Notes on Adjustment Computations
Part I

Based on Former Geodetic Science Courses
GS 650 and GS 651

Taught at
The Ohio State University

by Prof. Burkhard Schaffrin

by Kyle Snow

December 9, 2019

Contents

Acknowledgments

I would like to thank Dr. Shengjie Ge for his assistance with typesetting an early version with LaTeX. Also, Mr. Xiankun Wang, MSc., is acknowledged for his careful reading of the text and for pointing out several typographical errors that were corrected in the edition distributed in January, 2010. Dr. Dru Smith suggested additional formulas and the inclusion of summary formulas found at the end of most chapters beginning in 2018. Questions posed by Dr. Jihye Park resulted in improvements to the first edition posted in 2019. A thorough review by Prof. Chris Jekeli in early 2019 resulted in many improvements. Constructive comments from Prof. Clyde Goad were also helpful.

I would also like to express my thanks to the students of the Adjustment Computations course that I taught at Ohio State in the Fall Semester of 2019, during which time I gradually added 95 pages to these notes. These students had to endure my weekly updates to the notes, and their comments and question, as well as their keen eyes for my typographical errors, helped to improve the notes.

Of course, without the extraordinary teaching of Prof. Burkhard Schaffrin, these notes would not even exist.

Columbus, Ohio, December 2019 *Kyle Snow*

Introduction

This document is primarily based on notes taken by Kyle Snow in Geodetic Science adjustments courses GS 650 and GS 651 taught by Burkhard Schaffrin at The Ohio State University in 1997 and 1998. The appendix contains several matrix properties and identities used throughout the text. A bibliography at the end includes referenced material and material for suggested reading.

Notation

A few comments about the notation used in this document may be helpful. Matrices are displayed in uppercase. Vectors are lowercase and are set in bold-face type (bold face is not used for any other symbols). Scalar variables are generally lowercase. Greek letters are used for unknown, non-random parameters, while Latin letters are used for unknown, random variables. Symbols denoting estimates of non-random variables use Greek letters with a hat on top, while predictions of random variables are shown as Latin letters with tildes on top. Tables 1 and 2 list variables, mathematical operators, and abbreviations used herein.

Table 1: Variables and mathematical operators

Symbol	Description
A	coefficient (design) matrix in the Gauss-Markov Model
B	coefficient matrix in the Model of Condition Equations
$\boldsymbol{c}$	right-side vector in the system of normal equations $N\hat{\boldsymbol{\xi}} = \boldsymbol{c}$
$C\{\cdot\}$	covariance operator
$D\{\cdot\}$	dispersion operator
diag$[\cdot]$	a diagonal matrix with diagonal elements comprised of $[\cdot]$
dim	the dimension of a matrix
$\boldsymbol{e}$	unknown random error vector for the observations
$\tilde{\boldsymbol{e}}$	predicted random error (residual) vector for the observations

Continued on next page

Symbol	Description
$\boldsymbol{e}_0$	unknown random error vector associated with stochastic constraints
$\tilde{\boldsymbol{e}}_0$	predicted random error (residual) vector for $\boldsymbol{e}_0$
$E\{\cdot\}$	expectation operator
H_0	null hypothesis
H_A	alternative hypothesis
K	constraint matrix used in the Gauss-Markov Model with (stochastic) constraints
m	number of unknown parameters
$\mathrm{MSE}\{\cdot\}$	mean squared error operator
n	number of observations
N	normal-equations matrix in the system of normal equations $N\hat{\boldsymbol{\xi}} = \boldsymbol{c}$
$\mathcal{N}(\cdot)$	the nullspace (kernel) of a matrix or the normal distribution, depending on the context
P	weight matrix for the observations
P_0	weight matrix for stochastic constraints
q	rank of the coefficient (design) matrix A
Q	cofactor matrix for the observations
$Q_{\tilde{\boldsymbol{e}}}$	cofactor matrix for the predicted random errors (residuals)
r	redundancy of data model
$\mathbb{R}$	the field of real numbers
$\mathcal{R}(\cdot)$	the range (column) space of a matrix
rk	the rank of a matrix
tr	the trace of a matrix
U	matrix of eigenvectors
$\boldsymbol{w}$	constant vector in the Model of Condition Equations
$\boldsymbol{y}$	vector of observations (possibly in linearized form)
$\boldsymbol{z}$	vector of constraints used in the Gauss-Markov Model with stochastic constraints
α	significance level for statistical tests
$\boldsymbol{\alpha}$	observation coefficient vector in the Model of Direct Observations
β	a quantity associated with the power of a statistical test
χ^2	chi-square statistical distribution
δ	a small deviation or non-random error, as in δP denoting a non-random error in matrix P
Φ	Lagrange target function
$\boldsymbol{\eta}$	unit vector used in the Outlier Detection Model
$\boldsymbol{\kappa}_0$	vector of specified constants used in the Gauss-Markov Model with constraints
$\boldsymbol{\lambda}$	unknown vector of Lagrange multipliers
$\hat{\boldsymbol{\lambda}}$	estimated vector of Lagrange multipliers

Continued on next page

Symbol	Description
μ, $\boldsymbol{\mu}$	the expected value of a non-random variable, could be a scalar μ or vector $\boldsymbol{\mu}$
$\hat{\mu}$, $\hat{\boldsymbol{\mu}}$	the estimate of a non-random variable
$\hat{\boldsymbol{\mu}}_y$	vector of adjusted observations
ν	statistical degrees of freedom
θ	the orientation of a confidence ellipse
σ_0^2	unknown variance component
$\hat{\sigma}_0^2$	estimated variance component
Σ	dispersion (or covariance) matrix for the observations
$\boldsymbol{\tau}$	vector of ones (also called "summation vector")
Ω	(weighted) sum of squared residuals (unconstrained case)
$\boldsymbol{\xi}$	vector of unknown parameters
$\hat{\boldsymbol{\xi}}$	estimated parameter vector

Continued from previous page

Table 2: List of abbreviations

Abbrev.	Meaning
BLUUE	Best Linear Uniformly Unbiased Estimate
BLIP	Best LInear Prediction
cdf	cumulative distribution function
GHM	Gauss-Helmert Model
GMM	Gauss-Markov Model
LESS	LEast-Squares Solution
MSE	Mean Squared Error
pdf	probability density function
rms	root mean square

Chapter 1

Foundations: Observations, Parameters, Random Errors, and Essential Math

1.1 Parameters and Observations, Purpose of Adjustment Computations

In geodetic science, *observations* (measurements) are typically made for the purpose of estimating certain unknown quantities, for example, coordinates of GPS reference stations, or heights of benchmarks. These unknown quantities are often expressed as *parameters of an observational model*. In some cases an unknown parameter might be measured directly (say the length of a bridge), but often parameters are only "measured indirectly," for example by measuring angles and distances to determine coordinates of points. In any case, for our purposes we will consider these unknown quantities to be *fixed parameters*, rather than random parameters, which are treated in the Advanced Adjustment Computations notes. The terms fixed and random parameters refer to the statistical (stochastic) properties of the unknowns. Physically, one may think of a fixed parameter as representing a quantity that does not vary in time, at least not over the time span of interest.

While some observations can be made with the naked eye, for example by reading a tape to determine a distance between survey markers, often they are made by use of a more sophisticated instrument. Traditionally, in surveying most instruments were optical, such as a surveyor's level or transit. These instrument required scales or rods to be read with the aid of telescopes and magnifying eyepieces. Eventually, electro-optical instruments added electronic distance measuring functionality, while horizontal and vertical angles were still read by optical means. Later, with the advent of the total station, both angles and distances were measured electronically, and perhaps not even recorded manually if a data collector was used in conjunc-

tion with the instrument. Nowadays, robotic total stations, digital levels, GPS (or GNSS) receivers, and laser scanners, not to mention drones with GPS receivers, cameras, and LIDAR, remove most of the traditional elements of human observation. Nevertheless, we still refer to the quantities they measure and record (the data) as observations. *The import thing to know about observations, is that they always contain some level of unknown error*, whether they are made and recorded manually by a human, or made and recorded electronically with little or no human involvement.

Errors in observations differ by type (nature). The types we are most concerned with are random, systematic (bias), and blunders (mistakes). Blunders might belong to the statistical category of *outlier*. We will discuss these categories of errors in more detail in a later section. For now we simply assume that 1) all observations contain random errors, 2) that it is usually possible to account for systematic errors in some way (if we know they exist), and 3) that blunders must be avoided or found and removed.

Let us summarize these concepts with the following brief definitions:

Observation A measured quantity that has a numerical value and unit associated with it. Observations always contain unknown random error and might also be corrupted by systematic errors (biases) and blunders. Because of its random component, an observation is treated as a random variable.

Parameter An unknown quantity of interest that is to be estimated. Here we treat only fixed parameters, by which we mean they do not vary statistically.

Purpose of adjustment computations Let us assume for now that our observations are free of biases and blunders. They still contain unknown random errors. What's more, by design we usually have more observations than the minimum necessary to determine the unknown parameters of our observational model, if any. How then do we deal with these extra observations and their random errors? This is the subject of adjustment computations, the purpose of which is to adjust the observations in some way so that the difference between the given observations and their adjusted values (called *residuals*) is as small as possible according to a stated criterion. One particular method for doing so is the method of *least-squares adjustment*, which is the primary subject of these notes. The term "least squares" is due to the criterion of the method, which is summarized by the often-heard statement that

> *the sum of the squares of the (weighted) residuals is a minimum.*

1.2 Functional Relations and Stochastic Properties

As noted above, observations are typically made to determine the value of some unknown quantity (or quantities) of interest. In order to relate the observations with the unknown quantity, a mathematical function is specified. The function

may be linear or nonlinear depending on the complexity of the relation between the observations and the unknowns. In the case where the unknown quantity can be observed directly, a simple linear function might be suitable. In other cases, the chosen function may be highly nonlinear.

As an example, suppose a distance is measured between points p_1 and p_2 whose coordinates in the horizontal plane must be determined. In this case, the measured distance, call it y, is the observation. The the unknown quantities are the coordinate pairs of the two points, viz. (x_1, y_1) and (x_2, y_2). The functional relationship between the measured distance and unknown coordinates can be written as

$$y(x_1, y_1, x_2, y_2) \approx \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{1.1a}$$

Obviously, the function is nonlinear in the unknown variables x_1, y_1, x_2, and y_2. Note that the observation variable y is the dependent variable; it depends on the unknown coordinates, which are the independent variables of the function. The approximately-equals sign is used because the observation contains random error, and thus the unknown quantities do not fully explain the observation variable. Recall that the unknown quantities are considered to be nonrandom.

In order to change the approximately-equals sign to an equals sign, an additional term must be added to the function so that both sides of the equation have a random term (or, equivalently, the random term could be subtracted from the left side). The random term, call it e, accounts for the *unknown random error* in the observation. By introducing e, (1.1a) is then modified to read

$$y = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} + e, \tag{1.1b}$$

where the function arguments on the left side are dropped for simplicity.

Some authors would have placed the random error e on the left side of (1.1b) as a positive term. This is perhaps both a matter of convention and a matter of view point. *Adding* e to the left side says that an observation plus its random error is equal to some function of the unknown variables. We prefer to say that an observation *minus* its random error is equal to the function. Our convention seems to be predominant in the current literature, whereas the other convention shows up mostly in older publications. Besides being a matter of convention, we argue that it is more appealing mathematically to add the random error term e to the right side, because it makes the equation consistent in that an expression with random properties on the left then equals an expression with random properties on the right.

Equation (1.1b) is in the form of an (nonlinear) *observation equation*, which is what we call an equation that expresses an observation as a random variable that depends on unknown quantities that must be determined. Thus, we say that (1.1b) *models the observation* as a function of unknown variables; we call these unknowns *parameters* of the model. We want to determine (solve for) these parameters in some optimal way. As we will see later, the determination of the values of the parameters cannot be made with absolute certainty. Thus we use the statistical term *estimation* when we speak of determining numerical values for the parameters.

So, we may refer to (1.1b) as an observational model; however, it is not a complete model, because more needs be said about the stochastic nature of the random error e in order to exploit its random properties when estimating the unknown parameters. In the following sections we will discuss in some detail how the stochastic properties of random errors can be characterized. At this stage it is enough to say that only the *expectation* and the *variance* of the errors need to be specified in the model. Expectation is a statistical term that denotes the value we expect a random variable to take on, at least in an average sense. And in this context, variance is a statistical term that denotes our uncertainty about the expected value of a random variable, i.e., it puts bounds around the expected value we specify for the random variable (in this case random observation errors).

Unless otherwise noted, we will always specify the expectation of the random errors to be zero. The way the expectation and variance of the random error e is expressed mathematically is like this: $e \sim (0, \sigma^2)$, which reads, "e is distributed with zero expectation and sigma-squared variance." Thus, the observational model (1.1b) is made complete by extending it to

$$y = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} + e, \quad e \sim (0, \sigma^2). \tag{1.1c}$$

The observation equation is sometimes referred to as the functional part of the model (or functional model), while the statement $e \sim (0, \sigma^2)$ is sometimes referred to the stochastic part of the model (or stochastic model). We call the inverse of the variance the *weight* of the observation (weights are discussed in more detail in a later section). Note that the unit of measurement of the random error e is the same as that of the observation y, and the unit of measurement of the variance σ^2 is the square of the observation's unit.

The observational model (1.1c) is relevant to a particular problem, that is, to a particular type of observation (an observed distance) and to particular parameters (coordinates of two points). We would rather generalize it for use in a wide variety of geodetic adjustment problems. For that we collect the unknown parameters in the $m \times 1$ vector $\boldsymbol{\xi}$, the symbol used to represent a vector of m unknown parameters throughout these notes. Furthermore, (1.1c) contains only one observation; it must be generalized to handle any number of observations, possibly all with their own unique variances. For this we make use of matrices and vectors.

Suppose rather than a single observation y we are given an $n \times 1$ vector of observations $\boldsymbol{y} = [y_1, \ldots, y_n]^T$, which has an associated, unknown vector of random errors $\boldsymbol{e} = [e_1, \ldots, e_n]^T$. Our general model should allow each of the random errors to have its own unique variance, and it should allow for covariances between the random errors (covariances are defined in Section 1.5.2). Thus an $n \times n$ *cofactor matrix* Q is introduced, with its inverse $P \coloneqq Q^{-1}$ called weight matrix. When Q is multiplied by an unknown scalar σ_0^2 called *variance component*, the result is called *covariance matrix*, which is denoted by the symbol Σ, i.e., $\Sigma \coloneqq \sigma_0^2 Q = \sigma_0^2 P^{-1}$. Note that some authors call the variance matrix "variance-covariance matrix," and some authors call the variance component the "variance of unit weight." Putting these components together results in the following model:

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{\mathbb{R}^m \to \mathbb{R}^n}{\boldsymbol{f}(\boldsymbol{\xi})} + \underset{n\times 1}{\boldsymbol{e}}, \quad \boldsymbol{e} \sim (\underset{n\times 1}{\mathbf{0}}, \sigma_0^2 \underset{n\times n}{P^{-1}}). \tag{1.2a}$$

Note that the vector of functions $\boldsymbol{f}$ maps $\mathbb{R}^m$ into $\mathbb{R}^n$, denoted mathematically by $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^n$.

Now, if the vector of functions $\boldsymbol{f}$ is nonlinear in the unknown parameters $\boldsymbol{\xi}$, it can be linearized by a truncated Taylor series expansion (see Appendix B). Whether we have a linear form $\boldsymbol{f}(\boldsymbol{\xi})$ or a linearized form, we can represent it by an $n \times m$ *coefficient matrix* A, so that the model (1.2a) can be restated as

$$\boxed{\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \underset{n\times 1}{\boldsymbol{e}}, \quad \boldsymbol{e} \sim (\underset{n\times 1}{\mathbf{0}}, \sigma_0^2 \underset{n\times n}{P^{-1}}).} \tag{1.2b}$$

The development of the model (1.2b) is an important step in understanding the relations between observations, parameters, and random errors. The model is of type *Gauss-Markov*, which is an important model in geodetic science and one that is used extensively in Chapter 3, with particular extensions of it as the focus of Chapter 5 and Section 3.5. More details about random errors, covariances, and weights will follow, and the usefulness of model (1.2b) will become more apparent in later chapters. For now, we summarize with a basic description of each element of the model.

$\boldsymbol{y}$ is a *given* $n \times 1$ vector of observations.

A is a *given* $n \times m$ coefficient matrix that has full column rank, i.e, $\operatorname{rk} A = m$.

$\boldsymbol{\xi}$ is an $m \times 1$ vector of *unknown* parameters.

$\boldsymbol{e}$ is an $n \times 1$ vector of *unknown* random errors associated with the observations.

σ_0^2 is an *unknown* variance component (scalar quantity). Note that σ_0^2 is unitless.

P is an $n \times n$ weight matrix such that $P^{-1} \coloneqq Q$ for a *given* cofactor matrix Q, and where the covariance matrix Σ is defined as $\Sigma \coloneqq \sigma_0^2 P^{-1}$. Note that the diagonal elements of Q have units that are the square of the units of their associated observations.

1.3 Fundamentals of Matrix Algebra

Matrix algebra (or linear algebra) is fundamental to the mathematics of adjustment computations, and it is used extensively in these notes. Most of the concepts in matrix algebra used here are covered in a first course in linear algebra at the college or university level. Beyond that, there are many derivations in the chapters that follow that make use of certain matrix relations and identities involving inverses of sums and products of matrices that generally do not appear in a first course on linear algebra. These relations are helpful both for reducing complicated formulas to simpler forms and for showing alternative, but equivalent, solutions to the same problem. (Seeing more than one solution to a problem may help to provide greater insight into it, and we will find that sometimes one formula may be more or less efficient than another equivalent one depending on the problem at hand.)

A list of matrix relations and identities used in these notes is provided in Appendix A. The ones involving only a single line should be memorized. While it's not necessary to memorize the multi-line formulas to read these notes well, being able to recognize them or readily refer to them will make some of the derivations in later chapters easier to follow. To facilitate reading of the text, their equation numbers are usually referred to when they are used.

1.3.1 Important Concepts

Below is a list of the minimum concepts of linear algebra that the reader should be familiar with. Some are described briefly in the paragraphs that follow. These books are good sources for more complete descriptions: Strang (2006); Strang and Borre (1997).

- Gaussian elimination and back substitution
- Gauss-Jordan elimination
- The column space of a matrix
- The nullspace of a matrix
- The basis and dimension of a vector space
- The rank of a matrix
- Consistent and inconsistent systems of equations
- Eigenvalues and eigenvectors
- The properties of an invertible matrix
- The terms positive definite and positive semidefinite
- The term idempotent
- Choleskey's decomposition
- All other items in Appendix A

Vector spaces The space $\mathbb{R}^n$ consists of all vectors with n components. Two important vector spaces in adjustment computations are the column space and the nullspace of a matrix.

A basis of a vector space *A basis for a vector space is a sequence of vectors that are linearly independent and that span the space.* A vector space may have many different bases, but given a basis, every vector in the space can be expressed as a unique linear combination of the basis vectors. All bases for a vector space contain the same number of vectors. This number is the dimension of the space. The columns of any invertible $m \times m$ matrix provide a basis for $\mathbb{R}^m$.

Column space *The column space of a matrix A consists of all linear combinations of its columns.* It is denoted by $\mathcal{R}(A)$ and is also called the range of A. Its dimension equals the rank of A, which is also the number of linearly independent columns in the space. We say that the columns of A *span* the column space of A.

Note that the column space of the matrix product AB is contained in the column space of A, denoted mathematically by

$$\mathcal{R}(AB) \subset \mathcal{R}(A). \tag{1.3}$$

In words, it means that every column of the matrix product AB is a linear combination of the columns of A.

Nullspace *The nullspace of A consists of all solutions to $A\boldsymbol{x} = \boldsymbol{0}$.* It is denoted by $\mathcal{N}(A)$ and is also called the kernel of A. The dimension of the nullspace of A is the number of nonzero vectors in the space. Its dimension is

$$\dim \mathcal{N}(A) = m - \operatorname{rk} A, \text{ if } A \text{ has } m \text{ columns.} \tag{1.4}$$

This dimension is also called the nullity. If A is a square, nonsingular matrix, the only vector in its nullspace is $\boldsymbol{x} = \boldsymbol{0}$, and thus the dimension of its nullspace is zero.

The relationship between the dimensions of the column space and nullspace is given by

$$\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = \dim \mathbb{R}^m = m, \text{ if the size of } A \text{ is } n \times m. \tag{1.5}$$

The rank of a matrix The rank of a matrix A is the number of its independent rows, which is also the number of its independent columns.

Consistent and inconsistent systems of equations *A consistent system of equations is one that is solvable.* The equation $A\boldsymbol{x} = \boldsymbol{b}$ is only consistent if $\boldsymbol{b}$ is in the column space of A. For example, the equation in (1.2b) would not be consistent if the random error vector $\boldsymbol{e}$ were removed from it. That is because the observation vector $\boldsymbol{y}$ is not in the column space of the coefficient matrix A. Without $\boldsymbol{e}$, (1.2b) would be inconsistent.

Properties of an invertible matrix *A matrix A is invertible if there exists a matrix A^{-1} such that*

$$A^{-1}A = I \text{ and } AA^{-1} = I. \tag{1.6}$$

Only square matrices are possibly invertible. If matrix A is invertible:

- It is nonsingular (regular).
- Its inverse is unique.
- Its rank is equal to its dimension (size), i.e., $\operatorname{rk} A = m$ if the size of A is $m \times m$.
- Its rank is equal to the dimension of its column space, i.e., $\operatorname{rk} A = \dim \mathcal{R}(A)$.
- The vector $x = \mathbf{0}$ is the only vector in its nullspace. Therefore, $\dim \mathcal{N}(A) = 0$.
- All its eigenvalues are nonzero.

Positive definite and positive semidefinite matrices

Positive definite *A matrix A is positive definite if $\boldsymbol{x}^T A\boldsymbol{x} > 0$ for all nonzero vectors $\boldsymbol{x}$.* A positive definite matrix is nonsingular. All of its eigenvalues are greater than zero. If the matrix is also symmetric, it can be factored by the Cholesky decomposition. See page 25 for properties of a positive-definite matrix.

Positive semidefinite *A matrix A is positive semidefinite if $\boldsymbol{x}^T A\boldsymbol{x} \geq 0$ for all nonzero vectors $\boldsymbol{x}$.* A positive semidefinite matrix is singular. At least one of its eigenvalues is zero; the rest are greater than zero.

Idempotent matrices *An idempotent matrix equals its own square.* It is a square matrix, and it is singular unless it is the identity matrix.

$$\text{The } n \times n \text{ matrix } P \text{ is idempotent if } PP = P. \tag{1.7a}$$
$$\text{If the } n \times n \text{ matrix } P \text{ is idempotent, then so is } I_n - P. \tag{1.7b}$$
$$\text{If } P \text{ is idempotent, } \operatorname{tr} P = \operatorname{rk} P. \tag{1.7c}$$
$$\text{The eigenvalues of an idempotent matrix are 0 or 1.} \tag{1.7d}$$
$$\text{Projection matrices are idempotent.} \tag{1.7e}$$

1.3.2 Practice Problems

The reader should know how to solve the following problems:

1. Solve the following system of equations by Gaussian elimination and back substitution:

$$x_1 + 3x_2 - 2x_3 \qquad + 2x_5 \qquad = \quad 0$$

$$\begin{aligned} 2x_1 + 6x_2 - 5x_3 - \ 2x_4 + 4x_5 - \ 3x_6 &= -1 \\ 5x_3 + 10x_4 \qquad\quad + 15x_6 &= \ 5 \\ 2x_1 + 6x_2 \qquad + \ 8x_4 + 4x_5 + 18x_6 &= \ 6 \end{aligned}$$

2. Solve the preceding system of equations by Gauss-Jordan elimination.

3. Find a basis for the column space of

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 3 & 2 & 5 & 1 \\ 0 & 4 & 4 & -4 \end{bmatrix}.$$

4. Find a basis for the row space of

$$B = \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \\ 2 & -5 & -3 & -2 & 0 \\ 0 & 5 & 15 & 10 & 0 \\ 2 & 6 & 18 & 8 & 6 \end{bmatrix}.$$

5. Find a basis for the nullspace of matrix A and a basis for the row space of matrix B above. Confirm that the basis vectors in these nullspaces are orthogonal to the column space of A and the row space of B, respectively (see (A.19a)–(A.19d)).

6. What are the ranks of matrices A and B above.

7. Find the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 0 \end{bmatrix}.$$

   For each matrix, check that the sum of its eigenvalues equals its trace and that the product of its eigenvalues equals its determinant.

8. Compute the Cholesky factor of

$$N = \begin{bmatrix} 2 & 0 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 2 & 1 \\ 0 & -1 & 0 & 1 & 2 \end{bmatrix}$$

   and then compute the inverse of N using the Cholesky factor.

9. Assuming the partitioned matrix

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}$$

is nonsingular and that its sub-matrices N_{11} and N_{22} are also nonsingular, without referring to Appendix A, derive its inverse using elementary row operations.

10. If N is an $m \times m$ nonsingular matrix and K is an $l \times m$ matrix with $l < m$ such that $\text{rk}[N \mid K^T] = m$, then the matrix

$$\begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}$$

is nonsingular and so is the matrix $(N + K^T K)$. Note that 0 denotes an $l \times l$ matrix of zeros. Without reference to Chapter 5, derive its inverse using elementary row operations. Hint: start by multiplying the bottom row on the left by K^T and add the result to the top row.

11. With reference to Appendix A, derive (A.6b) from (A.6a).

12. If N is a nonsingular matrix defined by $N \coloneqq A^T P A$, show that the quantity $I - AN^{-1}A^T P$ is idempotent (see (1.7a)), where I is the identity matrix.

13. If the matrix P is idempotent, show that $I - P$ is too.

14. Can the dimension of the nullspace of a rectangular matrix ever be zero? Why or why not?

1.4 Random Variables

From here to the beginning of Section 1.5 we use notation consistent with textbooks in statistics for easy comparison to them. Accordingly, we use X to denote a random variable and x to denote a numerical value that the random variable could take on. After these sections, we resume use of notation consistent with the rest of these notes.

1.4.1 Review from statistics

According to Mikhail and Ackermann (1982), probabilities are associated with statistical events, which are the outcomes of statistical experiments. If an event has several possible outcomes, we associate with it a *stochastic or random variable* X, which can take on different numerical values x for different outcomes. The total of all possible outcomes of a statistical event associated with a random variable is called the *population*. Because of its large size, it is not practical, or even possible, to evaluate all the elements of a population. For this reason, we only select a small

number of them (by making observations), the set of which is called a *sample* of the population.

Let's associate these abstract statistical terms with a concrete example from geodetic science. Suppose the coordinates of a geodetic network are to be determined from data collected by GPS receivers. The act of collecting and processing those data is the experiment. The outcome is a set of coordinate differences between points in the network, which we take to be observations in this example. These coordinate differences could take on different values (i.e., no two experiments are likely to produce the same set of values). Therefore, each observed coordinate difference is considered to be a realization of a random variable. Obviously, we cannot obtain the entire population of observed coordinate differences among the network points, because there are an infinite number of them. Rather we must settle for a finite number of observations obtained from the experiment, which constitutes a sample.

Quoting Mikhail and Ackermann, "the total set of possible values of a random variable, X, together with their probabilities, constitute what is termed a *probability distribution* associated with the random variable." A probability distribution involves a function that assigns a probability to all possible values of the random variable it is associated with. The two types of probability distribution functions are *cumulative distribution function* and *probability density function*. These two distribution functions are defined in the following two sections for a single random variable (univariate case).

In general, the properties of cumulative distribution functions hold for both continuous and discrete random variables. However, probability density functions pertain only to continuous functions. Their discrete analog is the probability mass function. In the following, we will limit our discussion to continuous random variables. An important property of a continuous random variable is that the probability that it will take any particular value is zero. That is,

$$P\{X = x\} = 0 \tag{1.8}$$

for any number x, if X is a continuous random variable.

1.4.1.1 Cumulative Distribution Function

The cumulative distribution function, $F(x)$, gives the probability of the event $\{X \leq x\}$ for every number x, is given by

$$F(x) = P\{X \leq x\} = P\{-\infty < X \leq x\}. \tag{1.9}$$

In words, (1.9) says that the probability that the random variable X will take on a numerical value less than or equal to x is given by the function $F(x)$. By definition, the probabilities are limited to values between 0 and 1, i.e,

$$0 \leq P \leq 1, \text{ implying that } \lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1. \tag{1.10}$$

Finally, for any number x,

$$P\{x < X\} = 1 - F(x). \tag{1.11}$$

1.4.1.2 Probability Density Function

A probability density function $f(x)$ of a continuous random variable X provides a means to calculate the probabilities of *intervals* of numbers. It does not, however, give the probability that X equals a specific number x, because $P\{X = x\} = 0$ for all numbers x as stated in (1.8). The probability that X belongs to an interval $[a, b]$ is given by the integral

$$P\{a \leq X \leq b\} = \int_a^b f(x)\,dx. \tag{1.12}$$

It is the area under the curve $f(x)$ between a and b as shown in Figure 1.1.

If the lower limit a is replaced by $-\infty$, then a relationship between the cumulative distribution function and the probability density function can be written as

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \tag{1.13}$$

which, considering the Fundamental Theorem of Calculus, leads to the relation

$$\frac{d}{dx}F(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)\,dt = f(x), \tag{1.14}$$

for all values of x. It is important to note that a probability density function must satisfy the following two properties:

$$f(x) \geq 0 \text{ for all numbers } x. \tag{1.15a}$$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \tag{1.15b}$$

Any integrable function with satisfies these two properties is the probability density function of some random variable X.

It is also noted that, as a consequence of (1.8), $P\{X = a\} = 0$ and $P\{X = b\} = 0$, and therefore

$$P\{a \leq X \leq b\} = P\{a < X \leq b\} = P\{a \leq X < b\} = P\{a < X < b\}. \tag{1.16}$$

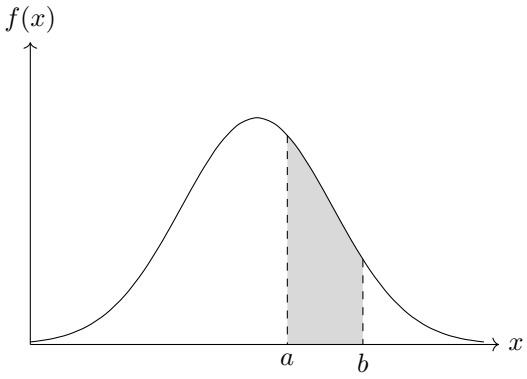

Figure 1.1: The shaded area under the curve of the density function $f(x)$ is the probability that a random variable takes on values in the interval $[a, b]$

1.4.2 Distributions for Adjustment Computations

The four distributions discussed in these notes are

1. The Gaussian or normal distribution.
2. The t (student) distribution.
3. The χ^2 distribution.
4. The F (Fisher) distribution.

For our purposes, these distributions are primarily used for hypothesis testing to validate statistically the results of various adjustment computations. Standard texts in statistics can be consulted for obtaining critical values of the distributions from tables. More details about these distributions can be found in Chapter 9. Here, we briefly describe the normal and standard normal distributions.

1.4.2.1 The Normal Distribution

The probability density function $f(x)$ of the normal distribution is defined by the equation

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty. \tag{1.17}$$

The parameters of the function are the mean $\mu(-\infty < \mu < \infty)$ and the variance $\sigma^2(\sigma^2 > 0)$. The graph of $f(x)$ is a bell-shaped curve that is symmetric about μ and that extends over the entire horizontal axis. The shorthand notation for indicating that a random variable X has a normal distribution with mean μ and variance σ^2 is

$$X \sim \mathcal{N}(\mu, \sigma^2). \tag{1.18}$$

Because $f(x)$ is symmetric about μ and reaches its maximum at $x = \mu$, the mean of the normal distribution is equal to its median and mode.

1.4.2.2 The Standard Normal Distribution

To avoid having to generate probability tables for many values of μ and σ, the random variable X is transformed to a standardized form, which can be done for any X by the transformation

$$Z = \frac{X - \mu}{\sigma}. \tag{1.19}$$

The resulting *standardized random variable* Z has mean $\mu_Z = 0$ and variance $\sigma_Z^2 = 1$. It expresses the distance of X from its mean μ in units of its standard deviation σ, as shown in Figure 1.2. Its probability density function is defined by the equation

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}. \tag{1.20}$$

In summary, we state that

> If the random variable X has a normal distribution with mean μ and variance σ^2, the the standardization $Z = (X - \mu)/\sigma$ of X has the standard normal distribution; i.e., $Z \sim \mathcal{N}(0, 1)$.

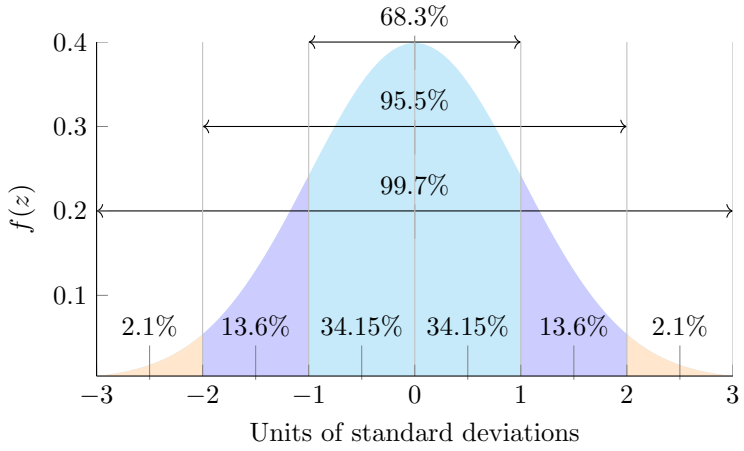

Figure 1.2: Normal distribution curve, with percent of areas under curve denoting probabilities. Image derived from Ti*k*Z code by John Canning, Senior Lecturer at the University of Brighton (http://johncanning.net/wp/?p=1202).

The reason that the curve in Figure 1.2 appears to peak near 0.4 is because $f(z = 0) = 1/\sqrt{2\pi} \approx 0.4$. The probabilities shown in the figure (as percentages) are due to the probability statements

$$P(-1 < z < 1) = P(\mu - \ \sigma < x < \mu + \ \sigma) = 0.683, \tag{1.21a}$$

$$P(-2 < z < 2) = P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.955, \tag{1.21b}$$
$$P(-3 < z < 3) = P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.997. \tag{1.21c}$$

The intervals associated with these probability statements are commonly referred to as the "1-sigma," "2-sigma," and "3-sigma" confidence intervals, respectively. Other commonly used intervals are the so-called 50%, 90%, 95%, and 99% confidence intervals. Their respective probability statements are given by

$$0.5 = P(-0.674 < z < 0.674), \tag{1.22a}$$
$$0.9 = P(-1.645 < z < 1.645), \tag{1.22b}$$
$$0.95 = P(-1.960 < z < 1.960), \tag{1.22c}$$
$$0.99 = P(-2.576 < z < 2.576). \tag{1.22d}$$

The probabilities associated with these statements can be obtained from Table C.1. For example, (1.21a) is obtained by subtracting $F(-1) = 0.1587$ from $F(1) = 0.8413$, which results in 0.6827, or using MATLAB: `normcdf(1)-normcdf(-1)`. For statements (1.22a)–(1.22d), the probability associated with the left side must be found in the table and then the corresponding z-value can be read (interpolated) from the table. Because the limits are centered around z, but the table lists $P[Z \leq z]$, one should determine the value to find in the table as follows: if P denotes the probability, the value $1 - (1 - P)/2 = (1 + P)/2$ is the value to find in the table to obtain the upper limit of z. For the lower limit, use $(1 - P)/2$, which only differs in sign from the upper one. For example, for (1.22a), find $(1 + .5)/2 = 0.75$ in the table. These limits can also be found by using the MATLAB function `norminv`. For example `norminv(0.25)` returns -0.6745, and `norminv(0.75)` returns 0.6745.

Further discussions about the standard normal distribution can be found in Section 9.3.1.

1.5 Random Variables in Adjustment Computations

In the following, we present some properties of random variables, which are also called stochastic variables by some authors (e.g., Bjerhammar, 1973). More particularly, we focus on variables that represent random observation errors. Such errors also have been called accidental errors *ibid*, pg. 5. Though we cannot know what values random errors will take on, we may state what we expect their values to be, and we may also specify their level of deviation or variance about their expected values.

In the following sections, the notions of expectation and dispersion are defined mathematically. We first start with the univariate case, where only one random error is considered. Then we proceed to the multivariate case, where a vector of n random errors is considered.

1.5.1 Univariate Case

The univariate case deals with a single random variable, i.e., it treats a scalar quantity rather than a vector quantity. Let us introduce the continuous random variable e with a given probability density function (pdf) $f(e_t)$, where e_t represents a realization of e, i.e., a possible value that e might take on.

Expectation The probabilistic mean of e is the value that we *expect* e to take on. We denote the *expectation* of e as μ_e and define it as follows:

$$\mu_e \coloneqq E\{e\} = \int_{-\infty}^{\infty} e_t \, f(e_t) \, de_t, \tag{1.23}$$

where E is called expectation operator. Equation (1.23) is also called the first moment of e. If the random variable e represents measurement error, then, ideally, $E\{e\} = 0$. If $E\{e\} \neq 0$, we say that the measurement error is biased.

Dispersion The *dispersion*, or *variance*, of e is denoted by σ_e^2 and is defined by

$$\sigma_e^2 \coloneqq E\{(e - E\{e\})^2\} = \int_{-\infty}^{\infty} (e_t - \mu_e)^2 f(e_t) \, de_t \, . \tag{1.24a}$$

If $E\{e\} = 0$, then obviously

$$\sigma_e^2 = \int_{-\infty}^{\infty} e_t^2 f(e_t) \, de_t \, . \tag{1.24b}$$

Equation (1.24a) is also called the second centralized moment of e. In addition to σ_e^2, The dispersion operator, $D\{e\}$, is also used to denote the dispersion (variance) of e, but usually we reserve this notation for the multivariate case. The terms dispersion and variance are used interchangeably throughout these notes. The square root of variance is called *standard deviation*.

Variance is an indicator of how closely the values taken on by the random variable e are to the expected value of e. It is reflective of measurement precision and is inversely proportional to it. Thus, a small variance indicates high precision, and a large variance indicates low precision. A succinct expression for the expectation and variance of the random variable e, when e is assumed to be unbiased random measurement error, is

$$e \sim (0, \sigma_e^2). \tag{1.25}$$

The expression (1.25) is said in words as "*e is distributed with zero mean and sigma-sub-e-squared variance.*" Note that (1.25) does not specify a pdf for e but only its expectation and dispersion (or variance).

1.5.1.1 Expectation and Variance Propagation

Consider the *observation equation*

$$y = \mu + e, \quad e \sim (0, \sigma_e^2), \tag{1.26}$$

where y is an observation (measurement), μ is an unknown *observable*, and e accounts for the random error inherent in the observation y. We want to find the expectation and variance of y. In other words, we want to know how the expectation and variance propagate from the random variable e to the random variable y. Note that μ is a constant, or *non-random*, variable. The expectation of a constant is the constant itself; i.e., $E\{\mu\} = \mu$.

Using (1.23), we can write the expectation of $y = \mu + e$ as

$$E\{y\} = \int_{-\infty}^{\infty} (\mu + e_t)\, f(e_t)\, de_t, \tag{1.27a}$$

where e_t was defined in the preceding section as a value that the random variable e can take on. The expectation operator is linear. Thus, the expectation of the sum of random variables is the sum of their individual expectations. And, as noted already, μ is a constant variable. Therefore

$$E\{y\} = \mu \cancelto{1}{\int_{-\infty}^{\infty} f(e_t)\, de_t} + \int_{-\infty}^{\infty} e_t\, f(e_t)\, de_t = \mu + E\{e\} = \mu + 0 = \mu. \tag{1.27b}$$

The first integral evaluates to one according to (1.15b); the second integral was defined as expectation in (1.23).

The following rules are useful when working with expectations, given random variables x and y and constant c:

$$E\{E\{x\}\} = E\{x\}, \tag{1.28a}$$

$$E\{x + y\} = E\{x\} + E\{y\}, \tag{1.28b}$$

$$E\{c\} = c, \tag{1.28c}$$

$$E\{cx\} = c \cdot E\{x\}, \tag{1.28d}$$

$$E\{x \cdot y\} = E\{x\} \cdot E\{y\} \quad \text{if and only if } x \text{ and } y \text{ are independent random variables,} \tag{1.28e}$$

$$E\{x^2\} \neq E\{x\}^2 \text{ in general.} \tag{1.28f}$$

These rules can be extended to the multivariate case by replacing random variables x and y with random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, and by replacing the constant c with a constant matrix A.

After introducing y_t as a variable of integration, as was done for e_t above, the dispersion (variance) of y is defined by

$$D\{y\} = \int_{-\infty}^{\infty} (y_t - E\{y\})^2 f(y_t)\, dy_t =$$

$$= \int_{-\infty}^{\infty} (\mu + e_t - \mu)^2 f(e_t)\, de_t = \int_{-\infty}^{\infty} e_t^2 f(e_t)\, de_t = \sigma_e^2. \quad (1.29)$$

Summarizing, the first and second moments (i.e., the mean and variance) of y can be written succinctly as $y \sim (\mu, \sigma_e^2)$.

Another useful formula for the dispersion of any random variable y expresses it as the difference of the expectation of the square of the variable and the square of the variable's expected value. It is derived as follows:

$$\begin{aligned} D\{y\} &= E\{(y - E\{y\})^2\} = \\ &= E\{y^2 - 2yE\{y\} + E\{y\}^2\} = \\ &= E\{y^2 - 2y\mu + \mu^2\} = \\ &= E\{y^2\} - 2\mu E\{y\} + E\{\mu^2\} = \\ &= E\{y^2\} - 2\mu^2 + \mu^2 = \\ &= E\{y^2\} - \mu^2 \Rightarrow \end{aligned}$$

$$\boxed{D\{y\} = E\{y^2\} - E\{y\}^2 = \sigma_y^2.} \quad (1.30a)$$

Given constants α and γ, the above formulas for expectation and dispersion can be summarized as follows:

$$\boxed{E\{\alpha y + \gamma\} = \alpha E\{y\} + \gamma,} \quad (1.31a)$$

$$\boxed{D\{\alpha y + \gamma\} = \alpha^2 D\{y\}.} \quad (1.31b)$$

Equation (1.31b) represents *the law of error propagation* (covariance propagation) in its simplest form. It shows that, in contrast to the expectation, the dispersion operator is not linear. Furthermore, it shows that dispersion is not affected by a constant offset.

1.5.1.2 Mean Squared Error

The mean squared error, or MSE, of y is the expectation of the square of the difference of y and its *true value* μ. It is defined as

$$\mathrm{MSE}\{y\} = E\{(y - \mu)^2\} \quad (1.32)$$

(compare to (1.30)). It is useful to express the MSE as a combination of the dispersion and a (squared) *bias* term. This is done via the following derivation:

$$\begin{aligned} \mathrm{MSE}\{y\} &= E\{(y - \mu)^2\} = E\{[(y - E\{y\}) - (\mu - E\{y\})]^2\} = \\ &= E\{(y - E\{y\})^2 - 2(y - E\{y\})(\mu - E\{y\}) + (\mu - E\{y\})^2\} = \end{aligned}$$

$$= E\{(y - E\{y\})^2\} - 2E\{(y - E\{y\})(\mu - E\{y\})\} + E\{(\mu - E\{y\})^2\}. \quad (1.33)$$

Note that while y is a random variable, $E\{y\}$ is not. So, in the middle term, the expectation operator only applies to y. Therefore, we may continue with

$$\mathrm{MSE}\{y\} = D\{y\} - 2(\cancelto{0}{E\{y\} - E\{y\}})(\mu - E\{y\}) + (\mu - E\{y\})^2 \Rightarrow$$
$$\mathrm{MSE}\{y\} = D\{y\} + \beta^2, \quad (1.34)$$

where bias is defined formally as

$$\beta \coloneqq E\{y - \mu\} = E\{y\} - \mu. \quad (1.35)$$

Thus, we see that the dispersion of y and the MSE of y are only equal in the absence of bias, or in other words, only if indeed $\mu = E\{y\}$.

We noted previously that dispersion (variance) is an indicator of precision. In contrast, MSE is a measure of accuracy; it includes both dispersion and bias terms. In general, it is harder to meet accuracy standards than precision standards. We can typically increase precision by making more observations (though this may come with additional costs in time and resources); however it might not be possible to reduce bias by making more observations, and it may be very difficult to determine the origin of bias.

Finally, we note that the square root of MSE is commonly called rms (root mean square). Thus, strictly speaking, standard deviation and rms are only equivalent in the absence of bias.

1.5.2 Multivariate Case

The multivariate case deals with multiple random variables, which are collected in a column vector. For example, multiple observations of the observable μ in (1.26) can be expressed in the following system of equations:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \boldsymbol{\tau}\mu + \boldsymbol{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \mu + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}, \quad (1.36)$$

where $\boldsymbol{\tau}$ is a "summation vector" defined as $\boldsymbol{\tau} \coloneqq [1, \ldots, 1]^T$.[1] In the case of unbiased observations, i.e. $E\{\boldsymbol{e}\} = \boldsymbol{0}$, the expectation of the random error vector $\boldsymbol{e}$ is written as

$$E\{\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}\} = \begin{bmatrix} E\{e_1\} \\ \vdots \\ E\{e_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (1.37)$$

[1] The phrase "summing vector" has also been used. The name comes from the fact that if the dot product is taken between a vector of ones and another vector, the result is the sum of the elements of the other vector.

showing that the expectation of a vector can be written component-wise. Likewise, for the dispersion of each element e_j of $\boldsymbol{e}$, we have

$$D\{e_j\} = E\{\big(e_j - \cancelto{0}{E\{e_j\}}\big)^2\} = E\{e_j^2\}. \tag{1.38}$$

For the multivariate case, we must introduce the concept of *covariance*, which is a measure of similar behavior between random variables, e.g., between elements e_j and e_k of $\boldsymbol{e}$. Formally, the definition of covariance is

$$C\{e_j, e_k\} = \sigma_{jk} \coloneqq E\{\big(e_j - E\{e_j\}\big)\big(e_k - E\{e_k\}\big)\}. \tag{1.39}$$

Obviously,

$$C\{e_j, e_k\} = C\{e_k, e_j\}. \tag{1.40}$$

Moreover, when $E\{\boldsymbol{e}\} = \mathbf{0}$, the covariance between two of its elements reduces to

$$C\{e_j, e_k\} = E\{e_j e_k\}, \tag{1.41}$$

since $E\{e_j\} = E\{e_k\} = 0$. Even though we see from the definition of the covariance (1.39) that it does not depend on bias, in practice we often find that bias appears as positive correlation (see (1.51) for the definition of correlation coefficient).

Two random variables are said to be independent if their joint probability distribution is equal to the product of their individual probability distributions. Mathematically, this is written as

$$f\{e_j, e_k\} = f(e_j) \cdot f(e_k) \Leftrightarrow e_j \text{ and } e_k \text{ are independent.} \tag{1.42}$$

If two random variables are independent, their covariance is zero. The converse is not true unless the random variables are jointly normally distributed.

In light of the concept of covariance, the dispersion of a vector of random variables is represented by a matrix. The jth diagonal element of the matrix is denoted by σ_j^2 (or σ_{jj}^2) and the j,k off-diagonal term is written as σ_{jk}. The matrix is called a *covariance matrix* and is represented by Σ. Due to (1.40), the covariance matrix is symmetric. An explicit representation of the covariance matrix Σ is given by

$$D\{\underset{n\times 1}{\boldsymbol{e}}\} = \begin{bmatrix} D\{e_1\} & C\{e_1,e_2\} & \dots & C\{e_1,e_n\} \\ C\{e_2,e_1\} & D\{e_2\} & \dots & C\{e_2,e_n\} \\ \vdots & \vdots & \ddots & \vdots \\ C\{e_n,e_1\} & C\{e_n,e_2\} & \dots & D\{e_n\} \end{bmatrix} =: \underset{n\times n}{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}. \tag{1.43}$$

Obviously, if the random variables are uncorrelated, the covariance matrix is diagonal.

An important property of a covariance matrix is that it must be at least positive semidefinite. A matrix is positive semidefinite if, and only if, all of its eigenvalues are non-negative. For many applications in geodetic science, the covariance matrix is positive definite, which means that all its eigenvalues are greater than zero. The following statements hold for any positive-definite matrix Σ:

- $\boldsymbol{\alpha}^T \Sigma \boldsymbol{\alpha} = 0 \Rightarrow \boldsymbol{\alpha} = \mathbf{0}$.
- Σ is a nonsingular matrix (also called a *regular* matrix).
- All eigenvalues of Σ are positive and non-zero.
- All principle submatrices of Σ are also positive definite.

In the following chapters, where we treat observational models, we factor out of the covariance matrix Σ a scalar term denoted by σ_0^2, called a *variance component*, with the resulting matrix called the *cofactor* matrix. We label the cofactor matrix as Q; its inverse is labeled P and is called *weight matrix*. The relations between these terms are written mathematically as

$$\Sigma = \sigma_0^2 Q = \sigma_0^2 P^{-1}. \tag{1.44}$$

The simplest form of a covariance matrix Σ is when the cofactor matrix Q is equal to the identity matrix I_n. Indeed, if Q is any multiple of the identity matrix, the data are said to be *homogeneously* distributed. Another term for that case is *independent and identically distributed*, abbreviated iid. If the covariance matrix is diagonal, but its diagonal elements are not all the same, the data are said to have a *heteroscedastic* distribution. These cases are illustrated as follows:

- Homogeneous case

$$D\{\boldsymbol{e}\} = \sigma_0^2 Q = \sigma_0^2 \begin{bmatrix} q & 0 & \cdots & 0 \\ 0 & q & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & q \end{bmatrix} \Rightarrow P = \begin{bmatrix} 1/q & 0 & \cdots & 0 \\ 0 & 1/q & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & 1/q \end{bmatrix} = \frac{1}{q} \cdot I_n \tag{1.45a}$$

- Heteroscedastic case

$$D\{\boldsymbol{e}\} = \sigma_0^2 Q = \sigma_0^2 \begin{bmatrix} q_{11} & 0 & \cdots & 0 \\ 0 & q_{12} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & q_{nn} \end{bmatrix} \Rightarrow P = \begin{bmatrix} 1/q_{11} & 0 & \cdots & 0 \\ 0 & 1/q_{12} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & 1/q_{nn} \end{bmatrix} \tag{1.45b}$$

- General case

$$D\{\boldsymbol{e}\} = \sigma_0^2 Q = \sigma_0^2 \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{12} & q_{1,3} & \vdots \\ \vdots & 0 & \ddots & \vdots \\ q_{n1} & \cdots & \cdots & q_{nn} \end{bmatrix}, \text{ with } q_{ij} = q_{ji} \Rightarrow P = Q^{-1} \tag{1.45c}$$

Note that for $P = [p_{ij}]$, $p_{ii} \neq 1/q_{ii}$. Thus, the inverse of the diagonal components of Q are not weights in this case!

1.5.2.1 Error Propagation with Matrices

The derivations of (1.31a), (1.31b) and (1.34) can easily be extended to the multivariate case. Here we show their matrix analogs without deriving them (though some are derived in the example problems that follow).

If $\boldsymbol{y}$ is a random vector, A a constant matrix, and $\boldsymbol{\gamma}$ a constant vector, then the formulas for propagation of expectation and dispersion (error or covariance propagation) are summarized as follows:

Expectation:

$$\boxed{E\{A\boldsymbol{y}+\boldsymbol{\gamma}\} = A{\cdot}E\{\boldsymbol{y}\}+\boldsymbol{\gamma}} \tag{1.46a}$$

Dispersion (law of error propagation):

$$\boxed{D\{A\boldsymbol{y}+\boldsymbol{\gamma}\} = A{\cdot}D\{\boldsymbol{y}\}{\cdot}A^T} \tag{1.46b}$$

Also, analogous to (1.30) and (1.30a) we have

$$\boxed{D\{\boldsymbol{y}\} = E\{(\boldsymbol{y}-E\{\boldsymbol{y}\})(\boldsymbol{y}-E\{\boldsymbol{y}\})^T\} = E\{\boldsymbol{y}\boldsymbol{y}^T\} - E\{\boldsymbol{y}\}E\{\boldsymbol{y}\}^T.} \tag{1.47}$$

Covariance: Given two random vectors, $\boldsymbol{y}$ and $\boldsymbol{z}$, their covariance is written as

$$\boxed{C\{\boldsymbol{z},\boldsymbol{y}\} = E\{(\boldsymbol{z}-\boldsymbol{\mu}_z)(\boldsymbol{y}-\boldsymbol{\mu}_y)^T\} = E\{\boldsymbol{z}\boldsymbol{y}^T\} - \boldsymbol{\mu}_z\boldsymbol{\mu}_y^T} \tag{1.48}$$

Mean Squared Error: If $\boldsymbol{y}$ is a random vector with true value $\boldsymbol{\mu}$, the MSE of $\boldsymbol{y}$ is written as

$$\text{MSE}\{\boldsymbol{y}\} = D\{\boldsymbol{y}\} + \boldsymbol{\beta}\boldsymbol{\beta}^T, \tag{1.49a}$$

where the bias vector $\boldsymbol{\beta}$ is defined formally as

$$\boldsymbol{\beta} \coloneqq E\{\boldsymbol{y}-\boldsymbol{u}\} = E\{\boldsymbol{y}\} - \boldsymbol{u}. \tag{1.49b}$$

Once again, we see that the mean squared error matrix of a random vector is only equal to the dispersion matrix of the random vector in the absence of bias, i.e., when $\boldsymbol{\mu} = E\{\boldsymbol{y}\} \Rightarrow \boldsymbol{\beta} = \boldsymbol{0}$.

1.5.2.2 Correlation Matrix

A measure of *correlation* can be derived from the Cauchy-Schwartz inequality, which is given by

$$\begin{aligned} C\{e_j,e_k\} &= \iint (e_t)_j \cdot (e_t)_k \cdot f((e_t)_j,(e_t)_k)\, d(e_t)_j\, d(e_t)_k = \sigma_{jk} \leq \\ &\leq \sqrt{\int (e_t)_j^2 \cdot f\big((e_t)_j\big)\, d(e_t)_j \cdot \int (e_t)_k^2 \cdot f\big((e_t)_k\big)\, d(e_t)_k} = \sqrt{\sigma_j^2\sigma_k^2}. \end{aligned} \tag{1.50}$$

Since σ_{jk} can take on a positive or a negative value, the above inequality leads to the notion of a *correlation coefficient*, defined as

$$\boxed{\rho_{jk} \coloneqq \frac{\sigma_{jk}}{\sqrt{\sigma_j^2 \sigma_k^2}}, \quad \text{with} \quad -1 \leq \rho_{jk} \leq 1.} \tag{1.51}$$

Analogous to the covariance matrix, we may form a matrix of correlation coefficients. Such a matrix is called a *correlation matrix* and is defined as

$$\underset{n\times n}{R} \coloneqq \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{bmatrix} = R^T. \tag{1.52}$$

Given a covariance matrix Σ, the correlation matrix can be generated easily by

$$R = \operatorname{diag}\big(\big[1/\sigma_1, \dots, 1/\sigma_n\big]\big) \cdot \Sigma \cdot \operatorname{diag}\big(\big[1/\sigma_1, \dots, 1/\sigma_n\big]\big). \tag{1.53}$$

A note on units: Units must be properly accounted for in covariance matrices. The following list clarifies the units of relevant terms.

σ_0^2 unitless

ρ_{jk} unitless

σ_j^2 has squared units of observation y_j

σ_{jk} has units of observation y_j multiplied by the units of observation y_k

A further discussion on observations and random errors is given in Section 2.1.1 in the context of data models and least-squares adjustments.

Examples of covariance propagation

1. Given $\boldsymbol{y}$ as an $n \times 1$ observation vector and $\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{y})$ as an $m \times 1$ vector of nonlinear functions of $\boldsymbol{y}$.

   Find the $m \times n$ covariance matrix $C\{\boldsymbol{z}, \boldsymbol{y}\}$.

   Solution: Let $\boldsymbol{\mu}$ be the true value of $\boldsymbol{y}$ and linearize about expansion point $\boldsymbol{\mu}_0$ to get $\boldsymbol{z} = \boldsymbol{\alpha}_0 + A(\boldsymbol{y} - \boldsymbol{\mu}_0)$, with $\boldsymbol{\alpha}_0 \coloneqq F(\boldsymbol{\mu}_0)$ and A as the Jacobian matrix of $\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{y})$.

   Law of covariance propagation:

$$\begin{aligned} \underset{m\times n}{C\{\boldsymbol{z}, \boldsymbol{y}\}} &= E\{\boldsymbol{z}\boldsymbol{y}^T\} - E\{\boldsymbol{z}\}E\{\boldsymbol{y}\}^T = \\ &= E\{\big[\boldsymbol{\alpha}_0 + A(\boldsymbol{y} - \boldsymbol{\mu}_0)\big]\, \boldsymbol{y}^T\} - E\{\boldsymbol{\alpha}_0 + A(\boldsymbol{y} - \boldsymbol{\mu}_0)\} \cdot E\{\boldsymbol{y}\}^T = \\ &= \boldsymbol{\alpha}_0 \cdot E\{\boldsymbol{y}\}^T + A \cdot E\{\boldsymbol{y}\boldsymbol{y}^T\} - A\boldsymbol{\mu}_0 \cdot E\{\boldsymbol{y}\}^T - \end{aligned}$$

$$
\begin{aligned}
&- \boldsymbol{\alpha}_0 \cdot E\{\boldsymbol{y}\}^T - A \cdot E\{\boldsymbol{y}\} \cdot E\{\boldsymbol{y}\}^T + A\boldsymbol{\mu}_0 \cdot E\{\boldsymbol{y}\}^T = \\
&= A\big[E\{\boldsymbol{y}\boldsymbol{y}^T\} - E\{\boldsymbol{y}\}E\{\boldsymbol{y}\}^T\big] = A \cdot D\{\boldsymbol{y}\} \Rightarrow \\
&\qquad\qquad C\{\boldsymbol{z}, \boldsymbol{y}\} = A \cdot D\{\boldsymbol{y}\}
\end{aligned}
$$

2. Rather than one variable $\boldsymbol{z}$ as above, suppose we have $\boldsymbol{z}_1$ of size $m_1 \times 1$ and $\boldsymbol{z}_2$ of size $m_2 \times 1$. Find the $m_1 \times m_2$ covariance matrix $C\{\boldsymbol{z}_1, \boldsymbol{z}_2\}$.

   Solution: After linearization

$$
C\{\boldsymbol{z}_1 = \boldsymbol{\alpha}_1^0 + A_1\boldsymbol{y}, \boldsymbol{z}_1 = \boldsymbol{\alpha}_2^0 + A_2\boldsymbol{y}\} = \underset{m_1 \times n}{A_1} \cdot \underset{n \times n}{D\{\boldsymbol{y}\}} \cdot \underset{n \times m_2}{A_2^T}
$$

3. Given the $m_1 \times 1$ random vector $\boldsymbol{z}_1$, the $m_2 \times 1$ random vector $\boldsymbol{z}_2$, constant vectors $\boldsymbol{\beta}_1(l_1 \times 1)$ and $\boldsymbol{\beta}_2(l_2 \times 1)$ and constant matrices $B_1(l_1 \times m_1)$ and $B_2(l_2 \times m_2)$.

   Find the covariance matrix of $\boldsymbol{x}_1 = \boldsymbol{\beta}_1 + B_1\boldsymbol{z}_1$ and $\boldsymbol{x}_2 = \boldsymbol{\beta}_2 + B_2\boldsymbol{z}_2$.

   Solution:

$$
C\{\boldsymbol{x}_1 = \boldsymbol{\beta}_1 + B_1\boldsymbol{z}_1, \boldsymbol{x}_2 = \boldsymbol{\beta}_2 + B_2\boldsymbol{z}_2\} = \underset{l_1 \times m_1}{B_1} \; \underset{m_1 \times m_2}{C\{\boldsymbol{z}_1, \boldsymbol{z}_2\}} \; \underset{m_2 \times l_2}{B_2^T}
$$

   Note that the matrix $C\{\boldsymbol{z}_1, \boldsymbol{z}_2\}$ is not necessarily symmetric.

4. What is the covariance of the random variable $\boldsymbol{y}$ with itself?

   Solution:

$$
C\{\boldsymbol{y}, \boldsymbol{y}\} = E\{\boldsymbol{y}\boldsymbol{y}^T\} - E\{\boldsymbol{y}\}E\{\boldsymbol{y}\}^T = D\{\boldsymbol{y}\}
$$

5. Given $n \times 1$ vectors $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{e}$ with $E\{\boldsymbol{e}\} = \boldsymbol{0}$, which implies that $E\{\boldsymbol{y}\} = \boldsymbol{\mu}$ and $D\{\boldsymbol{e}\} = E\{\boldsymbol{e}\boldsymbol{e}^T\}$.

   Find: The dispersion matrix $D\{\boldsymbol{y}\}$.

   Solution:

$$
\begin{aligned}
&D\{\boldsymbol{y}\} = E\{(\boldsymbol{y} - E\{\boldsymbol{y}\})(\boldsymbol{y} - E\{\boldsymbol{y}\})^T\} = E\{(\boldsymbol{y} - \boldsymbol{\mu})(\boldsymbol{y} - \boldsymbol{\mu})^T\} = \\
&= E\{\boldsymbol{y}\boldsymbol{y}^T - \boldsymbol{y}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{y}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T\} = E\{\boldsymbol{y}\boldsymbol{y}^T\} - \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T \Rightarrow \\
&\qquad\qquad D\{\boldsymbol{y}\} = E\{\boldsymbol{y}\boldsymbol{y}^T\} - \boldsymbol{\mu}\boldsymbol{\mu}^T
\end{aligned}
$$

6. Given random vectors $\boldsymbol{y}$ and $\boldsymbol{z}$, with corresponding expectations $E\{\boldsymbol{y}\} = \boldsymbol{\mu}_y$ and $E\{\boldsymbol{z}\} = \boldsymbol{\mu}_z$, find the covariance matrix $C\{\boldsymbol{z}, \boldsymbol{y}\}$.

   Solution:

$$
\begin{aligned}
C\{\boldsymbol{z}, \boldsymbol{y}\} &= E\{(\boldsymbol{z} - \boldsymbol{\mu}_z)(\boldsymbol{y} - \boldsymbol{\mu}_y)^T\} = E\{\boldsymbol{z}\boldsymbol{y}^T - \boldsymbol{z}\boldsymbol{\mu}_y^T - \boldsymbol{\mu}_z\boldsymbol{y}^T + \boldsymbol{\mu}_z\boldsymbol{\mu}_y^T\} = \\
&= E\{\boldsymbol{z}\boldsymbol{y}^T\} - \boldsymbol{\mu}_z\boldsymbol{\mu}_y^T - \boldsymbol{\mu}_z\boldsymbol{\mu}_y^T + \boldsymbol{\mu}_z\boldsymbol{\mu}_y^T = E\{\boldsymbol{z}\boldsymbol{y}^T\} - \boldsymbol{\mu}_z\boldsymbol{\mu}_y^T
\end{aligned}
$$

7. Suppose y_1, y_2, and y_3 are independent measurements with standard deviations $\sqrt{2}\,\text{cm}$, $2\,\text{cm}$, and $1\,\text{cm}$, respectively. The quantities x_1 and x_2 are computed from the measurements as follows

$$\begin{aligned} x_1 &= 2y_1 + y_2 + 2y_3, \\ x_2 &= y_1 - 2y_2 \quad . \end{aligned}$$

Evaluate the covariance matrix for the random vector $\boldsymbol{x} = [x_1,\ x_2]^T$.

Solution: The given equations can be written in matrix form as

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\boldsymbol{y} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix},\ D\{\boldsymbol{y}\} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{cm}^2 = \Sigma_{yy}$$

Now apply the law of error propagation (1.46b):

$$\begin{aligned} D\{\boldsymbol{x}\} = A{\cdot}D\{\boldsymbol{y}\}{\cdot}A^T = \begin{bmatrix} 2 & 1 & 2 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & -2 \\ 2 & 0 \end{bmatrix} \text{cm}^2 = \\ = \begin{bmatrix} 16 & -4 \\ -4 & 18 \end{bmatrix} \text{cm}^2 = \Sigma_{xx} \end{aligned}$$

$$\Rightarrow \sigma_{x_1} = 4\,\text{cm},\ \sigma_{x_2} = 3\sqrt{2}\,\text{cm},\ \sigma_{x_1x_2} = -4\,\text{cm}^2$$

$$\Rightarrow \rho_{x_1x_2} = \frac{\sigma_{x_1x_2}}{\sigma_{x_2}\sigma_{x_2}} = \frac{-4\,\text{cm}^2}{4\,\text{cm} \cdot 3\sqrt{2}\,\text{cm}} = -0.2357.$$

Correlation matrix:

$$\begin{aligned} R = \begin{bmatrix} \frac{1}{4\,\text{cm}} & 0 \\ 0 & \frac{1}{3\sqrt{2}\,\text{cm}} \end{bmatrix} \begin{bmatrix} 16 & -4 \\ -4 & 18 \end{bmatrix} \text{cm}^2 \begin{bmatrix} \frac{1}{4\,\text{cm}} & 0 \\ 0 & \frac{1}{3\sqrt{2}\,\text{cm}} \end{bmatrix} = \\ = \begin{bmatrix} 1 & -0.2357 \\ -0.2357 & 1 \end{bmatrix} \end{aligned}$$

8. An azimuth and distance were measured from known point C to point D to determine the coordinates of D (see Figure 1.3). Compute the coordinates of D and their covariance matrix, along with the correlation matrix, based on the given data.

Data:

$$\begin{aligned} x_c &= 2000.0\,\text{m}, & \sigma_{x_c} &= 1\,\text{cm} \\ y_c &= 3000.0\,\text{m}, & \sigma_{y_c} &= 1\,\text{cm} \\ \alpha &= 120^\circ 00' 00'', & \sigma_\alpha &= 10'' \\ s &= 1600.00\,\text{m}, & \sigma_s &= 5\,\text{cm} \end{aligned}$$

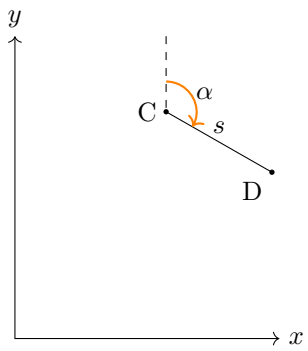

Figure 1.3: Azimuth α and distance s measured from point C to point D to determined the coordinates of point D.

Principle: covariance propagation $D\{A\boldsymbol{y} + \boldsymbol{\gamma}\} = A \cdot D\{\boldsymbol{y}\} \cdot A^T$

Let the random variable $\boldsymbol{y} := [x_c, y_c, \alpha, s]^T$ and the random variable $x := [x_D, y_D]^T$.

Functional relations:

$$x_D = x_C + s \cdot \sin\alpha$$
$$y_D = y_C + s \cdot \cos\alpha$$

$\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{y})$, $\boldsymbol{x}$ is a nonlinear function of $\boldsymbol{y}$. Under linearization

$$\boldsymbol{x} \approx \boldsymbol{f}(\boldsymbol{y}_0) + \left.\frac{\partial \boldsymbol{f}(\boldsymbol{y})}{\partial \boldsymbol{y}^T}\right|_{\boldsymbol{y}_0} (\boldsymbol{y} - \boldsymbol{y}_0)$$

Use values of observations for $\boldsymbol{y}_0$.

$$\boxed{\boldsymbol{f}(\boldsymbol{y}_0) \text{ gives: } x_D = 3385.64\,\text{m}, \quad y_D = 2200.00\,\text{m}}$$

$$\frac{\partial x_D}{\partial x_C} = 1, \quad \frac{\partial x_D}{\partial y_C} = 0, \quad \frac{\partial x_D}{\partial \alpha} = s \cdot \cos\alpha, \quad \frac{\partial x_D}{\partial s} = \sin\alpha,$$
$$\frac{\partial y_D}{\partial x_C} = 0, \quad \frac{\partial y_D}{\partial y_C} = 1, \quad \frac{\partial y_D}{\partial \alpha} = -s \cdot \sin\alpha, \quad \frac{\partial y_D}{\partial s} = \cos\alpha$$

$$\Rightarrow A = \begin{bmatrix} 1 & 0 & s \cdot \cos\alpha & \sin\alpha \\ 0 & 1 & -s \cdot \sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} 1 & 0 & -800.0 & 0.866 \\ 0 & 1 & -1385.64 & -0.5 \end{bmatrix}$$

Covariance matrix for given data:

$$\Sigma_{yy} = \begin{bmatrix} (0.01\,\mathrm{m})^2 & 0 & 0 & 0 \\ 0 & (0.01\,\mathrm{m})^2 & 0 & 0 \\ 0 & 0 & \left(\frac{10''}{3600''/1^\circ}\frac{\pi}{180^\circ}\right)^2 & 0 \\ 0 & 0 & 0 & (0.05\,\mathrm{m})^2 \end{bmatrix}$$

Covariance matrix for coordinates of point D:

$$D\{\begin{bmatrix} x_D \\ y_D \end{bmatrix}\} = A{\cdot}\Sigma_{yy}{\cdot}A^T = \begin{bmatrix} 0.0035 & 0.0015 \\ 0.0015 & 0.0052 \end{bmatrix} \mathrm{m}^2 = \Sigma_{xx}$$

Standard deviations for coordinates of point D:

$$\boxed{\Rightarrow \sigma_{x_D} = \pm 6\,\mathrm{cm}, \quad \sigma_{y_D} = \pm 7\,\mathrm{cm}}$$

Correlation matrix:

$$R = \begin{bmatrix} 1/\sigma_{x_D} & 0 \\ 0 & 1/\sigma_{y_D} \end{bmatrix} {\cdot}\Sigma_{xx}{\cdot} \begin{bmatrix} 1/\sigma_{x_D} & 0 \\ 0 & 1/\sigma_{y_D} \end{bmatrix} = \begin{bmatrix} 1 & 0.3568 \\ 0.3568 & 1 \end{bmatrix}$$

1.5.3 Practice Problems

1. Let X be a random variable with the following probability density function:

$$f(x) = \begin{cases} \frac{1}{8}(x-1) & \text{for } 1 < x < 5, \\ 0 & \text{otherwise.} \end{cases}$$

Derive the cumulative distribution function of X and evaluate $P[X < 2]$, $P[X > 4]$, and $P[1.5 < X < 4.5]$.

2. Let X be a random variable with the following probability density function:

$$f(x) = \frac{\sin x}{2} \quad \text{for } 0 < x < \pi.$$

Derive the cumulative distribution function of X and evaluate $P[X < \pi/4]$, $P[X > \pi/2]$, and $P[\pi/4 < X < \pi/2]$. Sum the three probabilities and comment on the result.

3. Evaluate the mean and variance of the random variable in the preceding problem (hint: integration by parts).

4. Two measurements are normally distributed with standard deviations of 0.55 m and 0.35 m, respectively. Compute the standard deviation of the sum and difference of the two measurements if the correlation coefficient of the two measurements is: (a) 0.5, (b) 0, (c) −0.5, (d) 1.0.

5. The X and Y coordinates of a survey point have standard deviations of $\sigma_x = 0.045\,\text{m}$ and $\sigma_y = 0.025\,\text{m}$, respectively. (a) Compute the correlation coefficient of X and Y if the covariance of X and Y is $0.00012\,\text{m}^2$. (b) Compute the covariance of X and Y if the correlation coefficient is 0.333.

6. Consider a linear transformation $Y = a + bX$, where X is a random variable having a normal distribution, i.e., $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.

   Show that $E\{Y\} = a + b{\cdot}\mu_X$ and $\sigma_Y^2 = b^2{\cdot}\sigma_X^2$. Then show that

   $$Z = \frac{X - \mu_X}{\sqrt{\sigma_X^2}}$$

   has zero mean and unit variance.

7. Consider the following system of equations

   $$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 & 2 \\ -1 & 3 & 2 & -1 \\ 1 & -1 & 6 & 7 \\ 2 & -2 & 14 & 20 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \boldsymbol{y} = A\boldsymbol{x},$$

   where y_1, y_2, y_3, and y_4 are independent and identically distributed (iid) each with the mean 0 and variance σ^2.

   (a) Express x_1, x_2, x_3, and x_4 in terms of y_1, y_2, y_3, and y_4.

   (b) Compute the covariance matrix for $\boldsymbol{x}$.

   (c) Suppose now that instead of being iid, the dispersion of $\boldsymbol{y}$ is given by the matrix

   $$\begin{bmatrix} \sigma^2 & \rho\sigma^2 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 \\ 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & \rho\sigma^2 & \sigma^2 \end{bmatrix}.$$

   Answer question (b) in this case.

8. Suppose three points A, B, and C are sequentially located on a straight line (Figure 1.4). A total station was used to measure distances between them, so that the total distance between A and C could be estimated. The data are listed in Table 1.1.

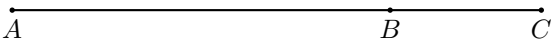

Figure 1.4: Points A, B, and C on a straight line

The variance of each observation is given by $\sigma^2 = (9\,\text{mm})^2 + (d/100)^2\,\text{mm}^2$, where d is distance in meters. Unbeknownst to the surveyor, a constant

Table 1.1: Observations of line segments

y_i	Segment	Observation [m]
y_1	$\overline{AB}$	$52.154 + 0.025 = 52.179$
y_2	$\overline{AB}$	$52.157 + 0.025 = 52.182$
y_3	$\overline{AB}$	$52.155 + 0.025 = 52.180$
y_4	$\overline{AC}$	$70.180 + 0.025 = 70.205$
y_5	$\overline{AC}$	$70.178 + 0.025 = 70.203$
y_6	$\overline{BC}$	$18.022 + 0.025 = 18.047$
y_7	$\overline{BC}$	$18.021 + 0.025 = 18.046$
y_8	$\overline{BC}$	$18.025 + 0.025 = 18.050$

bias $\beta = 2.5\,\text{cm}$ affected every observation, which is reflected as $+0.025\,\text{m}$ in Table 1.1.

Suppose the surveyor estimated the total distance $\overline{AC}$ by the formula

$$\overline{AC} = z = \frac{1}{6}(y_1 + y_2 + y_3 + y_6 + y_7 + y_8) + \frac{1}{4}(y_4 + y_5).$$

(a) Compute the standard deviation and rms (square root of MSE) of z using $\mu_{AC} = 70.179\,\text{m}$ as a hypothetical "true value" of the total distance $\overline{AC}$.

(b) Now use the same formula for z with the unbiased observations (i.e., remove the 25 cm bias from each observation). Compute its variance and compare to the variance of part (a). Do you expect the variances to be the same? Why or why not?

(c) Find a different combination of the measurements that would provide an estimate for the total distance that is not affected by the bias in the measurements. Compute the standard deviation and rms for this estimate (again using $\mu_{AC} = 70.179\,\text{m}$ as a hypothetical "true value"). Compare these results to those of parts (a) and (b) and comment on your findings.

9. Given a random variable y with expectation $E\{y\} = \mu_y$ and variance σ_y^2, suppose f and g are functions of y defined by $f = e^y$ and $g = y^3$, respectively.

(a) Using a Taylor series expansion, express the expectations and dispersions of f and g in terms of μ_y, σ_y^2, and $\delta = (\mu - \mu_0)$, where μ_0 is an approximation of μ.

(b) Assume that $E\{y\}$ coincides with the true value of μ_y of y, so that biases are due to the truncation of the Taylor series. What are the biases in the f and g due to the series truncation? Which bias is larger?

(c) Assume that the approximate value μ_0 coincides with the expectation μ_y of y. What are the expectations and dispersions now?

10. Sides a and b of the right-angled plane triangle in Figure 1.5 were measured. The values obtained are $a = 399.902\,\text{m}$ and $b = 300.098\,\text{m}$, with variances $\sigma_a^2 = (0.015\,\text{m})^2$ and $\sigma_a^2 = (0.020\,\text{m})^2$, respectively. The correlation coefficient is $\rho_{ab} = 0.2$. Compute side c and angle β and their standard deviations. Also determine the correlation, if any, between computed side c and angle β.

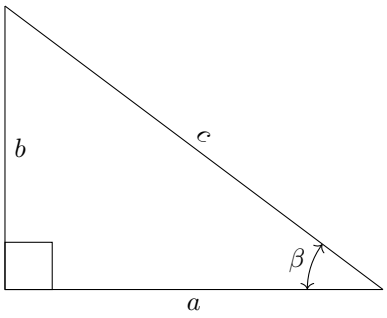

Figure 1.5: Right-angled plane triangle with measured sides a and b

11. The area of a trapezoidal parcel of land is computed by

$$A = \left(\frac{a_1 + a_2}{2}\right) b,$$

where a_1, a_2, and b were measured independently. The measurements and their standard deviations are $a_1 = 301.257 \pm 0.025\,\text{m}$, $a_2 = 478.391 \pm 0.045\,\text{m}$, and $b = 503.782 \pm 0.030\,\text{m}$. Compute the area of the parcel and the standard deviation of the computed area.

Chapter 2

The Model of Direct Observations

2.1 Model Definition

When an unknown parameter μ can be observed directly, the *model of direct observations* can be formed for the data by

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mu + e_1 \\ \vdots \\ \mu + e_n \end{bmatrix} = \boldsymbol{\tau}\mu + \boldsymbol{e}, \tag{2.1a}$$

$$\boldsymbol{e} \sim \left(\boldsymbol{0}, \sigma_0^2 Q\right), \;\; Q \coloneqq P^{-1}. \tag{2.1b}$$

The terms in the data model are defined as follows:

$\boldsymbol{y}$ is a *given* $n \times 1$ vector of observations with random properties.

μ is an unknown, non-random parameter to be *estimated*.

$\boldsymbol{\tau}$ is an $n \times 1$ vector of ones ("summation vector"), i.e., $\boldsymbol{\tau} \coloneqq [1, \ldots, 1]^T$.

$\boldsymbol{e}$ is an $n \times 1$ vector of unknown, random errors to be *predicted*.

Q is a *given* $n \times n$ cofactor matrix associated with $\boldsymbol{e}$. It is symmetric, positive-definite, and non-random.

P is an $n \times n$ positive-definite weight matrix, being the inverse of Q.

σ_0^2 is an unknown, non-random variance component that can be estimated.

Equation (2.1a) is called *observation equations*, while (2.1b) provides a *stochastic model* for the random observational errors. Together, these two equations comprise a complete *data model*.

2.1.1 Terminology: Observations, Redundancy, Residuals and Their Minimization

The observation vector $\boldsymbol{y}$ was described above as a given quantity. It is given in the sense that it consists of measurements (observations) that are typically made and recorded in the field. The measurements are considered to be a physical realization of an *observable* — the quantity ("the thing") being observed. An observable could be a dimension of an element of a physical object or a relationship between its elements, such as an angle between two connected edges of a geodetic network, the end points of which being accessible monuments in the ground. Or, and observable could be a property of an immaterial object, such as the phase of an electromagnetic wave. Another example of an observable is the length of a bridge from a mark scribed in concrete at its beginning to another at its end; then, an associated observation could be a distance measured and recorded with a surveyor's total station between those two marks. Being a measurement of an observable, an observation is a numerical value with an associated unit of measurement.

Even though the vector of observations $\boldsymbol{y}$ is given, it has *random* properties due to unavoidable *random errors* inherent both in making observations and in the instruments used to make them. These random errors are *unknown* quantities, and they are accounted for in the observation equations (2.1a) by the random error vector $\boldsymbol{e}$. Thus, we can say that we know the value of an observation, but we do not know the value of its random error constituent. However, we have already seen that we can say something about the expected values of the random errors (i.e., $E\{\boldsymbol{e}\} = \boldsymbol{0}$). Likewise, a statement can be made about the expectation of the observations, viz.

$$\boldsymbol{\mu}_y \coloneqq E\{\boldsymbol{y}\} = E\{\boldsymbol{\tau}\mu + \boldsymbol{e}\} = \boldsymbol{\tau}\mu. \tag{2.2a}$$

We may think of the vector $\boldsymbol{\mu}_y$ as the vector of *true observations*, the values of which are unknown, though they can be estimated via

$$\widehat{E\{\boldsymbol{y}\}} =: \hat{\boldsymbol{\mu}}_y = \boldsymbol{\tau}\hat{\mu}, \tag{2.2b}$$

where $\hat{\mu}$ is an estimate of the unknown parameter μ. The vector $\hat{\boldsymbol{\mu}}_y$ is called the vector of *adjusted observations*.

Because the given observations, $\boldsymbol{y}$, contain unknown random errors represented by $\boldsymbol{e}$, we cannot possibly expect that $\boldsymbol{y}$ will equal $\boldsymbol{\tau}\mu$, though we may usually hope that at least $\boldsymbol{y} \approx \boldsymbol{\tau}\mu$. The inequality $\boldsymbol{y} \neq \boldsymbol{\tau}\mu$ should be immediately evident from the symbols, since they imply that $\boldsymbol{y}$ is random and $\boldsymbol{\tau}\mu$ is not. (Recall the use of Latin characters for random variables and Greek characters for non-random variables as discussed on page 2.) The rule eluded to here is that when one side of an equation results in a random quantity, so must the other side. The incongruency reflected in $\boldsymbol{y} \neq \boldsymbol{\tau}\mu$ is rectified in (2.1a) by the addition of $\boldsymbol{e}$ on the right side. But practically speaking, $\boldsymbol{e}$ is not much help, since it is unknown. This is where least-squares adjustment theory and techniques can come to the rescue. For if there are more observations than parameters in the model (i.e., more than one observation for model (2.1)), we can use these *redundant* observations to *predict* values for $\boldsymbol{e}$

using a *predictor* derived from the principle of least-squares adjustment (see below for a brief discussion on predictors and estimators).

The number of independent, redundant observations is called the *redundancy* of the model. Another term for it is *degrees of freedom*, sometimes abbreviated df in the statistical literature. The vector of predicted random errors is denoted by $\tilde{\boldsymbol{e}}$, and it is also called the vector of *residuals*.

The idea behind least-squares adjustment is to predict the residuals so that the (weighted) sum of their squares is minimized, while still satisfying the relation $\boldsymbol{y} = \boldsymbol{\tau}\mu + \boldsymbol{e}$ shown in (2.1a), but now with the predicted random errors $\tilde{\boldsymbol{e}}$ and the estimated parameter $\hat{\mu}$ rather than their corresponding "true," but unknown, quantities. That is, the relation

$$\boldsymbol{y} = \boldsymbol{\tau}\hat{\mu} + \tilde{\boldsymbol{e}} \tag{2.3}$$

must hold after the adjustment, and the (weighted) sum of squared residuals $\Omega \coloneqq \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$ must be as small as possible. That is both the objective and the outcome of least-squares adjustments.

It is the data that are being adjusted in least-squares adjustments. They are adjusted so that the inconsistent equation $\boldsymbol{y} \neq \boldsymbol{\tau}\mu$ is replaced by the consistent equation $\hat{\boldsymbol{\mu}}_y = \boldsymbol{\tau}\hat{\mu}$. So, we speak of adjusted data, predicted residuals, and estimated parameters as the outcomes of a least-squares solution, which is derived in the next section.

Estimate vs. estimator In these notes we have hardly distinguished between the terms estimate and estimator. This is partly because sometimes the same symbol works for both terms depending on the context, though in some places we might have used estimate when we could have used estimator, and the same can be said for prediction and predictor. The distinction between these terms made by Tukey (1987, p. 633) is quoted as follows:

> An estimator is a function of the observations, a specific way of putting them together. It may be specified by an arithmetic formula, like $\bar{y} = \Sigma x_i/n$, or by words alone, as in directions for finding a sample median by ordering and counting. We distinguish between the estimator and its value, an estimate, obtained from the specific set of data. The variance estimator, $s^2 = \Sigma(x_i - \bar{x})^2/(n-1)$, yields the estimate 7 from the three observations 2, 3, 7. We say s^2 is an estimator for σ^2, and we call σ^2 the estimand. In the numerical example, 7 estimates σ^2.

2.2 The Least-Squares Solution

In order to minimize $\Omega \coloneqq \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$ while satisfying (2.3) we form the Lagrange target function

$$\Phi(\boldsymbol{e}, \boldsymbol{\lambda}, \mu) \coloneqq \boldsymbol{e}^T P \boldsymbol{e} + 2\boldsymbol{\lambda}^T(\boldsymbol{y} - \boldsymbol{\tau}\mu - \boldsymbol{e}), \tag{2.4}$$

where $\boldsymbol{\lambda}$ is an unknown $m \times 1$ vector of Lagrange multipliers. The target function is made stationary with respect to the unknown terms $\boldsymbol{e}$, $\boldsymbol{\lambda}$, and μ when its first

partial derivatives are set equivalent to zero, which is reflected in the following *Euler-Lagrange necessary conditions*:

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{e}} = \frac{1}{2}\left[\frac{\partial\Phi}{\partial e_j}\right]_{n\times 1} = P\tilde{\boldsymbol{e}} - \hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{2.5a}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\lambda}} = \frac{1}{2}\left[\frac{\partial\Phi}{\partial\lambda_j}\right]_{n\times 1} = \boldsymbol{y} - \boldsymbol{\tau}\hat{\mu} - \tilde{\boldsymbol{e}} \doteq \mathbf{0}, \tag{2.5b}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\mu} = \boldsymbol{\tau}^T\hat{\boldsymbol{\lambda}} \doteq 0. \tag{2.5c}$$

These necessary conditions are sometimes called first-order conditions due to the involvement of first partial derivatives. The sufficient condition for minimization is satisfied by the fact that the second partial derivative of Φ is $\partial\Phi^2/(\partial\boldsymbol{e}\partial\boldsymbol{e}^T) = 2P$, where the weight matrix P is positive definite according to (2.1). Therefore, the solution to the system of equations (2.5) yields the minimum of Φ, and thus the weighted sum of squared residuals (weighted SSR) $\Omega = \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$ is minimized. See Appendix A for comments on derivatives of quadratic functions with respect to column vectors.

Throughout these notes, we use a hat to denote an *estimate* of a non-random variable, whereas a tilde denotes a *prediction* of a random variable. The hat and tilde marks were introduced into (2.5) to distinguish between the unknown variables of the target function (2.4) and those particular quantities that satisfy the necessary conditions. This reflects that $\tilde{\boldsymbol{e}}$, $\hat{\boldsymbol{\lambda}}$, and $\hat{\mu}$ cannot take on just any values but rather only those that result from setting the first partial derivatives of the target function to zero (denoted by the $\doteq$ sign), which explains why it would not be logical to introduce the hat and tilde symbols in (2.4). Also note that, for the vector $\tilde{\boldsymbol{e}}$, we use the terms *residual* and *predicted random error* synonymously.

Now we must solve the system of equations (2.5) to obtain the least-squares solution (LESS) as follows:

$$\hat{\boldsymbol{\lambda}} = P\tilde{\boldsymbol{e}} = P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) \qquad \text{using (2.5a) and (2.5b)} \tag{2.6a}$$

$$\boldsymbol{\tau}^T\hat{\boldsymbol{\lambda}} = \boldsymbol{\tau}^T P\boldsymbol{y} - \big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)\hat{\mu} = 0 \qquad \text{using (2.6a) and (2.5c)} \tag{2.6b}$$

Equation (2.6b) leads to

$$\boxed{\hat{\mu} = \frac{\boldsymbol{\tau}^T P\boldsymbol{y}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}} \tag{2.7}$$

for the estimate of the unknown parameter μ. And, from (2.5b), we have

$$\boxed{\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{\tau}\hat{\mu} \Rightarrow} \tag{2.8a}$$

$$\tilde{\boldsymbol{e}} = \big[I_n - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\big]\boldsymbol{y} \tag{2.8b}$$

for the prediction of the random error vector $\boldsymbol{e}$. As stated already, the prediction $\tilde{\boldsymbol{e}}$ is also called residual vector. We say that the quantities $\hat{\mu}$, $\tilde{\boldsymbol{e}}$, and $\hat{\boldsymbol{\lambda}}$ belong to a LEast-Squares Solution (LESS) within the model of direct observations (2.1).

It turns out that $\hat{\mu}$ is an unbiased estimator of μ, since

$$E\{\hat{\mu}\} = E\{(\boldsymbol{\tau}^T P\boldsymbol{\tau})^{-1}\boldsymbol{\tau}^T P\boldsymbol{y}\} = (\boldsymbol{\tau}^T P\boldsymbol{\tau})^{-1}\boldsymbol{\tau}^T P\cdot E\{\boldsymbol{y}\} = (\boldsymbol{\tau}^T P\boldsymbol{\tau})^{-1}\boldsymbol{\tau}^T P\boldsymbol{\tau}\mu = \mu. \tag{2.9}$$

Likewise, the residual vector $\tilde{\boldsymbol{e}}$ is an unbiased predictor of the random error vector $\boldsymbol{e}$, since

$$E\{\tilde{\boldsymbol{e}}\} = \big[I_n - \boldsymbol{\tau}(\boldsymbol{\tau}^T P\boldsymbol{\tau})^{-1}\boldsymbol{\tau}^T P\big]\cdot E\{\boldsymbol{y}\} = \\ = \big[I_n - \boldsymbol{\tau}(\boldsymbol{\tau}^T P\boldsymbol{\tau})^{-1}\boldsymbol{\tau}^T P\big]\boldsymbol{\tau}\mu = \boldsymbol{\tau}\mu - \boldsymbol{\tau}\mu = \mathbf{0}. \tag{2.10}$$

The vectors $\boldsymbol{\tau}$ and $\tilde{\boldsymbol{e}}$ are said to be P-*orthogonal* since

$$\boldsymbol{\tau}^T P\tilde{\boldsymbol{e}} = \boldsymbol{\tau}^T P(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}) = \boldsymbol{\tau}^T P\big[I_n - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\big]\boldsymbol{y} = \\ = \boldsymbol{\tau}^T P\boldsymbol{y} - \boldsymbol{\tau}^T P\boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\boldsymbol{y} = 0. \tag{2.11}$$

This result reveals that the sum of the P-weighted residual vector within the model of direct observations is zero.

The adjusted observations, $\boldsymbol{\tau}\hat{\mu}$, on the right side of (2.8a) can also be expressed as

$$\boxed{\hat{\boldsymbol{\mu}}_y \coloneqq \widehat{E\{\boldsymbol{y}\}} = \boldsymbol{\tau}\hat{\mu} = \boldsymbol{y} - \tilde{\boldsymbol{e}}.} \tag{2.12}$$

Obviously, since $\boldsymbol{\tau}^T P\tilde{\boldsymbol{e}} = 0$, we also have

$$\big(\boldsymbol{\tau}\hat{\mu}\big)^T P\tilde{\boldsymbol{e}} = \hat{\boldsymbol{\mu}}_y^T P\tilde{\boldsymbol{e}} = 0. \tag{2.13}$$

Equation (2.13) reveals an important characteristic of LESS; viz., the vector of adjusted observations and the vector of P-weighted residuals are orthogonal to one another. From a geometric point of view (illustrated in Figure 2.1), the orthogonal relationship between these vectors means that the vector of observations $\boldsymbol{y}$ and the vector of adjusted observations $\hat{\boldsymbol{\mu}}_y$ are as close as possible to each other (considering the weights in P), which is exactly what we require from a least-squares adjustment: a minimum adjustment of the data that will satisfy the given observational model.

In addition to solving for the estimated parameter $\hat{\mu}$ and the predicted random error vector $\tilde{\boldsymbol{e}}$, we are typically interested in their dispersions (variances), which are an indicator of their precisions. To compute their dispersions, we apply the law of covariance propagation. First, for the dispersion of the estimated parameter $\hat{\mu}$ we have

$$D\{\hat{\mu}\} = \frac{\boldsymbol{\tau}^T P}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} D\{\boldsymbol{y}\} \frac{P\boldsymbol{\tau}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} = \frac{\boldsymbol{\tau}^T P\big(\sigma_0^2 P^{-1}\big)P\boldsymbol{\tau}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}\,\boldsymbol{\tau}^T P\boldsymbol{\tau}} \Rightarrow \\ \boxed{D\{\hat{\mu}\} = \frac{\sigma_0^2}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}.} \tag{2.14}$$

The $n \times n$ dispersion matrix for the residual vector $\tilde{\boldsymbol{e}}$ is derived by

$$D\{\tilde{\boldsymbol{e}}\} = D\{\big[I_n - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T P\big]\boldsymbol{y}\} =$$

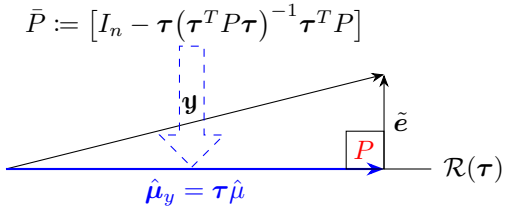

Figure 2.1: Depiction of P-orthogonality between residual vector $\tilde{\boldsymbol{e}}$ and vector of adjusted observations $\boldsymbol{\tau}\hat{\mu}$. The P in the box represents its roll in the orthogonality relationship. The projection matrix $\bar{P}$ is depicted by the big arrow as projecting the observation vector $\boldsymbol{y}$ onto the (one-dimensional) range space of $\boldsymbol{\tau}$. The vectors sum together as $\boldsymbol{y} = \boldsymbol{\tau}\hat{\mu} + \tilde{\boldsymbol{e}}$, just as they should.

$$
\begin{aligned}
&= \left[I_n - \boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T P\right] D\{\boldsymbol{y}\} \left[I_n - P\boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T\right] = \\
&= \sigma_0^2\left[P^{-1} - \boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T\right]\left[I_n - P\boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T\right] = \\
&= \sigma_0^2\left[P^{-1} - \boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T\right] - \sigma_0^2\boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T + \\
&\qquad + \sigma_0^2\boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T P \boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T \Rightarrow \\
&\boxed{D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2\left[P^{-1} - \boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T\right].} \qquad (2.15)
\end{aligned}
$$

It turns out that the last matrix in (2.15) involves the dispersion of the adjusted observations, since

$$\boxed{D\{\hat{\boldsymbol{\mu}}_y\} = \boldsymbol{\tau} D\{\hat{\mu}\}\boldsymbol{\tau}^T = \sigma_0^2\boldsymbol{\tau}\left(\boldsymbol{\tau}^T P \boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}^T.} \qquad (2.16)$$

Formally, neither (2.14) nor (2.15) nor (2.16) can be computed, since the variance component σ_0^2 is unknown, though it can be replaced by its estimate shown in (2.38). From (2.15) we see that the dispersion (variance) of the jth element of $\tilde{\boldsymbol{e}}$ is

$$\sigma^2_{\tilde{\boldsymbol{e}}_j} = \sigma_0^2\Big(\sigma^2_{jj} - \frac{1}{\boldsymbol{\tau}^T P \boldsymbol{\tau}}\Big), \qquad (2.17)$$

where σ^2_{jj} is the jth diagonal element of P^{-1}, and σ_0^2 is the variance component from the model (2.1). Thus it is apparent that the variance of the jth element of the residual vector $\tilde{\boldsymbol{e}}$ is smaller than the variance of the corresponding jth element of the true, but unknown, random error vector $\boldsymbol{e}$.

2.2.1 Equivalency to Arithmetic Mean and Weighted Arithmetic Mean

In the special case where the random errors are iid (i.e., the case of (1.45a)), the LESS (2.7) reduces to $\hat{\mu} = \boldsymbol{\tau}^T\boldsymbol{y}/(\boldsymbol{\tau}^T\boldsymbol{\tau})$, which is equivalent to the arithmetic mean. This is easily seen by noting that $\boldsymbol{\tau}^T\boldsymbol{y} = \sum_{i=1}^n y_i$ and $\boldsymbol{\tau}^T\boldsymbol{\tau} = n$. Therefore

$$\hat{\mu} = \frac{\sum y}{n}, \text{ if } \boldsymbol{e} \sim (\boldsymbol{0}, \text{iid}), \qquad (2.18)$$

which, obviously, is the formula for the arithmetic mean.

In the case where the random errors have a heteroscedastic distribution (i.e., the case of (1.45b) where the weight matrix P is diagonal), the LESS (2.7) is equivalent to the weighted arithmetic mean, since

$$\hat{\mu} = \frac{\boldsymbol{\tau}^T \operatorname{diag}([p_i, \ldots, p_n])\boldsymbol{y}}{\boldsymbol{\tau}^T \operatorname{diag}([p_i, \ldots, p_n])\boldsymbol{\tau}} = \frac{\sum_{i=1}^n p_i y_i}{\sum_{i=1}^n p_i}, \text{ if } \boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2 \operatorname{diag}([1/p_i, \ldots, 1/p_n]). \tag{2.19}$$

2.3 Observation Weighting and Weight Propagation

We start our discussion of observation weighting and weight propagation by showing examples of it. Following that, we give some definitions and rules for general cases.

Assume two measurements y_1 and y_2 with the *same (unknown) expectation* μ and given *variance* (precision) σ^2, i.e.

$$y_i \sim (\mu, \sigma^2) \text{ for } i = 1, 2. \tag{2.20}$$

One of the "most plausible" values for μ as derived from the measurements seems to the the *arithmetic mean*

$$\hat{\mu} := \frac{y_1 + y_2}{2}, \tag{2.21a}$$

which is unbiased since

$$E\{\hat{\mu}\} = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu. \tag{2.21b}$$

Its variance (dispersion) is given by

$$D\{\hat{\mu}\} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \frac{\sigma_1^2}{4} + \frac{\sigma_{12}}{2} + \frac{\sigma_2^2}{4} \tag{2.21c}$$

in general, or

$$D\{\hat{\mu}\} = \frac{\sigma^2}{2}, \text{ assuming } \sigma_{12} := 0 \text{ and } \sigma_1^2 = \sigma_2^2 =: \sigma^2. \tag{2.21d}$$

Now, if the result turns out to be insufficiently precise, i.e. the variance $\sigma^2/2$ is still too large, we are forced to perform a *third measurement* y_3. Assuming *independence* (implying $\sigma_{13} = 0 = \sigma_{23}$) and *identical variance* (implying $\sigma_3^2 = \sigma^2$), we are in the position to from *another arithmetic mean* via

(i) Simply averaging the first result $\hat{\mu}$ with the new observation y_3, i.e.

$$\bar{\hat{\mu}} := \frac{\hat{\mu} + y_3}{2}, \tag{2.22a}$$

which results in an unbiased estimate, since

$$E\{\bar{\hat{\mu}}\} = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu. \tag{2.22b}$$

Its variance is given by

$$D\{\bar{\hat{\mu}}\} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 & \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{2} \end{bmatrix} = \sigma^2 \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{4} \right) = \frac{3\sigma^2}{8}. \tag{2.22c}$$

(ii) Or we may use the arithmetic mean of all three observations via:

$$\hat{\hat{\mu}} := \frac{y_1 + y_2 + y_3}{3}, \tag{2.23a}$$

which is unbiased since

$$E\{\hat{\hat{\mu}}\} = \frac{1}{3}\mu + \frac{1}{3}\mu + \frac{1}{3}\mu = \mu. \tag{2.23b}$$

Its variance is given by

$$D\{\hat{\hat{\mu}}\} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 & \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \sigma^2 \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) = \frac{\sigma^2}{3}. \tag{2.23c}$$

We see that

$$D\{\hat{\hat{\mu}}\} = \frac{\sigma^2}{3} < \frac{3\sigma^2}{8} = D\{\bar{\hat{\mu}}\}, \tag{2.24}$$

and thus we claim that the estimate $\hat{\hat{\mu}}$ *is to be preferred over (is "better than")* $\bar{\hat{\mu}}$, since it is more precise, i.e. has smaller variance.

However, we can form a different linear combination of $\hat{\mu}$ and y_3 that will result in $\hat{\hat{\mu}}$, viz.

$$\hat{\hat{\mu}} = \frac{2 \cdot \hat{\mu} + 1 \cdot y_3}{2 + 1}. \tag{2.25a}$$

But, since

$$D\{\hat{\mu}\} = \frac{\sigma^2}{2} \text{ and } D\{y_3\} = \frac{\sigma^2}{1}, \tag{2.25b}$$

we can also write

$$\hat{\hat{\mu}} = \frac{D\{\hat{\mu}\}^{-1} \cdot \hat{\mu} + D\{y_3\}^{-1} \cdot y_3}{D\{\hat{\mu}\}^{-1} + D\{y_3\}^{-1}}, \tag{2.25c}$$

which is a properly *weighted (arithmetic) mean* of $\hat{\mu}$ and y_3.

Let's take our example one step further by assuming that the third measurement y_3 was performed *twice as precise* as the previous ones, i.e. $\sigma_{y_3} = \sigma/2 \Rightarrow y_3 \sim (\mu, \sigma^2/4)$. The corresponding "most plausible" value of μ would then be the weighted arithmetic mean according to

$$\hat{\hat{\mu}} := \frac{2\cdot\hat{\mu} + 4\cdot y_3}{2+4} = \frac{y_1 + y_2 + 4y_3}{6}, \tag{2.26a}$$

with

$$E\{\hat{\hat{\mu}}\} = \frac{1}{6}\mu + \frac{1}{6}\mu + \frac{4}{6}\mu = \mu, \tag{2.26b}$$

implying that $\hat{\hat{\mu}}$ is an unbiased estimate of μ. Its dispersion is provided by

$$D\{\hat{\hat{\mu}}\} = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 & \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2/4 \end{bmatrix} \begin{bmatrix} \frac{1}{6} \\ \frac{1}{6} \\ \frac{2}{3} \end{bmatrix} = \sigma^2 \left(\frac{1}{36} + \frac{1}{36} + \frac{4}{9}\cdot\frac{1}{4} \right) = \frac{\sigma^2}{6}. \tag{2.26c}$$

Definition: For a set of *uncorrelated random variables* $y_1, \ldots, y_n$, with variances $\sigma_1^2, \ldots, \sigma_n^2$, we define a set of corresponding weights by

$$\boxed{p_j := \frac{\text{const}}{\sigma_j^2} \quad \text{for all } j = 1, \ldots, n,} \tag{2.27}$$

where the constant is to be chosen arbitrarily, but fixed. In this case we obtain the *weight matrix* to be *diagonal* with

$$\boxed{P := \text{diag}(p_j) = \text{const}\cdot \text{diag}(\sigma_j^{-2}) = \text{const}\cdot\Sigma^{-1} \quad \text{for all } j = 1, \ldots, n.} \tag{2.28}$$

Definition: The arbitrarily chosen *constant* σ_0^2 is called *variance component* (or *variance of unit weight* by some authors), yielding the identities

$$\boxed{P := \sigma_0^2\cdot\Sigma^{-1} =: Q^{-1} \Leftrightarrow \Sigma = \sigma_0^2 Q = \sigma_0^2 P^{-1},} \tag{2.29}$$

with Q as $n \times n$ *cofactor matrix*.

Remarks:

(i) The variance component σ_0^2 is unitless by definition.

(ii) The preceding definition (2.29) is general enough for non-diagonal matrices $\Sigma = D\{\boldsymbol{y}\}$, or *correlated random variables* $y_1, \ldots, y_n$, respectively.

2.3.1 Choice of Best Weights

If we choose weights according to the rule (2.29), is that the *best* we can do? By *best* we mean a choice of weights that lead to a minimum variance for the estimate of μ. We also want to ensure that $\hat{\mu}$ remains unbiased. With these objectives in mind, consider the following:

(i) The *weighted (or general) arithmetic mean*

$$\bar{\mu} \coloneqq \sum_{j=1}^{n} \gamma_j y_j \quad \text{with} \quad \sum_{j=1}^{n} \gamma_j = 1, \tag{2.30a}$$

for $y_i \sim (\mu, \sigma_j^2)$ being *mutually uncorrelated, is unbiased since*

$$E\{\bar{\mu}\} = \sum_{j=1}^{n} \gamma_j E\{y_j\} = \mu \cdot \sum_{j=1}^{n} \gamma_j = \mu. \tag{2.30b}$$

This shows that *all* weighted averages are unbiased, implying that over infinitely many measurements they would provide the *true* value for μ.

(ii) The "best variance" of any weighted mean is determined by solving the following minimization problem.

$$D\{\bar{\mu}\} = \sum_{j=1}^{n} \gamma_j^2 \sigma_j^2 = \min_{\gamma_j} \{\sum_{j=1}^{n} \gamma_j = 1\}. \tag{2.31a}$$

The *Lagrange function*

$$\Phi(\gamma_j, \lambda) \coloneqq \sum_{j=1}^{n} \gamma_j^2 \sigma_j^2 - 2\lambda \cdot \big(\sum_{j=1}^{n} \gamma_j - 1\big) = \underset{\gamma_j, \lambda}{\text{stationary}} \tag{2.31b}$$

is formed for minimization of Φ, with λ introduced as a Lagrange multiplier. The Euler-Lagrange necessary conditions

$$\frac{1}{2} \frac{\partial \Phi}{\partial \gamma_j} = \sigma_j^2 \gamma_j - \lambda \doteq 0, \quad \text{for all } j \tag{2.31c}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \lambda} = -\sum_{j=1}^{n} \gamma_j + 1 \doteq 0 \tag{2.31d}$$

lead to a minimum of Φ, since the sufficient condition

$$\frac{1}{2} \frac{\partial^2 \Phi}{\partial \gamma_j^2} = \sigma_j^2 > 0 \tag{2.31e}$$

is fulfilled.

Equation (2.31c) implies that

$$\lambda = \sigma_j^2 \gamma_j \quad \text{for } \textit{all } j = 1, \ldots, n \tag{2.31f}$$

further implying that

$$\sigma_j^2 \gamma_j = \text{const} \Rightarrow \gamma_j = \frac{\text{const}}{\sigma_j^2}. \tag{2.31g}$$

From (2.31d), we have

$$1 = \sum_{j=1}^{n} \gamma_j = \text{const} \cdot \sum_{j=1}^{n} \sigma_j^{-2} \Rightarrow \text{const} = \big(\sum_{j=1}^{n} \sigma_j^{-2}\big)^{-1}, \tag{2.31h}$$

further implying that

$$\gamma_j = \frac{\sigma_j^{-2}}{\sum_{i=1}^{n} \sigma_i^{-2}}, \tag{2.31i}$$

which leads to

$$\boxed{\gamma_j = \frac{p_j}{\sum p_i} \quad \text{for } p_j \coloneqq \frac{1}{\sigma_j^2}.} \tag{2.31j}$$

as an expression for the jth weight γ_j.

Therefore, we can answer the question at the beginning of this section by saying

> If we choose the weights according to rule (2.29), we obtain that *weighted average* having a *minimum variance*, i.e. that which extracts the information out of the measurements in the "best" way.

2.3.2 Examples for Weighting

The following examples illustrate how weights are chosen as the reciprocals of variances when working with quantities that have been derived from observations and that might be combined with other data in an adjustment problem.

1. Leveling:

   Let σ^2 be the *variance* of one leveling setup between consecutive turning points. Then, assuming n different setups for the entire leveling run, we find the *height difference*

$$H_n - H_0 \coloneqq (H_n - H_{n-1}) + \ldots + (H_{j+1} - H_j) + \ldots + (H_1 - H_0) = \sum_{j=1}^{n} h_j,$$
$$\text{for } h_j \coloneqq H_j - H_{j-1}.$$

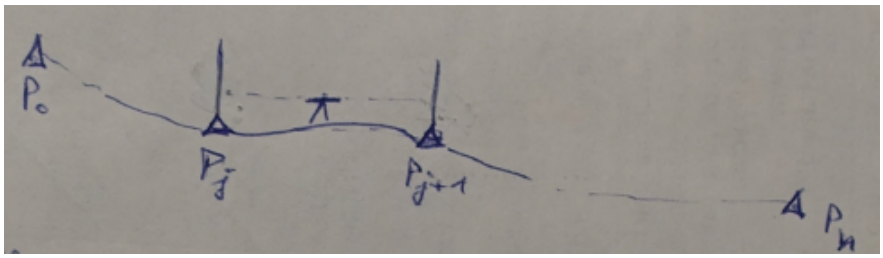

Figure 2.2: A single leveling run from point P_0 to point P_n

Further assuming *uncorrelated* observations, with variance $D\{h_j\} \coloneqq \sigma^2$ for *all* $j = 1, \ldots, n$, the *law of error propagation* gives

$$\boxed{D\{H_n - H_0\} = \sigma^2 \cdot n.}$$

Assuming *equal intervals* of length s_0 between consecutive turning points, we find the equivalent expression

$$\boxed{D\{H_n - H_0\} = (\sigma^2 s_0^{-1}) \cdot S,}$$

if S is the *distance* between point P_0 and P_n along the leveling run, implying that the weights are defined by one over the overall distance S, i.e.

$$\boxed{p \coloneqq S^{-1}.}$$

Here we assume that the interval s_0 is constant among all other leveling runs that may be combined in an adjustment, which is a common case when surveying standards are being adhered to. Thus the term $\sigma^2 s_0^{-1}$ is taken to be a constant "reference variance," and the weighting depends only on the length of the leveling run, which agrees with experience and intuition that suggests longer runs are less precise (thus lower weight) than shorter ones.

2. Horizontal directions: Let φ_j be the average of a set of measured directions to target j and n be the *number of rounds* (or sets) of directions measured. Further assume that the individual directions are uncorrelated and have variance σ^2. Then we find

$$\boxed{D\{\varphi_j\} = \sigma^2 \cdot n^{-1}}$$

as the variance of the averaged direction and

$$\boxed{p_j \coloneqq n}$$

as its corresponding weight. This agrees with experience and intuition that suggests that the more rounds that are measured, the greater the weight to be assigned when combined with other data in an adjustment.

Note, however, that angles from the same round are *correlated*, since they are essentially differences between two directions.

3. Electronic distance measurements: If S_j is a measured distance and ρ_1 and ρ_2 are coefficients from a calibration of the instrument ($\rho_1 > 0$, $\rho_2 > 0$), then the variance of S_j is

$$\boxed{D\{S_j\} = \sigma_0^2(\rho_1 + \rho_2 \cdot S_j^2),}$$

which implies that the corresponding weight is defined by

$$\boxed{p_j := \frac{1}{\rho_1 + \rho_2 \cdot S_j^2}.}$$

2.4 Estimated Variance Component

The variance component σ_0^2 is an unknown quantity in model (2.1). However, it can be estimated as a function of the P-weighted norm of the residual vector $\tilde{\boldsymbol{e}}$ and can be used as a "goodness of fit statistic," a concept discussed in Section 9.4. The estimated variance component is derived as follows:

The LEast-Squares Solution (LESS) within the model of direct observations is shown in (2.7) as

$$\hat{\mu} = \frac{\boldsymbol{\tau}^T P \boldsymbol{y}}{\boldsymbol{\tau}^T P \boldsymbol{\tau}} = \frac{\boldsymbol{\tau}^T \Sigma^{-1} \boldsymbol{y}}{\boldsymbol{\tau}^T \Sigma^{-1} \boldsymbol{\tau}}, \tag{2.32a}$$

so that the P-weighted norm of the residual vector

$$\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = \|\boldsymbol{y} - \boldsymbol{\tau} \cdot \hat{\mu}\|_P^2 = (\boldsymbol{y} - \boldsymbol{\tau} \cdot \hat{\mu})^T P (\boldsymbol{y} - \boldsymbol{\tau} \cdot \hat{\mu}) \tag{2.32b}$$

is a *random variable* with expectation

$$\begin{aligned}
&E\{(\boldsymbol{y} - \boldsymbol{\tau} \cdot \hat{\mu})^T P (\boldsymbol{y} - \boldsymbol{\tau} \cdot \hat{\mu})\} = \qquad (2.33a)\\
&\quad = E\{\boldsymbol{y}^T P \boldsymbol{y}\} - E\{\boldsymbol{y}^T P \boldsymbol{\tau} \cdot \hat{\mu}\} - \cancel{E\{\hat{\mu} \boldsymbol{\tau}^T P \boldsymbol{y}\}} + \cancel{E\{\hat{\mu}^2 \boldsymbol{\tau}^T P \boldsymbol{\tau}\}} =\\
&\quad = \operatorname{tr} E\{P \boldsymbol{y} \boldsymbol{y}^T\} - (\boldsymbol{\tau}^T P \boldsymbol{\tau})^{-1} \cdot \operatorname{tr} E\{\boldsymbol{\tau}^T P \boldsymbol{y} \boldsymbol{y}^T P \boldsymbol{\tau}\} =\\
&\quad = \operatorname{tr}\big[P \cdot E\{\boldsymbol{y} \boldsymbol{y}^T\}\big] - (\boldsymbol{\tau}^T P \boldsymbol{\tau})^{-1} \cdot \operatorname{tr}\big[P \boldsymbol{\tau} \boldsymbol{\tau}^T P \cdot E\{\boldsymbol{y} \boldsymbol{y}^T\}\big] =\\
&\quad = \operatorname{tr}\big[P \cdot D\{\boldsymbol{y}\}\big] + \operatorname{tr}\big[P \cdot E\{\boldsymbol{y}\} E\{\boldsymbol{y}\}^T\big] -\\
&\qquad - \operatorname{tr}\big[P \boldsymbol{\tau} (\boldsymbol{\tau}^T P \boldsymbol{\tau})^{-1} \boldsymbol{\tau}^T P (\sigma_0^2 P^{-1})\big] - \operatorname{tr}\big[P \boldsymbol{\tau} (\boldsymbol{\tau}^T P \boldsymbol{\tau})^{-1} \boldsymbol{\tau}^T P \boldsymbol{\tau} \mu^2 \boldsymbol{\tau}^T\big] =\\
&\quad = \sigma_0^2 \operatorname{tr} I_n + \cancel{\mu^2 \boldsymbol{\tau}^T P \boldsymbol{\tau}} - \sigma_0^2 \cdot \frac{\boldsymbol{\tau}^T P \cancel{\boldsymbol{\tau}}}{\cancel{\boldsymbol{\tau}^T} P \boldsymbol{\tau}} - \cancel{\mu^2 \cdot \boldsymbol{\tau}^T P \boldsymbol{\tau}} =\\
&\quad = \sigma_0^2 (n-1) \Rightarrow\\
&E\{(n-1)^{-1} (\boldsymbol{y} - \boldsymbol{\tau} \cdot \hat{\mu})^T P (\boldsymbol{y} - \boldsymbol{\tau} \cdot \hat{\mu})\} = \sigma_0^2 \qquad (2.33b)
\end{aligned}$$

The quantity $n - 1$ is called the *redundancy* or *degrees of freedom* of the model.

Now, we may take the argument of the expectation shown in (2.33b) and assign it the symbol $\hat{\sigma}_0^2$, implying that

$$\boxed{\begin{aligned}\hat{\sigma}_0^2 &= (n-1)^{-1}(\boldsymbol{y}-\boldsymbol{\tau}\cdot\hat{\mu})^T P(\boldsymbol{y}-\boldsymbol{\tau}\cdot\hat{\mu}) = && (2.34a)\\ &= (n-1)^{-1}(\boldsymbol{y}^T P\boldsymbol{y} - \hat{\mu}\cdot\boldsymbol{\tau}^T P\boldsymbol{y}) = && (2.34b)\\ &= \frac{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}}{n-1}, && (2.34c)\end{aligned}}$$

which is an *unbiased estimate* of σ_0^2, since $E\{\hat{\sigma}_0^2\} = \sigma_0^2$.

Remark In fact, $\hat{\sigma}_0^2$ is the "best" in a certain class of quadratic unbiased estimates of σ_0^2 (being invariant with respect to translations in μ) and has — under normality assumptions — dispersion

$$D\{\hat{\sigma}_0^2\} = 2(\sigma_0^2)^2(n-1)^{-1} = \mathrm{MSE}\{\hat{\sigma}_0^2\}. \tag{2.35}$$

In summary, we can write the so-called *sum of squared residuals (SSR)* as

$$\boxed{\Omega := \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}},} \tag{2.36}$$

which, together with the redundancy of the model

$$\boxed{r := n-1,} \tag{2.37}$$

comprises the formula

$$\boxed{\hat{\sigma}_0^2 := \frac{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}}{r}} \tag{2.38}$$

for the estimated variance component.

2.5 Computation Checks and an Example

2.5.1 Checks on Residuals

A statistical analysis of the results of various adjustment computations is the subject of Chapter 9, where tests for goodness of fit, detection of outliers, and for particular values of the estimated parameters are presented. But even before statistical analysis is employed, certain checks should be made on the residuals to confirm that they look reasonable and to assure that the computations were made correctly. Below is a minimal list of checks that should be made after any adjustment computation.

1. Inspect the elements of the residual vector $\tilde{\boldsymbol{e}}$ to make sure they look reasonable. As a general rule, if any residual is much greater than three times the square root of the given variance of its corresponding observation, the accuracy of the observation, or the validity of the given variance, might be questionable. In that case, the corresponding observation could be temporarily removed and

the adjustment computed again. Then, a residual could be *predicted* for the removed observation and the results inspected to decide if the observation should be retained or not. The method of outlier detection described in Section 9.7 is meant to help lead to decisions about the accuracy of a suspect observation and whether or not to admit it in the final adjustment.

2. Consider the magnitude of the estimated variance component $\hat{\sigma}_0^2$. Is it close to the value you expect it to take on (perhaps 1)? If it varies largely from the value you expect it to take on, it will generally indicate that either 1) the observational model is inaccurate, or 2) the weights (or variances) have not been accurately specified, or both.

   In the case of 1, the model may need to be revised to include more parameters so that the parameters of the model more accurately explain the observations. (Of course, then we would no longer have a model of direct observations with a single parameter μ.) Or, it could be that the observations contain some systematic errors that need to be removed so that the assumption that $E\{\boldsymbol{y}\} = \boldsymbol{\tau}\mu$ is made valid.

   In the case of 2, a relative small value of $\hat{\sigma}_0^2$ suggests that the specified observational variances (reflected in the cofactor matrix $Q = P^{-1}$) were too large (i.e. the observations are more precise than reflected in the cofactor matrix). Conversely, if $\hat{\sigma}_0^2$ turns out to be relatively large, the specified variances in Q might be too small (i.e. the observations are less precise than reflected in the cofactor matrix).

3. Provided the model redundancy is large enough, say greater than 10 or 20, we might expect that approximately half the residuals will be negative and about half positive. Certainly this would be the case if the random observation errors turned out to be normally distributed. So, it is prudent to check the ratio of negative to positive residuals and make sure the ratio is not greatly different than 1. Note that this check might not apply for adjustments within the model of condition equations discussed in Chapter 4.

4. Going beyond the previous item, if the redundancy is large enough, say greater than 10 or 20, a histogram of the residuals should be plotted to check how closely its shape resembles the pdf curve of a normal distribution, if it is assumed that the random observation errors are approximately normally distributed.

5. Compute the estimated variance component both by (2.34b) and (2.34c) and make sure they are equivalent up to the precision of the computations.

6. Compute the trace of the matrix of redundancy numbers as defined in (9.84a) and (9.84b) and confirm that the result is an integer that equals the redundancy of the model r.

2.5.2 Example of LESS Within the Model of Direct Observations

Given the following vector of observations $\boldsymbol{y}$ and its associated dispersion matrix $D\{\boldsymbol{y}\}$, compute the LESS for

1. The estimated parameter $\hat{\mu}$ and its estimated dispersion.
2. The estimated variance component $\hat{\sigma}_0^2$.
3. The vector of predicted residuals $\tilde{\boldsymbol{e}}$ and its estimated dispersion matrix.

$$\boldsymbol{y} = \begin{bmatrix} 100.02\,\mathrm{m} \\ 100.04\,\mathrm{m} \\ 99.97\,\mathrm{m} \end{bmatrix}, \quad D\{\boldsymbol{y}\} = \sigma_0^2 \begin{bmatrix} 1 & 1/2 & 0 \\ 1/2 & 1 & 0 \\ 0 & 0 & 9 \end{bmatrix} \mathrm{cm}^2.$$

Solution: To simplify the problem somewhat, we may subtract 100 m from the observations and solve for $\delta\hat{\mu}$ as an intermediate step, working with cm instead of meters. Then, the modified observation vector reads $\boldsymbol{y} \to \boldsymbol{y} = [2\,\mathrm{cm},\ 4\,\mathrm{cm},\ -3\,\mathrm{cm}]^T$. Weight matrix:

$$\Rightarrow P = \begin{bmatrix} 4/3 & -2/3 & 0 \\ -2/3 & 4/3 & 0 \\ 0 & 0 & 1/9 \end{bmatrix} \mathrm{cm}^{-2} = \frac{1}{9} \cdot \begin{bmatrix} 12 & -6 & 0 \\ -6 & 12 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathrm{cm}^{-2}$$

Estimated parameter and its estimated variance:

$$\delta\hat{\mu} = \frac{\boldsymbol{\tau}^T P \boldsymbol{y}}{\boldsymbol{\tau}^T P \boldsymbol{\tau}} = \frac{\left[2/3 \mid 2/3 \mid 1/9\right] \boldsymbol{y}}{\left[2/3 \mid 2/3 \mid 1/9\right] \boldsymbol{\tau}} = \frac{(11/3)\,\mathrm{cm}^{-1}}{(13/9)\,\mathrm{cm}^{-2}} = \frac{33}{13}\mathrm{cm}$$

$$\Rightarrow \boxed{\hat{\mu} = 100\,\mathrm{m} + \delta\hat{\mu} = 100.0254\,\mathrm{m}}$$

$$\text{with} \quad \boxed{D\{\hat{\mu}\} = \frac{\sigma_0^2}{\boldsymbol{\tau}^T P \boldsymbol{\tau}} = \frac{9}{13}\sigma_0^2\,\mathrm{cm}^2}$$

$$\text{Also} \quad \boldsymbol{y}^T P \boldsymbol{y} = \left[0 \mid 12/3 \mid -1/3\right] \boldsymbol{y} = 17 = 221/13$$

$$\text{and} \quad \delta\hat{\mu} \cdot \boldsymbol{\tau}^T P \boldsymbol{y} = \left(\frac{33}{13}\right) \cdot \left(\frac{11}{3}\right) = \frac{121}{13} \Rightarrow \boxed{\hat{\sigma}_0^2 = \frac{100/13}{2} = \frac{50}{13}}$$

$$\Rightarrow \hat{D}\{\hat{\mu}\} = \left(\frac{9}{13}\right) \cdot \left(\frac{50}{13}\right) = \frac{450}{169} = 2.66\,\mathrm{cm}^2$$

$$\boxed{\hat{\mu} = (100.025 \pm 0.016)\,\mathrm{m}}$$

Predicted residuals and their estimated covariance matrix:

$$\boxed{\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{\tau}\hat{\mu} = \left[-0.5385 \mid +1.4615 \mid -5.5385\right]^T \mathrm{cm}}$$

$$\text{with} \quad D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \Big(\begin{bmatrix} 1 & 1/2 & 0 \\ 1/2 & 1 & 0 \\ 0 & 0 & 9 \end{bmatrix} - \begin{bmatrix} 9/13 & 9/13 & 9/13 \\ 9/13 & 9/13 & 9/13 \\ 9/13 & 9/13 & 9/13 \end{bmatrix} \Big) \cdot \text{cm}^2 \Rightarrow$$

$$\boxed{D\{\tilde{\boldsymbol{e}}\} = \frac{\sigma_0^2}{13} \begin{bmatrix} 4 & -5/2 & -9 \\ -5/2 & 4 & -9 \\ -9 & -9 & 108 \end{bmatrix} \cdot \text{cm}^2}$$

Checks:

$$P\tilde{\boldsymbol{e}} = \begin{bmatrix} -1.6923 \\ +2.3077 \\ -0.6154 \end{bmatrix} \Rightarrow \boldsymbol{\tau}^T P\tilde{\boldsymbol{e}} = -0.00006 \checkmark \text{ and } \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} = 7.69 \approx 100/13 = \hat{\sigma}_0^2 \cdot 2 \checkmark$$

"Redundancy numbers" (see (9.84a) and (9.84b) for definition of redundancy numbers)

$$D\{\tilde{\boldsymbol{e}}\} \cdot P = \frac{\sigma_0^2}{13} \begin{bmatrix} 7 & -6 & -1 \\ -6 & 7 & -1 \\ -6 & -6 & 12 \end{bmatrix} \Rightarrow \text{tr}(D\{\tilde{\boldsymbol{e}}\} \cdot P)/\sigma_0^2 = \frac{26}{13} = 2 = 3 - 1 = r \checkmark$$

2.6 Best Linear Uniformly Unbiased Estimate

Here we take a statistical approach to estimating the unknown parameter μ. We want to find an estimate for μ, expressed as a linear combination of the observations $\boldsymbol{y}$, that extracts the "best" information from the data. The estimate is denoted by $\hat{\mu}$ and is characterized as the *Best Linear Uniformly Unbiased Estimate* (BLUUE) of μ. The three criteria used to derive the BLUUE are described as follows:

1. Linear criterion: The linear criterion states that the estimated parameter must be a linear combination of the data contained in $\boldsymbol{y}$, i.e.

$$\hat{\mu} = \boldsymbol{\alpha}^T \boldsymbol{y}, \tag{2.39a}$$

   where $\boldsymbol{\alpha}$ is an unknown vector to be determined.

2. Uniformly Unbiased criteria: An unbiased estimator is one for which its expectation is equal to the true, but unknown, quantity it estimates. Stated mathematically,

   $$\mu = E\{\hat{\mu}\} = E\{\boldsymbol{\alpha}^T \boldsymbol{y}\} = \boldsymbol{\alpha}^T E\{\boldsymbol{y}\} = \boldsymbol{\alpha}^T E\{\boldsymbol{\tau}\mu + \boldsymbol{e}\} = \boldsymbol{\alpha}^T \boldsymbol{\tau}\mu, \text{ for } \textit{any } \mu \in \mathbb{R},$$

   which implies

$$\boldsymbol{\alpha}^T \boldsymbol{\tau} = 1. \tag{2.39b}$$

   Requiring this condition to hold for *any* $\mu \in \mathbb{R}$ satisfies the "uniform" criterion, whereas the requirement that $\boldsymbol{\alpha}^T \boldsymbol{\tau} = 1$ satisfies the "unbiased" criterion.

3. Best criterion: The best criterion requires minimum MSE($\hat{\mu}$), or, equivalently, minimum dispersion, *since $\hat{\mu}$ is unbiased*. Mathematically, the criterion reads

$$\begin{aligned} &\min D\{\hat{\mu}\}, \quad \text{where } D\{\hat{\mu}\} = D\{\boldsymbol{\alpha}^T\boldsymbol{y}\} = \boldsymbol{\alpha}^T D\{\boldsymbol{y}\}\boldsymbol{\alpha} \Rightarrow \\ &\min D\{\hat{\mu}\} = \sigma_0^2 \boldsymbol{\alpha}^T Q \boldsymbol{\alpha}, \text{ subject to } \boldsymbol{\tau}^T\boldsymbol{\alpha} = 1. \end{aligned} \tag{2.39c}$$

Accordingly, a Lagrange target function is formed by

$$\Phi(\boldsymbol{\alpha}, \lambda) \coloneqq \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} + 2\lambda\big(\boldsymbol{\tau}^T\boldsymbol{\alpha} - 1\big). \tag{2.40}$$

The necessary conditions for stationarity are provided by the Euler-Lagrange equations, which are written as

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\alpha}} = Q\hat{\boldsymbol{\alpha}} + \boldsymbol{\tau}\hat{\lambda} \doteq 0, \tag{2.41a}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\lambda} = \boldsymbol{\tau}^T\hat{\boldsymbol{\alpha}} - 1 \doteq 0. \tag{2.41b}$$

The sufficient condition for minimization is satisfied by $\partial\Phi^2/(\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^T) = 2Q$, which is a positive definite matrix according to (2.1). Solving (2.41a) and (2.41b) simultaneously yields

$$\hat{\boldsymbol{\alpha}} = -Q^{-1}\boldsymbol{\tau}\hat{\lambda} = -P\boldsymbol{\tau}\hat{\lambda} \text{ using (2.41a)}, \tag{2.42a}$$

$$1 = \boldsymbol{\tau}^T\hat{\boldsymbol{\alpha}} = -\boldsymbol{\tau}^T P\boldsymbol{\tau}\hat{\lambda} \Rightarrow \hat{\lambda} = \frac{-1}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} \text{ using (2.41b) and (2.42a)}. \tag{2.42b}$$

Substituting (2.42b) into (2.42a) we get

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{\tau}^T P\boldsymbol{\tau})^{-1} P\boldsymbol{\tau}. \tag{2.42c}$$

Finally, substituting the transpose of (2.42c) into the linear requirement $\hat{\mu} = \boldsymbol{\alpha}^T\boldsymbol{y}$ yields the BLUUE of μ as

$$\boxed{\hat{\mu} = \frac{\boldsymbol{\tau}^T P\boldsymbol{y}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}.} \tag{2.43}$$

Equation (2.43) agrees with (2.7) derived for LESS. Thus we see that the LESS and the BLUUE are equivalent within the model of direct observations.

We may also prove mathematically that (2.43) fulfills the weighted LESS principle by showing that the P-weighted residual norm $\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$ for any other solution is larger than that obtained via BLUUE, which we do in the following: Suppose $\hat{\tilde{\mu}}$ is any other estimate for μ, then

$$\begin{aligned} \hat{\tilde{\boldsymbol{e}}}^T P\hat{\tilde{\boldsymbol{e}}} &= \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\tilde{\mu}}\big)^T P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\tilde{\mu}}\big) = \\ &= \big[\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) - \boldsymbol{\tau}\big(\hat{\tilde{\mu}} - \hat{\mu}\big)\big]^T P\big[\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) - \boldsymbol{\tau}\big(\hat{\tilde{\mu}} - \hat{\mu}\big)\big] = \\ &= \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)^T P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) - 2\big(\hat{\tilde{\mu}} - \hat{\mu}\big)\underbrace{\boldsymbol{\tau}^T P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)}_{0} + \big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)\big(\hat{\tilde{\mu}} - \hat{\mu}\big)^2 = \\ &= \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)^T P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) + \big(\boldsymbol{\tau}^T P\boldsymbol{\tau}\big)\big(\hat{\tilde{\mu}} - \hat{\mu}\big)^2 \geq \end{aligned}$$

$$\geq \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)^T P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) = \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$$

Q.E.D.

We have used the P-orthogonality relation (2.11) in the third line of the proof.

Let us briefly summarize these results by stating three important properties of the least-squares solution (LESS) of the unknown parameter μ within the model of direct observations.

The LESS (equivalently BLUUE) within the model of direct observations provides

1. An unbiased estimate $\hat{\mu}$ of the unknown parameter μ, i.e. $E\{\hat{\mu}\} = \mu$.
2. A minimum P-weighted norm of the residual vector, i.e. $\Omega := \|\tilde{\boldsymbol{e}}\|_P^2$ is minimized.
3. A minimum variance (dispersion) $D\{\hat{\mu}\}$.

2.7 Effects of a Wrongly Chosen Weight Matrix in the Model of Direct Observations

Assume that the weight matrix P has been wrongly chosen by an amount δP, where δP is assumed to be a small, positive (semi-)definite matrix that is *uncorrelated* with P. (Apparently δP itself would not have to be positive(semi-)definite as long as the sum $(P + \delta P)$ is positive definite.) Consequently, we have

$$P \to (P + \delta P) \Rightarrow \hat{\mu} \to (\hat{\mu} + \delta\hat{\mu}),\ D\{\hat{\mu}\} \to D\{\hat{\mu} + \delta\hat{\mu}\},\ \text{and}\ \hat{\sigma}_0^2 \to \hat{\sigma}_0^2 + \delta\hat{\sigma}_0^2. \quad (2.44)$$

2.7.1 Effect on the Parameter Estimate

The following shows the effect of a wrongly chosen weight matrix on the estimated parameter $\hat{\mu}$:

$$\big(\hat{\mu} + \delta\hat{\mu}\big) = \frac{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{y}}{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{\tau}} \Rightarrow$$

$$\delta\hat{\mu} = \frac{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{y}}{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{\tau}} - \hat{\mu} = \frac{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{y}}{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{\tau}} \cdot \frac{\boldsymbol{\tau}^T P\boldsymbol{\tau}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}} - \left(\frac{\boldsymbol{\tau}^T P\boldsymbol{y}}{\boldsymbol{\tau}^T P\boldsymbol{\tau}}\right) \cdot \frac{\boldsymbol{\tau}^T\big(P + \delta P\big)\boldsymbol{\tau}}{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{\tau}} =$$

$$= \frac{\cancel{\boldsymbol{\tau}^T P\boldsymbol{y}\boldsymbol{\tau}^T P\boldsymbol{\tau}} + \boldsymbol{\tau}^T\delta P\boldsymbol{y}\boldsymbol{\tau}^T P\boldsymbol{\tau} - \cancel{\boldsymbol{\tau}^T P\boldsymbol{y}\boldsymbol{\tau}^T P\boldsymbol{\tau}} - \boldsymbol{\tau}^T P\boldsymbol{y}\boldsymbol{\tau}^T\delta P\boldsymbol{\tau}}{(\boldsymbol{\tau}^T P\boldsymbol{\tau})\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{\tau}} =$$

$$= \frac{\boldsymbol{\tau}^T\delta P\boldsymbol{y}}{\boldsymbol{\tau}^T\big(P + \delta P\big)\boldsymbol{\tau}} - \frac{\boldsymbol{\tau}^T\delta P\boldsymbol{\tau}\hat{\mu}}{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{\tau}} = \frac{\boldsymbol{\tau}^T\delta P\big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big)}{\boldsymbol{\tau}^T(P + \delta P)\boldsymbol{\tau}}$$

Finally, we arrive at

$$\boxed{\delta\hat{\mu} = \frac{\boldsymbol{\tau}^T \delta P}{\boldsymbol{\tau}^T (P + \delta P)\boldsymbol{\tau}} \tilde{\boldsymbol{e}}.} \tag{2.45}$$

2.7.2 Effect on the Cofactor Matrix for the Estimated Parameter

The following shows the effect of a wrongly chosen weight matrix on the cofactor matrix $Q_{\hat{\mu}}$ for the estimated parameter $\hat{\mu}$, where $D\{\hat{\mu}\} = \sigma_0^2 Q_{\hat{\mu}}$ is the dispersion of $\hat{\mu}$:

$$\delta Q_{\hat{\mu}} = \big(Q_{\hat{\mu}} + \delta Q_{\hat{\mu}}\big) - Q_{\hat{\mu}} = \frac{1}{\boldsymbol{\tau}^T (P + \delta P)\boldsymbol{\tau}} - \frac{1}{\boldsymbol{\tau}^T P \boldsymbol{\tau}} =$$
$$= \frac{\boldsymbol{\tau}^T P \boldsymbol{\tau} - \boldsymbol{\tau}^T (P + \delta P)\boldsymbol{\tau}}{\big(\boldsymbol{\tau}^T P \boldsymbol{\tau}\big)\boldsymbol{\tau}^T (P + \delta P)\boldsymbol{\tau}} = \frac{-\boldsymbol{\tau}^T \delta P \boldsymbol{\tau}}{\big(\boldsymbol{\tau}^T P \boldsymbol{\tau}\big)\boldsymbol{\tau}^T (P + \delta P)\boldsymbol{\tau}}.$$

Thus we have

$$\boxed{\delta Q_{\hat{\mu}} = -\frac{\boldsymbol{\tau}^T \delta P \boldsymbol{\tau}}{\boldsymbol{\tau}^T (P + \delta P)\boldsymbol{\tau}} Q_{\hat{\mu}}.} \tag{2.46}$$

2.7.3 Effect on the Estimated Variance Component

The following shows the effect of a wrongly chosen weight matrix on the estimated variance component: First note that

$$\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = \big(\boldsymbol{y}^T - \hat{\mu}\boldsymbol{\tau}^T\big) P \big(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}\big) =$$
$$= \boldsymbol{y}^T P(\boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}) - \hat{\mu}\big(\boldsymbol{\tau}^T P \boldsymbol{y} - \cancel{\boldsymbol{\tau}^T P \boldsymbol{\tau} \frac{\boldsymbol{\tau}^T P \boldsymbol{y}}{\boldsymbol{\tau}^T P \boldsymbol{\tau}}}\big) =$$
$$= \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{y}^T P \boldsymbol{\tau}\hat{\mu} = \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{\tau}^T P \boldsymbol{y}\hat{\mu} = \boldsymbol{y}^T P \boldsymbol{y} - \hat{\mu}^2 \boldsymbol{\tau}^T P \boldsymbol{\tau}.$$

Following the above logic, we have

$$(n-1)\big(\hat{\sigma}_0^2 + \delta\hat{\sigma}_0^2\big) = \boldsymbol{y}^T (P + \delta P)\boldsymbol{y} - \boldsymbol{\tau}^T (P + \delta P)\boldsymbol{y}\big(\hat{\mu} + \delta\hat{\mu}\big) \Rightarrow$$
$$\Rightarrow (n-1)\delta\hat{\sigma}_0^2 = \boldsymbol{y}^T (\cancel{P} + \delta P)\boldsymbol{y} - \boldsymbol{\tau}^T (P + \delta P)\boldsymbol{y}(\hat{\mu} + \delta\hat{\mu}) - \cancel{\boldsymbol{y}^T P \boldsymbol{y}} + (\boldsymbol{\tau}^T P \boldsymbol{y})\hat{\mu} =$$

(Note: the last term will cancel one of the four terms in the binomial product.)

$$= \boldsymbol{y}^T (\delta P)\boldsymbol{y} - \boldsymbol{\tau}^T \delta P \boldsymbol{y}\big(\hat{\mu} + \delta\hat{\mu}\big) - \big(\boldsymbol{\tau}^T P \boldsymbol{y}\big)\delta\hat{\mu} =$$
$$= \boldsymbol{y}^T (\delta P)\boldsymbol{y} - \hat{\mu}\boldsymbol{\tau}^T (\delta P)\boldsymbol{y} - \boldsymbol{\tau}^T (P + \delta P)\boldsymbol{y}\delta\hat{\mu} =$$
$$= \big(\boldsymbol{y}^T - \hat{\mu}\boldsymbol{\tau}^T\big)(\delta P)\boldsymbol{y} - \boldsymbol{\tau}^T (P + \delta P)\boldsymbol{y}\delta\hat{\mu} =$$
$$= \tilde{\boldsymbol{e}}^T (\delta P)\boldsymbol{y} - \frac{\boldsymbol{\tau}^T (P + \delta P)\boldsymbol{y}}{\boldsymbol{\tau}^T (P + \delta P)\boldsymbol{\tau}} \boldsymbol{\tau}^T (\delta P)\tilde{\boldsymbol{e}} =$$

(Note that the previous results for $\delta\hat{\mu}$ have been substituted in the line above.)

$$= \boldsymbol{y}^T (\delta P)\tilde{\boldsymbol{e}} - \big(\hat{\mu} + \delta\hat{\mu}\big)\boldsymbol{\tau}^T (\delta P)\tilde{\boldsymbol{e}} =$$

$$\begin{aligned}(\text{Using } \boldsymbol{y}^T(\delta P)\tilde{\boldsymbol{e}} &= \big(\hat{\mu}\boldsymbol{\tau}^T + \tilde{\boldsymbol{e}}^T\big)\delta P\tilde{\boldsymbol{e}} = \tilde{\boldsymbol{e}}^T\delta P\tilde{\boldsymbol{e}} + \hat{\mu}\boldsymbol{\tau}^T\delta P\tilde{\boldsymbol{e}})\\ &= \tilde{\boldsymbol{e}}^T(\delta P)\tilde{\boldsymbol{e}} - \delta\hat{\mu}\boldsymbol{\tau}^T(\delta P)\tilde{\boldsymbol{e}} \Rightarrow\\ (n-1)\delta\hat{\sigma}_0^2 &= \tilde{\boldsymbol{e}}^T(\delta P)\tilde{\boldsymbol{e}} - \big(\delta\hat{\mu}\big)^2\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}\end{aligned}$$

Finally, we arrive at

$$\boxed{\delta\hat{\sigma}_0^2 = \frac{1}{n-1}\big[\tilde{\boldsymbol{e}}^T(\delta P)\tilde{\boldsymbol{e}} - \big(\delta\hat{\mu}\big)^2\boldsymbol{\tau}^T(P+\delta P)\boldsymbol{\tau}\big].} \tag{2.47}$$

2.7.4 Effect on the Estimated Dispersion

The the effect of a wrongly chosen weight matrix on the estimated dispersion of $\hat{\mu}$ is obviously given by

$$\boxed{\hat{D}\{\hat{\mu}+\delta\hat{\mu}\} = \big(\hat{\sigma}_0^2+\delta\hat{\sigma}_0^2\big)D\{\hat{\mu}+\delta\hat{\mu}\} = \big(\hat{\sigma}_0^2+\delta\hat{\sigma}_0^2\big)\big(Q_{\hat{\mu}}+\delta Q_{\hat{\mu}}\big).} \tag{2.48}$$

2.8 Practice Problems

1. Show that the LESS of (2.7) is an unbiased estimate of μ.
2. Show that the residual vector of (2.8a) is an unbiased prediction of $\boldsymbol{e}$.
3. Consider the problem of repeated measurements where an unknown distance μ between two points was directly observed n times. The observations are collected in the vector $\boldsymbol{y} = [y_1,\, y_2,\, \ldots, y_n]^T$. The distribution of their random errors is described by $\boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2\sigma^2 I_n)$; furthermore $E\{\boldsymbol{y}\} = \boldsymbol{\tau}\mu$.
   (a) If the random variable z is defined by $z = (y_1 + y_2 + \ldots + y_n)/n$, show that $E\{z\} = E\{\hat{\mu}\}$ as shown in (2.7) and that $D\{z\} = D\{\hat{\mu}\}$ as shown in (2.14).
   (b) Assuming that $\sigma_0^2 = 1$ and $\sigma^2 = 1\,\text{cm}^2$, graph the dispersion of $\hat{\mu}$ as a function of the number of observations n from $n = 2$ to $n = 100$.
   (c) Now suppose that there is correlation between successive observation errors described by the relations
   $$\rho_{e_i,e_{i+1}} = 0.001/\sigma^2,\;\; \rho_{e_i,e_{i+2}} = 0.0008/\sigma^2,$$
   $$\rho_{e_i,e_{i+3}} = -0.00006/\sigma^2,\;\; \text{for } i = 1,\ldots,n-3.$$
   Using assumed values $\sigma_0^2 = 1$ and $\sigma^2 = 1\,\text{cm}^2$, compute $D\{\hat{\mu}\}$ for $n = 100$.
   (d) Repeat item (b) for the case of item (c).
4. Twelve direct observations of one unknown parameter μ are listed in Table 2.1. The first set of five observations (I) were made at one time and have measurement variance $\sigma_I^2 = (0.05)^2$. The second set of seven observations (II) were made at a later time with measurement variance $\sigma_{II}^2 = (0.10)^2$. All random measurement errors are independent. No units are given.

Table 2.1: Twelve direct observations of one unknown parameter μ

Set I, $\sigma_I^2 = (0.05)^2$						
	y_1	y_2	y_3	y_4	y_5	
	9.99	10.04	9.93	9.88	9.93	
Set II, $\sigma_{II}^2 = (0.10)^2$						
y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}
10.03	10.04	10.05	9.99	10.02	9.95	10.09

(a) Using only data set I:

i. Compute the BLUUE (or LESS) $\hat{\mu}$.

ii. Compute the dispersion $D\{\hat{\mu}\}$ (no hat on D).

iii. Compute the residual vector $\tilde{\boldsymbol{e}}$.

iv. Compute the estimated variance component $\hat{\sigma}_0^2$.

(b) Now using only data set II, repeat items i–iv, this time denoting the computed values as $\hat{\hat{\mu}}$, $D\{\hat{\hat{\mu}}\}$, $\tilde{\tilde{\boldsymbol{e}}}$, and $\hat{\hat{\sigma}}_0^2$, respectively.

(c) Based on the discussion above about the weighted arithmetic mean, try to estimate the unknown parameter based on the linear combination

$$\hat{\hat{\hat{\mu}}} = \begin{bmatrix} \alpha_1 & \alpha_2 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\hat{\mu}} \end{bmatrix},$$

using the dispersions computed in the previous two items to determine the "weights" α_1 and α_2. Repeat items i and ii for this case, this time denoting the computed values as $\hat{\hat{\hat{\mu}}}$ and $D\{\hat{\hat{\hat{\mu}}}\}$, respectively.

(d) Now compute i–iv using all 12 observation simultaneously and compare your results to those computed in the preceding items. Comment on your findings.

5. Stellar observations were made in order to determine the astronomical azimuth between two geodetic control points. Table 2.2 shows the arc-minute and arc-second parts of each observation. The degrees part is 126° for all observations. The observations are considered to be *uncorrelated*. The first 12 observations were determined from sightings on Polaris with a precision of $\sigma_1 = 05''$. The remaining 18 observations were determined by sightings on the Sun with a less precise instrument than that used for the first 12. The precision of these observations is $\sigma_2 = 10''$.

(a) Compute the LESS within the model of direct observations for the estimated parameter $\hat{\mu}$, its estimated dispersion $\hat{D}\{\hat{\mu}\}$, and the estimated variance component $\hat{\sigma}_0^2$.

Table 2.2: Observations of the astronomical azimuth (in minutes and seconds of arc) between two points. Add 126° to all values.

No.	Direction	No.	Direction	No.	Direction
1	11′34″	11	11′34″	21	11′19″
2	11′30″	12	11′38″	22	11′22″
3	11′34″	13	11′35″	23	11′01″
4	11′29″	14	11′40″	24	11′44″
5	11′29″	15	11′37″	25	11′33″
6	11′37″	16	11′27″	26	11′23″
7	11′37″	17	11′33″	27	11′44″
8	11′37″	18	11′22″	28	11′13″
9	11′33″	19	11′39″	29	11′29″
10	11′24″	20	11′19″	30	10′38″

(b) Repeat the previous part 30 times ($i = 1, \ldots, 30$), removing one successive observation each time so that each ith solution is based on 29 observations. Tabulate your results and include in each line the difference between the removed observation y_{removed} and the estimated azimuth $\hat{\mu}_i$; let's refer to it as $e_{\text{predicted}} = y_{\text{removed}} - \hat{\mu}_i$. Highlight the solution that has the largest magnitude for $e_{\text{predicted}}$. Call it solution k for reference in the next part.

(c) Now repeat part (a) using all 30 observations, but this time modify the weight of the observation with the value for $e_{\text{predicted}}$ found in solution k of part (b). Use $1/(e_{\text{predicted}})_k^2$ for the new weight. Compare your solution to solution k from part (b). Are they close? Do you expect them to be? Why or why not?

Which of the 32 solutions that you computed would you adopt as the final solution? Give a justification for your choice.

6. Consider the weight matrix $P := P_{(5.a)}$ used in problem 5.a to have been "wrongly chosen" and the weight matrix used in 5.c to be legitimate. Let δP be their difference such that $P_{(5.c)} = P + \delta P$.

Compute the effects of the wrongly chosen weight matrix on the estimated parameter $\hat{\mu}$, its estimated dispersion $\hat{D}\{\hat{\mu}\}$, and the estimated variance component $\hat{\sigma}_0^2$.

Note that the root problem with 5.a is that its last observation appears to be an outlier, not that the weights were necessarily "chosen wrongly." However, it seems that the problem can be mitigated by an appropriate "de-weighting" of the suspect observation, which provides an opportunity to apply equations (2.45), (2.46), and (2.47).

2.9 Summary Formulas for the Least-Squares Solution Within the Model of Direct Observations

The model of direct observations is given by

$$\underset{n\times 1}{\boldsymbol{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mu + e_1 \\ \vdots \\ \mu + e_n \end{bmatrix} = \boldsymbol{\tau}\mu + \boldsymbol{e},$$
$$\boldsymbol{e} \sim \big(\mathbf{0}, \sigma_0^2 Q\big), \;\; Q \coloneqq P^{-1}.$$

Table 2.3: Summary formulas for the LESS within the model of direct observations

Quantity	Formula	Eq.
Model redundancy	$r = n - 1$	(2.37)
Estimated parameter	$\hat{\mu} = (\boldsymbol{\tau}^T P \boldsymbol{y})/(\boldsymbol{\tau}^T P \boldsymbol{\tau})$	(2.7)
Dispersion of estimated parameter	$D\{\hat{\mu}\} = \sigma_0^2/(\boldsymbol{\tau}^T P \boldsymbol{\tau})$	(2.14)
Vector of predicted residuals	$\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{\tau}\hat{\mu}$	(2.8a)
Dispersion matrix for residuals	$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot \big[P^{-1} - \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P \boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T\big]$	(2.15)
Sum of squared residuals (SSR)	$\Omega = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$	(2.36)
Estimated variance component	$\hat{\sigma}_0^2 = (\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}})/(n-1)$	(2.38)
Vector of adjusted observations	$\widehat{E\{\boldsymbol{y}\}} =: \hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}}$	(2.12)
Dispersion matrix for adjusted observations	$D\{\hat{\boldsymbol{\mu}}_y\} = \sigma_0^2 \cdot \boldsymbol{\tau}\big(\boldsymbol{\tau}^T P \boldsymbol{\tau}\big)^{-1}\boldsymbol{\tau}^T$	(2.16)

Chapter 3

The Gauss-Markov Model

3.1 Model Definition

The Gauss-Markov Model (GMM) is the underlying data model for many of the topics that follow. In presentation of the model, it is assumed that the observation equations (3.1a) have been linearized, if necessary. The model is written as follows:

$$\boldsymbol{y} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \;\; \operatorname{rk} A = m, \tag{3.1a}$$

$$\boldsymbol{e} \sim \left(\boldsymbol{0}, \sigma_0^2 P^{-1}\right). \tag{3.1b}$$

In the case of linearization, $\boldsymbol{y}$ is a vector of observations minus "zeroth-order" terms; A is a known $n \times m$ coefficient matrix (also called *design* or *information* matrix, or *Jacobian* matrix if partial derivatives are involved) relating the observations to the unknown parameters; $\boldsymbol{\xi}$ is a vector of unknown parameters to estimate (corrections to initial values in the case of linearization), and $\boldsymbol{e}$ is a vector of random observation errors, having zero expectation. Equation (3.1a) requires the $n \times m$ *coefficient matrix* A to have full column rank.

The $n \times n$ matrix P is symmetric. It contains weights of the observations, which may be correlated. The inverse of P shown in (3.1) implies that P is a positive-definite matrix; this inverse matrix is called the cofactor matrix and is denoted by Q. The symbol σ_0^2 represents a variance component, which is considered unknown but can be estimated. The dispersion matrix $D\{\boldsymbol{e}\} = \sigma_0^2 P^{-1}$ is called the variance-covariance matrix, or simply the covariance matrix, and is also denoted by Σ. In summary, we have the following relation between the dispersion, weight, and cofactor matrices of the unknown, random error vector $\boldsymbol{e}$:

$$D\{\boldsymbol{e}\} = \Sigma = \sigma_0^2 Q = \sigma_0^2 P^{-1}. \tag{3.2}$$

The redundancy r of the model (3.1a) is defined as

$$\boxed{r := n - \operatorname{rk} A = n - m.} \tag{3.3}$$

Redundancy is also called *degrees of freedom* in the context of statistical testing discussed in Chapter 9.

The GMM shown in (3.1) has two main components. The first component, (3.1a), contains the observation equations $\boldsymbol{y} = A\boldsymbol{\xi}+\boldsymbol{e}$, which show the functional relation between the observations, their random errors, and the unknown parameters that are to be estimated. The second component, (3.1b), shows a stochastic model, $\boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2 P^{-1})$, which expresses the expectation and dispersion of the random errors. These quantities are also called the first and second moments, respectively, of the random error vector $\boldsymbol{e}$.

If the rank of matrix A is less than the number of unknown parameters to estimate, we say that the problem is rank deficient. Such a problem cannot be solved based on the observations alone; additional information about the unknown parameters must be provided. The problem of rank deficiency is covered in Section 3.5 and, much more thoroughly, in the notes for the advanced adjustment computations course.

3.2 The Least-Squares Solution Within the Gauss-Markov Model

We now derive the LEast-Squares Solution (LESS) for the parameter estimate $\hat{\boldsymbol{\xi}}$ and the predicted random error (residual) vector $\tilde{\boldsymbol{e}}$, with their associated dispersion matrices, under the assumption that the coefficient matrix A has full column rank. For convenience, we define the $m \times m$ matrix N and the $m \times 1$ vector $\boldsymbol{c}$ as

$$\big[N,\, \boldsymbol{c}\big] \coloneqq A^T P \big[A,\, \boldsymbol{y}\big]. \tag{3.4}$$

The objective of least-squares minimization is to minimize the P-weighted sum of squared residuals, or, equivalently, to minimize the P-weighted random errors in the model (3.1). Thus the Lagrange target function

$$\Phi(\boldsymbol{\xi}) \coloneqq (\boldsymbol{y} - A\boldsymbol{\xi})^T P(\boldsymbol{y} - A\boldsymbol{\xi}) = \text{stationary} \tag{3.5}$$

should be minimized. Forming the the Euler-Lagrange necessary conditions (or first-order conditions) leads directly to the least-squares *normal equations*

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\xi}} = \big(A^T P A\big)\hat{\boldsymbol{\xi}} - A^T P \boldsymbol{y} = N\hat{\boldsymbol{\xi}} - \boldsymbol{c} \doteq \boldsymbol{0}. \tag{3.6}$$

The sufficient condition is satisfied by $(1/2)\cdot(\partial^2\Phi/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^T) = N$, which is positive-definite since matrix A has full column rank according to (3.1a). Equation (3.6) leads to the least-squares solution (LESS)

$$\boxed{\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c}} \tag{3.7}$$

for the unknown parameter vector $\boldsymbol{\xi}$, with its expectation computed by

$$E\{\hat{\boldsymbol{\xi}}\} = N^{-1}E\{\boldsymbol{c}\} = N^{-1}A^T P E\{\boldsymbol{y}\} = N^{-1}A^T P A\boldsymbol{\xi} = \boldsymbol{\xi}. \tag{3.8}$$

The predicted random error vector (also called *residual* vector) is then given by

$$\boxed{\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} = \big(I_n - AN^{-1}A^TP\big)\boldsymbol{y},} \tag{3.9}$$

with expectation

$$E\{\tilde{\boldsymbol{e}}\} = \big(I_n - AN^{-1}A^TP\big)E\{\boldsymbol{y}\} = \big(I_n - AN^{-1}A^TP\big)A\boldsymbol{\xi} = A\boldsymbol{\xi} - A\boldsymbol{\xi} = \boldsymbol{0}. \tag{3.10}$$

The expectation of the given observation vector is expressed as $E\{\boldsymbol{y}\} = \boldsymbol{\mu}_y$, where $\boldsymbol{\mu}_y$ is the true, but unknown, vector of observables. Thus we write the vector of adjusted observations as

$$\boxed{\widehat{E\{\boldsymbol{y}\}} =: \hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}} = A\hat{\boldsymbol{\xi}},} \tag{3.11}$$

with expectation

$$E\{\hat{\boldsymbol{\mu}}_y\} = AE\{\hat{\boldsymbol{\xi}}\} = A\boldsymbol{\xi}. \tag{3.12}$$

Equations (3.8), (3.10) and (3.12) show that the estimated parameters, the residuals, and the adjusted observations, respectively, are unbiased.

The corresponding dispersion matrices are computed by using the law of covariance propagation. The dispersion of the estimated parameters is computed by

$$\begin{aligned} D\{\hat{\boldsymbol{\xi}}\} = D\{N^{-1}A^TP\boldsymbol{y}\} &= \big(N^{-1}A^TP\big)D\{\boldsymbol{y}\}\big(PAN^{-1}\big) = \\ &= N^{-1}A^TP\big(\sigma_0^2P^{-1}\big)PAN^{-1} \Rightarrow \\ &\boxed{D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2N^{-1}.} \end{aligned} \tag{3.13}$$

And, the dispersion of the residual vector $\tilde{\boldsymbol{e}}$ is

$$\begin{aligned} D\{\tilde{\boldsymbol{e}}\} &= \big(I_n - AN^{-1}A^TP\big)D\{\boldsymbol{y}\}\big(I_n - PAN^{-1}A^T\big) = \\ &= \big(I_n - AN^{-1}A^TP\big)\big(\sigma_0^2P^{-1}\big)\big(I_n - PAN^{-1}A^T\big) = \\ &= \sigma_0^2\big(I_n - AN^{-1}A^TP\big)\big(P^{-1} - AN^{-1}A^T\big) \Rightarrow \\ &\boxed{D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big) =} && \text{(3.14a)} \\ &= D\{\boldsymbol{y}\} - D\{A\hat{\boldsymbol{\xi}}\} =: \sigma_0^2Q_{\tilde{\boldsymbol{e}}}, && \text{(3.14b)} \end{aligned}$$

where the matrix

$$Q_{\tilde{\boldsymbol{e}}} := P^{-1} - AN^{-1}A^T \tag{3.14c}$$

is the cofactor matrix of the residual vector $\tilde{\boldsymbol{e}}$. Equations (3.14a) to (3.14c) reveal that the variances of the residuals are smaller than the corresponding variances of the observations, since the matrix product $AN^{-1}A^T$ is positive-definite. Finally, the dispersion of the vector of adjusted observations is computed by

$$\boxed{D\{\hat{\boldsymbol{\mu}}_y\} = AD\{\hat{\boldsymbol{\xi}}\}A^T = \sigma_0^2AN^{-1}A^T.} \tag{3.15}$$

Summarizing the above equations, the respective distributions for the estimated parameter vector, the residual vector, and the vector of adjusted observations are succinctly expressed by

$$\hat{\boldsymbol{\xi}} \sim \big(\boldsymbol{\xi}, \sigma_0^2 N^{-1}\big), \tag{3.16a}$$

$$\tilde{\boldsymbol{e}} \sim \big(\mathbf{0}, \sigma_0^2 \big[P^{-1} - AN^{-1}A^T\big] =: \sigma_0^2 Q_{\tilde{\boldsymbol{e}}},\big), \tag{3.16b}$$

$$\hat{\boldsymbol{\mu}}_y \sim \big(A\boldsymbol{\xi}, \sigma_0^2 AN^{-1}A^T\big). \tag{3.16c}$$

Since the variance component σ_0^2 is an unknown quantity, the dispersions shown in (3.16) cannot be computed unless either σ_0^2 is estimated or a value is specified for it. In the case where the estimated variance component is used in lieu of the true, but unknown, variance component, we speak of an *estimated dispersion matrix* for the estimated parameter vector, which is provided by

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} = \hat{\sigma}_0^2 N^{-1}, \tag{3.17}$$

with obvious extension to other quantities, such as $\hat{D}\{\tilde{\boldsymbol{e}}\}$ and $\hat{D}\{\hat{\boldsymbol{\mu}}_y\}$. See Section 3.3 for the derivation of the variance component estimate $\hat{\sigma}_0^2$, the formula for which is given in (3.28).

3.2.1 Example — Fitting a Parabola

Suppose n observations were taken of data that, when plotted in 2D, appear to approximate a parabola (Figure 3.1). The y-coordinates represent measured data with random errors having zero mean and iid dispersion. The x-coordinates are assumed to be known without error. This is a classical *regression problem*.

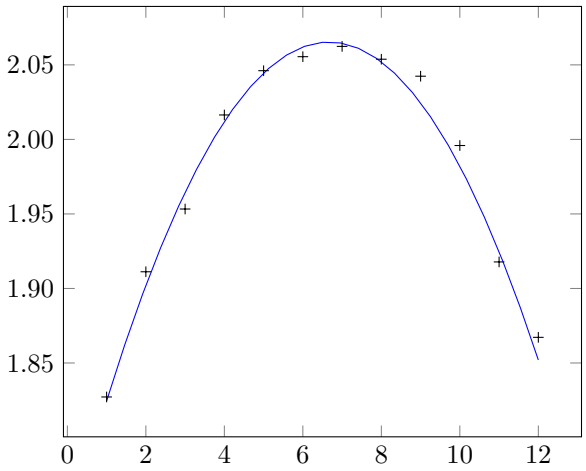

Figure 3.1: A fitted parabolic curve based on measured y-coordinates and given x-coordinates

The observations equations of the Gauss-Markov Model (GMM) are set up as follows: The ith observation equation, $i = 1, \ldots, n$,

$$y_i = ax_i^2 + bx_i + c + e_i, \tag{3.18}$$

can be extended to a system of equations in matrix from as

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}}_{\boldsymbol{\xi}} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}}_{\boldsymbol{e}}, \tag{3.19}$$

where $\boldsymbol{\xi} =: [a, b, c]^T$ is the vector of unknown parameters, which, together with the stochastic model $\boldsymbol{e} \sim (\mathbf{0}, \text{iid})$, constitutes a Gauss-Markov Model. Note that in other examples within the GMM, the the random observation errors could have a heteroscedastic distribution, or their dispersion could be represented by a full cofactor matrix Q.

3.2.2 Correlation of Adjusted Observations and Predicted Residuals

Equation (3.14b) implies that the covariance between the vector of *adjusted observations* $\hat{\boldsymbol{\mu}}_y = A\hat{\boldsymbol{\xi}}$ and the vector of residuals $\tilde{\boldsymbol{e}}$ is zero. Since, according to (3.7) and (3.9), both vectors are a function of the random vector $\boldsymbol{y}$, this can also be shown by applying the law of covariance propagation as follows:

$$\begin{aligned} C\{A\hat{\boldsymbol{\xi}}, \tilde{\boldsymbol{e}}\} &= AN^{-1}A^TP \cdot D\{\boldsymbol{y}\} \cdot \big(I_n - AN^{-1}A^TP\big)^T = \\ &= \sigma_0^2\big[AN^{-1}A^T - AN^{-1}\big(A^TPA\big)N^{-1}A^T\big] = \\ &= \sigma_0^2\big[AN^{-1}A^T - AN^{-1}A^T\big] = 0. \end{aligned} \tag{3.20}$$

Also, we have the following covariance between the adjusted and original observations:

$$\begin{aligned} C\{A\hat{\boldsymbol{\xi}}, \boldsymbol{y}\} &= AN^{-1}A^TPD\{\boldsymbol{y}\} = \sigma_0^2AN^{-1}A^TPP^{-1} = \\ &= \sigma_0^2AN^{-1}A^T = D\{A\hat{\boldsymbol{\xi}}\}. \end{aligned} \tag{3.21}$$

Zero correlation does not necessarily imply statistical independence, though the converse does hold. Analogous to (9.9a), the adjusted observations and predicted residuals are not statistically independent unless the expectation of their product is equal to the product of their expectations. The following shows that this property

is not satisfied: Since the trace of a scalar product is the scalar product itself, we start with

$$E\{(A\hat{\boldsymbol{\xi}})^T\tilde{\boldsymbol{e}}\} = E\{\mathrm{tr}\,\hat{\boldsymbol{\xi}}^T A^T\big(I_n - AN^{-1}A^TP\big)\boldsymbol{y}\}.$$

But the trace is invariant with respect to a cyclic transformation (see (A.5)). Thus,

$$E\{(A\hat{\boldsymbol{\xi}})^T\tilde{\boldsymbol{e}}\} = E\{\mathrm{tr}\big(A^T - A^TAN^{-1}A^TP\big)\boldsymbol{y}\hat{\boldsymbol{\xi}}^T\} =$$
$$= \mathrm{tr}\big(A^T - A^TAN^{-1}A^TP\big)E\{\boldsymbol{y}\hat{\boldsymbol{\xi}}^T\} \neq 0 = E\{(A\hat{\boldsymbol{\xi}})^T\}E\{\tilde{\boldsymbol{e}}\},\ \text{ since } E\{\tilde{\boldsymbol{e}}\} = \mathbf{0}.$$

3.2.3 P-Weighted Norm of the Residual Vector

The P-weighted norm of the residual vector $\tilde{\boldsymbol{e}}$ is an important quantity that can be used to check the overall ("global") fit of the adjustment. The norm is defined as

$$\boxed{\Omega \coloneqq \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}},} \tag{3.22}$$

and it is guaranteed to be a minimum, since $\tilde{\boldsymbol{e}}$ was obtained by minimizing $\boldsymbol{e}^TP\boldsymbol{e}$ (cf. (3.5)). In the special case where $P = I_n$, the quadratic form Ω is often called the *sum of squared residuals*, or SSR, in the statistical literature. We use the term SSR in the following chapters even when P is not the identity matrix. Substituting (3.9) into (3.22) leads to some commonly used alternative forms for Ω.

$$\begin{aligned}
\tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}} &= (\boldsymbol{y} - A\hat{\boldsymbol{\xi}})^TP(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) = && \text{(3.23a)}\\
&= \boldsymbol{y}^TP\boldsymbol{y} - \boldsymbol{y}^TPA\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}}^TA^TP\boldsymbol{y} + \hat{\boldsymbol{\xi}}^TA^TPA\hat{\boldsymbol{\xi}} = \\
&= \boldsymbol{y}^TP\boldsymbol{y} - 2\boldsymbol{c}^T\hat{\boldsymbol{\xi}} + \boldsymbol{c}^T\hat{\boldsymbol{\xi}} = \\
&= \boldsymbol{y}^TP\boldsymbol{y} - \boldsymbol{c}^T\hat{\boldsymbol{\xi}} = && \text{(3.23b)}\\
&= \boldsymbol{y}^TP\boldsymbol{y} - \boldsymbol{c}^TN^{-1}\boldsymbol{c} = && \text{(3.23c)}\\
&= \boldsymbol{y}^TP\boldsymbol{y} - (N\hat{\boldsymbol{\xi}})^TN^{-1}N\hat{\boldsymbol{\xi}} = \\
&= \boldsymbol{y}^TP\boldsymbol{y} - \hat{\boldsymbol{\xi}}^TN\hat{\boldsymbol{\xi}} = && \text{(3.23d)}\\
&= \boldsymbol{y}^T(P - PAN^{-1}A^TP)\boldsymbol{y} && \text{(3.23e)}
\end{aligned}$$

Note that the target function (3.5) could have been written explicitly as a function of the random error vector $\boldsymbol{e}$ with the introduction of a vector of Lagrange multipliers $\boldsymbol{\lambda}$ as follows:

$$\Phi(\boldsymbol{e}, \boldsymbol{\xi}, \boldsymbol{\lambda}) = \boldsymbol{e}^TP\boldsymbol{e} - 2\boldsymbol{\lambda}^T(\boldsymbol{y} - A\boldsymbol{\xi} - \boldsymbol{e}) = \text{stationary}. \tag{3.24}$$

This approach leads to the estimate of Lagrange multipliers as $-\hat{\boldsymbol{\lambda}} = P\tilde{\boldsymbol{e}}$ and thus leads to yet another expression for the P-weighted norm

$$\Omega = \tilde{\boldsymbol{e}}^TP\tilde{\boldsymbol{e}} = -\tilde{\boldsymbol{e}}^T\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}^TP^{-1}\hat{\boldsymbol{\lambda}}. \tag{3.25}$$

3.3 Estimated Variance Component Within the Gauss-Markov Model

As stated in Section 2.4, the variance component σ_0^2 is an unknown quantity in the Gauss-Markov Model (GMM). We now present the derivation of the estimated variance component $\hat{\sigma}_0^2$. As defined in (3.1), the dispersion matrix for the random error vector $\boldsymbol{e}$ is $D\{\boldsymbol{e}\} = \sigma_0^2 Q$. Also, by definition of dispersion we have $D\{\boldsymbol{e}\} = E\{(\boldsymbol{e} - E\{\boldsymbol{e}\})(\boldsymbol{e} - E\{\boldsymbol{e}\})^T\}$. But, for the error vector $E\{\boldsymbol{e}\} = \mathbf{0}$; therefore $D\{\boldsymbol{e}\} = E\{\boldsymbol{e}\boldsymbol{e}^T\} = \sigma_0^2 Q = \sigma_0^2 P^{-1}$.

The following steps lead to an expression for the variance component σ_0^2 in terms of the quadratic product $\boldsymbol{e}^T P \boldsymbol{e}$.

$$E\{\boldsymbol{e}\boldsymbol{e}^T\} = \sigma_0^2 Q \qquad \text{(by definition)}$$

$$PE\{\boldsymbol{e}\boldsymbol{e}^T\} = \sigma_0^2 I_n \qquad \text{(multiply both sides by } P\text{)}$$

$$\operatorname{tr} PE\{\boldsymbol{e}\boldsymbol{e}^T\} = \sigma_0^2 \operatorname{tr} I_n = n\sigma_0^2 \qquad \text{(apply the trace operator)}$$

$$\operatorname{tr} E\{P\boldsymbol{e}\boldsymbol{e}^T\} = n\sigma_0^2 \qquad \text{(move the constant matrix } P \text{ into the expectation)}$$

$$E\{\operatorname{tr} P\boldsymbol{e}\boldsymbol{e}^T\} = n\sigma_0^2 \qquad \text{(interchange the trace and expectation operators—both linear)}$$

$$E\{\operatorname{tr} \boldsymbol{e}^T P\boldsymbol{e}\} = n\sigma_0^2 \qquad \text{(the trace is invariant with respect to a cyclic transformation)}$$

$$E\{\boldsymbol{e}^T P\boldsymbol{e}\} = n\sigma_0^2 \qquad \text{(a quadratic product is a scalar; trace of scalar is scalar itself)}$$

$$\sigma_0^2 = E\{\frac{\boldsymbol{e}^T P\boldsymbol{e}}{n}\} \qquad \text{(dividing through by } n \text{ and placing } n \text{ inside } E\{\cdot\}\text{)}$$

$$\bar{\sigma}_0^2 \coloneqq \frac{\boldsymbol{e}^T P\boldsymbol{e}}{n} \qquad \text{(define a symbol for the term inside } E\{\cdot\}\text{)}$$

$$E\{\bar{\sigma}_0^2\} = \sigma_0^2 \qquad \text{(by substitution)}$$

Thus we can say that $\big(\boldsymbol{e}^T P\boldsymbol{e}\big)/n$ provides an unbiased representation of σ_0^2. However, we do not actually know the true random error vector $\boldsymbol{e}$, but we do know its predicted value $\tilde{\boldsymbol{e}}$.

We now work with the residual vector $\tilde{\boldsymbol{e}}$ to find an unbiased estimate of σ_0^2. Combining steps similar to those explained above, we can write

$$E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} = \operatorname{tr} E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} = \operatorname{tr} E\{\tilde{\boldsymbol{e}}\tilde{\boldsymbol{e}}^T\}P = \operatorname{tr} D\{\tilde{\boldsymbol{e}}\}P. \tag{3.26}$$

According to (3.14a), the dispersion of the residual vector is $D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big)$. Substituting this result into (3.26) gives

$$\begin{aligned} E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} &= \operatorname{tr} \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big)P = \\ &= \sigma_0^2\big(\operatorname{tr} I_n - \operatorname{tr} AN^{-1}A^T P\big) = \\ &= \sigma_0^2\big(\operatorname{tr} I_n - \operatorname{tr} N^{-1}A^T PA\big) = \qquad \text{(using (A.5))} \\ &= \sigma_0^2(n - \operatorname{rk} N) = \sigma_0^2(n - \operatorname{rk} A). \qquad \text{(using (1.7c))} \end{aligned}$$

Finally, we arrive at

$$E\{\frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n - \mathrm{rk}\, A}\} = \sigma_0^2. \tag{3.27}$$

Now, we simply label the argument of the expectation operator on the left side of (3.27) as $\hat{\sigma}_0^2$, which allows us to write the expression for the estimated variance component as

$$\boxed{\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n - \mathrm{rk}\, A}.} \tag{3.28}$$

Obviously, $\hat{\sigma}_0^2$ is a uniformly unbiased estimate of σ_0^2, since $E\{\hat{\sigma}_0^2\} = \sigma_0^2$. In the case of the Model of Direct Observations, we replace A with τ, which has rank of 1, and thus we have $\hat{\sigma}_0^2 \coloneqq \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}/(n-1)$, which verifies (2.38). Alternative expressions for $\hat{\sigma}_0^2$ can be reached by use of (3.23) and (3.25).

The above derivations imply the following relationship between $E\{\boldsymbol{e}^T P \boldsymbol{e}\}$ and $E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}\}$:

$$\frac{E\{\boldsymbol{e}^T P \boldsymbol{e}\}}{n} = \frac{E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}\}}{n - \mathrm{rk}\, A} = \sigma_0^2 \Rightarrow \tag{3.29a}$$

$$E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}\} < E\{\boldsymbol{e}^T P \boldsymbol{e}\} \tag{3.29b}$$

According to Grafarend and Schaffrin (1993, pg. 103), and Schaffrin (1997b), the dispersion, and estimated dispersion, respectively, of $\hat{\sigma}_0^2$ are given by

$$D\{\hat{\sigma}_0^2\} = (n-m)^{-1} \cdot 2\big(\sigma_0^2\big)^2 \tag{3.30}$$

and

$$\hat{D}\{\hat{\sigma}_0^2\} = (n-m)^{-1} \cdot 2\big(\hat{\sigma}_0^2\big)^2, \tag{3.31}$$

where it is assumed that $m = \mathrm{rk}\, A$.

3.4 Linearized Observation Equations and Algorithm

When the unknown parameters $\boldsymbol{\xi}$ are a nonlinear function of the observables, we can represent the observation equations by

$$E\{\boldsymbol{y}\} = \boldsymbol{a}(\boldsymbol{\xi}), \;\; D\{\boldsymbol{y}\} = \sigma_0^2 P^{-1} = D\{\boldsymbol{e}\}, \;\; \boldsymbol{e} \coloneqq \boldsymbol{y} - E\{\boldsymbol{y}\}, \tag{3.32a}$$

or

$$\boldsymbol{y} = \boldsymbol{a}(\boldsymbol{\xi}) + \boldsymbol{e}, \tag{3.32b}$$

where $\boldsymbol{a}(\boldsymbol{\xi})$ is a vector of functions that maps $\mathbb{R}^m$ to $\mathbb{R}^n$. Using an *approximate* vector of parameters $\boldsymbol{\xi}_{(0)}$ and a Taylor series expansion permits us to rewrite (3.32a) as

$$E\{\boldsymbol{y}\} = \boldsymbol{a}(\boldsymbol{\xi}_{(0)}) + \left.\frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\xi}^T}\right|_{\boldsymbol{\xi}=\boldsymbol{\xi}_{(0)}} \cdot (\boldsymbol{\xi} - \boldsymbol{\xi}_{(0)}) + \cdots \tag{3.33a}$$

$$\Rightarrow \boxed{E\{\boldsymbol{y} - \boldsymbol{a}(\boldsymbol{\xi}_{(0)})\} = A{\cdot}(\boldsymbol{\xi} - \boldsymbol{\xi}_{(0)})} + \text{higher order terms.} \tag{3.33b}$$

By truncating the Taylor series expansion (i.e. dropping the higher-order terms) and working with observation *increments* $\boldsymbol{y} - \boldsymbol{a}(\boldsymbol{\xi}_{(0)})$ and parameter increments $\boldsymbol{\xi} - \boldsymbol{\xi}_{(0)}$, we may form the system of least-squares *normal equations*

$$\boxed{(A^T P A)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_{(0)}) = A^T P(\boldsymbol{y} - \boldsymbol{a}(\boldsymbol{\xi}_{(0)})),} \tag{3.34}$$

leading to

$$\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}_{(0)} + N^{-1} A^T P(\boldsymbol{y} - \boldsymbol{a}(\boldsymbol{\xi}_{(0)})) \tag{3.35a}$$

and

$$D\{\hat{\boldsymbol{\xi}}\} = D\{\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_{(0)}\} = \sigma_0^2 (A^T P A)^{-1} \tag{3.35b}$$

for the estimate of $\boldsymbol{\xi}$ and its dispersion matrix, respectively.

The chosen approximate values for $\boldsymbol{\xi}_{(0)}$ may be less precise than we prefer, which, in turn, might affect the accuracy and precision of the computed values of $\hat{\boldsymbol{\xi}}$. In practice, $\boldsymbol{\xi}_{(0)}$ may be taken from a solution based on only a minimum subset of the observation equations (i.e., only m of them). Such approximate values could be improved upon by replacing them with the values obtained from a first computation of $\hat{\boldsymbol{\xi}}$. Then, the system of equations could be solved again, leading to a more precise values for $\hat{\boldsymbol{\xi}}$. This process could be repeated until the difference between $\hat{\boldsymbol{\xi}}$ and $\boldsymbol{\xi}_{(0)}$ becomes arbitrarily small. This approach is called an *iterative least-squares solution.*

For the jth iteration step of an iterative algorithm the approximate parameter vector $\boldsymbol{\xi}_{(j)}$ is specified by

$$\boldsymbol{\xi}_{(j)} \coloneqq \boldsymbol{\xi}_{(j-1)} + (\hat{\boldsymbol{\xi}}_{j-1} - \underset{\sim}{\mathbf{0}}), \tag{3.36a}$$

where the subtraction of a *random zero vector* $\underset{\sim}{\mathbf{0}}$ is a formality that ensures the approximate vector $\boldsymbol{\xi}_{(j)}$ is non-random, as it must be. Thus, we say that the subtraction of $\underset{\sim}{\mathbf{0}}$ strips $\hat{\boldsymbol{\xi}}_{j-1}$ of its randomness (note that subtracting a random zero does not change the numerical values of $\hat{\boldsymbol{\xi}}_{j-1}$). The iterations are repeated until

$$\left\| \hat{\boldsymbol{\xi}}_{j+1} - \hat{\boldsymbol{\xi}}_j \right\| < \epsilon \tag{3.36b}$$

for some chosen, small ϵ. Such an iterative algorithm is called a Gauss-Newton algorithm. It is summarized below.

Note that the symbol $\boldsymbol{\xi}_{(j)}$ (with parenthesis around the subscript) has been used to indicate the approximate Taylor-series expansion point, while $\boldsymbol{\xi}$ is used for the true (unknown) parameter vector as usual, and $\hat{\boldsymbol{\xi}}_j$ (without parenthesis on the subscript) denotes the vector of estimated parameters.

Iterative algorithm to solve a linearized system of normal equations:

1. Initialization: Specify initial values for $\boldsymbol{\xi}_{(0)}$, e.g., based on a minimum number of observation equations. Compute $\boldsymbol{a}(\boldsymbol{\xi}_{(0)})$ and form matrix $(A)_0$ based on $\boldsymbol{\xi}_{(0)}$. Then set iteration index to $j = 1$.

2. Compute the jth solution

$$\hat{\boldsymbol{\xi}}_j = \boldsymbol{\xi}_{(j-1)} + \left[(A)^T_{(j-1)} P (A)_{(j-1)}\right]^{-1} (A)^T_{(j-1)} P\left[\boldsymbol{y} - \boldsymbol{a}(\boldsymbol{\xi}_{(j-1)})\right]. \tag{3.37a}$$

Check for *convergence* using

$$\left\| \hat{\boldsymbol{\xi}}_{j+1} - \hat{\boldsymbol{\xi}}_j \right\| < \epsilon \tag{3.37b}$$

for some chosen ϵ. If the solution has converged go to step 4.

3. Update the expansion point $\boldsymbol{\xi}_{(j)}$ according to (3.36a). Update the partial derivatives in the Jacobian matrix $(A)_{(j)}$. Increment the iteration counter j by $j \to j+1$. Repeat step 2.

4. After convergence, compute the dispersion matrix

$$D\{\hat{\boldsymbol{\xi}}\} = D\{\hat{\boldsymbol{\xi}}_j\}, \tag{3.37c}$$

the residual vector

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{a}(\boldsymbol{\xi}_{(j)}), \tag{3.37d}$$

the vector of adjusted observations

$$\boldsymbol{y} - \tilde{\boldsymbol{e}} = \boldsymbol{a}(\boldsymbol{\xi}_{(j)}), \tag{3.37e}$$

and the estimated variance component

$$\hat{\sigma}_0^2 = (\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}})/r. \tag{3.37f}$$

Note that in this last step, the vector $\boldsymbol{\xi}_{(j)}$ is the "approximate" vector (expansion point) computed in step 3. Checks similar to those discussed in Section 2.5.1 should also be made.

3.5 Introduction of Datum Information to Treat the Rank-Deficient Gauss-Markov Model

A rank-deficient Gauss-Markov Model (GMM) is one in which the rank of the coefficient matrix A is less than its number of columns. This means that at least one column of A is either a scalar multiple of a different column or that it can be expressed as a linear combination of other columns. Such rank deficiency is expressed mathematically as $\operatorname{rk} A < m$, where m is the number of columns of A. It implies also that $\operatorname{rk} N = \operatorname{rk} A^T P A < m$, which means that the unknown parameters cannot be estimated by (3.7). Put another way, a rank-deficient model is one in which there are more parameters than can be estimated from the data. In fact, the rank of the coefficient matrix reveals the number of *estimable parameters* of the model.

Rank deficiency often arises in the context of network adjustments where station coordinates must be estimated but the observations do not contain sufficient information to define the underlying coordinate system, also called *datum* in this context. Thus we speak of a network *datum deficiency*. As noted in Chapter 5, a

2-D network where only angles and distance have been measured provides an example of a datum deficiency of three, owing to the unknown origin and orientation parameters of the network. However, if *certain values* (or known values) are provided for enough parameters, this "datum information" can be used to eliminate the rank deficiency of the model. The method is described in the following.

Consider the following (linearized) GMM with rank-deficient matrix A:

$$\boldsymbol{y} = A\boldsymbol{\xi} + \boldsymbol{e}, \;\; \boldsymbol{e} \sim \big(\boldsymbol{0}, \sigma_0^2 P^{-1}\big), \;\; \operatorname{rk} A =: q < m. \tag{3.38a}$$

We can partition the matrix A as

$$\underset{n\times m}{A} = \left[\begin{array}{c|c} \underset{n\times q}{A_1} & \underset{n\times(m-q)}{A_2} \end{array}\right], \;\; \text{with} \;\; \operatorname{rk} A_1 = q \coloneqq \operatorname{rk} A, \tag{3.38b}$$

so that matrix A_1 has full column rank. A compatible partitioning of the parameter vector $\boldsymbol{\xi}$, i.e.,

$$\boldsymbol{\xi} = \begin{bmatrix} \underset{q\times 1}{\boldsymbol{\xi}_1} \\ \\ \underset{(m-q)\times 1}{\boldsymbol{\xi}_2} \end{bmatrix}, \tag{3.38c}$$

leads to the following system of partitioned normal equations:

$$\begin{bmatrix} A_1^T \\ A_2^T \end{bmatrix} P \begin{bmatrix} A_1, & A_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} = \begin{bmatrix} A_1^T \\ A_2^T \end{bmatrix} P\boldsymbol{y} = \left[\begin{array}{c|c} A_1^T P A_1 & A_1^T P A_2 \\ \hline A_2^T P A_1 & A_2^T P A_2 \end{array}\right] \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} = \begin{bmatrix} A_1^T P\boldsymbol{y} \\ A_2^T P\boldsymbol{y} \end{bmatrix} =$$

$$= \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \hat{\boldsymbol{\xi}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix}. \tag{3.39}$$

The sub-scripted terms in (3.39) may be defined more succinctly as

$$\big[N_{ij}, \boldsymbol{c}_i\big] \coloneqq A_i^T P\big[A_j, \boldsymbol{y}\big], \;\; \text{for} \;\; i,j \in \{1,2\}. \tag{3.40}$$

Defining a datum for $m-q$ parameters means that values for them must be specified. Mathematically, a datum is defined by $\hat{\boldsymbol{\xi}}_2 \to \boldsymbol{\xi}_2^0$, where $\boldsymbol{\xi}_2^0$ is known. The rank of A_1 given in (3.38b) implies that the inverse of the $q \times q$ matrix N_{11} exists. Therefore, from the top row of (3.39), and with a *given datum* $\boldsymbol{\xi}_2^0$ substituted for $\hat{\boldsymbol{\xi}}_2$, we can write

$$N_{11}\hat{\boldsymbol{\xi}}_1 = \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \Rightarrow \tag{3.41a}$$

$$\boxed{\hat{\boldsymbol{\xi}}_1 = N_{11}^{-1}\big(\boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0\big).} \tag{3.41b}$$

Equation (3.41b) shows that datum values can be specified or modified after observations have been made and matrix N_{11} has been inverted. Moreover, since the only random component in (3.41b) is $\boldsymbol{c}_1$, we have

$$\boxed{D\{\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2 N_{11}^{-1}} \tag{3.42}$$

for the dispersion of the vector of estimated parameters $\hat{\boldsymbol{\xi}}_1$.

The predicted random error (residual) vector and its dispersion are then defined as follows:

$$\boxed{\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} = \boldsymbol{y} - \left[\begin{array}{c|c} A_1 & A_2 \end{array}\right] \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \boldsymbol{\xi}_2^0 \end{bmatrix} = \boldsymbol{y} - A_1\hat{\boldsymbol{\xi}}_1 - A_2\boldsymbol{\xi}_2^0,} \tag{3.43a}$$

$$\boxed{D\{\tilde{\boldsymbol{e}}\} = D\{\boldsymbol{y}\} - D\{A_1\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2\big(P^{-1} - A_1N_{11}^{-1}A_1^T\big).} \tag{3.43b}$$

Note that $C\{\boldsymbol{y}, \hat{\boldsymbol{\xi}}_1\} = 0$, which is implied by (3.43b). After computing the residuals, it is straightforward to compute the vector of adjusted observations and it dispersion matrix, respectively, by

$$\boxed{\widehat{E\{\boldsymbol{y}\}} =: \hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}} = A_1\hat{\boldsymbol{\xi}}_1 + A_2\boldsymbol{\xi}_2^0,} \tag{3.44a}$$

$$\boxed{D\{\hat{\boldsymbol{\mu}}_y\} = D\{A_1\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2 \cdot A_1N_{11}^{-1}A_1^T.} \tag{3.44b}$$

Here, $\hat{\boldsymbol{\mu}}_y$ is also interpreted as an estimate of the true, and thus unknown, vector of observables $\boldsymbol{\mu}_y$, where $E\{\boldsymbol{y}\} = \boldsymbol{\mu}_y$.

The sum of squared residuals (SSR) is given by

$$\boxed{\Omega = \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}},} \tag{3.45}$$

while the redundancy of the model is provided by

$$\boxed{r = n - \operatorname{rk} A = n - q.} \tag{3.46}$$

Substituting (3.43a) into (3.45), and considering (3.41a), leads to

$$\boxed{\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}}{r} = \frac{\boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}_1^T\hat{\boldsymbol{\xi}}_1 - \boldsymbol{c}_2^T\boldsymbol{\xi}_2^0}{n-q}} \tag{3.47}$$

as an estimate for the unknown variance component σ_0^2. Here, the relation $\hat{\boldsymbol{\xi}}_1^T N_{11}\hat{\boldsymbol{\xi}}_1$ $+\hat{\boldsymbol{\xi}}_1^T N_{12}\hat{\boldsymbol{\xi}}_2 = \hat{\boldsymbol{\xi}}_1^T\boldsymbol{c}_1$ has been used. However, since $\operatorname{rk} A_1 = \operatorname{rk} A = q$, the $n \times (m-q)$ submatrix A_2 must be in the column space of the $n \times q$ matrix A_1 so that

$$A_2 = A_1L \tag{3.48a}$$

for some $q \times (m-q)$ matrix L. Therefore,

$$N_{12} = A_1^T PA_2 = A_1^T PA_1L = N_{11}L \Rightarrow \tag{3.48b}$$

$$N_{11}^{-1}N_{12} = L. \tag{3.48c}$$

With this result, and using (3.41b), we have

$$\boldsymbol{c}_1^T\hat{\boldsymbol{\xi}}_1 + \boldsymbol{c}_2^T\boldsymbol{\xi}_2^0 = \boldsymbol{y}^T PA_1\big(N_{11}^{-1}\boldsymbol{c}_1 - N_{11}^{-1}N_{12}\boldsymbol{\xi}_2^0\big) + \boldsymbol{y}^T PA_2\boldsymbol{\xi}_2^0 =$$

$$
\begin{aligned}
&= \boldsymbol{y}^T P A_1 \big(N_{11}^{-1}\boldsymbol{c}_1 - L\boldsymbol{\xi}_2^0\big) + \boldsymbol{y}^T P A_2 \boldsymbol{\xi}_2^0 = \\
&= \boldsymbol{y}^T P A_1 N_{11}^{-1}\boldsymbol{c}_1 - \boldsymbol{y}^T P \big(A_1 L\big)\boldsymbol{\xi}_2^0 + \boldsymbol{y}^T P A_2 \boldsymbol{\xi}_2^0 = \\
&= \boldsymbol{y}^T P A_1 N_{11}^{-1}\boldsymbol{c}_1 = \boldsymbol{c}_1^T N_{11}^{-1}\boldsymbol{c}_1, \qquad (3.49)
\end{aligned}
$$

which, upon substitution into (3.47), leads to

$$
\boxed{\hat{\sigma}_0^2 = \frac{\boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}_1^T N_{11}^{-1}\boldsymbol{c}_1}{n-q}} \qquad (3.50)
$$

as an alternative form for the estimated variance component.

It is instructive to compare the dispersion of $\hat{\boldsymbol{\xi}}_1$ shown in (3.42) with the corresponding dispersion in the case that matrix A has full row rank, i.e., $\operatorname{rk} A = m$. In the full-rank case, we could invert the coefficient matrix of (3.39) and find the upper $q \times q$ block of the inverse, scaled by σ_0^2, to be the dispersion of $\hat{\boldsymbol{\xi}}_1$. Referring to (A.11) for the inverse of the partitioned matrix N, we find

$$
\begin{aligned}
\underbrace{D\{\hat{\boldsymbol{\xi}}_1\}}_{\text{no datum}} &= \sigma_0^2\big[N_{11}^{-1} + N_{11}^{-1}N_{12}\big(N_{22} - N_{21}N_{11}^{-1}N_{12}\big)^{-1}N_{21}N_{11}^{-1}\big] = \\
&= \sigma_0^2\big(N_{11} - N_{12}N_{22}^{-1}N_{21}\big)^{-1} > \sigma_0^2 N_{11}^{-1} = \underbrace{D\{\hat{\boldsymbol{\xi}}_1\}}_{\text{datum supplied}} . \qquad (3.51)
\end{aligned}
$$

The smaller dispersion in the last line of (3.51) shows that if a datum is introduced (increase in information), the unknown parameters $\boldsymbol{\xi}$ are estimated with smaller variance.

Minimally constrained adjustment The type of least-squares adjustment described in this section belongs to a class of *minimally constrained adjustment*, a subject treated in much greater detail in the notes for Advanced Adjustment Computations. The reason that the adjustment is of type minimally constrained is because the datum information only provides information on $m - q$ of the parameters, which is just enough to overcome the rank deficiency of the model. The result is a *unique solution* for the residual vector $\tilde{\boldsymbol{e}}$, the adjusted observations $A\hat{\boldsymbol{\xi}}$, and the estimated variance component $\hat{\sigma}_0^2$. This means that the specification for $\boldsymbol{\xi}_2^0$ will not affect the computed values of these quantities. Put another way, we can say that they are invariant to the choice of the datum. On the other hand, the vector of estimated parameters $\hat{\boldsymbol{\xi}}$ will not be unique; it does depend on the specification of $\boldsymbol{\xi}_2^0$.

3.6 Practice Problems

1. Starting with the Lagrange target function (3.24), derive the least-squares solution (LESS) within the Gauss-Markov Model for the unknown parameter vector $\boldsymbol{\xi}$ and the unknown vector of Lagrange multipliers $\boldsymbol{\lambda}$.

2. Based on your answer in the preceding problem, show that the identity $\Omega = \hat{\boldsymbol{\lambda}}^T P^{-1} \hat{\boldsymbol{\lambda}}$ in (3.25) holds.

3. In order to determine the height of point F, leveling measurements have been taken in forward and reverse directions from three different points A, B, and C, each with known height. The relevant data are given in Table 3.1.

Table 3.1: Leveling data for Problem 3.

Point	Height [m]	Forward obs. to F [m]	Length of path [km]	Reverse obs. from F [m]
A	100.055	10.064	2.5	−10.074
B	102.663	7.425	4	−7.462
C	95.310	14.811	6	−14.781

Assume that the standard deviations of the observations are $\sigma = 3\,\text{mm}$ per every one km of leveling and that all measurements are uncorrelated. Setup the Gauss-Markov Model and compute the LESS of:

(a) The height at point F and its estimated dispersion.

(b) The vector of residuals and its estimated dispersion matrix.

(c) The estimated variance component.

(d) Compute the trace of the product $\sigma_0^{-2} \cdot D\{\tilde{\boldsymbol{e}}\} \cdot P$ and confirm that it equals the redundancy of the model.

4. Elevations were observed with a digital level at nodes of a 2D-grid. The horizontal coordinates of the nodes $(\boldsymbol{X}, \boldsymbol{Y})$ are assumed be be known with certainty, while the random errors of the observed elevations have a homogeneous distribution with zero mean and $\sigma^2 = (10\,\text{mm})^2$ variance. The data are listed in Table 3.2.

(a) Use the LESS within the GMM to estimate the parameters of a fitted plane assuming the observation equations can be modeled by

$$E\{y_i\} = aX_i + bY_i + c, \quad i = 1, \ldots, n,$$

with unknown parameters $\boldsymbol{\xi} = [a,\, b,\, c]^T$.

(b) Use the LESS within the GMM to estimate parameters for a quadratic surface assuming the observation equations can be modeled by

$$E\{y_i\} = aX_i^2 + bY_i^2 + cX_iY_i + dX_i + eY_i + f, \quad i = 1, \ldots, n,$$

with unknown parameters $\boldsymbol{\xi} = [a,\, b,\, c,\, d,\, e,\, f]^T$.

Table 3.2: Elevation data y_i observed at known grid locations

i	X_i	Y_i	y_i	i	X_i	Y_i	y_i
1	−20	−20	9.869	14	0	10	10.019
2	−20	−10	9.920	15	0	20	10.037
3	−20	0	9.907	16	10	−20	9.946
4	−20	10	9.957	17	10	−10	9.988
5	−20	20	9.959	18	10	0	10.035
6	−10	−20	9.889	19	10	10	10.055
7	−10	−10	9.937	20	10	20	10.066
8	−10	0	9.973	21	20	−20	9.963
9	−10	10	10.025	22	20	−10	9.986
10	−10	20	10.026	23	20	0	10.037
11	0	−20	9.917	24	20	10	10.068
12	0	−10	10.000	25	20	20	10.069
13	0	0	10.007				

(c) Which of the two above observational models, the planar one or the quadratic one, fit the data best? Give the reason for your answer.

5. To determine the coordinates of an unknown point $P(x, y)$, some measurements were carried out from two given points $A(50, 30)$ and $B(100, 40)$ in meters.

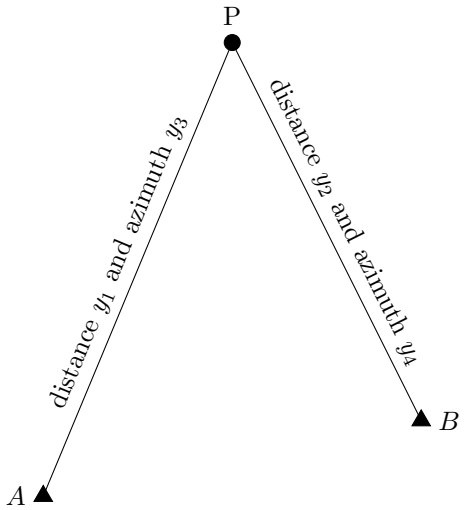

Figure 3.2: Two distances and two azimuths measured from known points A and B to determine coordinates of point P

Two distances were observed, the first from A to P and the second from B to P. The observed distances are $y_1 = 66.137\,\text{m}$ and $y_2 = 58.610\,\text{m}$, and they are considered to be uncorrelated with variance $\sigma^2 = (1\,\text{cm})^2$.

In addition, two azimuths were observed independently. The observed azimuth from A to P is $y_3 = 20°20'55''$ and the observed azimuth from B to P is $y_4 = 332°33'41''$. The standard deviation of both azimuths is $\pm\sigma_\alpha = 05''$.

Compute the following:

(a) The estimated coordinates of point P.

(b) The estimated variances of the coordinates and their correlation coefficient.

(c) The residual vector $\tilde{\boldsymbol{e}}$.

(d) The estimated variance component $\hat{\sigma}_0^2$.

6. To determine the coordinates of a new point P, distances were measured to four given points having known coordinates. One angle was also measured. The coordinates of the given points are listed in Table 3.3, and the observations, along with their standard deviations, are listed in Table 3.4.

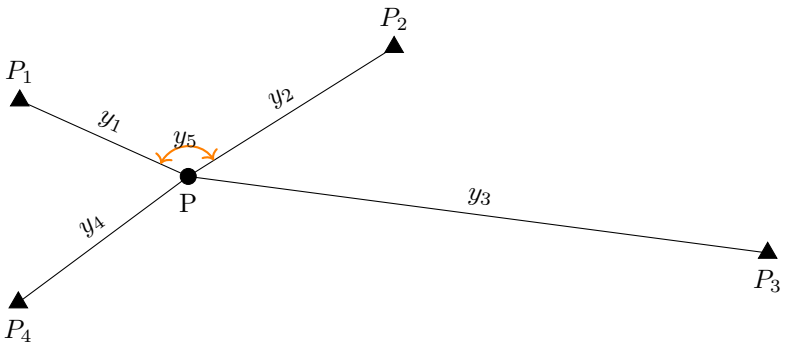

Figure 3.3: Four distances and one angle measured to determine point P

Table 3.3: Coordinates of known points in meters

Point	x_i [m]	y_i [m]
P_1	842.281	925.523
P_2	1337.544	996.249
P_3	1831.727	723.962
P_4	840.408	658.345

Table 3.4: Observations: distances and their standard deviation are in units of meters.

Observable	y_i	σ_i
P_1P	244.457	0.006
P_2P	321.622	0.010
P_3P	773.129	0.024
P_4P	280.019	0.080
$\angle P_1PP_2$	$123°38'20''$	$05''$

(a) Setup the observation equations and form the normal equations.

(b) Compute the LESS for the coordinates of points P and compute their variances and covariances.

(c) Compute the residual vector $\tilde{\boldsymbol{e}}$, the adjusted observations, and the dispersion matrices of both.

(d) Compute the estimated variance component $\hat{\sigma}_0^2$.

7. Pearson (1901) presented the data in Table 3.5 for a line-fitting problem. Considering the x-coordinates to be known with certainty and the y-coordinates to be observed with random errors having zero mean and iid dispersion, complete the following:

(a) Setup a Gauss-Markov Model to estimate the slope and y-intercept of a line and compute those estimates.

(b) Compute the residual vector $\tilde{\boldsymbol{e}}$ and the estimated variance component $\hat{\sigma}_0^2$.

(c) Plot the data along with the fitted line.

Table 3.5: Pearson (1901) data for a fitted line

Point i	x_i	y_i
1	0.0	5.9
2	0.9	5.4
3	1.8	4.4
4	2.6	4.6
5	3.3	3.5
6	4.4	3.7
7	5.2	2.8
8	6.1	2.8
9	6.5	2.4
10	7.4	1.5

8. The affine 2-D transformation is based on six unknown parameters:

   - ξ_1, ξ_2 for the *translation* of the origin of the coordinate frame,
   - β, $\beta + \epsilon$ for the *rotation* angles of the respective axes.
   - ω_1, ω_2 for the *scale* factors of the respective axes.

   For a point having coordinates (x_i, y_i) in the source coordinate frame and (X_i, Y_i) in the target coordinate frame, the transformation is described by

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} \omega_1 \cdot \cos\beta & -\omega_2 \cdot \sin(\beta+\epsilon) \\ \omega_1 \cdot \sin\beta & \omega_2 \cdot \cos(\beta+\epsilon) \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} e_{X_i} \\ e_{Y_i} \end{bmatrix}. \tag{3.52a}$$

   Here

   - (x_i, y_i) are given coordinates in the *source system*;
   - (X_i, Y_i) are observed coordinates in the *target system*;
   - i denotes the point number, $i \in \{1, 2, \ldots, n/2\}$.

   Making the substitutions

$$\xi_3 := \omega_1 \cos\beta,\ \xi_4 := \omega_2 \sin(\beta+\epsilon),\ \xi_5 := \omega_1 \sin\beta,\ \xi_6 := \omega_2 \cos(\beta+\epsilon) \tag{3.52b}$$

   results in the linear system of observation equations

$$\begin{aligned} X_i &= x_i \cdot \xi_3 - y_i \cdot \xi_4 + \xi_1 + e_{X_i}, \\ Y_i &= x_i \cdot \xi_5 + y_i \cdot \xi_6 + \xi_2 + e_{Y_i}, \end{aligned} \quad \begin{bmatrix} e_{X_i} \\ e_{Y_i} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_0^2 \begin{bmatrix} (Q_{XX})_{ii} & (Q_{XY})_{ii} \\ (Q_{XY}^T)_{ii} & (Q_{YY})_{ii} \end{bmatrix} \right), \tag{3.52c}$$

   where Q_{XX}, Q_{YY}, and Q_{XY} are given cofactor matrices.

   Using the data from Table 3.6, which is copied from Wolf (1983, p. 586), and assuming the random observation errors are iid, complete the following:

(a) Compute the least-squares estimates of $\hat{\boldsymbol{\xi}}$ and then the derived quantities $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\omega}_1$, and $\hat{\omega}_2$.

(b) Sketch a diagram showing the axes of both coordinate systems. Annotate the diagram with labels for the rotation angles between the axes and the translations between the two origins.

(c) Use the estimated parameters to compute coordinates in the xy system for points 1–3 shown in Table 3.6.

Table 3.6: Calibrated (known) and comparator (measured) coordinates from Wolf (1983, p. 586)

	Comparator coordinates		Calibrated coordinates	
Point	$\boldsymbol{X}$ [mm]	$\boldsymbol{Y}$ [mm]	$\boldsymbol{x}$ [mm]	$\boldsymbol{y}$ [mm]
Fiducial A	55.149	159.893	−113.000	0.000
Fiducial B	167.716	273.302	0.000	113.000
Fiducial C	281.150	160.706	113.000	0.000
Fiducial D	168.580	47.299	0.000	−113.000
1	228.498	105.029		
2	270.307	199.949		
3	259.080	231.064		

9. The spirit leveling data in Table 3.7 come from Rainsford (1968), where orthometric corrections have already been applied to the recorded observations. The weight of each observation was taken as the distance in miles divided by 100. All random observation errors are uncorrelated. The unknown parameters are the heights of points A, B, C, D, E, and F (Figure 3.4). Since the observations pertain to height differences, the model has a rank deficiency (datum deficiency) of one. Therefore, datum information is introduced as in Section 3.5 by specifying the height of point D as 1928.277 ft.

   Complete the following:

   (a) Set up a partitioned Gauss-Markov model and the corresponding partitioned least-squares normal equations according to Section 3.5.

   (b) Compute the LESS for the estimated heights of points A, B, C, E, and F.

   (c) Compute the residual vector and the estimated variance component.

   (d) Compute the adjusted observations and then sum them for each of the four closed loops in the network that pass through either point B or C. Also sum them for the closed perimeter loop that contains all points except B and C.

   (e) Repeat all your computations using a height of 1679.432 ft for point A as datum information. Which results are different and which are the

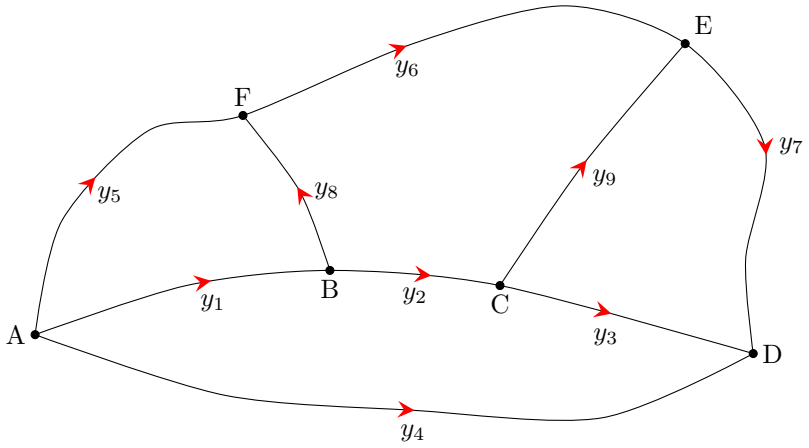

Figure 3.4: Leveling network after Rainsford (1968)

Table 3.7: Leveling data from Rainsford (1968)

From	To	No.	Observed height diff. [ft]	Length [miles]
A	B	1	+124.632	68
B	C	2	+217.168	40
C	D	3	−92.791	56
A	D	4	+248.754	171
A	F	5	−11.418	76
F	E	6	−161.107	105
E	D	7	+421.234	80
B	F	8	−135.876	42
C	E	9	−513.895	66

same between the two adjustments? Can you explain the differences and similarities?

3.7 Summary Formulas for the Introduction of Datum Information for the Least-Squares Solution Within the Rank Deficient Gauss-Markov Model

The rank deficient Gauss-Markov Model is given by

$$\underset{n\times 1}{\boldsymbol{y}} = \left[\begin{array}{c|c} \underset{n\times q}{A_1} & \underset{n\times(m-q)}{A_2} \end{array}\right] \begin{bmatrix} \underset{q\times 1}{\boldsymbol{\xi}_1} \\ \underset{(m-q)\times 1}{\boldsymbol{\xi}_2} \end{bmatrix} + \boldsymbol{e}, \quad \boldsymbol{e} \sim \left(\boldsymbol{0}, \sigma_0^2 P^{-1}\right),$$
$$\operatorname{rk} A =: q < m \quad \text{and} \quad \operatorname{rk} A_1 = q.$$

Table 3.8: Summary formulas for the introduction of datum information ($\hat{\boldsymbol{\xi}}_2 \rightarrow \boldsymbol{\xi}_2^0$) for the LESS within the rank deficient Gauss-Markov Model

Quantity	Formula	Eq.
Model redundancy	$r = n - \operatorname{rk} A = n - q$	(3.46)
Vector of estimated parameters, with given $\boldsymbol{\xi}_2^0$	$\hat{\boldsymbol{\xi}}_1 = N_{11}^{-1}\left(\boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0\right)$	(3.41b)
Dispersion matrix for estimated parameters	$D\{\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2 \cdot N_{11}^{-1}$	(3.42)
Vector of predicted residuals, with given $\boldsymbol{\xi}_2^0$	$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} = \boldsymbol{y} - A_1\hat{\boldsymbol{\xi}}_1 - A_2\boldsymbol{\xi}_2^0$	(3.43b)
Dispersion matrix for residuals	$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot \left(P^{-1} - A_1 N_{11}^{-1} A_1^T\right)$	(3.43b)
Sum of squared residuals (SSR)	$\Omega = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$	(3.45)
Estimated variance component, with given $\boldsymbol{\xi}_2^0$	$\hat{\sigma}_0^2 = (\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}})/r = (\boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}_1^T \hat{\boldsymbol{\xi}}_1 - \boldsymbol{c}_2^T \boldsymbol{\xi}_2^0)/(n-q)$	(3.47)

Continued on next page

Quantity	Formula	Eq.
Vector of adjusted observations	$\widehat{E\{\boldsymbol{y}\}} =: \hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}} = A_1\hat{\boldsymbol{\xi}}_1 + A_2\boldsymbol{\xi}_2^0$	(3.44a)
Dispersion matrix for adjusted observations	$D\{\hat{\boldsymbol{\mu}}_y\} = \sigma_0^2 \cdot A_1 N_{11}^{-1} A_1^T$	(3.44b)

Continued from previous page

3.8 Summary Formulas for the Least-Squares Solution Within the Gauss-Markov Model With Full Rank

The Gauss-Markov Model with full column rank coefficient matrix A is given by

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \quad \boldsymbol{e} \sim \left(\boldsymbol{0}, \sigma_0^2 P^{-1}\right), \qquad \operatorname{rk} A = m.$$

Table 3.9: Summary formulas for the LESS within the Gauss-Markov Model with full rank

Quantity	Formula	Eq.
Model redundancy	$r = n - \operatorname{rk} A = n - m$	(3.3)
Vector of estimated parameters	$\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c},\ [N,\, \boldsymbol{c}] \coloneqq A^T P[A,\, \boldsymbol{y}]$	(3.7)
Dispersion matrix for estimated parameters	$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 \cdot N^{-1}$	(3.13)
Vector of predicted residuals	$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} = \left(I_n - AN^{-1}A^T P\right)\boldsymbol{y}$	(3.9)
Dispersion matrix for residuals	$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot \left(P^{-1} - AN^{-1}A^T\right)$	(3.14a)
Sum of squared residuals (SSR)	$\Omega = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$	(3.22)

Continued on next page

Quantity	Formula	Eq.
Estimated variance component	$\hat{\sigma}_0^2 = (\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}})/(n - \operatorname{rk} A)$	(3.28)
Vector of adjusted observations	$\widehat{E\{\boldsymbol{y}\}} =: \hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}}$	(3.11)
Dispersion matrix for adjusted observations	$D\{\hat{\boldsymbol{\mu}}_y\} = \sigma_0^2 \cdot A N^{-1} A^T$	(3.15)

Continued from previous page

Chapter 4

The Model of Condition Equations

4.1 Model Definition

In the least-squares adjustment within the *model of condition equations*, the unknown parameters $\boldsymbol{\xi}$ are not estimated directly, rather the random error vector $\boldsymbol{e}$ is predicted. This approach might be taken if the parameters are of no particular interest, or it might be done to make the problem easy to formulate. An example of the latter is the adjustment of leveling networks, where the parameters (heights of the stations) are of primary interest, but because closed "level loops" within the network sum to zero (a necessary condition), it is convenient to difference the observations along these loops before performing the adjustment (see level-loop example in Section 4.4). Another motivation for using the model of condition equations is that the size of the matrix to invert in the least-squares solution (LESS) may be smaller than that in the corresponding LESS within the Gauss-Markov Model (GMM).

Let the $r \times n$ matrix B represent a difference operator such that when it is applied to the $n \times 1$ observation equations $\boldsymbol{y} = A\boldsymbol{\xi} + \boldsymbol{e}$, the parameters are eliminated. More specifically, we require that $BA = 0$, which implies that $B\boldsymbol{y} = B(A\boldsymbol{\xi} + \boldsymbol{e}) = B\boldsymbol{e}$. Therefore, by applying the difference operator B, the GMM is transformed to the following model of condition equations:

$$\underset{r\times 1}{\boldsymbol{w}} \coloneqq \underset{r\times n}{B}\,\boldsymbol{y} = B\boldsymbol{e},\ \underset{n\times 1}{\boldsymbol{e}} \sim (\boldsymbol{0}, \sigma_0^2 P^{-1}), \tag{4.1a}$$

$$r \coloneqq n - q = \operatorname{rk} B, \tag{4.1b}$$

where the variable r denotes the redundancy of the model, and q is the rank of the $n \times m$ matrix A from the GMM (3.1). Equation (4.1b) requires that matrix B has full row rank. Moreover, it shows that the redundancy of the model is not changed from that of the GMM by eliminating parameters.

4.2 The Least-Squares Solution Within the Model of Condition Equations

The least-squares criterion for minimizing the (P-weighted, squared) norm of the observation residuals is written as

$$\min \boldsymbol{e}^T P \boldsymbol{e} \text{ subject to } \boldsymbol{w} = B\boldsymbol{e}, \tag{4.2}$$

for which the Lagrange target function

$$\Phi(\boldsymbol{e}, \boldsymbol{\lambda}) := \boldsymbol{e}^T P \boldsymbol{e} + 2\boldsymbol{\lambda}^T(\boldsymbol{w} - B\boldsymbol{e}) \tag{4.3}$$

can be written, which must be made stationary with respect to the unknown terms $\boldsymbol{e}$ and $\boldsymbol{\lambda}$. Here, $\boldsymbol{\lambda}$ is an $r \times 1$ vector of Lagrange multipliers. Taking the first partial derivatives of (4.3) leads to the Euler-Lagrange necessary conditions

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{e}} = P\tilde{\boldsymbol{e}} - B^T\hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{4.4a}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = \boldsymbol{w} - B\tilde{\boldsymbol{e}} \doteq \boldsymbol{0}. \tag{4.4b}$$

The sufficient condition, required to ensure a minimum is reached, is satisfied by $\partial\Phi^2/\partial\boldsymbol{e}\partial\boldsymbol{e}^T = 2P$, which is positive definite since the weight matrix P is invertible according to (4.1a). The simultaneous solution of (4.4a) and (4.4b) leads to the *Best LInear Prediction* (BLIP) of $\boldsymbol{e}$ as derived in the following: Equation (4.4a) leads to

$$\tilde{\boldsymbol{e}} = P^{-1}B^T\hat{\boldsymbol{\lambda}}. \tag{4.5a}$$

Then, (4.4b) and (4.5a) allows

$$\boldsymbol{w} = B\tilde{\boldsymbol{e}} = \left(BP^{-1}B^T\right)\hat{\boldsymbol{\lambda}} \Rightarrow \tag{4.5b}$$

$$\hat{\boldsymbol{\lambda}} = \left(BP^{-1}B^T\right)^{-1}\boldsymbol{w} \Rightarrow \tag{4.5c}$$

$$\tilde{\boldsymbol{e}} = P^{-1}B^T\left(BP^{-1}B^T\right)^{-1}\boldsymbol{w}, \tag{4.5d}$$

finally leading to the predicted random error vector

$$\boxed{\tilde{\boldsymbol{e}} = P^{-1}B^T\left(BP^{-1}B^T\right)^{-1}B\boldsymbol{y}.} \tag{4.5e}$$

Note that the matrix product $BP^{-1}B^T$ results in a symmetric, positive definite matrix of size $r \times r$, since B has full row rank. The predicted random error vector $\tilde{\boldsymbol{e}}$ is also called the *residual* vector. The expectation of the given observation vector is expressed as $E\{\boldsymbol{y}\} = \boldsymbol{\mu}_y$, where $\boldsymbol{\mu}_y$ is the true, but unknown, vector of observables. Thus we write the vector of adjusted observations as

$$\boxed{\widehat{E\{\boldsymbol{y}\}} = \hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}}.} \tag{4.6}$$

Nota bene: Implicit in the term $B\boldsymbol{y}$ is the subtraction of a constant term $\boldsymbol{\kappa}$ if necessary, viz. $(B\boldsymbol{y} - \boldsymbol{\kappa}) - B\boldsymbol{e} = \boldsymbol{0}$, implying that $B\boldsymbol{y} \to B\boldsymbol{y} - \boldsymbol{\kappa}$. An example is the condition that the n interior angles of a simple polygon in a plane must sum to $\kappa = (n-2)180°$. Then the condition equation would read

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 - e_1 \\ y_2 - e_2 \\ \vdots \\ y_n - e_n \end{bmatrix} - (n-2)\pi = 0.$$

Thus, for numerical computations, we may need to modify (4.5e) to read

$$\tilde{\boldsymbol{e}} = P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}(B\boldsymbol{y} - \boldsymbol{\kappa}), \tag{4.7}$$

which has no affect on the dispersion formulas that follow.

The square of the P-weighted residual norm Ω, also called the sum of squared residuals (SSR), is computed by

$$\boxed{\begin{aligned} \Omega = \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} = \tilde{\boldsymbol{e}}^T B^T\hat{\boldsymbol{\lambda}} = \boldsymbol{w}^T\hat{\boldsymbol{\lambda}} = \boldsymbol{w}^T(BP^{-1}B^T)^{-1}\boldsymbol{w} = \qquad (4.8a) \\ = \boldsymbol{y}^T B^T(BP^{-1}B^T)^{-1}B\boldsymbol{y}, \qquad (4.8b) \end{aligned}}$$

leading to the estimated variance component

$$\boxed{\hat{\sigma}_0^2 = \frac{\Omega}{r} = \frac{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}}{r},} \tag{4.9}$$

with $r = \operatorname{rk} B$. In words, it is described as the squared P-weighted residual norm divided by the degrees of freedom (redundancy) of the model.

Applying the law of error propagation, the dispersion of the residual vector is computed by

$$\begin{aligned} D\{\tilde{\boldsymbol{e}}\} &= P^{-1}B^T(BP^{-1}B^T)^{-1}B \cdot D\{\boldsymbol{y}\} \cdot B^T(BP^{-1}B^T)^{-1}BP^{-1} = \\ &= P^{-1}B^T(BP^{-1}B^T)^{-1}B(\sigma_0^2P^{-1})B^T(BP^{-1}B^T)^{-1}BP^{-1} \Rightarrow \\ &\boxed{D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1}.} \end{aligned} \tag{4.10}$$

As we did earlier within the GMM (Section 3.2.2), we compute the covariance between the residual vector $\tilde{\boldsymbol{e}}$ and the vector adjusted observations $\hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}}$ as follows:

$$\begin{aligned} C\{\hat{\boldsymbol{\mu}}_y, \tilde{\boldsymbol{e}}\} &= C\{\big[I - P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{y},\, P^{-1}B^T\big(BP^{-1}B^T\big)B\boldsymbol{y}\} = \\ &= \big[I - P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] \cdot D\{\boldsymbol{y}\} \cdot \big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]^T = \\ &= \big[I - P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] \cdot \sigma_0^2P^{-1} \cdot B^T\big(BP^{-1}B^T\big)^{-1}BP^{-1} = \end{aligned}$$

$$= \sigma_0^2 \big[P^{-1} B^T \big(BP^{-1} B^T \big)^{-1} BP^{-1} - P^{-1} B^T \big(BP^{-1} B^T \big)^{-1} BP^{-1} B^T \cdot \\ \cdot \big(BP^{-1} B^T \big)^{-1} BP^{-1} \big] = 0 \tag{4.11}$$

Thus, it has been shown that the residuals and adjusted observations are uncorrelated, and therefore the dispersion of the adjusted observations can be written as

$$\boxed{D\{\hat{\boldsymbol{\mu}}_y\} = D\{\boldsymbol{y}\} - D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \big[P^{-1} - P^{-1} B^T (BP^{-1} B^T)^{-1} BP^{-1} \big].} \tag{4.12}$$

Note that B is not a unique matrix, but regardless of how B is chosen the results of the adjustment will be the same, provided the following necessary conditions for B are satisfied:

(i) *Dimensionality:* $\operatorname{rk} B = n - \operatorname{rk} A = n - q = r$, which means that $\operatorname{rk} B + \operatorname{rk} A = (n - q) + q = n$.

(ii) *Orthogonality:* $BA = 0$.

4.3 Equivalence Between LESS Within the Gauss-Markov Model and the Model of Condition Equations

To show the equivalence between the least-squares adjustments within the GMM and the model of condition equations, it must be shown that the predicted random error vectors (residuals) from both adjustments are equivalent. The residual vector $\tilde{\boldsymbol{e}}$ from each adjustment can be expressed as a projection matrix times the true random error vector $\boldsymbol{e}$ (or equivalently, times the observation vector $\boldsymbol{y}$) as shown below.

The residual vector within the GMM can be written as

$$\tilde{\boldsymbol{e}} = \big[I_n - AN^{-1} A^T P \big] \boldsymbol{e}. \tag{4.13}$$

And the residual vector within the model of condition equations can be written as

$$\tilde{\boldsymbol{e}} = \big[P^{-1} B^T \big(BP^{-1} B^T \big)^{-1} B \big] \boldsymbol{e}. \tag{4.14}$$

Note that the right sides of (4.13) and (4.14) cannot actually be computed since $\boldsymbol{e}$ is unknown, but the equations do hold since, for the GMM,

$$\begin{aligned} \tilde{\boldsymbol{e}} &= \big[I_n - AN^{-1} A^T P \big] \boldsymbol{y} = \\ &= \big[I_n - AN^{-1} A^T P \big] (A\boldsymbol{\xi} + \boldsymbol{e}) = \\ &= \big[A\boldsymbol{\xi} - AN^{-1} (A^T P A) \boldsymbol{\xi} \big] + \big[I_n - AN^{-1} A^T P \big] \boldsymbol{e} \Rightarrow \\ \tilde{\boldsymbol{e}} &= \big[I_n - AN^{-1} A^T P \big] \boldsymbol{e}, \end{aligned} \tag{4.15}$$

and, for the model of condition equations,

$$\tilde{\boldsymbol{e}} = P^{-1} B^T \big(BP^{-1} B^T \big)^{-1} B\boldsymbol{y} =$$

$$= P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B(A\boldsymbol{\xi}+\boldsymbol{e}) \Rightarrow$$
$$\tilde{\boldsymbol{e}} = \big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{e}, \tag{4.16}$$

using the fact that $BA = 0$.

To show that (4.13) and (4.14) are equivalent, it must be shown that the range spaces and the nullspaces are equivalent for their respective projection matrices

$$\bar{P}_1 \coloneqq [I_n - AN^{-1}A^TP] \text{ and } \bar{P}_2 \coloneqq [P^{-1}B^T(BP^{-1}B^T)^{-1}B].$$

(i) Equivalent range spaces: Show that

$$\mathcal{R}\big[I_n - AN^{-1}A^TP\big] = \mathcal{R}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big].$$

Proof: Since $A^TPP^{-1}B^T = A^TB^T = 0$, then

$$\big[I_n - AN^{-1}A^TP\big]\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{z} =$$
$$= \big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]\boldsymbol{z} - \mathbf{0} \text{ for any } \boldsymbol{z} \in \mathbb{R}^n,$$

which, according to (1.3), implies that

$$\mathcal{R}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] \subset \mathcal{R}\big[I_n - AN^{-1}A^TP\big].$$

Also:

$$\begin{aligned}
&\dim \mathcal{R}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] = \\
&= \operatorname{rk}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] = && \text{using (A.19a)}\\
&= \operatorname{tr}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big] = && \text{using (1.7c)}\\
&= \operatorname{tr}\big[BP^{-1}B^T\big(BP^{-1}B^T\big)^{-1}\big] = && \text{using (A.5)}\\
&= \operatorname{tr} I_r = r.
\end{aligned}$$

Furthermore:

$$\begin{aligned}
&\dim \mathcal{R}\big[I_n - AN^{-1}A^TP\big] = \\
&= \operatorname{rk}\big(I_n - AN^{-1}A^TP\big) = && \text{using (A.19a)}\\
&= \operatorname{tr}\big(I_n - AN^{-1}A^TP\big) = && \text{using (1.7c)}\\
&= \operatorname{tr} I_n - \operatorname{tr}\big(N^{-1}A^TPA\big) = && \text{using (A.5)}\\
&= n - \operatorname{rk} N = n - \operatorname{rk} A = \\
&= n - q = r,
\end{aligned}$$

which implies that

$$\boxed{\mathcal{R}\big[I_n - AN^{-1}A^TP\big] = \mathcal{R}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big],} \tag{4.17}$$

since one range space contains the other and both have the same dimension. Thus we have shown that the range spaces (column spaces) of $\bar{P}_1$ and $\bar{P}_2$ are equivalent. Now we turn to the nullspaces.

(ii) Equivalent Nullspaces: Show that

$$\mathcal{N}\big[I_n - AN^{-1}A^TP\big] = \mathcal{N}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big].$$

Proof:
First show that $\mathcal{N}\big[I_n - AN^{-1}A^TP\big] = \mathcal{R}(A)$.

We begin with

$$\big[I_n - AN^{-1}A^TP\big]A\boldsymbol{\alpha} = \mathbf{0} \text{ for all } \boldsymbol{\alpha},$$

which implies that

$$\mathcal{R}(A) \subset \mathcal{N}\big[I_n - AN^{-1}A^TP\big], \text{ since } A\boldsymbol{\alpha} \subset \mathcal{R}(A);$$

also

$$\dim \mathcal{R}(A) = \operatorname{rk} A = q.$$

Equations (A.19a) and (A.19b) reveal that the sum of the dimensions of the range space and nullspace of a matrix is equal to its number of columns. Using this property, and results from (i), we find that

$$\dim \mathcal{N}\big[I_n - AN^{-1}A^TP\big] =$$
$$= n - \dim \mathcal{R}\big[I_n - AN^{-1}A^TP\big] = n - r = q.$$

Therefore,

$$\mathcal{N}\big[I_n - AN^{-1}A^TP\big] = \mathcal{R}(A).$$

Also, we have

$$\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big]A = 0,$$

since $BA = 0$. The preceding development implies that

$$\mathcal{R}(A) = \mathcal{N}\big[I_n - AN^{-1}A^TP\big] \subset \mathcal{N}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big],$$

or

$$\mathcal{N}(\bar{P}_1) \subset \mathcal{N}(\bar{P}_2).$$

We showed in part (i) that the dimensions of the range spaces of the respective projection matrices are equivalent. And, since

$$\dim \mathcal{N}(\bar{P}_1) = n - \dim \mathcal{R}(\bar{P}_1) = n - \dim \mathcal{R}(\bar{P}_2),$$

it follows that

$$\dim \mathcal{N}(\bar{P}_1) = \dim \mathcal{N}(\bar{P}_2).$$

As already stated in part (i), if one vector space is a subset of another and both spaces have the same dimension, then the subspaces are equivalent. Therefore, we can say that

$$\mathcal{N}(\bar{P}_1) = \mathcal{N}(\bar{P}_2),$$

or

$$\boxed{\mathcal{N}\big[I_n - AN^{-1}A^TP\big] = \mathcal{N}\big[P^{-1}B^T\big(BP^{-1}B^T\big)^{-1}B\big].} \tag{4.18}$$

We have shown that both the range spaces and nullspaces of the projection matrices $\bar{P}_1$ and $\bar{P}_2$ are equivalent, proving that *the residual vectors from the two adjustments are the same and thus that the two adjustments are indeed equivalent.*

4.4 Examples — Linear and Nonlinear

4.4.1 Linear Example — a Small Leveling Network

The following example is borrowed from Mikhail and Gracie (1981, Problem 4-8). It involves a leveling network comprised of two closed loops as shown in Figure 4.1. The data are listed in Table 4.1.

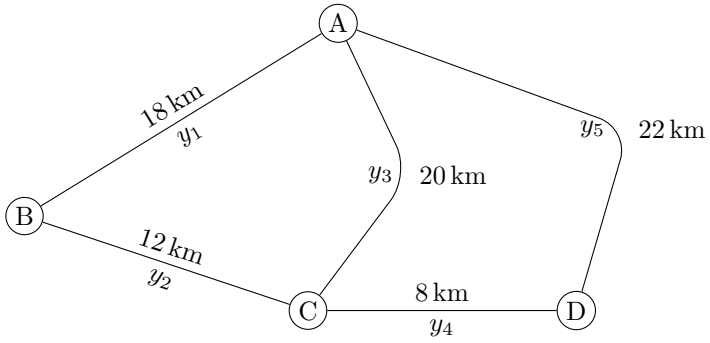

Figure 4.1: Example leveling network

Table 4.1: Leveling network data

Line	Element of $\boldsymbol{y}$	Observed elevation difference (m)	Length (km)
A to B	y_1	−12.386	18
B to C	y_2	−11.740	12
C to A	y_3	24.101	20
C to D	y_4	−8.150	8
D to A	y_5	32.296	22

In a leveling network, one condition equation can be written for each closed loop. Connecting observations in a counter-clockwise order, two condition equations may

be written as

$$(y_1 - e_1) + (y_2 - e_2) + (y_3 - e_3) = 0,$$
$$-(y_3 - e_3) + (y_4 - e_4) + (y_5 - e_5) = 0,$$

or, in matrix form, as

$$B\boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -12.386 \\ -11.740 \\ 24.101 \\ -8.150 \\ 32.296 \end{bmatrix} = B\boldsymbol{e}.$$

The observations weights are inversely proportional to the distances in km, so that

$$P^{-1} = 10^{-6} \cdot \mathrm{diag}(18, 12, 20, 8, 22) \cdot \mathrm{m}^2$$

appears to be a reasonable weight matrix. The residuals are then computed by

$$\tilde{\boldsymbol{e}} = P^{-1}B^T(BP^{-1}B^T)^{-1}B\boldsymbol{y} = \begin{bmatrix} -0.003 \\ -0.002 \\ -0.020 \\ 0.007 \\ 0.018 \end{bmatrix} \cdot \mathrm{m}.$$

The redundancy of the model is given by $r = \mathrm{rk}\, B = 2$. The adjusted observations are computed by

$$\hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}} = \begin{bmatrix} -12.383 \\ -11.738 \\ 24.121 \\ -8.157 \\ 32.278 \end{bmatrix} \cdot \mathrm{m}.$$

The dispersion matrix for the residuals is

$$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1} =$$
$$= \sigma_0^2 \cdot \begin{bmatrix} 7.7 & 5.1 & 5.1 & 1.4 & 3.8 \\ 5.1 & 3.4 & 3.4 & 0.9 & 2.5 \\ 5.1 & 3.4 & 11.4 & -2.3 & -6.3 \\ 1.4 & 0.9 & -2.3 & 1.5 & 4.2 \\ 3.8 & 2.5 & -6.3 & 4.2 & 11.5 \end{bmatrix} \mathrm{mm}^2.$$

The weighted sum of squared residuals is $\Omega := \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = (6.454972)^2$, leading to the estimated variance component $\hat{\sigma}_0^2 = \Omega/r = (4.564355)^2$. The estimated variance

component may be used to compute the *estimated dispersion matrix* for the residuals as

$$\hat{D}\{\tilde{\boldsymbol{e}}\} = \hat{\sigma}_0^2 \cdot P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1} = \\ = \begin{bmatrix} 1.61 & 1.07 & 1.07 & 0.29 & 0.79 \\ 1.07 & 0.71 & 0.71 & 0.19 & 0.52 \\ 1.07 & 0.71 & 2.38 & -0.48 & -1.31 \\ 0.29 & 0.19 & -0.48 & 0.32 & 0.87 \\ 0.79 & 0.52 & -1.31 & 0.87 & 2.40 \end{bmatrix} \text{cm}^2.$$

Now, if the same problem were to be modeled within the Gauss-Markov Model with the unknown parameters being the heights of the points denoted by $\boldsymbol{\xi} = [H_A, H_B, H_C, H_D]^T$, then the coefficient matrix would be written as

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

Obviously, the conditions $r = n - \operatorname{rk} A = \operatorname{rk} B = 2$ and $BA = 0$ are satisfied. Indeed, one can easily verify that the LESS within the GMM will produce the same residual vector and same estimated variance component as shown above.

4.4.2 Nonlinear Example — Observations of a Triangle

Table 4.2 lists distance observations for all sides of a triangle and two of its angles as depicted in Figure 4.2. The standard deviations of the observations are shown in the last column of the table. The observations are to be adjusted by computing the residual vector within the model of condition equations.

Table 4.2: Observations of sides and angles of a triangle

Obs. no.	Observation	Std. dev.
y_1	120.01 m	1 cm
y_2	105.02 m	1 cm
y_3	49.98 m	1 cm
y_4	$94°47'10''$	$20''$
y_5	$60°41'20''$	$20''$

The following two nonlinear condition equations can be written as a function of the unknown 5×1 random error vector $\boldsymbol{e}$, the first based on the law of sines and

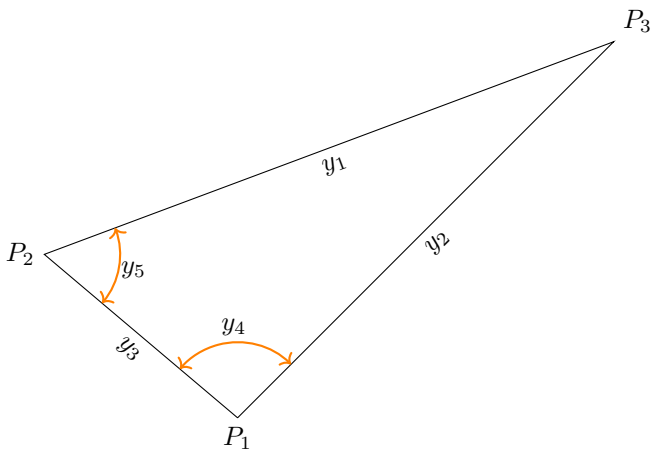

Figure 4.2: Observations of sides and angles of a triangle

the second on the law of cosines for a triangle in a plane:

$$f_1(\boldsymbol{e}) = (y_2 - e_2)\cdot\sin(y_4 - e_4) - (y_1 - e_1)\cdot\sin(y_5 - e_5) = 0 \tag{4.19a}$$

$$\begin{aligned} f_2(\boldsymbol{e}) = {} & (y_1 - e_1)^2 + (y_2 - e_2)^2 - (y_3 - e_3)^2 - \\ & - 2\cdot(y_1 - e_1)(y_2 - e_2)\cdot\cos(\pi - y_4 + e_4 - y_5 + e_5) = 0. \end{aligned} \tag{4.19b}$$

The following total derivatives are written for the sake of forming partial derivatives that are needed for linearization:

$$\begin{aligned} \mathrm{d}f_1 = {} & -\sin(y_4 - e_4)\mathrm{d}e_2 - (y_2 - e_2)\cos(y_4 - e_4)\mathrm{d}e_4 + \sin(y_5 - e_5)\mathrm{d}e_1 + \\ & + (y_1 - e_1)\cos(y_5 - e_5)\mathrm{d}e_5 = 0, \end{aligned} \tag{4.20a}$$

$$\begin{aligned} \mathrm{d}f_2 = {} & \big[-2(y_1 - e_1) + 2(y_2 - e_2)\cos(\pi - y_4 + e_4 - y_5 + e_5)\big]\mathrm{d}e_1 + \\ & + \big[-2(y_2 - e_2) + 2(y_1 - e_1)\cos(\pi - y_4 + e_4 - y_5 + e_5)\big]\mathrm{d}e_2 + \\ & + 2(y_3 - e_3)\mathrm{d}e_3 + \big[2(y_1 - e_1)(y_2 - e_2)\sin(\pi - y_4 + e_4 - y_5 + e_5)\big](\mathrm{d}e_4 + \mathrm{d}e_5). \end{aligned} \tag{4.20b}$$

From these equations we get the partial derivatives $\partial f_1/\partial e_2 = -\sin(y_4 - e_4)$, etc., leading to the Jacobian matrix

$$B = \begin{bmatrix} \frac{\partial f_1}{\partial e_1} & \frac{\partial f_1}{\partial e_2} & \frac{\partial f_1}{\partial e_3} & \frac{\partial f_1}{\partial e_4} & \frac{\partial f_1}{\partial e_5} \\ \frac{\partial f_2}{\partial e_1} & \frac{\partial f_2}{\partial e_2} & \frac{\partial f_2}{\partial e_3} & \frac{\partial f_2}{\partial e_4} & \frac{\partial f_2}{\partial e_5} \end{bmatrix}, \tag{4.20c}$$

which must have rank 2 (full row rank).

The problem is linearized by the truncated Taylor series

$$\boldsymbol{f}(\boldsymbol{e}) \approx \boldsymbol{f}(\boldsymbol{e}_0) + \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{e}^T}\right|_{\boldsymbol{e}=\boldsymbol{e}_0} \cdot(\boldsymbol{e} - \boldsymbol{e}_0) = \boldsymbol{0} \tag{4.21a}$$

about the expansion point $\boldsymbol{e}_0$, being an approximate value for the unknown vector of random errors $\boldsymbol{e}$. Using matrix B, evaluated at $\boldsymbol{e}_0$, to represent the partial derivatives, and introducing $\Delta\boldsymbol{e} \coloneqq \boldsymbol{e} - \boldsymbol{e}_0$ as an unknown, incremental vector of residuals, leads to the formula

$$-\boldsymbol{f}(\boldsymbol{e}_0) = B{\cdot}\Delta\boldsymbol{e}, \tag{4.21b}$$

which is in the form of

$$\boldsymbol{w} = B\boldsymbol{e} \tag{4.21c}$$

given in the model of condition equations. Therefore, we can setup an iterative algorithm to predict $\Delta\boldsymbol{e}$ as follows:

1. Set $\boldsymbol{e}_0 = \boldsymbol{0}$ and choose a convergence criterion ϵ.

2. Then for $j = 1, 2, \ldots$, while $\widetilde{\Delta\boldsymbol{e}}_j > \epsilon$, compute:

$$\widetilde{\Delta\boldsymbol{e}}_j = P^{-1}B_j^T(B_jP^{-1}B_j^T)^{-1}\boldsymbol{w}_j \tag{4.22a}$$

$$\tilde{\boldsymbol{e}}_j = \boldsymbol{e}_j + \widetilde{\Delta\boldsymbol{e}}_j. \tag{4.22b}$$

Then update the expansion point, the Jacobian matrix, and the vector $\boldsymbol{w}$ for the next iteration as follows:

$$\boldsymbol{e}_{j+1} = \tilde{\boldsymbol{e}}_j - \underset{\sim}{\boldsymbol{0}},\ \ B_{j+1} = B|_{\boldsymbol{e}_{j+1}},\ \text{ and }\ \boldsymbol{w}_{j+1} = -\boldsymbol{f}(\boldsymbol{e}_{j+1}). \tag{4.22c}$$

For the first iteration, the matrix B and vector $\boldsymbol{w}$ read

$$B = \begin{bmatrix} 0.08719744 & -0.09965131 & 0 & 8.762479 & 58.75108 \\ -48.92976 & 8.325453 & 99.96000 & 10463.14 & 10463.14 \end{bmatrix}$$

and

$$\boldsymbol{w} = \begin{bmatrix} -0.00816522 \\ -0.86019942 \end{bmatrix}.$$

Upon convergence the predicted residual vector turns out to be

$$\tilde{\boldsymbol{e}} = \begin{bmatrix} -0.0021\,\mathrm{m} \\ 0.0035\,\mathrm{m} \\ -0.0024\,\mathrm{m} \\ -05.6'' \\ -09.2'' \end{bmatrix}. \tag{4.23}$$

Note that when choosing a numerical value for the convergence criterion ϵ, one must be mindful of the units involved in the residual vector. In this example, we have units of meters, for which a change of less than 0.1 mm might be satisfactory, but we also have units of radians for the angles, for which a change of less than 5×10^{-6}rad might be required. In such cases it may be prudent to check the elements of $\widetilde{\Delta\boldsymbol{e}}_j$ individually, using separate convergence criterion for different observation types. Then, the algorithm would be considered to have converged when all the convergence criteria have been satisfied.

4.5 Generation of Equivalent Condition Equations When the Gauss-Markov Model is Rank Deficient

We may also wish to transform the rank-deficient model (3.38a) into a model of condition equations. To do so, consider the further splitting of the rank-deficient matrix A defined in (3.38b) as follows:

$$\underset{n\times m}{A} = \left[A_1 \;\middle|\; A_2\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{4.24a}$$

$$\dim(A_{11}) = q \times q \;\text{ and }\; \dim(A_{22}) = (n-q)\times(m-q). \tag{4.24b}$$

Also, we have $\operatorname{rk} A_{11} = q \coloneqq \operatorname{rk} A$. And, with the introduction of the $q \times (m-q)$ matrix L in (3.48a), satisfying $A_2 = A_1 L$, we may write

$$A_2 = \begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix} = A_1 L = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} L \Rightarrow A = \left[A_1 \;\middle|\; A_1 L\right]. \tag{4.25}$$

Now, the matrix B within the model of condition equations could be chosen as

$$\underset{r\times n}{B} \coloneqq \left[A_{21}A_{11}^{-1} \;\middle|\; -I_{n-q}\right], \tag{4.26}$$

with r being the redundancy of the model as shown in (3.46) and (4.1b). This is a legitimate choice for B as long as the two conditions discussed in Section 4.2, are satisfied, viz the dimensionality condition and the orthogonality condition.

The first condition requires that the dimensions of the column spaces of A and B sum to the number of observations n. The second condition requires that the rows of matrix B are orthogonal to the columns of A, i.e., $BA = 0$. Taken together, these conditions mean that A and B^T are orthogonal complements in n-dimensional space, or, stated more succinctly,

$$\mathcal{R}(A) \overset{\perp}{\oplus} \mathcal{R}(B^T) = \mathbb{R}^n. \tag{4.27}$$

Both conditions i and ii are satisfied for (4.26) as shown below.

i. Dimensionality condition:

$$\operatorname{rk} B = r = n - q = n - \operatorname{rk} A \Rightarrow \operatorname{rk} A + \operatorname{rk} B = n. \tag{4.28a}$$

ii. Orthogonality condition:

$$BA = B\left[A_1 \;\middle|\; A_2\right] = BA_1\left[I_q \;\middle|\; L\right], \tag{4.28b}$$

but

$$BA_1 = \left[A_{21}A_{11}^{-1} \;\middle|\; -I_{n-q}\right]\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} = A_{21}A_{11}^{-1}A_{11} - A_{21} = 0, \tag{4.28c}$$

and therefore

$$BA = 0. \tag{4.28d}$$

Note that as long as the rank of matrix A is known, we can always generate a splitting of A as shown in (4.24a); however, we may need to reorder the columns of A (tantamount to reordering the elements of the parameter vector) to ensure that A_{11} has full column rank.

4.6 Practice Problems

1. Practice deriving the formula for the residual vector $\tilde{\boldsymbol{e}}$ as shown in Section 4.2 until you can do it without referring to the notes.

2. Compute the residual vector of Problem 9 of Section 3.6 using the LESS within the model of condition equations. Confirm that the rank of matrix B is $n-5$ and that $BA = 0$, where A is the coefficient matrix from problem 9.

3. The observations listed in Table 4.3 are depicted in Figure 4.3. Assume that the listed angles were derived from differences of independently observed directions measured with a theodolite. For example, observation y_2 was derived from subtracting the observed direction from point P_2 to point P_3 from the direction from P_2 to P_4. The variance of each direction is $\sigma^2 = (10'')^2$.

   (a) Determine the variance of each of the six angles as well as the covariance between angles y_2 and y_3 and the covariance between angles y_4 and y_5. Based on these results, write down the covariance matrix Q.

   (b) Write down suitable condition equations and determine the redundancy of the model.

   (c) Using the LESS within the model of condition equations, compute the residual vector $\tilde{\boldsymbol{e}}$ and its dispersion matrix $D\{\tilde{\boldsymbol{e}}\}$.

   (d) Compute the estimated variance component $\hat{\sigma}_0^2$.

Table 4.3: Six measured angles between four points

Element of $\boldsymbol{y}$	Observation
y_1	$37°52'35''$
y_2	$46°56'10''$
y_3	$57°18'50''$
y_4	$37°52'40''$
y_5	$53°44'50''$
y_6	$31°03'20''$

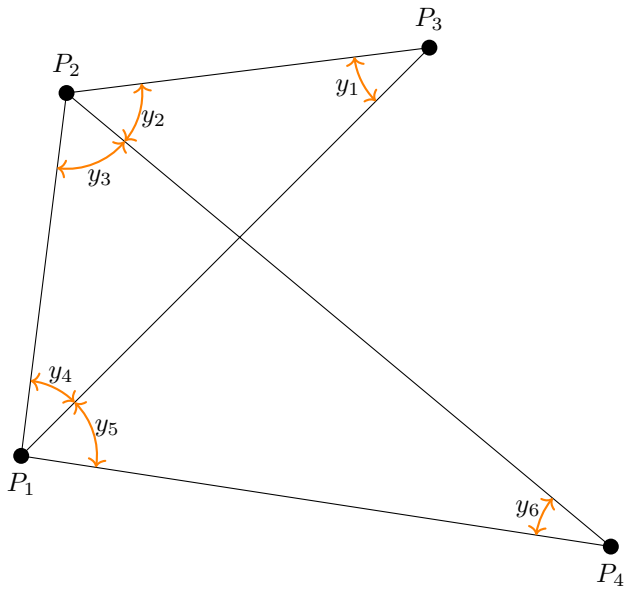

Figure 4.3: Six measured angles between four points

4. Four distances were measured between three points A, B, C as shown in Figure 4.4. The observed distances are $y_1 = 300.013\,\mathrm{m}$, $y_2 = 300.046\,\mathrm{m}$, $y_3 = 200.055\,\mathrm{m}$, and $y_4 = 500.152\,\mathrm{m}$. There are no correlations between the distances, and their standard deviations are defined by $\sigma = (5 + 10d)\,\mathrm{mm}$, where d is the measured distance in km. Perform a least-squares adjustment within the model of condition equations to find the adjusted distance between points A and C and its estimated variance.

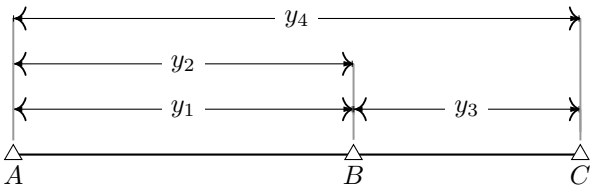

Figure 4.4: Four distances measured between three points A, B, C

5. Four angles are depicted in Figure 4.5. Angles y_1 and y_2 were derived from differencing among three observed directions. Angle y_3 was derived from an independent set of two directions. Likewise, angle y_4 was derived from yet another independent set of two directions. All directions are considered uncorrelated with standard deviation $\sigma = 10''$. The derived angles are $y_1 = 60°22'15''$, $y_2 = 75°39'20''$, $y_3 = 223°58'40''$, $y_4 = 136°01'30''$.

Note: The observed directions are uncorrelated, but some of the derived angles are not.

Use the LESS within the model of condition equations to compute the adjusted angles for y_1 and y_2. Also compute their variances.

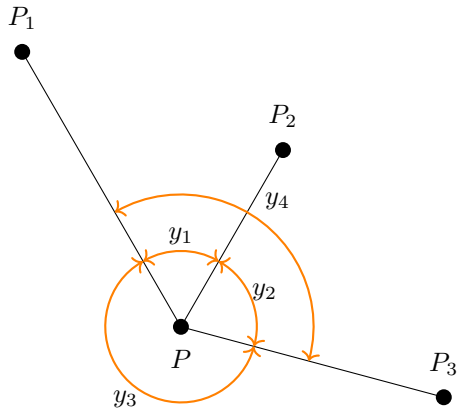

Figure 4.5: Four angles derived from three sets of directions

6. Using the data from problem 7 of Section 3.6, compute the residual vector $\tilde{\boldsymbol{e}}$ by using the LESS within the model of condition equations. Confirm that the rank of matrix B is $n-2$ and that $BA = 0$, where A is the coefficient matrix from problem 7.

   Hint: The slope between the first point and the jth point must equal the slope between the jth point and the $(j+1)$th point for $j = 2, 3, \ldots, n-1$.

4.7 Summary Formulas for the Least-Squares Solution Within the Model of Condition Equations

The model of condition equations is given by

$$\underset{r\times 1}{\boldsymbol{w}} := \underset{r\times n}{B}\,\boldsymbol{y} = B\boldsymbol{e}, \quad \boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2 P^{-1}),$$
$$r := \operatorname{rk} B.$$

Table 4.4: Summary formulas for the LESS within the model of condition equations

Quantity	Formula	Eq. No.
Model redundancy	$r = \operatorname{rk} B$	(4.1b)
Vector of predicted residuals	$\tilde{\boldsymbol{e}} = P^{-1}B^T\left(BP^{-1}B^T\right)^{-1}B\boldsymbol{y}$	(4.5e)
Dispersion matrix for residuals	$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1}$	(4.10)
Sum of squared residuals (SSR)	$\Omega = \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$	(4.8a)
Estimated variance component	$\hat{\sigma}_0^2 = \Omega/r$	(4.9)
Vector of adjusted observations	$\widehat{E\{\boldsymbol{y}\}} =: \hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}}$	(4.6)
Dispersion matrix for adjusted observations	$D\{\hat{\boldsymbol{\mu}}_y\} = \sigma_0^2 \cdot P^{-1} - D\{\tilde{\boldsymbol{e}}\}$	(4.12)

Chapter 5

The Gauss-Markov Model with Constraints

When *prior information* about the values of certain parameters, or about functional relationships between them, is known before the adjustment, those quantities can be maintained through the adjustment by application of *constraints*. For example, one may already know the height difference between two points in a leveling network that is to be adjusted, or it could be that the azimuth between two points in a 2D network to be adjusted must maintain a specified value. In both cases, the prior information can be preserved through constraints added to the Gauss-Markov Model (GMM). We say that such information is known *a priori*. The term a priori is a Latin phrase that literally means "from the earlier." In geodetic science, it refers to knowledge or information possessed before an experiment is conducted or an adjustment is computed.

One case where constraints might be useful is when the design matrix A does not have full column rank, implying that the inverse N^{-1} of the normal equation matrix does not exist, which means that the parameters of the model cannot be estimated using (3.7). This problem can occur, for example, when network observations must be adjusted in the estimation of point coordinates, but the observations themselves do not provide complete information about the network datum (i.e., its size, shape, orientation, and origin). For example, distance measurements provide information about the scale (size) of a network, and angle measurements provide information about its shape. But neither measurement type provides information about the origin or orientation of the network figure, which is necessary for estimating coordinates of network points. In such a case in 2-D, applying a constraint on two coordinates (i.e., on two parameters) and one azimuth (a function of four parameters) would provide the lacking information. In this case, the specified constraint values could be somewhat arbitrary, but we still may speak of them as being "known" (i.e., specified a priori) in the context of adjustments with constraints.

Of course, we have already seen in Section 3.5 how a minimum number of con-

straints on the unknown parameters can be imposed via datum information, thereby overcoming a datum (rank) deficiency in the model and permitting a minimally constrained adjustment of the observations. The model explored in this chapter can be used not only to handle datum deficiencies in a way that leads to a minimally constrained adjustment, it can also be used to handle a variety of fixed constraints, possibly leading to an *over-constrained adjustment*. The latter case is one in which the imposition of constraints will impact the values of the residual vector.

5.1 Model Definition and Minimization Problem

The Gauss-Markov Model (GMM) with constraints imposed on the unknown parameters (all or some of them) is written as

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \;\; \boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2 P^{-1}), \;\; \operatorname{rk} A =: q \le m, \tag{5.1a}$$

$$\underset{l\times 1}{\boldsymbol{\kappa}_0} = \underset{l\times m}{K}\,\boldsymbol{\xi}, \;\; \operatorname{rk} K =: l \ge m - q, \tag{5.1b}$$

where the rank condition

$$\operatorname{rk}\big[A^T,\, K^T\big] = m \tag{5.1c}$$

must be satisfied. The terms of the model are as defined on page 60, but now with the addition of a known $l \times m$ coefficient matrix K and an $l \times 1$ vector of specified constants $\boldsymbol{\kappa}_0$. Symbols for the normal equations were introduced in (3.4) and are repeated here for convenience:

$$\big[N,\, \boldsymbol{c}\big] \coloneqq A^T P \big[A,\, \boldsymbol{y}\big]. \tag{5.2}$$

Note that, in contrast to the model in (3.1), the coefficient matrix A in (5.1a) is not required to have full column rank, in which case the matrix inverse N^{-1} would not exist. However, the specified rank conditions imply that $(N+K^TK)^{-1}$ exists, and, if N^{-1} exists, so does $(KN^{-1}K^T)^{-1}$. This is because the range space of $[A^T,\, K^T]$ spans $\mathbb{R}^m$ as implied by the rank condition stated in (5.1c). The redundancy of the model is computed by

$$\boxed{r \coloneqq n - m + \operatorname{rk} K = n - m + l.} \tag{5.3}$$

Introducing an $l \times 1$ vector of Lagrange multipliers $\boldsymbol{\lambda}$, the Lagrange target function to minimize is

$$\Phi(\boldsymbol{\xi}, \boldsymbol{\lambda}) \coloneqq (\boldsymbol{y} - A\boldsymbol{\xi})^T P (\boldsymbol{y} - A\boldsymbol{\xi}) - 2\boldsymbol{\lambda}^T \big(\boldsymbol{\kappa}_0 - K\boldsymbol{\xi}\big) = \text{stationary} = \tag{5.4a}$$

$$= \boldsymbol{y}^T P \boldsymbol{y} - 2\boldsymbol{\xi}^T A^T P \boldsymbol{y} + \boldsymbol{\xi}^T A^T P A \boldsymbol{\xi} - 2\boldsymbol{\lambda}^T (\boldsymbol{\kappa}_0 - K\boldsymbol{\xi}). \tag{5.4b}$$

Its first partial derivatives are taken to form the following Euler-Lagrange necessary conditions:

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\xi}} = N\hat{\boldsymbol{\xi}} - \boldsymbol{c} + K^T \hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{5.5a}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = -\boldsymbol{\kappa}_0 + K\hat{\boldsymbol{\xi}} \doteq \mathbf{0}. \tag{5.5b}$$

In matrix form (5.5a) and (5.5b) are expressed as

$$\begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\kappa}_0 \end{bmatrix}, \tag{5.6}$$

where the vector on the left side contains $m + l$ unknowns to be estimated.

The sufficient condition, required for minimization, is satisfied by

$$(1/2)\big(\partial^2 \Phi / \partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T\big) = N, \tag{5.7}$$

which is positive-(semi)definite. We refer to the matrix on the left side of (5.6) as the least-squares *normal equation matrix*. It is invertible if, and only if, $\mathrm{rk}\big[A^T,\, K^T\big] = m$. This rank condition means that, for the normal equation matrix,

- among the first m columns, at least $m - l$ must be linearly independent, and
- the additional l columns are complementary, meaning that when combined with the first $m - l$ columns they span $\mathbb{R}^m$.

5.2 Estimation of Parameters and Lagrange Multipliers

In the following, we consider two cases: (1) N is invertible (nonsingular or regular), and (2) N is singular. The LEast-Squares Solution (LESS) is developed for both cases in the following:

Case 1: N is invertible, implying that matrix A has full column rank, i.e., $\mathrm{rk}\, A = m$. Equations (5.5a) and (5.5b) then imply

$$\hat{\boldsymbol{\xi}} = N^{-1}\big(\boldsymbol{c} - K^T\hat{\boldsymbol{\lambda}}\big), \tag{5.8a}$$

$$\boldsymbol{\kappa}_0 = K\hat{\boldsymbol{\xi}} = KN^{-1}\boldsymbol{c} - KN^{-1}K^T\hat{\boldsymbol{\lambda}} \tag{5.8b}$$

$$\Rightarrow \hat{\boldsymbol{\lambda}} = -\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big), \tag{5.8c}$$

finally leading to the LESS

$$\boxed{\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c} + N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big).} \tag{5.8d}$$

The vector difference $\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}$ in (5.8d) is called a *vector of discrepancies*. It shows the mismatch between the vector of specified constants $\boldsymbol{\kappa}_0$ and a linear combination (as generated by the matrix K) of the solution without constraints (i.e., $N^{-1}\boldsymbol{c}$). The estimated vectors $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\lambda}}$ may also be presented in terms of the inverse of the matrix in (5.6), viz.

$$\begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \left[\begin{array}{c|c} N^{-1} - N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}KN^{-1} & N^{-1}K^T\big(KN^{-1}K^T\big)^{-1} \\ \big(KN^{-1}K^T\big)^{-1}KN^{-1} & -\big(KN^{-1}K^T\big)^{-1} \end{array}\right] \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\kappa}_0 \end{bmatrix}. \tag{5.9}$$

Also, note that the expectation of the estimated vector of Lagrange multipliers is derived by

$$\begin{aligned} E\{\hat{\boldsymbol{\lambda}}\} &= -E\{\big(KN^{-1}K^T\big)^{-1}(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c})\} = \\ &= \big(KN^{-1}K^T\big)^{-1}\big[KN^{-1}A^TPE\{\boldsymbol{y}\} - \boldsymbol{\kappa}_0\big] = \\ &= \big(KN^{-1}K^T\big)^{-1}(K\boldsymbol{\xi} - \boldsymbol{\kappa}_0) = 0. \end{aligned} \tag{5.10}$$

Case 2: N is singular (i.e., not invertible), implying that matrix A does not have full column rank, i.e., $\operatorname{rk} A < m$.

Multiplying equation (5.5b) by K^T and adding the result to (5.5a), leads to

$$\begin{aligned} \big(N + K^TK\big)\hat{\boldsymbol{\xi}} &= \boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) \Rightarrow \\ \hat{\boldsymbol{\xi}} &= \big(N + K^TK\big)^{-1}\boldsymbol{c} + \big(N + K^TK\big)^{-1}K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big). \end{aligned} \tag{5.11}$$

Then from (5.5b) and (5.11) we have

$$\begin{aligned} \boldsymbol{\kappa}_0 = K\hat{\boldsymbol{\xi}} &= K\big(N + K^TK\big)^{-1}\boldsymbol{c} + K\big(N + K^TK\big)^{-1}K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) \Rightarrow \\ \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) &= \big[K\big(N + K^TK\big)^{-1}K^T\big]^{-1}\big[\boldsymbol{\kappa}_0 - K\big(N + K^TK\big)^{-1}\boldsymbol{c}\big]. \end{aligned} \tag{5.12}$$

Substituting (5.12) into (5.11) leads to the LESS

$$\boxed{\begin{aligned} \hat{\boldsymbol{\xi}} = &\big(N + K^TK\big)^{-1}\boldsymbol{c} + \big(N + K^TK\big)^{-1}K^T\cdot \\ &\cdot\big[K\big(N + K^TK\big)^{-1}K^T\big]^{-1}\big[\boldsymbol{\kappa}_0 - K\big(N + K^TK\big)^{-1}\boldsymbol{c}\big]. \end{aligned}} \tag{5.13}$$

The form of (5.13) is identical to (5.8d) except that all occurrences of matrix N in (5.8d) have been replaced by $N + K^TK$ in (5.13). Of course, (5.13) can be used for both the singular and nonsingular cases.

Also, note that the expectation of vector difference $\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}$ is derived by

$$\begin{aligned} E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\} &= E\{\big[K\big(N + K^TK\big)^{-1}K^T\big]^{-1}\big[\boldsymbol{\kappa}_0 - K\big(N + K^TK\big)^{-1}\boldsymbol{c}\big]\} = \\ &= \big[K\big(N + K^TK\big)^{-1}K^T\big]^{-1}\big[\boldsymbol{\kappa}_0 - K\big(N + K^TK\big)^{-1}A^TPE\{\boldsymbol{y}\}\big] = \\ &= \big[K\big(N + K^TK\big)^{-1}K^T\big]^{-1}K\big[I_m - \big(N + K^TK\big)^{-1}N\big]\boldsymbol{\xi} = \\ &= \big[K\big(N + K^TK\big)^{-1}K^T\big]^{-1}K\big[I_m - \big(N + K^TK\big)^{-1}\big(N + K^TK\big)\big]\boldsymbol{\xi} + \\ &\quad + \big[K\big(N + K^TK\big)^{-1}K^T\big]^{-1}K\big(N + K^TK\big)^{-1}K^T\cdot K\boldsymbol{\xi} = K\boldsymbol{\xi} \\ &\Rightarrow E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\} = K\boldsymbol{\xi} \;\text{ or }\; E\{\hat{\boldsymbol{\lambda}}\} = \boldsymbol{\kappa}_0 - K\boldsymbol{\xi} \Rightarrow \end{aligned} \tag{5.14a}$$

$$E\{\hat{\boldsymbol{\lambda}}\} = \mathbf{0}. \tag{5.14b}$$

5.3 Derivation of Dispersion Matrices

We now compute the formal dispersion matrices for the both the vector of estimated parameters $\hat{\boldsymbol{\xi}}$ and the vector of estimated Lagrange multipliers $\hat{\boldsymbol{\lambda}}$.

Case 1: For case 1, we start with (5.6), from which we have

$$\begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\kappa}_0 \end{bmatrix}. \tag{5.15}$$

Applying the law of covariance propagation, noting that $\boldsymbol{\kappa}_0$ is a non-random vector, and substituting the matrix from (5.9) implies that

$$\begin{aligned} D\{\begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix}\} &= \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} \cdot D\{\begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\kappa}_0 \end{bmatrix}\} \cdot \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} = \\ &= \sigma_0^2 \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} N & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} = \\ &= \sigma_0^2 \begin{bmatrix} N^{-1} - N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}KN^{-1} & 0 \\ 0 & \big(KN^{-1}K^T\big)^{-1} \end{bmatrix}, \end{aligned} \tag{5.16}$$

which, upon comparing to (5.9), reveals the relation

$$\begin{bmatrix} D\{\hat{\boldsymbol{\xi}}\} & X \\ X^T & -D\{\hat{\boldsymbol{\lambda}}\} \end{bmatrix} = \sigma_0^2 \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1}. \tag{5.17}$$

Here the symbol X represents a term of no particular interest. Note that $X \neq C\{\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}\} = 0$.

Case 2: The results for case 2 are slightly different, because we work with a system of equations involving $N + K^T K$ rather than N itself. Thus, rather than the system of equations shown in (5.15), we work with the modified system

$$\begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} N + K^T K & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{c} + K^T\boldsymbol{\kappa}_0 \\ \boldsymbol{\kappa}_0 \end{bmatrix}. \tag{5.18}$$

Note that the matrix in (5.15) has full rank even when matrix N is singular, so it is not necessary to use the modified system (5.18). However, this modified system has its own benefits, and it is consistent with equation (5.13) derived above.

Using the formulas for inverting a partitioned matrix (see (A.10) and (A.11)) and introducing the notation $N_K \coloneqq (N + K^T K)$ for the sake of brevity, we can write

$$\begin{bmatrix} N_K & K^T \\ K & 0 \end{bmatrix}^{-1} = \begin{bmatrix} N_K^{-1} - N_K^{-1}K^T\big(KN_K^{-1}K^T\big)^{-1}KN_K^{-1} & N_K^{-1}K^T\big(KN_K^{-1}K^T\big)^{-1} \\ \big(KN_K^{-1}K^T\big)^{-1}KN_K^{-1} & -\big(KN_K^{-1}K^T\big)^{-1} \end{bmatrix}. \tag{5.19}$$

Our goal is to express the inverted matrix in (5.19) in terms of the inverted matrix in (5.15). We start by multiplying the former by the inverse of the latter and then carry out some matrix algebra.

$$\begin{bmatrix} N+K^TK & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix} = \begin{bmatrix} N+K^TK & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} N+K^TK-K^TK & K^T \\ K & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} I_m & 0 \\ 0 & I_l \end{bmatrix} - \begin{bmatrix} N+K^TK & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} K^TK & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ 0 & I_l \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ K & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ K & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & I_l \end{bmatrix} \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix} =$$

$$= \left(\begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} - \begin{bmatrix} 0 & 0 \\ 0 & I_l \end{bmatrix} \right) \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}. \quad (5.20)$$

Multiplying the first and last products of (5.20) by the inverse of their last terms and considering (5.17) reveals that

$$\begin{bmatrix} N+K^TK & K^T \\ K & 0 \end{bmatrix}^{-1} = \begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix}^{-1} - \begin{bmatrix} 0 & 0 \\ 0 & I_l \end{bmatrix} = \begin{bmatrix} \sigma_0^{-2} D\{\hat{\boldsymbol{\xi}}\} & X \\ X^T & -\sigma_0^{-2} D\{\hat{\boldsymbol{\lambda}}\} - I_l \end{bmatrix}, \quad (5.21)$$

and therefore, in consideration of (5.19),

$$-\big[K\big(N+K^TK\big)^{-1}K^T\big]^{-1} = -\sigma_0^{-2} D\{\hat{\boldsymbol{\lambda}}\} - I_l \Rightarrow$$
$$D\{\hat{\boldsymbol{\lambda}}\} = \sigma_0^2 \big\{ \big[K\big(N+K^TK\big)^{-1}K^T\big]^{-1} - I_l \big\}. \quad (5.22)$$

Alternative derivation of dispersion matrix The following alternative approach to deriving the dispersion matrix for case 2 was recognized by Dru Smith and Kyle Snow during collaborative work, where, again, the abbreviation $N_K := (N + K^TK)$ is used: The law of linear covariance propagation (law of error propagation) allows us to write

$$D\{\begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}} \end{bmatrix}\} = \begin{bmatrix} N_K & K^T \\ K & 0 \end{bmatrix}^{-1} \cdot D\{\begin{bmatrix} \boldsymbol{c} + K^T\boldsymbol{\kappa}_0 \\ \boldsymbol{\kappa}_0 \end{bmatrix}\} \cdot \begin{bmatrix} N_K & K^T \\ K & 0 \end{bmatrix}^{-1} =$$

$$= \sigma_0^2 \begin{bmatrix} N_K & K^T \\ K & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} N & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} N_K & K^T \\ K & 0 \end{bmatrix}^{-1} =: \sigma_0^2 \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix}, \quad (5.23a)$$

where, upon substitution of (5.19), the block matrices Q_{11}, Q_{12}, and Q_{22} turn out to be

$$Q_{22} = \big(KN_K^{-1}K^T\big)^{-1} K \big(N_K^{-1} \cdot N \cdot N_K^{-1}\big) K^T \big(KN_K^{-1}K^T\big)^{-1}, \quad (5.23b)$$
$$Q_{12} = \big(N_K^{-1} \cdot N \cdot N_K^{-1}\big) K^T \big(KN_K^{-1}K^T\big)^{-1} - N_K^{-1}K^TQ_{22}, \quad (5.23c)$$

$$Q_{11} = \left[N_K^{-1} - N_K^{-1}K^T\left(KN_K^{-1}K^T\right)^{-1}KN_K^{-1}\right]\cdot NN_K^{-1} - Q_{12}KN_K^{-1}. \quad (5.23d)$$

Now, we wish to reduce these matrices to simpler forms, for which the following relationship is useful

$$N_K^{-1}\cdot N\cdot N_K^{-1} = N_K^{-1} - N_K^{-1}(K^TK)N_K^{-1}. \quad (5.24)$$

Substituting (5.24) into (5.23b) leads to

$$Q_{22} = \left(KN_K^{-1}K^T\right)^{-1}K\left[N_K^{-1} - N_K^{-1}(K^TK)N_K^{-1}\right]K^T\left(KN_K^{-1}K^T\right)^{-1} = \\ = \left(KN_K^{-1}K^T\right)^{-1} - I_l = \sigma_0^{-2}D\{\hat{\boldsymbol{\lambda}}\}. \quad (5.25a)$$

Then, substituting (5.25a) into (5.23c) results in

$$Q_{12} = \left(N_K^{-1}\cdot N\cdot N_K^{-1}\right)K^T\left(KN_K^{-1}K^T\right)^{-1} - N_K^{-1}K^T\left[\left(KN_K^{-1}K^T\right)^{-1} - I_l\right] = \\ = \left[N_K^{-1} - N_K^{-1}(K^TK)N_K^{-1}\right]K^T\left(KN_K^{-1}K^T\right)^{-1} - N_K^{-1}K^T\left(KN_K^{-1}K^T\right)^{-1} + N_K^{-1}K^T = \\ = 0 = C\{\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}\}, \quad (5.25b)$$

and, therefore, (5.23d) reduces to

$$Q_{11} = \left[N_K^{-1} - N_K^{-1}K^T\left(KN_K^{-1}K^T\right)^{-1}KN_K^{-1}\right]\cdot NN_K^{-1} = \\ = N_K^{-1} - N_K^{-1}(K^TK)N_K^{-1} - N_K^{-1}K^T(KN_K^{-1}K^T)^{-1}K\left[N_K^{-1} - N_K^{-1}(K^TK)N_K^{-1}\right] = \\ = N_K^{-1} - N_K^{-1}K^T\left(KN_K^{-1}K^T\right)^{-1}KN_K^{-1} = \sigma_0^{-2}D\{\hat{\boldsymbol{\xi}}\}. \quad (5.25c)$$

Summary of dispersion matrices For convenience, we summarize the dispersion matrices of the estimated parameters and the estimated Lagrange multipliers for both cases 1 and 2 as follows:

Case 1 (N nonsingular):

$$\boxed{\begin{aligned} D\{\hat{\boldsymbol{\xi}}\} &= \sigma_0^2\left[N^{-1} - N^{-1}K^T\left(KN^{-1}K^T\right)^{-1}KN^{-1}\right] \quad (5.26a)\\ D\{\hat{\boldsymbol{\lambda}}\} &= \sigma_0^2\left(KN^{-1}K^T\right)^{-1} \quad (5.26b)\end{aligned}}$$

Case 2 (N singular):

$$\boxed{\begin{aligned} D\{\hat{\boldsymbol{\xi}}\} &= \sigma_0^2\left(N + K^TK\right)^{-1} - \sigma_0^2\left(N + K^TK\right)^{-1}K^T\cdot \\ &\quad\cdot\left[K\left(N + K^TK\right)^{-1}K^T\right]^{-1}K\left(N + K^TK\right)^{-1} \quad (5.27a)\\ D\{\hat{\boldsymbol{\lambda}}\} &= \sigma_0^2\left\{\left[K\left(N + K^TK\right)^{-1}K^T\right]^{-1} - I_l\right\} \quad (5.27b)\end{aligned}}$$

Cases 1 and 2:

$$\boxed{C\{\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}\} = 0} \quad (5.28)$$

As with the parameter estimates, the dispersion matrices for both cases 1 and 2 have a similar form, with every occurrence of N in case 1 being replaced by $N+K^TK$ in case 2, the exception being the identity matrix I_l appearing in case 2. Also note that the dispersion matrices in (5.26a) and (5.27a) are nothing more than the coefficient matrices multiplying the vector $\boldsymbol{c}$ in (5.8d) and (5.13), respectively, multiplied by the (unknown) variance component σ_0^2. Finally, it is clear from the above that the constraints reduce the dispersion matrix of $\hat{\boldsymbol{\xi}}$ compared to the corresponding dispersion matrix within the GMM (without constraints) derived in Chapter 3 (cf. (3.13)).

5.4 Residuals and Adjusted Observations

For both cases 1 and 2, the residual vector $\tilde{\boldsymbol{e}}$ and vector of adjusted observations $\hat{\boldsymbol{\mu}}_y$ may be obtained in a straightforward way after the estimation of the parameters by use of the formulas

$$\boxed{\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}},} \tag{5.29}$$

and

$$\boxed{\widehat{E\{\boldsymbol{y}\}} = \hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}}.} \tag{5.30}$$

Here, $\hat{\boldsymbol{\mu}}_y$ is also interpreted as an estimate of the true, and thus unknown, vector of observables $\boldsymbol{\mu}_y$, where $E\{\boldsymbol{y}\} = \boldsymbol{\mu}_y$.

The dispersion matrix for the residual vector $\tilde{\boldsymbol{e}}$ can be derived from application of the law of covariance propagation as follows: Since

$$D\{\tilde{\boldsymbol{e}}\} = D\{\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\} = D\{\boldsymbol{y}\} + AD\{\hat{\boldsymbol{\xi}}\}A^T - 2C\{\boldsymbol{y}, A\hat{\boldsymbol{\xi}}\}, \tag{5.31}$$

we start by deriving the covariance matrix $C\{\boldsymbol{y}, A\hat{\boldsymbol{\xi}}\}$. For case 1 we have

$$C\{\boldsymbol{y}, A\hat{\boldsymbol{\xi}}\} = I_n \cdot D\{\boldsymbol{y}\} \cdot \big\{A\big[N^{-1}A^TP - N^{-1}K^T(KN^{-1}K^T)^{-1}KN^{-1}A^TP\big]\big\}^T = \tag{5.32a}$$

$$= \sigma_0^2 P^{-1}\big[PAN^{-1}A^T - PAN^{-1}K^T(KN^{-1}K^T)^{-1}KN^{-1}A^T\big] = \tag{5.32b}$$

$$= \sigma_0^2 A\big[N^{-1} - N^{-1}K^T(KN^{-1}K^T)^{-1}KN^{-1}\big]A^T = \tag{5.32c}$$

$$= AD\{\hat{\boldsymbol{\xi}}\}A^T = D\{A\hat{\boldsymbol{\xi}}\} = C\{\boldsymbol{y}, A\hat{\boldsymbol{\xi}}\}. \tag{5.32d}$$

Then, by substituting (5.32d) into (5.31), we arrive at

$$D\{\tilde{\boldsymbol{e}}\} = D\{\boldsymbol{y}\} - AD\{\hat{\boldsymbol{\xi}}\}A^T \Rightarrow \tag{5.33a}$$

$$\boxed{D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot \big\{P^{-1} - A\big[N^{-1} - N^{-1}K^T(KN^{-1}K^T)^{-1}KN^{-1}\big]A^T\big\}} \tag{5.33b}$$

and

$$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot \big[P^{-1} - AN^{-1}A^T + AN^{-1}K^T(KN^{-1}K^T)^{-1}KN^{-1}A^T\big]. \tag{5.33c}$$

Note that (5.33c) reveals that the dispersion matrix for the residuals within the GMM with constraints is larger than that for the GMM without constraints (cf. (3.14a)). For case 2, one only needs to replace the matrix N^{-1} with $(N + K^T K)^{-1}$ in formulas (5.32) and (5.33).

Obviously, the dispersion matrix for the adjusted observations is written as

$$\boxed{D\{\hat{\boldsymbol{\mu}}_y\} = D\{\boldsymbol{y} - \tilde{\boldsymbol{e}}\} = D\{A\hat{\boldsymbol{\xi}}\} = AD\{\hat{\boldsymbol{\xi}}\}A^T.} \tag{5.34}$$

5.4.1 A Numerical Example

A simple numerical example can be used to verify several of the equations derived above. We borrow our example from Smith et al. (2018), which is a small leveling network depicted in Figure 5.1.

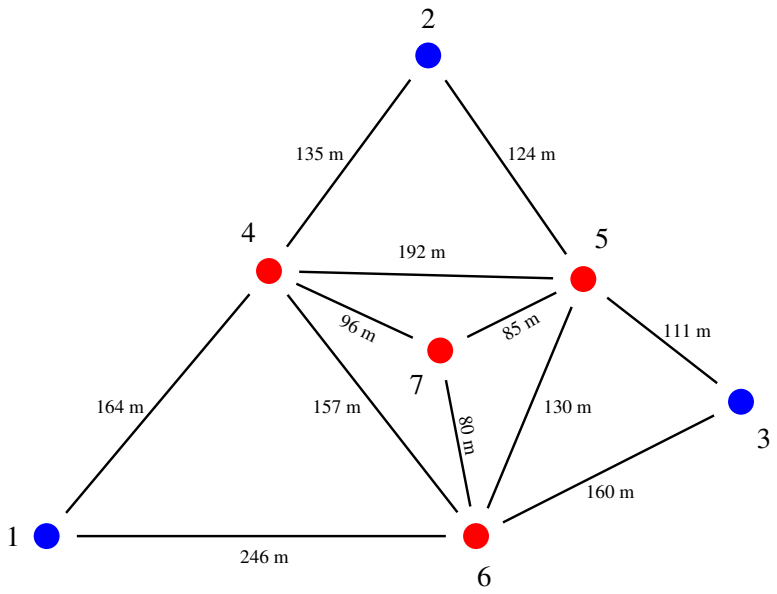

Figure 5.1: Small leveling network copied from Smith et al. (2018)

The matrices of interest are shown below. The unknown parameters (heights of stations) are ordered according to the seven numbered stations in Figure 5.1. The connection between the observations and the stations is reflected in the coefficient matrix A.

$$\underset{12\times 7}{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad \underset{12\times 1}{\boldsymbol{y}} = \begin{bmatrix} 0.333557 \\ 0.365859 \\ 2.850824 \\ -0.948661 \\ -1.040570 \\ -0.824317 \\ -1.989007 \\ -0.528043 \\ 2.517497 \\ -1.692892 \\ -0.296337 \\ -0.162582 \end{bmatrix} \text{m},$$

$$\underset{3\times 7}{K} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \underset{3\times 1}{\boldsymbol{\kappa}_0} = \begin{bmatrix} 68.8569 \\ 66.9471 \\ 68.1559 \end{bmatrix} \text{m},$$

$$P^{-1} = \text{diag}(\big[2.214,\ 1.440,\ 1.476,\ 1.215,\ 1.116,\ 0.720,\ 1.728,\ 1.170,\ 1.413,\ 0.864,\ 0.765,\ 0.999\big])\cdot 10^{-6}\,\text{m}^2$$

Since the matrix A does not have full column rank, the problem belongs to case 2.

5.5 Estimated Variance Component

The estimated variance component for the GMM with constraints is derived similar to that for the GMM without constraints as shown in Section 3.3. The estimation is based on the principle

$$\frac{\hat{\sigma}_0^2}{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}} = \frac{\sigma_0^2}{E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}\}}, \tag{5.35}$$

with the assumption $E\{\hat{\sigma}_0^2\} = \sigma_0^2$. Furthermore, for the purpose of validating the constraints via hypothesis testing, we wish to decompose the quadratic form $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$ into the sum of two quadratic forms, viz. $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = \Omega + R$, where Ω is the sum of squared residuals (SSR) associated with the LESS within the *GMM without constraints*. In the following, we derive these components for both cases 1 and 2.

5.5.1 Case 1 — Matrix N is invertible $\Rightarrow \text{rk}\, A = m$

$$\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = \big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\big)^T P \big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\big) =$$

$$
\begin{aligned}
&= \big[(\boldsymbol{y} - AN^{-1}\boldsymbol{c}) - AN^{-1}K^T\big(KN^{-1}K^T\big)^{-1}(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c})\big]^T P\cdot \\
&\qquad \cdot \big[(\boldsymbol{y} - AN^{-1}\boldsymbol{c}) - AN^{-1}K^T\big(KN^{-1}K^T\big)^{-1}(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c})\big] = \\
&= \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) - \\
&\qquad - \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T PAN^{-1}K^T\big(KN^{-1}K^T\big)^{-1}(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}) - \\
&\qquad - (\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c})^T\big(KN^{-1}K^T\big)^{-1}KN^{-1}A^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) + \\
&+ (\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c})^T\big(KN^{-1}K^T\big)^{-1}KN^{-1}\big(A^T PA\big)N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\cdot \\
&\qquad\qquad \cdot (\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}) = \\
&\qquad \text{(Note that } A^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) = 0.\text{)} \\
&= \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) + (\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c})^T\big(KN^{-1}K^T\big)^{-1}\cdot \\
&\qquad\qquad \cdot (\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}) = \\
&= \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) + \hat{\boldsymbol{\lambda}}^T\big(KN^{-1}K^T\big)\hat{\boldsymbol{\lambda}} = \Omega + R
\end{aligned}
\tag{5.36}
$$

The scalars Ω and R defined as

$$
\Omega \coloneqq \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) \tag{5.37a}
$$

and

$$
R \coloneqq (\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c})^T\big(KN^{-1}K^T\big)^{-1}(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}), \tag{5.37b}
$$

respectively.

Thus we have decomposed the quadratic form $\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$ into components Ω and R. Obviously, both Ω and R are random numbers since they are both functions of the random vector $\boldsymbol{c}$. It turns out that they are also uncorrelated. The random variable Ω is associated with the LESS within the GMM without constraints, whereas R is due to the addition of the constraints $\boldsymbol{\kappa}_0 = K\boldsymbol{\xi}$. From (5.37b) we see that R is always positive, revealing that the inclusion of constraints increases the value of $\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$. The random variables Ω and R are used for hypothesis testing as discussed in Chapter 9.

We now derive the expectation of $\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$.

$$
\begin{aligned}
E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} &= E\{\Omega\} + E\{R\} = \\
&= (n-m)\sigma_0^2 + E\{\hat{\boldsymbol{\lambda}}^T\big(KN^{-1}K^T\big)\hat{\boldsymbol{\lambda}}\} = \qquad \text{using (3.27) for } E\{\Omega\} \\
&= (n-m)\sigma_0^2 + \operatorname{tr}\big[\big(KN^{-1}K^T\big)E\{\hat{\boldsymbol{\lambda}}\ \hat{\boldsymbol{\lambda}}^T\}\big] = \\
&= (n-m)\sigma_0^2 + \operatorname{tr}\big[\big(KN^{-1}K^T\big)\big(D\{\hat{\boldsymbol{\lambda}}\} + E\{\hat{\boldsymbol{\lambda}}\}E\{\hat{\boldsymbol{\lambda}}\}^T\big)\big] = \\
&\qquad \text{(with } E\{\hat{\boldsymbol{\lambda}}\} = \mathbf{0} \text{ and } D\{\hat{\boldsymbol{\lambda}}\} = \sigma_0^2\big(KN^{-1}K^T\big)^{-1}\text{)} \\
&= (n-m)\sigma_0^2 + \operatorname{tr}\big[\big(KN^{-1}K^T\big)\sigma_0^2\big(KN^{-1}K^T\big)^{-1}\big] = \\
&= (n-m+l)\sigma_0^2
\end{aligned}
\tag{5.38}
$$

Substitution of (5.36) and (5.38) into (5.35) yields the following formula for the estimated variance component:

$$\boxed{\begin{aligned}\hat{\sigma}_0^2 = {} & \frac{\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)}{n-m+l} + \\ & + \frac{\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)}{n-m+l}.\end{aligned}} \tag{5.39}$$

Other useful forms of $\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$ are derived below starting with (5.36).

$$\begin{aligned}\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} &= \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) + \hat{\boldsymbol{\lambda}}^T\big(KN^{-1}K^T\big)\hat{\boldsymbol{\lambda}} = \\ &= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T N^{-1}\boldsymbol{c} - \big(\boldsymbol{\kappa}_0^T - \boldsymbol{c}^T N^{-1}K^T\big)\hat{\boldsymbol{\lambda}} = && \text{using (5.8c)} \\ &= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T N^{-1}\big(\boldsymbol{c} - K^T\hat{\boldsymbol{\lambda}}\big) - \boldsymbol{\kappa}_0^T\hat{\boldsymbol{\lambda}} = && \text{using (5.8a)} \\ &= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T\hat{\boldsymbol{\xi}} - \boldsymbol{\kappa}_0^T\hat{\boldsymbol{\lambda}} = \\ &= \boldsymbol{y}^T P\big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\big) - \boldsymbol{\kappa}_0^T\hat{\boldsymbol{\lambda}} = \\ &= \boldsymbol{y}^T P\tilde{\boldsymbol{e}} - \boldsymbol{\kappa}_0^T\hat{\boldsymbol{\lambda}} && (5.40)\end{aligned}$$

5.5.2 Case 2 — Matrix N is singular $\Rightarrow \operatorname{rk} A < m$

$$\begin{aligned}&\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} = \\ &= \big\{\boldsymbol{y} - A\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]\big\}^T P\big\{\boldsymbol{y} - A\big(N + K^T K\big)^{-1}\cdot \\ &\qquad\qquad \cdot\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]\big\} = \\ &= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T PA\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] - \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T\cdot \\ &\qquad \cdot\big(N + K^T K\big)^{-1}A^T P\boldsymbol{y} + \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T\big(N + K^T K\big)^{-1}\cdot \\ &\qquad\qquad \cdot\big(A^T PA + K^T K - K^T K\big)\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] = \\ &= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] - \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T\cdot \\ &\quad \cdot\big(N + K^T K\big)^{-1}\boldsymbol{c} + \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] - \\ &\quad - \big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]^T\big(N + K^T K\big)^{-1}K^T K\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big] = \\ &= \boldsymbol{y}^T P\boldsymbol{y} - \cancel{\boldsymbol{c}^T\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]} - \\ &\qquad - \cancel{\boldsymbol{c}^T\big(N + K^T K\big)^{-1}\boldsymbol{c}} - \cancel{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T K\big(N + K^T K\big)^{-1}\boldsymbol{c}} + \\ &+ \cancel{\boldsymbol{c}^T\big(N + K^T K\big)^{-1}\big[\boldsymbol{c} + K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big]} + \cancel{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T K\big(N + K^T K\big)^{-1}\boldsymbol{c}} + \\ &\qquad + \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T K\big(N + K^T K\big)^{-1}K^T\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) - \hat{\boldsymbol{\xi}}^T K^T K\hat{\boldsymbol{\xi}} = \\ &= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T\big(N + K^T K\big)^{-1}\boldsymbol{c} + \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T\big[K\big(N + K^T K\big)^{-1}K^T\big]\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) - \\ &\qquad\qquad - \boldsymbol{\kappa}_0^T\boldsymbol{\kappa}_0 \qquad (5.41)\end{aligned}$$

Now we compute the expectation for $\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$.

$$
\begin{aligned}
&E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} = \\
&= E\{\boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T\big(N + K^T K\big)^{-1}\boldsymbol{c}\big] + \\
&\quad + \big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T\big[K\big(N + K^T K\big)^{-1}K^T\big]\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) - \boldsymbol{\kappa}_0^T\boldsymbol{\kappa}_0\} = \\
&= E\{\boldsymbol{y}^T P\big[\boldsymbol{y} - A\big(N + K^T K\big)^{-1}\boldsymbol{c}\big]\big\} + E\{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T\big[K\big(N + K^T K\big)^{-1}K^T\big]\cdot \\
&\qquad\qquad \cdot\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big\} - E\{\boldsymbol{\kappa}_0^T\boldsymbol{\kappa}_0\} = \\
&= \operatorname{tr} P\big[I_n - A\big(N + K^T K\big)^{-1}A^T P\big]E\{\boldsymbol{y}\boldsymbol{y}^T\} + \operatorname{tr}\big[K\big(N + K^T K\big)^{-1}K^T\big]\cdot \\
&\qquad\qquad \cdot E\{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T\} - \operatorname{tr} E\{\boldsymbol{\kappa}_0^T\boldsymbol{\kappa}_0\} =
\end{aligned}
$$

(Note that $E\{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T\} = D\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\} + E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\}E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\}^T$ and $D\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\} = D\{\hat{\boldsymbol{\lambda}}\} = \sigma_0^2\big\{\big[K(N + K^T K)^{-1}K\big]^{-1} - I_l\big\}$, and $E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\} = K\boldsymbol{\xi}$, and $E\{\boldsymbol{y}\boldsymbol{y}^T\} = D\{\boldsymbol{y}\} + E\{\boldsymbol{y}\}E\{\boldsymbol{y}\}^T = \sigma_0^2 P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T$).

$$
\begin{aligned}
&= \operatorname{tr} P\big[I_n - A\big(N + K^T K\big)^{-1}A^T P\big]\big(\sigma_0^2 P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T\big) + \\
&\quad + \operatorname{tr}\big[K\big(N + K^T K\big)^{-1}K^T\big]\big[D\{\hat{\boldsymbol{\lambda}}\} + E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\}E\{\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\}^T\big] - \operatorname{tr} K\boldsymbol{\xi}\boldsymbol{\xi}^T K^T = \\
&= \operatorname{tr}\big[\sigma_0^2 I_n + PA\boldsymbol{\xi}\boldsymbol{\xi}^T A^T - PA\big(N + K^T K\big)^{-1}A^T\sigma_0^2 - PA\big(N + K^T K\big)^{-1}A^T PA\cdot \\
&\quad \cdot \boldsymbol{\xi}\boldsymbol{\xi}^T A^T\big] + \operatorname{tr}\big[K\big(N + K^T K\big)^{-1}K^T\big]\big(\big\{\big[K\big(N + K^T K\big)^{-1}K^T\big]^{-1} - I_l\big\}\sigma_0^2 + \\
&\qquad\qquad + K\boldsymbol{\xi}\boldsymbol{\xi}^T K^T\big) - \operatorname{tr} K\boldsymbol{\xi}\boldsymbol{\xi}^T K^T = \\
&= \sigma_0^2 n + \operatorname{tr}\boldsymbol{\xi}\boldsymbol{\xi}^T N - \sigma_0^2\operatorname{tr}\big(N + K^T K\big)^{-1}N - \operatorname{tr}\big(N + K^T K\big)^{-1}N\boldsymbol{\xi}\boldsymbol{\xi}^T N + \sigma_0^2 l - \\
&\qquad - \sigma_0^2\operatorname{tr}\big[K\big(N + K^T K\big)^{-1}K^T\big] + \operatorname{tr}\big[\big(N + K^T K\big)^{-1}K^T K\boldsymbol{\xi}\boldsymbol{\xi}^T K^T K\big] - \\
&\qquad\qquad - \operatorname{tr} K\boldsymbol{\xi}\boldsymbol{\xi}^T K^T = \\
&= \sigma_0^2 n - \sigma_0^2\operatorname{tr}\big(N + K^T K\big)^{-1}\big(N + K^T K\big) + \sigma_0^2 l + \\
&\quad + \operatorname{tr}\big[I_m - \big(N + K^T K\big)^{-1}N\big]\boldsymbol{\xi}\boldsymbol{\xi}^T N - \operatorname{tr}\big[I_m - \big(N + K^T K\big)^{-1}K^T K\big]\boldsymbol{\xi}\boldsymbol{\xi}^T K^T K = \\
&= \sigma_0^2(n - m + l) + \operatorname{tr}\big(N + K^T K\big)^{-1}K^T K\boldsymbol{\xi}\boldsymbol{\xi}^T N - \operatorname{tr}\big(N + K^T K\big)^{-1}N\boldsymbol{\xi}\boldsymbol{\xi}^T K^T K = \\
&= \sigma_0^2(n - m + l) + \operatorname{tr}\big[\big(N + K^T K\big)^{-1}K^T K\boldsymbol{\xi}\boldsymbol{\xi}^T N\big]^T - \\
&\qquad - \operatorname{tr} N\boldsymbol{\xi}\boldsymbol{\xi}^T K^T K\big(N + K^T K\big)^{-1} = \sigma_0^2(n - m + l)
\end{aligned}
$$

$$
\Rightarrow \boxed{E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}\} = \sigma_0^2(n - m + l)} \tag{5.42}
$$

Finally, substituting (5.41) and (5.42) into (5.35) yields

$$
\boxed{\begin{aligned}
\hat{\sigma}_0^2 = &\frac{\boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T\big(N + K^T K\big)^{-1}\boldsymbol{c}}{(n - m + l)} + \\
&+ \frac{\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big)^T\big[K\big(N + K^T K\big)^{-1}K^T\big]\big(\boldsymbol{\kappa}_0 - \hat{\boldsymbol{\lambda}}\big) - \boldsymbol{\kappa}_0^T\boldsymbol{\kappa}_0}{(n - m + l)},
\end{aligned}} \tag{5.43a}
$$

or, by use of (5.12) and with $N_K \coloneqq N + K^T K$ for compactness,

$$\boxed{\hat{\sigma}_0^2 = \frac{\boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T N_K^{-1} \boldsymbol{c} + \big(\boldsymbol{\kappa}_0 - K N_K^{-1} \boldsymbol{c}\big)^T \big(K N_K^{-1} K^T\big)\big(\boldsymbol{\kappa}_0 - K N_K^{-1} \boldsymbol{c}\big) - \boldsymbol{\kappa}_0^T \boldsymbol{\kappa}_0}{(n - m + l)},} \tag{5.43b}$$

or

$$\boxed{\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{(n - m + l)}.} \tag{5.43c}$$

We cannot directly identify Ω and R in (5.41) as we could in case 1. Therefore, we define Ω as

$$\Omega = \big(\boldsymbol{y} - A N^- \boldsymbol{c}\big)^T P \big(\boldsymbol{y} - A N^- \boldsymbol{c}\big), \tag{5.44}$$

and R as

$$R = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} - \Omega, \tag{5.45}$$

where $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$ is given in (5.41). The symbol N^- in (5.44) stands for a generalized inverse of the matrix N. While generalized inverses are beyond the scope of these notes, the following generalized inverse is shown so that readers unfamiliar with the topic can still make use of equation (5.44). First, assume that the matrix N and vector $\boldsymbol{c}$ have been partitioned as follows:

$$\underset{m \times m}{N} = \left[\begin{array}{c|c} \underset{q \times q}{N_{11}} & N_{12} \\ \hline N_{21} & N_{22} \end{array}\right] \quad \text{and} \quad \underset{m \times 1}{\boldsymbol{c}} = \left[\begin{array}{c} \underset{q \times 1}{\boldsymbol{c}_1} \\ \hline \boldsymbol{c}_2 \end{array}\right], \tag{5.46}$$

where the $q \times q$ submatrix N_{11} has full rank q, i.e., $\operatorname{rk} N_{11} = q \coloneqq \operatorname{rk} N$. Note that such a partitioning can always be formed, even if the parameters in $\boldsymbol{\xi}$ must be reordered to do so. Then, the $m \times m$ matrix $G \coloneqq \left[\begin{smallmatrix} N_{11}^{-1} & 0 \\ 0 & 0 \end{smallmatrix}\right]$ is a generalized inverse of N and thus can be used in (5.44) for N^-, which simplifies that equation to

$$\Omega = \boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}_1^T N_{11}^{-1} \boldsymbol{c}_1 \;\text{ if }\; \operatorname{rk} N_{11} = \operatorname{rk} N. \tag{5.47}$$

5.6 Hypothesis Test Using the Estimated Variance Component

The following ratio is formed for both cases 1 and 2 for the purposes of hypothesis testing (see Chapter 9 for more details on hypothesis testing):

$$T \coloneqq \frac{R/(l - m + q)}{\Omega/(n - q)} \sim F(l - m + q, n - q), \;\text{ with } q \coloneqq \operatorname{rk}(A). \tag{5.48}$$

The ratio T is called a *Test statistic*. It is assumed to have an F-distribution with $l-m+q$ and $n-q$ degrees of freedom.[1] The hypothesis test is then stated as

$$H_0 : K\boldsymbol{\xi} = \boldsymbol{\kappa}_0 \text{ versus } H_A : K\boldsymbol{\xi} \neq \boldsymbol{\kappa}_0, \tag{5.49}$$

where H_0 is called the *null hypothesis* and H_A is called the *alternative hypothesis*. For some chosen significance level α,

$$\begin{aligned} &\text{Accept } H_0 : \text{ if } T \leq F_{\alpha,l-m+q,n-q} \\ &\text{Reject } H_0 : \text{ if } T > F_{\alpha,l-m+q,n-q}, \end{aligned}$$

where $F_{\alpha,l-m+q,n-q}$ is taken from a table of *critical values* for the F-distribution. The critical values for certain values of r_1 and r_2 are listed in Appendix C. If MATLAB is available, the critical value may be generated by use of the MATLAB command `finv(`$1-\alpha$`,`r_1`,`r_2`)`.

Note that the redundancy $r_2 \coloneqq n-q$ represents the degrees of freedom for the system of equations if no constraints were applied, whereas the redundancy $r_1 \coloneqq l-m+q$ represents the *increase* in degrees of freedom due to the constraints, i.e.

$$r = r_1 + r_2 = (l-m+q) + (n-q) = n-m+l, \tag{5.50}$$

which agrees with (5.3). In the case that matrix A has full column rank (i.e., $\operatorname{rk} A = q = m$), then the redundancies reduce to $r_1 \coloneqq l$ and $r_2 \coloneqq n-m$, respectively.

5.7 Practice Problems

1. Derive the expectation of the vector of estimated parameters $\hat{\boldsymbol{\xi}}$ given in (5.8d). Is $\hat{\boldsymbol{\xi}}$ an unbiased estimator of the vector of unknown parameters $\boldsymbol{\xi}$?

2. Solve the following problems for the data given in Section 5.4.1:

   (a) Confirm that $N = A^T P A$ is rank deficient and that the rank condition (5.1c) is satisfied.

   (b) Compute the vector of estimated parameters $\hat{\boldsymbol{\xi}}$ by (5.13) and confirm that it agrees with that obtained by (5.18).

   (c) Compute the dispersion matrices of $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\lambda}}$ using (5.27a) and (5.27b), respectively, and compare to that obtained by (5.21).

   Note that the matrix to invert in (5.18) may be ill-conditioned (nearly singular) in this case due to the relative magnitude of the elements of matrices N and K. To increase numerical stability, you may need to scale matrix K and vector $\boldsymbol{\kappa}_0$ before using them in (5.18). Try scaling by 10×10^4. No scaling should be necessary for the other formulas.

[1] The assumption of F-distribution is based on an underlying assumption that the residuals are normally distributed, which means that functions of their squares, such as Ω and R, have a χ^2-distribution. The ratio of two independent variables that each have a χ^2-distribution is itself a random variable with F-distribution. Recall that no assumption about the probability density function of the random errors was required for the derivation of their least-squares prediction.

3. With reference to Section 5.4.1, we now simulate case 1 by changing the third element of the last row of matrix A from -1 to 0 and by changing the last element of vector $\boldsymbol{y}$ from -0.162582 to 67.992. All matrices involving N and vectors involving $\boldsymbol{y}$ must be recomputed accordingly.

   (a) Confirm that the revised matrix $N = A^T PA$ has full rank.

   (b) Compute the vector of estimated parameters $\hat{\boldsymbol{\xi}}$ by both (5.8d) and (5.13) and confirm that they are equal.

   (c) Compute the dispersion matrices of $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\lambda}}$ using (5.17) and (5.21) and confirm that they are equal.

   (d) Compute the dispersion matrices using the formulas for case 1, (5.26a) and (5.26b), and confirm that they agree with the respective formulas for case 2, (5.27a) and (5.27b).

   Note that the solution of this problem will not match that of the preceding problem; they are different problems.

4. Using the GMM with constraints, constrain the height of point D in problem 9 of Section 3.6 to 1928.277 ft and check that the LESS agrees with what you computed in parts (b), (c), and (d) of that problem.

5. By imposing certain constraints upon the unknown parameters, the affine transformation problem presented in Problem 8 of Section 3.6, can be converted to an orthogonality-preserving transformation (only one rotation angle instead of two), or, by a different set of constraints, it can be converted to a similarity transformation (one rotation angle and one scale factor). Using the data from Problem 8, setup the GMM with constraints, state the model redundancy, and compute the LESS for the unknown parameters and variance component in the following two cases:

   (a) Orthogonality-preserving transformation: Impose a constraint on the second rotation angle so that $\epsilon = 0$ via the following:

   $$\xi_4/\xi_6 = \xi_5/\xi_3 \Rightarrow \boxed{\xi_3\xi_4 - \xi_5\xi_6 = 0.}$$

   Note that linearization is required in this case.

   (b) Similarity transformation: Impose the constraints that $\epsilon = 0$ and $\omega_1 = \omega_2$ via the following:

   $$\boxed{\xi_3 - \xi_6 = 0 \text{ and } \xi_4 - \xi_5 = 0.}$$

6. To monitor deformation, points P_1 and P_2 were established between deforming and non-deforming regions, respectively. Distances were observed from three known points, A, B, and C, to both points P_1 and P_2 (see Figure 5.2) The 2D coordinates of the known points are listed in Table 5.1, and the observations are listed in Table 5.2. The variance of each observation is

$\sigma^2 = (0.005\,\mathrm{m})^2$. Distances observed from the same point have a correlation coefficient of $\rho = 0.4$. Otherwise the observations are uncorrelated. Suppose the baseline between points P_1 and P_2 is thought to be 251.850 m (perhaps determined from a previous survey), and it is decided to use this value as a constraint in a least-squares adjustment. Determine the following by use of the LESS within the GMM with constraints:

(a) The redundancy of the model.

(b) The 2D coordinates of points P_1 and P_2 and their dispersion matrix.

(c) The vector of observation residuals and its dispersion matrix.

(d) The estimated variance component.

(e) Setup a hypothesis test with significance level $\alpha = 0.05$ and determine if the constraint is consistent with the observations.

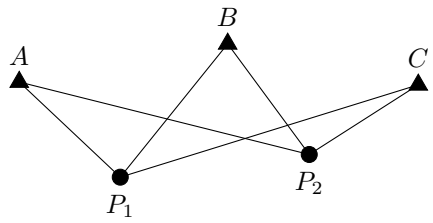

Figure 5.2: Observations from known points A, B, and C

Table 5.1: Coordinates of known points

Station	X [m]	Y [m]
A	456.351	500.897
B	732.112	551.393
C	984.267	497.180

Table 5.2: Observations from known points A, B, and C

From	To	Obs. [m]
A	P_1	183.611
A	P_2	395.462
B	P_1	226.506
B	P_2	181.858
C	P_1	412.766
C	P_2	171.195

7. The data plotted in Figure 3.1 are listed in Table 5.3 below, where the x-coordinates are assumed to be known and the y-coordinates were measured independently and have a common variance of $\sigma^2 = (1\,\text{cm})^2$.

   Now suppose a fitted parabola must pass through data point number 5 exactly. Compute the LESS within the GMM with constraints for the three unknown parameters of the parabola and form a hypothesis test to check the validity of the constraint.

Table 5.3: Known x-coordinates and measured y-coordinates plotted in Figure 3.1

No.	x_i [m]	y_i [m]
1	1.001	1.827
2	2.000	1.911
3	3.001	1.953
4	4.000	2.016
5	5.000	2.046
6	6.003	2.056
7	7.003	2.062
8	8.003	2.054
9	9.001	2.042
10	9.998	1.996
11	11.001	1.918
12	12.003	1.867

5.8 Summary Formulas for the Least-Squares Solution Within the Gauss-Markov Model with Constraints

The Gauss-Markov Model with constraints is given by

$$\boxed{\begin{aligned} \underset{n\times 1}{\boldsymbol{y}} &= \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e},\ \ \boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2 P^{-1}),\ \ \operatorname{rk} A =: q \le m, \\ \underset{l\times 1}{\boldsymbol{\kappa}_0} &= \underset{l\times m}{K}\,\boldsymbol{\xi},\ \ \operatorname{rk} K =: l \ge m - q,\ \ \operatorname{rk}\big[A^T,\, K^T\big] = m. \end{aligned}}$$

Table 5.4: Summary formulas for the LESS within the GMM with constraints

Quantity	Formula	Eq.
Model redundancy	$r = n - m + \operatorname{rk} K = n - m + l$	(5.3)
Vector of estimated parameters, when $\operatorname{rk} A = m$	$\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c} + N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{\kappa}_0 - KN^{-1}\boldsymbol{c}\big)$	(5.8d)
Dispersion matrix for estimated parameters, when $\operatorname{rk} A = m$	$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 \cdot \big[N^{-1} - N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}KN^{-1}\big]$	(5.26a)
Vector of estimated parameters, when $\operatorname{rk} A < m$	$\hat{\boldsymbol{\xi}} = \big(N + K^TK\big)^{-1}\boldsymbol{c} + \big(N + K^TK\big)^{-1}K^T\big[K\big(N + K^TK\big)^{-1}K^T\big]^{-1}\big[\boldsymbol{\kappa}_0 - K\big(N + K^TK\big)^{-1}\boldsymbol{c}\big]$	(5.13)
Dispersion matrix for estimated parameters, when $\operatorname{rk} A < m$	$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 \cdot \big[N_K^{-1} - N_K^{-1}K^T\big(KN_K^{-1}K^T\big)^{-1}KN_K^{-1}\big]$ with $N_K \coloneqq N + K^TK$	(5.27a)
Vector of predicted residuals	$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}$	(5.29)
Dispersion matrix for residuals, when $\operatorname{rk} A = m$	$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot \big\{P^{-1} - A\big[N^{-1} - N^{-1}K^T(KN^{-1}K^T)^{-1}KN^{-1}\big]A^T\big\}$	(5.33b)

Continued on next page

Quantity	Formula	Eq.
Dispersion matrix for residuals, when $\operatorname{rk} A < m$	$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot \big(P^{-1} - A\big\{(N+KTK)^{-1} - (N+KTK)^{-1}K^T[K(N+KTK)^{-1}K^T]^{-1}K(N+KTK)^{-1}\big\}A^T\big)$	(5.33b)
Sum of squared residuals (SSR)	$\mathrm{SSR} = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$	(5.40) (5.41)
Estimated variance component	$\hat{\sigma}_0^2 = (\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}})/r$	(5.43c)
Vector of adjusted observations	$\hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}}$	(5.30)
Dispersion matrix for adjusted observations	$D\{\hat{\boldsymbol{\mu}}_y\} = A \cdot D\{\hat{\boldsymbol{\xi}}\} \cdot A^T$	(5.34)

Continued from previous page

Chapter 6

The Gauss-Markov Model with Stochastic Constraints

6.1 Model Definition

The Gauss-Markov Model (GMM) with stochastic constraints is similar in form to the GMM with constraints shown in Chapter 5, with one important difference: the constraints in the stochastic case are specified with some level of uncertainty, expressed in the form of a given weight matrix P_0, or an associated cofactor matrix $Q_0 \coloneqq P_0^{-1}$. The model reads

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \quad \operatorname{rk} A =: q \leq \min\{m, n\}, \tag{6.1a}$$

$$\underset{l\times 1}{\boldsymbol{z}_0} = \underset{l\times m}{K}\,\boldsymbol{\xi} + \boldsymbol{e}_0, \quad \operatorname{rk} K =: l \geq m - q, \tag{6.1b}$$

$$\begin{bmatrix}\boldsymbol{e}\\ \boldsymbol{e}_0\end{bmatrix} \sim \Big(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2 \begin{bmatrix}P^{-1} & 0\\ 0 & P_0^{-1}\end{bmatrix}\Big). \tag{6.1c}$$

Note that in this model there is no correlation between the random error vectors $\boldsymbol{e}$ and $\boldsymbol{e}_0$. Also, the unknown variance component σ_0^2 is common to both cofactor matrices P^{-1} and P_0^{-1}. However, there may be correlations within one or both of the cofactor matrices, just not between them. Depending on the application, the data in the vector $\boldsymbol{y}$ can be thought of as new information, while the constraint information in the vector $\boldsymbol{z}_0$ can be thought of as prior information (for example, $\boldsymbol{z}_0$ could contain coordinates estimated from a previous adjustment, now considered as prior information).

It is required that the column space of the augmented matrix $[A^T \,|\, K^T]$ spans $\mathbb{R}^m$, which holds when the rank condition

$$\operatorname{rk}\big[A^T \,|\, K^T\big] = m \tag{6.2}$$

is satisfied. The redundancy of the model is give by

$$\boxed{r = n - m + l.} \tag{6.3}$$

In words, we can say that the redundancy is the number of observation equations minus the number of parameters to estimate plus the number of constraint equations.

6.2 Least-Squares Solution

According to Schaffrin (1995), the LEast-Squares Solution (LESS) for the unknown parameters $\boldsymbol{\xi}$ within model (6.1) may be derived by minimizing the Lagrange target function

$$\Phi(\boldsymbol{e}, \boldsymbol{e}_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\lambda}_0) = \boldsymbol{e}^T P \boldsymbol{e} + \boldsymbol{e}_0^T P_0 \boldsymbol{e}_0 + 2\big[\boldsymbol{\lambda}^T,\, \boldsymbol{\lambda}_0^T\big]\left(\begin{bmatrix} A \\ K \end{bmatrix}\boldsymbol{\xi} + \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e}_0 \end{bmatrix} - \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix}\right) = \\ = \underset{\boldsymbol{e},\boldsymbol{e}_0,\boldsymbol{\xi},\boldsymbol{\lambda},\boldsymbol{\lambda}_0}{\text{stationary}}. \tag{6.4}$$

Here we simply consider (6.1) as an extended GMM and apply *the addition theory of normal equations* as follows:

$$\begin{bmatrix} A^T & K^T \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P_0 \end{bmatrix}\begin{bmatrix} A \\ K \end{bmatrix}\cdot\hat{\boldsymbol{\xi}} = \begin{bmatrix} A^T & K^T \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & P_0 \end{bmatrix}\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix} \tag{6.5a}$$

or

$$\big(N + K^T P_0 K\big)\hat{\boldsymbol{\xi}} = \boldsymbol{c} + K^T P_0 \boldsymbol{z}_0, \tag{6.5b}$$

where

$$\big[N, \boldsymbol{c}\big] \coloneqq A^T P\big[A, \boldsymbol{y}\big]. \tag{6.6}$$

In the case where the matrix N is invertible, the Sherman-Morrison-Woodbury-Schur formula (A.6a) may be used to invert the matrix on the left side of (6.5b) as in the following:

$$\boxed{\hat{\boldsymbol{\xi}} = \big(N + K^T P_0 K\big)^{-1}\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big) =} \tag{6.7a}$$

$$= \big[N^{-1} - N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}\big]\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big) =$$

$$= N^{-1}\boldsymbol{c} + N^{-1}K^T P_0 \boldsymbol{z}_0 + N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\cdot$$

$$\cdot\big(-KN^{-1}\boldsymbol{c} - KN^{-1}K^T P_0 \boldsymbol{z}_0\big) =$$

$$= N^{-1}\boldsymbol{c} + N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big[\big(P_0^{-1} + KN^{-1}K^T\big)P_0\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c} -$$

$$- KN^{-1}K^T P_0 \boldsymbol{z}_0\big] \Rightarrow$$

$$\boxed{\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c} + N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big).} \tag{6.7b}$$

Thus, the LESS (6.7b) can be viewed as a weighted average between the prior and the new information. The vector $\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}$ is referred to as a *vector of discrepancies*. The solution can also be recognized as an update to the solution $\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c}$ within the GMM (3.1). It is also interesting to express it as an update to the LESS within the GMM with "fixed" constraints (5.1). This can be done by changing the symbols $\hat{\boldsymbol{\xi}}$ and $\boldsymbol{\kappa}_0$ in (5.8d) to $\hat{\boldsymbol{\xi}}_K$ and $\boldsymbol{z}_0$, respectively, solving for $N^{-1}\boldsymbol{c}$ in terms of these renamed variables, and substituting into (6.7b), which yields the following:

$$\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_K + N^{-1}K^T\big[\big(P_0^{-1} + KN^{-1}K^T\big)^{-1} - \big(KN^{-1}K^T\big)^{-1}\big]\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big). \quad (6.8)$$

Note that as $P_0^{-1} = Q_0 \to 0$, $\hat{\boldsymbol{\xi}} \to \hat{\boldsymbol{\xi}}_K$.

By applying the laws of covariance propagation to (6.7a), the dispersion matrix for the vector of estimated parameters $\hat{\boldsymbol{\xi}}$ is computed as follows:

$$\begin{aligned} D\{\hat{\boldsymbol{\xi}}\} &= \big(N + K^TP_0K\big)^{-1}D\{\boldsymbol{c} + K^TP_0\boldsymbol{z}_0\}\big(N + K^TP_0K\big)^{-1} = \\ &= \sigma_0^2\big(N + K^TP_0K\big)^{-1}\big(N + K^TP_0K\big)\big(N + K^TP_0K\big)^{-1} \Rightarrow \\ \boxed{D\{\hat{\boldsymbol{\xi}}\} &= \sigma_0^2\big(N + K^TP_0K\big)^{-1} = \sigma_0^2\big[N^{-1} - N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}\big].} \end{aligned} \quad (6.9)$$

The subtraction in (6.9) implies that our knowledge of the parameters has improved (variance decreased) by supplying the additional prior information, provided the estimated variance component $\hat{\sigma}_0^2$ does not change much in doing so. Indeed, if the new data, $\boldsymbol{y}$, is consistent with the old, $\boldsymbol{z}_0$, then $\hat{\sigma}_0^2$ is not expected to change very much when the data are combined. In contrast, $\hat{\sigma}_0^2$ is expected to increase if there is inconsistency between the old and new information. In such a case, it may be advisable to introduce a second variance component, one associated with the weight matrix P and the other with P_0. This is the purpose of the variance component model, which is introduced in the Advanced Adjustment Computations Notes.

We now present the residual vectors $\tilde{\boldsymbol{e}}$ and $\tilde{\boldsymbol{e}}_0$ (also called predicted random error vectors). The residual vector $\tilde{\boldsymbol{e}}$ for the observations $\boldsymbol{y}$ is computed by

$$\boxed{\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}.} \quad (6.10)$$

The residual vector $\tilde{\boldsymbol{e}}_0$ associated with the prior information $\boldsymbol{z}_0$ is

$$\boxed{\tilde{\boldsymbol{e}}_0 = \boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}} =} \quad (6.11a)$$

$$\begin{aligned} &= \big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) - \big(KN^{-1}K^T + P_0^{-1} - P_0^{-1}\big)\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\cdot \\ &\qquad\qquad \cdot \big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) = \\ &= \big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) - \Big[\big(KN^{-1}K^T + P_0^{-1}\big)\big(P_0^{-1} + KN^{-1}K^T\big)^{-1} - \\ &\qquad\qquad - P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\Big]\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) = \\ &= \big\{I_l - \big[I_l - P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big]\big\}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) = \end{aligned}$$

$$= P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big) \Rightarrow$$

$$\boxed{\tilde{\boldsymbol{e}}_0 = \big(I_l + KN^{-1}K^T P_0\big)^{-1}\big(\boldsymbol{z}_0 - KN^{-1}\boldsymbol{c}\big).} \tag{6.11b}$$

The dispersion matrix of the residual vectors is derived as follows (see also Practice Problem 2 in Section 6.6):

$$D\{\begin{bmatrix}\tilde{\boldsymbol{e}}\\ \tilde{\boldsymbol{e}}_0\end{bmatrix}\} = D\{\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}\} + D\{\begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\} - 2C\{\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}, \begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\} = D\{\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}\} - D\{\begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\} = \tag{6.12a}$$

$$= \sigma_0^2\begin{bmatrix}P^{-1} & 0\\ 0 & P_0^{-1}\end{bmatrix} - \sigma_0^2\begin{bmatrix}A\\ K\end{bmatrix}\big[N^{-1} - N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}\big]\cdot$$

$$\cdot\begin{bmatrix}A^T & K^T\end{bmatrix} =$$

$$= \sigma_0^2\begin{bmatrix}P^{-1} - AN^{-1}A^T & -AN^{-1}K^T\\ -KN^{-1}A^T & P_0^{-1} - KN^{-1}K^T\end{bmatrix} +$$

$$+\sigma_0^2\begin{bmatrix}AN^{-1}K^T\\ KN^{-1}K^T\end{bmatrix}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\begin{bmatrix}KN^{-1}A^T & KN^{-1}K^T\end{bmatrix}. \tag{6.12b}$$

From (6.12b), we can write the dispersion matrices for the residual vectors individually as

$$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big) + \sigma_0^2 AN^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}A^T \Rightarrow \tag{6.13a}$$

$$\boxed{D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2\big[P^{-1} - A\big(N + K^T P_0 K\big)^{-1}A^T\big] =} \tag{6.13b}$$

$$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 P^{-1} - D\{\hat{\boldsymbol{\mu}}_y\}, \tag{6.13c}$$

and

$$D\{\tilde{\boldsymbol{e}}_0\} = \sigma_0^2 P_0^{-1} - \sigma_0^2 KN^{-1}K^T + \sigma_0^2 KN^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1}K^T =$$

$$= \sigma_0^2 P_0^{-1} - \sigma_0^2 KN^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\cdot$$

$$\cdot\big(P_0^{-1} + KN^{-1}K^T - KN^{-1}K^T\big) =$$

$$= \sigma_0^2 P_0^{-1} - \sigma_0^2 KN^{-1}K^T\big(I_l + P_0 KN^{-1}K^T\big)^{-1} =$$

$$= \sigma_0^2 P_0^{-1}\big(I_l + P_0 KN^{-1}K^T\big)\big(I_l + P_0 KN^{-1}K^T\big)^{-1} - \sigma_0^2 KN^{-1}K^T\cdot$$

$$\cdot\big(I_l + P_0 KN^{-1}K^T\big)^{-1} =$$

$$= \sigma_0^2 P_0^{-1}\big(I_l + P_0 KN^{-1}K^T\big)^{-1} + \sigma_0^2 KN^{-1}K^T\big(I_l + P_0 KN^{-1}K^T\big)^{-1} -$$

$$- \sigma_0^2 KN^{-1}K^T\big(I_l + P_0 KN^{-1}K^T\big)^{-1} \Rightarrow$$

$$\boxed{D\{\tilde{\boldsymbol{e}}_0\} = \sigma_0^2 P_0^{-1}\big(I_l + P_0 KN^{-1}K^T\big)^{-1}.} \tag{6.14}$$

We summarize by listing a few equivalent formulas for $D\{\tilde{\boldsymbol{e}}_0\}$.

$$D\{\tilde{\boldsymbol{e}}_0\} = \sigma_0^2 P_0^{-1}\big(I_l + P_0 K N^{-1} K^T\big)^{-1} = \tag{6.15a}$$

$$= \sigma_0^2\big(I_l + K N^{-1} K^T P_0\big)^{-1} P_0^{-1} = \tag{6.15b}$$

$$= \sigma_0^2 P_0^{-1}\big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1} = \tag{6.15c}$$

$$= \sigma_0^2\big(P_0 + P_0 K N^{-1} K^T P_0\big)^{-1} = \tag{6.15d}$$

$$= \sigma_0^2\big[P_0^{-1} - K\big(N + K^T P_0 K\big)^{-1} K^T\big] = \tag{6.15e}$$

$$= \sigma_0^2 P^{-1} - D\{\hat{\boldsymbol{\mu}}_{z_0}\} \tag{6.15f}$$

The symmetry of the matrix $D\{\tilde{\boldsymbol{e}}_0\}$ has been exploited to get from (6.15a) to (6.15b), using the rule for the transpose of a matrix product (A.1) and the rule for the transpose of an inverse (A.2). Also (A.3) has been used in the above.

Now it remains to write a succinct form for the covariance matrix $C\{\tilde{\boldsymbol{e}}, \tilde{\boldsymbol{e}}_0\}$, beginning with the off-diagonal element of (6.12b).

$$C\{\tilde{\boldsymbol{e}}, \tilde{\boldsymbol{e}}_0\} = -\sigma_0^2 A N^{-1} K^T + \sigma_0^2 A N^{-1} K^T\big(P_0^{-1} + K N^{-1} K^T\big)^{-1} K N^{-1} K^T = \tag{6.16a}$$

$$= -\sigma_0^2 A N^{-1} K^T\big(P_0^{-1} + K N^{-1} K^T\big)^{-1}\big(P_0^{-1} + K N^{-1} K^T - K N^{-1} K^T\big) = \tag{6.16b}$$

$$= -\sigma_0^2 A N^{-1} K^T\big(I_l + P_0 K N^{-1} K^T\big)^{-1} = \tag{6.16c}$$

$$= -\sigma_0^2 A\big(I_m + N^{-1} K^T P_0 K\big)^{-1} N^{-1} K^T = \tag{6.16d}$$

$$= -\sigma_0^2 A\big(N + K^T P_0 K\big)^{-1} K^T \tag{6.16e}$$

The line following (6.16c) is based on relations shown in equations (A.8). To see how these equations are used, compare what follows the term $-\sigma_0^2 A$ in (6.16c) and (6.16d), with the first and last lines in (A.8).

We also note that in the GMM with stochastic constraints, the predicted residual vector $\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}$ by itself is no longer a projection of $\boldsymbol{y}$ onto the range space of A. However, the vector $\big[\tilde{\boldsymbol{e}}^T,\, \tilde{\boldsymbol{e}}_0^T\big]^T$ does represent a projection of $\big[\boldsymbol{y}^T,\, \boldsymbol{z}_0^T\big]^T$ onto the range space of $[A^T,\, K^T]^T$, since

$$\begin{bmatrix}\tilde{\boldsymbol{e}}\\ \tilde{\boldsymbol{e}}_0\end{bmatrix} = \begin{bmatrix}\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\\ \boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}\end{bmatrix} = \left\{\begin{bmatrix}I_n & 0\\ 0 & I_l\end{bmatrix} - \begin{bmatrix}A\\ K\end{bmatrix}\big(N + K^T P_0 K\big)^{-1}\begin{bmatrix}A^T P & K^T P_0\end{bmatrix}\right\}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}, \tag{6.17}$$

and the matrix in braces is idempotent, which can be verified by application of (1.7a).

The adjusted observations and adjusted constraint values are easily computed by

$$\boxed{\hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}} = A\hat{\boldsymbol{\xi}},} \tag{6.18}$$

and

$$\boxed{\hat{\boldsymbol{\mu}}_{z_0} = \boldsymbol{z}_0 - \tilde{\boldsymbol{e}}_0 = K\hat{\boldsymbol{\xi}}.} \tag{6.19}$$

Their respective dispersion matrices are derived by simple application of variance propagation as follows:

$$\boxed{D\{\hat{\boldsymbol{\mu}}_y\} = D\{A\hat{\boldsymbol{\xi}}\} = A \cdot D\{\hat{\boldsymbol{\xi}}\} \cdot A^T = \sigma_0^2 \cdot A\big(N + K^T P_0 K\big)^{-1} A^T,} \tag{6.20}$$

$$\boxed{D\{\hat{\boldsymbol{\mu}}_{z_0}\} = D\{K\hat{\boldsymbol{\xi}}\} = K \cdot D\{\hat{\boldsymbol{\xi}}\} \cdot K^T = \sigma_0^2 \cdot K\big(N + K^T P_0 K\big)^{-1} K^T.} \tag{6.21}$$

Here, $\hat{\boldsymbol{\mu}}_y$ is also interpreted as an estimate of the true, and thus unknown, vector of observables $\boldsymbol{\mu}_y$, where $E\{\boldsymbol{y}\} = \boldsymbol{\mu}_y$; likewise, $E\{\boldsymbol{z}_0\} = \boldsymbol{\mu}_{z_0}$.

Alternative derivation of normal equations Starting with the Lagrange target function (6.4), the vector of random errors $\boldsymbol{e}$ and the vector of Lagrange multipliers $\boldsymbol{\lambda}$ can be eliminated by substitution of $\boldsymbol{y} - A\boldsymbol{\xi}$ for $\boldsymbol{e}$. Furthermore, by introducing

$$-P_0^{-1}\boldsymbol{\lambda}_0 = \boldsymbol{e}_0 = \boldsymbol{z} - K\boldsymbol{\xi}, \tag{6.22a}$$

as in Schaffrin (1995), the target function can be expressed equivalently as

$$\Phi(\boldsymbol{\xi}, \boldsymbol{\lambda}_0) = (\boldsymbol{y} - A\boldsymbol{\xi})^T P(\boldsymbol{y} - A\boldsymbol{\xi}) + 2\boldsymbol{\lambda}_0^T\big(K\boldsymbol{\xi} - \boldsymbol{z}_0\big) - \boldsymbol{\lambda}_0^T P_0^{-1}\boldsymbol{\lambda}_0 = \underset{\boldsymbol{\xi},\boldsymbol{\lambda}_0}{\text{stationary.}} \tag{6.22b}$$

Minimizing the above target function leads to the following system of normal equations:

$$\begin{bmatrix} N & K^T \\ K & -P_0^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{z}_0 \end{bmatrix}. \tag{6.23}$$

Using (6.1b) and (6.23), we can express the predicted residual vector $\tilde{\boldsymbol{e}}_0$ as a function of the vector of Lagrange multipliers $\hat{\boldsymbol{\lambda}}_0$ as follows:

$$\boldsymbol{z}_0 = K\hat{\boldsymbol{\xi}} + \tilde{\boldsymbol{e}}_0 = K\hat{\boldsymbol{\xi}} - P_0^{-1}\hat{\boldsymbol{\lambda}}_0 \Rightarrow \tilde{\boldsymbol{e}}_0 = -P_0^{-1}\hat{\boldsymbol{\lambda}}_0. \tag{6.24}$$

Therefore, the dispersion of $\tilde{\boldsymbol{e}}_0$ is given also by

$$D\{\tilde{\boldsymbol{e}}_0\} = P_0^{-1} D\{\hat{\boldsymbol{\lambda}}_0)\} P_0^{-1}. \tag{6.25}$$

Assuming matrix N is invertible, from (6.23) we see that the dispersion of $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\lambda}}_0$ can be found from

$$\begin{aligned} D\{\begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\lambda}}_0 \end{bmatrix}\} &= \begin{bmatrix} N & K^T \\ K & -P_0^{-1} \end{bmatrix}^{-1} D\{\begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{z}_0 \end{bmatrix}\} \begin{bmatrix} N & K^T \\ K & -P_0^{-1} \end{bmatrix}^{-1} = \\ &= \sigma_0^2 \begin{bmatrix} N & K^T \\ K & -P_0^{-1} \end{bmatrix}^{-1} \begin{bmatrix} N & 0 \\ 0 & P_0^{-1} \end{bmatrix} \begin{bmatrix} N & K^T \\ K & -P_0^{-1} \end{bmatrix}^{-1} = \\ &= \sigma_0^2 \begin{bmatrix} N & K^T \\ K & -P_0^{-1} \end{bmatrix}^{-1} \begin{bmatrix} N^{-1} & 0 \\ 0 & P_0 \end{bmatrix}^{-1} \begin{bmatrix} N & K^T \\ K & -P_0^{-1} \end{bmatrix}^{-1} = \end{aligned}$$

$$= \sigma_0^2 \left[\begin{array}{c|c} N + K^T P_0 K & 0 \\ \hline 0 & P_0^{-1} + K N^{-1} K^T \end{array}\right]^{-1}. \tag{6.26}$$

The last line was reached by successively applying the rule for the product of two inverses (A.3). From (6.26) we see that

$$D\{\hat{\boldsymbol{\lambda}}_0\} = \sigma_0^2 \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} = \sigma_0^2 \big[P_0 - P_0 K \big(N + K^T P_0 K\big)^{-1} K^T P_0\big]. \tag{6.27}$$

Finally, substituting (6.27) into (6.25) and applying the product-of-inverses rule, we can write

$$D\{\tilde{\boldsymbol{e}}_0\} = \sigma_0^2 P_0^{-1} \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1} = \sigma_0^2 \big(P_0 + P_0 K N^{-1} K^T P_0\big)^{-1}. \tag{6.28}$$

Also, the off-diagonal blocks of (6.26) reveal that $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\lambda}}_0$ are uncorrelated, viz.

$$C(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}_0) = 0. \tag{6.29}$$

6.3 Variance Component Estimate

The derivation of the variance component estimate is shown here in detail. The trace operator is employed analogously to what was done in Section 3.3. We also make use of the following expectation and dispersion relationships:

$$E\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\} = \begin{bmatrix} A^T P & K^T P_0 \end{bmatrix} E\{\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix}\} =$$

$$= \begin{bmatrix} A^T P & K^T P_0 \end{bmatrix} \begin{bmatrix} A \\ K \end{bmatrix} \boldsymbol{\xi} = \big(N + K^T P_0 K\big)\boldsymbol{\xi}, \tag{6.30a}$$

$$D\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\} = D\{\begin{bmatrix} A^T P & K^T P_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix}\} =$$

$$= \sigma_0^2 \begin{bmatrix} A^T P & K^T P_0 \end{bmatrix} \begin{bmatrix} P^{-1} & 0 \\ 0 & P_0^{-1} \end{bmatrix} \begin{bmatrix} PA \\ P_0 K \end{bmatrix} = \sigma_0^2 \big(N + K^T P_0 K\big), \tag{6.31a}$$

as well as

$$E\{\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)^T\} = D\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\} + \\ + E\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\} E\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\}^T, \tag{6.32a}$$

$$E\{\boldsymbol{y}\boldsymbol{y}^T\} = D\{\boldsymbol{y}\} + E\{\boldsymbol{y}\} E\{\boldsymbol{y}\}^T = \sigma_0^2 P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T, \tag{6.32b}$$

$$E\{\boldsymbol{z}_0 \boldsymbol{z}_0^T\} = D\{\boldsymbol{z}_0\} + E\{\boldsymbol{z}_0\} E\{\boldsymbol{z}_0\}^T = \sigma_0^2 P_0^{-1} + K\boldsymbol{\xi}\boldsymbol{\xi}^T K^T. \tag{6.32c}$$

The estimated variance component is derived from the expectation of the combined quadratic forms of the residual vectors, $\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0$, based on the principle

$$\frac{\hat{\sigma}_0^2}{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0} = \frac{\sigma_0^2}{E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0\}}. \tag{6.33}$$

The derivation proceeds as follows:

$$
\begin{aligned}
&E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0\} = \\
&= E\{\Big(\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} - \begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\Big)^T \begin{bmatrix}P & 0\\ 0 & P_0\end{bmatrix}\Big(\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} - \begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\Big)\} = \\
&= E\{\boldsymbol{y}^T P\boldsymbol{y} + \boldsymbol{z}_0^T P_0 \boldsymbol{z}_0 - 2\hat{\boldsymbol{\xi}}^T\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big) + \hat{\boldsymbol{\xi}}^T\big(N + K^T P_0 K\big)\hat{\boldsymbol{\xi}}\} = \\
&= E\{\boldsymbol{y}^T P\boldsymbol{y} + \boldsymbol{z}_0^T P_0 \boldsymbol{z}_0 - 2\hat{\boldsymbol{\xi}}^T\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big) + \hat{\boldsymbol{\xi}}^T\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)\} = \\
&= E\{\boldsymbol{y}^T P\boldsymbol{y} + \boldsymbol{z}_0^T P_0 \boldsymbol{z}_0 - \hat{\boldsymbol{\xi}}^T\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)\} = \\
&= E\{\boldsymbol{y}^T P\boldsymbol{y} + \boldsymbol{z}_0^T P_0 \boldsymbol{z}_0 - \big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)^T\big(N + K^T P_0 K\big)^{-1}\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)\} = \\
&= E\{\mathrm{tr}\big(\boldsymbol{y}^T P\boldsymbol{y}\big) + \mathrm{tr}\big(\boldsymbol{z}_0^T P_0 \boldsymbol{z}_0\big) - \mathrm{tr}\big[\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)^T\big(N + K^T P_0 K\big)^{-1}\cdot \\
&\qquad\qquad \cdot\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)\big]\} = \\
&= E\{\mathrm{tr}\big(P\boldsymbol{y}\boldsymbol{y}^T\big) + \mathrm{tr}\big(P_0 \boldsymbol{z}_0 \boldsymbol{z}_0^T\big) - \mathrm{tr}\big[\big(N + K^T P_0 K\big)^{-1}\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)\cdot \\
&\qquad\qquad \cdot\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)^T\big]\} = \\
&= \mathrm{tr}\big(PE\{\boldsymbol{y}\boldsymbol{y}^T\}\big) + \mathrm{tr}\big(P_0 E\big\{\boldsymbol{z}_0 \boldsymbol{z}_0^T\big\}\big) - \mathrm{tr}\big[\big(N + K^T P_0 K\big)^{-1} E\{\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)\cdot \\
&\qquad\qquad \cdot\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)^T\}\big] = \\
&= \mathrm{tr}\big(PE\{\boldsymbol{y}\boldsymbol{y}^T\}\big) + \mathrm{tr}\big(P_0 E\big\{\boldsymbol{z}_0 \boldsymbol{z}_0^T\big\}\big) - \mathrm{tr}\big[\big(N + K^T P_0 K\big)^{-1} D\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\}\big] - \\
&\qquad\qquad - \mathrm{tr}\big[\big(N + K^T P_0 K\big)^{-1} E\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\} E\{\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\}^T\big] = \\
&= \mathrm{tr}\big[P\big(\sigma_0^2 P^{-1} + A\boldsymbol{\xi}\boldsymbol{\xi}^T A^T\big)\big] + \mathrm{tr}\big[P_0\big(\sigma_0^2 P_0^{-1} + K\boldsymbol{\xi}\boldsymbol{\xi}^T K^T\big)\big] - \\
&\qquad - \sigma_0^2\,\mathrm{tr}\big[\big(N + K^T P_0 K\big)^{-1}\big(N + K^T P_0 K\big)\big] - \\
&\quad - \mathrm{tr}\big[\big(N + K^T P_0 K\big)^{-1}\big(N + K^T P_0 K\big)\boldsymbol{\xi}\boldsymbol{\xi}^T\big(N + K^T P_0 K\big)\big] = \\
&= \sigma_0^2\,\mathrm{tr}\big(PP^{-1}\big) + \mathrm{tr}\big(PA\boldsymbol{\xi}\boldsymbol{\xi}^T A^T\big) + \sigma_0^2\,\mathrm{tr}\big(P_0 P_0^{-1}\big) + \mathrm{tr}\big(P_0 K\boldsymbol{\xi}\boldsymbol{\xi}^T K^T\big) - \\
&\qquad\qquad - \sigma_0^2\,\mathrm{tr}\big(I_m\big) - \mathrm{tr}\big(\boldsymbol{\xi}\boldsymbol{\xi}^T N + \boldsymbol{\xi}\boldsymbol{\xi}^T K^T P_0 K\big) = \\
&= \sigma_0^2\,\mathrm{tr}\big(I_n\big) + \mathrm{tr}\big(\boldsymbol{\xi}^T N\boldsymbol{\xi}\big) + \sigma_0^2\,\mathrm{tr}\big(I_l\big) + \mathrm{tr}\big(\boldsymbol{\xi}^T K^T P_0 K\boldsymbol{\xi}\big) - \\
&\qquad - \sigma_0^2\,\mathrm{tr}\big(I_m\big) - \mathrm{tr}\big(\boldsymbol{\xi}^T N\boldsymbol{\xi}\big) - \mathrm{tr}\big(\boldsymbol{\xi}^T K^T P_0 K\boldsymbol{\xi}\big) = \\
&= \sigma_0^2(n + l - m) \\
&\Rightarrow \sigma_0^2 = (n + l - m)^{-1}\cdot E\{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0\}
\end{aligned}
$$

From the preceding derivation, it follows that

$$\boxed{\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0}{n - m + l}} \tag{6.34}$$

provides an unbiased estimate of the variance component σ_0^2. Here, the numerator contains the sum of squared residuals

$$\boxed{\mathrm{SSR} : \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0,} \tag{6.35}$$

while the denominator contains the model redundancy, $r = n - m + l$, as specified in (6.3).

6.4 Hypothesis Test Using the Estimated Variance Component

Hypothesis testing can be used to validate that the least-squares solution is consistent with the stochastic constraints in the model (6.1). The test statistic to be computed is comprised of a ratio of two estimated, and therefore random, variances and thus has an F-distribution (see Section 9.4). The idea is to extract from the sum of the quadratic products in (6.35) the associated sum of squared residuals that would have been computed for the LESS within the unconstrained GMM solution, viz. $\hat{\boldsymbol{\xi}}_u = N^{-1}\boldsymbol{c}$, had it been estimated. We label this quantity Ω. What remains after extracting Ω from (6.35) is a quantity that depends on the weight matrix P_0. We denote this remaining portion as $R(P_0)$ to indicate that it is a function of P_0. Both Ω and $R(P_0)$ are scalars, and both have random properties. These two variables, which are used to form the test statistic, are defined as follows:

$$\Omega \coloneqq \big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big)^T P\big(\boldsymbol{y} - AN^{-1}\boldsymbol{c}\big) = \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T N^{-1}\boldsymbol{c}, \tag{6.36a}$$

$$R(P_0) \coloneqq \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0\tilde{\boldsymbol{e}}_0 - \Omega. \tag{6.36b}$$

Note: If the matrix N in (6.36a) is singular, than N^{-1} can be replaced with any generalized inverse of N as discussed on page 113.

Again we note that $\hat{\boldsymbol{\xi}}_u = N^{-1}\boldsymbol{c}$ represents the least-squares solution within model (6.1) had the stochastic constraints been omitted. In the following derivations, we also make use of (6.7b), (6.11a), and (6.11b) to write formulas for $\tilde{\boldsymbol{e}}_0$ and $\hat{\boldsymbol{\xi}}$ in terms of $\hat{\boldsymbol{\xi}}_u$ as follows:

$$\tilde{\boldsymbol{e}}_0 = \boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}} = \big(I_l + KN^{-1}K^T P_0\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big), \tag{6.37}$$

$$\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_u + N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0. \tag{6.38}$$

As long as N is non-singular (matrix A has full-column rank), we can determine a formula for $R(P_0)$ independent of Ω. To do so, we begin with the quadratic form for the full predicted residual vector appearing in (6.35) (also called sum of squared residuals, SSR) and decompose it into Ω and $R(P_0)$. Note that the crossed-out vector in the first line below is neglected since its contribution vanishes in the

quadratic product.

$$
\begin{aligned}
&\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0 = \Big(\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} - \begin{bmatrix}\cancel{A}\\ \cancel{K}\end{bmatrix}\hat{\boldsymbol{\xi}}\Big)^T \begin{bmatrix}P & 0\\ 0 & P_0\end{bmatrix}\Big(\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix} - \begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\Big) = \\
&= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T PA\hat{\boldsymbol{\xi}} + \boldsymbol{z}_0^T P_0 \boldsymbol{z}_0 - \boldsymbol{z}_0^T P_0 K\hat{\boldsymbol{\xi}} = \\
&= \boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T PA\big(\hat{\boldsymbol{\xi}}_u + N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0\big) + \boldsymbol{z}_0^T P_0\boldsymbol{z}_0 - \boldsymbol{z}_0^T P_0 K\big(\hat{\boldsymbol{\xi}}_u + N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0\big) = \\
&= \underbrace{\big(\boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{y}^T PA\hat{\boldsymbol{\xi}}_u\big)}_{\Omega} + \boldsymbol{z}_0^T P_0 \quad \underbrace{\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big)}_{(I_l + KN^{-1}K^T P_0)\tilde{\boldsymbol{e}}_0} \quad - \underbrace{\big(\boldsymbol{c} + K^T P_0\boldsymbol{z}_0\big)^T}_{\hat{\boldsymbol{\xi}}^T(N + K^T P_0 K)} N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0 = \\
&= \Omega + \boldsymbol{z}_0^T P_0\big(I_l + KN^{-1}K^T P_0\big)\tilde{\boldsymbol{e}}_0 - \hat{\boldsymbol{\xi}}^T\big(N + K^T P_0 K\big)N^{-1}K^T P_0\tilde{\boldsymbol{e}}_0 = \\
&= \Omega + \boldsymbol{z}_0^T\big(I_l + P_0 KN^{-1}K^T\big)P_0\tilde{\boldsymbol{e}}_0 - \big(K\hat{\boldsymbol{\xi}}\big)^T\big(I_l + P_0 KN^{-1}K^T\big)P_0\tilde{\boldsymbol{e}}_0 = \\
&= \Omega + \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}\big)^T\big(I_l + P_0 KN^{-1}K^T\big)P_0\tilde{\boldsymbol{e}}_0 = \\
&= \Omega + \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big)^T\big(I_l + P_0 KN^{-1}K^T\big)^{-1}\big(I_l + P_0 KN^{-1}K^T\big)\cdot \\
&\qquad\qquad \cdot\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big) = \\
&= \Omega + \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big)^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big) = \\
&= \Omega + R(P_0)
\end{aligned}
$$

Thus, $R(P_0)$ is defined as

$$
R(P_0) \coloneqq \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big)^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_u\big), \tag{6.39}
$$

with $\hat{\boldsymbol{\xi}}_u \coloneqq N^{-1}\boldsymbol{c}$ and assuming the inverse of N exists, in which case (6.36b) and (6.39) should yield identical results.

Finally, the test statistic T can be expressed as a ratio of $R(P_0)$ to Ω, viz.

$$
\boxed{
\begin{aligned}
T &= \frac{(\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0\tilde{\boldsymbol{e}}_0 - \Omega)/(l - m + q)}{\Omega/(n - q)} = \\
&= \frac{R\big(P_0\big)/(l - m + q)}{\Omega/(n - q)} \sim F(l - m + q, n - q).
\end{aligned}
} \tag{6.40}
$$

Recall from (6.2) that $l \coloneqq \mathrm{rk}(K)$ and $q \coloneqq \mathrm{rk}(A)$.

The following hypothesis test can now be performed, where $\mathcal{N}$ stands for the normal distribution and $\boldsymbol{z}_0$ is an unknown quantity:

$$
H_0 : \boldsymbol{z}_0 \sim \mathcal{N}\big(K\boldsymbol{\xi}, \sigma_0^2 P_0^{-1}\big) \text{ against } H_A : \boldsymbol{z}_0 \sim \mathcal{N}\big(\boldsymbol{z}_0 \neq K\boldsymbol{\xi}, \sigma_0^2 P_0^{-1}\big). \tag{6.41}
$$

The term H_0 is called the null hypothesis, and H_A is the alternative hypothesis. After choosing a level of significance α and taking $F_{\alpha, l-m+q, n-q}$ from a table of critical values for the F-distribution, the following logic can be applied:

$$
\text{If } T \leq F_{\alpha, l-m+q, n-q} \text{ accept } H_0\text{; else reject } H_0. \tag{6.42}
$$

If MATLAB is available, the critical value may be generated by use of the MATLAB command `finv(`$1-\alpha$`,`$l-m+q$`,`$n-q$`)`.

6.5 Some Comments on Reproducing Estimators

In this section we briefly discuss two estimators within the Gauss-Markov Model with stochastic constraints (6.1) that leave the constrained parameters unchanged, i.e., unchanged from the values specified in $\boldsymbol{z}_0$. Such estimators are called *reproducing estimators.* For example, in a network adjustment problem the a priori coordinates of a station might need to be left unchanged by the adjustment.

For simplicity, we restrict the discussion to models of full rank, i.e., rk $A = m$, where m is the number of columns of matrix A and also the number of parameters to estimate.

One approach that is sometimes taken to obtain a reproducing estimator is to simply adopt the estimator within the Gauss-Markov Model with fixed constraints shown in (5.8d), which is optimal for that model. Two points should be made regarding the use of that estimator within the model (6.1). First, it is not an optimal estimator within model (6.1), and, second, its dispersion matrix shown in (5.26a) and (5.27a) is not correct within model (6.1). In the following, we show the proper dispersion matrix for the reproducing estimator within model (6.1). First, we introduce different subscripts to denote various linear estimators for $\boldsymbol{\xi}$.

$\hat{\boldsymbol{\xi}}_U$ denotes the unconstrained estimator $\hat{\boldsymbol{\xi}}_U = N^{-1}\boldsymbol{c}$, which is not optimal within model (6.1).

$\hat{\boldsymbol{\xi}}_K$ denotes the reproducing estimator from equation (5.8d), which is not optimal within model (6.1).

$\hat{\boldsymbol{\xi}}_S$ denotes the estimator from equation (6.7a), which is optimal within model (6.1).

First we express the estimator $\hat{\boldsymbol{\xi}}_K$ as a function of the optimal estimator $\hat{\boldsymbol{\xi}}_S$. Using (6.5b), we can write

$$\big(N + K^T P_0 K\big)^{-1}\boldsymbol{c} = \hat{\boldsymbol{\xi}}_S - \big(N + K^T P_0 K\big)^{-1} K^T P_0 \boldsymbol{z}_0. \tag{6.43}$$

We then repeat (5.8d) for the estimator $\hat{\boldsymbol{\xi}}_K$ with N replaced by $\big(N + K^T P_0 K\big)$ and $\boldsymbol{\kappa}_0$ replaced by $\boldsymbol{z}_0$ according to the model (6.1). This is our starting point.

$$\begin{aligned}\hat{\boldsymbol{\xi}}_K &= \big(N + K^T P_0 K\big)^{-1}\boldsymbol{c} + \\ &+ \big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1}\big[\boldsymbol{z}_0 - K\big(N + K^T P_0 K\big)^{-1}\boldsymbol{c}\big]\end{aligned} \tag{6.44}$$

Now using (6.43) in (6.44), we can write

$$\begin{aligned}\hat{\boldsymbol{\xi}}_K &= \hat{\boldsymbol{\xi}}_S - \big(N + K^T P_0 K\big)^{-1} K^T P_0 \boldsymbol{z}_0 + \\ &+ \big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1}\big[\boldsymbol{z}_0 - K\big(N + K^T P_0 K\big)^{-1}\boldsymbol{c}\big].\end{aligned}$$

Factoring out $-\big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1}$ yields

$$\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S - \big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1} \cdot \\ \cdot \big\{\big[K\big(N + K^T P_0 K\big)^{-1} K^T\big] P_0 \boldsymbol{z}_0 - \boldsymbol{z}_0 + K\big(N + K^T P_0 K\big)^{-1} \boldsymbol{c}\big\}.$$

Now, from (6.7a) we recognize $K\hat{\boldsymbol{\xi}}_S$ in the above line; thus we write:

$$\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S + \big(N + K^T P_0 K\big)^{-1} K^T \big[K\big(N + K^T P_0 K\big)^{-1} K^T\big]^{-1} \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big). \quad (6.45)$$

We now have the fixed-constraint estimator $\hat{\boldsymbol{\xi}}_K$ expressed as a function of the optimal estimator for model (6.1), namely $\hat{\boldsymbol{\xi}}_S$. Using a familiar formula for $\big(N + K^T P_0 K\big)^{-1}$ and noting that

$$\big(N + K^T P_0 K\big)^{-1} K^T P_0 = N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1},$$

we can rewrite (6.45) as:

$$\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S + \big[N^{-1} - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} K N^{-1}\big] K^T \cdot \\ \cdot \big[K N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1}\big]^{-1} \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big). \quad (6.46)$$

Note the following useful relations:

$$\big[N^{-1} - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} K N^{-1}\big] K^T = \\ = N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1} \quad (6.47)$$

and

$$\big[K N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1}\big]^{-1} = \\ = P_0 \big(P_0^{-1} + K N^{-1} K^T\big) \big(K N^{-1} K^T\big)^{-1}. \quad (6.48)$$

Equation (6.47) is derived as follows:

$$\big[N^{-1} - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} K N^{-1}\big] K^T = \\ = N^{-1} K^T - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} \big(P_0^{-1} + K N^{-1} K^T - P_0^{-1}\big) = \\ = N^{-1} K^T - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} \big(P_0^{-1} + K N^{-1} K^T\big) - \\ - N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} \big(-P_0^{-1}\big) = \\ = N^{-1} K^T - N^{-1} K^T + N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1} = \\ = N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1}.$$

Successive application of the rule for the product of inverted matrices was used in equation (6.48). Substituting (6.47) and (6.48) into (6.46) yields:

$$\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_S + N^{-1} K^T \big(P_0^{-1} + K N^{-1} K^T\big)^{-1} P_0^{-1} P_0 \big(P_0^{-1} + K N^{-1} K^T\big) \cdot \\ \cdot \big(K N^{-1} K^T\big)^{-1} \big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big) =$$

$$= \hat{\boldsymbol{\xi}}_S + N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big). \tag{6.49}$$

Equation (6.49) gives an elegant expression of the fixed-constraint estimator $\hat{\boldsymbol{\xi}}_K$ in terms of the optimal estimator $\hat{\boldsymbol{\xi}}_S$. Realizing that the model with stochastic constraints (6.1) becomes the model with fixed constraints (5.1) when P_0^{-1} is replaced by zero, we can replace (6.49) with (6.50) below, which is also obvious from our starting equation (6.44). This also makes the appropriate dispersion matrix $D\{\hat{\boldsymbol{\xi}}_K\}$ under model (6.1) easier to compute.

$$\hat{\boldsymbol{\xi}}_K = \hat{\boldsymbol{\xi}}_U + N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_U\big) \tag{6.50}$$

Note that $C\{\boldsymbol{z}_0, \boldsymbol{y}\} = 0$, which allows us to apply the dispersion operator to (6.50) as follows:

$$\begin{aligned} D\{\hat{\boldsymbol{\xi}}_K\} &= D\{\hat{\boldsymbol{\xi}}_u - N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}K\hat{\boldsymbol{\xi}}_U\} + \\ &\quad + D\{N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}\boldsymbol{z}_0\} \Rightarrow \\ D\{\hat{\boldsymbol{\xi}}_S \to \hat{\boldsymbol{\xi}}_K\} &= \sigma_0^2 N^{-1} - \sigma_0^2 N^{-1}K^T\big(KN^{-1}K^T\big)^{-1}KN^{-1} + \\ &\quad + \sigma_0^2 N^{-1}K^T\big(KN^{-1}K^T P_0 KN^{-1}K^T\big)^{-1}KN^{-1}. \end{aligned} \tag{6.51}$$

Compare (6.51) to (5.16) to see that $D\{\hat{\boldsymbol{\xi}}_K\}$ increases by

$$\sigma_0^2 N^{-1}K^T\big(KN^{-1}K^T P_0 KN^{-1}K^T\big)^{-1}KN^{-1}$$

when the estimator $\hat{\boldsymbol{\xi}}_K$ is used for the model with stochastic constraints (6.1).

We already noted that $\hat{\boldsymbol{\xi}}_K$ is a sub-optimal (reproducing) estimator within model (6.1). We now give the optimal reproducing estimator without derivation (for details see Schaffrin (1997a)).

$$\hat{\boldsymbol{\xi}}_{opt-rep} = \hat{\boldsymbol{\xi}}_S + K^T\big(KK^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big) \tag{6.52}$$

The symbol $\hat{\boldsymbol{\xi}}_S$ on the right side of (6.52) represents the optimal ("non-reproducing") estimator. Equation (6.52) is identical to (6.49) when N^{-1} is replaced by I.

The dispersion matrix is given by

$$\begin{aligned} D\{\hat{\boldsymbol{\xi}}_{opt-rep}\} &= D\{\hat{\boldsymbol{\xi}}_S\} + D\{K^T\big(KK^T\big)^{-1}\big(\boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}_S\big)\} = \\ &= \sigma_0^2 N^{-1} - \sigma_0^2 N^{-1}K^T\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}KN^{-1} + \\ &\quad + \sigma_0^2 K^T\big(KK^T\big)^{-1}P_0^{-1}\big(P_0^{-1} + KN^{-1}K^T\big)^{-1}P_0^{-1}\big(KK^T\big)^{-1}K. \end{aligned} \tag{6.53}$$

Also note that

$$E\{\hat{\boldsymbol{\xi}}_{opt-rep}\} = \boldsymbol{\xi}, \tag{6.54a}$$

$$\boldsymbol{z}_0 - K\boldsymbol{\xi}_{opt-rep} = \mathbf{0}, \tag{6.54b}$$

$$D\{K\hat{\boldsymbol{\xi}}_{opt-rep}\} = D\{\boldsymbol{z}_0\} = \sigma_0^2 P_0^{-1}. \tag{6.54c}$$

6.6 Practice Problems

1. Given the target function

$$\Phi(\boldsymbol{\xi}, \boldsymbol{\lambda}_0) = (\boldsymbol{y} - A\boldsymbol{\xi})^T P(\boldsymbol{y} - A\boldsymbol{\xi}) + 2\boldsymbol{\lambda}_0^T(K\boldsymbol{\xi} - \boldsymbol{z}_0) - \boldsymbol{\lambda}_0^T P_0^{-1}\boldsymbol{\lambda}_0$$

   from (6.22b), complete the following:

   (a) With the help of (6.22a), show that equations (6.4) and (6.22b) are equivalent.

   (b) Formulate the Euler-Lagrange necessary conditions for the least-squares solution of the unknown parameters $\boldsymbol{\xi}$ and the unknown vector of Lagrange multipliers $\boldsymbol{\lambda}_0$.

   (c) Show how the sufficient condition for minimization is satisfied.

   (d) Using the Euler-Lagrange necessary conditions that you formulated in (a), derive the vector of estimated parameters $\hat{\boldsymbol{\xi}}$ and check that it agrees with (6.7b).

2. Confirm that (6.12a) is correct by showing that

$$C\{\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}_0\end{bmatrix}, \begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\} = D\{\begin{bmatrix}A\\ K\end{bmatrix}\hat{\boldsymbol{\xi}}\}.$$

3. Repeat Problem 4 of Section 5.7, this time using the following constraints:

   (a) Use 1928.277 ft as $\boldsymbol{z}_0$ and $\sigma^2 = (0.005\,\text{ft})^2$ for its variance. Compare your answers to those of Problem 4. Are they the same? If so, what is your explanation for that? Can a hypothesis test be formulated as described in Section 6.4?

   (b) Now add another constraint that requires the height of point D to be 248.750 ft greater than the height of point A, with variance σ^2=2(0.005^2) ft^2. Form a hypothesis test to check the consistency of the observation equations and the constraint equations.

4. Repeat Problem 6 of Section 5.7. This time use 251.850 m as $\boldsymbol{z}_0$ and $\sigma^2 = (0.005\,\text{m})^2$ for its variance. Compare your answers to those of Problem 6. What changes, what stays the same? Form a hypothesis test to check the consistency of the observation equations and the constraint equations.

5. Repeat Problem 7 of Section 5.7. This time use 2.046 m as $\boldsymbol{z}_0$ and $\sigma^2 = (1\,\text{cm})^2$ for its variance. Compare your answers to those of Problem 7. What changes, what stays the same? Form a hypothesis test to check the consistency of the observation equations and the constraint equations.

6. Referring to the example problem in Section 5.4.1, set the vector $\boldsymbol{\kappa}_0$ shown there equal to $\boldsymbol{z}_0$. Use the following matrix for P_0^{-1}:

$$P_0^{-1} = \begin{bmatrix} 2.84067584875257 & 0.533989733139618 & 0.535740019844372 \\ 0.533989733139618 & 2.14132575448909 & 0.531530384522843 \\ 0.535740019844372 & 0.531530384522843 & 2.19379908268108 \end{bmatrix} \cdot \cdot (10 \times 10^{-6})\,\mathrm{m}^2.$$

In addition, multiply the cofactor matrix P^{-1} by 0.017381 and the cofactor matrix P_0^{-1} by 8.709801 to account for the variance components estimated in Smith et al. (2018), which should result in a solution that agrees with the results shown therein.

Complete the following:

(a) Estimate the heights of all points.

(b) Form a hypothesis test to check the consistency of the observation equations and the constraint equations.

7. Show that the total residual vector $[\tilde{\boldsymbol{e}}^T,\, \tilde{\boldsymbol{e}}_0^T]^T$ results from a projection of $[\boldsymbol{y}^T,\, \boldsymbol{z}_0^T]^T$ onto the range space of $[A^T,\, K^T]^T$. Hint: see equation (6.17).

6.7 Summary Formulas for the Least-Squares Solution Within the Gauss-Markov Model with Stochastic Constraints

The Gauss-Markov Model with stochastic constraints is given by

$$\boxed{\begin{aligned} \underset{n\times 1}{\boldsymbol{y}} &= \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \\ \underset{l\times 1}{\boldsymbol{z}_0} &= \underset{l\times m}{K}\,\boldsymbol{\xi} + \boldsymbol{e}_0, \\ \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e}_0 \end{bmatrix} &\sim \Big(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & P_0^{-1} \end{bmatrix} \Big). \end{aligned}}$$

Table 6.1: Summary formulas for the LESS within the Gauss-Markov Model with stochastic constraints

Quantity	Formula	Eq.
Model redundancy	$r = n - m + l$	(6.3)
Vector of estimated parameters	$\hat{\boldsymbol{\xi}} = \big(N + K^T P_0 K\big)^{-1}\big(\boldsymbol{c} + K^T P_0 \boldsymbol{z}_0\big)$	(6.7a)

Continued on next page

Quantity	Formula	Eq.
Dispersion matrix for estimated parameters	$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 \cdot \left(N + K^T P_0 K\right)^{-1}$	(6.9)
Vector of predicted residuals	$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}$	(6.10)
Dispersion matrix for residuals	$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 \cdot \left[P^{-1} - A\left(N + K^T P_0 K\right)^{-1} A^T\right]$	(6.13b)
Vector of residuals of prior information	$\tilde{\boldsymbol{e}}_0 = \boldsymbol{z}_0 - K\hat{\boldsymbol{\xi}}$	(6.11a)
Dispersion matrix for residuals of prior information	$D\{\tilde{\boldsymbol{e}}_0\} = \sigma_0^2 \cdot P_0^{-1}\left(I_l + P_0 K N^{-1} K^T\right)^{-1}$	(6.14)
Sum of squared residuals (SSR)	$\Omega + R(P_0) = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0,\ \Omega \neq \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$	(6.35), (6.36b)
Estimated variance component	$\hat{\sigma}_0^2 = (\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0)/(n - m + l)$	(6.34)
Vector of adjusted observations	$\hat{\boldsymbol{\mu}}_y = \boldsymbol{y} - \tilde{\boldsymbol{e}}$	(6.18)
Dispersion matrix for adjusted observations	$D\{\hat{\boldsymbol{\mu}}_y\} = \sigma_0^2 \cdot A\left(N + K^T P_0 K\right)^{-1} A^T$	(6.20)
Vector of adjusted constraints	$\hat{\boldsymbol{\mu}}_{z_0} = \boldsymbol{z}_0 - \tilde{\boldsymbol{e}}_0$	(6.19)
Dispersion matrix for adjusted constraints	$D\{\hat{\boldsymbol{\mu}}_{z_0}\} = \sigma_0^2 \cdot K\left(N + K^T P_0 K\right)^{-1} K^T$	(6.21)

Continued from previous page

Chapter 7

Sequential Adjustments

A sequential adjustment might be called for when two successive data sets must be combined to estimate a single set of parameters. This type of adjustment is especially useful when only the parameter estimates and their dispersion matrix, but not the associated observations, are available from the first data set. Then, update formulas can be used that allow the second data set to be adjusted in a way that depends on the estimates from the first data set, with the results being equivalent to what would have been computed from a simultaneous adjustment of both data sets. Though we refer to the two data sets respectively as *first* and *second*, they could be any two successive data sets that must be treated by a sequential adjustment, e.g., they could be the ninth and tenth.

7.1 Model Definition

The data model for sequential adjustments is based on *two data sets*, denoted by subscripts 1 and 2, respectively. The first data set is comprised of n_1 observations, and the second is comprised of n_2. It is assumed that the observations from the first data set, $\boldsymbol{y}_1$, are *uncorrelated* with those from the second, $\boldsymbol{y}_2$, i.e., $C\{\boldsymbol{y}_1, \boldsymbol{y}_2\} = 0$. Moreover, all parameters associated with the second data set are also associated with the first data set. Thus, the data model is written as

$$\underset{n_1\times 1}{\boldsymbol{y}_1} = \underset{n_1\times m}{A_1}\,\boldsymbol{\xi} + \boldsymbol{e}_1, \tag{7.1a}$$

$$\underset{n_2\times 1}{\boldsymbol{y}_2} = \underset{n_2\times m}{A_2}\,\boldsymbol{\xi} + \boldsymbol{e}_2, \tag{7.1b}$$

$$\begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix} \sim \Big(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2 \begin{bmatrix}P_1^{-1} & 0\\ 0 & P_2^{-1}\end{bmatrix}\Big). \tag{7.1c}$$

The ranks of the coefficient (design) matrices A_1 and A_2 are such that

$$\operatorname{rk} A_1 = \operatorname{rk} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = m. \tag{7.2}$$

Note that the coefficient matrix A_1 has full column rank, that there is *no correlation* between the random error vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, and that both data sets share a *common variance component* σ_0^2. Also, the total number of observations from both data sets is defined as

$$n := n_1 + n_2. \tag{7.3}$$

The following notation is adopted for normal-equation variables used in Sections 7.1 to 7.3:

$$[N_{ii},\, \boldsymbol{c}_i] = A_i^T P_i \,[A_i,\, \boldsymbol{y}_i]\,, \quad i \in \{1, 2\}, \tag{7.4a}$$

so that

$$N_{11} = A_1^T P_1 A_1,\; N_{22} = A_2^T P_2 A_2,\; \boldsymbol{c}_1 = A_1^T P_1 \boldsymbol{y}_1,\; \text{and } \boldsymbol{c}_2 = A_2^T P_2 \boldsymbol{y}_2. \tag{7.4b}$$

> Subscripts on N and $\boldsymbol{c}$ have somewhat different meanings for sections that follow Section 7.3; careful attention should be paid to their definitions given there.

We use a single hat to denote estimates that are based only on the first data set and a double hat to denote estimates that are based on *both* data sets. For example, the estimate $\hat{\boldsymbol{\xi}}$ is based only on the first data set, whereas the estimate $\hat{\hat{\boldsymbol{\xi}}}$ is based on both data sets. This makes it convenient to show estimates based on both data sets as an update to estimates based on only the first data set.

We recognize a structural similarity between the data model shown in (7.1) and the Gauss-Markov Model with stochastic constraints shown in (6.1). Given this similarity, we may immediately write down a least-squares solution for $\boldsymbol{\xi}$, and its dispersion matrix, in the form of (6.7b) and (6.9), respectively, viewing the second data set as analogous to stochastic constraints.

$$\hat{\hat{\boldsymbol{\xi}}} = \hat{\boldsymbol{\xi}} + N_{11}^{-1} A_2^T \big(P_2^{-1} + A_2 N_{11}^{-1} A_2^T\big)^{-1} \big(\boldsymbol{y}_2 - A_2 \hat{\boldsymbol{\xi}}\big) = \tag{7.5}$$

$$= \hat{\boldsymbol{\xi}} + \big(N_{11} + A_2^T P_2 A_2\big)^{-1} A_2^T P_2 \big(\boldsymbol{y}_2 - A_2 \hat{\boldsymbol{\xi}}\big) \tag{7.6}$$

$$D\{\hat{\hat{\boldsymbol{\xi}}}\} = D\{\hat{\boldsymbol{\xi}}\} - \sigma_0^2 N_{11}^{-1} A_2^T \big(P_2^{-1} + A_2 N_{11}^{-1} A_2^T\big)^{-1} A_2 N_{11}^{-1} \tag{7.7}$$

Equation (A.8a) was used in going from (7.5) to (7.6). It is important to note that the matrix $\big(P_2^{-1} + A_2 N_{11}^{-1} A_2^T\big)$ is of size $n_2 \times n_2$; whereas the size of matrix $\big(N_{11} + A_2^T P_2 A_2\big)$ is $m \times m$. Therefore, if the second data set has only one observation, then $n_2 = 1$, and the update via (7.5) is very fast! This may be the case, for example, in a real-time application where one new observation is added at each epoch in time.

It is also noted that the matrix subtracted in (7.7) is positive-definite , which means that regardless of the precision of the second data set reflected in P_2, the dispersion of the parameters estimated from both data sets will be smaller than that estimated from only the first data set.

7.2 Verification of the Sequential Adjustment

In this section we discuss verification of the sequential adjustment, the aim of which is to confirm that the adjustment based on both data sets is consistent with an adjustment based only on the first data set. By consistent we mean that both the first data set only and the combined data sets fit the model well, implying that the residuals from an adjustment of the first data set would not change much in a sequential adjustment of both data sets.

We can make use of the work done in Chapter 6 to write the estimated variance component $\hat{\sigma}_0^2$ in a form composed of the sum of squared residuals (SSR) Ω based on an adjustment of the first data set only and an update $R(P_2)$ for the contribution to the SSR from the second data set, analogous to the derivation of (6.39). This facilitates hypothesis testing for the purpose of determining if the combined adjustment is consistent with an adjustment based only on the first data set. The decomposition of $\hat{\hat{\sigma}}_0^2$ into Ω and $R(P_0)$ is expressed as follows:

$$\hat{\hat{\sigma}}_0^2(n-m) = \Omega + R(P_2); \;\; \text{with} \;\; \Omega = \hat{\sigma}_0^2(n_1 - m) \tag{7.8a}$$

and where

$$R(P_2) = -\big(\boldsymbol{y}_2 - A_2\hat{\boldsymbol{\xi}}\big)^T\hat{\hat{\boldsymbol{\lambda}}} \;\; \text{with} \;\; \hat{\hat{\boldsymbol{\lambda}}} \coloneqq -\big(P_2^{-1} + A_2N_{11}^{-1}A_2^T\big)^{-1}\big(\boldsymbol{y}_2 - A_2\hat{\boldsymbol{\xi}}\big). \tag{7.8b}$$

Therefore, we can rewrite (7.8a) as

$$\boxed{\hat{\hat{\sigma}}_0^2(n-m) = \Omega + \big(\boldsymbol{y}_2 - A_2\hat{\boldsymbol{\xi}}\big)^T\big(P_2^{-1} + A_2N_{11}^{-1}A_2^T\big)^{-1}\big(\boldsymbol{y}_2 - A_2\hat{\boldsymbol{\xi}}\big),} \tag{7.8c}$$

where the form of $R(P_2)$ is obviously similar to that of $R(P_0)$ in (6.39).

Then, the test statistic

$$T = \frac{R/n_2}{\Omega/(n_1 - m)} \sim F(n_2, n_1 - m) \tag{7.9}$$

can be computed to verify the sequential adjustment, i.e., that both the first data set and the combined first and second data sets fit the model well. The test statistic has an F-distribution with n_2 and $n_1 - m$ degrees of freedom. For some specified significance level α, we may claim that the observations from the second data set are consistent with those from the first if $T \leq F_{\alpha,n_2,n_1-m}$. See Chapter 9 for more on hypothesis testing.

7.3 Alternative Solution for the Normal Equations

Using the *addition theory of normal equations*, we may find a matrix representation of the normal equations as follows, where again the double hats above $\boldsymbol{\xi}$ refer to a solution based on both data sets:

$$\big(A_1^TP_1A_1 + A_2^TP_2A_2\big)\hat{\hat{\boldsymbol{\xi}}} = \big(A_1^TP_1\boldsymbol{y}_1 + A_2^TP_2\boldsymbol{y}_2\big), \tag{7.10a}$$

or

$$\big(N_{11} + N_{22}\big)\hat{\hat{\boldsymbol{\xi}}} = \big(\boldsymbol{c}_1 + \boldsymbol{c}_2\big). \tag{7.10b}$$

These normal equations lead to

$$N_{11}\hat{\hat{\boldsymbol{\xi}}} + N_{22}\hat{\hat{\boldsymbol{\xi}}} - \boldsymbol{c}_2 = \boldsymbol{c}_1 \Rightarrow \tag{7.11a}$$

$$N_{11}\hat{\hat{\boldsymbol{\xi}}} + A_2^T\hat{\hat{\boldsymbol{\lambda}}}_2 = \boldsymbol{c}_1, \text{ with } \hat{\hat{\boldsymbol{\lambda}}} = P_2\big(A_2\hat{\hat{\boldsymbol{\xi}}} - \boldsymbol{y}_2\big) \Rightarrow \tag{7.11b}$$

$$\boldsymbol{y}_2 = A_2\hat{\hat{\boldsymbol{\xi}}} - P_2^{-1}\hat{\hat{\boldsymbol{\lambda}}}. \tag{7.11c}$$

Then, from (7.11b) and (7.11c), we can write the following system of least-squares normal equations:

$$\begin{bmatrix} N_{11} & A_2^T \\ A_2 & -P_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}} \\ \hat{\hat{\boldsymbol{\lambda}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{y}_2 \end{bmatrix}. \tag{7.12}$$

From the first row of (7.12) we get

$$\hat{\hat{\boldsymbol{\xi}}} = N_{11}^{-1}\boldsymbol{c}_1 - N_{11}^{-1}A_2^T\hat{\hat{\boldsymbol{\lambda}}} = \tag{7.13a}$$

$$= \hat{\boldsymbol{\xi}} - N_{11}^{-1}A_2^T\hat{\hat{\boldsymbol{\lambda}}}. \tag{7.13b}$$

Equation (7.13b) is an update formula as a function of the vector of estimated Lagrange multipliers $\hat{\hat{\boldsymbol{\lambda}}}$. Without further derivation, we can compare (7.13b) to (7.5) to get an expression for the estimated vector of Lagrange-multiplier as

$$\hat{\hat{\boldsymbol{\lambda}}} = -\big(P_2^{-1} + A_2N_{11}^{-1}A_2^T\big)^{-1}\big(\boldsymbol{y}_2 - A_2\hat{\boldsymbol{\xi}}\big), \tag{7.14}$$

which agrees with (7.8b). Applying covariance propagation to (7.13b), we find the dispersion matrix of $\hat{\hat{\boldsymbol{\xi}}}$ to be

$$\boxed{D\{\hat{\hat{\boldsymbol{\xi}}}\} = D\{\hat{\boldsymbol{\xi}}\} - \sigma_0^2N_{11}^{-1}A_2^T\big(P_2^{-1} + A_2N_{11}^{-1}A_2^T\big)^{-1}A_2N_{11}^{-1},} \tag{7.15}$$

where we used the fact that $C\{\boldsymbol{y}_2, \hat{\boldsymbol{\xi}}\} = 0$, which indicates that the observations from the second data set are uncorrelated with the estimated parameters based on the first data set only.

7.4 Sequential Adjustment, Rank-Deficient Case

7.4.1 First Data Set Only

Suppose matrix A_1 does not have full column rank, i.e. $\operatorname{rk} A_1 =: q_1 < m$. Then we may introduce a datum by further splitting the system of equations as was done in Section 4.5. Let us split A_1 into an $n_1 \times q_1$ part denoted A_{11} and an $n_1 \times (m - q_1)$

part denoted A_{12}. Accordingly, we also split the parameter vector $\boldsymbol{\xi}$ into a $q_1 \times 1$ part $\boldsymbol{\xi}_1$ and a $(m - q_1) \times 1$ part $\boldsymbol{\xi}_2$. Thus, we have

$$A_1 = [A_{11},\, A_{12}]\,, \;\; \operatorname{rk} A_{11} =: q_1, \;\; \text{and} \;\; \boldsymbol{\xi} = \left[\boldsymbol{\xi}_1^T,\, \boldsymbol{\xi}_2^T\right]^T. \tag{7.16a}$$

With this splitting, we introduce new terms for the normal equations, where it is stressed that the subscripts on matrix N and vector $\boldsymbol{c}$ are used differently than in the preceding sections. Most notably, the subscript 2 does not refer to the second data set, but only to the location of a block in the matrix N.

The terms N_{11}, N_{12}, and $\boldsymbol{c}_1$ are defined as follows:

$$\begin{bmatrix} A_{11}^T \\ A_{12}^T \end{bmatrix} P_1 \begin{bmatrix} A_{11} & A_{12} \end{bmatrix} = \left[\begin{array}{c|c} A_{11}^T P_1 A_{11} & A_{11}^T P_1 A_{12} \\ \hline A_{12}^T P_1 A_{11} & A_{12}^T P_1 A_{12} \end{array}\right] = \left[\begin{array}{c|c} N_{11} & N_{12} \\ \hline A_{12}^T P_1 A_{11} & A_{12}^T P_1 A_{12} \end{array}\right], \tag{7.16b}$$

and

$$\boldsymbol{c}_1 = A_{11}^T P_1 \boldsymbol{y}_1. \tag{7.16c}$$

Note that we purposely did not use symbols N_{21} and $\boldsymbol{c}_2$ here, because they will be defined in a different way in the next section.

Next we introduce datum information $\boldsymbol{\xi}_2^0$, such that $\boldsymbol{\xi}_2 \to \boldsymbol{\xi}_2^0$, where the subscript 2 now obviously refers to the datum, rather than a second data set. The formulas for the estimated parameters and their dispersion matrix based on the first data set only can be copied from (3.41b) and (3.42), respectively.

$$\boxed{\hat{\boldsymbol{\xi}}_1 = N_{11}^{-1}\big(\boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0\big)} \tag{7.17a}$$

$$\boxed{D\{\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2 N_{11}^{-1}} \tag{7.17b}$$

The estimated variance component $\hat{\sigma}_0^2$ is slightly different from that of (3.47) and (3.50) and is given by the formula

$$\boxed{\hat{\sigma}_0^2 = \frac{\boldsymbol{y}_1^T P_1 \big(\boldsymbol{y}_1 - A_{11}\hat{\boldsymbol{\xi}}_1 - A_{12}\boldsymbol{\xi}_2^0\big)}{(n_1 - q_1)}} \tag{7.17c}$$

or, equivalently,

$$\boxed{\hat{\sigma}_0^2 = \frac{\big(\boldsymbol{y}_1^T P_1 \boldsymbol{y}_1 - \boldsymbol{c}_1^T N_{11}^{-1} \boldsymbol{c}_1\big)}{(n_1 - q_1)}.} \tag{7.17d}$$

Note that the steps taken from (3.47) to (3.50) can be used to go from (7.17c) to (7.17d).

7.4.2 Both First and Second Data Sets

Now we introduce the second data set with a splitting analogous to the first, viz.

$$\boldsymbol{y}_2 = A_{21}\boldsymbol{\xi}_1 + A_{22}\boldsymbol{\xi}_2 + \boldsymbol{e}_2, \;\; \boldsymbol{e}_2 \sim \big(\boldsymbol{0}, \sigma_0^2 P_2^{-1}\big). \tag{7.18}$$

The matrix A_{21} is of size $n_2 \times q_1$, and A_{22} is of size $n_2 \times (m - q_1)$. No information in the second data set refers to the datum choice; it only adds to the redundancy provided by the first data set. Thus, the rank of the normal equation matrix is unchanged, which is true also for the 2×2-block coefficient matrix, i.e.,

$$\operatorname{rk} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} =: q = q_1. \tag{7.19}$$

The full least-squares normal equations are then written as

$$\left[\begin{array}{c|c} A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} & A_{11}^T P_1 A_{12} + A_{21}^T P_2 A_{22} \\ \hline A_{12}^T P_1 A_{11} + A_{22}^T P_2 A_{21} & A_{12}^T P_1 A_{12} + A_{22}^T P_2 A_{22} \end{array}\right] \begin{bmatrix} \hat{\boldsymbol{\xi}}_1 \\ \boldsymbol{\xi}_2^0 \end{bmatrix} = \\ = \left[\begin{array}{c|c} A_{11}^T P_1 & A_{21}^T P_2 \\ \hline A_{12}^T P_1 & A_{22}^T P_2 \end{array}\right] \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}. \tag{7.20}$$

From the first row of (7.20), we may write the least-squares solution for $\hat{\hat{\boldsymbol{\xi}}}_1$ directly, followed by its dispersion matrix, as

$$\boxed{\begin{aligned} \hat{\hat{\boldsymbol{\xi}}}_1 &= \big(A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21}\big)^{-1} \cdot \\ &\quad \cdot \big[\big(A_{11}^T P_1 \boldsymbol{y}_1 + A_{21}^T P_2 \boldsymbol{y}_2\big) - \big(A_{11}^T P_1 A_{12} + A_{21}^T P_2 A_{22}\big)\boldsymbol{\xi}_2^0\big], \end{aligned} \tag{7.21}$$

$$D\{\hat{\hat{\boldsymbol{\xi}}}_1\} = \sigma_0^2 \big(A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21}\big)^{-1}. \tag{7.22}$$

In order to derive update formulas, it is helpful to introduce an alternative expression for the normal equations analogous to what was done in (7.11a) through (7.12). From (7.17a), we can write

$$\big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 = \big(A_{11}^T P_1 \boldsymbol{y}_1\big) - \big(A_{11}^T P_1 A_{12}\big)\boldsymbol{\xi}_2^0, \tag{7.23a}$$

$$\text{or} \quad N_{11}\hat{\boldsymbol{\xi}}_1 = \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0, \tag{7.23b}$$

which, when subtracted from the first row of (7.20), leaves

$$\big(A_{21}^T P_2 A_{21}\big)\hat{\boldsymbol{\xi}}_1 = \big(A_{21}^T P_2 \boldsymbol{y}_2\big) - \big(A_{21}^T P_2 A_{22}\big)\boldsymbol{\xi}_2^0, \tag{7.23c}$$

$$\text{or} \quad N_{21}\hat{\boldsymbol{\xi}}_1 = \boldsymbol{c}_2 - N_{22}\boldsymbol{\xi}_2^0. \tag{7.23d}$$

Note that the symbols N_{11} and N_{12} are still being used as defined in (7.16b), whereas the definition of N_{22} and N_{21} becomes apparent by comparing (7.23c) to (7.23d).

Together, (7.23b) and (7.23d) comprise the first row of (7.20). Recombining (7.23b) and (7.23d) gives

$$\big(N_{11} + N_{21}\big)\hat{\hat{\boldsymbol{\xi}}}_1 = \boldsymbol{c}_1 + \boldsymbol{c}_2 - \big(N_{12} + N_{22}\big)\boldsymbol{\xi}_2^0, \tag{7.24a}$$

implying that

$$N_{11}\hat{\hat{\boldsymbol{\xi}}}_1 + A_{21}^T\hat{\hat{\boldsymbol{\lambda}}} = \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0, \text{ with } \hat{\hat{\boldsymbol{\lambda}}} \coloneqq P_2\big(A_{21}\hat{\hat{\boldsymbol{\xi}}}_1 - \boldsymbol{y}_2 + A_{22}\boldsymbol{\xi}_2^0\big). \tag{7.24b}$$

Note that in (7.23a)–(7.23d) a single hat was used for the estimate of $\boldsymbol{\xi}_1$ since each respective equation represents only one set of data. The double hat in (7.24a) denotes the estimate of $\boldsymbol{\xi}_1$ based on both data sets. From (7.24b) we can write the system of normal equations in matrix form as follows:

$$\begin{bmatrix} N_{11} & A_{21}^T \\ A_{21} & -P_2^{-1} \end{bmatrix}\begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 \\ \hat{\hat{\boldsymbol{\lambda}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \\ \boldsymbol{y}_2 - A_{22}\boldsymbol{\xi}_2^0 \end{bmatrix}. \tag{7.25}$$

The solution of (7.25) can be obtained by applying the inversion formula for a partitioned matrix as shown in (A.11), resulting in

$$\begin{aligned} \begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 \\ \hat{\hat{\boldsymbol{\lambda}}} \end{bmatrix} &= \begin{bmatrix} N_{11} & A_{21}^T \\ A_{21} & -P_2^{-1} \end{bmatrix}^{-1}\begin{bmatrix} \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \\ \boldsymbol{y}_2 - A_{22}\boldsymbol{\xi}_2^0 \end{bmatrix} = \\ &= \left[\begin{array}{c|c} N_{11}^{-1} - N_{11}^{-1}A_{21}^TS_2^{-1}A_{21}N_{11}^{-1} & N_{11}^{-1}A_{21}^TS_2^{-1} \\ \hline S_2^{-1}A_{21}N_{11}^{-1} & -S_2^{-1} \end{array}\right]\begin{bmatrix} \boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0 \\ \boldsymbol{y}_2 - A_{22}\boldsymbol{\xi}_2^0 \end{bmatrix}, \end{aligned} \tag{7.26}$$

where

$$S_2 \coloneqq P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T. \tag{7.27}$$

Finally, the estimated parameters and Lagrange multipliers are expressed as

$$\begin{aligned} \hat{\hat{\boldsymbol{\xi}}}_1 &= N_{11}^{-1}\big(\boldsymbol{c}_1 - N_{12}\boldsymbol{\xi}_2^0\big) + \\ &+ N_{11}^{-1}A_{21}^T\big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big)^{-1}\big[A_{21}N_{11}^{-1}\big(-\boldsymbol{c}_1 + N_{12}\boldsymbol{\xi}_2^0\big) + \boldsymbol{y}_2 - A_{22}\boldsymbol{\xi}_2^0\big] \Rightarrow \end{aligned} \tag{7.28a}$$

$$\boxed{\hat{\hat{\boldsymbol{\xi}}}_1 = \hat{\boldsymbol{\xi}}_1 + N_{11}^{-1}A_{21}^T\big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big)^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\big),} \tag{7.28b}$$

$$\hat{\hat{\boldsymbol{\lambda}}} = -\big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big)^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\big). \tag{7.28c}$$

The dispersion matrix of the estimated vector of Lagrange multipliers is

$$D\{\hat{\hat{\boldsymbol{\lambda}}}\} = \big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big)^{-1}D\{\boldsymbol{y} - A_{21}\hat{\boldsymbol{\xi}}_1\}\big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big)^{-1}, \tag{7.29}$$

since $D\{\boldsymbol{\xi}_2^0\} = 0$. The following relations also hold:

$$C\{\boldsymbol{y}_2, \hat{\boldsymbol{\xi}}_1\} = 0, \tag{7.30a}$$

$$D\{\boldsymbol{y} - A_{21}\hat{\boldsymbol{\xi}}_1\} = \sigma_0^2\big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big), \tag{7.30b}$$

$$D\{\hat{\hat{\boldsymbol{\lambda}}}\} = \sigma_0^2\big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big)^{-1}, \tag{7.30c}$$

$$D\{\hat{\hat{\boldsymbol{\xi}}}_1\} = D\{\hat{\boldsymbol{\xi}}_1\} - \sigma_0^2N_{11}^{-1}A_{21}^T\big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big)^{-1}A_{21}N_{11}^{-1}. \tag{7.30d}$$

The estimated variance component is expressed as follows:

$$\hat{\sigma}_0^2(n-q) = \hat{\sigma}_0^2\big(n_1-q_1\big) + \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\big)^T \cdot \\ \cdot \big(P_2^{-1} + A_{21}N_{11}^{-1}A_{21}^T\big)^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\big) \Rightarrow \quad (7.31a)$$

$$\boxed{\hat{\sigma}_0^2(n-q) = \hat{\sigma}_0^2\big(n_1-q_1\big) - \hat{\hat{\boldsymbol{\lambda}}}^T\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1 - A_{22}\boldsymbol{\xi}_2^0\big).} \quad (7.31b)$$

Once again, we note that we have used the definition $N_{11} \coloneqq A_{11}^T P_1 A_{11}$ in this section.

7.5 Sequential Adjustment with New Parameters

In this section we consider the case where the second data set refers to all the parameters of the first data set plus some additional new parameters. Thus we speak of m_1 parameters associated with the first data set and an additional m_2 introduced with the second data set, so that the total number of parameters involved in the combination of both data sets is given by $m = m_1 + m_2$. In the double subscripts used below, the first one refers to the data set, while the second one refers to the matrix splitting. For example, A_{21} is that part of the design matrix from the second data set that refers to the original parameters, whereas A_{22} is associated with the new parameters involved in the second data set. We could have adopted a new symbol to denote a "preprocessed observation" vector that includes datum information, e.g. $\bar{\boldsymbol{y}}$. However, we have elected to continue using $\boldsymbol{y}$ and simply note that it could include datum information in addition to the observations. The data model that follows implies that we have assumed there are no correlations between the observations of data-set one and those of data-set two; it also implies that both sets of observations share a common variance component σ_0^2.

$$\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix} = \begin{bmatrix}A_{11} & 0\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix}, \quad \begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix} \sim \Big(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2\begin{bmatrix}P_1^{-1} & 0\\ 0 & P_2^{-1}\end{bmatrix}\Big) \quad (7.32)$$

The size of the system of equations is implied by the following:

$$\boldsymbol{y}_1 \in \mathbb{R}^{n_1},\ \ \boldsymbol{y}_2 \in \mathbb{R}^{n_2},\ \ \boldsymbol{\xi}_1 \in \mathbb{R}^{m_1},\ \ \boldsymbol{\xi}_2 \in \mathbb{R}^{m_2},\ \ \big[\boldsymbol{\xi}_1^T, \boldsymbol{\xi}_2^T\big]^T \in \mathbb{R}^m, \quad (7.33a)$$

$$n = n_1 + n_2,\ \ m = m_1 + m_2. \quad (7.33b)$$

Now, using the addition theory of normal equations, we can write

$$\begin{bmatrix}A_{11}^T & A_{21}^T\\ 0 & A_{22}^T\end{bmatrix}\begin{bmatrix}P_1 & 0\\ 0 & P_2\end{bmatrix}\begin{bmatrix}A_{11} & 0\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\hat{\hat{\boldsymbol{\xi}}}_1\\ \hat{\hat{\boldsymbol{\xi}}}_2\end{bmatrix} = \begin{bmatrix}A_{11}^T P_1 & A_{21}^T P_2\\ 0 & A_{22}^T P_2\end{bmatrix}\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix} \Rightarrow \quad (7.34a)$$

$$\left[\begin{array}{c|c} A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} & A_{21}^T P_2 A_{22}\\ \hline A_{22}^T P_2 A_{21} & A_{22}^T P_2 A_{22}\end{array}\right]\begin{bmatrix}\hat{\hat{\boldsymbol{\xi}}}_1\\ \hat{\hat{\boldsymbol{\xi}}}_2\end{bmatrix} = \begin{bmatrix}A_{11}^T P_1 \boldsymbol{y}_1 + A_{21}^T P_2 \boldsymbol{y}_2\\ A_{22}^T P_2 \boldsymbol{y}_2\end{bmatrix}. \quad (7.34b)$$

Here again, the double-hats refer to estimates based on both data sets.

Now, the first data set may no longer be available, rather we may have only the estimates from the first adjustment. In this case we can use the bottom row of (7.34b) to solve for the estimates of the new parameters in terms of only the second set of observations, leading to

$$\boxed{\hat{\hat{\boldsymbol{\xi}}}_2 = \big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 \big(\boldsymbol{y}_2 - A_{21}\hat{\hat{\boldsymbol{\xi}}}_1\big).} \tag{7.35}$$

Then, from the normal equations based solely on the first data set, we may substitute

$$A_{11}^T P_1 \boldsymbol{y}_1 = \big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 \tag{7.36}$$

into the top row of the right side of (7.34b) and invert the normal-equation matrix on the left to solve for the parameter estimates. For convenience, we introduce the following symbols to use in the inverted matrix:

$$S_1 \coloneqq A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} - A_{21}^T P_2 A_{22}\big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 A_{21} = \tag{7.37a}$$

$$= A_{11}^T P_1 A_{11} + A_{21}^T \bar{P}_2 A_{21}, \tag{7.37b}$$

$$\bar{P}_2 \coloneqq P_2 - P_2 A_{22}\big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2, \tag{7.37c}$$

$$N_{22} = A_{22}^T P_2 A_{22}. \tag{7.37d}$$

We refer to $\bar{P}_2$ as a *reduced weight matrix*. Upon inverting the normal-equations matrix from (7.34b) (see (A.11) for the inverse of a partitioned matrix), we find the following solution for $\hat{\hat{\boldsymbol{\xi}}}_1$ and $\hat{\hat{\boldsymbol{\xi}}}_2$:

$$\begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 \\ \hat{\hat{\boldsymbol{\xi}}}_2 \end{bmatrix} = \left[\begin{array}{c|c} S_1^{-1} & -S_1^{-1}\big(A_{21}^T P_2 A_{22}\big) N_{22}^{-1} \\ \hline -N_{22}^{-1}\big(A_{22}^T P_2 A_{21}\big) S_1^{-1} & N_{22}^{-1} + N_{22}^{-1}\big(A_{22}^T P_2 A_{21}\big) S_1^{-1}\big(A_{21}^T P_2 A_{22}\big) N_{22}^{-1} \end{array}\right] \cdot$$

$$\cdot \begin{bmatrix} \big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 + A_{21}^T P_2 \boldsymbol{y}_2 \\ A_{22}^T P_2 \boldsymbol{y}_2 \end{bmatrix}. \tag{7.38}$$

We can continue by using (7.37b) and (7.37c) with the first row of (7.38) to arrive at

$$\hat{\hat{\boldsymbol{\xi}}}_1 = S_1^{-1}\big[\big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 + A_{21}^T P_2 \boldsymbol{y}_2 - \big(A_{21}^T P_2 A_{22}\big) N_{22}^{-1} A_{22}^T P_2 \boldsymbol{y}_2\big] = \tag{7.39a}$$

$$= S_1^{-1}\big\{\big[\big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 + A_{21}^T \bar{P}_2 \boldsymbol{y}_2\big] + \big[\big(A_{21}^T \bar{P}_2 A_{21}\big) - \big(A_{21}^T \bar{P}_2 A_{21}\big)\big]\hat{\boldsymbol{\xi}}_1\big\} = \tag{7.39b}$$

$$= S_1^{-1} A_{21}^T \bar{P}_2 \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) + S_1^{-1}\big(A_{11}^T P_1 A_{11} + A_{21}^T \bar{P}_2 A_{21}\big)\hat{\boldsymbol{\xi}}_1 = \tag{7.39c}$$

$$= S_1^{-1} A_{21}^T \bar{P}_2 \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) + \hat{\boldsymbol{\xi}}_1 \Rightarrow \tag{7.39d}$$

$$\boxed{\hat{\hat{\boldsymbol{\xi}}}_1 - \hat{\boldsymbol{\xi}}_1 = S_1^{-1} A_{21}^T \bar{P}_2 \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big),} \tag{7.39e}$$

where (7.39e) is in the form of an *update formula*.

We assume that P_2 is invertible, as implied in the given model (7.32). We now wish to check the rank of the reduced weight matrix $\bar{P}_2$. It is easy to check that the product $P_2^{-1}\bar{P}_2$ is idempotent. Then using (1.7c) and (A.4) we find

$$\operatorname{rk}\bar{P}_2 = \operatorname{rk}\big(P_2^{-1}\bar{P}_2\big) = \operatorname{tr}\big(P_2^{-1}\bar{P}_2\big) = \operatorname{tr}\big(I_{n_2} - A_{22}\big(A_{22}^T P_2 A_{22}\big)^{-1}A_{22}^T P_2\big) = \tag{7.40a}$$

$$= n_2 - \operatorname{tr}\big[A_{22}\big(A_{22}^T P_2 A_{22}\big)^{-1}A_{22}^T P_2\big] = n_2 - \operatorname{tr}\big[\big(A_{22}^T P_2 A_{22}\big)^{-1}A_{22}^T P_2 A_{22}\big] = \tag{7.40b}$$

$$= n_2 - m_2 < n_2. \tag{7.40c}$$

Thus there is a rank reduction that comes from modifying the original weight matrix P_2 to obtain $\bar{P}_2$. Moreover, we find that matrix $\bar{P}_2$ is singular.

The dispersion matrices for the estimated parameters, i.e., $D\{\hat{\hat{\boldsymbol{\xi}}}_1\}$ and $D\{\hat{\hat{\boldsymbol{\xi}}}_2\}$, are shown at the end of the next section.

7.6 Sequential Adjustment with New Parameters and Small Second Data Set

In (7.39e) we must invert the $m_1 \times m_1$ matrix S_1 to solve the system of equations. However, in some applications, the number of observations n_2 in the second data set may be significantly less than m_1. In this case we would like to reformulate the solution in (7.39e) so that only a matrix of size $n_2 \times n_2$ needs to be inverted.

We have an alternative expression for matrix S_1 in (7.37b), the inverse of which can be derived as follows:

$$S_1^{-1} = \big[\big(A_{11}^T P_1 A_{11}\big) + \big(A_{21}^T \bar{P}_2 A_{21}\big)\big]^{-1} = \tag{7.41a}$$

$$= \big\{\big[I_{m_1} + \big(A_{21}^T \bar{P}_2 A_{21}\big)\big(A_{11}^T P_1 A_{11}\big)^{-1}\big]\big(A_{11}^T P_1 A_{11}\big)\big\}^{-1} = \tag{7.41b}$$

$$= \big(A_{11}^T P_1 A_{11}\big)^{-1}\big[I_{m_1} + \big(A_{21}^T \bar{P}_2 A_{21}\big)\big(A_{11}^T P_1 A_{11}\big)^{-1}\big]^{-1}. \tag{7.41c}$$

Using (7.41c), we may rewrite (7.39e) as

$$\hat{\hat{\boldsymbol{\xi}}}_1 - \hat{\boldsymbol{\xi}}_1 =$$

$$= \big(A_{11}^T P_1 A_{11}\big)^{-1}\big[I_{m_1} + \big(A_{21}^T \bar{P}_2 A_{21}\big)\big(A_{11}^T P_1 A_{11}\big)^{-1}\big]^{-1} A_{21}^T \bar{P}_2\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) = \tag{7.42a}$$

$$= \big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big). \tag{7.42b}$$

Here, we have made use of (A.8a) in the step from (7.42a) to (7.42b), with matrices A and D in (A.8a) set to identity. Note that the matrix to invert inside the square brackets is of size $m_1 \times m_1$ in (7.42a) but is size $n_2 \times n_2$ in (7.42b). The choice of which equation to use will usually be determined by the smaller of m_1 and n_2. Also, we have the relation

$$-\hat{\hat{\boldsymbol{\lambda}}} = \big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big), \tag{7.43}$$

which means that the solution for the first subset of parameters may also be expressed as

$$\boxed{\hat{\hat{\boldsymbol{\xi}}}_1 - \hat{\boldsymbol{\xi}}_1 = -\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2 \hat{\hat{\boldsymbol{\lambda}}}.} \tag{7.44}$$

Now we begin with (7.35), and substitute (7.42b), to find a solution for the parameters $\hat{\hat{\boldsymbol{\xi}}}_2$ in terms of the Lagrange multipliers $\hat{\hat{\boldsymbol{\lambda}}}$:

$$\hat{\hat{\boldsymbol{\xi}}}_2 = \big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 \big(\boldsymbol{y}_2 - A_{21}\hat{\hat{\boldsymbol{\xi}}}_1\big) = \tag{7.45a}$$

$$= \big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 \cdot \big\{\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) - A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2 \cdot \cdot \big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big)\big\} = \tag{7.45b}$$

$$= \big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 \big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1}\big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) \tag{7.45c}$$

$$\boxed{\Rightarrow \hat{\hat{\boldsymbol{\xi}}}_2 = -\big(A_{22}^T P_2 A_{22}\big)^{-1} A_{22}^T P_2 \hat{\hat{\boldsymbol{\lambda}}}.} \tag{7.45d}$$

The inverse formula of (A.6a) was used to go from (7.45b) to (7.45c), with matrices T, W, and V in (A.6a) set to identity matrices of appropriate sizes.

To facilitate computing the parameter dispersion matrix we write the following system of normal equations, noting that (7.46b) is in the form of an update solution (cf. (7.34b) and (7.36)):

$$\left[\begin{array}{c|c} A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} & A_{21}^T P_2 A_{22} \\ \hline A_{22}^T P_2 A_{21} & A_{22}^T P_2 A_{22} \end{array}\right] \begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 \\ \hat{\hat{\boldsymbol{\xi}}}_2 \end{bmatrix} = \begin{bmatrix} \big(A_{11}^T P_1 A_{11}\big)\hat{\boldsymbol{\xi}}_1 + A_{21}^T P_2 \boldsymbol{y}_2 \\ A_{22}^T P_2 \boldsymbol{y}_2 \end{bmatrix}. \tag{7.46a}$$

Then substituting (7.35) leads to

$$\left[\begin{array}{c|c} A_{11}^T P_1 A_{11} + A_{21}^T P_2 A_{21} & A_{21}^T P_2 A_{22} \\ \hline A_{22}^T P_2 A_{21} & A_{22}^T P_2 A_{22} \end{array}\right] \begin{bmatrix} \hat{\hat{\boldsymbol{\xi}}}_1 - \hat{\boldsymbol{\xi}}_1 \\ \hat{\hat{\boldsymbol{\xi}}}_2 \end{bmatrix} = \begin{bmatrix} A_{21}^T P_2 \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) \\ A_{22}^T P_2 \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big) \end{bmatrix}. \tag{7.46b}$$

Note that (7.46a) is equivalent to (7.34b) shown earlier.

We have already inverted the normal-equation matrix in (7.38). Taking elements from (7.38), we may write the parameter dispersion and covariance matrices as follows:

$$\boxed{\begin{aligned} D\{\hat{\hat{\boldsymbol{\xi}}}_1\} &= \sigma_0^2 S_1^{-1} = \sigma_0^2 \big(A_{11}^T P_1 A_{11} + A_{21}^T \bar{P}_2 A_{21}\big)^{-1}, && (7.47a)\\ C\{\hat{\hat{\boldsymbol{\xi}}}_1, \hat{\hat{\boldsymbol{\xi}}}_2\} &= -D\{\hat{\hat{\boldsymbol{\xi}}}_1\}\big(A_{21}^T P_2 A_{22}\big)\big(A_{22}^T P_2 A_{22}\big)^{-1}, && (7.47b)\\ D\{\hat{\hat{\boldsymbol{\xi}}}_2\} &= \sigma_0^2 \big(A_{22}^T P_2 A_{22}\big)^{-1} - \big(A_{22}^T P_2 A_{22}\big)^{-1}\big(A_{22}^T P_2 A_{21}\big) C\{\hat{\hat{\boldsymbol{\xi}}}_1, \hat{\hat{\boldsymbol{\xi}}}_2\}. && (7.47c) \end{aligned}}$$

Each of the above covariance matrices (7.47a) through (7.47c) include the matrix S_1^{-1}, which implies that a matrix of size $m_1 \times m_1$ must be inverted. However,

with the insertion of I_{n_2} into (7.47a), and with appropriate matrix groupings, we may apply the inversion formula (A.6a) to find an inverse of smaller dimension as shown in the following:

$$D\{\hat{\hat{\boldsymbol{\xi}}}_1\} = \sigma_0^2\big[\big(A_{11}^T P_1 A_{11}\big) + \big(A_{21}^T \bar{P}_2\big) I_{n_2} A_{21}\big]^{-1} = \tag{7.48a}$$

$$= \sigma_0^2 N_{11}^{-1} - \sigma_0^2 N_{11}^{-1} A_{21}^T \bar{P}_2 \big(I_{n_2} + A_{21} N_{11}^{-1} A_{21}^T \bar{P}_2\big)^{-1} A_{21} N_{11}^{-1}. \tag{7.48b}$$

Here again, we have used $N_{11} \coloneqq A_{11}^T P_1 A_{11}$ for compactness. The parenthetical term that must be inverted in equation (7.48b) is an $n_2 \times n_2$ matrix, which, again, may be much smaller than an $m_1 \times m_1$ matrix, depending on the application. Of course, the matrix $(A_{11}^T P_1 A_{11})^{-1}$ is also size $m_1 \times m_1$, but it is assumed that this inverse had already been performed in the adjustment of the first data set and was saved for subsequent use.

The estimated variance component is expressed as

$$\hat{\hat{\sigma}}_0^2\big(n-m\big) = \hat{\sigma}_0^2\big(n_1 - m_1\big) - \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big)^T \bar{P}_2 \hat{\hat{\boldsymbol{\lambda}}}. \tag{7.49a}$$

Then, substituting (7.43) results in

$$\hat{\hat{\sigma}}_0^2\big(n-m\big) = \hat{\sigma}_0^2\big(n_1 - m_1\big) + \\ + \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big)^T \bar{P}_2 \big[I_{n_2} + A_{21}\big(A_{11}^T P_1 A_{11}\big)^{-1} A_{21}^T \bar{P}_2\big]^{-1} \big(\boldsymbol{y}_2 - A_{21}\hat{\boldsymbol{\xi}}_1\big). \tag{7.49b}$$

7.7 Practice Problems

1. Considering Problem 9 of Section 3.6, assume that a second observation campaign has been conducted, where the original observation scheme was repeated, except that the final three observations from the first campaign were not repeated in the second one. Both data sets are listed in Table 7.1, and a diagram of the leveling network is shown in Figure 3.4. Furthermore, assume that the weight of each observation in both data sets is defined as the distance in miles associated with the observation divided by 100. Introduce datum information so that the height of point D is fixed at 1928.277 ft.

   (a) Compute estimates for $\hat{\boldsymbol{\xi}}$, along with its cofactor matrix, and the estimated variance component $\hat{\sigma}_0^2$ based only on the first data set.

   (b) Using the results of the previous step, compute estimates for $\hat{\hat{\boldsymbol{\xi}}}$, $D\{\hat{\hat{\boldsymbol{\xi}}}\}$, and the estimated variance component $\hat{\hat{\sigma}}_0^2$ using update formulas that do not directly depend on the observations from the first data set.

Table 7.1: Leveling data from Rainsford (1968) as $\boldsymbol{y}_I$ and simulated second data set as $\boldsymbol{y}_{II}$. d stands for distance between stations.

From	To	No.	$\boldsymbol{y}_I$ [ft]	$\boldsymbol{y}_{II}$, [ft]	d [miles]
A	B	1	+124.632	+124.659	68
B	C	2	+217.168	+217.260	40
C	D	3	−92.791	−92.904	56
A	D	4	+248.754	+248.797	171
A	F	5	−11.418	−11.402	76
F	E	6	−161.107	−161.172	105
E	D	7	+421.234		80
B	F	8	−135.876		42
C	E	9	−513.895		66

2. Now consider the case where one new station, G, was added to the network during the second observation campaign as depicted in Figure 7.1. The data for the first observation campaign can be taken from Table 7.1. The data from the second observation campaign are listed in Table 7.2.

   Use (7.39e) followed by (7.35) to compute $\hat{\hat{\boldsymbol{\xi}}}_1$ and $\hat{\hat{\boldsymbol{\xi}}}_2$, respectively, or, instead, use (7.42b) for $\hat{\hat{\boldsymbol{\xi}}}_1$.

   Hint: Because of the network datum deficiency of one, you can modify the observation vector from the second data set by subtracting out the datum value of 1928.277 ft from observations y_3 and y_7 and then remove the parameter for the height of station D from the parameter vector. You may check your answers by combining both data sets into one, and then solve for the unknown parameters according to Section 3.5.

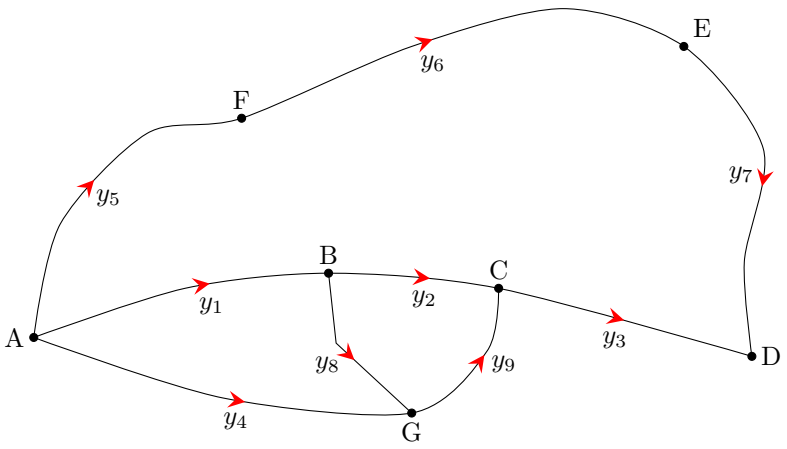

Figure 7.1: Simulated extension of a leveling network by Rainsford (1968)

Table 7.2: Leveling data for simulated second data set as $\boldsymbol{y}_{II}$. d stands for distance between stations.

From	To	No.	$\boldsymbol{y}_{II}$, [ft]	d [miles]
A	B	1	+124.659	68
B	C	2	+217.260	40
C	D	3	−92.904	56
A	G	4	+178.852	85
A	F	5	−11.402	76
F	E	6	−161.172	105
E	D	7	+421.212	80
B	G	8	+54.113	45
G	C	9	+162.992	45

Chapter 8

Condition Equations with Parameters: the Gauss-Helmert Model

8.1 Model Definition

Data models introduced prior to this chapter have either admitted observation equations with unknown parameters *or* condition equations without parameters, but not both. In contrast, the Gauss-Helmert Model (GHM) allows both condition equations *and* equations involving unknown parameters to be combined in the same model. Thus, the GHM can be viewed as being more flexible (or more general) than either the Gauss-Markov Model (GMM) (Chapter 3) or the Model of Condition Equations (Chapter 4), since it combines aspects of both. In some cases, the GHM might be useful for dealing with complicated observation equations, for example when multiple observations are related functionally to one or more parameters via specified (possibly nonlinear) equations.

In other cases, the LEast-Squares Solution (LESS) within the GHM is equivalent to that of orthogonal regression, or, more generally, to a total least-squares (TLS) solution. Such solutions are sought within models that have both independent and dependent random data variables. Examples are line and curve fitting in 2D when both x- and y-coordinates are measured. Coordinate transformation problems also fall in this category when the coordinates from both the source and target systems are measured quantities. We will learn how to treat these problems in this chapter.

8.1.1 An example Gauss-Helmert Model

We begin our discussion of the GHM with a leveling-network example in order to contrast the GMM with the Model of Condition Equations and to show how the

GHM combines the information used in those two models. The diagram in Figure 8.1 shows a leveling network with four points (P_1, P_2, P_3, P_4) that has been observed in two closed loops comprised of a total of five observations $(y_1, y_2, y_3, y_4, y_5)$. First we present a (rank deficient) partitioned GMM as

$$\boldsymbol{y} = A_1\boldsymbol{\xi}_1 + A_2\boldsymbol{\xi}_2 + \boldsymbol{e}, \tag{8.1a}$$

$$\boldsymbol{e} \sim \big(\mathbf{0}, \sigma_0^2 P^{-1}\big), \tag{8.1b}$$

$$\operatorname{rk} A_1 = \operatorname{rk}\big[A_1 \,|\, A_2\big] =: q < m, \tag{8.1c}$$

where the coefficient matrix A and the vector of unknown parameters $\boldsymbol{\xi}$ have been partitioned, respectively, as

$$A = \left[\begin{array}{c|c} A_1 & A_2 \\ {\scriptstyle n\times q} & {\scriptstyle n\times(m-q)} \end{array}\right] \quad \text{and} \quad \boldsymbol{\xi} = \left[\begin{array}{c|c} \boldsymbol{\xi}_1^T & \boldsymbol{\xi}_2^T \\ {\scriptstyle 1\times q} & {\scriptstyle 1\times(m-q)} \end{array}\right]^T. \tag{8.2}$$

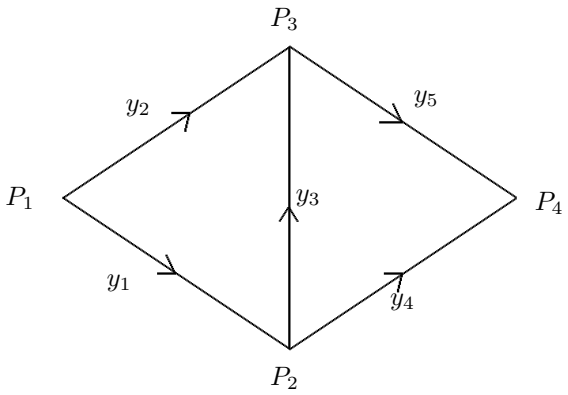

Figure 8.1: Leveling network. Arrows point in the direction of the level runs.

In this example, the number of unknown parameters is $m = 4$ (heights of four points). Since leveled height-differences supply no information about the height datum, we can only estimate the heights of three of the points with respect to the remaining fourth one. That explains why $\operatorname{rk} A =: q = 3 < m$, implying a *datum deficiency* of $m - q = 1$. Thus, the model has been partitioned so that $\boldsymbol{\xi}_1$ contains three estimable heights, and $\boldsymbol{\xi}_2$ is a single non-estimable height, which must be assigned a "datum value." In this example, we arbitrarily chose point P_4 for the non-estimable height. As was stated in Section 3.5, we have the relationship $A_2 = A_1 L$ for some $q \times (m - q)$ matrix L, which means that matrix A_2 is a linear combination of the columns of matrix A_1, reflecting the rank deficiency of matrix $A = [A_1 \,|\, A_2]$.

The problem could also be solved within the Model of Condition Equations introduced in Chapter 4, which reads

$$B\boldsymbol{y} = B\boldsymbol{e}, \;\; \boldsymbol{e} \sim (\mathbf{0}, \sigma_0^2 P^{-1}), \tag{8.3a}$$

with the orthogonality condition

$$\text{i. } B \cdot \Big[A_1 \;\Big|\; A_2 \Big] = 0, \tag{8.3b}$$

and the rank condition

$$\text{ii. } \operatorname{rk} B = r = n - \operatorname{rk} A_1. \tag{8.3c}$$

These two conditions ensure equivalent least-squares solutions within the models of (8.1) and (8.3a) as discussed in Section 4.3.

We have the following design (coefficient) matrices and parameter vectors for the example leveling network, for which it is easy to verify that both conditions i and ii are satisfied:

$$\begin{gathered} A_1 = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad A = \Big[A_1 \;\Big|\; A_2 \Big], \\ B = \begin{bmatrix} 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 \end{bmatrix}, \quad \boldsymbol{\xi}_1 = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}, \quad \boldsymbol{\xi}_2 = \big[h_4 \big], \end{gathered} \tag{8.4a}$$

with

$$q \coloneqq \operatorname{rk} A_1 = \operatorname{rk} A = 3, \;\; r \coloneqq \operatorname{rk} B = 2 = n - \operatorname{rk} A_1 = 5 - 3 = 2, \;\; \text{and} \;\; B \cdot A = 0. \tag{8.4b}$$

Here, h_i represents the height of point P_i.

Now we wish to introduce a new coefficient matrix B that does not contain matrix A in its nullspace, so that we can form a Model of Condition Equations with parameters. For now we use the symbol $\bar{B}$ in order to distinguish it from the coefficient matrix B used in the Model of Condition Equations, which does contain matrix A in its nullspace (i.e., $BA = 0$, but $\bar{B}A \neq 0$). Similarly, we introduce other bar-terms to form the following GHM:

$$\bar{\boldsymbol{y}} = \bar{B}\boldsymbol{y} = \bar{\boldsymbol{w}} = \bar{B}A_1\boldsymbol{\xi}_1 + \bar{B}A_2\boldsymbol{\xi}_2 + \bar{B}\boldsymbol{e}, \tag{8.5a}$$

$$\bar{B}\boldsymbol{e} \sim \big(\mathbf{0}, \sigma_0^2 \bar{B} P^{-1} \bar{B}^T \big), \tag{8.5b}$$

$$\operatorname{rk}\big(\bar{B}\big) =: \bar{r}. \tag{8.5c}$$

The size of $\bar{B}$ is $\bar{r} \times n$, implying that $\bar{B}$ has full row rank. The GHM in (8.5) is equivalent to the GMM in (8.1) if, and only if,

iii. $\bar{B}A_1$ has $n - \bar{r}$ columns of zeros, and

iv. $\operatorname{rk}(\bar{B}A_1) + r = \bar{r} \Leftrightarrow n = \bar{r} + q - \operatorname{rk}(\bar{B}A_1) = \operatorname{rk} \bar{B} + \operatorname{rk} A - \operatorname{rk}(\bar{B}A_1)$

Note that, through the matrix $\bar{B}$, one observation is eliminated for each eliminated parameter. Referring to the level network example, we may wish to eliminate

the height of point P_3 from the parameter list (perhaps it is a temporary benchmark of no particular interest). This can be done by introducing the following example matrix $\bar{B}$:

$$\bar{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \Rightarrow \bar{B}A_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \bar{B}A_1 = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \end{bmatrix}.$$

With these example matrices we have $n = 5$, $r = 2$, $\bar{r} = \operatorname{rk} \bar{B} = 4$, $q = \operatorname{rk} A_1 = 3$, and $\operatorname{rk}(\bar{B}A_1) = 2$. Since $n - \bar{r} = 1$, the single column of zeros in $\bar{B}A_1$ satisfies condition iii. Also, condition iv is satisfied since $n = 5 = \operatorname{rk} \bar{B} + \operatorname{rk} A - \operatorname{rk}(\bar{B}A_1) = 4 + 3 - 2$.

As an aside, we note that it is also possible to remove l estimable parameters via the splitting of the constraint equation introduced in (5.1), i.e.

$$\underset{l \times m}{K} \boldsymbol{\xi} = \begin{bmatrix} K_1, & K_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} \Rightarrow \tag{8.6a}$$

$$\boldsymbol{\xi}_1 = K_1^{-1}\boldsymbol{\kappa}_0 - K_1^{-1}K_2\boldsymbol{\xi}_2. \tag{8.6b}$$

Here, K_1 is a $l \times l$ invertible matrix, and K_2 is of size $l \times (m-l)$. Upon substitution for $\boldsymbol{\xi}_1$ of (8.6b) into (8.1), we find the following modified system of observation equations with l parameters eliminated:

$$\boldsymbol{y} = A_1\boldsymbol{\xi}_1 + A_2\boldsymbol{\xi}_2 + \boldsymbol{e} = A_1K_1^{-1}\boldsymbol{\kappa}_0 + (A_2 - A_1K_1^{-1}K_2)\boldsymbol{\xi}_2 + \boldsymbol{e}. \tag{8.7}$$

The $l \times 1$ vector $\boldsymbol{\xi}_1$ has vanished on the right side of (8.7). While this technique is possible, it might not be used frequently in practice.

8.2 Least-Squares Solution

We could derive the solution for $\boldsymbol{\xi}$ within the GHM (8.5) from statistical principles via BLUUE (Best Linear Uniformly Unbiased Estimate), but here we use the equivalent principle of LESS (LEast-Squares Solution) as was done in Sections 2.2, 3.2 and 4.3, etc. In the following, we recombine coefficient matrices A_1 and A_2 back into the single matrix A and recombine the partitioned parameter vector back into a single vector $\boldsymbol{\xi} = [\boldsymbol{\xi}_1^T, \boldsymbol{\xi}_2^T]^T$. Accordingly, we can rewrite (8.5) as

$$\bar{\boldsymbol{w}} = \bar{B}A_1\boldsymbol{\xi}_1 + \bar{B}A_2\boldsymbol{\xi}_2 + \bar{B}\boldsymbol{e} = \bar{A}\boldsymbol{\xi} + \bar{B}\boldsymbol{e}, \tag{8.8}$$

where another bar-symbol was introduced for convenience, viz. $\bar{A} \coloneqq \bar{B}A$.

Our target function should minimize a quadratic form in the random error vector $\boldsymbol{e}$ itself, rather than $B\boldsymbol{e}$; i.e., we minimize $\boldsymbol{e}^T P \boldsymbol{e}$. Then, the Lagrange target function is written as

$$\Phi(\boldsymbol{e}, \boldsymbol{\xi}, \boldsymbol{\lambda}) =: \boldsymbol{e}^T P \boldsymbol{e} + 2\boldsymbol{\lambda}^T\big(\bar{B}\boldsymbol{e} + \bar{A}\boldsymbol{\xi} - \bar{\boldsymbol{w}}\big), \tag{8.9}$$

which must be made stationary with respect to the unknown vectors $\boldsymbol{e}$, $\boldsymbol{\xi}$, and $\boldsymbol{\lambda}$. This is done by imposing the Euler-Lagrange necessary conditions, which results in the following system of equations:

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{e}} = P\tilde{\boldsymbol{e}} + \bar{B}^T\hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{8.10a}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\xi}} = \bar{A}^T\hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{8.10b}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\lambda}} = \bar{B}\tilde{\boldsymbol{e}} + \bar{A}\hat{\boldsymbol{\xi}} - \bar{\boldsymbol{w}} \doteq \mathbf{0}. \tag{8.10c}$$

The vectors of predicted random errors (residuals) and estimated parameters are then solved for as follows:

$$\begin{aligned}
&\tilde{\boldsymbol{e}} = -\big(P^{-1}\bar{B}^T\big)\hat{\boldsymbol{\lambda}} \Rightarrow && \text{from equation (8.10a)}\\
&-\big(\bar{B}P^{-1}\bar{B}^T\big)\hat{\boldsymbol{\lambda}} = \bar{\boldsymbol{w}} - \bar{A}\hat{\boldsymbol{\xi}} \Rightarrow && \text{multiplying by } \bar{B} \text{ and using (8.10c)}\\
&-\hat{\boldsymbol{\lambda}} = \big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\big(\bar{\boldsymbol{w}} - \bar{A}\hat{\boldsymbol{\xi}}\big) \Rightarrow && \big(\bar{B}P^{-1}\bar{B}^T\big) \text{ is invertible}\\
&-\bar{A}^T\hat{\boldsymbol{\lambda}} = \bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\big(\bar{\boldsymbol{w}} - \bar{A}\hat{\boldsymbol{\xi}}\big) = \mathbf{0} \Rightarrow && \text{mult. by } \bar{A}^T \text{ and using (8.10b)}\\
&\bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{A}\hat{\boldsymbol{\xi}} = \bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{\boldsymbol{w}}
\end{aligned}$$

Finally, we arrive at

$$\boxed{\hat{\boldsymbol{\xi}} = \big[\bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{A}\big]^{-1}\bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{\boldsymbol{w}}} \tag{8.11a}$$

and

$$\boxed{\tilde{\boldsymbol{e}} = P^{-1}\bar{B}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\big(\bar{\boldsymbol{w}} - \bar{A}\hat{\boldsymbol{\xi}}\big)} \tag{8.11b}$$

for the estimated parameters and predicted residuals, respectively. Equation (8.11a) has the same form as the LESS derived within the GMM in Section 3.2, and (8.11b) looks much like formula (4.5d) for the residual vector within the Model of Condition Equations.

Note that matrix $\bar{A}$ would need to have full column rank to use (8.11a). Thus, in the example problem in the preceding section, the datum deficiency would need to be handled first. This could be done, for example, by modifying ("pre-processing") the observation vector as mentioned in Section 7.5 and in Problem 2 of Section 7.7. In the remainder of this chapter, we will assume that there are no rank deficiencies in the the data models.

The dispersion matrix for the estimated parameter vector $\hat{\boldsymbol{\xi}}$ is expressed by

$$\boxed{D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2\big[\bar{A}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{A}\big]^{-1}.} \tag{8.12}$$

And the dispersion matrix for the residual vector reads

$$\boxed{\begin{aligned}D\{\tilde{\boldsymbol{e}}\} = P^{-1}\bar{B}^T\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\big[\bar{B}\cdot D\{\boldsymbol{e}\}\cdot\bar{B}^T - \\ - \bar{A}\cdot D\{\hat{\boldsymbol{\xi}}\}\cdot\bar{A}^T\big]\big(\bar{B}P^{-1}\bar{B}^T\big)^{-1}\bar{B}P^{-1},\end{aligned}} \tag{8.13}$$

with $D\{\boldsymbol{e}\} = \sigma_0^2 P^{-1}$ as stated in the model (8.1).

Notation change: For the remainder of the chapter we drop the bars from the symbols as a matter of convenience. Recall that the bars were introduced in the first place to distinguish between the matrix B introduced in (8.5) and that used in Chapter 4 for the Model of Condition Equations. Dropping the bars means that $\bar{B} \rightarrow B$, $\bar{\boldsymbol{w}} \rightarrow \boldsymbol{w}$, $\bar{A} \rightarrow BA$.

We make one more notation change by replacing the matrix product BA that was used in the derivations above with the symbol A itself to represents a more general form of the model. Recall that the matrix BA above included the coefficient matrix A from a Gauss-Markov Model (GMM). However the more general formulation of the least-squares adjustment within the GHM would not necessarily reference quantities used in a GMM.

With these simplified notations, we rewrite the solution (8.11a) as follows:

$$\boxed{\hat{\boldsymbol{\xi}} = \big[A^T\big(BP^{-1}B^T\big)^{-1}A\big]^{-1}A^T\big(BP^{-1}B^T\big)^{-1}\boldsymbol{w}.} \tag{8.14}$$

The dispersion of $\hat{\boldsymbol{\xi}}$ is derived in parts as follows:

$$\begin{aligned} D\{A^T\big(BP^{-1}B^T\big)^{-1}\boldsymbol{w}\} &= A^T\big(BP^{-1}B^T\big)^{-1}D\{\boldsymbol{w}\}\big(BP^{-1}B^T\big)^{-1}A = \\ &= A^T\big(BP^{-1}B^T\big)^{-1}B\cdot D\{\boldsymbol{y}\}\cdot B^T\big(BP^{-1}B^T\big)^{-1}A = \\ &= \big(BP^{-1}B^T\big)^{-1}\big(\sigma_0^2A^TBP^{-1}B^T\big)\big(BP^{-1}B^T\big)^{-1}A = \\ &= \sigma_0^2A^T\big(BP^{-1}B^T\big)^{-1}A; \end{aligned}$$

therefore

$$\begin{aligned} D\{\hat{\boldsymbol{\xi}}\} &= \big[A^T\big(BP^{-1}B^T\big)^{-1}A\big]^{-1}\cdot D\{A^T\big(BP^{-1}B^T\big)^{-1}\boldsymbol{w}\}\cdot\big[A^T\big(BP^{-1}B^T\big)^{-1}A\big]^{-1} = \\ &= \big[A^T\big(BP^{-1}B^T\big)^{-1}A\big]^{-1}\big[\sigma_0^2A^T\big(BP^{-1}B^T\big)^{-1}A\big]\big[A^T\big(BP^{-1}B^T\big)^{-1}A\big]^{-1}, \end{aligned}$$

finally resulting in

$$\boxed{D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2\big[A^T\big(BP^{-1}B^T\big)^{-1}A\big]^{-1}.} \tag{8.15}$$

8.3 Iteratively Linearized Gauss-Helmert Model

In this section we present the Gauss-Helmert Model (GHM) as an iteratively linearized model, showing how to form both the model and the least-squares solution within the model at each step of an iteration scheme. The developed algorithm is useful for a wide range of problems encountered in geodetic science and other disciplines.

The reader should be clear that the coefficient matrices A and B used in this section (and those that follow) are not the same as the coefficient matrices A and B used in the GMM and Model of Condition Equations, respectively. This should be obvious from the development that follows.

Suppose we are given a non-linear functional model that relates n observations $\boldsymbol{y}$ to m unknown parameters $\boldsymbol{\Xi}$ among $m + r$ non-linear condition equations $\boldsymbol{b}$ such that

$$\boldsymbol{b}(\underbrace{\boldsymbol{y} - \boldsymbol{e}}_{n\times 1}, \underbrace{\boldsymbol{\Xi}}_{m\times 1}) = \boldsymbol{0}, \;\; \boldsymbol{b} \in \mathbb{R}^{m+r}, \;\; \boldsymbol{e} \sim (\boldsymbol{0}, \sigma_0^2\underset{n\times n}{P^{-1}}), \tag{8.16}$$

representing also a mapping $\boldsymbol{b} : \mathbb{R}^{m+n} \to \mathbb{R}^{m+r}$. Equation (8.16) is a non-linear Gauss-Helmert Model.

By introducing the "true" $n \times 1$ vector of observables $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu} \coloneqq \boldsymbol{y} - \boldsymbol{e} = E\{\boldsymbol{y}\}, \tag{8.17}$$

the least-squares objective for model (8.16) is then defined by

$$\boldsymbol{e}^T P \boldsymbol{e} = \min, \;\; \text{subject to} \;\; \boldsymbol{b}(\boldsymbol{\mu}, \boldsymbol{\Xi}) = \mathbf{0}. \tag{8.18}$$

An iterative linearization of (8.16), together with the least-squares estimation of the unknown parameters $\boldsymbol{\Xi}$ and prediction of the unknown random errors $\boldsymbol{e}$, can be formed as follows.

Begin by assigning initial values $\boldsymbol{\mu}_0$ and $\boldsymbol{\Xi}_0$ to the unknowns $\boldsymbol{\mu}$ and $\boldsymbol{\Xi}$, respectively, e.g., $\boldsymbol{\mu}_0 = \boldsymbol{y} - \underset{\sim}{\mathbf{0}}$ and $\boldsymbol{\Xi}_0$ by some approximate method (perhaps using LESS within the GMM if linearization would not be required for that solution). Then execute the following conditional loop:

While

$$\delta < \|\hat{\boldsymbol{\xi}}_j\| \;\; \text{or} \;\; \epsilon < \|\tilde{\boldsymbol{e}}^{(j)} - \tilde{\boldsymbol{e}}^{(j-1)}\| \tag{8.19}$$

for chosen thresholds δ and ϵ, and $j \in \mathbb{N}$, perform the following steps:

(i) Use the truncated Taylor series about expansion point $(\boldsymbol{\mu}_j, \boldsymbol{\Xi}_j)$:

$$\left[\tfrac{\partial \boldsymbol{b}}{\partial \boldsymbol{\mu}^T}\big|_{\boldsymbol{\mu}_j, \boldsymbol{\Xi}_j}, \;\; \tfrac{\partial \boldsymbol{b}}{\partial \boldsymbol{\Xi}^T}\big|_{\boldsymbol{\mu}_j, \boldsymbol{\Xi}_j} \right] \cdot \begin{bmatrix} \boldsymbol{\mu} - \boldsymbol{\mu}_j \\ \boldsymbol{\Xi} - \boldsymbol{\Xi}_j \end{bmatrix} + \boldsymbol{b}(\boldsymbol{\mu}_j, \boldsymbol{\Xi}_j) = \mathbf{0}, \tag{8.20a}$$

and replace $\boldsymbol{\mu}$ with $\boldsymbol{y} - \boldsymbol{e}$ in accordance with (8.17), to introduce

$$\underset{m\times 1}{\boldsymbol{\xi}_{j+1}} \coloneqq \boldsymbol{\Xi} - \boldsymbol{\Xi}_j, \;\; \underset{(m+r)\times m}{A^{(j)}} \coloneqq -\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\Xi}^T}\big|_{\boldsymbol{\mu}_j, \boldsymbol{\Xi}_j}, \;\; \underset{(m+r)\times n}{B^{(j)}} \coloneqq \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\mu}^T}\big|_{\boldsymbol{\mu}_j, \boldsymbol{\Xi}_j}, \tag{8.20b}$$

$$\underset{(m+r)\times 1}{\boldsymbol{w}_j} \coloneqq \boldsymbol{b}(\boldsymbol{\mu}_j, \boldsymbol{\Xi}_j) + B^{(j)} \cdot (\boldsymbol{y} - \boldsymbol{\mu}_j), \tag{8.20c}$$

and to form the linearized Gauss-Helmert Model

$$\boldsymbol{w}_j = A^{(j)} \hat{\boldsymbol{\xi}}_{j+1} + B^{(j)} \boldsymbol{e}, \;\; \boldsymbol{e} \sim (\mathbf{0}, \sigma_0^2 P^{-1}). \tag{8.20d}$$

(ii) Produce the $(j+1)$th LEast-Squares Solution (LESS) for (8.20d), viz.

$$\begin{aligned} \hat{\boldsymbol{\xi}}_{j+1} = \big\{ (A^{(j)})^T \big[(B^{(j)}) P^{-1} (B^{(j)})^T \big]^{-1} (A^{(j)}) \big\}^{-1} \cdot \\ \cdot (A^{(j)})^T \big[(B^{(j)}) P^{-1} (B^{(j)})^T \big]^{-1} \boldsymbol{w}_j, \end{aligned} \tag{8.20e}$$

$$\tilde{\boldsymbol{e}}^{(j+1)} = P^{-1} (B^{(j)})^T \big[(B^{(j)}) P^{-1} (B^{(j)})^T \big]^{-1} (\boldsymbol{w}_j - (A^{(j)}) \hat{\boldsymbol{\xi}}_{j+1}). \tag{8.20f}$$

(iii) Obtain new approximate values (non-random) through

$$\boldsymbol{\Xi}_{j+1} \coloneqq \hat{\boldsymbol{\Xi}}^{(j+1)} - \underset{\sim}{\mathbf{0}} = \boldsymbol{\Xi}_j + \hat{\boldsymbol{\xi}}_{j+1} - \underset{\sim}{\mathbf{0}}, \tag{8.20g}$$

$$\boldsymbol{\mu}_{j+1} \coloneqq \hat{\boldsymbol{\mu}}^{(j+1)} - \underset{\sim}{\mathbf{0}} = \boldsymbol{y} - \tilde{\boldsymbol{e}}^{(j+1)} - \underset{\sim}{\mathbf{0}}, \tag{8.20h}$$

where $\underset{\sim}{\mathbf{0}}$ denotes a "random zero vector" of suitable size (following Harville, 1986). This means that the jth (approximate) estimates are stripped of their randomness while keeping their numerical values. The use of $\underset{\sim}{\mathbf{0}}$ is formally required to avoid the assignment of random values to a non-random quantity; however, its use is of no consequence in practice, since it does not affect the numerical results.

Repeat the cycle until convergence is reached.

As already suggested, the initial approximate values for $\boldsymbol{\mu}$ might be taken from the observation vector $\boldsymbol{y}$ via $\boldsymbol{\mu}_0 \coloneqq \boldsymbol{y} - \underset{\sim}{\mathbf{0}}$. Unfortunately, this has occasionally led to the misunderstanding that the so-called "misclosure vector" $\boldsymbol{w}_i$, in the ith iteration cycle, ought to be updated by $\boldsymbol{b}(\boldsymbol{\mu}_i, \boldsymbol{\Xi}_i)$ when, in fact, the *correct update* is described by (8.20c). Also, the expression for $\boldsymbol{w}_j$ in (8.20c) is approximately equal to $\boldsymbol{b}(\boldsymbol{y}, \boldsymbol{\Xi}_j)$ and sometimes may turn out to be precisely equal to it; however, in some cases its usage may lead to convergence to an inaccurate solution. An excellent treatment of potential pitfalls for solving non-linear least-squares problems, can be found in Pope (1972), which the reader is encouraged to read. See Schaffrin and Snow (2010) for a more detailed discussion of this topic.

8.4 Estimated Variance Component

The P-weighted norm of the residual vector $\tilde{\boldsymbol{e}}$ is defined as

$$\boxed{\Omega \coloneqq \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} =} \tag{8.21a}$$

$$= \big(\hat{\boldsymbol{\lambda}}^T B P^{-1}\big) P \big(P^{-1} B^T \hat{\boldsymbol{\lambda}}\big) = \tag{8.21b}$$

$$= \big[-\big(\boldsymbol{w} - A\hat{\boldsymbol{\xi}}\big)^T \big(BP^{-1}B^T\big)^{-1}\big] \big(BP^{-1}B^T\big)\hat{\boldsymbol{\lambda}} = \tag{8.21c}$$

$$= \big(\boldsymbol{w} - A\hat{\boldsymbol{\xi}}\big)^T \big(BP^{-1}B^T\big)^{-1} \big(\boldsymbol{w} - A\hat{\boldsymbol{\xi}}\big) = \tag{8.21d}$$

$$= \big(B\tilde{\boldsymbol{e}}\big)^T \big(BP^{-1}B^T\big)^{-1} \big(B\tilde{\boldsymbol{e}}\big). \tag{8.21e}$$

Thus it follows that, the *uniformly unbiased estimate* of the variance component σ_0^2 is given by

$$\boxed{\hat{\sigma}_0^2 = \frac{\big(B\tilde{\boldsymbol{e}}\big)^T \big(BP^{-1}B^T\big)^{-1} \big(B\tilde{\boldsymbol{e}}\big)}{r} = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{r} = \frac{-\boldsymbol{w}^T \hat{\boldsymbol{\lambda}}}{r},} \tag{8.22}$$

where the redundancy r is defined as

$$\boxed{r \coloneqq \operatorname{rk} B - \operatorname{rk} A,} \tag{8.23}$$

which is the number of rows of B minus the number of columns of A, assuming matrix B has full row rank and matrix A has full column rank.

8.5 Equivalent Normal Equations

From (8.10b) and the second equation following (8.10c), and considering the notation changed described on Page 155, we can recognize the following system of normal equations:

$$\begin{bmatrix} BP^{-1}B^T & -A \\ -A^T & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{w} \\ \mathbf{0} \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \left[\begin{array}{c|c} BP^{-1}B^T & -A \\ \hline -A^T & 0 \end{array}\right]^{-1} \begin{bmatrix} -\boldsymbol{w} \\ \mathbf{0} \end{bmatrix}. \quad (8.24)$$

We want to show that the solution to this system yields the same $\hat{\boldsymbol{\xi}}$ as that of (8.14). The formula for the inverse of a partitioned matrix (see (A.11)) leads to the following solution:

$$\begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \left[\begin{array}{c|c} X_1 & X_2 \\ \hline -W^{-1}A^T(BP^{-1}B^T)^{-1} & (0-W)^{-1} \end{array}\right] \begin{bmatrix} -\boldsymbol{w} \\ \mathbf{0} \end{bmatrix},$$

with $W \coloneqq A^T(BP^{-1}B^T)^{-1}A$, and finally to

$$\begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} -X_1\boldsymbol{w} \\ \left[A^T(BP^{-1}B^T)^{-1}A\right]^{-1}A^T(BP^{-1}B^T)^{-1}\boldsymbol{w} \end{bmatrix}. \quad (8.25)$$

Here the symbols X_1 and X_2 represent quantities of no interest. We see that the solution for the parameters $\hat{\boldsymbol{\xi}}$ is the same in (8.14).

8.6 Example Problems

The following example problems are meant to help illustrate the use of the Gauss-Helmert Model (GHM).

8.6.1 Example — Fitting a Parabola When Both x- and y-Coordinates are Observed

In this example, we show how the GHM can be used to fit a parabola when both the x- and y-coordinates have been observed. This is in contrast to the problem treated in Section 3.2.1 with the GMM, where only the dependent variables (y-coordinates) could be considered as measurements. Here, the observation vector $\boldsymbol{y}$ is comprised of all pairs of the $n/2$ measured points. For example, $\boldsymbol{y}$ could be defined as

$$\underset{n\times 1}{\boldsymbol{y}} = \left[x_1, x_2, \ldots, x_{n/2}, y_1, y_2, \ldots, y_{n/2}\right]^T. \quad (8.26)$$

Alternatively, the elements of $\boldsymbol{y}$ could be ordered by coordinate pairs, i.e., $\boldsymbol{y} = [x_1, y_1, \ldots, x_{n/2}, y_{n/2}]^T$. The key is that consistency of ordering must be maintained for the coefficient matrix B, the random error vector $\boldsymbol{e}$, and the observation cofactor matrices, too.

Denoting the true (unknown) variables as μ_{x_i} and μ_{y_i}, $i = 1, 2, \ldots, n/2$, the following equations can be written for the ith pair of observed variables (x_i, y_i):

$$x_i = \mu_{x_i} + e_{x_i},\ E\{e_{x_i}\} = 0 \Rightarrow E\{x_i\} = \mu_{x_i}, \tag{8.27a}$$
$$y_i = \mu_{y_i} + e_{y_i},\ E\{e_{y_i}\} = 0 \Rightarrow E\{y_i\} = \mu_{y_i}. \tag{8.27b}$$

For this example, we assume that the measurement errors are iid. Collecting the random error terms in vectors $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$, respectively, their stochastic nature can then be expressed succinctly as

$$\underset{n\times 1}{\boldsymbol{e}} = \begin{bmatrix}\boldsymbol{e}_x \\ \boldsymbol{e}_y\end{bmatrix} \sim \left(\begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2 \begin{bmatrix} I_{n/2} & 0 \\ 0 & I_{n/2}\end{bmatrix}\right). \tag{8.28}$$

The (nonlinear) function that relates the ith pair of variables (μ_{x_i}, μ_{y_i}) to the non-random parameters $\boldsymbol{\Xi} = [\Xi_1, \Xi_2, \Xi_3]^T$ is given by

$$b_i(\Xi_1, \Xi_2, \Xi_3, \mu_{x_i}, \mu_{y_i}) = \mu_{y_i} - \mu_{x_i}^2 \Xi_1 - \mu_{x_i}\Xi_2 - \Xi_3 = 0,\ \ i \in \{1, 2, \ldots, n/2\}, \tag{8.29a}$$

which can be linearized about $(\boldsymbol{u}_{i0}, \boldsymbol{\Xi}_0)$ by

$$b_i^0 + \mathrm{d}\mu_{y_i} - (2\mu_{x_i}^0 \Xi_1^0 + \Xi_2^0)\,\mathrm{d}\mu_{x_i} - (\mu_{x_i}^2)^0\,\mathrm{d}\Xi_1 - \mu_{x_i}^0\,\mathrm{d}\Xi_2 - \mathrm{d}\Xi_3 = 0, \tag{8.29b}$$

where higher order terms have been neglected. Here the superscript 0 denotes the expansion point for the variables and parameters that the derivatives are evaluated at, viz. $\boldsymbol{u}_{i0} = [\mu_{x_i}^0, \mu_{y_i}^0]^T$ and $\boldsymbol{\Xi}_0 = [\Xi_1^0, \Xi_2^0, \Xi_3^0]^T$. The argument list for b_i has been dropped for the sake of brevity. Now define $n/2$ equations with:

$$\boldsymbol{\xi} = \big[\mathrm{d}\Xi_1,\, \mathrm{d}\Xi_2,\, \mathrm{d}\Xi_3\big]^T = \boldsymbol{\Xi} - \boldsymbol{\Xi}_0, \tag{8.30a}$$
$$-A_i = \big[-(\mu_{x_i}^0)^2,\, -\mu_{x_i}^0,\, -1\big], \tag{8.30b}$$
$$B_i = \big[-2\mu_{x_i}^0 \Xi_1^0 - \Xi_2^0,\, 1\big], \tag{8.30c}$$

where A_i is the ith row of an $(m + r) \times m$ matrix A (with $m = 3$ and r being the redundancy of the model, and $n = 2(m + r)$ in this example). In contrast, B_i shows only the non-zero elements of a row of an $(m+r) \times n$ matrix B. Those two elements go in the ith and $2i$th columns, respectively, of the ith full row of B (assuming the ordering of observations shown in (8.26)). Further define

$$\mathrm{d}\mu_{x_i} = \mu_{x_i} - \mu_{x_i}^0 = x_i - \mu_{x_i}^0 - e_{x_i} \ \text{ and } \ \mathrm{d}\mu_{y_i} = \mu_{y_i} - \mu_{y_i}^0 = y_i - \mu_{y_i}^0 - e_{y_i}, \tag{8.30d}$$

along with vectors

$$\boldsymbol{e}_i = \big[e_{x_i},\, e_{y_i}\big]^T, \ \text{ and } \ w_i = b_i^0 + B_i\big[x_i - \mu_{x_i}^0,\, y_i - \mu_{y_i}^0\big]^T \tag{8.30e}$$

so that (8.29b) can be rewritten for the ith observed coordinate pair as

$$-A_i\boldsymbol{\xi} - B_i\boldsymbol{e}_i + w_i = 0. \tag{8.31}$$

Then the complete set of $n/2 = m + r$ equations can be expressed as

$$\boldsymbol{w} = A\boldsymbol{\xi} + B\boldsymbol{e}, \tag{8.32}$$

which is obviously in the form of a GHM, within which a least-squares solution can be computed using the algorithm described in Section 8.3. In order to avoid non-convergence or convergence to the wrong solution, one must pay careful attention to the comments in the last paragraph of Section 8.3. In practice, they mean that all the terms A, B, and $\boldsymbol{w}$ must be updated at each iteration, using numerical value computed from the previous iteration.

The preceding formulation can be applied to the fitting of many different kinds of functions in 2D- and 3D-space, including lines, planes, quadratic surfaces, etc. When the data are iid, these adjustments amount to solving so-called "orthogonal regression" problems, since the residuals pairs $(\tilde{e}_{x_i}, \tilde{e}_{y_i})$ define vectors that are orthogonal to the fitted curve (or surface in 3D). In the case of a general weight matrix P, we might prefer to say "P-weighted orthogonal regression," since the weights will influence the direction of the 2D and 3D residual vectors.

8.6.2 Example — Fitting a Ellipse When Both x- and y-Coordinates are Observed

An equation for an ellipse can be written as a function of its center point (z_1, z_2), the length of its semi-major axis a, the length of its semi-minor axis b, and the counter-clockwise angle α between the z_1-axis and the semi-major axis (see Figure 8.2). Accordingly, a (nonlinear) function that relates the ith pair of $n/2$ pairs of random variables (μ_{x_i}, μ_{y_i}) to the unknown (but non-random) parameters $(\mu_\alpha, \mu_a, \mu_b, \mu_{z_1}, \mu_{z_2})$ is provided by

$$\begin{aligned} & b_i\left(\mu_\alpha, \mu_a, \mu_b, \mu_{z_1}, \mu_{z_2}, \mu_{x_i}, \mu_{y_i}\right) = \\ & = \mu_b^2\big[\cos^2\mu_\alpha(\mu_{x_i} - \mu_{z_1})^2 + 2\cos\mu_\alpha \sin\mu_\alpha(\mu_{x_i} - \mu_{z_1})(\mu_{y_i} - \mu_{z_2}) + \\ & \qquad + \sin^2\mu_\alpha(\mu_{y_i} - \mu_{z_2})^2\big] + \\ & + \mu_a^2\big[\sin^2\mu_\alpha(\mu_{x_i} - \mu_{z_1})^2 - 2\sin\mu_\alpha\cos\mu_\alpha(\mu_{x_i} - \mu_{z_1})(\mu_{y_i} - \mu_{z_2}) + \\ & \qquad + \cos^2\mu_\alpha(\mu_{y_i} - \mu_{z_2})^2\big] - \mu_a^2\mu_b^2 = 0, \end{aligned} \tag{8.33}$$

with $i \in \{1, \ldots, n/2\}$. Collecting the unknown parameters in the vector $\boldsymbol{\Xi}$, viz. $\boldsymbol{\Xi} = [\mu_\alpha, \mu_a, \mu_b, \mu_{z_1}, \mu_{z_2}]^T$, their values can then be estimated via a least-squares solution within the GHM as outlined in Section 8.6.1, where (8.29a) would be replaced by (8.33).

8.6.3 Example — 2D Similarity Transformation When Coordinates Have Been Observed in Both the Source and Target Systems

If $n/2$ coordinate pairs (X_i, Y_i) and (x_i, y_i) have been observed in both target ("new") and source ("old") coordinate systems, respectively, then the following

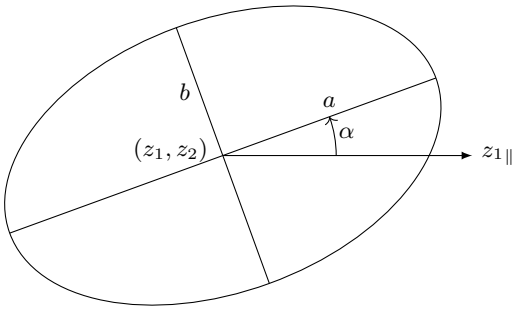

Figure 8.2: An ellipse with semi-major and semi-minor axes a and b, respectively, centered at (z_1, z_2) and rotated by angle α

GHM can be used to model a 2D similarity transformation:

$$\boldsymbol{b}(\boldsymbol{\mu}, \boldsymbol{\xi}) \coloneqq \begin{bmatrix} \dots \\ X_i \\ Y_i \\ \dots \end{bmatrix} - \begin{bmatrix} \dots \\ e_{X_i} \\ eY_i \\ \dots \end{bmatrix} - \begin{bmatrix} \dots & \dots & \dots & \\ 1 & 0 & x_i - e_{x_i} & -(y_i - e_{y_i}) \\ 0 & 1 & y_i - e_{y_i} & x_i - e_{x_i} \\ \dots & & & \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} = \mathbf{0}, \quad (8.34a)$$

where

$$\boldsymbol{y} \coloneqq \begin{bmatrix} \dots, X_i, Y_i, \dots, x_i, y_i, \dots \end{bmatrix}^T \text{ is a } 2n \times 1 \text{ vector of observed coordinates,} \quad (8.34b)$$

$$\boldsymbol{e} \coloneqq \begin{bmatrix} \dots, e_{X_i}, e_{Y_i}, \dots, e_{x_i}, e_{y_i}, \dots \end{bmatrix}^T \text{ is a } 2n \times 1 \text{ random error vector,} \quad (8.34c)$$

$$\boldsymbol{\mu} \coloneqq \boldsymbol{y} - \boldsymbol{e} \text{ is a } 2n \times 1 \text{ vector of actual ("true") coordinates, and} \quad (8.34d)$$

$$\boldsymbol{\xi} \coloneqq \begin{bmatrix} \xi_1, \xi_2, \xi_3, \xi_4 \end{bmatrix}^T \text{ is the } 4 \times 1 \text{ vector of unknown parameters, with} \quad (8.34e)$$

$$\xi_3 \coloneqq \omega \cos \alpha, \text{ and } \xi_4 \coloneqq \omega \sin \alpha. \quad (8.34f)$$

Here, ξ_1 and ξ_2 are translation parameters along the X- and Y-axis, respectively; ω is a scale factor, and α is a counter-clockwise rotation angle.

8.7 Some Published Examples

The following papers include numerical examples for the GHM that may be of interest to the reader:

1. Circle fitting: Schaffrin and Snow (2010).
2. Line fitting in 3D: Snow and Schaffrin (2016).
3. 2D similarity transformations: Neitzel and Petrovic (2008).

In addition to these sources, Pope's 1972 paper is highly recommended to understand how to best avoid potential pitfalls in adjusting data within iteratively linearized models.

8.8 Practice Problems

1. A circle is to be fitted to the set of measured coordinates shown in Table 8.1, which were presented in Schaffrin and Snow (2010). Both x- and y-coordinates were measured, and the associated random errors are considered to be iid.

Table 8.1: Measured coordinates for the fitting of a circle. Units are not given.

No.	$\boldsymbol{x}$	$\boldsymbol{y}$
1	0.7	4.0
2	3.3	4.7
3	5.6	4.0
4	7.5	1.3
5	6.4	−1.1
6	4.4	−3.0
7	0.3	−2.5
8	−1.1	1.3

   (a) Setup an appropriate Gauss-Helmert Model with the coordinates of the center of the circle and its radius as the three unknown parameters. What is the redundancy of the model?

   (b) Compute the least-squares estimates of the center of the circle and its radius. You may use the following initial approximations for the parameters: $\boldsymbol{\Xi}_0 = [3, 1, 4]^T$ (in order of x and y coordinates of the center point followed by the radius).

   (c) Compute the estimated variance component and the empirical rms of the estimated parameters (i.e., the square roots of the diagonal elements of the estimated dispersion matrix).

   (d) What is the geometrical relationship between the estimated center of the circle and each respective pair of observed and adjusted coordinates?

2. An ellipse is to be fitted to the set of measured coordinates plotted in Figure 8.3 and listed in Table 8.2. Both z_1- and z_2-coordinates were measured, and the associated random errors are considered to be iid.

   (a) Setup an appropriate Gauss-Helmert Model with the coordinates of the center of the ellipse (z_1, z_2), its semi-major and semi-minor axes lengths a and b, and the angle α between the z_1 axis and the semi-major axis as

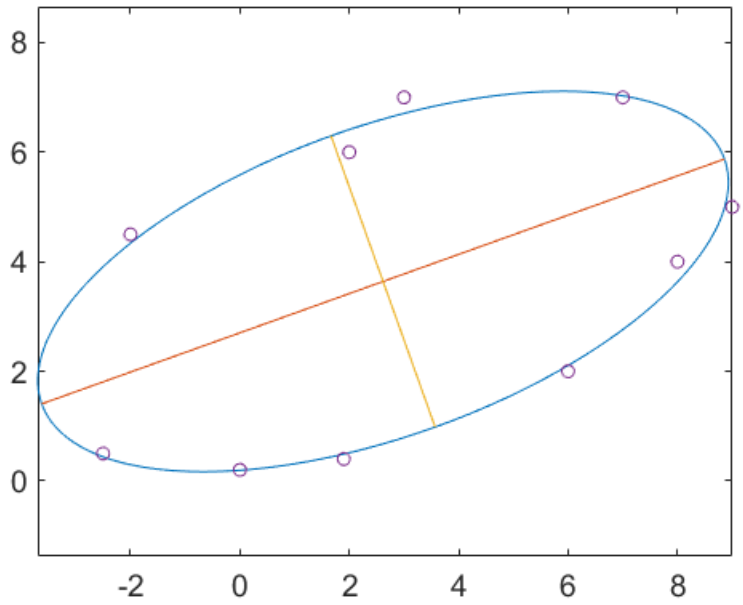

Figure 8.3: Fitted ellipse and measured coordinates in the 2-D plane (listed in Table 8.2)

five unknown parameters (see Section 8.6.2). What is the redundancy of the model?

Table 8.2: Measured coordinates for the fitting of an ellipse. Units are not given.

No.	z_1	z_2
1	2.0	6.0
2	7.0	7.0
3	9.0	5.0
4	3.0	7.0
5	6.0	2.0
6	8.0	4.0
7	−2.0	4.5
8	−2.5	0.5
9	1.9	0.4
10	0.0	0.2

(b) Compute the least-squares estimates of the unknown parameters of the ellipse. You may use the following initial approximations for the parameters: $\boldsymbol{\Xi}_0 = [0, 7, 3, 3, 4]^T$ (in order of μ_α^0, μ_a^0, μ_b^0, $\mu_{z_1}^0$, $\mu_{z_1}^0$).

(c) Compute the estimated variance component and the empirical rms of the estimated parameters (i.e., the square roots of the diagonal elements of the estimated dispersion matrix).

3. The data used for fitting the parabola shown in Figure 3.1 are listed in Table 8.3. Assume that the cofactor matrix for the x-coordinates is $Q_x = (0.010\,\mathrm{m})^2 \cdot I_n$ and that the cofactor matrix for the y-coordinates is $Q_y = (0.005\,\mathrm{m})^2 \cdot I_n$. Using the model presented in Section 8.6.1, compute the following:

(a) Estimates for the three unknown parameters of the parabola.

(b) The estimated variance component.

(c) The empirical rms of the estimated parameters (i.e., the square roots of the diagonal elements of the estimated dispersion matrix).

Table 8.3: Measured coordinates for the fitting of a parabola. The units are in meters.

No.	$\boldsymbol{x}$	$\boldsymbol{y}$
1	1.007	1.827
2	1.999	1.911
3	3.007	1.953
4	3.998	2.016
5	4.999	2.046
6	6.015	2.056
7	7.014	2.062
8	8.014	2.054
9	9.007	2.042
10	9.988	1.996
11	11.007	1.918
12	12.016	1.867

8.9 Summary Formulas for the Least-Squares Solution Within the Gauss-Helmert Model

See the last paragraph of Section 8.2 for comments about replacing the product BA used in earlier sections of this chapter with the matrix A itself, as was done in the

following:

The linearized Gauss-Helmert Model (GHM) is given by

$$\underset{(r+m)\times 1}{\boldsymbol{w}} = \underset{(r+m)\times m}{A}\,\boldsymbol{\xi} + \underset{(r+m)\times n}{B}\,\underset{n\times 1}{\boldsymbol{e}}, \quad \underset{n\times 1}{\boldsymbol{e}} \sim (\mathbf{0}, \sigma_0^2 P^{-1}).$$

Table 8.4: Summary formulas for the LESS within the Gauss-Helmert Model

Quantity	Formula	Eq.
Model redundancy	$r = \operatorname{rk} B - \operatorname{rk} A$	(8.23)
Vector of estimated parameters	$\hat{\boldsymbol{\xi}} = \left[A^T\left(BP^{-1}B^T\right)^{-1}A\right]^{-1}A^T\left(BP^{-1}B^T\right)^{-1}\boldsymbol{w}$	(8.11a)
Dispersion matrix for estimated parameters	$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 \cdot \left[A^T\left(BP^{-1}B^T\right)^{-1}A\right]^{-1}$	(8.12)
Vector of predicted residuals	$\tilde{\boldsymbol{e}} = P^{-1}B^T\left(BP^{-1}B^T\right)^{-1}\left(\boldsymbol{w} - A\hat{\boldsymbol{\xi}}\right)$	(8.11b)
Dispersion matrix for residuals	$D\{\tilde{\boldsymbol{e}}\} = P^{-1}B^T\left(BP^{-1}B^T\right)^{-1}\left[B \cdot D\{\boldsymbol{e}\} \cdot B^T - A \cdot D\{\hat{\boldsymbol{\xi}}\} \cdot A^T\right]\left(BP^{-1}B^T\right)^{-1}BP^{-1}$	(8.13)
Sum of squared residuals (SSR)	$\Omega = \tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}$	(8.21a)
Estimated variance component	$\hat{\sigma}_0^2 = \Omega / r$	(8.22)

Continued from previous page

Chapter 9

Statistical Analysis

It is assumed that the reader of these notes has had at least a first course in statistical methods or probability theory and thus has some familiarity with hypothesis testing in statistical analysis. Therefore, key terms and concepts will be described only briefly, and the main focus will be placed on the application of hypothesis testing to parameters estimated from least-squares adjustments as described in the preceding chapters. For a broader treatment of statistical methods, and an excellent refresher on hypothesis testing in particular, see Snedecor and Cochran (1980).

Consider a normally distributed random (scalar) variable y with the following first through fourth moments:

$$E\{y\} = \mu, \tag{9.1a}$$

$$E\{(y-\mu)^2\} = D\{y\} = \sigma^2, \tag{9.1b}$$

$$E\{(y-\mu)^3\} = 0, \tag{9.1c}$$

$$E\{(y-\mu)^4\} = 3(\sigma^2)^2. \tag{9.1d}$$

The third moment being zero in (9.1c) means there is no skewness in the distribution of the random variable. The right side of (9.1d) indicates that there is no kurtosis (peak) in the distribution.

If (9.1c) or (9.1d) are not satisfied, the variable is not normally distributed and can be characterized as follows:

$$E\{(y-\mu)^3\} > 0 \Leftrightarrow \text{the distribution is skewed to the positive side.} \tag{9.2a}$$

$$E\{(y-\mu)^3\} < 0 \Leftrightarrow \text{the distribution is skewed to the negative side.} \tag{9.2b}$$

$$E\{(y-\mu)^4\} - 3(\sigma^2)^2 > 0 \Leftrightarrow \text{the distribution has positive kurtosis.} \tag{9.2c}$$

$$E\{(y-\mu)^4\} - 3(\sigma^2)^2 < 0 \Leftrightarrow \text{the distribution has negative kurtosis.} \tag{9.2d}$$

Skewness appears in a graph of a sample of the random variable (e.g., a histogram) as a shift in the peak value from center. Positive kurtosis shows higher probability near the expected value μ, which results in a taller, narrower graph. Negative

kurtosis shows higher probability in the tails of the graph; thus the graph appears flatter than that of a normally distributed variable.

The pdf (probability density function, or density function) of a normally distributed random (scalar) variable y is

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(y-\mu)^2/2\sigma^2}, \tag{9.3}$$

where μ is the expectation of the distribution (population mean), σ is standard deviation, σ^2 is variance, and e is Euler's number (i.e., the base of the natural logarithm, $e \approx 2.71828$). Note that the term $1/\sqrt{2\pi\sigma^2} \approx 0.4/\sigma$ denotes the amplitude of the graph of the curve, μ shows the offset of the peak from center, and σ is the distance from the center to the inflection points of the curve.

The cdf (cumulative distribution function, or distribution function) of a *normally distributed* random variable is expressed as

$$F(y) = \int_{-\infty}^{y} f(t)\, dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} e^{-(t-\mu)^2/2\sigma^2}\, dt. \tag{9.4}$$

Figure 9.1 shows pdf and cdf plots for the normal distribution using various values for μ and σ_0^2. Line colors and types match between the pdf and cdf plots. The solid, green line represents the respective standard normal pdf and cdf curves.

Note that, in geodetic-science applications, the random variable y might be an observation, an adjusted observation, a predicted residual, etc. We can standardize the random variable y with the following transformation, which subtracts out the mean and divides by the standard deviation:

$$z = \frac{y - \mu}{\sigma}. \tag{9.5}$$

The standardized random variable z has the following moments and probability functions:

$$E\{z\} = 0, \tag{9.6a}$$

$$D\{z\} = 1, \tag{9.6b}$$

$$\text{pdf}: \; f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}, \tag{9.6c}$$

$$\text{cdf}: \; F(z) = \int_{-\infty}^{z} f(t)\, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt\,. \tag{9.6d}$$

A plot of the pdf of z is shown in Figure 9.2, along with example Student's t-distribution curves (discussed below).

In the multivariate case, the random variable $\boldsymbol{y}$ is an $n \times 1$ vector, with an $n \times n$ dispersion (covariance) matrix $\Sigma = D\{\boldsymbol{y}\}$ and expectation vector $\boldsymbol{\mu} = E\{\boldsymbol{y}\}$, which is also size $n \times 1$. The pdf is then written as

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}}\, e^{-(\boldsymbol{y}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{y}-\boldsymbol{\mu})/2}. \tag{9.7}$$

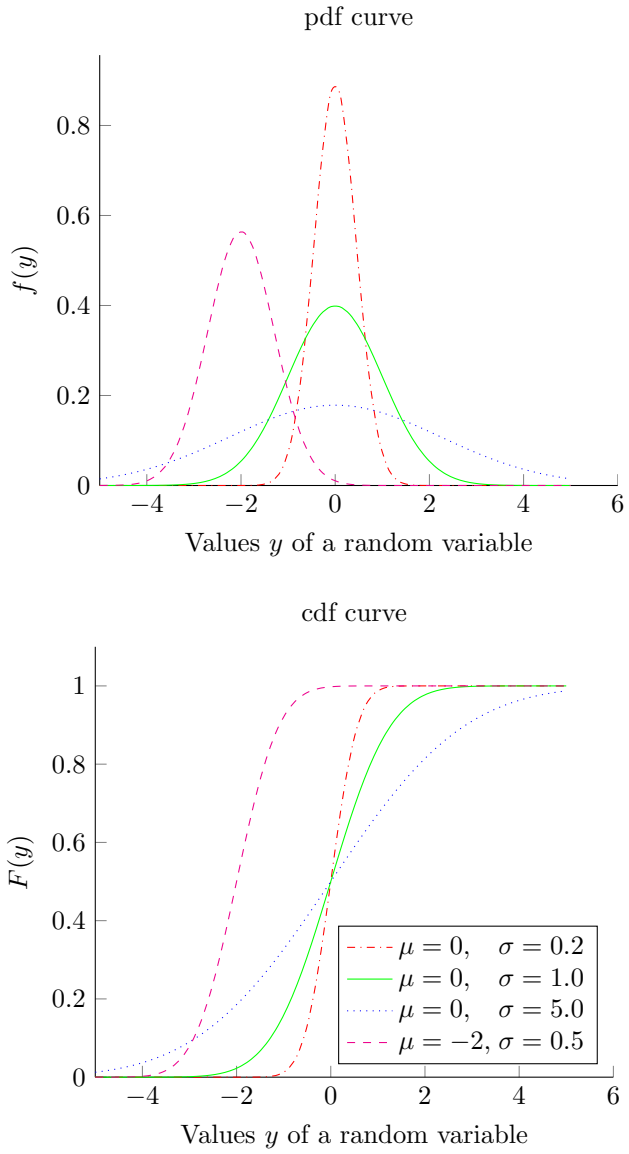

Figure 9.1: pdf curve (top) and cdf curve (bottom) for the normal distribution with matching line types and colors so that the legend pertains to both graphs

And the cdf is written as

$$F(y_1, \ldots, y_n) = \int_{-\infty}^{y_n} \cdots \int_{-\infty}^{y_1} f(t_1, \ldots, t_n)\, dt_1 \ldots dt_n\,. \tag{9.8}$$

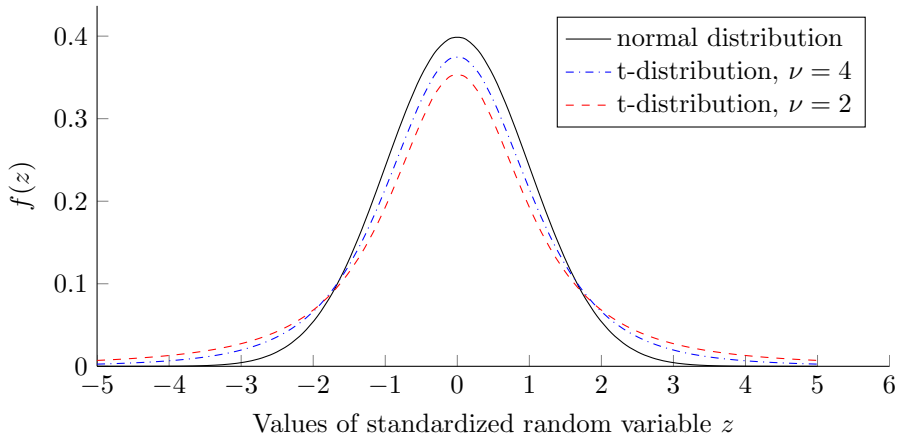

Figure 9.2: Curves of student's t- and normal distributions for a standardized random variable

The elements of $\boldsymbol{y}$, i.e. $y_1, \ldots, y_n$, are statistically independent if, and only if,

$$f(t_1, \ldots, t_n) = f(t_1) \cdot f(t_2) \cdot \ldots f(t_n), \tag{9.9a}$$

which implies

$$C\{y_i, y_j\} = 0 \text{ for } i \neq j. \tag{9.9b}$$

Equation (9.9b) states that there is no covariance between the elements of random vector $\boldsymbol{y}$.

The third and fourth moments for the multivariate case are given in (9.10a) and (9.10b), respectively.

$$E\{(y_i - \mu_i)(y_j - \mu_j)(y_k - \mu_k)\} = 0 \text{ for } i, j, k = \{1, \ldots, n\} \tag{9.10a}$$

$$E\{(y_i - \mu_i)(y_j - \mu_j)(y_k - \mu_k)(y_l - \mu_l)\} = 3(\sigma_i^2)\delta_{ijkl} \text{ for } i, j, k, l = \{1, \ldots, n\} \tag{9.10b}$$

In the following, we discuss *studentized residuals*, which have a t-distribution (or Student's t-distribution). The pdf for a (scalar) variable having a t-distribution and $\nu = n - 1$ degrees of freedom is defined as follows:

$$f(t) = \frac{1}{\sqrt{(n-1)\pi}} \cdot \frac{\Gamma(n/2)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}}, \tag{9.11}$$

where the gamma function is defined by

$$\Gamma(n) \coloneqq (n-1)\Gamma(n-1) = \int_0^\infty e^{-t} t^{n-1}\, dt = (n-1)! \text{ for } n \in \mathbb{N}. \tag{9.12}$$

As is known from introductory statistics, the pdf for the Student's t-distribution resembles the pdf of the normal distribution when n is around 30. A plot of the pdf

for the Student's t-distribution, with $\nu = 2, 4$, together with the pdf for the normal distribution, is shown in Figure 9.2.

9.1 Standardized and Studentized Residuals

We begin this section by restating the (full-rank) Gauss-Markov Model and writing the predicted vector of random errors within the model.

$$\boldsymbol{y} = A\boldsymbol{\xi} + \boldsymbol{e}, \quad \boldsymbol{e} \sim \big(\boldsymbol{0}, \sigma_0^2 P^{-1}\big), \quad \mathrm{rk}\, A = m \tag{9.13a}$$

$$\tilde{\boldsymbol{e}} = \big(I_n - AN^{-1}A^T P\big)\boldsymbol{y} = \big(I_n - AN^{-1}A^T P\big)\boldsymbol{e} \tag{9.13b}$$

As usual, the observation vector $\boldsymbol{y}$ is of size $n \times 1$, and the coefficient matrix A is of size $n \times m$. Obviously, the far-right side of (9.13b) cannot be computed since $\boldsymbol{e}$ is an unknown variable. However, the expression is useful for analytical purposes.

In the following, we assume that the random error vector $\boldsymbol{e}$ has a normal distribution expressed by $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma_0^2 P^{-1})$ (where the symbol $\mathcal{N}$ denotes normal distribution). This assumption is made for the sake of hypothesis testing in statistical analysis, which requires that test statistics[1] be computed as a function of a specified pdf. The justification of the assumption owes to the central limit theorem as stated by Bjerhammar (1973, p. 35) (see also the footnote on page 182). However, since $\boldsymbol{e}$ and σ_0^2 are unknown their respective prediction $\tilde{\boldsymbol{e}}$ and estimate $\hat{\sigma}_0^2$ are used instead; consequently, the student t-distribution is used in place of the normal distribution for formulating hypothesis tests.

The so-called *standardized residual* vector is a function of the residual vector $\tilde{\boldsymbol{e}}$ and its dispersion matrix $D\{\tilde{\boldsymbol{e}}\}$ as shown in the following:

$$D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2\big(P^{-1} - AN^{-1}A^T\big), \tag{9.14a}$$

$$\sigma_{\tilde{e}_j}^2 = \boldsymbol{\eta}_j^T D\{\tilde{\boldsymbol{e}}\}\boldsymbol{\eta}_j = E\{\tilde{e}_j^2\}, \tag{9.14b}$$

with

$$\boldsymbol{\eta}_j \coloneqq \big[0, \ldots, 0, \underset{j\text{th}}{1}, 0, \ldots, 0\big]^T, \tag{9.14c}$$

as a unit vector that serves to extract the jth diagonal element from the dispersion matrix. Then, the jth standardized residual is defined as

$$\boxed{\tilde{z}_j \coloneqq \tilde{e}_j / \sigma_{\tilde{e}_j}.} \tag{9.15}$$

Since the variance component σ_0^2 is considered unknown in the model (9.13a), we replace it with its estimate $\hat{\sigma}_0^2$, leading to the following analogous set of equations for the *studentized residual*:

$$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n - \mathrm{rk}(A)} = \frac{\boldsymbol{y}^T P \boldsymbol{y} - \boldsymbol{c}^T N^{-1} \boldsymbol{c}}{n - m}, \tag{9.16a}$$

$$\hat{D}\{\tilde{\boldsymbol{e}}\} = \hat{\sigma}_0^2\big(P^{-1} - AN^{-1}A^T\big), \tag{9.16b}$$

[1] The term *test statistic* is called *test criterion* by Snedecor and Cochran (1980, p. 65).

$$\hat{\sigma}^2_{\tilde{e}_j} = \boldsymbol{\eta}_j^T \hat{D}\{\tilde{\boldsymbol{e}}\}\boldsymbol{\eta}_j = \hat{E}\{\tilde{e}_j^2\}. \tag{9.16c}$$

Then the studentized residual is defined as

$$\boxed{\tilde{t}_j \coloneqq \tilde{e}_j/\hat{\sigma}_{\tilde{e}_j}.} \tag{9.17}$$

Note that the denominator in (9.15) is constant (due to the unknown but constant variance component σ_0^2), whereas the denominator of (9.17) is random due to the introduction of the estimate $\hat{\sigma}_0^2$, which is random. Of course the numerator is random in both cases.

Using Q to represent cofactor matrices in general, we can rewrite the standardized and studentized residuals in the following alternative forms:

$$\text{Standardized residual:} \quad \tilde{z}_j \coloneqq \tilde{e}_j \big/ \sqrt{\sigma_0^2 \big(Q_{\tilde{\boldsymbol{e}}}\big)_{jj}} \sim \mathcal{N}(0,1). \tag{9.18a}$$

$$\text{Studentized residual:} \quad \tilde{t}_j \coloneqq \tilde{e}_j \big/ \sqrt{\hat{\sigma}_0^2 \big(Q_{\tilde{\boldsymbol{e}}}\big)_{jj}} \sim t(n-1). \tag{9.18b}$$

Here $D\{\tilde{\boldsymbol{e}}\} = \sigma_0^2 Q_{\tilde{\boldsymbol{e}}}$, and $(Q_{\tilde{\boldsymbol{e}}})_{jj}$ denotes the jth diagonal element of the residual cofactor matrix $Q_{\tilde{\boldsymbol{e}}}$, and we have assumed that the standardized residuals are normally distributed, implying that the studentized residuals follow the student t-distribution. Again, it is noted that (9.18a) cannot be computed unless the variance component σ_0^2 is known.

Example: Direct observations of a single parameter μ with weight matrix $P = I_n$.

$$\begin{gathered}
\boldsymbol{y} = \boldsymbol{\tau}\mu + \boldsymbol{e}, \;\; \boldsymbol{e} \sim \mathcal{N}\big(\boldsymbol{0}, \sigma_0^2 I_n\big), \;\; \text{with } \boldsymbol{\tau} = [1, \ldots, 1]^T \\
\hat{\mu} = \frac{\boldsymbol{\tau}^T \boldsymbol{y}}{\boldsymbol{\tau}^T \boldsymbol{\tau}} = \frac{1}{n}\big(y_1 + \ldots + y_n\big) \sim \mathcal{N}\big(\mu, \sigma_0^2/n\big) \\
\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{\tau}\hat{\mu} \sim \mathcal{N}(\boldsymbol{0}, \sigma_0^2\big[I_n - n^{-1} \cdot \boldsymbol{\tau}\boldsymbol{\tau}^T\big]) \\
Q_{\tilde{\boldsymbol{e}}} = I_n - n^{-1} \cdot \boldsymbol{\tau}\boldsymbol{\tau}^T \\
\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T \tilde{\boldsymbol{e}}}{(n-1)}
\end{gathered}$$

The formula for $Q_{\tilde{\boldsymbol{e}}}$ in the above example means that $(Q_{\tilde{\boldsymbol{e}}})_{jj} = (n-1)/n$, which shows that the more observations we have (i.e., the larger n is), the more the dispersion of the *predicted* random error $D\{\tilde{\boldsymbol{e}}\}$ approaches the dispersion of the *true* random error $D\{\boldsymbol{e}\}$. In this example the standardized and studentized residuals are written as follows:

$$\text{Standardized:} \quad \tilde{z}_j = \frac{\tilde{e}_j}{\sqrt{\sigma_0^2 (Q_{\tilde{\boldsymbol{e}}})_{jj}}} = \frac{\tilde{e}_j \sqrt{n}}{\sigma_0 \sqrt{n-1}} \sim \mathcal{N}(0,1). \tag{9.19a}$$

$$\text{Or, alternatively:} \quad \tilde{z}_j = \frac{\tilde{e}_j}{\sqrt{(Q_{\tilde{\boldsymbol{e}}})_{jj}}} = \frac{\tilde{e}_j \sqrt{n}}{\sqrt{n-1}} \sim \mathcal{N}(0, \sigma_0^2). \tag{9.19b}$$

$$\text{Studentized:} \quad \tilde{t}_j = \frac{\tilde{e}_j}{\sqrt{\hat{\sigma}_0^2 (Q_{\tilde{\boldsymbol{e}}})_{jj}}} = \frac{\tilde{e}_j}{\sqrt{\tilde{\boldsymbol{e}}^T \tilde{\boldsymbol{e}}}} \sqrt{n} \sim t(n-1). \tag{9.19c}$$

We extend the example by including a hypothesis test for the parameter estimate $\hat{\mu}$ against a specified value μ_0 at a significance level α.

$$\begin{aligned} &\text{Hypothesis test:} && H_0 : E\{\hat{\mu}\} = \mu_0 \ \text{ against } H_A : E\{\hat{\mu}\} \neq \mu_0. \\ &\text{Test statistic:} && t = \frac{\hat{\mu} - \mu_0}{\sqrt{\hat{\sigma}_0^2}}\sqrt{n} \sim t(n-1). \end{aligned}$$

We accept the null hypothesis H_0 if $t_{-\alpha/2} \leq t \leq t_{\alpha/2}$; otherwise we reject H_0. We may perform a similar test $H_0 : E\{\tilde{e}_j\} = 0$ for the jth residual. In this case the test statistic is the studentized residual computed by (9.19c).

9.2 Hypothesis Testing Within the Gauss-Markov Model

The hypothesis test introduced in Section 9.1 for direct observations of a single parameter is now extended to the Gauss-Markov Model (GMM). In introducing the GMM in Chapter 3, a probability density function was not given for the random observation errors; only the first and second moments of the random errors were specified. This is indeed all that is necessary to formulate and solve the least-squares estimation problem within the GMM. However, in order to perform classical hypothesis testing after the least-squares estimate has been computed, a probability distribution must be specified. Typically, we assume that the observation errors have a *normal distribution*. Then, the (full rank) GMM is written succinctly as

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}, \ \ \mathrm{rk}\,A = m, \ \ \boldsymbol{e} \sim \mathcal{N}\big(\mathbf{0}, \sigma_0^2 P^{-1}\big). \tag{9.20}$$

where the symbol $\mathcal{N}$ denotes the normal distribution.

Minimization of the observation errors via a least-squares adjustment leads to the following vectors of parameter estimates and predicted random-errors, shown with their corresponding normal distributions (normal because the distribution of the observations were assumed to be normal for the sake of hypothesis testing):

$$\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c} \sim \mathcal{N}\big(\boldsymbol{\xi}, \sigma_0^2 N^{-1}\big), \tag{9.21a}$$

$$\tilde{\boldsymbol{e}} = \big(I_n - AN^{-1}A^T P\big)\boldsymbol{y} \sim \mathcal{N}\big(\mathbf{0}, \sigma_0^2\big[P^{-1} - AN^{-1}A^T\big]\big). \tag{9.21b}$$

Or equivalently, we could write for the predicted residual vector

$$\tilde{\boldsymbol{e}} = \big(I_n - AN^{-1}A^T P\big)\boldsymbol{e} = Q_{\tilde{\boldsymbol{e}}} P\boldsymbol{y} \sim \mathcal{N}\big(\mathbf{0}, \sigma_0^2 Q_{\tilde{\boldsymbol{e}}}\big), \tag{9.22a}$$

with its cofactor matrix provided by

$$Q_{\tilde{\boldsymbol{e}}} \coloneqq P^{-1} - AN^{-1}A^T. \tag{9.22b}$$

The jth standardized and studentized residuals are then written as

$$j\text{th standardized residual:} \quad \tilde{z}_j \coloneqq \tilde{e}_j \big/ \sqrt{\sigma_0^2 (Q_{\tilde{\boldsymbol{e}}})_{jj}} \sim \mathcal{N}(0,1), \tag{9.23}$$

$$j\text{th studentized residual:} \qquad \tilde{t}_j \coloneqq \tilde{e}_j \Big/ \sqrt{\hat{\sigma}_0^2 (Q_{\tilde{\boldsymbol{e}}})_{jj}} \sim t(n-m). \tag{9.24}$$

As shown in Chapter 3, we compute the estimated reference variance within the GMM by

$$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n-m}, \tag{9.25}$$

where $n-m$ is the redundancy of the model. The hypothesis test for the jth studentized residual then becomes

$$H_0 : E\{\tilde{e}_j\} = 0 \;\text{ versus }\; H_A : E\{\tilde{e}_j\} \neq 0. \tag{9.26}$$

Likewise, we may test individual elements of the estimated parameter vector $\hat{\boldsymbol{\xi}}$. For example, we may want to compare the jth element of the estimated parameter vector, $\hat{\xi}_j$, against some specified value $\xi_j^{(0)}$. In this case, the null hypothesis and computed test statistic are defined as follows:

$$H_0 : E\{\hat{\xi}_j\} = \xi_j^{(0)} \;\text{ versus }\; H_A : E\{\hat{\xi}_j\} \neq \xi_j^{(0)}, \tag{9.27a}$$

$$t_j = \frac{\hat{\xi}_j - \xi_j^{(0)}}{\sqrt{\hat{\sigma}_0^2 \big(N^{-1}\big)_{jj}}} \sim t(n-m), \tag{9.27b}$$

or

$$t_j^2 = \frac{\big(\hat{\xi}_j - \xi_j^{(0)}\big)\big[\big(N^{-1}\big)_{jj}\big]^{-1}\big(\hat{\xi}_j - \xi_j^{(0)}\big)/1}{(\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}})/(n-m)} \sim F(1, n-m). \tag{9.27c}$$

From (9.27b) and (9.27c) we see that in this case the square of the test statistic having a Student's t-distribution has an F-distribution.

For a given significance level α, we accept H_0 if $t_{-\alpha/2} \leq t_j \leq t_{\alpha/2}$; otherwise we reject H_0. We can use a cdf table for the t-distribution to find the value of $t_{\alpha/2}(n-m)$. Note that α is the probability of making a Type I error (also called the significance level of the test), and $n-m$ is the degrees of freedom associated with $\hat{\sigma}_0^2$; for the F-distribution, 1 is the degrees of freedom associated with the numerator.

9.3 Confidence Intervals for Ellipses, Ellipsoids, and Hyperellipsoids

After we estimate the mean of a population, or the parameter of a data model, we might then like to make a statement about the accuracy of the estimated value. In statistics, a *probability statement* gives the probability that the estimated quantity falls within a certain interval centered on the true, but unknown mean (or model parameter). Such an interval is called a *confidence interval*, and its upper and lower bounds are called *confidence limits*. Confidence ellipses, ellipsoids, and hyperellipsoids are the respective 2-D, 3-D, and n-D analogues to confidence intervals.

9.3.1 Confidence Intervals — Univariate Case

By definition, the cdf (cumulative distribution function) of a random variable X is

$$F_X(x) = P(X \leq x), \quad -\infty < x < \infty, \tag{9.28}$$

which provides the probability that the unknown quantity X is less than or equal to the sampled value x. It follows, then, that the probability that X lies within the interval $(a, b]$ is

$$P(a < X \leq b) = F_X(b) - F_X(a). \tag{9.29}$$

Applying (9.29) to the standard normal random variable z of (9.5), we can write the following probabilities for confidence intervals bounded by $\pm 1\sigma$, $\pm 2\sigma$, $\pm 3\sigma$, respectively, from the mean, where $\sigma = 1$ since $z \sim \mathcal{N}(0, 1)$ according to (9.6a) and (9.6b):

$$P(-1 < z \leq 1) = P(\mu - \ \sigma < y \leq \mu + \ \sigma) = 68.3\% \tag{9.30a}$$

$$P(-2 < z \leq 2) = P(\mu - 2\sigma < y \leq \mu + 2\sigma) = 95.5\% \tag{9.30b}$$

$$P(-3 < z \leq 3) = P(\mu - 3\sigma < y \leq \mu + 3\sigma) = 99.7\% \tag{9.30c}$$

The intervals associated with these probability statements are commonly referred to as the "1-sigma," "2-sigma," and "3-sigma" confidence intervals, respectively. Other commonly used intervals are the so-called 90%, 95%, and 99% confidence intervals. For a normally distributed random variable z, their respective probability statements are

$$90\% = P(-1.645 < z \leq 1.645), \tag{9.31a}$$

$$95\% = P(-1.960 < z \leq 1.960), \tag{9.31b}$$

$$99\% = P(-2.576 < z \leq 2.576). \tag{9.31c}$$

Probability limits correspond to the area under the graph of the associated pdf. For example, the area between $\pm\sigma$ under the graph of the standard normal distribution shown in Figure 9.2 is 0.683, and it is 0.997 for $\pm 3\sigma$. The regions beyond these areas are called the *tails* of the graph. Figure 1.2 depicts a graphical representation of the areas comprising $\pm\sigma$, $\pm 2\sigma$, and $\pm\ 3\sigma$. It is shown again in Figure 9.3 for convenience.

9.3.2 Confidence Ellipses — Bivariate Case

Now let us consider the bivariate (2-D) case where $\boldsymbol{y}$ is a random 2-D vector and $\boldsymbol{\mu}$ is its expected value; i.e., $\boldsymbol{\mu} = E\{\boldsymbol{y}\}$. Also, the dispersion of $\boldsymbol{y}$ is given by a 2×2 dispersion matrix Σ. In summary, we have

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = E\{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\}, \quad \Sigma \coloneqq D\{\boldsymbol{y}\} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}, \quad \sigma_{12} = \sigma_{21}. \tag{9.32}$$

When speaking of the elements of the vectors and matrix in (9.32), we say that μ_1 is the expected value of y_1; σ_1^2 is the variance of y_1 (with σ_1 called standard deviation), and σ_{12} is the covariance between y_1 and y_2.

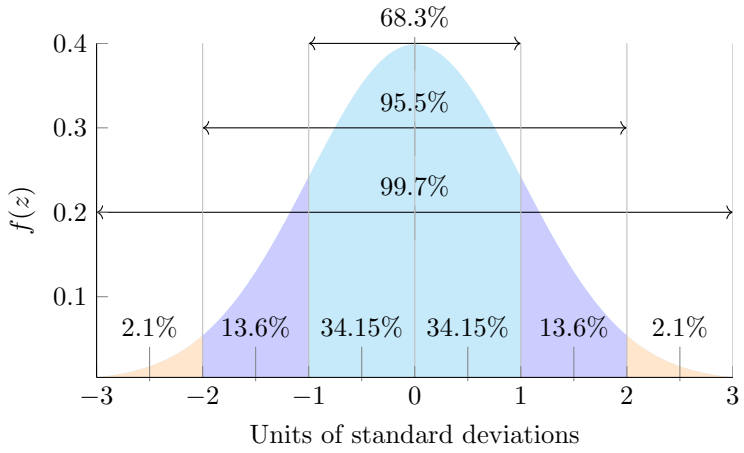

Figure 9.3: Normal distribution curve, with percent of areas under curve denoting probabilities. Image derived from Ti*k*Z code by John Canning, Senior Lecturer at the University of Brighton (http://johncanning.net/wp/?p=1202).

The 2-D analogue to a confidence interval is a confidence ellipse, which can be generated from

$$(\boldsymbol{y}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = \tag{9.33a}$$

$$= \frac{1}{(1-\rho_{12}^2)}\left(\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho_{12}\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right) = k^2, \tag{9.33b}$$

where k is a constant, and ρ is the *correlation coefficient* defined by

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2}. \tag{9.34}$$

By varying k, we generate a family of ellipses, each having an associated constant probability. Setting $k = 1$ results in the *standard confidence ellipse*. The ellipses actually originate by slicing the surface associated with a bivariate density function (pdf) with a plane parallel to the (y_1, y_2)-coordinate plane as described in the following.

Using the terms defined in (9.32), together with equation (9.7), we can write the *joint pdf* (or bivariate density function) of $\boldsymbol{y}$ explicitly as

$$f(\boldsymbol{y}) = f(y_1,y_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2-\sigma_{12}^2}}\cdot$$
$$\cdot\exp\left\{-\frac{\sigma_1^2\sigma_2^2}{2(\sigma_1^2\sigma_2^2-\sigma_{12}^2)}\left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\sigma_{12}\frac{(y_1-\mu_1)}{\sigma_1^2}\frac{(y_2-\mu_2)}{\sigma_2^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right]\right\} = \tag{9.35a}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \cdot$$
$$\cdot \exp\left\{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho_{12}\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right]\right\}, \tag{9.35b}$$

where exp stands for the exponential function, e.g., $\exp\{x\} = e^x$. The density function has the form of a bell-shaped surface over the (y_1, y_2)-coordinate plane, centered at (μ_1, μ_2). By ignoring ρ, the respective *marginal pdf's* $f(y_1)$ and $f(y_2)$ can be written as

$$f(y_1) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2\right\}, \tag{9.36a}$$

and

$$f(y_2) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right\}. \tag{9.36b}$$

The bivariate density function $f(y_1, y_2)$ and the marginal density functions $f(y_1)$ and $f(y_1)$ are depicted in Figure 9.4 with ellipses traced out by slicing planes.

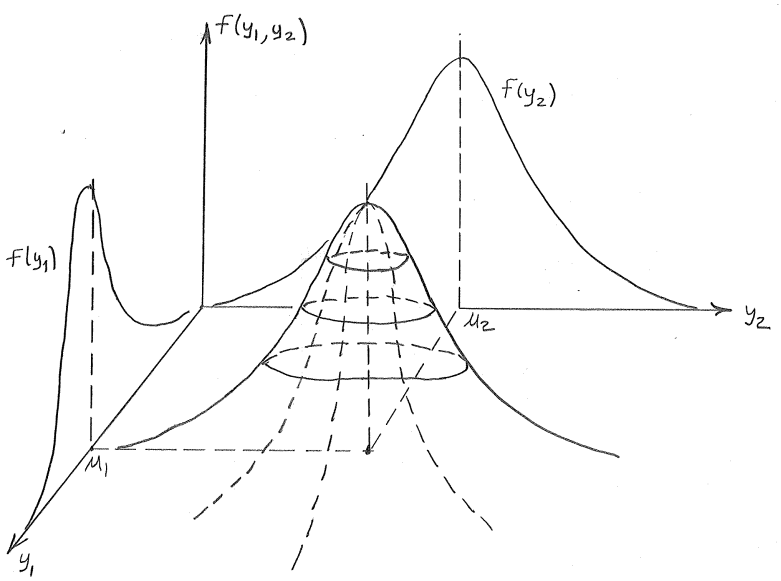

Figure 9.4: Bivariate and marginal density functions (pdf's) with ellipses traced from slicing planes, after Mikhail and Gracie (1981, p. 221)

Each element of the vector $\boldsymbol{y}$ may be normalized according to (9.5), so that the jth element of the normalized vector $\boldsymbol{z}$ is expressed in terms of the corresponding jth element of $\boldsymbol{y}$; that is $z_j = (y_j - \mu_j)/\sigma_j$, $j = 1, 2$. Substituting z_j into (9.35b)

we can write the following pdf for the normalized 2-D vector $\boldsymbol{z}$:

$$f(z_1, z_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \cdot \exp\left\{-\frac{1}{2\big(1-\rho_{12}^2\big)}\big(z_1^2 - 2\rho_{12}z_1z_2 + z_2^2\big)\right\}. \quad (9.37)$$

As noted above, a family of ellipses can be generated by slicing the bell-shaped surface generated by the density function (9.37) with planes parallel to the (y_1, y_2)-coordinate plane. The formula for the ellipse can be defined by setting the density function to a constant value related to the height of the slicing plane, which after some simplification results in an equation of the form (9.33b). According to Mikhail and Gracie (1981, p. 221), the relationship between the height h of the slicing plane above the (y_1, y_2)-coordinate plane and the constant k in (9.33b) is given by $k^2 = \ln[4\pi^2h^2\sigma_1^2\sigma_2^2(1-\rho_{12}^2)]^{-1}$. Setting $k = 1$ gives the equation for the standard confidence ellipse as follows:

$$z_1^2 - 2\rho_{12}z_1z_2 + z_2^2 = 1 - \rho_{12}^2. \quad (9.38)$$

The size, shape, and orientation of the confidence ellipse are determined by the eigenvalues and eigenvectors of the dispersion matrix Σ.

9.3.2.1 Eigenvector-eigenvalue decomposition of Σ

The eigenvector-eigenvalue decomposition of the 2×2 matrix Σ is described as follows: Denote the eigenvectors of Σ as $\boldsymbol{u}_j$ and the eigenvalues as λ_j, $j = 1, 2$. Then we have the relation

$$\Sigma\boldsymbol{u}_j = \lambda\boldsymbol{u}_j, \quad (9.39)$$

for which we write the following characteristic equation:

$$\det\big(\Sigma - \lambda I_2\big) = \big(\sigma_1^2 - \lambda\big)\big(\sigma_2^2 - \lambda\big) - \sigma_{12}^2 = \lambda^2 - \big(\sigma_1^2 + \sigma_2^2\big)\lambda + \big(\sigma_1^2\sigma_2^2 - \sigma_{12}^2\big) = 0. \quad (9.40)$$

In (9.40), λ has been used in general to represent either eigenvalue λ_1 or λ_2. By convention, we require $\lambda_1 \geq \lambda_2 > 0$ and write the following solution for the roots of the characteristic equation (9.40):

$$\lambda_{1 \text{ or } 2} = \frac{\sigma_1^2 + \sigma_2^2}{2} \pm \sqrt{\left(\frac{\sigma_1^2 + \sigma_2^2}{2}\right)^2 - \frac{1}{4}4\sigma_1^2\sigma_2^2 + \frac{4\sigma_{12}^2}{4}} \Rightarrow \quad (9.41a)$$

$$\lambda_{1 \text{ or } 2} = \frac{\sigma_1^2 + \sigma_2^2}{2} \pm \frac{1}{2}\sqrt{\big(\sigma_1^2 - \sigma_2^2\big)^2 + 4\sigma_{12}^2} > 0, \quad (9.41b)$$

which shows that the eigenvalues must be greater than zero, since Σ is positive definite.

Now we must find the two corresponding eigenvectors. Let the matrix U be comprised of the two eigenvectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ such that $U \coloneqq [\boldsymbol{u}_1,\ \boldsymbol{u}_2]$. Also define a diagonal matrix comprised of the corresponding eigenvalues $\Lambda \coloneqq \text{diag}(\lambda_1,\ \lambda_2)$. Then according to (9.39) we have

$$\Sigma U = U\Lambda = \quad (9.42a)$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \tag{9.42b}$$

$$= \left[\begin{array}{c|c} \sigma_1^2 u_{11} + \sigma_{12} u_{21} & \sigma_1^2 u_{12} + \sigma_{12} u_{22} \\ \hline \sigma_{12} u_{11} + \sigma_2^2 u_{21} & \sigma_{12} u_{12} + \sigma_2^2 u_{22} \end{array}\right] = \begin{bmatrix} \lambda_1 \cdot u_{11} & \lambda_2 \cdot u_{12} \\ \lambda_1 \cdot u_{21} & \lambda_2 \cdot u_{22} \end{bmatrix}. \tag{9.42c}$$

Starting by equating the first columns on each side of (9.42c), and then the second columns, we can write the following four equations in the four unknowns u_{11}, u_{12}, u_{21}, and u_{22}:

$$u_{21} = \frac{(\lambda_1 - \sigma_1^2)u_{11}}{\sigma_{12}}, \;\; u_{21} = \frac{\sigma_{12} u_{11}}{\lambda_1 - \sigma_2^2}, \;\; u_{12} = \frac{\sigma_{12} u_{22}}{\lambda_2 - \sigma_1^2}, \;\; u_{12} = \frac{(\lambda_2 - \sigma_2^2)u_{22}}{\sigma_{12}}. \tag{9.43}$$

The eigenvector $\boldsymbol{u}_1 = [u_{11}, u_{21}]^T$ defines the direction of the *semimajor axis* of the confidence ellipse, while the eigenvector $\boldsymbol{u}_2 = [u_{12}, u_{22}]^T$, orthogonal to $\boldsymbol{u}_1$, defines the *semiminor axis* direction. The square root of the eigenvalue λ_1 gives the semimajor-axis length, and the square root of the eigenvalue λ_2 gives the semiminor-axis length. Also, if θ is the angle measured counter clockwise from the positive z_1-axis to the semimajor axis of the confidence ellipse, then we can write the matrix U as

$$U = [\boldsymbol{u}_1, \; \boldsymbol{u}_2] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{9.44}$$

Using (9.43) and (9.44), the angle θ is derived as follows:

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{u_{21}}{u_{11}} = \frac{\lambda_1 - \sigma_1^2}{\sigma_{12}} = \frac{\sigma_{12}}{\lambda_1 - \sigma_2^2} = -\frac{u_{12}}{u_{22}} = \frac{\sigma_2^2 - \lambda_2}{\sigma_{12}} = \frac{\sigma_{12}}{\sigma_1^2 - \lambda_2} \;\; \text{and} \tag{9.45a}$$

$$\tan(2\theta) = \frac{2\tan\theta}{1 - \tan^2\theta} = \left(\frac{2\sigma_{12}}{\lambda_1 - \sigma_2^2}\right) \frac{1}{1 - \frac{\sigma_{12}^2}{(\lambda_1 - \sigma_2^2)^2}} \left(\frac{\lambda_1 - \sigma_2^2}{\lambda_1 - \sigma_2^2}\right) \Rightarrow \tag{9.45b}$$

$$\tan(2\theta) = \frac{2\sigma_{12}(\lambda_1 - \sigma_2^2)}{(\lambda_1 - \sigma_2^2)^2 - \sigma_{12}^2} = \frac{2\sigma_{12}(\lambda_1 - \sigma_2^2)4}{[2(\lambda_1 - \sigma_2^2)]^2 - 4\sigma_{12}^2}. \tag{9.45c}$$

By manipulating (9.41b), we have

$$2(\lambda_1 - \sigma_2^2) = (\sigma_1^2 - \sigma_2^2) \pm \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\sigma_{12}^2} \Rightarrow \tag{9.46a}$$

$$[2(\lambda_1 - \sigma_2^2)]^2 = 2(\sigma_1^2 - \sigma_2^2)^2 \pm 2(\sigma_1^2 - \sigma_2^2)\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\sigma_{12}^2} + 4\sigma_{12}^2. \tag{9.46b}$$

Substituting (9.46a) and (9.46b) into (9.45c) gives

$$\tan(2\theta) = \frac{4\sigma_{12}\left[(\sigma_1^2 - \sigma_2^2) \pm \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\sigma_{12}^2}\right]}{2(\sigma_1^2 - \sigma_2^2)\left[(\sigma_1^2 - \sigma_2^2) \pm \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\sigma_{12}^2}\right]} \Rightarrow \tag{9.47a}$$

$$\boxed{\tan(2\theta) = \frac{2\sigma_{12}}{\sigma_1^2 - \sigma_2^2}.} \tag{9.47b}$$

The sign of the numerical value of the right side of (9.47b) tells which quadrant the positive side of the semimajor axis falls in.

Returning now to the notion of ellipses of constant probability represented by (9.33a), probabilities for various values of k are most easily determined by using a transformed system of equations centered on $\boldsymbol{\mu}$ and rotated so that the y_1 and y_2 axes coincide with the axes formed by the eigenvectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Then, instead of correlated coordinates y_1 and y_2, we end up with uncorrelated coordinates u_1 and u_2 with respective variances λ_1 and λ_2 from (9.41b). And so the probability statement for being on or within an ellipse having semimajor and semiminor axes $k\sqrt{\lambda_1}$ and $k\sqrt{\lambda_2}$, respectively, is

$$P\left\{\frac{u_1^2}{\lambda_1} + \frac{u_2^2}{\lambda_2} < k^2\right\} = P\{\chi_2^2 < k^2\} = 1 - \alpha, \tag{9.48}$$

where α is a specified level of significance. Because it is assumed that u_1 and u_2 are sampled from a normal distribution, the sum of their squares has a χ_2^2 distribution. See Section 9.4.1 for a description of the χ_2^2 distribution.

Given a value for $P = 1 - \alpha$, the value of k (or visa verse) can be determined from a table of values for the χ^2 density function. Users of MATLAB® can generate P given k^2 by using $P = \text{chi2cdf}(k^2, 2)$, and k^2 given P can be generated by $k^2 = \text{chi2inv}(P, 2)$. Commonly used values are shown in Table 9.1. Compare the probability of 39.4% associated with the 1-sigma confidence ellipse to the value of 68.3% shown in (9.30a) for the 1-sigma confidence interval in the univariate case.

Table 9.1: "k-sigma" probabilities for various confidence ellipses. $P = 1 - \alpha$.

P	0.394	0.500	0.900	0.950	0.990
k	1.000	1.177	2.146	2.447	3.035

An empirical error ellipse differs from the confidence ellipse in that the matrix Σ is replaced by the estimated matrix $\hat{\Sigma}$ such that $\hat{\Sigma}^{-1} = \hat{\sigma}_0^{-2} P$, where $\hat{\sigma}_0^2$ is the estimated variance component. In this case, rather than (9.33a), the empirical error ellipse is described by

$$\frac{(\boldsymbol{y} - \hat{\boldsymbol{\mu}})^T P (\boldsymbol{y} - \hat{\boldsymbol{\mu}})}{\hat{\sigma}_0^2} = 1. \tag{9.49}$$

If we are evaluating $n/2$ number of 2-D points, so that P is of size $n \times n$, we may simply work with each of the ($n/2$ number of) 2×2 block diagonal matrices of $\hat{\sigma}_0^{-2} P$ independently to form the empirical error ellipse of each point. However, we must bear in mind that these block diagonal matrices do not tell the whole story since the off-block-diagonal elements have been ignored. In any case, it may be prudent to verify that the associated correlation-coefficients of the off-block-diagonal elements are relatively small in magnitude.

9.3.2.2 2-D examples within the Gauss-Markov Model

The following two examples apply to the Gauss-Markov Model (GMM):

1. Consider the GMM (9.20), with an associated least-squares solution and dispersion given in (9.21a). Assume that the parameter vector $\boldsymbol{\xi}$ is comprised of successive 2-D point coordinates such that $(\hat{\xi}_{2i-1},\ \hat{\xi}_{2i})$ represents the coordinate estimates of the ith point. Now, also assume that we wish to compare the estimates with given (fixed) values $(\xi^0_{2i-1},\ \xi^0_{2i})$, perhaps from published results of a previous adjustment. Then we may write the following equations for the null hypothesis and the standard error ellipse ($k = 1$), where, for convenience, $k \coloneqq 2i$ and $j \coloneqq k - 1$ are used for indices:

$$H_0 : E\{\begin{bmatrix}\hat{\xi}_j,\ \hat{\xi}_k\end{bmatrix}^T\} = \begin{bmatrix}\xi^0_j,\ \xi^0_k\end{bmatrix}^T, \tag{9.50a}$$

$$\frac{1}{\hat{\sigma}^2_0}\begin{bmatrix}\hat{\xi}_j - \xi^0_j \\ \hat{\xi}_k - \xi^0_k\end{bmatrix}^T \begin{bmatrix}N_{j,j} & N_{j,k} \\ N_{k,j} & N_{k,k}\end{bmatrix} \begin{bmatrix}\hat{\xi}_j - \xi^0_j \\ \hat{\xi}_k - \xi^0_k\end{bmatrix} = 1. \tag{9.50b}$$

2. Suppose that instead of comparing the solution to given, fixed values we want to compare the results (2-D coordinate estimates) of two adjustments. Using the previously defined indices, let the estimates of the ith point of the second adjustment be represented by $(\hat{\hat{\xi}}_j,\ \hat{\hat{\xi}}_k)$. We ask the question: is the outcome of the second adjustment statistically equivalent to the first? Unless there is statistically significant overlap of the respective error ellipses, the answer is no. The null hypothesis H_0 and the test statistic f are defined as follows:

$$H_0 : E\{\begin{bmatrix}\hat{\xi}_j,\ \hat{\xi}_k\end{bmatrix}^T\} = E\{\begin{bmatrix}\hat{\hat{\xi}}_j,\ \hat{\hat{\xi}}_k\end{bmatrix}^T\}, \tag{9.51a}$$

$$f \coloneqq \frac{1/2}{\hat{\sigma}^2_0/\sigma^2_0}\begin{bmatrix}\hat{\xi}_j - \hat{\hat{\xi}}_j \\ \hat{\xi}_k - \hat{\hat{\xi}}_k\end{bmatrix}^T D\{\begin{bmatrix}\hat{\xi}_j - \hat{\hat{\xi}}_j \\ \hat{\xi}_k - \hat{\hat{\xi}}_k\end{bmatrix}\}^{-1}\begin{bmatrix}\hat{\xi}_j - \hat{\hat{\xi}}_j \\ \hat{\xi}_k - \hat{\hat{\xi}}_k\end{bmatrix} \sim F(2, n - \operatorname{rk} A). \tag{9.51b}$$

   Here, 1/2 in the numerator reflects the first degrees of freedom, 2, owing to two elements of the parameter vector being tested. Also note that the unknown variance component σ^2_0 shown in the denominator cancels with the same term occurring in the dispersion matrix. Moreover, in computing the test statistic f, it is assumed that the estimated variance component $\hat{\sigma}^2_0$ is common to both adjustments. This assumption can be verified by a homogeneity test $H_0 : E\{\hat{\sigma}^2_0\} = E\{\hat{\hat{\sigma}}^2_0\}$, which is discussed in Section 9.4. Here, we also assume that the rank of matrix A is equivalent in both adjustments, which is equal to the number of unknown parameters m according to the model definition (9.20). Note that in the case that the two adjustments are uncorrelated, we could replace the inverted dispersion matrix of parameter differences with the inverse of the sum of the two respective dispersion matrices.

9.3.3 Confidence Ellipsoids and Hyperellipsoids — Multivariate Case

In the 3-D case, confidence ellipses are extended to confidence ellipsoids. But, in our general formulation of the GMM we may be working with any arbitrary higher-dimensional space, and thus we speak of confidence hyperellipsoids. Since 3-D and higher dimensions are natural extensions of the 2-D case, no further discussion is necessary. However, we do list probabilities associated with confidence ellipsoids for the 3-D case in Table 9.2. The table entries can be generated using the same MATLAB® commands shown in the previous section, except that the second argument must be 3 (degrees of freedom) instead of 2.

Table 9.2: "k-sigma" probabilities for various confidence ellipsoids. $P = 1 - \alpha$.

P	0.199	0.500	0.900	0.950	0.990
k	1.000	1.538	2.500	2.796	3.365

9.4 χ^2-distribution, Variance Testing, and F-distribution

This section includes the statistical topics of χ^2- and F-distributions as well as the topic of variance testing.

9.4.1 χ^2-distribution

The χ^2-distribution is attributed to the German geodesist F.R. Helmert from 1876. If we claim that the (unknown) random error vector $\boldsymbol{e}$ from the GMM is normally distributed as $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma_0^2 P^{-1})$, then the quadratic product $\boldsymbol{e}^T P \boldsymbol{e}$ has a χ^2-distribution with $\nu \coloneqq \operatorname{rk} P = n$ *degrees of freedom*, expressed by

$$\frac{\boldsymbol{e}^T P \boldsymbol{e}}{\sigma_0^2} \sim \chi^2(\nu). \tag{9.52}$$

Now, define $x \coloneqq \boldsymbol{e}^T P \boldsymbol{e}/\sigma_0^2$ (which cannot actually be computed since both $\boldsymbol{e}$ and σ_0^2 are unknown). Therefore, the pdf of x is written as

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu-2)/2} e^{-x/2} & \text{for } x > 0 \\ 0 & \text{for } x \le 0, \end{cases} \tag{9.53}$$

where e is Euler's number $2.71828\ldots$ The gamma function $\Gamma(\cdot)$ was defined in (9.12). Figure 9.5 shows plots of the χ^2-distribution for $\nu = \{1, 3, 5, 8, 10, 30\}$ with respective colors: black, magenta, cyan, red, green, blue. Note that the peaks of the curves move to the right as ν increases and that the curves appear to approximate

the normal-distribution curve as ν grows to 10 and larger. This agrees with our expectation that the χ^2-distribution is *asymptotically normal*, due to the *central limit theorem.* [2]

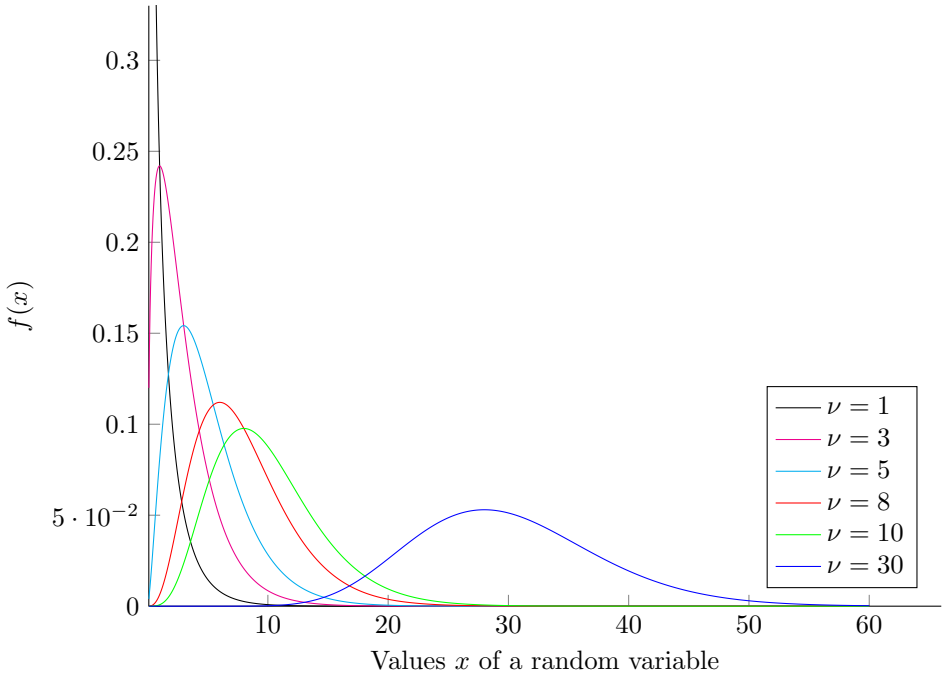

Figure 9.5: Curve of χ^2-distribution with various degrees of freedom ν

From the variance component derivations in Section 3.3, we can write

$$E\{\boldsymbol{e}^T P \boldsymbol{e}/\sigma_0^2\} = \operatorname{tr}\big(P \cdot E\{\boldsymbol{e}\boldsymbol{e}^T\}\big) = \operatorname{tr} I_n = n, \tag{9.54a}$$

$$E\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}/\sigma_0^2\} = \operatorname{tr}\big(P \cdot E\{\tilde{\boldsymbol{e}}\tilde{\boldsymbol{e}}^T\}\big) = \operatorname{tr}\big(I_n - AN^{-1}A^T P\big) = n - \operatorname{rk} A = n - m. \tag{9.54b}$$

Equations (9.25) and (9.54b) lead to

$$\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}/\sigma_0^2 = \nu\hat{\sigma}_0^2/\sigma_0^2 \sim \chi^2(\nu), \tag{9.55a}$$

with

$$\nu \coloneqq n - m \tag{9.55b}$$

as the degrees of freedom (usually denoted r for redundancy elsewhere in these notes).

[2] According to Bjerhammar (1973, Section 2.15), the central limit theorem says that "the sum of n independent stochastic variables having equal expectation and variance will have a distribution that converges towards the normal distribution for $n \to \infty$."

Note that though we have been discussing the random error vector $\boldsymbol{e}$ and the predicted residual $\tilde{\boldsymbol{e}}$, the relations expressed in (9.55a) apply to all quadratic forms in normally distributed variables. Thus, when we have a vector of normally distributed variables, the corresponding quadratic form will have a χ^2-distribution.

9.4.2 Variance Testing

Suppose we want to compare the estimated variance component $\hat{\sigma}_0^2$ to a given quantity σ^2 (in the latter, the 0-subscript is not used so as not to confuse the given value with the unknown "true value"). We do so by performing the following hypothesis test at a chosen significance level α (e.g., $\alpha = 0.05$):

$$H_0 : E\{\hat{\sigma}_0^2\} \leq \sigma^2 \text{ vs. } H_A : E\{\hat{\sigma}_0^2\} > \sigma^2 \tag{9.56a}$$

$$t := (n-m) \cdot \left(\hat{\sigma}_0^2/\sigma^2\right) \sim \chi^2(n-m) \tag{9.56b}$$

$$\text{If } t \leq \chi^2_{\alpha,n-m} \text{ accept } H_0\text{; else reject } H_0. \tag{9.56c}$$

The test as shown is referred to as a one-tailed test, because the null hypothesis only states that the expectation of the estimated quantity is less than or equal to a given value (the use of $\geq$ in H_0 would also constitute a one-tailed test). In contrast, a two-tailed test would require an equals sign in the null hypothesis. The jargon one- and two-tailed comes from the fact that $1-\alpha$ represents the area under the pdf curve left of the right tail in the one-tailed case, and it represents the area between both the left and right tails (each of which have area $\alpha/2$) in the two-tailed case.

Under the assumption that the data model is correct, if the estimate $\hat{\sigma}_0^2$ turns out statistically to be less than the given value σ^2, we deem our measurements to be more precise than that reflected in the weight matrix P. On the other hand, if $\hat{\sigma}_0^2$ proves statistically to be greater than the given value, we deem our measurements to be less precise. Usually our main concern is that $\hat{\sigma}_0^2$ reflects that our measurements are at least as precise as what is reflected by the elements of the weight matrix P, thus the use of a single-tailed hypothesis may be more commonly used in practice.

On the other hand, if we need to test for equality between the estimated variance component $\hat{\sigma}_0^2$ and a chosen value σ^2, the above hypothesis test should be modified to depend on $\alpha/2$ as follows:

$$H_0 : E\{\hat{\sigma}_0^2\} = \sigma^2 \text{ vs. } H_A : E\{\hat{\sigma}_0^2\} \neq \sigma^2 \tag{9.57a}$$

$$t := (n-m) \cdot \left(\hat{\sigma}_0^2/\sigma^2\right) \sim \chi^2(n-m) \tag{9.57b}$$

$$\text{If } \chi^2_{1-\alpha/2,n-m} < t < \chi^2_{\alpha/2,n-m} \text{ accept } H_0\text{; else reject } H_0. \tag{9.57c}$$

Note: Some tables of the χ^2 distribution list *percentiles* that equal the area under the curve less than $\chi^2_{p,\mathrm{df}}$ rather than the area under the curve right of $\chi^2_{\alpha,\mathrm{df}}$ shown in other tables (where df stands for degrees of freedom, sometimes denoted as ν). Either type of table can be used as long as the relationship $p = 1-\alpha$ is considered.

In the case where we need to compare two estimated reference variances $\hat{\sigma}^2_{0,1}$ and $\hat{\sigma}^2_{0,2}$ from two independent adjustments, we must compute a ratio of test statistics, which has an F-distribution (assuming both the numerator and denominator

have χ^2-distributions). Let t_1 and t_2 be the test statistics from the respective adjustments; then we can write

$$\frac{t_1/(n_1-m_1)}{t_2/(n_2-m_2)} = \hat{\sigma}_{0,1}^2/\hat{\sigma}_{0,2}^2 \sim F(n_1-m_1, n_2-m_2), \tag{9.58}$$

where $n_i - m_i$, $i = 1, 2$, are the respective degrees of freedom of the two independent adjustments.

9.4.3 F-distribution

The F-distribution was named for its discover R.A. Fisher (1925) by G.W. Snedacor (1935). It is a distribution for the ratio of two mutually independent random variables that have χ^2-distributions with degrees of freedom $v_1 \coloneqq m$ and $v_2 \coloneqq n-m$, respectively. The pdf of such a variable is given by

$$f(w) = \frac{\Gamma\big(\frac{m}{2}+\frac{n-m}{2}\big)m^{m/2}(n-m)^{(n-m)/2}w^{(m/2)-1}}{\Gamma(\frac{m}{2})\Gamma\big(\frac{n-m}{2}\big)(n-m+mw)^{(m/2+(n-m)/2)}} = \tag{9.59a}$$

$$= \frac{(v_1/v_2)^{v_1/2}\Gamma\big((v_1+v_2)/2\big)w^{(v_1/2)-1}}{\Gamma(v_1/2)\Gamma(v_2/2)\big(1+v_1w/v_2\big)^{(v_1+v_2)/2}}. \tag{9.59b}$$

As n becomes large compared to m, the curve of the F-distribution approaches the curve of the normal distribution.

9.5 Hypothesis Testing on the Estimated Parameters

In the GMM, we may wish to perform a global model-check by comparing a specified parameter vector $\boldsymbol{\xi}^0$ to the estimated vector $\hat{\boldsymbol{\xi}}$. In such a case, we may use as the test statistic the ratio of weighted norms of the difference vector $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0$ and the predicted residual vector $\tilde{\boldsymbol{e}}$ as follows:

$$w \coloneqq \frac{(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}^0)^T A^T P A(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}^0)}{\sigma_0^2 m} \cdot \frac{\sigma_0^2(n-m)}{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}} \sim F(m, n-m). \tag{9.60}$$

Here we have assumed that matrix A has full rank, i.e., $\operatorname{rk} A = m$. Since the numerator and denominator are statistically independent of one another, the test statistic w has an F-distribution with m and $n-m$ degrees of freedom, as shown in (9.60). Therefore, our global model-check is made by the following hypothesis test:

$$H_0 : E\{\hat{\boldsymbol{\xi}}\} = \boldsymbol{\xi}^0 \;\text{ vs. }\; H_A : E\{\hat{\boldsymbol{\xi}}\} \neq \boldsymbol{\xi}^0 \tag{9.61a}$$

$$\text{If } w \leq F_{\alpha,m,n-m} \;\text{ accept } H_0;\; \text{else reject } H_0. \tag{9.61b}$$

We now show that the numerator and denominator of w are indeed independent, as required for use of the F-distribution. To do so, we only need to show that

$$C\{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}, (\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^T (A^T P A)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})\} = 0. \tag{9.62}$$

Note that, without loss of generality, we have replaced $\boldsymbol{\xi}^0$ with $\boldsymbol{\xi}$. From (4.5e) we have $\tilde{\boldsymbol{e}} = [I_n - AN^{-1}A^TP]\boldsymbol{e}$. Therefore,

$$\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} = \boldsymbol{e}^T \big[I_n - PAN^{-1}A^T\big] P \big[I_n - AN^{-1}A^TP\big]\boldsymbol{e} = \\ = \boldsymbol{e}^T \big[P - PAN^{-1}A^TP\big]\boldsymbol{e} =: \boldsymbol{e}^T M_1 \boldsymbol{e}. \quad (9.63a)$$

Also

$$A(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \boldsymbol{e} - \tilde{\boldsymbol{e}} = \boldsymbol{e} - \big(I_n - AN^{-1}A^TP\big)\big(A\boldsymbol{\xi} + \boldsymbol{e}\big) = \big(AN^{-1}A^TP\big)\boldsymbol{e} \Rightarrow \quad (9.63b)$$

$$(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^T \big(A^TPA\big)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \boldsymbol{e}^T \big(PAN^{-1}A^T\big) P \big(AN^{-1}A^TP\big)\boldsymbol{e} = \quad (9.63c)$$

$$= \boldsymbol{e}^T \big(PAN^{-1}A^TP\big)\boldsymbol{e} =: \boldsymbol{e}^T M_2 \boldsymbol{e}. \quad (9.63d)$$

By substitution of (9.63a) and (9.63d), the condition (9.62) is equivalent to the condition that $\boldsymbol{e}^T M_1 \boldsymbol{e}$ and $\boldsymbol{e}^T M_2 \boldsymbol{e}$ are independent, which holds if, and only if,

$$\boldsymbol{e}^T M_1 D\{\boldsymbol{e}\} M_2 \boldsymbol{e} = 0, \quad (9.63e)$$

which is true since

$$\boldsymbol{e}^T \big(P - PAN^{-1}A^TP\big)\big(\sigma_0^2 P^{-1}\big)\big(PAN^{-1}A^TP\big)\boldsymbol{e} = 0. \quad (9.63f)$$

9.6 Checking an Individual Element (or 2-D or 3-D Point) of the Parameter Vector

We may use an $l \times m$ matrix K to select a subset of size l from the $m \times 1$ vector of estimated parameters $\hat{\boldsymbol{\xi}}$ for hypothesis testing as follows:

$$H_0 : E\{K\hat{\boldsymbol{\xi}}\} = K\boldsymbol{\xi}^0 = \boldsymbol{\kappa}_0, \quad (9.64a)$$

$$H_A : E\{K\hat{\boldsymbol{\xi}}\} = K\boldsymbol{\xi}^0 \neq \boldsymbol{\kappa}_0. \quad (9.64b)$$

If $l = 1$, K is a unit row vector that extracts the relevant element from the parameter vector, in which case κ_0 is simply a scalar quantity. The following examples show the matrix K used for extracting a single element, a 2-D point, and a 3-D point, respectively:

$$K \coloneqq \big[0, \ldots, 0, 1, 0, \ldots, 0\big], \text{ where 1 appears at the } j\text{th element;} \quad (9.65a)$$

$$K \coloneqq \big[0_2, \ldots, 0_2, I_2, 0_2, \ldots, 0_2\big], \text{ where } K \text{ is size } 2 \times m; \quad (9.65b)$$

$$K \coloneqq \big[0_3, \ldots, 0_3, I_3, 0_3, \ldots, 0_3\big], \text{ where } K \text{ is size } 3 \times m. \quad (9.65c)$$

For 2-D and 3-D points, the subscripts denote the dimension of the square sub-matrices (zero matrix or identity matrix), and I_n ($n \in \{2, 3\}$) is the jth sub-matrix of K, which means it "selects" the jth point from $\hat{\boldsymbol{\xi}}$.

The test statistic is then defined as

$$w := \frac{\big[K\big(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0\big)\big]^T D\{K\big(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0\big)\}^{-1}\big[K\big(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0\big)\big]/\operatorname{rk} K}{\hat{\sigma}_0^2/\sigma_0^2} = \quad (9.66a)$$

$$= \frac{\left[K\hat{\boldsymbol{\xi}} - \boldsymbol{\kappa}_0\right]^T \left[KN^{-1}K^T\right]^{-1} \left[K\hat{\boldsymbol{\xi}} - \boldsymbol{\kappa}_0\right]/l}{\hat{\sigma}_0^2} =: \frac{R/l}{(\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}})/(n-m)}. \tag{9.66b}$$

Note that σ_0^2 appears in the denominator of (9.66a) in order to cancel out the same term hidden inside the dispersion matrix in the numerator. Also note that since $\boldsymbol{\xi}^0$ is a specified (and therefore non-random) quantity to test against, the dispersion is not affected by it, i.e.,

$$D\{K(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^0)\} = D\{K\hat{\boldsymbol{\xi}}\} = \sigma_0^2 KN^{-1}K^T. \tag{9.67}$$

The symbols R and Ω are used for convenience and are analogous to the symbols introduced in Sections 5.5 and 6.4, respectively. They are statistically independent of one another and have the following distributions:

$$R \sim \chi^2(l), \;\; \Omega \sim \chi^2(n-m). \tag{9.68}$$

Statistical independence between the random variables R and Ω means that their joint pdf is equivalent to the product of their individual pdf's: $f(R,\Omega) = f(R)\cdot f(\Omega)$. Independence can be shown by following the same line of thought as that used at the end of the previous section, where M_1 remains unchanged and M_2 is now $PAN^{-1}K^T\left[KN^{-1}K^T\right]^{-1}KN^{-1}A^TP$. Therefore, the test statistic (9.66b) has an F-distribution represented by

$$w \sim F(l, n-m). \tag{9.69}$$

An alternative, more compact, form for w when $l = 1$ is given by

$$w = \frac{(\hat{\xi}_j - (\kappa_0)_j)^2}{\hat{\sigma}_0^2 \left(N^{-1}\right)_{jj}} \sim F(1, n-m). \tag{9.70}$$

The decision to accept or reject the null hypothesis is made analogous to (9.61b).

9.6.1 Non-central F-distribution

If the null hypothesis H_0 is false, the test statistic w is said to have a non-central F-distribution (denoted here as F'), which requires a non-centrality parameter θ so that $w \sim F'\left(v_1, v_2, \theta\right)$ under H_A, where v_1 and v_2 have been used to denote the degrees of freedom, in general. The qualification "under H_A" implies that we must pose a specific alternative hypothesis H_A in this case, rather than just the negation of H_0. For a one-tailed test, the area under the non-central F-distribution curve and to the right of F_α (from the F-distribution table) is denoted as β. The value of β is also the probability of making an *error of the second kind*, namely to accept the null hypothesis H_0 when the specified alternative hypothesis H_A is actually true. The quantity $1-\beta$ is known as the *power of the test*. As the value of θ increases, so does the value $1-\beta$. Below we have rewritten (9.69) for the non-central case, with the theoretical formula for 2θ following.

$$w \sim F'(l, n-m, \theta) \tag{9.71a}$$

$$2\theta = \big(K\boldsymbol{\xi} - \boldsymbol{\kappa}_0\big)^T \big(KN^{-1}K^T\big)^{-1} \big(K\hat{\boldsymbol{\xi}} - \boldsymbol{\kappa}_0\big) \tag{9.71b}$$

Note that the non-centrality property is reflected in (9.71b) by including both the true (unknown) vector of parameters $\boldsymbol{\xi}$ and its estimate $\hat{\boldsymbol{\xi}}$ in bilinear form.

9.7 Detection of a Single Outlier in the Gauss-Markov Model

A model that expresses the jth observation as a potential outlier can be written as

$$y_j = \boldsymbol{a}_j^T \boldsymbol{\xi}^{(j)} + \xi_0^{(j)} + e_j. \tag{9.72}$$

The terms of the model are described as follows

y_j is the jth element of the $n \times 1$ observation vector $\boldsymbol{y}$.

$\boldsymbol{a}_j$ is an $m \times 1$ column vector that is comprised of the m elements of the jth row of matrix A so that $[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]^T \coloneqq A$.

$\boldsymbol{\xi}^{(j)}$ denotes the $m \times 1$ vector of unknown parameters associated with that set of observations whose jth element is considered an outlier, as opposed to $\boldsymbol{\xi}$, which is associated with the same set of observations except that the jth one is not considered as an outlier.

e_j is the jth element of the unknown random error vector $\boldsymbol{e}$.

$\xi_0^{(j)}$ is an unknown (scalar) parameter that accounts for an outlier. In other words, it accounts for a non-random error in the observation. The formula for its estimate is developed below.

The following example may be illustrative: Suppose the observation y_j should have been 100 m but only a value of 10 m was recorded, then $\xi_0^{(j)}$ accounts for a 90 m blunder.

A modified GMM whose jth observation might be deemed an outlier is expressed as

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\boldsymbol{\xi}^{(j)} + \underset{n\times 1}{\boldsymbol{\eta}_j}\,\xi_0^{(j)} + \boldsymbol{e}, \quad \boldsymbol{\eta}_j \coloneqq \big[0, \ldots, 0, 1, 0, \ldots, 0\big]^T, \tag{9.73a}$$

$$\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma_0^2 P^{-1}). \tag{9.73b}$$

Note that the number 1 in $\boldsymbol{\eta}_j$ appears at the jth element; all other elements are 0. We must compare the model in (9.73) with the original GMM (3.1), which is assumed to not include an outlier. Since the model (9.73) assumes only one outlier in the data set, n comparisons of the two models are necessary in order to test all y_i $(i = 1, \ldots, n)$ observations independently. For each comparison we introduce the constraint equation

$$\xi_0^{(j)} = K \begin{bmatrix} \boldsymbol{\xi}^{(j)} \\ \xi_0^{(j)} \end{bmatrix} = \kappa_0 = 0. \tag{9.74}$$

Here $K \coloneqq [0, 0, \ldots, 1]$ is of size $1 \times (m+1)$. When we impose the constraint (9.74) upon the model (9.73), we obtain a model equivalent to the original GMM (3.1) that does not include an additional parameter to model an outlier.

Note: For the remainder of this section, we will assume that the weight matrix P is diagonal: $P = \operatorname{diag}(p_1, \ldots, p_n)$, where p_i is the weight of the ith observation. See Schaffrin (1997b) for a treatment of outlier detection with correlated observations.

Now, we begin with the following Lagrange target function to derive a least-squares estimator in the unconstrained model (9.73):

$$\Phi\big(\boldsymbol{\xi}^{(j)}, \xi_0^{(j)}\big) = \big(\boldsymbol{y} - A\boldsymbol{\xi}^{(j)} - \boldsymbol{\eta}_j \xi_0^{(j)}\big)^T P \big(\boldsymbol{y} - A\boldsymbol{\xi}^{(j)} - \boldsymbol{\eta}_j \xi_0^{(j)}\big), \tag{9.75}$$

which is made stationary with respect to $\boldsymbol{\xi}^{(j)}$ and $\xi_0^{(j)}$ by setting the first partial derivatives of (9.75) to zero, resulting in the following Euler-Lagrange necessary conditions:

$$\frac{1}{2}\left[\frac{\partial \Phi}{\partial \boldsymbol{\xi}^{(j)}}\right]^T = -A^T P \boldsymbol{y} + A^T P \boldsymbol{\eta}_j \hat{\xi}_0^{(j)} + A^T P A \hat{\boldsymbol{\xi}}^{(j)} \doteq \mathbf{0}, \tag{9.76a}$$

$$\frac{1}{2}\frac{\partial \Phi}{\partial \xi_0^{(j)}} = -\boldsymbol{\eta}_j^T P \boldsymbol{y} + \boldsymbol{\eta}_j^T P A \hat{\boldsymbol{\xi}}^{(j)} + \boldsymbol{\eta}_j^T P \boldsymbol{\eta}_j \hat{\xi}_0^{(j)} \doteq \mathbf{0}. \tag{9.76b}$$

Of course the second partial derivatives are functions of P, which is positive-definite by definition, thereby satisfying the sufficient condition required for obtaining the minimum of (9.75). In matrix form we have

$$\begin{bmatrix} N & A^T P \boldsymbol{\eta}_j \\ \boldsymbol{\eta}_j^T P A & \boldsymbol{\eta}_j^T P \boldsymbol{\eta}_j \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}^{(j)} \\ \hat{\xi}_0^{(j)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{\eta}_j^T P \boldsymbol{y} \end{bmatrix}, \tag{9.77a}$$

or, because P was assumed to be diagonal,

$$\begin{bmatrix} N & \boldsymbol{a}_j p_j \\ p_j \boldsymbol{a}_j^T & p_j \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}}^{(j)} \\ \hat{\xi}_0^{(j)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ p_j y_j \end{bmatrix}. \tag{9.77b}$$

Here, as in previous chapters, we have used the definition $[N,\, \boldsymbol{c}] \coloneqq A^T P [A,\, \boldsymbol{y}]$. Using (A.11) for the inverse of a partitioned matrix, and decomposing the resulting inverse into a sum of two matrices, results in

$$\begin{aligned} \begin{bmatrix} \hat{\boldsymbol{\xi}}^{(j)} \\ \hat{\xi}_0^{(j)} \end{bmatrix} = \begin{bmatrix} N^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ p_j y_j \end{bmatrix} + \begin{bmatrix} N^{-1} \boldsymbol{a}_j p_j \\ -1 \end{bmatrix} \big(p_j - p_j \boldsymbol{a}_j^T N^{-1} \boldsymbol{a}_j p_j\big)^{-1} \cdot \\ \cdot \begin{bmatrix} p_j \boldsymbol{a}_j^T N^{-1} & -1 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ p_j y_j \end{bmatrix}, \end{aligned} \tag{9.78a}$$

or

$$\begin{bmatrix} \hat{\boldsymbol{\xi}}^{(j)} \\ \hat{\xi}_0^{(j)} \end{bmatrix} = \begin{bmatrix} N^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{c} \\ p_j y_j \end{bmatrix} - \begin{bmatrix} N^{-1} \boldsymbol{a}_j p_j \\ -1 \end{bmatrix} \big(p_j - p_j^2 \boldsymbol{a}_j^T N^{-1} \boldsymbol{a}_j\big)^{-1} p_j \big(y_j - \boldsymbol{a}_j^T N^{-1} \boldsymbol{c}\big). \tag{9.78b}$$

From (9.78b), and recalling that $\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c}$ is based on a data set assumed to have no outliers, we can write the following difference between estimations:

$$\hat{\boldsymbol{\xi}}^{(j)} - \hat{\boldsymbol{\xi}} = -N^{-1}\boldsymbol{a}_j\left(\frac{y_j - \boldsymbol{a}_j^T\hat{\boldsymbol{\xi}}}{p_j^{-1} - \boldsymbol{a}_j^T N^{-1}\boldsymbol{a}_j}\right) = -N^{-1}\boldsymbol{a}_j\frac{\tilde{e}_j}{(Q_{\tilde{\boldsymbol{e}}})_{jj}}, \tag{9.79}$$

where $(Q_{\tilde{\boldsymbol{e}}})_{jj}$ is the jth diagonal element of the cofactor matrix for the residual vector $\tilde{\boldsymbol{e}}$. For the estimated non-random error in y_j we have

$$\hat{\xi}_0^{(j)} = \frac{y_j - \boldsymbol{a}_j^T\hat{\boldsymbol{\xi}}}{1 - p_j\boldsymbol{a}_j^T N^{-1}\boldsymbol{a}_j} = \frac{\tilde{e}_j}{(Q_{\tilde{\boldsymbol{e}}}P)_{jj}} = \frac{\tilde{e}_j/p_j}{(Q_{\tilde{\boldsymbol{e}}})_{jj}}. \tag{9.80}$$

The hypothesis test for the jth observation being an outlier is then written as

$$H_0 : E\{\hat{\xi}_0^{(j)}\} = 0 \;\text{ versus }\; H_A : E\{\hat{\xi}_0^{(j)}\} \neq 0. \tag{9.81}$$

The test statistic has an F-distribution and is computed by

$$T_j = \frac{R_j/1}{(\Omega - R_j)/(n-m-1)} \sim F(1, n-m-1). \tag{9.82}$$

The definition of R_j, in terms of $\hat{\xi}_0^{(j)}$, is

$$R_j \coloneqq \frac{\left(\hat{\xi}_0^{(j)} - 0\right)^2}{KN_1^{-1}K^T} = \frac{\left(\hat{\xi}_0^{(j)}\right)^2}{\left(p_j - p_j^2\boldsymbol{a}_j^T N^{-1}\boldsymbol{a}_j\right)^{-1}} = \frac{\tilde{e}_j^2}{(Q_{\tilde{\boldsymbol{e}}}P)_{jj}^2}p_j(Q_{\tilde{\boldsymbol{e}}}P)_{jj} = \frac{\tilde{e}_j^2}{(Q_{\tilde{\boldsymbol{e}}})_{jj}}. \tag{9.83}$$

It is important to note that the symbols $\tilde{\boldsymbol{e}}$ and $Q_{\tilde{\boldsymbol{e}}}$ represent the residual vector and its cofactor matrix, respectively, as predicted within the GMM (3.1) — see (3.9) and (3.14c). As was already mentioned, when we impose the constraint (9.74) on model (9.73b) we reach a solution identical to the LESS within model (3.1). It is also important to understand the terms in the denominator of (9.82). As stated previously, the symbol R is used to account for that portion of the P-weighted residual norm due to the constraints. The first parenthetical term in the denominator, $(\Omega - R_j)$, accounts for that part of the norm coming from the unconstrained solution. Here we have used $\Omega \coloneqq \tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}}$, with $\tilde{\boldsymbol{e}}$ belonging to the constrained solution (determined within the model (3.1)). Therefore, we must subtract R from Ω, as it is defined here, to arrive at the portion of the norm coming from the unconstrained LESS computed within model (9.73).

We note again that the equations from (9.77b) to (9.83) hold only in the case of a diagonal weight matrix P. Regardless of whether or not P is diagonal, the quantity

$$r_j \coloneqq (Q_{\tilde{\boldsymbol{e}}}P)_{jj} \tag{9.84a}$$

is the jth so-called *redundancy number*, for the unconstrained solution in this case. The following properties hold for r_j:

$$0 < r_j \leq 1 \;\text{ for } i = \{1, \ldots, n\} \;\text{ and }\; \sum_j r_j = n - \operatorname{rk} A. \tag{9.84b}$$

Note that $(Q_{\tilde{e}}P)_{jj} = (Q_{\tilde{e}})_{jj}p_j$ for the case that matrix P is diagonal.

Finally, the matrix N_1 in (9.83) is defined as

$$N_1 = \begin{bmatrix} N & \boldsymbol{a}_j p_j \\ p_j \boldsymbol{a}_j^T & p_j \end{bmatrix}, \tag{9.85}$$

which appears in (9.77b). Pre- and post-multiplying N_1^{-1} by K extracts only its last diagonal element, which, according to the formula for inverting a partitioned matrix, turns out to be the scalar quantity $(p_j - p_j^2 \boldsymbol{a}_j^T N^{-1} \boldsymbol{a}_j)^{-1}$, also appearing in (9.83).

We comment that outlier detection at the 2-D and 3-D level can also be performed, for example, in testing whether observed 2-D and 3-D points are outliers. The 3-D case is also appropriate for GPS baseline adjustments. Its development is shown by Snow (2002); see also Snow and Schaffrin (2003).

A strategy for outlier detection Since the model (9.73) only accounts for an outlier at the jth observation. A strategy is needed to check for outliers at all observations including the case where more than one outlier might be present in the observation vector $\boldsymbol{y}$. The usual way of approaching this problem is to perform n independent outlier tests, allowing j to run from 1 to n. If the null hypothesis cannot be accepted for one or more of the tests, the observation associated with the largest value for the test statistic T_j is flagged as a potential outlier and removed from the observation vector $\boldsymbol{y}$. The entire process is repeated until the null hypothesis can be accepted for all remaining observations, with n being reduced by 1 for each successive set of tests.

To be more conservative, after each set of tests that results in an observation being flagged as a potential outlier and removed, the previously removed observations are added back in one at a time (in the opposite order they were removed) to see if they can remain in the observation vector or if they once again must be removed. Eventually, one would hope to reach a point where all outliers have been detected and removed, implying that finally the null hypothesis can be accepted for all remaining residuals.

The reason for this conservative step is that an outlier at the jth element of the observation vector may result in a larger test statistic for some residuals other than $\tilde{e}_j$. To see how this could be, we repeat the formula for the vector or residuals shown in (3.9):

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} = \big(I_n - AN^{-1}A^T P\big)\boldsymbol{y} = Q_{\tilde{e}}P\boldsymbol{y} =: R\boldsymbol{y}, \tag{9.86a}$$

where the symbol R has been used to denote the matrix whose diagonal elements are the so-called redundancy numbers as shown in (9.84a). If R is expressed as matrix of column vectors, viz. $R = [\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n]$, then it is easy to see that

$$\tilde{\boldsymbol{e}} = \boldsymbol{r}_1 \cdot y_1 + \boldsymbol{r}_2 \cdot y_2 + \cdots + \boldsymbol{r}_n \cdot y_n, \tag{9.86b}$$

revealing that each element of $\tilde{\boldsymbol{e}}$ is potentially a linear combination of all the elements of $\boldsymbol{y}$ (since R is not expected to be a diagonal matrix, in general). This means

that an outlier at the jth element of $\boldsymbol{y}$ could "bleed into" residuals other than $\tilde{e}_j$, perhaps giving the impression that some different observation is an outlier, when in fact it is not. This linear relationship between the residuals and the observations attests to the challenge of successful outlier detection after an adjustment. It may or may not succeed in identifying all outliers and in avoiding wrongly misidentifying some observations as outliers when they in fact are not. The challenge of successful outlier detection underscores the importance of avoiding making errors in observations and in finding strategies to find any blunders in the data before an adjustment is performed.

Appendix A

Useful Matrix Relations and Identities

Product of transposes:

$$A^T B^T = (BA)^T \tag{A.1}$$

Transpose of inverse:

$$(A^T)^{-1} = (A^{-1})^T \tag{A.2}$$

Product of inverses:

$$A^{-1} B^{-1} = (BA)^{-1} \tag{A.3}$$

Rank of triple product: Given: $A(m \times n)$, $B(m \times m)$, $C(n \times n)$:

$$B, C \text{ nonsingular} \Rightarrow \mathrm{rk}(BAC) = \mathrm{rk}(A) \;\text{ or }\; \mathrm{rk}(BA) = \mathrm{rk}(A) \;\text{ if }\; C = I \tag{A.4}$$

Trace invariant with respect to a cyclic permutation of factors: If the product ABC is square, then the following trace operations are equivalent:

$$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB). \tag{A.5}$$

Sherman-Morrison-Woodbury-Schur formula:

$$(T - UW^{-1}V)^{-1} = T^{-1} + T^{-1}U(W - VT^{-1}U)^{-1}VT^{-1} \tag{A.6a}$$

Multiplying on the right by U and rearranging leads to the so-called push-through identity

$$T^{-1}U(W - VT^{-1}U)^{-1}W = (T - UW^{-1}V)^{-1}U. \tag{A.6b}$$

The origin of the phrase "push-through" is illustrated by the special case where $T = tI$ and $W = wI$, leading to

$$U(tI - (1/w)VU)^{-1} = (tI - (1/w)UV)^{-1}U. \tag{A.6c}$$

As a consequence of (A.6a), we also have:

$$(I \pm UW^{-1}V)^{-1} = I \mp U(W \pm VU)^{-1}V, \tag{A.7a}$$
$$(I \pm UV)^{-1} = I \mp U(I \pm VU)^{-1}V, \tag{A.7b}$$
$$(I \pm W^{-1}V)^{-1} = I \mp (W \pm V)^{-1}V, \tag{A.7c}$$
$$(I \pm V)^{-1} = I \mp (I \pm V)^{-1}V, \tag{A.7d}$$
$$(I \pm W^{-1})^{-1} = I \mp (W \pm I)^{-1}. \tag{A.7e}$$

Equations (39–43) of "Useful Matrix Equalities" (handout from Prof. Schaffrin, possibly originating from Urho A. Uotila).

$$DC(A + BDC)^{-1} = (D^{-1} + CA^{-1}B)^{-1}CA^{-1} = \tag{A.8a}$$
$$= D(I + CA^{-1}BD)^{-1}CA^{-1} = \tag{A.8b}$$
$$= DC(I + A^{-1}BDC)^{-1}A^{-1} = \tag{A.8c}$$
$$= DCA^{-1}(I + BDCA^{-1})^{-1} = \tag{A.8d}$$
$$= (I + DCA^{-1}B)^{-1}DCA^{-1} \tag{A.8e}$$

Suppose the matrices A and B in (A.8) are identity matrices, then we have

$$DC(I + DC)^{-1} = (D^{-1} + C)^{-1}C = \tag{A.9a}$$
$$= D(I + CD)^{-1}C = \tag{A.9b}$$
$$= (I + DC)^{-1}DC. \tag{A.9c}$$

Inverse of the partitioned normal equation matrix: Assume the matrix N is of full rank and is partitioned as follows:

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}. \tag{A.10}$$

The following steps lead to the inverse of N expressed in terms of the partitioned

blocks:

$$\left[\begin{array}{cc|cc} N_{11} & N_{12} & I & 0 \\ N_{21} & N_{22} & 0 & I \end{array}\right] \rightarrow \left[\begin{array}{c|c|c|c} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ \hline N_{21} & N_{22} & 0 & I \end{array}\right] \rightarrow$$

$$\left[\begin{array}{c|c|c|c} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ \hline 0 & N_{22} - N_{21}N_{11}^{-1}N_{12} & -N_{21}N_{11}^{-1} & I \end{array}\right] \rightarrow$$

$$\left[\begin{array}{cc|cc} I & N_{11}^{-1}N_{12} & N_{11}^{-1} & 0 \\ 0 & I & -\big(N_{22} - N_{21}N_{11}^{-1}N_{12}\big)^{-1}N_{21}N_{11}^{-1} & \big(N_{22} - N_{21}N_{11}^{-1}N_{12}\big)^{-1} \end{array}\right] \rightarrow$$

$$\left[\begin{array}{cc|c|c} I & 0 & N_{11}^{-1} + N_{11}^{-1}N_{12} \cdot W \cdot N_{21}N_{11}^{-1} & -N_{11}^{-1}N_{12} \cdot W \\ 0 & I & -W \cdot N_{21}N_{11}^{-1} & W \end{array}\right],$$

with $W \coloneqq (N_{22} - N_{21}N_{11}^{-1}N_{12})^{-1}$. Finally we may write

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \left[\begin{array}{c|c} N_{11}^{-1} + N_{11}^{-1}N_{12} \cdot W \cdot N_{21}N_{11}^{-1} & -N_{11}^{-1}N_{12} \cdot W \\ \hline -W \cdot N_{21}N_{11}^{-1} & W \end{array}\right]. \quad \text{(A.11)}$$

Note that other equivalent representations of this inverse exist. Taking directly from the Useful Matrix Equalities handout mentioned above, we write some additional expressions for the inverse.

$$\begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \quad \text{(A.12)}$$

$$\begin{aligned} Q_{11} &= \big(N_{11} - N_{12}N_{22}^{-1}N_{21}\big)^{-1} = && \text{(A.13a)} \\ &= N_{11}^{-1} + N_{11}^{-1}N_{12}\big(N_{22} - N_{21}N_{11}^{-1}N_{12}\big)^{-1}N_{21}N_{11}^{-1} = && \text{(A.13b)} \\ &= N_{11}^{-1} + N_{11}^{-1}N_{12}Q_{22}N_{21}N_{11}^{-1} && \text{(A.13c)} \end{aligned}$$

$$\begin{aligned} Q_{22} &= \big(N_{22} - N_{21}N_{11}^{-1}N_{12}\big)^{-1} = && \text{(A.14a)} \\ &= N_{22}^{-1} + N_{22}^{-1}N_{21}\big(N_{11} - N_{12}N_{22}^{-1}N_{21}\big)^{-1}N_{12}N_{22}^{-1} = && \text{(A.14b)} \\ &= N_{22}^{-1} + N_{22}^{-1}N_{21}Q_{11}N_{12}N_{22}^{-1} && \text{(A.14c)} \end{aligned}$$

$$\begin{aligned} Q_{12} &= -\big(N_{11} - N_{12}N_{22}^{-1}N_{21}\big)^{-1}N_{12}N_{22}^{-1} = -Q_{11}N_{12}N_{22}^{-1} = && \text{(A.15a)} \\ &= -N_{11}^{-1}N_{12}\big(N_{22} - N_{21}N_{11}^{-1}N_{12}\big)^{-1} = -N_{11}^{-1}N_{12}Q_{22} && \text{(A.15b)} \end{aligned}$$

$$Q_{21} = -N_{22}^{-1}N_{21}\big(N_{11} - N_{12}N_{22}^{-1}N_{21}\big)^{-1} = -N_{22}^{-1}N_{21}Q_{11} = \tag{A.16a}$$

$$= -\big(N_{22} - N_{21}N_{11}^{-1}N_{12}\big)^{-1}N_{21}N_{11}^{-1} = -Q_{22}N_{21}N_{11}^{-1} \tag{A.16b}$$

In the case that $N_{22} = 0$, we have:

$$Q_{22} = -\big(N_{21}N_{11}^{-1}N_{12}\big)^{-1} \tag{A.17a}$$

$$Q_{11} = N_{11}^{-1} + N_{11}^{-1}N_{12}Q_{22}N_{21}N_{11}^{-1} \tag{A.17b}$$

$$Q_{12} = -N_{11}^{-1}N_{12}Q_{22} \tag{A.17c}$$

$$Q_{21} = -Q_{22}N_{21}N_{11}^{-1} \tag{A.17d}$$

Schur Complement: the parenthetical term $\big(N_{22} - N_{21}N_{11}^{-1}N_{12}\big)$ shown above is called the Schur Complement of N_{11}. In general, given the partitioned matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{A.18a}$$

if matrix D is invertible, the Schur complement of D is

$$S_1 = A - BD^{-1}C. \tag{A.18b}$$

Likewise, if matrix A is invertible, the Schur complement of A is

$$S_2 = D - CA^{-1}B. \tag{A.18c}$$

Fundamental Theorem of Linear Algebra: If A is a matrix of size $n \times m$ and the rank of A is $q \coloneqq \mathrm{rk}(A)$, then:

1. $\mathcal{R}(A) =$ column space of A; dimension q (A.19a)
2. $\mathcal{N}(A) =$ nullspace of A; dimension $m - q$ (A.19b)
3. $\mathcal{R}(A^T) =$ row space of A; dimension q (A.19c)
4. $\mathcal{N}(A^T) =$ left nullspace of A; dimension $n - q$ (A.19d)

Derivative of quadratic form:
While some authors write the derivative of a quadratic form (a scalar-valued vector function) with respect to a column vector as a row vector, we write such a derivative as a column vector. This is in agreement with the following authors: Grafarend and Schaffrin (1993); Harville (2000, pg. 295); Koch (1999, pg. 69); Lütkepohl (1996, pg. 175); Strang and Borre (1997, pg. 300). For example, given $\boldsymbol{x} \in \mathbb{R}^n$ and $Q \in \mathbb{R}^{n\times n}$, we have

$$\Phi(\boldsymbol{x}) = \boldsymbol{x}^T Q\boldsymbol{x} \Rightarrow \frac{\partial \Phi}{\partial \boldsymbol{x}} = 2Q\boldsymbol{x}. \tag{A.20}$$

Vector and matrix norms For a real number $p \geq 1$, the p-norm, or L^p-norm, of the $n \times 1$ vector $\boldsymbol{x}$ is defined by

$$\|\boldsymbol{x}\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}. \tag{A.21}$$

Special cases:

1. $p = 1$, 1-norm or L^1-norm:

$$\|\boldsymbol{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n| \tag{A.22a}$$

2. $p = 2$, 2-norm or L^2-norm (Euclidean distance/norm):

$$\boxed{\|\boldsymbol{x}\|_2 = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}} \tag{A.22b}$$

3. $p = \infty$, ∞-norm or L^∞-norm ("infinity norm"):

$$\|\boldsymbol{x}\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\} \tag{A.22c}$$

In a similar way, entry-wise matrix norms for a $n \times m$ matrix A are defined by

$$\|A\|_p = \|\operatorname{vec} A\|_p = \left(\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|^p\right)^{1/p}, \tag{A.23}$$

where vec is the operator that turns a matrix into a vector by stacking its columns on top of each other from the first to the last.

Special cases:

1. $p = 2$, "Frobenius norm":

$$\boxed{\|A\|_2 = \|A\|_F = \sqrt{\operatorname{tr}(A^T A)}} \tag{A.24a}$$

2. $p = \infty$, Max norm:

$$\|A\|_\infty = \|A\|_{\max} = \max_{i,j}[a_{ij}] \tag{A.24b}$$

Determinants and inverses of 2×2 and 3×3 matrices

For a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the determinant is defined by

$$\det A = |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \tag{A.25a}$$

The inverse of A can be found by

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{A.25b}$$

Writing a 3×3 matrix A as

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, \tag{A.26a}$$

the determinant of A is found by

$$\begin{aligned} \det A = |A| &= \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = \\ &= +a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} = \\ &= -d \begin{vmatrix} b & c \\ h & i \end{vmatrix} + e \begin{vmatrix} a & c \\ g & i \end{vmatrix} - f \begin{vmatrix} a & b \\ g & h \end{vmatrix} = \\ &= +g \begin{vmatrix} b & c \\ e & f \end{vmatrix} - h \begin{vmatrix} a & c \\ d & f \end{vmatrix} + i \begin{vmatrix} a & b \\ d & e \end{vmatrix}. \end{aligned} \tag{A.26b}$$

The inverse of A is found by

$$\begin{aligned} A^{-1} = \frac{1}{|A|} \begin{bmatrix} +\begin{vmatrix} e & f \\ h & i \end{vmatrix} & -\begin{vmatrix} d & f \\ g & i \end{vmatrix} & +\begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ -\begin{vmatrix} b & c \\ h & i \end{vmatrix} & +\begin{vmatrix} a & c \\ g & i \end{vmatrix} & -\begin{vmatrix} a & b \\ g & h \end{vmatrix} \\ +\begin{vmatrix} b & c \\ e & f \end{vmatrix} & -\begin{vmatrix} a & c \\ d & f \end{vmatrix} & +\begin{vmatrix} a & b \\ d & f \end{vmatrix} \end{bmatrix}^T &= \frac{1}{|A|} \begin{bmatrix} \begin{vmatrix} e & f \\ h & i \end{vmatrix} & \begin{vmatrix} c & b \\ i & h \end{vmatrix} & \begin{vmatrix} b & c \\ e & f \end{vmatrix} \\ \begin{vmatrix} f & d \\ i & g \end{vmatrix} & \begin{vmatrix} a & c \\ g & i \end{vmatrix} & \begin{vmatrix} c & a \\ f & d \end{vmatrix} \\ \begin{vmatrix} d & e \\ g & h \end{vmatrix} & \begin{vmatrix} b & a \\ h & g \end{vmatrix} & \begin{vmatrix} a & b \\ d & f \end{vmatrix} \end{bmatrix} = \\ &= \frac{1}{|A|} \begin{bmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{bmatrix}. \end{aligned} \tag{A.26c}$$

Appendix B

Linearization

A truncated Taylor series is frequently used to *linearize* a nonlinear function. Reader's will remember the series for the univariate case from calculus. As a review, we present both Taylor's theorem and series, as well as quadratic and linear approximations to functions based on truncations of the series. Then we show the extension of the liner approximation to the multivariate cases using matrices.

B.1 Taylor's Theorem and Series for the Univariate Case

If the function f and its first n derivatives f', f'', $\ldots$, $f^{(n)}$ are continuous on the interval $[a, b]$ and if $f^{(n)}$ is differentiable on (a, b), then there exists a number c_{n+1} between a and b such that

$$f(b) = f(a) + f'(a)(b-a) + \frac{f''(a)}{2}(b-a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(b-a)^n + \\ + \frac{f^{(n+1)}(c_{n+1})}{(n+1)!}(b-a)^{n+1}. \quad \text{(B.1)}$$

Taylor series The Taylor series itself, for f at $x = a$, is given by

$$f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(b-a)^n + \ldots \quad \text{(B.2)}$$

Quadratic approximation A quadratic approximation of $f(x)$ near $x = a$ is

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2, \tag{B.3a}$$

with an error $e_2(x)$ that satisfies

$$\left|e_2(x)\right| \leq \frac{\left|\max f'''(c)\right|}{6}|x-a|^3, \quad c \text{ between } a \text{ and } x. \tag{B.3b}$$

Linear approximation Likewise, a linear approximation of $f(x)$ near $x = a$ is

$$f(x) \approx f(a) + f'(a)(x-a), \tag{B.4a}$$

with an error $e_1(x)$ that satisfies

$$\left|e_1(x)\right| \leq \frac{\left|\max f''(c)\right|}{2}(x-a)^2, \quad c \text{ between } a \text{ and } x. \tag{B.4b}$$

B.2 Linearization: A Truncated Taylor's Series for the Multivariate Case

Let $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\Xi})$ represent an $n \times 1$ set of non-linear functions of the independent $m \times 1$ vector $\boldsymbol{\Xi}$. Assume that the functions $\boldsymbol{f}$ are continuous over the interval $[\boldsymbol{\Xi}, \boldsymbol{\Xi}_0]$ and that their first derivatives exist over the interval $(\boldsymbol{\Xi}, \boldsymbol{\Xi}_0)$. Then, a linear approximation of $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\Xi})$ near $\boldsymbol{\Xi} = \boldsymbol{\Xi}_0$ is given by

$$\boldsymbol{y} \approx \boldsymbol{f}(\boldsymbol{\Xi}_0) + \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\Xi}^T}\right|_{\boldsymbol{\Xi}_0} \cdot(\boldsymbol{\Xi} - \boldsymbol{\Xi}_0), \tag{B.5a}$$

which, after introduction of the incremental vector $\boldsymbol{\xi} \coloneqq \boldsymbol{\Xi} - \boldsymbol{\Xi}_0$ and the $n \times m$ matrix $A \coloneqq \partial \boldsymbol{f}/\partial \boldsymbol{\Xi}^T$, can be rewritten as

$$\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\Xi}_0) \approx A\boldsymbol{\xi}. \tag{B.5b}$$

More detailed representations of $\boldsymbol{f}(\boldsymbol{\Xi}_0)$ and A are as follows:

$$\underset{n\times 1}{\boldsymbol{f}(\boldsymbol{\Xi}_0)} = \begin{bmatrix} f_1(\Xi_1^0, \ldots, \Xi_m^0) \\ \vdots \\ f_n(\Xi_1^0, \ldots, \Xi_m^0) \end{bmatrix}, \quad \underset{n\times m}{A} = \begin{bmatrix} \left.\frac{\partial f_1}{\partial \Xi_1}\right|_{\Xi_1^0} & \cdots & \left.\frac{\partial f_1}{\partial \Xi_m}\right|_{\Xi_m^0} \\ \vdots & & \vdots \\ \left.\frac{\partial f_n}{\partial \Xi_1}\right|_{\Xi_1^0} & \cdots & \left.\frac{\partial f_n}{\partial \Xi_m}\right|_{\Xi_m^0} \end{bmatrix}. \tag{B.6}$$

Example Distances y_1, y_2, and y_3 in the horizontal plane are given from three points with known horizontal coordinates to one new point with unknown horizontal coordinates (u, v). Using (u_1, v_1) as the coordinates of the first known point, etc., and (u_0, v_0) as an *approximation* for the unknown coordinates (u, v), linearize the distance functions $y_1 = f_1(u, v) = \sqrt{(u_1 - u)^2 + (v_1 - v)^2}$, etc.

Solution

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_{\boldsymbol{y}} - \underbrace{\begin{bmatrix} \sqrt{(u_1-u_0)^2+(v_1-v_0)^2} \\ \sqrt{(u_2-u_0)^2+(v_2-v_0)^2} \\ \sqrt{(u_3-u_0)^2+(v_3-v_0)^2} \end{bmatrix}}_{\boldsymbol{f}(\boldsymbol{\Xi}=\boldsymbol{\Xi}_0)} \approx$$

$$\approx \underbrace{\begin{bmatrix} \frac{(u_0-u_1)}{\sqrt{(u_1-u_0)^2+(v_1-v_0)^2}} & \frac{(v_0-v_1)}{\sqrt{(u_1-u_0)^2+(v_1-v_0)^2}} \\ \frac{(u_0-u_2)}{\sqrt{(u_2-u_0)^2+(v_2-v_0)^2}} & \frac{(v_0-v_2)}{\sqrt{(u_2-u_0)^2+(v_2-v_0)^2}} \\ \frac{(u_0-u_3)}{\sqrt{(u_3-u_0)^2+(v_3-v_0)^2}} & \frac{(v_0-v_3)}{\sqrt{(u_3-u_0)^2+(v_3-v_0)^2}} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} u-u_0 \\ v-v_0 \end{bmatrix}}_{\boldsymbol{\xi}}. \quad \text{(B.7)}$$

Note that we have not included a random error vector $\boldsymbol{e}$, as the focus here is on linearization, not modeling of random errors — we did not say that $\boldsymbol{y}$ is a vector of observations; we used the term given distances.

Appendix C

Statistical Tables

C.1 Values of the Standard Normal Cumulative Distribution Function

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = P[Z \leq z]$$

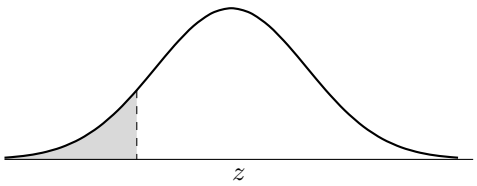

Table C.1: Probabilities $P[Z \leq z]$ computed by the MATLAB function `normcdf(z)` over the domain $[-3.09, 3.09]$, at an interval of 0.01

z	0	1	2	3	4	5	6	7	8	9
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143

Continued on next page

z	0	1	2	3	4	5	6	7	8	9
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Continued from previous page

C.2 Percentiles (Critical Values) of the t-Distribution

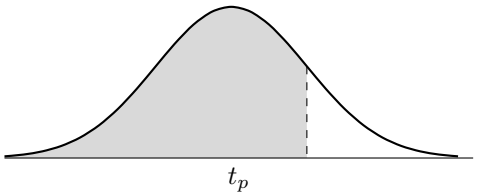

Table C.2: Percentiles (critical values) of the t-distribution computed by the MATLAB function `tinv(`p`,`ν`)` for percentile p and degrees of freedom ν

ν	$t_{0.55}$	$t_{0.60}$	$t_{0.65}$	$t_{0.70}$	$t_{0.75}$	$t_{0.80}$	$t_{0.85}$
1	0.1584	0.3249	0.5095	0.7265	1.0000	1.376	1.963
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.061	1.386
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.250
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.9410	1.190
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.156
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.134
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.8960	1.119
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.108
9	0.1293	0.2610	0.3979	0.5435	0.7027	0.8834	1.100
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.093
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.088
12	0.1283	0.2590	0.3947	0.5386	0.6955	0.8726	1.083
13	0.1281	0.2586	0.3940	0.5375	0.6938	0.8702	1.079
14	0.1280	0.2582	0.3933	0.5366	0.6924	0.8681	1.076
15	0.1278	0.2579	0.3928	0.5357	0.6912	0.8662	1.074
16	0.1277	0.2576	0.3923	0.5350	0.6901	0.8647	1.071
17	0.1276	0.2573	0.3919	0.5344	0.6892	0.8633	1.069
18	0.1274	0.2571	0.3915	0.5338	0.6884	0.8620	1.067
19	0.1274	0.2569	0.3912	0.5333	0.6876	0.8610	1.066
20	0.1273	0.2567	0.3909	0.5329	0.6870	0.8600	1.064
21	0.1272	0.2566	0.3906	0.5325	0.6864	0.8591	1.063
22	0.1271	0.2564	0.3904	0.5321	0.6858	0.8583	1.061
23	0.1271	0.2563	0.3902	0.5317	0.6853	0.8575	1.060
24	0.1270	0.2562	0.3900	0.5314	0.6848	0.8569	1.059
25	0.1269	0.2561	0.3898	0.5312	0.6844	0.8562	1.058
26	0.1269	0.2560	0.3896	0.5309	0.6840	0.8557	1.058
27	0.1268	0.2559	0.3894	0.5306	0.6837	0.8551	1.057
28	0.1268	0.2558	0.3893	0.5304	0.6834	0.8546	1.056
29	0.1268	0.2557	0.3892	0.5302	0.6830	0.8542	1.055
30	0.1267	0.2556	0.3890	0.5300	0.6828	0.8538	1.055
40	0.1265	0.2550	0.3881	0.5286	0.6807	0.8507	1.050
60	0.1262	0.2545	0.3872	0.5272	0.6786	0.8477	1.045
80	0.1261	0.2542	0.3867	0.5265	0.6776	0.8461	1.043
100	0.1260	0.2540	0.3864	0.5261	0.6770	0.8452	1.042
200	0.1258	0.2537	0.3859	0.5252	0.6757	0.8434	1.039
400	0.1257	0.2535	0.3856	0.5248	0.6751	0.8425	1.038
600	0.1257	0.2535	0.3855	0.5247	0.6749	0.8422	1.037

Continued on next page

ν	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$	$t_{0.9995}$	
800	0.1257	0.2534	0.3855	0.5246	0.6748	0.8421	1.037
1000	0.1257	0.2534	0.3854	0.5246	0.6747	0.8420	1.037
∞	0.1257	0.2533	0.3853	0.5244	0.6745	0.8416	1.036

ν	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$	$t_{0.9995}$
1	3.078	6.314	12.71	31.82	63.66	36.62
2	1.886	2.920	4.303	6.965	9.925	31.60
3	1.638	2.353	3.182	4.541	5.841	12.92
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
80	1.292	1.664	1.990	2.374	2.639	3.416
100	1.290	1.660	1.984	2.364	2.626	3.390
200	1.286	1.653	1.972	2.345	2.601	3.340
400	1.284	1.649	1.966	2.336	2.588	3.315
600	1.283	1.647	1.964	2.333	2.584	3.307
800	1.283	1.647	1.963	2.331	2.582	3.303
1000	1.282	1.646	1.962	2.330	2.581	3.300
∞	1.282	1.645	1.960	2.326	2.576	3.291

Continued from previous page

C.3 Critical Values of the χ^2-Distribution

Table C.3: Critical values of the χ^2-distribution computed by the MATLAB function `chi2inv`$(1-\alpha,\nu)$ for significance level α and degrees of freedom ν

ν	α=0.999	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.001
1	0.000	0.000	0.000	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.841	5.024	6.635	7.879	10.828
2	0.002	0.010	0.020	0.051	0.103	0.211	0.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.816
3	0.024	0.072	0.115	0.216	0.352	0.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.467
5	0.210	0.412	0.554	0.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.833	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.458
7	0.598	0.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.322
8	0.857	1.344	1.646	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.152	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	1.479	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588
11	1.834	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	21.920	24.725	26.757	31.264
12	2.214	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	23.337	26.217	28.300	32.909
13	2.617	3.565	4.107	5.009	5.892	7.042	9.299	12.340	15.984	19.812	22.362	24.736	27.688	29.819	34.528
14	3.041	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	26.119	29.141	31.319	36.123
15	3.483	4.601	5.229	6.262	7.261	8.547	11.037	14.339	18.245	22.307	24.996	27.488	30.578	32.801	37.697
16	3.942	5.142	5.812	6.908	7.962	9.312	11.912	15.338	19.369	23.542	26.296	28.845	32.000	34.267	39.252
17	4.416	5.697	6.408	7.564	8.672	10.085	12.792	16.338	20.489	24.769	27.587	30.191	33.409	35.718	40.790
18	4.905	6.265	7.015	8.231	9.390	10.865	13.675	17.338	21.605	25.989	28.869	31.526	34.805	37.156	42.312
19	5.407	6.844	7.633	8.907	10.117	11.651	14.562	18.338	22.718	27.204	30.144	32.852	36.191	38.582	43.820
20	5.921	7.434	8.260	9.591	10.851	12.443	15.452	19.337	23.828	28.412	31.410	34.170	37.566	39.997	45.315
21	6.447	8.034	8.897	10.283	11.591	13.240	16.344	20.337	24.935	29.615	32.671	35.479	38.932	41.401	46.797
22	6.983	8.643	9.542	10.982	12.338	14.041	17.240	21.337	26.039	30.813	33.924	36.781	40.289	42.796	48.268
23	7.529	9.260	10.196	11.689	13.091	14.848	18.137	22.337	27.141	32.007	35.172	38.076	41.638	44.181	49.728
24	8.085	9.886	10.856	12.401	13.848	15.659	19.037	23.337	28.241	33.196	36.415	39.364	42.980	45.559	51.179
25	8.649	10.520	11.524	13.120	14.611	16.473	19.939	24.337	29.339	34.382	37.652	40.646	44.314	46.928	52.620
26	9.222	11.160	12.198	13.844	15.379	17.292	20.843	25.336	30.435	35.563	38.885	41.923	45.642	48.290	54.052
27	9.803	11.808	12.879	14.573	16.151	18.114	21.749	26.336	31.528	36.741	40.113	43.195	46.963	49.645	55.476
28	10.391	12.461	13.565	15.308	16.928	18.939	22.657	27.336	32.620	37.916	41.337	44.461	48.278	50.993	56.892
29	10.986	13.121	14.256	16.047	17.708	19.768	23.567	28.336	33.711	39.087	42.557	45.722	49.588	52.336	58.301
30	11.588	13.787	14.953	16.791	18.493	20.599	24.478	29.336	34.800	40.256	43.773	46.979	50.892	53.672	59.703
31	12.196	14.458	15.655	17.539	19.281	21.434	25.390	30.336	35.887	41.422	44.985	48.232	52.191	55.003	61.098
32	12.811	15.134	16.362	18.291	20.072	22.271	26.304	31.336	36.973	42.585	46.194	49.480	53.486	56.328	62.487
33	13.431	15.815	17.074	19.047	20.867	23.110	27.219	32.336	38.058	43.745	47.400	50.725	54.776	57.648	63.870
34	14.057	16.501	17.789	19.806	21.664	23.952	28.136	33.336	39.141	44.903	48.602	51.966	56.061	58.964	65.247
35	14.688	17.192	18.509	20.569	22.465	24.797	29.054	34.336	40.223	46.059	49.802	53.203	57.342	60.275	66.619
36	15.324	17.887	19.233	21.336	23.269	25.643	29.973	35.336	41.304	47.212	50.998	54.437	58.619	61.581	67.985
37	15.965	18.586	19.960	22.106	24.075	26.492	30.893	36.336	42.383	48.363	52.192	55.668	59.893	62.883	69.346
38	16.611	19.289	20.691	22.878	24.884	27.343	31.815	37.335	43.462	49.513	53.384	56.896	61.162	64.181	70.703
39	17.262	19.996	21.426	23.654	25.695	28.196	32.737	38.335	44.539	50.660	54.572	58.120	62.428	65.476	72.055
40	17.916	20.707	22.164	24.433	26.509	29.051	33.660	39.335	45.616	51.805	55.758	59.342	63.691	66.766	73.402
41	18.575	21.421	22.906	25.215	27.326	29.907	34.585	40.335	46.692	52.949	56.942	60.561	64.950	68.053	74.745
42	19.239	22.138	23.650	25.999	28.144	30.765	35.510	41.335	47.766	54.090	58.124	61.777	66.206	69.336	76.084
43	19.906	22.859	24.398	26.785	28.965	31.625	36.436	42.335	48.840	55.230	59.304	62.990	67.459	70.616	77.419
44	20.576	23.584	25.148	27.575	29.787	32.487	37.363	43.335	49.913	56.369	60.481	64.201	68.710	71.893	78.750
45	21.251	24.311	25.901	28.366	30.612	33.350	38.291	44.335	50.985	57.505	61.656	65.410	69.957	73.166	80.077
46	21.929	25.041	26.657	29.160	31.439	34.215	39.220	45.335	52.056	58.641	62.830	66.617	71.201	74.437	81.400
47	22.610	25.775	27.416	29.956	32.268	35.081	40.149	46.335	53.127	59.774	64.001	67.821	72.443	75.704	82.720

Continued on next page

48	23.295	26.511	28.177	30.755	33.098	35.949	41.079	47.335	54.196	60.907	65.171	69.023	73.683	76.969	84.037
49	23.983	27.249	28.941	31.555	33.930	36.818	42.010	48.335	55.265	62.038	66.339	70.222	74.919	78.231	85.351
50	24.674	27.991	29.707	32.357	34.764	37.689	42.942	49.335	56.334	63.167	67.505	71.420	76.154	79.490	86.661
60	31.738	35.534	37.485	40.482	43.188	46.459	52.294	59.335	66.981	74.397	79.082	83.298	88.379	91.952	99.607
70	39.036	43.275	45.442	48.758	51.739	55.329	61.698	69.334	77.577	85.527	90.531	95.023	100.43	104.22	112.32
80	46.520	51.172	53.540	57.153	60.391	64.278	71.145	79.334	88.130	96.578	101.88	106.63	112.33	116.32	124.84
90	54.155	59.196	61.754	65.647	69.126	73.291	80.625	89.334	98.650	107.57	113.15	118.14	124.12	128.30	137.21
100	61.918	67.328	70.065	74.222	77.929	82.358	90.133	99.334	109.14	118.50	124.34	129.56	135.81	140.17	149.45

Continued from previous page

C.4 Critical Values of the F-Distribution

The tables below list critical values of the F-distribution computed by the MATLAB function `finv(`$1-\alpha$`,`r_1`,`r_2`)` for level of significance α and degrees of freedom $r_1 = \{1, 2, 3\}$ and r_2, where $\alpha(2)$ pertains to two tails and $\alpha(1)$ pertains to a single tail. The critical values for ∞ were generated by `finv(`$1-\alpha$`,`r_1`,1.0e6)`.

Table C.4: Critical values of the F-distribution with numerator degrees of freedom $r_1 = 1$

$\alpha(2)$:	0.5000	0.2000	0.1000	0.0500	0.0200	0.0100	0.0050	0.0020	0.0010
$\alpha(1)$:	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0025	0.0010	0.0005
r_2									
1	5.828	39.86	161.4	647.8	4052.	16210.	64840	405400.	1621000.
2	2.571	8.526	18.51	38.51	98.50	198.5	398.5	998.5	1998.
3	2.024	5.538	10.13	17.44	34.12	55.55	89.58	167.0	266.5
4	1.807	4.545	7.709	12.22	21.20	31.33	45.67	74.14	106.2
5	1.692	4.060	6.608	10.01	16.26	22.78	31.41	47.18	63.61
6	1.621	3.776	5.987	8.813	13.75	18.63	24.81	35.51	46.08
7	1.573	3.589	5.591	8.073	12.25	16.24	21.11	29.25	36.99
8	1.538	3.458	5.318	7.571	11.26	14.69	18.78	25.41	31.56
9	1.512	3.360	5.117	7.209	10.56	13.61	17.19	22.86	27.99
10	1.491	3.285	4.965	6.937	10.04	12.83	16.04	21.04	25.49
11	1.475	3.225	4.844	6.724	9.646	12.23	15.17	19.69	23.65
12	1.461	3.177	4.747	6.554	9.330	11.75	14.49	18.64	22.24
13	1.450	3.136	4.667	6.414	9.074	11.37	13.95	17.82	21.14
14	1.440	3.102	4.600	6.298	8.862	11.06	13.50	17.14	20.24
15	1.432	3.073	4.543	6.200	8.683	10.80	13.13	16.59	19.51
16	1.425	3.048	4.494	6.115	8.531	10.58	12.82	16.12	18.89
17	1.419	3.026	4.451	6.042	8.400	10.38	12.55	15.72	18.37
18	1.413	3.007	4.414	5.978	8.285	10.22	12.32	15.38	17.92
19	1.408	2.990	4.381	5.922	8.185	10.07	12.12	15.08	17.53
20	1.404	2.975	4.351	5.871	8.096	9.944	11.94	14.82	17.19
21	1.400	2.961	4.325	5.827	8.017	9.830	11.78	14.59	16.89
22	1.396	2.949	4.301	5.786	7.945	9.727	11.64	14.38	16.62
23	1.393	2.937	4.279	5.750	7.881	9.635	11.51	14.20	16.38
24	1.390	2.927	4.260	5.717	7.823	9.551	11.40	14.03	16.17
25	1.387	2.918	4.242	5.686	7.770	9.475	11.29	13.88	15.97
26	1.384	2.909	4.225	5.659	7.721	9.406	11.20	13.74	15.79
27	1.382	2.901	4.210	5.633	7.677	9.342	11.11	13.61	15.63
28	1.380	2.894	4.196	5.610	7.636	9.284	11.03	13.50	15.48
29	1.378	2.887	4.183	5.588	7.598	9.230	10.96	13.39	15.35
30	1.376	2.881	4.171	5.568	7.562	9.180	10.89	13.29	15.22
35	1.368	2.855	4.121	5.485	7.419	8.976	10.61	12.90	14.72
40	1.363	2.835	4.085	5.424	7.314	8.828	10.41	12.61	14.35
45	1.358	2.820	4.057	5.377	7.234	8.715	10.26	12.39	14.08
50	1.355	2.809	4.034	5.340	7.171	8.626	10.14	12.22	13.86
60	1.349	2.791	4.001	5.286	7.077	8.495	9.962	11.97	13.55
70	1.346	2.779	3.978	5.247	7.011	8.403	9.838	11.80	13.33
80	1.343	2.769	3.960	5.218	6.963	8.335	9.747	11.67	13.17
90	1.341	2.762	3.947	5.196	6.925	8.282	9.677	11.57	13.05
100	1.339	2.756	3.936	5.179	6.895	8.241	9.621	11.50	12.95
120	1.336	2.748	3.920	5.152	6.851	8.179	9.539	11.38	12.80

Continued on next page

$\alpha(2)$:	0.5000	0.2000	0.1000	0.0500	0.0200	0.0100	0.0050	0.0020	0.0010
$\alpha(1)$:	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0025	0.0010	0.0005
r_2									
140	1.334	2.742	3.909	5.134	6.819	8.135	9.480	11.30	12.70
160	1.333	2.737	3.900	5.120	6.796	8.102	9.437	11.24	12.63
180	1.332	2.734	3.894	5.109	6.778	8.077	9.403	11.19	12.57
200	1.331	2.731	3.888	5.100	6.763	8.057	9.377	11.15	12.52
300	1.328	2.722	3.873	5.075	6.720	7.997	9.297	11.04	12.38
500	1.326	2.716	3.860	5.054	6.686	7.950	9.234	10.96	12.28
∞	1.323	2.706	3.841	5.024	6.635	7.879	9.141	10.83	12.12

Continued from previous page, $r_1 = 1$

Table C.5: Critical values of the F-distribution with numerator degrees of freedom $r_1 = 2$

$\alpha(2)$:	0.5000	0.2000	0.1000	0.0500	0.0200	0.0100	0.0050	0.0020	0.0010
$\alpha(1)$:	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0025	0.0010	0.0005
r_2									
1	7.500	49.50	199.5	799.5	5000.	20000.	80000.	500000.	2000000.
2	3.000	9.000	19.00	39.00	99.00	199.0	399.0	999.0	1999.0
3	2.280	5.462	9.552	16.04	30.82	49.80	79.93	148.5	236.6
4	2.000	4.325	6.944	10.65	18.00	26.28	38.00	61.25	87.44
5	1.853	3.780	5.786	8.434	13.27	18.31	24.96	37.12	49.78
6	1.762	3.463	5.143	7.260	10.92	14.54	19.10	27.00	34.80
7	1.701	3.257	4.737	6.542	9.547	12.40	15.89	21.69	27.21
8	1.657	3.113	4.459	6.059	8.649	11.04	13.89	18.49	22.75
9	1.624	3.006	4.256	5.715	8.022	10.11	12.54	16.39	19.87
10	1.598	2.924	4.103	5.456	7.559	9.427	11.57	14.91	17.87
11	1.577	2.860	3.982	5.256	7.206	8.912	10.85	13.81	16.41
12	1.560	2.807	3.885	5.096	6.927	8.510	10.29	12.97	15.30
13	1.545	2.763	3.806	4.965	6.701	8.186	9.839	12.31	14.43
14	1.533	2.726	3.739	4.857	6.515	7.922	9.475	11.78	13.73
15	1.523	2.695	3.682	4.765	6.359	7.701	9.173	11.34	13.16
16	1.514	2.668	3.634	4.687	6.226	7.514	8.918	10.97	12.69
17	1.506	2.645	3.592	4.619	6.112	7.354	8.701	10.66	12.29
18	1.499	2.624	3.555	4.560	6.013	7.215	8.513	10.39	11.94
19	1.493	2.606	3.522	4.508	5.926	7.093	8.349	10.16	11.64
20	1.487	2.589	3.493	4.461	5.849	6.986	8.206	9.953	11.38
21	1.482	2.575	3.467	4.420	5.780	6.891	8.078	9.772	11.16
22	1.477	2.561	3.443	4.383	5.719	6.806	7.965	9.612	10.95
23	1.473	2.549	3.422	4.349	5.664	6.730	7.863	9.469	10.77
24	1.470	2.538	3.403	4.319	5.614	6.661	7.771	9.339	10.61
25	1.466	2.528	3.385	4.291	5.568	6.598	7.687	9.223	10.46
26	1.463	2.519	3.369	4.265	5.526	6.541	7.611	9.116	10.33
27	1.460	2.511	3.354	4.242	5.488	6.489	7.542	9.019	10.21
28	1.457	2.503	3.340	4.221	5.453	6.440	7.478	8.931	10.09
29	1.455	2.495	3.328	4.201	5.420	6.396	7.419	8.849	9.992
30	1.452	2.489	3.316	4.182	5.390	6.355	7.365	8.773	9.898
35	1.443	2.461	3.267	4.106	5.268	6.188	7.145	8.470	9.519
40	1.435	2.440	3.232	4.051	5.179	6.066	6.986	8.251	9.247
45	1.430	2.425	3.204	4.009	5.110	5.974	6.865	8.086	9.042
50	1.425	2.412	3.183	3.975	5.057	5.902	6.770	7.956	8.883
60	1.419	2.393	3.150	3.925	4.977	5.795	6.632	7.768	8.651

Continued on next page

$\alpha(2)$:	0.5000	0.2000	0.1000	0.0500	0.0200	0.0100	0.0050	0.0020	0.0010
$\alpha(1)$:	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0025	0.0010	0.0005
r_2									
70	1.414	2.380	3.128	3.890	4.922	5.720	6.535	7.637	8.489
80	1.411	2.370	3.111	3.864	4.881	5.665	6.463	7.540	8.371
90	1.408	2.363	3.098	3.844	4.849	5.623	6.409	7.466	8.281
100	1.406	2.356	3.087	3.828	4.824	5.589	6.365	7.408	8.209
120	1.402	2.347	3.072	3.805	4.787	5.539	6.301	7.321	8.103
140	1.400	2.341	3.061	3.788	4.760	5.504	6.255	7.260	8.029
160	1.398	2.336	3.053	3.775	4.740	5.478	6.222	7.215	7.974
180	1.397	2.332	3.046	3.766	4.725	5.457	6.195	7.180	7.931
200	1.396	2.329	3.041	3.758	4.713	5.441	6.175	7.152	7.897
300	1.393	2.320	3.026	3.735	4.677	5.393	6.113	7.069	7.797
500	1.390	2.313	3.014	3.716	4.648	5.355	6.064	7.004	7.718
∞	1.386	2.303	2.996	3.689	4.605	5.298	5.992	6.908	7.601

Continued from previous page, $r_1 = 2$

Table C.6: Critical values of the F-distribution with numerator degrees of freedom $r_1 = 3$

$\alpha(2)$:	0.5000	0.2000	0.1000	0.0500	0.0200	0.0100	0.0050	0.0020	0.0010
$\alpha(1)$:	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0025	0.0010	0.0005
r_2									
1	8.200	53.59	215.7	864.2	5403.0	21610.	86460.	540400.2	2162000.
2	3.153	9.162	19.16	39.17	99.17	199.2	399.2	999.2	1999.
3	2.356	5.391	9.277	15.44	29.46	47.47	76.06	141.1	224.7
4	2.047	4.191	6.591	9.979	16.69	24.26	34.96	56.18	80.09
5	1.884	3.619	5.409	7.764	12.06	16.53	22.43	33.20	44.42
6	1.784	3.289	4.757	6.599	9.780	12.92	16.87	23.70	30.45
7	1.717	3.074	4.347	5.890	8.451	10.88	13.84	18.77	23.46
8	1.668	2.924	4.066	5.416	7.591	9.596	11.98	15.83	19.39
9	1.632	2.813	3.863	5.078	6.992	8.717	10.73	13.90	16.77
10	1.603	2.728	3.708	4.826	6.552	8.081	9.833	12.55	14.97
11	1.580	2.660	3.587	4.630	6.217	7.600	9.167	11.56	13.65
12	1.561	2.606	3.490	4.474	5.953	7.226	8.652	10.80	12.66
13	1.545	2.560	3.411	4.347	5.739	6.926	8.242	10.21	11.89
14	1.532	2.522	3.344	4.242	5.564	6.680	7.910	9.729	11.27
15	1.520	2.490	3.287	4.153	5.417	6.476	7.634	9.335	10.76
16	1.510	2.462	3.239	4.077	5.292	6.303	7.403	9.006	10.34
17	1.502	2.437	3.197	4.011	5.185	6.156	7.205	8.727	9.989
18	1.494	2.416	3.160	3.954	5.092	6.028	7.035	8.487	9.686
19	1.487	2.397	3.127	3.903	5.010	5.916	6.887	8.280	9.424
20	1.481	2.380	3.098	3.859	4.938	5.818	6.757	8.098	9.196
21	1.475	2.365	3.072	3.819	4.874	5.730	6.642	7.938	8.994
22	1.470	2.351	3.049	3.783	4.817	5.652	6.539	7.796	8.816
23	1.466	2.339	3.028	3.750	4.765	5.582	6.447	7.669	8.657
24	1.462	2.327	3.009	3.721	4.718	5.519	6.364	7.554	8.515
25	1.458	2.317	2.991	3.694	4.675	5.462	6.289	7.451	8.386
26	1.454	2.307	2.975	3.670	4.637	5.409	6.220	7.357	8.269
27	1.451	2.299	2.960	3.647	4.601	5.361	6.158	7.272	8.163
28	1.448	2.291	2.947	3.626	4.568	5.317	6.100	7.193	8.066
29	1.445	2.283	2.934	3.607	4.538	5.276	6.048	7.121	7.977
30	1.443	2.276	2.922	3.589	4.510	5.239	5.999	7.054	7.894

Continued on next page

$\alpha(2)$:	0.5000	0.2000	0.1000	0.0500	0.0200	0.0100	0.0050	0.0020	0.0010
$\alpha(1)$:	0.2500	0.1000	0.0500	0.0250	0.0100	0.0050	0.0025	0.0010	0.0005
r_2									
35	1.432	2.247	2.874	3.517	4.396	5.086	5.802	6.787	7.565
40	1.424	2.226	2.839	3.463	4.313	4.976	5.659	6.595	7.329
45	1.418	2.210	2.812	3.422	4.249	4.892	5.551	6.450	7.151
50	1.413	2.197	2.790	3.390	4.199	4.826	5.466	6.336	7.013
60	1.405	2.177	2.758	3.343	4.126	4.729	5.343	6.171	6.812
70	1.400	2.164	2.736	3.309	4.074	4.661	5.256	6.057	6.673
80	1.396	2.154	2.719	3.284	4.036	4.611	5.193	5.972	6.571
90	1.393	2.146	2.706	3.265	4.007	4.573	5.144	5.908	6.493
100	1.391	2.139	2.696	3.250	3.984	4.542	5.105	5.857	6.432
120	1.387	2.130	2.680	3.227	3.949	4.497	5.048	5.781	6.341
140	1.385	2.123	2.669	3.211	3.925	4.465	5.008	5.728	6.277
160	1.383	2.118	2.661	3.199	3.906	4.441	4.977	5.689	6.230
180	1.381	2.114	2.655	3.189	3.892	4.423	4.954	5.658	6.193
200	1.380	2.111	2.650	3.182	3.881	4.408	4.936	5.634	6.164
300	1.377	2.102	2.635	3.160	3.848	4.365	4.881	5.562	6.078
500	1.374	2.095	2.623	3.142	3.821	4.330	4.838	5.506	6.010
∞	1.369	2.084	2.605	3.116	3.782	4.279	4.773	5.422	5.910

Continued from previous page, $r_1 = 3$

Appendix D

Answers to Practice Problems

The following list contains partial answers to selected practice problems.

Chapter 2 TODO

Chapter 3

3.a; 3.c $\hat{\xi} = 110.1176\,\text{m}$; $\hat{\sigma}_0^2 = 2.205883$.

4.a; 4.b $\hat{a} = 0.00252$, $\hat{b} = 0.00288$, $\hat{c} = 9.98620$, $\hat{\sigma}_0^2 = (1.987)^2$; $\hat{a} = -6.1 \times 10^{-5}$, $\hat{b} = -5.6 \times 10^{-5}$, $\hat{c} = 9.9 \times 10^{-6}$, $\hat{d} = 2.52 \times 10^{-3}$, $\hat{e} = 2.88 \times 10^{-3}$, $\hat{f} = 10.010$, $\hat{\sigma}_0^2 = 1.407^2$.

5.a; 5.d $\hat{P}_x = 72.997\,\text{m}$, $\hat{P}_y = 92.009\,\text{m}$; $\hat{\sigma}_0^2 = (0.690)^2$.

6.b; 6.b $\hat{P}_x = 1065.201\,\text{m}$, $\hat{P}_y = 825.198\,\text{m}$; $\hat{\sigma}_0^2 = (1.758)^2$.

7.a; 7.b $\hat{\xi}_1 = -0.5396$ (slope), $\hat{\xi}_2 = 5.7612$ (y-intercept); $\hat{\sigma}_0^2 = (0.316)^2$.

8.a $\hat{\xi}_1 = 168.149\,\text{mm}$, $\hat{\xi}_2 = 160.300\,\text{mm}$, $\hat{\omega}_1 = 1.000011$, $\hat{\omega}_2 = 1.000021$, $\hat{\beta} = 00°12'22.0''$, $\widehat{\beta + \epsilon} = 00°13'08.5''$.

9.b; 9.c $\hat{H}_A = 1679.509\,\text{ft}$, $\hat{H}_B = 1804.043\,\text{ft}$, $\hat{H}_C = 2021.064\,\text{ft}$, $\hat{H}_E = 1507.075\,\text{ft}$, $\hat{H}_F = 1668.148\,\text{ft}$, $H_D^0 = 1928.277\,\text{ft}$; $\hat{\sigma}_0^2 = (0.081)^2$.

9.e $\hat{H}_B = 1803.966\,\text{ft}$, $\hat{H}_C = 2020.986\,\text{ft}$, $\hat{H}_D = 1928.200\,\text{ft}$, $\hat{H}_E = 1506.998\,\text{ft}$, $\hat{H}_F = 1668.071\,\text{ft}$, $H_A^0 = 1679.432\,\text{ft}$, $\hat{\sigma}_0^2 = (0.081)^2$.

Chapter 4

3. $\tilde{\boldsymbol{e}} = [8.1, 8.8, -5.3, 3.4, -8.8, -9.4]^T$arcsec, $\hat{\sigma}_0^2 = (0.879453)^2$,

$$Q = \begin{bmatrix} 200 & 0 & 0 & 0 & 0 & 0 \\ 0 & 200 & -100 & 0 & 0 & 0 \\ 0 & -100 & 200 & 0 & 0 & 0 \\ 0 & 0 & 0 & 200 & -100 & 0 \\ 0 & 0 & 0 & -100 & 200 & 0 \\ 0 & 0 & 0 & 0 & 0 & 200 \end{bmatrix} \text{arcsec}^2 \text{ (to be converted).}$$

4. $\hat{\mu}_{y_4} = 500.214\,\text{m} \pm 5\,\text{mm}$.

5. $\hat{\sigma}_0^2 = (1.1321)^2$, $Q = \begin{bmatrix} 200 & -100 & 0 & 0 & 0 \\ -100 & 200 & 0 & 0 & 0 \\ 0 & 0 & 200 & 0 & 0 \\ 0 & 0 & 0 & 200 & 0 \\ 0 & 0 & 0 & 0 & 200 \end{bmatrix}$ arcsec^2 (to be converted).

Chapter 5

5.a; 5.b $r = 3$, $\hat{\sigma}_0^2 = (0.015)^2$; $r = 4$, $\hat{\sigma}_0^2 = (0.013)^2$.

6.a; 5.b $r = 3$, $\hat{\boldsymbol{P}}_1 = (589.979, 374.998)\,\text{m}$.

7. $\hat{a} = -0.00735466$, $\Omega = 7.57541$, $R = 0.162439$.

Chapter 6

3.a; 3.b $r = 4$, $\hat{\sigma}_0^2 = (0.08063)^2$; $r = 5$, $\hat{\sigma}_0^2 = (0.07305)^2$, $T = 0.104487$.

4. $r = 3$, $\hat{\sigma}_0^2 = (4.599140)^2$, $T = 33.07538$.

5. $\hat{a} = -0.00729396$, $\Omega = 7.57541$, $R = 0.0234899$.

6.a $\hat{\boldsymbol{\xi}}^T = \begin{bmatrix} 68.8534 & 66.9512 & 68.1542 & 66.0026 & 67.9917 & 68.5199 & 67.6955 \end{bmatrix}^T$m, $\hat{\sigma}_0^2 = (1.00036)^2$.

Chapter 7

1.a See answers to Problems 9.b and 9.c of Chapter 3.

1.b $\hat{\hat{\boldsymbol{\xi}}} = [1679.497, 1804.053, 2021.126, 1507.062, 1668.156, 1928.277]^T$ft, $\hat{\hat{\sigma}}_0^2 = (0.08197)^2 = 0.006719$.

2. Estimated height in feet: $\hat{H}_A = 1679.493$, $\hat{H}_B = 1804.072$, $\hat{H}_C = 2021.150$, $\hat{H}_E = 1507.068$, $\hat{H}_F = 1668.159$, $\hat{H}_G = 1858.255$.

Chapter 8

1.b; 1.c $\hat{\boldsymbol{\xi}} = [3.04324,\ 0.74568,\ 4.10586]^T$; $\hat{\sigma}_0^2 = (0.243289)^2 = 0.059190$.

2.b; 2.c $\hat{\boldsymbol{\xi}} = [19.700\,975^\circ,\ 6.6284,\ 2.8227,\ 2.6177,\ 3.6400]^T$; $\hat{\sigma}_0^2 = (0.263559)^2 = 0.069463$.

3.a; 3.b $\hat{\boldsymbol{\xi}} = [1.73586328,\ 0.098057768,\ -0.0072771964]^T$; $\hat{\sigma}_0^2 = (1.830478)^2 = 3.350650$.

Bibliography

Bjerhammar, A. (1973). *Theory of Errors and Generalized Matrix Inverses*. Elsevier Scientific Publishing Company, Amsterdam.

Björck, A. (1996). *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, Philadelphia.

Christensen, R. (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer Texts in Statistics. Springer, New York, fourth edition.

Grafarend, E. W. and Schaffrin, B. (1993). *Ausgleichungsrechnung in linearen Modellen*. Bibliographisches Institut, Mannheim.

Harville, D. A. (1986). Using ordinary least squares software to compute combined intra-interblock estimates of treatment contrasts. *The American Statistician*, 40(2):153–157.

Harville, D. A. (2000). *Matrix Algebra From a Statistician's Perspective*. Springer, corrected edition.

Heumann, C., Nittner, T., Rao, C., Scheid, S., and Toutenburg, H. (2013). *Linear Models: Least Squares and Alternatives*. Springer Series in Statistics. Springer New York.

Koch, K.-R. (1999). *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer-Verlag, Berlin, second edition. Translated from the third German edition (1997).

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons, Chichester.

Mikhail, E. and Ackermann, F. (1982). *Observations and least squares*. University Press of America.

Mikhail, E. M. and Gracie, G. (1981). *Analysis and Adjustment of Survey Measurements*. Van Nostrand Reinhold Company, New York, Toronto.

Neitzel, F. and Petrovic, S. (2008). Total Least-Squares (TLS) im Kontext der Ausgleichung nach kleinsten Quadraten am Beispiel der ausgleichenden Geraden (Total Least-Squares in the context of least-squares adjustment of a straight line). *Z. für Vermessungswesen*, 133:141–148.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572.

Pope, A. (1972). Some pitfalls to be avoided in the iterative adjustment of nonlinear problems. In *Proc. of the 38th Annual Meeting of the ASPRS*, pages 449–477, Falls Church, VA. American Society for Photogrammetry & Remote Sensing.

Rainsford, H. F. (1968). *Survey Adjustments and Least Squares*. Constable & Company Ltd., London.

Schaffrin, B. (1995). A generalized lagrange functions approach to include fiducial constraints. *Zeitschrift für Vermessungswessen*, 120(3):325–333.

Schaffrin, B. (1997a). On suboptimal geodetic network fusion. Presented at the IAG General Meeting, Rio de Janeiro, Brazil.

Schaffrin, B. (1997b). Reliability measures for correlated observations. *J. of Surveying Engineering*, 123(3):126–137.

Schaffrin, B. and Snow, K. (2010). Total Least-Squares regularization of Tykhonov type and an ancient racetrack in Corinth. *Linear Algebra and its Applications*, 432(8):2061–2076.

Smith, D., Heck, J., Gillins, D., and Snow, K. (2018). On least-squares adjustments within the Variance Component Model with Stochastic Constraints. *NOAA Technical Memorandum*, NOS NGS 74.

Snedecor, G. and Cochran, W. (1980). *Statistical Methods*. Iowa State University Press, Ames, Iowa.

Snow, K. (2002). Applications of parameter estimation and hypothesis testing to GPS network adjustments. Master's thesis, The Ohio State University.

Snow, K. and Schaffrin, B. (2003). Three-dimensional outlier detection for GPS networks and their densification via the BLIMPBE approach. *GPS Solutions*, 7(2):130–139.

Snow, K. and Schaffrin, B. (2016). Line fitting in Euclidean 3D space. *Studia Geophysica et Geodaetica*, 60(2):210–227.

Strang, G. (2006). *Linear Algebra and Its Applications*. Thomson, Brooks/Cole.

Strang, G. and Borre, K. (1997). *Linear Algebra, Geodesy, and GPS*. Wellesley-Cambridge Press, Wellesley, MA.

Tukey, J. W. (1987). *The Collected Works of John W. Tukey: Philosophy and Principles of Data Analysis 1965-1986*. Number 4. Taylor & Francis.

Wolf, P. R. (1983). *Elements of Photogrammetry: With Air Photo Interpretation and Remote Sensing*. McGraw-Hill.