Theory of Errors and Least Squares Adjustment

Part 2

Huaan Fan

2024

Royal Institute of Technology
Department of Geodesy and Photogrammetry
Stockholm, Sweden

January 1996

Theory of Errors and Least Squares Adjustment

Part 2

Huaan Fan

Royal Institute of Technology
Department of Geodesy and Photogrammetry
Stockholm, Sweden

January 1996

~ 1998 KTH ~
~ Stockholm ~
Curso 6. Física
Autocurso T. Errores +
~ Gen. Matrix ~

Preface

Theory of errors and adjustment has been a well defined subject within geodesy, photogrammetry and surveying. Currently there exist several masterpieces on this subject, such as the famous works by Bjerhammar (1973), Mikhail (1976) and Koch (1981), just to mention a few of them. However, while teaching theory of errors at KTH since 1992, 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, photogrammetry and surveying.

To meet the above demand, an attempt was started in the Spring of 1995, which has resulted in this compendium in its preliminary version. Naturally, it is neither the author's ambition nor within his capability to challenge those great masterpieces mentioned before. The compendium will primarily be used as teaching materials for courses on theory of errors and adjustment at KTH.

The compendium consists of four chapters. Chapter §1 deals with basic concepts in theory of errors, such as error types, standard error and its propagation, error ellipse, linear equation systems as well as some elementary concepts of probability theory. 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.

Chapter §4 deals with diverse topics based on recent development in the field of theory of errors. These topics include generalized matrix inverses and their applications in geodesy; 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 basic 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 Gauss Table for solving normal equations and related concepts) 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. For those examples with more realistic data, the presented solutions are obtained on an Intel Pentium PC using Lahey Fortran Compiler (version 5.1). Minor decimal differences may occur if the same examples are calculated using other hardware and software.

It is a great pleasure for the author to acknowledge kind help and encouragement received during the preparation of this compendium. I would like to thank Professor Lars E Sjöberg for providing the opportunity to write the compendium

and Tech.Lic. Kjell Almgren for proof–reading the manuscript. Of particular benefit have been my students in the class L92–MK at KTH, who have worked through the raw materials as the compendium evolved. To all of them, I express my sincere thanks !

Last but not least, I wish to thank my wife Helene and my son Daniel for sacrificing many evenings and weekends, as a significant part of the manuscript was written on my home PC during the Spring of 1995.

Stockholm, January 1996.

Huaan Fan

§4 Advanced Topics

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 many new advancement in geodetic theory and practices, including geodetic data processing. Classical geodetic adjustment techniques have been given a more sound theoretical foundation, mostly due to application of statistical and probability theory. New problems have been identified and new methods developed for processing geodetic data with the help of modern electronic computers. To reflect these new developments, we will devote this chapter to four advanced topics in theory of errors and geodetic adjustment.

The first topic concerns generalized matrix inverses and their application in geodetic adjustment. Here we extend the classical concepts of matrix inverses in order to obtain general solutions to various kinds of linear equation systems. Traditional geodetic adjustment methods are then treated as specific types of linear equation systems.

The second topic aims at a posteriori estimation of the accuracy of observation data. The normal procedure in geodetic adjustment is to start with some a priori statistical model of observation accuracy together with the well–defined analytical model (e.g. condition equations or observation equations). By introducing the concepts of variance–covariance components, it will be shown that one can estimate the accuracy of the existing observation data.

The next topic deals with gross error detection and reliability problems. With the help of statistical theory and electronic computing techniques, we try to design automatic procedures for detecting possible gross errors or other abnormal, non–random errors in our measurement data. The concept of reliability is introduced in order to better describe the quality of observation data and computed results, not only their accuracy but also their strength against possible gross errors or systematical errors.

The last topic comes from the need to process large amount of observation data, collected likely in real time and often related to dynamic systems or processes. We will discuss the classical theory of Wiener–Hopf prediction as well as the well–known Kalman filtering technique. This topic is of great importance, not only in geodetic data processing (such as satellite orbit computation, kinematic positioning using the Global Positioning System) but also in many other engineering fields like automatic control, electronic communications and space technology.

§4.1 Generalized Inverses and Geodetic Adjustment

NO VA MAS

Numerically, geodetic adjustment may be regarded as solving different types of linear linear equation systems like:

$$\underset{n\,x\,m}{A}\ \underset{m\,x\,1}{X} = \underset{n\,x\,1}{L}\ \left[- \underset{n\,x\,1}{\epsilon} \right] \qquad (4.1.1a)$$

where the unspecified term inside the bracket is needed when an inconsistent equation system (i.e. $AX \neq L$) is concerned. If the above system is consistent and A is a non–singular square matrix of nxn, the solution can be uniquely expressed by the inverse matrix of A :

$$\underset{n\,x\,1}{X} = \underset{n\,x\,n}{A^{-1}}\ \underset{n\,x\,1}{L} \qquad (4.1.1b)$$

The inverse matrix A satisfies the classical definition in matrix algebra (I = unit matrix):

$$A\,A^{-1} = A^{-1}\,A = I \qquad (4.1.1c)$$

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. Nevertheless, one may still wish to express the solution of an arbitrary equation system (4.1.1a) in a simple, intuitive and elegant way just as (4.1.1b) for the case of a non–singular square matrix A:

$$X = G\,L \qquad (4.1.1d)$$

where G is a matrix different from the classical inverse of A, but functions like an inverse matrix for the purpose of constructing the solution of (4.1.1a).

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.

In this section, we are going to define the generalized inverse matrix of any arbitrary matrix. Using the generalized matrix inverses, solutions to different types of linear equation systems (e.g. condition equations, observation equations) will be discussed under different criteria. Historically, the former Department of Geodesy at KTH has played an important role in the field of generalized inverses, mainly due to prof Arne Bjerhammar's pioneer work in this 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).

§4.1.1 Generalized Matrix Inverses A^-

For any arbitrary matrix $\underset{n\,x\,m}{A}$, a matrix A^- is called a generalized inverse of A iff:

$$\underset{n\times m}{A}\ \underset{}{A^{-}}\ \underset{n\times m}{A} = \underset{n\times m}{A} \qquad (4.1.1)$$

The traditional inverse of a non–singular matrix is automatically a generalized inverse, but not vice versa. A^{-} must have dimension mxn in order for matrix multiplications to be meaningful. A generalized inverse defined by (4.1.1) is also called a *g–inverse*. A g–inverse of A is normally denoted as A^{-} to distinguish from an ordinary inverse A^{-1} when $|A| \neq 0$. Exceptions are those unique generalized inverses in Section 4.1.2 where the superscript $^{-1}$ is still used to denote the generalized inverses.

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, i.e.

$$r(AB) \leq \min\{r(A), r(B)\} \qquad (4.1.2)$$

Applying Eq. (4.1.2) on Eq. (4.1.1), we have:

$$r(A) = r(AA^{-}A) \leq \min\{r(A), r(A^{-})\}$$

which implies:

$$r(A^{-}) \geq r(A) \qquad (4.1.3)$$

The generalized inverse A^{-} is not unique. This can be seen from Example 4.1.1 below.

Example 4.1.1

Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $C = \begin{bmatrix} 0 & 0 \\ 0 & 1/4 \end{bmatrix}$

We then have: $ABA = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = A\,!$

and $ACA = \begin{bmatrix} 0 & \frac{1}{2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = A\,!$

which means that both B and C are generalized inverse of A although $B \neq C$! ∎

Example 4.1.2

Assume that matrix $\underset{n\times m}{A}$ has rank $r \equiv r(A) \leq \min(n, m)$ 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:

$$\underset{n\,x\,m}{A} = \begin{bmatrix} \underset{r\,x\,r}{A_{11}} & \underset{r\,x\,s}{A_{12}} \\ \underset{t\,x\,r}{A_{21}} & \underset{t\,x\,s}{A_{22}} \end{bmatrix} \tag{4.1.4a}$$

where $n = r + t$, $m = r + s$, $r(A) = r(A_{11}) = r$ and $|A_{11}| \neq 0$. Then one particular generalized inverse of A is :

$$\underset{m\,x\,n}{A^-} = \begin{bmatrix} \underset{r\,x\,r}{A_{11}^{-1}} & \underset{r\,x\,t}{0} \\ \underset{s\,x\,r}{0} & \underset{s\,x\,t}{0} \end{bmatrix} \tag{4.1.4b}$$

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). Let α be a matrix of dimension txr, then we have:

$$(A_{21}, A_{22}) = \alpha\,(A_{11}, A_{12})$$

which leads to $\alpha = A_{21}\,A_{11}^{-1}$ and $A_{22} = A_{21}\,A_{11}^{-1}\,A_{12}$. Finally we have:

$$AA^-A = \begin{bmatrix} \underset{r\,x\,r}{I} & \underset{r\,x\,t}{0} \\ \underset{t\,x\,r}{A_{21}\,A_{11}^{-1}} & \underset{t\,x\,t}{0} \end{bmatrix} \begin{bmatrix} \underset{r\,x\,r}{A_{11}} & \underset{r\,x\,s}{A_{12}} \\ \underset{t\,x\,r}{A_{21}} & \underset{t\,x\,s}{A_{22}} \end{bmatrix} = \begin{bmatrix} \underset{r\,x\,r}{A_{11}} & \underset{r\,x\,s}{A_{12}} \\ \underset{t\,x\,r}{A_{21}} & \underset{t\,x\,r\;\;r\,x\,r\;\;r\,x\,s}{A_{21}\,A_{11}^{-1}\,A_{12}} \end{bmatrix} = \begin{bmatrix} \underset{r\,x\,r}{A_{11}} & \underset{r\,x\,s}{A_{12}} \\ \underset{t\,x\,r}{A_{21}} & \underset{t\,x\,s}{A_{22}} \end{bmatrix} = A$$

which proves that (4.1.4b) is a generalized inverse of A. ∎

Example 4.1.3 (*Rank factorization*)

Let $\underset{n\,x\,m}{A}$ 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:

$$\underset{n\,x\,m}{A} = \underset{n\,x\,r}{B}\;\underset{r\,x\,m}{C}, \qquad r(A) = r(B) = r(C) = r \tag{4.1.5a}$$

Then a particular generalized inverse of A is given by

$$A^- = C^T\,(CC^T)^{-1}\,(B^TB)^{-1}\,B^T \tag{4.1.5b}$$

Proof

As $r(B) = r(C) = r$, B^TB and CC^T are square matrices with full rank r and should have ordinary inverses. Thus we have:

$$A\,A^-\,A = BC\,\underbrace{C^T\,(CC^T)^{-1}}_{I}\,\underbrace{(B^TB)^{-1}B^T}_{I}\,BC = BC = A.$$

One way to do rank factorization (4.1.5a) on a symmetrical matrix A of dimension nxn and rank r is to apply Cholesky triangular decomposition on A:

$$\underset{n\times n}{A} = \underset{n\times n}{T^{\top}}\ \underset{n\times n}{T}$$

where T is a upper–triangular matrix. If $r < n$, there must exist $(n-r)$ lines in T which contain only zeros. If we denote by $\bar{T}$ the matrix after removing all lines of zeros in T, we have:

$$\underset{n\times n}{A} = \underset{n\times r}{\bar{T}^{\top}}\ \underset{r\times n}{\bar{T}} = \underset{n\times r}{B}\ \underset{r\times n}{C} \qquad (4.1.5c)$$

where $B=\bar{T}^{\top}$ and $C=\bar{T}$. (4.1.5c) is also called truncated triangular decomposition. A numerical example is given below. ∎

Example 4.1.4 (*rank factorization with truncated triangular decomposition*)

Let: $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ with $r(A) = 1$

Triangular decomposition gives: $A = T^{\top}T = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$

After truncating, we have: $A = \bar{T}^{\top}\bar{T} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}[1\ 2] = BC$

where: $B^{\top} = C = [\,1\ \ 2\,]$, $B^{\top}B = C^{\top}C = 5$, $(B^{\top}B)^{-1} = (CC^{\top})^{-1} = 1/5$.

From (4.1.5b), we get a particular generalized inverse of A:

$$A^{-} = C^{\top}(CC^{\top})^{-1}(B^{\top}B)^{-1}B^{\top} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}\frac{1}{5}\frac{1}{5}[1\ 2] = \frac{1}{25}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

Check: $A\,A^{-}A = \frac{1}{25}\begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = A$! ∎ ✓

Previous examples show that the generalized inverse defined by (4.1.1) is not unique. In other words, for an arbitrary matrix A there exist more than one generalized inverses A^{-} which all satisfy definition (4.1.1). The following theorem shows that once a particular generalized inverse A^{-} is found, the complete set of generalized inverses of A (i.e. containing all possible inverses) can be represented in a general form.

Theorem 4.1

Let A^{-} be a particular generalized inverse of matrix $\underset{n\times m}{A}$. Then the complete set of generalized inverses of A, denoted as $\underset{m\times n}{\tilde{A}^{-}}$, is given by:

$$\underset{m\times n}{\tilde{A}^{-}} = \underset{m\times n}{A^{-}} + \underset{m\times n}{N}\left(\underset{n\times n}{I} - \underset{n\times m}{A}\ \underset{m\times n}{A^{-}}\right) + \left(\underset{m\times m}{I} - \underset{m\times n}{A^{-}}\ \underset{n\times m}{A}\right)\underset{m\times n}{M} \qquad (4.1.6)$$

where N and M are two arbitrary matrices with compatible dimensions.

Proof

Let $\tilde{A}^-$ denote the complete set of generalized inverse of A, A^- a particular inverse, and A_1^- a sub–set of $\tilde{A}^-$ defined as:

$$A_1^- = A^- + N(I - A A^-) + (I - A^- A) M$$

Considering $AA^-A = A$, we have:

$$A A_1^- A = A [A^- + N(I - A A^-) + (I - A^- A) M] A = A$$

which means A_1^- is a generalized inverse, i.e. $A_1^- \in \tilde{A}^-$ ✓

Letting $(M = \tilde{A}^- A A^-)$ and $(N = \tilde{A}^- - A^-)$ we obtain:

$$A_1^- = A^- + (\tilde{A}^- - A^-)(I - AA^-) + (I - A^-A)\tilde{A}^- AA^- = \tilde{A} \quad ✓$$

which means $\tilde{A}^- \in A_1^-$. So we should have $A_1^- \in \tilde{A}^- \in A_1^-$ which implies $\tilde{A}^- \equiv A_1^-$ or:

$$\tilde{A}^- = A^- + N(I - A A^-) + (I - A^- A) M. \quad ■ \quad ✓ (4.16)$$

Example 4.1.5

For $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$, we have a particular generalized inverse: $A^- = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Then the complete set of g–inverses of A can be obtained as follows:

$$AA^- = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}, \qquad I - AA^- = \begin{bmatrix} 0 & 0 \\ -2 & 1 \end{bmatrix}, \qquad I - A^-A = \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix}$$

$$\tilde{A}^- = A^- + N(I - A A^-) + (I - A^- A) M$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}. \quad ■ \quad ✓$$

§4.1.2 Special Generalized Inverses

For a matrix A, (4.1.1) defines the general g–inverse A^-. Among all possible g–inverses of A, one can identify several special g–inverses. In this subsection, we first present a useful theorem on generalized matrix inverses and then introduce several special generalized inverses.

Theorem 4.2

Let matrix $\underset{n \times m}{A}$ have rank r and let R , S be two arbitrary matrices of compatible

dimensions. $(RA)^-R$ and $S(AS)^-$ are both generalized inverse of A if $r(RA)=r(AS)=r$.

Proof:

We first use rank factorization to prove the theorem for a particular generalized inverse of RA :

$$\underset{n\times m}{A} = \underset{n\times r}{B}\ \underset{r\times m}{C}, \qquad r(A) = r(B) = r(C) = r$$

$$RA = RB\,C = B_1\,C, \quad B_1 \equiv RB, \qquad r(A) = r(B_1) = r(C) = r$$

$$|B_1^TB_1| \neq 0,\ |CC^T| \neq 0$$

$$(RA)^- = C^T(CC^T)^{-1}(B_1^TB_1)^{-1}B_1^T$$

$$(RA)^-R = C^T(CC^T)^{-1}(B_1^TB_1)^{-1}B_1^T\,R$$

falta $(B_1^TB_1)^{-1}$

aplico def. (A^-)

$$A\,(RA)^-R\,A = A\,C^T(CC^T)^{-1}B_1^T\,R\,A = B\,CC^T(CC^T)^{-1}\,(B_1^TB_1)^{-1}B_1^TR\,BC$$

($A = B\,CC^T$)

$$= B\,(B_1^TB_1)^{-1}B_1^TB_1\,C = BC = A\,. \quad \checkmark \quad I$$

i.e.: $(RA)^-R$ is a generalized inverse of A for the particular g–inverse $(RA)^-$ of RA. Then we prove the theorem for any arbitrary g–inverse $R\tilde{A}$ of RA:

$$(\widetilde{RA})^-R = \left[\,(RA)^- + N[I-(RA)(RA)^-] + [I-(RA)^-(RA)]M\,\right] R$$

$$A\,(\widetilde{RA})^-R\,A = A\left\{(RA)^- + N[I-(RA)(RA)^-] + [I-(RA)^-(RA)]M\right\} R\,A$$

$$= A\,(RA)^-R\,A + AN[RA - (RA)(RA)^-(RA)] + [A - A\,(RA)^-R\,A)]\,MRA = A$$

($(RA)^-$ = A^-; $N[RA-(RA)(RA)^-(RA)] = \phi$; $A-A\,(RA)^-R\,A = \phi$)

i.e. $(RA)^-R$ is a generalized inverse of A for any arbitrary choice of $(RA)^-$. ∎ ✓

In similar way, one can prove the theorem for $S(AS)^-$. Now we are ready to present three types of special generalized inverses, namely the minimum norm inverses, the least squares inverses and the minimum norm–least squares inverse.

Minimum Norm Inverses

Let $\underset{n\times m}{A}$ denote a matrix of rank r and $\underset{m\times m}{Q}$ a positive definite matrix. A minimum norm generalized inverse of A is defined as:

$$A_{QO}^- \equiv QA^T(AQA^T)^-\,, \qquad (4.1.7)$$

where $(AQA^T)^-$ is any generalized inverse of AQA^T.

As Q is positive definite, we have $r(QA^T) = r(A) = r$. From Theorem 4.2, it follows that A_{QO}^- is a g–inverse of A (here $S \equiv QA^T$). As the g–inverse $(AQA^T)^-$ in (4.1.7) is normally not unique, the minimum norm inverse A_{QO}^- consequently is not unique

es de la forma

either. The complete set of A_{QO}^{-} can be obtained by using the complete set of g–inverses of AQA^T in Eq. (4.1.7). Letting $(AQA^T)^-$ be one particular g–inverse of AQA^T, the complete set of the g–inverse of AQA^T follows from Eq. (4.1.6):

$$(A\tilde{Q}A^T)^- = (AQA^T)^- + \bar{N}\,[I - (AQA^T)\,(AQA^T)^-] + [I - (AQA^T)^-\,(AQA^T)]\,\bar{M}$$

where $\bar{N}$ and $\bar{M}$ are arbitrary matrices of compatible dimensions.

The complete set of minimum norm inverses follows:

$$\begin{aligned}
\tilde{A}_{QO}^{-} &= QA^T(A\tilde{Q}A^T)^- \\
&= QA^T\left[(AQA^T)^- + \bar{N}\,[I - (AQA^T)\,(AQA^T)^-] + [I - (AQA^T)^-\,(AQA^T)]\,\bar{M}\right] \\
&= A_{QO}^{-} + N\,[I - A\,A_{QO}^{-}] + [QA^T - (QA^T)\,(QA^T)^-\,(QA^T)]\bar{M} \\
&= A_{QO}^{-} + N\,[I - A\,A_{QO}^{-}]\,, \qquad (4.1.8)
\end{aligned}$$

where $N \equiv QA^T\bar{N}$ denotes a new arbitrary matrix. In the derivations above, we have used Theorem 4.2 to obtain a g–inverse of QA^T.

The general minimum norm inverses given by (4.1.8) is dependent only on one arbitrary matrix N, instead of two arbitrary matrices as in Eq. (4.1.6), which indicates that A_{QO}^{-} belongs to a special sub–set of A^-. Why we call A_{QO}^{-} minimum norm inverses is related to the property of solutions of a linear equation system that has A as coefficient matrix. This will be discussed in §4.1.5.

If the rank of matrix $\underset{n\,x\,m}{A}$ 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 $r(AQA^T) = n$ or $|AQA^T| \neq 0$ and consequently the classical inverse $(AQA^T)^{-1}$ is well defined. In this case [i.e.: r(A)=n], we can write A_{QO}^{-1} as:

$$A_{QO}^{-1} = QA^T(AQA^T)^{-1} \qquad (4.1.9a)$$

for which we get:

$$A\,A_{QO}^{-1} = I \qquad (4.1.9b)$$

Least Squares Inverses

Let $\underset{n\times n}{P}$ be a positive definite matrix. The least squares inverses of matrix $\underset{n\times m}{A}$ are defined as:

$$A_{OP}^{-} = (A^TPA)^{-} A^TP \tag{4.1.10}$$

where $(A^TPA)^{-}$ denotes any arbitrary g–inverse of (A^TPA).

As $(A^TPA)^{-}$ is usually non–unique, A_{OP}^{-} is normally not unique either. The complete set of A_{OP}^{-} can be derived in the same way as the derivation of Eq. (4.1.8). Let $(A^TPA)^{-}$ and $(\widetilde{A^TPA})^{-}$ denote a particular g–inverse and the arbitrary g–inverse of A^TPA, respectively. Furthermore, let A_{OP}^{-} and $\tilde{A}_{OP}^{-}$ denote a particular and the arbitrary least squares inverse of A, respectively. Then we have:

$$\begin{aligned}
\tilde{A}_{OP}^{-} &= (\widetilde{A^TPA})^{-} A^TP \\
&= \left[(A^TPA)^{-} + \overline{N}\,[I - (A^TPA)\,(A^TPA)^{-}] + [I - (A^TPA)^{-}\,(A^TPA)]\,\overline{M} \right] A^TP \\
&= A_{OP}^{-1} + \overline{N}\,[A^TP - (A^TP)\,(A^TP)^{-}\,(A^TP)] + [I - (A^TPA)^{-}(A^TPA)]\,\overline{M}A^TP \\
&= A_{OP}^{-} + [I - A_{OP}^{-}\,A]\,M
\end{aligned} \tag{4.1.11}$$

where $M \equiv \overline{M}A^TP$ denotes a new arbitrary matrix.

If $r(A) = m$ holds, then $r(A^TPA) = m$ and $|A^TPA| \neq 0$. Consequently, $(A^TPA)^{-1}$ is well defined and A_{OP}^{-1} becomes a unique g–inverse:

$$A_{OP}^{-1} = (A^TPA)^{-1}A^TP \tag{4.1.12}$$

with:

$$A_{OP}^{-1}\,A = I \tag{4.1.13}$$

Minimum Norm–Least Squares Inverse

Theorem 4.3

For any matrix $\underset{n\times m}{A}$, if S and R are two arbitrary matrices such that

$$r(A) = r(RA) = r(AS) = r(RAS) = r, \tag{4.1.14}$$

then the following g–inverse of A is unique:

$$A_{SR}^{-1} = S\,(RAS)^{-}\,R \;=\; S\,(AS)^{-}\,A\,(RA)^{-}\,R \qquad (4.1.15)$$

Proof:

We first prove that A_{SR}^{-1} in (4.1.15) is a g–inverse of A:

$$A\,S(RAS)^{-}R\,A = A\,(RA)^{-}\,R\,A = A\,A^{-}\,A = A \quad \checkmark$$
$$A\,S\,(AS)^{-}\,A\,(RA)^{-}\,R\,A = A\,A^{-}\,A\,A^{-}\,A = A\,A^{-}\,A = A \quad \checkmark$$

Provar que é uma A^{-}!

We then prove that $A_{SR}^{-1} = S\,(RAS)^{-}\,R$ is unique. Let A_{SR}^{-1} and $\tilde{A}_{SR}^{-1}$ denote a particular and the general form of g–inverse given by Eq. (12b), respectively. Let $(RAS)^{-}$ and $(R\tilde{A}S)^{-}$ denote a particular and the general g–inverse of RAS, respectively. Then we have:

$$\tilde{A}_{SR}^{-1} = S(R\tilde{A}S)^{-}R$$
$$= S\left\{(RAS)^{-} + N\,[I - (RAS)\,(RAS)^{-}] + [I - (RAS)^{-}\,(RAS)]\,M\right\} R$$
$$= S(RAS)^{-}R + SN\,[R - (RAS)\,(RAS)^{-}R] + [S - S\,(RAS)^{-}\,(RAS)]\,MR$$
$$= A_{SR}^{-1} + SN\,[R - R\,R^{-}\,R] + [S - S\,S^{-}\,S]\,MR$$
$$\tilde{A}_{SR}^{-1} \equiv A_{SR}^{-1}. \quad \checkmark$$

Similarly, one can prove $S\,(AS)^{-}\,A\,(RA)^{-}\,R$ is a unique g–inverse of A. Details of the proof are left to the reader. ∎

As a special case of Theorem 4.3, let $S = QA^{\top}$ and $R = A^{\top}P$, where $\underset{n\times n}{P}$ and $\underset{m\times m}{Q}$ are two positive definite matrices. It is trivial to see that such S and R satisfy the condition (4.1.14). By inserting S and R into Eq. (4.1.15), we get a unique generalized inverse of A:

$$A_{QP}^{-1} = QA^{\top}(A^{\top}PAQA^{\top})^{-}A^{\top}P \doteq QA^{\top}(AQA^{\top})^{-}A(A^{\top}PA)^{-}A^{\top}P \qquad (4.1.16)$$

(e também)

As another special case, let $\underset{n\times n}{P} = \underset{n\times n}{I}$ and $\underset{m\times m}{Q} = \underset{m\times m}{I}$ and we get from Eq. (4.1.15) the so called *pseudo–inverse* or *Moore–Penrose inverse*:

$$A_{II}^{-1} = A^{\top}(A^{\top}AA^{\top})^{-}A^{\top} = A^{\top}(AA^{\top})^{-}A(A^{\top}A)^{-}A^{\top} \qquad (4.1.17)$$

(1ª versão) (2ª versão)

In some literature, A_{II}^{-1} is denoted by A^{+}.

A_{QP}^{-1} can be written in another form:

$$A_{QP}^{-1} = A_{QI}^{-1}\,A\,A_{IP}^{-1} \qquad (4.1.18)$$

To prove (4.1.18), we insert separately $Q = I$ and $P = I$ into (4.1.16) to obtain:

$$A_{QI}^{-1} = QA^T (AQA^T)^- A (A^TA)^- A^T$$

$$A_{IP}^{-1} = A^T (AA^T)^- A (A^TPA)^- A^TP$$

$$A_{QI}^{-1} A A_{IP}^{-1} = QA^T (AQA^T)^- A (A^TA)^- A^T A A^T (AA^T)^- A (A^TPA)^- A^TP$$

$$= QA^T (AQA^T)^- A A^- A A^- A (A^TPA)^- A^TP$$

$$= QA^T (AQA^T)^- A (A^TPA)^- A^TP$$

$$= A_{QP}^{-1}. \blacksquare$$

4.1.3 Computation of Minimum Norm–Least Squares Inverses

For given P and Q, A_{QP}^{-1} can be computed directly from (4.1.16). In addition, it can be computed by the following alternative methods.

A Orthogonal Bordering

Let $\underset{n\,x\,m}{A}$ have rank r and rank defect $k \equiv \min(n,m) - r$.

$$A_{QP}^{-1} = A_{QI}^{-1} A A_{IP}^{-1} \tag{4.1.19}$$

where:

$$A_{QI}^{-1} = QA^T(AQA^T + CC^T)^{-1},$$

$$A_{IP}^{-1} = (A^TPA + D^TD)^{-1}A^TP$$

and $\underset{n\,x\,k}{C}$ and $\underset{m\,x\,k}{D^T}$ are two matrices such that:

$$A^T \underset{n\,x\,k}{C} = 0, \; |AQA^T + CC^T| \neq 0$$

and $A \underset{m\,x\,k}{D^T} = 0 \;\; |A^TPA + D^TD| \neq 0$

B Limiting Value Method

$$A_{QP}^{-1} = \lim_{\delta \to 0} QA^T(AQA^T + \delta P^{-1})^{-1} \tag{4.1.20}$$

$$A_{QP}^{-1} = \lim_{\delta \to 0} (A^TPA + \delta Q^{-1})^{-1}A^TP \tag{4.1.21}$$

C Truncated Triangular Decomposition

$$A_{QP}^{-1} = A_{QI}^{-1} A A_{IP}^{-1} \tag{4.1.22}$$

$$A_{QI}^{-1} = QA^T(AQA^T)_{II}^{-1} = QA^TT^T(TT^T)^{-1}(TT^T)^{-1}T, \; AQA^T = T^TT \tag{4.1.23}$$

$$A_{IP}^{-1} = (A^TPA)_{II}^{-1}A^TP = T^T(TT^T)^{-1}(TT^T)^{-1}TA^TP, \; A^TPA = T^TT \tag{4.1.24}$$

Matrix T in (4.1.23) and (4.1.24) is obtained after truncated triangular decomposition (Cf Example 4.1.3 and Example 4.1.4).

D Partitioning Method

i) Assume that matrix $\underset{n\,x\,m}{A}$ can be partitioned **by columns** into several sub–matrices, one of which (say A_i) has the same rank as A:

$$\underset{n\,x\,m}{A} = (A_1, A_2, \cdots, A_i, \cdots), \qquad r(A_i) = r(A) = r$$

$$\text{then:} \quad A_{QP}^{-1} = QB^T(BQB^T)^{-1} A_i^T P, \qquad B = A_i^T PA \tag{4.1.25}$$

ii) Assume that matrix $\underset{n\,x\,m}{A}$ can be partitioned **by rows** into several sub–matrices, one of which (say A_i) has the same rank as A:

$$\underset{n\,x\,m}{A} = \begin{bmatrix} A_1 \\ A_2 \\ \cdots \\ A_i \\ \cdots \end{bmatrix}, \quad r(A_i) = r(A) = r$$

$$\text{then:} \quad A_{QP}^{-1} = QA_i^T(B^TPB)^{-1} B^TP, \qquad B = AQA_i^T \tag{4.1.26}$$

E Iterative Method (for computing A_{II}^{-1} or A^+)

Initial inverse: $A_0^+ = \delta A^T$ where δ is a very small positive number.

Correction: $V_{k+1} = A_k^+ (I - AA_k^+), \; k = 0, 1, 2, \cdots$ (4.1.27)

New inverse: $A_{k+1}^+ = A_k^+ + V_{k+1}$ (4.1.28)

The iteration will continue until the matrix norm of the correction V, $\|V\| \equiv [tr(V^TV)]^{\frac{1}{2}}$, becomes sufficiently small.

4.1.4 General Solutions of A Consistent Linear Equation System

A linear equation system is said to be consistent if it has at least one set of solutions. Theorem 4.4 below tells us how to judge whether an equation system is consistent.

Theorem 4.4

A linear equation system, $\underset{n\times m}{A}\ \underset{m\times 1}{X} = \underset{n\times 1}{L}$, is consistent iff (≡ if and only if)

$$A\,A^{-}L = L \tag{4.1.29}$$

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 by the fact that if $AX = L$ has a solution, say X_1, we have $AX_1 = L$ which implies $AA^{-}A\,X_1 = L$ or $AA^{-}L = L$. ∎

Example 4.1.6

For equation system: $\begin{bmatrix} 2 & -1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ <=> $AX = L$, we can find:

$$A^{-} = \begin{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}^{-1} & 0 \\ 0 & \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 0 \\ -1 & 2 & 0 \end{bmatrix}$$

$$AA^{-}L = \frac{1}{3}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ -1 & 2 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = L$$

According to Theorem 4.4, $AX = L$ is not consistent. This can be shown by looking at the solution of the first two equations of $AX=L$:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

which is inconsistent with the third equation which gives: $x_2 = 2$.

Theorem 4.5

The general solution or the complete set of solutions of a consistent linear equation system $AX = L$ is given by

$$\tilde{X} = A^{-}L + (I - A^{-}A)\,V = \tilde{A}^{-}L \tag{4.1.30}$$

where A^{-}, $\tilde{A}^{-}$ denote a particular and the general g–inverse of A, respectively, and V denotes an arbitrary vector.

Proof

Let $\hat{X}$ 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 $AA^{-}L = L$ according to Theorem 4.4. Thus one has

$$A\,X_1 = AA^{-}L + (A - AA^{-}A)\, V = AA^{-}L = L$$

That is to say, X_1 is a solution of $AX = L$ or $X_1 \in \hat{X}$.

Setting $V \equiv \hat{X}$ in the defining equation of X_1 above gives:

$$X_1 = A^{-}L + (I - A^{-}A)\, \hat{X} = A^{-}L + \hat{X} - A^{-}L = \hat{X}$$

which implies: $\hat{X} \in X_1$. Thus we have $X_1 \in \hat{X} \in X_1$ which leads to $X_1 \equiv \hat{X}$.

To prove $\hat{X} = \hat{A}^{-}L$, we simply multiply Eq. (4.1.6) by L:

$$\begin{aligned} \hat{A}^{-}L &= [\, A^{-} + N\,(I - AA^{-}) + (I - A^{-}A)\, M\,]\, L \\ &= A^{-}L + N\,(I - AA^{-})\, AX + (I - A^{-}A)\, ML = A^{-}L + (I - A^{-}A)\, ML \\ &= A^{-}L + (I - A^{-}A)\, V = \hat{X} \qquad (V \equiv ML) \quad \blacksquare \end{aligned}$$

Example 4.1.7

Assume we have the following equation system:

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \qquad <=>\ A\,X\ =\ L$$

We then have:

$$A^{-} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\quad A^{-}A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},\quad I - A^{-}A = \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix}$$

$$AA^{-}L = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} = L \implies AX = L \text{ is consistent.}$$

The general solution of AX=L is given by Theorem 4.5:

$$\hat{X} = A^{-}L + (I - A^{-}A)\, V = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} -v \\ v \end{bmatrix} = \begin{bmatrix} 2-v \\ v \end{bmatrix}. \quad \blacksquare$$

4.1.5 Minimum Norm Solution

Assume that we have a consistent condition equation system:

$$\underset{n\times m}{A}\ \underset{m\times 1}{X} = \underset{n\times 1}{L}, \qquad m > n \geq r \equiv r(A) \tag{4.1.31}$$

(4.1.31) corresponds to the functional model of condition adjustment (2.1.2). More precisely, matrix A above corresponds to matrix B in (2.1.2), X corresponds to the residual vector and L corresponds to the discrepancy vector W. In condition adjustment, we also assume a statistical model for the residual vector X, namely the mathematical expectation and the variance matrix (Cf Eqs. 2.1.3 and 2.1.4):

$$E(X) = 0\ ; \quad E(XX^T) = \sigma_0^2 \underset{m\times m}{Q} \tag{4.1.32}$$

where σ_0^2 denotes the variance factor and Q the cofactor matrix or inverse weight matrix of X, which is assumed to be positive definite. As discussed in Section §2.1, the main objective of condition adjustment is to find an optimal solution of X such that (4.1.31) and the following minimization requirement are satisfied:

$$\hat{X}^T Q^{-1} \hat{X} = \text{minimum} \quad (\text{least squares}) \tag{4.1.33}$$

(4.1.33) is equivalent to minimizing the Euclidean norm of vector $\hat{X}$ (with respect to Q):

$$\|X\| \equiv (X^T Q^{-1} X)^{\frac{1}{2}} = \text{minimum} \tag{4.1.34}$$

From the relation $m > n \geq r \equiv r(A)$, one can show that (4.1.31) is consistent and that thus has infinite number of solutions. Among all possible solutions to $AX = L$, the solution that attains the minimum Euclidean norm (4.1.34) can be called the *minimum norm solution* of $AX = L$. This solution should be the same as the solution of condition adjustment, as both are defined by (4.1.31) and (4.1.34). In Section §2.1, the solution of condition adjustment has been derived using Lagrange method. Below, we will derive the minimum norm solution based on the general solution of consistent linear equation system and with the help of generalized matrix inverses.

From Theorem 4.5, the general solution of (4.1.31) can be written as:

$$\hat{X} = A^- L + (I - A^- A)\, V\,,$$

where A^- can be any arbitrary g–inverse of A. As the minimum norm inverse A_{QO}^- is a g–inverse of A, we replace A^- in the general solution above by A_{QO}^- and obtain:

$$\hat{X} = A_{QO}^- L + (I - A_{QO}^- A)\, V$$

$$\hat{X}^T Q^{-1} \hat{X} = [A_{QO}^- L + (I - A_{QO}^- A)\, V]^T Q^{-1} [A_{QO}^- L + (I - A_{QO}^- A)\, V]$$

$$= (A_{QO}^- L)^T Q^{-1} (A_{QO}^- L) + V^T (I - A_{QO}^- A)^T Q^{-1} (I - A_{QO}^- A)\, V + 2\, U$$

where:

$$U \equiv (A_{QO}^- L)^T Q^{-1} (I - A_{QO}^- A)V = L^T [(AQA^T)^-]^T AQ\, Q^{-1} (I - A_{QO}^- A)V$$

$$= L^T[(AQA^T)^-]^T[A - A\, A_{QO}^-\, A]\, V = 0 \qquad (\text{note: } r(QA^T) = r(AQA^T))$$

which leads to:

$$\tilde{X}^T Q^{-1} \tilde{X} = (A_{QO}^- L)^T Q^{-1} (A_{QO}^- L) + V^T (I - A_{QO}^- A)^T Q^{-1} (I - A_{QO}^- A) V$$

Since Q^{-1} is positive definite, the two quadratic forms of L and V above are non–negative, i.e.

$$\tilde{X}^T Q^{-1} \tilde{X} \geq (A_{QO}^- L)^T Q^{-1} (A_{QO}^- L)$$

where the identity holds if and only if $V = 0$. Hence, the minimum norm solution $\hat{X}$ is obtained:

$$\hat{X} = \tilde{X}\Big|_{V=0} = A_{QO}^- L = QA^T (AQA^T)^- L \qquad (4.1.35)$$

As pointed out in Section 3.1, the minimum norm inverse A_{QO}^- is not unique. However, the minimum norm solution given by (4.1.35) is unique. This can be proved by the fact that for any arbitrary minimum norm inverse $\tilde{A}_{QO}^-$ as defined in (4.1.8), $\tilde{A}_{QO}^- L$ is equal to $A_{QO}^- L$ where A_{QO}^- is any particular minimum norm inverse:

$$\hat{X} = \tilde{A}_{QO}^- L = [A_{QO}^- + N(I - AA_{QO}^-)]L = A_{QO}^- L + N(I - AA_{QO}^-)AX = A_{QO}^- L \qquad (4.1.36)$$

In traditional condition adjustment, we normally use only independent conditions which imply a coefficient matrix A with full rank $r(\underset{n \times m}{A}) = n$. In this case the coefficient matrix of the normal equation, $AQ^{-1}A^T$, is non–singular and thus has an ordinary inverse $(AQA^T)^{-1}$. The minimum norm solution now becomes:

$$\hat{X} = QA^T (AQA^T)^{-1} L = A_{QO}^{-1} L, \qquad \text{for } r(\underset{n \times m}{A}) = n \qquad (4.1.37)$$

$$A_{QO}^{-1} = QA^T (AQA^T)^{-1}, \qquad \text{for } r(\underset{n \times m}{A}) = n \qquad (4.1.38)$$

However, if we have used linearly dependent condition equations, we will then get a singular normal equation. The minimum norm solution (4.1.35) is still valid in this case and what we need to do is to use a generalized inverse of AQA^T. No matter which generalized inverse of AQA^T we choose, the minimum norm solution $\hat{X}$ remains the same according to (4.1.36).

In summary, one can say that condition adjustment by least squares principle is equivalent to find the minimum norm solution of a consistent linear equation system. This minimum norm solution is constructed using the generalized inverse A_{QO}^- of the coefficient matrix A, which explains why A_{QO}^- is called the minimum norm generalized inverse. For matrix A with full rank, the minimum norm solution and the solution of

condition adjustment (Cf. 2.1.10) are identical. However, for singular A matrix, the traditional condition adjustment is no longer valid and generalized matrix inverses should be used. In other words, the minimum norm inverse A_{QO}^{-} has provided us with a mathematical tool to do condition adjustment even if the coefficient matrix A does not have full rank (e.g. when dependent condition equations are used).

Example 4.1.8

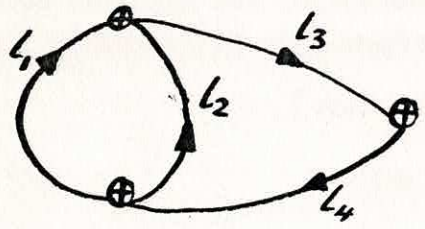


Fig. 4.1

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.

For this network, we have n=4 observations, m=2 necessary observations (one of the 3 benchmarks must be assumed to be fixed) and n–m=2 independent condition equations. Based on the two non–overlapping closed loops, we can obtain the following condition equations:

$$\underset{2\times 4}{B}\ \underset{4\times 1}{\epsilon} = \underset{2\times 1}{W} = \underset{2\times 4}{B}\ \underset{4\times 1}{L}$$

where: $B = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$; $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$; $L = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{bmatrix}$; $W = BL = \begin{bmatrix} l_1 - l_2 \\ l_2 + l_3 + l_4 \end{bmatrix}$.

As the weight matrix of L can be regarded as a unit matrix and BB^T is non–singular, the least squares estimate of ϵ follows directly from (2.1.10) or equivalently (4.1.37):

$$\hat{\epsilon} = B^T(BB^T)^{-1}W = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}^{-1} W = \frac{1}{5} \begin{bmatrix} 3 & 1 \\ -2 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} W$$

$$= \frac{1}{5} \begin{bmatrix} 3 & 1 \\ -2 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} l_1 - l_2 \\ l_2 + l_3 + l_4 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3l_1 - 2l_2 + l_3 + l_4 \\ -2l_1 + 3l_2 + l_3 + l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \end{bmatrix} \qquad \text{(a)}$$

If we use all three closed loops in Fig. 4.1, then we will have the following linearly correlated condition equations:

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} = \begin{bmatrix} l_1 - l_2 \\ l_2 + l_3 + l_4 \\ l_1 + l_3 + l_4 \end{bmatrix} \quad \Leftrightarrow \quad \underset{3\times 4}{B}\ \underset{4\times 1}{\epsilon} = \underset{3\times 1}{W}$$

As $r(B) = 2$, BB^T is singular and thus does not have an ordinary inverse. The minimum norm solution of ϵ is given by (4.1.35):

$$\hat{\epsilon} = B_{IO}^{-} W = B^T(BB^T)^{-} W$$

where: $$BB^T = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix} \text{ with } |BB^T| = 0.$$

The g–inverse $|BB^T|^{-}$ is not unique and consequently the minimum norm inverse B_{IO}^{-} is not unique either. Below, we will choose three different inverses for $|BB^T|^{-}$ to see whether they may affect the minimum norm solution $\hat{\epsilon}$.

The first choice of $|BB^T|^{-}$ is obtained from (4.1.4b):

$$|BB^T|^{-} = \begin{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}^{-1} & 0 \\ 0 \quad 0 & 0 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_{IO}^{-} = B^T(BB^T)^{-} = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3 & 1 & 0 \\ -2 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$

$$\hat{\epsilon} = B_{IO}^{-} W = \frac{1}{5} \begin{bmatrix} 3 & 1 & 0 \\ -2 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} l_1 - l_2 \\ l_2 + l_3 + l_4 \\ l_1 + l_3 + l_4 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3l_1 - 2l_2 + l_3 + l_4 \\ -2l_1 + 3l_2 + l_3 + l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \end{bmatrix} \qquad \text{(b)}$$

The second choice of $|BB^T|^{-}$ and the resulted estimate $\hat{\epsilon}$ are as follows:

$$|BB^T|^{-} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^{-1} \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & -2 \\ 0 & -2 & 3 \end{bmatrix}$$

$$B_{IO}^{-} = B^T(BB^T)^{-} = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & -2 \\ 0 & -2 & 3 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 0 & -2 & 3 \\ 0 & 3 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\hat{\epsilon} = B_{IO}^{-} W = \frac{1}{5} \begin{bmatrix} 0 & -2 & 3 \\ 0 & 3 & -2 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} l_1 - l_2 \\ l_2 + l_3 + l_4 \\ l_1 + l_3 + l_4 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3l_1 - 2l_2 + l_3 + l_4 \\ -2l_1 + 3l_2 + l_3 + l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \end{bmatrix} \qquad \text{(c)}$$

The third choice of $(BB^T)^{-}$ is obtained after replacing A^{-} in (4.1.6) by the first choice of $(BB^T)^{-}$ above, and letting N in (4.1.6) be 1/5 I (unit matrix) and finally letting M in (4.1.6) be zero:

$$(BB^T)^- = \frac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{1}{5}I\left[I - \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}\frac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right] = \frac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ -1 & -1 & 1 \end{bmatrix}$$

$$B_{IO}^- = B^T(BB^T)^- = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}\frac{1}{5}\begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ -1 & -1 & 1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 2 & 0 & 1 \\ -2 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\hat{\epsilon} = B_{IO}^- W = \frac{1}{5}\begin{bmatrix} 2 & 0 & 1 \\ -2 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} l_1 - l_2 \\ l_2 + l_3 + l_4 \\ l_1 + l_3 + l_4 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 3l_1 - 2l_2 + l_3 + l_4 \\ -2l_1 + 3l_2 + l_3 + l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \end{bmatrix} \qquad \text{(d)}$$

Comparing the four solutions in Eqs. (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. The latter case is accomplished with the help of generalized inverses. The last three solutions also shows that the minimum norm solution $\hat{\epsilon}$ is independent of the choice of generalized inverse $(BB^T)^-$, in consistency with the general conclusion expressed in (4.1.36). Further more, one can notice that the first choice of $(BB^T)^-$ corresponds to condition adjustment using only the first two linearly independent conditions, while the second choice of $(BB^T)^-$ implies the use of the last two linearly independent conditions.

4.1.6 Least Squares Solutions

Assume that we have an inconsistent linear equation system:

$$\underset{n\,x\,m}{A}\ \underset{m\,x\,1}{X} \neq \underset{n\,x\,1}{L}, \quad n>m\geq r(A)=r \qquad (4.1.39)$$

One can always make (4.1.39) consistent by adding an additional unknown vector ϵ to the right–hand side:

$$\underset{n\,x\,m}{A}\ \underset{m\,x\,1}{X} = \underset{n\,x\,1}{L} - \underset{n\,x\,1}{\epsilon} \qquad (4.1.40)$$

(4.1.40) corresponds to the functional model of adjustment by elements, where X, L, ϵ and A denote the unknown parameter, the observation vector, the residual vector and the coefficient matrix (design matrix), respectively. In adjustment by elements, one also assumes the following statistical model for the residual vector ϵ:

$$E(\epsilon) = 0\,; \quad E(\epsilon\epsilon^T) = \sigma_0^2\, P^{-1} \qquad (4.1.41)$$

where σ_0^2 stands for the variance factor and P the weight matrix of L.

Adjustment by elements tries to find a solution X to the inconsistent equation system (4.1.39) such that the sum of weighted residuals squared, $\epsilon^T P \epsilon$, is minimized. In linear algebra, such a solution is called the *least squares solution.* In Chapter 3, the least

squares solution has been derived for design matrix A of full rank, i.e. r(A)=m. Below, we are going to discuss the least squares solution for design matrix A of any rank. First, let us express all possible solutions to (4.1.40) through substitution of the arbitrary inverse A^- in Theorem 4.5 by the least squares inverse A_{OP}^- :

$$\tilde{X} = A_{OP}^- (L - \epsilon) + (I - A_{OP}^- A) V$$

As $AA_{OP}^-A = A$, we then have:

$$A \tilde{X} = A A_{OP}^- (L-\epsilon) = L-\epsilon, \quad \text{or:} \quad (I-AA_{OP}^-) \epsilon = (I-AA_{OP}^-) L$$

It is trivial to show that the unit matrix $\underset{nxn}{I}$ is a generalized inverse of $(I-AA_{OP}^-)$. This leads to the general solution of ϵ :

$$\tilde{\epsilon} = I (I-AA_{OP}^-)L + [I - I (I-AA_{OP}^-)] V = (I-AA_{OP}^-) L + AA_{OP}^- V$$

$$\begin{aligned}\tilde{\epsilon}^T P \tilde{\epsilon} &= [(I-AA_{OP}^-) L + AA_{OP}^- V]^T P [(I-AA_{OP}^-) L + AA_{OP}^- V] \\ &= [(I-AA_{OP}^-) L]^T P [(I-AA_{OP}^-) L] + V^T (AA_{OP}^-)^T P (AA_{OP}^-) V \\ &\geq [(I-AA_{OP}^-) L]^T P [(I-AA_{OP}^-) L]\end{aligned}$$

where the identity holds iff V=0. Then we get the least squares solution for ϵ:

$$\hat{\epsilon} = \left\{(I-AA_{OP}^-) L + AA_{OP}^- V\right\} \Big|_{V=0} = (I-AA_{OP}^-) L \qquad (4.1.42)$$

and the least squares solution for X:

$$\begin{aligned}\hat{X} &= \left\{A_{OP}^-(L-\hat{\epsilon}) + (I-A_{OP}^- A)V\right\}\Big|_{V=0} = A_{OP}^- [L - (I-AA_{OP}^-)L] \\ &= A_{OP}^- A A_{OP}^- L \\ &= (A^TPA)^- A^TP A (A^TPA)^- A^TPL = (A^TPA)^- A^TP(A^TP)^- A^TP L \\ &= (A^TPA)^- A^TP L = A_{OP}^- L \qquad (4.1.43)\end{aligned}$$

Due to the non–uniqueness of A_{OP}^-, the above solution is not unique. From (4.1.11), we get the complete set of the least squares solutions:

$$\tilde{X} = \tilde{A}_{OP}^- L = A_{OP}^- L + (I-A_{OP}^- A) V_x, \qquad (4.1.44)$$

where $\tilde{A}_{OP}^-$ denotes the general least squares inverse of A, A_{OP}^- denotes a particular least squares inverse and V_x denotes an arbitrary vector.

However, in spite of the non–uniqueness of A_{OP}^{-}, it can be shown that the residual vector $\hat{\epsilon}$ given in (4.1.42) is unique:

$$\begin{aligned}\hat{\epsilon} &= (I - A\,\tilde{A}_{OP}^{-})\,L = L - A\,[A_{OP}^{-} + (I - A_{OP}^{-}A)V_{\epsilon}]\,L \\ &= L - A\,A_{OP}^{-}L - [A - A\,A_{OP}^{-}\,A]\,V_{\epsilon}\,L = (I - AA_{OP}^{-})\,L \end{aligned} \qquad (4.1.45)$$

It should be noted that the least squares solution presented above applies to both singular and non–singular design matrix A. In traditional adjustment by elements where the design matrix A has full rank r(A) = m, the least squares solutions (4.1.43) and (4.1.42) reduce to (3.1.13) and (3.1.15), respectively:

$$\hat{X} = (A^{\top}PA)^{-1}A^{\top}PL \qquad (|A^{\top}PA| \neq 0) \qquad (4.1.46)$$

$$\hat{\epsilon} = (I - A\,A_{OP}^{-})L = L - A\hat{X} \qquad (4.1.47)$$

In summary, the least squares solutions (4.1.43) and (4.1.42), constructed using the least squares generalized inverses A_{OP}^{-}, are applicable for both singular and non–singular design matrix A. For design matrix A of full rank [i.e. r(A)=m], the above solutions are identical to the solutions of the traditional adjustment by elements. In the case of singular A, the least squares solution for X is non–unique while the solution for ϵ is still unique. An intuitive explanation to this is as follows.

In traditional geodetic adjustment, the rank defect of A [i.e. r(A)<m] is often caused by the lack of reference datum, e.g. without any fixed point, or without any fixed orientation, or for triangulation network without any fixed length to define the scale of the network. Therefore, the unknown parameters, which often are the coordinates of network points, cannot be uniquely determined. One can find many different solutions to X, which all satisfy the internal geometry defined by the measurements (e.g. angles, distances, height differences), corresponding to the different choices of the generalized inverse A_{OP}^{-}. On the other hand, 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. Therefore, the least squares solutions $\hat{\epsilon}$ can be unique even if the absolute coordinates $\hat{X}$ are not unique.

Among all possible solutions of $\hat{X}$, 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 4.1.9 EJEMPLO!

We consider the adjustment of the levelling network shown in Fig. 4.1 using the method adjustment by elements.

With no fixed point to define the height reference datum, the classical approach to network adjustment is to assume one benchmark, say P_1, is fixed with given height (say

$x_1 = 0$). Choosing the heights (x_2, x_3) of P_1 and P_2 as unknown parameters, our observation equations become:

$$\begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{bmatrix} - \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \quad \Leftrightarrow \quad \underset{4\times1}{L} - \underset{4\times1}{\epsilon} = \underset{4\times2}{A}\ \underset{2\times1}{X}$$

As in Example 4.1.8, the weight matrix of observations L is a unit matrix. As $r(A) = 2$, a unique least squares solution $\hat{X}$ can be obtained from (3.1.13) or (4.1.46):

$$A_{OI}^{-1} = (A^T A)^{-1} A^T = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 2 & 2 & -1 & -1 \\ 1 & 1 & 2 & -3 \end{bmatrix}$$

$$\hat{X} = A_{OI}^{-1} L = \frac{1}{5} \begin{bmatrix} 2l_1 + 2l_2 - l_3 - l_4 \\ l_1 + l_2 + 2l_3 - 3l_4 \end{bmatrix} \qquad \text{(e)}$$

$$I - A A_{OI}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 2 & 2 & -1 & -1 \\ 1 & 1 & 2 & -3 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3 & -2 & 1 & 1 \\ -2 & 3 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}$$

$$\hat{\epsilon} = L - A\hat{X} = (I - A A_{OI}^{-1}) L = \frac{1}{5} \begin{bmatrix} 3l_1 - 2l_2 + l_3 + l_4 \\ -2l_1 + 3l_2 + l_3 + l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \\ l_1 + l_2 + l_3 + l_4 \end{bmatrix} \qquad \text{(f)}$$

The variance–covariance matrix of $\hat{X}$ is given by (3.1.14):

$$C_{\hat{X}\hat{X}} = \sigma_0^2 (A^T A)^{-1} = \sigma_0^2 \frac{1}{5} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

As P_1 is assumed to be fixed, its height x_1 should have zero variance and zero covariances with x_2 and x_3. Let Y denote vector $(x_1, \hat{x}_2, \hat{x}_3)^T$. Then we can write the variance–covariance matrix of Y as:

$$C_{YY} = \begin{bmatrix} 0 & 0 \\ 0 & C_{\hat{X}\hat{X}} \end{bmatrix} = \sigma_0^2 \frac{1}{5} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix} \qquad \text{(g)}$$

and $\operatorname{tr}(C_{YY}) = \sigma_0^2$ (h)

If we choose the heights of all three benchmarks as unknown parameters, the observations become:

$$\begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{bmatrix} - \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

or in matrix form:
$$\underset{4\times1}{L} - \underset{4\times1}{\epsilon} = \underset{5\times3}{A}\ \underset{3\times1}{X} \qquad \text{(i)}$$

The least squares solution of X follows from (4.1.43):

$$A^TA = \begin{bmatrix} -1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & -2 & -1 \\ -2 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad |A^TA| = 0$$

$$A_{OI}^- = (A^TA)^-A^T = \left[\begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix}^{-1} \ 0\right] \begin{bmatrix} -1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -1 & -1 & -2 & 3 \\ 1 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\hat{X} = A_{OI}^- L = \frac{1}{5}\begin{bmatrix} -l_1-l_2-2l_3+3l_4 \\ l_1+l_2-3l_3+2l_4 \\ 0 \end{bmatrix} \qquad \text{(j)}$$

Since the g–inverse $(A^TA)^-$ as well as A_{OI}^- are not unique, $\hat{X}$ in (j) is only one particular least squares solution. The complete set of least squares solutions can be obtained from (4.1.44):

$$I - A_{OI}^- A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} -1 & -1 & -2 & 3 \\ 1 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{X} = A_{OI}^- L + (I - A_{OI}^- A)\, V_x = \hat{X} + (I - A_{OI}^- A)\, V_x = \hat{X} + \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

$$= \hat{X} + \begin{bmatrix} w \\ w \\ w \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -l_1-l_2-2l_3+3l_4 \\ l_1+l_2-3l_3+2l_4 \\ 0 \end{bmatrix} + \begin{bmatrix} w \\ w \\ w \end{bmatrix} \qquad \text{(k)}$$

where $V_x = (u, v, w)^T$ denotes an arbitrary vector. With different choices of arbitrary constant w, we get accordingly different least squares solutions. Below are two such solutions:

For $w = 0$, we have:
$$X = \hat{X} = \frac{1}{5}\begin{bmatrix} -l_1-l_2-2l_3+3l_4 \\ l_1+l_2-3l_3+2l_4 \\ 0 \end{bmatrix},$$

which corresponds to traditional adjustment by elements with benchmark P_3 assumed to be fixed and have height zero.

For $w = (l_1+l_2+2l_3-3l_4)/5$, we have:
$$X = \frac{1}{5}\begin{bmatrix} 2l_1 + 2l_2 - l_3 - l_4 \\ l_1 + l_2 + 2l_3 - 3l_4 \\ 0 \end{bmatrix},$$

which is equivalent to traditional adjustment by elements with P_1 assumed to be fixed and have height zero [Cf Eq. (e)].

The estimated residual $\hat{\epsilon}$ can be obtained from $\tilde{X}$ in (k):

$$\hat{\epsilon} = L - A\tilde{X} = L - A[\hat{X} + (I - A_{OP}^{-}A)V_x] = L - A\hat{X} - (A - A)V_x$$

$$= L - A\hat{X} = (I - AA_{OI}^{-})L = \frac{1}{5}\begin{bmatrix} 3 & -2 & 1 & 1 \\ -2 & 3 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix} L$$

$$= \frac{1}{5}\begin{bmatrix} 3l_1 - 2l_2 + l_3 + l_4 \\ -2l_1 + 3l_2 + l_3 + l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \\ l_1 + l_2 + l_3 + l_4 \end{bmatrix} \qquad (1)$$

The above results have confirmed that when the design matrix A is singular, the least squares solution is not unique while the residuals $\hat{\epsilon}$ can be still uniquely determined. In traditional geodetic networks (like the levelling network illustrated in Fig. 4.1), the singularity of A or A^TA is mostly caused by the lack of a coordinate reference datum. The consequence becomes that coordinates of network points cannot be uniquely determined. In this situation, 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 three different height datum for the whole network. In Example 4.1.10 at the end of sub–section §4.1.7, another generalized inverse will be chosen, resulting in a third height datum. This may remind us of an important fact that any mathematical treatment always has certain physical meaning or geodetic implication. It is very dangerous in geodetic practice if one just blindly applies mathematic theory without clearly understanding its geodetic implications.

4.1.7 Minimum Norm–Least Squares Solution

Assume that from the adjustment model (4.1.39) and (4.1.40), non–unique solutions (4.1.44) to X have been obtained by the least squares principle $\epsilon^T P\epsilon$ = minimum. A unique solution among all possible solutions (4.1.44) can be defined by assuming an extra constraint on X:

$$X^T \underset{m\,x\,m}{Q^{-1}} X = \text{minimum} \qquad (4.1.48)$$

where Q is a known symmetrical positive definite matrix. As constraints (4.1.48) implies the minimization of the Euclidean norm of vector X, the unique solution from (4.1.44) which satisfies constraint (4.1.48) is called a *minimum norm–least squares solution.*

Denoting the general least squares solution in Eq. (4.1.44) by $\tilde{X}$ and multiplying both sides of (4.1.44) by A, we obtain a new consistent equation system:

$$A\tilde{X} = A A_{OP}^{-} L \tag{4.1.49}$$

where A_{OP}^{-} denotes any particular least squares inverse of A.

The minimum norm–least squares solution can be obtained from the minimum norm solution of (4.1.49) (Cf 4.1.35):

$$\hat{X} = A_{QO}^{-} A A_{OP}^{-} L = Q A^T (A Q A^T)^{-} A (A^T P A)^{-} A^T P\, L = A_{QP}^{-1} L \tag{4.1.50}$$

As described in Section 3.2, the generalized inverse A_{QP}^{-1} is unique and therefore the minimum norm–least squares solution given by Eq. (4.1.50) is also unique. The solution for the residual vector ϵ follows directly:

$$\hat{\epsilon} = L - A\hat{X} = (I - A A_{QP}^{-1}) L \tag{4.1.51}$$

If $Q \equiv I$, which may be interpreted as if all components of X have equal weights, the minimum norm–least squares solution becomes:

$$\hat{X} = A_{IP}^{-1} L \tag{4.1.52}$$

$$\hat{\epsilon} = L - A\hat{X} = (I - A A_{IP}^{-1}) L \tag{4.1.53}$$

Using the orthogonal bordering method (Cf §4.1.3), A_{IP}^{-1} can be calculated as:

$$A_{IP}^{-1} = (A^T P A + D^T D)^{-1} A^T P \tag{4.1.54}$$

where D is defined such that:

$$A D^T = 0 \quad \text{and} \quad |D D^T| \neq 0 \tag{4.1.55}$$

Let $U \equiv A^T P A + D^T D$ and we have:

$$DU = (AD^T)^T PA + DD^T D = DD^T D \implies (DD^T)^{-1} DU = D$$

or: $$DU^{-1} = (DD^T)^{-1} D \tag{4.1.56}$$

Using (4.1.52) and (4.1.55) and (4.1.56), we can derive a very interesting result for the minimum norm–least squares solution $\hat{X}$:

$$\begin{aligned} D\hat{X} &= D A_{IP}^{-1} L = D(A^T P A + D^T D)^{-1} A^T P\, L = DU^{-1} A^T P L \\ &= (DD^T)^{-1} D A^T P L = (DD^T)^{-1} (AD^T)^T P L = (DD^T)^{-1} 0^T P L \\ &= 0 \end{aligned} \tag{4.1.57}$$

Theoretically, matrix D should have dimension dxm, where d is the rank defect of A

or rank defect of the geodetic network concerned and m is the number of parameters. (4.1.55) indicates that D is defined by the design matrix A, i.e. the configuration of the network. This gives us a possibility to find the elements of D for different types of network:

For levelling network:

$$\underset{1\times m}{D} = [1,\ 1,\ \cdots,\ 1] \tag{4.1.58}$$

where m denotes the number of unknown point in the levelling network.

For trilateration network with distance measurements

$$\underset{3\times m}{D} = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\ -y_1^0 & x_1^0 & -y_2^0 & x_2^0 & \cdots & -y_k^0 & x_k^0 \end{bmatrix} \tag{4.1.59}$$

where m = number of unknown parameters, k = m/2 = number of unknown points and $(x_i^0,\ y_i^0)$ denotes the approximate coordinates of the i–th unknown point (i=1, 2, $\cdots$, k).

For triangulation network with angle measurements

$$\underset{4\times m}{D} = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\ -y_1^0 & x_1^0 & -y_2^0 & x_2^0 & \cdots & -y_k^0 & x_k^0 \\ x^0 & y^0 & x^0 & y^0 & \cdots & x^0 & y^0 \end{bmatrix} \tag{4.1.60}$$

where m = number of unknown parameters, k = m/2 = number of unknown points and $(x_i^0,\ y_i^0)$ denotes the approximate coordinates of the i–th unknown point (i=1, 2, $\cdots$, k).

Geodetic networks with rank defect are often called *free networks*. This is due to the fact that rank defect is almost always caused by the lack of or insufficient geometrical reference datum so that the networks may freely move, or/and rotate, or/and change in scale. Least squares estimates in free networks are not unique, as shown in (4.1.44). By imposing minimization condition on both $\epsilon^T P \epsilon$ and $X^T Q^{-1} X$, a unique solution (4.1.50) can be obtained mathematically, using the minimum–norm least squares inverse. However, as (4.1.57) indicates, this implies that special conditions or requirements are imposed on the parameters such that a reference datum is implicitly defined.

More concretely, for levelling networks where D is given by (4.1.58), the condition (4.1.57) becomes:

$$\hat{x}_1 + \hat{x}_2 + \cdots + \hat{x}_m = 0,$$

which implies that the level which corresponds to the average height of all benchmarks

after adjustment is defined as the zero height level (reference datum). If approximate heights of benchmarks have been used introduced in the observation equations, X in (4.1.52) is actually estimate of the correction to the approximate height and the zero height is defined such that the average height of all benchmarks after the adjustment remains the same as the average value of all approximate heights.

In the case of triangulation networks with angle measurements and with D given by (4.1.60), approximate coordinates are always introduced for linearization purpose and thus the condition (4.1.57) now implies that *before and after the adjustment* :

a) the network center (the geometric center O of all triangulation points), remains unchanged, so that the absolute position of the network is uniquely defined;

b) the weighted means of azimuths from the network center O to all triangulation points remains unchanged, where the weight is the squared distance from O to each triangulation point, so that the orientation of the network is uniquely defined;

c) the sum of squared distances from O to all triangulation points remains unchanged, so that the scale of the network is uniquely defined.

Example 4.1.10

We consider once again the levelling network illustrated in Fig. 4.1 with observation equations (i) given in Example 4.1.9. Now we are going to calculate the minimum norm–least squares solution $\hat{X}$ with both P and Q being unit matrices.

First we calculate the generalized inverse A_{II}^{-1} using Eqs. (4.1.54) and (4.1.58):

$$D = [\,1 \quad 1 \quad 1\,]$$

$$A^TA + D^TD = \begin{bmatrix} 3 & -2 & -1 \\ -2 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$|A^TA + D^TD| = 45 \neq 0\,; \qquad (A^TA + D^TD)^{-1} = \frac{1}{15}\begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

$$A_{II}^{-1} = (A^TA + D^TD)^{-1}A^T = \frac{1}{15}\begin{bmatrix} -3 & -3 & -1 & 4 \\ 3 & 3 & -4 & 1 \\ 0 & 0 & 5 & -5 \end{bmatrix}$$

$$\hat{X} = A_{II}^{-1}L = \frac{1}{15}\begin{bmatrix} -3l_1 - 3l_2 - l_3 + 4l_4 \\ 3l_1 + 3l_2 - 4l_3 + l_4 \\ 5l_3 - 5l_4 \end{bmatrix} \qquad \text{(m)}$$

$$D^T\hat{X} = \hat{x}_1 + \hat{x}_2 + \hat{x}_3 = 0 \qquad \text{(n)}$$

$$I - A\,A_{II}^{-1} = \frac{1}{5}\begin{bmatrix} 3 & -2 & 1 & 1 \\ -2 & 3 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}$$

$$\hat{\epsilon} = L - A\,\hat{X} = (I - A\,A_{II}^{-1})\,L = \frac{1}{5}\begin{bmatrix} 3l_1 - 2l_2 + l_3 + l_4 \\ -2l_1 + 3l_2 + l_3 + l_4 \\ l_1 + l_2 + 2l_3 + 2l_4 \\ l_1 + l_2 + l_3 + l_4 \end{bmatrix} \qquad (o)$$

$$C_{\hat{X}\hat{X}} = \sigma_0^2\,A_{II}^{-1}\,(A_{II}^{-1})^T = \sigma_0^2\,\frac{1}{45}\begin{bmatrix} 7 & -2 & -5 \\ -2 & 7 & -5 \\ -5 & -5 & 10 \end{bmatrix} \qquad (p)$$

$$\mathrm{tr}\{C_{\hat{X}\hat{X}}\} = \frac{8}{15}\,\sigma_0^2 \qquad (q)$$

Some Remarks

* Comparing results in Examples 4.8, Examples 4.9 and Examples 4.10, we see that the estimated residuals are the same for all methods, as it should be.

* Let us define the weight center of the levelling network as a point with height equal to the mean height of all three benchmarks of the network. Eq. (n) above implies that the weight center has zero height and thus defines the height datum for the whole network.

* A comparison of Eq. (q) above with Eq. (g) in Example 4.1.9 show that the variance–covariance matrix of the benchmark heights 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 ($8/15\ \sigma_0^2$) in the former is smaller than the total variance (σ_0^2) in the latter case.

In Chapter 2 and Chapter 3, we have discussed various methods of least sqaures adjustment. In order to carry out the adjustment, one needs both the functional model (condition equations or observation equations) and the a priori statistical model, namely the expectation and variance–covariance matrix of the measurement errors. As commented at the beginning of §2.1.2, least squares adjustment is dependent more on the relative accuracy of the observations expressed by the weight matrix P, than on the absolute accuracy expressed by the variance–covariance matrix. In other words, we are able to do least squares adjustment with a given weight matrix, even though the variance–covariance matrix of the observations is unknown.

Assume that we have estimated the residuals $\hat{\epsilon}$ from condition adjustment [Cf Eq. (2.1.10)]:

$$\hat{\epsilon} = P^{-1}B^{T}(BP^{-1}B^{T})^{-1}W \qquad (4.2.1)$$

where B and W denote the coefficient matrix and the constant vector of condition equations (2.1.2). Suppose that we have now another weight matrix P_1 which differs from the above weight matrix only by a factor κ ($\kappa \neq 0$):

$$P = \kappa P_1 \qquad (4.2.2)$$

Inserting (4.2.2) into (4.2.1) gives:

$$\hat{\epsilon} = P^{-1}B^{T}(BP^{-1}B^{T})^{-1}W = (\kappa P_1)^{-1}B^{T}[B(\kappa P_1)^{-1}B^{T}]^{-1}W$$
$$= 1/\kappa\, P_1^{-1}B^{T}\,\kappa\,(BP_1^{-1}B^{T})^{-1}W = P_1^{-1}B^{T}(BP_1^{-1}B^{T})^{-1}W \qquad (4.2.3)$$

(4.2.3) shows that the estimated residuals are identical using two different weight matrices which differ only by a constant factor.

In geodetic practice, one is often able to obtain the weight matrix of the observations (i.e. a relative measure of accuracy) without directly involving the absolute variances or covariances. For instance, in angle and direction measurements one used to assign weight to each observation according to the number of rounds by which an angle or direction is measured with a theodolite. In the case of levelling networks, one may define the weight of each levelled height difference as inversely proportional to the length of the levelling line concerned or to the approximate height difference. Though one may question these empirical weighting models, they generally work pretty well in practice.

Problems or difficulties occur when there are correlations among the observations or when there are different types of observations, e.g. angle measurements and distance measurements. In the former case, it is difficult to describe correlation without knowing the absolute values of the corresponding covariances. In the latter case, we can assign separately one weight matrix to all angles and another weight matrix to all distances. However, to form a unified weight matrix for all observations we need to know the absolute variances of angles and distances. Or more precisely speaking, we need to know the relative accuracy between angles and distances, which is very difficult or impossible to get from empirical weighting procedures.

One way to overcome the above difficulties is to estimate the variance–covariance matrix of the observations based on the obtained observation data and the given functional model (condition equations or observation equations). As it is impossible to estimate all elements in the variance–covariance matrix of the observations, we decompose the total, unknown variance–covariance matrix into a linear combination of several given matrices with some unknown constants, called the *variance–covariance components*. After the decomposition, we then try to estimate these unknown variance–covariance components based on the results of a preliminary adjustment using either the method of condition adjustment or adjustment by elements. Since the estimation is carried out after field measurements (i.e. the stochastic events happened already), the estimated values of the variance–covariance components are called *a posteriori estimates*.

The estimation of variance–covariance components is not only a geodetic problem, but also an important subject for applied mathematics and particularly statistical inferences. In this section, we first describe the decomposition of variance–covariance matrix with examples from geodesy and surveying. Then we present the classical estimate by the famous German geodesist Helmert, both for condition adjustment model and for Gauss–Markov model (i.e. adjustment by elements). Finally we will describe the **B**est **Q**uadratic **U**nbiased **E**stimate (BQUE) and its equivalent, namely the **MI**nimum **N**orm **Q**uadratic **U**nbiased **E**stimate (MINQUE).

There are numerous literatures, both geodetic and pure mathematical, on variance–covariance component estimation. Interested readers may consult e.g. Helmert (1924), Rao and Mitra (1971), Rao (1973), Grafarend et al. (1980), Sjöberg (1983), Fan and Sjöberg (1986), Egeltoft (1992).

§4.2.1 Decomposition of the Variance–Covariance Matrix

Assume that measurement error (true value of residual) ϵ_i has zero expectation and variance q_{ii} $(1 \le i \le n)$. Assume also that errors ϵ_i and ϵ_j have covariance q_{ij} $(1 \le i,j \le n)$. Then the variance–covariance matrix Q of error vector $\epsilon \equiv (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^T$ can be written as:

$$E[\underset{n\times 1}{\epsilon}\ \underset{1\times n}{\epsilon^T}] \equiv \underset{n\times n}{Q} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{12} & q_{22} & \cdots & q_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} \qquad (4.2.4a)$$

Note that in this section, Q is used to denote the variance–covariance matrix of the observations, not the cofactor matrix as defined in previous chapters. Now let us partition the above nxn symmetrical matrix Q into $q=k(k+1)/2$ sub–matrices where k is the number of sub–matrices on each row or column and q is the total number of sub–matrices. Each sub–matrix is expressed as the product of an unknown number σ_{ij} and a given matrix Q_{ij} $(1 \le i,j \le k)$:

$$\underset{n\,x\,n}{Q} = \begin{bmatrix} \sigma_{11}\,Q_{11} & \sigma_{12}\,Q_{12} & \cdots & \sigma_{1k}\,Q_{1k} \\ \sigma_{21}\,Q_{21} & \sigma_{22}\,Q_{22} & \cdots & \sigma_{2k}\,Q_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{k1}\,Q_{k1} & \sigma_{k2}\,Q_{k2} & \cdots & \sigma_{kk}\,Q_{kk} \end{bmatrix} \qquad (4.2.4b)$$

where $\sigma_{ij} = \sigma_{ji}$ is assumed. The partitioned matrix Q can be further rewritten as a linear combination of q given matrices (Q_j):

$$\underset{n\,x\,n}{Q} = \sum_{j=1}^{q} \sigma_j \underset{n\,x\,n}{Q_j} \qquad (4.2.4c)$$

where σ_j's are the so called variance–covariance components consisting of those $q=k(k+1)/2$ two–indexed constants in (4.2.4b):

$$\underset{q\,x\,1}{\sigma} \equiv \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \cdots \\ \sigma_k \\ \sigma_{k+1} \\ \cdots \\ \sigma_q \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \cdots \\ \sigma_{1k} \\ \sigma_{22} \\ \cdots \\ \sigma_{kk} \end{bmatrix} \qquad (4.2.4d)$$

and Q_j is an nxn square matrix associated with σ_j. As example, we give Q_j for j = 1, 2, k, k+1, q :

$$Q_1 = \begin{bmatrix} Q_{11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}; \quad Q_2 = \begin{bmatrix} 0 & Q_{12} & \cdots & 0 \\ Q_{21} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}; \quad Q_k = \begin{bmatrix} 0 & 0 & \cdots & Q_{1k} \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ Q_{k1} & 0 & \cdots & 0 \end{bmatrix}$$

$$Q_{k+1} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & Q_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}; \quad \cdots\cdots; \quad Q_q = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & Q_{kk} \end{bmatrix}$$

(4.2.4c) is the general form of the decomposed variance–covariance matrix with q variance–covariance components. Our task is to estimate these q components from the measurement data. As (4.2.4c) may appear too abstract, we present two geodetic examples to show how the variance–covariance matrix can be decomposed with some variance–covariance components.

Example 4.2.1

We consider the case when the observations can be divided into two groups (e.g. angle measurements and distance measurements) which are independent of each other. By empirical weighting, we may obtain the sub–weight matrices P_1 and P_2 for group one and group two, respectively. These two weight matrices refer to two different

unit–weight variance factors, σ_1^2 and σ_2^2, respectively. The total variance–covariance matrix can then be written as:

$$Q = \begin{bmatrix} \sigma_1^2 P_1^{-1} & \mathcal{O} \\ \mathcal{O} & \sigma_2^2 P_2^{-1} \end{bmatrix} = \sigma_1^2 Q_1 + \sigma_2^2 Q_2 \qquad (4.2.5a)$$

where: $$Q_1 = \begin{bmatrix} P_1^{-1} & \mathcal{O} \\ \mathcal{O} & \mathcal{O} \end{bmatrix}; \quad Q_2 = \begin{bmatrix} \mathcal{O} & \mathcal{O} \\ \mathcal{O} & P_2^{-1} \end{bmatrix} \qquad (4.2.5b)$$

(4.2.5a) and (4.2.5b) are the variance–covariance model for observations of two independent types. It is a special case of the general form (4.2.4c), with two variance components corresponding to the variance factors of the two independent groups. ∎

Example 4.2.2

meterlo

Let us now look at distance measurements using Electromagnetic Distance Measurement (EDM) instruments. Most EDM manufacturers define the distance accuracy of their EDM instruments by a constant factor plus a factor proportional to the length, such as 5 mm ± 3 ppm, where ppm stands for part per million. Thus one may assume that the variance of a measured distance l_i may be written as:

$$s_i^2 = \sigma_1^2 + l_i \sigma_2^2 \qquad (4.2.6a)$$

If we assume that any two distance measurements are independent of each other, the variance–covariance matrix of n distance measurements then becomes a diagonal matrix:

$$Q = \begin{bmatrix} s_1^2 & & & \\ & s_2^2 & & \\ & & \cdots & \\ & & & s_n^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2+l_1\sigma_2^2 & & & \\ & \sigma_1^2+l_2\sigma_2^2 & & \\ & & \cdots & \\ & & & \sigma_1^2+l_n\sigma_2^2 \end{bmatrix}$$

$$= \sigma_1^2 Q_1 + \sigma_2^2 Q_2 \qquad (4.2.6b)$$

where $$Q_1 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \cdots & \\ & & & 1 \end{bmatrix}; \quad Q_2 = \begin{bmatrix} l_1 & & & \\ & l_2 & & \\ & & \cdots & \\ & & & l_n \end{bmatrix} \qquad (4.2.6c)$$

§4.2.2 Helmert's Method in Adjustment by Elements

Suppose that the variance–covariance matrix of the observations have been decomposed by (4.2.4c) with u variance–covariance components: $\sigma_1, \sigma_2, \cdots, \sigma_q$. Let σ_j^0 denote

an approximate value of σ_j ($1 \leq j \leq q$). Then an approximate variance–covariance matrix can be obtained from (4.2.4c):

$$Q_0 \equiv \sum_{j=1}^{q} \sigma_j^0 \underset{n \times n}{Q_j} \tag{4.2.7a}$$

Furthermore, let us assume that the inverse matrix of Q_0 can be written as the sum of q matrices $C_1, C_2, \cdots, C_q$:

$$Q_0^{-1} = \left[\sum_{j=1}^{q} \sigma_j^0 \underset{n \times n}{Q_j} \right]^{-1} = \sum_{j=1}^{q} C_j \tag{4.2.7b}$$

The decomposition (4.2.7b) is generally not unique but normally possible. A special choice of C_j's is :

$$C_j = Q_0^{-1} \sigma_j^0 Q_j Q_0^{-1} \tag{4.2.7c}$$

It is trivial to show that C_j above satisfies (4.2.7b):

$$\sum_{j=1}^{q} C_j = \sum_{j=1}^{q} \left[Q_0^{-1} \sigma_j^0 Q_j Q_0^{-1} \right] = Q_0^{-1} \left[\sum_{j=1}^{q} \sigma_j^0 Q_j \right] Q_0^{-1} = Q_0^{-1} Q_0 Q_0^{-1} = Q_0^{-1}$$

Now we are ready to present Helmert's method for estimating the variance–covariance components σ_j ($j = 1, 2, \cdots, u$) using the functional model of adjustment by elements. Assume that we have the following observation equations [Cf (3.1.2)]:

$$\underset{n \times 1}{L} - \underset{n \times 1}{\epsilon} = \underset{n \times m}{A} \underset{m \times 1}{X} \tag{4.2.8a}$$

Using the approximate variance–covariance matrix Q_0 , the least squares estimate of the residuals follows [Cf (3.1.15)]:

$$\hat{\epsilon} = L - A(A^T Q_0^{-1} A)^{-1} A^T Q_0^{-1} L = M_0 L \tag{4.2.8b}$$

where: $$M_0 \equiv I - A(A^T Q_0^{-1} A)^{-1} A^T Q_0^{-1} \tag{4.2.8c}$$

with $$M_0 A \equiv 0 \tag{4.2.8d}$$

Now let us calculate the expectation of the following quadratic form of $\hat{\epsilon}$:

$$g_i \equiv \hat{\epsilon}^T C_i \hat{\epsilon} = L^T M_0^T C_i M_0 L \tag{4.2.9a}$$

$$\begin{aligned} E(g_i) &= E[\hat{\epsilon}^T C_i \hat{\epsilon}] = E[L^T M_0^T C_i M_0 L] = E[\epsilon^T M_0^T C_i M_0 \epsilon] \\ &= E[\epsilon^T M_0^T C_i M_0 \epsilon] = tr[M_0^T C_i M_0 E(\epsilon\epsilon^T)] = tr[M_0^T C_i M_0 Q] \\ &= \sum_{j=1}^{q} tr[M_0^T C_i M_0 Q_j] \sigma_j = \sum_{j=1}^{q} h_{ij} \sigma_j \end{aligned} \tag{4.2.9b}$$

where $h_{ij} \equiv tr[M_0^T C_i M_0 Q_j]$. (4.2.9c)

For i = 1, 2, 3, $\cdots$, q, (4.2.9b) leads to the following linear equation system:

$$\underset{q\times q}{H}\ \underset{q\times 1}{\sigma} = E(\underset{q\times 1}{g}) \tag{4.2.10}$$

where: $$H \equiv \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1q} \\ h_{21} & h_{22} & \cdots & h_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ h_{q1} & h_{q2} & \cdots & h_{qq} \end{bmatrix}; \quad \sigma \equiv \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \cdots \\ \sigma_q \end{bmatrix}; \quad g \equiv \begin{bmatrix} g_1 \\ g_2 \\ \cdots \\ g_q \end{bmatrix}$$

and h_{ij} in H is defined by (4.2.9c).

From (4.2.10), a Quadratic Unbiased Estimate (QUE) $\hat{\sigma}$ of σ is obtained:

$$H\ \hat{\sigma} = g \tag{4.2.11}$$

When $|H| \neq 0$, we have:

$$\hat{\sigma} = H^{-1} g \quad \text{and} \quad E(\hat{\sigma}) = H^{-1} E(g) = \sigma. \tag{4.2.12}$$

Eqs. (4.2.9a), (4.2.9c) and (4.2.11) are basic formulas for calculating a quadratic unbiased estimate of variance–covariance components using Helmert's method in Gauss–Markov model.

The above description can be regarded as a general formulation of Helmert's method. Historically, Helmert's method was derived for a specific case when the observations consist of two independent groups of measurements, e.g. angles and distances. The detailed formulas for this special case will be given below. Another special estimate of Helmert type is related to the special decomposition in (4.2.7c).

Helmert's Method for Two Independent Groups of Observations

Suppose that our observations can be divided into two independent groups L_1 and L_2 with different types of observations. We also assume that there are n_1 measurements in L_1 and n_2 measurements in L_2. The observation equations may read as:

$$\begin{bmatrix} L_1 \\ L_2 \end{bmatrix} - \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} X$$

Our variance–covariance model is defined as in (4.2.5a) with two variance components σ_1^2 and σ_2^2:

$$Q = E(\epsilon\epsilon^T) = \begin{bmatrix} \sigma_1^2 P_1^{-1} & 0 \\ 0 & \sigma_2^2 P_2^{-1} \end{bmatrix} = \sigma_1^2 Q_1 + \sigma_2^2 Q_2 \tag{4.2.13}$$

where P_1 and P_2 are the weight matrices of L_1 and L_2, respectively. Q_1 and Q_2 are as defined in (4.2.5b). Let σ_1^0 and σ_2^0 denote approximate values of σ_1 and σ_2, respectively. Then an approximate variance–covariance matrix can be obtained:

$$Q_0 = \begin{bmatrix} \sigma_1^{0\,2}\, P_1^{-1} & 0 \\ 0 & \sigma_2^{0\,2}\, P_2^{-1} \end{bmatrix} = \sigma_1^{0\,2}\, Q_1 + \sigma_2^{0\,2}\, Q_2 \qquad (4.2.14)$$

Similar to (4.2.7b), Q_0^{-1} may be written as:

$$Q_0^{-1} = \begin{bmatrix} \sigma_1^{0\,2}\, P_1^{-1} & 0 \\ 0 & \sigma_2^{0\,2}\, P_2^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_1^{0\,-2}\, P_1 & 0 \\ 0 & \sigma_2^{0\,-2}\, P_2 \end{bmatrix} = C_1 + C_2 \qquad (4.2.15)$$

$$\text{where:} \quad C_1 \equiv \begin{bmatrix} \sigma_1^{0\,-2}\, P_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}; \quad C_2 \equiv \begin{bmatrix} 0 & 0 \\ 0 & \sigma_2^{0\,-2}\, P_2^{-1} \end{bmatrix} \qquad (4.2.16)$$

As we have only two variance–covariance components in (4.2.12), g in (4.2.11) will be a 2–dimensional vector and H will be a 2x2 matrix. Inserting (4.2.12) and (4.2.13) into (4.2.9a) and (4.2.9c), we can get all elements g_1, g_2 in g and h_{11}, h_{12}, h_{21}, h_{22} in H. The linear equation system (4.2.11) is now reduced to the following equations:

$$\begin{bmatrix} \bar{h}_{11} & \bar{h}_{12} \\ \bar{h}_{21} & \bar{h}_{22} \end{bmatrix} \begin{bmatrix} \hat{\sigma}_1^2/\sigma_1^{0\,2} \\ \hat{\sigma}_2^2/\sigma_2^{0\,2} \end{bmatrix} = \begin{bmatrix} \hat{\epsilon}_1^{\,T}\, \sigma_1^{0\,-2} P_1\, \hat{\epsilon}_1 \\ \hat{\epsilon}_2^{\,T}\, \sigma_2^{0\,-2} P_2\, \hat{\epsilon}_2 \end{bmatrix} \qquad (4.2.17)$$

where $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$ are the least squares estimates of residuals to L_1 and L_2, respectively. $\bar{h}_{ij}$ are defined as follows:

$$\left.\begin{aligned} \bar{h}_{11} &= n_1 - 2\ \mathrm{tr}(N^{-1}N_{11}) + \mathrm{tr}(N^{-1}N_{11}N^{-1}N_{11}) \\ \bar{h}_{12} &= \mathrm{tr}(N^{-1}N_{11}N^{-1}N_{22}) = \bar{h}_{21} \\ \bar{h}_{22} &= n_2 - 2\ \mathrm{tr}(N^{-1}N_{22}) + \mathrm{tr}(N^{-1}N_{22}N^{-1}N_{22}) \end{aligned}\right\} \qquad (4.2.18)$$

$$\left.\begin{aligned} N &= A^T Q_0^{-1} A = N_{11} + N_{22} \\ \text{where:} \quad N_{11} &= \sigma_1^{0\,-2}\, A_1^{\,T} P_1 A_1 \\ N_{22} &= \sigma_2^{0\,-2}\, A_2^{\,T} P_2 A_2 \end{aligned}\right\} \qquad (4.2.19)$$

The quadratic unbiased estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ can be directly solved from (4.2.17):

$$\begin{bmatrix} \hat{\sigma}_1^2/\sigma_1^{0\,2} \\ \hat{\sigma}_2^2/\sigma_2^{0\,2} \end{bmatrix} = \begin{bmatrix} \bar{h}_{11} & \bar{h}_{12} \\ \bar{h}_{21} & \bar{h}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\epsilon}_1^{\,T}\, \sigma_1^{0\,-2} P_1\, \hat{\epsilon}_1 \\ \hat{\epsilon}_2^{\,T}\, \sigma_2^{0\,-2} P_2\, \hat{\epsilon}_2 \end{bmatrix} \qquad (4.2.20)$$

Helmert's Method for the Special Decomposition (4.2.7c)

As mentioned before, the decomposition (4.2.7b) of the approximate variance–covariance matrix Q_0^{-1} is generally not unique. Therefore, for arbitrary choices of C_j's it is not generally guaranteed that matrix H in (4.2.10) as defined by (4.2.9c) is symmetrical. However, for the special decomposition (4.2.15), the resulted linear equation system for $\hat{\sigma}$ turns out to be symmetrical [Cf (4.2.17) and (4.2.18)]. Furthermore, as shown below, the particular decomposition (4.2.7c) will also lead to a symmetrical equation system for $\hat{\sigma}$.

Substitution of (4.2.7c) in (4.2.9a) and (4.2.9c) leads to:

$$g_i = \sigma_i^0 \, \hat{\epsilon}^T Q_0^{-1} Q_i Q_0^{-1} \hat{\epsilon} = \sigma_j^0 \, u_i \tag{4.2.21}$$

and $$h_{ij} = \sigma_i^0 \, \mathrm{tr}[M_0^T Q_0^{-1} Q_i Q_0^{-1} M_0 Q_j] = \sigma_i^0 \, \mathrm{tr}[R_0 Q_i R_0 Q_j]$$

$$= \sigma_i^0 \, s_{ij} \tag{4.2.22}$$

where:

$$u_i \equiv \hat{\epsilon}^T Q_0^{-1} Q_j Q_0^{-1} \hat{\epsilon} \tag{4.2.23}$$

$$s_{ij} \equiv \mathrm{tr}[R_0 Q_i R_0 Q_j] \tag{4.2.24}$$

$$R_0 \equiv Q_0^{-1} M_0 = R_0^T \tag{4.2.25}$$

Inserting (4.2.21) and (4.2.22) into (4.2.11), we obtain:

$$S \, \hat{\sigma} = u \tag{4.2.26}$$

where: $$S \equiv \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1q} \\ s_{21} & s_{22} & \cdots & s_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ s_{q1} & s_{q2} & \cdots & s_{qq} \end{bmatrix}; \quad \hat{\sigma} \equiv \begin{bmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \cdots \\ \hat{\sigma}_q \end{bmatrix}; \quad u \equiv \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_q \end{bmatrix}$$

The unbiased estimate $\hat{\sigma}$ is then solved from (4.2.26):

$$\hat{\sigma} = S^{-1} u \tag{4.2.27}$$

From (4.2.24), we can easily show:

$$s_{ij} \equiv s_{ji} \tag{4.2.28}$$

and $$S = S^T \tag{4.2.29}$$

Example 4.2.3

To further demonstrate Helmert's method for estimation of variance–covariance components, we apply the general formulas (4.2.9a), (4.2.9c) and (4.2.11) to estimate

the variance factor σ_0^2 in (3.1.6). In this case, $u = 1$, i.e. we have only one variance component:

$$Q = \sigma_0^2 P^{-1} = \sigma_1 Q_1,$$

where P is the weight matrix, $\sigma_1 \equiv \sigma_0^2$ and $Q_1 \equiv P^{-1}$. Let α_0^2 denote an approximate value of σ_0^2 and Q_0 denote an approximate variance–covariance matrix such that:

$$Q_0 = \alpha_0^2 P^{-1} \text{ and } Q_0^{-1} = \alpha_0^{-2} P \equiv C_1$$

Then from (4.2.8c), we have:

$$M_0 = I - A (A^T Q_0^{-1} A)^{-1} A^T Q_0^{-1} = I - A(A^T P A)^{-1} A^T P$$

According to (4.2.9a) and (4.2.9c), g_1 and h_{11} are as follows:

$$\begin{aligned} g_1 &\equiv \hat{\epsilon}^T C_1 \hat{\epsilon} = \hat{\epsilon}^T \alpha_0^{-2} P \hat{\epsilon} = \alpha_0^{-2} \hat{\epsilon}^T P \hat{\epsilon} \\ h_{11} &= \text{tr}[M_0^T C_1 M_0 Q_1] = \text{tr}\left[M_0^T \alpha_0^{-2} P M_0 P^{-1}\right] = \alpha_0^{-2} \text{tr}\left[M_0^T P M_0 P^{-1}\right] \\ &= \alpha_0^{-2} \text{tr}\left[[I - PA(A^T PA)^{-1} A^T] P [I - A(A^T PA)^{-1} A^T P] P^{-1}\right] \\ &= \alpha_0^{-2} \text{tr}[I - PA(A^T PA)^{-1} A^T] = \alpha_0^{-2} (n - m) \end{aligned}$$

where n is the number of observations and m is the number of unknown parameters. From (4.2.11), the quadratic unbiased estimate of σ_0^2 is then obtained:

$$\hat{\sigma}_0^2 = g_1/h_{11} = \hat{\epsilon}^T P \hat{\epsilon}/(n-m),$$

which is identical as (3.1.21).

§4.2.3 Helmert's Method in Condition Adjustment

Assume that we have the following condition equations:

$$B \epsilon = W \qquad (4.2.30)$$

Assume also that the theoretical variance–covariance matrix Q of ϵ is defined by (4.2.4c), that an approximate variance–covariance matrix Q_0 using approximate variance–covariance components σ_j^0 is defined by (4.2.7a) and that Q_0^{-1} can be decomposed as in (4.2.7b).

The least squares estimate of $\hat{\epsilon}$ with the approximate variance–covariance matrix Q_0 is then given by (2.1.10):

$$\hat{\epsilon} = Q_0 B^T (B Q_0 B^T)^{-1} W = G_0 W \tag{4.2.31}$$

where: $$G_0 \equiv Q_0 B^T (B Q_0 B^T)^{-1} . \tag{4.2.32}$$

Now let us calculate the following quadratic form of $\hat{\epsilon}$ and its expectation:

$$g_i \equiv \hat{\epsilon}^T C_i \hat{\epsilon} = W^T G_0^T C_i G_0 W \tag{4.2.33}$$

$$E[\hat{\epsilon}^T C_i \hat{\epsilon}] = tr[G_0^T C_i G_0 E(WW^T)] = tr[G_0^T C_i G_0 B Q B^T]$$

$$= tr\left[G_0^T C_i G_0 B \left[\sum_{j=1}^{q} \sigma_j Q_j\right] B^T\right] = \sum_{j=1}^{q} \left[tr[G_0^T C_i G_0 D_j]\sigma_j\right] \tag{4.2.34}$$

where $D_j \equiv B Q_j B^T$. For $i = 1, 2, 3, \cdots, q$, (4.2.34) leads to the following linear equation system:

$$H \sigma = E(g) \tag{4.2.35}$$

where: $$H \equiv \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1q} \\ h_{21} & h_{22} & \cdots & h_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ h_{q1} & h_{q2} & \cdots & h_{qq} \end{bmatrix}; \quad \sigma \equiv \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \cdots \\ \sigma_q \end{bmatrix}; \quad g \equiv \begin{bmatrix} g_1 \\ g_2 \\ \cdots \\ g_q \end{bmatrix}$$

with $$h_{ij} \equiv tr[G_0^T C_i G_0 D_j] \qquad (1 \leq i,j \leq q) \tag{4.2.36}$$

From (4.2.35), we then obtain an unbiased quadratic estimate $\hat{\sigma}$ for σ:

$$H \hat{\sigma} = g \tag{4.2.37}$$

When $|H| \neq 0$, we have:

$$\hat{\sigma} = H^{-1} g \tag{4.2.38}$$

Eqs. (4.2.33), (4.2.36) and (4.2.38) are the basic formulas for calculating the Helmert–type estimate of variance–covariance components using condition adjustment model. Naturally, for the same geodetic problem and same variance–covariance model (4.2.4c), (4.2.12) and (4.2.38) should give the same quadratic unbiased estimate to the variance–covariance components.

If we use the special decomposition in (4.2.7c), (4.2.37) will become a symmetrical linear equation system:

$$S \hat{\sigma} = u \tag{4.2.39}$$

where: $$S \equiv \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1q} \\ s_{21} & s_{22} & \cdots & s_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ s_{q1} & s_{q2} & \cdots & s_{qq} \end{bmatrix}; \quad \hat{\sigma} \equiv \begin{bmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \cdots \\ \hat{\sigma}_q \end{bmatrix}; \quad u \equiv \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_q \end{bmatrix}$$

$$s_{ij} \equiv \mathrm{tr}[D_0^{-1} D_i D_0^{-1} D_j] = s_{ji} \qquad (1 \le i,j \le q) \qquad (4.2.40)$$

$$D_0 \equiv B Q_0 B^T \qquad (4.2.41)$$

$$u_i \equiv \hat{\epsilon}^T Q_0^{-1} Q_i Q_0^{-1} \hat{\epsilon} \qquad (1 \le i \le q) \qquad (4.2.42)$$

$$\hat{\epsilon} = Q_0 B^T (B Q_0 B^T)^{-1} W \qquad (4.2.43)$$

Similar to Helmert's method in Gauss–Markov model, the quadratic unbiased estimate in (4.2.38) can be applied to the special case with two types of independent observations L_1 and L_2. Assume that our condition equations are as follows:

$$[B_1 \; B_2] \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = W$$

where ϵ_1 and ϵ_2 are residuals of L_1 and L_2, respectively. Using the notations in Eqs. (4.2.13), (4.2.14), (4.2.15) and (4.2.16), (4.2.37) can be reduced to:

$$\begin{bmatrix} \bar{h}_{11} & \bar{h}_{12} \\ \bar{h}_{21} & \bar{h}_{22} \end{bmatrix} \begin{bmatrix} \hat{\sigma}_1^2/\sigma_1^{0\,2} \\ \hat{\sigma}_2^2/\sigma_2^{0\,2} \end{bmatrix} = \begin{bmatrix} \hat{\epsilon}_1^T \sigma_1^{0\,-2} P_1 \hat{\epsilon}_1 \\ \hat{\epsilon}_2^T \sigma_2^{0\,-2} P_2 \hat{\epsilon}_2 \end{bmatrix} \qquad (4.2.44)$$

where $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$ are the least squares estimates of residuals to L_1 and L_2, respectively.

$\bar{h}_{ij}$ are defined as follows:

$$\left.\begin{aligned} \bar{h}_{11} &= \mathrm{tr}(N^{-1} N_{11} N^{-1} N_{11}) \\ \bar{h}_{12} &= \mathrm{tr}(N^{-1} N_{11} N^{-1} N_{22}) = \bar{h}_{21} \\ \bar{h}_{22} &= \mathrm{tr}(N^{-1} N_{22} N^{-1} N_{22}) \end{aligned}\right\} \qquad (4.2.45)$$

where:

$$\left.\begin{aligned} N &= B\, Q_0\, B^T = N_{11} + N_{22} \\ N_{11} &= \sigma_1^{0\,2}\, B_1 P_1^{-1} B_1^T \\ N_{22} &= \sigma_2^{0\,2}\, B_2 P_2^{-1} B_2^T \end{aligned}\right\} \qquad (4.2.46)$$

The quadratic unbiased estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ can be directly solved from (4.2.44).

§4.2.4 BQUE in Adjustment by Elements

From (4.2.9a) and (4.2.12), we can see that the Helmert–type estimate $\hat{\sigma}$ is an unbiased estimate of the variance–covariance components σ based on the quadratic forms of the observations or their residuals. As the choice of matrices C_j's is arbitrary and meanwhile no optimal selection has been attempted, one may wonder whether this method can guarantee a good estimate of σ. In statistical inference and estimation

theory, one often defines the optimal estimate as the one which has minimum variance. Applying this concept to the problem of variance–covariance component estimation, we may define our optimal estimate: *Among all quadratic unbiased estimates (QUE) of variance–covariance components, the estimate that attains the minimum variance is called the Best Quadratic Unbiased Estimate (BQUE).*

In order to simplify the discussion of the variances of the estimated variance–covariance components, we consider a linear combination of q variance–covariance components:

$$\rho \equiv p^T \sigma \equiv p_1 \sigma_1 + p_2 \sigma_2 + \cdots + p_q \sigma_q \equiv \sum_{j=1}^{q} p_i \sigma_j \qquad (4.2.47)$$

where p_j's $(1 \leq j \leq q)$ denote some constants from which vector p is constructed. $p^T \sigma$ can be any one of the q components, say σ_k, if one chooses $p_k = 1$ and $p_j = 0$ for $j \neq k$.

Assume that we have the following observation equations:

$$\underset{n \times 1}{L} - \underset{n \times 1}{\epsilon} = \underset{n \times m}{A} \; \underset{m \times 1}{X} \qquad (4.2.48)$$

It is also assumed that the variance–covariance matrix of ϵ can be decomposed as in (4.2.4c):

$$\underset{n \times n}{Q} = \sum_{j=1}^{q} \sigma_j \underset{n \times n}{Q_j} \qquad (4.2.4c)$$

and that from a set of approximate variance–covariance components σ_j^0, an approximate variance–covariance matrix Q_0 is obtained as in (4.2.7a):

$$Q_0 \equiv \sum_{j=1}^{q} \sigma_j^0 \underset{n \times n}{Q_j} \qquad (4.2.7a)$$

We now look at the quadratic form $L^T ML$, where M is an arbitrary symmetrical matrix:

$$E\left[L^T ML\right] = \mathrm{tr}\left[M\, E(LL^T)\right] = \mathrm{tr}\left[M\, E\left[\epsilon\epsilon^T + AXX^TA^T + 2\, AX\epsilon^T\right]\right]$$

$$= \mathrm{tr}\left[M\,(Q + AXX^TA^T)\right] = \sum_{j=1}^{q} \mathrm{tr}(MQ_j)\,\sigma_j + X^T A^T MA\, X \equiv \sum_{j=1}^{q} p_i \sigma_j$$

The last identity above holds for any arbitrary vector p iff:

$$\left.\begin{aligned} A^T MA &= 0 \\ \mathrm{tr}(MQ_j) &= p_j \quad (1 \leq j \leq q) \end{aligned}\right\} \qquad (4.2.49)$$

(4.2.49) is the sufficient and necessary conditions for that the quadratic form $L^T ML$ is an unbiased estimate of $\rho \equiv p^T \sigma$.

In geodetic adjustment, one often replace the unknown parameters X by $X_0+\delta X$ where X_0 stands for approximate values of X and δX the corresponding correction, so that the observation equations actually used are :

$$\underset{n\times 1}{\bar{L}} - \underset{n\times 1}{\epsilon} = \underset{n\times m}{A} \cdot \underset{m\times 1}{\delta X}$$

where $\bar{L} \equiv L - AX_0$. This replacement is necessary when the non–linear observation equations are to be linearized. In the case of linear observation equations, this replacement is not needed but allowed, as it does not affect the final estimates $\hat{\epsilon}$ and $\hat{X}$ for both cases (disregarding errors caused by neglecting higher order terms during linearization). In other words, the least squares estimates $\hat{\epsilon}$ and $\hat{X}$ are *invariant* with respect to the shift in the unknown parameters (such as a shift caused by different approximate values X_0).

Similarly, we may demand that our QUE above, L^TML, is also invariant with the parameter shift, which implies that for any arbitrary shift X_0 , the following equation:

$$\bar{L}^TM\bar{L} \equiv L^TML$$

Considering that fact that L^TML satisfies (4.2.48), we have:

$$(L-AX)^TM(L-AX_0) = L^TML + 2\,L^TMAX_0 \equiv L^TML$$

which holds iff $MA \equiv 0$. As $MA = 0$ automatically leads to $A^TMA = 0$, we may conclude that the sufficient and necessary conditions for that L^TML is an Invariant Quadratic Unbiased Estimate (IQUE) of $p^T\sigma$ is:

$$\left.\begin{aligned} M\,A &= 0 \\ tr(MQ_j) &= p_j \qquad (1\le j\le q) \end{aligned}\right\} \tag{4.2.50}$$

Now let us try to find the variance of the above IQUE, L^TMA. Under (4.2.50), we have:

$$L^TML = (\epsilon+AX)^T\,M\,(\epsilon+AX) = \epsilon^TM\epsilon + 2\,\epsilon^T\,MA\,X + X^T\,A^TMA\,X = \epsilon^TM\epsilon$$

For normally distributed random vector ϵ with variance–covariance matrix Q, it can be proved (Koch, 1981) that the variance of the quadratic form $\epsilon^TM\epsilon$ (and hence the variance of L^TML) is as follows:

$$Var(\epsilon^TM\epsilon) \equiv Var(L^TML) = 2\,tr(MQMQ) \tag{4.2.51}$$

Considering (4.2.50), (4.2.51) and the definition of BQUE, the problem of finding a

BQUE of $\rho \equiv p^T \sigma$ is to find a symmetrical matrix M such that:

$$\left\{ \begin{array}{lcl} MA & = & 0 \\ tr(MQ_j) & = & p_j \\ tr(MQMQ) & = & \text{minimum} \end{array} \right\} \qquad (4.2.52)$$

This is a conditional minimization problem with matrix M as variables. However, the variance–covariance components and consequently the variance–covariance matrix Q above are also unknown. To overcome this dilemma, we may replace the theoretical variance–covariance matrix Q by the approximate one Q_0 defined in (4.2.7a). Our problem becomes now:

$$\left\{ \begin{array}{lcl} MA & = & 0 \\ tr(MQ_j) & = & p_j \\ tr(MQ_0MQ_0) & = & \text{minimum} \end{array} \right\} \qquad (4.2.53)$$

In Appendix A1.3, it has been proved that the solution of (4.2.53) is as follows:

$$M = \sum_{i=1}^{q} \lambda_i R_0 Q_i R_0 \qquad (4.2.54)$$

$$R_0 \equiv Q_0^{-1} - Q_0^{-1} A (A^T Q_0^{-1} A)^{-1} A^T Q_0^{-1} \qquad (4.2.55)$$

λ_i's $(1 \le i \le q)$ are solved from the following equation system:

$$\underset{qxq}{S} \; \underset{qx1}{\lambda} = \underset{qx1}{p} \qquad \text{or:} \qquad \lambda = S^{-1} p \qquad (4.2.56)$$

where:
$$\underset{qxq}{S} \equiv \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1q} \\ s_{21} & s_{22} & \cdots & s_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ s_{q1} & s_{q2} & \cdots & s_{qq} \end{bmatrix}; \quad \lambda \equiv \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_q \end{bmatrix}; \quad p \equiv \begin{bmatrix} p_1 \\ p_2 \\ \cdots \\ p_q \end{bmatrix}$$

with:
$$s_{ij} = tr(R_0 Q_i R_0 Q_j) = s_{ji} \qquad (1 \le i,j \le q) \qquad (4.2.57)$$

Finally, the BQUE of $\rho = p^T \sigma$ is obtained :

$$\hat{\rho} = L^T \hat{M} L \qquad (4.2.58)$$

with M computed from (4.2.54).

The BQUE of each variance–covariance component can be obtained by specifying vector p. For component σ_j $(1 \le j \le q)$, we have accordingly $p^T = (p_1, p_2, \cdots, p_q)$ with $p_j = 1$ and $p_k = 0$ $(k \ne j)$. Denoting the inverse matrix of S in (4.2.56) by G:

$$S^{-1} \equiv G = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1q} \\ G_{21} & G_{22} & \cdots & G_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ G_{q1} & G_{q2} & \cdots & G_{qq} \end{bmatrix}$$

the BQUE , $\hat{\sigma}_j$, can be obtained from (4.2.56) and (4.2.54):

$$\lambda = S^{-1} p = G\ (0, 0, \cdots, 0, 1, 0, \cdots, 0)^T = \begin{bmatrix} G_{1j} \\ G_{2j} \\ \cdots \\ G_{qj} \end{bmatrix}$$

$$\hat{\sigma}_j = L^T M L = L^T \Big[\sum_{i=1}^{q} \lambda_i R_0 Q_i R_0 \Big] L = \sum_{i=1}^{q} G_{ij} u_i \qquad (4.2.59)$$

where: $$u_i \equiv L^T R_0 Q_i R_0 L = \hat{\epsilon}^T Q_0^{-1} Q_i Q_0^{-1} \hat{\epsilon} \qquad (4.2.60)$$

$$\hat{\epsilon} = \left[I - A(A^T Q_0^{-1} A)^{-1} A^T Q_0^{-1} \right] L \qquad (4.2.61)$$

Letting j run from 1 to q, (4.2.59) leads to the following linear equation system for vector $\hat{\sigma}$, the BQUE of variance–covariance components:

$$\hat{\sigma} \equiv \begin{bmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \cdots \\ \hat{\sigma}_q \end{bmatrix} = G u = S^{-1} u \qquad (4.2.62)$$

where $u \equiv (u_1, u_2, \cdots, u_q)^T$ and u_j's are defined as in (4.2.60).

The BQUE $\hat{\sigma}$ in (4.2.62) is obtained with the initial, approximate values σ_j^0 $(1 \le j \le q)$. Therefore, we may call (4.2.62) a *locally best quadratic unbiased estimate* – around the initial values σ_j^0 $(1 \le j \le q)$. Accordingly, one may call the solution to the theoretical problem in (4.2.52) the *globally best quadratic unbiased estimate*, which is hardly possible to estimate in practice.

Eqs. (4.2.60), (4.2.57) and (4.2.62) are the basic formulas for calculating the BQUE $\hat{\sigma}$ of variance–covariance components using the functional model of adjustment by elements. Comparing these three basic formulas with Eqs. (4.2.23), (4.2.24) and (4.2.27), we can find that our (locally) BQUE $\hat{\sigma}$ is actually the same as a QUE of Helmert type for a special decomposition of Q_0^{-1} as given in (4.2.7c). In other words, BQUE is a special QUE of Helmert type with minimum variance, while the general estimate of Helmert type given in (4.2.12) is only an unbiased estimate without any optimization.

Example 4.2.4

We consider the BQUE of the unit–weight standard error σ_0 in adjustment by elements [Cf (4.2.48)]. Our variance–covariance model now consists of only one variance component σ_0^2 :

$$Q = E(\epsilon\epsilon^T) = \sigma_0^2 P^{-1} = \sigma_1 Q_1$$

where P = the weight matrix of L

$\sigma_1 \equiv \sigma_0^2$

$Q_1 \equiv P^{-1}$

Let α^2 denote an approximate value of σ_0^2 . The approximate variance–covariance matrix Q_0 is then obtained:

$$Q_0 = \alpha^2 Q_1 = \alpha^2 P^{-1}$$

From (4.2.60), (4.2.61) and (4.2.57), we have:

$$u_1 = \hat{\epsilon}^T Q_0^{-1} Q_1 Q_0^{-1} \hat{\epsilon} = \alpha^{-4} \hat{\epsilon}^T P \hat{\epsilon}$$

$$\hat{\epsilon} = \left[I - A(A^T Q_0^{-1} A)^{-1} A^T Q_0^{-1}\right] L = \left[I - A(A^T PA)^{-1} A^T P\right] L$$

$$s_{11} = \mathrm{tr}\left[\alpha^{-2}[P - PA(A^T PA)^{-1} A^T P]\, P^{-1}\, \alpha^{-2}[P - PA(A^T PA)^{-1} A^T P]\, P^{-1}\right]$$

$$= \alpha^{-4}\, \mathrm{tr}[I - PA(A^T PA)^{-1} A^T] = \alpha^{-4}(n - m)$$

From (4.2.62), we get the BQUE of ϵ_0^2:

$$\hat{\sigma}_0^2 = u_1/s_{11} = \frac{\hat{\epsilon}^T P \hat{\epsilon}}{n - m}$$

which is identical to (3.1.21). ∎

§4.2.5 BQUE in Condition Adjustment

Using similar notations as before, we assume condition equations,

$$B \epsilon = W$$

a variance–covariance model as in (4.2.4c):

$$\underset{n \times n}{Q} = \sum_{j=1}^{q} \sigma_j \underset{n \times n}{Q_j} \qquad (4.2.4c)$$

and an approximate variance–covariance matrix Q_0 as in (4.2.7a):

$$Q_0 \equiv \sum_{j=1}^{q} \sigma_j^0 \underset{n \times n}{Q_j} \qquad (4.2.7a)$$

Now we want to use a quadratic form W^TMW, where M is a symmetrical matrix to be specified, to estimate the linear combination of variance–covariance components as given in (4.2.47):

$$\rho \equiv p^T\sigma \equiv \sum_{j=1}^{q} p_i\sigma_j \qquad (4.2.47)$$

First, we look at the expectation of the proposed quadratic form:

$$E(W^TMW) = tr[M\,E(WW^T)] = tr(M\,BQB^T) = \sum_{j=1}^{q} tr(MD_j)\,\sigma_j ,$$

where $D_j \equiv BQ_jB^T$. Comparing the above equation with (4.2.7), we find that the sufficient and necessary condition for W^TMW to be an unbiased estimate of $p^T\sigma$ is:

$$tr(MD_j) = p_j , \quad (1 \le j \le q) \qquad (4.2.63)$$

Secondly, we calculate the variance of W^TMW under the assumption that ϵ has multi–variate normally distribution (Koch, 1980):

$$Var(W^TMW) = 2\,tr[M\,E(WW^T)\,M\,E(WW^T)] = 2tr(MDMD) \qquad (4.2.64)$$

where: $D \equiv BQB^T = \sum_{j=1}^{q} D_j\,\sigma_j$.

The best quadratic unbiased estimate of $p^T\sigma$ is one particular quadratic form $W^T\hat{M}W$ which satisfies the condition (4.2.63) and minimizes $var(W^TMW)$. To find such an estimate is equivalent to find a symmetrical matrix $\hat{M}$ such that:

$$\left.\begin{aligned} tr(\hat{M}D_j) &= p_j \qquad (1 \le j \le q) \\ tr(\hat{M}D\hat{M}D) &= \text{minimum} \end{aligned}\right\} \qquad (4.2.65)$$

Similar to (4.2.52), (4.2.66) is not possible to solve as both $\hat{M}$ and D are unknown. Therefore, we choose to search for *locally best quadratic unbiased estimates* for ρ and also σ with respect to a set of approximate variance–covariance components σ_j^0 $(1 \le j \le q)$. Replacing D in (4.2.65) by the approximate variance–covariance matrix Q_0 , the theoretical problem (4.2.65) now reduces to :

$$\left.\begin{aligned} tr(\hat{M}D_j) &= p_j \qquad (1 \le j \le q) \\ tr(\hat{M}D_0\hat{M}D_0) &= \text{minimum} \end{aligned}\right\} \qquad (4.2.66)$$

The solution to the above conditional minimization problem has been derived in

Appendix A1.3 and listed below:

$$\hat{\rho} = W^T \hat{M} W \tag{4.2.67}$$

$$\hat{M} = \sum_{i=1}^{q} \lambda_i D_0^{-1} D_i D_0^{-1} \tag{4.2.68}$$

where: $D_0 \equiv B Q_0 B^T$ (4.2.69)

λ_i's in (4.2.68) are to be solved from the following equations:

$$\underset{qxq}{S} \underset{qx1}{\lambda} = \underset{qx1}{p} \qquad \text{or:} \quad \lambda = S^{-1} p \tag{4.2.70}$$

where: $$\underset{qxq}{S} \equiv \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1q} \\ s_{21} & s_{22} & \cdots & s_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ s_{q1} & s_{q2} & \cdots & s_{qq} \end{bmatrix}; \quad \lambda \equiv \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_q \end{bmatrix}; \quad p \equiv \begin{bmatrix} p_1 \\ p_2 \\ \cdots \\ p_q \end{bmatrix}$$

with: $s_{ij} = \mathrm{tr}(D_0^{-1} D_i D_0^{-1} D_j) = s_{ji} \qquad (1 \le i,j \le q)$ (4.2.71)

It can be easily shown that the best quadratic unbiased estimate of $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_q)^T$ becomes:

$$\underset{qx1}{\hat{\sigma}} = \underset{qxq}{S^{-1}} \underset{qx1}{u} \tag{4.2.72}$$

where: $$u \equiv \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_q \end{bmatrix}$$

and $u_i \equiv W^T D_0^{-1} D_i D_0^{-1} W = \hat{\epsilon}^T Q_0^{-1} Q_i Q_0^{-1} \hat{\epsilon}$ (4.2.73)

$$\hat{\epsilon} = Q_0 B^T (B Q_0 B^T)^{-1} W \tag{4.2.74}$$

If we compare (4.2.73) and (4.2.71) with (4.2.42) and (4.2.40), respectively, we can see that the BQUE $\hat{\sigma}$ in condition adjustment is identical to the unbiased estimate $\hat{\sigma}$ of Helmert type [Cf (4.2.39)] for the special decomposition (4.2.7c).

For same geodetic network and same set of observations, the two BQUEs of σ given in (4.2.62) and (4.2.72) should be identical, as adjustment by elements and condition adjustment are two parallel but equivalent models.

§4.2.6 MINQUE

The MINQUE (MInimum Norm Quadratic Unbiased Estimate) theory in variance–covariance component estimation is closely associated with the well–known Indian mathematician C.R. Rao[1]. This estimate is based on the decomposition of the residual vector ϵ in the same manner as the variance–covariance matrix Q is decomposed. Among all quadratic unbiased estimates, MINQUE is obtained after minimizing the Euclidean norm of some selected matrix. Similar to Helmert's method and BQUE, adjustment by elements and condition adjustment will lead to the same MINQUE of variance–covariance components. Therefore, we choose below to describe the MINQUE theory only for the functional model of adjustment by elements.

Our adjustment model and variance–covariance model are the same as in (4.2.48), (4.2.4c) and (4.2.7a):

$$\underset{n\times 1}{L} - \underset{n\times 1}{\epsilon} = \underset{n\times m}{A}\ \underset{m\times 1}{X}$$

$$\underset{n\times n}{Q} = \sum_{j=1}^{q} \sigma_j \underset{n\times n}{Q_j}$$

$$Q_0 = \sum_{j=1}^{q} \sigma_j^0 \underset{n\times n}{Q_j}$$

Corresponding to decomposition (4.2.4c), we decompose the residual ϵ as follows:

$$\underset{n\times 1}{\epsilon} = \sum_{j=1}^{q} U_j \xi_j = \underset{n\times f}{U}\ \underset{f\times 1}{\xi} \qquad (4.2.75)$$

where: U_j = some given matrices of dimension $n\times f_j$

ξ_j = a stochastic vector such that:

$$E(\xi_j) = 0$$

$$E(\xi\xi_j^T) = \sigma_j I_j \qquad (I_j \text{ is a unit matrix of dimension } f_j \times f_j)$$

$$U \equiv (U_1, U_2, \cdots, U_q)$$

$$\xi \equiv (\xi_1^T, \xi_2^T, \cdots, \xi_q^T)^T$$

$$f = f_1 + f_2 + \cdots + f_q .$$

From the definition of ξ_j $(1\le j\le q)$, we can find the expectation and variance–covariance matrix of ξ :

---

[1] Rao had research cooperations with the Department of Geodesy and Photogrammetry at KTH. In 1977 the former professor in geodesy at KTH, prof Arne Bjerhammar, visited the Indian Statistical Instititue where Rao has been working. Prof Lars E Sjöberg also had personal communications with Rao's associate, Dr S.K. Mitra.

$$\left.\begin{aligned} E(\xi) &= 0 \\ E(\xi\xi^T) \equiv Q_{\xi\xi} &= \begin{bmatrix} \sigma_1 I_1 & & & \\ & \sigma_2 I_2 & & \\ & & \cdots & \\ & & & \sigma_q I_q \end{bmatrix} \end{aligned}\right\} \qquad (4.2.76)$$

where $Q_{\xi\xi}$ is a diagonal matrix and I_j is unit matrix of $f_j x f_j$. Combining (4.2.75) and (4.2.76), we have:

$$E(\epsilon\epsilon^T) = Q = U\, Q_{\xi\xi}\, U^T = \sum_{j=1}^{q} \sigma_j\, U_j U_j^T$$

Comparing the above equation with (4.2.4c) leads to:

$$Q_j \equiv U_j U_j^T \qquad (1 \le j \le q) \qquad (4.2.77)$$

Now let Δ denote the following diagonal matrix:

$$\underset{f\,x\,f}{\Delta} = \begin{bmatrix} p_1/f_1 I_1 & & & \\ & p_2/f_2 I_2 & & \\ & & \cdots & \\ & & & p_q/f_q I_q \end{bmatrix} \qquad (4.2.78)$$

It is easy to show that the quadratic form $\xi^T\Delta\xi$ is a unbiased estimate of $p^T\sigma$ [Cf. (4.2.47)]:

$$E(\xi^T\Delta\xi) = tr[\Delta\, E(\xi\xi^T)] = tr[\Delta\, Q_{\xi\xi}] = \sum_{j=1}^{q} p_j\, \sigma_j = p^T\sigma$$

In Section §4.2.3, we have already shown that the quadratic form L^TML for M satisfying (4.2.51) is also an unbiased estimate of $p^T\sigma$. Taking into account conditions in (4.2.51), the difference of these two estimates can be obtained :

$$L^TML - \xi^T\Delta\xi = \epsilon^TM\epsilon - \xi^T\Delta\xi = \xi^T\, G\, \xi \qquad (4.2.78)$$

where: $G \equiv U^TMU - \Delta$ (4.2.79)

The Euclidean norm $\|G\|$ of matrix G is defined as:

$$\|G\|^2 \equiv tr(GG^T) = tr(M\overline{Q}M\overline{Q}) - \|\Delta\|^2 \qquad (4.2.80)$$

where: $\overline{Q} \equiv UU^T = \sum_{j=1}^{q} U_j U_j^T = \sum_{j=1}^{q} Q_j$ (4.2.81)

$$\|\Delta\|^2 = \sum_{j=1}^{q} p_j^2/f_j$$

The invariant MInimum Norm Quadratic Unbiased Estimate (MINQUE) of $\rho=p^T\sigma$ is defined as a quadratic form L^TML such that the symmetrical matrix M satisfies the

invariance condition (4.2.50) and that the matrix norm $\|G\|$ attains minimum value. As $\|\Delta\|$ in (4.2.80) is independent of M, the problem of finding an invariant MINQUE of ρ is to find a symmetrical M such that:

$$\left.\begin{aligned} MA &= 0 \\ \mathrm{tr}(MQ_j) &= p_j \\ \mathrm{tr}(M\overline{Q}M\overline{Q}) &= \text{minimum} \end{aligned}\right\} \qquad (4.2.82)$$

This is the same conditional minimization problem as in (4.2.53), except that Q_0 in (4.2.53) is now replaced by $\overline{Q}$. The solution, i.e. MINQUE, is given by (4.2.58) for $\rho=p^T\sigma$ and by (4.2.62) for σ with Q_0 replaced by $\overline{Q}$. Detailed formulas are omitted here, as we want to derive the MINQUE using a priori information of the variance components, i.e. the approximate values of variance components $\sigma_1^0, \sigma_2^0, \cdots, \sigma_q^0$.

First we rewrite the decomposition in (4.2.75) as:

$$\epsilon = \sum_{j=1}^{q} U_j \xi_j = \sum_{j=1}^{q} V_j \eta_j = V\eta \qquad (4.2.83)$$

where:

$$\left.\begin{aligned} V_j &\equiv (\sigma_j{}^0)^{\frac{1}{2}} U_j \\ \eta_j &\equiv (1/\sigma_j{}^0)^{\frac{1}{2}} \xi_j \\ V &\equiv (V_1, V_2, \cdots, V_q) \\ \eta &\equiv (\eta_1{}^T, \eta_2{}^T, \cdots, \eta_q{}^T) \\ V_j V_j{}^T &= \sigma_j Q_j \\ VV^T &= Q_0 \\ E(\eta\eta^T) &= I \end{aligned}\right\} \qquad (4.2.84)$$

Letting Ω define the following diagonal matrix,

$$\underset{f\times f}{\Omega} = \begin{bmatrix} p_1\sigma_1{}^0/f_1 I_1 & & & \\ & p_2\sigma_2{}^0/f_2 I_2 & & \\ & & \cdots & \\ & & & p_q\sigma_q{}^0/f_q I_q \end{bmatrix} \qquad (4.2.85)$$

one can show that $\eta^T\Omega\eta$ is an unbiased estimate of $\rho\equiv p^T\sigma$ and that the difference between L^TML and $\eta^T\Omega\eta$ is:

$$L^TML - \eta^T\Omega\eta \equiv \epsilon^TM\epsilon - \eta^T\Omega\eta = \eta^T F \eta \qquad (4.2.86)$$

where:

$$F \equiv V^TMV - \Omega\,. \qquad (4.2.87)$$

Now our MINQUE of $\rho\equiv p^T\sigma$ is given by L^TML such that M satisfies (4.2.50) and that the Euclidean norm of matrix F is minimized. As the Euclidean norm of F can be

found as:

$$\|F\|^2 \equiv \mathrm{tr}(FF^T) = \mathrm{tr}(MQ_0MQ_0) - \|\Omega\|^2 \tag{4.2.88}$$

the searching of a MINQUE of $p^T\sigma$ is now equivalent to solving the following conditional minimization problem:

$$\left.\begin{aligned} MA &= 0 \\ \mathrm{tr}(MQ_j) &= p_j \\ \mathrm{tr}(MQ_0MQ_0) &= \text{minimum} \end{aligned}\right\} \tag{4.2.89}$$

which is exactly the same problem as defined as (4.2.53). Consequently the solution to our MINQUE follows from (4.2.58) and (4.2.62):

$$\hat{\rho} = L^T M L \tag{4.2.90}$$

$$M = \sum_{i=1}^{q} \lambda_i R_0 Q_i R_0 \tag{4.2.91}$$

$$R_0 \equiv Q_0^{-1} - Q_0^{-1}A(A^TQ_0^{-1}A)^{-1}A^TQ_0^{-1} \tag{4.2.92}$$

λ_i's $(1 \le i \le q)$ are solved from the following equation system:

$$\underset{qxq}{S}\ \underset{qx1}{\lambda} = \underset{qx1}{p} \qquad \text{or:} \qquad \lambda = S^{-1}p \tag{4.2.93}$$

with:

$$\underset{qxq}{S} \equiv \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1q} \\ s_{21} & s_{22} & \cdots & s_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ s_{q1} & s_{q2} & \cdots & s_{qq} \end{bmatrix}; \quad \lambda \equiv \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_q \end{bmatrix}; \quad p \equiv \begin{bmatrix} p_1 \\ p_2 \\ \cdots \\ p_q \end{bmatrix}$$

$$s_{ij} = \mathrm{tr}(R_0Q_iR_0Q_j) = s_{ji} \qquad (1 \le i,j \le q) \tag{4.2.94}$$

and

$$\hat{\sigma} \equiv \begin{bmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \cdots \\ \hat{\sigma}_q \end{bmatrix} = S^{-1}u \tag{4.2.95}$$

with:

$$u \equiv \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_q \end{bmatrix}$$

$$u_i \equiv L^T R_0 Q_i R_0 L = \hat{\epsilon}^T Q_0^{-1} Q_i Q_0^{-1} \hat{\epsilon} \tag{4.2.96}$$

$$\hat{\epsilon} = \left[I - A(A^TQ_0^{-1}A)^{-1}A^TQ_0^{-1}\right] L \tag{4.2.97}$$

Comparing MINQUE with BQUE described in §4.2.4 and §4.2.5, we can see that

(a) MINQUE does not require that the residual vector ϵ is normally distributed as BQUE does;

(b) MINQUE does not require a priori information of the variance components as BQUE does, though the approximate variance components can be used in calculating MINQUE;

(c) using the same set of approximate variance components, MINQUE and BQUE are identical.

Due to remark (c) above, some geodetic literatures do not distinguish MINQUE and BQUE in practical applications.

§4.3 Detection of Gross and Systematic Errors

In §1.1.1, we have classified measurement errors into random errors, systematic errors and gross errors. Optimal estimates outlined in §1.2.6 and various adjustment methods described in previous chapters are all designed under the assumption that our measurements contain only random or stochastic errors. In this section, we will instead discuss systematic errors and gross errors.

As discussed before, systematic errors can be partly reduced or totally avoided by e.g. carefully calibrating instruments to be used, designing 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 as some kind of biases so that the measurement error ϵ does not attain zero expectation. Naturally one may try to model the systematic errors in the adjustment by introducing extra parameters, whenever such modelling is possible. However, one should make sure that such modelling is correct and the estimated systematic effects are statistically significant. Another way to detect systematic effects, even though maybe small, in the measurements is to do statistical tests.

As for gross errors, in principle they should be avoided through e.g. carefulness of survey personnel, rigorous and efficient checking routines, etc. All these security measures against gross errors should have been considered at the planning stage before the field work as well as during field survey. Therefore, in the following discussions we concentrate on how to detect, after the field work, still remaining gross errors by analyzing the measurement results. Mathematically, gross errors may be treated as random errors which are extra–ordinarily larger than (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 estimated residuals of the measurements.

From probability and statistical theory, we know that if measurement errors are stochastic errors of some distribution, quantities derived from these measurements will also be stochastic with certain probability distributions. This leads us to the following procedure for detecting non–random errors :

(a) We start from the assumption that the measurement errors are random errors with certain specified distributions. This is actually the basis of our zero hypothesis (H_0). The alternative hypothesis (H_1) is simply that the measurement errors are not random errors of assumed distribution;

(b) We then construct some quantities which would have well–defined statistical distribution when the above assumption holds;

(c) Now we test the computed values of the above statistics (sample values) against the theoretical critical values for certain risk level. If the hypothesis test is passed, we then accept the assumption in (a). Otherwise, we may suspect there exist systematic or gross errors in the measurement results.

The above statistical approach based on hypothesis test can be carried out either directly on the measurement errors as shown in §4.3.1, or after a preliminary least squares adjustment as described in §4.3.2. §4.3.3 deals with simultaneous estimation of variance components and detection of gross errors. Other approaches without directly using hypothesis test are briefly outlined in §4.3.4.

§4.3.1 Randomness Test

Before we discuss how to find out whether measurement errors are random or not, we should know what properties random errors may have. If measurement errors ϵ_i (i=1, 2, 3, $\cdots$, n) are (or approximately) normally distributed, one can deduct that these random errors should have the following characters:

(a) The arithmetic mean of ϵ_i should approach zero when the number (n) of observations approaches infinity, i.e.:

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \epsilon_i}{n} = 0$$

(b) Positive errors and negative errors with same absolute magnitude should have equal chance to occur;

(c) Errors of smaller absolute magnitudes should have larger probability to occur than errors of larger absolute magnitudes;

(d) Under specified measurement condition, the absolute magnitude of errors should be within some limit.

Considering the above properties of random errors, one can construct various statistics to test whether the measurement errors are random or not. Below, five different tests will be outlined.

(i) Testing the number of positive versus negative errors

Let s_+ denote the number of positive errors. If ϵ_i is random, s_+ should be of binomial distribution with expectation $\mu=\frac{1}{2}n$ and variance $\sigma^2=n/4$. When n is very large, s_+ will approach a normal distribution. In other words, the following statistics is approximately of standard normal distribution:

$$\frac{s_+ - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \sim N(0,1) \qquad (4.3.1a)$$

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:
$$P\left\{|s_+ - s_-| < \sqrt{n}\, c_{\frac{1}{2}\alpha}\right\} = 1-\alpha \qquad (4.3.1b)$$

where $s_- \equiv n - s_+$ denotes the number of negative errors 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 zero hypothesis (H_0) and alternative hypothesis (H_1):

$$H_0: \ P(\epsilon_i>0) = \tfrac{1}{2}; \qquad H_1: \ P(\epsilon_i>0) \neq \tfrac{1}{2}. \tag{4.3.1c}$$

For a chosen risk level α, if $|s_+ - s_-| < \sqrt{n}\, c_{\frac{1}{2}\alpha}$ we conclude that there is no significant difference between s_+ and s_-. Otherwise, we accept H_1, i.e. there are systematic effects in ϵ_i.

(ii) Testing the order of positive versus negative errors

Sometimes, ϵ_i contains systematic effects which 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. Therefore, we need to test the order of the positive versus negative errors.

Let s_1 denote the number of adjacent error pairs with same sign, as defined by:

$$s_1 = x_1 + x_2 + \cdots + x_{n-1} \tag{4.3.2a}$$

where:
$$x_i = \begin{cases} 1 & \text{if } \epsilon_i \text{ and } \epsilon_{i+1} \text{ have the same sign} \\ 0 & \text{otherwise} \end{cases} \tag{4.3.2b}$$

If ϵ_i are random errors, x_i should have the same probability ($\frac{1}{2}$) to be 1 or 0. In this case, s_1 is of binomial distribution which approaches the normal distribution when n is very large, i.e.:

$$\frac{s_1 - \frac{1}{2}(n-1)}{\frac{1}{2}\sqrt{n-1}} \equiv \frac{s_1 - s_0}{\sqrt{n-1}} \sim N(0,1) \tag{4.3.2c}$$

where $s_0 = (n-1) - s_1$ denotes the number of adjacent error pairs with opposite sign. Similar to (4.3.1c), we define our hypothesis as follows:

$$H_0: \ P(x_i=1) = \tfrac{1}{2}; \qquad H_1: \ P(x_i=1) \neq \tfrac{1}{2}. \tag{4.3.2d}$$

For a chosen risk level α, if $|s_1 - s_0| < \sqrt{n-1}\, c_{\frac{1}{2}\alpha}$, we conclude that the order of positive and negative errors is random. Otherwise, we accept H_1, i.e. there are systematic effects in ϵ_i.

(iii) 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, as defined by:

$$s^2 \equiv \lambda_1 \epsilon_1^2 + \lambda_2 \epsilon_2^2 + \cdots + \lambda_n \epsilon_n^2 \qquad (4.3.3a)$$

where:
$$\lambda_i = \begin{cases} +1 & \text{if } \epsilon_i > 0 \\ -1 & \text{if } \epsilon_i < 0 \end{cases} \qquad (4.3.3b)$$

Note that the notation s^2 above does not denote the square of quantity s. If ϵ_i are random errors, s^2 should be close to zero and $\lambda_i=+1$ and $\lambda_i=-1$ have equal probability, i.e. $P(\lambda_i=+1) = P(\lambda_i=-1) = \frac{1}{2}$. The expectation and variance of λ_i can be found:

$$\left.\begin{aligned} E(\lambda_i) &= (+1)\,\tfrac{1}{2} + (-1)\,\tfrac{1}{2} = 0 \\ E(\lambda_i^2) &= (+1)^2\,\tfrac{1}{2} + (-1)^2\,\tfrac{1}{2} = 1 \end{aligned}\right\} \qquad (4.3.3c)$$

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 uncorrelated with normal distribution $N(0,\sigma^2)$, the expectation and variance of $\lambda_i\,\epsilon_i^2$ can be obtained :

$$\left.\begin{aligned} E[\lambda_i \epsilon_i^2] &= E(\lambda_i)\; E(\epsilon_i^2) = 0 \\ E[(\lambda_i \epsilon_i^2)^2] &= E(\lambda_i^2)\; E(\epsilon_i^4) = E(\epsilon_i^4) = 3\,\sigma^4 \end{aligned}\right\} \qquad (4.3.3d)$$

Finally the expectation and variance of s^2 can be obtained:

$$\left.\begin{aligned} E(s^2) &= \sum_{i=1}^{n} E(\lambda_i \epsilon_i^2) = 0 \\ E(s^4) &= E\{[\sum_{i=1}^{n} \lambda_i^2 \epsilon_i^4]^2\} = 3n\sigma^4 \end{aligned}\right\} \qquad (4.3.3e)$$

When n is very large, s^2 will approach normal distribution $N(0, 3n\sigma^4)$. Thus for risk level α, we have:

$$P\left\{|s^2/(\sqrt{3n}\,\sigma^2)| < c_{\frac{1}{2}\alpha}\right\} = 1-\alpha \qquad (4.3.3f)$$

where $c_{\frac{1}{2}\alpha}$ is the critical value of N(0,1) at risk level α. If $|s^2| < \sqrt{3n}\sigma^2\, c_{\frac{1}{2}\alpha}$, we accept that s^2 is not significantly different from zero. Otherwise, it indicates the existence of systematic effects in ϵ_i.

(iv) Testing the sum of errors

Let s denote the sum of the n errors:

$$s \equiv \epsilon_1 + \epsilon_2 + \cdots + \epsilon_n \qquad (4.3.4a)$$

If $\epsilon_i \sim N(0, \sigma^2)$ are uncorrelated with each other, it can be easily shown that $s \sim N(0, n\sigma^2)$. For risk level α, we then have:

$$P\left\{ \left| \frac{s}{\sqrt{n}\,\sigma} \right| < c_{\frac{1}{2}\alpha} \right\} = 1 - \alpha \qquad (4.3.4b)$$

where $c_{\frac{1}{2}\alpha}$ is the critical value of N(0,1) at risk level α. If $|s| < \sqrt{n}\,\sigma\, c_{\frac{1}{2}\alpha}$, we accept $\epsilon_i \sim N(0, \sigma^2)$. Otherwise, we may suspect systematic effects in ϵ_i.

(v) Testing the maximum absolute value of errors

Let ϵ_m denote the error of maximum absolute value. If $\epsilon_i \sim N(0, \sigma^2)$, we have for risk level α:

$$P\left\{ |s/\sigma| < c_{\frac{1}{2}\alpha} \right\} = 1 - a \qquad (4.3.5b)$$

Thus, if $|\epsilon_m| < \sigma\, c_{\frac{1}{2}\alpha}$, we accept $\epsilon_i \sim N(0, \sigma^2)$. Otherwise, we may suspect systematic effects in ϵ_i.

Example 4.3.1

In a geodetic triangulation network, 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

From w_i listed above, we can estimate the standard error of misclosures:

$$\sigma^2 = \frac{1}{30} \sum_{i=1}^{30} w_i^2 = 25.86/30 \qquad \Rightarrow \qquad \sigma = 0.93"$$

At risk level $\alpha = 4.55\%$ (or confidence level $1-\alpha = 95.45\%$), $c_{\frac{1}{2}\alpha} = 2$. The five different tests described before are performed below:

(i) *Testing the number of positive versus negative errors*

$|s_+ - s_-| = |14 - 16| = 2 < c_{\frac{1}{2}\alpha}\sqrt{n} \approx 11$ $\Longrightarrow$ test is passed !

(ii) *Testing the order of positive versus negative errors*

$|s_1 - s_0| = |18-11| = 7 < \sqrt{n-1}\, c_{\frac{1}{2}\alpha} \approx 11$ $\Longrightarrow$ test is passed !

(iii) *Testing the sum of positive versus negative errors squared*

$|s^2| = |3.40| = 3.4 < \sqrt{3n}\sigma^2 c_{\frac{1}{2}\alpha} \approx 16.41$ $\Longrightarrow$ test is passed !

(iv) *Testing the sum of errors*

$|s| = |2.6| = 2.6 < \sqrt{n}\,\sigma\, c_{\frac{1}{2}\alpha} \approx 10.2$ $\Longrightarrow$ test is passed !

(v) *Testing the maximum absolute value of errors*

$|\epsilon_m| = |-2.0| = 2.0 > \sigma\, c_{\frac{1}{2}\alpha} \approx 1.86$ $\Longrightarrow$ test is not passed !

$\Longrightarrow$ misclosure w_{11} might contain gross error.

AMBOS MUY UTILES Y USADOS COMO EERRs

AHORA HAY QUE METERSE NO CON SERIES SINO REDES

§4.3.2 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). The starting point of data snooping, also called the B–method, is that for a set of observations with assumed distribution, the statistical distribution of the derived quantities can be derived and tested against the assumed distribution. Furthermore, data snooping assumes that only one observation contains gross errors which can be located by statistically testing the least squares residuals of all observations against the estimated standard error of the residuals.

The zero hypothesis (H_0) in data snooping is based on the following functional and statistical models free of gross errors:

$$H_0: \quad \left\{ \begin{array}{l} \underset{n\,x\,1}{L} - \underset{n\,x\,1}{\epsilon} = \underset{n\,x\,m}{A}\ \underset{m\,x\,1}{X} \\ E(\epsilon) = 0 \\ E(\epsilon\epsilon^T) = \sigma_0^2\, Q = \sigma_0^2\, P^{-1} \end{array} \right\} \qquad (4.3.6)$$

where P and Q denote the weight matrix and the cofactor matrix of L and ϵ, respectively. In this case, the least squares estimate of ϵ and its cofactor $Q_{\hat{\epsilon}\hat{\epsilon}}$ can be easily obtained [Cf. (3.1.15), (3.1.31)]:

$$\left.\begin{aligned} \hat{\epsilon} &= (I - A\,N^{-1}A^{T}P)\,L, \qquad (N = A^{T}PA) \\ Q_{\hat{\epsilon}\hat{\epsilon}} &= P^{-1} - P^{-1}AN^{-1}A^{T}P^{-1} \\ \Omega &= \hat{\epsilon}^{T}P\hat{\epsilon} \end{aligned}\right\} \qquad (4.3.7)$$

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 (4.3.6) while the functional model is complemented by the gross error Δ_i:

$$H_1: \left\{\begin{aligned} \underset{n\times 1}{L} - \underset{n\times 1}{\epsilon} &= \underset{n\times m}{A}\ \underset{m\times 1}{X} + \underset{n\times 1}{e_i}\ \underset{1\times 1}{\Delta_i} \\ E(\epsilon) &= 0 \\ E(\epsilon\epsilon^{T}) &= \sigma_0^{2}\,Q = \sigma_0^{2}\,P^{-1} \end{aligned}\right\} \qquad (4.3.8)$$

where e_i denotes a column vector with zero elements except the i–th element which equals 1: $e_i \equiv (0, 0, \cdots, 0, 1, 0, \cdots, 0)^{T}$. Let $\hat{\epsilon}_1$ denote the least squares estimate of residual vector ϵ from (4.3.8). It can be proved that $\hat{\epsilon}_1^{T}P\hat{\epsilon}_1$ is related to Ω in (4.3.6) by:

$$\Omega_1 \equiv \hat{\epsilon}_1^{T}P\hat{\epsilon}_1 = \hat{\epsilon}^{T}P\hat{\epsilon} - \Delta\Omega = \Omega - \Delta\Omega \qquad (4.3.9a)$$

where:
$$\Delta\Omega = \hat{\epsilon}^{T}\,Pe_i(e_i^{T}PQ_{\hat{\epsilon}\hat{\epsilon}}Pe_i)^{-1}e_i^{T}P\,\hat{\epsilon} \qquad (4.3.9b)$$

For a diagonal weigh matrix like $P = \mathrm{diag}(p_1, p_2, \cdots, p_n)$, (4.3.9b) will reduce to:

$$\Delta\Omega = \frac{\hat{\epsilon}_i^{2}}{q_{ii}} \qquad (4.3.9c)$$

where q_{ii} denotes the corresponding element of $\hat{\epsilon}_i$ in the cofactor matrix $Q_{\hat{\epsilon}\hat{\epsilon}}$. It can be proved that

$$\Omega/\sigma_0^{2} \sim \chi^{2}(n-m) \qquad (4.3.10)$$

and
$$R/\sigma_0^{2} \sim \chi^{2}(1) \qquad \text{or:} \qquad w_i \equiv \sqrt{R}/\sigma_0 \equiv \frac{\hat{\epsilon}_i}{\sigma_{\hat{\epsilon}_i}} \sim N(0,1) \qquad (4.3.11)$$

where $\hat{\epsilon}_i$, $\sigma_{\hat{\epsilon}_i}$ denote least squares estimate of ϵ_i and its standard error.

When the a priori variance factor σ_0^{2} is given, the above two statistics can be used to test whether there exist gross error in the observation l_i.

Overall test (multi–dimensional)

If $\Omega/\sigma_0^2 < \chi_{n-m,\ 1-\alpha}$, accept H_0, i.e. there is no gross error. Otherwise, reject H_0. Here $\chi_{n-m,\ 1-\alpha}$ is the critical value of $\chi^2(n-m)$ at risk level α.

Test of each individual observation (one dimensional)

If $|w_i| < c_{\frac{1}{2}\alpha}$, we accept H_0, i.e. there exists no significant gross error in l_i. Otherwise, reject H_0. This testing procedure can be performed subsequently for each observation l_i (i=1, 2, $\cdots$, n) and thus called a data snooping strategy.

It should be pointed out that from the adjustment results ($\hat{\epsilon}$, Ω, Ω_1, $\Delta\Omega$, etc.), one can construct other statistics than given in (4.3.10) and (4.1.11). See e.g. Pope (1975), Heck (1980), Ashkenazi and Crane (1982).

There are some limitations with data snooping technique for gross error detection. Firstly, it assumes that there is only one gross error Δ_i. If there are more than one gross errors in the observations, the method cannot guarantee the detection. Secondly, tests using (4.3.10) and (4.3.11) assume that the a priori variance factor σ_0^2 and the correct weight matrix are known, which is often hardly true.

Thirdly, test with (4.3.11) is based implicitly on the anticipation that the observation with large gross error will also have large residual after the least squares adjustment. However, many studies have shown that least squares adjustment has the tendency to hide large residuals (e.g. due to gross error) and distribute their effects to other observations. Consequently, observations with large (gross) errors may not necessarily obtain large residuals after least squares adjustment. In this case, detection and location of gross errors becomes very difficult.

4.3.3 Simultaneous Study of Gross Errors and Variance Components

As mentioned before, the success of data snooping requires that the correct weight matrix of the observations is known. This is not always true, especially when several different types of observations are involved. On the other hand, successful a posteriori estimation of variance components requires that the observations are free of gross errors and systematic errors. A way out of this dilemma may be to detect gross errors and estimate variance components simultaneously as attempted in e.g. Persson (1982) and Li (1983). Below, we only outline the strategy proposed in Li (1983).

Assume that the observations can be divided into q groups:

$$\underset{n\times 1}{L} = \begin{bmatrix} L_1 \\ L_2 \\ \cdots \\ L_q \end{bmatrix}, \qquad \underset{n\times 1}{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_q \end{bmatrix} \tag{4.3.12}$$

Within each group, all observations are assumed to have same accuracy and weights, while different groups have different accuracy:

$$E(\epsilon\epsilon^T) = \sigma_0^2 \begin{bmatrix} P_1^{-1} & & & \\ & P_2^{-1} & & \\ & & \cdots & \\ & & & P_q^{-1} \end{bmatrix} \tag{4.3.13}$$

As an initial step, we carry out a preliminary least squares adjustment. Let $\hat{\epsilon}_i$ denote the estimated residuals for the i–th group. Förstner (1979) showed that a variance estimate for the i–th group can be obtained from:

$$\hat{\sigma}_i^2 = \frac{\hat{\epsilon}_i^T \hat{\epsilon}_i}{r_i} \qquad (i = 1, 2, \cdots, q) \tag{4.3.14}$$

where $r_i \equiv tr(Q_{\hat{\epsilon}_i\hat{\epsilon}_i} P_i)$ and $Q_{\hat{\epsilon}_i\hat{\epsilon}_i}$, P_i denote the submatrices in the cofactor matrix and weight matrix of L for the i–th group of observations. Based on $\hat{\sigma}_i^2$ and the overall estimate of the variance factor:

$$\hat{\sigma}_0^2 = \frac{\hat{\epsilon}^T P \hat{\epsilon}}{n-m} \tag{4.3.15}$$

where m = number of parameters and $\hat{\epsilon}$ denotes the least squares residuals, a new set of weights for each group can be obtained:

$$\hat{p}_i = \hat{\sigma}_i^2 / \hat{\sigma}_0^2 \tag{4.3.16}$$

Meanwhile, one can estimate the variance of the j–th observation in the i–th group from the corresponding residual:

$$\hat{\sigma}_{ij}^2 = \hat{\epsilon}_{ij}^2 / r_{ij} \qquad (j = 1, 2, 3, \cdots,) \tag{4.3.17}$$

where r_{ij} is the corresponding element of ϵ_{ij} in the cofactor matrix of $\hat{\epsilon}$. If there is no gross error in l_{ij}, the two variances in (4.3.14) and (4.17) should have no significant difference. To test this, we construct the following statistics:

$$z_{ij} = \hat{\sigma}_{ij}^2 / \hat{\sigma}_i^2 \tag{4.3.18}$$

which approximately has distribution $F(1, r_i)$. Let $F_{1,r_i,\alpha}$ denote the critical value of $F(1,r_i)$ for risk level α. If $z_{ij} \geq F_{1,r_i,\alpha}$, then l_{ij} possibly contains gross error. We can now assign new weights to all observations in the i–th group:

$$\hat{p}_{ij} = \begin{cases} \hat{p}_i & \text{if } z_{ij} < F_{1,r_i,\alpha} \\ \hat{\sigma}_0^2/\hat{\sigma}_{ij}^2 & \text{if } z_{ij} \geq F_{1,r_i,\alpha} \end{cases} \tag{4.3.19}$$

With the new set of weights, a new adjustment can be performed to estimate new variances and weights. The iteration continues until it converges. Observations with very small weights will be observations which possibly contain gross errors.

4.3.4 Other Methods

There are other methods which do not explicitly use statistical tests to detect and locate observations with gross errors. Below, we outline the main ideas of three methods.

Least Sum Method

To overcome the drawback of least squares method in smoothing out residuals, one may use an alternative adjustment method, called *least sum adjustment*, which tolerates larger residuals, thus facilitating the detection and location of large gross errors in the measurements. The least sum method uses the same observation equations as in (4.3.6) but obtains the optimal estimates for X and ϵ by minimizing the sum of the weighted absolute values of residuals:

$$\left.\begin{aligned} L - \epsilon &= AX \\ \sum_{i=1}^{n} p_i \, |\epsilon_i| &= \text{minimum} \end{aligned}\right\} \tag{4.3.20}$$

where p_i denotes the weight of l_i. (4.3.22) represents a linear programming problem which can be solved numerically using e.g. simplex algorithm. Details are omitted here. When residuals are estimated, one can easily identify observations possibly with gross errors as they are expected to have relatively large residuals.

Robust Estimators

Robust estimators are estimators which are insensitive to limited variations in the distribution function of the observations, e.g. due to presence of gross errors. They are based on other types of object functions than least squares principle or least sum principle. The robust estimator by Huber is obtained by minimizing the following object function:

$$f(\epsilon) = \sum_{i=1}^{n} f(\epsilon_i) \tag{4.3.21}$$

where:
$$f(\epsilon_i) = \begin{cases} \epsilon_i^2 & \text{if } |\epsilon_i| \geq 2\sigma \\ 2\sigma\,(2|\epsilon_i| - 2\sigma) & \text{if } |\epsilon_i| > 2\sigma \end{cases} \tag{4.3.22}$$

Another type of robust estimator has the following object function:

$$f(\epsilon) = \sum_{i=1}^{n} |\epsilon_i|^k \qquad (4.3.23)$$

where k is chosen as a number between 1.2 and 1.5.

The Danish Method

The Danish method is an extension of the robust estimators. It is an iterative approach which starts with a traditional least squares adjustment where all observations are assumed to have equal weight 1. After the initial adjustment, new weights are assigned to each observation according to the size of residual from the initial adjustment:

$$p_i = \begin{cases} 1 & \text{if } |\hat{\epsilon}_i| \leq 2\sigma \\ c_1 \exp(-c_2\, \epsilon_i^2) & \text{if } |\hat{\epsilon}_i| > 2\sigma \end{cases} \qquad (4.3.24)$$

where c_1 and c_2 are two positive numbers chosen empirically.

Using (4.3.24), a new set of weights are obtained from which a new least squares adjustment can be is performed, leading to a new set of residuals $\hat{\epsilon}$ and subsequently a new set of weights according to (4.3.16). The iteration continues until convergence is achieved when observations affected by gross errors have weights zero. The magnitude of their estimated residuals is a measure for the magnitude of gross errors.

§4.4 Filtering and Prediction

The orginal meaning of word "*filtering*" is to extract the desired information (signals) from the received electromagnetic signals which are contaminated by undesired noises. In a general sense, filtering is to obtain optimal estimates of the wanted parameters by filtering out the effects of various errors (i.e. noises). At a first look, this is quite similar to geodetic adjustment. However, there are essential differences: in filtering problems the parameters to be estimated are also stochastic quantities just like the measurement errors.

In many engineering fields (e.g. electronic communication and system control), the wanted signals or parameters are time–dependent functions. This is especially true, when the wanted parameters are associated with a dynamical system. This makes another difference compared with classical geodetic adjustment: filtering often deals with a continuous problem with parameters and measurements being continuous functions of e.g. time t. Today, geodetic applications may also encounter continuous problems, such as kinematic navigation using GPS or inertial survey system. Therefore, it is worthwhile to broaden our theories and methods of geodetic adjustment.

In practical applications, most continuous problems can be approximately linearized so that they can be formulated and solved in a discrete way. Nevertheless, understanding well the continuous filtering problem would be very helpful for correctly understanding and solving the discrete counterpart.

If our task is to derive an estimate to a time function $x(t+\tau)$ (at time epoch $t+\tau$) from measurements $l(t)$, three specific problems can be distinguished:

(a) when $\tau < 0$, the problem is called *smoothing* or *interpolation*;
(b) when $\tau = 0$, the problem is called *filtering*;
(c) when $\tau > 0$, the problem is called *prediction* or *extrapolation*.

It should be pointed out that the variables involved are not necessarily time t but can be e.g. xyz–coordinates of points, or other types of variables.

When discussing prediction and filtering problems, one often use the concept of stochastic functions or stochastic processes. Therefore, we first introduce some basic concepts of stochastic processes in §4.4.1. Then we will discuss Wiener–Hopf prediction for both continuous time series as well as discrete data. The mathematical models of linear dynamical systems and the well–known Kalman filtering procedure in discrete case will be described in §4.4.4 and §4.4.5.

§4.4.1 Basic Concepts of Stochastic processes

If a random variable x also depends on one or more non–random variables, x is then called a stochastic function or process. A stochastic function depending on time t is often written as x(t). If t varies continuously, x(t) is called a continuous stochastic process. If t takes only discrete values $t_1, t_2, t_3, \cdots$, x is called a discrete stochastic process, denoted as x_k $(k = 1, 2, 3, \cdots)$.

The expectation and auto–covariance function of a continuous stochastic process in the time domain, $x(t)$, are defined as follows:

$$\left.\begin{aligned} E\{\, x(t)\,\} &= 0 \\ E\{\, [x(t_1)-\mu(t_1)]\,[x(t_2)-\mu(t_2)]\,\} &= C_x(t_1,t_2) \end{aligned}\right\} \qquad (4.4.1a)$$

In general, $\mu(t)$ is a function of t and $C(t_1, t_2)$ is a function of both t_1 and t_2. If $\mu(t)$ is equal to a constant independent of t and $C(t_1,t_2)$ is depends only on the difference t_2-t_1, $x(t)$ is then called a *weakly stationary stochastic process.* A *strictly stationary stochastic process* $x(t)$ is obtained if the n–th order statistics of $x(t)$:

$$E\left\{ \prod_{k=1}^{n}\left[x(t_k)-E[x(t_k)]\right] \right\} = E\left\{\left[x(t_1)-E[x(t_1)]\right]\left[x(t_2)-E[x(t_2)]\right]\cdots\left[x(t_n)-E[x(t_n)]\right]\right\}$$

is invariant with respect to a shift in time origin for any arbitrary $n < +\infty$.

For a weakly stationary stochastic process $x(t)$, we can write its auto–covariance function in a simple way:

$$E\{[x(t)-\mu(t)]\,[x(t+\tau)-\mu(t+\tau)]\} = C_{xx}(\tau) \qquad (4.4.1b)$$

as $t_2-t_1 \equiv \tau$ for $t_1=t$ and $t_2=t+\tau$.

If $y(t)$ is another stochastic process with expectation $E\{y(t)\} \equiv \nu(t)$, the cross–covariance function between $x(t)$ and $y(t)$ is defined as:

$$E\{[x(t_1)-\mu(t_1)]\,[y(t_2)-\nu(t_2)]\} = C_{xy}(t_1, t_2) \qquad (4.4.1c)$$

Similarly, one can define expectation and covariances for a discrete time series x_k (k, j $= 0, 1, 2, 3, \cdots$):

$$\left.\begin{aligned} E\{\, x_k\,\} &= \mu_k \\ E\{\, [x_k-\mu_k]\,[x_j-\mu_j]\} &= C_{xx}(k, j) \end{aligned}\right\} \qquad (4.4.2a)$$

If $C_{xx}(k,j)$ is dependent only on $j–k$, x_k is called a *discrete weakly stationary stochastic process* and its covariance can be written alternatively as:

$$E\{\, [x_k - \mu_k]\,[x_{k+m} - \mu_{k+m}]\,\} \equiv C_{xx}(m) \qquad (4.4.2b)$$

where $m = 0, 1, 2, 3, \cdots$.

Example 4.4.1

Assume that a and b are two independent random variables of normal distribution N(0,1). A stochastic process is defined as:

$$x(t) = a\cos\omega t + b\sin\omega t$$

where ω is a constant and t is the time variable. The expectation and auto–covariance function of $x(t)$ are:

$$\mu(t) = E\{ x(t) \} = E(a)\cos\omega t + E(b)\sin\omega t = 0$$

$$C(t_1, t_2) = E\{ [x(t_1)-\mu(t_1)]\,[x(t_2)-\mu(t_2)] \} = E\{ x(t_1)\,x(t_2) \}$$

$$= E(a^2)\cos\omega t_1 \cos\omega t_2 - E(a)\,E(b)\sin\omega(t_2-t_1) + E(b^2)\sin\omega t_1 \sin\omega t_2$$

$$= \cos\omega(t_2-t_1)\,.$$

Therefore, $x(t)$ is a weakly stationary stochastic process.

§4.4.2 Continuous Wiener–Hopf Prediction

Wiener–Hopf prediction aims at predict one or more stochastic quantities from a set of stochastic observations, using the criteria that the variance of the prediction error attains minimum. We first discuss prediction in continuous time series and then in next section, present the discrete Wiener–Hopf prediction. As examples, we will once more look at the classical condition adjustment described in §2.1 as well as the adjustment model (3.4.1), using the principle of discrete Wiener–Hopf prediction.

Assume that the measurement $l(t)$ and the wanted signal $x(t)$ are both weakly stationary stochastic time series. Their expectations, auto–covariance functions and the cross–covariance function between them are as follows:

$$E\{x(t)\} = 0\,; \qquad C_{xx}(\tau) = E\{x(t)\,x(t+\tau)\} \tag{4.4.3a}$$

$$E\{l(t)\} = 0\,; \qquad C_{ll}(\tau) = E\{l(t)\,l(t+\tau)\} \tag{4.4.3b}$$

$$C_{xl}(\tau) = E\{x(t)\,l(t+\tau)\} \tag{4.4.3c}$$

Now we want to derive an estimate of $x(t)$ from $l(t)$:

$$\hat{x}(t) = \int_{-\infty}^{+\infty} h(t-\tau)\,l(\tau)\,d\tau \tag{4.4.4}$$

where $h(t)$ denotes an unknown weight function. Wiener–Hopf prediction chooses the optimal weight function by minimizing the variance of the prediction error $\epsilon(t)$:

$$\Delta(t) = \hat{x}(t) - x(t) \tag{4.4.5a}$$

$$\sigma^2(t) = E\{\Delta(t)\,\Delta(t)\} = \text{minimum} \tag{4.4.5b}$$

The minimization condition (4.4.5b) leads to the so called Wiener–Hopf integral:

$$C_{xl}(\lambda) = \int_{-\infty}^{+\infty} h(\tau)\, C_{ll}(\lambda-\tau)\, d\tau \equiv h(\lambda) * C_{ll}(\lambda) \tag{4.4.6}$$

(4.4.6) is an equation in the time domain and can be solved in the frequency domain by Fourier transformation. Let $S_{xl}(\omega)$, $H(\omega)$ and $S_{ll}(\omega)$ denote the Fourier transform of $C_{xl}(\lambda)$, $h(\lambda)$ and $C_{ll}(\lambda)$, respectively:

$$S_{xl}(\omega) = \int_{-\infty}^{+\infty} C_{xl}(\lambda)\, e^{-i\omega\lambda}\, d\lambda \tag{4.4.7a}$$

$$H(\omega) = \int_{-\infty}^{+\infty} h(\omega)\, e^{-i\omega\lambda}\, d\lambda \tag{4.4.7b}$$

$$S_{ll}(\omega) = \int_{-\infty}^{+\infty} C_{ll}(\lambda)\, e^{-i\omega\lambda}\, d\lambda \tag{4.4.7c}$$

Applying convolution theorem of Fourier transformation on (4.4.6), we obtain :

$$S_{xl}(\omega) = H(\omega) \cdot S_{ll}(\omega) \tag{4.4.8a}$$

or: $$H(\omega) = S_{xl}(\omega)/S_{ll}(\omega) \tag{4.4.8b}$$

Finally the optimal weighting function $h(\lambda)$ in time domain is obtained by the inverse Fourier transform of $H(\omega)$:

$$h(\lambda) = \int_{-\infty}^{+\infty} H(\omega)\, e^{i\omega\lambda}\, d\omega \tag{4.4.9}$$

Example 4.4.2 (*Wiener–Kolmogorov Problem*)

Let the measurement $l(t)$ is the sum of the signal $x(t)$ and a noise $\epsilon(t)$:

$$l(t) = x(t) + \epsilon(t) \tag{4.4.10}$$

The statistical property of $x(t)$ is given in (4.4.3a) and that of $\epsilon(t)$ is as follows:

$$E\{\ \epsilon(t)\ \} = 0, \qquad E\{\ \epsilon(t)\ \epsilon(t+\tau)\ \} = C_{\epsilon\epsilon}(\tau) \tag{4.4.11}$$

Then it can be proved that $l(t)$ will have the auto–covariance function and cross–covariance function with $x(t)$:

$$E\{l(t)\ l(t+\tau)\} = C_{xx}(\tau) + C_{\epsilon\epsilon}(\tau) \tag{4.4.12}$$

$$E\{x(t)\ l(t+\tau)\} = C_{xx}(\tau) = C_{xl}(\tau) \tag{4.4.13}$$

Let $S_{xx}(\omega)$ and $S_{\epsilon\epsilon}(\omega)$ denote the Fourier transform of $C_{xx}(\tau)$ and $C_{\epsilon\epsilon}(\tau)$, respectively. Then the Fourier transform of C_{xl} and C_{ll} can be derived:

$$S_{xl}(\omega) = S_{xx}(\omega), \qquad S_{ll}(\omega) = S_{xx}(\omega) + S_{\epsilon\epsilon}(\omega)$$

$$H(\omega) = \frac{S_{xx}(\omega)}{S_{xx}(\omega) + S_{\epsilon\epsilon}(\omega)} \tag{4.4.14}$$

The right side of (4.4.14) can be viewed as a weighting factor in the spectral domain. Inserting (4.4.14) into (4.4.9) will give the optimal weighting function needed to obtain the optimal estimate of $x(t)$ from $l(t)$ as expressed by (4.4.4).

§4.4.3 Discrete Wiener–Hopf Prediction

Let $\underset{n\,x\,1}{L} = (l_1, l_2, \cdots, l_n)^T$ denote the vector of n observations which have the following expectation and variance matrix:

$$\left.\begin{aligned} E\{\ L\ \} &= \mu_1 \\ E\{\ (L-\mu_1)\ (L-\mu_1)^T\ \} &= \underset{n\,x\,n}{C_{ll}} \end{aligned}\right\} \tag{4.4.15}$$

Let x denote a random variable such that:

$$\left.\begin{aligned} E\{\ x\ \} &= \mu_x \\ E\{\ [x-\mu_x]^2\ \} &= \underset{1\,x\,1}{\sigma_x^2} \\ E\{\ [x-\mu_x]\ [L-\mu_l]^T\ \} &= \underset{1\,x\,n}{C_{xl}} = C_{lx}{}^T \end{aligned}\right\} \tag{4.4.16}$$

Now we are looking for an estimate to x of the form:

$$\hat{x} = \alpha^T L + \beta \tag{4.4.17}$$

where $\alpha \equiv (\alpha_1, \alpha_2, \cdots, \alpha_n)^T$ denote a vector of unknown constants to de determined and β another non–random constant. If we require that $\hat{x}$ is an unbiased estimate of x, we have:

$$E\{\hat{x}\} = \alpha^T \mu_l + \beta \equiv \mu_x \qquad \text{or } \beta \equiv \mu_x - \alpha^T \mu_l$$

$$\text{i.e} \quad \hat{x} = \mu_x + \alpha^T (L - \mu_l) \tag{4.4.18}$$

The estimation error of $\hat{x}$ and its variance are:

$$\Delta = \hat{x} - x = \alpha^T (L - \mu_l) - (x - \mu_x) = [\alpha^T, -I] \begin{bmatrix} L - \mu_l \\ x - \mu_x \end{bmatrix}$$

$$\sigma_\Delta^2 = E\{\Delta^2\} = [\alpha^T, -I] \begin{bmatrix} C_{ll} & C_{lx} \\ C_{xl} & \sigma_x^2 \end{bmatrix} \begin{bmatrix} \alpha \\ -I \end{bmatrix}$$

$$= \alpha^T C_{ll} \alpha - 2\, C_{xl}\, \alpha + \sigma_x^2 \tag{4.4.19}$$

The discrete Wiener–Hopf prediction chooses the unknown constants in α in such a way that the error variance of $\hat{x}$ given by (4.4.19) is minimized. Therefore, we let the derivative of σ_Δ^2 with respect to α vanish:

$$d\{\sigma_\Delta^2\}/d\alpha = 2\alpha^T C_{ll} - 2\, C_{xl} = 0 \qquad \Longrightarrow: \quad \alpha^T = C_{xl}\, C_{ll}^{-1}$$

Inserting the above α^T into (4.4.18) and (4.4.19), we obtain our optimal prediction $\hat{x}$, its variance $\sigma_{\hat{x}}^2$ (given by the error propagation law) and its error variance σ_Δ^2:

$$\left.\begin{aligned} \hat{x} &= \mu_x + C_{xl}\, C_{ll}^{-1} (L - \mu_l) \\ \sigma_{\hat{x}}^2 &= C_{xl}\, C_{ll}^{-1}\, C_{lx} \\ \sigma_\Delta^2 &= \sigma_x^2 - C_{xl}\, C_{ll}^{-1}\, C_{lx} \end{aligned}\right\} \tag{4.4.20}$$

The above prediction can be easily extended to the case where a vector $\underset{m\,x\,1}{X} \equiv (x_1, x_2, \cdots, x_m)^T$ is to be predicted from the observation vector L. Let $\underset{m\,x\,1}{\mu_x}$, $\underset{m\,x\,m}{C_{xx}}$ denote the a priori expectation and variance–covariance matrix of X, respectively, and let $\underset{m\,x\,n}{C_{xl}}$ denote the covariance matrix between X and L. Then based on (4.4.20), we can easily derive the Wiener–Hopf prediction $\hat{X}$, its variance–covariance matrix $C_{\hat{X}\hat{X}}$ and the variance–covariance matrix $C_{\Delta\Delta}$ of the prediction error vector ($\Delta \equiv \hat{X} - X$) :

$$\hat{X} = \mu_x + C_{xl}\, C_{ll}^{-1} (L - \mu_x) \tag{4.4.21}$$

$$C_{\hat{X}\hat{X}} = C_{xl}\, C_{ll}^{-1}\, C_{lx} \tag{4.4.22}$$

$$C_{\Delta\Delta} = C_{xx} - C_{xl}\, C_{ll}^{-1}\, C_{lx} \tag{4.4.23}$$

Example 4.4.3 (*Condition Adjustment*)

We look at the condition adjustment model [see (2.1.2)]:

$$\underset{t\,x\,n}{B}\ \underset{nx1}{\epsilon} = \underset{t\,x\,1}{W}$$

with statistical information [see (2.1.3), (2.1.4), (2.1.14) and (2.1.15)]:

$$E\{\epsilon\} = 0\,, \qquad E\{\epsilon\epsilon^T\} = \sigma_0^2 P^{-1} = C_{\epsilon\epsilon}$$

$$E\{W\} = B\,E(\epsilon) = 0, \qquad E\{WW^T\} = \sigma_0^2\, BP^{-1}B^T = C_{ww}$$

$$E\{\epsilon W^T\} = \sigma_0^2 P^{-1}B^T = C_{\epsilon w}$$

From (4.4.22), we can predict $\hat{\epsilon}$ from the observed W:

$$\hat{\epsilon} = E\{\epsilon\} + [W - E\{W\}] = P^{-1}B^T(BP^{-1}B^T)^{-1}\, W \tag{4.4.24}$$

The variance–covariance matrix of $\hat{\epsilon}$ and the variance–covariance matrix of the prediction error $\Delta \equiv \hat{\epsilon} - \epsilon$ are:

$$C_{\hat{\epsilon}\hat{\epsilon}} = C_{\epsilon w}\, C_{ww}^{-1}\, C_{w\epsilon} = \sigma_0^2\, P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1} \tag{4.4.25}$$

$$C_{\Delta\Delta} = C_{\epsilon\epsilon} - C_{\epsilon w}\, C_{ww}^{-1}\, C_{w\epsilon} = \sigma_0^2\, [P^{-1} - P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1}] \tag{4.4.26}$$

Comparing (4.4.24) and (4.4.25) with (2.1.10) and (2.1.18), we can see that *the Wiener–Hopf prediction in condition adjustment is identical to the least squares estimate of the residuals* $\hat{\epsilon}$. In other words, the least squares estimate $\hat{\epsilon}$ has the minimum error variance among all linear estimates and thus can be called a *Best Linear Unbiased Estimate (BLUE)*. In addition, a comparison of (4.4.26) with (2.1.19) shows that the error variance matrix of $\hat{\epsilon}$ is the same the variance–covariance matrix of the adjusted observations $\hat{L} = L - \hat{\epsilon}$.

Example 4.4.4 (*Adjustment by Elements with Random Parameters*)

We now consider the adjustment model in (3.4.1):

$$\underset{n\times 1}{L} - \underset{n\times 1}{\epsilon} = \underset{n\times m}{A}\ \underset{m\times 1}{X}$$

where both ϵ and X are random variables with the following a priori statistical properties [cf (3.4.2), (3.4.3) and (3.4.4)]:

$$\begin{aligned} &E\{\epsilon\} = 0; && E(\epsilon\epsilon^T) = \sigma_0^2 P^{-1} && \equiv C_{\epsilon\epsilon} = \sigma_0^2 P^{-1} \\ &E\{X\} = \mu_x; && E\{(X-\mu_x)(X-\mu_x)^T\} && \equiv C_{xx} = \sigma_0^2 P_x^{-1} \\ & && E\{(X-\mu_x)\epsilon^T\} = 0 && \equiv C_{x\epsilon} \end{aligned}$$

The expectation and variance matrix of L as well as covariance matrices between L and ϵ, X can be derived:

$$\begin{aligned} \mu_l &= E\{L\} = A\,\mu_x \\ C_{ll} &= E\{[L-\mu_l]\,[L-\mu_l]^T\} = AC_{xx}A^T + C_{\epsilon\epsilon} = \sigma_0^2\,(AP_x^{-1}A^T+P^{-1}) \\ C_{xl} &= E\{[X-\mu_x]\,[L-\mu_l]^T\} = C_{xx}A^T = \sigma_0^2\,P_x^{-1}A^T \\ C_{\epsilon l} &= E\{\,\epsilon\,[L-\mu_l]^T\} = C_{\epsilon\epsilon} = \sigma_0^2\,P^{-1} \end{aligned}$$

Using (4.4.21), we can predict $\hat{X}$ from observations L:

$$\begin{aligned} \hat{X} &= \mu_x + C_{xl}\,C_{ll}^{-1}\,(L-\mu_l) = \mu_x + C_{xx}A^T(AC_{xx}A^T+C_{\epsilon\epsilon})^{-1}\,(L-A\,\mu_x) \\ &= \mu_x + P_x^{-1}A^T\,(AP_x^{-1}A^T+P^{-1})^{-1}\,(L-A\,\mu_x) \end{aligned} \tag{4.4.27}$$

Considering identity (A1.1) in Appendix A1, one can show:

$$P_x^{-1}A^T\,(AP_x^{-1}A^T + P^{-1})^{-1} = (A^TPA + P_x)^{-1}\,A^TP \tag{4.4.28}$$

Thus the prediction $\hat{X}$ can be rewritten in an alternative but equivalent way:

$$\hat{X} = \mu_x + (A^TPA + P_x)^{-1}A^TP\,\{L-A\,\mu_x\} \tag{4.4.29}$$

Similarly, one can derive the Wiener–Hopf prediction $\hat{\epsilon}$:

$$\begin{aligned} \hat{\epsilon} &= E\{\epsilon\} + C_{\epsilon l}\,C_{ll}^{-1}\,(L-\mu_l) \\ &= (L-A\mu_x) - AP_x^{-1}A^T(AP_x^{-1}A^T + P^{-1})^{-1}\,\{\,L-A\mu_x\,\} \end{aligned} \tag{4.4.30}$$

$$= (L - A\mu_x) - A(A^T P A + P_x)^{-1} A^T P \{ L - A\mu_x \} \qquad (4.4.31)$$

where (4.4.30) and (4.4.31) correspond to (4.4.27) and (4.4.29), respectively. The above two Wiener–Hopf predictions ($\hat{X}$, $\hat{\epsilon}$) are identical to results given in (3.4.10) and (3.4.11). Therefore, the least squares adjustment using pseudo–observations as described in §3.4 is equivalent to the best linear unbiased estimation.

The variance–covariance matrix of $\hat{X}$ can be obtained from (4.4.24):

$$C_{\hat{X}\hat{X}} = C_{xx} - C_{xl}\, C_{ll}^{-1}\, C_{lx} = C_{xx} - C_{xx}A^T(AC_{xx}A^T + C_{\epsilon\epsilon})^{-1}AC_{xx} \qquad (4.4.32)$$

$$= (A^T C_{\epsilon\epsilon}^{-1} A + C_{xx}^{-1})^{-1} = \sigma_0^2\,(A^T P A + P_x)^{-1} = \sigma_0^2\, P_{\hat{X}}^{-1} \qquad (4.4.33)$$

where $P_{\hat{X}}$ denotes the weight matrix of the prediction $\hat{X}$ and is given by:

$$P_{\hat{X}} = A^T P A + P_x \qquad (4.4.34)$$

4.4.4 Mathematical Models of Linear Dynamical Systems

A dynamical system can be described by a number of parameters. For instance, the motion of an artificial satellite in space can be described by its 3–D coordinates (x,y,z) and their time derivatives. A vector consisting of these parameters are often called the *state vector* of the dynamical system. The dynamical characteristics of a *linear* system can be described by a vectorized differential equation like:

$$\underset{m\,x\,1}{\dot{X}(t)} = \underset{m\,x\,m}{G(t)}\ \underset{m\,x\,1}{X(t)} + \underset{m\,x\,q}{H(t)}\ \underset{q\,x\,1}{v(t)} \qquad (4.4.35)$$

where:

- $X(t)$ = the state vector of the dynamical system
- $\dot{X}(t)$ = the time derivative of the state vector
- $G(t)$ = transition matrix of the state vector
- $H(t)$ = transition matrix of the system noise vector
- $v(t)$ = system noise vector with the following expectation and covariance matrix (time dependent in general):

$$E\{v(t)\} = 0; \qquad E\{v(t_1)\, v(t_2)^T\} = \underset{q\,x\,q}{D(t)}\ \delta(t_2 - t_1) \qquad (4.4.36)$$

where $\delta(t)$ is the so called Dirac–function as defined by:

$$\left.\begin{aligned} \delta(\tau) &= 0 \qquad \text{for } \tau \neq 0 \\ \int_{-\infty}^{+\infty} \delta(\tau)\, d\tau &= 1 \end{aligned}\right\} \qquad (4.4.37)$$

The observations can be related to the state vector in the following way:

$$\underset{n\times 1}{L}(t) = \underset{n\times m}{A}(t)\ \underset{m\times 1}{X}(t) + \underset{n\times 1}{\epsilon}(t) \qquad (4.4.38)$$

where:

$L(t)$ = the observation vector (time dependent in general)

$A(t)$ = the design matrix

$\epsilon(t)$ = the observation error vector with the following statistical property:

$$E\{\epsilon(t)\} = 0, \qquad E\{\epsilon(t_1)\ \epsilon(t_2)^T\} = \underset{n\times n}{C}(t)\ \delta(t_2-t_1) \qquad (4.4.39)$$

The system noise $v(t)$ and the observation error $\epsilon(t)$ are normally assumed to be uncorrelated with each other, i.e.:

$$E\{\ v(t_1)\ \epsilon(t_2)^T\ \} = \underset{q\times n}{0} \qquad (4.4.40)$$

Equation (4.4.35) is called the *state equation* while (4.4.38) is called the *observation equation.* (4.4.36), (4.4.39) and (4.4.40) form the associated *statistical models.* As all quantities in the above equations are in general continuous functions of time t, they together represent the *mathematical models of a continuous linear dynamical system.* The problem of estimating $X(t)$ from $L(t)$ using the continuous state equation and the observation equation may be solved analytically. However, since some dynamical systems originally have discrete characters and meanwhile most continuous problems can be reduced into discrete ones, we do not present here the solution of the continuous case. Instead, we concentrate on the discrete problem defined below.

The state equation and observation equation of a discrete linear dynamical system can be written as follows:

$$\underset{m\times 1}{X_{k+1}} = \underset{m\times m}{G_{k+1\,k}}\ \underset{m\times 1}{X_k} + \underset{m\times q}{H_{k+1}}\ \underset{q\times 1}{v_{k+1}} \qquad (k= 0, 1, 2, \cdots) \qquad (4.4.41)$$

$$\underset{n\times 1}{L_{k+1}} = \underset{n\times m}{A_{k+1}}\ \underset{m\times 1}{X_{k+1}} + \underset{n\times 1}{\epsilon_{k+1}} \qquad (k= 0, 1, 2, \cdots) \qquad (4.4.42)$$

The definition of various notations in (4.4.41) and (4.4.42) are similar to those in (4.4.35) and (4.4.38), except that all quantities are discrete, depending on index k. The associated statistical models are:

$$\left.\begin{aligned} E\{\ v_k\ \} &= 0 \\ E\{\ \epsilon_k\ \} &= 0 \\ E\{\ v_k\ v_j^T\ \} &= \underset{q\times q}{D_k}\ \delta_{kj} \\ E\{\ \epsilon_k\ \epsilon_j^T\ \} &= \underset{n\times n}{C_k}\ \delta_{kj} \\ E\{\ v_k\ \epsilon_j^T\ \} &= 0 \end{aligned}\right\} \qquad (4.4.43)$$

where $\delta_{kj} = \begin{cases} 1 & \text{for } k = j \\ 0 & \text{for } k \neq j \end{cases}$.

The initial value of X_k and its variance matrix Q_{kk} for k=0 are denoted by $\hat{X}_{00}$ and Q_{00}, respectively. The weight matrix of $\hat{X}_{00}$ is then: $P^x_{00} = \sigma_0^2 Q^{-1}_{00}$, where σ_0^2 denotes the variance factor (unit–weight standard error squared).

§4.4.5 Discrete Kalman Filtering

We now discuss the optimal estimation in the discrete problem represented by (4.4.41), (4.4.42) and (4.4.43). Our objective is to derive the optimal estimate of X_{k+1} and its variance–covariance matrix based on the available observations $L_1, L_2, L_3, \cdots, L_k, L_{k+1}$. We start with the first two steps for $k = 0$ and $k = 1$, before presenting the general solutions for any arbitrary k+1.

Step 1 ($k = 0$)

From the state equation (4.4.41) and the initial values $\hat{X}_{00}$ and Q_{00}, we can predict the initial value and the initial variance–covariance matrix of X_1 :

$$\left.\begin{aligned} \hat{X}_{10} &= G_{10} X_{00} \\ Q_{10} &= G_{10} Q_{00} G_{10}^T + H_1 D_1 H_1^T = \sigma_0^2 P^{x-1}_{10} \end{aligned}\right\} \quad (4.4.44)$$

where P^x_{10} denotes the weight matrix of $\hat{X}_{10}$. The subscript "$_{10}$" indicates that the parameter X_1 is concerned and no observation is used to predict $\hat{X}_{10}$. When L_1 is available, an adjustment by elements with observation equation:

$$L_1 - \epsilon_1 = A_1 X_1,$$

can be made using $\hat{X}_{10}$, Q_{10} and P^x_{10} as the a priori expectation, variance matrix and weight matrix, respectively. The variance–covariance matrix of L_1 and covariance matrix between X_1 and L_1 are:

$$\left.\begin{aligned} C_{l_1 l_1} &= A_1 Q_{10} A_1^T + C_1 \\ C_{x_1 l_1} &= Q_{10} A_1^T \end{aligned}\right\} \quad (4.4.45)$$

From (4.4.27) and (4.4.29), we then obtain the best linear unbiased estimate $\hat{X}_{11}$:

$$\hat{X}_{11} = \hat{X}_{10} + C_{x_1 l_1} C_{l_1 l_1}^{-1} \{ L_1 - A_1 \hat{X}_{10} \}$$

$$= \hat{X}_{10} + Q_{10} A_1^T (A_1 Q_{10} A_1^T + C_1)^{-1} \{ L_1 - A_1 \hat{X}_{10} \} \quad (4.4.46)$$

$$= \hat{X}_{10} + (A_1^T P_1 A_1 + P_{10}^x)^{-1} A_1^T P_1 \{ L_1 - A_1 \hat{X}_{10} \} \quad (4.4.47)$$

The error variance matrix of $\hat{X}_{11}$, denoted as Q_{11}, is given by (4.4.32) and (4.4.33):

$$Q_{11} = Q_{10} - Q_{10} A_1^T (A_1 Q_{10} A_1^T + C_1)^{-1} A_1 Q_{10} \quad (4.4.48)$$

$$= (A_1^T C_1 A_1 + Q_{10})^{-1} = \sigma_0^2 \, P_{11}^{x-1} \quad (4.4.49)$$

where P_{11}^x denotes the weight matrix of $\hat{X}_{11}$ and is defined as:

$$P_{11}^x = A_1^T P_1 A_1 + P_{10}^x \quad (4.4.50)$$

Step 2 ($k = 1$)

First we obtain a prediction of X_2 and its error variance matrix using the state equation:

$$\left.\begin{aligned} \hat{X}_{21} &= G_{21} \hat{X}_{11} \\ Q_{21} &= G_{21} Q_{11} G_{21}^T + H_2 D_2 H_2^T = \sigma_0^2 \, P_{21}^{x-1} \end{aligned}\right\} \quad (4.4.51)$$

where P_{21}^x denotes the weight matrix of $\hat{X}_{21}$. The subscript "$_{21}$" indicates that the parameter X_2 is concerned and only observation L_1 is used to predict $\hat{X}_{21}$. When L_2 is available, the observation equation:

$$L_2 - \epsilon_2 = A_2 X_2,$$

can be used together with $\hat{X}_{21}$, Q_{21} and P_{21}^x to obtain the updated estimate of $\hat{X}_2$ [see also (4.4.46) and (4.4.47)] :

$$\hat{X}_{22} = \hat{X}_{21} + Q_{21} A_2^T (A_2 Q_{21} A_2^T + C_2)^{-1} \{ L_2 - A_2 \hat{X}_{21} \} \quad (4.4.52)$$

$$= \hat{X}_{21} + (A_2^T P_2 A_2 + P_{21}^x)^{-1} A_2^T P_2 \{ L_2 - A_2 \hat{X}_{21} \} \quad (4.4.53)$$

The error variance matrix of $\hat{X}_{22}$ is as follows:

$$Q_{22} = Q_{21} - Q_{21}A_2^T (A_2 Q_{21} A_2^T + C_2)^{-1} A_2 Q_{21} \quad (4.4.54)$$

$$= (A_2^T C_2 A_2 + Q_{21})^{-1} = \sigma_0^2 P_{22}^{x-1} \quad (4.4.55)$$

where P_{22}^x denotes the weight matrix of $\hat{X}_{22}$:

$$P_{22}^x = A_2^T P_2 A_2 + P_{21}^x \quad (4.4.56)$$

Step k+1 (k = k)

Now let us consider the general case for any arbitrary k+1. Assume that from L_1, L_2, $\cdots$, L_k, we have obtained estimate $\hat{X}_{kk}$, its error variance matrix Q_{kk} and weight matrix P_{kk}^x . From the state equation (4.4.41), we get an prediction of X_{k+1} and its error variance matrix $Q_{k+1\,k}$ and weight matrix $P_{k+1\,k}^x$:

$$\hat{X}_{k+1\,k} = G_{k+1\,k} \hat{X}_{k\,k} \quad (4.4.57)$$

$$Q_{k+1\,k} = G_{k+1\,k} Q_{k\,k} G_{k+1\,k}^T + H_{k+1} D_{k+1} H_{k+1}^T = \sigma_0^2 P_{k+1\,k}^{x\,-1} \quad (4.4.58)$$

With the new observation equation $L_{k+1} - \epsilon_{k+1} = A_{k+1} X_{k+1}$, the updated estimate of X_{k+1} is obtained from (4.4.27) and (4.4.29):

$$\hat{X}_{k+1\,k+1} = \hat{X}_{k+1\,k} + K_{k+1} (L_{k+1} - A_{k+1} \hat{X}_{k+1\,k}) \quad (4.4.59)$$

where matrix K_{k+1} denotes the so called *Kalman gain* as defined by:

$$K_{k+1} = Q_{k+1\,k} A_{k+1}^T (A_{k+1} Q_{k+1\,k} A_{k+1}^T + C_{k+1})^{-1} \quad (4.4.60)$$

$$= (A_{k+1}^T C_{k+1}^{-1} A_{k+1} + P_{k+1\,k}^{x\,-1})^{-1} A_{k+1}^T C_{k+1}^{-1}$$

$$= (A_{k+1}^T P_{k+1} A_{k+1} + P_{k+1\,k}^x)^{-1} A_{k+1}^T P_{k+1} \quad (4.4.61)$$

The error variance matrix and weight matrix of $\hat{X}_{k+1\,k+1}$ follow from (4.4.32) and (4.4.33):

$$Q_{k+1\,k+1} = Q_{k+1\,k} - Q_{k+1\,k} A_{k+1}^T (A_{k+1} Q_{k+1\,k} A_{k+1}^T + C_{k+})^{-1} A_{k+1} Q_{k+1\,k} \quad (4.4.62)$$

$$= (A_{k+1}^T C_{k+1} A_{k+1} + Q_{k+1\,k})^{-1} = \sigma_0^2 P_{k+1\,k+1}^{x\,-1} \quad (4.4.63)$$

$$P_{k+1\,k+1}^x = A_{k+1}^T C_{k+1} A_{k+1} + Q_{k+1\,k} \quad (4.4.64)$$

Formula pairs, (4.4.60) and (4.4.61), (4.4.62) and (4.4.63), are due to identity (4.4.28). As the first formula in each pair involves the inverse of a matrix of n x n while the

second formula involves matrix inverse of dimension m x m, it could be favorable to use the second formula in each pair when the number of observations in L_{k+1} is larger than the number of parameters in X_{k+1} (i.e. $n > m$).

One of the main advantages of the discrete Kalman filtering is that at each processing step, not all observations $L_1, L_2, \cdots, L_{k+1}$ need to be stored and processed in the computer's internal memory . Instead, only one group of observation data need to be stored and processed at each step, while information from previous groups of observations is not lost, rather transferred forward in the form of the parameter estimate obtained at the step before. This procedure is especially efficient for dynamical systems where index k corresponds to different time epoch and observations are collected in (almost) real time. Using the strategy of Kalman filtering, processing of observation data and updating of parameter estimates can be easily done in real time.

Appendix A1: Some Formulas in Matrix Algebra

A1.1 Two Identities of Matrix Inversion

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

$$AB\,(D + CAB)^{-1} = (A^{-1} + BD^{-1}C)^{-1}\,BD^{-1} \tag{A1.1}$$

$$(D + CAB)^{-1} = D^{-1} - D^{-1}C\,(A^{-1} + BD^{-1}C)^{-1}\,BD^{-1} \tag{A1.2}$$

Proof:

Let $\underset{s\,x\,s}{G} \equiv D + CAB$ and we have:

$$(D + CAB)\,G^{-1} = \underset{s\,x\,s}{I} = DG^{-1} + CABG^{-1}$$

$$G^{-1} = D^{-1} - D^{-1}\,CABG^{-1} \tag{A1.3}$$

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

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

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

Considering $G = D + CAB$, the last equation gives immediately (A1.1). Inserting the last equation into (A1.3) then leads to (A1.2).

A1.2 Inverse of A Partitioned Square Matrix

Let $\underset{n\,x\,n}{N}$ denote a non–singular square matrix which can be partitioned into four submatrices:

$$\underset{n\,x\,n}{N} = \begin{bmatrix} \underset{r\,x\,r}{N_{11}} & \underset{r\,x\,s}{N_{12}} \\ \underset{s\,x\,r}{N_{21}} & \underset{s\,x\,s}{N_{22}} \end{bmatrix} \quad (n=r+s) \tag{A1.4}$$

where both N_{11} and N_{22} are non–singular. Then the inverse matrix of N is given by:

$$\underset{n\,x\,n}{N}^{-1} = \begin{bmatrix} N_{11}^{-1} + N_{11}^{-1}\,N_{12}\,\overline{N}_{22}^{-1}\,N_{21}\,N_{11}^{-1} & -N_{11}^{-1}\,N_{12}\,\overline{N}_{22}^{-1} \\ -\overline{N}_{22}^{-1}\,N_{21}\,N_{11}^{-1} & \overline{N}_{22}^{-1} \end{bmatrix} \tag{A1.5}$$

$$= \begin{bmatrix} \overline{N}_{11}^{-1} & -\overline{N}_{11}^{-1}\,N_{12}\,N_{22}^{-1} \\ -N_{22}^{-1}\,N_{21}\,\overline{N}_{11}^{-1} & N_{22}^{-1} + N_{22}^{-1}\,N_{21}\,\overline{N}_{11}^{-1}\,N_{12}\,N_{22}^{-1} \end{bmatrix} \tag{A1.6}$$

where:

$$\left.\begin{aligned} \overline{N}_{11} &\equiv N_{11} - N_{12}\, N_{22}^{-1}\, N_{21} \\ \overline{N}_{22} &\equiv N_{22} - N_{21}\, N_{11}^{-1}\, N_{12} \end{aligned}\right\} \qquad \text{(A1.7)}$$

Proof:

Assume that the inverse matrix of N, Q, is also partitioned :

$$\underset{n\,x\,n}{N}^{-1} \equiv \underset{n\,x\,n}{Q} \equiv \begin{bmatrix} \underset{r\,xr}{Q_{11}} & \underset{r\,xs}{Q_{12}} \\ \underset{s\,xr}{Q_{21}} & \underset{s\,xs}{Q_{22}} \end{bmatrix}$$

Since $N\,Q = I$, i.e.:

$$\begin{bmatrix} \underset{r\,xr}{N_{11}} & \underset{r\,xs}{N_{12}} \\ \underset{s\,xr}{N_{21}} & \underset{s\,xs}{N_{22}} \end{bmatrix} \begin{bmatrix} \underset{r\,xr}{Q_{11}} & \underset{r\,xs}{Q_{12}} \\ \underset{s\,xr}{Q_{21}} & \underset{s\,xs}{Q_{22}} \end{bmatrix} = \begin{bmatrix} \underset{rxr}{I} & \underset{sxr}{0} \\ \underset{sxr}{0} & \underset{sxs}{I} \end{bmatrix}$$

we then have :

$$\begin{aligned} N_{11}\, Q_{11} + N_{12}\, Q_{21} &= I && \text{(a)} \\ N_{11}\, Q_{12} + N_{12}\, Q_{22} &= 0 && \text{(b)} \\ N_{21}\, Q_{11} + N_{22}\, Q_{21} &= 0 && \text{(c)} \\ N_{21}\, Q_{12} + N_{22}\, Q_{22} &= I && \text{(d)} \end{aligned}$$

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

$$\begin{aligned} Q_{11} &= N_{11}^{-1} - N_{11}^{-1}\, N_{12}\, Q_{21} \\ N_{21}\, Q_{11} + N_{22}\, Q_{21} &= N_{21}(N_{11}^{-1} - N_{11}^{-1}\, N_{12}\, Q_{21}) + N_{22}\, Q_{21} = 0 \\ Q_{21} &= -(N_{22} - N_{21}\, N_{11}^{-1}\, N_{12})^{-1}\, N_{21}\, N_{11}^{-1} = -\overline{N}_{22}^{-1}\, N_{21}\, N_{11}^{-1} \\ Q_{11} &= N_{11}^{-1} + N_{11}^{-1}\, N_{12}\, \overline{N}_{22}^{-1}\, N_{21}\, N_{11}^{-1} \\ N_{21}\, Q_{12} + N_{22}\, Q_{22} &= I \end{aligned}$$

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

$$\begin{aligned} Q_{12} &= -N_{11}^{-1}\, N_{12}\, Q_{22} \\ N_{21}\, Q_{12} + N_{22}\, Q_{22} &= N_{21}\,(-N_{11}^{-1}\, N_{12}\, Q_{22}) + N_{22}\, Q_{22} = \overline{N}_{22}\, Q_{22} = I \\ Q_{22} &= \overline{N}_{22}^{-1} \\ Q_{12} &= -N_{11}^{-1}\, N_{12}\, \overline{N}_{22}^{-1} \end{aligned}$$

(A1.6) can be proved in similar way but using the relation $Q\,N = I$.

As the inverse matrix of N is unique, the partitioned inverse matrix in (A1.5) and (A1.6) should be identical which implies:

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

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

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

or more explicitly:

$$(N_{11} - N_{12} N_{22}^{-1} N_{21})^{-1} \equiv N_{11}^{-1} + N_{11}^{-1} N_{12} \overline{N}_{22}^{-1} N_{21} N_{11}^{-1} \qquad \text{(A1.8)}$$

$$(N_{22} - N_{21} N_{11}^{-1} N_{12})^{-1} = N_{22}^{-1} + N_{22}^{-1} N_{21} \overline{N}_{11}^{-1} N_{12} N_{22}^{-1} \qquad \text{(A1.9)}$$

$$N_{11}^{-1} N_{12} (N_{22} - N_{21} N_{11}^{-1} N_{12})^{-1} = (N_{11} - N_{12} N_{22}^{-1} N_{21})^{-1} N_{12} N_{22}^{-1} \qquad \text{(A1.10)}$$

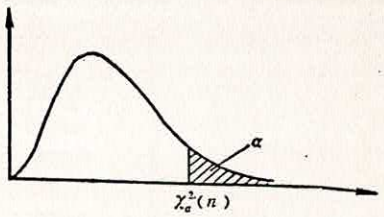
(A1.10) corresponds to the first inverse formula of matrices (A1.1), while (A1.8) and (A1.9) correspond to the inverse formula (A1.2).

The Standard Normal Distribution

ε X	$f(\varepsilon)$ *Frequency*	$P(\varepsilon<X)$ *Distribution*
0.0	0.39894	0.50000
0.1	0.39695	0.53983
0.2	0.39104	0.57926
0.3	0.38139	0.61791
0.4	0.36827	0.65542
0.5	0.35207	0.69146
0.6	0.33322	0.72575
0.7	0.31225	0.75804
0.8	0.28969	0.78814
0.9	0.26609	0.81594
1.0	0.24197	0.84134
1.1	0.21785	0.86433
1.2	0.19419	0.88493
1.3	0.17137	0.90320
1.4	0.14973	0.91924
1.5	0.12952	0.93319
1.6	0.11092	0.94520
1.7	0.09405	0.95543
1.8	0.07895	0.96407
1.9	0.06562	0.97128
2.0	0.05399	0.97725
2.1	0.04398	0.98214
2.2	0.03547	0.98610
2.3	0.02833	0.98928
2.4	0.02239	0.99180
2.5	0.01753	0.99379
2.6	0.01358	0.99534
2.7	0.01042	0.99653
2.8	0.00792	0.99744
2.9	0.00595	0.99813
3.0	0.00443	0.99865
3.1	0.00327	0.99903
3.2	0.00238	0.99931
3.3	0.00172	0.99952
3.4	0.00123	0.99966
3.5	0.00087	0.99977
3.6	0.00061	0.99984
3.7	0.00042	0.99989
3.8	0.00029	0.99993
3.9	0.00020	0.99995
4.0	0.00013	0.99997

The χ^2-Distribution

$$P\{\chi^2(n) > \chi_\alpha^2(n)\} = \alpha$$



n	$\alpha=0.995$	0.99	0.975	0.95	0.90	0.75
1	—	—	0.001	0.004	0.016	0.102
2	0.010	0.020	0.051	0.103	0.211	0.575
3	0.072	0.115	0.216	0.352	0.584	1.213
4	0.207	0.297	0.484	0.711	1.064	1.923
5	0.412	0.554	0.831	1.145	1.610	2.675
6	0.676	0.872	1.237	1.635	2.204	3.455
7	0.989	1.239	1.690	2.167	2.833	4.255
8	1.344	1.646	2.180	2.733	3.490	5.071
9	1.735	2.088	2.700	3.325	4.168	5.899
10	2.156	2.558	3.247	3.940	4.865	6.737
11	2.603	3.053	3.816	4.575	5.578	7.584
12	3.074	3.571	4.404	5.226	6.304	8.438
13	3.565	4.107	5.009	5.892	7.042	9.299
14	4.075	4.660	5.629	6.571	7.790	10.165
15	4.601	5.229	6.262	7.261	8.547	11.037
16	5.142	5.812	6.908	7.962	9.312	11.912
17	5.697	6.408	7.564	8.672	10.085	12.792
18	6.265	7.015	8.231	9.390	10.865	13.675
19	6.844	7.633	8.907	10.117	11.651	14.562
20	7.434	8.260	9.591	10.851	12.443	15.452
21	8.034	8.897	10.283	11.591	13.240	16.344
22	8.643	9.542	10.982	12.338	14.042	17.240
23	9.260	10.196	11.689	13.091	14.848	18.137
24	9.886	10.856	12.401	13.848	15.659	19.037
25	10.520	11.524	13.120	14.611	16.473	19.939
26	11.160	12.198	13.844	15.379	17.292	20.843
27	11.808	12.879	14.573	16.151	18.114	21.749
28	12.461	13.565	15.308	16.928	18.939	22.657
29	13.121	14.257	16.047	17.708	19.768	23.567
30	13.787	14.954	16.791	18.493	20.599	24.478
31	14.458	15.655	17.539	19.281	21.434	25.390
32	15.134	16.362	18.291	20.072	22.271	26.304
33	15.815	17.074	19.047	20.867	23.110	27.219
34	16.501	17.789	19.806	21.664	23.952	28.136
35	17.192	18.509	20.569	22.465	24.797	29.054
36	17.887	19.233	21.336	23.269	25.643	29.973
37	18.586	19.960	22.106	24.075	26.492	30.893
38	19.289	20.691	22.878	24.884	27.343	31.815
39	19.996	21.426	23.654	25.695	28.196	32.737
40	20.707	22.164	24.433	26.509	29.051	33.660
41	21.421	22.906	25.215	27.326	29.907	34.585
42	22.138	23.650	25.999	28.144	30.765	35.510
43	22.859	24.398	26.785	28.965	31.625	36.436
44	23.584	25.148	27.575	29.787	32.487	37.363
45	24.311	25.901	28.366	30.612	33.350	38.291

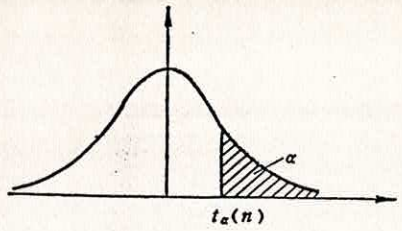
The χ^2-Distribution

$$P\{\chi^2(n) > \chi^2_\alpha(n)\} = \alpha$$

n	α=0.25	0.10	0.05	0.025	0.01	0.005
1	1.323	2.706	3.841	5.024	6.635	7.879
2	2.773	4.605	5.991	7.378	9.210	10.597
3	4.108	6.251	7.815	9.348	11.435	12.838
4	5.385	7.779	9.488	11.143	13.277	14.860
5	6.626	9.236	11.071	12.833	15.086	16.750
6	7.841	10.645	12.592	14.449	16.812	18.548
7	9.037	12.017	14.067	16.013	18.475	20.278
8	10.219	13.362	15.507	17.535	20.090	21.955
9	11.389	14.684	16.919	19.023	21.666	23.589
10	12.549	15.987	18.307	20.483	23.209	25.188
11	13.701	17.275	19.675	21.920	24.725	26.757
12	14.845	18.549	21.026	23.337	26.217	28.299
13	15.984	19.812	22.362	24.736	27.688	29.819
14	17.117	21.064	23.685	26.119	29.141	31.319
15	18.245	22.307	24.996	27.488	30.578	32.801
16	19.369	23.542	26.296	28.845	32.000	34.267
17	20.489	24.769	27.587	30.191	33.409	35.718
18	21.605	25.989	28.869	31.526	34.805	37.156
19	22.718	27.204	30.144	32.852	36.191	38.582
20	23.828	28.412	31.419	34.170	37.566	39.997
21	24.935	29.615	32.671	35.479	38.932	41.401
22	26.039	30.813	33.924	36.781	40.289	42.796
23	27.141	32.007	35.172	38.076	41.638	44.181
24	28.241	33.196	36.415	39.364	42.980	45.559
25	29.339	34.382	37.652	40.646	44.314	46.928
26	30.435	35.563	38.885	41.923	45.642	48.290
27	31.528	36.741	40.113	43.194	46.963	49.645
28	32.620	37.916	41.337	44.461	48.278	50.993
29	33.711	39.087	42.557	45.722	49.588	52.336
30	34.800	40.256	43.773	46.979	50.892	53.672
31	35.887	41.422	44.985	48.232	52.191	55.003
32	36.973	42.585	46.194	49.480	53.486	56.328
33	38.058	43.745	47.400	50.725	54.776	57.648
34	39.141	44.903	48.602	51.966	56.061	58.964
35	40.223	46.059	49.802	52.203	57.342	60.275
36	41.304	47.212	50.998	54.437	58.619	61.581
37	42.383	48.363	52.192	55.668	59.892	62.883
38	43.462	49.513	53.384	56.896	61.162	64.181
39	44.539	50.660	54.572	58.120	62.428	65.476
40	45.616	51.805	55.758	59.342	63.691	66.766
41	46.692	52.949	56.942	60.561	64.950	68.053
42	47.766	54.090	58.124	61.777	66.206	69.336
43	48.840	55.230	59.304	62.990	67.459	70.616
44	49.913	56.369	60.481	64.201	68.710	71.893
45	50.985	57.505	61.656	65.410	69.957	73.166

The t-Distribution

$P\{t(n) > t_\alpha(n)\} = \alpha$



n	α=0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896

$$P\{ F(n,m) > F_{\alpha}(n,m) \} = \alpha = 0.05$$

m \ n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	161	200	216	225	230	234	237	239	241	242	243	244	245	245	246	246	247	247
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.23	2.20	2.18	2.15	2.13	2.11	2.09	2.07
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18	2.15	2.13	2.11	2.09	2.07	2.05
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.10	2.07	2.04	2.01	1.99	1.97	1.95	1.94
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.08	2.05	2.02	1.99	1.97	1.95	1.93	1.92
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03	2.00	1.98	1.95	1.93	1.92	1.90
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.99	1.96	1.94	1.92	1.90	1.88
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92	1.90	1.89	1.87
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.03	1.99	1.96	1.93	1.91	1.89	1.87	1.86
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.95	1.92	1.90	1.88	1.86	1.84
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	2.00	1.97	1.94	1.91	1.89	1.87	1.85	1.83
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.99	1.96	1.93	1.90	1.88	1.86	1.84	1.82
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.92	1.89	1.87	1.85	1.83	1.81
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.97	1.93	1.90	1.88	1.85	1.83	1.81	1.79
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84	1.82	1.80	1.78
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.94	1.90	1.87	1.85	1.82	1.80	1.78	1.76
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.86	1.84	1.81	1.79	1.77	1.75
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.84	1.82	1.79	1.77	1.75	1.73
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.90	1.86	1.83	1.80	1.78	1.76	1.74	1.72
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82	1.79	1.77	1.75	1.73	1.71
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91	1.87	1.83	1.80	1.77	1.75	1.72	1.70	1.69
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.79	1.76	1.73	1.71	1.69	1.67
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.77	1.74	1.72	1.69	1.67	1.66
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.82	1.78	1.75	1.72	1.70	1.68	1.66	1.64
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.77	1.74	1.71	1.69	1.66	1.64	1.62
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76	1.73	1.70	1.68	1.65	1.63	1.61
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.72	1.69	1.67	1.64	1.62	1.60

References

Baarda, W. (1967). *Statistical Concepts in Geodesy.* Neth. Geod. Comm. Publ. on Geodesy, New Series 2, No.4.

Baarda, W. (1968). *A Testing procedure for Use in Geodetic Networks.* Neth. Geod. Comm. Publ. on Geodesy, New Series 2, No.5.

Bjerhammar, A. (1973). *Theory of Errors and Generalized Matrix Inverses.* Elsevier, Amsterdam.

Fan, H and L E Sjöberg (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.

Heck, B. (1980). *Statistische Ausreisserkriterien zur Kontrolle Geodätischer Beobachtugen.* Inte. Kurs für Ingenieursvermessung, Zürich, 1980.

Koch, K.R. (1980). *Parameterschätzung und Hypothesentests in Linearen Modellen.* Dümmler–Bonn.

Krarup, Torben, Jens Juhl and Kurt Kubik (198?). *Götterdämmerung over Least Squares Adjustment.* Presented at the 14th congress of the International Society of Photogrammetry.

Kubik, Kurt (1970). *The Estimation of the Weights of the Measured Quantities within the Method of Least Squares.* Bull. Geod. No. 95.

Mikhail, E.M. (1976). *Observations and Least Squares.* Harper and Row, 1976.

Pope, A.J. (1975). *The Statistics of Residuals and the Detection of Outliers.* XVI–th IUGG General Assembly, Grenoble, 1975.

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.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications.* 2nd Edition. John Wiley & Sons.

Rao, C.R. and S.K. Mitra (1971). *Generalized Inverses of Matrices and Its Apllications.* John Wiley & Sons.

Sjöberg, L.E. (1983). *Unbiased Estimation of Variance–Covariance Components in Condition Adjustment with Unknowns – A MINQUE Approach.* ZfV., 108.

Sjöberg, L.E. (1984). *Lecture Notes on General Matrix Calculus, Adjustment and Variance–Covariance Component Estimation.* Nordiska forskarkurser 12/1984 "Optimization of Geodetic Operations". Norges Geogr. Oppm. Publ. 3/1984.

Srivastava, M.S. and C.G. Khatri (1979). *An Introduction to Multivariate Statistics.* Elsevier North Holland, Inc.

Yu, Z. and L. Lu. (1982). *Surveying Adjustment* (in Chinese). Surveying and Mapping Publishing House, Beijing, China.