i + c\ \delta z_i + d\ \delta t_{ik} \tag{3.98}$$

where:

$$\left.\begin{aligned} s_{ik}^0 &= \sqrt{(x_k - x_i^0)^2 + (y_k - y_i^0)^2 + (z_k - z_i^0)^2} \\ a &= \frac{\partial \ell_{ik}}{\partial x_i} = -\frac{x_k^0 - x_i^0}{s_{ik}^0} \\ b &= \frac{\partial \ell_{ik}}{\partial y_i} = -\frac{y_k^0 - y_i^0}{s_{ik}^0} \\ c &= \frac{\partial \ell_{ik}}{\partial z_i} = -\frac{z_k^0 - z_i^0}{s_{ik}^0} \\ d &= c_0 \end{aligned}\right\} \tag{3.99}$$

Observation Equation of the Carrier Phase Measurement

The geometric distance s_{ik} between satellite k and receiver i on the ground can also be expressed as the product of the wavelength λ and the total number of cycles of the radio signal travelling from the satellite to the receiver. The total number of cycles can be split into an integer part N and a decimal part ϕ :

$$s_{ik} = \lambda \cdot (\phi_{ik} + N_{ik}) \tag{3.100}$$

N_{ik} is also called *integer ambiguity* or simply *ambiguity*, while ϕ_{ik} is called the *carrier phase measurement* which can be made with very high precision. The ambiguity N_{ik} itself can be determined from the measurements (pseudorange and carrier phase measurements).

Taking into account the combined clock error (δt_{ik}) of the satellite k and receiver i, the observation equation of the phase measurement becomes:

$$\phi_{ik} - \varepsilon_{ik} = \frac{1}{\lambda}(s_{ik} + c_0\ \delta t_{ik}) - N_{ik} = \frac{1}{\lambda}\sqrt{(x_k - x_i)^2 + (y_k - y_i)^2 + (z_k - z_i)^2} + \frac{c_0}{\lambda}\ \delta t_{ik} - N_{ik} \quad (3.101)$$

Using the approximate coordinates ($x_i^0,\ y_i^0,\ z_i^0$) of receiver P_i , the above observation equation can be linearized:

$$(\phi_{ik} - \phi_{ik}^0) - \varepsilon_{ik} = a\ \delta x_i + b\ \delta y_i + c\ \delta z_i + d\ \delta t_{ik} + e\ N_{ik} \quad (3.102)$$

where:

$$\left.\begin{aligned} &\phi_{ik}^0 = \tfrac{1}{\lambda} s_{ik}^0 = \tfrac{1}{\lambda}\sqrt{(x_k - x_i^0)^2 + (y_k - y_i^0)^2 + (z_k - z_i^0)^2} \\ &a = \tfrac{\partial \ell_{ik}}{\partial x_i} = -\tfrac{x_k^0 - x_i^0}{s_{ik}^0} \\ &b = \tfrac{\partial \ell_{ik}}{\partial y_i} = -\tfrac{y_k^0 - y_i^0}{s_{ik}^0} \\ &c = \tfrac{\partial \ell_{ik}}{\partial z_i} = -\tfrac{z_k^0 - z_i^0}{s_{ik}^0} \\ &d = \tfrac{1}{\lambda} c_0 \\ &e = -1 \end{aligned}\right\} \quad (3.103)$$

Eqs. (3.98) and (3.102) indicate that the unknown parameters in the observation equations can include not only the unknown coordinate corrections ($\delta x_i,\ \delta y_i,\ \delta z_i$), but also other types of parameters such as $\delta t_{ik},\ N_{ik}$.

3.3 Adjustment by Elements with Constraints

In some geodetic networks, one or more quantities (such as distances, azimuths, angles, or height differences) between unknown points are fixed. In this case, the number of necessary observations (m_0) will normally be smaller than the number of unknown points multiplied by the dimension of the network. However, sometimes it may be easier to obtain observation equations if the coordinates of all unknown points are regarded as unknown parameters. As pointed out before, choosing more than m_0 parameters will result in functional dependency among the parameters. Then, one need to use the method of adjustment by elements with constraints to be described below.

Assume that we have the following observation equations:

$$\underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} = \underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} \tag{3.104}$$

When m is larger than the number (m_0) of necessary observations, the m parameters in vector X will satisfy the following $r = m - m_0$ constraints (matrix equations):

$$\underset{r\cdot m}{A_x}\ \underset{m\cdot 1}{X} = \underset{r\cdot 1}{d} \tag{3.105}$$

where L, ε and A are defined as in (3.6), (3.7) and (3.8), and A_x, d are given matrices (or vector). The residual ε is assumed to have the following a priori statistical properties:

$$E(\varepsilon) = 0\ , \quad E(\varepsilon\varepsilon^\top) = \sigma_0^2\ P^{-1} \tag{3.106}$$

The task of adjustment by elements with constraints is to find estimates $\widehat{X}$ and $\widehat{\varepsilon}$ such that the following three conditions are satisfied:

$$\widehat{\varepsilon}^\top P\widehat{\varepsilon} = \ minimum \tag{3.107}$$

$$L - \widehat{\varepsilon} = A\widehat{X} \tag{3.108}$$

$$A_x\widehat{X} = d \tag{3.109}$$

This is again a conditional minimization problem. To find the solution, we form the Lagrange function $F(X)$ and let its derivative with respect to X be zero :

$$F(X) = \varepsilon^\top P\varepsilon + 2\lambda^\top (A_x X - d)$$

$$\frac{\partial F}{\partial X}\ |_{X=\widehat{X},\ \varepsilon=\widehat{\varepsilon}} = 2\ \widehat{\varepsilon}^\top P\,(-A) + 2\lambda^\top A_x = 0$$

i.e.:

$$A^\top P\widehat{\varepsilon} - A_x^\top\lambda = 0 \tag{3.110}$$

where λ denotes the Lagrangian multipliers. After inserting (3.108) into (3.110), the normal equation of adjustment by elements with constraints is obtained :

$$\begin{bmatrix} A^\top PA & A_x^\top \\ A_x & 0 \end{bmatrix} \begin{bmatrix} \widehat{X} \\ \lambda \end{bmatrix} = \begin{bmatrix} A^\top PL \\ d \end{bmatrix} \tag{3.111}$$

From the first equation of Eq.(3.111), one can get :

$$\widehat{X} = N^{-1}A^\top P\ L - N^{-1}A_x^\top\lambda \tag{3.112}$$

where:

$$N = A^\top PA \tag{3.113}$$

Inserting (3.112) into (3.109), λ is finally found:

$$\lambda = N_x^{-1}(A_xN^{-1}A^\top P\ L - d)$$

where:

$$N_x = A_xN^{-1}A_x^\top \tag{3.114}$$

The least squares estimate $\widehat{X}$ is then obtained by putting back λ into (3.112):

$$\widehat{X} = N^{-1}(I - A_x^\top N_x^{-1} A_x N^{-1})A^\top PL + N^{-1}A_x^\top N_x^{-1} d \tag{3.115}$$

The estimated residual $\widehat{\varepsilon}$ and the adjusted observation $\widehat{L}$ are obtained in the usual way as before :

$$\widehat{\varepsilon} = L - A\widehat{X} \tag{3.116}$$

$$\widehat{L} = L - \widehat{\varepsilon} = A\widehat{X} \tag{3.117}$$

The a posteriori estimate of σ_0^2 can be obtained based on the estimated $\widehat{\varepsilon}$:

$$\widehat{\sigma}_0^2 = \frac{\widehat{\varepsilon}^\top P \widehat{\varepsilon}}{n - m + r} \tag{3.118}$$

where $n - m + r$ denotes the number of over-determinations, as the number of independent parameters is $m - r$.

Using the error propagation law, the variance matrices of $\widehat{X}$ and $\widehat{L}$ can be derived :

$$\begin{aligned} C_{\widehat{X}\widehat{X}} &= N^{-1}(I - A_x^\top N_x^{-1} A_x N^{-1})A^\top P \cdot \sigma_0^2 P^{-1} \cdot \left[N^{-1}(I - A_x^\top N_x^{-1} A_x N^{-1})A^\top P\right]^\top \\ &= \sigma_0^2 (N^{-1} - N^{-1} A_x^\top N_x^{-1} A_x N^{-1}) \end{aligned} \tag{3.119}$$

$$C_{\widehat{L}\widehat{L}} = A C_{\widehat{X}\widehat{X}} A^\top = \sigma_0^2 \, A(N^{-1} - N^{-1} A_x^\top N_x^{-1} A_x N^{-1})A^\top \tag{3.120}$$

If the observation equations and the constraints in (3.105) are non-linear functions of the parameters X to be determined, one can use the same procedures described in Subsection 3.2.1 to linearize the non-linear equations. Assume that we have the following n non-linear observation equations:

$$\left.\begin{array}{l} \widetilde{\ell}_1 = \ell_1 - \varepsilon_1 = \; f_1(x_1, x_2, \cdots, x_m) \\ \widetilde{\ell}_2 = \ell_2 - \varepsilon_2 = \; f_2(x_1, x_2, \cdots, x_m) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \\ \widetilde{\ell}_i = \ell_i - \varepsilon_i \; = \; f_i(x_1, x_2, \cdots, x_m) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \\ \widetilde{\ell}_n = \ell_n - \varepsilon_n = f_n(x_1, x_2, \cdots, x_m) \end{array}\right\} \tag{3.121}$$

Let x_i^0 and δx_i denote an approximate value of x_i and its corresponding correction such that $x_i = x_i^0 + \delta x_i$ $(1 \le i \le n)$:

$$\underset{m\cdot 1}{X} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix} = \underset{m\cdot 1}{X}^0 + \underset{m\cdot 1}{\delta X}, \quad \underset{m\cdot 1}{X}^0 = \begin{bmatrix} x_1^0 \\ x_2^0 \\ \cdots \\ x_m^0 \end{bmatrix}, \quad \underset{m\cdot 1}{\delta X} = \begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \cdots \\ \delta x_m \end{bmatrix} \tag{3.122}$$

As described in Subsection 3.2.2, the non-linear observation equations can be linearized to :

$$\underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} = \underset{n\cdot m}{A} \; \underset{m\cdot 1}{\delta X} \tag{3.123}$$

where:

$$\underset{n\cdot 1}{L} = L' - c, \quad L' = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \cdots \\ \ell_n \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{bmatrix}, \quad c_i = f_i(x_1^0,\, x_2^0,\, \cdots,\, x_m^0) \tag{3.124}$$

$$\underset{n\cdot m}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad a_{ij} = \frac{\partial f_i}{\partial x_j} \quad (1 \le i \le n;\ 1 \le j \le m) \tag{3.125}$$

Now let us consider the $r = m - m_0$ non-linear constraints among the unknown parameters :

$$\left.\begin{array}{l} g_1(x_1, x_2, \cdots, x_m) = 0 \\ g_2(x_1, x_2, \cdots, x_m) = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ g_i(x_1, x_2, \cdots, x_m) = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ g_r(x_1, x_2, \cdots, x_m) = 0 \end{array}\right\} \tag{3.126}$$

where g_i $(1 \le i \le r)$ denotes an arbitrary non-linear function. Replacing x_j $(1 \le j \le m)$ in (3.126) by $x_j^0 + \delta x_j$ and expanding each function into a Taylor series
(around x_j^0) truncated to first-order terms, the above r non-linear equations reduce to :

$$\underset{r\cdot m}{A_x}\ \underset{m\cdot 1}{\delta X} = \underset{r\cdot 1}{d} \tag{3.127}$$

where :

$$\underset{r\cdot m}{A_x} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ g_{r1} & g_{r2} & \cdots & g_{rm} \end{bmatrix}, \quad g_{ij} = \frac{\partial g_i(x_1^0, x_2^0, \cdots, x_m^0)}{\partial x_j^0} \quad (1 \le i \le r,\ \ 1 \le j \le m) \tag{3.128}$$

$$\underset{r\cdot 1}{d} = \begin{bmatrix} d_1 \\ d_2 \\ \cdots \\ d_r \end{bmatrix}, \quad d_i = -g_i(x_1^0, x_2^0, \cdots, x_m^0) \quad (1 \le i \le r) \tag{3.129}$$

After linearization of both the observation equations and the constraints, the least squares estimates of δX can be obtained from Eq.(3.115):

$$\delta\widehat{X} = N^{-1}(I - A_x^\top N_x^{-1} A_x N^{-1})A^\top PL + N^{-1}A_x^\top N_x^{-1} d \tag{3.130}$$

which also leads to the least squares estimate of X :

$$\widehat{X} = X^0 + \delta\widehat{X} \tag{3.131}$$

The least squares estimates of ε and L, the a posteriori estimate of σ_0^2 and the variance-covariance matrices of $\widehat{X}$, $\widehat{L}$ are given by Eqs.(3.116), (3.117), (3.118), (3.119) and (3.120), respectively. Details are omitted here.

Chapter 4

Generalized Matrix Inverses in Least Squares Adjustment

Since the last world war, science and engineering have undergone tremendous development. Revolutionary developments in the fields of *e.g.* computer engineering, electronic communication and space technology have contributed to new advancement in theory of errors. Classical adjustment techniques have been given a more sound theoretical foundation, mostly due to application of statistical theory and linear algebra. New problems have been identified and new methods developed for processing measurement data with the help of modern electronic computers. To reflect these new developments, we will introduce several new topics in theory of errors and least squares adjustment, starting from this chapter. The first topic concerns generalized matrix inverses and their application in surveying adjustment. Here we extend the classical concepts of matrix inverses in order to obtain general solutions to various kinds of linear equation systems. Traditional adjustment methods are then treated as specific types of linear equation systems.

Numerically, surveying adjustment may be regarded as solving different types of linear equation systems like:

$$\underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} = \underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} \tag{4.1}$$

where the last term above is needed when an inconsistent equation system is concerned (*i.e.* $AX \neq L$). If the above system is consistent, then the residual term ε is not needed. When A is a non-singular square matrix of dimension $n \times n$, i.e. having non-zero determinant:

$$|A| \neq 0 \tag{4.2}$$

then the solution can be uniquely expressed by the inverse matrix of A^{-1} :

$$X = A^{-1}L \tag{4.3}$$

where the inverse matrix A^{-1} satisfies the classical definition in matrix algebra (I = unit matrix):

$$A \cdot A^{-1} = A^{-1} \cdot A = I \tag{4.4}$$

However, for a rectangular matrix A (*i.e.* $n \neq m$) such as in condition adjustment ($n < m$) or in adjustment by elements ($n > m$), the above inverse matrix is not defined according to classical matrix algebra. Nevertheless, one may still wish to express the solution of an arbitrary equation system Eq.(4.1) in a simple, intuitive and elegant way :

$$X = G\ L \tag{4.5}$$

where matrix G functions like an inverse matrix, just as A^{-1} in Eq.(4.3).

Another situation, where the classical definition of matrix inverses needs to be modified, occurs when matrix A is a singular square matrix. This happens with the coefficient matrix of our normal equation,

when linearly-correlated condition equations are used or when the network datum is not sufficiently defined in adjustment by elements.

Historically, the former Department of Geodesy at KTH has played an important role in the theoretical development of generalized inverses, mainly due to Professor Arne Bjerhammar's pioneer work in this mathematical field. To interested readers, the following literatures on generalized inverses may be recommended: Bjerhammar (1973); Rao and Mitra (1971); Rao (1973), and finally Sjöberg (1984), Sjöberg (1990).

4.1 Generalized Matrix Inverses

4.1.1 $g-$Inverses

For any arbitrary matrix $\underset{n\cdot m}{A}$, a matrix A^- of dimension $m \times n$ is called a generalized inverse of A iff :

$$A \underset{m\cdot n}{A}^- A = A \tag{4.6}$$

The generalized inverse A^- defined above is also called a $g-$inverse of A. The traditional inverse of a non-singular matrix is automatically a generalized inverse, but not vice versa. It can be proved that for any arbitrary matrices A and B, the rank of AB is not larger than the rank of A or B,

$$r(AB) \leq \min\{r(A),\, r(B)\} \tag{4.7}$$

Applying Eq.(4.7) on Eq.(4.6), we can deduct :

$$r(A) = r(AA^-A) \leq \min\left\{r(A),\, r(A^-)\right\}$$

which implies:

$$r(A^-) \geq r(A) \tag{4.8}$$

Example 4.1

Let A, B, C denote the following three matrices:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$$

By direct calculations, we find:

$$ABA = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = A$$

and

$$ACA = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = A$$

According to definition Eq.(4.6), both B and C are generalized inverses of A, although $B \neq C$. ∎

Example 4.2

Assume that matrix $\underset{n\cdot m}{A}$ has rank $r = r(A) \leq \min(n,\, m)$ and that A can be partitioned so that one of the diagonal sub-matrix (say, A_{11}) is a non-singular square matrix with full rank r :

$$\underset{n\cdot m}{A} = \begin{bmatrix} \underset{r\cdot r}{A_{11}} & \underset{r\cdot t}{A_{12}} \\ \underset{s\cdot r}{A_{21}} & \underset{s\cdot t}{A_{22}} \end{bmatrix} \quad (n = r+s, \;\; m = r+t) \tag{4.9}$$

Then one particular generalized inverse of A is :

$$\underset{m\cdot n}{A}^{-} = \begin{bmatrix} \underset{r\cdot r}{A_{11}^{-1}} & \underset{r\cdot s}{0} \\ \underset{t\cdot r}{0} & \underset{t\cdot s}{0} \end{bmatrix} \tag{4.10}$$

where A_{11}^{-1} denotes the ordinary inverse of A_{11}, as $|A_{11}| \neq 0$.

Proof. Since $r(A) = r \leq \min(n, m)$ and A_{11} has full rank, some rows (or columns) of A are linear combinations of other rows (or columns). This means that there exists a matrix, α, of dimension $s \times r$, such that :

$$\begin{bmatrix} A_{21} & A_{22} \end{bmatrix} = \alpha \begin{bmatrix} A_{11} & A_{12} \end{bmatrix}$$

which leads to

$$\alpha = A_{21}\, A_{11}^{-1} \quad \text{and} \quad A_{22} = \alpha A_{12} = A_{21}\, A_{11}^{-1} A_{12}$$

Now we can insert $A_{22} = A_{21}\, A_{11}^{-1} A_{12}$ into Eq.(4.9) to obtain :

$$AA^{-}A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{21}\, A_{11}^{-1} A_{12} \end{bmatrix} \cdot \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{21}\, A_{11}^{-1} A_{12} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = A$$

which proves that A^{-} given by Eq.(4.10) is a generalized inverse of A. ∎

For the following matrix A which has rank $r = 2$:

$$\underset{4\times 3}{A} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & 4 \\ 4 & 2 & 4 \\ 6 & 3 & 6 \end{bmatrix}$$

one particular generalized inverse of A can be obtained using Eq.(4.10):

$$\underset{3\times 4}{A^{-}} = \begin{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} & \underset{2\times 2}{0} \\ \underset{1\times 2}{0} & \underset{1\times 2}{0} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Example 4.3 (*rank factorization*)

Let $\underset{n\cdot m}{A}$ be a matrix with rank $r(A) = r \leq \min(n, m)$. By rank factorization, A can be written as a product of two matrices with same rank r :

$$\underset{n\cdot m}{A} = \underset{n\cdot r}{B}\; \underset{r\cdot m}{C}, \qquad [r(A) = r(B) = r(C) = r] \tag{4.11}$$

Then a particular generalized inverse of A is given by:

$$A^{-} = C^{\top}(CC^{\top})^{-1}(B^{\top}B)^{-1}B^{\top} \tag{4.12}$$

where both $CC^{\top}$ and $B^{\top}B$ have full rank, r, and thus have conventional inverses. That A^{-} is a $g-$inverse of A can be proved directly:

$$A\, A^{-}A = BC \cdot C^{\top}(CC^{\top})^{-1}(B^{\top}B)^{-1}B^{\top} \cdot BC = BC = A$$

One way to make rank factorization in Eq.(4.11) on a symmetrical matrix A of dimension $n \times n$ and rank r is to apply Cholesky's triangular decomposition on A:

$$A = T^\top T$$

where T is an upper-triangular matrix. If $r < n$, there must exist $(n - r)$ rows in T which contain only zeros. If we denote by $\overline{T}$ the matrix obtained from T after all zero-rows have been removed, we have:

$$A = \overline{T}^\top \overline{T} \tag{4.13}$$

Eq.(4.13) is known as *truncated triangular decomposition*. A numerical example of truncated triangular decomposition is given below for a matrix A defined as:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad r(A) = 1$$

Cholesky's triangular decomposition gives :

$$A = T^\top T = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

which leads to the truncated triangular decomposition:

$$A = \overline{T}^\top \overline{T} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = BC$$

where: $B^\top = C = \begin{bmatrix} 1 & 2 \end{bmatrix}$, $B^\top B = C^\top C = 5$, $(B^\top B)^{-1} = (CC^\top)^{-1} = 1/5$. By Eq.(4.12), we get a particular generalized inverse of A:

$$A^- = C^\top (CC^\top)^{-1} (B^\top B)^{-1} B^\top = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \frac{1}{5} \frac{1}{5} \begin{bmatrix} 1 & 2 \end{bmatrix} = \frac{1}{25} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

As a check, we may calculate:

$$AA^-A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \cdot \frac{1}{25} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = A \ ! \quad \blacksquare$$

Example 4.1 and **4.3** show that the generalized inverse defined by Eq.(4.6) is not unique. In other words, for an arbitrary matrix A there exist more than one generalized inverses A^- which all satisfy the definition in Eq.(4.6). The following theorem shows that once a particular generalized inverse A^- is found, the complete set of generalized inverses of A (i.e. all possible generalized inverses of A) can be represented in a general form.

Theorem 4.1

Let A^- be a particular generalized inverse of matrix $\underset{n\cdot m}{A}$. Then the complete set of generalized inverses of A, denoted as $\widetilde{A}^-$, is given by:

$$\underset{m\cdot n}{\widetilde{A}^-} = \underset{m\cdot n}{A^-} + \underset{m\cdot n}{N} \left[\underset{n\cdot n}{I} - \underset{n\cdot m}{A} \cdot \underset{m\cdot n}{A^-} \right] + \left[\underset{m\cdot m}{I} - \underset{m\cdot n}{A^-} \cdot \underset{n\cdot m}{A} \right] \underset{m\cdot n}{M} \tag{4.14}$$

where N and M are two arbitrary matrices with compatible dimensions.

Proof. Let $\widetilde{A}^-$ denote the complete set of generalized inverse of A, A^- a particular inverse, and A_1^- a matrix defined as:

$$A_1^- = A^- + N\left[I - AA^-\right] + \left[I - A^-A\right] M$$

Then we have:

$$AA_1^-A = A\left\{A^- + N\left[I - AA^-\right] + \left[I - A^-A\right] M\right\} A = A$$

which means A_1^- is a generalized inverse, i.e.:

$$A_1^- \in \widetilde{A}^-$$

For a special choice, $M = \widetilde{A}^- A\ A^-$ and $N = \widetilde{A}^- - A^-$, we have:

$$A_1^- = A^- + \left(\widetilde{A}^- - A^-\right)\left[I - AA^-\right] + \left[I - A^-A\right] \cdot \widetilde{A}^- A\ A^- = \widetilde{A}^-$$

which implies:

$$\widetilde{A}^- \in A_1^-$$

So we should have

$$A_1^- \in \widetilde{A}^- \in A_1^-$$

which holds iff :

$$\widetilde{A}^- = A_1^- = A^- + N\left[I - AA^-\right] + \left[I - A^-A\right] M$$

Example 4.4

In **Example 4.1**, matrix A and one particular generalized inverse of A are as follows:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad A^- = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

from which one can find:

$$AA^- = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}, \quad I - AA^- = \begin{bmatrix} 0 & 0 \\ -2 & 1 \end{bmatrix}, \quad I - A^-A = \begin{bmatrix} 0 & -2 \\ 0 & 1 \end{bmatrix}$$

By Eq.(4.14), the complete set of g-inverses of A is :

$$\widetilde{A}^- = A^- + N(I - A\ A^-) + (I - A^-A)M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + N \begin{bmatrix} 0 & 0 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -2 \\ 0 & 1 \end{bmatrix} M$$

If the arbitrary matrices $N,\ M$ are:

$$N = \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix}, \quad M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

$\widetilde{A}^-$ can then be written as:

$$\widetilde{A}^- = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}. \quad \blacksquare$$

Before we turn to special generalized inverses, we present a very useful theorem on the $g-$inverses which will be used in later discussions of this chapter.

Theorem 4.2

Let matrix A have rank r and let R , S be two arbitrary matrices of compatible dimensions. If $r(RA) = r(AS) = r$, two generalized inverses of A are :

$$A^- = (RA)^- R \tag{4.15}$$

$$A^- = S(AS)^- \tag{4.16}$$

Proof. We first make rank factorization on A [see Eq.(4.11)]:

$$\underset{n\cdot m}{A} = \underset{n\cdot r}{B}\ \underset{r\cdot m}{C}, \qquad [r(A) = r(B) = r(C) = r]$$
$$RA = RB\ C = B_1\ C, \text{ where } B_1 = RB,\ r(A) = r(B_1) = r(C) = r,\ \left|B_1^\top B_1\right| \neq 0,\ \left|CC^\top\right| | \neq 0$$

From Eq.(4.12), we can obtain one particular generalized inverse for RA :

$$(RA)^- = C^\top(CC^\top)^{-1}(B_1^\top B_1)^{-1}B_1^\top$$
$$(RA)^- R = C^\top(CC^\top)^{-1}(B_1^\top B_1)^{-1}B_1^\top R$$

That Eq.(4.15) holds, for the particular g-inverse $(RA)^-$ of RA given above, can be easily proved:

$$A\cdot\ (RA)^- R\cdot\ A = B\ C\cdot C^\top(CC^\top)^{-1}(B_1^\top B_1)^{-1}B_1^\top R\cdot BC = BC = A$$

We then prove that Eq.(4.15) holds for any arbitrary g-inverse $\widetilde{(RA)}^-$ of RA. From Eq.(4.14), $\widetilde{(RA)}^-$ can be written as:

$$\widetilde{(RA)}^- = (RA)^- + N[I - (RA)(RA)^-] + [I - (RA)^-(RA)]M$$
$$\widetilde{(RA)}^- R = \{(RA)^- + N[I - (RA)(RA)^-] + [I - (RA)^-(RA)]M\}\cdot R$$
$$A\cdot\widetilde{(RA)}^- R\cdot A = A\{(RA)^- + N[I - (RA)(RA)^-] + [I - (RA)^-(RA)]M\}\ R\ A$$
$$= A(RA)^- R\ A + AN[RA - (RA)(RA)^-(RA)] + [A - A(RA)^- R\ A)]MRA$$
$$= A\cdot A^-\cdot A + AN[RA - RA] + [A - A\cdot A^-\cdot A)]MRA = A$$

This proves that $(RA)^- R$ is a generalized inverse of A for any arbitrary choice of $(RA)^-$. Eq.(4.16) can be proved in a similar way.

4.1.2 Minimum-Norm Inverses

For an arbitrary matrix A, Eq.(4.14) defines the general g-inverse A^-. Among all possible g-inverses of A, one can identify several special g-inverses. Let $\underset{n\cdot m}{A}$ denote a matrix of rank r and $\underset{m\cdot m}{Q}$ a positive definite matrix. A minimum-norm generalized inverse of A is defined as:

$$A_{QO}^- = QA^\top(AQA^\top)^- \tag{4.17}$$

where $(AQA^\top)^-$ is one particular generalized inverse of $AQA^\top$. The name ″minimum-norm inverses″ is due to the property of the solution of a linear equation system obtained using A_{QO}^- (See **Subsection 4.2.2**).

As Q is positive definite, we have $r(QA^\top) = r(A) = r$. Considering Eq.(4.16) in **Theorem 4.2**, it follows that A_{QO}^- is a g-inverse of A (for $S = QA^\top$). As the g-inverse $(AQA^\top)^-$ in Eq.(4.17) is normally not unique, the minimum-norm inverse A_{QO}^- is not unique either. The complete set of A_{QO}^- can be obtained by using the complete set of g-inverses of $AQA^\top$ in Eq. Eq.(4.17). Letting $(AQA^\top)^-$ be one particular g-inverse of $AQA^\top$, the complete set of the g-inverse of $AQA^\top$ follows from Eq.(4.14) :

$$\widetilde{(AQA^\top)}^- = (AQA^\top)^- + \overline{N}\left[I - \left(AQA^\top\right)\left(AQA^\top\right)^-\right] + \left[I - \left(AQA^\top\right)^-\left(AQA^\top\right)\right]\overline{M}$$

where $\overline{N}$ and $\overline{M}$ are arbitrary matrices of compatible dimensions.

The complete set of minimum-norm inverses of A follows:

$$\widetilde{A}_{QO}^- = QA^\top\widetilde{(AQA^\top)}^-$$
$$= QA^\top\cdot\left\{(AQA^\top)^- + \overline{N}\left[I - \left(AQA^\top\right)\left(AQA^\top\right)^-\right] + \left[I - \left(AQA^\top\right)^-\left(AQA^\top\right)\right]\overline{M}\right\}$$

$$= A_{QO}^{-} + N\left[I - A\ A_{QO}^{-}\right] + \left[QA^{\top} - QA^{\top}\cdot\left(AQA^{\top}\right)^{-}A\cdot QA^{\top}\right]\overline{M}$$
$$= A_{QO}^{-} + N\left[I - A\ A_{QO}^{-}\right] + \left[QA^{\top} - QA^{\top}\cdot\left(QA^{\top}\right)^{-}\cdot QA^{\top}\right]\overline{M}$$

where $N = QA^{\top}\overline{N}$ denotes a new arbitrary matrix. Considering Eq.(4.15), we notice that $\left(AQA^{\top}\right)^{-}A$ is a g-inverse of $QA^{\top}$. The above equation then becomes:

$$\widetilde{A}_{QO}^{-} = A_{QO}^{-} + N\left[I - A\ A_{QO}^{-}\right] \tag{4.18}$$

Right-multiplying both sides of (4.18) by A gives :

$$\widetilde{A}_{QO}^{-}\cdot A = A_{QO}^{-}\cdot A \tag{4.19}$$

which means that *the product of the minimum-norm inverse A_{QO}^{-} and A itself is unique*, even though A_{QO}^{-} is generally not unique.

If the rank of matrix $\underset{n\cdot m}{A}$ is equal to dimension n (i.e. the rows of matrix A have full rank), the minimum-norm inverse A_{QO}^{-} becomes a unique inverse. This is because $r(A) = n$ leads to $r\left(AQA^{\top}\right) = n$ and $|AQA^{\top}| \neq 0$. Consequently the conventional inverse $(AQA^{\top})^{-1}$ is well defined. In this case [i.e.: $r(A) = n$], we can write the minimum-norm inverse of A by A_{QO}^{-1} :

$$A_{QO}^{-1} = QA^{\top}(AQA^{\top})^{-1} \tag{4.20}$$

for which one has :

$$A\ A_{QO}^{-1} = I \tag{4.21}$$

Note that Eq.(4.21) does not hold for any arbitrary A_{QO}^{-} with $r(A) < n$.

4.1.3 Least Squares Inverses

Let $\underset{n\cdot n}{P}$ be a positive definite matrix. The least squares inverses of matrix $\underset{n\cdot m}{A}$ are defined as:

$$A_{OP}^{-} = (A^{\top}PA)^{-}A^{\top}P \tag{4.22}$$

where $(A^{\top}PA)^{-}$ denotes one particular g-inverse of $A^{\top}PA$.

As $(A^{\top}PA)^{-}$ is usually non-unique, A_{OP}^{-} is not unique either. The complete set of A_{OP}^{-} can be derived in the same way as the derivation of Eq.(4.18). Let $(A^{\top}PA)^{-}$ and $(\widetilde{A^{\top}PA})^{-}$ denote a particular g-inverse and the arbitrary g-inverse of $A^{\top}PA$, respectively. Furthermore, let A_{OP}^{-} and $\widetilde{A}_{OP}^{-}$ denote a particular and the arbitrary least squares inverse of A, respectively. Then we have:

$$\widetilde{A}_{OP}^{-} = (\widetilde{A^{\top}PA})^{-}A^{\top}P$$
$$= \left\{(A^{\top}PA)^{-} + N'\left[I - \left(A^{\top}PA\right)\left(A^{\top}PA\right)^{-}\right] + \left[I - \left(A^{\top}PA\right)^{-}\left(A^{\top}PA\right)\right]M'\right\}\cdot A^{\top}P$$
$$= A_{OP}^{-} + N'\left[A^{\top}P - A^{\top}P\cdot A\left(A^{\top}PA\right)^{-}\cdot A^{\top}P\right] + \left[I - A_{OP}^{-}A\right]M'A^{\top}P$$

where M' and N' denote two arbitrary matrice of suitable dimensions.

Applying Eq.(4.16) for $S = A$ shows that $A\left(A^{\top}PA\right)^{-}$ is a $g-$inverse of $A^{\top}P$. Thus the complete set of the least squares generalized inverses of A becomes now :

$$\widetilde{A}_{OP}^{-} = A_{OP}^{-} + \left[I - A_{OP}^{-}A\right]M \tag{4.23}$$

where $M = M'A^\top P$ is a (new) arbitrary matrix. Left-multiplying both sides of (4.23) by A gives :

$$A \cdot \widetilde{A}_{OP}^{-} = A \cdot A_{OP}^{-} \tag{4.24}$$

which means that *the product of matrix A and its least squares inverse A_{OP}^{-} is unique*, even though A_{OP}^{-} is generally not unique.

If $r(A) = m$ holds, then $r(A^\top PA) = m$ and $|A^\top PA| \neq 0$. Consequently, $(A^\top PA)^{-1}$ is well defined and Eq.(4.22) represents a unique $g-$inverse of A:

$$A_{OP}^{-1} = (A^\top PA)^{-1}A^\top P \tag{4.25}$$

which satisfies :

$$A_{OP}^{-1}A = I \tag{4.26}$$

4.1.4 Minimum-Norm Least Squares Inverse

The minimum-norm least squares inverse is a special generalized inverse which defines the so called minimum-norm least squares solution for linear equation systems (see **Subsection 4.2.4**). The characteristics of this type of inverses are described by the following theorem.

Theorem 4.3

For any matrix $\underset{n\cdot m}{A}$, if S and R are two arbitrary matrices such that :

$$r(A) = r(RA) = r(AS) = r(RAS) = r \tag{4.27}$$

then the following g-inverse of A is unique:

$$A_{SR}^{-1} = S(RAS)^{-}R = S(AS)^{-}A(RA)^{-}R \tag{4.28}$$

Proof. We first apply **Theorem 4.2** to prove that A_{SR}^{-1} in Eq.(4.28) is a g-inverse of A:

$$\begin{aligned} A \cdot S(RAS)^{-}R \cdot A &= A(RA)^{-}R\ A = A\ A^{-}A = A \\ A \cdot S(AS)^{-}A(RA)^{-}R \cdot A &= A\ A^{-}A\ A^{-}A = A\ A^{-}A = A \ ^{1} \end{aligned}$$

We then prove that $A_{SR}^{-1} = S(RAS)^{-}R$ is unique. Let A_{SR}^{-1} and $\widetilde{A}_{SR}^{-1}$ denote a particular inverse and the general inverse as defined by Eq.(4.28), respectively. Let $(RAS)^{-}$ and $\widetilde{(RAS)}^{-}$ denote a particular and the general g-inverse of RAS, respectively. Then we have:

$$\begin{aligned} \widetilde{A}_{SR}^{-1} &= S \cdot \widetilde{(RAS)}^{-} \cdot R \\ &= S\left\{(RAS)^{-} + N[I - (RAS)(RAS)^{-}] + [I - (RAS)^{-}(RAS)]M\right\}R \\ &= S(RAS)^{-}R + SN\left[R - (RAS)(RAS)^{-}R\right] + \left[S - S(RAS)^{-}(RAS)\right]MR \\ &= A_{SR}^{-1} + SN[R - R\ R^{-}R] + [S - S\ S^{-}S]MR \\ &= A_{SR}^{-1} \end{aligned}$$

Similarly, one can prove $S(AS)^{-}A(RA)^{-}R$ is a unique g-inverse of A . Note that the uniqueness of A_{SR}^{-1} is with respect to the chosen S and R.

[1] Note that both $S(AS)^{-}$ and $(RA)^{-}R$ are $g-$inverses of A. But they are not necessarily the same $g-$inverse of A, even though both inverses are denoted by A^{-} here.

As a special case of **Theorem 4.3**, let $S = QA^\top$ and $R = A^\top P$, where $\underset{m\cdot m}{Q}$ and $\underset{n\cdot n}{P}$ are two positive definite matrices. Obviously, such S and R satisfy the condition in Eq.(4.27). By inserting S and R into Eq.(4.28), we get the unique minimum-norm least squares inverse of A:

$$A_{QP}^{-1} = QA^\top (A^\top PAQA^\top)^- A^\top P = QA^\top (AQA^\top)^- A(A^\top PA)^- A^\top P \tag{4.29}$$

It can be proved that the following relation holds:

$$A_{QP}^{-1} = A_{QI}^{-1}\, A\, A_{IP}^{-1} \tag{4.30}$$

To prove Eq.(4.30), we insert separately $Q = I$ and $P = I$ into Eq.(4.29) to obtain:

$$A_{QI}^{-1} = QA^\top (AQA^\top)^- A(A^\top A)^- A^\top$$
$$A_{IP}^{-1} = A^\top (AA^\top)^- A(A^\top PA)^- A^\top P$$

$$A_{QI}^{-1} A\, A_{IP}^{-1} = QA^\top (AQA^\top)^- A(A^\top A)^- A^\top \cdot A \cdot A^\top (AA^\top)^- A(A^\top PA)^- A^\top P$$
$$= QA^\top (AQA^\top)^- A \cdot A^- \cdot A \cdot A^- \cdot A(A^\top PA)^- A^\top P$$
$$= QA^\top (AQA^\top)^- A(A^\top PA)^- A^\top P$$
$$= A_{QP}^{-1}$$

Another special case is defined if we let $Q = \underset{m\cdot m}{I}$ and $P = \underset{n\cdot n}{I}$. Then we obtain the so called *pseudo-inverse* or *Moore-Penrose inverse*:

$$A_{II}^{-1} = A^\top (A^\top AA^\top)^- A^\top = A^\top (AA^\top)^- A(A^\top A)^- A^\top \tag{4.31}$$

In some literature, A_{II}^{-1} is denoted by A^+. The pseudo-inverse A^+ can also be defined by the following four conditions:

$$\left.\begin{array}{l} AA^+A = A \\ A^+AA^+ = A^+ \\ (AA^+)^\top = AA^+ \\ (A^+A)^\top = A^+A \end{array}\right\} \tag{4.32}$$

One can prove that the $g-$inverse of A given by Eq.(4.12) is a pseudo-inverse of A :

$$A^+ = C^\top (CC^\top)^{-1} (B^\top B)^{-1} B^\top \tag{4.33}$$

where B, C are obtained from rank factorization:

$$A = BC\ , \qquad [r(A) = r(B) = r(C) = r] \tag{4.34}$$

The proof is easily done by the fact that A^+ given by Eq.(4.33) satisfies all four conditions in Eq.(4.32).

4.1.5 Computation of Minimum-Norm Least Squares Inverses

For given P and Q, A_{QP}^{-1} can be computed directly from Eq.(4.29). In addition, it can be computed by several alternative methods outlined below (Sjöberg, 1991). For simplicity, detailed derivations are omitted.

Orthogonal Bordering

Let $\underset{n\cdot m}{A}$ have rank defect $d = \min(n, m) - r(A)$, and let P, Q be two positive definite square matrices of compatible dimensions. $\underset{n\cdot m}{A}$ can be computed according to Eq.(4.30):

$$A_{QP}^{-1} = A_{QI}^{-1}\ A\ \ A_{IP}^{-1} \tag{4.35}$$

with :

$$A_{QI}^{-1} = QA^\top (AQA^\top + CC^\top)^{-1} \tag{4.36}$$

$$A_{IP}^{-1} = (A^\top PA + D^\top D)^{-1} A^\top P \tag{4.37}$$

where $\underset{n\cdot k}{C}$ denotes a matrix such that :

$$\left.\begin{array}{l} A^\top C = 0 \\ \left|C^\top C\right| \neq 0 \\ \left|AQA^\top + CC^\top\right| \neq 0 \end{array}\right\} \tag{4.38}$$

and $\underset{k\cdot m}{D}$ denotes a matrix such that :

$$\left.\begin{array}{l} AD^\top = 0 \\ \left|DD^\top\right| \neq 0 \\ \left|A^\top PA + D^\top D\right| \neq 0 \end{array}\right\} \tag{4.39}$$

When determining matrices C and D, the first equation in (4.38) and (4.39) does not provide unique solutions to C and D, respectively. However, this should not affect the computation of A_{QP}^{-1}. In geodetic and photogrammetric applications, matrix D can be determined in advance for various types of geodetic and photogrammetric networks (see **Subsection 4.3.2**).

Limiting Value Method

By the limiting value method, the minimum-norm least-squares inverse A_{QP}^{-1} can be computed as the following limiting values:

$$A_{QP}^{-1} = \lim_{\delta \to 0} \left\{ QA^\top (AQA^\top + \delta\ P^{-1})^{-1} \right\} \tag{4.40}$$

$$A_{QP}^{-1} = \lim_{\delta \to 0} \left\{ (A^\top PA + \delta\ Q^{-1})^{-1} A^\top P \right\} \tag{4.41}$$

The above limiting values can be calculated either analytically (see **Example 4.5** below), or numerically for a very small $\delta\ (> 0)$.

Truncated Triangular Decomposition

We still use Eq.(4.30) to calculate A_{QP}^{-1}. But A_{QI}^{-1} and A_{IP}^{-1} are computed from the following equations:

$$A_{QI}^{-1} = QA^\top (AQA^\top)_{II}^{-1} \tag{4.42}$$

$$A_{IP}^{-1} = (A^\top PA)_{II}^{-1} A^\top P \tag{4.43}$$

where the two pseudo-inverses, $(AQA^\top)_{II}^{-1}$ and $(A^\top PA)_{II}^{-1}$, are calculated using Eq.(4.33).

To obtain $(AQA^\top)_{II}^{-1}$, we first decomposed $AQA^\top$ into the product of two triangular matrices using Cholesky's triangular decomposition method :

$$AQA^\top = T^\top T$$

After truncating zero rows in T and zero columns in $T^\top$, we have the truncated triangular decomposition:

$$AQA^\top = \overline{T}^\top \overline{T} \tag{4.44}$$

Then $(AQA^\top)_{II}^{-1}$ is obtained by Eq.(4.33):

$$\left(AQA^\top\right)_{II}^{-1} = \overline{T}^\top \left(\overline{T} \cdot \overline{T}^\top\right)^{-1} \left(\overline{T} \cdot \overline{T}^\top\right)^{-1} \overline{T} \tag{4.45}$$

$(A^\top PA)_{II}^{-1}$ can be obtained in a similar way.

Partitioning Method

The partition method can be applied in two ways :

- Assume that matrix $\underset{n\cdot m}{A}$ can be partitioned **by columns** into several sub-matrices $A_1, A_2, \cdots$, one of which (say A_i) has the same rank as A:

$$\underset{n\cdot m}{A} = \left[\begin{array}{cccccc} A_1 & A_2 & \cdots & A_i & \cdots \end{array}\right], \quad r(A) = r(A_i) \tag{4.46}$$

  then A_{QP}^{-1} can be computed from:

$$A_{QP}^{-1} = QB^\top (BQB^\top)^{-1} A_i^\top P \tag{4.47}$$

  where B is defined as:

$$B = A_i^\top PA \tag{4.48}$$

- Assume that matrix $\underset{n\cdot m}{A}$ can be partitioned **by rows** into several sub-matrices $A_1, A_2, \cdots$, one of which (say A_i) has the same rank as A:

$$\underset{n\cdot m}{A} = \left[\begin{array}{c} A_1 \\ A_2 \\ \cdots \\ A_i \\ \cdots \end{array}\right], \quad r(A) = r(A_i) \tag{4.49}$$

  then A_{QP}^{-1} can be computed from:

$$A_{QP}^{-1} = QA_i^\top (B^\top PB)^{-1} B^\top P \tag{4.50}$$

  where B is defined as:

$$B = AQA_i^\top \tag{4.51}$$

Iterative Method for computing A_{II}^{-1} or A^+

The iterative method for computing A^+ starts with the initial (approximate) solution for A^+ :

$$A_0^+ \approx \delta \cdot A^\top \tag{4.52}$$

where δ is a very small positive number. A correction for the approximate A_0^+ can be :

$$V_1 = A_0^+ \left(I - AA_0^+\right) \tag{4.53}$$

and the corrected solution for A^+ becomes:

$$A_1^+ = A_0^+ + V_1 \tag{4.54}$$

One can repeat the above correcting process. At $k-$th iteration step the correction is:

$$V_{k+1} = A_k^+(I - AA_k^+), \quad k = 0, 1, 2, 3, \cdots \tag{4.55}$$

and the new solution for A^+ is:

$$A_{k+1}^+ = A_k^+ + V_{k+1} \tag{4.56}$$

The iteration may stop when the norm of the last correction V_{k+1} is smaller than a pre-defined small number ε (> 0):

$$\|V_{k+1}\| = \sqrt{tr\left(V_{k+1}^\top V_{k+1}\right)} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}^2} < \varepsilon \tag{4.57}$$

where $tr\,(\cdot)$ denotes the trace operator and v_{ij} denotes elements of matrix V_{k+1} of dimensions $n \times m$.

Numerical tests indicated that the above iteration does not converge correctly for any arbitrary choice of the small positive number δ. Therefore the value of δ need to be chosen with great care.

Example 4.5

Let matrices A, P, Q be defined as follows:

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \quad Q = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$$

We want to calculate A_{QP}^{-1} and A_{II}^{-1} using the definition formulas as well as the five methods described before. As preparations, we calculated first the following matrices:

$$P^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A^\top P = \begin{bmatrix} 4 & 1 \\ 2 & \frac{1}{2} \end{bmatrix}, \quad A^\top PA = \begin{bmatrix} 18 & 9 \\ 9 & \frac{9}{2} \end{bmatrix}, \quad (A^\top PA)^- = \frac{1}{18}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$AQ = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}, \quad AQA^\top = \begin{bmatrix} 12 & 6 \\ 6 & 3 \end{bmatrix}, \quad (AQA^\top)^- = \frac{1}{12}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$A^\top A = 5\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} = AA^\top, \quad (A^\top A)^- = \frac{1}{20}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = (AA^\top)^-$$

1. *Using the definition formulas* :

$$A_{QP}^{-1} = QA^\top(AQA^\top)^- A(A^\top PA)^- A^\top P$$
$$= \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \cdot \frac{1}{12}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \cdot \frac{1}{18}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 1 \\ 2 & \frac{1}{2} \end{bmatrix} = \frac{1}{27}\begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$$

$$A_{II}^{-1} = A^\top (AA^\top)^- A (A^\top A)^- A^\top$$
$$= \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \cdot \tfrac{1}{20} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \cdot \tfrac{1}{20} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} = \tfrac{1}{25} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

2. *Using the method of orthogonal bordering*

Here we have: $r(A) = 1$ which implies $d = 2 - 1 = 1$. A_{QI}^{-1} can be computed as follows:

$$\underset{2\cdot 2}{A}\ \underset{2\cdot 1}{C} = 0 \ \rightarrow \ \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ \rightarrow c_2 = -2c_1 \quad \rightarrow \quad \underset{2\cdot 1}{C} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \ \left|C^\top C\right| = 5 \neq 0$$
$$AQA^\top + CC^\top = \begin{bmatrix} 12 & 6 \\ 6 & 3 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 13 & 4 \\ 4 & 7 \end{bmatrix}, \quad \left|AQA^\top + CC^\top\right| = 75 \neq 0$$
$$\left(AQA^\top + CC^\top\right)^{-1} = \tfrac{1}{15} \begin{bmatrix} 7 & -4 \\ -4 & 13 \end{bmatrix}$$
$$A_{QI}^{-1} = QA^\top \left(AQA^\top + CC^\top\right)^{-1} = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \cdot \tfrac{1}{15} \begin{bmatrix} 7 & -4 \\ -4 & 13 \end{bmatrix} = \tfrac{1}{15} \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

Then we compute A_{IP}^{-1} :

$$\underset{2\cdot 2}{A}\ \underset{2\cdot 1}{D^\top} = 0 \ \rightarrow \ \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ \rightarrow d_2 = -2d_1 \ \rightarrow \ \underset{1\cdot 2}{D} = \begin{bmatrix} -1 & 2 \end{bmatrix}, \ \left|DD^\top\right| = 5 \neq 0$$
$$A^\top PA + D^\top D = \begin{bmatrix} 18 & 9 \\ 9 & \frac{9}{2} \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 19 & 7 \\ 7 & \frac{17}{2} \end{bmatrix}, \quad \left|A^\top PA + D^\top D\right| = \tfrac{225}{2} \neq 0$$
$$\left(A^\top PA + D^\top D\right)^{-1} = \tfrac{1}{225} \begin{bmatrix} 17 & -14 \\ -14 & 38 \end{bmatrix}$$
$$A_{IP}^{-1} = \left(A^\top PA + D^\top D\right)^{-1} A^\top P = \tfrac{1}{225} \begin{bmatrix} 17 & -14 \\ -14 & 38 \end{bmatrix} \cdot \begin{bmatrix} 4 & 1 \\ 2 & \frac{1}{2} \end{bmatrix} = \tfrac{1}{45} \begin{bmatrix} 8 & 2 \\ 4 & 1 \end{bmatrix}$$

Finally we obtain:

$$A_{QP}^{-1} = A_{QI}^{-1} \cdot A \cdot A_{IP}^{-1} = \tfrac{1}{15} \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \cdot \tfrac{1}{45} \begin{bmatrix} 8 & 2 \\ 4 & 1 \end{bmatrix} = \tfrac{1}{27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$$

$$AA^\top + CC^\top = 5 \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 21 & 8 \\ 8 & 9 \end{bmatrix}$$
$$A_{II}^{-1} = A^\top \left(AA^\top + CC^\top\right)^{-1} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \cdot \tfrac{1}{125} \begin{bmatrix} 9 & -8 \\ -8 & 21 \end{bmatrix} = \tfrac{1}{25} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

3. *Using limiting value method:*

Let δ denote an arbitrary small constant. We first calculate A_{QP}^{-1} using Eq.(4.41):

$$A^\top PA + \delta Q^{-1} = \begin{bmatrix} 18 & 9 \\ 9 & \frac{9}{2} \end{bmatrix} + \delta \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2\delta + 18 & 9 \\ 9 & \delta + \frac{9}{2} \end{bmatrix}$$
$$\left|A^\top PA + \delta Q^{-1}\right| = (2\delta + 18)\left(\delta + \tfrac{9}{2}\right) - 9 \times 9 = 2\delta^2 + 27\delta$$
$$\left(A^\top PA + \delta Q^{-1}\right)^{-1} = \tfrac{1}{2\delta^2 + 27\delta} \begin{bmatrix} \delta + \frac{9}{2} & -9 \\ -9 & 2\delta + 18 \end{bmatrix}$$
$$\left(A^\top PA + \delta Q^{-1}\right)^{-1} A^\top P = \tfrac{1}{2\delta^2 + 27\delta} \begin{bmatrix} \delta + \frac{9}{2} & -9 \\ -9 & 2\delta + 18 \end{bmatrix} \cdot \begin{bmatrix} 4 & 1 \\ 2 & \frac{1}{2} \end{bmatrix}$$
$$= \tfrac{1}{2\delta^2 + 27\delta} \begin{bmatrix} 4\delta & \delta \\ 4\delta & \delta \end{bmatrix} = \tfrac{1}{2\delta + 27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$$
$$A_{QP}^{-1} = \lim_{\delta \to 0} \left\{ \left(A^\top PA + \delta Q^{-1}\right)^{-1} A^\top P \right\} = \lim_{\delta \to 0} \left\{ \tfrac{1}{2\delta + 27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix} \right\} = \tfrac{1}{27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$$

Then we compute A_{QP}^{-1} by Eq.(4.40):

$$AQA^\top + \delta P^{-1} = \begin{bmatrix} 12 & 6 \\ 6 & 3 \end{bmatrix} + \delta \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} \delta + 12 & 6 \\ 6 & 2\delta + 3 \end{bmatrix}$$

$$\left|AQA^\top + \delta P^{-1}\right| = (\delta + 12)(2\delta + 3) - 6 \times 6 = 2\delta^2 + 27\delta$$

$$\left(AQA^\top + \delta P^{-1}\right)^{-1} = \tfrac{1}{2\delta^2+27\delta} \begin{bmatrix} 2\delta + 3 & -6 \\ -6 & \delta + 12 \end{bmatrix}$$

$$QA^\top \left(AQA^\top + \delta P^{-1}\right)^{-1} = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \cdot \tfrac{1}{2\delta^2+27\delta} \begin{bmatrix} 2\delta + 3 & -6 \\ -6 & \delta + 12 \end{bmatrix} = \tfrac{1}{2\delta+27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$$

$$A_{QP}^{-1} = \lim_{\delta\to 0} \left\{ QA^\top \left(AQA^\top + \delta P^{-1}\right)^{-1} \right\} = \lim_{\delta\to 0} \left\{ \tfrac{1}{2\delta+27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix} \right\} = \tfrac{1}{27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$$

4. *Using truncated triangular decomposition :*

$$AQA^\top = \begin{bmatrix} 12 & 6 \\ 6 & 3 \end{bmatrix} = T^\top T = \begin{bmatrix} \sqrt{12} & 0 \\ \frac{1}{2}\sqrt{12} & 0 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{12} & \frac{1}{2}\sqrt{12} \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{12} \\ \frac{1}{2}\sqrt{12} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{12} & \frac{1}{2}\sqrt{12} \end{bmatrix} = \overline{T}^\top \cdot \overline{T}, \quad \overline{T} = \begin{bmatrix} \sqrt{12} & \frac{1}{2}\sqrt{12} \end{bmatrix}, \ \overline{T} \cdot \overline{T}^\top = 15, \ \left(\overline{T} \cdot \overline{T}^\top\right)^{-1} = \tfrac{1}{15}$$

$$A_{QI}^{-1} = QA^\top \overline{T}^\top \left(\overline{T} \cdot \overline{T}^\top\right)^{-1} \left(\overline{T} \cdot \overline{T}^\top\right)^{-1} \overline{T}$$

$$= \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{12} \\ \frac{1}{2}\sqrt{12} \end{bmatrix} \cdot \tfrac{1}{15} \cdot \tfrac{1}{15} \cdot \begin{bmatrix} \sqrt{12} & \frac{1}{2}\sqrt{12} \end{bmatrix} = \tfrac{1}{15} \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

$$A^\top PA = \begin{bmatrix} 18 & 9 \\ 9 & \frac{9}{2} \end{bmatrix} = T^\top T = \begin{bmatrix} \sqrt{18} & 0 \\ \frac{1}{2}\sqrt{18} & 0 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{18} & \frac{1}{2}\sqrt{18} \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{18} \\ \frac{1}{2}\sqrt{18} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{18} & \frac{1}{2}\sqrt{18} \end{bmatrix} = \overline{T}^\top \overline{T}, \ \overline{T} = \begin{bmatrix} \sqrt{18} & \frac{1}{2}\sqrt{18} \end{bmatrix}, \ \overline{T} \cdot \overline{T}^\top = \tfrac{45}{2}, \ \left(\overline{T} \cdot \overline{T}^\top\right)^{-1} = \tfrac{2}{45}$$

$$A_{IP}^{-1} = \overline{T}^\top \left(\overline{T} \cdot \overline{T}^\top\right)^{-1} \left(\overline{T} \cdot \overline{T}^\top\right)^{-1} \overline{T} A^\top P$$

$$= \begin{bmatrix} \sqrt{18} \\ \frac{1}{2}\sqrt{18} \end{bmatrix} \cdot \tfrac{2}{45} \cdot \tfrac{2}{45} \cdot \begin{bmatrix} \sqrt{18} & \frac{1}{2}\sqrt{18} \end{bmatrix} \cdot \begin{bmatrix} 4 & 1 \\ 2 & \frac{1}{2} \end{bmatrix} = \tfrac{1}{45} \begin{bmatrix} 8 & 2 \\ 4 & 1 \end{bmatrix}$$

$$A_{QP}^{-1} = A_{QI}^{-1} \cdot A \cdot A_{IP}^{-1} = \tfrac{1}{15} \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \cdot \tfrac{1}{45} \begin{bmatrix} 8 & 2 \\ 4 & 1 \end{bmatrix} = \tfrac{1}{27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$$

5. *Using the partitioning method :*

- *partitioning by columns*

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \quad r(A_1) = r(A) = 1$$

$$A_1^\top P = \begin{bmatrix} 4 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 4 & 1 \end{bmatrix}$$

$$B = A_1^\top PA = \begin{bmatrix} 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} = 9 \begin{bmatrix} 2 & 1 \end{bmatrix}$$

$$QB^\top = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \cdot 9 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 9 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$BQB^\top = 9 \cdot 9 \cdot 3 = 243$$

$$A_{QP}^{-1} = QB^\top (BQB^\top)^{-1} A_1^\top P = 9 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \tfrac{1}{243} \cdot \begin{bmatrix} 4 & 1 \end{bmatrix} = \tfrac{1}{27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$$

- *partioning by rows*

$A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 4 & 2 \end{bmatrix}, \quad r(A_1) = r(A) = 1$

$QA_1^\top = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

$B = AQA_1^\top = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 6 \end{bmatrix}$

$B^\top P = \begin{bmatrix} 12 & 6 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 12 & 3 \end{bmatrix}$

$B^\top PB = \begin{bmatrix} 12 & 3 \end{bmatrix} \cdot \begin{bmatrix} 12 \\ 6 \end{bmatrix} = 162$

$A_{QP}^{-1} = QA_1^\top \left(B^\top PB\right)^{-1} B^\top P = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \cdot \frac{1}{162} \cdot \begin{bmatrix} 12 & 3 \end{bmatrix} = \frac{1}{27} \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}$

6. *Using the iterative method:*

$A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}, \quad \| A \| = \sqrt{25} = 5$

$\delta = 0.03, \quad \varepsilon = 10^{-3}$

Initial solution:
$A_0^+ = \delta A^\top = \delta A = 0.03 \cdot A$

First iteration:
$V_1 = A_0^+ \left(I - AA_0^+\right) = \delta A \cdot (I - 5\delta A) = 0.007\,5 \cdot A$
$\| V_1 \| = 0.0075 \cdot \| A \| = 0.037\,5 > \varepsilon$
$A_1^+ = A_0^+ + V_1 = 0.037\,5 \cdot A$

Second iteration:
$V_2 = A_1^+ \left(I - AA_1^+\right) = 0.002\,343\,74 \cdot A$
$\| V_2 \| = 0.00234374 \cdot \| A \| = 0.011\,718\,7 > \varepsilon$
$A_2^+ = A_1^+ + V_2 = 0.039\,843\,74\ A$

Third iteration:
$V_3 = A_2^+ \left(I - AA_2^+\right) = 0.000\,155\,64\ A$
$\| V_3 \| = 0.000\,155\,64 \cdot \| A \| = 0.000\,778\,20\ < \varepsilon$

Final solution:
$A_{II}^{-1} = A^+ \approx A_2^+ = 0.039\,843\,74\ A \approx 0.04 \cdot A = \frac{1}{25} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}.$ ∎

4.2 Solutions of Linear Equation Systems

Using generalized matrix inverses, the solutions of any linear equation system:

$$\underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} = \underset{n\cdot 1}{L} \tag{4.58}$$

can be expressed by the general formula:

$$X = A^{-}L \tag{4.59}$$

where A^{-} denotes a generalized matrix inverse of A. Depending on the character of A, several types of solutions may be distinguished :

- $r(A) = n = m$: the system is non-singular and its solution can be obtained using the conventional inverse A^{-1}:
$$X = A^{-1}L \quad for\ |A| \neq 0 \tag{4.60}$$
- $r(A) \leq n < m$: the system is under-determined and there exist non-unique solutions, from which a minimum-norm solution can be constructed such that:
$$X^{\top}Q^{-1}X = \text{minimum} \tag{4.61}$$
where Q is a pre-defined, non-singular, square matrix.
- $r(A) = m < n$: the system is over-determined with full rank and most likely inconsistent due to data error. A unique least squares solution can be defined such that
$$(L - AX)^{\top} P (L - AX) = \text{minimum} \tag{4.62}$$
where P is a pre-defined, non-singular, square matrix.
- $r(A) < m < n$: the system is over-determined with rank defect, and most likely inconsistent due to data error. The least squares solution satisfying Eq.(4.62) is not unique. But the minimum-norm least-squares solution satisfying both Eq.(4.61) and Eq.(4.62) is unique.

4.2.1 Consistent Linear Equation Systems

The linear equation system, as given in Eq.(4.58), is said to be consistent if it has at least one set of solutions.

Theorem 4.4

Eq.(4.58) is consistent iff :

$$AA^{-}L = L \tag{4.63}$$

Proof. If $A\ A^{-}L = L$, then $A^{-}L$ is a solution of $AX = L$, which proves the sufficiency part. The necessity can be proved as follows: if $AX = L$ has a solution, say X_1, we then have $AX_1 = L$ which implies $AA^{-}A\ X_1 = L$ or $AA^{-}L = L$. ∎

Example 4.6

Consider the following linear equation system:

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \quad \Longleftrightarrow \quad AX = L$$

By direct calculations, we can find:

$$A^- = \tfrac{1}{3}\begin{bmatrix} 1 & 1 & 0 \\ -1 & 2 & 0 \end{bmatrix}$$

$$AA^- = \tfrac{1}{3}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ -1 & 2 & 0 \end{bmatrix}$$

$$AA^-L = \tfrac{1}{3}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ -1 & 2 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \neq L$$

According to the sufficiency and necessity condition Eq.(4.63), the above linear equation system is not consistent. This can be shown by looking at the solution of the first two equations :

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

which is inconsistent with the third equation which gives: $x_2 = 2$ ∎

Theorem 4.5

If the linear equation system in Eq.(4.58) is consistent, then its general solution or the complete set of solutions can be expressed as :

$$\widetilde{X} = A^-L + (I - A^-A)V = \widetilde{A}^-L \tag{4.64}$$

where A^-, $\widetilde{A}^-$ denote a particular and the general g-inverse of A, respectively, and V denotes an arbitrary vector.

Proof. Let $\widetilde{X}$ denote the general solution of $AX = L$ and let X_1 be defined as

$$X_1 = A^-L + (I - A^-A)V$$

That $AX = L$ is consistent implies $AA^-L = L$, which leads to:

$$AX_1 = AA^-L + (A - AA^-A)V = AA^-L = L$$

That is to say, X_1 is one particular solution of $AX = L$ or $X_1 \in \widetilde{X}$. Setting $V = \widetilde{X}$ in the defining equation of X_1 above gives:

$$X_1 = A^-L + (I - A^-A)\widetilde{X} = A^-L + \widetilde{X} - A^-L =$$

which implies: $\widetilde{X} \in X_1$. Thus we have $X_1 \in \widetilde{X} \in X_1$ which holds iff $X_1 = \widetilde{X}$.

To prove the second identity in Eq.(4.64), we simply multiply Eq.(4.14) by L :

$$\widetilde{A}^-L = \left[A^- + N(I - AA^-) + (I - A^-A)M\right]L = A^-L + N(I - AA^-)AX + (I - A^-A)ML$$

$$= A^-L + (I - A^-A)ML = A^-L + (I - A^-A)V$$

where $V = ML$. ∎

Example 4.7

Consider the following linear equation system:

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \Longleftrightarrow \quad AX = L$$

We then have:

$$A^- = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad AA^- = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}, \quad A^-A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad I - A^-A = \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix}$$

$$A^-L = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$AA^-L = \begin{bmatrix} 2 \\ 4 \end{bmatrix} = L \quad \rightarrow \text{the system is consistent !}$$

$$\widetilde{X} = A^-L + (I - A^-A)V = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2 - v_2 \\ v_2 \end{bmatrix}. \quad \blacksquare$$

4.2.2 Minimum-Norm Solution

Consider a consistent linear equation system:

$$\underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} = \underset{n\cdot 1}{L} \qquad [r(A) \leq n < m] \tag{4.65}$$

By Eq.(4.64), the general solution of the above system is obtained:

$$\widetilde{X} = A^-L + (I - A^-A)V$$

where A^- can be one particular g-inverse of A and V is an arbitrary vector. As the minimum-norm inverse A^-_{QO} ,

$$A^-_{QO} = QA^\top \left(AQA^\top\right)^- \tag{4.66}$$

is a g-inverse of A, we replace A^- in the general solution above by A^-_{QO} and obtain:

$$\widetilde{X} = A^-_{QO}L + (I - A^-_{QO}A)V \tag{4.67}$$

where Q is a pre-defined, non-singular square matrix. With respect to matrix Q, one can define the norm of vector $\widetilde{X}$ as the square root of the corresponding quadratic form:

$$\| \widetilde{X} \| = \sqrt{\widetilde{X}^\top Q^{-1} \widetilde{X}} \tag{4.68}$$

Inserting Eq.(4.67) into Eq.(4.68), we find :

$$\| \widetilde{X} \|^2 = \widetilde{X}^\top Q^{-1} \widetilde{X} = \left[A^-_{QO}L + (I - A^-_{QO}A)V\right]^\top Q^{-1} \left[A^-_{QO}L + (I - A^-_{QO}A)V\right]$$

$$= \left[A^-_{QO}L\right]^\top Q^{-1} \left[A^-_{QO}L\right] + \left[(I - A^-_{QO}A)V\right]^\top Q^{-1} \left[(I - A^-_{QO}A)V\right] + 2U$$

where U denotes the following cross-product :

$$U = (A^-_{QO}L)^\top Q^{-1}(I - A^-_{QO}A)V = L^\top \left[\left(AQA^\top\right)^-\right]^\top AQ \cdot Q^{-1} \cdot (I - A^-_{QO}A)V = 0$$

The norm of $\widetilde{X}$ becomes now :

$$\widetilde{X}^\top Q^{-1} \widetilde{X} = \left[A^-_{QO}L\right]^\top Q^{-1} \left[A^-_{QO}L\right] + \left[(I - A^-_{QO}A)V\right]^\top Q^{-1} \left[(I - A^-_{QO}A)V\right] \geq \left[A^-_{QO}L\right]^\top Q^{-1} \left[A^-_{QO}L\right]$$

where the identity holds if and only if $V = 0$. Hence, the minimum-norm solution $\widehat{X}$, i.e. the solution that has smallest norm according to Eq.(4.68), is obtained from Eq.(4.67) with $V = 0$:

$$\widehat{X} = A^-_{QO}L = QA^\top \left(AQA^\top\right)^- L \tag{4.69}$$

where A^-_{QO} is any arbitrary minimum-norm inverse of A and $(AQA^\top)^-$ is any arbitrary $g-$inverse of $AQA^\top$. Both A^-_{QO} and $(AQA^\top)^-$ are non-unique. But the minimum-norm solution $\widehat{X}$ is unique, which can be proved from the consistency condition Eq.(4.63) and the general expression for an arbitrary minimum-norm inverse given in Eq.(4.18):

$$\widetilde{A}^-_{QO} \cdot L = \left\{A^-_{QO} + N\left[I - A\ A^-_{QO}\right]\right\} \cdot L = A^-_{QO}L + N\left[I - A\ A^-_{QO}\right] \cdot AA^-L = A^-_{QO}L \tag{4.70}$$

When A has full rank, i.e. $r(A) = n$, $AQA^\top$ in Eq.(4.69) becomes non-singular and the minimum-norm solution $\widehat{X}$ can then be expressed using conventional matrix inverse:

$$\widehat{X} = QA^\top \left(AQA^\top\right)^{-1} L \qquad [r(A) = n < m] \tag{4.71}$$

Let us recall the condition adjustment with the following functional model:

$$\underset{t\cdot n}{B}\ \underset{n\cdot 1}{\varepsilon} = \underset{t\cdot 1}{W} \qquad [r(B) = t < n] \tag{4.72}$$

and the *a priori* statistical model:

$$E(\varepsilon) = 0, \quad Var(\varepsilon) = E\left(\varepsilon\varepsilon^\top\right) = \sigma_0^2 Q \tag{4.73}$$

where Q denotes the cofactor matrix of ε (or the observation vector). The least squares adjustment is to find estimate to ε such that Eq.(4.72) is satisfied and that $\varepsilon^\top Q^{-1}\varepsilon$ is minimized. Considering Eq.(4.68), we can immediately conclude that condition adjustment by least squares principle is equivalent to find a minimum-norm solution to the consistent linear equation system Eq.(4.72). From Eq.(4.71), we get the minimum-norm solution $\widehat{\varepsilon}$:

$$\widehat{\varepsilon} = B_{QO}^- W = QB^\top \left(BQB^\top\right)^{-1} W \qquad [r(B) = t < n] \tag{4.74}$$

where B_{QO}^- denotes the unique minimum-norm inverse of B :

$$B_{QO}^- = QB^\top \left(BQB^\top\right)^{-1} \qquad [r(B) = t < n] \tag{4.75}$$

The minimum-norm solution in Eq.(4.69) applies for any matrix A such that $r(A) \le n < m$. Consequently, the minimum-norm solution of the observation error ε should also exist when $r(B) < t < n$, which happens if the condition equations in Eq.(4.72) are linearly dependent or the number (t) of condition equations is larger than the number of over-determinations. In this case, condition adjustment can still be carried out with the help of generalized matrix inverses. According to Eq.(4.69), our least squares estimate (i.e. minimum-norm solution) $\widehat{\varepsilon}$ will be:

$$\widehat{\varepsilon} = B_{QO}^- W = QB^\top \left(BQB^\top\right)^- W \qquad [r(B) < t < n] \tag{4.76}$$

where B_{QO}^- denotes the non-unique minimum-norm inverse of B :

$$B_{QO}^- = QB^\top \left(BQB^\top\right)^- \qquad [r(B) < t < n] \tag{4.77}$$

As Eq.(4.70) shows, the minimum-norm solution $\widehat{\varepsilon}$ is unique, even though B_{QO}^- and $\left(BQB^\top\right)^-$ are not. This is demonstrated in **Example 4.8** below.

Example 4.8

In a simple levelling network with 3 benchmarks (Cf Fig. 4.1), four height differences have been measured and all measurements are assumed to have the same weight $p = 1$.

For this network, we have $n = 4$ observations, $m = 2$ necessary observations (one of the 3 benchmarks must be assumed to be fixed) and $t = n - m = 2$ independent condition equations. Based on the two non-overlapping closed loops, we can obtain the following condition equations:

$$\underset{2\cdot 4}{B}\ \underset{4\cdot 1}{\varepsilon} = \underset{2\cdot 1}{W} \qquad [r(B) = 2 = t < n = 4]$$

where:

$$B = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}, \quad L = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \end{bmatrix}, \quad W = BL = \begin{bmatrix} \ell_1 - \ell_2 \\ \ell_2 + \ell_3 + \ell_4 \end{bmatrix}$$

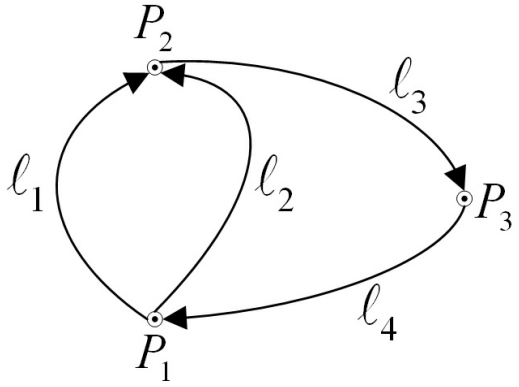

Figure 4.1: A simple levelling network

This is the full-rank case as shown in Eq.(4.72). Since the weight matrix of L can be regarded as a unit matrix and $BB^\top$ is non-singular, the least squares estimate of ε follows directly:

$$\begin{aligned}
\widehat{\varepsilon} &= B^\top(BB^\top)^{-1}W = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} \ell_1 - \ell_2 \\ \ell_2 + \ell_3 + \ell_4 \end{bmatrix} \\
&= \frac{1}{5}\begin{bmatrix} 3 & 1 \\ -2 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} \ell_1 - \ell_2 \\ \ell_2 + \ell_3 + \ell_4 \end{bmatrix} \\
&= \frac{1}{5}\begin{bmatrix} 3\ell_1 - 2\ell_2 + \ell_3 + \ell_4 \\ -2\ell_1 + 3\ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \end{bmatrix} \qquad (a)
\end{aligned}$$

If we use all three closed loops in Fig. 4.1, then we will have the following linearly correlated condition equations:

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} = \begin{bmatrix} \ell_1 - \ell_2 \\ \ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_3 + \ell_4 \end{bmatrix} \quad \Longleftrightarrow \quad \underset{3\cdot 4}{B}\ \underset{4\cdot 1}{\varepsilon} = \underset{3\cdot 1}{W}$$

Now we have $r(B) = 2 < t = 3$. Thus $BB^\top$ is singular and does not have a conventional inverse. The minimum-norm solution of ε is given by Eq.(4.69):

$$\widehat{\varepsilon} = B_{IO}^- W = B^\top(BB^\top)^- W$$

where:

$$BB^\top = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \quad \text{with } |BB^\top| = 0$$

The g-inverse $(BB^\top)^-$ is not unique and consequently the minimum norm inverse B_{IO}^- is not unique either. Below, we will choose three different inverses for $|BB^\top|^-$ to see whether they may affect the minimum norm solution $\widehat{\varepsilon}$.

The first choice of $(BB^\top)^-$ is obtained from Eq.(4.10):

$$(BB^\top)^- = \begin{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$B^-_{IO} = B^\top(BB^\top)^- = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tfrac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ -2 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$

$$\widehat{\varepsilon} = B^-_{IO}W = \tfrac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ -2 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}\begin{bmatrix} \ell_1 - \ell_2 \\ \ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_3 + \ell_4 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 3\ell_1 - 2\ell_2 + \ell_3 + \ell_4 \\ -2\ell_1 + 3\ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \end{bmatrix} \qquad (b)$$

The second choice of $(BB^\top)^-$ and the resulted estimate $\widehat{\varepsilon}$ are as follows:

$$(BB^\top)^- = \begin{bmatrix} 0 & 0 \\ 0 & \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^{-1} \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & -2 \\ 0 & -2 & 3 \end{bmatrix}$$

$$B^-_{IO} = B^\top(BB^\top)^- = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \tfrac{1}{5}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & -2 \\ 0 & -2 & 3 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 0 & -2 & 3 \\ 0 & 3 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\widehat{\varepsilon} = B^-_{IO}W = \tfrac{1}{5}\begin{bmatrix} 0 & -2 & 3 \\ 0 & 3 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \ell_1 - \ell_2 \\ \ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_3 + \ell_4 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 3\ell_1 - 2\ell_2 + \ell_3 + \ell_4 \\ -2\ell_1 + 3\ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \end{bmatrix} \qquad (c)$$

The third choice of $(BB^\top)^-$ is obtained after replacing A^- in Eq.(4.18) by the first choice of $(BB^\top)^-$ above, and letting N in Eq.(4.18) be $\frac{1}{5}\ I$ (unit matrix):

$$(BB^\top)^- = \tfrac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \tfrac{1}{5}I \cdot \left\{ I - \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix} \cdot \tfrac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} = \tfrac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ -1 & -1 & 1 \end{bmatrix}$$

$$B^-_{IO} = B^\top(BB^\top)^- = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \tfrac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ -1 & -1 & 1 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 2 & 0 & 1 \\ -2 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\hat{\epsilon} = B^-_{IO}W = \tfrac{1}{5}\begin{bmatrix} 2 & 0 & 1 \\ -2 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \ell_1 - \ell_2 \\ \ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_3 + \ell_4 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 3\ell_1 - 2\ell_2 + \ell_3 + \ell_4 \\ -2\ell_1 + 3\ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \end{bmatrix} \qquad (d)$$

Comparing the four solutions given in (a), (b), (c) and (d), we can see that condition adjustment using only 2 independent conditions gives the same solution to ε as using 3 linearly correlated conditions. In the latter case with 3 correlated condition equations, three different choices of the $g-$inverse for $BB^\top$ have been used and they all lead to exactly same minimum-norm solution $\widehat{\varepsilon}$, in consistency with the general conclusion expressed by Eq.(4.70). Furthermore, one can notice that the first choice of $(BB^\top)^-$ corresponds to condition adjustment using only the first two linearly independent conditions, while the second choice of $(BB^\top)^-$ implies the use of the last two linearly independent conditions.

4.2.3 Least Squares Solutions

Assume that we have an inconsistent linear equation system due to measurement errors (ε) in L :

$$\underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} \neq \underset{n\cdot 1}{L} \qquad [r(A) \leq m < n] \tag{4.78}$$

One can always make Eq.(4.78) consistent by adding the unknown vector ε to the right side of Eq.(4.78):

$$\underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} = \underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} \qquad [r(A) \leq m < n] \tag{4.79}$$

There is no unique solution to Eq.(4.79), as both X and ε are unknown. The least squares solution to X is so defined that the resulted residual vector ε satisfies:

$$\varepsilon^{\top} P \varepsilon = \text{minimum} \tag{4.80}$$

where P is a pre-defined positive definite, square matrix. To derive the least squares solution, we start from the general solution of Eq.(4.79) as given by Eq.(4.64) :

$$\widetilde{X} = A^{-}(L - \varepsilon) + (I - A^{-}A)V_x$$

where A^{-} is one particular $g-$inverse of A and V_x an arbitrary vector. Choosing A^{-} as the least squares inverse defined by Eq.(4.22), the general solution becomes:

$$\widetilde{X} = A_{OP}^{-}(L - \varepsilon) + (I - A_{OP}^{-}A)V_x \tag{4.81}$$

As $\widetilde{X}$ must satisfy Eq.(4.79), we obtain then:

$$L - \varepsilon = A\widetilde{X} = A \cdot A_{OP}^{-}(L - \varepsilon) + 0$$

or:

$$\left(I - AA_{OP}^{-}\right)\varepsilon = \left(I - AA_{OP}^{-}\right)L \tag{4.82}$$

The following identity :

$$\left(I - AA_{OP}^{-}\right) \cdot I \cdot \left(I - AA_{OP}^{-}\right) = \left(I - AA_{OP}^{-}\right) \tag{4.83}$$

implies that the unit matrix I is a $g-$inverse of the coefficient matrix $\left(I - AA_{OP}^{-}\right)$. Thus according to Eq.(4.64), the general solution to Eq.(4.82) can be written as :

$$\widetilde{\varepsilon} = I \cdot \left(I - AA_{OP}^{-}\right)L + \left[I - I \cdot \left(I - AA_{OP}^{-}\right)\right]V = \left(I - AA_{OP}^{-}\right)L + AA_{OP}^{-}V \tag{4.84}$$

where V is an arbitrary vector. The quadratic form of the above $\widetilde{\varepsilon}$ with respect to P can be found :

$$\begin{aligned}
\widetilde{\varepsilon}^{\top} P \widetilde{\varepsilon} &= \left[\left(I - AA_{OP}^{-}\right)L + AA_{OP}^{-}V\right]^{\top} P \left[\left(I - AA_{OP}^{-}\right)L + AA_{OP}^{-}V\right] \\
&= \left[\left(I - AA_{OP}^{-}\right)L\right]^{\top} P \left[\left(I - AA_{OP}^{-}\right)L\right] + \left[AA_{OP}^{-}V\right]^{\top} P \left[AA_{OP}^{-}V\right] \\
&\geq \left[\left(I - AA_{OP}^{-}\right)L\right]^{\top} P \left[\left(I - AA_{OP}^{-}\right)L\right]
\end{aligned}$$

where the identity holds iff $V = 0$, which gives the least squares solution of ε:

$$\widehat{\varepsilon} = \widetilde{\varepsilon}\,|_{V=0} = \left(I - AA_{OP}^{-}\right)L \tag{4.85}$$

The least squares solution to X is then obtained after substitution of $\widehat{\varepsilon}$ into Eq.(4.81):

$$\overline{X} = \widetilde{X}\,|_{\varepsilon=\widehat{\varepsilon}} = A_{OP}^{-} \cdot AA_{OP}^{-}L + (I - A_{OP}^{-}A)V_x = A_{OP}^{-}L + (I - A_{OP}^{-}A)V_x \tag{4.86}$$

where we have used the following relation:

$$A_{OP}^{-}\, A\, A_{OP}^{-} = A_{OP}^{-} \tag{4.87}$$

Because of V_x in Eq.(4.86), the least squares solution $\overline{X}$ is not uniquely determined. However, it can be shown that the least squares solution $\widehat{\varepsilon}$ for the residual is unique. This can be proved by replacing the

particular least squares inverse A_{OP}^{-} by the general expression of the least squares inverse as given in Eq.(4.23):

$$\left(I - A \cdot \widetilde{A}_{OP}^{-}\right) L = \left\{I - A \cdot \left[A_{OP}^{-} + \left(I - A_{OP}^{-} A\right) V_{\varepsilon}\right]\right\} L = \left(I - A A_{OP}^{-}\right) L = \widehat{\varepsilon} \tag{4.88}$$

where V_{ε} denotes an arbitrary vector.

In the case that A has full rank, *i.e.* $r(A) = m$, $A^{\top} P A$ becomes non-singular and A_{OP}^{-} is uniquely defined:

$$\left.\begin{aligned} A_{OP}^{-} &= A_{OP}^{-1} = \left(A^{\top} P A\right)^{-1} A^{\top} P \qquad \left(\left|A^{\top} P A\right| \neq 0\right) \\ A_{OP}^{-} A &= A_{OP}^{-1} A = I \end{aligned}\right\} \tag{4.89}$$

as well as the least squares solutions $\overline{X}$ and $\widehat{\varepsilon}$:

$$\left.\begin{aligned} \overline{X} &= A_{OP}^{-1} L = \left(A^{\top} P A\right)^{-1} A^{\top} P L \\ \widehat{\varepsilon} &= \left(I - A A_{OP}^{-1}\right) L = L - A \overline{X} \end{aligned}\right\} \tag{4.90}$$

One can observe that solving the linear equation system Eq.(4.79) under the condition of Eq.(4.80) is actually the same problem as adjustment by elements under the least squares principle, where L is the (reduced) observations, A is the design matrix, ε is the observation error and P the weight matrix of L. When A has full rank, the least squares solution of Eq.(4.79) and the results of adjustment by elements are identical, as given by Eq.(4.90). What makes the difference between the least squares solution using least-squares inverses and adjustment by elements is that, matrix A in Eqs.(4.79), (4.85) and (4.86) can be singular, i.e. $r(A) < m$. Therefore, by using the generalized matrix inverses A_{OP}^{-} we have now extended the traditional adjustment by elements to the case of rank-defected design matrix.

The cause to rank defect in A is often caused by the lack of reference datum, e.g. without any (or sufficient) necessary initial data in the network. These necessary initial data can be given coordinates of fixed points, fixed orientation (azimuth), fixed scale factor (fixed baseline). In this case, the unknown parameters $\overline{X}$ given in Eq.(4.86), which often are the coordinates of network points, cannot be uniquely determined. However, Eq.(4.88) shows that the least squares solutions $\widehat{\varepsilon}$ are unique, even if the absolute coordinates $\overline{X}$ are not unique. This is due to the fact that the residuals ε reflect only errors in the measurements (angles, distances, height differences) which are independent of the absolute position and orientation of the network. One may find many different solutions to X, which all satisfy the internal geometry defined by the measurements (angles, distances, height differences, etc.), corresponding to the different choices of the generalized inverse A_{OP}^{-}.

Among all least squares solutions $\overline{X}$, one can identify special unique solutions by putting extra conditions on the solution. One such unique solution is the so called minimum-norm least squares solution, to be described in the next subsection.

Example 4.9

We consider the adjustment of the levelling network (shown in Fig. 4.1 and 4.2 below) using the method of adjustment by elements.

If no fixed point exists for defining the height reference datum, traditional least squares adjustment is to assume one benchmark, say P_1, as fixed with given height (say $x_1 = 0$). Choosing the heights (x_2, x_3) of P_2 and P_3 as unknown parameters, our observation equations become:

$$\begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \end{bmatrix} - \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \quad \Longleftrightarrow \quad \underset{4\cdot 1}{L} - \underset{4\cdot 1}{\varepsilon} = \underset{4\cdot 2}{A}\ \underset{2\cdot 1}{X}$$

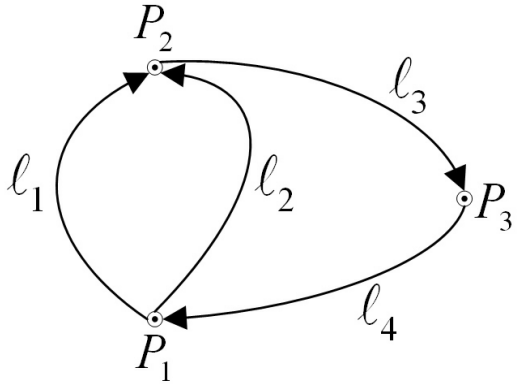

Figure 4.2: A simple levelling network

As in **Example 4.8**, the weight matrix of the observations L is take as the unit matrix:

$$\underset{4\times4}{P} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

As $r(A) = 2$, a unique least squares solution $\widehat{X}$ can be obtained :

$$(A^\top A)^{-1}A^\top = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}^{-1} \times \begin{bmatrix} 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 2 & 2 & -1 & -1 \\ 1 & 1 & 2 & -3 \end{bmatrix}$$

$$\widehat{X} = (A^\top A)^{-1}A^\top L = \tfrac{1}{5}\begin{bmatrix} 2\ell_1 + 2\ell_2 - \ell_3 - \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 - 3\ell_4 \end{bmatrix}$$

$$I - A\cdot(A^\top A)^{-1}A^\top = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \tfrac{1}{5}\begin{bmatrix} 2 & 2 & -1 & -1 \\ 1 & 1 & 2 & -3 \end{bmatrix} = \tfrac{1}{5}\begin{bmatrix} 3 & -2 & 1 & 1 \\ -2 & 3 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}$$

$$\widehat{\varepsilon} = L - A\widehat{X} = \left[I - A\cdot(A^\top A)^{-1}A^\top\right]L = \tfrac{1}{5}\begin{bmatrix} 3\ell_1 - 2\ell_2 + \ell_3 + \ell_4 \\ -2\ell_1 + 3\ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \end{bmatrix} \qquad (a)$$

The variance-covariance matrix of $\widehat{X}$ is as follows:

$$C_{\widehat{X}\widehat{X}} = \sigma_0^2\cdot(A^\top A)^{-1} = \sigma_0^2\cdot\frac{1}{5}\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

As P_1 is assumed to be fixed, its height x_1 can be regarded as having zero variance and zero covariances with x_2 and x_3. Let Y denote vector $(x_1, \widehat{x}_2, \widehat{x}_3)^\top$ and we then have:

$$Y = \begin{bmatrix} x_1 \\ \widehat{x}_2 \\ \widehat{x}_3 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 0 \\ 2\ell_1 + 2\ell_2 - \ell_3 - \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 - 3\ell_4 \end{bmatrix} \qquad (b)$$

$$C_{YY} = \sigma_0^2\cdot\frac{1}{5}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix} \qquad (c)$$

and the total variance of all three heights is the trace of C_{YY} :

$$tr(C_{YY}) = \sigma_0^2\cdot\frac{0+2+3}{5} = \sigma_0^2 \qquad (d)$$

If we choose the heights of all three benchmarks as unknown parameters, the observations become:

$$\begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \end{bmatrix} - \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \Longleftrightarrow \quad \underset{4\cdot 1}{L} - \underset{4\cdot 1}{\varepsilon} = \underset{4\cdot 3}{A}\ \underset{3\cdot 1}{X}$$

Now we have a singular design matrix A :

$$r(A) = 2 < m = 3$$

A particular least squares inverse of A can be found (see **Example 4.2**):

$$A^\top A = \begin{bmatrix} -1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & -2 & -1 \\ -2 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad \left|A^\top A\right| = 0$$

$$\left(A^\top A\right)^- = \begin{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \tfrac{1}{5} \begin{bmatrix} 3 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A_{OI}^- = \left(A^\top A\right)^- A^\top = \tfrac{1}{5} \begin{bmatrix} 3 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \tfrac{1}{5} \begin{bmatrix} -1 & -1 & -2 & 3 \\ 1 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Using A_{OI}^- , a particular least squares solution of X can be obtained:

$$\overline{X}' = A_{OI}^- L = \tfrac{1}{5} \begin{bmatrix} -\ell_1 - \ell_2 - 2\ell_3 + 3\ell_4 \\ \ell_1 + \ell_2 - 3\ell_3 + 2\ell_4 \\ 0 \end{bmatrix}$$

The complete set of all possible least squares solutions can be obtained from Eq.(4.86):

$$I - A_{OI}^- A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \tfrac{1}{5} \begin{bmatrix} -1 & -1 & -2 & 3 \\ 1 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\widetilde{X} = A_{OI}^- L + (I - A_{OI}^- A) V_x = \overline{X}' + (I - A_{OI}^- A) \cdot V_x = \overline{X}' + \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

$$= \tfrac{1}{5} \begin{bmatrix} -\ell_1 - \ell_2 - 2\ell_3 + 3\ell_4 \\ \ell_1 + \ell_2 - 3\ell_3 + 2\ell_4 \\ 0 \end{bmatrix} + \begin{bmatrix} w \\ w \\ w \end{bmatrix}$$

where $V_x = (u,\, v,\, w)^\top$ denotes an arbitrary vector.

With different choices of the arbitrary constant w, we get accordingly different least squares solutions. Below are two examples of such solutions:

- For $w = \frac{1}{5}(\ell_1 + \ell_2 + 2\ell_3 - 3\ell_4)$, we have :

$$\overline{X}_1 = \frac{1}{5} \begin{bmatrix} 0 \\ 2\ell_1 + 2\ell_2 - \ell_3 - \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 - 3\ell_4 \end{bmatrix} \qquad (e)$$

  which corresponds to assuming benchmark P_1 as fixed with zero height [see Eq.(b) above];

- For $w = 0$, we have :

$$\overline{X}_2 = \frac{1}{5}\begin{bmatrix} -\ell_1 - \ell_2 - 2\ell_3 + 3\ell_4 \\ \ell_1 + \ell_2 - 3\ell_3 + 2\ell_4 \\ 0 \end{bmatrix} \qquad (f)$$

  which corresponds to assuming P_3 as fixed with zero height;

The estimated residual $\widehat{\varepsilon}$ can be uniquely obtained from Eq.(4.88) :

$$\widehat{\varepsilon} = L - A\overline{X}_1 = L - A\overline{X}_2 = L - A\overline{X}' = \left(I - AA_{OI}^{-}\right) \cdot L$$

$$= \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \cdot \frac{1}{5} \begin{bmatrix} -1 & -1 & -2 & 3 \\ 1 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right\} \cdot \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \end{bmatrix}$$

$$= \frac{1}{5}\begin{bmatrix} 3 & -2 & 1 & 1 \\ -2 & 3 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix} \cdot \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 3\ell_1 - 2\ell_2 + \ell_3 + \ell_4 \\ -2\ell_1 + 3\ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \end{bmatrix} \qquad (g)$$

The above results have confirmed that when the design matrix A is singular, the least squares solution of the unknown parameters is not unique; while the residuals $\widehat{\varepsilon}$ can be still uniquely determined and they are identical to the results of traditional adjustment without datum defect [compare Eq.(g) with Eq.(a) above]. The use of generalized inverses makes it possible to define the complete set of all least squares solutions. The two different choices of generalized inverses in this practical example, simply a mathematical treatment, have implicitly defined two different height datums for the whole network. In **Example 4.10** at the end of Section 4.3, another generalized inverse will be chosen, resulting in a third height datum. This reminds us that any mathematical treatment may have certain physical or geodetic implications. It could be dangerous if one blindly applies mathematic theory without clearly understanding its geodetic meanings.

4.2.4 Minimum-Norm Least-Squares Solution

In Subsection 4.2.3, we have shown that the complete set of least squares solutions to the linear equation system:

$$\underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} = \underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon}, \qquad [r(A) \le m < n] \tag{4.91}$$

under the condition:

$$\varepsilon^{\top} P \varepsilon = \text{minimum} \tag{4.92}$$

are:

$$\overline{X} = A_{OP}^{-} L + (I - A_{OP}^{-} A) V_x \tag{4.93}$$

where V_x is an arbitrary vector and A_{OP}^{-} denotes a particular least squares inverse of A :

$$A_{OP}^{-} = \left(A^{\top} P A\right)^{-} A^{\top} P \tag{4.94}$$

Obviously, the solutions $\overline{X}$ in Eq.(4.93) are not unique, as A_{OP}^{-} is not unique when $r(A) < m$. Among all possible least squares solution $\overline{X}$, a unique solution $\widehat{X}$ can be defined by assuming an extra constraint :

$$\widehat{X}^{\top} Q^{-1} \widehat{X} = minimum \tag{4.95}$$

where Q is a given symmetrical positive definite matrix. This unique solution $\widehat{X}$ which satisfies both Eq.(4.92) and Eq.(4.95) is called the *minimum-norm least-squares solution* to the consistent linear equation system Eq.(4.91) or the inconsistent linear equation system Eq.(4.78).

Multiplying both sides of Eq.(4.93) by A, we obtain a new consistent equation system which contains only $\overline{X}$ as unknowns :

$$A\overline{X} = AA_{OP}^{-}L \tag{4.96}$$

The minimum norm-least squares solution $\widehat{X}$ of Eq.(4.91) can then be obtained as the minimum-norm solution of Eq.(4.96). From Eqs.(4.65), (4.68) and (4.69) in Subsection 4.2.2, we get directly:

$$\widehat{X} = A_{QO}^{-} \cdot AA_{OP}^{-}L = QA^{\top}(AQA^{\top})^{-} \cdot A(A^{\top}PA)^{-}A^{\top}P\ L = A_{QP}^{-1} \cdot L \tag{4.97}$$

where A_{QP}^{-1} is the unique minimum-norm least-squares inverse of A :

$$A_{QP}^{-1} = A_{QO}^{-}AA_{OP}^{-} = QA^{\top}(AQA^{\top})^{-}A(A^{\top}PA)^{-}A^{\top}P \tag{4.98}$$

The solution $\widehat{\varepsilon}$ for the residual vector, which is always unique, follows immediately :

$$\widehat{\varepsilon} = L - A\overline{X} = L - A\widehat{X} = (I - AA_{QP}^{-1})L \tag{4.99}$$

and finally the adjusted observations $\widehat{L}$ can be computed as:

$$\widehat{L} = L - \widehat{\varepsilon} = A\widehat{X} = AA_{QP}^{-1}L \tag{4.100}$$

The variance-covariance matrices of $\widehat{X}$ and $\widehat{\varepsilon}$ can be obtained by applying error propagation law on Eqs.(4.97), (4.99) and (4.100), respectively:

$$C_{\widehat{X}\widehat{X}} = A_{QP}^{-1} \cdot \sigma_0^2 P^{-1} \cdot \left(A_{QP}^{-1}\right)^{\top} \tag{4.101}$$

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = (I - AA_{QP}^{-1}) \cdot \sigma_0^2 P^{-1} \cdot (I - AA_{QP}^{-1})^{\top} = \sigma_0^2 \cdot \left(P^{-1} - AA_{QP}^{-1}A^{\top}\right) \tag{4.102}$$

$$C_{\widehat{L}\widehat{L}} = AA_{QP}^{-1} \cdot \sigma_0^2 \cdot \left(AA_{QP}^{-1}\right)^{\top} = \sigma_0^2 \cdot AA_{QP}^{-1}A^{\top} \tag{4.103}$$

where σ_0^2 denotes the theoretical variance factor, which can be replaced by the a posteriori estimate:

$$\widehat{\sigma}_0^2 = \frac{\widehat{\varepsilon}^{\top}P\widehat{\varepsilon}}{n-m} \tag{4.104}$$

If $Q = I$, which may be interpreted as if all components of X have equal weights, the minimum-norm least-squares solution becomes:

$$\left.\begin{array}{l} \widehat{X} = A_{IP}^{-1}L \\ \widehat{\varepsilon} = L - A\widehat{X} = (I - AA_{IP}^{-1})L \end{array}\right\} \tag{4.105}$$

The unique generalized inverse A_{IP}^{-1} used above can be calculated either using the general definition formulas in Eq.(4.29), or some other methods outlined in **Subsection 4.1.5**. For instance, A_{IP}^{-1} can be calculated using the orthogonal bordering method (see **Subsection 4.1.5**) :

$$A_{IP}^{-1} = (A^{\top}PA + D^{\top}D)^{-1}A^{\top}P \tag{4.106}$$

where D is defined such that :

$$\left.\begin{array}{l} AD^{\top} = 0 \\ \left|DD^{\top}\right| \neq 0 \\ \left|A^{\top}PA + D^{\top}D\right| \neq 0 \end{array}\right\} \tag{4.107}$$

Let $U = A^{\top}PA + D^{\top}D$ and we have :

$$DU = (AD^{\top})^{\top}PA + DD^{\top}D = DD^{\top}D$$

from which, one can get :

$$DU^{-1} = (DD^{\top})^{-1}D \tag{4.108}$$

Combining Eqs.(4.105) with Eqs.(4.106) and (4.108), the following relation can be derived :

$$D\widehat{X} = DU^{-1} \cdot A^{\top}PL = (DD^{\top})^{-1}D \cdot A^{\top}PL = (DD^{\top})^{-1} \cdot \left(AD^{\top}\right)^{\top} \cdot PL$$

Considering the second equation in (4.107), the above relation becomes:

$$D\widehat{X} = 0 \tag{4.109}$$

Eq.(4.109) describes the conditions that the minimum-norm least-squares solution $\widehat{X}$ automatically satisfies.

Further more, if P is also a unit matrix, our minimum-norm least-squares solution can then be expressed by the pseudo inverse $A^{+} = A_{II}^{-1}$:

$$\left.\begin{array}{l} \widehat{X} = A^{+}L \\ \widehat{\varepsilon} = L - A\widehat{X} = (I - AA^{+})L \end{array}\right\} \tag{4.110}$$

4.3 Free Network Adjustment

4.3.1 Free Networks

In geodetic or photogrammetric problems, the unknown parameters in adjustment by elements are often chosen as the coordinates of network points. The absolute position, orientation and scale of the network concerned are defined by the necessary initial data, i.e. datum parameters defining the coordinate reference system in which all network points are defined. Networks, which lack sufficient initial datum parameters, are called *free networks*. A completely free network may freely translate, rotate and change in network scale.

For a completely free network, the design matrix $\underset{n\cdot m}{A}$ in the observation equations

$$L - \varepsilon = \underset{n\cdot m}{A}\, X \tag{4.111}$$

will be singular, i.e.

$$r(\underset{n\cdot m}{A}) = r < m < n \tag{4.112}$$

The rank defect of A :

$$d = m - r \tag{4.113}$$

is actually the same as the datum defect of the free network. For completely free networks that has no initial data at all, the number (d) of rank defect (or datum defect) is equal to the number of necessary initial data (see **Subsection 2.2.1**). Under the least squares principle:

$$\varepsilon^\top P\varepsilon = \text{ minimum}$$

the coefficient matrix $A^\top PA$ of the normal equation will be singular, as $r(A^\top PA) = r(A)$ for any positive definite weight matrix P. Consequently, the least squares solution $\overline{X}$ given by Eq.(4.81) is not unique, although the least squares estimate $\widehat{\varepsilon}$ of the residuals given in Eq.(4.85) is unique. By applying the minimum-norm condition :

$$\widehat{X}^\top Q^{-1}\widehat{X} = \text{minimum}$$

the unique minimum-norm least-squares solution $\widehat{X}$ can then be obtained. For $Q = I$, $\widehat{X}$ can be written as [see (4.106)] :

$$\widehat{X} = A_{IP}^{-1}L = (A^\top PA + D^\top D)^{-1}A^\top PL \tag{4.114}$$

where D is defined by Eqs.(4.107) and has dimension $d \times m$.

In **Subsection 4.2.4**, it has been proved that the derived minimum-norm least-squares solution $\widehat{X}$ also satisfies:

$$\underset{d\cdot m}{D}\ \underset{m\cdot 1}{\widehat{X}} = 0 \tag{4.115}$$

In order to calculate $\widehat{X}$ using (4.114) and understand the geometrical and geodetic meaning of Eq.(4.115), we need to find out the structure of this matrix for geodetic and photogrammetric networks.

4.3.2 Structure of Matrix D

From the defining equations (4.107), one can see that matrix D must be dependent on the structure of the design matrix A, i.e. the configuration of the network. This gives us a possibility to find out the elements of D for different types of free networks where the coordinates of all unknown points are treated as unknown parameters in the observation equations (4.91).

One of the simplest free networks in geodesy and surveying may be levelling networks without any fixed benchmark. As the rank defect of one-dimensional free levelling network is equal to $d = 1$, matrix D will then be a row vector :

$$\underset{1\cdot m}{D} = \begin{bmatrix} d_1 & d_2 & d_3 & \cdots & d_m \end{bmatrix}$$

where m denotes the number of unknown parameters (equal to the number of unknown benchmarks). From Eq.(3.76) in **Subsection 3.2.3**, we find that any levelled height difference observation ℓ_{ij} between unknown benchmarks P_i and P_j has the following observation equations:

$$\ell_{ij} - \varepsilon_{ij} = x_j - x_i$$

In other words, each row in the A matrix of a free levelling network consists mostly zero elements except two. These two non-zero elements are $+1$ and -1, respectively. The second equation in Eq.(4.107) means the product of each row in A with the column vector $D^\top$ must be zero. This can hold iff all elements d_i are equal to a constant, say d_0. The first equation in Eq.(4.107) requires that this constant d_0 should be non-zero. Therefore, one may choose the simplest numerical value for d_i as just 1, i.e.

$$d_i = 1 \quad (i = 1, 2, 3, \cdots, m)$$

Similarly one can analyse the observation equations of other types of free networks. below we summarize the expressions of matrix D for five types of networks:

- *Levelling networks (1-D) with $d = 1$:*

$$\underset{1\cdot m}{D} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \tag{4.116}$$

  where m denotes the number of unknown points in the levelling network.

- *2-D triangulation networks with distance measurements ($d = 3$) :*

$$\underset{3\cdot m}{D} = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\ -y_1^0 & x_1^0 & -y_2^0 & x_2^0 & \cdots & -y_k^0 & x_k^0 \end{bmatrix} \tag{4.117}$$

  where m is the number of unknown parameters, $k = m/2$ is the number of unknown points, and $(x_i^0,\, y_i^0)$ denotes the approximate coordinates of the $i-th$ point $(i = 1, 2, \cdots, k)$.

- *2-D triangulation networks with angle measurements ($d = 4$)* :

$$\underset{4\cdot m}{D} = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\ -y_1^0 & x_1^0 & -y_2^0 & x_2^0 & \cdots & -y_k^0 & x_k^0 \\ x_1^0 & y_1^0 & x_2^0 & y_2^0 & \cdots & x_k^0 & y_k^0 \end{bmatrix} \tag{4.118}$$

  where m is the number of unknown parameters, $k = m/2$ is the number of unknown points, and $(x_i^0,\, y_i^0)$ denotes the approximate coordinates of the $i-th$ point $(i = 1, 2, \cdots, k)$.

- *2-D triangulation networks with direction measurements ($d = 4$)* :

$$\underset{4\cdot m}{D} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & -y_1^0 & x_1^0 & -y_2^0 & x_2^0 & \cdots & -y_k^0 & x_k^0 \\ 0 & 0 & \cdots & 0 & x_1^0 & y_1^0 & x_2^0 & y_2^0 & \cdots & x_k^0 & y_k^0 \end{bmatrix} \tag{4.119}$$

  where m is the number of unknown parameters, $k = (m - k_1)/2$ is the number of unknown points, k_1 is the number of theodolite station, and $(x_i^0,\, y_i^0)$ denotes the approximate coordinates of the $i-th$ point $(i = 1, 2, \cdots, k)$. The first k_1 columns refer to the orientation angles at theodolite stations, which are a part of the unknown parameter vector X.

- *3-D free network without distance measurements ($d = 7$) :*

$$\underset{7\cdot m}{D} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & \cdots & 0 & 0 & 1 \\ 0 & -z_1^0 & y_1^0 & 0 & -z_2^0 & y_2^0 & \cdots & 0 & -z_k^0 & y_k^0 \\ z_1^0 & 0 & -x_1^0 & z_2^0 & 0 & -x_2^0 & \cdots & z_k^0 & 0 & -x_k^0 \\ -y_1^0 & x_1^0 & 0 & -y_2^0 & x_2^0 & 0 & \cdots & -y_k^0 & x_k^0 & 0 \\ x_1^0 & y_1^0 & z_1^0 & x_2^0 & y_2^0 & z_2^0 & \cdots & x_k^0 & y_k^0 & z_k^0 \end{bmatrix} \tag{4.120}$$

where m is the number of unknown parameters, $k = m/3$ is the number of unknown points, and $(x_i^0,\, y_i^0, z_i^0)$ denotes the approximate coordinates of the $i-th$ point $(i = 1, 2, \cdots, k)$.

The seven initial data in a three -dimensional free network correspond to a 3-D Helmert transformation with seven transformation parameters. Following this idea, a detailed proof of Eq.(4.120) will be given below. Let $(\delta x_i, \delta y_i, \delta z_i)$ denote the coordinate corrections to the approximate coordinates $(x_i^0,\, y_i^0, z_i^0)$ of point i $(i = 1, 2, 3 \cdot \cdot, k)$. For this 3-D free network, the minimum norm condition when $Q = I$ becomes:

$$F = \widehat{X}^\top \widehat{X} = \sum_{i=1}^{k} \left\{ \delta x_i^2 + \delta y_i^2 + \delta z_i^2 \right\} = \text{minimum} \tag{4.121}$$

The new coordinates $(x_i^0 + \delta x_i, y_i^0 + \delta y_i, z_i^0 + \delta z_i)$ of point i after network adjustment can be written as a Helmert transformation of the approximate coordinates $(x_i^0,\, y_i^0, z_i^0)$. The seven Helmert transformation parameters are the three translation parameters (x_0, y_0, z_0), three small rotation angles $(\delta_1, \delta_2, \delta_3)$ about the $x-$, $y-$ and $z-$axis respectively, and finally a small scale change s :

$$\begin{bmatrix} x_i^0 + \delta x_i \\ y_i^0 + \delta y_i \\ z_i^0 + \delta z_i \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + (1+s) \begin{bmatrix} 1 & \delta_3 & -\delta_2 \\ -\delta_3 & 1 & \delta_1 \\ \delta_2 & -\delta_1 & 1 \end{bmatrix} \begin{bmatrix} x_i^0 \\ y_i^0 \\ z_i^0 \end{bmatrix} \tag{4.122}$$

Omitting the second-order terms $s \cdot \delta_i$ $(i = 1, 2, 3)$, the above equation becomes:

$$\begin{bmatrix} \delta x_i \\ \delta y_i \\ \delta z_i \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + \begin{bmatrix} 0 & \delta_3 & -\delta_2 \\ -\delta_3 & 0 & \delta_1 \\ \delta_2 & -\delta_1 & 0 \end{bmatrix} \begin{bmatrix} x_i^0 \\ y_i^0 \\ z_i^0 \end{bmatrix} + s \cdot \begin{bmatrix} x_i^0 \\ y_i^0 \\ z_i^0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & -z_i^0 & y_i^0 & x_i^0 \\ 0 & 1 & 0 & z_i^0 & 0 & -x_i^0 & y_i^0 \\ 0 & 0 & 1 & -y_i^0 & x_i^0 & 0 & z_i^0 \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ y_0 \\ z_0 \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ s \end{bmatrix} \tag{4.123}$$

or:

$$\left.\begin{aligned} \delta x_i &= x_0 - z_i^0 \cdot \delta_2 + y_i^0 \cdot \delta_3 + s \cdot x_i^0 \\ \delta y_i &= y_0 + z_i^0 \cdot \delta_1 - x_i^0 \cdot \delta_3 + s \cdot y_i^0 \\ \delta z_i &= z_0 - y_i^0 \cdot \delta_1 + x_i^0 \cdot \delta_2 + s \cdot z_i^0 \end{aligned}\right\} \tag{4.124}$$

The minimum norm in Eq.(4.121) can be achieved if we insert Eq.(4.124) into Eq.(4.121) and let the derivatives of F with respect to each of the seven transformation parameters be zero:

$$\frac{\partial F}{\partial x_0} = \sum_{i=1}^{k} \left\{ 2\delta x_i \cdot \frac{\partial \delta x_i}{\partial x_0} \right\} = 2 \sum_{i=1}^{k} \delta x_i = 0$$

$$\frac{\partial F}{\partial y_0} = 2 \sum_{i=1}^{k} \delta y_i = 0$$

$$\frac{\partial F}{\partial z_0} = 2\sum_{i=1}^{k} \delta z_i = 0$$

$$\frac{\partial F}{\partial \delta_1} = \sum_{i=1}^{k} \left\{ 2\delta y_i \frac{\partial \delta y_i}{\partial \delta_1} + 2\delta z_i \frac{\partial \delta z_i}{\partial \delta_1} \right\} = -2\sum_{i=1}^{k} \left\{ -z_i^0 \delta y_i + y_i^0 \delta z_i \right\} = 0$$

$$\frac{\partial F}{\partial \delta_2} = \sum_{i=1}^{k} \left\{ 2\delta x_i \frac{\partial \delta x_i}{\partial \delta_2} + 2\delta z_i \frac{\partial \delta z_i}{\partial \delta_2} \right\} = -2\sum_{i=1}^{k} \left\{ +z_i^0 \delta x_i - x_i^0 \delta z_i \right\} = 0$$

$$\frac{\partial F}{\partial \delta_3} = \sum_{i=1}^{k} \left\{ 2\delta x_i \frac{\partial \delta x_i}{\partial \delta_3} + 2\delta y_i \frac{\partial \delta y_i}{\partial \delta_3} \right\} = -2\sum_{i=1}^{k} \left\{ -y_i^0 \delta x_i + x_i^0 \delta y_i \right\} = 0$$

$$\frac{\partial F}{\partial s} = \sum_{i=1}^{k} \left\{ 2\delta x_i \frac{\partial \delta x_i}{\partial s} + 2\delta y_i \frac{\partial \delta y_i}{\partial s} + 2\delta z_i \frac{\partial \delta z_i}{\partial s} \right\} = 2\sum_{i=1}^{k} \left\{ x_i^0 \delta x_i + y_i^0 \delta y_i + z_i^0 \delta z_i \right\} = 0$$

Neglecting constants $+2$ and -2 before the summation signs in the above seven equations and rewriting all equations in matrix notations will lead finally to:

$$\underset{7\cdot m}{D}\ \underset{m\cdot 1}{\widehat{X}} \ = \ 0$$

where D is given in Eq.(4.120).

4.3.3 Interpretation of Free Network Solution

Although a free network lacks sufficient reference datum, the uniquely defined minimum-norm least-squares solution implies that a reference datum has somehow been established mathematically. By looking at the characters of matrix D, explicitly given in Eqs.(4.116) - (4.120), one may find out the geodetic implications of the minimum-norm least-squares solution $\widehat{X}$, and the mathematical reference datum adopted for the solution.

Levelling Networks

Substitution of D given by Eq.(4.116) in Eq.(4.109) leads to :

$$\widehat{x}_1 + \widehat{x}_2 + \cdots + \widehat{x}_m = 0$$

or:

$$x_0 = \frac{\sum_{i=1}^{m} \widehat{x}_i}{m} = 0 \tag{4.125}$$

where x_0 stands for the average height of all benchmarks after a minimum-norm least-squares solution. That $x_0 = 0$ can be interpreted as that *the minimum-norm least-squares solution has chosen the average height level of all benchmarks after adjustment as the zero height level* (height reference datum).

If approximate heights of the m benchmarks, $x_1^0,\ x_2^0,\ \cdots,\ x_m^0$, have been introduced in the observation equations, $\widehat{X}$ is actually estimate of the corrections $\delta\widehat{x}_1,\ \delta\widehat{x}_2,\ \cdots,\ \delta\widehat{x}_m$ to the approximate heights. Similar to (4.125), the average height correction δx_0 will now be zero :

$$\delta x_0 = \frac{\sum_{i=1}^{m} \delta\widehat{x}_i}{m} = 0$$

This implies that the reference height level is so defined that *the average height x_0 of all benchmarks after the free network adjustment remains the same as the average height before the adjustment* :

$$x_0 = \frac{\sum_{i=1}^{m} \widehat{x}_i}{m} = \frac{\sum_{i=1}^{m} x_i^0}{m} \qquad \text{with} \quad \widehat{x}_i = x_i^0 + \delta\widehat{x}_i \tag{4.126}$$

2-D Triangulation Networks

In triangulation networks, approximate coordinates $(x_i^0,\ y_i^0)$ are always introduced for the purpose of linearization and the unknown parameter vector $\widehat{X}$ given in (4.114) actually contains the coordinate corrections $(\delta x_i, \delta y_i)$ to the approximate coordinates $(i = 1,\ 2,\ 3,\ \cdots,\ k)$:

$$\underset{m\cdot 1}{\widehat{X}} = \begin{bmatrix} \delta\widehat{x}_1 \\ \delta\widehat{y}_1 \\ \delta\widehat{x}_2 \\ \delta\widehat{y}_2 \\ \cdots \\ \delta\widehat{x}_k \\ \delta\widehat{y}_k \end{bmatrix}$$

where k denotes the number of free, unknown network points while m is the total number of adjusted parameters.

Assume that point P_i has approximate coordinates $(x_i^0,\ y_i^0)$ before the adjustment (see **Figure 4.3**). After free network adjustment, P_i is moved slightly to a near-by location P_i' with the adjusted coordinates $(\widehat{x}_i, \widehat{y}_i)$:

$$\begin{bmatrix} \widehat{x}_i \\ \widehat{y}_i \end{bmatrix} = \begin{bmatrix} x_i^0 \\ y_i^0 \end{bmatrix} + \begin{bmatrix} \delta\widehat{x}_i \\ \delta\widehat{y}_i \end{bmatrix}$$

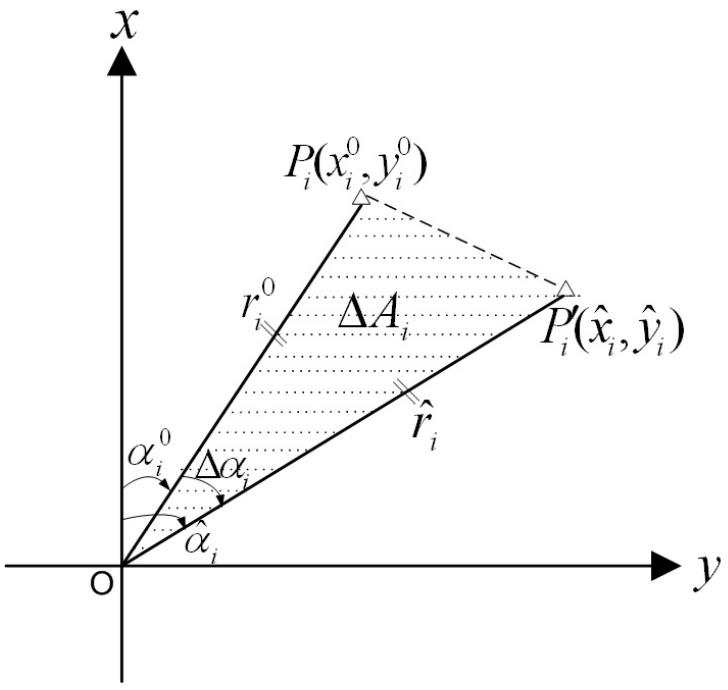

Figure 4.3: Reference datum in free network adjustment

The radial distances from the coordinate centre $O(0, 0)$ to P_i and P_i', respectively, are:

$$r_i^0 = \sqrt{(x_i^0)^2 + (y_i^0)^2} \qquad \text{and} \qquad \widehat{r}_i = \sqrt{(\widehat{x}_i)^2 + (\widehat{y}_i)^2} \tag{4.127}$$

and the corresponding azimuths of the two radial vectors, $\overrightarrow{OP_i}$ and $\overrightarrow{OP_i'}$, are:

$$\alpha_i^0 = \arctan\left(\frac{y_i^0}{x_i^0}\right) \qquad \text{and} \qquad \widehat{\alpha}_i = \arctan\left(\frac{\widehat{y}_i}{\widehat{x}_i}\right) \tag{4.128}$$

Now let us insert the first two rows of the D-matrix, given in Eq.(4.117) or (4.118), into Eq.(4.115) and arrive at :

$$\sum_{i=1}^{k} \delta\widehat{x}_i = 0 \quad \text{and} \quad \sum_{i=1}^{k} \delta\widehat{y}_i = 0 \tag{4.129}$$

from which the coordinates (x_0, y_0) of the geometric centre of all network points can be found :

$$x_0 = \frac{\sum_{i=1}^{k} \widehat{x}_i}{k} = \frac{\sum_{i=1}^{k} x_i^0}{k} + \frac{\sum_{i=1}^{k} \delta\widehat{x}_i}{k} = \frac{\sum_{i=1}^{k} x_i^0}{k} \tag{4.130}$$

$$y_0 = \frac{\sum_{i=1}^{k} \widehat{y}_i}{k} = \frac{\sum_{i=1}^{k} y_i^0}{k} + \frac{\sum_{i=1}^{k} \delta\widehat{y}_i}{k} = \frac{\sum_{i=1}^{k} y_i^0}{k} \tag{4.131}$$

The last two identities indicate that *the geometrical centre of the triangulation network remains fixed in the free network adjustment* so that the absolute location of the network is implicitly established.

Substitution of the third row of the $D-$matrix given in Eq.(4.117) or (4.118) into Eq.(4.115) gives :

$$\sum_{i=1}^{k} \left(-y_i^0 \delta\widehat{x}_i + x_i^0 \delta\widehat{y}_i\right) = 0 \tag{4.132}$$

To interpret the above equation, we look at the relation between $\widehat{\alpha}_i$ and α_i [see Eq.(4.128)] :

$$\widehat{\alpha}_i = \arctan\left(\frac{\widehat{y}_i}{\widehat{x}_i}\right) = \arctan\left(\frac{y_i^0 + \delta\widehat{y}_i}{x_i^0 + \delta\widehat{x}_i}\right) \approx \alpha_i^0 + \frac{-y_i^0 \delta\widehat{x}_i + x_i^0 \delta\widehat{y}_i}{\left(r_i^0\right)^2} \tag{4.133}$$

After free network adjustment, the radial vector $\overrightarrow{OP_i}$ has been rotated by an angle $\Delta\alpha_i = \widehat{\alpha}_i - \alpha_i^0$ to become the new radial vector $\overrightarrow{OP_i'}$. The area ΔA_i swept by vector $\overrightarrow{OP_i}$ during the rotation is :

$$\Delta A_i = \frac{1}{2}\left(r_i^0\right)^2 \cdot \Delta\alpha_i = \frac{1}{2}\left(r_i^0\right)^2 \cdot \left(\widehat{\alpha}_i - \alpha_i^0\right) = \frac{1}{2}\left(-y_i^0 \delta\widehat{x}_i + x_i^0 \delta\widehat{y}_i\right)$$

And the total area A swept by all radial vectors $\overrightarrow{OP_i}$ $(i = 1, 2, \cdots, k)$ in the network follows from Eq.(4.132):

$$A = \sum_{i=1}^{k} \Delta A_i = \frac{1}{2} \sum_{i=1}^{k} \left(-y_i^0 \delta\widehat{x}_i + x_i^0 \delta\widehat{y}_i\right) = 0 \tag{4.134}$$

Eq.(4.134) shows that *the whole network has zero rotation during free network adjustment, although each individual radial vector may have undertaken a slight rotation.*

The last row of the $D-$matrix given in Eqs.(4.118), inserted into Eq.(4.115), gives :

$$\sum_{i=1}^{k} \left(x_i^0 \cdot \delta\widehat{x}_i + y_i^0 \cdot \delta\widehat{y}_i\right) = 0 \tag{4.135}$$

Now we consider the squared radial distances [see Eq.(4.128)] :

$$(\widehat{r}_i)^2 = (x_i^0 + \delta\widehat{x}_i)^2 + (y_i^0 + \delta\widehat{y}_i)^2 \approx (x_i^0)^2 + (y_i^0)^2 + 2x_i^0 \cdot \delta\widehat{x}_i + 2y_i^0 \cdot \delta\widehat{y}_i = \left(r_i^0\right)^2 + 2\left(x_i^0 \cdot \delta\widehat{x}_i + y_i^0 \cdot \delta\widehat{y}_i\right)$$

Taking into account (4.135), we immediately obtain:

$$\sum_{i=1}^{k} (\widehat{r}_i)^2 = \sum_{i=1}^{k} \left(r_i^0\right)^2 + 2\sum_{i=1}^{k} \left(x_i^0 \cdot \delta\widehat{x}_i + y_i^0 \cdot \delta\widehat{y}_i\right) = \sum_{i=1}^{k} \left(r_i^0\right)^2 \tag{4.136}$$

which implies that the *sum of all squared radial distances remains fixed in free network adjustment* and the scale of the network is then defined.

Other Types of Networks

Trilateration networks contain distance measurements and thus do not lack scale datum. Therefore, free network adjustment for this kind of networks establishes the reference datum by fixing the network centre and by allowing zero rotation for all radial vectors together.

As for the three-dimensional free network, the first three rows of matrix D in Eq.(4.120) corresponds to zero translation in $x-$, $y-$ and $z-$direction, respectively, after free network adjustment. In other words, the network centre remains fixed to define the network's absolute position in space. The next three rows of D represents zero rotations along $x-$, $y-$ and $z-$direction, respectively, which establishes the orientation of the network. The last row of D corresponds to zero scale change after the adjustment.

The three-dimensional free network can be viewed as the general case which includes even the 1-D levelling networks as well as the 2-D triangulation/trilateration networks. Photogrammetric bundle adjustment with relative orientation and without absolute orientation is one example where completely free 3-D networks are involved (see e.g. Papo, et al. 1982).

The reference datum (position, orientation and scale) in traditional networks is normally defined by the fixed coordinates of one or more fixed points and the fixed azimuth and length of one or more fixed baselines. In free network adjustment, no point, azimuth or distance is held fixed and all coordinates of networks points, all azimuths and lengths of baselines are subject to changes (or corrections). The position, orientation and scale of the network is maintained by fixing the network centre and the sum of the radial distances squared, and furthermore by allowing zero total rotation .

Example 4.10

We consider once again the levelling network treated in **Example 4.8** and **4.9**. Now we are going to calculate the minimum norm-least squares solution $\hat{X}$ with both P and Q being unit matrices.

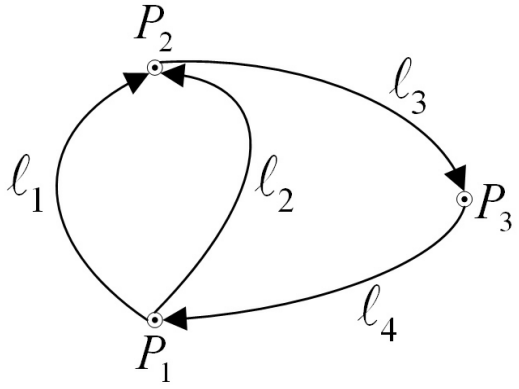

Figure 4.4: A simple levelling network

First we calculate the generalized inverse A_{II}^{-1} using Eqs.(4.106) and (4.116):

$$D = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

$$A^\top A + D^\top D = \begin{bmatrix} 3 & -2 & -1 \\ -2 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$|A^\top A + D^\top D| = 45 \neq 0$$

$$(A^\top A + D^\top D)^{-1} = \tfrac{1}{15}\begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

$$A_{II}^{-1} = (A^\top A + D^\top D)^{-1} A^\top = \tfrac{1}{15}\begin{bmatrix} -3 & -3 & -1 & 4 \\ 3 & 3 & -4 & 1 \\ 0 & 0 & 5 & -5 \end{bmatrix}$$

$$\widehat{X} = A_{II}^{-1} L = \tfrac{1}{15}\begin{bmatrix} -3 & -3 & -1 & 4 \\ 3 & 3 & -4 & 1 \\ 0 & 0 & 5 & -5 \end{bmatrix} \cdot \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_3 \\ \ell_4 \end{bmatrix} = \tfrac{1}{15}\begin{bmatrix} -3\ell_1 - 3\ell_2 - \ell_3 + 4\ell_4 \\ 3\ell_1 + 3\ell_2 - 4\ell_3 + \ell_4 \\ 5\ell_3 - 5\ell_4 \end{bmatrix} \qquad (h)$$

Check: $D\widehat{X} = \widehat{x}_1 + \widehat{x}_2 + \widehat{x}_3 = 0$ (!)

$$I - A\, A_{II}^{-1} = \tfrac{1}{5}\begin{bmatrix} 3 & -2 & 1 & 1 \\ -2 & 3 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}$$

$$\widehat{\varepsilon} = L - A\widehat{X} = (I - A\, A_{II}^{-1})L = \tfrac{1}{5}\begin{bmatrix} 3\ell_1 - 2\ell_2 + \ell_3 + \ell_4 \\ -2\ell_1 + 3\ell_2 + \ell_3 + \ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \\ \ell_1 + \ell_2 + 2\ell_3 + 2\ell_4 \end{bmatrix} \qquad (i)$$

$$C_{\widehat{X}\widehat{X}} = \sigma_0^2\, A_{II}^{-1}(A_{II}^{-1})^\top = \sigma_0^2 \cdot \tfrac{1}{45}\begin{bmatrix} 7 & -2 & -5 \\ -2 & 7 & -5 \\ -5 & -5 & 10 \end{bmatrix} \qquad (j)$$

$$tr\left[C_{\widehat{X}\widehat{X}}\right] = \tfrac{8}{15}\sigma_0^2 \qquad (k)$$

Based on the above results, the following remarks can be made:

- Comparing results in **Example 4.8**, **Example 4.9** and **Example 4.10**, we see that the estimated residuals are the same for all methods, as it should be.
- Eq.(h) above implies that the geometric centre has zero height and thus defines the height datum for the whole network.
- A comparison of Eq.(j) above with Eq.(c) in **Example 4.9** indicates that the variance-covariance matrix of the benchmark heights after adjustment is more homogeneous in the minimum-norm least-squares solution than the solution of traditional adjustment by elements with one benchmark artificially held fixed. The total variance ($\frac{8}{15}\sigma_0^2$) in the former is smaller than the total variance (σ_0^2) in the latter case [see Eq.(k) and Eq.(d)].

4.3.4 Alternative Formulations

In the previous subsections, the minimum-norm least-squares solution of the free network has been derived using the unique generalized inverse A_{QP}^{-1} or A_{IP}^{-1}. One can also derive the same solution using alternative formulations. These formulations may help us to get more insight on the characteristics of free network adjustment.

Extra Constraints

Assume that we have the following rank-defected observation equations:

$$\underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} = \underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon}, \qquad [r(A) \le m < n] \tag{4.137}$$

We now interpret Eq.(4.115) as the constraints that the estimate $\widehat{X}$ should satisfies:

$$D\widehat{X} = 0 \tag{4.138}$$

Our free network adjustment is then reduced to a classical adjustment by elements with constraints (4.138) under the least squares principle:

$$\widehat{\varepsilon}^{\top} P \widehat{\varepsilon} = minimum \tag{4.139}$$

where P is the weight matrix of L (ε). From Eq.(3.111) in **Section 3.3**, we obtain directly the normal equation for our free network:

$$\begin{bmatrix} A^{\top}PA & D^{\top} \\ D & 0 \end{bmatrix} \begin{bmatrix} \widehat{X} \\ \lambda \end{bmatrix} = \begin{bmatrix} A^{\top}PL \\ 0 \end{bmatrix} \tag{4.140}$$

As $A^{\top}PA$ in Eq.(4.140) is singular, we cannot follow the derivations in **Section 3.3** to derive $\widehat{X}$. Instead, we left-multiply the second equation in Eq.(4.140) by $D^{\top}$ and add the resulted equation to the first one in (4.141), which leads to a new normal equation system:

$$\begin{bmatrix} \overline{N} & D^{\top} \\ D & 0 \end{bmatrix} \begin{bmatrix} \widehat{X} \\ \lambda \end{bmatrix} = \begin{bmatrix} A^{\top}PL \\ 0 \end{bmatrix} \tag{4.141}$$

where:

$$\overline{N} = A^{\top}PA + D^{\top}D, \qquad \left|\overline{N}\right| \neq 0 \tag{4.142}$$

$\widehat{X}$ can then be solved from Eq.(4.141) [see Eq.(4.99)]:

$$\widehat{X} = \overline{N}^{-1}A^{\top}PL - \overline{N}^{-1}D^{\top}\left(D\overline{N}^{-1}D^{\top}\right)^{-1}D\overline{N}^{-1}A^{\top}PL \tag{4.143}$$

As the product $D\overline{N}^{-1}A^{\top}$ is equal to zero[2], the parameter estimate $\widehat{X}$ of the free network becomes:

$$\widehat{X} = \overline{N}^{-1}A^{\top}PL = \left(A^{\top}PA + D^{\top}D\right)^{-1}A^{\top}PL \tag{4.144}$$

which is identical to the minimum-norm least-squares solution (4.114). The variance-covariance matrix of $\widehat{X}$ is:

$$C_{\widehat{X}\widehat{X}} = \left(A^{\top}PA + D^{\top}D\right)^{-1}A^{\top}P \cdot \sigma_0^2 P^{-1} \cdot \left[\left(A^{\top}PA + D^{\top}D\right)^{-1}A^{\top}P\right]^{\top} = \sigma_0^2 \cdot \left(\overline{N}^{-1} - \overline{N}^{-1}D^{\top}D\overline{N}^{-1}\right) \tag{4.145}$$

---

[2] **Proof.** As $\overline{N} \cdot \overline{N}^{-1} = I$, we have:

$$\left(A^{\top}PA + D^{\top}D\right) \cdot \overline{N}^{-1} = I$$

Multiplying both sides by D and taking into account Eq.(4.107) leads to:

$$0 + DD^{\top}D\overline{N}^{-1} = D \quad or \quad D\overline{N}^{-1} = \left(DD^{\top}\right)^{-1}D$$

and thus:

$$D\overline{N}^{-1}A^{\top} = \left(DD^{\top}\right)^{-1}D \cdot A^{\top} = \left(DD^{\top}\right)^{-1}\left(A^{\top}D\right)^{\top} = 0$$

Pseudo-Observations

We may interpret the parameter estimate $\widehat{X}$ that satisfies both the least squares principle Eq.(4.139) and the minimum norm condition $\widehat{X}^\top \widehat{X}$=minimum, as pseudo-observations with observed values equal to zero. The observation equation of such pseudo-observations are:

$$\underset{d\cdot 1}{0} - \underset{d\cdot 1}{\varepsilon_x} = \underset{d\cdot m}{D}\,\underset{m\cdot 1}{\widehat{X}} \tag{4.146}$$

Together with the rank-defected, real observation equations Eq.(4.137), we have the total observation equations:

$$\begin{bmatrix} L \\ 0 \end{bmatrix} - \begin{bmatrix} \varepsilon \\ \varepsilon_x \end{bmatrix} = \begin{bmatrix} A \\ D \end{bmatrix} \cdot \widehat{X} \tag{4.147}$$

where ε_x denotes the residual of the pseudo-observations. The joint weight matrix for both types of observations is:

$$\begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix}$$

The least squares solution from the above observation equations is given by Eq.(3.18):

$$\widehat{X} = \left\{ \begin{bmatrix} A \\ D \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A \\ D \end{bmatrix} \right\}^{-1} \begin{bmatrix} A \\ D \end{bmatrix}^\top \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} L \\ 0 \end{bmatrix} = \left(A^\top P A + D^\top D\right)^{-1} A^\top P L \tag{4.148}$$

which is again identical to (4.114).

Chapter 6

Detection of Gross and Systematic Errors

Measurement errors can be classified into random errors, systematic errors and gross errors. Theory of errors mostly deals with random or stochastic errors.

Systematic errors can be reduced or avoided by e.g. carefully calibrating instruments to be used, using optimal measurement procedures, choosing favourable physical environment for the field measurement, and finally correcting measurement results after field work if reliable corrections are available. Mathematically, systematic errors behave like some kind of biases so that the measurement error ε does not attain zero expectation. If possible, one may try to model the systematic errors in the adjustment by introducing extra parameters. Naturally, one should make sure that such modelling is correct and the estimated systematic effects are statistically significant. Another way to detect small systematic effects in the measurements is statistical tests.

Gross errors should be avoided, e.g. through carefulness, rigorous and efficient checking routines, etc. Technically, gross errors may be treated as random errors which are extra-ordinarily large (e.g. 5-20 times their theoretical standard errors). This way of thinking is the basis of detecting and locating large gross errors from the preliminarily estimated residuals of the measurements.

Hypothesis test can be carried out either directly on the measurement data as shown in **Section 6.1**, or after a preliminary least squares adjustment as described in **Section 6.2**. The minimum detectable gross errors lead us to define in **Section 6.3** the internal and external reliability of the observations. The simultaneous treatment of gross error detection and variance components estimation will be discussed in **Section 6.4**. Other approaches suitable for gross error detection and without directly using hypothesis test are briefly outlined in **Section 6.5**.

6.1 Randomness Test

In **Chapter 1**, we have summarized the characteristics of the random errors :

1. The arithmetic mean of measurement errors ε_i $(i = 1, 2, \cdots, n)$ should approach zero when the number (n) of observations approaches infinity:

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} (\varepsilon_i)}{n} = 0 \tag{6.1}$$

2. Positive errors and negative errors with same absolute magnitude should have equal chance to occur;

3. Errors of smaller magnitude should have larger probability to occur than larger errors;

4. Under specified measurement condition, the magnitude of errors should be within some limit.

Considering the above characteristics of random errors, one can construct various statistics to directly test whether a set of measurement errors $\varepsilon_1,\ \varepsilon_2, \cdots, \varepsilon_n$ are random or not. Below, five different tests will be outlined.

The general procedure of statistical tests for detecting non-random errors can be summarized as follows :

- We start from the assumption that the measurement errors are random errors with given distributions. This is actually our null hypothesis (H_0). The alternative hypothesis (H_1) is simply that the measurement errors are not from the assumed distribution;
- We then construct some quantities which can be computed from observation data and which have well-defined statistical distribution when the above assumption holds;
- Now we test the computed sample values of the above statistics against the theoretical critical values at certain risk level α. If the test is passed, we then accept H_0. Otherwise, we may suspect there might exist systematic or gross errors in the measurement results.

(i) Testing the Maximum Absolute Value of Errors

Let ε_m denote the absolute value of the error of maximum magnitude :

$$\varepsilon_m = \max_{1 \le i \le n} (|\varepsilon_i|) \tag{6.2}$$

If all ε_i are random variables and have normal distribution, $\varepsilon_i \sim N(0,\, \sigma^2)$, we have for risk level α :

$$P\left\{ \frac{\varepsilon_m}{\sigma} < c_{\frac{1}{2}\alpha} \right\} = 1 - \alpha \tag{6.3}$$

Thus, if $\varepsilon_m < \sigma\ c_{\frac{1}{2}\alpha}$, we accept the assumption $\varepsilon_i \sim N(0,\, \sigma^2)$. Otherwise, we may suspect systematic effects in ε_i .

(ii) Testing the Sum of Errors

Let s denote the sum of the n errors :

$$s = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n$$

If $\varepsilon_i \sim N(0,\, \sigma^2)$ are uncorrelated with each other, s will have normal distribution $N(0,\, n\sigma^2)$. For risk level α, we then have:

$$P\left\{ \left| \frac{s}{\sqrt{n}\sigma} \right| < c_{\frac{1}{2}\alpha} \right\} = 1 - \alpha \tag{6.4}$$

where $c_{\frac{1}{2}\alpha}$ is the critical value of $N(0,1)$ at risk level α. If $|s| < \sqrt{n} \cdot \sigma \cdot c_{\frac{1}{2}\alpha}$, we accept $\varepsilon_i \sim N(0,\, \sigma^2)$. Otherwise, we may suspect systematic effects in ε_i .

(iii) Testing the Number of Positive versus Negative Errors

Let $s_+,\ s_-$ denote the number of positive errors and negative errors, respectively. If ε_i is random, s_+ should be of binomial distribution with expectation and variance as follows (see **Subsection 1.4.1**):

$$\mu = \frac{1}{2}n, \quad \sigma^2 = \frac{n}{4} \tag{6.5}$$

When n is very large, s_+ will approach the normal distribution. In other words, the following statistics is approximately of standard normal distribution:

$$\frac{s_+ - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \sim N(0,1) \tag{6.6}$$

At risk level α, we then have:

$$P\left\{\left|\frac{s_+ - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}\right| < c_{\frac{1}{2}\alpha}\right\} = P\left\{\left|\frac{s_+ - s_-}{\sqrt{n}}\right| < c_{\frac{1}{2}\alpha}\right\} = 1-\alpha$$

or:

$$P\left\{|s_+ - s_-| < \sqrt{n}c_{\frac{1}{2}\alpha}\right\} = 1-\alpha \tag{6.7}$$

where $s_- = n - s_+$, and $c_{\frac{1}{2}\alpha}$ denotes the critical value of $N(0,1)$ at risk level α. To test whether the number of positive errors is statistically equal to the number of negative errors, we have the following null hypothesis (H_0) and alternative hypothesis (H_1) :

$$H_0 : \ s_+ = s_- \qquad \text{and} \qquad H_1 : \ s_+ \neq s_- \tag{6.8}$$

For a chosen risk level α, if $|s_+ - s_-| < \sqrt{n}c_{\frac{1}{2}\alpha}$ we accept H_0. Otherwise, we reject H_0, i.e. s_+ and s_- are significantly different and there may be systematic effects in ε_i.

(iv) Testing the Signs of the Errors

Sometimes, systematic effects lead to that positive (or negative) errors follow each other, even though the total number of positive errors is statistically equal to the number of negative ones. The following statistical test tries to check whether or not the signs of the errors are random.

Let s_1 denote the number of adjacent error pairs which have the same signs, as defined by:

$$s_1 = x_1 + x_2 + \cdots + x_{n-1} \tag{6.9}$$

where :

$$x_i = \begin{cases} 1 & \text{if } \varepsilon_i \text{ and } \varepsilon_{i+1} \text{ have the same sign} \\ 0 & \text{otherwise} \end{cases} \qquad (i = 1, 2, \cdots, n-1) \tag{6.10}$$

If ε_i are random errors, x_i should have the same probability $p = \frac{1}{2}$ to be 1 or 0. Thus s_1 is of binomial distribution which approaches the normal distribution when n is very large. This gives us a statistics of standard normal distribution :

$$\frac{s_1 - \frac{1}{2}(n-1)}{\frac{1}{2}\sqrt{n-1}} = \frac{s_1 - s_0}{\sqrt{n-1}} \sim N(0,1)$$

where $s_0 = (n-1) - s_1$ denotes the number of adjacent error pairs with opposite signs. Our hypotheses to be tested are as follows:

$$H_0 : \ s_1 = s_0 \qquad \text{and} \qquad H_1 : \ s_1 \neq s_0 \tag{6.11}$$

For a chosen risk level α, if $|s_1 - s_0| < \sqrt{n-1}c_{\frac{1}{2}\alpha}$, we accept H_0 . Otherwise, we reject H_0, *i.e.* there may be systematic effects in ε_i $(i = 1, 2, \cdots, n)$.

(v) Testing the Sum of Positive versus Negative Errors Squared

Let s^2 denote the difference between the sum of positive errors squared and the sum of negative errors squared :

$$s^2 = \lambda_1\varepsilon_1^2 + \lambda_2\varepsilon_2^2 + \cdots + \lambda_n\varepsilon_n^2 \tag{6.12}$$

where :

$$\lambda_i = \begin{cases} +1 & \text{if } \varepsilon_i > 0 \\ -1 & \text{if } \varepsilon_i < 0 \end{cases} \tag{6.13}$$

Note that the notation s^2 above does not denote the square of quantity s. If ε_i $(i = 1, 2, \cdots, n)$ are random errors, s^2 should be close to zero and that $\lambda_i = +1$ and $\lambda_i = -1$ have equal probability to happen:

$$P(\lambda_i = +1) = P(\lambda_i = -1) = \frac{1}{2} \qquad (i = 1, 2, \cdots, n) \tag{6.14}$$

The expectation and variance of λ_i can be found :

$$\left.\begin{aligned} E(\lambda_i) &= (+1) \cdot \tfrac{1}{2} + (-1) \cdot \tfrac{1}{2} = 0 \\ E(\lambda_i^2) &= (+1)^2 \cdot \tfrac{1}{2} + (-1)^2 \cdot \tfrac{1}{2} = 1 \end{aligned}\right\} \tag{6.15}$$

As the sign of ε_i is independent of its absolute value, λ_i and ϵ_i should be uncorrelated with each other. Furthermore, when ε_i $(i = 1, 2, \cdots, n)$ are uncorrelated with normal distribution $N(0, \sigma^2)$, the expectation and variance of $\lambda_i \epsilon_i^2$ can be obtained :

$$\left.\begin{aligned} E(\lambda_i \varepsilon_i^2) &= E(\lambda_i) \cdot E(\varepsilon_i^2) = 0 \\ E\left[(\lambda_i \varepsilon_i^2)^2\right] &= E(\lambda_i^2) \cdot E(\varepsilon_i^4) = E(\varepsilon_i^4) = 3\sigma^4 \end{aligned}\right\} \tag{6.16}$$

which lead to the expectation and variance of s^2 :

$$\left.\begin{aligned} E(s^2) &= \textstyle\sum_{i=1}^{n} \left[E(\lambda_i \varepsilon_i^2)\right] = 0 \\ E\left[(s^2)^2\right] &= \textstyle\sum_{i=1}^{n} \left[E(\lambda_i^2 \varepsilon_i^4)\right] = \sum_{i=1}^{n} \left[E(\lambda_i^2) \cdot E(\varepsilon_i^4)\right] = 3n\sigma^4 \end{aligned}\right\} \tag{6.17}$$

When n is very large, s^2 will approach normal distribution $N(0,\, 3n\sigma^4)$. Thus for risk level α, we have :

$$P\left\{\left|\frac{s^2}{\sqrt{3n}\sigma^2}\right| < c_{\frac{1}{2}\alpha}\right\} = 1 - \alpha \tag{6.18}$$

where $c_{\frac{1}{2}\alpha}$ is the critical value of $N(0, 1)$ at risk level α. If the computed s^2 satisfies:

$$\left|s^2\right| < \sqrt{3n}\sigma^2 \; c_{\frac{1}{2}\alpha} \tag{6.19}$$

we accept that s^2 is statistically equal to zero. Otherwise, s^2 is significantly different from zero and there may exist systematic effects in ε_i.

Example 6.1

In a geodetic triangulation network, $n = 30$ triangles have been observed with the following triangular misclosures (w_i) :

i	w_i ($''$)	i	w_i ($''$)	i	w_i ($''$)
1	+1.5	11	-2.0	21	-1.1
2	+1.0	12	-0.7	22	-0.4
3	+0.8	13	-0.8	23	-1.0
4	-1.1	14	-1.2	24	-0.5
5	+0.6	15	+0.8	25	+0.2
6	+1.1	16	-0.3	26	+0.3
7	+0.2	17	+0.6	27	+1.8
8	-0.3	18	+0.8	28	+0.6
9	-0.5	19	-0.3	29	-1.1
10	+0.6	20	-0.9	30	-1.3

The standard error of w_i is assumed to be $\sigma = \pm 0.93''$. We want to test the randomness of this group of triangular misclosures using the five different tests described above. The risk level is chosen to be: $\alpha = 4.55\%$, for which the critical value of the standard normal distribution is $c_{\frac{1}{2}\alpha} = 2$.

(*i*) *Testing the maximum absolute value of errors*

$|w_m| = |w_{11}| = |-2.0| = 2.0'' > \sigma \cdot c_{\frac{1}{2}\alpha} = 1.86''$

$\Longrightarrow$ test is not passed !

(ii) Testing the sum of errors

$|s| = |2.6| = 2.6 < \sqrt{n}\sigma \cdot c_{\frac{1}{2}\alpha} \approx 10.2$

$\Longrightarrow$ test is passed !

(iii) Testing the number of positive versus negative errors

$|s_+ - s_-| = |14 - 16| = 2 < \sqrt{n} \cdot c_{\frac{1}{2}\alpha} \approx 11$

$\Longrightarrow$ test is passed !

(iv) Testing the order of positive versus negative errors

$|s_1 - s_0|\,| = |18 - 11| = 7 < \sqrt{n-1} \cdot c_{\frac{1}{2}\alpha} \approx 11$

$\Longrightarrow$ test is passed !

(v) Testing the sum of positive versus negative errors squared

$|s^2| = |3.40| = 3.4 < \sqrt{3n}\sigma^2 \cdot c_{\frac{1}{2}\alpha} \approx 16.41$

$\Longrightarrow$ test is passed !

The first test above has been rejected, which implies that w_{11} might not have the normal distribution $N(0, \sigma^2)$. However, since the last four tests have been passed and $|w_{11}| = 2.0''$ is quite close to the tolerance level $1.86''$, further investigations may be nedded before we draw a definite conclusion on w_{11}.

6.2 Data Snooping

Data snooping is designed to detect and locate gross errors in the observations based on a preliminary least squares adjustment and statistical tests. It was developed during 1960's by professor Baarda at the Delft University of Technology in the Netherlands (Baarda, 1967, 1968, 1973). Data snooping assumes that only one observation contains a gross error which is then located by statistically testing the least squares residuals against the standard errors of the residuals. One can repeat the above test, one by one, for each observation.

The null hypothesis (H_0) in data snooping is that there is no gross error in our observations and that our functional and statistical models are as follows :

$$H_0 : \left\{ \begin{array}{l} \underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} = \underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} \\ E\{\varepsilon\} = 0 \\ E\left\{\varepsilon\varepsilon^\top\right\} = \sigma_0^2 Q = \sigma_0^2 P^{-1} \end{array} \right\} \tag{6.20}$$

where P and Q denote the weight matrix and the cofactor matrix of ε (or L), respectively. The least squares estimates of X and ε are then given by Eq.(3.18) and Eq.(3.20), respectively :

$$\widehat{X} = \left(A^\top P A\right)^{-1} A^\top P L \tag{6.21}$$

$$\widehat{\varepsilon} = \left[I - A(A^\top P A)^{-1} A^\top P\right] L \tag{6.22}$$

and its variance-covariance matrix $C_{\widehat{\varepsilon}\widehat{\varepsilon}}$ and cofactor matrix $Q_{\widehat{\varepsilon}\widehat{\varepsilon}}$ by Eq.(3.34):

$$C_{\widehat{\varepsilon}\widehat{\varepsilon}} = \sigma_0^2 \cdot Q_{\widehat{\varepsilon}\widehat{\varepsilon}} \tag{6.23}$$

$$Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = P^{-1} - A(A^\top P A)^{-1} A^\top = \left[\begin{array}{cccc} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{array} \right] \tag{6.24}$$

One can also calculate the sum of the weighted squares of $\widehat{\varepsilon}$ by :

$$\Omega = \widehat{\varepsilon}^\top P\, \widehat{\varepsilon} = L^\top \left[P - PA(A^\top P A)^{-1} A^\top P\right] L \tag{6.25}$$

The alternative hypothesis H_1 is that there is one observation, say ℓ_i, which contains gross error Δ_i. The statistical model is the same as in Eq.(6.20), while the functional model is complemented by the gross error Δ_i :

$$H_1 : \left\{ \begin{array}{l} \underset{n\cdot 1}{L} - \underset{n\cdot 1}{\varepsilon} = \underset{n\cdot m}{A}\ \underset{m\cdot 1}{X} + \underset{n\cdot 1}{e_i} \cdot \Delta_i \\ E\{\varepsilon\} = 0 \\ E\left\{\varepsilon\varepsilon^\top\right\} = \sigma_0^2 Q = \sigma_0^2 P^{-1} \end{array} \right\} \tag{6.26}$$

where e_i denotes a column vector of zero elements except the $i-th$ element which equals 1:

$$\underset{n\cdot 1}{e_i} = \left[\begin{array}{c} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \\ 0 \\ \cdots \\ 0 \end{array} \right] \tag{6.27}$$

Let $\widehat{\varepsilon}_\Delta$ denote the least squares estimate of residual vector ε based on the hypothesis H_1 in Eq.(6.26). It can be proved that the sum of the weighted squares of $\widehat{\varepsilon}_\Delta$ is related to the the sum of the weighted squares of $\widehat{\varepsilon}$ given in Eq.(6.25) :

$$\Omega_\Delta = \widehat{\varepsilon}_\Delta^\top P \widehat{\varepsilon}_\Delta = \widehat{\varepsilon}^\top P \widehat{\varepsilon} - \Delta\Omega = \Omega - \Delta\Omega \tag{6.28}$$

where $\Delta\Omega$ is as follows:

$$\Delta\Omega = \widehat{\varepsilon}^\top P e_i \left(e_i^\top P Q_{\widehat{\varepsilon}\widehat{\varepsilon}} P e_i\right)^{-1} e_i^\top P \widehat{\varepsilon} \tag{6.29}$$

with $\widehat{\varepsilon}$ and $Q_{\widehat{\varepsilon}\widehat{\varepsilon}}$ given by Eq.(6.22) and Eq.(6.24), respectively. The least squares estimate of the gross error Δ_i from the adjustment model in Eq.(6.26) is:

$$\widehat{\Delta}_i = \left(e_i^\top P Q_{\widehat{\varepsilon}\widehat{\varepsilon}} P e_i\right)^{-1} e_i^\top P \widehat{\varepsilon} \tag{6.30}$$

and the least squares estimate of X is now :

$$\widehat{X}_\Delta = \widehat{X} - \left(A^\top P A\right)^{-1} A^\top P e_i \cdot \widehat{\Delta}_i \tag{6.31}$$

where $\widehat{X}$ is the least squares estimate under H_0, as given in Eq.(6.21).

If the weight matrix P is diagonal (i.e. all observations are uncorrelated with each other),

$$\underset{n\cdot n}{P} = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \cdots & \\ & & & p_n \end{bmatrix} \tag{6.32}$$

Eqs.(6.29) and (6.30) can be simplified :

$$\Delta\Omega = \frac{\widehat{\varepsilon}_i^2}{q_{ii}} \tag{6.33}$$

$$\widehat{\Delta}_i = \frac{\widehat{\varepsilon}_i}{p_i \cdot q_{ii}} \tag{6.34}$$

where q_{ii} denotes the $i-th$ diagonal element in the cofactor matrix $Q_{\widehat{\varepsilon}\widehat{\varepsilon}}$ as given in Eq.(6.24) .

If the a priori variance factor σ_0^2 is known, we obtain the following two statistics under the null hypothesis H_0 :

$$\frac{\Omega}{\sigma_0^2} \sim \chi^2(n-m) \tag{6.35}$$

$$\frac{\Delta\Omega}{\sigma_0^2} \sim \chi^2(1) \quad \text{or equivalently:} \quad \frac{\sqrt{\Delta\Omega}}{\sigma_0} \sim N(0,1) \tag{6.36}$$

In the case of a diagonal weight matrix P, we insert Eq.(6.33) into Eq.(6.36) to obtain :

$$u_i = \frac{\sqrt{\Delta\Omega}}{\sigma_0} = \frac{\widehat{\varepsilon}_i}{\sigma_0\sqrt{q_{ii}}} = \frac{\widehat{\varepsilon}_i}{\sigma_{\widehat{\varepsilon}_i}} \sim N(0,1) \tag{6.37}$$

where $\sigma_{\widehat{\varepsilon}_i}$ denotes the standard error of $\widehat{\varepsilon}_i$.

Eqs.(6.35), (6.37) can then be used to test whether there exists gross error (Δ_i) in the observation ℓ_i :

- *Overall test (multi-dimensional)*

  If $\Omega < \sigma_0^2 \cdot \chi_{n-m,\,\alpha}$, we accept H_0 (i.e. there is no gross error) and reject H_0 otherwise. Here $\chi_{n-m,\,\alpha}$ is the critical value of $\chi^2(n-m)$ at risk level α.

- *Test of each individual observation (one dimensional)*

  If $|u_i| < c_{\frac{1}{2}\alpha}$, we accept H_0 and otherwise reject H_0. Here $c_{\frac{1}{2}\alpha}$ denotes the critical value of $N(0,1)$ at risk level α.

If the a priori variance factor σ_0^2 is not known, an a posteriori estimate may be computed from the estimated $\widehat{\varepsilon}$:

$$\widehat{\sigma}_0^2 = \frac{\Omega}{n-m} = \frac{\widehat{\varepsilon}^\top P \widehat{\varepsilon}}{n-m} \tag{6.38}$$

Replacing σ_0 in Eq.(6.37) by the estimated $\widehat{\sigma}_0$, we obtain a new statistics of $t-$distribution (see Subsection 1.4.1) :

$$w_i = \frac{\widehat{\varepsilon}_i}{\widehat{\sigma}_0 \sqrt{q_{ii}}} \sim t\,(n-m-1) \tag{6.39}$$

The above statistics can be used to test whether ℓ_i contains a gross error: if $|w_i| < t_{\frac{1}{2}\alpha}(n-m-1)$ we accept H_0 and reject H_0 otherwise. Here $t_{\frac{1}{2}\alpha}(n-m-1)$ denotes the critical value of the $t-$distribution for $n-m-1$ degrees of freedom at risk level α.

The above testing procedure can be performed subsequently for each observation $\ell_1,\ \ell_2,\ \cdots,\ \ell_n$, separately, and thus called a data snooping strategy. The minimum gross error $|\Delta_i|$ that can be detected by hypothesis tests using Eq.(6.35), and/or Eq.(6.37) or Eq.(6.39) is dependent on the risk level α and the testing power of the tests. This aspect will be described in more details in **Section 6.3**, as the minimum detectable error is directly related to the reliability of the observations and the network concerned.

There are some limitations with data snooping technique for gross error detection. Firstly, it assumes that there is only one gross error. If there are more than one gross errors in the observations, the method cannot guarantee the detection. Secondly, tests using Eq.(6.37) or Eq.(6.39) implicitly assume that the observation with large gross error will also have large residual from the preliminary least squares adjustment. However, many studies have shown that least squares adjustment has the tendency to smooth the residuals, i.e. hide large residuals (e.g. due to gross error) and distribute their effects to other observations. Consequently, observations with large (gross) errors may not clearly be identified based on the residuals.

Example 6.2

This example uses the method of data snooping to search for gross errors in the five observations, adjusted separately in **Example 2.4** and **Example 3.3**.

For the given *a priori* unit-weight standard error $\sigma_0 = 3\ mm$, the used weight matrix P is :

$$\underset{5\cdot 5}{P} = \begin{bmatrix} \frac{1}{4} & & & & \\ & \frac{1}{4} & & & \\ & & \frac{1}{4} & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$$

and the least squares residuals $\widehat{\varepsilon}$ have been estimated to

$$\widehat{\varepsilon} = \begin{bmatrix} +6.45\,'' \\ -3.40\,'' \\ +2.95\,'' \\ +4.82\,^{mm} \\ -3.98\,^{mm} \end{bmatrix}$$

The cofactor matrix of the adjusted observations $\widehat{L}$ has been found in **Example 2.4** :

$$Q_{\widehat{L}\widehat{L}} = \begin{bmatrix} +1.8664 & -1.1792 & -0.6872 & -0.1377 & +0.7148 \\ -1.1792 & +1.8664 & -0.6872 & +0.7148 & -0.1377 \\ -0.6872 & -0.6872 & +1.3744 & -0.5771 & -0.5771 \\ -0.1377 & +0.7148 & -0.5771 & +0.3616 & +0.1230 \\ +0.7148 & -0.1377 & -0.5771 & +0.1230 & +0.3616 \end{bmatrix}$$

The cofactor matrix of the least squares residuals $\widehat{\varepsilon}$ can then be obtained from :

$$Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = P^{-1} - Q_{\widehat{L}\widehat{L}} = \begin{bmatrix} +2.1336 & +1.1792 & +0.6872 & +0.1377 & -0.7148 \\ +1.1792 & +2.1336 & +0.6872 & -0.7148 & +0.1377 \\ +0.6872 & +0.6872 & +2.6256 & +0.5771 & +0.5771 \\ +0.1377 & -0.7148 & +0.5771 & +0.6384 & -0.1230 \\ -0.7148 & +0.1377 & +0.5771 & -0.1230 & +0.6384 \end{bmatrix}$$

The diagonal elements of $Q_{\widehat{\varepsilon}\widehat{\varepsilon}}$, denoted by q_{ii} $(1 \leq i \leq 5)$, are then computed from the above equation and listed in the table below.

Using the *a priori* σ_0 as well as the *a posteriori* $\widehat{\sigma}_0$, data snooping are then performed on the five residuals $\widehat{\varepsilon}_i$. The risk level α has been chosen to 5% for both the normal distribution test and the $t-$test, with the following critical values :

$$c_{\frac{1}{2}\alpha} = 1.96, \quad t_{\frac{1}{2}\alpha}(n-m-1) = t_{\frac{1}{2}\alpha}(2) = 4.3$$

The detailed results of data snooping are listed in the table below.

i	$\widehat{\varepsilon}_i$	q_{ii}	$\sqrt{q_{ii}}$	$u_i = \frac{\widehat{\varepsilon}_i}{\sigma_0 \times \sqrt{q_{ii}}}$ $(\sigma_0 = 3\ ^{mm})$	$w_i = \frac{\widehat{\varepsilon}_i}{\widehat{\sigma}_0 \times \sqrt{q_{ii}}}$ $(\widehat{\sigma}_0 = 4.26\ ^{mm})$
1	$+6.45''$	2.1336	1.46	$+1.47$	$+1.04$
2	$-3.40''$	2.1336	1.46	-0.78	-0.55
3	$+2.95''$	2.6256	1.62	$+0.61$	$+0.43$
4	$+4.82^{mm}$	0.6384	0.80	$+2.01$	$+1.41$
5	-3.98^{mm}	0.6384	0.80	-1.66	-1.17
				$\alpha = 5\%, \quad c_{\frac{1}{2}\alpha} = 1.96$	$\alpha = 5\%,\ t_{\frac{1}{2}\alpha}(2) = 4.3$

We then compare the computed u_i and w_i with $c_{\frac{1}{2}\alpha}$ and $t_{\frac{1}{2}\alpha}(n-m-1)$, respectively. All observations pass the tests except in the case of u_4 which is slightly larger than but very close to the critical value $c_{\frac{1}{2}\alpha}$. With so few observations $(n = 5)$ and low redundancy $(n - m = 3)$, it seems that we do not have any strong reason to suspect gross errors in the five observations.

6.3 Reliability of Observations

The concept of quality concerns not only the accuracy and precision measures, characterized e.g. by the variance-covariance matrices of the various observations and adjusted quantities, but also the capability of the network (or the adjustment process) to detect possible gross errors. The latter aspect of the quality concept is often known as the *reliability* of the observations and the network concerned.

6.3.1 Local Redundancies

Let $\underset{n\cdot n}{R}$ be a square matrix defined as the product of the cofactor matrix $Q_{\widehat{\varepsilon}\widehat{\varepsilon}}$ of the estimated least squares residuals $\widehat{\varepsilon}$ and the weight matrix P of the observations :

$$\underset{n\cdot n}{R} = Q_{\widehat{\varepsilon}\widehat{\varepsilon}} \cdot P = \underset{n\cdot n}{I} - A\left(A^\top PA\right)^{-1} A^\top P = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \tag{6.40}$$

The trace of R can be found :

$$tr(R) = tr\left[\underset{n\cdot n}{I}\right] - tr\left[A\left(A^\top PA\right)^{-1} A^\top P\right] = n - tr\left[\left(A^\top PA\right)^{-1} A^\top PA\right] = n - m = r \tag{6.41}$$

where n, m are the number of observations and unknown parameters, respectively, and r is the total number of redundancies or over-determinations of the network concerned. Matrix R is called the *redundancy matrix* of the observation vector L, while the $i-th$ diagonal element of R, designated as r_i $(i = 1, 2, \cdots, n)$, is called the *local redundancy* of the $i-th$ observation ℓ_i:

$$r_i = (R)_{ii} = r_{ii} \qquad (i = 1, 2, \cdots, n) \tag{6.42}$$

Eq.(6.41) implies:

$$r_1 + r_2 + \cdots + r_n = r \tag{6.43}$$

that is to say, the sum of all local redundancies is equal to the overall redundancy of the network.

If P is a diagonal matrix as shown in Eq.(6.32) and $Q_{\widehat{\varepsilon}\widehat{\varepsilon}}$ has elements defined by Eq.(6.24), we have:

$$r_i = q_{ii} \cdot p_i \tag{6.44}$$

From Eq.(3.35),

$$Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = P^{-1} - Q_{\widehat{L}\widehat{L}}$$

we have:

$$q_{ii} = \frac{1}{p_i} - \frac{1}{\widehat{p}_i} \tag{6.45}$$

where $\widehat{p}_i$ denotes the weight of the adjusted value of the $i-th$ observation ℓ_i and $\frac{1}{\widehat{p}_i}$ is the $i-th$ diagonal element of the cofactor matrix $Q_{\widehat{L}\widehat{L}}$ of the adjusted observations $\widehat{L}$. Substitution of Eq.(6.45) into Eq.(6.44) leads to :

$$r_i = 1 - \frac{p_i}{\widehat{p}_i} \tag{6.46}$$

As the weight of the adjusted observation $\widehat{\ell}_i$ is always larger than (or in the worst case equal to) the weight of the unadjusted observation ℓ_i, i.e. $\widehat{p}_i \geq p_i$, Eq.(6.46) implies :

$$0 \leq r_i \leq 1 \tag{6.47}$$

$r_i = 1$ corresponds the ideal case when $\widehat{\ell_i}$ has been perfectly determined (i.e. $\widehat{p}_i \to \infty$) through the adjustment, while $r_i = 0$ represents the worst case that the adjustment does not at all done any improvement to ℓ_i (i.e. $\widehat{p}_i = p_i$). Therefore, the local redundancy r_i can be regarded as a measure of accuracy, i.e. the degree of accuracy improvement accomplished by the least squares adjustment for each individual observation ℓ_i $(i = 1, 2, \cdots, n)$.

6.3.2 Errors of Hypothesis Tests

Hypothesis tests usually suffer from two types of testing errors. As an example, we look at a standard normal distribution u_i. Our null hypothesis (H_0) and alternative hypothesis (H_1) can be:

$$H_0 \ : \quad u_i \sim N(0,1); \qquad H_1 \ : \quad u_i \sim N(\delta_0, 1) \tag{6.48}$$

where δ_0 stands for a non-central shift due to e.g. a gross error (see Figure 6.1).

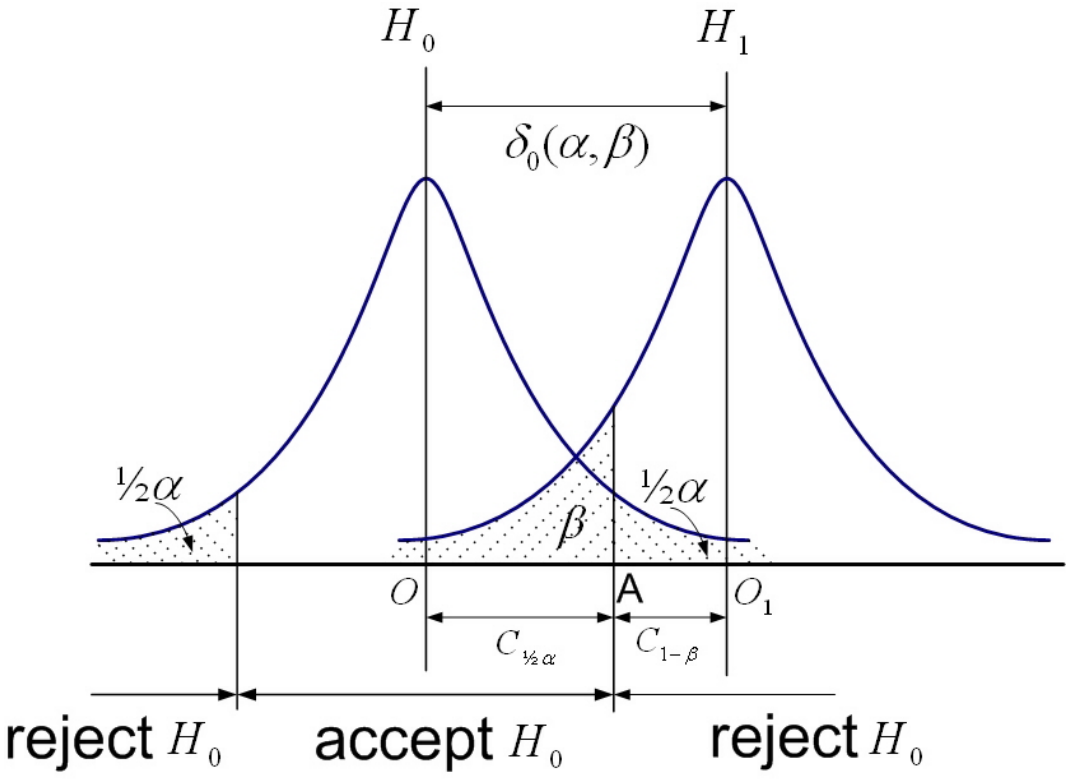

Figure 6.1: Testing errors of standard normal distributions

There are two types of testing errors:

- *Type I error* : H_0 is true but is *incorrectly* rejected when the computed sample value $|u_i|$ is larger than the critical value $c_{\frac{1}{2}\alpha}$ of the standard normal distribution for risk level α. Obviously, the probability that the Type I error happens is just the risk level α;

- *Type II error* : H_0 is false (i.e. H_1 is true) but is *incorrectly* accepted when the computed sample value $|u_i|$ is smaller than $c_{\frac{1}{2}\alpha}$. The probability that the Type II error occurs is designated by β. Then $1-\beta$ corresponds to the probability that H_0 is false and is *correctly* rejected. In probability theory, $1-\beta$ is called the *power* of a statistical test.

In Figure 6.1, the dotted areas represent the probability of the Type I error (at risk level α), while the shaded area represents the probability β of the Type II error. The total shift δ_0 can be divided into two parts: a partial shift from the central point O under H_0 to the critical point A of H_0 for risk level $\frac{1}{2}\alpha$; and a partial shift from the central point O_1 under H_1 to A. The former shift is nothing else but $c_{\frac{1}{2}\alpha}$, while the second shift can be found to be the critical value $c_{1-\beta}$ of the standard normal distribution for risk level $1-\beta$. Therefore, the total shift δ_0 between H_0 and H_1 is a function of α and β :

$$\delta_0\left(\alpha,\beta\right) = c_{\frac{1}{2}\alpha} + c_{1-\beta} \tag{6.49}$$

For specified risk level α and testing power β, δ_0 corresponds to the smallest shift of the standard normal distribution that is possible to detect. Normally one chooses $\alpha = 5\%$ and $\beta = 20\%$, which gives $\delta_0 \approx 2.80$.

6.3.3 Internal and External Reliability

Now let us reconsider the statistics of standard normal distribution defined in Eq.(6.37) and the estimated gross error in the observation ℓ_i. Comparing Eq.(6.34) and Eq.(6.37), we find :

$$\widehat{\Delta}_i = \frac{\sigma_0 \cdot \sqrt{q_{ii}}}{p_i \cdot q_{ii}} \cdot u_i = \frac{\sigma_0/\sqrt{p_i}}{\sqrt{p_i \cdot q_{ii}}} = \frac{\sigma_i}{\sqrt{r_i}} \cdot u_i \tag{6.50}$$

where r_i denotes the local redundancy of ℓ_i [see Eq.(6.44)] and σ_i denotes the standard error ℓ_i :

$$\sigma_i = \frac{\sigma_0}{\sqrt{p_i}} \tag{6.51}$$

The minimum detectable gross error $|\Delta_i|$ can then be found :

$$\theta_i = |\Delta_i|_{\min} = \frac{\sigma_i}{\sqrt{r_i}} \cdot \delta_0(\alpha,\beta) \tag{6.52}$$

where $\delta_0(\alpha,\beta)$ is the non-central shift in the standard normal distribution, as defined in Eq.(6.49). θ_i is called the *internal reliability* of the observation ℓ_i. It is a measure of our capability or possibility to detect gross errors in the observations.

Correspondingly, one can define the *external reliability* of ℓ_i as the effect of the undetected gross error Δ_i in ℓ_i on the adjustment results, which include the estimated parameters $\widehat{X}$, the adjusted observations $\widehat{L}$ and any linear functions of the estimated parameters [1]. The effect of Δ_i on $\widehat{X}$ is the second term in Eq.(6.31) :

$$\Delta_i^{(X)} = \widehat{X}_\Delta - \widehat{X} = -\left(A^\top P A\right)^{-1} A^\top P e_i \cdot \widehat{\Delta}_i \tag{6.53}$$

The effect of Δ_i on the adjusted observations $\widehat{L}$ can be derived as follows :

$$\begin{aligned}\Delta_i^{(\widehat{L})} = A\left(\widehat{X}_\Delta - \widehat{X}\right) = -A\left(A^\top P A\right)^{-1} A^\top P e_i \widehat{\Delta}_i = e_i\widehat{\Delta}_i - \left[I - A\left(A^\top P A\right)^{-1} A^\top P\right] \cdot e_i\widehat{\Delta}_i \\ = (I-R)\cdot e_i\widehat{\Delta}_i \end{aligned} \tag{6.54}$$

where R is the redundancy matrix as defined by Eq.(6.40). From (6.54), we can find the effect of Δ_i on the adjusted $\widehat{\ell}_i$:

$$\Delta_i^{(\widehat{\ell}_i)} = (1-r_i)\cdot\widehat{\Delta}_i \tag{6.55}$$

[1] This classical definition of external reliability was orginally due to Baarda. See Baarda (1967), pp.13, Eq.(2.2.15) and Baarda (1968), pp.68, Eq.(11.17).

Combining Eq.(6.52) with Eq.(6.55), we arrive at the least detectable effect of Δ_i on the adjusted $\widehat{\ell}_i$:

$$\varphi_i = \left|\Delta_i^{(\widehat{\ell}_i)}\right|_{\min} = (1 - r_i) \cdot \frac{\sigma_i}{\sqrt{r_i}} \delta_0(\alpha, \beta) = (1 - r_i) \cdot \theta_i \tag{6.56}$$

In many geodetic literatures, φ_i defined above is simply called the *external reliability* of the observation ℓ_i . Eqs.(6.52) and (6.56) indicate that the local redundancy r_i functions also as reliability measures for the observations, in addition to its role as the accuracy measure.

Example 6.3

Assuming a theoretical variance factor $\sigma_0^2 = 22.5\ mm^2$, we can use the adjustment results of **Example 3.2** to calculate the local redundancy, internal and external reliability for each of the five observations.

According to **Example 3.2**, the weight matrix of L or ε is :

$$\underset{5\cdot 5}{P} = \begin{bmatrix} 2 & & & & \\ & 2 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 2 \end{bmatrix}$$

and the cofactor matrix of the least squares residuals $\widehat{\varepsilon}$ is :

$$Q_{\widehat{\varepsilon}\widehat{\varepsilon}} = \frac{1}{28} \begin{bmatrix} 5 & 5 & 2 & 2 & -4 \\ 5 & 5 & 2 & 2 & -4 \\ 2 & 2 & 12 & 12 & 4 \\ 2 & 2 & 12 & 12 & 4 \\ -4 & -4 & 4 & 4 & 6 \end{bmatrix}$$

From Eq.(6.40), the redundancy matrix follows:

$$R = Q_{\widehat{\varepsilon}\widehat{\varepsilon}} P = \frac{1}{28} \begin{bmatrix} 5 & 5 & 2 & 2 & -4 \\ 5 & 5 & 2 & 2 & -4 \\ 2 & 2 & 12 & 12 & 4 \\ 2 & 2 & 12 & 12 & 4 \\ -4 & -4 & 4 & 4 & 6 \end{bmatrix} \begin{bmatrix} 2 & & & & \\ & 2 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 2 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 5 & 5 & 1 & 1 & -4 \\ 5 & 5 & 1 & 1 & -4 \\ 2 & 2 & 6 & 6 & 4 \\ 2 & 2 & 6 & 6 & 4 \\ -4 & -4 & 2 & 2 & 6 \end{bmatrix}$$

The local redundancy of each observation is :

$$\begin{cases} r_1 = \frac{5}{14} \\ r_2 = \frac{5}{14} \\ r_3 = \frac{6}{14} \\ r_4 = \frac{6}{14} \\ r_5 = \frac{6}{14} \end{cases} \qquad \left(check : \sum_{i=1}^{5} r_i = 2 = n - m \right)$$

The standard error of each observation can be calculated from Eq.(6.51) :

$$\begin{cases} \sigma_1 = \sigma_0/\sqrt{p_1} = 3.35\ ^{mm} \\ \sigma_2 = \sigma_0/\sqrt{p_2} = 3.35\ ^{mm} \\ \sigma_3 = \sigma_0/\sqrt{p_3} = 4.74\ ^{mm} \\ \sigma_4 = \sigma_0/\sqrt{p_4} = 4.74\ ^{mm} \\ \sigma_5 = \sigma_0/\sqrt{p_5} = 3.35\ ^{mm} \end{cases}$$

Taking $\alpha = 5\%$, $\beta = 20\%$ and $\delta_0 \approx 2.80$, we obtain the internal reliability from Eq.(6.52) :

$$\begin{cases} \theta_1 = 2.8 \times \sigma_1/\sqrt{r_1} = 4.7\sigma_1 = 15.7\ ^{mm} \\ \theta_2 = 2.8 \times \sigma_2/\sqrt{r_2} = 4.7\sigma_2 = 15.7\ ^{mm} \\ \theta_3 = 2.8 \times \sigma_3/\sqrt{r_3} = 4.3\sigma_3 = 20.3\ ^{mm} \\ \theta_4 = 2.8 \times \sigma_4/\sqrt{r_4} = 4.3\sigma_4 = 20.3\ ^{mm} \\ \theta_5 = 2.8 \times \sigma_5/\sqrt{r_5} = 4.3\sigma_5 = 14.3\ ^{mm} \end{cases}$$

and finally the external reliability from Eq.(6.56) :

$$\begin{cases} \varphi_1 = (1 - r_1)\theta_1 = 10.1 \ ^{mm} \\ \varphi_2 = (1 - r_2)\theta_2 = 10.1 \ ^{mm} \\ \varphi_3 = (1 - r_3)\theta_3 = 11.6 \ ^{mm} \\ \varphi_4 = (1 - r_4)\theta_4 = 11.6 \ ^{mm} \\ \varphi_5 = (1 - r_5)\theta_5 = \ 8.2 \ ^{mm} \end{cases}$$

Bibliography

[1] Baarda, W. (1967). *Statistical Concepts in Geodesy.* Neth. Geod. Comm. Publ. on Geodesy, Vol. 2, No.4.

[2] Baarda, W. (1968). *A Testing procedure for Use in Geodetic Networks.* Neth. Geod. Comm. Publ. on Geodesy, Vol. 2, No.5.

[3] Baarda, W. (1973). *S-Transformations and Criterion Matrices.* Neth. Geod. Comm. Publ. on Geodesy, Vol. 5, No.1.

[4] Baarda, W. (1977). *Measures for the Accuracy of Geodetic Networks.* Presented at the IAG Symposium on Optimization of Design and Computation of Control Networks, July 3-9, 1977, Sopron, Hungary.

[5] Bjerhammar, A. (1973). *Theory of Errors and Generalized Matrix Inverses.* Elsevier, Amsterdam.

[6] Dahlquist, G., and Björk, Å. (1974). *Numerical Methods.* Prentice-Hall, New Jersey.

[7] Ebner, H. (1972). *A posteriori Varianzschätzungen für die Koordinaten unabhängiger Modelle.* ZfV, 4, pp.166-172.

[8] Enger, Jan. (1994). *Linjära modeller inklusive regressions- och variansanalysmodeller.* Department of Mathematical Statistics, KTH, Stockholm.

[9] Fan, H. and and Sjöberg, L.E. (1986). *On the Estimation of Variance-Covariance Components with and without Invariance in Gauss-Helmert Model.* TRITA GEOD 1007, Department of Geodesy, Royal Institute of Technology, December 1986.

[10] Förstner, W. (1979). *Konvergenzbeschleunigung ber der a posteriori Varianzschätzung.* ZfV, 4, pp.149-156.

[11] Grafarend, E., Kleusberg, A., and Schaffrin, B. (1980). *An Introduction to the Variance-Covariance Component Estimation of Helmert Type.* ZfV, 4, pp. 161-180.

[12] Granshaw, S.I. (1980). *Bundle Adjustment Methods in Engineering Photogrammetry.* Photogrammetric Records, 10(56): 181-207. October 1980.

[13] Hald, A. (1952). *Statistical Theory with Engineering Applications.* John Wiley & Sons.

[14] Heck, B. (1980). *Statistische Ausreisserkriterien zur Kontrolle Geodätischer Beobachtugen.* Inte. Kurs für Ingenieursvermessung, Zürich, 1980.

[15] Koch, K.R. (1980). *Parameterschätzung und Hypothesentests in linearen Modellen.* Dümmler-Bonn.

[16] Krarup, Torben, Jens Juhl and Kurt Kubik (198?). *Götterdämmerung over Least Squares Adjustment.* Presented at the 14th congress of the International Society of Photogrammetry.

[17] Kuang, Shanlong. (1996). *Geodetic Network Analysis and Optimal Design: Concepts and Applications.* Ann Arbor Press, Chelsea, Michigan.

[18] Kubik, Kurt. (1967). *Schätzung der Gewichte der Fehlergleichungen beim Ausgleichsproblem nach vermittelnden beobachtungen.* ZfV, 5, pp.173-178.

[19] Kubik, Kurt (1970). *The Estimation of the Weights of the Measured Quantities within the Method of Least Squares.* Bull. Geod. No. 95.

[20] Li, Deren (1983). Bildmessung und Luftbildwessen, No 5, 1983.

[21] Luenberger, D.G. (1969). *Optimization by Vector Space Method*. John Willy and Sons.

[22] Mikhail, E.M. (1976). *Observations and Least Squares*. Harper and Row, 1976.

[23] Mikhail, E.M., and Gracie, G. (1981). *Analysis and Adjustment of Survey Measurements.* Van Nostrand Reinhold, New York.

[24] Papo, H.B. and Perelmuter, A. (1982). *Free Net Analysis in Close-Range Photogrammetry.* Photogrammetric Engineering and Remote Sensing, Vol.48, No.4, April 1982, pp.571-576.

[25] Pope, A.J. (1975). *The Statistics of Residuals and the Detection of Outliers.* XVI-th IUGG General Assembly, Grenoble, 1975.

[26] Persson, Clas-Göran. (1981). *On the Estimation of Variance Components in Linear Models and Related Models with Special Reference to Geodetic Applications.* octoral Dissertation. Department of Geodesy, Royal Institute of Technology, Stockholm.

[27] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications.* 2nd Edition. John Wiley & Sons.

[28] Rao, C.R., and Kleffe, J. (1988). *Estimation of Variance Components and Applications.* Elsevier Science Publishers.

[29] Rao, C.R. and S.K. Mitra (1971). *Generalized Inverses of Matrices and Its Apllications.* John Wiley & Sons.

[30] Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components.* John Wiley & Sons.

[31] Sjöberg, L.E. (1983). *Unbiased Estimation of Variance-Covariance Components in Condition Adjustment with Unknowns — A MINQUE Approach.* ZfV., 108.

[32] Sjöberg, L.E. (1984a). *Lecture Notes on General Matrix Calculus, Adjustment and Variance-Covariance Component Estimation.* Nordiska forskarkurser ″Optimization of Geodetic Operations″. Norges Geogr. Oppm. Publ. 3/1984.

[33] Sjöberg, L.E. (1984b). *Non-negative Variance Component Estimation in Linear Adjustment Models.* Manuscripta Geodaetica, Vol.9, pp.247-280, 1984.

[34] Sjöberg, L.E. (1990). *General Matrix Calculus and Prediction Theory.* TRITA GEOD 2010, Department of Geodesy, KTH, Stockholm, December 1990.

[35] Srivastava, M.S. and C.G. Khatri (1979). *An Introduction to Multivariate Statistics.* Elsevier North Holland, Inc.

[36] Yu, Z. and L. Lu. (1982). *Surveying Adjustment* (in Chinese). Surveying and Mapping Publishing House, Beijing, China.