Adjustment Computations

by

Alfred Leick

Department of Spatial Information Science and Engineering
University of Maine
5711 Boardman Hall
Orono, ME 04469-5711

Adjustments by Least Squares is the most frequently used estimation procedure in Surveying Engineering and related disciplines. Anyone trying to gather information about specific quantities, either by direct or indirect measurement has to be fundamentally concerned with concepts such as random measurement errors, systematic measurement errors, propagation of errors and variances, precision and accuracy estimates, hypothesis testing, etc., in order to ensure a fulfillment of accuracy specification. Moreover, the successful Surveying Engineer will carry out optimization studies for each project, regardless of whether it is in Photogrammetry, Engineering Surveying, Control Surveying, Geodesy or Land Surveying. Such efforts include consideration regarding instrumentation and geometry of the design. With the help of a simulation study, in which case only the anticipated measurement accuracy is required, an optimal design can be readily established long before the actual field work starts, thus making a meaningful planning of the field work possible. This in turn can lead to a decisive reduction of expenses.

Least squares estimation is but one of several estimation techniques regularly taught in statistical courses. The appeal of the least squares approach in Physical Science is that the required statistical information about observations is minimal and that the normal equations are linear algebraic equations which are conceptually easy to solve.

An elegant derivation of the adjustment models and a thorough understanding of subsequent analyses and statistical testing require a profound knowledge of linear algebra and statistics. In this course, therefore, adjustments is not considered an "exclusive domain of surveying" with its own jargon, but rather a general estimation technique applicable

to a variety of disciplines. This understanding is reflected in the choice of symbols, in the terminology and in the generality of the derivations. Application of the models to specific surveying engineering problems is the subject of laboratories, which in itself is a sufficient justification for a separate course in a Surveying Engineering curriculum. This general approach makes it possible to fully utilize the students' prerequisite knowledge of linear algebra and statistics.

Since a complete course in adjustments must include a discussion on statistical hypothesis testing, and thus must inevitably be concerned with statistical terminology and interpretation, it is logical to base the treatment of adjustments right from the beginning upon statistical concepts; i.e., the observations are said to be corrupted by random noise and have a variance-covariance matrix with respect to some unknown probability distribution, rather than referring to weighted observations where no statistical concepts are needed at all. For the estimate itself this distinction is not important since the algebraic expressions are identical in both cases. In hypothesis testing, however, assumptions regarding the probability distribution have to be made, unless so-called distribution-free tests are employed, which is rather unusual.

It is important to realize that the prerequisites of this course include at least one course in statistics and linear algebra respectively. Only such a prerequisite makes it possible to limit adjustments to one course i.e., 45 lecture hours, and yet present a complete derivation of the most important models, study their use in carefully selected laboratories, and cover the necessary statistical testing procedures. In

fact, the derivations make total use of matrices which provide a clear and easily comprehensible means of derivation. In order to refresh the students' knowledge on statistics and linear algebra, the basic concepts are reviewed during the various phases of the course. However, such a review cannot be, and by no means is intended to be a substitute for a separate course.

As for the review of linear algebra it is assumed that the student is familiar with the essential matrix operations. In a minor attempt the concepts of vector spaces are reviewed in order to give occasionally a geometric representation of the principles of least squares estimation. But in general, the "vector space methods" are de-emphasized for the benefit of the theory of matrices and determinants. The review includes a thorough discussion of quadratic forms and eigenvectors and eigenvalues. The Cholesky algorithm for inverting symmetric and positive definite matrices is discussed in detail so that the students are able to program the algorithm. This section also includes a discussion on the linearizations of multi-dimensional functions, which is of particular importance since most models in Surveying Engineering are inherently non-linear.

In the first part of the review of statistics basic concepts such as random variables, probability density functions, cumulative distribution functions, mathematical expectations, mean and variance are discussed. The review is extended to multivariate distributions, variance-covariance matrices and correlation matrices. A key part of this review is the derivation of the law of variance-covariance propagation in matrix notation for linear functions of random variables.

Finally, some of the characteristics of estimation (unbiased, minimum variance) are examined. Note that thus far, no specific distribution has been examined.

By far the most extensive portion of the course is devoted to the derivation of the various adjustment models. The standard methods of observation equations, condition equations and the mixed model, in which the observations and parameters are related implicitly, are treated exhaustively. Then follows a discussion of the cases of conditions between parameters, sequential solutions and weighted (observed) parameters. In the latter instances the derivations are carried out for the mixed model, and the expressions for the cases of observation equations and condition equations are formally obtained by specifying certain matrices. The iterations which are necessary due to the non-linearity of the mathematical model are explained for each case. In laboratories all adjustments have to be iterated until satisfactory conversion occurs.

In the second part of the review of statistics those elements which pertain to post-adjustment hypothesis testing are exposed. The review includes normal distributions, t-distribution, F-distribution, the principles of hypothesis testing and confidence intervals including Type I and Type II errors. The goodness-of-fit test and the distribution-free sign test are also part of the review. Particularly important for post-adjustment analyses are the multivariate normal distribution and the distribution of certain quadratic forms. An example is the test on the a-posteriori variance of unit weight which is an indicator of the distortion of the adjustment and which is based on a quadratic form.

Applying this test systematically to sequential solutions provides a powerful tool for the detection of blunders and even systematic errors. Next, the generalized linear hypothesis is discussed in detail which permits statistical tests on the parameters. This portion of the course is concluded with a discussion of error ellipses and aspects of the analysis of adjusted residuals.

Horizontal network adjustment based on distance, direction and angular measurements, adjustments of level networks and station directions, all of which occupy a large portion of the laboratory program, are systematically used to introduce the student to the concepts of minimal and inner constraints. The course contents includes an introductory discussion on generalized inverses and estimable functions. However, most of the emphasis is placed on the laboratory work, not on the teaching of some beautiful generalized theory. The latter would be more appropriate for a graduate course.

Laboratories are an indispensible part of this course. There is one extensive laboratory each week. Laboratories are intended to supplement the lecture and to provide an opportunity to use the various adjustment models and techniques in solving typical problems from the general area of Surveying Engineering. The laboratories generally require extensive computer programming in order to handle practical and relevant examples. Each student is given free computer time and can use in-house terminals. In an effort to lessen the programming load the sub-routines for elementary matrix operations are provided.

This course has been developed essentially on the basis of class notes of adjustments and statistics courses taken by the author while a

student at Ohio State University. The notation used for representing the mathematical expressions of the adjustment models was the one used by Prof. U.A. Uotila, Department of Geodetic Science, Ohio State University, at the time when the course was taken.

June, 1980 A. Leick

TABLE OF CONTENTS

APPENDIX

1. Review of Some Concepts from Linear Algebra

This section contains some of the elements of linear algebra which are needed in adjustments. In general, proofs are not given here. The student may find the proofs and a more detailed exposition of linear algebra in the standard literature, e.g., Lipschutz (1968), or Graybill (1969). A summary on matrix operations can also be found in Mikhail (1976, Appendix A).

1.1 Matrix Operations

Definitions: A matrix is an array of real numbers which are subdivided into rows and columns:

$$ {}^{n}A_{u} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1u} \\ a_{21} & a_{22} & \cdots & a_{2u} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{n1} & a_{n2} & \cdots & a_{nu} \end{pmatrix} = \left(A_1 A_2 \cdots A_u\right) \qquad (1.1) $$

a_{ij} is an element of the matrix A with the location at the ith row and the jth column. n and u denote the total number of rows and columns respectively. If n=u, the matrix is called a square matrix, and if n≠u, the matrix is called a rectangular matrix. A square matrix for which $a_{ij}=0$ for all $i \neq j$ is called a diagonal matrix. If, in addition, all diagonal elements are the same, we speak of a scalar matrix. In the case that the diagonal elements are unity, the matrix is called an identity matrix or unit matrix. Finally, a matrix is called symmetric if $a_{ij} = a_{ji}$ for all i and j.

Matrix Transpose: If B is the transpose of A, i.e.,

$$ B = A^T $$

then $b_{ij} = a_{ji}$ for all i and j.

The transpose of a sum or product of matrices is respectively,

$$ (A + B)^T = A^T + B^T $$
$$ (AB)^T = B^T A^T $$

It is obvious that $(A^T)^T = A$. It follows, that if B is symmetric, then the products

$$ABA^T \text{ and } A^TBA$$

are also symmetric.

Orthogonal Matrix: Let P be an n x n square matrix. P is defined to be an orthogonal matrix if and only if $P^{-1} = P^T$.

Linear dependency: The columns of a matrix are said to be linearly independent if there is no set of scalars $\alpha_1, \alpha_2, \ldots \alpha_u$, not all zero, such that

$$\sum_{1}^{u} \alpha_i A_i = 0 \qquad (1.2)$$

Determinants: Each u x u square matrix A has a uniquely defined scalar which is called the determinant of A. The following notation is used for the determinant:

$$\det A, \text{ or } |A|$$

u is the order of the determinant. If we have a 1x1 matrix then the determinant is equal to the matrix element,

$$1A_1 = a_{11} \qquad (1.3)$$

$$|A| = a_{11}$$

In general, the determinant of a u x u matrix is expressed in terms of determinants of submatrices of size (u - 1) x (u - 1), etc. Consider the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1u} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2u} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ a_{u1} & a_{u2} & & \ldots & a_{uu} \end{pmatrix}$$

For each element we can compute the minor, which is equal to the determinant of A after the respective row and column were deleted. The minor for i=1 and j=2 is,

$$m_{12} = \begin{vmatrix} a_{21} & a_{23} & \cdots & a_{2u} \\ a_{31} & a_{33} & \cdots & a_{3u} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{u1} & a_{u2} & \cdots & a_{uu} \end{vmatrix} \tag{1.4}$$

The cofactor c_{ij} is equal to plus or minus the minor depending on the subscripts i and j:

$$c_{ij} = (-1)^{i+j} m_{ij} \tag{1.5}$$

Finally, the determinant of A can be expressed as

$$|A| = \sum_{j=1}^{u} a_{kj} c_{kj} \tag{1.6}$$

The subscript k is fixed in equation (1.6) but can be any value between 1 and u; i.e., the determinant can be computed based on the minors for any of the u rows.

The determinant of a matrix and its transpose are the same, i.e., $|A| = |A^T|$. Also, the determinant of a matrix is not changed if the elements of the i-th column are multiplied by a scalar and the result is added to the corresponding elements of the h-th column, $h \neq i$. It follows that the determinant of a matrix is zero if the columns are linearly dependent. In that case the matrix is said to be singular.

In the case of u=2 the determinant is

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{1.7}$$

Rank of a Matrix: The rank of a matrix is the order of the largest non-singular square submatrix (largest non-zero determinant) that can be found.

For finding the rank of a matrix it is useful to know that elementary row and column operations do not change the rank of a matrix, such as interchanging two rows (columns), multiplying a row (column) by a constant, and adding one row (column) to another.

The following are some rules pertaining to the rank of matrices:

a) The rank of an n x u rectangular matrix A with u<n is

$$R({}_nA_u) \leq u$$

b) The rank of a product of matrices cannot exceed the rank of the matrix with lowest rank,

$$R(A_1B_1 \ldots D) \leq \min [R(A), R(B), \ldots R(D)]$$

c) Multiplication of a matrix by a non-singular matrix does not change the rank.

d) If $R({}_nA_u) = u$ and $R({}_uB_n) = u$ then $R(AB) = u$.

Trace of a Matrix: The trace of a square matrix is

$$Tr({}_uA_u) = \sum_{i=1}^{u} a_{ii} \qquad (1.8)$$

The following relations for a trace of a square matrix can easily be verified:

a) $Tr(A^T) = Tr(A)$

b) $Tr(A + B) = Tr(A) + Tr(B)$

c) $Tr({}_pE_q\ {}_qF_p) = Tr({}_qF_p\ {}_pE_q)$

d) If $\lambda_1, \lambda_2 \ldots \lambda_u$ are the eigenvalues of a u x u matrix A, then

$$Tr(A) = \sum_{i=1}^{u} \lambda_i$$

and

$$Tr(A^{-1}) = \sum_{i=1}^{u} \frac{1}{\lambda_i}$$

Algebra of Matrices and Determinants: Here only a few remarks regarding the multiplication of matrices will be made. The multiplication of matrices is defined by

$$ {}_mA_k \; {}_kB_n = {}_mC_n \tag{1.9} $$

where the elements are related as

$$ c_{ij} = \sum_{r=1}^{r=k} a_{ir} b_{rj} \tag{1.10} $$

We observe that the number of columns of the first matrix must be equal to the number of rows of the second matrix, and that the size of the new matrix is determined by the number of rows of the first matrix and the number of columns of the second matrix. The following relationships hold:

$A(BC) = (AB)C = ABC$ associative law

$A(B+C) = AB + AC$
$(A+B)C = AC + BC$ } distributive law

$AB \neq BA$ in general: commutative law

It is important to note that the product of two matrices can be zero,

$$ {}_rA_u \; {}_uB_p = {}_nO_p \tag{1.11} $$

although neither A or B is the zero matrix. In fact, if the rank of A is $r \leq u$, then there always exists a matrix B of rank u-r such that the product is zero. The rank of B cannot exceed u-r. If A is a square matrix, then there exists a non-zero matrix B such that AB=0 only if A has a rank defect. See also the section on vector spaces.

Finally, the determinant of the product of two square matrices is equal to the product of the determinants of each matrix:

$$ |{}_nA_n \; {}_nB_n| = |A| \cdot |B| \tag{1.12} $$

The algebraic operations explained above can be performed on submatrices as if they were elements of a matrix, provided precautions are given to the dimensions. An example of partitioning is

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \cdots & a_{1u} \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots & a_{2u} \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots & a_{3u} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} & \cdots & a_{nu} \end{pmatrix} = \begin{pmatrix} {}_{2}A_{11_{2}} & {}_{2}A_{12_{u-2}} \\ {}_{n-2}A_{21_{2}} & {}_{n-2}A_{22_{u-2}} \end{pmatrix}$$

$$= \begin{pmatrix} {}_{2}B_{1_{u}} \\ {}_{n-2}B_{2_{u}} \end{pmatrix} = \begin{pmatrix} {}_{n}C_{1_{2}} & {}_{n}C_{2_{u-2}} \end{pmatrix}$$

The manner in which a matrix is being partitioned depends on the circumstances of the particular problem. In order to demonstrate the multiplication of submatrices two examples are given:

Example 1: The matrices A and B are partitioned as follows:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$$

$$\text{then} \quad AB = \begin{pmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{pmatrix}$$

Example 2: If $A = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}$ and $B = \begin{pmatrix} B_1 B_2 \end{pmatrix}$

$$\text{then} \quad AB = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \begin{pmatrix} B_1 B_2 \end{pmatrix} = \begin{pmatrix} A_1B_1 & A_1B_2 \\ A_2B_1 & A_2B_2 \\ A_3B_1 & A_3B_2 \end{pmatrix}$$

It is emphasized again that the partitioning has to be such as not to violate the rules of dimensioning as expressed in equation (1.9).

1.2 Vector Spaces

This section contains a summary of definitions and theorems on vector spaces as can be found in Graybill (1969, Sections 2 and 5).

Definition: Let V_n be a set of n-dimensional vectors such that for every two vectors in V_n, the sum of the two vectors is also in V_n, and for each vector in V_n and each scalar, the product is in V_n. This set V_n is called a vector space.

The vector space R_n: Let R_n be the set of all n x 1 vectors for a fixed positive n, that is

$$R_n = \{Y : Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \quad -\infty < y_i < \infty, \; i = 1, 2 \ldots n\}$$

then R_n is a vector space. For n=3, R_3 is the space we generally think of in three dimensional geometry.

Subspace: Let S_n be a subset of vectors in the vector space V_n. If the set S_n is itself a vector space, then S_n is called a subspace of the vector space V_n.

Basis: Let V_n be a vector space. If each vector in V_n can be obtained by a linear combination of the vectors $\{A_1 A_2 \cdots A_r\}$ then the set $\{A_1 A_2 \cdots A_r\}$ is said to generate (or span) V_n. If, in addition, the vectors $\{A_1 A_2 \cdots A_r\}$ are linearly independent then the set is called a basis for V_n. There are many bases for a given vector space. Recall that any two linearly independent vectors in a plane form a basis for that plane. The number of vectors in any basis for V_n is unique. Let r be the number of base vectors for V_n, then r is called the dimension of V_n.

Example: Let V_3 be a vector space spanned by

$$A_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$$

Then the dimension of V_3 is 2 since the rank of $A = (A_1A_2A_3A_4)$ is two. Each of the sets (A_1A_2), (A_1A_3), (A_1A_4), (A_2A_3), and (A_2A_4) is a basis for V_3.

Generally, if $r>0$ is the rank of the matrix of the vectors $A_1 \cdots A_n$ that span the vector space V_n, then there are exactly r linearly independent vectors in the set, and every vector in V_n can be expressed uniquely as a linear combination of these r vectors.

It follows that R_n, as defined above, has the dimension n since all n x 1 vectors contain exactly n linearly independent vectors. For example, a basis for R_n is

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad \cdots \quad A_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

If the basis vectors are mutually orthogonal, that is the inner product $A_i^T A_j = 0$ is zero for all $i \neq j$, then the basis is called orthogonal. If, in addition, $A_i^T A_i = 1$ for all i, then the basis is called orthonormal.

Every vector space V_n has an orthogonal base.

Euclidian Space: The vector space R_n is defined to be the Euclidian Space, denoted by E_n, if the distance between any two points (vectors) A and B in R_n is defined to be

$$d = \left[\sum_1^n (a_i - b_i)^2 \right]^{\frac{1}{2}} = \left[(A-B)^T (A-B) \right]^{\frac{1}{2}} \qquad (1.13)$$

Intersection of two vector spaces: Let S_1 and S_2 be two vector subspaces in E_n. The intersection of these two subspaces, which we denote by $S = S_1 \cap S_2$, is defined as the set of vectors A that belong to both S_1 and S_2, that is

$$S = \{A : A \in S_1, A \in S_2\}$$

Sum of Vector Subspaces: Let S_1 and S_2 be two subspaces of the vector space E_n. The set of vectors S, denoted by $S = S_1 \oplus S_2$, is called the sum of the vector spaces S_1 and S_2 and is defined by

$$S = \{A : A = B_1 + B_2, B_1 \in S_1 \text{ and } B_2 \in S_2\}$$

S is a vector subspace of E_n.

Orthogonal Vector Subspace in E_n: Let S_1 and S_2 be two subspaces in E_n. If $A^T B = 0$ for each vector A in S_1 and for each vector B in S_2, then S_1 and S_2 are defined to be orthogonal subspaces in E_n and we denote this by $S_1 \perp S_2$.

Orthogonal Complement of a Vector Subspace in E_n: Let S_1 be a vector subspace in E_n. The vector subspace S_2 in E_n is defined as the orthogonal complement of S_1 in E_n if and only if $S_1 \perp S_2$ and $S_1 \oplus S_2 = E_n$.

We sometimes denote the orthogonal complement of a subspace S_1 by $S_1^\perp$. For a given subspace S_1 in E_n, the orthogonal complement $S_1^\perp$ always exists and is unique.

Let S_1 be a subspace of E_n and let L be any vector in E_n, then L can be written as a sum of two vectors $L = \hat{Z} + V$ where $\hat{Z}$ is in S_1 and V is in the orthogonal complement of S_1.

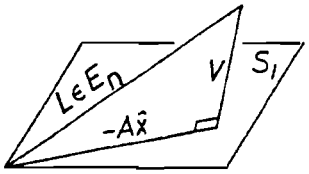


Figure 1.1 Example for Geometric Representations in Estimation Theory.

The subspace S_1 is a plane in E_n of dimension $R(A) = r \leq u$, i.e.,

$$S_1 = \{Z : Z = {}_nA_u X, \; A = (A_1 A_2 \cdots A_u); \text{ for all } {}_uX_1\}$$

$\hat{Z}$ is the orthogonal projection of L (line in E_n through origin O) onto the subspace S_1, which is spanned by the columns of the A matrix. The orthogonal projection of L onto the r-dimensional subspace is unique, as can be seen from simple geometric considerations. In adjustments we will determine a set $\hat{X}$ such that

$$\hat{Z} = A \hat{X} \tag{1.14}$$

It may be mentioned already at this point, that there is only a unique $\hat{X}$ fulfilling equation (1.14) if $R({}_nA_u) = u$. In all other cases a minimum number of u-R(A) conditions will have to be used to find an $\hat{X}$. It is clear that the resulting $\hat{X}$ depends on the conditions imposed.

We observe that V is completely contained in the n - R(A) dimensional orthogonal complement of S_1. Moreover, the length of V, i.e.,

$$d = V^T V$$

is a minimum and is unique for the given subspace S_1 as can be seen from simple geometric considerations.

Column Space of a Matrix: Let A be an n x u matrix. We denote the u columns of A as vectors in E_n, so that $A = (A_1 A_2 \cdots A_u)$. The vector space spanned by these u column vectors of A is called the column space of A. The dimension of the column space is equal to the rank of A, that is the number of linearly independent columns. The column space of A and the column space of AA^T are the same.

Null Space of a Matrix: Let A be an n x u matrix. The null space of the matrix A is defined to be the set of vectors S where

$$S = \{Y : AY = 0, \; Y \in E_u\}$$

The null space of an n x u matrix A is a vector subspace of E_u with dimension u - R(A).

The null space of A^T and the orthogonal complement of the column space of A are the same.

1.3 Quadratic Forms:

Let X be a u x 1 vector and A a u x u symmetric matrix. Then

$$ {}_1V_1 = X^T A X $$

is called a quadratic form

If $A = A^T$

and $X^T A X \geq 0$

then the matrix A and the quadratic form are called positive semi-definite.

If $A = A^T$

and $X^T A X > 0$

then the matrix A and the quadratic form are called positive definite.

The following properties hold for an u x u positive definite matrix A:

a) $R(A) = u$ (full rank)

b) $a_{ii} > 0$ for all i

c) A^{-1} is positive definite.

d) Let B be an n x u matrix with rank u < n. Then the matrix $B^T AB$ is positive definite. If $R(B) = r < u$ then $B^T AB$ is positive semi-definite.

e) Let D be a q x q matrix formed by deleting u - p rows and the corresponding u - p columns of A. Then D is positive definite.

Necessary and sufficient conditions for a symmetric matrix to be positive definite are

a) Every principal minor determinant is positive

$$ a_{11} > 0, \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0, \cdots |A| > 0 $$

b) The eigenvalues of A are all real and positive.

c) There exists a u x u matrix D such that $D^T D = A$ (see next section).

1.4 Eigenvalues and Eigenvectors

Consider the equation

$$_uA_{uu}X_1 = \lambda_u X_1 \tag{1.16}$$

A is a u x u matrix. The u x 1 vector X is called the eigenvector, and the scalar λ is the eigenvalue. The equation (1.16) can be rewritten as

$$(A - \lambda I)\, X = 0 \tag{1.17}$$

This is a homogeneous equation. If X_o is a solution of (1.17) then $\alpha\, X_o$ is also a solution, where α is a scalar. Thus, the solution of equation (1.17) provides only the direction of the eigenvector. There exists a non-trivial solution for X if

$$|A - \lambda I| = 0 \tag{1.18}$$

This is the characteristic equation. It is a polynomial of the u-th order in λ providing u solutions λ_i, i = 1 ··· u. Some of the eigenvalues can be zero, equal (multiple solution), or even complex.

In adjustments we are mainly interested in eigenvalues of symmetric and positive definite matrices. For a symmetric matrix all eigenvalues are real. There may be multiple solutions. The number of zero eigenvalues equals the rank defect. Thus, computing the eigenvalues provides one method of finding out the rank defect of a matrix. For a symmetric matrix the eigenvectors are mutually orthogonal, that is

$$X_i X_j = 0. \tag{1.19}$$

The eigenvectors are, therefore, suitable for spanning the column space of the matrix. The eigenvectors corresponding to the non-zero eigenvalues span the subspace in E_u with dimension r=R(A). The eigenvectors corresponding to the zero eigenvalues span the orthogonal complement of the vector subspace.

If the matrix is positive definite then all eigenvalues are positive. For a positive semidefinite matrix, all nonzero eigenvalues are positive.

In actual computations the eigenvalues are computed first from equation (1.18), and then the eigenvector can be computed for each eigenvalue from equation (1.17). The eigenvectors can be normalized as follows:

$$E_i = \frac{1}{|X_i|} X_i$$

with $$|X_i| = \left[\sum_i^u x_i^2 \right]^{\frac{1}{2}}$$

If we combine all u normalized eigenvectors to form a matrix E_1

$$E = (E_1 E_2 \cdots E_u)$$

then E is an orthonormal matrix, or a rotation matrix, for which

$$E^T = E^{-1} \tag{1.20}$$

holds.

The Use of Eigenvalues and Eigenvectors in Adjustments:

1) Using the relations (1.16) and (1.19) we can write the product AE as

$$AE = \left(AE_1\ AE_2 \cdots AE_u\right)$$

$$= \left(\lambda_1 E_1\ \lambda_2 E_2 \cdots \lambda_u E_u\right)$$

$$= \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1u} \\ e_{21} & e_{22} & \cdots & e_{2u} \\ \vdots & \vdots & \cdots & \vdots \\ e_{u1} & e_{u2} & \cdots & e_{uu} \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_u \end{pmatrix}$$

$$= E \Lambda$$

where Λ is a diagonal matrix with λ_i as elements at the diagonal. Multiplying this equation by E^T from the left and making use of equation (1.20) gives

$$E^T A E = \Lambda \tag{1.21}$$

Taking the inverse of both sides gives

$$E^T A^{-1} E = \Lambda^{-1} \tag{1.22}$$

Both relations are frequently used in adjustments. Equation (1.21) simply states that if a matrix A is premultiplied by E^T and postmultiplied by E, where the columns of E are the normalized eigenvectors, then the product is a diagonal matrix whose diagonal elements are the eigenvalues of A.

From equation (1.21) we see that

$$\Lambda^{-\frac{1}{2}} E^T A E \Lambda^{-\frac{1}{2}} = I,$$

or $$D^T A D = I \qquad (1.23)$$

where $D = E\Lambda^{-\frac{1}{2}}$

is a nonsingular matrix.

2) The eigenvectors and eigenvalues allow an important geometrical interpretation with respect to quadratic forms. If ${}_uA_u$ is positive definite then

$$X^T A X = V$$

is the equation of a u-dimensional ellipsoid expressed in a Cartesian coordinate system (X). The center of the ellipsoid is at X = 0. Transforming (rotating) the coordinate system (X) by

$$X = EY \qquad (1.24)$$

gives for the quadratic form the expression

$$Y^T E^T A E Y = V,$$

or

$$Y^T \Lambda Y = V \qquad (1.25)$$

The latter expression can be written as

$$y_1^2 \lambda_1 + y_2^2 \lambda_2 + \cdots y_u^2 \lambda_u = V,$$

or

$$\frac{y_1^2}{\frac{V}{\lambda_1}} + \frac{y_2^2}{\frac{V}{\lambda_2}} + \cdots\cdots + \frac{y_u^2}{\frac{V}{\lambda_u}} = 1 \qquad (1.26)$$

But this is the equation for the u-dimensional ellipsoid in the principal axes form, i.e., the coordinate system (Y) coincides with the principal axes of the ellipsoid, and the lengths of the principal axes are proportional to the reciprocal of the square root of the eigenvalues. The situation is demonstrated in the following example:

Let A be a 2 x 2 positive definite matrix

$$A = \begin{pmatrix} 6 & 2 \\ 2 & 2 \end{pmatrix}$$

then $X^TAX = V$ is the equation of an ellipse. The characteristic equation $|A - \lambda I| = 0$ has the solutions

$$\lambda_1 = 4 + \sqrt{8}$$
$$\lambda_2 = 4 - \sqrt{8}$$

The normalized eigenvectors are

$$E_1 = C_1 \begin{pmatrix} 2 + \sqrt{8} \\ 2 \end{pmatrix} \text{ and } E_2 = C_2 \begin{pmatrix} 2 - \sqrt{8} \\ 2 \end{pmatrix}$$

where the constants C_1 and C_2 are chosen accordingly. Thus, the equation of the ellipse in the coordinate system (X) and (Y) is:

$$X^TAX \equiv 6x_1^2 + 4x_1x_2 + 2x_2^2 = V$$

$$Y^T\Lambda Y \equiv \frac{y_1^2}{\frac{1}{4 + \sqrt{8}}} + \frac{y_2^2}{\frac{1}{4 - \sqrt{8}}} = V$$

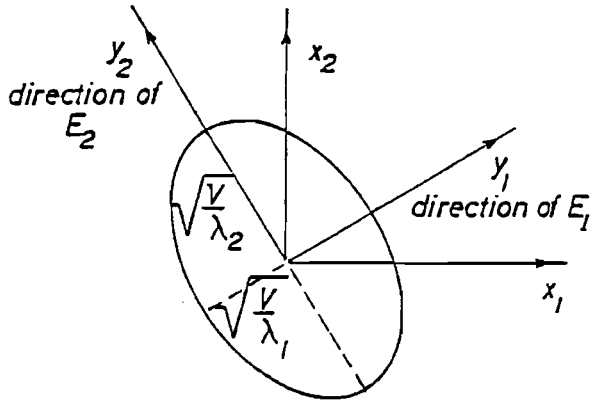


Figure 1.2 Geometric Significance of Eigenvectors and Eigenvalues.

3) If the u x u matrix A is positive semidefinite with R(A) = r<u, equations similar to (1.21) and (1.26) can be found. The matrix E is of size n x r, and Λ is an r x r diagonal matrix containing the non-zero eigenvalues at the diagonal; and the ellipsoid, as expressed by the quadratic form, has the dimension r.

In general, for any positive semi-definite matrix ${}_uA_u$ with $R(A)=r<u$, we can find a matrix $\overline{D}$ such that

$$\overline{D}^T A \overline{D} = \begin{pmatrix} {}_rI_r & 0 \\ 0 & 0 \end{pmatrix} \tag{1.27}$$

where the identity matrix is of the order r. Consider the matrix $({}_uF_r \quad {}_uG_{u-r})$, where the columns of G span the null space of $A = A^T$, i.e., $AG = 0$, and F is chosen such that the matrix $(F \vdots G)$ is non-singular. One obvious choice for F is to use the r eigenvectors of A as columns, although this is not the only possible choice. Any set of r vectors spanning the column space of A will be sufficient. For applications in Surveying, a matrix G can usually be found simply by inspection of the design matrix. This aspect of choosing G will be discussed in a later section. With the stated properties of F and G we obtain

$$\begin{pmatrix} F^T \\ G^T \end{pmatrix} A \quad (FG) = \begin{pmatrix} F^TAF & 0 \\ 0 & 0 \end{pmatrix}$$

This matrix is of rank r and, therefore, F^TAF is an r x r nonsingular submatrix. If F consists of the eigenvectors of A then $F^TAF = \Lambda$ is a diagonal matrix with r non-zero eigenvalues at the diagonal. The matrix $\overline{D}$ becomes

$$\overline{D} = (F \Lambda^{-\frac{1}{2}} \vdots G)$$

In the case that F has been chosen differently, we can find a matrix $\overline{E}$ such that

$$\overline{E} (F^TAF) \overline{E} = \overline{\Lambda},$$

where $\overline{E}$ is the eigenvector matrix of F^TAF and $\overline{\Lambda}$ is the diagonal matrix with the corresponding eigenvalues. Finally,

$$\overline{\Lambda}^{-\frac{1}{2}} \overline{E}^T (F^TAF) \overline{E} \overline{\Lambda}^{-\frac{1}{2}} = I,$$

so that

$$\overline{D} = (F\overline{E} \overline{\Lambda}^{-\frac{1}{2}} \vdots G)$$

is a non-singular matrix.

1.5 Inverse of a Matrix (Caley-Inverse)

There are several techniques for computing an inverse of a matrix. In this section we will only discuss the inversion of non-singular square matrices, i.e., $|A| \neq 0$ (Caley-Inverse). Such inverses have the property,

$$A A^{-1} = I$$
$$A^{-1} A = I$$

It is sometimes useful to know that

$$(AB)^{-1} = B^{-1} A^{-1},$$

where both A and B are nonsingular.

1.5.1 Inverse by Adjoint Matrix

Replace each element a_{ij} of A by the cofactor c_{ij}. The cofactor has been explained in Section (1.1). Thus we have a matrix,

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1u} \\ c_{21} & & & \vdots \\ \vdots & & & \vdots \\ c_{ul} & & \cdots & c_{uu} \end{pmatrix}$$

The adjoint matrix, denoted by adj A, is,

$$\text{adj } A = C^T$$

The inverse becomes,

$$A^{-1} = \frac{\text{adj } A}{|A|} = \frac{C^T}{|A|} \tag{1.28}$$

The method of inverting a matrix is explained in any course on linear algebra. The proof of equation (1.28) can therefore be found in the standard literature. This method is applicable to a non-singular square matrix. However, since this method requires a large computational effort other methods are usually preferred.

For the 2 x 2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

we find

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix}$$

That A^{-1} is indeed the inverse of A can be verified by straight forward matrix multiplication.

1.5.2 Inverse by Matrix Partitioning

Partition the nonsingular square matrix N, which is to be inverted, as follows:

$$N = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix}, \qquad (1.29)$$

where N_{11} and N_{22} are square matrices, although not necessarily of the same size. Denote the inverse matrix by Q and partition it accordingly, i.e.,

$$Q = N^{-1} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \qquad (1.30)$$

so that the size of N_{11} and Q_{11}, N_{12} and Q_{12}, etc., are respectively the same. Since Q is the inverse of N the following equations must be fulfilled:

$$N\,Q = I \qquad (1.31)$$

and

$$Q\,N = I \qquad (1.32)$$

Each of the equations (1.31) and (1.32) is itself a set of 4 equations. Each set can be solved for the unknown submatrices (Q_{11}, Q_{12}, Q_{21}, Q_{22}) in terms of the known submatrices (N_{11}, N_{12}, N_{21}, N_{22}). Thus, we find two solutions for Q, each solution being a different function of the submatrices N_{ij}. The actual solution for the submatrices Q_{ij} is carried out according to the standard rules for solving a system of linear equations, with the restriction that the inverse is defined only for square submatrices. The solution is

$$Q = \left(\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right)$$

$$= \left(\begin{array}{c|c} N_{11}^{-1} + N_{11}^{-1} N_{12} (N_{22}-N_{21}N_{11}^{-1}N_{12})^{-1} N_{21} N_{11}^{-1} & -N_{11}^{-1}N_{12}(N_{22}-N_{21}N_{11}^{-1}N_{12})^{-1} \\ \hline -(N_{22}-N_{21}N_{11}^{-1}N_{12})^{-1}N_{21}N_{11}^{-1} & (N_{22}-N_{21}N_{11}^{-1}N_{12})^{-1} \end{array}\right)$$

$$= \left(\begin{array}{c|c} (N_{11}-N_{12}N_{22}^{-1}N_{21})^{-1} & -(N_{11}-N_{12}N_{22}^{-1}N_{21})^{-1}N_{12}N_{22}^{-1} \\ \hline -N_{22}^{-1}N_{21}(N_{11}-N_{12}N_{22}^{-1}N_{21})^{-1} & N_{22}^{-1}+N_{22}^{-1}N_{21}(N_{11}-N_{12}N_{22}^{-1}N_{21})^{-1}N_{12}N_{22}^{-1} \end{array}\right) \qquad (1.33)$$

We observe that through systematic application of the matrix partitioning technique the inverse can be computed by multiplication only and by taking the inverse of 1 x 1 matrices. Usually the above technique is used in order to reduce the size of large matrices which should be inverted. The actual inversions of the smaller submatrices is done with other methods.

The respective submatrices in equation (1.33) have to be the same since there is only one inverse Q. These identities will be used frequently in Section 3. It is therefore recommended that the students fully understand the derivation of equation (1.33).

1.5.3 The Algorithms of Gauss and Cholesky.

The Gauss algorithm for solving a system of linear equations and inverting matrices is assumed to be known to the student from courses in linear algebra. The solution is simply found through elementary row and column operations. A complete description of the methods involved can be found, e.g., in Graybill (1969, p. 276 and 289). The Gauss algorithm is very efficient for inverting large matrices. It involves fewer multiplications than the method of computing the adjoint matrices.

The Cholesky algorithm, sometimes called the "Square Root Method", is a modification of the Gauss algorithm. This procedure is tailored to positive definite matrices. If it is used for computing the inverse of a non-positive definite matrix imaginary

numbers might occur. A description of the Square Root Method can be found again in Graybill (1969, p. 298). Since in adjustments mostly positive definite matrices have to be inverted, the principle steps for using the Square Root Method are repeated here. In those rare instances, as e.g., in Section 3.5, where non-positive definite matrices may occur, any of the previously discussed methods can be used.

Assume the system of equations

$$ {}_uN_u \; {}_uX_1 = {}_uU_1 \tag{1.34} $$

has to be solved. N is a positive definite matrix. We are also interested in the quantity

$$ v = U^T N^{-1} U \tag{1.35} $$

Since N is positive definite we can write

$$ N = L\,L^T \tag{1.36} $$

where L is a lower triangular matrix,

$$ L = \begin{vmatrix} l_{11} & & & \\ l_{21} & l_{22} & & 0 \\ l_{31} & l_{32} & \ddots & \\ \vdots & \vdots & & \\ l_{u1} & l_{u2} & \cdots & l_{uu} \end{vmatrix} $$

Multiplication of (1.34) from the left by L^{-1} gives

$$ L^{-1} N X = L^{-1} U $$

Substitution of expression (1.36) for N results in

$$ L^T X = L^{-1} U \tag{1.37} $$

The computation of the elements of L and $L^{-1}U$ is referred to as the "reduction of the equation system (1.34)" and the solution of the system (1.37) is referred to as the "back-solution".

The elements of L are

$$ l_{jk} = \left[n_{jk} - \sum_{m=1}^{k-1} l_{jm}\, l_{km} \right] / l_{kk} \qquad k<j $$

$$ l_{jj} = \sqrt{n_{jj} - \sum_{m=1}^{j-1} l_{jm}^2} \qquad k=j \tag{1.38} $$

n_{jk} is, of course, an element of N.

The complete solution involves the following steps:

1) Compute the elements l_{jk} from the elements l_{nm} with n<j and m<k.
2) Compute l_{jk} either column-wise or row by row.
3) As soon as l_{jk} is computed it can be stored in the position n_{jk} in the computer since n_{jk} will not be used anymore. This saves storage.
4) The computation of L is now completed. The column $L^{-1}U$ is computed by applying the algorithm (1.38) to the augmented matrix N_o:

$$u+1\ N_{o\,u+1} = \left(\begin{array}{c|c} N & U \\ \hline U^T & d \end{array}\right) \tag{1.39}$$

It is clear that in this step only the last row has to be reduced since the other rows had already been reduced in steps 1) to 3).

d is a fictitious diagonal element. From the algorithm (1.38) we find that

$$l^2_{u+1,\,u+1} = d - \sum_{m=1}^{u} l^2_{u+1,m}$$

$$= d - U^T(L^T)^{-1}L^{-1}U$$

$$= d - U^T N^{-1} U$$

Hence, if we choose d=0, then

$$l^2_{u+1,u+1} = -U^T N^{-1} U \tag{1.40}$$

is a quantity which is needed in adjustments.

So far we have computed

$$\left(\begin{array}{c|c} L & \\ \hline (L^{-1}U)^T & -U^T N^{-1} U \end{array}\right) \tag{1.41}$$

and the back solution of (1.37) can be started.

5) Note that the vector U in (1.39) can be replaced by any other "right hand side". Only the reduction of the last row in (1.39) has to be done again.

6) Usually we are not only interested in the solution of equation (1.34) but also in the inverse N^{-1}. The inverse is computed column by column. Replace the U vector in (1.39) by the vector

$$C^{(\alpha)} = \begin{vmatrix} 0 \\ 0 \\ \cdot \\ 0 \\ 1 \\ 0 \\ \cdot \\ 0 \end{vmatrix}$$

which consists of zeros except at the position α where there is a one. Compute $L^{-1}C^{(\alpha)}$ by reducing

$$\left(\begin{array}{c|c} N & C^{(\alpha)} \\ \hline (C^{(\alpha)})^T & \end{array}\right),$$

giving

$$\left(\begin{array}{c|c} L & \\ \hline (L^{-1}C^{(\alpha)})^T & \end{array}\right),$$

and then solve the back solution

$$L^T q^{(\alpha)} = L^{-1}C^{(\alpha)}$$

for $q^{(\alpha)}$ which is the α-th column of N^{-1}.

The following observations are made regarding Cholesky's algorithm:

1) From (1.38) it is seen that the reduction does not have to start before column α since all reduced elements $l_{u+1,i}$ with $i < \alpha$ are zero. Yet, the computation of the inverse requires a much higher computational effort than is needed for the mere solution of (1.34). Since the columns of the inverse can be computed separately, one might, in the case of very large systems, compute only those columns and elements which are really needed. The meaning of this statement will become clear in Section 3 where the inverse N^{-1} is related to the variance-covariance matrix of the adjusted parameters.

2) From (1.38) it is also seen that if N has zeros in the upper part of the columns, the matrix L^T has zeros at the corresponding positions (up to the first non-zero position in the respective column). This fact should be used in any computer programming if the matrix N is diagonally banded in order to avoid too many zero multiplications.

1.6 Matrix Differentiation

Case 1: A matrix is differentiated by differentiating each of the elements:

$$\frac{dA(s)}{ds} = \begin{vmatrix} \frac{da_{11}}{ds} & \frac{da_{21}}{ds} & \cdots & \frac{da_{1u}}{ds} \\ \frac{da_{21}}{ds} & & & \vdots \\ \vdots & & & \\ \frac{da_{n1}}{ds} & & \cdots & \frac{da_{nu}}{ds} \end{vmatrix} \tag{1.42}$$

If the elements of the matrix are a function of several variables, then the partial derivative of the matrix is obtained by taking the partials of each element.

Case 2: The differential of an n-dimensional vector function

$$Y = F(X) = \begin{vmatrix} f_1(X) \\ f_2(X) \\ \vdots \\ f_n(X) \end{vmatrix}$$

where each component is a function of u variables $X = (x_1, x_2, \ldots x_u)$ is

$$dY = \frac{\partial F}{\partial X} dX = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_u} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & & \frac{\partial f_2}{\partial x_u} \\ \vdots & & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_u} \end{vmatrix} \begin{vmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_u \end{vmatrix} \tag{1.43}$$

Denoting the n x u matrix which contains partials by G, then

$$dY = {}_nG_u dX$$

G has as many columns as there are parameters. The number of rows is equal to the dimension of Y.

Case 3: In this case the columns of A are constant.

a) ${}_nY_1 = {}_nA_u{}_uX_1$

The differential is $dY = \frac{\partial(AX)}{\partial X} dX = AdX$,

or

$$\frac{dY}{dX} = \frac{\partial Y}{\partial X} = A \qquad (1.44)$$

b) ${}_1Y_n = {}_1X^T_{nn}A_u$. The differential is

$$dY = dX^TA$$

c) $u = {}_1X^T_uA_uY_1$ Since u is a 1x1 matrix we can write

$$u = X^TAY = Y^TA^TX$$

It follows that the total differential

$$du = \frac{\partial u}{\partial X} dX + \frac{\partial u}{\partial Y} dY$$

becomes

$$du = Y^TA^TdX + X^TAdY \qquad (1.45)$$

since

$$\frac{\partial u}{\partial X} = Y^TA^T \quad \text{and} \quad \frac{\partial u}{\partial Y} = X^TA$$

d) $v = {}_1X^T_uA_uX_1$. The differential of the quadratic form follows from the expression (1.45) by specifying that Y = X and that A be symmetric:

$$dv = 2X^TAdX$$

or

$$\frac{dv}{dX} = \frac{\partial v}{\partial X} = 2X^TA$$

1.7 Linearization

In surveying, the parameters and the observations are usually related by a non-linear function. In order to perform an adjustment these relationships have to be linearized. This is accomplished by expanding the function in a Taylor series and retaining only the linear terms.

Consider the non-linear function

$$y = f(x) \tag{1.47}$$

which has only one variable x. The Taylor series expansion of this function is

$$y = f(x_o) + \left.\frac{dy}{dx}\right|_{x_o} dx + \frac{1}{2!}\left.\frac{d^2y}{dx^2}\right|_{x_o} dx^2 + \ldots. \tag{1.48}$$

The linearized form of (1.47) is

$$\bar{y} = f(x_o) + \left.\frac{dy}{dx}\right|_{x_o} dx \tag{1.49}$$

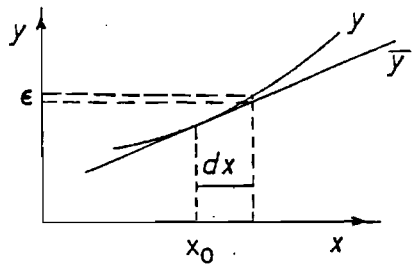


Figure 1.3 Linearization

x_o is called the point of expansion. At that point the linearized and the non-linear functions coincide. If the non-linear relationship is replaced by the linear function then the error $\varepsilon = y - \bar{y}$ increases as the free parameter x departs from the expansion point x_o. Therefore, if we would like the linearized form (1.49) to be a sufficiently accurate representation of the non-linear relation (1.47) for a given

y (observation) and the corresponding x (parameter, i.e., coordinate of a station), we have to make sure that the expansion point x_0 is close to x. Since x is itself an unknown it will be necessary to use iterative procedures to ensure that the point of expansion x_0 converges toward x.

Usually the observation is related to several parameters. The Taylor series of

$$z = f(x, y) \tag{1.50}$$

is

$$z = f(x_0 y_0) + \frac{\partial z}{\partial x}\Big|_{x_0 y_0} dx + \frac{\partial z}{\partial y}\Big|_{x_0 y_0} dy + \varepsilon\,(dx^2, dy^2, dxdy, \ldots)$$

The point of expansion is P ($x=x_0$, $y=y_0$). The linearized form,

$$z = f(x_0 y_0) + \frac{\partial z}{\partial x}\Big|_{x_0 y_0} dx + \frac{\partial z}{\partial y}\Big|_{x_0 y_0} dy \tag{1.51}$$

represents the tangent plane on the surface (1.50) at the expansion point.

In the most general case there are observed n functions with u parameters:

$$Y = F(X) = \begin{pmatrix} f_1(X) \\ f_2(X) \\ \vdots \\ f_n(X) \end{pmatrix} = \begin{pmatrix} f_1(x_1 x_2 \cdots x_u) \\ f_2(x_1 x_2 \cdots x_u) \\ \vdots \\ f_n(x_1 x_2 \ldots x_u) \end{pmatrix}$$

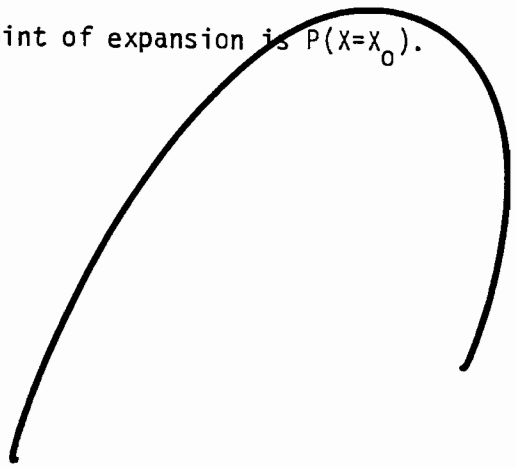
The linearized form is

$$Y = F(X_0) + \frac{\partial F}{\partial X}\Big|_{X_0} dX \tag{1.52}$$

where

$$\frac{\partial F}{\partial X} \equiv {}_{n}G_{u} = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_u} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_u} \\ \vdots & & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_u} \end{vmatrix}$$

The point of expansion is $P(X=X_o)$.



2. Review of Statistical Concepts (I)

The relevancy of statistical concepts with respect to adjustments can be divided into two parts. A first set of concepts is applicable to the set-up of the adjustment itself, whereas a second set is needed in analyzing the results of the adjustment.

In this section only the first set of concepts will be discussed. Some of the important items are the categorization of observational errors, the definition of accuracy and precision, and of mean and variance. Since in adjustments usually several parameters (unknowns) are involved it is necessary to introduce some aspects from multivariate statistics which lead to the concept of the variance-covariance matrix. Essential for adjustments are the techniques of propagating the mean and the variance-covariances. Finally some of the properties in estimation will be discussed in general. In Section 3, when the least squares estimates have been obtained, we will verify some of their properties.

2.1 Some Definitions

Event (≡observation) The outcome of a statistical experiment, e.g., throwing a dice, measuring an angle, a distance, etc.

Random Variable Denoted by $\tilde{x}$ or $\tilde{X}$; the name for the outcome of an event. It takes on the values of the respective outcomes. $\tilde{X}$ denotes a vector of random variables. Note that a linear function of a random variable, such a

$$\tilde{Y} = A\tilde{X},$$

is also a random variable.

Population The totality of all events. It includes all possible values of the random variable. The population is described by a finite set of parameters (population parameters).

Sample A subset of the population. For example, if the same distance is measured 10 times, then these 10 measurements are a sample of all possible measurements.

Statistics The sample is used to compute one or more statistics which represent estimates of the population parameters or functions of the latter. Thus, with the 10 measurements of one and the same distance we compute statistics which are estimates of those parameters which describe in some way the outcome of all possible measurements of the same distance.

Probability Related to the frequency of occurrence of a specific event. Thus, each value of a random variable is associated with a probability.

Probability Distribution describes the various probabilities as related to the possible values of the random variable.

Observational Error The difference between the sample and the population parameter (true value)

Blunder Usually large errors are due to a careless observer; the experiment should be designed such that blunders are discovered. Example: 5m error in distance measurements.

Constant Errors Errors having the same sign (effect) for all observations. Example: a tape is short by 10 cm.

Systematic Errors Errors which vary systematically in sign and/or magnitude. Example: lateral refraction when measuring angles.

Random (Accidental) Errors The probability of a positive or negative error of a given magnitude is the same (equal frequency of occurrence). The error is usually small.

Blunders, constant and systematic errors, can largely be avoided through a careful observer, adequate instrument calibration, and observing under various conditions (weather, etc.). The constant and systematic errors are particularly dangerous since they tend to accumulate. Random errors are unavoidable. In adjustments it is assumed that the only errors present are the random errors.

Independent Sampling: The previous event does not influence the subsequent event.

Histogram In the usual measuring processes in Surveying smaller random errors occur with higher frequencies than larger errors.

Assume that the same distance has been measured n times. Let us plot along the x-axis the numerical value of each of the measurements. Then subdivide the x-axis into intervals of equal length Δx and count the numbers n_i of measurements within each interval. The relative

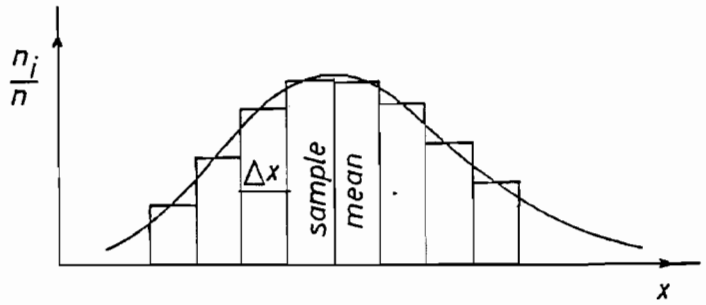


Figure 2.1 Histogram

frequency n_i/n is plotted along the ordinate. Next, a rectangle whose height is a function of the relative frequency, is constructed as shown in Fig. 2.1. This plot is called a histogram. Provided that the measurements are only affected by random error, the smoothed step function of the histogram has a bell-like shape. The maximum occurs around the sample mean which is an estimate of the true length of the distance. The larger the deviation from the sample mean, the smaller is the relative frequency.

It is emphasized again, that not all histograms of any conceivable statistical experiments have the same characteristics as described above. But, repeated measurements of distances, angles, etc. result in such bell-shaped histograms.

Accuracy and Precision Accuracy refers to the closeness of observations to the true value of a quantity. Precision refers to the closeness of repeated observations around the sample mean.

Consider again the smoothed curve of the histogram as a characterization of an observation sequence. In the figure below several such curves are shown.

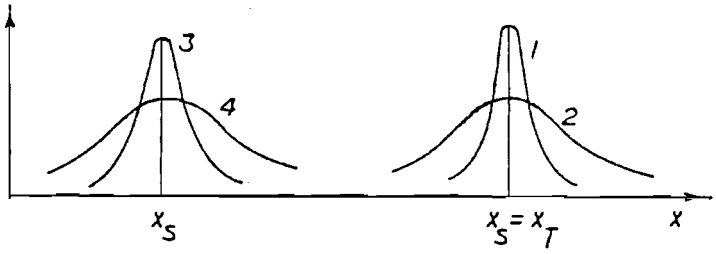


Figure 2.2 Accuracy and Precision

We see that the curves 1 and 2 are symmetric with respect to the true value. We say that the measurements have a high accuracy. But the shape of both curves is quite different. Curve 1 is tall and narrow; i.e., most of the random errors are very small. We say that the measurements of sequence 1 have also a high precision, whereas the precision of sequence 2 is lower since the relative frequency of larger random errors is larger than in case 1. The curves 3 and 4 are symmetric with respect to the sample mean x_s, which is different from the true value x_T. Both sequences have a low accuracy although the precision of sequence 3 is high. The difference $x_T - x_s$ is due to constant and systematic errors (bias). It is clear that an increase in the number of observations does not reduce the bias.

The above definition of accuracy and precision does not only refer to direct observations but also to linear functions of the observations. This is important in adjustments since the least squares estimates are linear functions of the observations.

2.2 Probability Density Function and Cumulative Distribution Function

2.2.1 Discrete Case

Statistical experiments with discrete probabilities do not occur very frequently in Surveying. This case is, therefore, only discussed briefly here.

The probability that x occurs is p_x. We write

$$P(\tilde{x} = x) = p_x \tag{2.1}$$

and define

$$a) \quad p_x \geq 0 \tag{2.2}$$

$$b) \quad \sum_{pop.} p_x = 1 \tag{2.3}$$

The definition of 2.2 is quite logical since a negative probability has no meaning. Either a certain event occurs with a certain finite probability (frequency) or it does not occur. In the latter case the probability is zero. The definition 2.3 says that the sum of all probabilities is one.

As an example for discrete probability consider the throwing of two dice. The random variable $\tilde{x}$ is the sum of the two dice for one throwing. All possible combinations of the two dice constitute the population:

Population = {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

The probability p_x that a certain event occurs, e.g., throwing x = 5, is equal to the number n_x of possible combinations giving this event x, divided by the total number of combinations of the two dice,

$$p_x = \frac{n_x}{\sum n_x}$$

It follows that the probability density function for this experiment is

$$p_x = \begin{cases} 1/36 \text{ if } \tilde{x} = 2,12 \\ 2/36 \text{ if } \tilde{x} = 3,11 \\ 3/36 \text{ if } \tilde{x} = 4,10 \\ 4/36 \text{ if } \tilde{x} = 5,9 \\ 5/36 \text{ if } \tilde{x} = 6,8 \\ 6/36 \text{ if } \tilde{x} = 7 \end{cases}$$

We see that each $p_x \geq 0$ and that the sum of all probabilities is one.

The cumulative distribution function $F(x)$ gives the probability that the random variable is equal to or smaller than x,

$$F(x) = P(\tilde{x} \leq x) \tag{2.4}$$

In the case of a discrete probability density function the probability $P(\tilde{x} \leq x)$ is obtained by simply adding the individual probabilities.

For example

$$P(\tilde{x} \leq 4) = p_2 + p_3 + p_4 = \frac{6}{36}$$

The cumulative distribution function for the experiment of throwing two dice becomes:

$$F(x) = \begin{cases} 0 & \text{if } \tilde{x} < 2 \\ 1/36 & \text{if } 2 \leq \tilde{x} < 3 \\ 3/36 & \text{if } 3 \leq \tilde{x} < 4 \\ 6/36 & \text{if } 4 \leq \tilde{x} < 5 \\ 10/36 & \text{if } 5 \leq \tilde{x} < 6 \\ 15/36 & \text{if } 6 \leq \tilde{x} < 7 \\ 21/36 & \text{if } 7 \leq \tilde{x} < 8 \\ 26/36 & \text{if } 8 \leq \tilde{x} < 9 \\ 30/36 & \text{if } 9 \leq \tilde{x} < 10 \\ 33/36 & \text{if } 10 \leq \tilde{x} < 11 \\ 35/36 & \text{if } 11 \leq \tilde{x} < 12 \\ 36/36 & \text{if } \tilde{x} \geq 12 \end{cases}$$

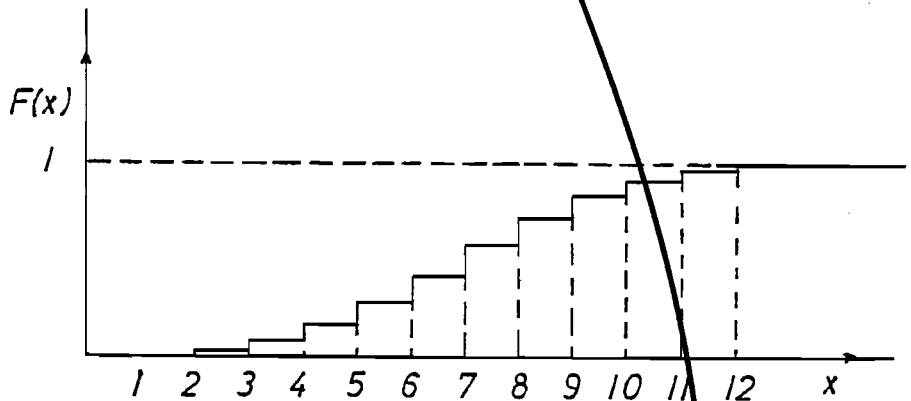


Figure 2.3 Example of Probability Density Function and Cumulative Distribution Function in the Discrete Case

There are two important parameters which help in characterizing the distribution of random variables. They are the mean and the variance.

The mean, also called the expected value, or expectation, is a parameter of location. In the case of discrete probability the expected value is given by

$$\mu_x = E(\tilde{x}) = \frac{\Sigma\, x_i\, p_{x_i}}{\Sigma p_{x_i}} = \Sigma\, x_i p_{x_i} \qquad (2.5)$$

where the definition (2.3) was used. Note the similarity between the expected value and the conventional weighted mean. For our example of throwing two dice the expected value is $\mu_x = 7$.

The variance is a parameter of dispersion. It measures the dispersion of the distribution around the mean. In the case of discrete probability the variance is

$$\sigma_x^2 = \sum_{pop.} (x_i - \mu_x)^2 p_{x_i} \qquad (2.6)$$

For the example of throwing two dice we find $\sigma_x^2 = 5.83$.

2.2.2. Continuous Case

Consider an outcome of an experiment which falls within the interval [a, b] as shown in Fig. 2.4 Subdivide this interval into n equal subintervals of length Δx. We define the probability that the outcome falls

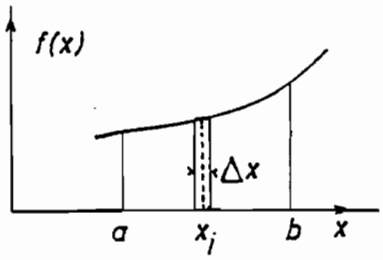


Figure 2.4 Probability Density Function in the Continuous Case as a Limit of a Step Function

in that subinterval which contains x_i as

$$f(x_i)\,\Delta x$$

The probability that the outcome falls in the interval [a,b] is

$$P(a \leq \tilde{x} \leq b) = \sum_{1}^{n} f(x_i)\,\Delta x$$

Taking the limit $\Delta x \to 0$ we obtain

$$P(a \leq \tilde{x} \leq b) = \int_a^b f(x)dx \tag{2.7}$$

The definition (2.7) presupposes the existence of some function f which, integrated from a to b ($b \geq a$), gives the probability that the random variable assumes a value in the interval [a,b] . Note that f(x) itself does not give the probability that the random variable assumes the variable x. In the continuous case the probability is defined by an integral. It follows that

$$P(x = a) = \int_a^a f(x)dx = 0 \tag{2.8}$$

Thus, the probability is zero that the random variable takes on a certain value x. This does not imply that it is impossible to obtain the value x in an actual experiment due to limitations in our ability to measure. Say that we have a tape with cm-division and that the reading for a certain distance is 100.06 m. In this case, we actually mean the interval [100.055, 100.065] . Therefore, equation (2.8) implies an infinitely accurate measurement technique. Repeated measurements of a distance is a sample from a continuous distribution and not a discrete distribution although our reading capability is finite.

In order for f(x) to be a probability density function it has to fulfill certain conditions. Firstly, f(x) has to be a non-negative function since there is always an outcome of an experiment. Secondly, we would like the probability that x is one of <u>all</u> possible outcomes to be 1. Thus, f(x) has to fulfill the following conditions:

a) $f(x) \geq 0$ (2.9)

b) $\int_{-\infty}^{\infty} f(x)dx = 1$ (2.10)

Note that the integration is taken over the whole range (population) of the random variable. The condition (2.10) also implies that $f(-\infty) = f(\infty) = 0$.

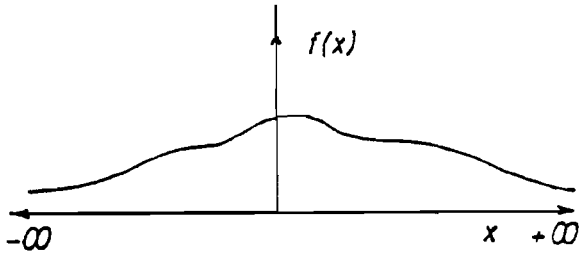


Figure 2.5 The Probability Density Function in the Continuous Case

From equation (2.7) it follows that the probability of the outcome to be smaller or equal to x is

$$P(\tilde{x} \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt \qquad (2.11)$$

F(x) is the cumulative distribution function. It is a non-decreasing function since $f(x) \geq 0$.

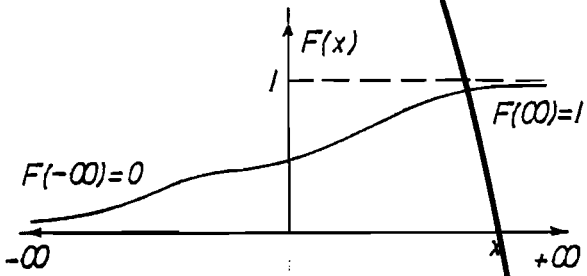


Figure 2.6 Cumulative Distribution Function

2.2.2.1 Mean

The mean, or the expected value, in the case of the continuous distribution is defined by

$$\mu_x = E(\tilde{x}) = \int_{-\infty}^{\infty} x f(x)dx \tag{2.12}$$

We see that the mean is a function of the density function of the random variable. The concept of the mean is very important in the theory of estimation. Note that the integration in (2.12) is extended over the whole population. Equation (2.12) is the analogy to the weighted mean in the case of continuous distributions. $E(\tilde{x})$ is the abscissa of the center of gravity of the area under $f(x)$.

2.2.2.2 Variance

The variance is defined by

$$\sigma_x^2 = E(\tilde{x} - \mu_x)^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)dx \tag{2.13}$$

The variance measures the spread of the probability density in the sense that it gives the expected value of the squared deviations from the mean. A small variance, therefore, indicates that most of the probability density is around the mean. This is a very desireable situation. As for sampling, a small variance indicates high precision. For example, interpreting the curves in Fig. 2.2 as density functions, the variances of the curves 1 and 3 will be smaller than the variances of the curves 2 and 4.

2.3 Multivariate and Marginal Distributions

Since in adjustments usually more than one random variable is involved, we have to review some of the concepts related to multivariate (multi-dimensional) statistics.

Any function $f(x_1, x_2, \cdots x_n)$, of n continuous random variables $\tilde{x}_1$, $\tilde{x}_2$, ... $\tilde{x}_n$ can be a joint density function provided that

$$\begin{aligned} &a)\quad f(x_1, x_2, \ldots x_n) \geq 0 \\ &b)\quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots x_n)dx_1 dx_2 \ldots dx_n = 1 \end{aligned} \tag{2.14}$$

Thus, the probability

$$P(\tilde{x}_1<a_1,\ \tilde{x}_2<a_2,\ \ldots\tilde{x}_n<a_n)$$

$$= \int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\cdots\int_{-\infty}^{a_n} f(x_1,x_2,\cdots x_n)\, dx_1\, dx_2\cdots dx_n \qquad (2.15)$$

is computed in the usual manner.

2.3.1 Mean

The expected value for the parameter $\tilde{x}_i$ is:

$$\mu_{x_i} = E(\tilde{x}_i) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_i\, f(x_1, x_2 \ldots x_n)\, dx_1 dx_2\cdots dx_n \qquad (2.16)$$

Using vector notation we can write the expected value of all parameters as

$$E(\tilde{X}) = \begin{pmatrix} E(\tilde{x}_1) \\ E(\tilde{x}_2) \\ \vdots \\ E(\tilde{x}_n) \end{pmatrix} \qquad (2.16a)$$

2.3.2 Variance-Covariance Matrix, Correlation Matrix

The variance of the random variable $\tilde{x}_i$ is computed by

$$\sigma_{x_i}^2 = E\,(\tilde{x}_i - \mu_i)^2 = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (x_i - \mu_i)^2\, f(x_1, x_2,\ldots x_n)\, dx_1\ldots dx_n \qquad (2.17)$$

$$\geq 0$$

In the case of multivariate distributions we can compute another number called the covariance, which describes the statistical relationship between two random variables. The covariance is computed by

$$\sigma_{x_i x_j} = E\,[\,(\tilde{x}_i - \mu_i)\,(\tilde{x}_j - \mu_j)\,]$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (x_i-\mu_i)\,(x_j-\mu_j)\, f(x_1\ldots x_n)\, dx_1\ldots dx_n \qquad (2.18)$$

Whereas the variance is always larger or equal to zero the covariance can be negative, positive, or even zero. Immediately related to the variances and covariance of two random variables is the correlation coefficient

$$\rho_{x_i x_j} = \frac{E\,[(\tilde{x}_i-\mu_i)\,(\tilde{x}_j-\mu_j)]}{\sigma_{x_i}\,\sigma_{x_j}} \equiv \frac{\sigma_{x_i x_j}}{\sigma_{x_i}\sigma_{x_j}} \qquad (2.19)$$

$\sigma_{x_i} = \sqrt{\sigma_{x_i}^2}$ is called the standard deviation. Thus, the correlation coefficient of any two random variables is defined to be the covariance of the two variables divided by their standard deviations. It can be shown that,

$$-1 \leq \rho_{x_i x_j} \leq 1 \tag{2.20}$$

This is an important relationship.

The covariance and the correlation coefficient are helpful in finding a geometric interpretation of the density function. Here, we will state only some characteristics without giving the mathematical derivation. Figure 2.7 shows a typical density function for two random variables $\tilde{x}_1$ and $\tilde{x}_2$.

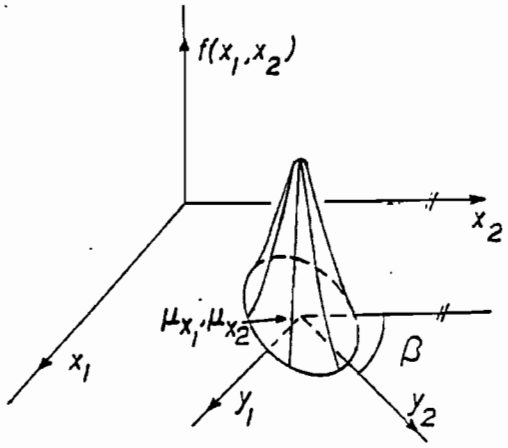


Figure 2.7 Two-Dimensional Density Function

The point (μ_{x_1}, μ_{x_2}) is the origin of a new coordinate system (y_1, y_2) whose orientation differs from the (x_1, x_2) system by the angle β. The following properties can be found:

1) There is always a coordinate system (y_1, y_2) for which $\rho_{y_1 y_2} = 0$.

   Thus, we can always transform the random variables $\tilde{x}_1$ and $\tilde{x}_2$ into a set $\tilde{y}_1$ and $\tilde{y}_2$ for which the covariance is zero. In later sections we will make frequent use of this property.

2) The variances σ_{y_1} and σ_{y_2} are either maximum or minimum.

3) The orientation of the trace of $f(x_1,x_2)$ in the plane (x_1,x_2), which is an ellipse in our case, is a function of the magnitude and sign of $\sigma_{x_1x_2}$.

4) The shape of the ellipse is a function of the correlation coefficient. For $\rho_{x_1x_2} = \pm 1$ the ellipse degenerates into a line. Therefore, the correlation coefficient can be looked upon as a measure of the intensity of the concentration of the probability about the semi-major axis. For example, in the case of a large positive correlation between two random variables we can say that there is the same tendency for the outcome of both random variables, i.e., as x_1 gets larger, x_2 gets larger. Only if the correlation is unity is there an exact linear relationship between the outcomes. In that case the random variables are said to be linearly dependent.

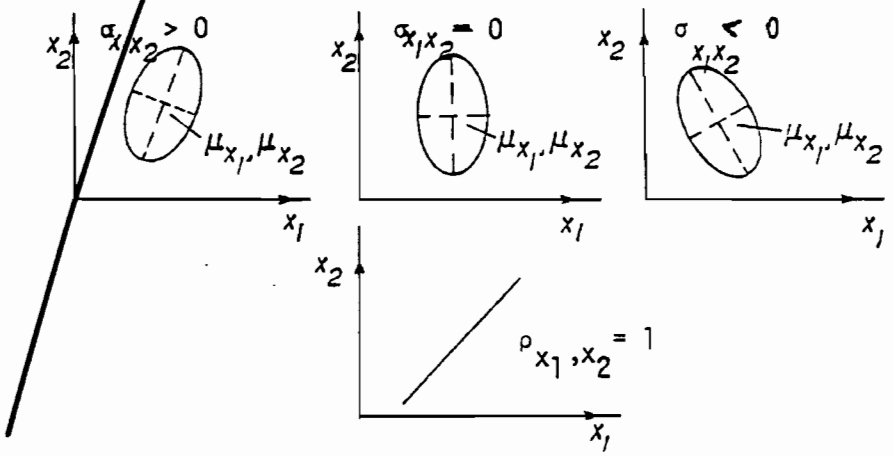


Figure 2.8 Geometrical Interpretation of the Covariance and the Correlation Coefficient for two Random Variables.

In the case of n random variable we can compute n variances as per equation (2.17) and one covariance according to (2.18) for each pair of random variables. The result is arranged in the form of a matrix, called the variance-covariance matrix. Let us write

$$(\tilde{X} - \mu) = \begin{pmatrix} \tilde{x}_1 - \mu_1 \\ \tilde{x}_2 - \mu_2 \\ \vdots \\ \tilde{x}_n - \mu_n \end{pmatrix}$$

Then the n x n variance-covariance matrix becomes

$$\Sigma_X = E\left[(X-\mu_X)(X-\mu_X)^T\right] \equiv \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_n} \\ & \sigma_{x_2}^2 & \cdots & \vdots \\ \text{sym.} & & \ddots & \\ & & & \sigma_{x_n}^2 \end{pmatrix} \tag{2.21}$$

This is a symmetric matrix since $\sigma_{x_ix_j} = \sigma_{x_jx_i}$. Computing the correlations according to equation (2.19) and arranging them in a matrix gives

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ & 1 & \rho_{23} & \cdots & \rho_{2n} \\ \text{sym.} & & \ddots & & \vdots \\ & & & & 1 \end{pmatrix} \tag{2.22}$$

This is the correlation matrix. It is an n x n symmetric matrix whose diagonal elements are one.

2.3.3 Marginal Density Function

Given the joint density function for n random variables we need also the density function, called the marginal density function, of a subset of random variables. Let $f(x_1, x_2 \ldots x_n)$ be the joint density function, and $g(x_1, x_2 \ldots x_p)$ be the marginal density of $x_1, x_2, \ldots x_p$; then

$$g(x_1, x_2 \ldots x_p) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots x_n)\, dx_{p+1}\, dx_{p+2} \ldots dx_n \tag{2.23}$$

The marginal density of any other subset of variables is obtained in an obviously similar fashion.

2.4. Propagation of Mean and Variance

The contents of this section has wide-spread application in adjustments and should therefore be studied carefully. Usually we are more interested in a linear function of the observations than in the observations themselves. In such cases a propagation of variances becomes necessary.

From equation (2.12) it is found that for a constant c:

$$\begin{aligned} &\text{a)}\quad E(c) = c \int_{-\infty}^{\infty} f(x)dx = c \\ &\text{b)}\quad E(c\tilde{x}) = cE(\tilde{x}) \end{aligned} \tag{2.24}$$

Thus the expected value of a constant is equal to the same constant, and the expected value of a multiple of a random variable is equal to the same multiple of the expected value. It follows, in particular, that

$$E(E(\tilde{x})) = \mu_x.$$

The relations (2.24) also hold if we have a multivariate density function as can be seen from equation (2.16). Let $\tilde{y} = \tilde{x}_1 + \tilde{x}_2$ be a simple linear function of the random variables; then

$$\begin{aligned} E(\tilde{x}_1+\tilde{x}_2) &= \int_{-\infty}^{\infty}\int (x_1 + x_2)\, f(x_1,x_2)dx_1dx_2 \\ &= \int_{-\infty}^{\infty}\int x_1 f(x_1,x_2)dx_1dx_2 + \int_{-\infty}^{\infty}\int x_2 f(x_1,x_2)dx_1dx_2 \\ &= E(\tilde{x}_1) + E(\tilde{x}_2) \end{aligned} \tag{2.25}$$

Thus, the expected value of the sum of two random variables is equal to the sum of the individual expected values. Combining the results of (2.24) and (2.25) we can find the expected value in the case of a general linear transformation of random variables:

$$\begin{aligned} \tilde{y}_1 &= a_{10} + a_{11}\tilde{x}_1 + a_{12}\tilde{x}_2 + \ldots a_{1n}\tilde{x}_n \\ \tilde{y}_2 &= a_{20} + a_{21}\tilde{x}_1 + a_{22}\tilde{x}_2 + \ldots a_{2n}\tilde{x}_n \\ &\vdots \\ \tilde{y}_m &= a_{m0} + a_{m1}\tilde{x}_1 + a_{m2}\tilde{x}_2 + \ldots a_{mn}\tilde{x}_n \end{aligned}$$

This transformation can be written in matrix notation as

$$_m\tilde{Y}_1 = A_o + {}_mA_n\tilde{X} \tag{2.26}$$

where elements a_{ij} of A are constants.

It follows that

$$\boxed{E(\tilde{Y}) = A_o + AE(\tilde{X})} \tag{2.27}$$

This is the law for propagating the mean. A corresponding expression for propagating the variances can be found as follows:

Consider the simple transformation

$$\begin{aligned}\tilde{y}_1 &= a_{10} + a_{11}\tilde{x}_1 + a_{12}\tilde{x}_2\\ \tilde{y}_2 &= a_{20} + a_{21}\tilde{x}_1 + a_{22}\tilde{x}_2\end{aligned} \tag{2.28}$$

with the given probability density function $f(x_1,x_2)$. We find that

$$\begin{aligned}\sigma_{y_1}^2 &= E(\tilde{y}_1 - \mu_{y_1})^2 = E(a_{10}+a_{11}\tilde{x}_1+a_{12}\tilde{x}_2 - a_{10}-a_{11}\mu_{x_1}-a_{12}\mu_{x_2})^2\\ &= E\{a_{11}(\tilde{x}_1-\mu_{x_1}) + a_{12}(\tilde{x}_2-\mu_{x_2})\}^2\\ &= E\{a_{11}^2(\tilde{x}_1-\mu_{x_1})^2 + a_{12}^2(\tilde{x}_2-\mu_{x_2})^2 + 2a_{11}a_{12}(\tilde{x}_1-\mu_1)(\tilde{x}_2-\mu_2)\}\\ &= a_{11}^2\sigma_{x_1}^2 + a_{12}^2\sigma_{x_2}^2 + 2a_{11}a_{12}\sigma_{x_1x_2}\end{aligned}$$

Similarly,

$$\sigma_{y_2}^2 = a_{21}^2\sigma_{x_1}^2 + a_{22}^2\sigma_{x_2}^2 + 2a_{21}a_{22}\sigma_{x_1x_2}$$

Note that the constant term drops out. For the covariance we obtain:

$$\begin{aligned}\sigma_{y_1y_2} &= E[(\tilde{y}_1-\mu_{y_1})(\tilde{y}_2-\mu_{y_2})]\\ &= E\{[a_{11}(\tilde{x}_1-\mu_{x_1}) + a_{12}(\tilde{x}_2-\mu_{x_2})][a_{21}(\tilde{x}_1-\mu_1) + a_{22}(\tilde{x}_2-\mu_{x_2})]\}\\ &= E\{a_{11}a_{21}(\tilde{x}_1-\mu_1)^2 + (a_{11}a_{22}+a_{12}a_{21})(\tilde{x}_1-\mu_1)(\tilde{x}_2-\mu_2) + a_{12}a_{22}(\tilde{x}_2-\mu_{x_2})^2\}\\ &= a_{11}a_{21}\sigma_{x_1}^2 + (a_{11}a_{22}+a_{12}a_{21})\sigma_{x_1x_2} + a_{12}a_{22}\sigma_{x_2}^2\end{aligned}$$

It can be easily verified that the last three expressions can be written as

$$\begin{pmatrix}\sigma_{y_1}^2 & \sigma_{y_1y_2}\\ \sigma_{y_2y_1} & \sigma_{y_2}^2\end{pmatrix} = \begin{pmatrix}a_{11}a_{12}\\ a_{21}a_{22}\end{pmatrix}\begin{pmatrix}\sigma_{x_1}^2 & \sigma_{x_1x_2}\\ \sigma_{x_2x_1} & \sigma_{x_2}^2\end{pmatrix}\begin{pmatrix}a_{11}a_{12}\\ a_{21}a_{22}\end{pmatrix}^T \tag{2.29}$$

Thus, the variance-covariance matrix of the transformed variables y_i is obtained by pre- and post-multiplying the variance-covariance matrix of the original random variables by a matrix whose coefficients are the constant elements of the transformation. The relationships (2.28) and (2.29) can readily be generalized. It follows that given the linear transformation (2.26) and the variance-covariance matrix Σ_X, the variance-covariance matrix of the transformed variables $\tilde{Y}$ is

$$\boxed{\Sigma_Y = A\,\Sigma_X A^T} \tag{2.30}$$

This is the law for propagating variances and covariances. It is extremely important in adjustments. Please note, that the constant term A_0 has no effect on the propagation of variance-covariances.

The relationship (2.30) can be derived more elegantly by using matrix notation right from the outset. Given the linear transformation (2.26) and the expected value (2.27) we find the covariance matrix, upon using the definition (2.21), to be

$$\begin{aligned}
\Sigma_Y &= E\left[(\tilde{Y}-\mu_Y)(\tilde{Y}-\mu_Y)^T\right] \\
&= E\left[\tilde{Y} - A_0 - AE(\tilde{X})\right]\left[\tilde{Y} - A_0 - AE(\tilde{X})\right]^T \\
&= E\left[A\tilde{X} - AE(\tilde{X})\right]\left[A\tilde{X} - AE(\tilde{X})\right]^T \\
&= AE[(\tilde{X} - E(\tilde{X}))(\tilde{X} - E(\tilde{X}))^T]A^T \\
&= A\,\Sigma_X A^T
\end{aligned}$$

As an example for the application of the above law we compute the variance-covariance matrix of

$$\tilde{Y}_1 = A_0 + A\tilde{X}_1$$

$$\tilde{Y}_2 = B_0 + B\tilde{X}_2$$

with

$$\Sigma_X = \begin{pmatrix} \Sigma_{X_1} & \Sigma_{X_1X_2} \\ \Sigma_{X_2X_1} & \Sigma_{X_2} \end{pmatrix}$$

It follows that

$$\Sigma_Y = \begin{pmatrix} \Sigma_{Y_1} & \Sigma_{Y_1Y_2} \\ \Sigma_{Y_2Y_1} & \Sigma_{Y_2} \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} \Sigma_{X_1} & \Sigma_{X_1X_2} \\ \Sigma_{X_2X_1} & \Sigma_{X_2} \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}^T$$

$$= \begin{pmatrix} A\Sigma_{X_1}A^T & A\Sigma_{X_1X_2}B^T \\ B\Sigma_{X_2X_1}A^T & B\Sigma_{X_2}B^T \end{pmatrix}$$

The relation

$$\Sigma_{Y_1Y_2} = A\Sigma_{X_1X_2}B$$

is sometimes referred to as the covariance propagation. Note, although $\tilde{Y}_1$ and $\tilde{Y}_2$ are functionally independent they are correlated because $\tilde{X}_1$ and $\tilde{X}_2$ are correlated.

Finally, an alternative expression for the variance-covariance matrix is given. It follows from the propagation of the mean that

$$\begin{aligned} \Sigma_X &= E\,(\tilde{X}-\mu_X)(\tilde{X}-\mu_X)^T \\ &= E\,[\tilde{X}\tilde{X}^T - \tilde{X}\mu_X^T - \mu_X\tilde{X}^T + \mu_X\mu_X^T] \\ &= E(\tilde{X}\tilde{X}^T) - \mu_X\mu_X^T \end{aligned} \qquad (2.31)$$

2.5 Cofactor Matrix and Weight Matrix

Given the variance-covariance matrix Σ_X and

$$\sigma_o^2 Q_X = \Sigma_X \qquad (2.32)$$

where σ_o^2 is a scalar, than Q_X is called the cofactor matrix. The scalar can have units. If Σ_X is non-singular then the weight matrix P is given by

$$P_X = Q_X^{-1} = \sigma_o^{-2}\Sigma_X^{-1} \qquad (2.33)$$

The general weight matrix P can be full, diagonal, or can even be the identity matrix. Usually the weight matrix of the observations will be diagonal in which case the observations are uncorrelated. Thus,

$$P = \begin{pmatrix} p_1 & & 0 \\ & p_2 & \\ & & \ddots \\ 0 & & p_n \end{pmatrix} = \sigma_o^2 \begin{pmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \frac{1}{\sigma_2^2} & \\ & & \ddots \\ 0 & & \frac{1}{\sigma_n^2} \end{pmatrix}$$

It is seen that in this particular case the weights are proportional to the reciprocal of the variances. If the variance $\sigma_i^2 = \sigma_o^2$ then the weight is unity. It is because of this relation that σ_o^2 and σ_o are called the variance of unit weight and the standard deviation of unit weight.

2.6 Measures of Position and Dispersion of a Sample

It was mentioned previously that the expected value (mean) and the variance are measures of position and dispersion for the population. The integrations involved in computing these quantities must be taken over the whole population. However, we do not have immediate access to the population since our sample can only consist of a finite number of events (measurements). Hence, we have to rely on the sample in order to determine the parameters of the population. We can also derive specific quantities from the sample, called statistics, which are estimates of the population parameters.

In the following a few statistics are given which can be used to estimate the parameter of position. Assume that we have a sample (x_1, $x_2 \ldots x_n$) of one and the same random variable $\tilde{x}$.

a) Sample mean:

$$\bar{x} = \frac{1}{n} \sum_1^n x_i \qquad (2.34)$$

b) Median: Arrange the sample according to size. If n is odd, the median is the observation numbered (n + 1)/2; if n is even, the median is defined as the average of the observations numbered n/2 and (n + 2)/2.

The median has a disadvantage in that it does not utilize fully all the positional information contained in the sample. It tends to fluctuate.

c) Mode: The mode is the sample value which occurs most often. The mode and the mean are not necessarily the same. As will be seen later, in the case of the normal distribution the mode and the mean coincide.

d) Midrange: The midrange is simply the mean of the maximum and minimum sample value:

$$\text{midrange} = (x_{max} + x_{min})/2$$

Statistics for estimating the dispersion of a distribution can be:

a) Range: The range is the difference between the maximum and minimum value,

$$\text{range} = x_{max} - x_{min}$$

b) Mean (average) deviation: The average deviation is defined by

$$\text{average deviation} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}| \quad \text{where } \bar{x} \text{ is the sample mean.}$$

c) Variance:

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (2.35)$$

The sample variance measures the average of the squared deviation from the mean. The divisor is n-1 and not n! There are only n-1 deviations independent for a sample of size n. The n-th deviation can be expressed in terms of the other n-1 deviations and the mean.

It is instructive to emphasize again that in the above discussion we assumed a sample size of n for one and the same parameter. This is quite different from the usual case in adjustments where the sample size is one and we would like to estimate a number of unknown linear functions of the observation, based on several samples of as many populations; thus each sample is of size one. For example, we measure a certain distance with a specific instrument. The measure itself is a random sample of the random variable "distance". If only one measurement is taken then this sample is identical to the sample mean and is as such an estimate for the population mean. But the instrument can provide the distance with only a finite accuracy.

From experiences of previous experiments, comparison of other instruments, or from measuring a known distance (calibration) we usually have an idea about the measurement accuracy of the instrument used. This information constitutes a sample and as such is an estimate of a second population parameter, i.e., the dispersion. It will be shown in Section 3 that it is sufficient to have for each population the estimates for these two population parameters. Sometimes the sample size is larger than one, e.g., we measure the same distance several times in order to eliminate blunders, systematic errors, etc. In that case the sample mean and variance are computed according to equations (2.34) and (2.35) and used as estimates of the population mean and variance in the adjustment. It is interesting to note that computing the sample mean according to (2.34) is itself a least squares solution. The student will be able to verify this fact as soon as the basic models in Section 3 have been studied.

2.7 Types of Estimation

Estimation means making inferences regarding parameters of population (distributions) on the basis of statistics, which in turn are functions of the samples (observations). But any estimation is made with respect to its own criteria. Recall the statement of the last section that even the sample mean is a "least squares" solution. In general, then, we will obtain different estimates for population parameters depending on the criteria chosen for the estimation. Naturally, we would like to impose reasonable criteria so as to obtain a "useable" solution with a reasonable amount of computational effort and which has some optimal properties. Any function of the samples used to estimate the population parameters and having certain properties is called an estimator. If the function is linear we speak of a linear estimator.

Unbiased Estimation: Any estimator $\tilde{\theta}$ whose mathematical expectation is equal to a parameter θ is called an unbiased estimator for the parameter θ,

$$E(\tilde{\theta}) = \theta \tag{2.36}$$

The concept of unbiased estimation is important in Surveying. This results from the fact that a given measurement tool might not be able to observe certain parameters. Typical examples are adjustments of station directions, leveling nets, and horizontal networks. Let us consider the latter case. Assume that

a trilateration network has been observed. The extension of the network may be relatively small so that a "plane" computation can be carried out. We would like to describe the net by Cartesian coordinates; that is, Θ denotes the vector of coordinates. It is obvious that the distance observations do not contain information about a unique set of Cartesian station coordinates. Although the relative positions of the stations are determined by the distance measurement it is possible to translate or rotate all stations by the same amount without changing any of the station distances. This fact shows up in the design matrix of the adjustment which has a rank defect of 3 in this particular case; and, consequently, the normal matrix is singular. It can be shown that the relation (2.36) is not fulfilled and that the coordinates are therefore not unbiasedly estimable. We can fix a coordinate system arbitrarily, say, by choosing the origin of the coordinate system at one particular station and letting one of the coordinate axes pass through another station (minimal constraint). We are then able to compute the coordinates of all other stations. These coordinates are, of course, relative to the arbitrarily fixed system. Assume that two surveying teams have observed the same net independently and would like to compare their adjusted coordinates. A comparison makes sense only if they also use the same definition of the coordinate system (same minimal constraint). If the two teams agree to compare, for example, the adjusted angles of the net, they could do this at once, although both might define the coordinate system differently. This is a consequence of the angles being unbiasedly estimable. If Θ denotes the angles, then the relation (2.36) can be shown to be fulfilled.

We thus recognize the importance of properly identifying unbiased estimable parameters. This subject will be treated in great detail in laboratories.

Minimum Variance Estimation: The minimum variance estimator gives an estimate which has a minimum variance compared to other estimators. This is, of course, a desireable property since the precision increases as the variance decreases.

In Section 3 the above two properties will be verified for the least squares estimator. As a special case we show here that the sample mean and variance are unbiased estimates of the population mean and variance. Consider n random variables, each describing the same population with mean μ and variance σ_x^2, then

$$E(\tilde{\bar{x}}) = E\{\frac{1}{n}\ (\tilde{x}_1 + \tilde{x}_2 + \ldots \tilde{x}_n)\}$$

$$= \frac{1}{n}\{E(\tilde{x}_1) + E(\tilde{x}_2) + \ldots E(\tilde{x}_n)\} = \frac{1}{n}\, n\mu = \mu$$

Thus,

$$E(\tilde{\bar{x}}) = \mu,$$

The mean of a sample is an unbiased estimate for the population mean. In the case of the sample variance we obtain

$$E(s_x^2) = \frac{1}{n-1}\ E\ \{\Sigma(\tilde{x}_i - \tilde{\bar{x}})^2\}$$

After some rearrangement, which can be found in any introductory book on statistics, we find

$$E(s_x^2) = \sigma_x^2$$

3. Adjustments by Least Squares

In the first three subsections the three basic adjustment models are introduced. There is sufficient reason to treat each of these models separately since, for a given adjustment problem, one of the models may appear more readily applicable than the other. It is shown how the models of observation equations and condition equations can formally be derived from the mixed model through specifying certain matrices. Advantage is taken of this fact in Subsections 3.4 to 3.6 where the derivations are principally done for the mixed model, and the solutions of the other model are found formally through the specifications mentioned above. This method provides a good means of systematizing the various solutions. Of course, each of the solutions can be derived separately starting with the respective minimum condition.

Since in Surveying most of the relationships are inherently non-linear special attention is given to the linearization and the iteration techniques for each of the cases.

Although the review of statistical concepts in Section 2 may appear lengthy and in much detail, it is worth noting that, thus far, the density distribution has not been specified. In fact, the mere existence of the mean and the variance-covariance matrix is sufficient to perform a least squares adjustment. A specific density function is not needed.

3.1 Method of Observation Equations

This is probably the most often used method in Surveying. It is easy to implement on the computer. The main characteristic of this model is the explicit functional relationship between the observations and the unknown parameters.

The following symbols will be used throughout this section:

u X_o 1 vector of approximate values for the variables (≡parameters ≡ unknown). These values are assumed to be known a priori, and they should be as close as possible to the true values.

X_a vector of parameters as obtained from the adjustment (adjusted parameters)

$$X = X_a - X_o \tag{3.1}$$

$$X_a = X_o + X \tag{3.2}$$

$_n{L_b}_1$ vector of observation; (sample)

L_o numerical value of the observed quantity as computed from the adopted mathematical model and the approximate values X_o

L_a adjusted values of observed quantities

$$L = L_o - L_b \tag{3.3}$$

$$V = L_a - L_b \tag{3.4}$$

V is the vector of "Residuals". They result from the random observational errors. Finally, the mathematical model for adjustment by observation equations is

$$\boxed{L_a = F(X_a)} \tag{3.5}$$

Note that the mathematical model relates the true values, i.e., the adjusted observations and the adjusted parameters!

With the symbols explained above we can rewrite the relation (3.5) as

$$L_b + {}_nV_1 = {}_nF_1\,(X_a) \tag{3.6}$$

or $$L_b + V = F(X_o + X) \tag{3.7}$$

Linearizing this relationship with the expansion point at X_o gives

$$L_b + V = F(X_o) + AX \tag{3.8}$$

where

$${}_nA_u = \left.\frac{\partial F(X)}{\partial X}\right/_{X_o} \tag{3.9}$$

and $$L_o = F(X_o) \tag{3.10}$$

The linear equation (3.8) can be written in the form

$$\boxed{{}_nV_1 = {}_nA_u\ X_1 + {}_nL_1} \tag{3.11}$$

These are the observation equations in matrix form. The matrix A is called the "design matrix" or the "coefficient matrix".

Besides the functional model (3.8) we must have a stochastic model, that is, the variance-covariance matrix Σ_{L_b} of the observations. According to the equation (2.33) we write,

$$P_{L_b} = \sigma_o^2\ \Sigma_{L_b}^{-1} \tag{3.12}$$

From the law for propagating the variance we find

$$P_{L_b} = P_L \equiv {}_nP_n \tag{3.13}$$

As far as notation is concerned we recognize that L in equation (3.11) denotes the sample value of $\tilde{L}$. Since most of this section is more concerned with the numerical aspect of the estimation than with the statistical aspect the symbol "~", denoting a random variable, will only be used when absolutely necessary. The complete observation equation model can be written as follows:

$$E(\tilde{L}) = AX \tag{3.14}$$

$$E[\{\tilde{L} - E(\tilde{L})\}\{\tilde{L} - E(\tilde{L})\}^T] = \Sigma_{L_b} = \sigma_o^2 P^{-1} \tag{3.15}$$

implying $$E(\tilde{V}) = 0 \tag{3.16}$$

Since $L = L_o - L_b$ we have $\Sigma_L = \Sigma_{L_b}$

Note that the unknown X is a fixed parameter and not a random variable. The variance-covariance can have the following special forms:

a) identity matrix
b) diagonal matrix
c) full matrix
d) non-singular
e) singular

In case e) the observations are functionally dependent; that is, at least one observation can be expressed as a function of the others. This case will not be considered any further here. If the variance-covariance matrix is non-singular the observations are functionally independent. We will only deal with this case. If the variance-covariance matrix is full, (case c), then the observations are stochastically dependent, or, correlated. In the cases a) and b) we are dealing with uncorrelated observations.

Having the functional and the stochastic model we can proceed with the solution by applying the "principle of least squares" :

$$\phi(X) \equiv V^T P V = \text{minimum} \tag{3.16}$$

Thus the solution for X is found by minimizing the quadratic form (3.16). This condition is quite general in that it is valid for both a full and a diagonal weight matrix. In the special case that P is the identity matrix it is seen that (3.16) requires the minimization of the sum of the squares of the residuals. It is from this particular case that the name "least squares" was derived. We also observe that the principle of least squares only requires a weight matrix and not necessarily the knowledge of the a-priori variance of unity weight σ_0^2 or the variance-covariance matrix. We can, therefore, perform a least squares adjustment without invoking the concept of a variance-covariance matrix at all. We only have to assign weights to the observations in some manner. However, this type of interpretation does not allow a statistical evaluation of the results and it, therefore, is not persued any further. The estimates of the parameters themselves are independent of such interpretation.

From equations (3.11) and (3.13) we obtain

$$V^TPV = X^TA^TPAX + X^TA^TPL + L^TPAX + L^TPL \tag{3.17}$$

The 3rd and 4th term are just numbers (matrices of size 1), so that

$$X^TA^TPL = L^TPAX$$

Using the notation

$$N \equiv A^TPA \tag{3.18}$$

$$U \equiv A^TPL \tag{3.19}$$

equation (3.17) becomes

$$\phi \equiv V^TPV = X^TNX + 2X^TU + L^TPL \tag{3.20}$$

The solution for X, denoted by $\hat{X}$, is found by minimizing the quadratic form ϕ. A stationary point of the function $\phi = \phi(X)$ is obtained by setting the partial derivations with respect to the parameters equal to zero. Using the rules for matrix differentiation we obtain

$$\frac{1}{2}\frac{\partial\phi(X)}{\partial X} = \hat{X}^TN + U^T = 0 \tag{3.21}$$

The estimate $\hat{X}$ of the parameters follows from the solution of (3.21). Statistically, the estimate of X is itself a random variable $\tilde{\hat{X}}$ since it is a function of the random variables $\tilde{L}$. Thus, if the symbol "~" is used for $\tilde{L}$ it should also be used for $\tilde{\hat{X}}$ and vice versa depending on whether one intends to emphasize the random nature of the variable or the sample value.

Re-writing equation (3.21) in the form

$$\boxed{{}_nN_n\hat{X} + U = 0} \tag{3.22}$$

gives the n normal equations for the n parameters. N is called the normal matrix. It is a positive definite matrix assuming that A has a full column rank. The solution of equation (3.22) is

$$\hat{X} = -N^{-1}U \tag{3.23}$$

$$= -N^{-1}A^{T}PL$$

The adjusted parameters are, of course,

$$X_a = \hat{X} + X_o$$

and the adjusted observations are

$$L_a = L_b + \hat{V} = L_o + A\hat{X}$$

The actual computation of the minimum obtained for $V^{T}PV$ is very important for later use in analysis. Substituting equation (3.23) in (3.20) gives

$$V^{T}PV = -U^{T}N^{-1}U + L^{T}PL$$

$$= -\hat{X}^{T}N\hat{X} + L^{T}PL \tag{3.24}$$

It is now possible to compute an estimate for the a priori variance of unit weight σ_o^2. Recognizing that $\tilde{\phi}$, $\tilde{\hat{V}}$, $\tilde{\hat{X}}$ are random functions since each is a function of $\tilde{L}$ we can compute the expected value of

$$E(\tilde{\hat{Q}}) = E(\tilde{\hat{V}}^{T}P\tilde{\hat{V}}) \tag{3.25a}$$

The symbol $\tilde{\hat{V}}$ denotes the random variable of the adjusted residuals,

$$\tilde{\hat{V}} = -AN^{-1}A^{T}P\tilde{L} + \tilde{L} \tag{3.25b}$$

which must be distinguished from

$$\tilde{V} = AX + \tilde{L} \tag{3.25c}$$

Making use of the properties of a trace of a matrix and of equation (3.24) we find

$$E(\tilde{\hat{V}}^T P \tilde{\hat{V}}) = E\{Tr(\tilde{L}^T P \tilde{L}) - Tr(\tilde{X}^T N \tilde{X})\}$$

$$= E\{Tr(\tilde{L}\tilde{L}^T P) - Tr(\tilde{X}\tilde{X}^T N)\}$$

$$= Tr\{E(\tilde{L}\tilde{L}^T P)\} - Tr\{E(\tilde{X}\tilde{X}^T N)\}$$

$$= Tr\{E(\tilde{L}\tilde{L}^T P)\} - Tr\{E(N^{-1}A^T P \tilde{L}\tilde{L}^T P A N^{-1} N)\} \quad (3.25)$$

From equation (3.15) and (2.31) we know that

$$E(\tilde{L}\tilde{L}^T) = \Sigma_{L_b} + E(\tilde{L})E(\tilde{L}^T)$$

Substitution of equation (3.14) for the expected value of $\tilde{L}$ gives

$$E(\tilde{L}\tilde{L}^T) = \Sigma_{L_b} + AXX^T A^T$$

$$= \sigma_o^2 P^{-1} + AXX^T A^T$$

Inserting this expression into (3.25) gives

$$E(\tilde{\hat{V}}^T P \tilde{\hat{V}})$$

$$= Tr\{\sigma_o^2 \, n I_n + AXX^T A^T P\} - Tr\{\sigma_o^2 \, u I_u + XX^T A^T P A\}$$

$$= n\sigma_o^2 + Tr(AXX^T A^T P) - u\sigma_o^2 - Tr(AXX^T A^T P)$$

$$= \sigma_o^2 (n-u) \quad (3.26)$$

This derivation could have been carried out somewhat more generally permitting a rank defect on the design matrix, e.g. $R(A) = R(N) = r<u$. In that case the result would be

$$E(\tilde{\hat{V}}^T P \tilde{\hat{V}}) = \sigma_o^2 \{n - R(A)\} \quad (3.27)$$

This is an important result. If we choose

$$\tilde{\hat{\sigma}}_o^2 = \frac{\tilde{\hat{V}}^T P \tilde{\hat{V}}}{n-R(A)} \quad (3.28)$$

then

$$E(\tilde{\hat{\sigma}}_o^2) = \sigma_o^2$$

i.e., the expression (3.28) is an unbiased estimate of the unknown a priori variance of unit weight σ_o^2. Evaluating (3.28) for the specific sample L gives

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{n - R(A)} \tag{3.29}$$

$\hat{\sigma}_o^2$ is also called the a-posteriori variance of unit weight. It plays an important role in the statistical analysis in later sections. It reflects the overall behavior of the adjustment since all observations, and the stochastic and functional model, enter into its computation.

Using the law for propagating variances it is now possible to compute the variance-covariance matrix for a number of quantities. We had earlier

$$\Sigma_L = \Sigma_{L_b} = \sigma_o^2 P^{-1}$$

From equation (3.23) it follows that the variance-covariance of $\tilde{X}$ is

$$\hat{\Sigma}_X = \hat{\sigma}_o^2 N^{-1} \tag{3.30}$$

If one is interested in the variance of a linear function of the adjusted parameter one can use again the law of propagating the variances. In the case of a large system (many parameters u) one might compute only the variances and covariances of those parameters which are included in the function in question. In such cases the possibilities given by the Cholesky algorithm can be fully utilized.

For the residuals one obtains

$$\begin{aligned} \hat{V} &= A\hat{X} + L \\ &= -AN^{-1}A^T PL + L \\ &= (I - AN^{-1}A^T P)L \end{aligned}$$

$$\Sigma_V = (I - AN^{-1}A^TP)\,\sigma_o^2P^{-1}\,(I - PAN^{-1}A^T)$$

$$= \sigma_o^2\,(P^{-1} - AN^{-1}A^T)$$

The estimated variance-covariance matrix for the residuals is

$$\boxed{\hat{\Sigma}_V = \hat{\sigma}_o^2\,(P^{-1} - AN^{-1}A^T)} \qquad (3.31)$$

For the adjusted observations we get

$$\Sigma_{V-L} \equiv \Sigma_{L_a - L_o} = \Sigma_{L_a} = A\Sigma_X A^T = \sigma_o^2 AN^{-1}A^T$$

and

$$\hat{\Sigma}_{L_a} = \hat{\sigma}_o^2\,AN^{-1}A^T \qquad (3.32)$$

From the equations above we find the relationship

$$\boxed{\Sigma_{L_a} = \Sigma_{L_b} - \Sigma_V} \qquad (3.33)$$

The variances of the adjusted observations are obtained by subtracting the variances of the residuals from those of the observation. This difference is exactly the amount of improvement achieved by the adjustment.

Finally, we would like to compute the covariance matrix for the vector

$$Z = \begin{vmatrix} X \\ V \\ L_a \end{vmatrix}$$

Using the law of propagation of variances:

$$\Sigma \begin{vmatrix} X \\ V \\ L_a \end{vmatrix} = \sigma_o^2 \begin{pmatrix} N^{-1} & 0 & N^{-1}AT \\ 0 & P^{-1} - AN^{-1}A^T & 0 \\ AN^{-1} & 0 & AN^{-1}A^T \end{pmatrix} \qquad (3.34)$$

In the following some remarks are made as to the correct use of the adjustment model with observation equations, and some of the properties involved:

Note: 1) The expression $-U^T N^{-1} U$ needed in computing $V^T PV$ is obtained immediately from Cholesky's algorithm (See Section 1.5.3).

2) Comparing $U^T \hat{X} = -U^T N^{-1} U$ is a useful check (programming error, etc.). Compute the left side with the result of eq.(3.23) and take the right side from Cholesky's algorithm.

3) Compute the residuals from eq. (3.11), form $V^T PV$ and compare with (3.24). This is another check for programming errors, etc.

4) Compute the residuals from the non-linear equation (3.6) and form $V^T PV$. Comparing this result with the corresponding value of 3) gives a check on the sufficiency of the linearization. Remember that a successful linearization depends on good approximate parameters X_o, i.e., whether or not the linear function (3.11) approximates sufficiently well the non-linear function (3.6). One should at least carry out one iteration.

Iteration (method of observation equation)
For the i-th adjustment we have the following algorithm:

$$X_i = X_{i-1} + X^{i-1}$$

$$V_i = A_i X^i + (L_i - L_b)$$

$$A_i = \frac{\partial F(X)}{\partial X} \Big/ X_i$$

$$L_i = F(X_i)$$

Step 1: X_o; Compute A_o, $L_o = F(X_o)$, $L = L_o - L_b$

Adjustment gives: X^o, V_o, $V_o^T P V_o$

Step 2: $X_1 = X_o + X^o$, $L_1 = F(X_1)$, A_1, $L = L_1 - L_b$

Adjustment gives: X^1, V_1, $V_1^T P V_1$

Step 3: If $|V_1^T P V_1 - V_o^T P V_o| < \varepsilon$,

where ε is sufficiently small, the adjustment has converged. A satisfactory conversion might require several iterations. Repeat Step 2 and Step 3.

Note: $X^i \to 0$

$V_i \to V = L_a - L_b$

as i gets larger

5) Assume that in the condition (3.16) the weight matrix P is replaced by $\bar{P} = \alpha P$; that is the weight matrix is scaled.

a) $\hat{X}$ remains unchanged (equation 3.23)

b) $\hat{V}$ remains unchanged

c) $V^T \bar{P} V = \alpha V^T P V$

d) $\bar{\sigma}_o^2 = \alpha \sigma_o^2$

One frequently uses this technique of scaling all the weights in order to avoid extreme large or small numbers in the normal matrix which might cause rounding errors on the computer or might even make the inversion of N impossible using standard inversion routines.

6) Another method for improving the condition of the normal matrix N is to change the units of individual parameters. The observation equations are:

$$V = (A_1\ A_2\ \cdots\ A_u) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_u \end{pmatrix} + L$$

Assume that X_2 is a parameter for an angle given in radians. We can change the units to arcsec (") by

$$V = (A_1\ \frac{1}{\rho''}\ A_2\ \cdots\ A_u) \begin{pmatrix} X_1 \\ X_2('') \\ \vdots \\ X_u \end{pmatrix} + L$$

where $\rho'' \approx 2 \times 10^5$. This causes the 2nd diagonal element of N to be divided by $(\rho'')^2$:

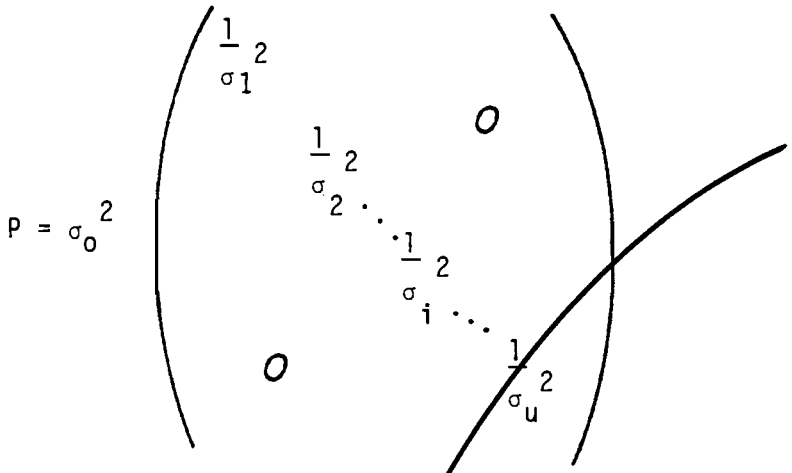
$$N = (A^T P A) = \begin{pmatrix} A_1^T \\ \frac{1}{\rho} A_2^T \\ \vdots \\ A_u^T \end{pmatrix} P \left(A_1 \tfrac{1}{\rho} A_2 \cdots A_u\right)$$

$$= \begin{pmatrix} A_1^T P A_1 & \frac{1}{\rho} A_1^T P A_2 & \cdots & A_1^T P A_u \\ \frac{1}{\rho} A_2^T P A_1 & \frac{1}{\rho} A_2 P A_2 & \cdots & \frac{1}{\rho} A_2^T P A_u \\ \vdots & & & \\ A_u^T P A_1 & \frac{1}{\rho} A_u^T P A_2 & \cdots & A_u^T P A_u \end{pmatrix}$$

There is a corresponding change in $A^T PL$:

$$A^T PL = \begin{pmatrix} A_1^T PL \\ \frac{1}{\rho} A_2^T PL \\ \vdots \\ A_u^T PL \end{pmatrix}$$

7) Finally, one can change the units of the observations in order to improve the condition of the normal matrix. Remember that the units in the weight matrix (inverse of variance-covariance matrix) and those of the observations L have to correspond respectively. Take the simplified case where P is a diagonal matrix:



$$P = \sigma_o^2 \begin{pmatrix} \frac{1}{\sigma_1^2} & & & & 0 \\ & \frac{1}{\sigma_2^2} & & & \\ & & \ddots & & \\ & & & \frac{1}{\sigma_i^2} & \\ & & & & \ddots \\ 0 & & & & \frac{1}{\sigma_u^2} \end{pmatrix}$$

The unit of σ_i has to be the same as that of the observation l_i, e.g., meters (m). The weight p_i has then the dimension (m^{-2}), e.g.,

$$p_i = \frac{1}{.5m^2}$$

Introducing the observation l_i with the unit (mm) into the adjustment requires a corresponding change in p_i by

$$p_i = \frac{1}{.5 \times 10^6 (mm)^2}$$

Usually a good numerical inversion on the computer can be assured if the diagonal elements of the normal matrix are nearly of the same order of magnitude. Please note that any numerical instability caused by nearly dependent columns of the A-matrix cannot be avoided through the procedures 5 through 7.

8) The structure of the variance covariance matrix Σ_X of the adjusted parameters does not depend on the observations themselves but it depends only on the design matrix A and the weight matrix P. Thus, the analysis of the covariance matrix Σ_X can be done prior to any actual field work by simply assuming the geometry (mathematical model or, equivalently, the design matrix A) and

the weights, which, of course, should reflect the anticipated measurement accuracy. Such a procedure is called a simulation study. One can also study the optimal solution before starting the field work. Firstly, one can change the geometry as to reduce the magnitude of the variances of the adjusted parameters. In the ideal situation the columns of A (or N) are orthogonal to each other. Secondly, one can plan the observation scheme in advance. Certain observations contribute more than others in reducing the variance of the adjusted parameters. Thus, through simulation studies we can find the optimal geometry and optimal observation scheme.

9) $A^T P \hat{V} = 0$

Prove by substitution:

$$A^T P (A\hat{X} + L)$$

$$= -A^T P A N^{-1} A^T P L + A^T P L$$

$$= -A^T P L + A^T P L = 0$$

10) Alternative method for deriving the least squares equation: Use $AX - V + L$ as a condition, and use the method of the Lagrange multiplier.

$$\phi = \phi(V, K, X) = V^T P V - 2K^T (AX - V + L)$$

Minimize ϕ:

$$\frac{1}{2}\frac{\partial \phi}{\partial V} = V^T P - K^T = 0$$

$$\frac{1}{2}\frac{\partial \phi}{\partial X} = -K^T A = 0 \qquad (3.35)$$

$$\frac{1}{2}\frac{\partial \phi}{\partial K^T} = A\hat{X} - V + L = 0$$

From the first two equations we get the condition

$$A^T PV = 0$$

Multiplying the third equation from the left by $A^T P$ gives

$$A^T PA\,\hat{X} - \underbrace{A^T PV}_{0} + A^T PL = 0$$

If follows that,

$$\hat{X} = -N^{-1}A^T PL$$

11) In deriving the least squares solution we have placed the first derivative to zero as is seen from equation (3.21) and (3.35). But such a procedure assures only that a stationary point of the function $\phi = \phi(X)$ will be obtained. That indeed a minimum was found is seen as follows: Compute $V^T PV$ for any value $\tilde{\tilde{X}}$ being different from $\hat{X}$ by ΔX:

$$\tilde{\tilde{X}} = \hat{X} + \Delta X$$

$$(V^T PV)_{\tilde{\tilde{X}}} = (L^T + \tilde{\tilde{X}}^T A^T)\, P(A\tilde{\tilde{X}} + L)$$

$$= L^T PA\tilde{\tilde{X}} + L^T PL + \tilde{\tilde{X}}^T A^T PAX + \tilde{\tilde{X}}^T A^T PL$$

$$= \tilde{\tilde{X}}^T A^T PA\tilde{\tilde{X}} + 2\tilde{\tilde{X}}^T A^T PL + L^T PL$$

$$= \hat{X}^T N\hat{X} + \hat{X}^T N\Delta X + \Delta X^T N\hat{X} + \Delta X^T N\Delta X + 2\hat{X}^T A^T PL + 2\Delta X^T A^T PL + L^T PL$$

$$= \hat{X}^T N\hat{X} + 2\hat{X}A^T PL + L^T PL + 2\Delta X^T N\hat{X} + 2\Delta X^T A^T PL + \Delta X^T N\Delta X$$

$$= (V^T PV)_{\hat{X}} + \Delta X^T N\Delta X + 2\Delta X^T \underbrace{(N\hat{X} + A^T PL)}_{0}$$

$$= (V^T PV)_{\hat{X}} + \underbrace{\Delta X^T N\Delta X}_{\geq 0}$$

because N is positive definite.

Thus,

$$(V^T PV)_{\tilde{X}} \geq (V^T PV)_{\hat{X}}$$

12) The least squares estimate $\hat{\tilde{X}}$ is an unbiased estimate. Since

$$\hat{\tilde{X}} = -N^{-1} A^T P \tilde{L}$$

the expected value is

$$E(\hat{\tilde{X}}) = -N^{-1} A^T P E(\tilde{L})$$

$$= N^{-1} A^T P A X = X$$

where use was made of equation (3.14).

13) Variance-covariance matrix of the adjusted parameters:

$$V(\hat{X}) = E\{\hat{X} - E(\hat{X})\}\{\hat{X} - E(\hat{X})\}^T$$

$$= E[(-N^{-1} A^T P L - N^{-1} A^T P A X)(-N^{-1} A^T P L - N^{-1} A^T P A X)^T]$$

$$= E[\{N^{-1} A^T P (L + AX)\}\{(L + AX)^T P A N^{-1}\}]$$

$$= N^{-1} A^T P E[(L + AX)(L + AX)^T] P A N^{-1}$$

$$= N^{-1} A^T P E[\{L - E(L)\}\{L - E(L)\}] P A N^{-1}$$

$$= N^{-1} A^T P \cdot \sigma_o^2 P^{-1} P A N^{-1} = \sigma_o^2 N^{-1}$$

14) The least squares solution is also a minimum variance solution. Change the least squares estimate $\tilde{X}$ by a $\Delta\tilde{X}$ and compute the variance of the linear function

$${}_1G_n^T \tilde{\tilde{X}} = G^T(\hat{\tilde{X}} + \Delta\tilde{X})$$

with $\hat{\tilde{X}} = -N^{-1} A^T P \tilde{L}$

and $\Delta\tilde{X} = {}_uS_n \tilde{L}$

$$\sigma^2_{G\dot{X}} = (-G^T N^{-1} A^T P \vdots G^T S) \; \Sigma_L \begin{pmatrix} -PAN^{-1}G \\ S^T G \end{pmatrix}$$

$$= \sigma_o^2 \, G^T N^{-1} A^T P P^{-1} P A N^{-1} G + \sigma_o^2 G^T S P^{-1} S^T G$$

$$= G^T \Sigma_X G + \sigma_o^2 \, {}_1G^T S_n \, P^{-1} \, {}_nS^T G_1$$

Since the second term is a positive definite quadratic form it follows that

$$\sigma^2_{G\dot{X}} \geq G^T \Sigma_X G$$

Taking various values for G we see that the parameters $\tilde{\hat{X}}$ have minimum variance.

15) The variance-covariance matrix of $\tilde{V}$ is Σ_{L_b}. Since $E(\tilde{V}) = 0$ we have

$$E(\tilde{V}\tilde{V}^T) = E(AX + \tilde{L})(AX + \tilde{L})^T$$

$$= E(AXL^T + AXX^TA^T + LL^T + LX^TA^T)$$

$$= -AXX^TA + AXX^TA^T + \Sigma_{L_b} + E(L)\,E(L^T) - AXX^TA^T$$

$$= \Sigma_{L_b}$$

In this derivation we have used the relations (2.31) and (3.14). Note that Σ_{L_b} is not the variance-covariance matrix of the linear function

$$\tilde{\ddot{V}} = A\tilde{\hat{X}} + \tilde{L}$$

where the parameter X has been replaced by its estimate, which is a random variable itself.

3.2 Method of Condition Equations

From the n observation equations of the equation system (3.11) we can eliminate the u parameters, having left $r = n - u$ equations in which the residuals are the only unknowns. Thus, we have transformed the n observation equations into r condition equations for the residuals.

Equation (3.11) is

$$V = AX + L$$

Multiply this system from the left by a matrix T_i such that each additional multiplication results in the elimination of another parameter. After u multiplication from the left one arrives at

$$T_{n-u}\,T_{n-u+1}\cdots T_{n-2}T_2\;T_{n-1}T_1\;{}_nV_1 = T_u\cdots T_2T_1AX + T_u\cdots T_2T_1L \qquad (3.36)$$

in which all parameters X are eliminated. Denoting

$$B \equiv T_u\cdots T_2\cdot T_1$$

and

$$BL = -W \qquad (3.37)$$

the condition equations become

$$\boxed{{}_rB_n\,V + W = 0} \qquad (3.38)$$

W is called the vector of discrepancies or the misclosure. Equation (3.38) implies the following important relation

$$\boxed{B\,A = 0} \qquad (3.39)$$

Provided the same criterion of minimization (as in the use of observation equation) is used when solving (3.38) it is clear from the pure algebraic point of view that both adjustment methods have to yield the same results. In practical application one never proceeds as described above. In fact, the problem which is

to be adjusted dictates most of the time the adjustment method to be used. It is, therefore, desireable to derive separately a set of solution equations for the method of condition equations.

The non-linear mathematical (functional) model for condition equations is

$$\boxed{F(L_a) = 0} \tag{3.40}$$

or

$$F(L_b + V) = 0 \tag{3.41}$$

Linearizing this model around the point L_b gives

$$\boxed{r B_{n\,n} V_1 + r W_1 = 0} \tag{3.42}$$

with

$$B = \left.\frac{\partial F(L)}{\partial L}\right|_{L_b} \tag{3.43}$$

and

$$W = F(L_b) \tag{3.44}$$

The residuals are defined as usual

$$V = L_a - L_b$$

The stochastic model is again given by the variance-covariance matrix of the matrix of the observations Σ_{L_b}, with

$$P = \sigma_o^2 \, \Sigma_{L_b}^{-1} \tag{3.45}$$

From (3.38) it is seen that W is a random variable with $E(\tilde{W}) = 0$ since $E(\tilde{V}) = 0$. The variance-covariance matrix of $\tilde{W}$ is

$$\Sigma_W = B \Sigma_{L_b} B^T \tag{3.46}$$

Equation (3.42) is the linearized condition equation. W is the vector of misclosures, B is the coefficient matrix evaluated at the expansion point L_b.

The least squares principle requires that V^TPV be minimized subject to the condition of equation (3.42). This is a standard minimization problem in the calculus of variations. It is solved by introducing a new variable, called the LAGRANGE multiplier, and minimizing

$$\phi(V,K) = V^TPV - 2_1K_r^T\,(_rB_n\,V_1 + W) \qquad (3.47)$$

The stationary point of Q (V,K) is obtained from

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial V}\right) = V^TP - K^TB$$

$$-\frac{1}{2}\left(\frac{\partial \phi}{\partial K}\right) = V^TB^T + W^T$$

Transposing both expressions and equating them to zero gives:

$$PV - B^TK = 0 \qquad (3.48)$$

$$BV + W = 0 \qquad (3.49)$$

This system can be solved for the residuals V and the Lagrange multiplier K. From equation (3.48) it follows that

$$V = P^{-1}B^TK \qquad (3.50)$$

Inserting this expression in (3.49) gives

$$BP^{-1}B^TK + W = 0$$

Using the notation

$$M \equiv BP^{-1}B^T \qquad (3.51)$$

we get

$$\boxed{M\,K + W = 0} \tag{3.52}$$

and

$$K = -M^{-1}W \tag{3.53}$$

Equations (3.52) are called the normal equations. Using equations (3.50), (3.51), and (3.53) the quadratic form becomes

$$\begin{aligned} V^TPV &= K^TMK = -K^TW \\ &= W^TM^{-1}W \end{aligned} \tag{3.54}$$

Finally, the a-posteriori variance of unit weight is computed by

$$\hat{\sigma}_o^2 = \frac{V^TPV}{r} \tag{3.55}$$

where r is the degree of freedom of the adjustment. It is equal to the number of condition equations. The expression (3.55) is, of course, an unbiased estimate of the a priori variance of unit weight σ_o^2. A special proof is not needed.

The variance-covariance matrix for the Lagrange multiplier can be derived from equation (3.53):

$$\begin{aligned} \Sigma_K &= \sigma_o^2\, M^{-1}\, M\, M^{-1} \\ &= \sigma_o^2\, M^{-1} \end{aligned} \tag{3.57}$$

For the variance-covariance matrix of the residuals we obtain from eq. (3.50) and (3.57)

$$\Sigma_V = \sigma_o^2\, P^{-1}B^TM^{-1}BP^{-1} \tag{3.58}$$

The estimated variance-covariance matrix based on the sample L_b, of course, is

$$\hat{\Sigma}_V = \hat{\sigma}_o^2 \, P^{-1} B^T M^{-1} B P^{-1} \tag{3.59}$$

Finally, the variance covariance matrix of the adjusted observations is derived from equations (3.50)

$$\begin{aligned} L_a &= L_b + V \\ &= L_b - P^{-1} B^T M^{-1} W \\ &= L_b + P^{-1} B^T M^{-1} B L \\ &= L_b - P^{-1} B^T M^{-1} B L_b + P^{-1} B^T M^{-1} B L_o \end{aligned}$$

Since using the law of variance propagation (L_o is a constant):

$$\Sigma_{L_a} = \sigma_o^2 \, (P^{-1} - P^{-1} B^T M^{-1} B P^{-1}) \tag{3.60}$$

Comparing (3.58) and (3.60) we obtain again

$$\Sigma_{L_a} = \Sigma_{L_b} - \Sigma_V \tag{3.61}$$

For the vector $Z = \begin{pmatrix} K \\ V \\ L_a \end{pmatrix}$

one obtains the variance-covariance matrix

$$\Sigma \begin{pmatrix} K \\ V \\ L_a \end{pmatrix} = \sigma_o^2 \begin{pmatrix} M^{-1} & M^{-1} B P^{-1} & 0 \\ P^{-1} B M & P^{-1} B^T M^{-1} B P^{-1} & 0 \\ 0 & 0 & P^{-1} - P^{-1} B^T M^{-1} B P^{-1} \end{pmatrix} \tag{3.62}$$

In the following a few of the properties of the method of conditions equations are listed:

1) The expression $-W^T M^{-1} W$ which is needed to compute $V^T PV$ is obtained immediately from Cholesky's algorithm.

2) Comparing the result of 1) with $K^T W$ as computed from equation (3.53) provides a useful check for programming errors.

3) Since the functional model (3.40) is usually non-linear, the adjustment has to be iterated. Note that the non-linear relation is exactly zero only for the adjusted observations (true values) L_a. The point of expansion in (3.43) is L_b. Therefore, the linearization might not be sufficient if the residuals $V = L_a - L_b$ are large. For the i-th adjustment let us choose as expansion point L_i which is equal to the adjusted observation of the previous adjustment:

$$L_i = L_{i-1} + \overline{V}_{i-1}$$

with

$$\overline{V}_o = 0 \text{ and } L_o = L_b$$

then

$$B_i \overline{V}_i + W_i = 0$$

$$B_i = \left.\frac{\partial F(L)}{\partial L}\right|_{L_i}$$

$$W_i = F(L_i)$$

$$\overline{V}_i = L_{i+1} - L_i$$

$\overline{V}_i$ is not identical to the conventional residuals V. It is clear that, as the number of iterations increases, $\overline{V}_i$ tends toward zero since the point of expansion L_i moves toward L_a. Introducing V in the algorithm above we write

$$V_i = L_{i+1} - L_b$$

Thus

$$\bar{V}_i = V_i + (L_b - L_i)$$

with

$$L_1 = L_b$$

The linearized functional model becomes:

$$B_i V_i + W_i + B_i(L_b - L_i) = 0$$

or

$$B_i V_i + W_i' = 0$$

The discrepancy vector W_i' consists now of two terms. The second term results from the fact that we have preferred to replace $\bar{V}_i$ by V_i. The latter quantity has the advantage that it converges toward V as i increases. In the the first adjustment, i=1, the second term in W' is zero since $L_1 = L_b$.

Step 1: $L_1 = L_b$; compute B_1, $W_1' = W_1 - 0 = F(L_b)$
adjustment gives V_1, $V_1^T P V_1$

Step 2: $L_2 = V_2 + L_b$; compute B_2, $W_2' = F(L_2) + B_2(L_b - L_2)$;
adjustment gives V_2, $V_2^T P V_2$

Step 3: If $V_2^T P V_2 - V_1^T P V_1 < |\varepsilon|$, where ε is sufficiently small then the adjustment has converged. Repeat the iterations if necessary.

Note: $V_i \to V = L_a - L_b$
as i gets larger.

4) As for scaling the weight matrix and changing the units of the observations, the same rules apply as in the case of observation equations.

5) If one adjustment is carried out by both methods then the adjusted observations in both cases have to be the same. From equations (3.32) and (3.60) the following equality is obtained:

$$AN^{-1}A^T = P^{-1} - P^{-1}B^T M^{-1} BP^{-1} \tag{3.63}$$

For proving this relation one should multiply from the left and then from the right by the matrix B and use the relation BA = 0.

Multiplying (3.63) by the non-singular symmetric matrix P either from the left or right gives

$$AN^{-1}A^T P = I - P^{-1}B^T M^{-1} B \tag{3.64}$$

Using the symbols

$$H_A = AN^{-1}A^T P$$

$$H_B = P^{-1}B^T M^{-1} B$$

then H_H, H_B, $I - H_A$ and $I - H_B$ are idempotent matrices because the product of each of the matrices with itself is the same matrix again.

6) Using the relation (3.63) it can be easily shown that $V^T PV$ is the same for both adjustment methods.

3.3. Mixed Models

In this case the observations and the parameters have an implicit functional relationship. Referring again to the system of observation equations (3.11) we now eliminate only a subset of p unknowns. There are left n-p = n' equations in which the residuals and the remaining parameters are related implicitly. In analogy to (3.36) we write

$$B'V = A'X + B'L \tag{3.65}$$

with

$$A' = B'A$$

These are n'=n-p equations containing u'=u-p unknowns. Note that the degree of freedom is DF = n-u=n'-u'. Applying the same criterion of minimization as in the case of observation equations, we must obtain an identical solution for the remaining parameters. The above technique can be used to eliminate so-called nuisance parameters.

For those cases where the relationship between the observations and the parameters is implicit by nature it is beneficial to derive the complete algorithm separately. It may be worthwhile to know that in cases in which the B' matrix is of a simple form, one might create artificial observations for which the B' matrix becomes the identity matrix so that the algorithm for observation equations is applicable.

In the subsequent development the symbols A, B, W, N, u and r will be used. Please note that identical symbols in the previous two models have only "formally" the same meaning. The number of equations is denoted by r. This includes all types of equations, i.e., those in which the observations and the parameters occur implicitly or explicitly, and those containing only parameters. It follows from the observation equation model that n + u > r since it is principally possible to add u-1 conditions about the parameters to the n observation equations. On the other side, r > u since the possible elimination of the u parameters must result in at least one condition equation; this ensures a degree of freedom of at least one. Thus

$$n + u > r > u \tag{3.66}$$

The functional model is

$$F(L_a, X_a) = 0 \tag{3.67}$$

Linearizing this function around the point (L_b, X_o)

$$\boxed{{}_rB_n V + {}_rA_u X_1 + W = 0} \tag{3.68}$$

whereby

$${}_rB_n = \left.\frac{\partial F}{\partial L}\right|_{X_o, L_b} \tag{3.69}$$

$${}_rA_u = \left.\frac{\partial F}{\partial X}\right|_{X_o, L_b} \tag{3.70}$$

$${}_rW_1 = F(L_b, X_o) \tag{3.71}$$

r: Number of equations of the type indicated in (3.67), in which observations and parameters may appear implicitly or explicitly, or which may contain only parameters or observations

u: Number of parameters

n: Number of observations

$\tilde{V}, \tilde{W}$: Random variables

A least squares solution is found by minimizing $V^T P V$ subject to the condition (3.68). This leads to the minimizing of the function

$$\Phi(V, K, X) = V^T P V - 2K^T(BV + AX + W) \tag{3.71}$$

where K is the vector of the Lagrange multipliers. The partial derivatives are

$$\frac{1}{2}\left(\frac{\partial \Phi}{\partial V}\right) = V^T P - K^T B$$

$$\frac{1}{2}\left(\frac{\partial \Phi}{\partial X}\right) = -K^T A$$

$$\frac{1}{2}\frac{\partial \phi}{\partial K} = V^T B^T + \hat{X}^T A^T + W^T$$

Transposing the above expressions and equating them to zero gives

$$n \quad -PV + B^T K = 0 \tag{3.72}$$

$$r \quad BV + A\hat{X} = -W \tag{3.73}$$

$$u \quad A^T K = 0 \tag{3.74}$$

These three equations can be written in a matrix as follows:

$$\begin{pmatrix} {}^{n}-P_{n} & {}^{n}B_{r}^{T} & {}^{n}0_{u} \\ {}^{r}B_{n} & {}^{r}0_{r} & {}^{r}A_{u} \\ {}^{u}0_{n} & {}^{u}A_{r}^{T} & {}^{u}0_{u} \end{pmatrix} \begin{pmatrix} {}^{n}V_{1} \\ {}^{r}K_{1} \\ {}^{u}\hat{X}_{1} \end{pmatrix} = \begin{pmatrix} {}^{n}0_{1} \\ {}^{r}-W_{1} \\ {}^{u}0_{1} \end{pmatrix} \tag{3.75}$$

Equation (3.72) yields

$$\boxed{V = P^{-1} B^T K} \tag{3.76}$$

Using this expression we can eliminate the residuals V from equation (3.73) giving

$$\begin{pmatrix} -BP^{-1}B^T & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} -K \\ \hat{X} \end{pmatrix} = \begin{pmatrix} -W \\ 0 \end{pmatrix} \tag{3.77}$$

Using the notation

$$\boxed{{}^{r}M_{r} \equiv {}^{r}B_{n}\, P_{n}^{-1}\, {}_{n}B_{r}^{T}} \tag{3.78}$$

the solution for the Lagrange multiplier is

$$\boxed{K = -M^{-1}(A\hat{X} + W)} \tag{3.79}$$

The adjusted parameters are

$$-A^T M^{-1} A\hat{X} - A^T M^{-1} W = 0$$

or

$$\boxed{\hat{X} = -(A^T M^{-1} A)^{-1} A^T M^{-1} W} \qquad (3.80)$$

An estimate of the a posteriori variance of unit weight is computed in the usual way. Equation (3.76) provides an expression for the quadratic form

$$V^T P V = K^T B P^{-1} P P^{-1} B^T K = K^T M K$$

Using equation (3.79)

$$V^T P V = -K^T M M^{-1} (A\hat{X} + W) = -K^T A\hat{X} - K^T W$$

Substituting eq. (3.74) gives

$$V^T P V = -K^T W \qquad (3.81)$$

$$= W^T M^{-1} W + W^T M^{-1} A\hat{X}$$

The estimate of the a-posteriori variance of unit weight is

$$\boxed{\hat{\sigma}_o^2 = \frac{V^T P V}{r-u}} \qquad (3.82)$$

whereby r-u is the degree of freedom.

The cofactor matrix Q_W is readily derived from equation (3.71) and the law of variance propagation:

$$Q_W = \left|\frac{\partial F}{\partial L}\right| P^{-1} \left|\frac{\partial F}{\partial L}\right|^T = B P^{-1} B^T = M \qquad (3.83)$$

Note that the approximate coordinates X_o in equation (3.71) are constants. Applying the law of error propagation to equation (3.80) gives the cofactor for the adjusted parameters,

$$Q_X = (A^T M^{-1} A)^{-1} A^T M^{-1} M M^{-1} A (A^T M^{-1} A)^{-1} \tag{3.84}$$

$$\boxed{Q_X = (A^T M^{-1} A)^{-1}}$$

and the variance-covariance matrix is

$$\Sigma_X = \sigma_o^2 Q_X \tag{3.85}$$

The estimated variance-covariance matrix is

$$\boxed{\hat{\Sigma}_X = \hat{\sigma}_o^2 Q_X} \tag{3.86}$$

We can also invert the matrix of equation (3.77) and compare the corresponding submatrix with Q_X in equation (3.84). Using the expression of equation (1.33) for inverting a partitioned matrix gives

$$Q_{22} = (A^T M A)^{-1}$$

which is identical to the cofactor matrix (3.84).

The expression (3.82) for the a posteriori-variance of unit weight can easily be derived using the expectation operator:

$$\hat{V}^T P \hat{V} = -K^T W$$

$$= (A\hat{X} + W)^T M^{-1} W$$

$$= \hat{X}^T A^T M^{-1} W + W^T M^{-1} W$$

$$= -W^T M^{-1} A (A^T M^{-1} A)^{-1} A^T M^{-1} W + W^T M^{-1} W$$

$$= W^T (M^{-1} - M^{-1} A (A^T M^{-1} A)^{-1} A^T M^{-1}) W$$

$$E(\tilde{\hat{V}}^T P \tilde{\hat{V}}) = E(\mathrm{Tr}\, \tilde{\hat{V}}^T P \tilde{\hat{V}})$$

$$= E\{\mathrm{Tr}(M^{-1} - M^{-1} A (A^T M^{-1} A)^{-1} A^T M^{-1}) \tilde{W}\tilde{W}^T\}$$

$$= \mathrm{Tr}\{(M^{-1} - M^{-1}A(A^TM^{-1}A)^{-1}A^TM^{-1}\}\, M\, \sigma_o^2$$

since $E(\tilde{W}\tilde{W}^T) = \sigma_o^2 M + E(\tilde{W})\,E(\tilde{W}^T) = \sigma_o^2 M + AXX^TA^T$. Thus,

$$E(\tilde{V}^TP\tilde{V}) = \mathrm{Tr}\{{}_rI_r - M^{-1}A(A^TM^{-1}A)^{-1}A^T\}$$

$$= \mathrm{Tr}\{{}_rI_r - (A^TM^{-1}A)^{-1}(A^TM^{-1}A)\}$$

$$= \mathrm{Tr}\{{}_rI_r - {}_uI_u\}$$

$$= r - u$$

Therefore $\tilde{\sigma}_o^2$ is an unbiased estimate of σ_o^2.

Finally, we need the variance-covariance of the residuals and the adjusted observations. From equation (3.76)

$$\hat{V} = -P^{-1}B^TM^{-1}(A\hat{X} + W)$$

$$= -P^{-1}B^TM^{-1}\{-A(A^TM^{-1}A)^{-1}A^TM^{-1}W + W\}$$

$$= \{P^{-1}B^TM^{-1}A(A^TM^{-1}A)^{-1}A^TM^{-1} - P^{-1}B^TM^{-1}\}W$$

With $\Sigma_W = \sigma_o^2 M$ the law of variance propagation gives:

$$\boxed{\Sigma_V = \sigma_o^2\, P^{-1}B^TM^{-1}\{M - A(A^TM^{-1}A)^{-1}A^T\}M^{-1}BP^{-1}} \qquad (3.87)$$

For the adjusted observations we obtain

$$L_a = L_b + \hat{V}$$

$$= L_b + (P^{-1}B^TM^{-1}AM^{-1}A^TM^{-1} - P^{-1}B^TM^{-1})\,W$$

$$\frac{\partial L_a}{\partial L_b} = I + P^{-1}B^TM^{-1}AM^{-1}A^TM^{-1}B - P^{-1}B^TM^{-1}B$$

$$\Sigma_{L_a} = \left|\frac{\partial L_a}{\partial L_b}\right| \sigma_o P^{-1} \left|\frac{\partial L_a}{\partial L_b}\right|^T$$

$$= \sigma_o P^{-1} - \Sigma_V$$

Thus,

$$\boxed{\Sigma_{L_a} = \Sigma_{L_b} - \Sigma_V} \qquad (3.88)$$

Remarks:

1) The mixed model includes the previously discussed models. With A≡0 we obtain formally the usual condition equations. Taking B≡-I and L≡W leads formally to the observation equation model.

2) Equation (3.54) provides two methods for computing V^TPV. They are a useful check for programming errors.

3) The mixed model requires also iterations. Recall that the nonlinear model is only fulfilled for (L_a, X_a). Let's take (L_i, X_i) as the point of expansion for the i-th adjustment:

$$L_i = L_{i-1} + \overline{V}_{i-1}$$

$$\overline{V}_o = 0, \quad L_o = L_b$$

$$X_i = X_{i-1} + x^{i-1}$$

$$x^o = 0$$

Using this notation the linearized model becomes

$$B_{X_i L_i} \overline{V}_i + A_{X_i L_i} x^i + W_{X_i L_i} = 0$$

The quantity $\overline{V}_i$ tends toward zero as the expansion point moves toward L_a. Introducing the residuals $V = L_a - L_b$ gives

$$V_i = L_{i+1} - L_b$$

and

$$\overline{V}_i = V_i + (L_b - L_i)$$

The linearized functional model becomes now

$$B_{X_i L_i} V_i + A_{X_i L_i} X^i + \{W_{X_i L_i} + B_{X_i L_i} (L_b - L_i)\} = 0$$

This formulation permits the point of expansion (L_i, X_i) to move toward (L_a, X_a). Note that a second term is added to the discrepancy vector. This term results from the introduction of the residuals V_i instead of the quantity $\overline{V}_i$.

Step 1: $L_1 = L_b$, $X_1 = X_o$.

Compute A_1, B_1, $W_1 = F(X_1, L_1)$

Adjustment gives X^1, V_1, $V_1^T P V_1$

Step 2: $L_2 = V_1 + L_b$, $X_2 = X_1 + X^1$

Compute A_2, B_2, $W_2 + B_2(L_b - L_2)$

Adjustment gives X^2, V_2, $V_2^T P V_2$

Step 3: If $V_2^T P V_2 - V_1^T P V_1 < |\varepsilon|$

then the adjustment has converged;

otherwise repeat the iteration.

4) The iteration procedure for the models of condition or observation equations is readily derived from the above procedure.

For condition equations

$$F(L_a) = 0 \text{ set } A = \frac{\partial F(L_a)}{\partial X} \equiv 0$$

$$B_{L_i} V_i + W_{L_i} + B_{L_i} (L_b - L_i) = 0$$

The B matrix is only a function of the observations in this particular case.

For observation equations

$$F(X_a) - L_a = 0 \text{ set } B_i = \frac{\partial (F(X_a) - L_a)}{\partial L_a} = -I$$

and $W_i = F(X_i) - L_i - I(L_b - L_i) = F(X_i) - L_b = L_i$

$$-V_i + A_{X_i} X^i + L_i = 0$$

The A matrix is only a function of the parameters in this particular case.

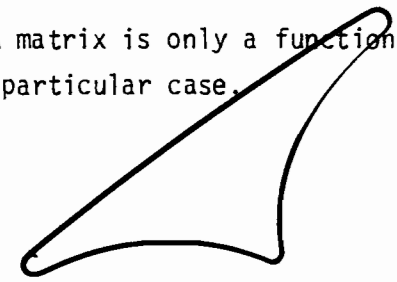


Table 3.1 Expressions of the Three Standard Adjustment Models

$F(X_a, L_a) = 0$	$L_a = F(X_a)$	$F(L_b) = 0$
$BV + AX + W = 0$	$B = -I \quad L=W,\ r=n$ $V = AX + L$	$A = 0$ $BV + W = 0$
$M = BP^{-1}B^T$ $N = A^TM^{-1}A$ $U = A^TM^{-1}W$	$M = P^{-1}$ $N = A^TPA$ $U = A^TPL$	$M = BP^{-1}B^T$
$N\hat{X} = -U \quad \hat{X} = -N^{-1}U$	$N\hat{X} = -U \quad \hat{X} = -N^{-1}U$	
$K = -M^{-1}(A\hat{X} + W)$		$K = -M^{-1}W$
$\hat{V}^TP\hat{V} = -K^TW$ $= U^T\hat{X} + W^TM^{-1}W$ $= -U^TN^{-1}U + W^TM^{-1}W$	$\hat{V}^TP\hat{V} = U^T\hat{X} + L^TPL$ $= -U^TN^{-1}U + L^TPL$	$\hat{V}^TP\hat{V} = -K^TW$ $= W^TM^{-1}W$
$\hat{V} = P^{-1}B^TK$	$\hat{V} = A\hat{X} + L$	$\hat{V} = P^{-1}B^TK$
$\hat{\sigma}_o^2 = \frac{\hat{V}^TP\hat{V}}{r-u}$	$\hat{\sigma}_o^2 = \frac{\hat{V}^TP\hat{V}}{r-u} = \frac{\hat{V}^TP\hat{V}}{n-u}$	$\hat{\sigma}_o^2 = \frac{\hat{V}^TP\hat{V}}{r}$
$\hat{\Sigma}_X = \hat{\sigma}_o^2 N^{-1}$	$\hat{\Sigma}_X = \hat{\sigma}_o^2 N^{-1}$	
$\hat{\Sigma}_V = \hat{\sigma}_o^2 P^{-1}B^T(M-AN^{-1}A^T)\,M^{-1}BP^{-1}$	$\hat{\Sigma}_V = \hat{\sigma}_o^2 (P^{-1}-AN^{-1}A^T)$	$\hat{\Sigma}_V = \hat{\sigma}_o^2 P^{-1}B^TM^{-1}BP^{-1}$
$\hat{\Sigma}_{L_a} = \hat{\Sigma}_{L_b} - \hat{\Sigma}_V$	$\hat{\Sigma}_{L_a} = \hat{\Sigma}_{L_b} - \hat{\Sigma}_V$	$\hat{\Sigma}_{L_a} = \hat{\Sigma}_{L_b} - \hat{\Sigma}_V$

3.4 Conditions between Parameters

In order to obtain more flexibility in our adjustment procedures, we would like to include additional conditions between some or all parameters. The conditions which the adjusted parameters should fulfill are s equations of the type

$$G(X_a) = 0 \tag{3.89}$$

The number of conditions must be less than the number of parameters, i.e., $s < u$. The linearization gives

$$\left.\frac{\partial G}{\partial X}\right|_{X_o} X + G(X_o) = 0 \tag{3.89}$$

or

$$\boxed{C\,X + W_C = 0} \tag{3.90}$$

where

$$C = \left.\frac{\partial G}{\partial X}\right|_{X_o} \quad \text{and} \quad W_C = G(X_o) \tag{3.91}$$

To be most general we consider the mixed model plus some extra conditions:

$$\begin{aligned} F(L_a, X_a) &= 0 \\ G(X_a) &= 0 \end{aligned} \tag{3.92}$$

The linear form is

$$BV + AX + W = 0 \tag{3.93}$$

$$CX + W_C = 0 \tag{3.94}$$

Note that W_C is no random variable!

The least squares solution of (3.93) and (3.94) follows immediately from the solution of the mixed model in Section (3.3) using submatrices. We can write

$$\bar{B}V + \bar{A}X + \bar{W} = 0$$

with $\bar{B} = \begin{vmatrix} B \\ 0 \end{vmatrix}$ $\bar{A} = \begin{vmatrix} A \\ C \end{vmatrix}$ and $\bar{W} = \begin{vmatrix} W \\ W_C \end{vmatrix}$

However, in order to gain more experience with the principle of least squares we can derive the solution from the beginning by minimizing V^TPV subject to the conditions (3.93) and (3.94)

$$\Phi(V, X, K_L, K_C) = V^TPV - 2K_L^T(BV + AX + W) - 2K_C^T(CX + W_C)$$

where K_L and K_C are vectors of Lagrange multipliers. Taking the partial derivatives and equating them to zero gives

$$\begin{aligned} \frac{1}{2}\left(\frac{\partial\Phi}{\partial V}\right)^T &\equiv -P\hat{V} + B^TK_L = 0 \\ \frac{1}{2}\left(\frac{\partial\Phi}{\partial X}\right)^T &\equiv A^TK_L + C^TK_C = 0 \\ \frac{1}{2}\left(\frac{\partial\Phi}{\partial K_L}\right)^T &\equiv B\hat{V} + A\hat{X} + W = 0 \\ \frac{1}{2}\left(\frac{\partial\Phi}{\partial K_C}\right)^T &\equiv C\hat{X} + W_C = 0 \end{aligned} \qquad (3.95)$$

These equations can be written in matrix form as

$$\begin{pmatrix} \bar{n}P_n & {}_nB_r^T & {}_nO_u & {}_nO_s \\ {}_rB_n & {}_rO_r & {}_rA_u & {}_rO_s \\ {}_uO_n & {}_uA_r^T & {}_uO_u & {}_uC_s^T \\ {}_sO_n & {}_sO_r & {}_sC_u & {}_sO_s \end{pmatrix} \begin{pmatrix} {}_n\hat{V}_1 \\ {}_rK_{L_1} \\ {}_u\hat{X}_1 \\ {}_sK_{C_1} \end{pmatrix} = \begin{pmatrix} {}_nO_1 \\ \bar{r}W_1 \\ {}_uO_1 \\ {}_sW_{C_1} \end{pmatrix} \qquad (3.96)$$

The dimensions must fulfill the unequality

$$n + u > r + s > u \qquad (3.97)$$

or

$$n + (u-s) > r > (u-s)$$

The first equation of (3.96) gives for the residuals

$$\boxed{\hat{V} = P^{-1}B^T K_L} \tag{3.98}$$

The equations for the remaining three parameters are

$$\begin{pmatrix} BP^{-1}B^T & A & 0 \\ A^T & 0 & C^T \\ 0 & C & 0 \end{pmatrix} \begin{pmatrix} K_L \\ \hat{X} \\ K_C \end{pmatrix} = \begin{pmatrix} -W \\ 0 \\ -W_C \end{pmatrix} \tag{3.99}$$

The first equation of (3.99) gives

$$K_L = -(BP^{-1}B^T)^{-1} W - (BP^{-1}B^T)^{-1} A\hat{X}$$

Denoting

$$M \equiv BP^{-1}B^T \tag{3.100}$$

then

$$\boxed{K_L = -M^{-1}W - M^{-1}A\hat{X}} \tag{3.101}$$

Substituting this expression in the 2nd equation of (3.99) gives

$$-A^T M^{-1} W - A^T M^{-1} A\hat{X} + C^T K_C = 0$$

or

$$-A^T M^{-1} A\hat{X} + C^T K_C = A^T M^{-1} W \tag{3.102}$$

We are left with the following system:

$$\begin{pmatrix} -A^T M^{-1} A & C^T \\ C & 0 \end{pmatrix} \begin{pmatrix} \hat{X} \\ K_C \end{pmatrix} = \begin{pmatrix} A^T M^{-1} W \\ -W_C \end{pmatrix}$$

or

$$\begin{pmatrix} A^T M^{-1} A & C^T \\ C & 0 \end{pmatrix} \begin{pmatrix} \hat{X} \\ K_C \end{pmatrix} = \begin{pmatrix} -A^T M^{-1} W \\ -W_C \end{pmatrix} \tag{3.103}$$

This equation system shows how the normal equations can be augmented to incorporate a constraint (condition)!

The first equation of (3.103) gives an expression for the parameters

$$A^T M^{-1} A \hat{X} - C^T K_C = -A^T M^{-1} W$$

having the solution

$$\boxed{\hat{X} = (A^T M^{-1} A)^{-1} (C^T K_C - A^T M^{-1} W)} \qquad (3.104)$$

Inserting this expression into the second equation of (3.103) gives

$$C(A^T M^{-1} A)^{-1} C^T K_C - C(A^T M^{-1} A)^{-1} A^T M^{-1} W = -W_C$$

The solution for the Lagrange multiplier becomes

$$\boxed{K_C = \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \{-W_C + C(A^T M^{-1} A)^{-1} A^T M^{-1} W\}} \qquad (3.105)$$

The equations (3.98), (3.101), (3.104), and (3.105) constitute the least squares solution for the system (3.92).

It is sometimes of interest to solve the above least squares problem in a manner such that we first find the solution without the additional conditions and then compute the effect of the conditions on the solution separately (sequential algorithm). Substituting the solution (3.105) for K_C into (3.104) gives

$$\hat{X} = (A^T M^{-1} A)^{-1} C^T K_C - (A^T M^{-1} A)^{-1} A^T M^{-1} W \qquad (3.106)$$

The second term is exactly the solution of (3.92) <u>without</u> the conditions. Denoting this solution by the symbol "*" we get

$$\hat{X} = X^* + \delta X \qquad (3.107)$$

with $$X^* = -(A^T M^{-1} A)^{-1} A^T M^{-1} W \qquad (3.108)$$

$$\delta X = (A^T M^{-1} A)^{-1} C^T K_C \qquad (3.109)$$

From the equations (3.108) and (3.105) it follows that

$$K_C = -\{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \{W_C + CX^*\} \tag{3.110}$$

and thus

$$\boxed{\delta X = -(A^T M^{-1} A)^{-1} C^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \{W_C + CX^*\}} \tag{3.111}$$

Finally, we can compute the quadratic form $V^T PV$. From equation (3.98) we have

$$\begin{aligned} \hat{V}^T P \hat{V} &= K_L^T B P^{-1} P P^{-1} B^T K_L \\ &= K_L^T B P^{-1} B^T K_L \end{aligned} \tag{3.112}$$

Transposing the equations (3.99), multiplying as indicated below

$$\begin{array}{lll} K_L^T BP^{-1}B^T + \hat{X}^T A^T + W^T = 0 & \Big| K_L \\ -X^T \Big| \quad A^T K_L + C^T K_C = 0 \\ \hat{X}^T C^T + W_C^T = 0 & \Big| K_C \end{array}$$

and adding, gives

$$K_L^T BP^{-1}B^T K_L + \hat{X}^T A^T K_L + W^T K_L - \hat{X}^T A^T K_L - \hat{X}^T C^T K_C + \hat{X}^T C^T K_C + W_C^T K_C = 0$$

Comparing this expression with (3.112) gives

$$\hat{V}^T P \hat{V} = -W^T K_L - W_C^T K_C \tag{3.113}$$

Substituting the respective expressions for K_L and K_C yields an alternative expression

$$\hat{V}^T P \hat{V} = W^T M^{-1}(W + A\hat{X}) - W_C^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \{-W_C - CX^*\} \tag{3.114}$$

This expression can be changed further to identify separately the contribution resulting from the constraints. Remember that for the first system only, i.e. the mixed model without constraints, we had according to equation (3.81)

$$V^T PV^* = -W^T K$$

with $K \neq K_L$. The analogous expression for the quadratic form in the case of constraints is included in (3.113)

$$V^T PV = -WK_L^T - W_C^T K_C$$

The Lagrange multiplier K_L can be written according to (3.101)

$$K_L = -M^{-1}(W + A\hat{X})$$

$$= -M^{-1}\{W + A(X^* + \delta X)\}$$

$$= \underbrace{-M^{-1}W - M^{-1}AX^*}_{K} \underbrace{- M^{-1}A\delta X}_{\Delta K}$$

Comparing this expression with (3.79) we see that

$$K_L = K + \Delta K$$

with $\Delta K = -M^{-1}A\delta X.$

The quadratic form (3.113) can now be given in the form

$$V^T PV = V^T PV^* + \Delta V^T PV$$

$$= -W^T K - W^T \Delta K - W_C^T K_C \qquad (3.115)$$

with $\Delta V^T PV = -W^T \Delta K - W_C^T K_C$

$$= W^T M^{-1} A\delta X - W_C^T K_C$$

$$= -W^T M^{-1} A \,(A^T M^{-1} A)^{-1} \; C^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \{W_C + CX^*\}$$

$$+ W_C^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \{W_C + CX^*\}$$

$$\boxed{\Delta V^T PV = \{CX^* + W_C\}^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \{CX^* + W_C\}} \qquad (3.116)$$

We can observe the following properties:

1) ΔVPV can be computed without having computed $\hat{X}$ (we need only X^*).

2) Since $\{C(A^TM^{-1}A)^{-1}C^T\}$ is a positive definite matrix, the additive term is a positive definite quadratic form,

$$\Delta V^TPV \geq 0$$

Thus, any additional constraint increases V^TPV.

The degree of freedom for the combined solution is r + s - u. Thus, the estimate of the a posteriori variance of unit weight is

$$\hat{\sigma}_o^2 = \frac{\hat{V}^TP\hat{V}}{r+s-u} \tag{3.117}$$

The variance-covariance matrix of the parameters can be derived as usual by applying the law of propagation of variances. The parameters X are, according to equations (3.105) and (3.106),

$$\hat{X} = (A^TM^{-1}A)^{-1}\left[C^T\{C(A^TM^{-1}A)^{-1}C^T\}^{-1}\{-W_C + C(A^TM^{-1}A)^{-1}A^TM^{-1}W\} - A^TM^{-1}W\right] \tag{3.118}$$

The only random variable is W since W_C is a constant vector. The cofactor matrix of W is again $Q_W = M$. Thus

$$\frac{\partial\hat{X}}{\partial W} = G = (A^TM^{-1}A)^{-1}\left[C^T\{C(A^TM^{-1}A)^{-1}C^T\}^{-1}C(A^TM^{-1}A)^{-1}A^TM^{-1} - A^TM^{-1}\right]$$

Using the notation

$$N \equiv A^TM^{-1}A$$

$$S \equiv \{C(A^TM^{-1}A)^{-1}C^T\}^{-1} = (CN^{-1}C^T)^{-1}$$

We obtain

$$Q_X = GQ_WG^T$$

$$= N^{-1}(C^TSCN^{-1}A^TM^{-1} - A^TM^{-1})\,M\,(M^{-1}AN^{-1}C^TSC - M^{-1}A)\,N^{-1}$$

$$= N^{-1} - N^{-1}C^TSCN^{-1}$$

Thus

$$Q_X = (A^T M^{-1} A)^{-1} - (A^T M^{-1} A)^{-1} C^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} C(A^T M^{-1} A)^{-1} \quad (3.119)$$

Note that the first term in (3.119) is the cofactor matrix of X*. We can, therefore, write

$$Q_X = Q_{X*} - \Delta Q_X \quad (3.120)$$

Each term in the diagonal of ΔQ_X is positive since $\{C (A^T M^{-1} A)^{-1} C^T\}$ is a positive definite matrix. Therefore $q_{x_i} < q_{x_i}^*$, which confirms the general statement that the introduction of conditions on the parameters reduces their variances.

The expressions for $\hat{X}$, K_C and Q_X derived above can also be obtained using matrix partitioning techniques on equation (3.103). Using equation (1.33) we find, e.g. for the upper left submatrix

$$Q_{11} = (A^T M^{-1} A)^{-1} - (A^T M^{-1} A)^{-1} C^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} C(A^T M^{-1} A)^{-1}$$

We find that $Q_{11} \equiv Q_X$. The parameters X can be derived in a similar manner by matrix partitioning.

The variance-covariance matrix of the residuals follows from equations (3.98), (3.108), and (3.111)

$$\hat{V} = -P^{-1} B^T M^{-1} (W + AX^* + A\delta X)$$

$$= -P^{-1} B^T M^{-1} (W - AN^{-1} A^T M^{-1} W + AN^{-1} C^T S C N^{-1} A^T M^{-1} W) + \text{constant term}$$

where

$$N \equiv A^T M^{-1} A$$

and

$$S \equiv \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} = (CN^{-1} C^T)^{-1}$$

The partial derivative is

$$\frac{\partial \hat{V}}{\partial W} = -P^{-1} B^T M^{-1} (I - AN^{-1} A^T M^{-1} + AN^{-1} C^T S C N^{-1} A^T M^{-1})$$

With $Q_W = M$ it follows that

$$Q_V = \left|\frac{\partial V}{\partial W}\right| Q_W \left|\frac{\partial V}{\partial W}\right|^T$$

$$= P^{-1}B^T M^{-1} (M - AN^{-1}A^T + AN^{-1}C^T SCN^{-1}A^T) M^{-1}B P^{-1}$$

Comparing this expression with equation (3.87), which gives the cofactor matrix without the condition, denoted here by Q_{V*}, we can write

$$Q_V = Q_{V*} + \Delta Q_V$$

with

$$\Delta Q_V = P^{-1}B^T M^{-1}(AN^{-1}C^T SCN^{-1}A^T) M^{-1}B P^{-1}$$

Substituting ΔQ_X from equations (3.119) and (3.120) this expression becomes

$$\boxed{\Delta Q_V = P^{-1}B^T M^{-1} (A \Delta Q_X A^T) M^{-1}B P^{-1}}$$

Note that the non-linear combined system (3.92) must be iterated. If we first find the solution of $F(L_a, X_a) = 0$, including the iterations, then for computing the effects of the condition $G(X_a) = 0$ we have to evaluate the matrices C and W_C at the last expansion point X_i! In the final solution the expansion point must be the same for all matrices.

Alternative derivation: It seems instructive to repeat part of the previous derivation in a slightly different manner. Consider the models (3.93) and (3.94)

$$BV + AX + W = 0$$

$$CX + W_C = 0$$

The solution of the first system only, is

$$X* = -(A^T M^{-1} A)^{-1} A^T M^{-1} W$$

$$\Sigma_{X*} = \sigma_o^2 (A^T M^{-1} A)^{-1}$$

In a second step we find the solution of the condition equation (3.94) considering only X* as observations. Thus

$$\left.\begin{aligned} & C(X^* + \Delta X) + W_C = 0 \\ \text{or} \quad & C\Delta X + \overline{W}_C = 0 \\ \text{with} \quad & \overline{W}_C = W_C + CX^* \\ \text{and} \quad & \Sigma_{X^*} = \sigma_o^2 (A^T M^{-1} A)^{-1} \end{aligned}\right\} \qquad (3.121)$$

The "parameter correction" takes on the role of "residual" with a weight matrix

$$P_{X^*} = (A^T M^{-1} A)$$

The least squares solution of (3.121) is given by equations (3.50) and (3.51) of Section 3.2

$$\Delta X = -(A^T M^{-1} A)^{-1} C^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \overline{W}_C \qquad (3.122)$$

which is identical to the solution (3.111). We can now compute the "$V^T PV$" for the second step:

$$\begin{aligned} \Delta X^T P_{X^*} \Delta X &= \overline{W}_C{}^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} \overline{W}_C \qquad (3.123) \\ &= \Delta V^T PV \end{aligned}$$

which is identical to the extra sum $\Delta V^T PV$ due to the conditions as seen from equation (3.116). From equations (3.58) and (3.61) the cofactor matrix for the adjusted "residuals" (parameters) becomes

$$Q_{\Delta X} = (A^T M^{-1} A)\, C^T \{C(A^T M^{-1} A)^{-1} C^T\}^{-1} C(A^T M^{-1} A)^{-1}$$

which is identical to ΔQ in expression (3.120), and

$$Q_X = Q_{X^*} - Q_{\Delta X}$$

We may also compute the change in the residuals which is caused by the condition (3.94). Let V* denote the residual as computed from the first system only. The final residuals V are

$$V = -P^{-1}B^T M^{-1}(A\hat{X} + W)$$

$$= -P^{-1}B^T M^{-1}\{A(\hat{X}^* + \Delta X) + W\}$$

$$= V^* + \Delta V$$

with

$$(\Delta V) = -P^{-1}B^T M^{-1} A\Delta X$$

Furthermore, using this expression for the change in residuals, we can verify the following relation:

$$\Delta X^T P_{X^*} \Delta X = (\Delta V)^T P \quad (\Delta V) = \Delta V^T P V \qquad (3.124)$$

Table 3.2: Expressions for the Case of Conditions Between Parameters

$F(X_a, L_a) = 0 \quad G(X_a) = 0$	$L_a = F(X_a) \quad G(X_a) = 0$
$BV + AX + W = 0$ $CX + W_C = 0$	$V = AX + L$ $CX + W_C = 0$
	$B = -I \quad L = W \quad r = n$
$M = BP^{-1}B^T$ $N = A^TM^{-1}A$ $U = A^TM^{-1}W$ $\begin{pmatrix} N & C^T \\ C & 0 \end{pmatrix}\begin{pmatrix} \hat{X} \\ -K_C \end{pmatrix} = \begin{pmatrix} -U \\ -W_C \end{pmatrix}$	$M = P^{-1}$ $N = A^TPA$ $U = A^TPL$ $\begin{pmatrix} N & C^T \\ C & 0 \end{pmatrix}\begin{pmatrix} \hat{X} \\ -K_C \end{pmatrix} = \begin{pmatrix} -U \\ -W_C \end{pmatrix}$
$\hat{X} = X^* + \delta X$ $X^* = -N^{-1}U$ $\delta X = -N^{-1}C^TS\,(W_C + CX^*)$ $S = (CN^{-1}C^T)^{-1}$	$\hat{X} = X^* + \delta X$ $X^* = -N^{-1}U$ $\delta X = -N^{-1}C^TS(W_C + CX^*)$ $S = (CN^{-1}C^T)^{-1}$
$\hat{V} = -P^{-1}B^TM^{-1}(W + A\hat{X})$	$\hat{V} = A\hat{X} + L$
$\hat{V}^TP\hat{V} = V^TPV^* + \Delta V^TPV$ $\hat{V}^TP\hat{V}^* = -U^TN^{-1}U + W^TM^{-1}W$ $\Delta V^TPV = (CX^* + W_C)\,S\,(CX^* + W_C)$	$\hat{V}^TP\hat{V} = V^TPV^* + \Delta V^TPV$ $\hat{V}^TP\hat{V}^* = -U^TN^{-1}U + L^TPL$ $\Delta V^TPV = (CX^* + W_C)^T\,S\,(CX^* + W_C)$
$\hat{\sigma}_o^2 = \frac{\hat{V}^TP\hat{V}}{r+s-u}$	$\hat{\sigma}_o^2 = \frac{\hat{V}^TP\hat{V}}{n+s-u}$
$\hat{\Sigma}_X = \hat{\sigma}_o^2\,(Q_{X^*} - \Delta Q_X)$ $= \hat{\sigma}_o^2\,(N^{-1} - N^{-1}C^TSCN^{-1})$	$\hat{\Sigma}_X = \hat{\sigma}_o^2\,(Q_{X^*} - \Delta Q_X)$ $= \hat{\sigma}_o^2\,(N^{-1} - N^{-1}C^TSCN^{-1})$
$\hat{\Sigma}_V = \hat{\sigma}_o^2\,(Q_{V^*} + \Delta Q_V)$ $= \hat{\sigma}_o^2\,P^{-1}B^TM^{-1}(M - AN^{-1}A^T + AN^{-1}C^TSCN^{-1}A^T)\,M^{-1}B\,P^{-1}$	$\hat{\Sigma}_V = \hat{\sigma}_o^2\,(Q_{V^*} + \Delta Q_V)$ $= \hat{\sigma}_o^2\,(P^{-1} - AN^{-1}A^T + AN^{-1}C^TSCN^{-1}A^T)$

3.5 Sequential Solutions

Assume the observations are done in two groups with the second group consisting of one or several observations.

For a more general exposition the mixed model is considered:

$$F_1(L_a^1, X_a) = 0 \tag{3.125}$$

$$F_2(L_a^2, X_a) = 0 \tag{3.126}$$

Let the observations between both groups be uncorrelated

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} = \sigma_o^2 \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & \Sigma_2^{-1} \end{pmatrix} \tag{3.127}$$

Both groups of weights refer to the same σ_o^2.

The linear adjustment model is

$$B_1 V_1 + A_1 X + W_1 = 0 \tag{3.128a}$$

$$B_2 V_2 + A_2 X + W_2 = 0 \tag{3.128b}$$

with

$$B_1 = \frac{\partial F_1}{\partial L^1}\Big|_{L_b^1, X_o} \qquad A_1 = \frac{\partial F_1}{\partial X}\Big|_{L_b^1, X_o}$$

$$B_2 = \frac{\partial F_2}{\partial L^2}\Big|_{L_b^2, X_o} \qquad A_2 = \frac{\partial F_2}{\partial X}\Big|_{L_b^2, X_o}$$

$$W_1 = F_1(L_b^1, X_o)$$

$$W_2 = F_2(L_b^2, X_o)$$

The function to be minimized is

$$\phi = V_1^T P_1 V_1 + V_2^T P_2 V_2 - 2K_1^T(B_1 V_1 + A_1 X + W_1) - 2K_2^T(B_2 V_2 + A_2 X + W_2)$$

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial V_1}\right)^T = P_1 \hat{V}_1 - B_1^T K_1 = 0 \tag{3.129}$$

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial V_2}\right)^T = P_2 \hat{V}_2 - B_2^T K_2 = 0 \tag{3.130}$$

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial X}\right)^T = -A_1^T K_1 - A_2^T K_2 = 0 \tag{3.131}$$

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial K_1}\right)^T = B_1 \hat{V}_1 + A_1 \hat{X} + W_1 = 0 \tag{3.132}$$

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial K_2}\right)^T = B_2 \hat{V}_2 + A_2 \hat{X} + W_2 = 0 \tag{3.133}$$

From equations (3.129) and (3.130) the residuals of both groups are

$$\hat{V}_1 = P_1^{-1} B_1^T K_1 \tag{3.134}$$

$$\hat{V}_2 = P_2^{-1} B_2^T K_2 \tag{3.135}$$

Combining (3.134) and (3.132)

$$B_1 P_1^{-1} B_1^T K_1 + A_1 \hat{X} + W_1 = 0$$

with

$$M_1 \equiv B_1 P_1^{-1} B_1^T \tag{3.136}$$

The Lagrange multiplier becomes

$$K_1 = -M_1^{-1} A_1 \hat{X} - M_1^{-1} W_1 \tag{3.137}$$

The equations (3.131) and (3.133) become, after combination with equations (3.135) and (3.137)

$$A^T M_1^{-1} A_1 \hat{X} + A_1^T M_1^{-1} W_1 - A_2^T K_2 = 0$$

$$B_2 P_2^{-1} B_2^T K_2 + A_2 \hat{X} + W_2 = 0$$

Using

$$M_2 \equiv B_2 P_2^{-1} B_2^T \tag{3.138}$$

both equations can be written in matrix form

$$\begin{pmatrix} A_1^T M_1^{-1} A_1 & A_2^T \\ A_2 & -M_2 \end{pmatrix} \begin{pmatrix} \hat{X} \\ -K_2 \end{pmatrix} = \begin{pmatrix} -A_1^T M_1^{-1} W_1 \\ -W_2 \end{pmatrix} \tag{3.139}$$

This equation shows how the normal matrix of the first system must be augmented in order to find the solution of both groups. It is, of course, possible to invert the whole matrix in (3.139). But this matrix might become quite big. Also, assume we add only one observation to the first system. Then we would have to invert the whole system once again. It is, therefore, more convenient to compute the contribution of the 2nd set of observations to the solution of the 1st set.

Applying matrix partitioning techniques as expressed in equation (1.33) to equation (3.139) gives

$$\hat{X} = -Q_{11} A_1^T M_1^{-1} W_1 - Q_{12} W_2 \tag{3.140}$$

$$K_2 = Q_{21} A_1^T M_1^{-1} W_1 + Q_{22} W_2 \tag{3.141}$$

with

$$Q_{11} = (A_1^T M_1^{-1} A_1)^{-1} - (A_1^T M_1^{-1} A_1)^{-1} A_2^T \{ M_2 + A_2 (A_1^T M_1^{-1} A_1)^{-1} A_2^T \} A_2 (A_1^T M_1^{-1} A_1)^{-1} \tag{3.142}$$

$$Q_{12} = +(A_1^T M_1^{-1} A_1)^{-1} A_2^T \{ M_2 + A_2 (A_1^T M_1^{-1} A_1)^{-1} A_2^T \}^{-1} \tag{3.143}$$

$$Q_{21} = Q_{12}^T \tag{3.144}$$

$$Q_{22} = -\{ M_2 + A_2 (A_1^T M_1^{-1} A_1)^{-1} A_2^T \}^{-1} \tag{3.145}$$

Substituting the Q_{11} and Q_{12} into equation (3.140) gives the solution for the parameters in sequential form:

$$\hat{X} = X^* + \Delta X \tag{3.146}$$

with

$$X^* = -(A_1^T M_1^{-1} A_1)^{-1} A_1^T M_1^{-1} W_1 \tag{3.147}$$

$$\Delta X = -(A_1^T M_1^{-1} A_1)^{-1} A_2^T \{M_2 + A_2 (A_1^T M_1^{-1} A_1)^{-1} A_2^T\}^{-1} (A_2 X^* + W_2) \tag{3.148}$$

where X^* is the solution of the first group only. Similarly,

$$K_2 = -\{M_2 + A_2 (A_1^T M_1^{-1} A_1)^{-1} A_2^T\}^{-1} (A_2 X^* + W_2) \tag{3.149}$$

Using K_2 the parameter correction ΔX can be written as

$$\Delta X = (A_1^T M_1^{-1} A_1)^{-1} A_2^T K_2 \tag{3.150}$$

There is a third way of giving the solution for X. Write the equations (3.128a,b) as

$$\begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} + \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} X + \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = 0 \tag{3.151}$$

which is the partitioned form of

$$B V + A X + W = 0$$

The least squares solution of the latter system is known to be

$$\hat{X} = -(A^T M^{-1} A)^{-1} A^T M^{-1} W$$

In terms of submatrices we have:

$$M = B P^{-1} B^T = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix} \begin{pmatrix} P_1^{-1} & 0 \\ 0 & P_2^{-1} \end{pmatrix} \begin{pmatrix} B_1^T & 0 \\ 0 & B_2^T \end{pmatrix}$$

$$= \begin{pmatrix} B_1 P_1^{-1} B_1^T & 0 \\ 0 & B_2 P_2^{-1} B_2^T \end{pmatrix} = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}$$

$$A^T M^{-1} A = (A_1^T A_2^T)\begin{pmatrix} M_1^{-1} & 0 \\ 0 & M_2^{-1} \end{pmatrix}\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$$

$$= A_1^T M_1^{-1} A_1 + A_2^T M_2^{-1} A_2$$

$$A^T M^{-1} W = (A_1^T A_2^T)\begin{pmatrix} M_1^{-1} & 0 \\ 0 & M_2^{-1} \end{pmatrix}\begin{pmatrix} W_1 \\ W_2 \end{pmatrix}$$

$$= A_1^T M_1^{-1} W_1 + A_2^T M_2^{-1} W_2$$

The solution for the parameters becomes

$$\hat{X} = (A_1^T M_1^{-1} A_1 + A_2^T M_2^{-1} A_2)^{-1} \quad (A_1^T M_1^{-1} W_1 + A_2^T M_2^{-1} W_2) \tag{3.152}$$

Equation (3.152) shows another method of how the solution of the first group can be altered so as to give the solution of the complete system. The method is sometimes called the method of adding the normal equations. Note that this form of the solution could have been obtained immediately from equation (3.139) upon applying the alternative form of expression (1.33). The cofactor matrix is

$$Q_X = (A^T M^{-1} A)^{-1} = (A_1^T M_1^{-1} A_1 + A_2^T M_2^{-1} A_2)^{-1} \tag{3.153}$$

Using again the alternative expressions of (1.33) it is seen that Q_X of (3.153) and Q_{11} of (3.142) are identical. We can, therefore, write the cofactor matrix Q_X of (3.153) in the following form:

$$Q_X = Q_{X*} + \delta Q_X$$

$$= (A_1^T M_1 A_1 + A_2 M_2^{-1} A_2)^{-1}$$

$$= (A_1^T M_1^{-1} A_1)^{-1} - (A_1^T M_1^{-1} A_1)^{-1} A_2^T \{M_2 + A_2 (A_1^T M_1^{-1} A_1)^{-1} A_2^T\}^{-1} A_2 (A_1^T M_1^{-1} A_1)^{-1}$$

$$= Q_{X*} - Q_{X*} A_2^T (M_2 + A_2 Q_{X*} A_2^T)^{-1} A_2 Q_{X*} \tag{3.154}$$

where Q_{X*} is the cofactor matrix of the first group of observations only,

$$\boxed{\delta Q_x = -Q_{x*}\, A_2^T (M_2 + A_2 Q_{x*} A_2^T)^{-1}\, A_2 Q_{x*}} \qquad (3.155)$$

is the contribution of the additional observations (2nd group) to the cofactor matrix of the first group.

It is interesting to see that δQ_x can be computed without having the actual observations of the second group. This can be of great help in practical application. Assume that a complex network has been observed. In the subsequent adjustment it is found that certain parameters are not determined with sufficient accuracy. It is then possible to find through simulation study those additional observations which increase the accuracy of the parameters in question most (optimization techniques).

Equation (3.155) shows that each additional observation reduces the diagonal elements of the cofactor matrix. The diagonal terms of $(-\delta Q_x)$ are positive definite quadratic forms.

The computation of $V^T PV$ proceeds as usual

$$\hat{V}^T P \hat{V} = \hat{V}_1^T P_1 \hat{V}_1 + \hat{V}_2^T P_2 \hat{V}_2$$

Substituting equations (3.134) and (3.135) gives

$$\hat{V}^T P \hat{V} = K_1^T B_1 P_1^{-1} P_1 P_1^{-1} B_1^T K_1 + K_2^T B_2 P_2^{-1} P_2 P_2^{-1} B_2^T K_2$$

$$= K_1^T B_1 P_1^{-1} B_1^T K_1 + K_2^T B_2 P_2^{-1} B_2^T K_2$$

In order to develop this expression further we substitute the equations (3.134) and (3.135) in (3.132) and (3.133). The resulting equations are

$$\begin{array}{lllll} K_1^T \mid & B_1 P_1^{-1} B_1^T K_1 + & A_1 \hat{X} + & W_1 = 0 & \\ K_2^T \mid & B_2 P_2^{-1} B_2^T K_2 + & A_2 \hat{X} + & W_2 = 0 & \\ & -A_1^T K_1 - A_2^T K_2 = 0 & & & \mid \hat{X} \end{array}$$

to which we have added equation (3.131). Carrying out the multiplications as indicated and adding all three equations gives

$$\hat{V}^T P \hat{V} = -K_1^T W_1 - K_2^T W_2 \tag{3.156}$$

Substituting the expression (3.137) for K_1,

$$K_1 = -M_1^{-1} A_1 (X^* + \Delta X) - M_1^{-1} W_1$$

gives

$$\begin{aligned} \hat{V}^T P \hat{V} &= (X^*)^T A_1^T M_1^{-1} W_1 + \Delta X^T A_1^T M_1^{-1} W_1 + W_1^T M_1^{-1} W_1 - K_2^T W_2 \\ &= V^T P V^* + \Delta X^T A_1^T M_1^{-1} W_1 - K_2^T W_2 \end{aligned} \tag{3.157}$$

The final form for $V^T PV$ is reached by substituting (3.148) and (3.149) for ΔX and K_2:

$$\boxed{\begin{aligned} V^T P V &= V^T P V^* + \Delta V^T P V \\ &= V^T P V^* + (A_2 X^* + W_2)^T \{M_2 + A_2 (A_1^T M_1^{-1} A_1)^{-1} A_2^T\}^{-1} (A_2 X^* + W_2) \end{aligned}} \tag{3.158}$$

with $\Delta V^T P V \geq 0$ because it is a positive definite quadratic form.

<u>Generalization of sequential solutions</u>: Consider the case of K sets of observations:

$$\begin{aligned} F_1 (L_a^1, X_a) &= 0 \\ F_2 (L_a^2, X_a) &= 0 \\ F_3 (L_a^3, X_a) &= 0 \\ &\vdots \\ F_K (L_a^K, X_a) &= 0 \end{aligned} \tag{3.159}$$

Let the stochastic model be

$$P = \begin{pmatrix} P_1 & & & & 0 \\ & P_2 & & & \\ & & P_3 & & \\ & & & \ddots & \\ 0 & & & & P_K \end{pmatrix} \tag{3.160}$$

i.e., there is no correlation between the various groups of observations. All the weights refer to the same a-priori variance of unit weight σ_o^2. The point of expansion for each subgroup is (L_b^i, X_o). The least squares solution of the first system is

$$X_1 = -Q_1 A_1^T M_1^{-1} W_1$$

with

$$Q_1 = Q_{X_1} = (A_1^T M_1^{-1} A_1)^{-1} \equiv N_1^{-1} \tag{3.161}$$

and

$$\hat{V}_1^T P \hat{V}_1 = W_1^T M_1^{-1} W_1 + W_1^T M_1^{-1} A_1 \hat{X}_1$$

Equation (3.154) shows how the cofactor matrix changes as another group of observations is added:

$$\begin{aligned} Q_2 &= Q_1 - Q_1 A_2^T (M_2 + A_2 Q_1 A_2^T)^{-1} A_2 Q_1 \\ Q_3 &= Q_2 - Q_2 A_3^T (M_3 + A_3 Q_2 A_3^T)^{-1} A_3 Q_2 \\ &\vdots \\ Q_i &= Q_{i-1} - Q_{i-1} A_i^T (M_i + A_i Q_{i-1} A_i^T)^{-1} A_i Q_{i-1} \end{aligned} \tag{3.162}$$

The change in parameters as a result of a new group of observations is, according to equation (3.146),

$$\Delta X_i = -Q_{i-1} A_i^T (M_i + A_i Q_{i-1} A_i^T)^{-1} (A_i X_{i-1} + W_i) \qquad (3.163)$$

For the quadratic form we obtain from equation (3.158)

$$V^T P V_i = V^T P V_{i-1} + \Delta V^T P V_i$$

$$\Delta V^T P V_i = (A_i \hat{X}_{i-1} + W_i)^T \{M_i + A_i Q_{i-1} A_i^T\}^{-1} (A_i \hat{X}_{i-1} + W_i) \qquad (3.164)$$

This is the general formulation of the sequential algorithm. It expresses the changes a new group of observations has on the result of the previous adjustment containing all previous observations. There is no restriction on the number of equations in the new group.

Occasionally, it is desirable to remove a set of observations from an existing adjustment. Consider again the uncorrelated case, i.e., the set of observations to be removed is uncorrelated to the other observations. The solution is readily seen from equation (3.152) which shows how normal equations are added when a new set of observations is incorporated. In case observations are removed the corresponding part of the normal matrix and the right-hand-side term have to be subtracted. In this case equation (3.152) becomes

$$\hat{X} = -(A_1^T M_1^{-1} A_1 - A_2 M_2^{-1} A_2)^{-1} (A_1^T M_1^{-1} W_1 - A_2^T M_2^{-1} W_2)$$

$$= -\{A_1^T M_1^{-1} A_1 + A_2^T (-M_2^{-1}) A_2\}^{-1} \{A_1^T M_1^{-1} W_1 + A_2^T (-M_2^{-1}) W_2\} \qquad (3.165)$$

It is seen from (3.165) that one only has to assign a negative sign to the weight matrix of the group of observations which is being removed, since

$$-M_2 = B_2 (-P_2^{-1}) B_2^T$$

The same sign change should be made in the expressions (3.163) and (3.164) for the sequential solution. It is noted that in this case the matrix

$$\{-M_i + A_i Q_{i-1} A_i^T\}$$

is still symmetric but not positive definite anymore. The inversion cannot be done by Cholesky's algorithm. A more general inversion procedure must be used.

Finally, we would like to derive the contribution to the residuals due to the additional group. From equations (3.134), (3.137), and (3.146) we obtain

$$V_1 = -P_1^{-1} B_1^T M_1^{-1} \{A_1(X^* + \Delta X) + W_1\}$$

$$\frac{\partial V_1}{\partial W_1} = -P_1^{-1} B_1^T M_1^{-1} (I - A_1 N_1^{-1} A_1^T M_1^{-1} + A_1 N_1^{-1} A_2^T T A_2 N_1^{-1} A_1^T M_1^{-1})$$

$$\frac{\partial V_1}{\partial W_2} = P_1^{-1} B_1^T M_1^{-1} A_1 N_1^{-1} A_2^T T$$

Since
$$Q_{W_1,W_2} = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix} \tag{3.166}$$

we obtain from the law of variance propagation, after some algebraic rearrangements,

$$Q_{V_1} = Q_{V_1}{}^* + \Delta Q_{V_1}$$

with

$$Q_{V_1}{}^* = (P_1^{-1} B_1^T) M_1^{-1} (P_1^{-1} B_1^T)^T - (P_1^{-1} B_1^T M_1^{-1} A_1) N_1^{-1} (P_1^{-1} B_1^T M_1^{-1} A_1)^T$$

$$\boxed{\Delta Q_{V_1} = (P_1^{-1} B_1^T M_1^{-1} A_1 N_1^{-1} A_2^T) T (P_1^{-1} B_1^T M_1^{-1} A_1 N_1^{-1} A_2^T)^T}$$

The matrix T is

$$T \equiv \{M_2 + A_2 N_1^{-1} A_2^T\}^{-1} \tag{3.168}$$

Note that $\hat{\Sigma}_{V_1} = \hat{\sigma}_o^2 Q_{V_1}$

where $\hat{\sigma}_o^2$ is computed from all groups included in the adjustment. The expression for V_2 follows from equations (3.135) and (3.149):

$$V_2 = -P_2^{-1} B_2^T T (A_2 X^* + W_2)$$

$$\frac{\partial V_2}{\partial W_1} = P_2^{-1} B_2^T T A_2 N_1^{-1} A_1^T M_1^{-1}$$

$$\frac{\partial V_2}{\partial W_2} = -P_2^{-1} B_2^T T$$

Using again the law of variance propagation and the relation (3.166) the cofactor for V_2 becomes

$$\boxed{Q_{V_2} = P_2^{-1} B_2^T T B_2 P_2^{-1}} \qquad (3.169)$$

The estimated variance-covariance matrix is

$$\hat{\Sigma}_{V_2} = \hat{\sigma}_o^2 Q_{V_2}$$

The variance-covariance matrix of the adjusted observations is, as usual,

$$\Sigma_{L_a}^{i} = \Sigma_{L_b}^{i} - \Sigma_{V_i} \qquad (3.170)$$

As for iterations one has to make sure that all groups are evaluated for the same approximate parameters. If the first system is iterated, the approximate coordinates for the last iteration have to be used as expansion points for the other groups. Since there are no common observations between groups, the iteration with respect to the observations can be done individually for each group. When a group is removed, the same expansion point must be used as was used in the first place.

The sequential algorithm described in this section includes the special case of direct observations on the parameters. If all parameters are observed we have $P_2=P_x$, $B_2= -I$, and $A_2=I$. If the observed value of the parameter is also the point of expansion then we have, in addition, $W_2 = 0$, or $L_2=0$. If only a subset of parameters is observed directly then B_2 is still an identity matrix but A_2 is a rectangular matrix with "ones" at the appropriate positions. The expansion point for the parameters does not have to coincide with the observed value of the parameter. This situation arises if the first system has been iterated. If it is decided that the first system is not going to be iterated at all, one probably will select the observed coordinates as the coordinates of the expansion point.

Table 3.3 Expressions for the Sequential Solution

$F_1(L_a^1, X_a) = 0$ $F_2(L_a^2, X_a) = 0$ $P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$	$L_a^1 = F_1(X_a)$ $L_a^2 = F_2(X_a)$ $P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$
$B_1 V_1 + A_1 X + W_1 = 0$ $B_2 V_2 + A_2 X + W_2 = 0$	$V_1 = A_1 X + L_1$ $V_2 = A_2 X + L_2$
	$B_1 = -I,\ W_1 = L_1,\ r_1 = n_1,\ B_2 = I,\ W_2 = L_2,\ r_2 = n_2$
$M_i = B_i P_i^{-1} B_i^T$ $N_i = A_i^T M_i^{-1} A_i$ $U_i = A_i^T M_i^{-1} W_i$ $\begin{pmatrix} N_1 & A_2^T \\ A_2 & -M_2 \end{pmatrix} \begin{pmatrix} \hat{X} \\ -K_2 \end{pmatrix} = \begin{pmatrix} -U_1 \\ -W_2 \end{pmatrix}$	$M_i = P_i^{-1}$ $N_i = A_i^T P_i A_i$ $U_i = A_i^T P_i L_i$ $\begin{pmatrix} N_1 & A_2^T \\ A_2 & -P_2^{-1} \end{pmatrix} \begin{pmatrix} \hat{X} \\ -K_2 \end{pmatrix} = \begin{pmatrix} -U_1 \\ -L_2 \end{pmatrix}$
$\hat{X} = -(N_1+N_2)^{-1}(U_1+U_2)$	$\hat{X} = -(N_1+N_2)^{-1}(U_1+U_2)$
$\hat{X} = X^* + \Delta X$ $X^* = -N_1^{-1} U_1$ $T = (M_2 + A_2 N_1^{-1} A_2^T)^{-1}$ $\Delta X = -N_1^{-1} A_2^T (A_2 X^* + W_2)$	$\hat{X} = X^* + \Delta X$ $X^* = -N_1^{-1} U_1$ $T = (M_2 + A_2 N_1^{-1} A_2^T)^{-1}$ $\Delta X = -N_1^{-1} A_2^T T (A_2 X^* + L_2)$
$V_1 = V_1^* + \Delta V_1$ $V_1^* = -P_1^{-1} B_1^T M_1^{-1} (A_1 X^* + W_1)$ $\Delta V_1 = -P_1^{-1} B_1^T M_1^{-1} A_1 \Delta X$	$V_1 = V_1^* + \Delta V_1$ $V_1^* = A_1 X^* + L_1$ $\Delta V_1 = A_1 \Delta X$
$\hat{V}^T P \hat{V} = \hat{V}^T P \hat{V}^* + \Delta V^T P V$ $\hat{V}^T P \hat{V}^* = W_1^T M_1^{-1} W_1 - U_1^T M_1^{-1} U_1$ $\Delta V^T P V = (A_2 X^* + W_2)^T T (A_2 X^* + W_2)$	$\hat{V}^T P \hat{V} = \hat{V} P \hat{V}^* + \Delta V^T P V$ $\hat{V} P \hat{V}^* = L_1^T P_1 L_1 - U_1^T N^{-1} U_1$ $\Delta V^T P V = (A_2 X^* + L_2)^T T (A_2 X^* + L_2)$
$Q_X = Q_{X^*} - \Delta Q$ $Q_{X^*} = N_1^{-1}$ $\Delta Q = N_1^{-1} A_2^T T A_2 N_1^{-1}$	$Q_X = Q_{X^*} - \Delta Q$ $Q_{X^*} = N_1^{-1}$ $\Delta Q = N_1^{-1} A_2^T T A_2 N_1^{-1}$

Table 3.3 Continued

$\hat{\sigma}_o^2 = \frac{\hat{v}^T P \hat{v}}{r_1 + r_2 - u}$	$\hat{\sigma}_o^2 = \frac{\hat{v}^T P \hat{v}}{n_1 + n_2 - u}$
$Q_{v_1} = Q_{v_1}{}^* + \Delta Q_{v_1}$ $Q_{v_1}{}^* = (P_1^{-1} B_1^T) M_1^{-1} (P_1^{-1} B_1^T)^T$ $-(P_1^{-1} B_1^T M_1^{-1} A_1) N_1^{-1} (P_1^{-1} B_1^T M_1^{-1} A_1)^T$ $\Delta Q_{v_1} = (P_1^{-1} B_1^T M_1^{-1} A_1 N_1^{-1} A_2^T) T \times$ $(P_1^{-1} B_1^T M_1^{-1} A_1 N_1^{-1} A_2^T)^T$ $Q_{v_2} = P_2^{-1} B_2^T T B_2 P_2^{-1}$	$Q_{v_1} = Q_{v_1}{}^* + \Delta Q_{v_1}$ $Q_{v_1}{}^* = P_1^{-1} - A_1 N_1^{-1} A_1^T$ $\Delta Q_{v_1} = A_1 N_1^{-1} A_2 T A_2^T N_1^{-1} A_1^T$ $Q_{v_2} = P_2^{-1} T P_2^{-1}$

3.6 Weighted Parameters

This section is an extension of the remarks made at the end of the previous section regarding direct observations on the parameters. Sometimes it is possible to obtain information directly on the parameters through some additional measurements which differ from the type of measurements used thus far. Since these new measurements are known only to a certain accuracy they must be introduced into the adjustment in such a way that their contributions to the final result are in accordance with their accuracy expressed by their covariance matrix. We are, thus, led to the method of weighted parameters, or to the case with special weighted constraints. We will have to handle the parameters as "observations" having a variance-covariance matrix associated with them. The following extreme cases can arise regarding the weight of an observation:

weight = ∞	the observation becomes a constant (fixed variable); the associated residual is zero.
weight = 0	unknown in the classical meaning.

It is understood that the weights of both groups of observations refer to the same variance of unit weight.

Let

L_{L_a}	be the adjusted observations of the first group.
L_{X_a}	be the adjusted observations of the second group. (One may think of them as being the group of "observed parameters".)

Generally, both groups of observations differ significantly in their respective weights:

$$P_L = \sigma_o^2 \Sigma_L^{-1} \tag{3.171}$$

$$P_X = \sigma_o^2 \Sigma_X^{-1} \tag{3.172}$$

We assume that there is no correlation between both groups of observations.

The matrix P_X may be full, diagonal, diagonal with some zeros on the diagonal, or even completely zero.

The mathematical model for the general case discussed above is the method of condition equations

$$F(L_{L_a}, L_{X_a}) = 0 \tag{3.173}$$

The stochastic model is expressed by equations (3.171) and (3.172). Let there be

n observations in L_L
u observations in L_X
r equations in (3.173).

The linearized form of (3.173) is

$$B_L V_L + B_X V_X + W = 0 \tag{3.174}$$

with

$$B_L = \frac{\partial F}{\partial L_L} \Big/ L_{L_b}, L_{X_b}$$

$$B_X = \frac{\partial F}{\partial L_X} \Big/ L_{L_b}, L_{X_b}$$

$$W = F(L_{L_b}, L_{X_b})$$

where L_{L_b} and L_{X_b} are the observations. The least squares principle requires minimization of

$$\phi(V_L, V_X, K) = V_L^T P_L V_L + V_X^T P_X V_X - 2K^T(B_L V_L + B_X V_X + W_L)$$

The partial derivatives are

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial V_L}\right)^T = P_L \hat{V}_L - B_L^T K = 0 \tag{3.175}$$

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial V_X}\right)^T = P_X \hat{V}_X - B_X^T K = 0 \tag{3.176}$$

$$\frac{1}{2}\left(\frac{\partial \phi}{\partial K}\right)^T = B_L V_L + B_X V_X + W = 0 \qquad (3.177)$$

which can be written in matrix form as follows:

$$\begin{pmatrix} -P_L & 0 & B_L^T \\ 0 & -P_X & B_X^T \\ B_L & B_X & 0 \end{pmatrix} \begin{pmatrix} \hat{V}_L \\ \hat{V}_X \\ K \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -W \end{pmatrix} \qquad (3.178)$$

For a least squares adjustment to exist it must be

$$n + u > r$$

The residuals of the first group of observations follows from equation (3.175)

$$\hat{V}_L = P_L^{-1} B_L^T K \qquad (3.179)$$

Substituting this expression into the remaining two equations gives

$$\begin{pmatrix} -P_X & B_X^T \\ B_X & M_L \end{pmatrix} \begin{pmatrix} \hat{V}_X \\ K \end{pmatrix} = \begin{pmatrix} 0 \\ -W \end{pmatrix} \qquad (3.180)$$

where

$$M_L \equiv B_L P_L^{-1} B_L^T \qquad (3.181)$$

The solution is easily found by computing the inverse using matrix partitioning techniques (1.33):

$$V_X = Q_{12} W$$

$$K = -Q_{22} W$$

with

$$Q_{12} = (P_X+B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1} = P_X^{-1} B_X^T (M_L+B_X P_X^{-1} B_X^T)^{-1} \quad (3.182)$$

$$Q_{22} = M_L^{-1} - M_L^{-1} B_X (P_X+B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1} \quad (3.183)$$

$$= (M_L+B_X P_X^{-1} B_X^T)^{-1}$$

Thus

$$\boxed{\hat{V}_X = -(P_X+B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1} W} \quad (3.184)$$

and

$$K = \{-M_L^{-1} + M_L^{-1} B_X (P_X+B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1}\} W \quad (3.185)$$

The expression (3.184) shows how the weight of the second set of observations enters into the corrections (residuals) for the second group of observations (parameters). If $P_X = 0$ there is no contribution at all, and the result is identical to the one of the mixed model of Section 3.3 in which V_X is considered to be the vector of parameters.

The computation of $V^T PV$ proceeds as usual. Using equation (3.179) the quadratic form becomes

$$\hat{V}^T P \hat{V} = \hat{V}_L^T P_L \hat{V}_L + \hat{V}_X^T P_X \hat{V}_X$$

$$= K^T M_L K + \hat{V}_X^T P_X \hat{V}_X$$

Note the contribution of V_X to the quadratic form $V^T PV$. Substituting the expression for the Lagrange multiplier

$$K = -M_L^{-1} W - M_L^{-1} B_X \hat{V}_X$$

gives

$$\hat{V}^T P \hat{V} = -K^T B_X \hat{V}_X - K^T W + \hat{V}_X^T P_X \hat{V}_X$$

Making use of equation (3.176) the quadratic form finally becomes

$$\hat{V}^T P \hat{V} = -K^T W$$

$$= W^T M_L^{-1} W - W^T M_L^{-1} B_X (P_X+B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1} W \quad (3.186)$$

An estimate of the a-posteriori variance of unit weight is

$$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{r'} \tag{3.187}$$

where r' equals the number of equations minus the number of observations with zero weight (parameters).

The variance-covariance matrix of the residuals of the second group can be found from the law of variance propagation. Since both groups are uncorrelated the cofactor for W is

$$\begin{aligned} Q_W &= B_L P_L^{-1} B_L^T + B_X P_X^{-1} B_X^T \\ &= M_L + M_X \end{aligned} \tag{3.188}$$

From equations (3.182), (3.184) and (3.188) it follows that

$$\begin{aligned} Q_{V_X} &= Q_{12} Q_W Q_{12}^T \\ &= (P_X + B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1} (M_L + M_X) M_L^{-1} B_X (P_X + B_X^T M_L^{-1} B_X)^{-1} \\ &= (P_X + B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1} \underbrace{(M_L + M_X)(M_L + B_X P_X^{-1} B_X^T)^{-1}}_{I} B_X P_X^{-1} \\ &= (P_X + B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1} B_X P_X^{-1} \\ &= P_X^{-1} B_X^T (M_L + B_X P_X^{-1} B_X)^{-1} B_X P_X^{-1} \end{aligned}$$

Above use has been made twice of the two identical expressions for Q_{12} as given in (3.182). The expressions for Q_{V_X} can further be simplified by using the relation

$$(P_X + B_X^T M_L^{-1} B_X)^{-1} = P_X^{-1} - P_X^{-1} B_X^T (M_L + B_X P_X^{-1} B_X^T)^{-1} B_X P_X^{-1}$$

which can be verified by the identities in (1.33). The final expression for the cofactor matrix is

$$Q_{V_X} = P_X^{-1} - (P_X + B_X^T M_L^{-1} B_X)^{-1} \tag{3.189}$$

The second term indicates the improvement due to the adjustment.

The variance-covariance matrix of the adjusted observations is, as usual,

$$\Sigma_{L_{X_a}} = \sigma_o^2 (Q_{L_{X_b}} - Q_{L_{V_X}})$$

$$= \sigma_o^2 \{P_X^{-1} - P_X^{-1} + (P_X + B_X^T M_L^{-1} B_X)^{-1}\}$$

$$\Sigma_{L_{X_a}} = \sigma_o^2 (P_X + B_X^T M_L^{-1} B_X)^{-1} \tag{3.190}$$

and the estimate is

$$\boxed{\hat{\Sigma}_{L_{X_a}} = \hat{\sigma}_o^2 (P_X + B_X^T M_L^{-1} B_X)^{-1}} \tag{3.191}$$

The cofactor matrix for the observations of the first group can be derived according to the same scheme, applying the law of variance propagation and using repeatedly the matrix identities (1.33). After considerable algebraic rearrangements the result is

$$Q_{V_L} = P_L^{-1} B_L^T M_L^{-1} \{M_L - B_X (P_X + B_X^T M_L^{-1} B_X)^{-1} B_X^T\} M_L^{-1} B_L P_L^{-1} \tag{3.192}$$

This adjustment model has to be iterated along the same lines as explained for the model of condition equations in Section 3.2:

$$L_L^1 = L_{L_b}, \quad L_X^1 = L_{X_b}$$

$$V_L^i = L_L^{i+1} - L_{L_b}$$

$$V_X^i = L_X^{i+1} - L_{X_b}$$

$$B_L^i V_L^i + B_X^i V_X^i + \overline{W}^i = 0$$

with

$$\overline{W}^i = W^i + B_L^i (L_{L_b} - L_L^i) + B_X^i (L_{X_b} - L_X^i)$$

Note that the discrepancy vector $\bar{W}$ consists of three terms in this case. It follows that

$$V_L^i \rightarrow V_L$$

$$V_X^i \rightarrow V_X$$

as i increases. This iteration procedure is valid regardless of whether L_X or only a subset (i.e., the adjustment contains some parameters in the classical sense) is observed. If we consider the approximate coordinate of the parameters as observations with zero weight then the above algorithm can be implemented without any change. The respective subset of V_X which denotes the parameters, will not tend toward zero as it does in Sections 3.1 and 3.3, yet the point of expansion moves toward the adjusted value. In this case the parameters are updated similarly to the observations. The corrections to the approximate values of the parameters, of course, must not contribute to V^TPV. But this is accomplished automatically since the weight matrix has zeros at the respective places. If it is desirable to update the parameters in the classical sense, i.e., the corrections tend toward zero, the above algorithm has to be modified only slightly.

In Table 4.3 a summary of equations for the case of weighted parameters is given. From the first and the third columns we see that the equations are readily applicable even for the case in which only a partial set of L_X is observed. The inverse of P_X is nowhere computed directly, thus avoiding any numerical difficulties on the computer.

The second column of Table 3.4 contains the expression of the mixed model in order to demonstrate the formal correspondence between the mixed model and the more general model with weighted parameters. We only have to specify that $P_X \equiv 0$ and $L_{X_b} = X_o$.

Another important correspondence exists with respect to the adjustment in groups as explained in the previous section. Emphasis was placed in Section 3.4 on the sequential algorithm. But considering equation (3.152), which shows how the normal equations of two groups are combined, the relationship with the model of weighted parameters becomes obvious. Specifying the second group by $B_2=-I$, $A_2=I$, and $W_2=L_2=L_o^2-L_b^2 = X_o-L_{X_b}$ and setting $X_o=L_{X_b}$,

Table 3.4 Expressions for the Case of Weighted Parameters

$F(L_a, L_{X_a}) = 0 \quad P = \begin{vmatrix} P_L & 0 \\ 0 & P_X \end{vmatrix}$	$F(L_a, X_a) = 0$	$L_a = F(L_{X_a}) \quad \begin{vmatrix} P & 0 \\ 0 & P_X \end{vmatrix}$
$B_L V_L + B_X V_X + W = 0$	$BV + AX + W = 0$	$V = AX + L \quad L = L_o - L_b$
	$B = B_L,\ A = B_X,\ P_X = 0$	$B_L = -I,\ A = B_X,\ L = W,\ P = P_L$
$M_L = B_L P_L^{-1} B_L^T$ $\hat{V}_X = -(P_X + B_X^T M_L^{-1} B_X)^{-1} B_X^T M_L^{-1} W$ $K = -M_L^{-1}(B_X \hat{V}_X + W)$ $\hat{V}_L = P_L^{-1} B_L^T K$	$M = BP^{-1}B^T$ $\hat{X} = -(A^T M^{-1} A)^{-1} A^T M^{-1} W$ $K = -M^{-1}(A\hat{X} + W)$ $\hat{V} = P^{-1} B^T K$	$M = P^{-1}$ $\hat{X} = -(P_X + A^T P A)^{-1} A^T P L$ $\hat{V} = A\hat{X} + L$
$Q_{\hat{L}_X} = (P_X + B_X^T M_L^{-1} B_X)^{-1}$ $Q_{\hat{V}_L} = P_L^{-1} B_L^T M_L^{-1} B_L P_L^{-1} - P_L^{-1} B_L^T$ $\cdot M_L^{-1} B_X (P_X + B_X^T M_L^{-1} B_X)^{-1}$ $\cdot B_X^T M_L^{-1} B_L P_L^{-1}$ $Q_{\hat{V}_{L_a}} = P_L^{-1} - Q_{\hat{V}_L}$	$Q_{\hat{X}} = (A^T M^{-1} A)^{-1}$ $Q_{\hat{V}} = P^{-1} B^T M^{-1} B P^{-1}$ $-P^{-1} B^T M^{-1} A (A^T M^{-1} A)^{-1} A^T M^{-1} B P^{-1}$ $Q_{\hat{L}_a} = P^{-1} - Q_{\hat{V}}$	$Q_{\hat{X}} = (P_X + A^T P A)^{-1}$ $Q_{\hat{V}} = P^{-1} - A(P_X + A^T P A)^{-1} A^T$ $Q_{\hat{L}_a} = P^{-1} - Q_{\hat{V}}$
$\hat{\sigma}_o^2 = \dfrac{\hat{V}_X^T P_X \hat{V}_X + \hat{V}_L^T P_L \hat{V}_L}{r}$	$\hat{\sigma}_o^2 = \dfrac{\hat{V}^T P \hat{V}}{r-u}$	$\hat{\sigma}_o^2 = \dfrac{\hat{V}^T P \hat{V}}{n_1 + n_X - u}$ n_1 = # of elements in L n_X = # of non-zero diagonal elements in P_X u = # of elements in X

the equation (3.152) is identical to the expression $\hat{V}_X$ and $\hat{X}$ in the columns 1 and 3 of Table 3.4. The necessity to equate $X_o = L_{X_b}$ in order to obtain the identical expressions stems from the different approaches in both cases. In the model with weighted parameters, the parameters are observations by assumption, and, consequently, there is no conceptual need to introduce an approximate value X_o. The identity of the least squares expression also holds if only a subset of parameters is observed. In this case A has as many rows as there are observed parameters. There is a "one" at the column which corresponds to the observed parameter. All other elements are zero.

Finally, it is pointed out that the expressions developed in this section permit a convenient way to constrain the parameters to their approximate values. One only has to add a large number (weight) in the corresponding diagonal element of the normal matrix. This is one way to impose minimum constraint so that the normal matrix is not singular anymore.

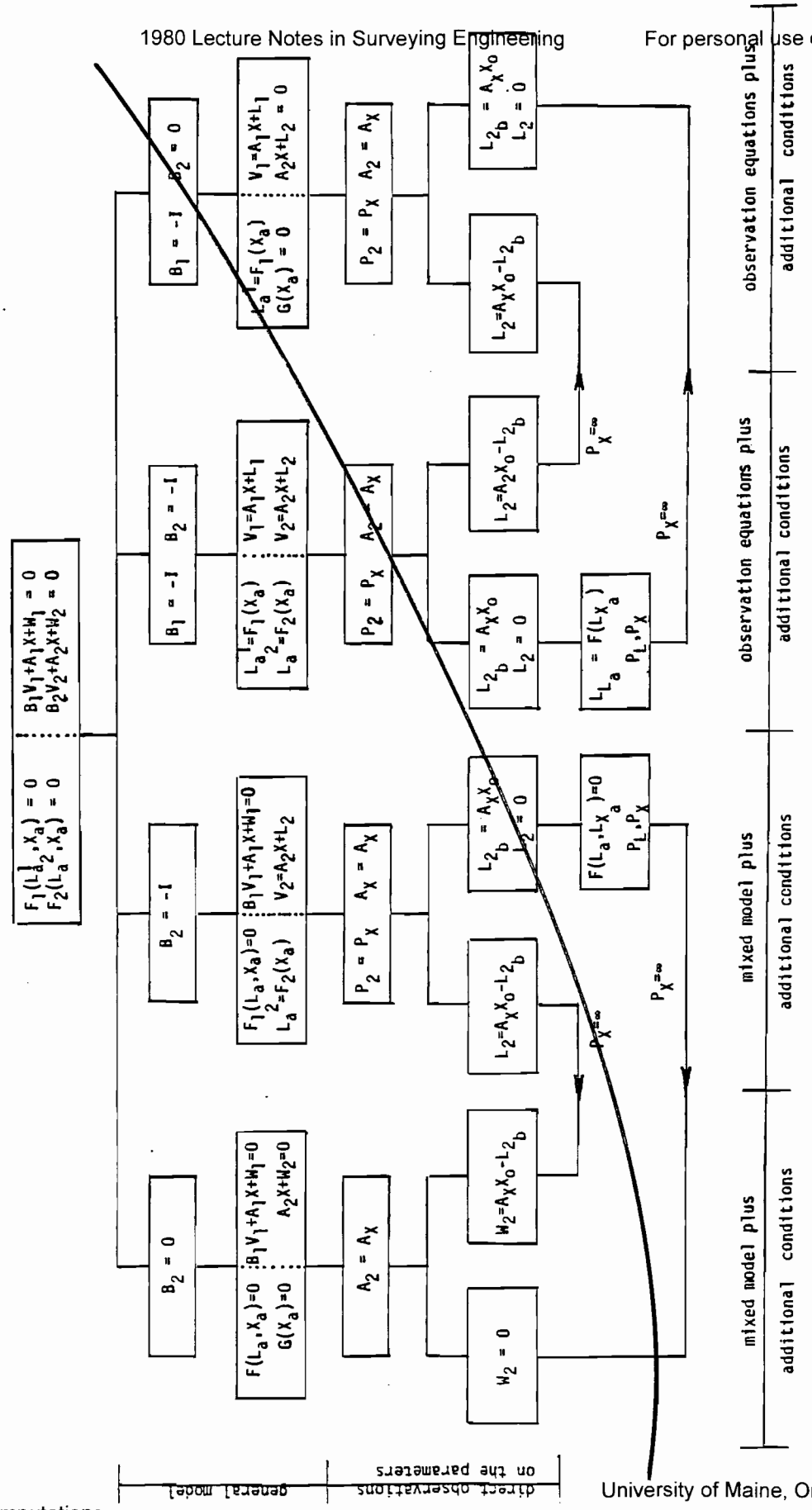
In Table 3.5 the expressions are given for direct observations on parameters in the case of the observation equation model. These equations follow immediately from those of Table 3.3 and include the special case of column 3 of Table 3.4. Note that the number of rows of A_X and the size of P_X is equal to the number of observed parameters. A_X has "ones" at the appropriate positions.

Table 3.5 Direct Observations on Parameters (Observation Equation Model)

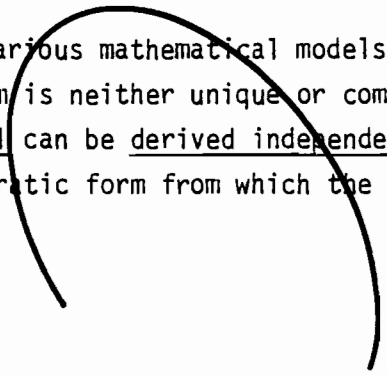
$L_a = F(X_a)$ $\quad \begin{pmatrix} P & 0 \\ 0 & P_X \end{pmatrix}$ $L_{X_a} = A_X X_a$	$L_a = F(X_a)$ $\quad \begin{pmatrix} P & 0 \\ 0 & P_X \end{pmatrix}$ $L_{X_a} = X_a$
$V = AX + L$ $V_X = A_X X + L_X$ $L_X = A_X X_o - L_{X_b}$	$V = AX + L$ $V_X = X + L$ $L_X = X_o - L_{X_b}$
$N = A^T P A$ $U = A^T P L$ $U_X = A_X^T P_X L_X$	$N = A^T P A$ $U = A^T P L$ $U_X = P_X L_X$
$X_a = X_o + \hat{X} = X_o + X^* + \Delta X$ $X^* = -N^{-1} U$ $T = (P_X^{-1} + A_X N^{-1} A_X^T)^{-1}$ $\Delta X = -N^{-1} A_X^T T (A_X X^* + L_X)$ $\hat{X} = -(N + A_X P_X A_X^T)^{-1} (U + A_X^T P_X L_X)$	$X_a = X_o + \hat{X} = X_o + X^* + \Delta X$ $X^* = -N^{-1} U$ $T = (P_X^{-1} + N^{-1})^{-1}$ $\Delta X = -N^{-1} T (X^* + L_X)$ $\hat{X} = -(N + P_X)^{-1} (U + P_X L_X)$
$\hat{V}_1 = V_1^* + \Delta V$ $\hat{V}_1^* = AX^* + L$ $\Delta V = A \Delta X$	$\hat{V}_1 = V_1^* + \Delta V$ $\hat{V}_1^* = AX^* + L$ $\Delta V = A \Delta X$
$\hat{V}^T P \hat{V} = \hat{V}^T P \hat{V}^* + \Delta V^T P V$ $\hat{V}^T P \hat{V}^* = L^T P L - U^T N^{-1} U$ $\Delta V^T P V = (A_X X^* + L_X)^T T (A_X X^* + L_X)$	$\hat{V}^T P \hat{V} = \hat{V}^T P \hat{V}^* + \Delta V^T P V$ $V^T P \hat{V}^* = L^T P L - U^T N^{-1} U$ $\Delta V^T P V = (X^* + L_X)^T T (X^* + L_X)$

Table 3.5 Continued

$Q_X = Q_X^* - \Delta Q$ $Q_X^* = N^{-1}$ $\Delta Q = N^{-1}A_X^T T A_X N^{-1}$	$Q_X = Q_X^* - \Delta Q$ $Q_X^* = N^{-1}$ $\Delta Q = N^{-1}TN^{-1}$
$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{n-u+n_x}$	$\hat{\sigma}_o^2 = \frac{\hat{V}^T P \hat{V}}{n}$
$Q_{V_1} = Q_{V_1}^* + \Delta Q_{V_1}$ $Q_{V_1}^* = P^{-1} - AN^{-1}A^T$ $\Delta Q_{V_1} = AN^{-1}A_X T A_X^T N^{-1} A_1^T$ $Q_{V_2} = P_X^{-1} T P_X^{-1}$	$Q_{V_1} = Q_{V_1}^* + \Delta Q_{V_1}$ $Q_{V_1}^* = P^{-1} - AN^{-1}A^T$ $\Delta Q_{V_1} = AN^{-1}TN^{-1}A^T$ $Q_{V_2} = P_X^{-1} T P_X^{-1}$



The diagram shows how various mathematical models can be linked to the general model. The diagram is neither unique or complete. It is important to note that each model can be derived independently, i.e., for each model one can find a quadratic form from which the solution can be derived through minimization.



4. Review of Statistical Concepts (II)

The least squares estimation, as discussed in the previous sections, does not require any specific distribution for the residuals or the observations. It is only necessary that the variance covariance matrix of the observations exists. The model for observation equations is

$$\left.\begin{aligned} &\tilde{V} = AX + \tilde{L} \\ &E(\tilde{V}) = 0; \quad E(\tilde{V}\tilde{V}^T) = \Sigma_{L_b} = \sigma_o^{2}P^{-1} \end{aligned}\right\} \tag{4.1}$$

or

$$E(-\tilde{L}) = AX; \; E(\tilde{L}-E(\tilde{L})(\tilde{L}-E(\tilde{L}))^T = \Sigma_{L_b}$$

Note that X itself is a fixed parameter and not a stochastic variable, whereas the estimate of X, $\hat{X}$, is a stochastic variable. However, if we do want to compute interval estimates for the least squares estimates and test hypotheses the specific form of the population distribution must be known.

4.1 Moment-Generating Function

The special expectation

$$E(e^{t\tilde{x}}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \tag{4.2}$$

is called the moment generating function (mgf) of the random variable $\tilde{x}$. The following notation is used:

$$M(t) = E(e^{t\tilde{x}}) \tag{4.3}$$

For t = 0 we have

$$M(0) = E(1) = \int_{-\infty}^{\infty} f(x)dx = 1 \tag{4.4}$$

since f(x) is a density function.

The importance of the mgf stems from the fact that it is unique and completely determines the distribution of the random variable. Thus, if two random variables have the same mgf, they also have the same distribution.

The mean and the variances are special cases of the mgf:

$$\frac{dM(t)}{dt} = M'(t) = \int_{-\infty}^{\infty} x e^{tx} f(x)dx$$

$$M'(0) = E(\tilde{x}) = \mu \tag{4.5}$$

$$M''(t) = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)dx$$

$$M''(0) = E(\tilde{x}^2)$$

Therefore the variance can be expressed as

$$\sigma^2 = E(\tilde{x}^2) - \mu^2 = M''(0) - \{M'(0)\}^2 \tag{4.6}$$

4.2 Stochastic Independence

The concept of stochastic independence is needed for deriving some of the distributions which are functions of several random variables.

Two random variables $\tilde{x}_1$ and $\tilde{x}_2$ having a joint pdf $f(x_1,x_2)$ are independent if

$$f(x_1,x_2) \equiv f_1(x_1) \; f_2(x_2) \tag{4.7}$$

where $f_1(x_1)$ and $f_2(x_2)$ are the marginal probability density functions. In this case

$$P(a<\tilde{x}_1<b \;,\; c<\tilde{x}_2<d) = P(a<\tilde{x}_1<b) \cdot P(c<\tilde{x}_2<d) \tag{4.8}$$

since

$$P(a<\tilde{x}_1<b,\; c<\tilde{x}_2<d) = \int_a^b \int_c^d f_1(x_1)\, f_2(x_2)\, dx_1\, dx_2$$

$$= \left[\int_a^b f_1(x_1)\, dx_1\right]\left[\int_c^d f_2(x_2)\, dx_2\right]$$

$$= P(a<\tilde{x}_1<b) \cdot P(c<\tilde{x}_2<d)$$

Furthermore, the expected value of the product of linear functions $u(\tilde{x}_1)$ and $v(\tilde{x}_2)$ is equal to the product of their expected values:

$$E\{u(\tilde{x}_1) \cdot v(\tilde{x}_2)\} = E(u(\tilde{x}_1))\, E(v(\tilde{x}_2)) \tag{4.9}$$

Since

$$E\{u(\tilde{x}_1)\, v(\tilde{x}_2)\} = \int_{-\infty}^{\infty}\int u(x_1)\, v(x_2)\, f_1(x_1)\, f_2(x_2)\, dx_1 dx_2$$

$$= \left\{\int_{-\infty}^{\infty} u(x_1)\, f_1(x_1) dx_1\right\}\left\{\int_{-\infty}^{\infty} v(x_2)\, f_2(x_2) dx_2\right\}$$

$$= E\{u(x_1)\}\, E\{v(x_2)\}$$

This theorem can be used to demonstrate an important result about the correlation coefficients:

Let $\tilde{x}_1$ and $\tilde{y}_1$ be independent having the means (μ_1, μ_2) and the variances (σ_1^2, σ_2^2); then

$$E\{(\tilde{x}-\mu_1)(\tilde{y}-\mu_2)\} = \underbrace{E(\tilde{x}-\mu_1)}_{0} \cdot \underbrace{E(\tilde{y}-\mu_2)}_{0} = 0$$

That is, the stochastic independence of $\tilde{x}$ and $\tilde{y}$ implies that the correlation coefficient of $\tilde{x}$ and $\tilde{y}$ is zero. (The converse is not necessarily true).

For the stochastically independent random variables $\tilde{x}_1$ and $\tilde{y}_1$ the moment generating function of the joint pdf is equal to the product of the mgf of the marginal distributions.

$$M(t_1, t_2) = M(t_1, 0) \cdot M(0, t_2) \tag{4.10}$$

since

$$M(t_1, t_2) = E(e^{t_1\tilde{x}_1 + t_2\tilde{x}_2})$$

$$= E(e^{t_1\tilde{x}_1}\, e^{t_2\tilde{x}_2})$$

$$= E(e^{t_1\tilde{x}_1})\, E(e^{t_2\tilde{x}_2})$$

$$= M(t_1, 0) M(0, t_2)$$

4.3 Special One-Dimensional Distributions

4.3.1 Gamma and Chi-Square Distribution

The function

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy \tag{4.11}$$

is called a gamma function. It is proven in calculus that the integral exists for $\alpha>0$ and that its value is a positive number. For $\alpha=1$:

$$\Gamma(1) = \int_0^\infty e^{-y}\, dy = 1 \tag{4.12}$$

If $\alpha>1$ we obtain through integration by parts:

$$\Gamma(\alpha) = (\alpha-1) \int_0^\infty y^{\alpha-2} e^{-y}\, dy = (\alpha-1)\cdot\Gamma(\alpha-1) \tag{4.13}$$

Accordingly, if α is a positive integer greater than one, the gamma function is

$$\Gamma(\alpha) = (\alpha-1)\,(\alpha-2)\,\ldots\,(3)\,(2)\,(1) \tag{4.13}$$

If we define

$$0! = 1$$

then,

$$\Gamma(\alpha) = (\alpha-1)! \tag{4.14}$$

Let us change the variables as

$$y = \frac{x}{\beta} \qquad \text{with} \quad \beta>0$$

with

$$dy = \frac{1}{\beta}\, dx$$

Substituting the new variable into equation (4.11) gives

$$\Gamma(\alpha) = \int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \left(\frac{1}{\beta}\right) dx$$

Dividing both sides by $\Gamma(\alpha)$ gives

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx$$

Since $\alpha>0$, $\beta>0$, and $\Gamma(\alpha)>0$

$$\boxed{f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \qquad 0<x<\infty \\ = 0 \text{ elsewhere}} \tag{4.15}$$

is a positive function whose integration over the domain is one. $f(x)$, therefore, can serve as a probability density function. It is said to be a distribution of the gamma type.

The mean and the variance of this distribution are obtained by evaluating the mgf:

$$M(t) = \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx$$

$$= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x(1-\beta t)/\beta} dx$$

Change of variables: $y = x(1-\beta t)/\beta$ with $t < \frac{1}{\beta}$

$$x = y\beta/(1-\beta t)$$

$$dy = \frac{1-\beta t}{\beta} dx$$

$$M(t) = \int_0^\infty \frac{\beta/(1-\beta t)}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1} e^{-y} dy$$

$$= \left(\frac{1}{1-\beta t}\right)^\alpha \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy}_{1}$$

$$= \left(\frac{1}{(1-\beta t)}\right)^{\alpha} \qquad t < \frac{1}{\beta}$$

since the integral is one according to equation (4.11)

Thus,

$$M'(t) = (-\alpha)\,(1-\beta t)^{(-\alpha-1)}\,(-\beta)$$

$$M''(t) = (-\alpha)\,(-\alpha-1)\,(1-\beta t)^{-\alpha-2}\,(-\beta)^2$$

$$\mu = M'(0) = \alpha\beta \tag{4.16}$$

$$\sigma^2 = M''(0) - \mu^2 = \alpha(\alpha+1)\,\beta^2 - \alpha^2\beta^2 = \alpha\beta^2 \tag{4.17}$$

Consider the special case of the gamma distribution for which

$$\alpha = \frac{r}{2} \tag{4.18}$$

$$\beta = 2$$

r is a positive integer and will be called the degree of freedom. The density function has now the form

$$f(x) = \frac{1}{\Gamma(r/2)\,2^{r/2}}\; x^{r/2-1}\; e^{-x/2}\; dx \qquad 0<x<\infty$$
$$= 0 \quad \text{elsewhere} \tag{4.19}$$

The mgf is

$$M(t) = (1-2t)^{-r/2} \qquad t<\tfrac{1}{2} \tag{4.20}$$

giving a mean and variance of

$$\mu = r \tag{4.21}$$

$$\sigma^2 = 2r$$

The distribution (4.19) is called the Chi-Square distribution of r degrees of freedom. We use the symbol:

$$\tilde{x} \sim \chi^2_r \tag{4.22}$$

The χ^2 distribution is, for small degrees of freedom, very unsymmetric.

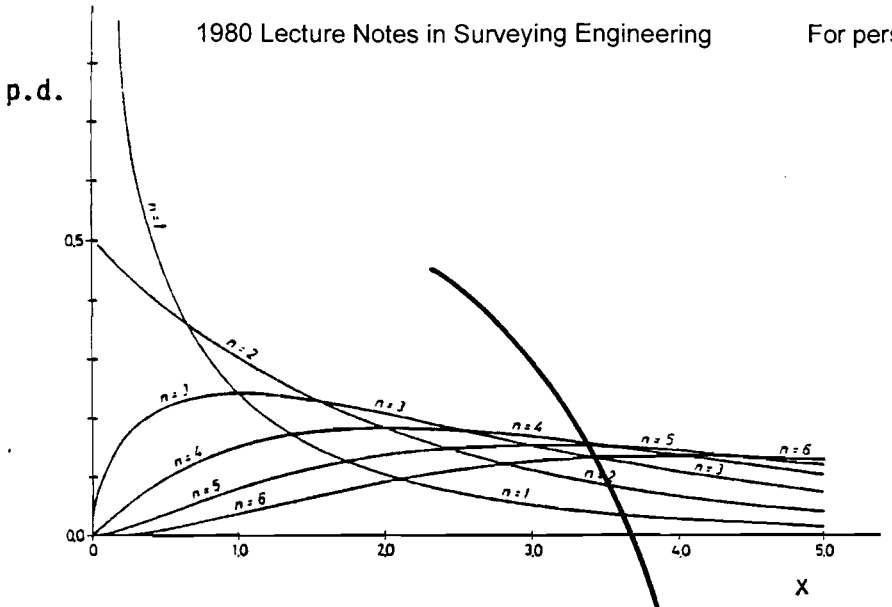


Figure 4.1 χ^2 Distribution for Various Degrees of Freedom

The area under the curve expresses the probability. In most text books on statistics, tables are given for the integral

$$P(\tilde{x}<x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2} dw \qquad (4.23)$$

The degree of freedom is sufficient to completely describe the χ^2 distribution.

4.3.2. Normal Distribution

The normal distribution is of great importance to Survey Engineering. It has been observed that the values of variables describing many naturally occurring phenomena do indeed follow a normal distribution closely. Examples are repeated measurement of angles, distances, etc.

Consider the integral

$$\int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi} \qquad (4.24)$$

whose evaluation can be found in advanced books on calculus. Thus,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy = 1 \qquad (4.25)$$

The integrand can be used as a density function since it is a positive function and the integration over the domain is unity. Change the variables in (4.25) as follows:

$$y = \frac{x-a}{b} \qquad b>0$$

$$dy = \frac{dx}{b}$$

Thus,

$$\int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} e^{\frac{-(x-a)^2}{2b^2}} dx = 1$$

Since b>0

$$f(x) = \frac{1}{b\sqrt{2\pi}} e^{\frac{-(x-a)^2}{2b^2}} \qquad -\infty<x<\infty \qquad (4.26)$$

satisfies the conditions for a pdf. This is the normal distribution. It is readily seen that f(x) in (4.26) is symmetric since

$$f(x) = f(-x)$$

The mgf is

$$M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{b\sqrt{2\pi}} e^{\frac{-(x-a)^2}{2b^2}} dx$$

After changing the variables by

$$x = by + b^2t + a$$

$$dx = bdy$$

and subsequent integration the mgf becomes

$$M(t) = e^{at + \frac{b^2 t^2}{2}} \tag{4.27}$$

The mean and the variance follow from

$$M'(t) = (a + b^2 t)\, M(t)$$

$$M''(t) = M(t)\, b^2 + M(t)\, (a + b^2 t)^2$$

$$\mu = M'(0) = a$$

$$\sigma^2 = M''(0) - \mu^2 = b^2 + a^2 - a^2 = b^2$$

Expressing the density function in terms of the mean and variance we get

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \tag{4.28}$$

Usually the following notation

$$\tilde{x} \sim n(\mu, \sigma^2) \tag{4.29}$$

is used. The two parameters μ and σ^2 completely describe the normal distribution.

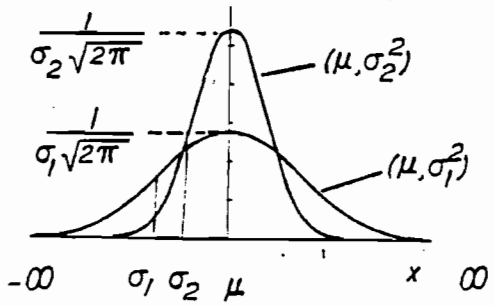


Figure 4.2 Normal Density Function

The following characteristics of the normal distribution can be observed:

a) The distribution is symmetric about the mean.

b) The maximum value of the density function is at $x = \mu$:

$$f(x=\mu) = f_{max} = \frac{1}{\sigma\sqrt{2\pi}}$$

c) For small σ^2(variances) f_{max} is larger and the slope is steeper for values around the mean than in the case of large variances. Values close to the mean have higher density than others.

d) The density function is assymptotically zero at $\pm \infty$.

e) The inflection point is at $x = \mu \pm \sigma$

Set $y = x-\mu$

then $f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2}$

$$f'(y) = \frac{1}{\sigma\sqrt{2\pi}} \left(\frac{-2y}{2\sigma^2}\right) e^{-y^2/2\sigma^2}$$

$$f'(y) = \frac{1}{\sigma^3\sqrt{2\pi}} y\, e^{-y^2/2\sigma^2}$$

$$f''(y) = \frac{1}{\sigma^3\sqrt{2\pi}} \left[e^{-y^2/2\sigma^2} + y\left(\frac{-2y}{2\sigma^2}\right) e^{-y^2/2\sigma^2} \right]$$

$$f''(y) = \frac{1}{\sigma^3\sqrt{2\pi}} \left(1 - \frac{y^2}{\sigma^2}\right) e^{-y^2/2\sigma^2}$$

For the inflection point the second derivative must be zero

$$f''(y) = 0$$

It follows that

$$1 - \frac{y^2}{\sigma^2} = 0$$

must be zero. Thus,

$$y = \pm\sigma \rightarrow x = \mu \pm \sigma$$

4.3.2.1 Standardized Normal Distribution

Consider a random variable which is normally distributed

$$\tilde{x} \sim n(\mu,\sigma^2) \ ;$$

then the transformed variable

$$\tilde{w} = \frac{\tilde{x}-\mu}{\sigma} \sim n(0,1) \qquad (4.30)$$

has a normal distribution with zero mean and unit variance (= standardized normal distribution). The probability that the random variable $\tilde{w}$ is smaller than a certain value w is

$$P(\tilde{w}<w) = P(\frac{\tilde{x}-\mu}{\sigma} < w)$$

$$= P(\tilde{x} < w\sigma+\mu)$$

$$= \int_{-\infty}^{w\sigma+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

Change the variables of integration

$$y = (x-\mu)/\sigma$$

$$dy = dx/\sigma$$

then

$$P(\tilde{w}<w) = \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy$$

But

$$g(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \qquad -\infty < w < \infty \tag{4.31}$$

in n(0,1). This fact considerably simplifies calculations of probabilities concerning normally distributed values.

Suppose $\tilde{x} \sim n(\mu,\sigma^2)$

and $c_1 < c_2$

then

$$P(c_1 < \tilde{x} < c_2) = P(\tilde{x} < c_2) - P(\tilde{x} < c_1)$$

$$= P\left(\frac{\tilde{x}-\mu}{\sigma} < \frac{c_2-\mu}{\sigma}\right) - P\left(\frac{\tilde{x}-\mu}{\sigma} < \frac{c_1-\mu}{\sigma}\right)$$

$$= \int_{-\infty}^{(c_2-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \, dw - \int_{-\infty}^{(c_1-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \, dw$$

since

$$\tilde{w} = \frac{\tilde{x}-\mu}{\sigma} \sim n(0,1)$$

That is, the probability concerning $\tilde{x} \sim n(\mu,\sigma^2)$ can be expressed in terms of probability concerning $\tilde{w} \sim n(0,1)$. Usually, the integral

$$N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \, dw \tag{4.32}$$

is tabulated. Since the normal density function is symmetric, the relation

$$N(-x) = 1 - N(x) \tag{4.33}$$

holds

Thus, if

$$\tilde{x} \sim n\ (\mu, \sigma^2)$$

$$P(c_1 < \tilde{x} < c_2) = P\left(w_1 < \tilde{w} < w_2\right) = P\left(\frac{c_2-\mu}{\sigma} < \tilde{w} < \frac{c_1-\mu}{\sigma}\right) \qquad (4.34)$$

$$= N\left(\frac{c_2-\mu}{\sigma}\right) - N\left(\frac{c_1-\mu}{\sigma}\right)$$

Example 1 $\tilde{x} \sim n\ (2, 25)$

$$P(0 < \tilde{x} < 10) = N\left(\frac{10-2}{5}\right) - N\left(\frac{0-2}{5}\right)$$

$$= N(1.6) - N(-0.4)$$

$$= 0.945 - (1-0.655) = 0.600$$

Example 2

$$\tilde{x} \sim n\ (\mu, \sigma^2)$$

What is the probability that the outcome lies within one σ from the mean?

$$P(\mu-\sigma < \tilde{x} < \mu+\sigma) = N\left(\frac{\mu+\sigma-\mu}{\sigma}\right) - N\left(\frac{\mu-\sigma-\mu}{\sigma}\right)$$

$$= N(1) - N(-1)$$

$$= 0.841 - (1 - 0.841) = 0.682$$

Example 3

$$\tilde{x} \sim n\ (\mu, \sigma^2)$$

$$P(\mu-3\sigma < \tilde{x} < \mu+3\sigma) = N(3) - N(-3)$$

$$= 0.9973$$

It is important to note that in the case of normal distribution the observations fall in about 68% of all cases within one standard deviation from the mean, and only every 370th observation will deviate from the mean by more than 3σ. Therefore, 3σ is sometimes taken as the limit to what can be regarded as random error. Any larger deviation from the mean is usually considered a blunder. Principally speaking, large errors cannot be avoided, but their occurence is unlikely.

The values in the following tables are quite often used:

x	1σ	2σ	3σ	4σ
N(x) - N(-x)	0.6827	0.9544	0.9973	0.99994

x	0.674σ	1.645σ	1.960σ	2.326σ	2.576σ	3.291σ
N(x) - N(-x)	0.5	0.90	0.95	0.98	0.99	0.999

Map Accuracy Standard (MAS): The limit of 1.645 σ is sometimes referred to as the Map Accuracy Standard(MAS). It is the limit which contains 90% of the probability:

$$P(-x_{MAS} < \tilde{x} < x_{MAS}) = P(-w_{MAS} < \tilde{w} < w_{MAS}) = 0.90$$

$$2\int_0^{w_{MAS}} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \, dw = 0.9$$

From integration we find

$$w_{MAS} = 1.645$$

giving

$$x_{MAS} = \mu \pm 1.645\ \sigma \qquad (4.35)$$

Probable Error (PE): Another measure of precision is the probable error (P.E.). It is defined such that its range contains 50% probability. Thus,

$$P(-x_{PE} < \tilde{x} < x_{PE}) = P(-w_{PE} < \tilde{w} < w_{PE}) = 1/2$$

$$\int_0^{w_{PE}} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \, dw = 1/4$$

From numerical integration we obtain

$$w_{PE} = 0.674$$

Thus
$$x_{PE} = \mu \pm w_{PE}\,\sigma$$
$$= \mu \pm 0.674\,\sigma \qquad (4.36)$$

Average Error (AE): The average error is defined as

$$AE = \frac{\sum_{i}^{n} |x_i - \mu|}{n}$$

We can combine the errors of the same magnitude (but with opposite sign):

$$AE = \frac{m_1}{n} |x_1 - \mu| + \frac{m_2}{n} |x_2 - \mu| + \ldots$$

where m_i is the number of occurence of the error i. If we associate probabilities to the m_i then we obtain

$$AE = p_1 |x_1 - \mu| + p_2 |x_2 - \mu| + \ldots = \Sigma\, p_i |x_i - \mu|$$

The definition of the AE can be extended to the continuous case. Denote

$$\varepsilon = (x_i - \mu) \sim n(0, \sigma^2)$$

Thus

$$AE = \int_{-\infty}^{\infty} |\varepsilon| \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\varepsilon}{2\sigma^2}} d\varepsilon$$

Since this distribution is symmetric with respect to zero we can limit the integration to one half of the range and replace $|\varepsilon|$ by ε:

$$AE = 2\int_{0}^{\infty} \varepsilon \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\varepsilon^2}{2\sigma^2}} d\varepsilon$$
$$= \frac{2}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} \varepsilon e^{-\frac{\varepsilon^2}{2\sigma^2}} d\varepsilon$$

Changing the variables of integration:

$$u = \frac{\varepsilon^2}{2\sigma^2}$$

$$du = \frac{\varepsilon}{\sigma^2}\, d\varepsilon$$

we get

$$AE = \frac{2\sigma^2}{\sigma\sqrt{2\pi}} \int_0^\infty e^{-u}\, du$$

$$= \frac{2\sigma}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}\,\sigma = 0.7978\,\sigma$$

where use is made of equation (4.12).

Thus $$X_{AE} = \mu \pm 0.7986\,\sigma \qquad (4.37)$$

The three measures of precision are related as follows:

$$\boxed{PE < AE < \text{ST. Dev. } (\sigma)} \qquad (4.38)$$

In case of the normal distribution it is easy to identify location parameters other than the mean.

The mode of a distribution of one random variable $\tilde{x}$ is that value of x that maximizes the pdf f(x). In case of the normal distribution the mode coincides with the mean.

The median of a distribution of one random variable $\tilde{x}$ is a value of x such that $P(\tilde{x}<x) < \frac{1}{2}$ and $P(\tilde{x}>x) \geq \frac{1}{2}$. In case of the normal distribution the median and the mean coincide.

4.4 Distribution of Functions

4.4.1 Square of the Standardized Normal Distribution

As a simple example of a function of a random variable we compute the distribution of the square of a random variable which is normally distributed.

Given that

$$\tilde{x} \sim n(\mu, \sigma^2)$$

then $$\tilde{w} = \frac{(\tilde{x} - \mu)}{\sigma} \sim n(0,1)$$

Find the distribution of

$$\tilde{v} = \tilde{w}^2 = \frac{(\tilde{x}-\mu)^2}{\sigma^2} \qquad (4.40)$$

$$P(\tilde{w}^2 \leq v)$$

$$= P(-\sqrt{v} \leq \tilde{w} \leq \sqrt{v})$$

$$= 2\int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw \qquad 0 \leq v \leq \infty$$

Change the variable of integration: $w = \sqrt{y} = \sqrt{v}$

$$dw = \frac{1}{2\sqrt{y}} dy = \frac{1}{2w} dy$$

then

$$P(\tilde{w} \leq v) = \int_0^v \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-\frac{y}{2}} dy \qquad 0 \leq v \leq \infty$$

The probability density function is

$$f(v) = F'(v)$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{v}} e^{-\frac{v}{2}} = \frac{1}{\sqrt{\pi}\sqrt{2}} v^{\frac{1}{2}-1} e^{-\frac{v}{2}} \qquad 0 \leq v \leq \infty$$

Since this is a pdf it must be that

$$\int_0^\infty f(v)\, dv = 1$$

Comparing the density function with the χ^2 distribution we see that

$$\tilde{v} \sim \chi^2_1 \tag{4.41}$$

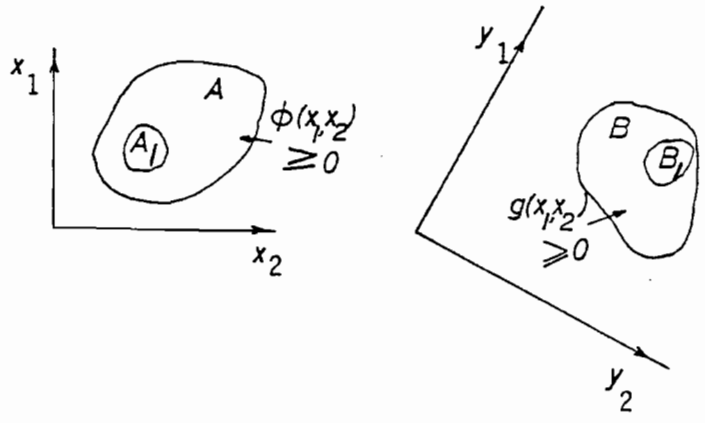
with one degree of freedom and that

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

Therefore, the square of the standardized normal distribution is a chi-square distribution with one degree of freedom.

4.4.2 Transformation by a One-to-One Mapping

Consider two continuous random variables $\tilde{x}_1$ and $\tilde{x}_2$ with a joint density function $\phi\,(x_1,x_2)$. The region in the (x_1, x_2) plane in which the density function is non-zero is denoted by A.



A: region of non-zero pdf ϕ B: region of non-zero pdf g

Figure 4.3 Transformation of Random Variables

What is the density function of a function of the two random variables

$$\tilde{y}_1 = u_1 (\tilde{x}_1, \tilde{x}_2) \tag{4.42}$$

If we have two functions

$$\begin{aligned} y_1 &= u_1 (x_1, x_2) \\ y_2 &= u_2 (x_1, x_2) \end{aligned} \tag{4.43}$$

which define a one-to-one transformation which maps the region A onto the region B, then we can find the joint pdf of

$$\begin{aligned} \tilde{y}_1 &= u_1 (\tilde{x}_1, \tilde{x}_2) \\ \tilde{y}_2 &= u_2 (\tilde{x}_1, \tilde{x}_2) \; . \end{aligned} \tag{4.44}$$

Let A_1 and B_1 be subsets of A and B such that A_1 is mapped onto B_1; then the events

$$(\tilde{x}_1, \tilde{x}_2) \in A \text{ and } (\tilde{y}_1, \tilde{y}_2) \in B$$

are equivalent. It follows that

$$\begin{aligned} P[(\tilde{y}_1, \tilde{y}_2) \in B] &= P[(\tilde{x}_1, \tilde{x}_2) \in A] \\ &= \iint_A \phi (x_1, x_2) \, dx_1 \, dx_2 \end{aligned}$$

In order to carry out this integration we change the variables of integration. The inverse function of (4.43) is

$$\begin{aligned} x_1 &= w_1 (y_1, y_2) \\ x_2 &= w_2 (y_1, y_2) \end{aligned}$$

From analysis it is known that

$$\iint_A \phi(x_1,x_2) dx_1 dx_2 = \iint_B \phi[w_1(y_1,y_2), w_2(y_1,y_2)] \cdot |J| \, dy_1 dy_2 \tag{4.45}$$

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

is the Jacobian determinant. Thus

$$P\,[(\tilde{y}_1,\tilde{y}_2)\,\varepsilon B] = \iint_B \phi[w_1(y_1,y_2),w_2(y_1,y_2)]\;|J|\;dy_1\,dy_2, \qquad (4.46)$$

which implies that the joint pdf of $(\tilde{y}_1,\tilde{y}_2)$ is

$$g(y_1,y_2) = \phi\;[w_1(y_1,y_2),\; w_2(y_2,y_1)]\;|J| \quad (y_1,y_2)\;\varepsilon B \qquad (4.47)$$

$$= \text{zero elsewhere}$$

The marginal pdf of $\tilde{y}_1$ can be obtained from the joint pdf of $g(y_1,y_2)$ in the usual manner by integrating on y_2.

This method of transforming random variables can, of course, be extended to more than two random variables.

4.4.3 The t - Distribution

Assume that $\tilde{w}$ and $\tilde{v}$ are two stochastically independent random variables with unit normal and χ^2_r distribution respectively:

$$\begin{aligned} &\tilde{w} \sim n\,(0,1) \\ &\tilde{v} \sim \chi^2_r \\ &\tilde{w},\,\tilde{v} \text{ stochastically independent} \end{aligned} \qquad (4.48)$$

Because both random variables are stochastically independent, the joint density function is the product of the individual probability density functions:

$$\phi(w,v) = \frac{1}{\sqrt{2\pi}}\;e^{-\frac{w^2}{2}} \cdot \frac{1}{\Gamma(\frac{r}{2})2^{r/2}}\;v^{\frac{r}{2}-1}\;e^{-\frac{v}{2}}$$

$$-\infty < w < \infty$$
$$0 < v < \infty$$

We are interested in the density function of the new random variable

$$\tilde{t} = \frac{\tilde{w}}{\sqrt{\frac{\tilde{v}}{r}}} \qquad (4.49)$$

Consider the transformation

$$t = \frac{w}{\sqrt{\frac{v}{r}}}$$

$$u = v$$

This is a one-to-one transformation of $A = \{(w,v) : -\infty<w<\infty,\ 0<v<\infty\}$ to $B = \{(t, u) : -\infty<t<\infty,\ 0<u<\infty\}$. The inverse transformation is

$$w = t\sqrt{\frac{u}{r}}$$

$$v = u$$

The second transformation equation ($u = v$) is arbitrary as long as the transformation is one-to-one. The Jacobian determinant which is needed to compute the density function $f(t,u)$ is

$$|J| = \begin{vmatrix} \frac{\partial w}{\partial t} & \frac{\partial w}{\partial u} \\ \frac{\partial w}{\partial t} & \frac{\partial w}{\partial u} \end{vmatrix} = \frac{\sqrt{u}}{\sqrt{r}}$$

Thus, according to equation (4.47) the joint density function is:

$$f(t,u) = \phi(t\frac{\sqrt{u}}{\sqrt{r}}, u)\ |J|$$

$$= \frac{1}{\sqrt{2\pi}\ \Gamma\left(\frac{r}{2}\right) 2^{r/2}}\ u^{\frac{r}{2}-1}\ e^{\left[-\frac{u}{2}\left(1+\frac{t^2}{r}\right)\right]}\ \cdot\sqrt{\frac{u}{r}}$$

$$-\infty<t<\infty$$
$$0<u<\infty$$

Since we are only interested in the parameter t we compute the marginal probability density function

$$f_1(t) = \int_{-\infty}^{\infty} f(t,u)\,du$$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\;\Gamma\left(\frac{r}{2}\right)2^{r/2}}\; u^{(r+1)/2-1}\; e^{\left[-\frac{u}{2}\left(1+\frac{t^2}{r}\right)\right]}\,du$$

Changing the variables of integration by

$$Z = \frac{u}{2}\left(1 + \frac{t^2}{r}\right)$$

gives

$$f_1(t) = \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\;\Gamma\left(\frac{r}{2}\right)2^{r/2}} \left(\frac{2Z}{1+\frac{t^2}{r}}\right)^{\left(\frac{r+1}{2}\right)-1} e^{-Z} \left(\frac{2}{1+\frac{t^2}{r}}\right) dZ$$

$$= \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\;\Gamma\left(\frac{r}{2}\right)} \frac{1}{\left(1+\frac{t^2}{r}\right)^{(r+1)/2}} \qquad -\infty<t<\infty \qquad (4.50)$$

where use has been made of equation (4.11). $f_1(t)$ is the probability density function of the t - distribution. It is completely determined by the the degree of freedom, r.

The probability as related to the t - distribution is usually given in tables:

where

$$\boxed{\begin{aligned} &P\left(\tilde{t}_r > t_{\alpha,r}\right) = 1 - T_{\alpha,r} = \alpha \\ &T_{\alpha,r} = \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\;\Gamma\left(\frac{r}{2}\right)} \int_{-\infty}^{t_{\alpha,r}} \frac{dZ}{\left(1+\frac{Z^2}{r}\right)^{(r+1)/2}} \end{aligned}} \qquad (4.51)$$

From equation (4.50) it is seen that the density function is symmetric with respect to zero. Furthermore if $r = \infty$, the t distribution is identical to the standardized normal distribution n (0, 1).

when $t_r \xrightarrow[r\to\infty]{} n(0,1)$

Based on the definition of exponential functions one has

$$\lim_{r\to\infty} \frac{1}{\left(1 + \frac{t^2}{r}\right)^{\frac{r+1}{2}}} = e^{-\frac{t^2}{2}} \tag{4.52}$$

Using the relation of equation (4.13) the two gamma functions in (4.50) can be written as:

$$\frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} = \frac{\frac{r-1}{2}\,\Gamma\left(\frac{r-1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} = \frac{\frac{n-1}{2}\,\Gamma\left(\frac{r-1}{2}\right)}{\left(\frac{r}{2}-1\right)\Gamma\left(\frac{r}{2}-1\right)} \tag{4.53}$$

Denoting

$$\nu_r \equiv \frac{\Gamma\left(\frac{r}{2}\right)}{\Gamma\left(\frac{r-1}{2}\right)}$$

then (4.53) becomes

$$\nu_{r+1} = \frac{\frac{r-1}{2}}{\nu_r} = \frac{\frac{r-1}{2}\cdot\nu_{r-1}}{\left(\frac{r}{2}-1\right)}$$

or, equivalently,

$$\nu_r\,\nu_{r+1} = \frac{r}{2} - \frac{1}{2}$$

$$\nu_r\,\nu_{r-1} = \frac{r}{2} - 1$$

These expressions yield for a sufficiently large r:

$$\nu_r^2 = \frac{r}{2} - \frac{6}{4} \tag{4.54}$$

The expression needed in the density function (4.50) is

$$\nu_{r+1} = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} \approx \sqrt{\frac{r}{2}} \tag{4.55}$$

which is approximately equal to the square root of one half of the degree of freedom in the case of a large r. Substituting equations (4.55) and (4.52) in equation (4.50) yields a density distribution

$$f(t) \underset{r=\text{large}}{=} \frac{\sqrt{\frac{r}{2}}}{\sqrt{\pi r}} e^{-\frac{t^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

which is identical to the standardized normal distribution.

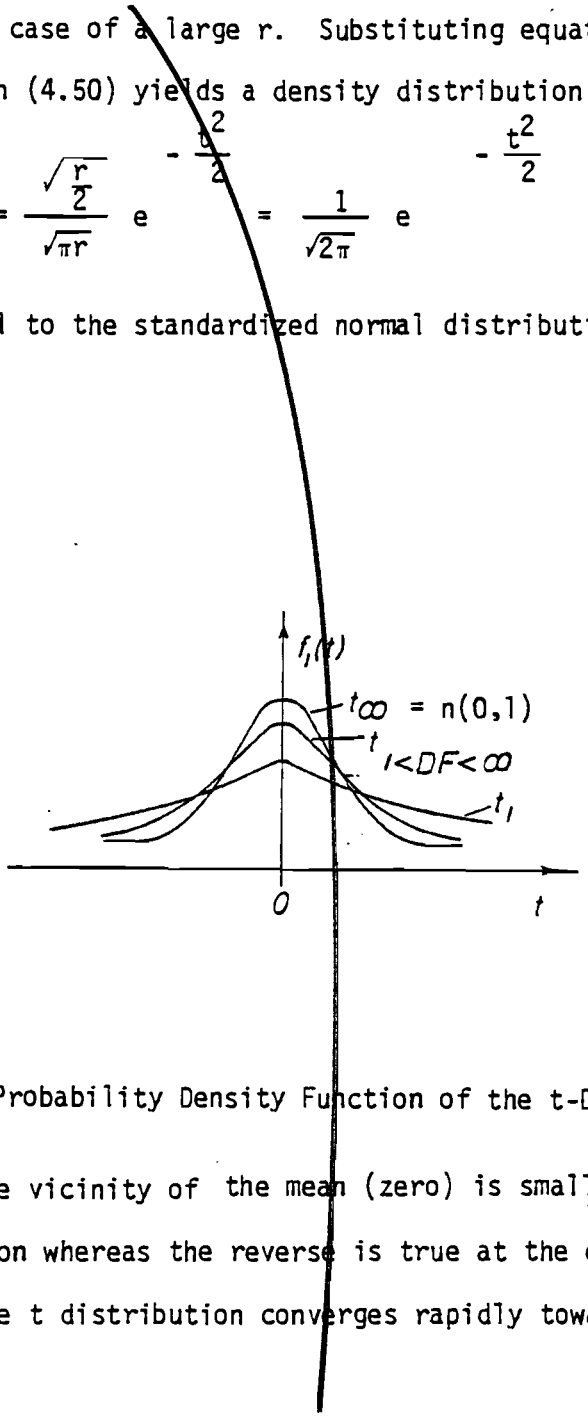


Figure 4.4 The Probability Density Function of the t-Distribution

The density in the vicinity of the mean (zero) is smaller than for the unit normal distribution whereas the reverse is true at the extremities of the distribution. The t distribution converges rapidly toward the normal distribution.

4.4.4 The F - distribution

Consider two stochastically independent random variables having a χ^2 distribution with r_1 and r_2 degrees of freedom respectively

$$\begin{aligned} \tilde{u} &\sim \chi^2_{r_1} \\ \tilde{v} &\sim \chi^2_{r_2} \end{aligned} \tag{4.56}$$

The joint pdf is

$$\phi(u,v) = \frac{1}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)2^{\frac{r_1+r_2}{2}}}\, u^{\frac{r_1}{2}-1}\, v^{\frac{r_2}{2}-1}\, e^{-\frac{(u+v)}{2}}$$

$$0<u<\infty$$
$$0<v<\infty$$

We are interested in the distribution of the random variable

$$\tilde{F} = \frac{\frac{\tilde{u}}{r_1}}{\frac{\tilde{v}}{r_2}} \tag{4.57}$$

The equations

$$f = \frac{\frac{u}{r_1}}{\frac{v}{r_2}}$$

$$z = v$$

define a one-to-one transformation which maps the set $A = \{(u,v): 0<u<\infty, 0<v<\infty\}$ onto the set $B = \{(f,z): 0<f<\infty, 0<z<\infty\}$. The inverse transformation is

$$u = \frac{r_1}{r_2} zf$$

$$v = z$$

has the Jacobian determinant

$$|J| = \frac{r_1}{r_2} z$$

After substitution, the joint p.d.f. becomes

$$\phi(f,z) = \frac{1}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right) 2^{\frac{(r_1+r_2)}{2}}} \left(\frac{r_1 zf}{r_2}\right)^{\frac{r_1}{2}-1} z^{\frac{r_2}{2}-1} e^{-\frac{z}{2}\left(\frac{r_1 f}{r_2}+1\right)} \cdot \frac{r_1 z}{r_2}$$

Integration over the variable z gives the marginal p.d.f. g(f)

$$g(f) = \int_{-\infty}^{\infty} \phi(f,z)\, dz$$

$$= \int_0^{\infty} \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} f^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right) 2^{\frac{r_1+r_2}{2}}} z^{\frac{r_1+r_2}{2}-1} e^{-\frac{z}{2}\left(\frac{r_1 f}{r_2}+1\right)} dz$$

Changing the variable of integration by

$$y = \frac{z}{2}\left(\frac{r_1}{r_2} f + 1\right)$$

provides the final expression for the p.d.f.

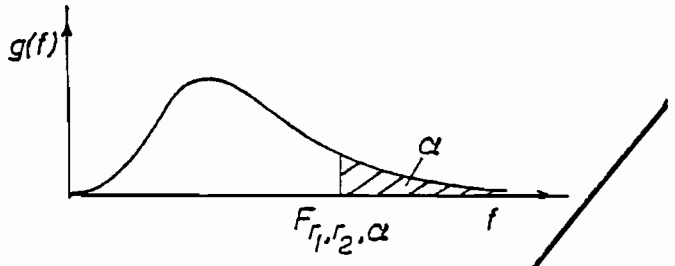
$$g_1(f) = \int_0^\infty \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} f^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right) 2^{\frac{r_1+r_2}{2}}} \left(\frac{2y}{\frac{r_1 f}{r_2}+1}\right)^{\frac{r_1+r_2}{2}-1} e^{-y} \frac{2}{\frac{r_1 f}{r_2}+1} dy$$

$$= \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \frac{f^{\frac{r_1}{2}-1}}{\left(1+\frac{r_1 f}{r_2}\right)^{\frac{r_1+r_2}{2}}}$$

with $0 < f < \infty$ (4.58)

where use has been made of equation (4.11). This is the F-distribution. It is completely determined by the two parameters r_1 and r_2. The area under the F density function is usually given in tables, e.g.:

$$P(\tilde{F}_{r_1,r_2} > F_{r_1,r_2,\alpha}) = \int_{F_{r_1,r_2,\alpha}}^{\infty} g_1(f)df \quad (4.59)$$



where $F_{r_1 r_2 \alpha}$ is that point on the absissa to the right of which lies the area α.

Sometimes tables are given in a slightly different form. Care should always be taken to identify the degrees of freedom properly since the density function is not symmetric in these variables. It can also be shown that the following relation holds

$$F_{r_1,r_2,\alpha} = \frac{1}{F_{r_2,r_1,1-\alpha}} \tag{4.60}$$

Figure 4.5 shows the density function for various degrees of freedom.

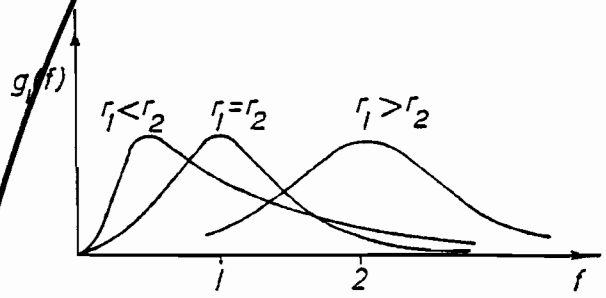


Figure 4.5 The F-Distribution

In case both degrees of freedom are the same

$$P(\tilde{F}_{r_i r_i} \leq 1) = 0.5$$

for any r_i.

4.4.5 Moment Generating Function Techniques

Previously the technique of one-to-one transformations was used in order to derive the distribution of non-linear functions. In case of linear functions the moment generating function technique is more convenient. Assume:

$$\phi(x_1, x_2, \ldots x_n)$$

is the joint p.d.f. of n random variables $(\tilde{x}_1, \tilde{x}_2 \ldots \tilde{x}_n)$. Let

$$\tilde{y} = u(\tilde{x}_1, \tilde{x}_2 \ldots \tilde{x}_n)$$

be a function of the random variables and we are interested in the density function g(y). The moment generating function is

$$M(t) = E(e^{t\tilde{y}}) = \int_{-\infty}^{\infty} e^{ty} \; g(y)dy$$

$$= E\left(e^{tu(\tilde{x}_1 \tilde{x}_2, \ldots \tilde{x}_n)}\right)$$

$$= \int\cdots\int_{-\infty}^{\infty} e^{tu\,(x_1 \ldots x_n)} \; \phi\,(x_1 \ldots x_n)dx_1 \ldots dx_n$$

The procedure is that we compute $E(e^{tu(\tilde{x}_1 \ldots \tilde{x}_n)})$ based on $\phi(x_1 \ldots x_n)$. This gives us the mgf $E(e^{t\tilde{y}})$ without having to compute the density function g(y) first. If the mgf is seen to be that of a certain kind of distribution, the uniqueness property makes it certain that $\tilde{y}$ has that kind of distribution.

4.4.6 Distribution of Some Linear Functions

The following linear functions of random variables are of special importance in statistical analysis.

Case 1: Assume that $(\tilde{x}_1, \tilde{x}_2 \ldots \tilde{x}_n)$ are n stochastically independent variables, each having a normal distribution:

$$\tilde{x}_1 \sim n(\mu_1, \sigma_1^2)$$

$$\tilde{x}_2 \sim n(\mu_2, \sigma_2^2)$$

$$\vdots$$

$$\tilde{x}_n \sim n(\mu_n, \sigma_n^2)$$

Then the random variable

$$\tilde{y} = k_1\tilde{x}_1 + k_2\tilde{x}_2 + \ldots k_n\tilde{x}_n$$

is distributed as

$$\boxed{\tilde{y} \sim n\left(\sum_{i}^{n} k_i\mu_i, \sum_{i}^{n} k_i^2\sigma_i^2\right)} \tag{4.61}$$

where the k_i are constants. The proof is based on the mgf. Since the random variables are stochastically independent we can write

$$M(t) = E\left(e^{t(k_1\tilde{x}_1 + k_2\tilde{x}_2 + \ldots k_n\tilde{x}_n)}\right)$$

$$= E\left(e^{tk_1\tilde{x}_1}\right) \; E\left(e^{tk_2\tilde{x}_2}\right) \; \ldots \; E\left(e^{tk_n\tilde{x}_n}\right) \tag{4.62}$$

The mgf for the normal distribution is, according to equation (4.27),

$$M(t) = E\left(e^{tx_i}\right) = e^{\mu_i t + \frac{\sigma_i^2 t}{2}}$$

This expression is valid for all real t's. Therefore, it is also valid for $t' = tk_i$:

$$E(e^{tk_i x_i}) = e^{\mu_i (k_i t) + \frac{\sigma_i^2 (k_i t)^2}{2}}$$

Substituting this expression in equation (4.62) gives

$$M(t) = \prod_{i=1}^{u} e^{(\mu_i k_i)t + \frac{(k_i^2 \sigma_i^2)\, t^2}{2}}$$

$$= e^{(\Sigma k_i \mu_i)t + \frac{(\Sigma k_i \sigma_i^2)\, t^2}{2}}$$

This is the mgf for $n(\Sigma k_i \mu_i, \Sigma k_i^2 \sigma_i^2)$.

Therefore, the sum of mutually stochastically independent normally distributed variables has a normal distribution.

Case 2: Assume $(\tilde{x}_1, \tilde{x}_2 \ldots \tilde{x}_n)$ are n stochastically independent random variables each having a chi-square distribution:

$$\tilde{x}_1 \sim \chi^2_{r_1}$$

$$\tilde{x}_2 \sim \chi^2_{r_2}$$

$$\vdots$$

$$\tilde{x}_n \sim \chi^2_{r_n}$$

Then the random variable

$$\tilde{y} = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 \ldots + \tilde{x}_n$$

is distributed

$$\boxed{\tilde{y} \sim \chi^2_{\Sigma r_i}} \qquad (4.63)$$

Since the random variables are stochastically independent the mgf is

$$M(t) = E(e^{t(\tilde{x}_1+\tilde{x}_2+\ldots\tilde{x}_n)})$$

$$= E(e^{t\tilde{x}_1}) \; E(e^{t\tilde{x}_2}) \ldots E(e^{t\tilde{x}_n})$$

The mgf for the chi-square distribution is according to equation (4.20)

$$E(e^{tx_i}) = (1-2t)^{-r_i/2}$$

with $t<\frac{1}{2}$. Using this expression the mgf for $\tilde{y}$ becomes

$$M(t) = (1-2t)^{-(r_1+r_2+\ldots r_n)/2} \qquad t<\tfrac{1}{2}$$

But this is the mgf of $\chi^2_{\Sigma r_i}$

Case 3: Assume $(\tilde{x}_1,\tilde{x}_2,\ldots\tilde{x}_n)$ are n stochastically independent random variables each having a normal distribution:

$$\tilde{x}_1 \sim n(\mu_1,\sigma_1^2)$$

$$\tilde{x}_2 \sim n(\mu_2,\sigma_2^2)$$

$$\vdots$$

$$\tilde{x}_n \sim n(\mu_n,\sigma_n^2)$$

Then $$\boxed{\tilde{y} = \sum^{n}\left(\frac{\tilde{x}_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2_n} \qquad (4.64)$$

The proof follows from case 2 and the fact that

$$\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2 \sim \chi^2_1$$

4.4.7 Distribution of Sample Mean and Variance

Assume $(x_1, x_2 \ldots x_n)$ are n stochastically independent random variables each having the same normal distribution:

$$\tilde{x}_1 \sim n(\mu, \sigma^2)$$

$$\tilde{x}_2 \sim n(\mu, \sigma^2)$$

$$\vdots$$

$$\tilde{x}_n \sim n(\mu, \sigma^2)$$

The mean is $\bar{\tilde{x}} = \dfrac{\sum_{i}^{n} \tilde{x}_i}{n}$

The distribution of the mean is readily derived from Case 1 of the previous section:

We have

$$\mu_1 = \mu_2 = \ldots . \mu_n = \mu$$

$$\sigma_1^2 = \sigma_2^2 = \ldots \sigma_n^2 = \sigma^2$$

$$k_1 = k_2 = \ldots . k_n = 1/n$$

Hence

$$\bar{\tilde{x}} \sim n\left(\sum_{i}^{n} \frac{\mu}{n}, \sum_{i}^{n} \frac{1}{n^2} \sigma_i^2\right)$$

or

$$\boxed{\bar{\tilde{x}} \sim n\left(\mu, \frac{\sigma^2}{n}\right)} \qquad (4.65)$$

Before the distribution of the sample variance can be derived the stochastic independence of the mean and variance has to be established. First the stochastic independence of $\tilde{\bar{x}}$ and $(\tilde{x}_1 - \tilde{\bar{x}},\ \tilde{x}_2 - \tilde{\bar{x}},\ \ldots \tilde{x}_n - \tilde{\bar{x}})$ is shown. Let $M(t, t_1 \ldots t_n)$ be the mgf of $(\tilde{\bar{x}},\ \tilde{x}_1 - \tilde{\bar{x}},\ \tilde{x}_2 - \tilde{\bar{x}}, \ldots \tilde{x}_n - \tilde{\bar{x}})$, then

$$M(t,t_1 \ldots t_n) = \int_{-\infty}^{\infty} \cdots \int e^{\{t\bar{x} + t_1(x_1-\bar{x}) + \ldots + t_n(x_n-\bar{x})\}} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-\Sigma(x_i-\mu)^2}{2\sigma^2}} dx_1 \ldots dx_n \qquad (4.66)$$

Note that $\bar{x}$ is a function of the random variables $\tilde{x}_1 \ldots \tilde{x}_n$. We, therefore, only need the joint density function $f(x_1 \ldots x_n)$ which can be written as

$$f_1(x_1)\ f_2(x_2) \ldots f_n(x_n)$$

since the $\tilde{x}_i$ are stochastically independent. Consider in (4.66) the integral on x_i only

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\{t+nt_i-(t_1+\ldots+t_n)\}\frac{x_i}{n} - \frac{(x_i-\mu)^2}{2\sigma^2}} dx_i \qquad (4.67)$$

This expression is valid for any $t, t_1, \ldots t_n$.

If we set

$$\dot{t} = \{t + nt_i - (t_1+\ldots+t_n)\}\ \frac{1}{n}$$

the expression (4.67) is seen to be the mgf of the normal distribution $n(\mu,\sigma^2)$, i.e.,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\dot{t}x_i - \frac{(x_i-\mu)^2}{2\sigma}} dx_i = e^{\mu\dot{t} + \frac{\sigma^2\dot{t}^2}{2}} \qquad (4.68)$$

The expression for $\dot{t}$ can be modified as

$$\dot{t} = \frac{1}{n}\{t + nt_i - \Sigma t_i\} = \frac{1}{n}\{t + n(t_i - \bar{t})\}$$

with $\bar{t} = \frac{\Sigma t_i}{n}$.

Using these expressions the mgf (4.66) can be written as the product of individual mgf's of the type (4.58) as follows:

$$M(t,t_1 \ldots t_n) = e^{\sum_1^n \left[\frac{\mu}{n}t + \mu(t_i - \bar{t}) + \sigma^2 \frac{\{t + n(t_i-\bar{t})\}^2}{2n^2}\right]}$$

Since $\sum_1^n (t_i - \bar{t}) = 0$ this expression simplifies as

$$M(t,t_1 \ldots t_n) = e^{\left[\mu t + \frac{\sigma^2 t^2}{2n} + \frac{\sigma^2 \Sigma (t_i-\bar{t})^2}{2}\right]}$$

$$= e^{\mu t + \frac{\sigma^2 t}{2n}} \cdot e^{\frac{\sigma^2 \Sigma (t_i-\bar{t})^2}{2}}$$

$$= M(t,0) \cdot M(0,t_1 \ldots t_n)$$

The first factor is the moment generating function of $\tilde{\bar{x}}$. The second factor does not depend on t. According to (4.10), $\tilde{\bar{x}}$ and $(\tilde{x}_1 - \tilde{\bar{x}}, \tilde{x}_2 - \tilde{\bar{x}}, \ldots \tilde{x}_n - \tilde{\bar{x}})$ are stochastically independent. It follows that $\tilde{\bar{x}}$ and

$$(n-1)\tilde{s}^2 = \sum_1^n (\tilde{x}_i - \tilde{\bar{x}})^2 \qquad (4.69)$$

are also stochastically independent.

We can now compute the distribution of the sample variance.

$$\Sigma (\tilde{x}_i - \mu)^2 = \Sigma(\tilde{x}_i - \tilde{\bar{x}} + \tilde{\bar{x}} - \mu)^2$$

$$= \Sigma(\tilde{x}_i - \tilde{\bar{x}})^2 + n(\tilde{\bar{x}} - \mu)^2$$

since $2(\tilde{\bar{x}}-\mu) \cdot \underbrace{\Sigma(\tilde{x}_i-\tilde{\bar{x}})}_{0} = 0$

Thus,

$$\frac{\sum_{i}^{n}(\tilde{x}_i-\mu)^2}{\sigma^2} = \frac{n(\tilde{\bar{x}}-\mu)^2}{\sigma^2} + \frac{(n-1)\tilde{s}^2}{\sigma^2} \qquad (4.70)$$

where

$$\frac{\sum_{i}^{n}(\tilde{x}_i-\mu)^2}{\sigma^2} \sim \chi^2_n$$

according to case 3 of Section 4.4.6. Since

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

we have

$$\frac{(\tilde{\bar{x}} - \mu)\sqrt{n}}{\sigma} \sim n(0,1)$$

and

$$\frac{(\tilde{\bar{x}} - \mu)^2 n}{\sigma^2} \sim \chi^2_1$$

Since $\tilde{x}$ and $\tilde{s}$ are stochastically independent, the variables

$$\frac{n(\tilde{\bar{x}}-\mu)^2}{\sigma^2} \quad \text{and} \quad \frac{(n-1)s^2}{\sigma^2}$$

are also stochastically independent. The mgf

(4.70) is

$$E\left(e^{\frac{t\Sigma(\tilde{x}_i-\mu)^2}{\sigma^2}}\right) = E\left(e^{\frac{tn(\bar{\tilde{x}}-\mu)^2}{\sigma^2} + \frac{t(n-1)\tilde{s}^2}{\sigma^2}}\right)$$

$$= E\left(e^{\frac{tn(\bar{\tilde{x}}-\mu)^2}{\sigma^2}}\right) E\left(e^{\frac{t(n-1)\tilde{s}^2}{\sigma^2}}\right)$$

Since $E\left(e^{\frac{t\Sigma(\tilde{x}_i-\mu)^2}{\sigma^2}}\right) = (1-2t)^{-n/2}$

and $E\left(e^{\frac{tn(\bar{\tilde{x}}-\mu)^2}{\sigma^2}}\right) = (1-2t)^{-1/2}$

the mgf

$$E\left(e^{\frac{t(n-1)\tilde{s}^2}{\sigma^2}}\right) = (1-2t)^{-(n-1)/2}$$

which is the mgf of χ^2_{n-1}. Thus,

$$\boxed{\frac{(n-1)\tilde{s}^2}{\sigma^2} \sim \chi^2_{n-1}} \qquad (4.71)$$

Note that s^2 is an unbiased estimate of σ^2 and not the variance of the mean which is σ^2/n.

The derivation of the distribution of the sample mean and variance is typically treated in a first course in statistics. The derivation was repeated here in order to gain expertise in handling various statistical concepts in preparation for Section 5, where multivariate concepts will be discussed. There we will treat the case of several random variables having a sample size of "one" for each variable.

4.5 Hypothesis Tests and Confidence Intervals

4.5.1 Hypothesis Tests

A test of a statistical hypothesis is a rule which, when the experimental sample values have been obtained, leads to a decision to accept or to reject the hypothesis under consideration. A hypothesis is simply a statement about the population (parameters). If the statistical hypothesis completely specifies the distribution, it is called a "simple statistical hypothesis"; if it specifies less than necessary to describe the distribution it is called a "composite statistical hypothesis". Examples:

a) H: The population is normal and has mean μ and variance σ.

b) H: The population is normal with mean μ.

c) H: The population is normal.

d) H: The population is normal with a mean larger than a specified constant.

Case a) is a simple statistical hypothesis and cases b) to d) are composite statistical hypotheses. Each hypothesis has an alternative hypothesis, e.g.,

Null hypothesis H_0: population parameter $\Theta = \Theta_0$

Alternative Hypothesis H_1: $\Theta = \Theta_1$

For testing a hypothesis we first have to draw a sample of the population and compute the values for one or more sample statistics. on the basis of the sample values and the specifications of the hypothesis. If the sample statistics falls within a "critical region", the null hypothesis under consideration is rejected.

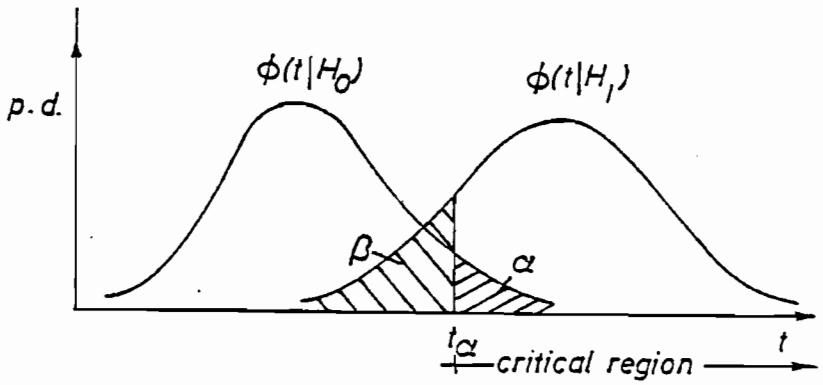


Figure 4.6 Example for Critical Region of a Statistical Test.

Figure 4.6 displays the probability density functions for the case when the null hypothesis H_0 is true and for the case when the alternative hypothesis H_1 is true. As a rule for testing the hypothesis we reject the null hypothesis and accept the alternative hypothesis if the sample statistics exceeds the value t_α, i.e.,

$$t_{H_0} > t_\alpha \quad \text{or} \quad t_{H_1} > t_\alpha \tag{4.72}$$

Thus, for $\phi(t|H_0)$ we reject H_0 and accept H_1 if $t_{H_0} > t_\alpha$, and, for $\phi(t|H_1)$ is true, we reject H_0 and accept H_1 if $t_{H_1} > t_\alpha$

The power function of a test of a statistical hypothesis H_0 against an alternative hypothesis H_1 is that function, defined for all distributions under consideration, which yields the probability that the sample falls in the critical region, i.e., a function which yields the probability of rejecting the null hypothesis. The value of the power function of a certain parameter point is called the power of the test at that point.

The area under the probability density function $\phi(t|H_0)$ to the right of t_α is α. Thus,

$P(H_0$ is true, but the test has rejected it)

$$= \int_{t_\alpha}^{\infty} \phi(t|H_0)dt = \alpha \qquad (4.73)$$

α is the power of the test with respect to the null hypothesis. This value is usually called the "significance level". The error committed in (4.73), (i.e., the null hypothesis was rejected although it is true), is called a Type I error. We can say that the significance level of a test is equal to the probability of a Type I error.

The area under the probability density function $\phi(t|H_1)$ to the right of t_α is $1-\beta$. Thus, $P(H_1$ is true, test has rejected $H_0)$

$$= \int_{t_\alpha}^{\infty} \phi(t|H_1)dt = 1-\beta \qquad (4.74)$$

$1-\beta$ is equal to the power of the test under the alternative hypothesis. The area to the left of t_α under the pdf $\phi(t|H_1)$ is equal to the probability that H_0 is accepted although H_1 is true:

$P(H_1$ is true, test has accepted $H_0) = \beta$.
The latter statement is equivalent to:

$$P(H_0 \text{ is false, but test has accepted it}) = \beta \qquad (4.75)$$

This is, of course, an erroneous conclusion. The acceptance of the null hypothesis when it is actually false is called a Type II error.

Usually we would like to minimize the probabilities of the Type I and Type II errors. However, during this course we will not compute the probabilities of the Type II error, since such a computation requires the knowledge of the distribution under the alternative hypothesis, which is, in general, a non-central distribution, i.e., a type of distribution we have not discussed. In our case the null hypothesis will be a simple hypothesis, and the alternative hypothesis will be a composite. By not computing the distribution under

the alternative hypothesis, we do not have to deal with the additional complications resulting from composite hypotheses. In actual testing it will be at our disposal to choose the significance level. We should bear in mind that, as we select a decreasing significant level, i.e., the probability of committing Type I error, the probability of the Type II error is increasing. Since we do not compute the latter probability in this course we have no way of minimizing the probability of both types of errors. As a compromising working rule we simply adopt a significance level of 0.05 most of the time.

The following table gives a summary of possible decisions with respect to the null-hypothesis:

Hypothesis H_0	Test Result	Remark
True	Accept H_0	O.K.
True	Reject H_0	Type I Error
False	Reject H_0	O.K.
False	Accept H_0	Type II Error

The rule (4.72) for testing the hypothesis is referred to as a "one tail test" in the upper end of the distribution. Frequently, a "two tail test" is employed. In such cases the null hypothesis under consideration is rejected if

$$|t| > t_{\alpha},$$

i.e.

$$P(|t| > t_{\alpha}) = \alpha \qquad (4.76)$$

Note that no subscript has been attached to t. Since we do not compute t under the alternative hypothesis in this course we will be concerned only with t_{H_0}. As a matter of simplification the subscript H_0 can be neglected. If we reject the hypothesis H_0 in case (4.76) holds, we subject ourselves to the possibility of rejecting the hypothesis H_0 when it is true $100\alpha\%$ of the cases (Type I error). The rule (4.76) applies to the case when the probability distribution under the null-hypothesis is symmetric.

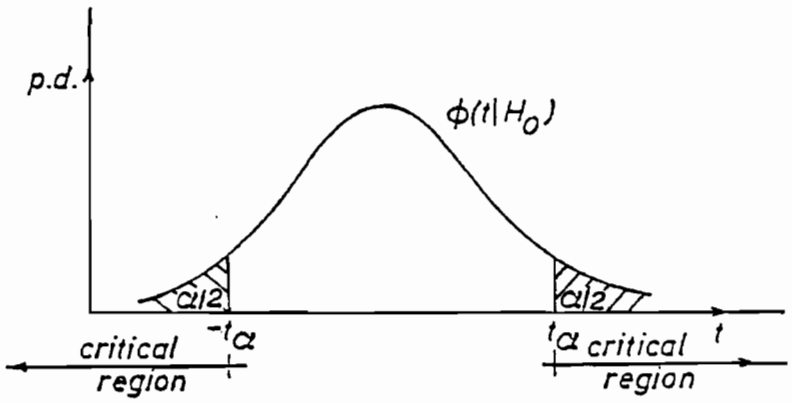


Figure 4.7 Two Tail Test (Symmetric Distribution)

The critical region is split into two parts. They are located at the tails of the distribution, each covering an area of $\alpha/2$ under the probability curve. In the case that the distribution under the null-hypothesis is not symmetric the rule (4.76) is replaced by

$$t > t_{\alpha/2}$$
$$t < t_{1-\alpha/2} \qquad (4.77)$$

i.e., we would like to have the probability $\alpha/2$ at each end of the tail.

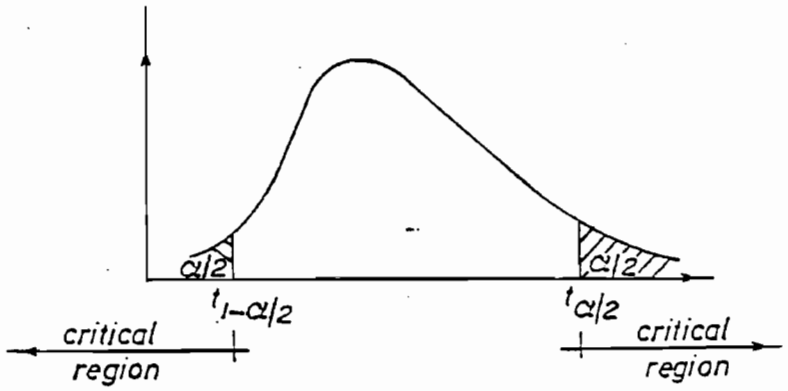


Figure 4.8 Two Tail Test (General Distribution)

Occasionally we would like to perform a statistical test at the lower end of the distribution. The rule for rejecting the null hypothesis under consideration is

$$t < t_{1-\alpha} \tag{4.78}$$

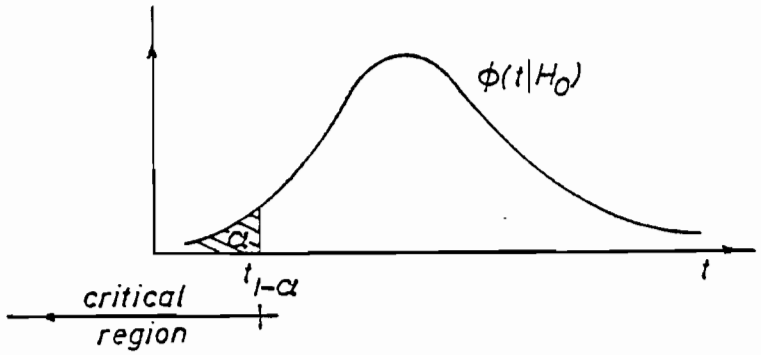


Figure 4.9 One Tail Test

Note, that the symbol t_α defines the value of t to the right of which is the area α under the probability density curve.

4.5.2 Confidence Interval:

Let Θ be an unknown parameter of a population and t_1 and t_2 two statistics (functions of the sample only) and

$$P(t_1<\Theta<t_2) = 1-\alpha \tag{4.79}$$

then $\{t_1, t_2\}$ is a <u>100(1-α)% confidence interval</u> for the population parameter Θ.

Consider the case of the random variable $\bar{x}$ having a normal distribution

$n(\mu,\sigma^2)$. Assume that the variance σ^2 is known. We would like to find a confidence interval for the mean μ. Since $\tilde{x} \sim n(\mu,\sigma^2)$ the random variable $\tilde{w} = (\tilde{x}-\mu)/\sigma \sim n(0,1)$ has a standardized normal distribution. Thus,

$$P(-a<\tilde{x}<b) = 1-\alpha$$

or

$$P(-w_1<\tilde{w}<w_2) = 1-\alpha$$

$$P\left(-w_1 < \frac{\tilde{x}-\mu}{\sigma} < w_2\right) = 1-\alpha$$

$$P(-w_1\sigma - \tilde{x} < -\mu < -\tilde{x} + w_2\sigma) = 1-\alpha$$

$$P(\tilde{x}-w_2\sigma<\mu<\tilde{x} + w_1\sigma) = 1-\alpha \qquad (4.80)$$

The values w_1 and w_2 can be found from the tables of the normal distribution, given the probability $1-\alpha$. Since σ is assumed to be known the interval

$$\{\tilde{x}-w_2\sigma,\ \tilde{x}+w_1\sigma\} \qquad (4.81)$$

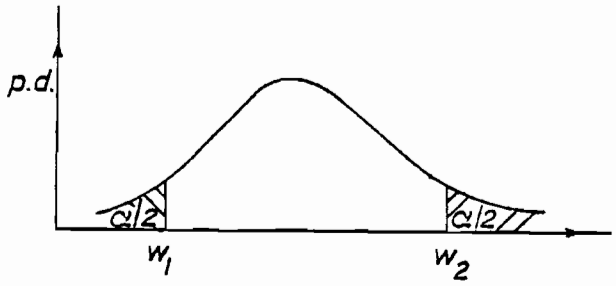
is a function of the random variable $\tilde{x}$. It is, therefore, called a random interval. The probability statement (4.80) can be read as follows: Prior to the performance of the random experiment, the probability is $1-\alpha$ that the random interval (4.81) includes the unknown fixed parameter μ. Suppose the experiment yields $\tilde{x} = x$. Then the interval $\{x-w_2\sigma,\ x+w_1\sigma\}$ has known end points. Obviously we cannot say that $1-\alpha$ is the probability that the particular interval $\{x-w_2\sigma,\ x+w_1\sigma\}$ includes the parameter μ, since μ, although unknown, is some constant, and this particular interval either does or does not include μ. However, the fact that we had a $1-\alpha$ (α is small) probability, prior to the performance of the experiment, that the random interval $\{\tilde{x} - w_2\sigma,\ x+w_1\sigma\}$ includes the fixed mean μ, leads us to have some reliance on the particular interval $(x-w_2\sigma,\ x+w_1\sigma)$. The reliance is reflected by calling the known interval $(x-w_2\sigma,\ x+w_1\sigma)$ a $100(1-\alpha)\%$ confidence interval. Thus, replacing $\tilde{x}$ by the measured value x in expression (4.80) gives for the two statistics

$$t_1 = x - w_2\sigma$$

$$t_2 = x + w_1\sigma$$

Generally, the limits w_1 and w_2 are selected such that the probability to the left of w_1 is equal to the probability to the right of w_2, i.e.,

$$P(\tilde{w}<w_1) = P(\tilde{w}>w_2) = \frac{1}{2}\alpha \tag{4.82}$$



If the distribution under consideration is symmetric, then

$$w_2 = -w_1$$

4.5.3 Standard Cases

4.5.3.1 Case 1: Population Mean (Variance Known)

Given:

$$\left.\begin{aligned} \tilde{x}_1 &\sim n(\mu,\sigma^2) \\ \tilde{x}_2 &\sim n(\mu,\sigma^2) \\ &\vdots \\ \tilde{x}_n &\sim n(\mu,\sigma^2) \end{aligned}\right\} \tag{4.83}$$

The distribution of the sample mean $\bar{\tilde{x}}$ is

$$\bar{\tilde{x}} \sim n(\mu,\sigma^2/n)$$

Hypothesis: $H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

$$\tilde{w} = \frac{(\tilde{\bar{x}}-\mu)\sqrt{n}}{\sigma} \sim n(0,1)$$

Replace the random variable of the statistics by its sample value, substitute the population parameters as specified under H_0, and compute

$$w = \frac{(\bar{x}-\mu_0)\sqrt{n}}{\sigma}$$

Reject H_0 at a $100\alpha\%$ significance level if $|w| > w_{\alpha/2}$

where $\int_{-\infty}^{w_{\alpha/2}} n(0,1)dw = 1 - \alpha/2$

Confidence interval: Since $n(0,1)$ is symmetric we have

$$P(-w_{\alpha/2} < \frac{(\tilde{\bar{x}}-\mu)\sqrt{n}}{\sigma} < w_{\alpha/2}) = 1-\alpha$$

$$P(\frac{-\sigma w_{\alpha/2}}{\sqrt{n}} < \tilde{\bar{x}} - \mu < \frac{w_{\alpha/2}\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(-\tilde{\bar{x}} - \frac{\sigma w_{\alpha/2}}{\sqrt{n}} < -\mu < -\tilde{\bar{x}} + \frac{w_{\alpha/2}\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(\tilde{\bar{x}} - \frac{w_{\alpha/2}\sigma}{\sqrt{n}} < \mu < \tilde{\bar{x}} + \frac{\sigma w_{\alpha/2}}{\sqrt{n}}) = 1-\alpha$$

Replacing the random variable $\tilde{\bar{x}}$ by the sample value $\bar{x}$ gives a $100(1-\alpha)\%$ confidence interval for the population mean

$$[\bar{x} - \frac{w_{\alpha/2}\sigma}{\sqrt{n}}, \quad \bar{x} + \frac{w_{\alpha/2}\sigma}{\sqrt{n}}]$$

4.5.3.2 Case 2: Population Mean (Variance Unknown)

Given: Sequence (4.83); μ and σ unknown

Then $\tilde{\bar{x}} \sim n(\mu, \sigma^2/n)$

$$\tilde{v} = \frac{(n-1)\tilde{s}^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\tilde{w} = \frac{(\tilde{\bar{x}}-\mu)\sqrt{n}}{\sigma} \sim n(0,1)$$

$$\tilde{t} = \frac{\tilde{w}}{\sqrt{\tilde{v}/r}} = \frac{(\tilde{\bar{x}}-\mu)\sqrt{n}}{\tilde{s}} \sim t_{n-1}$$

$\tilde{s}$ is the sample variance

Hypothesis: H_o: $\mu = \mu_o$

H_1: $\mu \neq \mu_o$

$$\tilde{t} = \frac{(\tilde{\bar{x}}-\mu)}{\tilde{s}}\sqrt{n} \sim t_{n-1}$$

Replace the random variables of the statistics by the sample values and substitute the population parameters of H_o:

$$t = \frac{(\bar{x}-\mu_o)}{s}\sqrt{n}$$

Reject H_o at 100α% significance level if $|t| > t_{n-1,\alpha/2}$

where $\int_{-\infty}^{t_{n-1,\alpha/2}} t_{n-1}\,dt = 1-\alpha/2$

Confidence interval: Since t_{n-1} is symmetric we have

$$P(-t_{n-1,\alpha/2} < \frac{(\tilde{\bar{x}}-\mu)}{\tilde{s}}\sqrt{n} < t_{n-1,\alpha/2}) = 1-\alpha$$

or $$P(\tilde{\bar{x}} - \frac{t_{n-1,\alpha/2}\tilde{s}}{\sqrt{n}} < \mu < \tilde{\bar{x}} + \frac{t_{n-1,\alpha/2}\tilde{s}}{\sqrt{n}}) = 1-\alpha$$

Replacing the random variables $\tilde{\bar{x}}$ and $\tilde{s}$ by the sample values $\bar{x}$ and s gives a $100(1-\alpha)\%$ confidence interval for the population mean.

$$\left[\bar{x} - \frac{t_{n-1,\alpha/2}\, s}{\sqrt{n}} , \quad \bar{x} + \frac{t_{n-1,\alpha/2}\, s}{\sqrt{n}} \right]$$

4.5.3.3 Case 3: Population Variance

Given: Sequence (4.83)

Then: $\frac{(n-1)}{\sigma^2}\, \tilde{s}^2 \sim \chi^2_{n-1}$

Hypothesis: $H_0: \sigma^2 = \sigma_0^2$

$H_1: \sigma^2 \neq \sigma_0^2$

Replace the random variable by the sample values and substitute the population parameters of H_0:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Reject H_0 at $100\alpha\%$ significance level if

$$\chi^2 < \chi^2_{n-1,1-\alpha/2}$$

$$\chi^2 > \chi^2_{n-1,\alpha/2}$$

where

$$\int_0^{\chi^2_{n-1,1-\alpha/2}} \chi^2_{n-1}\, dx = \alpha/2$$

$$\int_0^{\chi^2_{n-1,\alpha/2}} \chi^2_{n-1}\, dx = 1-\alpha/2$$

Confidence interval: Since the χ^2_{n-1} distribution is not symmetric we have:

$$P(\chi^2_{n-1,1-\alpha/2} < \frac{(n-1)\,\tilde{s}^2}{\sigma^2} < \chi^2_{n-1,\alpha/2}) = 1-\alpha$$

Note, that we have followed the convention to select equal probability at the tails of the χ^2 distribution.

$$P\left(\frac{\chi^2_{n-1,1-\alpha/2}}{(n-1)\tilde{s}^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{n-1,\alpha/2}}{(n-1)\tilde{s}^2}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)\tilde{s}^2}{\chi^2_{n-1,1-\alpha/2}} > \sigma^2 > \frac{(n-1)\,\tilde{s}^2}{\chi^2_{n-1,\alpha/2}}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)\tilde{s}^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 > \frac{(n-1)\tilde{s}^2}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha$$

Replacing the random variable $\tilde{s}$ by the sample value s gives a $100(1-\alpha)\%$ confidence interval for the population variance.

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \quad , \quad \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right]$$

4.5.3.4 Case 4: Means of Two Populations

Case 4.1: Given: Two sequences of stochastically independent random variables, each having a normal distribution which has the same but unknown variance.

$$\left.\begin{array}{ll} \tilde{x}_1 \sim n(\mu,\sigma^2) & \tilde{y}_1 \sim n(\mu_2,\sigma^2) \\ \tilde{x}_2 \sim n(\mu,\sigma^2) & \tilde{y}_2 \sim n(\mu_2,\sigma^2) \\ \vdots & \vdots \\ \tilde{x}_n \sim n(\mu,\sigma^2) & \tilde{y}_m \sim n(\mu_2,\sigma^2) \end{array}\right\} \qquad (4.84)$$

We have

$$\bar{\tilde{x}} \sim n\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\bar{\tilde{y}} \sim n\left(\mu, \frac{\sigma^2}{m}\right)$$

$$\frac{(n-1)\tilde{s}_x}{\sigma^2} \sim \chi^2_{n-1}$$

$$\frac{(m-1)\tilde{s}_y}{\sigma^2} \sim \chi^2_{m-1}$$

All four variables ($\bar{\tilde{x}}, \bar{\tilde{y}}, \tilde{s}_x$ and $\tilde{s}_y$) are stochastically independent since the mean and the variance of each group, and the groups are stochastically independent. Thus,

$$\bar{\tilde{x}} - \bar{\tilde{y}} \sim n\left(\mu_1-\mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right)$$

and

$$\frac{(\bar{\tilde{x}}-\bar{\tilde{y}}) - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim n(0,1)$$

Also,

$$\frac{(n-1)\tilde{s}_x^2}{\sigma^2} + \frac{(m-1)\tilde{s}_y^2}{\sigma^2} \sim \chi^2_{n+m-2}$$

It follows that

$$\frac{\tilde{\bar{x}} - \tilde{\bar{y}} - (\mu_1-\mu_2)}{\tilde{R}} = \frac{\tilde{\bar{x}} - \tilde{\bar{y}} - (\mu_1-\mu_2)}{\sqrt{\frac{(n-1)\tilde{s}_x^2+(m-1)\tilde{s}_y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim t_{n+m-2}$$

Hypothesis:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Replace the random variables by the sample values

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)}}$$

Reject H_0 at a 100 $(1-\alpha)$% significance level if

$$|t| > t_{n+m-2,\alpha}$$

Confidence interval: Since the t_{n+m-2} distribution is symmetric we have

$$P\left(-t_{n+m+2,\frac{\alpha}{2}} < \frac{\tilde{\bar{x}}-\tilde{\bar{y}}-(\mu_1-\mu_2)}{\tilde{R}} < t_{n+m-2,\frac{\alpha}{2}}\right) = 1-\alpha$$

or

$$P(\tilde{\bar{x}}-\tilde{\bar{y}}-t_{n+m+2,\alpha/2}\ \tilde{R} < \mu_1-\mu_2 < \tilde{\bar{x}}-\tilde{\bar{y}}+t_{n+m+2,\alpha/2}\ \tilde{R}) = 1-\alpha$$

Replacing the random variables $\tilde{\bar{x}}$, and $\tilde{\bar{y}}$, and $\tilde{R}$ by the sample values $\bar{x}$, $\bar{y}$ and R gives a $100(1-\alpha)\%$ confidence interval for the difference of the population mean:

$$[\bar{x}-\bar{y}-t_{n+m-2}R, \quad \bar{x}-\bar{y}+t_{n+m-2}R]$$

Case 4.2: Usually it will not be known that the variances of the two populations are equal, and yet we would like to test the hypothesis concerning the means, having available only the sample variances. This is the so-called Fisher-Behrens problem; see Hamilton, p.92. It can be shown that

$$t = \frac{(\tilde{\bar{x}}-\tilde{\bar{y}}) - (\mu_1-\mu_2)}{\left[\frac{s_x^2}{n} + \frac{s_y^2}{m}\right]^2}$$

has an approximate t distribution with

$$\frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2}{\frac{s_x^4}{n(n+1)} + \frac{s_y^4}{m(m+1)}} - 2$$

degrees of freedom. Testing the statistical hypothesis and finding the confidence interval is done as explained in case 4.1.

4.5.3.5 Case 5: Variance of Two Populations

Given: Sequence (4.84)

We have

$$\tilde{\bar{x}} \sim n(\mu_1, \sigma_1^2/n)$$

$$\tilde{\bar{y}} \sim n(\mu_2, \sigma_2^2/m)$$

$$\frac{(n-1)\tilde{s}_x^2}{\sigma_1^2} \sim \chi^2_{n-1}$$

$$\frac{(m-1)\tilde{s}_y^2}{\sigma_2^2} \sim \chi^2_{m-1}$$

All 4 variables $\tilde{\bar{x}}$, $\tilde{\bar{y}}$, $\tilde{s}_x$ and $\tilde{s}_y$ are stochastically independent.

$$\tilde{F} = \frac{(n-1)\tilde{s}_x^2 \,/\, [\sigma_1^2(n-1)]}{(m-1)\tilde{s}_y^2 \,/\, [\sigma_2^2(m-1)]} = \frac{\tilde{s}_x^2\,\sigma_2^2}{\tilde{s}_y^2\,\sigma_1^2} \sim F_{n-1,m-1}$$

Hypothesis:

$$H_0: \ \sigma_1^2 = \sigma_2^2$$

$$H_1: \ \sigma_1^2 \neq \sigma_2^2$$

Compute statistics under H_0: $F = \frac{s_x^2}{s_y^2}$

Reject H_0 at a $100(\alpha)\%$ significance level if

$$F < F_{n-1,m-1,1-\frac{\alpha}{2}}$$

$$F > F_{n-1,m-1,\frac{\alpha}{2}}$$

It is important to identify the degrees of freedom in the numerator and denominator since

$$F_{r_1,r_1,\alpha} \neq F_{r_2,r_1,\alpha}$$

The probability in the tails is again equal, i.e. $\frac{\alpha}{2}$

Confidence interval: Since the F distribution is not symmetric we have:

$$P(F_{n-1,m-1,1-\frac{\alpha}{2}} < \frac{\tilde{s}_x^2 \sigma_2^2}{\tilde{s}_y^2 \sigma_1^2} < F_{n-1,m-1,\frac{\alpha}{2}}) = 1-\alpha$$

$$P(\frac{\tilde{s}_y^2}{\tilde{s}_x^2} F_{n-1,m-1,1-\frac{\alpha}{2}} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{\tilde{s}_y^2}{\tilde{s}_x^2} F_{n-1,m-1,\frac{\alpha}{2}}) = 1-\alpha$$

Replacing the random variables $\tilde{s}_y$ and $\tilde{s}_x$ by the sample variances s_y and s_x gives a $100(1-\alpha)\%$ confidence interval for the ratio of the population variances:

$$[\frac{s_y^2}{s_x^2} F_{n-1,m-1,1-\frac{\alpha}{2}} , \frac{s_y^2}{s_x^2} F_{n-1,m-1,\frac{\alpha}{2}}]$$

4.5.3.6 Case 6: A Goodness-of-Fit Test

Assume that we have a series of observations and we wish to test whether the observations come from a certain population with a specified distribution.

We subdivide the observation series into n classes. n_i is the number of observations in class i. The subdivision should be such that $n_i \geq 5$. Compute for each class the expected number d_i of observations based on the hypothetical distribution. It can be shown that,

$$\chi^2 = \sum_{i=1}^{n} \frac{(n_i - d_i)^2}{d_i}$$

is distributed approximately as χ^2_{n-1}. If the population parameters are estimated from the sample then the degree of freedom should be reduced for each parameter estimated.

Hypothesis:

H_0: sample is from specified population

H_1: sample is not from specified population

Reject H_0 at a $100\alpha\%$ significance level if

$$\chi^2 > \chi^2_{DF,\alpha}$$

where DF is the degree of freedom.

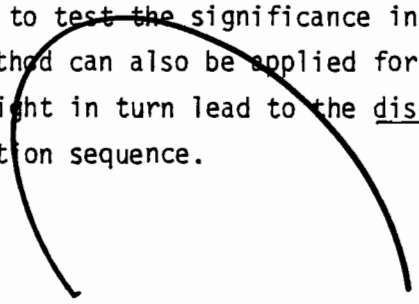
The test is more accurate the more observations are available.

4.5.3.7 Sign Tests

In deriving the statistic χ in the previous section use was made of the hypothetical distribution H_0. A test of a hypothesis H_0 which is based upon a statistics whose distribution, under H_0, does not depend upon the specific distribution nor any parameter of that distribution is called a distribution-free or a non-parametric test.

One application of such tests is to analyze the signs of an observation sequence. Assume that you have an observation sequence x_i and form the differences $x_i - \bar{x}$ where $\bar{x}$ is the sample mean. Define a new random variable to be +1 if the difference is positive and -1 if the difference is negative. If the observations are drawn from a normal population, then the sum is expected to be zero. It is possible to derive a distribution for the

sum and to test the significance in case the sum deviates from zero. This method can also be applied for investigating the sequence of signs, which might in turn lead to the discovery of a systematic effect in the observation sequence.



5. Statistical Inference in Least Squares Adjustments

5.1 Multivariate Normal Distribution

In Section 4 we dealt briefly with random variables from two different normal distributions. At this time we would like to introduce the multivariate normal distribution. This is a multi-dimensional distribution where each variable has a normal distribution and where the variables might or might not be stochastically independent. Let X be a random

$$\tilde{X} = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{pmatrix} \tag{5.1}$$

vector with n random components $\tilde{x}_i$. Let the expected value, or the mean, of $\tilde{x}$ be

$$E(\tilde{X}) = \begin{pmatrix} E(\tilde{x}_1) \\ E(\tilde{x}_2) \\ \vdots \\ E(\tilde{x}_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = {}_n\mu_1 \tag{5.1}$$

and the variance-covariance matrix be

$$E[(\tilde{X}-\mu)\ (\tilde{X}-\mu)^T] = {}_n\Sigma_n \tag{5.2}$$

Then the density of the multidimensional normal distribution is

$$f(x_1,x_2\ldots x_n) = \frac{1}{(2\pi)^{n/2}\ |\Sigma|^{\frac{1}{2}}}\ e^{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)} \tag{5.3}$$

The similarity of the multivariate normal density (5.3) and the univariate normal density (4.28) is clear. We observe that $f(x_1,x_2 \ldots x_n)$ is a positive function. Since Σ^{-1} is positive definite,

$$(X-\mu)^T \Sigma^{-1} (X-\mu) \geq 0$$

and the density is bounded, that is,

$$f(x_1,x_2 \ldots x_n) \leq \frac{1}{(2\pi)^{n/2} |\Sigma|^{\frac{1}{2}}}$$

It can also be verified that

$$\int_{-\infty}^{\infty} \ldots \int f(x_1,x_2 \ldots x_n) dx_1 \ldots dx_n = 1$$

If a random vector $\tilde{X}$ has a distribution (5.3) we say that $\tilde{X}$ is multivariate normal with mean μ and variance Σ. We use the following notation

$$\boxed{{}_n\tilde{X}_1 \sim N_n({}_n\mu_1, {}_n\Sigma_n)} \qquad (5.4)$$

Note that μ is an n dimensional vector. The mean and the variance-covariance matrix completely describe the multivariate normal distribution.

In the following we state some useful theorems on multivariate normal distributions. Proofs, many of which are not given here, can be found in the standard literature.

Theorem: If $\tilde{X}$ is multivariate normal

$$\tilde{X} \sim N_n(\mu,\Sigma)$$

and $\tilde{Z} = {}_mD_n\tilde{X}$ is a linear function of the random variable where D is a m x n matrix of rank $m \leq n$ then

$$\tilde{Z} \sim N_m (D\mu, D\Sigma D^T) \qquad (5.5)$$

Thus, the mean and the variance of the new distribution are

$$E(\tilde{Z}) = E(D\tilde{X}) = DE(\tilde{X}) = D\mu$$

$$E[\tilde{Z}-DE(\tilde{X})]\,[\tilde{Z}-DE(\tilde{X})]^T = E[D\tilde{X}-DE(\tilde{X})]\,[D\tilde{X}-DE(\tilde{X})]^T$$

$$= DE[\tilde{X}-E(\tilde{X})]\,[\tilde{X}-E(\tilde{X})]^T D^T$$

$$= D\Sigma D^T$$

The same result follows from the law of variance propagation (2.30). Since $R(D) = m$ the variance covariance matrix $D\Sigma D^T$ is non-singular. The same law of transformation holds if Σ is singular or the rank of D is less than m. However, we do not pursue this aspect any further, since it leads to "singular" or "degenerate" normal distributions.

Theorem: If $\tilde{X}$ is multivariate normal

$$\tilde{X} \sim N(\mu,\Sigma)$$

the marginal distribution of any set of components of $\tilde{X}$ is multivariate normal with means, variances, and covariances obtained by taking the proper component of μ and Σ respectively. For example:

$$\tilde{X} = \begin{pmatrix}\tilde{X}_1\\ \tilde{X}_2\end{pmatrix} \sim N\left[\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right] \tag{5.6}$$

then the marginal distribution of $\tilde{X}_2$ is

$$\tilde{X}_2 \sim N(\mu_2, \Sigma_{22}) \tag{5.7}$$

The same law holds, of course, if the set contains only one component, say x_i. The marginal distribution of $\tilde{x}_i$ is then

$$\tilde{x}_i \sim n(\mu_i, \sigma_i^2) \tag{5.8}$$

Theorem: If $\tilde{X}$ is multivariate normal, a necessary and sufficient condition that a subset of the random variables and the subset of the remaining variables be stochastically independent, is that each covariance of a variable from one set and a variable from the other set be zero. For example,

$$\begin{pmatrix}\tilde{X}_1\\ \tilde{X}_2\end{pmatrix} \sim N\left[\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22}\end{pmatrix}\right] \tag{5.9}$$

then $\tilde{X}_1$ and $\tilde{X}_2$ are stochastically independent. Since $\sigma_{ij} = \sigma_i\sigma_j\rho_{ij}$, and $\sigma_i \neq 0$, the condition $\sigma_{ij} = 0$ is equivalent to $\rho_{ij} = 0$. Thus, if one set of normally distributed random variables is uncorrelated with the remaining variables, the two sets are independent. The proof of the above theorem follows from the fact that the multidimensional density function of the normal distribution can be written in case (5.9) as

$$f(X_1,X_2) = f_1(X_1)\, f_2(X_2)$$

or, in the case that all covariances are zero,

$$f(x_1x_2\ldots x_n) = f_1(x_1)f_2(x_2)\ldots f_n(x_n),$$

as can be verified easily.

5.2 Distribution of some Quadratic Forms

Statistical tests in least squares adjustments require the distribution of some quadratic forms. One of them is V^TPV. In order to emphasize the statistical nature of this derivation the symbol "~" will consistently be used to denote random variables. The adjustment model of observation equations is written as

$$E(\tilde{L}) = AX$$

and

$$E[\tilde{L}-E(\tilde{L})][\tilde{L}-E(\tilde{L})]^T = \Sigma_{L_b} = \sigma_o^2 P^{-1} \tag{5.10}$$

Note that X itself is a fixed parameter although the estimate $\hat{X}$ is a random variable and will be denoted by $\tilde{\hat{X}}$. We now assume that the observations have a multivariate normal distribution,

$$\tilde{L} \sim N_n (-AX, \Sigma_{L_b}) \tag{5.11}$$

The model (5.10) can also be written as

$$\tilde{V} = AX + \tilde{L}$$

with

$$E(\tilde{V}) = 0 \text{ and } E(\tilde{V}\tilde{V}^T) = \Sigma_{L_b} = \sigma_o^2 P^{-1} \tag{5.11}$$

Thus,

$$\tilde{V} \sim N_n(0,\Sigma_{L_b}) = N_n(0,\sigma_o^2 P^{-1}) \tag{5.12}$$

If the variance covariance matrix of the observations is diagonal, then the residuals are normally distributed and stochastically independent, i.e.,

$$\tilde{V}_i \sim n(0,\sigma_i^2) = n(0,\sigma_o^2 \frac{1}{p_i}) \tag{5.12}$$

In the case that Σ_{L_b} is non-diagonal we can always transform the adjustment model such that the transformed residuals are stochastically independent and have a unit variate normal distribution. In Section 1.4 it was shown that for a positive definite matrix, say P^{-1}, there exists a non-singular matrix D such that

$$D^T P^{-1} D = I$$

with $D = E\Lambda^{-\frac{1}{2}}$. The columns of the orthogonal matrix E consist of the eigenvectors of P^{-1}, and Λ is a diagonal matrix having the eigenvalues of P^{-1} at the diagonal. We perform the transformation,

$$\tilde{L} = D^T\tilde{L}$$

giving

$$D^T\tilde{V} = D^TAX + D^T\tilde{L}$$

or

$$n\bar{V}_1 = n\bar{A}_u X + n\bar{L}_1 \tag{5.13}$$

whereby the notation used is self-explanatory. The transformed residuals are distributed as

$$\tilde{\bar{V}} \sim N_n(0,\sigma_o^2 I) \tag{5.14}$$

since

$$E(\tilde{\bar{V}}) = D^TE(\tilde{V}) = 0, \quad \Sigma_{\bar{V}} = D^TP^{-1}D\sigma_o^2 = \sigma_o^2 I$$

The quadratic form V^TPV remains invariant under this transformation:

$$\tilde{V}^TP\tilde{V} = \tilde{\bar{V}}^TD^TPD\tilde{\bar{V}} \tag{5.15}$$

$$= \tilde{\bar{V}}^T\tilde{\bar{V}}$$

A and $\bar{A}$ have the same rank since multiplication by a non-singular matrix does not change the rank.

The least squares principle requires the minimization of $\tilde{\bar{V}}^T\tilde{\bar{V}}$. Therefore, we have to find the distribution of

$$\tilde{R} = \tilde{\bar{V}}^T\tilde{\bar{V}}_{min} \tag{5.16}$$

This minimization is solved somewhat differently than in Section 3. We simply perform an appropriate orthogonal transformation. Let r be the rank of the design matrix $\bar{A}$, i.e. $r=R(\bar{A})$. Let the matrix F be an (n x r) matrix whose columns constitute an orthonormal basis for the column space of $\bar{A}$ (one such choice for the columns of F may be to take the normalized eigenvectors of $\bar{A}\bar{A}^T$). Let G be an (n x n-r) matrix such that (F:G) is orthogonal; that is, the columns of G span the orthogonal complement to the column space of $\bar{A}$ or, equivalently, G spans the null space of $\bar{A}^T$, then

$$\begin{pmatrix} F^T \\ G^T \end{pmatrix} (FG) = \begin{pmatrix} F^TF & F^TG \\ G^TF & G^TG \end{pmatrix} = \begin{pmatrix} {}_rI_r & 0 \\ 0 & {}_{n-r}I_{n-r} \end{pmatrix} \tag{5.17}$$

$$(FG)\begin{pmatrix} F^T \\ G^T \end{pmatrix} = FF^T + GG^T = I \tag{5.18}$$

and

$$\bar{A}^TG = 0, \quad G^T\bar{A} = 0 \tag{5.19}$$

Now we perform an orthogonal transformation:

$$\begin{pmatrix} F^T \\ G^T \end{pmatrix} \tilde{\tilde{V}} = \begin{pmatrix} F^T \\ G^T \end{pmatrix} \bar{A}X + \begin{pmatrix} F^T \\ G^T \end{pmatrix} \tilde{\tilde{L}}$$

or

$$\begin{pmatrix} F^T\tilde{\tilde{V}} \\ G^T\tilde{\tilde{V}} \end{pmatrix} = \begin{pmatrix} F^T\bar{A}X \\ 0 \end{pmatrix} + \begin{pmatrix} F^T\tilde{\tilde{L}} \\ G^T\tilde{\tilde{L}} \end{pmatrix} \tag{5.20}$$

Introducing new variables

$$\tilde{Z}_1 = F^T\tilde{\tilde{L}} \quad \text{and} \quad \tilde{Z}_2 = G^T\tilde{\tilde{L}}$$

equation (5.20) can be written as

$$\tilde{\tilde{V}}_Z = \begin{pmatrix} \tilde{\tilde{V}}_{Z_1} \\ \tilde{\tilde{V}}_{Z_2} \end{pmatrix} = \begin{pmatrix} F^T\bar{A}X \\ 0 \end{pmatrix} + \begin{pmatrix} \tilde{Z}_1 \\ \tilde{Z}_2 \end{pmatrix} \tag{5.21}$$

The quadratic form remains again invariant under the orthogonal transformation:

$$\begin{aligned} \tilde{\tilde{V}}_Z^T\tilde{\tilde{V}}_Z &= \tilde{\tilde{V}}^T(FF^T + GG^T)\tilde{\tilde{V}} \\ &= \tilde{\tilde{V}}^T\tilde{\tilde{V}} \end{aligned}$$

upon using equation (5.18). The actual quadratic form is obtained from equation (5.21):

$$\tilde{\tilde{V}}_Z^T\tilde{\tilde{V}}_Z = (F^T\bar{A}X + \tilde{Z}_1)^T(F^T\bar{A}X + \tilde{Z}_1) + \tilde{Z}_2^T\tilde{Z}_2 \tag{5.22}$$

This expression must be minimized through variation of the parameter X. If we choose X such that

$$-F^T\bar{A}\,\tilde{\hat{X}} = \tilde{Z}_1 \tag{5.23}$$

then

$$\tilde{R} \equiv \tilde{V}_Z{}^T\,\tilde{V}_{Z_{min}} \equiv \tilde{V}^T\,\tilde{V}_{min} \equiv \tilde{V}^T P\tilde{V}_{min}$$
$$= \tilde{Z}_2{}^T\,\tilde{Z}_2 \tag{5.24}$$

There always exists a solution for $\tilde{\hat{X}}$ in equation (5.23). If $\bar{A}$ has no rank defect then F^TA is a non-singular (u x u) matrix. In the case of a rank defect on $\bar{A}$, a solution can be found by imposing u-r conditions on the parameters (minimal constraints, inner constraints). It is interesting to realize that the random variable $\tilde{R}$ does not depend on the particular solution of $\tilde{\hat{X}}$ since it is only a function of $\tilde{Z}_2$. This proves the fact that $\tilde{R}$ is invariant with respect to the selection of the minimal constraint. Moreover, the residuals themselves are independent of the minimal constraint. Substituting the solution (5.23) into (5.20) gives

$$\begin{pmatrix} F^T \\ G^T \end{pmatrix} \tilde{\hat{V}} = \begin{pmatrix} 0 \\ G^T\tilde{L} \end{pmatrix}$$

Since (FG) is an orthogonal matrix the residuals become

$$\tilde{\hat{V}} = (FG)\begin{pmatrix} 0 \\ G^T\tilde{L} \end{pmatrix} = GG^T\tilde{L} \tag{5.25}$$

This expression is independent of the solution on $\tilde{\hat{X}}$ as far as the selection of the minimal constraint is concerned. The matrix G only has to span $\bar{A}^T$.

The distribution of the random variable $\tilde{Z}$ is

$$\tilde{Z} \sim N_n\left[\begin{pmatrix} -F^TAX \\ 0 \end{pmatrix}, \sigma_o{}^2\begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r}I_{n-r} \end{pmatrix}\right] \tag{5.26}$$

since $E(\tilde{Z}) = \begin{pmatrix} -F^TAX \\ 0 \end{pmatrix}$

and $$\Sigma_Z = \sigma_o^2 \begin{pmatrix} F^T \\ G^T \end{pmatrix} I \, (FG) = \sigma_o^2 \begin{pmatrix} F^TF & F^TG \\ G^TF & G^TG \end{pmatrix} = \sigma_o^2 \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

as is seen from equation (5.20). The random variables $\tilde{Z}_1$ and $\tilde{Z}_2$ are stochastically independent, giving

$$\tilde{Z}_2 \sim N_{n-r}(0, \sigma_o^2 I) \tag{5.27}$$

The components of $\tilde{Z}_2$ are also stochastically independent, i.e.

$$\tilde{z}_{2i} \sim n(0, \sigma_o^2)$$

It follows that

$$\frac{\tilde{z}_{2i}}{\sigma_o} \sim n(0,1)$$

and

$$\frac{\tilde{R}}{\sigma_o^2} = \frac{\tilde{Z}_2^T \tilde{Z}_2}{\sigma_o^2} = \sum_1^{n-r} \frac{z_{2i}^2}{\sigma_o^2} \sim \underbrace{\chi^2_{n-r}}_{DF} \tag{5.28}$$

Thus $\tilde{V}^T P \tilde{V}/\sigma_o^2$ is distributed as a χ^2 with a degree of freedom equal to the degree of freedom of the adjustment. Since

$$\hat{\sigma}_o^2 = \frac{\tilde{V}^T P \tilde{V}}{DF}$$

it follows that

$$\tilde{\chi}^2 = \frac{\tilde{V}^T P \tilde{V}}{\sigma_o^2} = \frac{\hat{\sigma}_o^2}{\sigma_o^2} \cdot DF \sim \chi^2_{DF} \tag{5.29}$$

Fundamental Test in Least Squares Adjustments: Based on the statistics (5.29) an important test can be performed to find out whether the adjustment is distorted. Given the variance-covariance matrix of the observations we have to decide at the beginning of the adjustment on the value of the a-priori variance of unit weight, σ_o^2, since

$$P = \sigma_o^2 \Sigma_{L_b}^{-1}$$

We can now test whether or not the a-posteriori variance of unit weight is in agreement with the a-priori value.

H_o: $\sigma_o = \sigma_T$ (≡value of adopted a-priori variance of unit weight). Adjustment is not distorted.

H_1: $\sigma_o \neq \sigma_T$ (≡adopted a-priori variance of unit weight). Adjustment is distorted.

Replace the random variable $\tilde{\sigma}_o^2$ by the sample value $\hat{\sigma}_o^2$ and compute

$$\chi^2 = \frac{\hat{\sigma}_o^2}{\sigma_T^2} DF$$

Reject H_o at a $100\alpha\%$ significance level if

$$\chi^2 < \chi^2_{DF,1-\alpha/2}$$

$$\chi^2 > \chi^2_{DF,\alpha/2}$$

The causes which lead to a distorted adjustment will be discussed later. The two tail test is frequently replaced by a one tail test at the upper end of the distribution. Note that the degree of freedom does not change when we add minimal constraints. The degree of freedom always is $n-R(\bar{A}) = n-r$. If there are u parameters it was stated earlier also that the degree of freedom is n-u. However, in case of rank deficiency we have to add $s = u-r$ (minimal) constraint in order to solve the system. This gives a degree of freedom of $n-u+s = n-u+(u-r) = n-r$.

The previous test can also be applied when testing additional observations in the sequential mode. In those cases we use the latest sample $\hat{\sigma}_o^2$, i.e. the one which includes the additional observations, for testing the hypothesis

$$H_o: \sigma_o = \sigma_T$$

The test statistic is

$$\chi^2 = \frac{\tilde{\hat{v}}^T P \tilde{\hat{v}} + \Delta\tilde{\hat{v}}^T P \tilde{\hat{v}}}{\sigma_o^2} = \frac{\tilde{\tilde{\sigma}}_{o(2)}^2}{\sigma_o^2} DF_2 \sim \chi^2_{DF_2} \qquad (5.30)$$

where DF_2 is the degree of freedom after adding the new observations. It is important that in this case we cannot perform a variance ratio test (Case 5 of section 4.5.3.5) since the numerator and denominator in

$$\frac{(V^T PV + \Delta V^T PV)/ DF_2}{V^T PV/ DF_1}$$

are not stachastically independent.

However, the statistics

$$F = \frac{\Delta\tilde{\hat{v}}^T P \tilde{\hat{v}} \quad DF_1}{\tilde{\hat{v}}^T P \tilde{\hat{v}} \quad (DF_2 - DF_1)} \sim F_{DF_2-DF_1,\ DF_1} \qquad (5.31)$$

has a F-distribution. $\Delta\tilde{\hat{v}}^T P \tilde{\hat{v}}/\sigma_o^2$ has a χ^2 distribution of a degree of freedom equal to the number of new observations. This follows from the previous derivation of the distribution of $\tilde{\hat{v}}^T P \tilde{\hat{v}}$. It is understood that the new observations are uncorrelated with the first group and that both groups refer to the same a-priori variance of unit weight σ_o^2. The stochastic independence of $\Delta\tilde{\hat{v}}^T P \tilde{\hat{v}}$ and $\tilde{\hat{v}}^T P \tilde{\hat{v}}$ can be shown as follows:

Let the solution of the first set

$$\tilde{V}_1 = A_1 X + \tilde{L}_1$$

be $\tilde{\hat{X}}^* = -N^{-1} A_1^T P_1 \tilde{L}_1$

with $Q_{X^*} = N^{-1} = (A_1^T P_1 A_1)^{-1}$

Then the final solution for X is obtained by solving the least squares problem

$$\tilde{V}_2 = A_2(\tilde{\hat{X}}^* + \Delta X) + \tilde{L}_2,$$

or $\tilde{V}_2 = A_2 \Delta X + \tilde{L}_2'$,

with $\tilde{L}_2' = A_2 \tilde{\hat{X}}^* + \tilde{L}_2$

and $P_{L_2'}^{-1} = Q_{L_2'} = P_{L_2}^{-1} + A_2 N^{-1} A_2^T$

according to the principles of least squares. The solution

$$\Delta \tilde{\hat{X}} = -(A_2^T P_{L_2'} A_2)^{-1} A_2^T P_{L_2'} \tilde{L}_2'$$

is in agreement with the result given in Section 3.4. Since $\Delta V^T P V \equiv \tilde{\hat{V}}_2^T P_{L_2} \tilde{\hat{V}}_2$ it is necessary to show the stochastic independence of $\tilde{\hat{V}}_2$ and $\tilde{\hat{V}}_1$. It is sufficient to show that the covariance between both sets is zero. Since $E(\tilde{\hat{V}}_2) = 0$ and $E(\tilde{\hat{V}}_1) = 0$ we have

$$E(\tilde{\hat{V}}_2 \tilde{\hat{V}}_1^T) = E\{(A_2 \Delta \tilde{\hat{X}} + \tilde{L}_2')\tilde{\hat{V}}_1^T\}$$

$$= E\{[-A_2 (A_2^T P_{L_2'} A_2)^{-1} A_2^T P_{L_2'} \tilde{L}_2 + \tilde{L}_2]\tilde{\hat{V}}_1^T + [-A_2 (A_2^T P_{L_2'} A_2)^{-1} A_2^T P_{L_2'} A_2 X^* + A_2 X^*]\tilde{\hat{V}}_1^T$$

Substituting $\tilde{\hat{V}}_1 = A_1 \tilde{\hat{X}}^* + \tilde{L}_1$ it is seen that the first term is zero, since $E(\tilde{L}_2 \tilde{L}_1^T) = 0$. Since the second term is zero, the stochastic independence has been established.

There is another quadratic form important in the analysis of least squares. The least squares solution vector for the parameter is

$$\tilde{\hat{X}} = -(A^T P A)^{-1} A^T P \tilde{L} \tag{5.32}$$

Note that the solution for the parameters is a random variable. Since

$$E(\tilde{\hat{X}}) = X$$

and

$$E(\tilde{\hat{X}} - X)(\tilde{\hat{X}} - X)^T = \sigma_o^2 (A^T P A)^{-1}$$

we have

$$\tilde{\hat{X}} \sim N_u(X, \sigma_o^2 {}_u N_u^{-1}) \tag{5.33}$$

We are interested in the distribution of

$$\tilde{Q} = \frac{1}{\sigma_o^2} (\tilde{\hat{X}} - X)^T N (\tilde{\hat{X}} - X) \tag{5.34}$$

In order to find the distribution we perform again an orthogonal transformation. First, we see that

$$\tilde{Y} = \tilde{\hat{X}} - X \sim N_u(0, \sigma_o^2 N^{-1}) \tag{5.35}$$

Since N is positive definite we know that there exists a non-singular matrix D such that

$$D^T N D = {}_u I_u \tag{5.36}$$

With the transformation

$$\tilde{\bar{Y}} = D^{-1}\tilde{Y} \tag{5.37}$$

we obtain

$$\tilde{\bar{Y}}^T N \tilde{\bar{Y}} = \tilde{Y}^T D^T N D \tilde{Y} = \tilde{Y}^T \tilde{Y} \tag{5.38}$$

The new random variable $\tilde{\bar{Y}}$ is multi-variate normal with

$$\tilde{\bar{Y}} \sim N_u(0, \sigma_o^2 I), \tag{5.39}$$

since

$$E(\tilde{\bar{Y}}) = D^{-1}E(\tilde{Y}) = 0$$

and $$\Sigma_{\tilde{\bar{Y}}} = D^{-1}M^{-1}(D^{-1})^T\sigma_o^2 = (D^T N D)^{-1}\sigma_o^2 = \sigma_o^2 I$$

The components of the random variable $\tilde{\bar{Y}}$ are stochastically independent and each is distributed like $n(0,\sigma_o^2)$. The variable $\tilde{Q}$ of equation (5.34) becomes, based on equations (5.35) and (5.38)

$$\tilde{Q} = \frac{\tilde{\bar{Y}}^T\tilde{\bar{Y}}}{\sigma_o^2} = \sum_i^u \frac{\tilde{\bar{y}}_i^2}{\sigma_o^2} \tag{5.40}$$

Since $\tilde{\bar{y}}_i/\sigma_o \sim n(0,1)$ we finally obtain

$$\tilde{Q} \sim \chi^2_u$$

or

$$\boxed{\frac{1}{\sigma_o^2}(\tilde{\bar{X}} - X)^T N (\tilde{\bar{X}} - X) \sim \chi^2_u} \tag{5.41}$$

In the preceeding derivation the implied assumption was that the normal matrix

N in nonsingular, or, equivalently, that there is no rank defect on the design matrix A. In the case that there is a rank defect on A, say $R(A) = r < u$, then it is easy to show that $\tilde{Q} \sim \chi^2_r$. The expression (5.41) could be used to test the parameters if the a-priori variance of unit weight were known. But we only have an estimate of σ_0, which is $\hat{\sigma}_0$. Therefore, in order to perform any testing we have to find a new random variable in which σ_0^2 cancels. Consider the two random variables:

$$\frac{\tilde{\sigma}_0^2}{\sigma_0^2} DF \sim \chi^2_{DF}$$

and

$$\frac{(\tilde{X} - X)^T N (\tilde{X} - X)}{\sigma_0^2} \sim \chi^2_r$$

It has previously been shown that $\tilde{\sigma}_0$ or, equivalently $\tilde{V}^T P \tilde{V}$, is stochastically independent of $\hat{X}$. Consequently

$$\frac{\frac{1}{\sigma_0^2}(\tilde{X} - X)^T N (\tilde{X} - X) \frac{1}{r}}{\frac{\tilde{\sigma}_0^2}{\sigma_0^2} DF \cdot \frac{1}{DF}} \sim F_{r,DF}$$

or,

$$\boxed{\tilde{F} \equiv \frac{(\tilde{X} - X)^T N (\tilde{X} - X)}{\tilde{\sigma}_0^2 \, r} \sim F_{r,DF}} \qquad (5.42)$$

The expression F is the equation of an r-dimensional ellipsoid with the center at $\hat{X}$. It defines a random region of constant probability

$$P(\tilde{F} \leq F_\alpha) = \int_0^F F_{r,DF} = 1 - \alpha \qquad (5.43)$$

In the case of F = 1 the region is called the "random r-dimensional ellipsoid of standard deviation." The principle axes of this ellipsoid are inversely

proportional to the eigenvalues of N. This is seen immediately if we replace the transformation (5.37) by

$$\tilde{X} = E^{-1}\tilde{Y}$$

such that $E^T N E = \Lambda$

where Λ is the diagonal matrix of eigenvalues and E is the eigenvector matrix of N. Thus, we obtain

$$\frac{\tilde{Z}^T \Lambda \tilde{Z}}{r\hat{\sigma}_o^2} = \frac{1}{\hat{\sigma}_o^2 r}\sum^r \tilde{z}_i^2 \lambda_i = \frac{1}{\hat{\sigma}_o^2 r}\sum^r \frac{\tilde{z}_i^2}{\left(\frac{1}{\sqrt{\lambda_i}}\right)^2} = 1 \qquad (5.44)$$

But this is the principle axes form of the equation of the r-dimensional ellipsoid. From (5.44) it is seen that the ellipsoid becomes smaller the larger the eigenvalues of the normal matrix. Since the probability is $1-\alpha$ that the ellipsoid contains the true parameter $\tilde{X}$, we naturally would like the ellipsoid to be as small as possible. It is exactly this relationship which makes us choose a design such that the eigenvalues of N are large, provided we have a choice at all.

With (5.42) we can test hypotheses of the parameters.

Hypothesis: H_o: $X = X_T$ (some numerical value)

H_1: $X \neq X_T$

Compute the statistics F under the null hypothesis and replace the random variable by its sample value

$$F = \frac{(\hat{X} - X_T)^T N (\hat{X} - X_T)}{\hat{\sigma}_o^2 \; r}$$

Reject H_o at a 100α % significance level if

$$F > F_{r,DF,\alpha}$$

Note that the critical region is taken at the upper trail of the distribution. A small value for F indicates acceptance of H_o (knowing that σ_o already has been tested and accepted).

5.3 General Linear Hypothesis

Consider the model

$$\tilde{V} = AX + \tilde{L},$$

$$CX + W = 0 \qquad (5.45)$$

$$E[\tilde{L} - E(\tilde{L})][\tilde{L} - E(\tilde{L})]^T = \Sigma_{L_b}$$

which is recognized as the model of observation equations with conditions on the parameters. The solution is

$$\tilde{\hat{X}} = \tilde{\hat{X}}^* + \Delta\tilde{\hat{X}}$$

with $$\tilde{\hat{X}}^* = -(A^TPA)^{-1}A^TP\tilde{L} = -N^{-1}A^TP\tilde{L}$$

$$\Delta\tilde{\hat{X}} = N^{-1}C^T[CN^{-1}C^T]^{-1}[-W-C\tilde{\hat{X}}^*], \qquad (5.46)$$

and $$\tilde{\hat{V}}^TP\tilde{\hat{V}} = \tilde{\hat{V}}^TP\tilde{\hat{V}}^* + \Delta\tilde{\hat{V}}^TP\tilde{\hat{V}}$$

with $$\Delta\tilde{\hat{V}}^TP\tilde{\hat{V}} = [C\tilde{\hat{X}}^* + W]^T[CN^{-1}C^T]^{-1}[C\tilde{\hat{X}}^* + W]$$

We observe that this is a sequential solution. The whole procedure can also be interpreted in such a way that we are given the least squares solution of

the first system and would like to test the linear hypothesis

$$H_o: \quad CX + W = 0 \tag{5.47}$$

$$H_1: \quad CX + W \neq 0$$

H_o consists of s linear equations which the parameters have to fulfill. The test is going to be based on a random variable $\tilde{F}$ which relates $\tilde{\tilde{V}}^T P \tilde{\tilde{V}}$ and $\Delta\tilde{\tilde{V}}^T P \tilde{\tilde{V}}$. First we confirm the independence of both quadratic forms.

Consider the observation equations only:

$$\tilde{V} = AX + \tilde{L} \tag{5.48}$$

with $E(\tilde{V}) = 0$ or $E(-\tilde{L}) = AX$

The least squares solution of (5.48) is

$$\tilde{\tilde{X}}^* = -N^{-1} A^T P \tilde{L},$$

having a distribution

$$\tilde{\tilde{X}}^* \sim N_u[X, \sigma_o^2 N^{-1}], \tag{5.49}$$

i.e. the expected value of the estimate is an unbiased estimate for X. Next we consider the condition equation

$$CX + W = 0 \tag{5.50}$$

The parameters in (5.50) take on the role of "observation". The adjusted parameters of the observation equations, $\tilde{\tilde{X}}^*$, are the "observed values". Thus

$$C(\tilde{\tilde{X}}^* + \Delta\tilde{\tilde{X}}) + W = 0 \tag{5.51}$$

where $\Delta\tilde{\tilde{X}}$ are the "residuals" with the expected value of zero. Since $E(\tilde{\tilde{X}}) = X$

it must be that

$$E(\Delta\tilde{\tilde{X}}) = 0$$

It follows that $\Delta\tilde{\tilde{X}}$ is multivariate normal with

$$\Delta\tilde{X} \sim N(0,\sigma_0^2 N^{-1}) \tag{5.52}$$

Also, the least squares solution of the condition equation (5.51) results in identical expressions as given in (5.46). The change in the residuals due to the conditions (5.51) can be computed as follows:

$$\tilde{\tilde{V}}^* = A\tilde{\tilde{X}}^* + \tilde{L}$$

$$\tilde{\tilde{V}} = A(\tilde{\tilde{X}}^* + \Delta\tilde{\tilde{X}}) + \tilde{L} \tag{5.53}$$

Differencing both expressions gives

$$\Delta\tilde{V} = \tilde{\tilde{V}} - \tilde{\tilde{V}}^* = A\Delta\tilde{\tilde{X}} \tag{5.54}$$

with $E(\Delta\tilde{V}) = AE(\Delta\tilde{\tilde{X}}) = 0$

and $E(\tilde{\tilde{V}}^*) = E[A\tilde{\tilde{X}}^* + \tilde{L}] = AX - AX = 0 \tag{5.55}$

The covariance between $\Delta\tilde{V}$ and $\tilde{\tilde{V}}^*$ is

$$E[(\Delta\tilde{V})\tilde{\tilde{V}}^{*T}] = E(A\Delta\tilde{\tilde{X}}\tilde{\tilde{V}}^{*T})$$

$$= [-AN^{-1}C^T(CN^{-1}C^T)^{-1}(W + C\tilde{\tilde{X}}^*)\tilde{\tilde{V}}^{*T}]$$

$$= -AN^{-1}C^T(CN^{-1}C^T)^{-1}\{E(W\tilde{\tilde{V}}^{*T}) + CE(\tilde{\tilde{X}}^*\tilde{\tilde{V}}^{*T})\}$$

Both random variables are stochastically independent if the covariance is zero.

We have

$$E(\tilde{\tilde{W}}\tilde{\tilde{V}}^{*T}) = 0 \text{ since } E(\tilde{\tilde{V}}^{*}) = 0$$

For the second term it follows that

$$\begin{aligned} E(\tilde{\tilde{X}}^{*}\tilde{\tilde{V}}^{*T}) &= E[\tilde{\tilde{X}}^{*}(\tilde{\tilde{X}}^{*T}A^{T} + \tilde{L}^{T})] \\ &= E[\tilde{\tilde{X}}^{*}\tilde{\tilde{X}}^{*T}A^{T}] + E[\tilde{\tilde{X}}^{*}\tilde{L}^{T}] \\ &= E(N^{-1}A^{T}P\tilde{L}\tilde{L}^{T}PAN^{-1}A^{T}) + E(-N^{-1}A^{T}P\tilde{L}\tilde{L}^{T}) \\ &= \sigma_o^2 N^{-1}A^{T} + XX^{T}A - \sigma_o^2 N^{-1}A^{T} - XX^{T}A \\ &= 0 \end{aligned}$$

This completes the proof that $E(\Delta\tilde{\tilde{W}}\tilde{\tilde{V}}^{*}) = 0$ and, therefore, both types of residuals are stochastically independent. It is only necessary to show that $\Delta\tilde{\tilde{V}}^{T}P\tilde{\tilde{V}}$ is related to $\Delta\tilde{V}$. From equation (5.46) we find the following relation

$$\Delta\tilde{\tilde{X}}N\Delta\tilde{\tilde{X}} = \Delta\tilde{\tilde{V}}^{T}P\tilde{\tilde{V}}, \qquad (5.56)$$

and equation (5.54) yields

$$\begin{aligned} (\Delta\tilde{V})^{T}P(\Delta\tilde{V}) &= \Delta\tilde{\tilde{X}}^{T}A^{T}PA\Delta\tilde{\tilde{X}} \\ &= \Delta\tilde{\tilde{X}}^{T}N\Delta\tilde{\tilde{X}} \qquad (5.57) \end{aligned}$$

Comparing (5.56) and (5.56) gives

$$\Delta\tilde{\tilde{V}}^{T}P\tilde{\tilde{V}} = (\Delta\tilde{V})^{T}P(\Delta\tilde{V}) \qquad (5.58)$$

Since $(\Delta\tilde{V})$ and $\tilde{\tilde{V}}^{*}$ are stochastically independent, the same must be true for $\Delta\tilde{\tilde{V}}^{T}P\tilde{\tilde{V}}$ and $\hat{V}^{T}P\tilde{V}^{*}$.

The distribution of $\Delta V^T PV$ can be obtained in the usual manner.

$$\Delta\tilde{\hat{V}}^T P\tilde{\hat{V}} = \tilde{Y}^T(CN^{-1}C^T)^{-1}\tilde{Y}$$

with $$\tilde{Y} = W + C\tilde{\hat{X}}^*$$

Since $$E(\tilde{Y}) = E(W + C\tilde{\hat{X}}^*) = W + CX = 0.$$

$$\Sigma_Y = C\Sigma_{X^*}C^T = \sigma_o^2 CN^{-1}C^T$$

the distribution of the random variable Y is

$$\tilde{Y} \sim N_s[0,\sigma_o^2(CN^{-1}C^T)].$$

If the s conditions are linearly independent then the variance-covariance matrix has full rank, $R(\Sigma_Y) = s$. Through a non-singular transformation we can find a random variable $\tilde{\bar{Y}}$ such that

$$\Delta\tilde{\hat{V}}^T P\tilde{\hat{V}} = \tilde{Y}^T(CN^{-1}C^T)^{-1}\tilde{Y} = \tilde{\bar{Y}}^T\tilde{\bar{Y}} = \sum^{s}\tilde{\bar{y}}_i^2$$

and $$\tilde{\bar{Y}} \sim N_s(0,\sigma_o^2 I)$$

Since $$\frac{\tilde{\bar{y}}_i}{\sigma_o} \sim n(0,1)$$

the distribution for the quadratic form becomes

$$\frac{\Delta\tilde{\hat{V}}^T P\tilde{\hat{V}}}{\sigma_o^2} = \sum^{s}\frac{\tilde{\bar{y}}_i^2}{\sigma_o^2} \sim \chi_s^2 \qquad (5.59)$$

Because of the stochastic independence of $\Delta\tilde{\hat{V}}^T P\tilde{\hat{V}}$ and $\tilde{\hat{V}}^{*T} P\tilde{\hat{V}}^*$ we can find the new random variable

$$\tilde{F} = \frac{\frac{\Delta\tilde{\hat{V}}^T P \tilde{\hat{V}}}{\sigma_o^2} \cdot \frac{1}{s}}{\frac{\tilde{\hat{V}}^T P \tilde{V}^*}{\sigma_o^2} \cdot \frac{1}{DF}} \sim F_{s,DF}$$

$$\boxed{\tilde{F} = \frac{\Delta\tilde{\hat{V}}^T P \tilde{\hat{V}}}{(\tilde{\hat{\sigma}}_o^*)^2} \cdot \frac{1}{s} = \frac{\Delta\tilde{\hat{V}}^T P \tilde{\hat{V}}}{\tilde{\hat{V}}^T P \tilde{V}^*} \cdot \frac{DF}{s} \sim F_{s,DF}} \qquad (5.60)$$

DF is the degree of freedom for the first system only. Note that the unknown σ_o cancels. A small $\Delta V^T PV$ implies that the null hypothesis (5.47) is acceptable, i.e. the conditions are in "agreement with the information" expressed in the observations of the first model. The conditions do not cause any "distortions" of the first model. We, therefore, usually apply only a one tail test at the upper end of the distribution. Thus, reject H_o at a $100\alpha\%$ significance level if

$$F > F_{s,DF,\alpha}$$

Examples for generalized hypothesis testing:

Case 1: Test on all parameters:

$$H_o: X = X_T$$

$$H_1: X \neq X_T$$

The test is to find out whether the parameters can have a certain numerical value X_T. We compute the statistics with the numerical values of H_o and replace the random variables by the sample values. With

$$C = {}_uI_u$$

and $W = -X_T$

the statistics (5.60) becomes, together with equation (5.46),

$$F = \frac{(\hat{X}^* - X_T)^T N (\hat{X}^* - X_T)}{\hat{V}^T P \hat{V}^*} \cdot \frac{DF}{u}$$

which is to be compared with $F_{u,DF,\alpha}$.

Case 2: Test on a single parameter:

H_o: $x_i = x_{iT}$

H : $x_i \neq x_{iT}$

Since $C = (0 \ldots 1 \ldots 0)$

and $W = -x_{iT}$

the statistic becomes

$$F = \frac{(x_i^* - x_{iT})^2}{q_{ii} \cdot \hat{V}^T P \hat{V}^*} \cdot DF$$

which is to be compared with $F_{1,DF,\alpha} \equiv t^2_{DF,\alpha/2}$

q_{ii} is the ith diagonal element of N^{-1}.

Case 3: Test on a subset of parameters.

Arrange the normal matrix such that the parameters to be tested appear at the lower end of the parameter vector. $X = [X_1, X_2]^T$

Hypothesis:

$$H_o: \quad {}_q X_{2_1} = X_{2T}$$

$$H_1: \quad X_2 \neq X_{2T}$$

with $q < u$. The constraint reads:

$$(0 \; I)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} - X_{2T} = 0$$

Compute the statistics

$$F = \frac{(\hat{X}_2 - X_{2T})^T (N_{22} - N_{21} N_{11}^{-1} N_{12})(\hat{X}_2 - X_{2T})}{\hat{V}^T P \hat{V}^*} \cdot \frac{DF}{q}$$

and compare it with $F_{q,DF,\alpha}$. The matrices N_{22}, N_{21}, N_{11} and N_{12} are the submatrices of the normal matrix corresponding to the subset of parameters X_1 and X_2. We observe that the test can, of course, be carried out without the particular arrangement of the parameters. But in certain instances, e.g. when the inverse of the normal matrix is computed using partitioning techniques, computational efforts can be reduced since the computation of the statistics does not require any further matrix inversions if the parameters to be tested are arranged as indicated.

Case 4: Test of equality of parameters.

The general linear hypothesis can also be used to test the equality of parameters:

$$H_o: \quad x_i = x_k$$

$$H_1: \quad x_i \neq x_k$$

The linear condition equation becomes in this case

$$C = (0\ldots010\text{-}10\ldots0)$$

$$W = 0$$

Evaluating the statistics under H_o and substituting the sample values for the random variables gives

$$F = \frac{(\hat{x}_j - \hat{x}_k)^2 (CN^{-1}C)^{-1}}{\hat{V}^T P \hat{V}^*} \quad DF$$

which is to be compared with $F_{1,DF,\alpha} \equiv t^2_{DF,\alpha/2}$

5.4 Ellipse of Standard Deviation

Ellipses of standard deviation can be computed for any pair of parameters. They are a graphical representation of the marginal standard deviation of the parameters and the covariance between them. In surveying, ellipses of standard deviation are usually computed in adjustments of horizontal networks since they allow a convenient interpretation of the directional station position accuracy. This concept can be extended to more dimensions. In the case that three parameters are involved we speak of "ellipsoids of standard deviation", or "hyperellipsoids of standard deviation", if more than three parameters are involved.

Recall that the random variable $\tilde{\tilde{X}}$ is distributed multivariate normal as

$$\tilde{\tilde{X}} \sim N(X, \sigma_o^2 N^{-1})$$

$$\sim N(X, \sigma_o^2 Q_X) \qquad (5.61)$$

where Q_X is the cofactor matrix of X.

According to theorems of Section 5.1 the marginal distribution of a subset of $\tilde{X}$ is multivariate normal with the mean, variance and covariance taken from the respective positions of X and Q_X. Thus,

$$\begin{pmatrix} \tilde{x}_i \\ \tilde{x}_j \end{pmatrix} \sim N_2\left[\begin{pmatrix} x_i \\ x_j \end{pmatrix}, \sigma_o^2 \begin{pmatrix} q_{ii} & q_{ij} \\ q_{ji} & q_{jj} \end{pmatrix}\right] \tag{5.62}$$

or

$$\begin{pmatrix} \tilde{x}_i \\ \tilde{x}_j \end{pmatrix} \sim N_2[0, \sigma_o^2 Q_i] \tag{5.63}$$

Since Q_X is a positive definite matrix, any submatrix Q_i, which is created as explained above, is also positive definite. Just imagine that the parameters are renumbered and that x_i and x_j become x_1 and x_2. In that case Q_i is the upper principal submatrix which has to be positive according to the definition of positive definite matrices.

Recognizing the fact that a marginal distribution like (5.63) can be given for <u>any</u> two parameters we interpret x_i and x_j, in view of the application in Surveying, as the two dimensional Cartesian station coordinates of points. In an attempt to simplify notation let us write

$$\tilde{Y}_i = (\tilde{X}_i - X_i) = \begin{pmatrix} \tilde{x}_i - x_i \\ \tilde{x}_j - x_j \end{pmatrix} \tag{5.64}$$

where X_i is a two dimensional vector containing the Cartesian coordinates of station i. Also, if there is no reason for confusion the subscript i will be omitted and we call $x_1 = x_i$ and $x_2 = x_j$. (Y) denotes a coordinate system parallel to the system (X) and with its origin at X.

Consider the quadratic form

$$\tilde{S} = \frac{1}{\sigma_o^2} (\tilde{X}_i - X_i)^T Q_i^{-1} (\tilde{X}_i - X_i) \sim \chi_2^2 \tag{5.65}$$

From the discussion of previous sections it is known that $\tilde{S}$ has a Chi-squared distribution of two degrees of freedom. Since Q_i is positive definite the quadratic form (5.65) is the equation of an ellipse with its center at $\tilde{\hat{X}}_i$. The ellipse is the curve of constant probability α,

$$P\left[\frac{1}{\sigma_o^2}\ (\tilde{\hat{X}}_i - X_i)^T Q_i^{-1} (\tilde{\hat{X}}_i - X_i) \leq \gamma_\alpha\right] = \int_0^{\gamma_\alpha} \chi_2^2\, dx = 1-\alpha \qquad (5.66)$$

Strictly speaking the area defined by

$$\frac{1}{\sigma_o^2}\ (\tilde{\hat{X}}_i - X_i)^T Q_i^{-1} (\tilde{\hat{X}}_i - X_i) \leq \gamma$$

is a random region (analogous to the random interval in unit variate distributions) since its center is given by a random vector. Applying an orthogonal transformation (rotation) the equation for the ellipse can be transformed to the (Z) coordinate system whose axes are parallel to the principal axis of the ellipse. Let R_i be the orthogonal matrix whose columns are the eigenvectors of Q_i; then the transformation

$$\tilde{Z}_i = R_i^T \tilde{Y}_i$$

or

$$\tilde{Y}_i = R_i \tilde{Z}_i$$

leads to the "principal axes" form

$$\sigma_o^2 \tilde{S} = Z_i^T R_i^T Q_i^{-1} R_i \tilde{Z}_i = \tilde{Z}_i^T \Lambda_i^{-1} \tilde{Z}_i = \frac{\tilde{z}_1^2}{\lambda_1 Q} + \frac{\tilde{z}_2^2}{\lambda_2 Q} \qquad (5.67)$$

Since R_i is the eigenvector matrix of Q_i we have $R_i^T Q_i R_i = \Lambda_i$ and $R_i^T Q_i^{-1} R_i = \Lambda_i^{-1}$, where Λ_i is a diagonal matrix with the eigenvalues of Q_i

at the diagonal. Both eigenvalues are positive. Combining equations (5.66) and (5.67) gives

$$P\left[\frac{\tilde{z}_1^2}{\left(\sigma_0\sqrt{\lambda_1^Q\,\gamma_\alpha}\right)^2}+\frac{\tilde{z}_2^2}{\left(\sigma_0\sqrt{\lambda_2^Q\,\gamma_\alpha}\right)^2}\le 1\right]=\int_0^{\gamma_\alpha}\chi_2^2\,dx=1-\alpha \qquad (5.68)$$

The random region (ellipse) is given here in the coordinate system of the principal axis. We observe that the size of the ellipse increases as the eigenvalues λ_i^Q increase. This is in agreement with the statement made earlier that the size of the random hyper ellipsoid decreases as the eigenvalues of the normal matrix increase. The eigenvalues of the normal matrix and those of the full sized cofactor matrix are reciprocal. γ serves as a magnification of the ellipse. If γ is equal to unity the region is called "random ellipse of standard deviation".

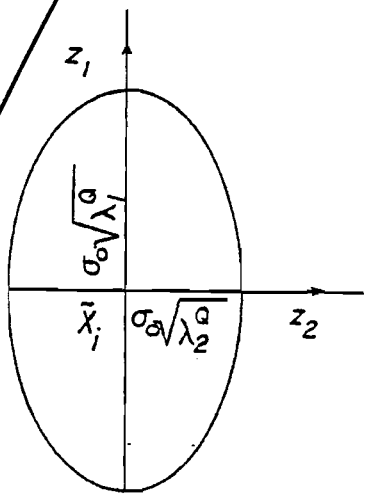


Figure 5.1 Random Ellipse of Standard Deviation.

Specific values for the integral (5.68) can be found in the following tables:

$\sqrt{\gamma}$	1	2	3
$1-\alpha$	39.35	86.47	98.89

$\sqrt{\gamma}$	2.146	2.448	3.035
$1-\alpha$	90	95	99

It is seen that in only 39% of all cases the random ellipse of the standard deviation contains the parameters x_1 and x_2. A magnification by three is necessary to ensure a probability of 99%.

The size of the ellipse in Fig. (5.1) is a function of the constant but unknown a-priori variance of unit weight. In order to overcome this dependency we recall that the residuals $\tilde{V}$ and the parameters $\tilde{X}$ are stochastically independent. We can, therefore, form a new random variable

$$\tilde{F} = \frac{\left[\frac{\tilde{z}_1^2}{\left(\sigma_0 \sqrt{\lambda_1^Q}\right)^2} + \frac{\tilde{z}_2^2}{\left(\sigma_0 \sqrt{\lambda_2^Q}\right)^2} \right] \frac{1}{2}}{\frac{\tilde{\hat{V}}^T P \tilde{\hat{V}}}{\sigma_0^2} \cdot \frac{1}{DF}} \sim F_{2,DF}, \qquad (5.69)$$

which has an F distribution of 2 and DF degrees of freedom. Since

$$\tilde{\sigma}_0^2 = \frac{\tilde{\hat{V}}^T P \tilde{\hat{V}}}{DF}$$

the expression above can be further simplified as

$$\tilde{F} = \left[\frac{\tilde{z}_1^2}{\left(\tilde{\sigma}_0 \sqrt{\lambda_1^Q}\right)^2} + \frac{\tilde{z}_2^2}{\left(\tilde{\sigma}_0 \sqrt{\lambda_2^Q}\right)^2} \right] \frac{1}{2} \sim F_{2,DF}$$

We find

$$P(\tilde{F} \leq F_\alpha) = \int_0^{F_\alpha} F_{2,DF} = 1-\alpha$$

or

$$P\left[\frac{\tilde{z}_1^2}{\left(\tilde{\sigma}_0 \sqrt{\lambda_1^Q 2F_\alpha}\right)^2} + \frac{\tilde{z}_2^2}{\left(\tilde{\sigma}_0 \sqrt{\lambda_2^Q 2F_\alpha}\right)^2} \leq 1 \right] = \int_0^{F_\alpha} F_{2,DF} = 1-\alpha \qquad (5.70)$$

Although the unknown variance of unit weight was eliminated, the region defined in equation (5.70) is an ellipse whose size is itself a random variable.

The special case for which $\sqrt{2F_\alpha} = 1$ is again called the random ellipse of standard deviation. The probability that the population parameters x_1 and x_2 fall within this particular ellipse is a function of the degree of freedom of the adjustment. In the table below some magnification factors for the random ellipse of standard deviation are given as a function of degree of freedom and probability. It is seen that in the range of small degrees of freedom an increase in the degree of freedom rapidly decreases the magnification factor; whereas, in the case of a large degree of freedom, any additional observations cause only a minor reduction of the magnification factor. A degree of freedom of ten appears desirable.

Table 5.1 Magnification factors $\sqrt{2\,F_{2,DF,\alpha}}$ for the random ellipse of standard deviation in order to gain a probability $1-\alpha$ that the magnified ellipse contains the population parameters x_1 and x_2.

n-u	Probability 1-α		
	95%	98%	99%
1	20.0	50.0	100
2	6.16	9.90	14.1
3	4.37	6.14	7.85
4	3.73	4.93	6.00
5	3.40	4.35	5.15
6	3.21	4.01	4.67
8	2.99	3.64	4.16
10	2.86	3.44	3.89
12	2.79	3.32	3.72
15	2.71	3.20	3.57
20	2.64	3.09	3.42
30	2.58	2.99	3.28
50	2.52	2.91	3.18
100	2.49	2.85	3.11
∞	2.45	2.80	3.03

In analogy to the case of random intervals of Section 4.5.2, we can replace the random variables $\tilde{x}_i$ and $\tilde{\sigma}_0$ by the sample values $\hat{x}_i$ and $\hat{\sigma}_0$ and thus fix the center and the size of the ellipse. This action is accompanied by replacing the adjective "random" by the noun "confidence". For reasons of simplification the term "confidence ellipse of standard deviation" is

further simplified to "ellipse of standard deviation".

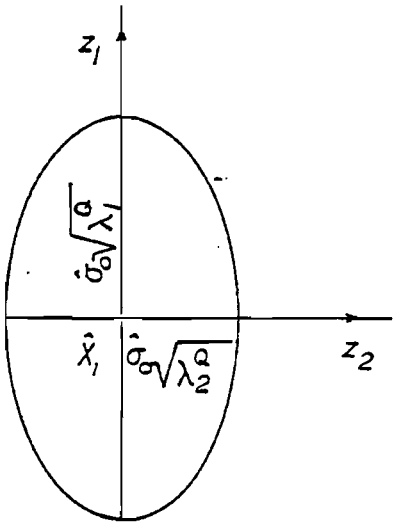


Fig. 5.2 Ellipse of Standard Deviation. (The center is at the adjusted point $\hat{x}_1$ and the size is a function of the a-posteriori variance of unit weight.)

The eigenvalues are found from the characteristic equation

$$|Q_i - \lambda I| = \begin{vmatrix} q_{x_1} - \lambda & q_{x_1x_2} \\ q_{x_1x_2} & q_{x_2} - \lambda \end{vmatrix} = (q_{x_1} - \lambda)(q_{x_2} - \lambda) - q_{x_1x_2}^2 = 0$$

The solution is

$$\lambda_{1,2}Q_i = \frac{q_{x_1} + q_{x_2}}{2} \pm \frac{1}{2} W \qquad (5.71)$$

where $$W = \sqrt{(q_{x_1} - q_{x_2})^2 + 4q_{x_1x_2}^2} \qquad (5.72)$$

Since q_{x_1} and q_{x_2} are both positive the length of the semi-major and minor axes

will be obtained depending on the sign of W.

Although the size of the ellipse of standard deviation is known now, the orientation of the ellipse with respect to the (X) system has to be found and a criterion has to be established to decide which of the Z_i-axes coincides with the minor or major axis of the ellipse. Both problems are most easily solved by comparing the equation for the ellipse in different coordinate systems. The $(Z)_i$ and the $(Y)_i$ systems are both related by a rotation

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = R_i^T \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{5.73}$$

where R_i is the eigenvector matrix of Q_i. Let the rotation angle be ϕ; then

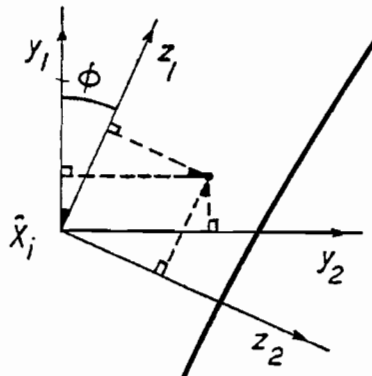


Figure 5.3 Rotation of Coordinate System.

the following relations can be seen from Figure 5.3.

$$\begin{aligned} z_1 &= y_2 \sin\phi + y_1 \cos\phi \\ z_2 &= y_2 \cos\phi - y_1 \sin\phi \end{aligned} \tag{5.74}$$

In equation (5.67) we saw that

$$Y_i^T Q_i^{-1} Y_i = Z_i^T \Lambda_i^{-1} Z_i$$

This expression can be rewritten as

$$Y_i^T Q_i Y_i = Z_i^T \Lambda_i Z_i = z_1^2 \lambda_1^Q + z_2^2 \lambda_2^Q \tag{5.75}$$

since $R_i^T Q_i R_i = \Lambda_i$. From (5.74) we obtain

$$\begin{aligned} z_1^2 \lambda_1^Q + z_2^2 \lambda_2^Q = {} & y_2^2 (\lambda_1^Q \sin^2\phi + \lambda_2^Q \cos^2\phi) \\ & + y_1^2 (\lambda_1^Q \cos^2\phi + \lambda_2^Q \sin^2\phi) \\ & + y_1 y_2 (\lambda_1^Q - \lambda_2^Q) 2\sin\phi\cos\phi \end{aligned} \tag{5.76}$$

This expression is equivalent to

$$\begin{aligned} Y_i^T Q_i Y_i &= (y_1 y_2) \begin{pmatrix} q_{x_1} & q_{x_1 x_2} \\ q_{x_1 x_2} & q_{x_2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= y_2^2 q_{x_2} + y_1^2 q_{x_1} + y_1 y_2 2 q_{x_1 x_2} \end{aligned} \tag{5.77}$$

Comparison of some of the coefficients of (5.76) and (5.77) gives

$$2(\lambda_1^Q - \lambda_2^Q)\sin\phi\cos\phi = 2 q_{x_1 x_2}$$

or

$$\sin 2\phi = \frac{2 q_{x_1 x_2}}{\lambda_1^Q - \lambda_2^Q} = \frac{2 q_{x_1 x_2}}{W} \tag{5.78}$$

and

$$\cos 2\phi = 1 - \sin^2 2\phi = \frac{q_{x_1} - q_{x_2}}{W} \tag{5.79}$$

The equations (5.78) and (5.79) are sufficient to determine uniquely the angle ϕ. The angle refers to the semi-major axis or semi-minor axis depending on whether W is taken positive or negative.

The following is a summary of relations needed to determine the shape and orientation of the ellipse of standard deviation and some properties related to it.

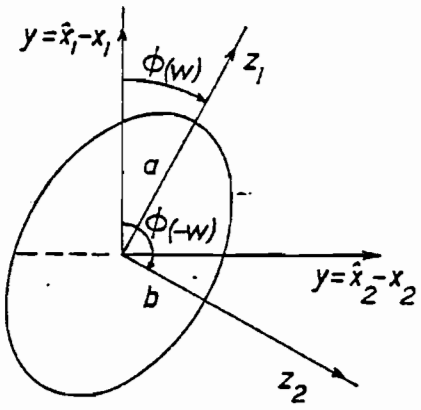


Figure 5.4 Defining Elements of the Ellipse of Standard Deviation

$$W = +\sqrt{(q_{x_1} - q_{x_2})^2 + 4q_{x_1x_2}^2}$$

$$a = \hat{\sigma}_0\sqrt{\frac{q_{x_1} + q_{x_2}}{2} + \frac{1}{2}W} = \hat{\sigma}_0\sqrt{\lambda_1 Q}$$

$$b = \hat{\sigma}_0\sqrt{\frac{q_{x_1} + q_{x_2}}{2} - \frac{1}{2}W} = \hat{\sigma}_0\sqrt{\lambda_2 Q}$$

$$\sin 2\phi = \frac{2q_{x_1x_2}}{W}$$

$$\cos 2\phi = \frac{q_{x_1} - q_{x_2}}{W}$$

(5.80)

In order to study the properties of the ellipse of standard deviation let us compute the positional error with respect to a general direction ψ.

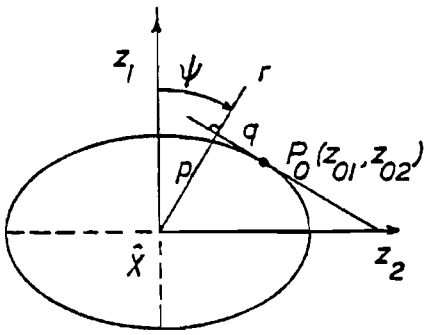


Figure 5.5 Positional Error p

The distribution of the random variable $\tilde{Z}$ is multivariate normal with

$$\begin{pmatrix}\tilde{z}_1\\ \tilde{z}_2\end{pmatrix} \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \sigma_0^2\begin{pmatrix}\lambda_1^Q & 0\\ 0 & \lambda_2^Q\end{pmatrix}\right] \tag{5.81}$$

i.e. both variables are stochastically independent. Consider the random variable $\tilde{r}$ which is the projection of the random point $(\tilde{z}_1, \tilde{z}_2)$ onto a line whose azimuth is ψ:

$$\tilde{r} = \tilde{z}_1\cos\psi + \tilde{z}_2\sin\psi, \tag{5.82}$$

The estimated variance of this random variable is

$$\hat{\sigma}_r^2 = a^2\cos^2\psi + b^2\sin^2\psi \tag{5.83}$$

as can be verified by using the law of variance propagation. This variance has a geometrical interpretation. Let the ellipse be projected onto the line $\tilde{r}$. The directions of r and of the tangent on the ellipse differ by

90^0. Since the equation of the ellipse is given by

$$\frac{z_1^2}{a^2} + \frac{z_2^2}{b^2} = 1$$

the total differential becomes

$$\frac{2z_1 dz_1}{a^2} + \frac{2z_2 dz_2}{b^2} = 0,$$

or

$$\frac{dz_1}{dz_2} = -\frac{z_2}{z_1}\frac{a^2}{b^2}$$

Since the slope of the tangent is $(-\psi)$, this equation becomes

$$-\frac{z_2}{z_1}\frac{a^2}{b^2} = -\tan\psi$$

or

$$\frac{z_{01}}{a^2}\sin\psi - \frac{z_{02}}{b^2}\cos\psi = 0 \qquad (5.84)$$

This equation relates the coordinates of an elliptical point P_0 to the slope of the tangent on the ellipse at that point. The length p of the projection of the ellipse is according to Figure 5.5.

$$p = z_{01}\cos\psi + z_{02}\sin\psi \qquad (5.85)$$

In order to implement the condition (5.84) we square it and multiply the results by $a^2 b^2$, giving

$$\frac{z_{01}^2 b^2}{a^2}\sin^2\psi - 2z_{01}z_{02}\cos\psi\sin\psi + \frac{z_{02}^2 a^2}{b^2}\cos^2\psi = 0$$

Adding this expression to the square of p in (5.85),

$$p^2 = z_{01}^2\cos^2\psi + 2z_{01}z_{02}\cos\psi\sin\psi + z_{02}^2\sin^2\psi,$$

gives

$$p^2 = \frac{z_{01}^2}{a^2}(b^2\sin\psi + a^2\cos\psi)^2 + \frac{z_{02}^2}{b^2}(a^2\cos^2\psi + b^2\sin^2\psi)$$

$$= \left(\frac{z_{01}^2}{a^2} + \frac{z_{02}^2}{b^2}\right)(b^2\sin^2\psi + a^2\cos^2\psi)$$

$$= a^2\cos^2\psi + b^2\sin^2\psi \qquad (5.86)$$

Comparing this expression with (5.83) we recognize that $\hat{\sigma}_r = p$, i.e. the standard deviation in a certain direction is equal to the projection of the ellipse in that particular direction. Therefore, the ellipse of standard deviation is not a "standard deviation curve". Figure (5.6) exhibits the standard deviations for several directions. It is seen that for narrow

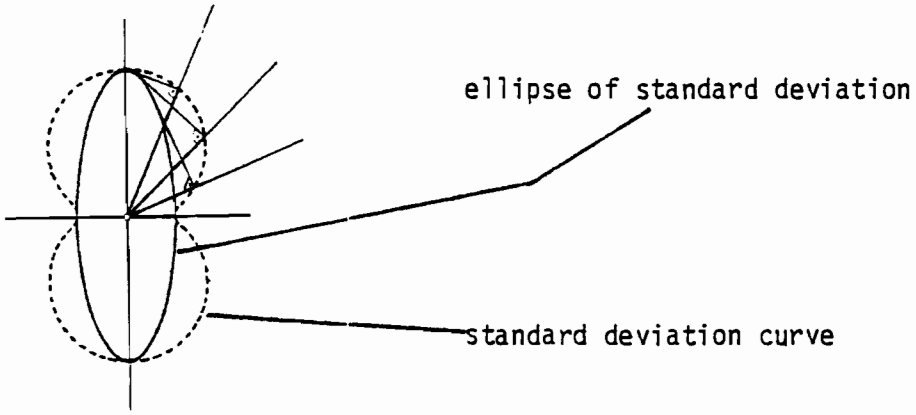


Figure 5.6 Standard Deviation Curve

ellipses there is only a small segment for which the standard deviations are close to the length of the semi-minor axis. The standard deviation increases rapidly as the direction ψ moves away from the minor principal

axis. Therefore, an extremely narrow ellipse is not very desirable if the over-all accuracy for the station position is important.

As a by-product of the property discussed above it is seen that the marginal standard deviations $\hat{\sigma}_{x_1}$ and $\hat{\sigma}_{x_2}$ are the projections of the ellipse onto the directions of the x_1 and x_2 axes, and that a and b must be the maximum and minimum standard deviations of the point. The rectangle formed by the semi-sides $\hat{\sigma}_{x_1}$ and $\hat{\sigma}_{x_2}$ encloses the ellipse. In many cases this rectangle can be used as a substitute in order to obtain an idea of the

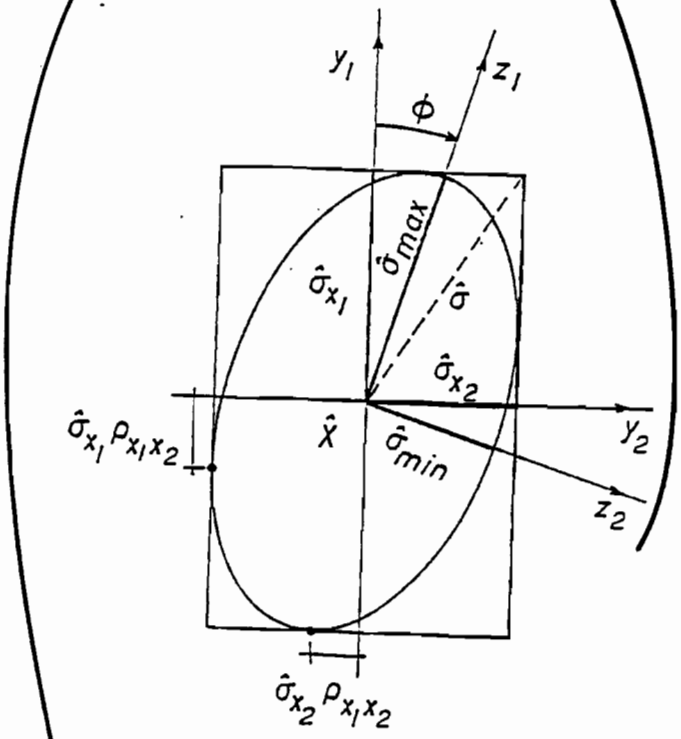


Figure 5.7 Characteristics of the Ellipse of Standard Deviation

shape of the ellipse of standard deviations. The semi-major axis must always be smaller than the diagonal of the rectangle. The diagonal itself is sometimes referred to as the "mean position error σ",

$$\hat{\sigma} = \sqrt{\hat{\sigma}_{x_1}^2 + \hat{\sigma}_{x_2}^2} = \hat{\sigma}_0\sqrt{\mathrm{Tr}Q_{\hat{l}}} = \hat{\sigma}_0\sqrt{q_{x_1} + q_{x_2}} \qquad (5.87)$$

This error has an interesting property in that it is invariant with respect to the specific orientation of the coordinate system (X)! From equation (5.71) we see that

$$q_{x_1} + q_{x_2} = \lambda_1^Q + \lambda_2^Q. \qquad (5.88)$$

But the eigenvalues are independent of the choise of the coordinate system. They are related to the lengths of the principal axes which are a characteristic of the distribution, and not of the particular coordinate system chosen to describe the distribution. From equations (5.87) and (5.88) it follows that

$$\hat{\sigma} = \hat{\sigma}_o \sqrt{\lambda_1^Q + \lambda_2^Q} = \sqrt{\hat{\sigma}^2_{max} + \hat{\sigma}^2_{min}} \tag{5.89}$$

The points of contact of ellipse and rectangle are a function of the correlation coefficients. For the points of contact the tangent on the ellipse is either horizontal or vertical. The equation of the ellipse in the (Y)-system is:

$$(y_1 y_2) \begin{pmatrix} q_{x_1} & q_{x_1x_2} \\ q_{x_1x_2} & q_{x_2} \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 1$$

Computing the inverse, the expression becomes

$$(y_1 y_2) \begin{pmatrix} q_{x_2} & -q_{x_1x_2} \\ -q_{x_1x_2} & q_{x_1} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = q_{x_1} q_{x_2} - q^2_{x_1x_2}$$

Evaluating the left side and dividing both sides by $q_{x_1} q_{x_2}$ gives

$$\frac{y_1^2}{q_{x_1}} + \frac{y_2^2}{q_{x_2}} - \frac{2y_1y_2q_{x_1x_2}}{q_{x_1} q_{x_2}} = 1 - \rho^2_{x_1x_2}, \tag{5.90}$$

from which it follows that

$$\frac{dy_1}{dy_2} = \frac{\dfrac{2y_2}{q_{x_2}} - \dfrac{2y_1 q_{x_1x_2}}{\sqrt{q_{x_1} q_{x_2}}}}{\dfrac{2y_2 \rho_{x_1x_2}}{\sqrt{q_{x_1} q_{x_2}}} - \dfrac{2y_1}{q_{x_1}}} \tag{5.91}$$

Case 1: Slope $= \infty$; $y_2 = \hat{\sigma}_0\sqrt{q_{x_2}}$

It follows from the denominator of (5.91):

$$\frac{\hat{\sigma}_0\sqrt{q_{x_2}}\,\rho_{x_1x_2}}{\sqrt{q_{x_1}q_{x_2}}} - \frac{y_1}{q_{x_1}} = 0$$

or $$y_1 = \hat{\sigma}_0\sqrt{q_{x_1}}\,\rho_{x_1x_2} = \hat{\sigma}_{x_1}\rho_{x_1x_2}$$

Case 2: Slope $= 0$; $y_1 = \hat{\sigma}_0\sqrt{q_{x_1}}$

It follows from the numerator of (5.91):

$$y_2 = \hat{\sigma}_0\sqrt{q_{x_2}}\,\rho_{x_1x_2} = \hat{\sigma}_{x_2}\rho_{x_1x_2}$$

We observe that the ellipse becomes narrower the larger the correlation coefficient. If the correlation is one (linear dependence of the two parameters), the ellipse of standard deviation degenerates into the diagonal of the rectangle. The ellipse becomes a circle if a = b or $\sigma_{x_1} = \sigma_{x_2}$ and $\rho_{x_1x_2} = 0$.

APPENDIX A

General Matrix Inverses (Minimal and Inner Constraint)

1.) Generalized Inverses

Consider the system of n equations with u parameters

$$ {}_nA_u x = \ell \qquad (A1) $$

The rank of the A matrix is R(A) = r with r < min (u,n); i.e. A has less than full column rank. The general solution of (A1) can be written as

$$ x = G\ell + (I - GA)z \qquad (A2) $$

where z is an arbitrary uxl vector. The matrix G is called the generalized inverse of A. It has the property

$$ {}_nA_{uu}G_{nn}A_u = A \qquad (A3) $$

The matrix G is not unique; there are several G's which fulfill (A2) and (A3). $G\ell$ is the particular solution of (A1) whereas (I - GA)z is the solution of the homogeneous system Ax = 0 since A(I - GA) = A - AGA = 0.

The number of independent solutions in (A2) is equal to the rank defect u - r. The matrices I - GA and GA are idempotent, as can easily be verified with the help of (A3). Thus

$$ R(I - GA) = Tr(I - GA) = u - Tr(GA) = u - R(GA) = u - r $$

This follows from $R(GA) \leq R(A)$ and $R(GA) \geq R(A)$ because A(GA) = A. For an arbitrary z vector there are, thus, u-r linearly independent vectors in (I - GA)z.

A special generalized inverse of A is the pseudo inverse A^+. This inverse is unique. It fulfills the conditions (defining conditions):

$$\left.\begin{aligned} AA^+A &= A \\ A^+AA^+ &= A^+ \\ (AA^+)^T &= AA^+ \\ (A^+A)^T &= A^+A \end{aligned}\right\} \qquad (A5)$$

Let the u x u matrix G now be the generalized inverse of the symmetric matrix A^TA. Instead of (A3) we then have

$$A^TAGA^TA = A^TA \qquad (A6)$$

Transposing (A6) we see that G^T is also a generalized inverse of A^TA. We further find

$$AGA^TA = A \quad \text{and} \quad A^TAGA^T = A^T \qquad (A7)$$

In order to verify these relations let's write:

$$\begin{aligned} E^TE &\equiv [AGA^TA - A]^T[AGA^TA - A] \\ &= [A^TAG - I]A^T[AGA^TA - A] \\ &= [A^TAG - I][A^TAGA^TA - A^TA] \\ &= [A^TAG - I][A^TA - A^TA] \\ &= 0 \end{aligned}$$

If $E^TE = 0$ then $E = 0$ and thus $AGA^TA - A = 0$ which proves the first relation in (A7). The second relation is proven similarly.

Let G and F be two different generalized inverses of A^TA. Then according to expression (A7) we have

$$AGA^TA = AFA^TA$$

Multiplying this relation from the right by GA^T then

$$AGA^TAGA^T = AFA^TAGA^T$$

becomes

$$AGA^T = AFA^T$$

because of (A7). Thus, the matrix

$$AGA^T \tag{A8}$$

is invariant with respect to the choice of the inverse G. Recall that G is the generalized inverse of A^TA and not of A in this context.

The matrix $G_1 = GA^TAG^T$ is also a generalized inverse of A^TA, since together with (A7) we have

$$A^TA(GA^TAG^T)A^TA = A^TA \tag{A9}$$

Let N^+ be the pseudo inverse of $N = A^TA$; then the relations (A5) become

$$\left.\begin{aligned} A^TAN^+A^TA &= A^TA \\ N^+A^TAN^+ &= N^+ \\ (A^TAN^+)^T &= A^TAN^+ \\ (N^+A^TA)^T &= N^+A^TA \end{aligned}\right\} \tag{A10}$$

Transposing these relations we observe that $(N^+)^T$ is a pseudo inverse of A^TA; therefore, since the pseudo inverse is unique,

$$(N^+)^T = N^+ \tag{A11}$$

i.e. the matrix is symmetric.

2.) Estimable Functions of the Parameters.

Consider the singular system of normal equations

$$A^TA\hat{X} = -A^TL \tag{A12}$$

which follows from the adjustment model

$$E(L) = -AX$$

$$\Sigma_L = \sigma^2 I \tag{A13}$$

Let the rank of the design matrix be R(A) = r < u < n. The fact that we have assumed a diagonal variance-covariance matrix for the observations does not impair the generality of this discussion since the case above can always be achieved by linear transformation. Let the u x u matrix G be a generalized inverse of A^TA such that (A6) holds. With (A2) the particular solution becomes

$$\hat{X}_g = -GA^TL \tag{A14}$$

Since G is not unique the estimate $\hat{X}_g$ is not unique either. It depends on the selection of G. Furthermore

$$E(\hat{X}_g) = -GA^TE(L) = GA^TAX \neq X \tag{A15}$$

since $GA^TA \neq I$ in general. Thus $\hat{X}_g$ is a <u>biased</u> estimate of X. However, $\hat{X}_g$ is an unbiased estimate of the linear function $X_g = GA^TAX$ of the

parameters since

$$E(\hat{X}_g) = E(GA^TAX) = GA^TAX = X_g \tag{A16}$$

In general, a linear function c^TX of the parameters is unbiasedly estimable if a vector d exists such that $c^TX = E(-d^TL)$. With equation (A13) we obtain for the vector c

$$c^T = d^TA \tag{A17}$$

A necessary and sufficient condition for the function c^TX to be unbiasedly estimable is

$$c^TGA^TA = c^T \tag{A18}$$

If c^TX is estimable then $c^T = d^TA$ and

$$c^TGA^TA = d^TAGAA^T = d^TA = c^T$$

The estimable function $x_g = c^TX$ is determined uniquely, since with expressions (A14) and (A17)

$$c^TX_g = d^TAX_g = -d^TAGA^TL$$

is invariant with respect to the selection of G according to (A8). Further, the unbiased estimate is

$$\hat{x}_g = c^T\hat{X}_g \tag{A19}$$

since

$$E(\hat{x}_g) = c^TE(\hat{X}_g) = c^TGA^TAX = d^TAGA^TAX = d^TAX = c^TX \tag{A20}$$

For the variance we obtain:

$$V(x_g) = E\{[(c^T GA^T L - E(c^T GA^T L)][c^T GA^T L - E(c^T GA^T L)]^T\}$$

$$= \sigma_o^2 c^T GA^T AG^T c = \sigma_o^2 c^T GA^T AG^T A^T d$$

$$= \sigma_o^2 c^T GA^T d = \sigma_o^2 d\, AGA^T d \qquad (A21)$$

$$= \sigma_o^2 c^T Gc$$

With (A8) it follows that $V(\hat{x}_g)$ is also independent of the choice of G. It can, furthermore, be shown that the variance of the estimable function is minimal. Examples of estimable functions are angles and distance ratios in the case of a horizontal network in which the parameters X are the Cartesian coordinates. See Grafarend and Schaffrin (1974).

It can easily be seen that the particular vector

$$X_g = GA^T AX \qquad (A22)$$

is an unbiasedly estimable function of X since each of its components is unbiasedly estimable. The necessary and sufficient condition (A18) is fulfilled:

$$GA^T AGA^T A = GA^T A$$

The unbiased estimate X_g is (A14).

Out of the many possibilities for finding unbiasedly estimable functions by the transformation (A22) we pick the one in which the covariance matrix has a minimal trace. This is accomplished by the pseudo inverse N^+. The proof follows later. The unbiasedly estimable function of the parameter X is

$$X_p = N^+ A^T A X \tag{A23}$$

with N^+ being the pseudo inverse of $N = A^T A$. The unbiased estimate of X_p is

$$\hat{X}_p = -N^+ A^T A N^+ A^T L = -N^+ A^T L \tag{A24}$$

And the variance-covariance matrix is

$$\Sigma_{X_p} = \sigma_o^2 N^+ \tag{A25}$$

Thus, the use of the pseudo inverse in solving the singular normal equations (A12) leads directly to an unbiased estimate of the linear function X_p. It will be necessary to investigate the meaning of the linear function X_p for each application.

The estimate $\hat{X}_p$ is distributed as

$$\hat{\tilde{X}}_p \sim N_r(-N^+ A^T A X, \sigma_o^2 N_u N_u^+)$$

according to (A24) and (A25). The rank of N^+ is equal to r. Analagous to the transformations performed in Section 5 we can find a linear transformation

$$_r\tilde{Y}_1 = J\hat{\tilde{X}}_p$$

such that the r components of $\tilde{Y}$ are stochastically independent and distributed as n(0,1), and that the quadratic form

$$(\hat{\tilde{X}}_p - N^+A^TAX)^TN(\hat{\tilde{X}}_p - N^+A^TAX) = \tilde{\tilde{Y}}^T\tilde{\tilde{Y}}$$

remains invariant. It follows that

$$\boxed{\frac{(\hat{\tilde{X}}_p - N^+A^TAX)^TN(\hat{\tilde{X}}_p - N^+A^TAX)}{\sigma_o^2} \sim \chi^2_r} \qquad \text{(A26)}$$

Using the relations (A10) for the pseudo inverse the expression (A26) can be readily simplified as to

$$\frac{1}{\sigma_o^2}(\hat{\tilde{X}}_p - X)^TN(\hat{\tilde{X}}_p - X) \sim \chi^2_r \qquad \text{(A27)}$$

This expression is, of course, identical to (5.41) as applied to the case of a rank deficiency.

3.) Minimal Constraint and Inner Constraint.

Generalized inverses, including the pseudo inverse, can be derived with the help of a minimal number of conditions which are necessary to avoid the singularity of the normal equations. Consider the following adjustment problem:

$$E(L) = -AX_b$$

$$BX_b = 0$$

$$V(L) = \sigma_o^2 I \qquad \text{(A28)}$$

with $R({}_nA_u) = r < u < n$

The augmented normal matrix

$$\begin{pmatrix} A^TA & B^T \\ B & 0 \end{pmatrix}^{-I} = \begin{pmatrix} \Sigma_b & P^T \\ P & R \end{pmatrix} \tag{A29}$$

is non-singular for every $(u - r)$ x u matrix B which has full row rank and whose rows are linearly independent of A. In that case the rows of B are also linearly independent of A^TA. The matrix $\begin{pmatrix} A^TA \\ B \end{pmatrix}$ has, thus, full row rank and the normal matrix (A29) has a full rank since the columns of B^T are linearly independent of A^TA.

From the theorem of Section 1 we know that there exists a non-singular matrix M,

$${}_uM_u = ({}_uF_r \; {}_uE^T_{u-r}) \tag{A30}$$

such that

$$M^T(A^TA)M = \begin{pmatrix} {}_rF^T(A^TA)F_r & 0 \\ 0 & 0 \end{pmatrix} \tag{A31}$$

where the non-zero submatrix $F^T(A^TA)F$ is of size and rank r. The relation (A31) implies

$$E(A^TA)E^T = 0 \tag{A32}$$

which, in turn, implies

$$AE^T = 0 \tag{A33}$$

The matrix E^T spans the Null space of the design or the normal matrix since

$$A^T A E^T = 0 \tag{A34}$$

Having a particular solution X_b of (A28) the (u - r) linearly independent general solutions of the singular normal equation system can be written as

$$\bar{X} = X_b + E^T f \tag{A35}$$

where f is an (u - r) x 1 arbitrary vector. With the help of the matrix E all possible solutions in the parameters can be expressed based on one and the same set of observations.

If the rows of B in (A29) are linearly independent of A, then the matrix BE^T has a full rank, i.e.

$$R(BE^T) = u - r \tag{A36}$$

With the help of (A36) the inverse in (A29) can now be computed. From

$$\begin{pmatrix} A^T A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \Sigma_b & P^T \\ P & R \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

we obtain

$$\begin{aligned} &\text{a)} \quad A^T A \Sigma_b + B^T P = I \\ &\text{b)} \quad A^T A P^T + B^T R = 0 \\ &\text{c)} \quad B \Sigma_b = 0 \\ &\text{d)} \quad B P^T = I \end{aligned} \tag{A37}$$

Multiplying a) by E from the left and using $EA^T = 0$ we get

$$P = (EB^T)^{-1} E$$

The latter expression fulfills equation d). Using the above expression it follows from b):

$$A^T A E^T (B E^T)^{-1} + B^T R = 0$$

or

$$B^T R = 0$$

Since B has full rank it follows that R = 0. Thus,

$$\begin{pmatrix} A^T A & B^T \\ B & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_b & E^T (B E^T)^{-1} \\ (E B^T)^{-1} E & 0 \end{pmatrix} \tag{A38}$$

Substituting the expression for P in relation a) of (A37) gives

$$\boxed{A^T A \Sigma_b = I - B^T (E B^T)^{-1} E \equiv T_b} \tag{A39}$$

or

$$(A^T A + B^T B)\{\Sigma_b + E^T (B E^T)^{-1} (E B^T)^{-1} E\} = I$$

The validity of this expression follows from (A39) and (A37). It can be solved for Σ_b:

$$\boxed{\Sigma_b = (A^T A + B^T B)^{-1} - E^T (E B^T B E^T)^{-1} E} \tag{A40}$$

Thus, the least squares solution of $\hat{X}_b$ subject to the condition $BX_b = 0$ is

$$\boxed{\hat{X}_b = -\Sigma_b A^T L} \tag{A41}$$

The variance-covariance matrix of the parameters follows from the law

of propagating variances:

$$\Sigma_{X_b} = \sigma_o^2 \Sigma_b A^T A \Sigma_b$$

$$= \sigma_o^2 \Sigma_{\hat{b}} \qquad \text{(A42)}$$

The latter part of (A42) follows from (A39) upon multiplying from the left by Σ_b. We also find that $A^T A \Sigma_b A^T A = A^T A$. Thus Σ_b is a symmetric reflexive inverse of $A^T A$. Note that the products $A^T A \Sigma_b$ and $\Sigma_b A^T A$ are not symmetric. Hence Σ_b is not a pseudo inverse. T_b is called the transformation matrix since $X_b = T_b{}^T X$.

We have imposed u - r conditions on the parameters in order to solve the least squares problem, i.e. the minimal number of conditions is equal to the rank defect of the design (or normal) matrix. Any solution derived in this manner is called a minimal constraint solution. There are obviously several different sets of minimal constraints possible for one and the same adjustment. The only prerequisite on the B matrix is that it has full row rank and that its rows are linearly independent of A.

The estimate $\hat{X}_g$ from (A14), which is a function of the reflexive generalized inverse and the estimate $\hat{X}_b$ of (A41), which depends on the conditions, agree. Therefore, the transformation (A22) of the non-estimable parameters X into the estimable functions X_g corresponds to the introduction of the conditions $BX_b = 0$ in (A28).

The transformation to other minimal constraints is straightforward. From (A14) and (A40) we conclude that for conditions $CX_c = 0$ in (A28), provided C is a (u - r) x u matrix of full rank whose rows are linearly independent of A , and provided T_c and Σ_c are defined analogously to (A39) and (A40),

$$X_c = T_c{}^T X$$

$$\hat{X}_c = -\Sigma_c A^T L \qquad \text{(A43)}$$

$$\Sigma_{X_c} = \sigma_o^2 \Sigma_c$$

Thus, we have found two linear functions of the parameters

$$X_b = T_b^T X$$

and

$$X_c = T_c^T X$$

which are estimable. From the definition (A39) of the transformation matrices T_b and T_c, and the relations

$$A^T A \Sigma_b A^T A = A^T A$$

$$\Sigma_b A^T A \Sigma_b = \Sigma_b$$

$$A^T A \Sigma_c A^T A = A^T A$$

$$\Sigma_c A^T A \Sigma_c = \Sigma_c$$

the following solutions can be readily verified:

$$T_b = T_c T_b$$

$$T_c = T_b T_c$$

and

$$\Sigma_c = T_c^T \Sigma_b T_c$$

$$\Sigma_b = T_b^T \Sigma_c T_b \qquad \text{(A44)}$$

Using these expressions we find the following direct transformation between the estimable functions

$$X_b = T_b{}^T X_c$$
$$X_c = T_c{}^T X_b \qquad \text{(A45a)}$$

The same relationship holds for the estimates

$$\hat{X}_b = T_b{}^T \hat{X}_c$$
$$\hat{X}_c = T_c{}^T \hat{X}_b \qquad \text{(A45b)}$$

with

$$\Sigma_{X_b} = \sigma_o{}^2 \Sigma_b$$
$$\Sigma_{X_c} = \sigma_o{}^2 \Sigma_c$$

The equations (A44) and (A45) show how the parameters and their variance-covariance matrices transform under minimal constraint.

Instead of using the general condition $BX_b = 0$ one can make use of the specific condition

$$EX_g = 0 \qquad \text{(A46)}$$

The rows of E are linearly independent of A because $AE^T = 0$. Thus, we can substitute in (A38) to (A40) the matrix B = E. Using the notation Σ_g instead of Σ_b we obtain:

$$\begin{pmatrix} A^TA & E^T \\ E & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_g & E^T(EE^T)^{-1} \\ (EE^T)^{-1}E & 0 \end{pmatrix} \tag{A47}$$

$$A^TA\Sigma_g = \Sigma_g A^TA = I - E^T(EE^T)^{-1}E \equiv T \tag{A48}$$

$$\boxed{\Sigma_g = (A^TA + E^TE)^{-1} - E^T(EE^TEE^T)^{-1}E} \tag{A49}$$

Since $E\Sigma_g = 0$ according to (A37) the equation (A48) gives $\Sigma_g A^TA\Sigma_g = \Sigma_g$ and $A^TA\Sigma_g A^TA = A^TA$. We also find that the matrices $A^TA\Sigma_g$ and $\Sigma_g A^TA$ are symmetric. These four conditions are necessary and sufficient for Σ_g to be the pseudo inverse of N, i.e.

$$\boxed{\Sigma_g = N^+}$$

The pseudo inverse N^+ can be computed from (A47), (A49) or (A44) according to

$$N^+ = T\Sigma_B T \tag{A50}$$

The unbiased estimates of the linear parametric function

$$X_p = -N^+A^TAX$$

become

$$\hat{X}_p = -N^{+}A^{T}L \tag{A51}$$

with the variance-covariance matrix being

$$\Sigma_{X_p} = \sigma_o^{2} N^{+} \tag{A52}$$

This special solution is called the inner constraint solution. It is obtained by computing either the pseudo inverse of the normal matrix based on available algorithms or, equivalently, by applying the condition $EX = 0$ to the parameters. Examples of the latter case are given in Leick and Tyler (1980).

4) Properties of Inner Constraint Solution:

A) Consider the general solution (A35)

$$\bar{X} = \hat{X}_b + E^{T}f \tag{A53}$$

which gives all linearly independent solutions of the singular normal equations. The system (A53) can be considered a set of observations equations. The parameters X_b are the "observations" having a unit matrix as a weight matrix; f are the parameters which are to be determined by least squares. $\bar{X}$ denotes the residuals to the coordinates. The least squares solution with

$$\bar{X}^{T}\bar{X} = \min.$$

gives

$$f = -(EE^{T})^{-1}E\hat{X}_b \tag{A54}$$

and

$$\hat{\bar{X}} = [I - E^T(EE^T)^{-1}E]\hat{X}_b \qquad (A55)$$

$$= -T\Sigma_b A^T L$$

$$= -N^+ A^T L$$

$$= \hat{X}_p$$

Note that $N^+A^T = T\Sigma_b TA^T = T\Sigma_b A^T AN^+A^T = T\Sigma_b A^T$ which follows from the respective expressions given earlier. The solution (A55) is identical to the pseudo inverse solution. Thus, if the pseudo inverse is chosen the resulting parameters have a minimum norm

$$\hat{\bar{X}}^T\hat{\bar{X}}_{min} = X_p^T X_p \leq X_b^T X_b$$

compared to any minimal constraint solution X_b.

We realize that in the case of a non-linear adjustment the adjusted parameters are a function of the approximate parameter values used in forming the normal equations since $X = X_a - X_o$. Therefore, the actual minimum of X^TX is also a function of the approximate parameter values.

In many problems in Surveying Engineering it is relatively easy to find the matrix E and to give a geometric interpretation of the estimable parameter function X_p. Examples are given in Leick and Tyler (1980).

B) TrN^+ = minimum compared to any other minimal constraint solution. From equation (A50) we get

$$TrN^+ = Tr(T\Sigma_b T)$$

$$= TrTT\Sigma_b$$

$$= TrT\Sigma_b$$

$$= Tr\Sigma_b - TrE^T(EE^T)^{-1}E\Sigma_b$$

Here we have used the fact that $TT = T$. The matrix $(EE^T)^{-1}$ can be factorized as $S^TS = (EE^T)^{-1}$. Thus

$$\begin{aligned} \mathrm{Tr}N^+ &= \mathrm{Tr}\Sigma_b - \mathrm{Tr}E^TS^TSE\Sigma_b \\ &= \mathrm{Tr}\Sigma_b - \mathrm{Tr}SE\Sigma_bE^TS^T \\ &\leq \mathrm{Tr}\Sigma_b \end{aligned}$$

since $\mathrm{Tr}SE\Sigma_bE^TS^T \geq 0$

C) The residuals are invariant with respect to the specific choice of the minimal constraint. Consider again equation (A53)

$$\bar{X} = X_b + E^Tf$$

where $\bar{X}$ represents any of the possible linearly independent solutions. Multiplying this equation from the left by A gives

$$A\bar{X} = AX_b + AE^Tf$$

or

$$A\bar{X} = AX_b$$

The product of the coefficient matrix and the parameter vector is invariant. This implies that the residuals

$$V = AX + L_b$$

are invariant with respect to the specific choice of X. Thus, L_a is also invariant.

D) Since the residuals are invariant, the quadratic norm V^TPV is also invariant with respect to the choice of the minimal constraint.

E) $\Sigma_V = \Sigma_o^2(P^{-1} - A\Sigma_x A^T)$ is invariant since $A\Sigma_x A^T$ is invariant according to (A8).

F) From E) it follows that

$$\Sigma_{L_\alpha} = \sigma_o^2 A\Sigma_x A^T$$

is invariant.

4) The Pseudo-Inverse of the Design Matrix

The least squares solution of

$$V = AX + L$$

$$\Sigma_{L_b} = \sigma_o I \qquad \text{(A56)}$$

with

$$R({}_nA_u) = r < u < n$$

is given by equation (A51) as follows:

$$\hat{X}_p = -N^+ A^T L \qquad \text{(A57)}$$

N^+ is the pseudo inverse of the normal matrix $N = A^TA$. It can be computed from the expression

$$N^+ = (A^TA + E^TE)^{-1} - E^T(EE^TEE^T)^{-1}E \qquad \text{(A58)}$$

The columns of the matrix E^T span the Null space of the design matrix A or the normal matrix A^TA.

The elements of the E-matrix can usually be found readily by

inspecting the general form of the elements of the A-matrix. Based on the verifications given previously in this Appendix, the properties of the solution (A57) can be summarized as follows

a) V^TV = minimum for all X

b) $X_p^TX_p$ = minimum for all X for which V^TV = minimum

The identical solution to (A57) can also be given in terms of the pseudo inverse of the design matrix. The solution to be discussed in the following is computationally more cumbersome than the previous one. This is so because the E-matrix can be found easily in most applications. The subsequent derivations, therefore, should be regarded as an exercise in the use of pseudo inverses. The method to be suggested will depend explicitly on the use of eigenvectors. To start the derivation let us recall the least squares solution expressed by equation (5.23), which reads in the present notation

$$-_rF_{nn}^TA_u\hat{X} = F^TL \qquad \text{(A59)}$$

The matrix F is (n x r) and its columns constitute an orthonormal basis for the column space of the design matrix A. We also note that one choice for the columns of F could be the normalized eigenvectors of AA^T. The system (A59) has various solutions each of which minimizes V^TV. A solution can simply be found by imposing u-r conditions on the parameters, e.g. $x_1 = x_2 = \ldots x_{u-r} = 0$, and solving for the remaining r parameters. Alternatively we may consider (A59) a condition equation which is to be solved under the condition

$$X^TX = \text{minimum}$$

According to the solution given in Section 3.2 for condition equation adjustments we obtain

$$\hat{X}_p = -A^T F(F^T A A^T F)^{-1} F^T L \qquad (A60)$$

The matrix $F^T A A^T F$ is of size and rank r, and, therefore, non-singular. It can readily be verified that the pseudo inverse of the design matrix is

$$A^+ = A^T F(F^T A A^T F)^{-1} F^T \qquad (A61)$$

since the four defining conditions (A5) are fulfilled. Equation (A61) shows how the pseudo inverse can be computed using the eigenvectors of AA^T.

In summary we can say that the solution of the inconsistent equation system (observation equations)

$$AX = -L$$

subject to

a) $(AX + L)^T(AX + L) = \text{minimum}$

b) $X^T X = \text{minimum}$

is

$$X = -A^+ L \qquad (A62)$$

For the sake of completeness we present here also a slightly modified form for A^+ using the singular value decomposition (Rao, 1973, p. 42). Let

$$R({}_n A_u) = r$$

then

$$R({}_u A^T A_u) = r$$

$$R({}_n A A^T{}_n) = r$$

If

$_{u}V_{r}$ is the orthonormal eigenvector matrix of A^TA

and

$_{n}F_{r}$ is the orthonormal eigenvector matrix of AA^T

then

$$V^TV = {}_{r}I_{r}$$

and

$$F^TF = {}_{r}I_{r} \tag{A63}$$

Let $_{r}\Lambda_{r}$ denote the diagonal matrix whose diagonal elements consists of the non-zero eigenvalues of A^TA or AA^T (note that both matrices have the same set of non-zero eigenvalues) then A can be factorized as follows

$$A = F\Lambda^{\frac{1}{2}}V^T$$

$$\Lambda^{\frac{1}{2}} = F^TAV \tag{A64}$$

yielding

$$A^TA = V\Lambda V^T$$

$$AA^T = F\Lambda F^T$$

The pseudo inverse of A is

$$A^{+} = V\Lambda^{-\frac{1}{2}}F^T \tag{A65}$$

It can be readily verified that the matrices (A64) and (A65) fulfill the four defining conditions (A5) for the pseudo inverse. The equations (A64) are the analog to similar expressions given in Section 1 for positive definite matrices. The identity of (A65) and (A61) can be established with the help of (A63) and (A64):

$$A^T F(F^T A A^T F)^{-1} F^T$$

$$= A^T F \Lambda^{-1} F^T$$

$$= V \Lambda^{\frac{1}{2}} F^T F \Lambda^{-1} F^T$$

$$= V \Lambda^{-\frac{1}{2}} F^T$$

We conclude that the pseudo inverse of the design matrix A can be computed from either (A65), (A61) or (A57). The first two cases involve the eigenvectors of the normal matrix where as the latter expressions, in combination of (A58) makes use of the Null space.

REFERENCES

Ayres, Frank, Jr., 1962. Matrices. Schaum's Outline Series, McGraw-Hill Book Company.

Faddeev, D.K. and Faddeeva, V.N., 1963. Computational Methods of Linear Algebra, translated from the Russian by Robert C. Williams. Freeman and Co., San Francisco.

Grafarend, E. and B. Schaffrin, 1974. "Unbiased Free Net Adjustment". Survey Review XXII, 171.

Grafarend, E. and B. Schaffrin, 1976. "Equivalence of Estimable Quantities and Invariants in Geodetic Networks". Zeitschrift fuer Vermessungswesen, No. 11.

Graybill, Franklin A., 1969. Introduction to Matrices with Applications in Statistics. Wadsworth Publishing Company, Inc., Belmont, California.

Graybill, Franklin A., 1976. Theory and Application of the Linear Model. Duxbury Press; North Scituate, Massachusetts.

Hamilton, Walter C., 1964. Statistics in Physical Science -- Estimation, Hypothesis Testing and Least Squares. Ronald Press, New York.

Koch, K.R., 1980. Parameterschätzung and Hypothesentests in linearen Modellen. Dümmler, Bonn.

Leick, Alfred and David Tyler, 1980. Lecture Notes of the Winter Institute in Surveying Engineering; University of Maine at Orono, January 14-18.

Lipschutz, Seymour, 1968. Linear Algebra. Schaum's Outline Series, McGraw-Hill Book Company.

Milhail, Edward M., 1976. Observations and Least Squares, Dun Donnelley Publisher, New York.

Pope, Allen Y., 1972. "Some Pitfalls to be Avoided in the Iterative Adjustment of Non-linear Problems". Proceedings of the 38th Annual Meeting of the American Society of Photogrammetry, March.

Pope, Allen Y., 1974. The Statistics of Residuals and the Detection of Outliers, NOAA Technical Reports NOS65 NGS1, Washington D.C.

Rao, C.R., 1973. Linear Statistical Interference and its Applications. John Wiley & Sons.

Uotila, U.A., 1967. "Introduction to Adjustment Computations with Matrices". Unpublished class notes. The Ohio State University.

Uotila, U.A., 1973. "Useful Statistics for Land Surveyors". Surveying and Mapping (ACSM), March.

Uotila, U.A., 1975. "Statistical Tests as Guidelines in Analysis of Adjustment of Control Nets". Surveying and Mapping (ACSM), March.

Vanicek, Peter, 1974. Introduction to Adjustment Calculus. Lecture Notes No. 18. Department of Surveying Engineering, University of New Brunswick, Fredericton.

Veress, S.A., 1974. Adjustment by Least Squares (Monograph). American Congress on Surveying and Mapping, Washington, D.C.

Wells, D.E. and E.Y. Krakiwsky, 1971. "The Method of Least Squares". Lecture Notes No. 18. Department of Surveying Engineering, University of New Brunswick, Fredericton.

Wolf, Paul R., 1980. Adjustment Computations. P.B.L. Publishing Co., Monona, Wisconsin.