Markus Vennebusch

Singular Value Decomposition and Cluster Analysis as Regression Diagnostics Tools in Geodetic VLBI

Markus Vennebusch • **Singular Value Decomposition and Cluster Analysis as Regression Diagnostics Tools in Geodetic VLBI**

universität**bonn** igg

Institut für
Geodäsie und Geoinformation

Schriftenreihe

3

Markus Vennebusch

Singular Value Decomposition and
Cluster Analysis as Regression
Diagnostics Tools in Geodetic VLBI

Vorwort

Seit mehr als zwanzig Jahren gehört die geodätische Radiointerferometrie, auch kurz *VLBI* genannt (von Very Long Baseline Interferometry) zu den tragenden Beobachtungsverfahren der geodätischen Forschung, da sie die direkte Verbindung zwischen dem himmelsfesten, quasi-inertialen Referenzsystem der Himmelskörper am Rand des uns bekannten Universums und dem erdfesten Referenzsystem der Beobachtungsstationen herstellt. Die "Zielmarken" sind hier quasi-stellare Objekte (Quasare) oder andere Radiostrahlung emittierende Galaxien, die für die Beobachter auf der Erde als punktförmig erscheinen und die mit Radioteleskopen beobachtet werden.

Während die Referenzsysteme die geodätische Säule der Geometrie repräsentieren, verbindet die VLBI diese Referenzsysteme über die hypothesenfreie Bestimmung der Erdrotation, die die zweite Säule der Geodäsie darstellt. Auch wenn die dritte Säule der Geodäsie, die Bestimmung des Schwerefeldes der Erde, nicht durch die *VLBI* abgedeckt wird, so stellt sich das Verfahren doch als fundamental für die Erdsystemforschung dar, da das Rotationsverhalten der Erde in all seinen Komponenten nur durch die VLBI bestimmt werden kann.

Die VLBI benötigt für die Durchführung der Messungen einen detaillierten Beobachtungsplan, in dem die Abfolge der Beobachtungen der einzelnen Radioteleskope festgelegt werden muss, wobei sowohl Beobachtungen aller Teleskope des Netzwerkes zu einem einzelnen Quasar als auch Beobachtungen von Subnetzen notwendig sind. Die Geometrie der Beobachtungen und damit der Beobachtungsplan legen die Stabilität der Kleinste-Quadrate-Lösung fest. Zur Charakterisierung der geometrischen Eigenschaften eines Ensembles von Beobachtungen wurden bisher in der Hauptsache Kovarianzinformationen herangezogen, die es jeweils nur erlaubte, die Gesamtheit aller Beobachtungen zu betrachten und zu interpretieren.

Die vorliegende Dissertation von Herrn Vennebusch stellt insofern eine bemerkenswerte Neuerung dar, als dass nun mit Hilfe der Cluster-Analyse die Bedeutung einzelner Gruppen von Beobachtungen differenziert interpretiert werden kann. Herr Vennebusch hat dazu mit der Singulärwertzerlegung eine geeignete Schnittstelle zwischen der Lösung eines größeren Gleichungssystems und der Cluster-Analyse gefunden. Damit hat er einen entscheidenden Schritt zur Identifikation von (Gruppen von) Hebelbeobachtungen ermöglicht, den es im nächsten Schritt in die Software zur Erstellung von Beobachtungsplänen einzubinden gilt.

Bonn, im April 2008

Zusammenfassung

Es ist bekannt, dass Hebelpunkt-Beobachtungen die Schätzung von Parametern stark beeinflussen. Bisher wurden Redundanzanteile von Beobachtungen verwendet, um *einzelne* Hebelpunkt-Beobachtungen von *einzelnen* redundanten (bzw. weniger wichtigen) Beobachtungen zu unterscheiden. In dieser Arbeit wird ein objektives Verfahren zur Aufdeckung von *Gruppen* von wichtigen und weniger wichtigen (und somit redundanten) Beobachtungen entwickelt. Außerdem wird bestimmt, welche Parameter hauptsächlich von diesen Beobachtungsgruppen beeinflusst werden.

Der hier vorgeschlagene Ansatz basiert auf geometrischen Aspekten der Ausgleichungsrechnung und verwendet die Singulärwertzerlegung der Designmatrix eines Ausgleichungsproblems und Cluster Analyse-Verfahren zur Regressionsanalyse.

Obwohl der hier vorgeschlagene Ansatz auf beliebige geodätische Ausgleichungsprobleme angewendet werden kann, werden in dieser Arbeit nur Anwendungen bezogen auf die geodätische Langbasis-Interferometrie (VLBI) gezeigt. Allgemein ist der hier vorgeschlagene Ansatz dazu geeignet, (Gruppen von) Beobachtungen aufzudecken, die die geschätzten Parameter signifikant beeinflussen oder nur vernachlässigbaren Einfluss haben (und somit auf diese Beobachtungen am ehesten verzichtet werden kann).

In dieser Arbeit wird zunächst der theoretische Hintergrund der geometrischen Aspekte der Ausgleichungsrechnung zusammengefasst. Dann wird die Singulärwertzerlegung der Designmatrix des zugehörigen Ausgleichungsproblems verwendet, um Kenngrößen für den Einfluss und die Ähnlichkeit von Beobachtungen zu bestimmen. Gruppen von Beobachtungen mit ähnlichem Informationsgehalt werden anschließend mit Hilfe von Cluster Analyse-Algorithmen gebildet. Nach einer kurzen Wiederholung der Grundlagen der geodätischen Langbasis-Interferometrie wird der vorgeschlagene Ansatz sowohl auf fiktive als auch auf reale Ein-Basislinien-Sessionen angewendet. Damit werden die Tauglichkeit und die Fähigkeiten des hier entwickelten Regressionsdiagnose-Werkzeuges unter Beweis gestellt.

Summary

It is well known that high-leverage observations significantly affect the estimation of parameters. So far, mainly redundancy numbers have been used for the detection of *single* high-leverage observations or of *single* redundant observations. In this thesis an objective method for the detection of *groups* of important and less important (and thus redundant) observations is developed. In addition, the parameters which are mainly affected by these groups of observations are identified.

The method proposed in this thesis is based on geometric aspects of adjustment theory and uses the singular value decomposition of the design matrix of an adjustment problem and cluster analysis methods for regression diagnostics.

Although the proposed method can be applied to any geodetic adjustment problem, in this thesis only applications to geodetic very long baseline interferometry (VLBI) are shown. In general, the method is well suited for the detection of (groups of) observations that significantly affect the estimated parameters or that are of negligible impact (and are thus candidates for observations that can be omitted).

In this thesis, at first the theoretical background of the geometrical aspects of geodetic adjustment theory is summarized. Then the singular value decomposition of the design matrix of an adjustment problem is used for the computation of measures of the impact and similarity of observations. Groups of observations with a similar information content are then identified by statistical cluster analysis algorithms. After a short review of geodetic very long baseline interferometry the proposed method is applied to artificial and real single-baseline sessions in order to show the capabilities of the regression diagnostics tool developed in this thesis.

Contents

0. Introduction

Since the 1970ies Very Long Baseline Interferometry (VLBI) has been used to determine station coordinates on earth as well as parameters of earth rotation with very high precision. With an accuracy of $2 - 3 \cdot 10^{-9}$ VLBI belongs to the most precise measurement techniques in geodesy.

VLBI observations consist of the differences of arrival times of signals of extragalactic radio sources received at two radio telescopes. In contrast to the Global Positioning System (GPS) in VLBI the analyst can define the type of observations by manually selecting the two observing sites and the radio sources both radio telescopes point at. Depending on the duration of an observation session and the size of the observing network, many thousands of observations and hundreds of unknown parameters accumulate (MA 1990). The adjustment of VLBI observations and the determination of the target parameters is a typical least-squares problem as it occurs in many scientific and engineering tasks.

It is well-known that the VLBI technique is very sensitive to variations in the choice of observations as well as to small variations in the observed time delay. Also, the choice of the functional model (i.e., of the unknown parameters) and the choice of constraints strongly affect the estimated parameters. This has also been recognized by the International VLBI Service for Astrometry and Geodesy (IVS) and has been summarized in the IVS-WG3 Report on Data Analysis (SCHUH, H. ET AL. 2006):

> Robustness and reliability of VLBI solutions are key elements of the quality of VLBI results. Therefore, improved analysis strategies together with observation scheduling will have to be developed which reduce the influence of single observations on the results.

In addition, many authors recognized the sensitivity of their results to small variations in both the network geometry and the observation geometry (see e.g. FISCHER 2006). In statistical terminology this problem can be summarized as weak quality of the design of a (VLBI-)experiment (FÖRSTNER 1987). In order to overcome this problem, methods need to be developed to identify observations with a similar content of information and to separate important (groups of) observations from less important (groups of) observations. The 'importance' of observations is closely related to the redundancy of observations. Since important observations (or observations with a low redundancy) significantly affect the estimated parameters, the detection of influencial observations (or influencial observation groups) is of great benefit for the improvement of both the precision and reliability of (VLBI) results. In addition, the omission of less important observations obviously bears economical advantages.

Investigations of the sensitivity of VLBI solution parameters (such as e.g. baseline components or earth orientation parameters) have been performed in the 1970ies and 1980ies by e.g. MA 1978, LUNDQVIST 1984 or BROUWER 1985. These authors analysed the partial derivatives of the VLBI observation equation with respect to the most common parameters or investigated the variations in the covariance matrix of the unknown parameters after including or omitting individual observations (DERMANIS and GRAFAREND 1981). Optimal observation geometries for the determination of polar motion have been derived by NOTHNAGEL 1991. Software for optimal observation schedules has been developed by e.g. STEUFMEHL 1994 and attempts for an improvement of the stochastic model of VLBI have been performed by TESMER 2004.

In this thesis a method is presented which serves as a *regression diagnostics tool*[1] by taking into account the geometry (of both the network and the observations) of a VLBI session by investigating the entire design matrix of the associated adjustment problem by using existing algebraic and statistical tools. The objective of this thesis is the development of an analysis tool for an automatic and objective separation of important and less important (groups of) observations and for the determination of the impact of these observations (or observation groups) on each parameter involved. In contrast to methods for the generation of observation

---

[1] In geodesy, the term 'regression' is commonly used for the determination of the parameters of a regression line or a regression polynomial. Here (as well as in geophysical literature or in statistical literature), the term 'regression' is used in a more general context, describing the procedure of parameter estimation in linear models (cf. BELSLEY et al. 1980).

schedules (as developed by e.g. STEUFMEHL 1994) this method can be used for the analysis of existing observation schedules and for the detection of important and thus influential observation groups. It is well-known that influential (or high-leverage) observations significantly affect the estimated parameters and thus should be controlled (or supported) by appropriate (independent) observations. On the other hand, redundant observations only have negligible influence on the estimates and are thus candidates for observations that can be omitted.

Since the regression diagnostics tools is directly applied to the design matrix of the associated adjustment problem, the computation of normal equations (and thus the magnification of numerical sensitivity) is avoided. Hence, the proposed method is quite insensitive to round-off errors and loss-of-digits problems.

Although the proposed method is not limited to the analysis of VLBI observations, in this thesis only applications for geodetic VLBI are shown.

In addition to the presentation of the theoretical background of the regression diagnostics tool, a user-friendly software package for the analysis of several kinds of adjustment problems and especially for the analysis of VLBI observation schedules has been implemented and tested for its practical use.

In order to solve the tasks described above, mainly two methods will be used: On the one hand an algebraic tool, called singular value decomposition, is used to provide geometrical insight into the system of linear equations associated with the adjustment problem to be solved. The geometrical aspects of adjustment theory (or the 'vector space approach') offer a different perspective of least-squares methods than the calculus approaches (as used in e.g. KOCH 1999 or NIEMEIER 2002). The geometrical approach (as described in e.g. MEISSL 1982 or TEUNISSEN 1985) additionally provides a 'geometrical insight' into geodetic adjustment problems.

On the other hand a statistical tool, called cluster analysis, is being used for the detection of groups of observations with a similar content of information. Although cluster analysis is usually applied to attributes of real physical objects, it can also be used for the generation of observation groups. Chapters 1 to 3 of this thesis deal with both the singular value decomposition and cluster analysis methods.

The general structure of this thesis is as follows:

- Chapter 1 repeats and summarizes the algebraic background necessary for the understanding of geometrical interpretations of systems of linear equations. Therefore, vector spaces and projections onto subspaces are used to derive the method of least-squares and to understand the use of the singular value decomposition for algebraic problems (see e.g. MEYER 2000 or LAY 2003).

- Chapter 2 describes the relations between the geometrical 'vector space approach' and the calculus approach of estimating parameters of linear models (also known as Gauss-Markov model). Again, emphasis is put on geometrical aspects such as angles between vectors or subspaces in order to derive regression diagnostics tools that can be interpreted geometrically.

- Chapter 3 provides the basics of cluster analysis methods, i.e., statistical methods for the detection of similarities of objects and thus for the detection of groups of observations/information with a similar impact. These methods will be of relevance for the practical investigations performed in chapter 5.

- Chapter 4 gives a short overview of the VLBI principle and describes the basic methods for the determination of the most common parameters usually estimated from VLBI observations.

- Chapter 5 shows applications of the methods developed in the previous chapters for the analysis of measurements with plane and spatial interferometers. In this chapter, examples for practical applications of the VLBI observation schedule analysis software developed by the author of this thesis are shown. The main intention of this chapter is to show the capabilities of the regression diagnostics tool and to compare its results with existing strategies of schedule generation.

- Chapter 6 summarizes the capabilities of the regression diagnostics tool developed in the first chapters and provides an outlook to further possible applications.

Each chapter of this thesis can be read separately. Thus, especially chapter 2 contains some aspects which have already been treated in chapter 1.

1. Fundamental Linear Algebra

1.1 Introduction

Since geodetic adjustment theory consists of both linear algebra and statistical methods, some fundamental basics about linear algebra have to be reviewed in order to solve the tasks described in the introduction. Linear algebra provides the theoretical background for understanding the nature of systems of linear equations and offers methods to solve over-determined and inconsistent systems of linear equations.

The analysis and the geometrical interpretation of such over-determined systems of linear equations is the main content of this chapter and will lead to the concepts of vector spaces and projections onto subspaces. Furthermore, the fundamental concept of least-squares solutions of over-determined systems of linear equations and their geometrical interpretation are derived. These methods are closely related to the singular value decomposition (SVD) of the coefficient matrix of the linear system. The singular value decomposition will be of main importance for the analysis of geodetic adjustment problems in general and for the analysis of VLBI observation schedules as described in the following chapters in particular.

Most of the issues described in this chapter are of fundamental nature, and can be found in a variety of literature. Here only the concepts will be described, more details can be found e.g. in LAY 2003, MEYER 2000 or STRANG 2003.

1.2 Systems of linear equations

In many scientific research areas large systems of (linear) equations have to be solved or analyzed in order to get a deeper understanding of the corresponding adjustment problem. Any kind of such a **system of linear equations** (or a **linear system**) as

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1u}x_u &= y_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2u}x_u &= y_2 \\
\cdots\cdots\cdots\cdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nu}x_u &= y_n
\end{aligned}
\tag{1.1}
$$

can be expressed in matrix notation as

$$
\mathbf{Ax} = \mathbf{y},
\tag{1.2}
$$

with $\mathbf{A}$ being a $n \times u$ coefficient matrix, $\mathbf{y}$ being a $n \times 1$ vector of known constants (also known as 'right-hand-side of the system') and $\mathbf{x}$ being a $u \times 1$ vector of unknown constants. Depending on the entries of $\mathbf{y}$ the system is either called homogenous (for $\mathbf{y} = \mathbf{0}$) or inhomogeneous (for $\mathbf{y} \neq \mathbf{0}$). Arranging the coefficient matrix $\mathbf{A}$ and the right-hand side of the system $\mathbf{y}$ in a common matrix yields the **augmented matrix** $[\mathbf{A} \mid \mathbf{y}]$ of the system.

A linear system (1.1) has either

- no solution,

- exactly one solution, or

- infinitely many solutions.

If a system has no solution it is called **inconsistent**; or **consistent** if it has at least one solution. In addition, there are three possible classes of linear systems (valid for linear systems of full rank):

- uniquely determined systems $(u = n)$,

- under-determined systems $(u > n)$ and

- over-determined systems $(u < n)$,

with each of them either being consistent or inconsistent. In addition, the rank of the linear system has to be taken into account (see below).

The determination of the solution set(s) of a linear system is aided by visualisation tools called **row picture** and **column picture**: The row picture is generated by visualising the rows of the augmented matrix (i.e., each equation) as n (hyper-)planes in $\mathbb{R}^u$. If these hyperplanes have one common point of intersection the linear system has only one (unique) solution. Parallel hyperplanes indicate no solution while coinciding hyperplanes or intersection lines indicate infinitely many solutions. Figure 1.1 (a) shows an example for a consistent 2×2 linear system.
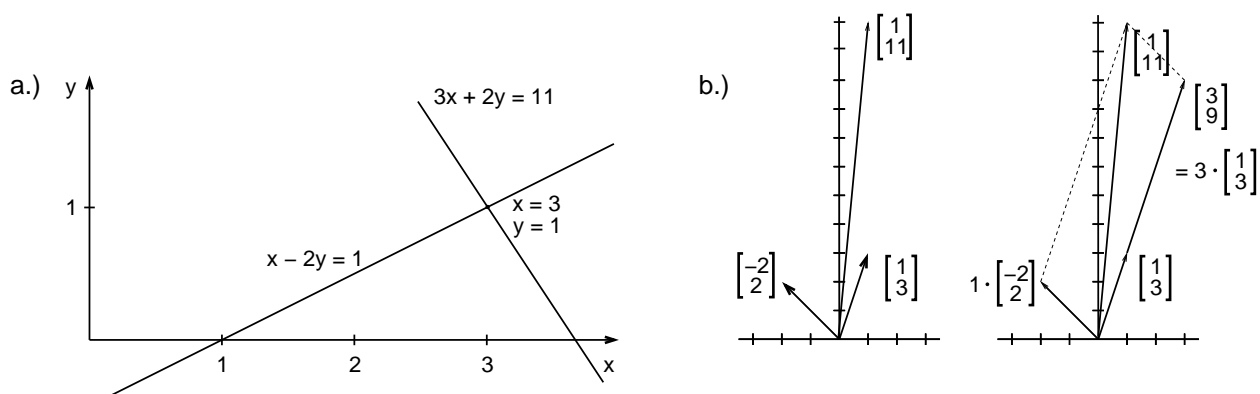
Figure 1.1: Row picture and column picture of a 2×2 linear system

For the column picture, equation (1.2) is interpreted columnwise, i.e., by visualising each column of the augmented matrix (STRANG 2003). The solution of the system (if any) is formed by determining the weights of that linear combination of the columns of $\mathbf{A}$ that yields the right hand side $\mathbf{y}$ of the system. Figure 1.1 (b) shows the column picture for the same 2×2 linear system as depicted in figure 1.1 (a).

The row picture can be used to explain the important term of the **condition** of a linear system: A linear system (and thus its solution) might be more or less sensitive to small perturbations caused by e.g. roundoff errors or loss-of-digits. Graphically this is displayed in figure 1.2 which shows the effect of small changes of the coefficient matrix or the right hand side of the system on the solution of a 2×2 linear system. Depending on the 'geometry' of the linear system (i.e., the intersection angle of the hyperplanes) the solution might change significantly. This sensitivity is inherent to the problem to be solved and cannot be overcome by any numerical 'tricks' (MEYER 2000). Thus a system is named **ill-conditioned** when even small changes produce relatively large changes in the solution. Otherwise, the system is said to be **well-conditioned**. The condition of a linear system is described by the **condition number** which -in the ideal case- is close to one and thus indicates (almost) orthogonal hyperplanes (for the computation of the condition number see section 1.6.4.1 on page 26).

As described in many fundamental books about Linear Algebra (see e.g. LAY 2003, STRANG 2003 or MEYER 2000) solutions of linear systems are easily determined by applying **Gaussian elimination** to the augmented matrix of the system. Using **elementary row operations** the system $[\mathbf{A} \mid \mathbf{y}]$ is transformed into a (**row equivalent**) triangular form $[\mathbf{E} \mid \mathbf{c}]$ by eliminating all elements below the **pivotal element** (=**forward step**). After **triangularisating** the coefficient matrix the solution is computed by **back-substitution** (= **backward step**) until each unknown has been determined.
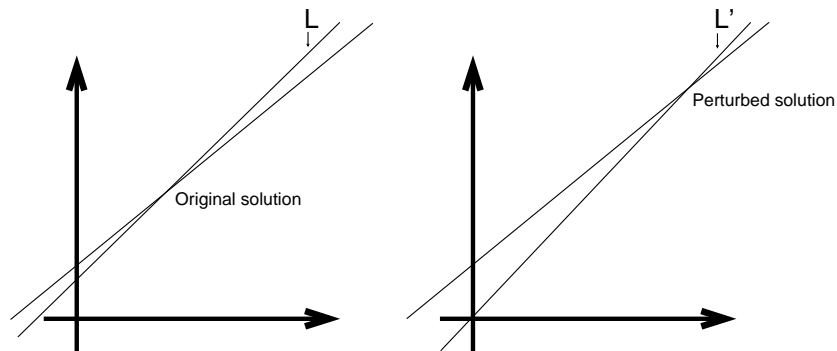
Figure 1.2: Ill-conditioned 2×2 linear system (MEYER 2000)

In the ideal case performing Gaussian Elimination on the coefficient matrix yields a complete triangular form, i.e., there never occurs a row of the form

$$\left(\begin{array}{ccccc} 0 & 0 & \cdots & 0 & \mid & \alpha \end{array} \right), \text{ with } \alpha \neq 0. \tag{1.3}$$

However, in many situations $\alpha \neq 0$, indicating an equation like

$$0x_1 + 0x_2 + \cdots + 0x_n = \alpha, \tag{1.4}$$

occurs and thus the back substitution process can not be completed. Equations as (1.4) with $\alpha \neq 0$ indicate an **inconsistent** system of linear equations which can not be solved exactly. Otherwise the system is said to be **consistent** and the system has (at least) one solution.

1.2.1 Solutions of linear systems

The general solution $\mathbf{x}$ of a linear system $\mathbf{Ax} = \mathbf{y}$ is composed by the sum of

1. the solution of the corresponding homogeneous system and

2. a particular solution of the non-homogeneous system.

Thus, at first, the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has to be solved:

(1.) The **trivial solution** (i.e., $x_1 = x_2 = \cdots = x_n = 0$) is always a solution of a homogeneous system. Thus, all solutions different from the trivial solution have to be determined by applying the Gaussian algorithm to the system $[\mathbf{A} \mid \mathbf{0}]$ yielding the system $[\mathbf{E} \mid \mathbf{0}]$ with $\mathbf{E}$ having the general form:

$$E = \begin{pmatrix} \circledast & * & * & * & * & * & * & * \\ 0 & 0 & \circledast & * & * & * & * & * \\ 0 & 0 & 0 & \circledast & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \circledast & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In many cases $\mathbf{E}$ (also known as **row echelon form**) is not of purely triangular form but rather of a 'stair-step' type of triangular form (MEYER 2000) caused by linear dependencies of some columns of the coefficient matrix $\mathbf{A}$. Although the entries of $\mathbf{E}$ are not unique the shape of $\mathbf{E}$ is unique. The first non-zero entries in each row (circled elements) denote pivot elements and thus indicate independent column vectors (=basic

columns). The respective variables are also known as **basic variables**. Non-basic columns can be expressed as linear combinations of basic columns and thus reveal **free variables** whose values have to be chosen. Whenever a system is consistent, the solution set can be described explicitly by solving the reduced system for the basic variables in terms of the free variables. Thus, in the case of a purely triangular matrix $\mathbf{E}$ no free variables exist. On the other hand, if at least one free variable exists there is an infinite number of solutions. Consequently, the trivial solution is the only solution if and only if there are no free variables.

In general, the basic variables can be expressed in terms of the free variables. All solutions of the homogeneous system can be described by successively setting one free variable to one and the remaining free variables to zero. For each case a **particular solution $\mathbf{h_i}$** is obtained. The general solution $\mathbf{x}$ of the homogeneous system $\mathbf{Ax} = \mathbf{0}$ is generated by all possible linear combinations of the particular solutions $\mathbf{h_i}$, i.e., by

$$\mathbf{x} = x_{f_1}\mathbf{h_1} + x_{f_2}\mathbf{h_2} + \cdots + x_{f_{n-r}}\mathbf{h_{n-r}} \tag{1.5}$$

with $x_{f_1}, x_{f_2}, \ldots, x_{f_{n-r}}$ denoting the free variables and the $n \times 1$ vectors $\mathbf{h_1}, \mathbf{h_2}, \ldots, \mathbf{h_{n-r}}$ representing particular solutions of the system. As the free variables x_{f_i} range over all possible values, the general solution generates all possible solutions. Thus, for each non-basic column of $\mathbf{E}$ (i.e., for each free variable) one particular solution $\mathbf{h_i}$ exists (see e.g. MEYER 2000).

(2.) In order to solve the (non-homogeneous) system $\mathbf{Ax} = \mathbf{y}$, equation (1.5) has to be extended by a particular solution $\mathbf{p}$ generated by setting the free variables to $x_{f_1} = x_{f_2} = \cdots = x_{f_n} = 0$.

The general solution of a non-homogeneous system is given by:

$$\mathbf{x} = \mathbf{p} + x_{f_1}\mathbf{h_1} + x_{f_2}\mathbf{h_2} + \cdots + x_{f_{n-r}}\mathbf{h_{n-r}}. \tag{1.6}$$

Thus, the general solution of the associated homogeneous system is a part of the general solution of the original non-homogeneous system.

1.2.2 Rank of a matrix

The most basic definition of the **rank r of a matrix $\mathbf{A}$** is given by the number of pivot elements of a matrix $\mathbf{A}$. Thus, if $\mathbf{A}$ is of dimension $n \times u$ the rank r can never exceed $min(n, u)$. Furthermore, r equals the number of basic columns in $\mathbf{A}$ and thus equals the number of non-zero rows in $\mathbf{E}$. Other rank definitions can be found in algebraic literature.

1.3 Vector spaces

Considering linear systems as linear combinations of the columns of the coefficient matrix $\mathbf{A}$ (with $\mathbf{x}$ being the weights of that particular linear combination which generates the right hand side $\mathbf{y}$ of the system via $\mathbf{Ax} = \mathbf{y}$) led to the column picture introduced above. A generalisation of the column picture from $\mathbb{R}^2$ or $\mathbb{R}^3$ to $\mathbb{R}^n$ leads to the theory of vector spaces which provides a very elegant way of investigating linear systems. A general vector space definition is given in table 1.1.

1.3.1 Subspaces and sums of subspaces

Subsets of a vector space $\mathcal{V}$ which fulfil the closure properties **(A1)** and **(M1)** of table 1.1 are said to be **subspaces** of $\mathcal{V}$. Thus, every vector through the origin as well as linear combinations of such vectors form a subspace. The zero vector is called the **trivial subspace**. In addition, an entire vector space is a subspace of its own.

Two subspaces might be 'added' to generate another subspace. Formally,

$$\mathcal{X} + \mathcal{Y} = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathcal{X} \ \text{ and } \ \mathbf{y} \in \mathcal{Y}\}, \tag{1.7}$$

with $\mathcal{X}$ and $\mathcal{Y}$ denoting subspaces of $\mathcal{V}$. Then the sum (also denoted as $\mathcal{X} \oplus \mathcal{Y}$) is again a subspace of $\mathcal{V}$.

Vector Space Definition

A set $\mathcal{V}$ is called **vector space over** $\mathcal{F}$ when the vector addition and scalar multiplication operations satisfy the following properties:

(A1) $\mathbf{x} + \mathbf{y} \in \mathcal{V}$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$. This is called the **closure property for vector addition**.

(A2) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ for every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$.

(A3) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for every $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.

(A4) There is an element $\mathbf{0} \in \mathcal{V}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{V}$.

(A5) For each $\mathbf{x} \in \mathcal{V}$, there is an element $(-\mathbf{x}) \in \mathcal{V}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.

(M1) $\alpha\mathbf{x} \in \mathcal{V}$ for all $\alpha \in \mathcal{F}$ and $\mathbf{x} \in \mathcal{V}$. This is the **closure property for scalar multiplication**.

(M2) $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$ for all $\alpha, \beta \in \mathcal{F}$ and every $\mathbf{x} \in \mathcal{V}$.

(M3) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ for every $\alpha \in \mathcal{F}$ and all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.

(M4) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ for all $\alpha, \beta \in \mathcal{F}$ and every $\mathbf{x} \in \mathcal{V}$.

(M5) $1\mathbf{x} = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{V}$.

$\mathcal{F}$ denotes a field of scalars. Since in the following investigations and analyses only real vector spaces $\mathbb{R}^n$ are of interest, $\mathcal{F}$ is the field $\mathbb{R}$ of real numbers.

Table 1.1: Vector space definition (MEYER 2000)

1.3.2 Spanning sets

All possible linear combinations of a set of vectors $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ from a vector space $\mathcal{V}$ are called $span(\mathcal{S})$, i.e.,

$$span(\mathcal{S}) = \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_r\mathbf{v}_r \mid \alpha_i \in \mathcal{F}\}. \tag{1.8}$$

Thus, the subspace $\mathcal{V} = span(\mathcal{S})$ generated by forming all linear combinations of vectors from $\mathcal{S}$ is called the space spanned by $span(\mathcal{S})$. Then $span(\mathcal{S})$ is called the **spanning set** for $\mathcal{V}$. Thus, $\mathcal{V}$ might be spanned by many different spanning sets. Furthermore, $span(\mathcal{S})$ might contain redundant vectors which do not contribute to the generation of $\mathcal{V}$.

1.3.3 Bases of vector spaces

1.3.3.1 Linear independence, bases and dimension

Any set of vectors is said to be **linearly independent** if only the trivial solution $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$ is a solution of the homogeneous equation

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}. \tag{1.9}$$

Any linear independent spanning set for a vector space $\mathcal{V}$ is called a **basis for** $\mathcal{V}$. As shown in e.g. Meyer 2000 a vector space might be generated by many different bases. Unlike spanning sets, bases do not contain redundant vectors. According to Meyer 2000 a linearly independent spanning set for a vector space $\mathcal{V}$ is called a **basis** for $\mathcal{V}$. If $\mathcal{V}$ denotes a subspace of $\mathbb{R}^m$ and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\} \subseteq \mathcal{V}$, then

- $\mathcal{B}$ is a basis for $\mathcal{V}$,

- $\mathcal{B}$ is a minimal spanning set for $\mathcal{V}$ and

- $\mathcal{B}$ is a maximal linearly independent subset of $\mathcal{V}$.

The number of vectors in any basis for $\mathcal{V}$ is called **dimension (dim) of a vector space** $\mathcal{V}$. It should not be confused with the number of components contained in the individual vectors of $\mathcal{V}$!

1.3.3.2 Coordinates

An important reason for specifying a basis $\mathcal{B}$ for a vector space $\mathcal{V}$ is to generate a 'coordinate system' for $\mathcal{V}$: If $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis for a vector space $\mathcal{V}$, then each vector $\mathbf{x}$ in $\mathcal{V}$ can be expressed uniquely by a set of scalars $c_1, \ldots, c_n$ such that

$$\mathbf{x} = c_1 \mathbf{b}_1 + \cdots + c_n \mathbf{b}_n. \tag{1.10}$$

The scalars (or weights) $c_1, \ldots, c_n$ are the **coordinates of x relative to the basis** $\mathcal{B}$, or the $\mathcal{B}$-**coordinates of x**. For the **standard basis**, i.e., for basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ with

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \tag{1.11}$$

the coordinates of a vector $\mathbf{x}$ are just the components of $\mathbf{x}$.

1.3.4 The four subspaces of a matrix

1.3.4.1 Column space and row space

As introduced in section 1.3 all possible linear combinations of certain vectors form a vector space. This means that applying $\mathbf{Ax}$ to any $n \times u$ Matrix $\mathbf{A}$ with an $u \times 1$-vector $\mathbf{x}$ generates a subspace of $\mathbb{R}^n$ (also known as **range** $R(\mathbf{A})$ **of A**). Since every matrix-vector product $\mathbf{Ax}$ is a linear combination of the columns of $\mathbf{A}$, $\mathbf{Ax}$ is the space spanned by the columns of $\mathbf{A}$. This space is called **column space of A** and formally reads

$$R(\mathbf{A}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^u\} \subseteq \mathbb{R}^n = \text{col } \mathbf{A}. \tag{1.12}$$

Likewise, the space spanned by the rows of $\mathbf{A}$ (i.e., $R(\mathbf{A}')$ or the space spanned by the columns of $\mathbf{A}'$) is called **row space of A**:

$$R(\mathbf{A}') = \{\mathbf{A}'\mathbf{y} \mid \mathbf{y} \in \mathbb{R}^n\} \subseteq \mathbb{R}^u = \text{row } \mathbf{A}. \tag{1.13}$$

1.3.4.2 Left and right nullspaces

In addition to the row and the column space of a matrix $\mathbf{A}$ a matrix possesses two other vector spaces: The set of all possible solutions of the homogeneous system $\mathbf{Ax} = \mathbf{0}$, i.e.,

$$N(\mathbf{A}) = \{\mathbf{x} \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^u \tag{1.14}$$

for any $n \times u$ matrix $\mathbf{A}$, forms the **right nullspace (or kernel) of A**. The set of vector spaces associated with any matrix $\mathbf{A}$ is completed by the **left-hand nullspace $\mathbf{N(A')}$ of A**, i.e., by

$$N(\mathbf{A'}) = \{\mathbf{y} \mid \mathbf{A'y} = \mathbf{0}\} \subseteq \mathbb{R}^m. \tag{1.15}$$

Figure 1.3 summarizes the four vector spaces of a matrix (STRANG 2003).
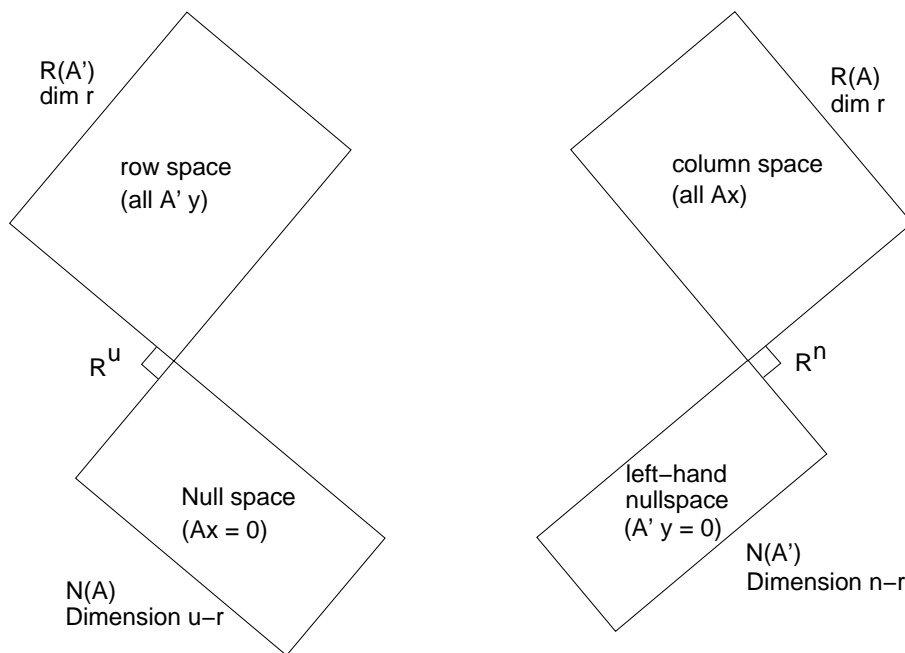
Figure 1.3: The four subspaces of an $n \times u$ matrix $\mathbf{A}$ (little squares indicate orthogonality of subspaces)

1.3.4.3 Dimensions of subspaces

In the general case of an $n \times u$ matrix $\mathbf{A}$ of rank r the columns of $\mathbf{A}$ do not form a basis if there are dependencies between some of the columns. However, the basic columns form an independent set and thus form a basis for $R(\mathbf{A})$. Thus, the dimension of the column space equals $dim \, R(\mathbf{A}) = r = rank(\mathbf{A})$. As shown by e.g. MEYER 2000 both the dimensions of the column space and the row space equal r. Consequently, the dimensions of the nullspace and the left nullspace equals $u - r$ and $n - r$, respectively. Table 1.2 summarizes the four vector spaces of a general $n \times u$ matrix of rank r.

Subspace:		Dimension:
Range or column space	$R(\mathbf{A}) = \{\mathbf{Ax}\} \subseteq \mathbb{R}^n$	$dim \, R(\mathbf{A}) = r$
Row space	$R(\mathbf{A'}) = \{\mathbf{A'y}\} \subseteq \mathbb{R}^u$	$dim \, R(\mathbf{A'}) = r$
Nullspace	$N(\mathbf{A}) = \{\mathbf{x} \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^u$	$dim \, N(\mathbf{A}) = u - r$
Left nullspace	$N(\mathbf{A'}) = \{\mathbf{y} \mid \mathbf{A'y} = \mathbf{0}\} \subseteq \mathbb{R}^n$	$dim \, N(\mathbf{A'}) = n - r$

Table 1.2: Summary of the four subspaces of an $n \times u$ matrix $\mathbf{A}$

1.4   Linear transformations

Any linear transformation **T** from one (finite-dimensional) vector space to another (finite-dimensional) vector space (as e.g. rotations, projections or reflections as well as the identity transformation and the zero transformation) can be expressed in matrix form. *In general every $n \times u$ matrix* **A** *acts as a (linear) mapping from $\mathbb{R}^u$ to $\mathbb{R}^n$* (an example is shown in figure 1.4). Therefore, a proper basis has to be chosen in either vector space. Then **A** is called the **coordinate matrix of the linear transformation** (MEYER 2000). One of the main aspects of Linear Algebra is to analyse special properties of such transformations (see e.g. LAY 2003 or STRANG 2003).
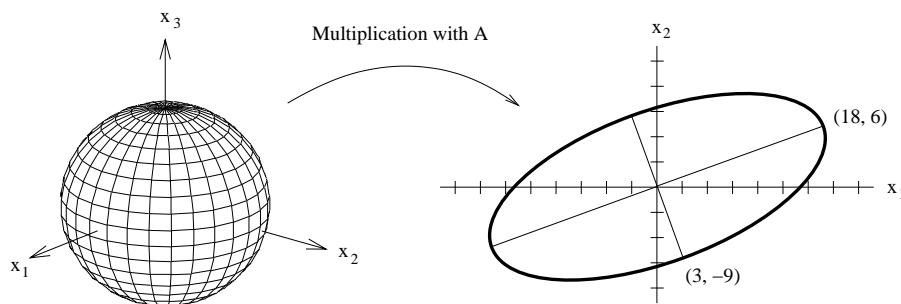
Figure 1.4: Transformation from $\mathbb{R}^3$ into $\mathbb{R}^2$ (maps a sphere onto an ellipse)

1.4.1   Change of basis

Due to the base dependency of the matrix representation of such transformations some properties might not be visible when using the initial (standard) basis. The solution of a problem (which might be initially described using a basis $\mathcal{B}$) is generally solved easier after changing to a new basis $\mathbb{C}$. Thus a **change of basis** might reveal special properties of a linear transformation so that the problem becomes more lucid (DERMANIS and RUMMEL 2000). The new basis might consist of orthogonal basis vectors and might yield a diagonal structure of the coefficient matrix, which is obviously easy to solve.

1.4.1.1   Matrix-vector product as a change of basis operation

With $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ being two bases of a vector space $\mathcal{V}$ the $n \times n$ **change-of-coordinate matrix** $P_{\mathcal{C} \leftarrow \mathcal{B}}$ which transforms a vector from $\mathcal{B}$ to $\mathcal{C}$ via

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C} \leftarrow \mathcal{B}} [\mathbf{x}]_{\mathcal{B}} \tag{1.16}$$

is computed by arraging the $\mathcal{C}$-coordinates of the vectors in the basis $\mathcal{B}$ as

$$P_{\mathcal{C} \leftarrow \mathcal{B}} = \left[\ [\mathbf{b}_1]_{\mathcal{C}} \quad [\mathbf{b}_2]_{\mathcal{C}} \quad \cdots \quad [\mathbf{b}_n]_{\mathcal{C}} \ \right]. \tag{1.17}$$

Formula (1.16) can be generalised to the dimension $n \times u$. In any case the coordinate vectors of the old basis have to be expressed in terms of the new basis to compute $P_{\mathcal{C} \leftarrow \mathcal{B}}$ via equation (1.17) (LAY 2003).

1.4.1.2 Change of basis for coordinate matrices

Due to a change of the underlying basis from basis $\mathcal{B}$ to $\mathcal{B}'$ the change of the coordinate matrix $\mathbf{A}$ of a linear transformation on $\mathcal{V}$ is computed by (MEYER 2000):

$$[\mathbf{A}]_{\mathcal{B}} = \mathbf{P}^{-1} [\mathbf{A}]_{\mathcal{B}'} \mathbf{P}, \quad \text{with} \quad \mathbf{P} = [\mathbf{I}]_{\mathcal{B}\mathcal{B}'}. \tag{1.18}$$

Equivalently,

$$[\mathbf{A}]_{\mathcal{B}'} = \mathbf{Q}^{-1} [\mathbf{A}]_{\mathcal{B}} \mathbf{Q}, \quad \text{with} \quad \mathbf{Q} = [\mathbf{I}]_{\mathcal{B}'\mathcal{B}} = \mathbf{P}^{-1}, \tag{1.19}$$

being the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$. A proof can be taken from MEYER 2000.

In general, applying left-multiplication of a coordinate matrix $\mathbf{A}$ with a change-of-basis matrix $\mathbf{P}$ is effectively a sequential application of matrix-vector multiplications and thus results in a change-of-basis operation for every *column* of $\mathbf{A}$. Therefore, left-multiplication with a change-of-basis matrix introduces a new basis to the *column space* of $\mathbf{A}$. On the other hand, right multiplication of a coordinate matrix with a change-of-basis matrix (which is different from the one mentioned above) results in a change-of-basis operation for the *row space* of $\mathbf{A}$.

1.4.2 Eigenvalues and Eigenvectors

The application of a linear transformation $\mathbf{T}$ on a vector $\mathbf{u}$ (via $\mathbf{u}' = \mathbf{A}\mathbf{u}$) usually results in a change of the direction of $\mathbf{u}$. On the other hand, there might be vectors which keep their direction (probably with a change of sign) after a linear transformation. A two-dimensional example is shown in figure 1.5. Vectors (or vector spaces) which do not change their direction after applying a linear transformation are called **invariant subspaces** and are important since they are used to simplify coordinate matrix representations of $\mathbf{T}$.
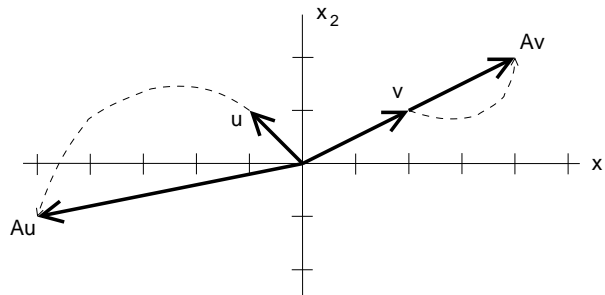
Figure 1.5: Effects of multiplication by $\mathbf{A}$

Invariant subspaces are identified by determining eigenvectors and eigenvalues of the coordinate matrix $\mathbf{A}$. According to LAY 2003 an **eigenvector** of an $n \times n$ matrix $\mathbf{A}$ is a non-zero vector $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called **eigenvalue** of $\mathbf{A}$ if there is a non-trival solution $\mathbf{x}$ of $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an *eigenvector corresponding to* λ. All possible linear combinations of the eigenvectors are called **eigenspace**.

Thus, $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ shows that under a transformation by $\mathbf{A}$ the eigenvectors experience only changes in magnitude or sign. The orientation of $\mathbf{A}\mathbf{x}$ in $\mathbb{R}^n$ is the same as that of $\mathbf{x}$. The eigenvalue λ indicates the amount of 'stretch' or 'shrink' to which the eigenvector $\mathbf{x}$ is subjected when transformed by $\mathbf{A}$.

Eigenvalues and eigenvectors can be used to factorize an $n \times n$-matrix $\mathbf{A}$ into

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \tag{1.20}$$

with $\mathbf{P}$ being a matrix containing n eigenvectors of $\mathbf{A}$ and $\mathbf{D}$ being a diagonal matrix containing n eigenvalues of $\mathbf{A}$ on its main diagonal. $\mathbf{A}$ is said to be **diagonalizable** if such a factorization exists, i.e., only if $\mathbf{A}$ has n linearly independent eigenvectors. In other words, $\mathbf{A}$ is diagonalizable if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. Such a basis is called **eigenvector basis**. In this case, the $\mathcal{B}$-matrix of the linear transformation T is diagonal. Diagonalising $\mathbf{A}$ is effectively finding a diagonal matrix representation of the linear transformation $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$.

Rearranging equation (1.20) to $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$ shows that $\mathbf{A}$ is diagonalized by applying the change-of-basis operators $\mathbf{P}$ (and $\mathbf{P}^{-1}$) to $\mathbf{A}$ and thus by changing to a new basis for $\mathbb{R}^n$. Equation (1.20) is also known as *spectral decomposition or eigenvalue decomposition (EVD)* and is of great importance for statistical applications and regression problems.

Multiple eigenvalues / non-diagonalizable matrices

Problems may occur when $\mathbf{A}$ does not possess n distinct eigenvalues. As shown in LAY 2003 or MEYER 2000 a matrix is only diagonalizable if and only if it possesses a complete set of eigenvectors and thus only if it possesses n distinct eigenvalues. Matrices that fail to possess complete sets of eigenvectors are called *deficient*.

In the case of several identical eigenvalues λ_i (called algebraic multiplicity of the eigenvalue λ_i) the number of associated eigenvectors (called geometrical multiplicity of λ_i) can be smaller than the algebraic multiplicity. Geometrically, this means that no unique basis vector for the eigenspace can be found.

1.5   Orthogonality and Least-squares

Orthogonality of vectors or vector spaces and projections onto vector spaces provide a very elegant and geometrically comprehensible way of deriving methods for solving over-determined linear systems in a least-squares sense without using the usual calculus approach (as described e.g. in KOCH 1999). Both the geometric 'vector space approach' and the calculus approach lead to the well-known normal equation approach. In addition, the vector space approach provides further methods and analysis tools to get deeper insight into the adjustment problem.

1.5.1   Inner products, norms and metric of a vector space

The **inner product**, **dot product** or **scalar product** of two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as (TREFETHEN 1997)

$$\mathbf{u}'\mathbf{v} = \mathbf{u} \cdot \mathbf{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n. \tag{1.21}$$

Furthermore it is used to compute the angle θ between two vectors in $\mathbb{R}^n$ via

$$\cos\theta = \frac{\mathbf{u}' \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}. \tag{1.22}$$

Thus, any two vectors in $\mathbb{R}^n$ are **orthogonal** if their inner product equals zero. Any vector space that is equipped with an inner product is called an **inner-product space**.

In formula (1.22) the norm operator $\|\cdot\|$ has been used. As described in e.g. VANICEK and KRAKIWSKY 1986 the (general) norm operator is used to measure distances $\rho(a, b)$ between any two elements a, b

of a vector space. For any vector space, the way of formulating this distance, or **metric**, can be chosen in many ways. The most common norm (or length) of a vector **u** -also known as Euclidean norm or 2-Norm- is defined as

$$\|\mathbf{u}\| = \sqrt{\mathbf{u}' \cdot \mathbf{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}. \tag{1.23}$$

The general properties and the different types of norms can be found e.g. in MEYER 2000. A vector space in which a metric has been defined is called a **metric space**.

1.5.1.1 Orthogonal complements

If a vector **u** is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then **u** is said to be orthogonal to W. The set of all vectors **u** that are orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^\perp$ (STRANG 2003). Thus, the nullspace $N(\mathbf{A})$ of a matrix **A** is an orthogonal complement of the row space of **A** while the nullspace of $\mathbf{A}'$ is an orthogonal complement of the column space of **A**, formally:

$$(\text{Row } \mathbf{A})^\perp = \text{Nul } \mathbf{A} \quad \text{and} \quad (\text{Col } \mathbf{A})^\perp = \text{Nul } \mathbf{A}'.$$

The orthogonality of the four subspaces of a matrix **A** is visualised (by little squares) in figure 1.3 on page 15.

1.5.1.2 Orthogonal projections

Orthogonal sets (e.g., orthogonal bases) are in particular helpful in simplifying calculations. This becomes obvious when considering the problem of projecting vectors onto certain orthogonal subspaces. An example is the decomposition of a vector **y** (in $\mathbb{R}^n$) into the sum of two vectors, i.e., $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ with $\hat{\mathbf{y}}$ being a multiple of a nonzero vector **u** and **z** being orthogonal to **u** (both in $\mathbb{R}^n$). As shown on the left of figure 1.6 (for the $\mathbb{R}^2$-case) the decomposition is given by **orthogonally projecting y** onto **u** and **z**, respectively. As derived in e.g. LAY 2003 $\hat{\mathbf{y}}$ and **z** are computed as follows:

$\hat{\mathbf{y}} = \dfrac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$ is the orthogonal projection of **y** onto **u** and

$\mathbf{z} = \mathbf{y} - \dfrac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$ is the component of **y** orthogonal to **u**. $\tag{1.24}$
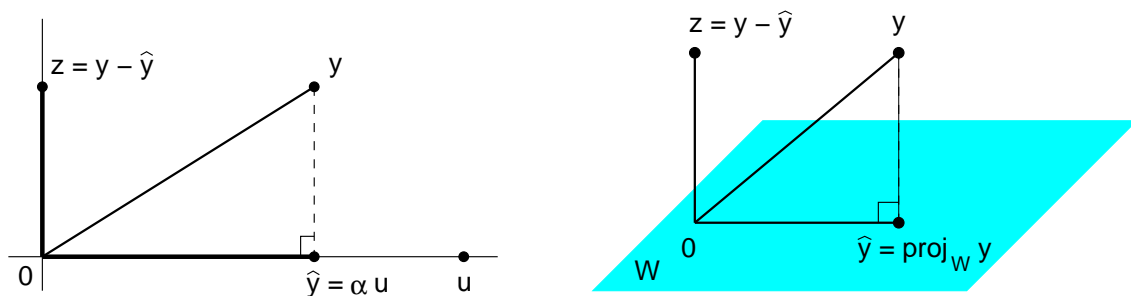
Figure 1.6: Decomposition into orthogonal complements, left: $\mathbb{R}^2$, right: $\mathbb{R}^n$

For $\mathbb{R}^n$ orthogonal projections can be generalised to the *Orthogonal Decomposition Theorem* (LAY 2003):

Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^\perp$. In fact, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p \tag{1.25}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

Each term in (1.25) is an orthogonal projection of $\mathbf{y}$ onto a one-dimensional subspace spanned by one of the $\mathbf{u}$-vectors in the basis for W. The orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto W is the sum of the projections of $\mathbf{y}$ onto one-dimensional subspaces which are orthogonal to each other (as shown on the right-hand side of figure 1.6 on the preceding page). This principle is of fundamental importance for the derivation of the least-squares algorithm in the next section.

1.5.1.3 Properties of orthogonal projectors

Some properties of (orthogonal) projection matrices will be relevant in the following chapters. Hence, a brief summary is given below (CASPARY and WICHMANN 1994, and MEYER 2000). For any projection matrix $\mathbf{P}$ holds:

- $\mathbf{P}$ is idempotent, i.e., $\mathbf{P}^2 = \mathbf{P}$,

- $\mathbf{Px} = \mathbf{x}$, i.e., further projection does not alter the previous projection result,

- since $\mathbf{P}$ is idempotent, the eigenvalues of $\mathbf{P}$ are either 0 or 1 and

- the trace and the rank of $\mathbf{P}$ are identical, i.e., $tr(\mathbf{P}) = rk(\mathbf{P})$.

For orthogonal projection matrices also

- $\mathbf{P}' = \mathbf{P} = \mathbf{P}^2$

applies.

1.5.2 Least-squares problems

For an over-determined system of linear equations $\mathbf{Ax} = \mathbf{y}$ the observation vector $\mathbf{y}$ almost certainly lies outside the column space $R(\mathbf{A})$ of the coefficient matrix $\mathbf{A}$, i.e., the system is almost certainly inconsistent. Nevertheless, the system can be solved (at least approximately) by finding a vector inside $R(\mathbf{A})$ *and* with minimal distance to $\mathbf{y}$. Following the so-called *Closest Point Theorem* (MEYER 2000) such a vector is given by the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto the column space $R(\mathbf{A})$ of the coefficient matrix $\mathbf{A}$. Thus, $\hat{\mathbf{y}}$ is an approximation of $\mathbf{y}$ which minimizes the distance $\| \mathbf{y} - \mathbf{Ax} \|$ (usually measured by using the 2-norm).

Consequently, the **general least-squares problem** is to find an **x** that leads to the smallest length of the vector $\mathbf{v} = \mathbf{y} - \mathbf{A}\mathbf{x}$ (also known as *residuals*). The vector $\hat{\mathbf{x}}$, which fulfils the (consistent) system

$$\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{y}}, \tag{1.26}$$

with $\hat{\mathbf{y}}$ being the projection of **y** onto $col(\mathbf{A})$, is called the **least-squares solution of Ax = y** (see figure 1.7).

Thus, the elements of $\hat{\mathbf{x}}$ denote the coordinates of $\hat{\mathbf{y}}$ with respect to the basis formed by the columns of the coefficient matrix **A**. As long as there is no rank deficiency, i.e., as long as there are no free variables $\hat{\mathbf{x}}$ is a unique vector[1].
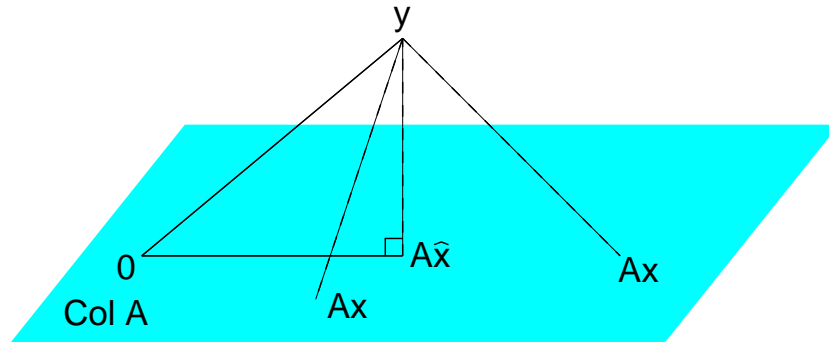
Figure 1.7: Least-squares principle

1.5.2.1 Least-squares solutions based on normal equations

As shown by e.g. STRANG 2003 the vector space based least-squares approach also leads to the well-known **normal equations**: Since $\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$ is orthogonal to the column space of **A**, the following equation holds:

$$\mathbf{A}'(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0}$$

$$\mathbf{A}'\mathbf{y} - \mathbf{A}'\mathbf{A}\hat{\mathbf{x}} = \mathbf{0}$$

$$\mathbf{A}'\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}'\mathbf{y} \tag{1.27}$$

Formula (1.27) yields a consistent (!) but probably rank-deficient system of linear equations for a compact least-squares solution of the original linear system $\mathbf{A}\mathbf{x} = \mathbf{y}$. As many authors (GRAMLICH and WERNER 2000, LAWSON and HANSON 1995, or KALMAN 1996) show, especially for ill-conditioned systems the solution of the normal equations becomes very sensitive to round-off errors and loss-of-digits since any errors in the entries of **A** are squared in the entries of $\mathbf{A}'\mathbf{A}$. Thus, the computation of the normal equations $\mathbf{A}'\mathbf{A}$ should be avoided. Alternative approaches for the solution of over-determined systems of linear equations in a least-squares sense are based on numerically more stable algorithms such as e.g. the QR-decomposition or the singular value decomposition of the coefficient matrix **A**. These methods are based on (vector length preserving) orthogonal transformations such as Householder transformations or Givens rotations (see e.g. GRAMLICH and WERNER 2000).

Accounting for different accuracies of the elements on the right-hand side of $\mathbf{A}\mathbf{x} = \mathbf{y}$ leads to the weighted least-squares principle. Applications of the least-squares principle in *adjustment theory* or *linear regression* will be further discussed in chapter 2.

---

[1] Here, correlations between parameters can be recognized by inspecting the angles between the column vectors of **A**: If there exists an (almost) linear dependency of the column vectors, no separation between the individual coordinate axes is possible and thus the coordinates of $\hat{\mathbf{y}}$ can not be well separated.

1.6   Singular Value Decomposition

One of the main aspects of linear algebra is the factorisation of linear systems, i.e., the decomposition of matrices into matrices with special properties. For example, the results of the Gauss algorithm can also be obtained by LU decomposition of the coefficient matrix or a linear system can be solved by performing a QR decomposition of its coefficient matrix (see e.g. STRANG 2003, GRAMLICH and WERNER 2000).

According to LAY 2003 singular value decomposition is 'one of the most useful matrix decompositions in applied linear algebra'. As described by STEWART 1993 the singular value decomposition has been developed by E. Beltrami and C. Jordan in 1873. Due to its outstanding relevance in linear algebra a variety of derivations can be found in the literature (e.g. BLANK, S.J. et al. 1989 or STEWART 1993). A complete definition of the singular value decomposition reads (MEYER 2000):

Singular Value Decomposition

For each $\mathbf{A} \in \mathbb{R}^u$ of rank r, there are orthogonal matrices $\mathbf{U}_{n \times n}$, $\mathbf{V}_{u \times u}$ and a diagonal matrix $\mathbf{S}_{r \times r} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ such that

$$\mathbf{A} = \mathbf{U} \left(\begin{array}{cc} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right)_{n \times u} \mathbf{V}' \quad \text{with } \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0. \tag{1.28}$$

The σ_i's are called the nonzero **singular values** of $\mathbf{A}$. When $r < p = \min\{n, u\}$, $\mathbf{A}$ is said to have $p - r$ additional zero singular values. The factorisation in (1.28) is called a **singular value decomposition** of $\mathbf{A}$, and the columns in $\mathbf{U}$ and $\mathbf{V}$ are called left-hand and right-hand **singular vectors** for $\mathbf{A}$, respectively.

1.6.1   Geometrical derivation of the Singular Value Decomposition

Contrary to more mathematical derivations (as given above) here a geometrical approach is used, as presented by TREFETHEN 1997. At first, the so-called reduced Singular Value Decomposition is derived.

1.6.1.1   Reduced Singular Value Decomposition

As described in chapter 1.4 every $n \times u$ matrix $\mathbf{A}$ acts as a linear mapping from $\mathbb{R}^u$ to $\mathbb{R}^n$. Thus, every matrix $\mathbf{A}$ maps a unit sphere S in $\mathbb{R}^u$ into a hyperellipse $\mathbf{A}S$ in $\mathbb{R}^n$. An example can be found in figure 1.4 on page 16. The hyperellipse in $\mathbb{R}^n$ can be obtained by stretching a unit sphere in $\mathbb{R}^n$ by some factors $\sigma_1, \ldots, \sigma_n$ (some of which might be zero) in orthogonal directions expressed by unit vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^n$. The vectors $\sigma_i \mathbf{u}_i$ are called principal semiaxes of the hyperellipse. As indicated on the right hand side of figure 1.8, these factors are the *singular values* $\sigma_1, \ldots, \sigma_n$ of $\mathbf{A}$ and indicate the lengths of the u semiaxes of the hyperellipse $\mathbf{A}S$. Usually these values are sorted in decreasing order, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.

The u unit vectors $\mathbf{u}_i$ of the principal semiaxes of $\mathbf{A}S$ are defined to be the *left singular vectors* of $\mathbf{A}$, numbered to correspond with the singular values.

Furthermore, u *right singular vectors* $\mathbf{v}_i$ of unit length are defined and correspond to the preimages of the principal semiaxes of $\mathbf{A}S$. These vectors are displayed on the left hand side of figure 1.8. Mathematically, the action of a matrix $\mathbf{A}$ on the right singular vectors $\mathbf{v}_i$ is

$$\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j \quad \text{for} \quad 1 \leq j \leq u, \tag{1.29}$$

which can also be expressed in matrix notation as

$$\mathbf{A}\mathbf{V} = \hat{\mathbf{U}}\hat{\mathbf{S}}. \tag{1.30}$$

Here, $\hat{\mathbf{S}}$ is a $u \times u$ diagonal matrix with the positive real singular values σ_j on its main diagonal, $\hat{\mathbf{U}}$ is an $n \times u$ matrix with orthonormal columns and $\mathbf{V}$ is an $u \times u$ matrix with orthonormal columns. Since $\mathbf{V}$ is orthonormal, $\mathbf{V}^{-1} = \mathbf{V}'$, and thus equation (1.30) can be written as

$$\mathbf{A} = \hat{\mathbf{U}}\hat{\mathbf{S}}\mathbf{V}'. \tag{1.31}$$

Factorisation (1.31) is also known as *reduced singular value decomposition* of $\mathbf{A}$ and is the basis for the more standard 'full' singular value decomposition of $\mathbf{A}$ described below.
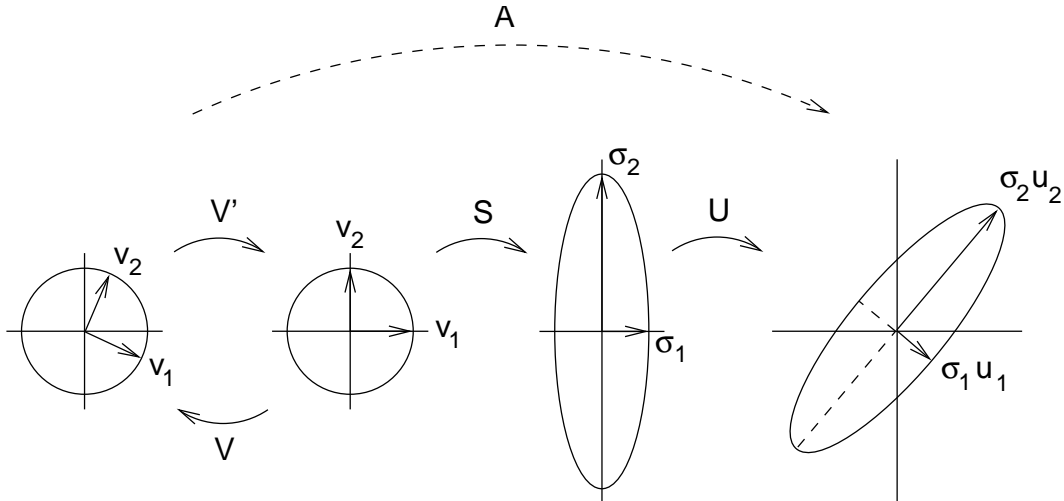
Figure 1.8: Linear mapping by $\mathbf{A}$ and singular value decomposition of $\mathbf{A}$ (STRANG 2003)

1.6.1.2   Full Singular Value Decomposition

For an overdetermined system (i.e., $n > u$) the columns of $\hat{\mathbf{U}}$ are just u orthonormal vectors in $\mathbb{R}^n$ which do not form a complete basis for $\mathbb{R}^n$. Thus, $n - u$ additional vectors have to be found to generate a complete basis and to extend $\hat{\mathbf{U}}$ to an orthonormal $n \times n$ matrix $\mathbf{U}$. The additional vectors can be constructed e.g. by using the Gram-Schmidt approach to make the columns of $\mathbf{U}$ form a complete (orthogonal) basis for $\mathbb{R}^n$ (STRANG 2003, or MEISSL 1982).

In addition, $\hat{\mathbf{S}}$ has to be modified in such a way that the $n - u$ columns of $\mathbf{U}$ are multiplied by zero so that the product (1.30) remains unchanged. The new $n \times u$ matrix $\mathbf{S}$ consists of $\hat{\mathbf{S}}$ extended by $n - u$ rows of zeros.

Since $\mathbf{V}$ remains unchanged, the *full Singular Value Decomposition* of $\mathbf{A}$ now reads:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}', \tag{1.32}$$

with $\mathbf{U}$ being an orthonormal $n \times n$ matrix containing the left singular vectors of $\mathbf{A}$, $\mathbf{S}$ being an $n \times u$ diagonal matrix with the singular values of $\mathbf{A}$ on its main diagonal and $\mathbf{V}$ an $u \times u$ orthonormal matrix containing the right singular vectors of $\mathbf{A}$. Graphically this factorisation can be visualised as shown in figure 1.9.

Singular Value Decomposition is neither limited to matrices with full rank nor to matrices containing more rows than columns (i.e., over-determined linear systems). Instead, any arbitrary $n \times u$ matrix of rank r can be factorised using formula (1.31) or (1.32) with $\mathbf{S}$ containing r nonnegative diagonal entries σ_i.

The singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ correspond to the eigenvectors of $\mathbf{A}\mathbf{A}'$ and $\mathbf{A}'\mathbf{A}$, respectively. The eigenvalues of $\mathbf{A}\mathbf{A}'$ and of $\mathbf{A}'\mathbf{A}$ are the same and are the squared singular values of $\mathbf{A}$. The eigenvalue decomposition (or diagonalisation) of $\mathbf{A}\mathbf{A}'$ and $\mathbf{A}'\mathbf{A}$ is always possible since $\mathbf{A}\mathbf{A}'$ and $\mathbf{A}'\mathbf{A}$ are symmetric matrices. For the same reason, the singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ form a complete, orthogonal basis (see e.g. STRANG 2003).
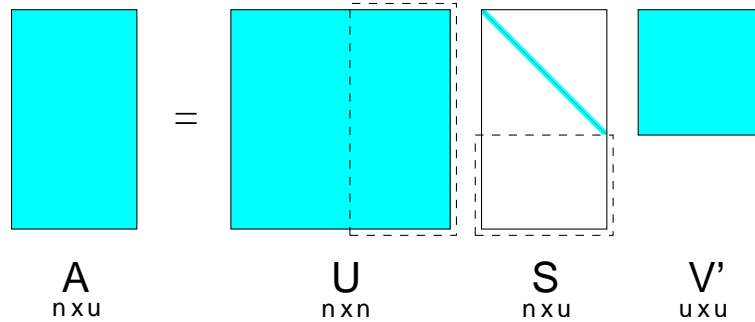
Figure 1.9: Graphical visualisation of the Singular Value Decomposition (SVD) of an $n \times u$ matrix $\mathbf{A}$ (for $n > u$, dashed lines indicate the differences between reduced and full SVD)

1.6.1.3   Geometrical analogies of the singular value decomposition

As shown in figure 1.8 on the preceding page the components of the factorisation of $\mathbf{A}$ into $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ also decompose the mapping represented by $\mathbf{A}$ (Strang 2003 and Trefethen 1997):

- $\mathbf{V}$ does not change the form of the unit sphere but introduces a new basis for $\mathbb{R}^u$ (also known as domain space of the mapping),

- $\mathbf{S}$ stretches the unit sphere into a hyperellipse and finally

- $\mathbf{U}$ rotates or reflects the hyperellipse without changing its shape (within the so-called range space of the mapping).

$\mathbf{A}$ and $\mathbf{S}$ represent the same mapping with respect to different bases: $\mathbf{A}$ describes the mapping with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^u$, $\mathbf{S}$ with respect to the bases formed by the left and right singular vectors. Thus, singular values reveal some information about the geometry of linear transformations since they show how much distortion can occur under a transformation by a matrix $\mathbf{A}$ (Meyer 2000). On the other hand, the singular value decomposition shows that any rectangular matrix can be diagonalised if appropriate bases for the domain and range space are chosen.

1.6.2   Canonical form and least-squares solutions

1.6.2.1   New Bases for the four fundamental subspaces of a matrix

For an arbitrary matrix $\mathbf{A}$ the full singular value decomposition determines new bases for the four fundamental subspaces. For the special case of an over-determined linear system (with $n > u$ and rank $r < u$)

- the first r left singular vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$ form an orthonormal basis for the column space of $\mathbf{A}$ (Col $\mathbf{A}$) and

- the remaining $n - r$ left singular vectors $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n$ form an orthonormal basis for the nullspace of A' (Nul $\mathbf{A}' = (\text{Col } \mathbf{A})^\perp$).

- An orthonormal basis for the row space of $\mathbf{A}$ is given by the first r right singular vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ and

- an orthonormal basis for the nullspace of $\mathbf{A}$ (Nul $\mathbf{A}$) is formed by the remaining (if any) $u - r$ right singular vectors $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_u$.

As shown in e.g. Meyer 2000 or Lay 2003 the singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are not unique (while the singular values σ_i are unique). The relations between the four bases are shown in figure 1.10 (Lay 2003).
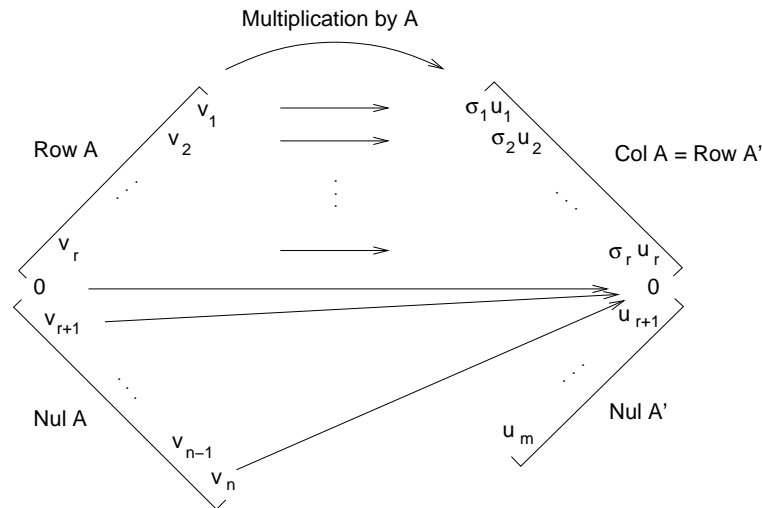
Figure 1.10: New bases for the four fundamental subspaces of a matrix **A** generated by singular value decomposition of **A** (Singular values are unique, singular vectors are not unique!).

1.6.2.2   Canonical form

The computation of new bases for the fundamental subspaces of **A** actually transforms **A** into its diagonal form **S**. This corresponds to changing the associated linear system **Ax** = **y** from the standard basis to new orthonormal bases using change-of-basis operations (see section 1.4.1). The linear system is said to be transformed into its 'canonical form' which significantly simplifies the corresponding least-squares problem (STRANG and BORRE 1997):

The change-of-basis is performed by expanding $\mathbf{y} \in \mathbb{R}^n$ in the basis of left singular vectors of **A** (columns of **U**) and by expanding $\mathbf{x} \in \mathbb{R}^u$ in the basis of right singular vectors of **A** (columns of **V**). The coordinate vectors for these expansions are

$$\bar{\mathbf{y}} = \mathbf{U}' \, \mathbf{y}, \quad \text{and} \quad \bar{\mathbf{x}} = \mathbf{V}' \, \mathbf{x}. \tag{1.33}$$

Using $\mathbf{A} = \mathbf{USV}'$, the relation **Ax** = **y** can be expressed in terms of $\bar{\mathbf{y}}$ and $\bar{\mathbf{x}}$:

$$\mathbf{y} = \mathbf{Ax} \iff \mathbf{U}'\mathbf{y} = \mathbf{U}'\mathbf{Ax} = \mathbf{U}'\mathbf{USV}'\mathbf{x} \iff \bar{\mathbf{y}} = \mathbf{S}\bar{\mathbf{x}}. \tag{1.34}$$

Applications and interpretations of the canonical form of **Ax** = **y** can be found in STRANG and BORRE 1997 and will be treated in more detail within the context of techniques for parameter estimation in linear models in chapter 2.

1.6.2.3   Pseudoinverse

Arranging the left singular vectors and right singular vectors of an $n \times u$ matrix **A** (with $n > u$ and rank $\mathbf{A} = r$) as

$$\mathbf{U} = [\mathbf{U}_r \quad \mathbf{U}_{n-r}], \quad \text{with} \quad \mathbf{U}_r = [\mathbf{u}_1 \dots \mathbf{u}_r] \text{ and}$$

$$\mathbf{V} = [\mathbf{V}_r \quad \mathbf{V}_{u-r}], \quad \text{with} \quad \mathbf{V}_r = [\mathbf{v}_1 \dots \mathbf{v}_r]$$

the *pseudoinverse (or Moore-Penrose inverse)* of **A** can be computed by

$$\mathbf{A}^+ = \mathbf{V}_r \mathbf{S}_r^{-1} \mathbf{U}_r' = \sum_{i=1}^{r} \frac{1}{\sigma_i} \cdot \mathbf{v}_i \cdot \mathbf{u}_i', \tag{1.35}$$

which in the case of a square matrix with full rank (i.e., $n = u = r$) equals the common matrix inverse $\mathbf{A}^{-1}$ (STRANG 2003). Thus, the pseudoinverse of $\mathbf{A}$ can be computed after singular value decomposition of $\mathbf{A}$ by simply inverting its singular values. Geometrically $\mathbf{A}^{+}$ depicts the inverse mapping of $\mathbf{A}$, i.e., $\mathbf{A}^{+}$ maps from $\mathbb{R}^{n}$ to $\mathbb{R}^{u}$. Details about different kind of matrix inverses can be found in e.g. CASPARY and WICHMANN 1994.

1.6.2.4   Least-squares solution by singular value decomposition

Using the pseudoinverse, the overdetermined linear system $\mathbf{Ax} = \mathbf{y}$ can be solved in a least-squares sense by

$$\hat{\mathbf{x}} = \mathbf{A}^{+}\mathbf{y} = \mathbf{V}_r\mathbf{S}_r^{-1}\mathbf{U}_r'\mathbf{y}. \tag{1.36}$$

Left-multiplication by $\mathbf{A}$ yields

$$\begin{aligned}
\mathbf{A}\hat{\mathbf{x}} &= (\mathbf{U}_r\mathbf{S}_r\mathbf{V}_r')(\mathbf{V}_r\mathbf{S}_r^{-1}\mathbf{U}_r'\mathbf{y}) \\
&= \mathbf{U}_r\mathbf{S}_r\mathbf{S}_r^{-1}\mathbf{U}_r'\mathbf{y} \quad \text{(since } \mathbf{V}_r \text{ is orthonormal and so } \mathbf{V}_r'\mathbf{V}_r = \mathbf{I}_r) \\
&= \mathbf{U}_r\mathbf{U}_r'\mathbf{y}.
\end{aligned}$$

Here, $\mathbf{U}_r\mathbf{U}_r'\mathbf{y}$ is the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto the column space of $\mathbf{A}$. Thus $\hat{\mathbf{x}}$ is a least-squares solution of $\mathbf{Ax} = \mathbf{y}$ (LAY 2003). In general, using the pseudoinverse for the solution of a linear system yields a solution of minimal norm (MEYER 2000).

1.6.3   Computational aspects

For the computation of the singular value decomposition of a matrix $\mathbf{A}$ as in equations (1.31) or (1.32) sophisticated and highly-optimised algorithms and implementations exist (GOLUB 1965, or GOLUB and REINSCH 1970). These algorithms are variants of algorithms used for the computation of eigenvalues and are given in e.g. GOLUB and KAHAN 1965, or PRESS et al. 1986. Other computation methods are described in TREFETHEN 1997. Implementations of fast and efficient algorithms with minimum memory requirements can be found in numerical libraries such as LAPACK or the GNU Scientific Library (GSL).

1.6.4   Applications of the Singular Value Decomposition

Singular value decomposition is used in a variety of sciences such as e.g. statistics, image processing or data compression. The application of singular value decomposition within parameter estimation techniques will be treated in more detail in chapter 2. Below a few general applications of the singular value decomposition are given.

1.6.4.1   Condition number of a linear system

Based on the singular value decomposition of a matrix $\mathbf{A}$ a new definition of the (2-norm) condition of a matrix (or for the associated linear system) can be given: The degree of distortion of the unit sphere under a transformation by $\mathbf{A}$ is measured by $\kappa = \sigma_1/\sigma_u$, i.e., the ratio of the largest singular value to the smallest singular value (MEYER 2000). A matrix is singular if its condition number is infinite (i.e., if there exists at least one zero singular value), and it is ill-conditioned if its condition number is very large (indicated by at least one very small singular value) (cf. section 1.2).

1.6.4.2   Rank determination

Singular Value Decomposition also serves as a robust tool for rank determination since the number of non-zero singular values equals the rank of a matrix $\mathbf{A}$ (e.g., LAWSON and HANSON 1995, or LAY 2003). Roundoff errors often lead to wrong rank determination results so that in practice very small singular value are assumed to be zero and the remaining non-zero singular values are used for the determination of the *effective rank* of a matrix $\mathbf{A}$ (GRAMLICH and WERNER 2000, or TREFETHEN 1997).

1.6.4.3   Lower rank approximations

Based on the singular value decomposition of a matrix $\mathbf{A}$ this matrix can be expressed as an outer product expansion:

$$\mathbf{A} = \sum_{i=1}^{u} \sigma_i \mathbf{u}_i \mathbf{v}_i'. \tag{1.37}$$

Equation (1.37) represents $\mathbf{A}$ as a sum of rank-one matrices (as does formula (1.35) for the pseudo-inverse). As shown by e.g. TREFETHEN 1997 or KALMAN 1996 equation (1.37) can be used to approximate any kind of (data) matrix by a sum of less than u 'slices', i.e., by a certain number of rank-one matrices computed by $\sigma_i \mathbf{u}_i \mathbf{v}_i'$. Applications of lower-rank approximations in statistics and for data compression in image processing can be found in e.g. LAY 2003.

The close relationship between singular value decomposition, principal component analysis (PCA) and the reduction of the dimension of multivariate data can be found in e.g. LAY 2003, or JACKSON 2003.

2. Parameter Estimation in Linear Models

2.1 Introduction

Parameter estimation in linear models (or adjustment theory or inverse theory) plays a central role in many scientific areas in order to condense or summarize data by fitting it to a mathematical function that depends on adjustable parameters which describe physical phenomena. Since VLBI data analysis is also based on these methods, parameter estimation techniques are of main importance for the development of regression diagnostics tools for improving the design of an experiment (DEHLERT 2000).

In the following chapter the relation (and equivalence) of commonly used calculus approaches (i.e., best linear unbiased estimators (BLUEs), etc.) and vector space based geometrical approaches are given. As will be shown, the non-geometric approaches bear some disadvantages. This has already been recognized by DERMANIS 1977, who mentions:

> Usually adjustment algorithms are derived from variational principles, as solutions to the problem of minimizing a quadratic form. Such an approach solves the problem but has little to offer to the understanding of its mathematical context and its relation to other techniques.

Thus, many authors only use the geometric or vector space approach to develop parameter estimation techniques. In the following chapter, geometrical concepts will be used to supply the calculus approach. These concepts will be used to provide geometrical interpretations of e.g. adjusted observations or residuals and the elements of projection operators will be used for the visualisation of redundancy numbers or 'impact factors'. The latter will be used in chapter 3 to develop methods for finding groups of observations and to separate important (groups of) observations from less important (groups of) observations.

2.2 Modeling of data

Modeling of data is used to describe measurement results (observations) by a convenient class of functions, such as appropriate linear combinations of polynomials or other so-called basis functions. Based on experiences and assumptions the mathematical formulation of the relation between observations and unknowns has often to be guessed and its correctness has to be verified by real observations. Adjustment theory tries to fit the observations to those functions and determines the coefficients of the (assumed) model (CASPARY and WICHMANN 1994).

A model is an image of the reality, expressed in mathematical terms, in a way, which involves a certain degree of abstraction and simplification. In general, a model consists of (DERMANIS and RUMMEL 2000)

- a set of observable objects (observations),

- a set of objects to be determined (unknowns) and

- a mathematical relation f, forming a connection between unknowns and observations.

On the one hand a model should be an appropriate (linear or linearised) description of the behaviour of a system, while, on the other hand, for practical reasons it should not be too complex. The degree of complexity of a model also depends on the particular purpose of the measurement[1].

---

[1]When using original observations the mathematical model has to be formulated in a very general way. The more reductions are applied to the original observations, the less general the mathematical model needs to be formulated. The final form of the mathematical model also depends on the purpose of the experiment.

In order to assess the agreement between the data and the model a so-called 'merit function' has to be chosen. Conventionally small values of the merit function indicate close agreement between the data and the model. The parameters of the model are then determined while minimizing the merit function, yielding *best-fit parameters* (PRESS et al. 1986).

A general parameter estimation procedure consists of the following steps:

- Estimation of parameters,

- Determination of formal errors of estimated parameters and

- Statistical goodness-of-fit test.

2.2.1 Mathematical Models

In general, an adjustment problem consists of two, equally important components: the functional *and* the stochastic model. Both of them are summarized by the general term *mathematical model* (see figure 2.1, LEICK 1990).

In general, measurements do not fit the mathematical model even if the mathematical model is correct. Furthermore, the model is not set up for the observed values but for (functions of) the observations, e.g. for the expectations $E(\mathbf{y})$ or for their variances $\mathbf{\Sigma_{yy}}$, cf. FÖRSTNER 1987.
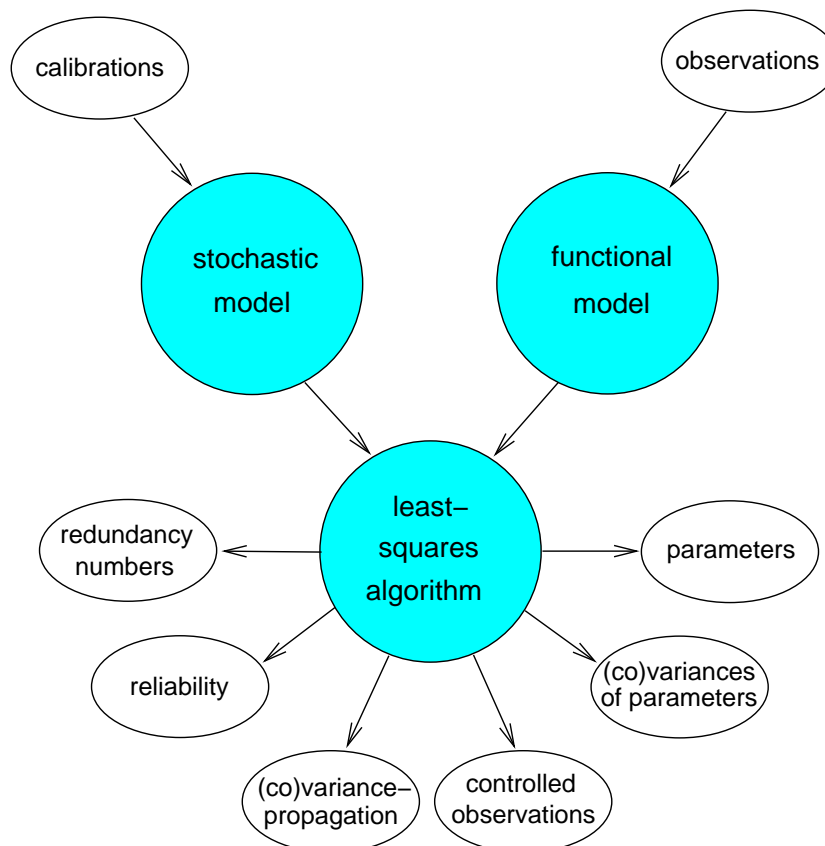
Figure 2.1: Elements of least-squares adjustment

Functional models

The functional model is the linear (or linearised) and simplified mathematical formulation of existing physical reality (LEICK 1990). It expresses the relations between observations and (unknown) parameters. In the most general case the mathematical formulation is an implicit non-linear function as

$$f(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = 0, \tag{2.1}$$

which is also known as mixed adjustment model or (together with a stochastic model) as *Gauss-Helmert model* (cf. KOCH 1999). Here, the variables denote:

$$
\begin{aligned}
\hat{\mathbf{y}} &= \quad \text{vector of } n \text{ adjusted observations} \\
\hat{\mathbf{x}} &= \quad \text{vector of } u \text{ adjusted parameters} \\
f &= \quad r \text{ non-linear mathematical functions.}
\end{aligned}
$$

In many cases, equation (2.1) can be simplified if the observations can be expressed explicitly in terms of the unknown parameters, i.e., if

$$\hat{\mathbf{y}} = f(\hat{\mathbf{x}}). \tag{2.2}$$

This model is also known as *observation equation model* or (together with a stochastic model) as *Gauss-Markov model*. Model (2.2) is of particular importance in geodetic adjustment problems since its parameters can be determined by standard algorithms and without special computational requirements. Since most of the VLBI data analysis software packages are based on model (2.2), it will be of main importance for the investigations carried out in chapter 5.

For the sake of completeness a third model has to be mentioned which does not contain any parameters at all:

$$f(\hat{\mathbf{y}}) = 0. \tag{2.3}$$

Model (2.3) is also known as *condition equation model* and has been of importance before fast computers have been available (see e.g. KOCH 1999).

Observation Equation Model

Usually model (2.2) has to be linearised by applying Taylor's theorem, i.e.,

$$
\begin{aligned}
f_i(x_1, \ldots, x_u) &= f_i(x_{1_0} + \Delta x_1, \ldots, x_{u_0} + \Delta x_u) \\
&= f_i(x_{1_0}, \ldots, x_{u_0}) + \left.\frac{\partial f_i}{\partial x_1}\right|_{\mathbf{x_0}} \Delta x_1 + \cdots + \left.\frac{\partial f_i}{\partial x_u}\right|_{\mathbf{x_0}} \Delta x_u + O(\Delta x^2)
\end{aligned}
\tag{2.4}
$$

leading to the linear system

$$
\begin{aligned}
y_1 &= a_{11}\Delta x_1 + a_{12}\Delta x_2 + \cdots + a_{1u}\Delta x_u \\
y_2 &= a_{21}\Delta x_1 + a_{22}\Delta x_2 + \cdots + a_{2u}\Delta x_u \\
&\quad \ldots\ldots\ldots\ldots\ldots \\
y_n &= a_{n1}\Delta x_1 + a_{n2}\Delta x_2 + \cdots + a_{nu}\Delta x_u
\end{aligned}
\tag{2.5}
$$

with a_{ij} being the partial derivative of the ith observation equation with respect to the jth parameter. The linear system (2.5) can be expressed in matrix notation as

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{2.6}$$

with the $n \times u$ design matrix $\mathbf{A}$ and the $u \times 1$-vector of unknowns $\mathbf{x}$

$$\mathbf{A} = \left[\begin{array}{ccc} a_{11} & \dots & a_{1u} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nu} \end{array} \right] \qquad x = \left[\begin{array}{c} \Delta x_1 \\ \vdots \\ \Delta x_u \end{array} \right], \tag{2.7}$$

which consists of the corrections Δx_i to the apriori values x_{i_0}, i.e.,

$$\mathbf{x} = \mid \Delta x_1, \ \dots, \ \Delta x_u \mid'. \tag{2.8}$$

Finally, the $n \times 1$-'observation' vector $\mathbf{y}$ (also known as 'observed minus computed'-vector) is computed by

$$\mathbf{y} = \mid y_1 - f_1(x_{1_0}, \ \dots, \ x_{u_0}), \dots, \ y_n - f_n(x_{1_0}, \dots, \ x_{u_0}) \mid'. \tag{2.9}$$

Stochastic models

In model (2.2) the design matrix $\mathbf{A}$ and the parameter vector $\mathbf{x}$ are assumed to be deterministic. The stochastic nature of the remaining components of a (general) mathematical model is comprised by the *stochastic model* (CASPARY and WICHMANN 1994) which consists of a variance-covariance matrix

$$\mathbf{\Sigma_{yy}} = \sigma_0^2 \ \mathbf{P_{yy}^{-1}} \tag{2.10}$$

of the observations $\mathbf{y}$ with an unknown factor σ^2. The factor σ^2 is also known as *variance of unit weight* and can be estimated within the adjustment process. The inverse of the weight matrix $\mathbf{P_{yy}}$ is commonly referred to as *cofactor matrix* $\mathbf{Q_{yy}}$ (KOCH 1999).

Considering a stochastic model (i.e., $\mathbf{P_{yy}} \neq \mathbf{I}$) leads to the weighted least-squares approach which can also be interpreted geometrically: From a geometrical point of view the inclusion of the weight matrix $\mathbf{P}$ generalises the standard inner product from $\mathbf{x}' \ \mathbf{I} \ \mathbf{y}$ (and thus the euclidean norm) to $\mathbf{x}' \ \mathbf{P_{yy}} \ \mathbf{y}$ and thus defines a new metric for $\mathbb{R}^n$ (CASPARY and WICHMANN 1994, or ÁDÁM 1982). As a result, the associated *orthogonal* projections become *oblique* and lead to some extended formulations for least-squares estimators (see e.g. TEUNISSEN 2003).

Since in the following investigations only the design of the experiment (i.e., the observation geometry) is of interest, no stochastic models will be included and thus $\mathbf{P_{yy}} = \mathbf{I}$. However, even in the case of $\mathbf{P_{yy}} \neq \mathbf{I}$ the same algorithms and procedures can be used after *homogenisation* of the design matrix $\mathbf{A}$ and of the observation vector $\mathbf{y}$ (see e.g. KOCH 1999) by using the Cholesky factorisation $\mathbf{P_{yy}} = \mathbf{GG}'$ of $\mathbf{P_{yy}}$ and

$$\bar{\mathbf{A}} = \mathbf{G}'\mathbf{A}, \bar{\mathbf{y}} = \mathbf{G}'\mathbf{y}. \tag{2.11}$$

2.3 Parameter estimation techniques

2.3.1 Forward and inverse problems

In geology or geophysics the problem of determining parameters from a linear model after performing observations is derived by introducing the terms 'forward problem' and 'inverse problem'. A 'forward problem' is defined to be a process of predicting the results of measurements on the basis of some general principle or model and a set of specific conditions relevant to the problem. On the other hand, 'inverse theory' is a set of mathematical techniques for reducing data to obtain useful information about the physical world on the basis of inferences drawn from observations (MENKE 1984). Inverse theory is used to provide information about the unknown parameters of a model, it does not provide the model itself. Thus, the physical model has to be specified beforehand. Schematically, the terms 'forward problem' and 'inverse problem' can be described as follows:

Forward problem:

$$\text{model parameter} \longrightarrow \text{model} \longrightarrow \text{prediction of data}$$

Inverse problem:

$$\text{data} \longrightarrow \text{model} \longrightarrow \text{estimates of model parameters}$$

'Inverse theory' is commonly used in geophysical or geological literature and forms the basis for some of the terms used below. In this context, 'inverse theory' and 'adjustment theory' might be used synonymously.

2.3.2 Linear Unbiased Estimators (LUEs and BLUEs)

The objective of inverse theory (and thus of parameter estimation in linear models) is to determine the unknown parameters of a linear model or at least to estimate linear combinations of those parameters, that can be estimated (KSHIRSAGAR 1983). The most common methods for parameter estimation are either based on probabilistic notions (such as probability functions, expectation values, unbiased estimators, etc. for the Maximum likelihood (ML) method or the Best Linear Unbiased Estimators (BLUE) approach) or on geometric notions such as the weighted least squares estimation principle as described in chapter 1.

This section reviews the necessary basics of Best Linear Unbiased Estimators and shows the relations (and equivalence) of the probabilistic and the geometric approach.

2.3.2.1 Properties of Estimators

Instead of deriving the least-squares principle and thus the solution of the model (2.2) by using vector spaces, most authors of standard geodetic literature (see e.g. KOCH 1999 or NIEMEIER 2002) make use of the conditions of linearity, unbiasedness and optimality.

As described in e.g. TEUNISSEN 2003, MEISSL 1982, or KOCH 1999, the assumption of a linear relationship between observations $\mathbf{y}$ and unknown parameters $\mathbf{x}$ can also be expressed in terms of the expectation of $\mathbf{y}$ via:

$$\mathbf{y} + \boldsymbol{\epsilon} = \mathbf{ax} \quad \text{or} \quad E(\mathbf{y}) = \mathbf{ax}. \tag{2.12}$$

Since the estimator $\hat{\mathbf{x}}$ of the unknown parameters $\mathbf{x}$ should be a linear function of the observations $\mathbf{y}$, $\hat{\mathbf{x}}$ must be of the form:

$$\hat{\mathbf{x}} = \mathbf{ly} \qquad\qquad\qquad \text{L-property} \qquad (2.13)$$

Another condition requires $\hat{\mathbf{x}}$ to be an unbiased estimator of $\mathbf{x}$ (STRANG 2003), i.e.,

$$E(\hat{\mathbf{x}}) = \mathbf{x} \quad \text{for every } \mathbf{x}. \qquad\qquad \text{U-property} \qquad (2.14)$$

Finally, the class of all possible estimators $\hat{\mathbf{x}}$ that satisfy properties (2.13) and (2.14) should be restricted to those having minimum variance, i.e.,

$$\sigma_{\hat{\mathbf{x}}}^2 \text{ minimal in the class of LU-estimators.} \qquad\qquad \text{B-property} \qquad (2.15)$$

2.3.2.2 Derivation of Estimators

In order to derive best linear unbiased estimators for the model (2.12) consider the associated linear system (containing n observations and u unknowns):

$$\mathbf{A}\mathbf{x} = E(\mathbf{y}), \tag{2.16}$$

with $\mathbf{y}$ being decomposed into the adjusted (or true) observations $\hat{\mathbf{y}}$ and the observation errors $\boldsymbol{\epsilon}$:

$$E(\mathbf{y}) = \hat{\mathbf{y}} + \boldsymbol{\epsilon}. \tag{2.17}$$

Assuming that the expectation of the observation errors $\boldsymbol{\epsilon}$ equals zero, i.e.,

$$E(\boldsymbol{\epsilon}) = \mathbf{0} \tag{2.18}$$

implies that

$$E(\mathbf{y}) = \hat{\mathbf{y}} \quad \text{and} \quad E(\mathbf{y}) = \mathbf{A}\mathbf{x}. \tag{2.19}$$

2.3.2.3 Linear Unbiased Estimators (LUEs) of estimable functions

Generalisation of the estimation of individual parameters leads to the concept of estimability of a linear function φ on the column space $R(\mathbf{A})$ of the design matrix $\mathbf{A}$ of a linear system. Since any vector in $R(\mathbf{A})$ can be represented by its coordinates $\mathbf{x} = (x_1, \ldots, x_u)$ (full rank assumed) with respect to the bases represented by the columns of $\mathbf{A}$, the function φ can be expressed as a linear function of the unknown parameters:

$$\varphi = \boldsymbol{\varphi}'\mathbf{x} = \varphi_1\mathbf{x}_1 + \varphi_2\mathbf{x}_2 + \cdots + \varphi_u\mathbf{x}_u. \tag{2.20}$$

As a special case, equation (2.20) contains the estimation of the individual parameters $\mathbf{x}_i$. Thus, any component x_i of the parameter vector $\mathbf{x}$ may be viewed as a function on $R(\mathbf{A})$. Examples for estimable functions can be found in MEISSL 1982 or KSHIRSAGAR 1983.

Furthermore, any component $\hat{\mathbf{y}}_i$ of the adjusted observations $\hat{\mathbf{y}}$ may be viewed as a function on $R(\mathbf{A})$, since

$$\hat{y}_i = a_{i1}x_1 + \cdots + a_{iu}x_u \tag{2.21}$$

and so $\boldsymbol{\varphi}$ is represented by

$$\boldsymbol{\varphi}' = (a_{i1}, \ldots, a_{iu}), \tag{2.22}$$

i.e., by the i-th row of the design matrix $\mathbf{A}$ (MEISSL 1982).

Since the parameters $\mathbf{x}$ are unknown, the functional $\varphi(x) = \boldsymbol{\varphi}'\mathbf{x}$ is unknown and has to be determined by a linear function of the observations $\mathbf{y}$. As derived in e.g. MEISSL 1982, a linear unbiased estimate (LUE) $\hat{\varphi}$ for the functional φ is derived by finding the coefficients β_i of the linear function:

$$\hat{\varphi} = \beta_1 y_1 + \beta_2 y_2 + \cdots + \beta_n y_n \qquad \text{or} \qquad \hat{\varphi} = \boldsymbol{\beta}'\,\mathbf{y} \quad \text{with} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \tag{2.23}$$

which is the $1 \times n$ representation of a linear function defined on $\mathbb{R}^n$. The expectation of $\hat{\varphi}$ reads

$$E(\hat{\varphi}) = E(\boldsymbol{\beta}'\mathbf{y}) = \boldsymbol{\beta}' E(\mathbf{y}) = \boldsymbol{\beta}'\hat{\mathbf{y}} = \boldsymbol{\beta}'\mathbf{A}\mathbf{x}, \tag{2.24}$$

which is a linear function of the unknown parameters.

A definition of the estimability of a linear function is based on the requirement that a linear function of the unknown parameters is said to be estimable if there exists at least one linear function of the observations $\boldsymbol{\beta}'\mathbf{y}$, such that $E(\boldsymbol{\beta}'\mathbf{y})$ equals $\boldsymbol{\varphi}'\mathbf{x}$ (KSHIRSAGAR 1983), i.e.,

$$E(\boldsymbol{\beta}'\mathbf{y}) = \boldsymbol{\varphi}'\mathbf{x} \quad \text{or} \quad \boldsymbol{\beta}'\mathbf{A}\mathbf{x} = \boldsymbol{\varphi}'\mathbf{x}, \tag{2.25}$$

which is equivalent to

$$\boldsymbol{\beta}'\mathbf{A} = \boldsymbol{\varphi}'. \tag{2.26}$$

Equation (2.26) shows that a necessary and sufficient condition for a linear function $\boldsymbol{\varphi}'\mathbf{x}$ for the model (2.12) to be estimable is that $\boldsymbol{\varphi}'$ is a linear combination of the row vectors of $\mathbf{A}$. Thus, only if the row vectors of $\mathbf{A}$ generate a complete basis for the row space of $\mathbf{A}$ every parameter can be estimated separately. Otherwise only linear combinations of parameters can be estimated. In other words: The row space of $\mathbf{A}$ indicates all possible estimable parameters or linear combinations of estimable parameters (!).

2.3.2.4   Best Linear Unbiased Estimators (BLUEs) of estimable functions

The definition of estimability only guarantees the existence of at least one unbiased estimate of an estimable function. Neither does it provide a method of obtaining an explicit formula for an estimator nor does it give a 'best' estimate (as stated in condition (2.15)) (SEARLE 1982).

According to KSHIRSAGAR 1983 the definition of a best linear unbiased estimator (BLUE) reads:

Definition of a BLUE

A linear function $\mathbf{b}'\mathbf{y}$ of the observations $\mathbf{y}$ in the model $\mathbf{A}\mathbf{x} = \mathbf{y} + \boldsymbol{\epsilon}$ is said to be the Best Linear Unbiased Estimate (BLUE) of a function $\boldsymbol{\varphi}'\mathbf{x}$, if it is unbiased for $\boldsymbol{\varphi}'\mathbf{x}$ and its variance is the smallest among all linear estimates of $\boldsymbol{\varphi}'\mathbf{x}$.

2.3.3   Gauss-Markov model

A method of obtaining the best linear unbiased estimate of any estimable function $\boldsymbol{\varphi}'\mathbf{x}$ of the unknown parameters $\mathbf{x}$ is provided by the following fundamental definition (KOCH 1999):

Gauss-Markov Model

Let $\mathbf{A}$ be an $n \times u$ matrix of given coefficients, $\mathbf{x}$ a $u \times 1$ vector of unknown, fixed parameters, $\mathbf{y}$ an $n \times 1$ random vector of observations and $\boldsymbol{\Sigma}_{\mathbf{yy}} = \sigma^2 \mathbf{P}_{\mathbf{yy}}^{-1}$ the $n \times n$ covariance matrix of $\mathbf{y}$, where the weight matrix $\mathbf{P}_{\mathbf{yy}}$ of the observations $\mathbf{y}$ is known and the positive factor σ^2 is unknown. Let $\mathbf{A}$ have full column rank, i.e., $rank\ \mathbf{A} = u$, and let the weight matrix $\mathbf{P}_{\mathbf{yy}}$ be positive definite. Then

$$\mathbf{A}\mathbf{x} = E(\mathbf{y}) \quad \text{with} \quad \boldsymbol{\Sigma}_{\mathbf{yy}} = \sigma^2 \mathbf{Q}_{\mathbf{yy}} = \sigma^2 \mathbf{P}_{\mathbf{yy}}^{-1} \tag{2.27}$$

is said to be a *Gauss-Markov model* with full rank. $E(\mathbf{y})$ denotes the expectation of $\mathbf{y}$.

Equation (2.27) can also be formulated in terms of the real observations. Since $\mathbf{y}$ almost certainly lies outside the column space of $\mathbf{A}$ equation (2.27) can also be formulated as a consistent system by adding a $n \times 1$ random vector $\mathbf{e}$ of errors. Equation (2.27) then reads

$$\mathbf{A}\mathbf{x} = \mathbf{y} + \mathbf{e} \quad \text{with} \quad E(\mathbf{e}) = \mathbf{0} \quad \text{and} \quad \boldsymbol{\Sigma}_{\mathbf{ee}} = \boldsymbol{\Sigma}_{\mathbf{yy}} = \sigma^2 \mathbf{P}_{\mathbf{yy}}^{-1}. \tag{2.28}$$

Solution of the Gauss-Markov model

The best linear unbiased estimator $\hat{\mathbf{x}}$ of the unknown parameters $\mathbf{x}$ in model (2.27) and its covariance matrix $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}$ is given by (see KOCH 1999)

$$\hat{\mathbf{x}} = (\mathbf{A'PA})^{-1}\mathbf{A'Py} \quad \text{and} \quad \boldsymbol{\Sigma}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = \sigma^2(\mathbf{A'PA})^{-1} \tag{2.29}$$

and thus agrees with the weighted least squares solution of a linear system as derived in chapter 1.

A proof can be found in any literature about linear models, such as KOCH 1999, KSHIRSAGAR 1983 or TOUTENBURG 2003. An algebraic proof can be found in MEYER 2000. General derivations of the variances of best estimators can be found in MEISSL 1982.

2.4 Geometric aspects of parameter estimation

As explained in chapter 1, the least-squares solution of a linear system can be interpreted geometrically. The column space $R(\mathbf{A})$ of the design matrix $\mathbf{A}$ is defined by all possible linear combinations of the columns of $\mathbf{A}$. For a general $n \times u$-matrix $\mathbf{A}$ of rank r the column space is an r-dimensional subspace of $\mathbb{R}^n$. Figure 2.2 gives an example for a 3×2-system of full rank. As mentioned above, only the unweighted case (i.e., $\mathbf{P_{yy}} = \mathbf{I}$) will be considered. For weighted least-squares estimation the projections will become oblique and thus more complex (interpretations for this case can be found in e.g. TEUNISSEN 2003).
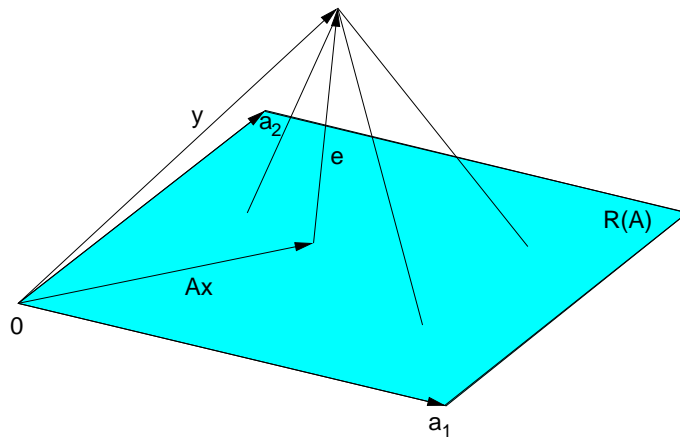
Figure 2.2: Least-squares geometry (example for a 3×2-linear system)

From the geometry of figure 2.2 it seems intuitively appealing to estimate $\mathbf{x}$ as $\hat{\mathbf{x}}$, such that $\mathbf{A}\hat{\mathbf{x}}$ is as close as possible to the observation vector $\mathbf{y}$. $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}$ are the 'adjusted observations' and $\hat{\mathbf{x}}$ denotes the coordinates of $\hat{\mathbf{y}}$ with respect to the basis formed by the columns of $\mathbf{A}$. $\hat{\mathbf{y}}$ is computed by orthogonally projecting the observations $\mathbf{y}$ onto the column space of $\mathbf{A}$. On the other hand, projecting $\mathbf{y}$ onto $R(\mathbf{A})^\perp$, i.e., the orthogonal complement of the column space $R(\mathbf{A})$, yields the residual vector $\hat{\mathbf{e}}$ (see fig. 2.3 on the next page). Thus, $\hat{\mathbf{e}}$ is orthogonal to the 'plane' spanned by the columns of $\mathbf{A}$, i.e.,

$$\mathbf{A'}(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0}, \tag{2.30}$$

which is equivalent to the well-known normal equations

$$\mathbf{A'A}\hat{\mathbf{x}} = \mathbf{A'y} \tag{2.31}$$

and thus

$$\hat{\mathbf{x}} = (\mathbf{A'A})^{-1}\mathbf{A'y}, \tag{2.32}$$

which agrees with equation (2.29) (for $\mathbf{P} = \mathbf{I}$).

Using the projection operator onto the column space $R(\mathbf{A})$:

$$\mathbf{H} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \tag{2.33}$$

and the projection operator onto the orthogonal complement $R(\mathbf{A})^{\perp}$ of $R(\mathbf{A})$

$$\mathbf{H}^{\perp} = \mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}', \tag{2.34}$$

the vector $\hat{\mathbf{y}}$ of adjusted observations and the residual vector $\hat{\mathbf{e}}$ can also be computed via

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} = \mathbf{A}\hat{\mathbf{x}} \quad \text{and} \quad \hat{\mathbf{e}} = -(\mathbf{I} - \mathbf{H})\mathbf{y} = -\mathbf{H}^{\perp}\mathbf{y} \tag{2.35}$$

to decompose the observation vector $\mathbf{y}$ into the orthogonal complements

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}} = \mathbf{H}\mathbf{y} + (\mathbf{I} - \mathbf{H})\mathbf{y}. \tag{2.36}$$

Figure 2.3 gives an illustration of the decomposition of $\mathbf{y}$ into orthogonal complements (ÁDÁM 1982).
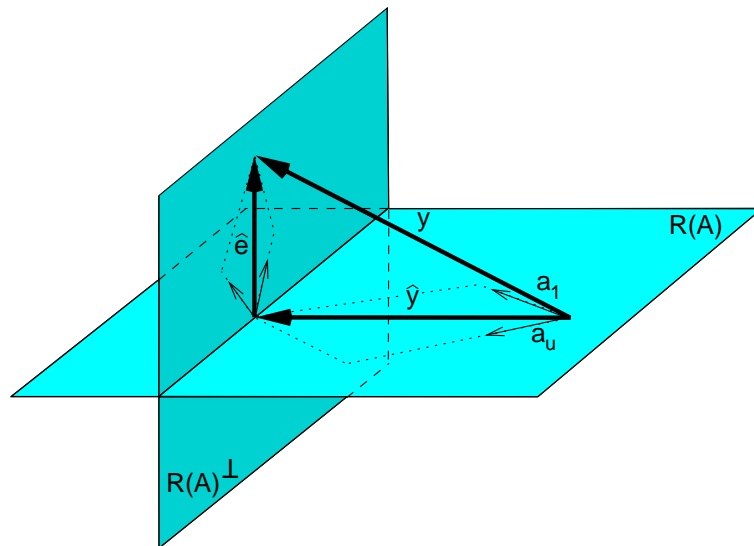
Figure 2.3: Decomposition of the observation vector $\mathbf{y}$ (in $\mathbb{R}^n$) into orthogonal complements

2.4.1   Data and model space

Orthogonal decompositions of $\mathbb{R}^u$ and $\mathbb{R}^n$

The concepts derived above can be further generalised by using the four fundamental vector spaces of a matrix to provide a comprehensive geometric explanation of parameter estimation in linear models and to show applications of the singular value decomposition for 'regression diagnostics'. The following derivations are based on SNIEDER and TRAMPERT 2000.

The geometric aspects of a general parameter estimation problem can be visualised as shown in figure 2.4 on the facing page which includes both the decomposition of the observation vector $\mathbf{y}$ into orthogonal complements (in $\mathbb{R}^n$) and the orthogonal decomposition of the parameter vector $\mathbf{x}$ (in $\mathbb{R}^u$). Again, the observation vector $\mathbf{y}$ is decomposed into a component $\hat{\mathbf{y}}$ belonging to the column space $R(\mathbf{A})$ of the design matrix $\mathbf{A}$ (denoted by $\mathbf{U}_r$ in fig. 2.4) and a component $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{A}\mathbf{x}$ which belongs to the orthogonal complement $R(\mathbf{A})^{\perp}$

of the column space of the design matrix (denoted by $\mathbf{U}_0$). In addition, the parameter vector $\mathbf{x}$ is decomposed into a component $\mathbf{x}_r$ belonging to a subspace of $\mathbb{R}^u$ (denoted by $\mathbf{V}_r$) and a component $\mathbf{x}_0 = \mathbf{x} - \mathbf{x}_r$ belonging to the orthogonal complement of $\mathbf{V}_r$ (denoted as $\mathbf{V}_0$).

As shown in fig. 2.4 the design matrix $\mathbf{A}$ serves as a (linear) mapping from $\mathbb{R}^u$ to $\mathbb{R}^n$; its pseudo-inverse $\mathbf{A}^+$ (see section 1.6.2.3 on page 25) thus maps from $\mathbb{R}^n$ to $\mathbb{R}^u$.
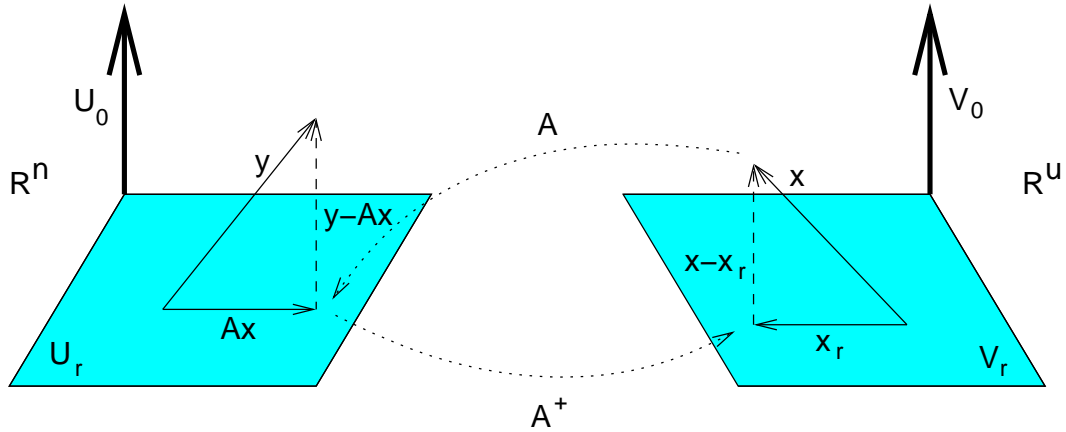
Figure 2.4: Geometrical interpretation of parameter estimation in linear models (left: Data space and data null space, right: model space and model null space, the design matrix acts as a mapping between $\mathbb{R}^u$ and $\mathbb{R}^n$) (SNIEDER and TRAMPERT 2000)

Eigenvalue decomposition of square linear systems (EVD)

Figure 2.4 might be best understood by first considering the eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ of a symmetric $u \times u$ linear system $\mathbf{Ax} = \mathbf{y}$ of full rank. Assuming that the eigenvectors form an orthonormal set, a vector $\mathbf{x}$ can be projected on these eigenvectors and can thus be expressed as

$$\mathbf{x} = \sum_{i=1}^{u} \mathbf{v}_i \left(\mathbf{v}_i' \cdot \mathbf{x} \right). \tag{2.37}$$

The product $\mathbf{Ax}$ can now be written as:

$$\mathbf{Ax} = \sum_{i=1}^{u} \lambda_i \mathbf{v}_i \left(\mathbf{v}_i' \cdot \mathbf{x} \right) = \mathbf{y}. \tag{2.38}$$

After expressing the vector $\mathbf{y}$ with respect to the same eigenvector basis as

$$\mathbf{y} = \sum_{i=1}^{u} \mathbf{v}_i \left(\mathbf{v}_i' \cdot \mathbf{y} \right), \tag{2.39}$$

equation (2.38) yields the following expansion for the solution vector $\mathbf{x}$:

$$\mathbf{x} = \sum_{i=1}^{u} \frac{1}{\lambda_i} \mathbf{v}_i \left(\mathbf{v}_i' \cdot \mathbf{y} \right). \tag{2.40}$$

Equation (2.40) shows that small eigenvalues (e.g. caused by a bad condition or even rank deficiencies of the linear system) lead to unreasonable contributions to the solution and (depending on the current application) should be omitted (SCHWARZ 1997).

Singular value decomposition of rectangular systems

For a rectangular linear system of rank r the dimensions of the observation vector $\mathbf{y}$ (in $\mathbb{R}^n$) and the solution vector $\mathbf{x}$ (in $\mathbb{R}^u$) differ. Thus, for the derivation of rectangular bases for $\mathbb{R}^n$ and $\mathbb{R}^u$ the singular value decomposition (cf. section 1.6.1.2 on page 23)

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}' \tag{2.41}$$

of the design matrix $\mathbf{A}$ has to be used. Instead of using one common basis, a new basis for $\mathbb{R}^n$ (consisting of n left-singular vectors $\mathbf{u}_i$) and a new basis for $\mathbb{R}^u$ (consisting of u right-singular vectors $\mathbf{v}_i$) is used. The relation between the two bases reads

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i, \tag{2.42}$$

with σ_i being the u singular values of $\mathbf{A}$. The product $\mathbf{A}\mathbf{x}$ can now be written as:

$$\mathbf{A}\mathbf{x} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \left(\mathbf{v}_i' \cdot \mathbf{x} \right) = \mathbf{y}, \tag{2.43}$$

with r being the rank of the linear system and thus being the number of non-zero singular values σ_i.

As described in chapter 1 the left singular vectors $\mathbf{u}_i$ and the right singular vectors $\mathbf{v}_i$ can be arranged to form the following matrices $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$. The matrix $\mathbf{U}$

$$\mathbf{U} = \begin{pmatrix} \vdots & \vdots & & \vdots & \vdots & & \vdots \\ u_1 & u_2 & \dots & u_r & u_{r+1} & \dots & u_n \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ & \underbrace{\hphantom{xxxxxxxx}}_{U_r} & & & \underbrace{\hphantom{xxxxxxxx}}_{U_0} & & \end{pmatrix}, \tag{2.44}$$

contains the left singular vectors $\mathbf{u}_i$ corresponding to the order of the singular values σ_i, which are (usually) arranged in descending order, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$, in a matrix $\mathbf{S}$:

$$\mathbf{S} = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{2.45}$$

The right singular vectors $\mathbf{v}_i$ form the columns of the matrix $\mathbf{V}$:

$$\mathbf{V} = \begin{pmatrix} \vdots & \vdots & & \vdots & \vdots & & \vdots \\ v_1 & v_2 & \dots & v_r & v_{r+1} & \dots & v_u \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ & \underbrace{\hphantom{xxxxxxxx}}_{V_r} & & & \underbrace{\hphantom{xxxxxxxx}}_{V_0} & & \end{pmatrix}. \tag{2.46}$$

Equation (2.43) shows that the singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ for $i > r$ do not contribute when $\mathbf{A}$ acts on a vector. Thus, the matrix $\mathbf{A}$ can be constructed from $\mathbf{U}_r$, $\mathbf{S}_r$ and $\mathbf{V}_r$ alone. According to (SNIEDER and TRAMPERT 2000) '$\mathbf{U}_0$ and $\mathbf{V}_0$ are dark spots of the space not illuminated by the operator $\mathbf{A}$'.

The subspace formed by the left singular vectors $\mathbf{u}_i$ of $\mathbf{U}_r$ corresponds to the column space $R(\mathbf{A})$ of $\mathbf{A}$ (as introduced in chapter 1) and is called *data space*. The subspace formed by the right singular vectors $\mathbf{v}_i$ of $\mathbf{V}_r$ corresponds to the row space $R(\mathbf{A}')$ of $\mathbf{A}$ and is called *model space*.

	Linear Algebra (Chapter 1)	Parameter Estimation in Linear Models (Chapter 2)
$\mathbf{U}_r = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ $\mathbf{U}_0 = \{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$	basis for column space $R(\mathbf{A})$ basis for orthogonal complement $R(\mathbf{A})^\perp$ of column space $R(\mathbf{A})$	basis for data space basis for data-null-space
$\mathbf{V}_r = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ $\mathbf{V}_0 = \{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_u\}$	basis for row space $R(\mathbf{A}')$ basis for orthogonal complement $R(\mathbf{A}')^\perp$ of row space $R(\mathbf{A}')$ (= null space of $\mathbf{A}$)	basis for model space basis for model-null-space

Table 2.1: Relations between terms used in Linear Algebra (chapter 1) and Parameter Estimation (chapter 2)

Since the adjusted observations $\mathbf{Ax}$ are orthogonal to $\mathbf{U}_0$, i.e., $\mathbf{U}_0'\mathbf{Ax} = \mathbf{0}$, any component of the observation vector $\mathbf{y}$ that lies in $\mathbf{U}_0$ cannot be explained by the (current) functional model. These components thus correspond to errors in the data or to errors in the functional model expressed by the operator $\mathbf{A}$. Therefore, $\mathbf{U}_0$ is called the *data-null-space* of $\mathbf{A}$.

On the other hand, limiting the summation in equation (2.43) to non-zero singular values restricts the estimated parameter vector $\mathbf{x}$ to the subspace $\mathbf{V}_r$ (model space). In other words: As shown in section 2.3.2.3 on page 33, the row space contains estimable (functions of the) unknown parameters and so the model parameters do not contain any components of the subspace $\mathbf{V}_0$ (which is also called *model-null-space*). This is a geometrical visualisation of the general solution of a linear system, i.e., $\mathbf{x}$ is decomposed into the particular solution $\mathbf{x}_r$ of the inhomogeneous system (and thus a component of $\mathbf{V}_r$) and a solution $\mathbf{x} - \mathbf{x}_r$ of the corresponding homogeneous system which is a component of the model null-space $\mathbf{V}_0$. Since $\mathbf{V}_0$ is the null space of $\mathbf{A}$ and so $\mathbf{AV}_0 = \mathbf{0}$, any parameter of the model that lies within $\mathbf{V}_0$ does not affect the data. According to SNIEDER and TRAMPERT 2000 'the data have no bearing on the components of the model vector that lie in $\mathbf{V}_0$'.

The data space and the data-null-space thus span $\mathbb{R}^n$, while the model space and the model-null-space span $\mathbb{R}^u$. Table 2.1 shows the equivalence of the algebraic terms used in chapter 1 with the terms used for parameter estimation in linear models within this chapter.

Restricting the solution vector to $\mathbf{V}_r$ and expanding $\mathbf{x}$ in the basis formed by $\mathbf{V}_r$ and expanding $\mathbf{y}$ in the basis formed by $\mathbf{U}_r$ yields the general least-squares solution of an over-determined linear system as

$$\hat{\mathbf{x}} = \sum_{i=1}^{r} \frac{1}{\sigma_i} \mathbf{v}_i \left(\mathbf{u}_i' \cdot \mathbf{y}\right) \tag{2.47}$$

or in matrix notation

$$\hat{\mathbf{x}} = \underbrace{\mathbf{V}_r \mathbf{S}_r^{-1} \mathbf{U}'_r}_{\mathbf{A}^+} \mathbf{y}. \tag{2.48}$$

$\mathbf{V}_r \mathbf{S}_r^{-1} \mathbf{U}'_r$ denotes the pseudo-inverse of $\mathbf{A}$ (SNIEDER and TRAMPERT 2000).

2.4.2 Resolution in parameter estimation

The geometrical concept of projections onto the model space and the data space can be used to derive indicators of how precisely the model parameters can be determined from the data and how well neighbouring data can be independently predicted, or *resolved* (SCALES et al. 2001).

Model resolution matrix

The *model resolution matrix (MRM)* indicates to what extent the model parameters can be independently retrieved from the estimation process (MENKE 1984). From $\hat{\mathbf{x}} = \mathbf{A}^+\mathbf{y}$, $\mathbf{Ax} = \mathbf{y}$ and the singular value decomposition of $\mathbf{A}$ follows

$$
\begin{aligned}
\hat{\mathbf{x}} &= \mathbf{A}^+\mathbf{y} \\
&= \mathbf{A}^+\mathbf{Ax} \\
&= \mathbf{V}_r\mathbf{S}_r^{-1}\mathbf{U}'_r\mathbf{U}_r\mathbf{S}_r\mathbf{V}'_r\mathbf{x} \\
&= \underbrace{\mathbf{V}_r\mathbf{V}'_r}_{MRM}\mathbf{x},
\end{aligned}
\tag{2.49}
$$

with $\mathbf{V}_r\mathbf{V}'_r$ being a $u \times u$ projection matrix onto the model space (SCALES et al. 2001). Only in the case of full rank the model resolution matrix equals the identity matrix and every parameter can be determined independently. Otherwise some parameters can only be estimated as linear combinations of remaining parameters. The more non-zero terms appear in the rows of the model resolution matrix, the more broadly averaged the inferences of the model parameters are.

The advantage of the model resolution matrix is that it can be computed even in the case of exactly dependent parameters (i.e., in the rank deficient case) and thus, if the computation of the correlation matrix fails (see e.g. KOCH 1999). In statistical terms, the model resolution matrix is used for the detection of so-called multicollinearity, i.e., linear dependencies of the columns of the design matrix $\mathbf{A}$ and thus for the detection of correlations between the estimated parameters (see e.g. BELSLEY et al. 1980, or TOUTENBURG 2003).

Data resolution matrix

In a similar way a *data resolution matrix (DRM)* $\mathbf{H}$ can be computed (see eq. 2.33). This matrix indicates how well the adjusted (or predicted) observations match the data or how well the data is predicted by the estimated model parameters (MENKE 1984).

The adjusted observations can be derived by

$$
\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{A}\hat{\mathbf{x}} \\
&= \mathbf{AA}^+\mathbf{y} \\
&= \mathbf{U}_r\mathbf{S}_r\mathbf{V}'_r\mathbf{V}_r\mathbf{S}_r^{-1}\mathbf{U}'_r\mathbf{y} \\
&= \underbrace{\mathbf{U}_r\mathbf{U}'_r}_{DRM}\mathbf{y},
\end{aligned}
\tag{2.50}
$$

with

$$
\mathbf{H} = \mathbf{U}_r\mathbf{U}'_r
\tag{2.51}
$$

being an $n \times n$ projection operator onto the data space of $\mathbf{A}$. Other definitions are solely based on the design matrix $\mathbf{A}$ (see also eq. 2.33):

$$
\mathbf{H} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}',
\tag{2.52}
$$

or (if the metric of the data space is also included by considering $\mathbf{\Sigma}_{\mathbf{yy}}^{-1}$):

$$
\mathbf{H} = \mathbf{A}(\mathbf{A}'\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{\Sigma}_{\mathbf{yy}}^{-1}.
\tag{2.53}
$$

The general definition of $\mathbf{H}$ (eq. 2.53, see also e.g. FÖRSTNER 1987) thus also takes into account the stochastic model of the observations by including the covariance matrix $\mathbf{\Sigma}_{\mathbf{yy}}$ of the observations $\mathbf{y}$ and thus accounts for the metric of the vector spaces involved. Since the following investigations do not contain any metric

aspects, $\mathbf{\Sigma_{yy}}$ equals the identity matrix (see page 31). As shown by e.g. TOUTENBURG 2003, this yields a symmetric data resolution matrix.

Since equation (2.52) contains the computation and inversion of normal equations $\mathbf{A'A}$, it bears some numerical problems. Using the singular value decomposition of $\mathbf{A}$, numerically more stable derivations of the data resolution matrix are given below.

The Data Resolution Matrix $\mathbf{H}$ is also known as 'Hat-Matrix' or 'prediction matrix' and serves as a regression diagnostics tool in many sciences such as statistics (TOUTENBURG 2003, COOK and WEISBERG 1982) or geophysics (PARKER 1994). According to (HOAGLIN and WELSCH 1978) 'a look at the hat matrix can reveal sensitive points in the design, i.e., points at which the value of y_i has a large impact on the fit' and is thus used to identify 'high-leverage points'. The general data resolution matrix (see eq. 2.53) depends both on the geometry (i.e., the design) of the experiment and the covariance matrix of the observations. It does not depend on the individual observations. Since the data resolution matrix is a projection matrix it has the following properties (FÖRSTNER 1987):

- symmetry (only if $\mathbf{\Sigma_{yy}} = \mathbf{I}$),

- idempotence (i.e., $\mathbf{H}^2 = \mathbf{H}$),

- eigenvalues are either 1 or 0 and

- the trace of $\mathbf{H}$ equals the rank of $\mathbf{H}$, i.e., $tr(\mathbf{H}) = rk(\mathbf{H}) = u$.

As shown in TOUTENBURG 2003 or COOK and WEISBERG 1982, the range of the elements of the data resolution matrix (for $\mathbf{\Sigma_{yy}} = \mathbf{I}$) is:

$$0 \leq h_{ii} \leq 1 \quad \text{and} \quad -0.5 \leq h_{ij} \leq 0.5. \tag{2.54}$$

Based on the representation

$$\hat{y}_i = h_{ii}y_i + \sum_{j=1}^{n} h_{ij}y_j \tag{2.55}$$

some authors (e.g. COOK and WEISBERG 1982) show that h_{ii} is the amount of leverage or influence exerted on $\hat{y}_i$ by y_i. A large main diagonal element with $h_{ii} \approx 1$ thus indicates that $\hat{y}_i$ is almost completely determined by y_i alone. Thus, observations y_i with large values h_{ii} can exert an undue effect on the least squares results (DODGE and JURECKOVÁ 2000). A small element ($h_{ii} \approx 0$) also leads to a small impact of the remaining observations, i.e., if a diagonal element h_{ii} equals zero the corresponding row of $\mathbf{H}$ is $\mathbf{0}$, which indicates that the ith observation does not affect the fit (for a proof, see TOUTENBURG 2003, or DODGE and JURECKOVÁ 2000). For a linear regression (with x_i indicating the x-component of an observation y_i), a main diagonal element h_{ii} can also be computed by

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2}, \tag{2.56}$$

which shows that h_{ii} mostly depends on the 'distance' $|x_i - \bar{x}|$ of an observation x_i to the centre of mass $\bar{x}$ of all observations. Therefore, some authors call the data resolution matrix a 'distance measure matrix' (BELSLEY et al. 1980).

Another interpretation of the data resolution matrix is based on the fact that the covariance matrix $\mathbf{\Sigma_{\hat{y}\hat{y}}}$ of the adjusted observations $\hat{\mathbf{y}}$ equals $\mathbf{\Sigma_{\hat{y}\hat{y}}} = \sigma^2 \cdot \mathbf{Q_{\hat{y}\hat{y}}} = \sigma^2 \cdot \mathbf{H} = \sigma^2 \cdot \mathbf{A}\mathbf{\Sigma_{\hat{x}\hat{x}}}\mathbf{A'}$. The cofactor matrix $\mathbf{Q_{\hat{y}\hat{y}}}$ of the adjusted observations will be of relevance in the next section. In addition, the covariance matrix $\mathbf{\Sigma_{vv}}$ of the residuals $\mathbf{v}$ equals $\mathbf{\Sigma_{vv}} = \sigma^2 \cdot (\mathbf{I} - \mathbf{H})$ (TOUTENBURG 2003) and thus controls the variations in the residuals (COOK and WEISBERG 1982). Due to this fact, the data resolution matrix (or the hat matrix) is extensively used for residual analyses in statistical applications (SAVILLE and WOOD 1997, TOUTENBURG 2003).

2.4.3   Impact factors and impact co-factors

Since the elements of the data resolution matrix indicate how much weight each observation has on the adjusted observations, the main-diagonal elements of $\mathbf{H}$ are called *impact factors h_{ii}* (or h_i), i.e.,

$$\text{impact factors} = \mathbf{h} = diag(\mathbf{H}), \tag{2.57}$$

while the off-diagonal elements of $\mathbf{H}$ are referred to as *impact co-factors h_{ij}*.

A close relation between impact factors and partial redundancies exists, since (for $\mathbf{P} = \mathbf{I}$)

$$\mathbf{\Sigma}_{\hat{\mathbf{v}}\hat{\mathbf{v}}} = \mathbf{I} - \mathbf{H}, \tag{2.58}$$

which is used for the computation of redundancy numbers r_i (LEICK 1990, ACKERMANN 1981):

$$r_i = 1 - h_{ii} = (\mathbf{I} - \mathbf{H})_{ii}. \tag{2.59}$$

The redundancy numbers r_i indicate the percentage of how much a gross error is shown in the residuals of the least squares fit (FÖRSTNER 1987). High leverage points (or observations with a small partial redundancy) are thus weakly controlled and complicate the detection of blunders (NIEMEIER 2002). As in geodetic networks, weakly controlled observations (or observations with a large impact factor) significantly affect the accuracy of the estimated parameters but degrade the reliability of the entire adjustment. Recent examples for geodetic applications such as redundancy analysis in plane networks can be found in e.g. EVEN-TZUR 2006.

Since the average size of a diagonal element h_{ii} of the data resolution matrix is u/n, some authors (see e.g. HOAGLIN and WELSCH 1978) recommend to mark observations as 'high-leverage points' if their impact factors exceed twice the average size, i.e., if $h_{ii} > 2 \cdot \frac{u}{n}$. For the generation of experiments which are insensitive to outliers, COOK and WEISBERG 1982 recommend experiment designs yielding small impact factors (i.e., high redundancies) of approximately the same size.

Since the data resolution matrix contains information given by the design matrix, leverage reflects only the *potential* effect of an observation on the regression. The determination of the actual effect of an observation on the regression results must also take the observations into account. Thus, many outlier detection methods are based on the residuals *and* the impact factors (or redundancies) of the observations (EEG 1986 or TOUTENBURG 2003).

Increase of uncertainty

FÖRSTNER 1992 shows the relationship of impact factors and the maximum effect of the rejection of an observation y_i onto the result $\hat{\mathbf{x}}$. Based on the projection operator onto the column space of the design matrix $\mathbf{A}$ (i.e., the data resolution matrix $\mathbf{H}$) and based on the application of sequential least-squares, sensitivity analysis for outliers can be performed. The empirical sensitivity of the results with respect to an outlier in y_i mainly depends on the *influence factor*

$$\mu_i = \sqrt{\frac{h_{ii}}{1 - h_{ii}}}, \tag{2.60}$$

which is thus only a function of the impact factor h_{ii}.

As also shown in FÖRSTNER 1992 the influence factor μ_i also measures the *relative increase of uncertainty* if the ith observation is omitted from the estimation process. The increase of uncertainty can be determined both for a group of observations and for the case where only a subset of parameters is analysed.

2.4.4 Geometrical interpretations of impact factors and impact co-factors

A geometrical interpretation and thus a graphical representation of impact factors and impact co-factors will be of importance in the next chapter and can be obtained by analysing the expression

$$h_{ij} = \mathbf{e}_i' \, \mathbf{H} \, \mathbf{e}_j, \qquad (2.61)$$

which extracts the element h_{ij} of the ith row and the jth column of the data resolution matrix using vectors $\mathbf{e}_i$ and $\mathbf{e}_j$ of the natural basis of $\mathbb{R}^n$. In general, impact co-factors h_{ij} (and as a special case impact factors h_{ii}) are obtained by the following steps:

1. At first, the projection $\mathbf{He}_j$ has to be performed. Since the length of a projected vector $\mathbf{y}$ onto a vector $\mathbf{a}$ equals $\|\mathbf{y}\| \cos \varphi_1$ (see figure 2.5) and since in this case $\mathbf{y} = \mathbf{e}_j$, the length of the projection simplifies to $\cos \varphi_1$.

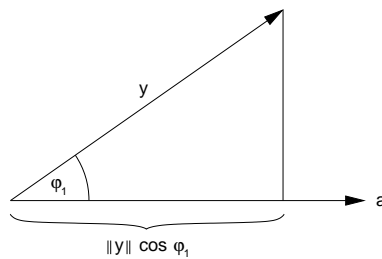Figure 2.5: Projection of vectors

2. In a second step, the projected vector (of length $\cos \varphi_1$) resulting from step 1 is projected onto $\mathbf{e}_i$ (which, in the case of impact factors h_{ii}, is a back-projection onto the same natural basis vector used in step 1). Due to the definition of the inner (or scalar) product, the length of the projection of $\mathbf{He}_j$ onto $\mathbf{e}_i$ can be expressed by using the cosine of the angle φ_2 between the projected vector from step 1 (which is contained in the column space of $\mathbf{A}$) and $\mathbf{e}_i$. The length of the vector resulting from this projection is thus $\cos \varphi_1 \cdot \cos \varphi_2$.

This leads to the general interpretation of impact co-factors h_{ij} as the product of the cosines of the angles φ_i and φ_j between the ith and jth vector of the natural basis and their respective projections onto the column space of $\mathbf{A}$, i.e.,

$$h_{ij} = \cos \varphi_i \cdot \cos \varphi_j. \qquad (2.62)$$

As a special case, eq. (2.62) contains an interpretation for the impact factors h_{ii} as

$$h_{ii} = \cos^2 \varphi_i. \qquad (2.63)$$

Due to the close relationship of projections and the cosine of the angle φ between the two vectors involved, the impact factors h_{ii} are thus proportional to $\cos^2 \varphi$. Since the angle between a spatial vector and its projection is always smaller than 90°, an almost linear descending relation between the angle and the corresponding impact factor exists (see figure 2.6).

For a 2×1-design matrix, figure 2.7 shows that h_{ii} equals the squared cosine of the angle φ_i between the $\mathbf{e}_i$-vector and its projection $\mathbf{H} \, \mathbf{e}_i$ onto the column space of $\mathbf{A}$. Furthermore, figure 2.7 shows a visualisation of impact co-factors h_{ij} and the close relation of impact factors h_{ii} and redundancy numbers r_i.

In a similar way, figure 2.8 gives a visualisation of impact factors and impact co-factors of 3×2-systems of linear equations. The first case occurs when three equally spaced observations are used for the computation of a regression line. In the second case, the third observation depicts a high-leverage observation due to its
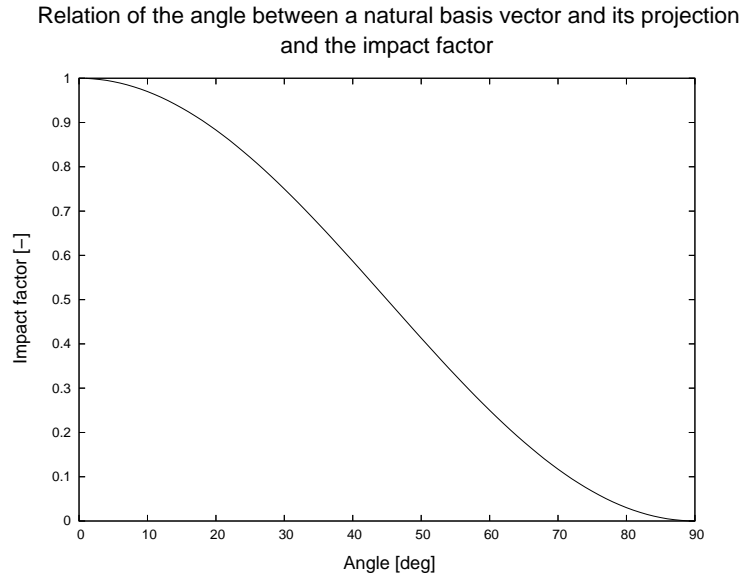
Figure 2.6: Relation of the angle between the ith vector of the natural basis and its projection and the importance factor of the ith observation

large distance from the remaining two observations (i.e., $x = 7$). This can also be recognized (in figure 2.8) by the small angle between $\mathbf{e}_3$ and $\mathbf{He}_3$ and thus the length of $\mathbf{e}'_3\mathbf{He}_3 = 0.99 \approx 1$.

A similar derivation for the interpretation of redundancy numbers is given by EEG 1986. Geometrically the respective angles in equations (2.63) and (2.62) have to be replaced by its 90°-complements since for the derivation of redundancy numbers the angles between the unit vectors of the natural basis and its projections onto the orthogonal complement of the column space of $\mathbf{A}$ have to be used. Therefore, the data resolution matrix $\mathbf{H}$ in equation (2.61) has to be replaced by its complementary operator $\mathbf{I} - \mathbf{H}$.

Impact co-factors as similarity measures

In the next chapter, co-factors h_{ij} will be used for the detection of groups of observations. Therefore, it is necessary to emphasize that only the relation of the (complementary) angles $\varphi_1, \ldots, \varphi_n$ is of importance for the size of h_{ij}. As shown in figure 2.7 on the next page, for the two-dimensional case, both angles φ_1 and φ_2 add to 90°. In general:

- identical angles $\varphi_1, \ldots, \varphi_n$ lead to impact co-factors of 0.5 (case 1). As shown in chapter 2.4.2 on page 39 and as visualised in figure 2.7 (as projections) the absolute values of h_{ij} can never exceed 0.5.

- large differences in $\varphi_1, \ldots, \varphi_n$ (e.g. one of them being small and the other one automatically being large) yield small impact co-factors h_{ij}.

Since the data resolution matrix corresponds to the (standardised) cofactor matrix $\mathbf{Q}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}$ of the adjusted observations, the impact co-factors (or the off-diagonal elements of $\mathbf{Q}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}$) show correlations of the observations. In terms of the 'observation geometry' (such as e.g. x-values of observations when determining a regression line or the orientation of both the baseline and the radio source two VLBI-telescopes are pointing at simultaneously), small impact co-factors h_{ij} indicate a significantly distinct information content of the respective observations. On the other hand, large impact co-factors h_{ij} (i.e., $\|h_{ij}\| \approx 0.5$) show that the respective observations have been performed under similar 'geometric' conditions.

Figure 2.7: Graphical visualisation of impact factors h_{ii}, impact co-factors h_{ij} and redundancies r_i for 2×1-systems of linear equations. Dotted line = first projection (onto the column space), dashed line = second projection (onto the vectors of the natural basis).
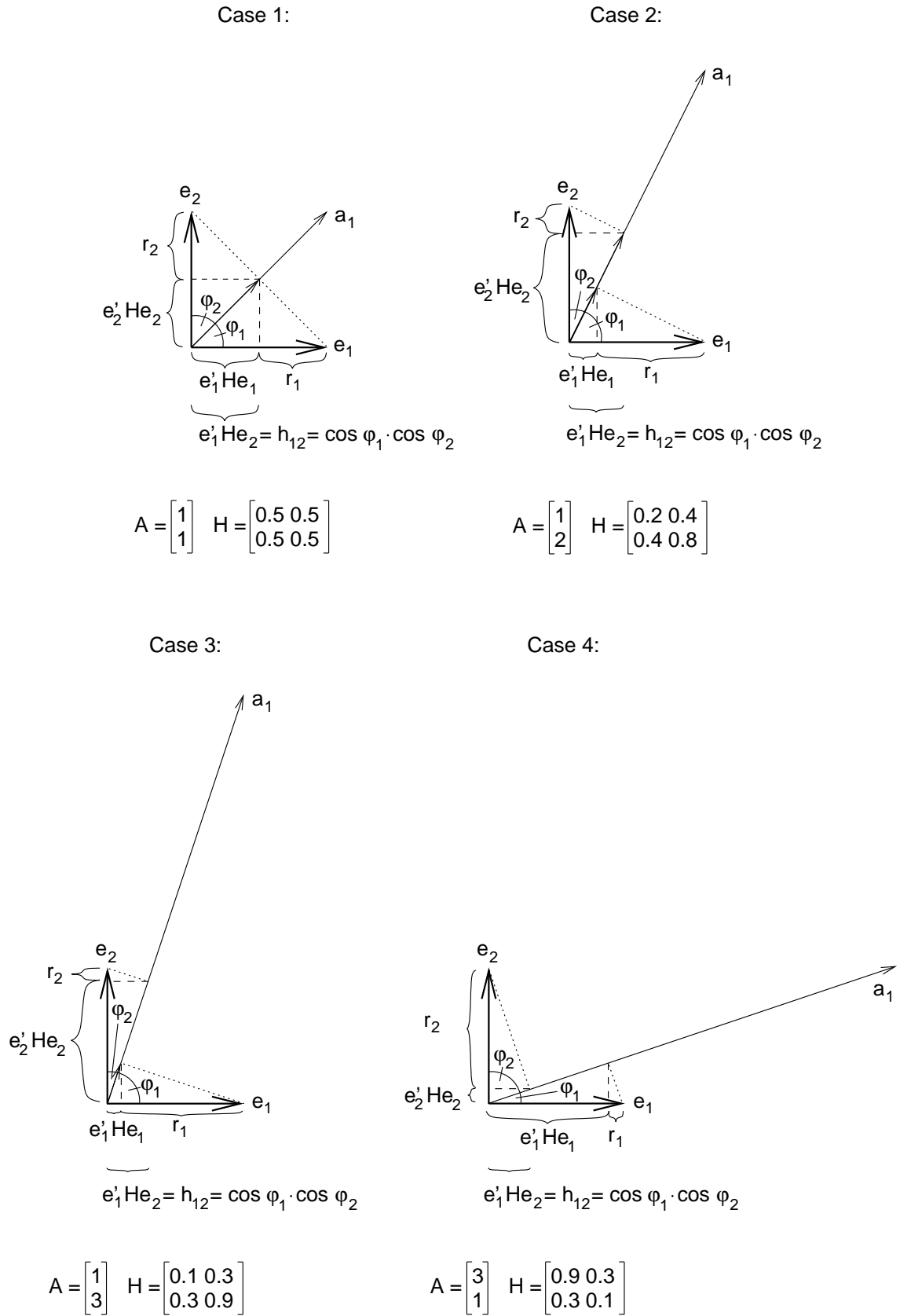
Figure 2.8: Graphical visualisation of impact factors h_{ii} and impact co-factors h_{ij} for 3×2-systems of linear equations. Dotted line = first projection (onto the column space), dashed line = second projection (onto the vectors of the natural basis).
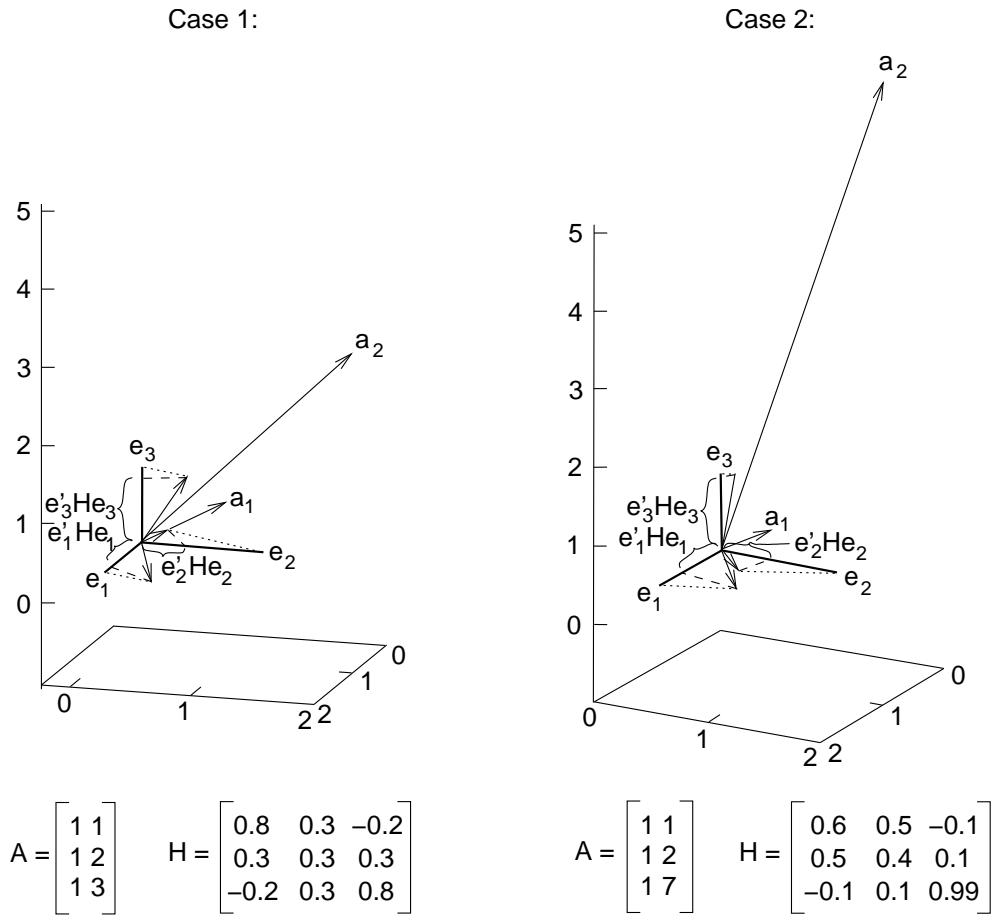
3. Cluster Analysis

3.1 Introduction

In order to detect groups of (VLBI-)observations the impact factors and impact co-factors as derived in the previous chapter can be used. The objective of this chapter is to apply objective grouping methods for an automatic detection of jointly influential groups of observations. Since the off-diagonal elements of the data resolution matrix can be interpreted as similarity measures, statistical methods for grouping similar objects can also be used in adjustment theory. A well-known method for the identification of groups of similar objects (or of observations with a similar information content) is the *cluster analysis*-approach as described in e.g. Hoaglin and Welsch 1978, or Gray and Ling 1984. These methods have been developed for statistical analyses since the 1980ies (see e.g. Belsley et al. 1980, or Romesburg 2004) or pattern recognition (see e.g. Duda et al. 2000).

In the following chapter the principles of cluster analysis methods are explained and interpretation guidelines for cluster analysis results are given. These methods will be applied to plane and spatial interferometers in chapter 5.

3.2 Cluster Analysis

One of the most common approaches for estimating similarities (or dissimilarities) between objects is given by cluster analysis methods. Based on measurable attributes, objects (such as persons, animals, pieces of land or patterns in digital photographs) can be objectively classified in groups (Romesburg 2004). However, cluster analysis can only reveal *candidates* for influential subsets (Gray and Ling 1984). Detailed descriptions and computational aspects can be found in e.g. Romesburg 2004, Duda et al. 2000 or Belsley et al. 1980. Here, only practical aspects are given. The examples given below are taken from Romesburg 2004.

Practical cluster analysis

In general, cluster analysis consists of the six steps listed in table 3.1. Some of these steps, however, cannot be transferred directly or are of no relevance for regression diagnostics.

In general cluster analysis the first two steps consist of the generation and an optional standardisation of a data matrix. The data matrix consists of a collection of attributes of objects. Figure 3.1 shows an example of five objects whose two attributes are displayed on the x- and y-axis. In a second step so-called resemblance

	General Cluster Analysis	Cluster Analysis for Regression Diagnostics
1.	Obtaining the data matrix	Setting up of design matrix
2.	Standardizing of the data matrix (optional)	-
3.	Computation of resemblance matrix	Computation of data resolution matrix
4.	Execute the clustering method	Execute the clustering method
5.	Rearrange the data and resemblance matrices	-
6.	Compute the cophenetic correlation matrix	-

Table 3.1: Terminology of general cluster analysis steps and for cluster analysis as used for regression diagnostics

coefficients are computed to measure the degree of similarity between each pair of objects. The resemblance coefficients are always either *dissimilarity coefficients* or *similarity coefficients*:

- The smaller the dissimilarity coefficient, the more similar two objects are. The larger this coefficient, the more distinct two objects are. Dissimilarity coefficients are also known as *Euclidean distance coefficients* and can be visualised geometrically (see below).

- On the other hand, large similarity coefficients indicate that two objects are very similar. A graphical representation is difficult.

In order to introduce the basic concepts of cluster analysis methods, most authors use the graphical representation of dissimilarity coefficients. The concept, however, can also be used to understand the use of similarity coefficients (as used in the investigations below). As shown in ROMESBURG 2004, 'the difference between a similarity coefficient and a dissimilarity coefficient is merely a difference in which direction the scale runs'.

Resemblance coefficients can be obtained in different ways (see e.g. DUDA et al. 2000). In the most simple case they can be interpreted as the euclidean 'distance' between each pair of objects (displayed as dashed lines in figure 3.1). Thus, small coefficients indicate a high similarity between two objects. Resemblance coefficients are arranged into a symmetric matrix of dimension *number of objects × number of objects*.
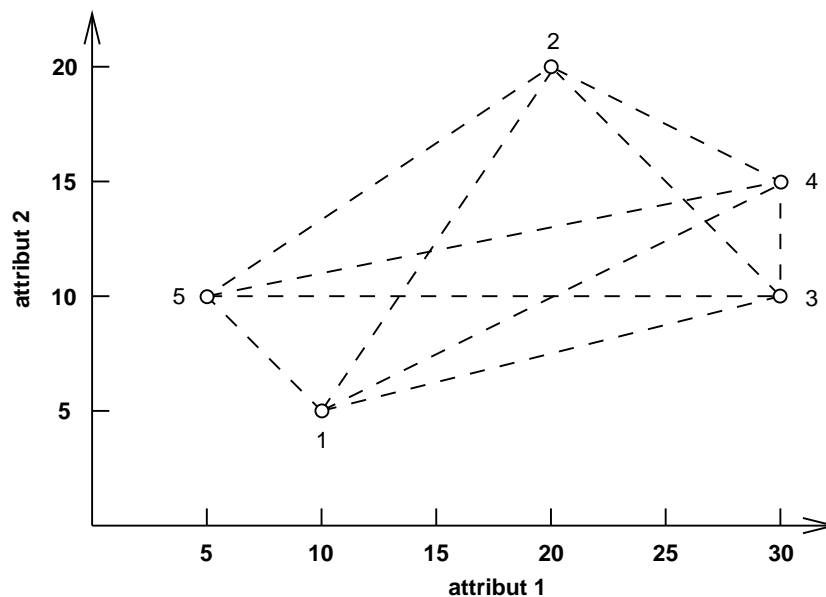
Figure 3.1: Cluster Analysis example: Graphical representation of two attributes of five objects

Geometrically, cluster analysis consists of a step-by-step forming of sets (clusters) of one or more objects of similar properties. At first, each object is regarded as an individual cluster. After the last step all objects are merged into one common cluster. Figure 3.2 on the next page gives an impression of the four clustering steps for the example given in figure 3.1.

From a computational point of view, cluster analysis consists of an iterative computation of 'distances' between each newly formed cluster. The 'distance' to a cluster with more than one object is computed by the (unweighted) average 'distance' to each of its objects. This clustering method is also known as 'unweighted pair-group method using arithmetic averages (UPGMA)' (ROMESBURG 2004). Other methods are described in DUDA et al. 2000 or RIPLEY 1996. The basic steps of a general cluster analysis algorithm are shown in figure 3.3.
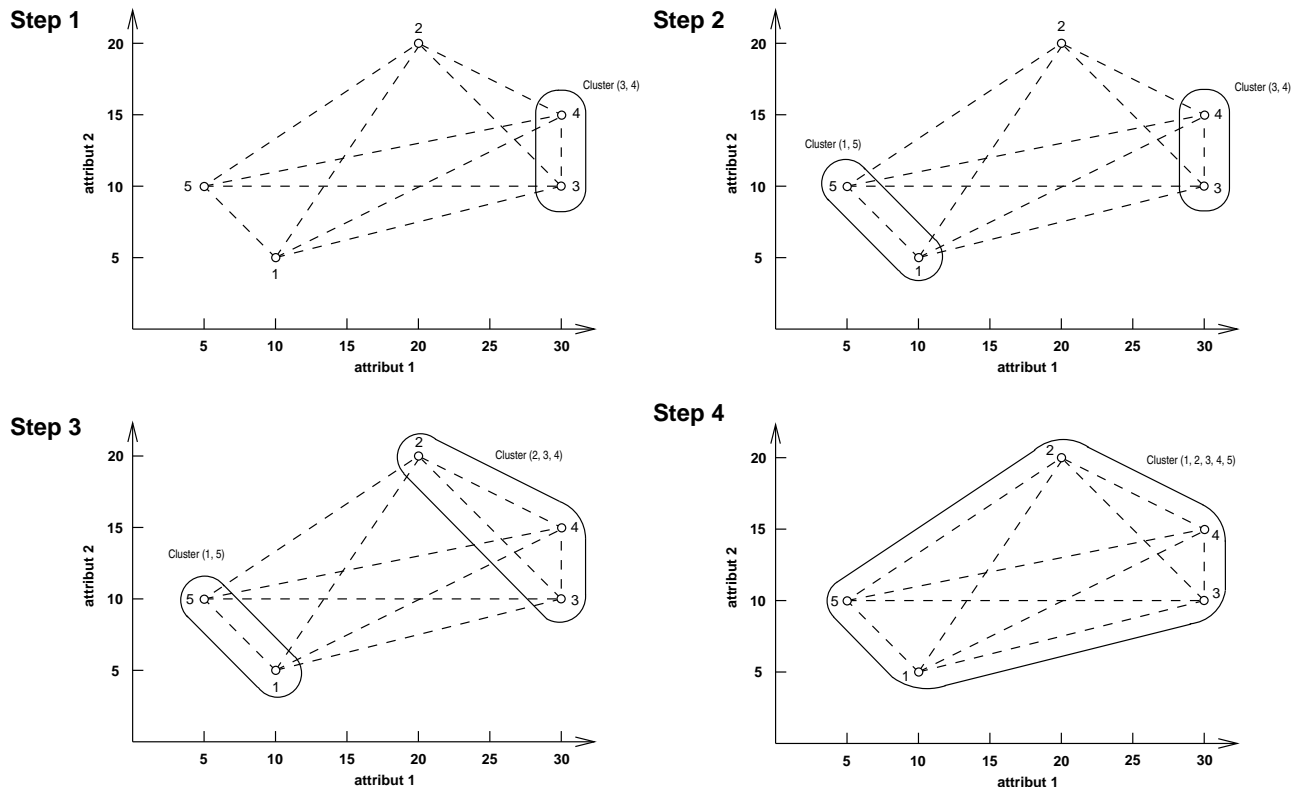
Figure 3.2: Cluster Analysis example: Clustering steps

Algorithm: Basic agglomerative hierarchical clustering algorithm	
1.	Compute similarity matrix, if necessary.
2.	**repeat**
3.	Merge the closest two clusters.
4.	Update the similarity matrix to reflect the similarity between the new cluster and the previous clusters.
5.	**until** Only one cluster remains.

Figure 3.3: Basic cluster analysis algorithm

Dendrograms

The results of a clustering process can be used to generate a map of sorts, called tree or dendrogram, to show the degrees of similarity between all pairs of objects. The x-axis shows the objects; the y-axis shows the similarity coefficient at which the previous clusters had been merged into a new cluster. The dendrogram for the example can be found in figure 3.4.

In order to actually classify all objects into clusters the dendrogram needs to be subdivided by 'cutting' at a reasonable similarity level. This step depicts the only subjective (and thus analyst-dependent) part of a cluster analysis procedure. A reasonable position for a 'tree cut' is given by a large 'gap' in the dendrogram indicating the clustering of two previously significantly distinct clusters. Thus, instead of a software based decision a visual inspection of the dendrogram is highly recommended.
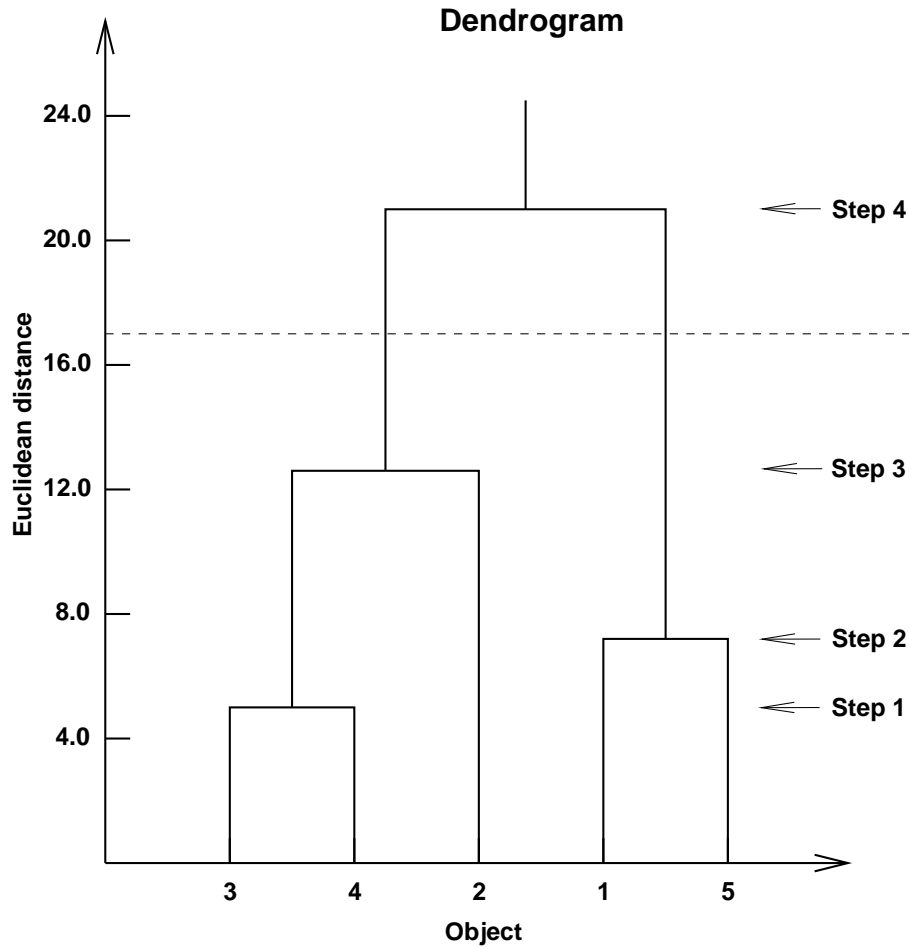
Figure 3.4: Cluster Analysis example: Dendrogram

Influential subsets and counter-acting observations

As explained in HOAGLIN and WELSCH 1978 or GRAY and LING 1984 the generalisation of high-leverage observations is a group of high-leverage observations or *influential subsets* of observations. According to GRAY and LING 1984 'interest is focused on the effects of jointly influential cases (i.e. observations), particularly those subsets whose individual cases interact to produce a high influence that is not accounted for by the main effects of their single cases (observations)'. Figure 3.5 on the next page gives a graphical visualisation of jointly influential subsets for a regression line.

In this thesis, a further distinction of jointly influential subsets is used:

- *Jointly influential (or reinforcing) observations* are those observations which affect the regression results in a similar way. Numerically these observations are indicated by large positive impact co-factors h_{ij}. Geometrically, these observations are performed under similar conditions. For a regression line (see figure 3.5), influential observations have similar x values, which are significantly distinct from the center of mass of the remaining observations.

- *Counter-acting observations* are those observations which have been performed under similar but opposite conditions. Numerically, this is shown by large negative impact co-factors h_{ij}. Geometrically, (for the regression line in figure 3.5) both observations are located on opposite sites, but with similar distances to the center of mass of the remaining observations.

Practical examples for both types of observations will be given in chapter 5.

Figure 3.5: Configurations of jointly influential and counter-acting observations. Case A: large positive h_{ij}, case B: large negative h_{ij}, case C: submatrix of $\mathbf{H}$ corresponding to $\{i, j, k, l, m\}$ contains several large positive and negative elements (GRAY and LING 1984).

3.3 Cluster analysis for parameter estimation problems

Due to the fact that the impact factors and impact co-factors of the observations can be interpreted geometrically using cosines of the angles between vectors of the natural basis and their projections onto the column space of the design matrix $\mathbf{A}$ they can be regarded as similarity measures. In contrast to most statistical cluster analysis applications (which perform clustering steps by grouping observations or objects with smallest distance, i.e., using dissimilarity coefficients), here observations with large similarity coefficients (impact co-factors) are clustered.

Hence, the main idea of using cluster analysis for parameter estimation problems is to replace the resemblance matrix by the data resolution matrix and to interpret off-diagonal elements of the data resolution matrix (impact co-factors) as similarity coefficients (GRAY and LING 1984) (see table 3.1). Based on this approach, *candidates* for influential subsets of observations can be found.

Parameter reduction

In order to determine the impact of each cluster of observations on the estimated parameters the concept of projections onto subspaces of the data space is used (see e.g. TEUNISSEN 2003). This approach is also known as reduction of parameters by estimating only a subset of the original parameters without changing the original functional model. The original linear system $\mathbf{Ax} = \mathbf{y}$ is partitioned into

$$[\mathbf{A}_1 \vdots \mathbf{A}_2] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \tag{3.1}$$

with $\mathbf{x}_1$ being a u_1-vector containing parameters of interest and $\mathbf{x}_2$ being the u_2-vector describing the parameters to be reduced. As shown in e.g. TEUNISSEN 2003 the design matrix $\bar{\mathbf{A}}_1$ of the *reduced system* can be computed using the orthogonal projector

$$\mathbf{P}_{\mathbf{A}_2}^{\perp} = \mathbf{I} - \mathbf{A}_2(\mathbf{A}'_2\mathbf{A}_2)^{-1}\mathbf{A}'_2 \tag{3.2}$$

and

$$\bar{\mathbf{A}}_1 = \mathbf{P}_{\mathbf{A}_2}^{\perp}\mathbf{A}_1. \tag{3.3}$$

As for the original system $\mathbf{Ax} = \mathbf{y}$, a data resolution matrix $\bar{\mathbf{H}}$ for the reduced system can be derived by

$$\bar{\mathbf{H}} = \bar{\mathbf{A}}_1(\bar{\mathbf{A}}'_1\bar{\mathbf{A}}_1)^{-1}\bar{\mathbf{A}}'_1. \tag{3.4}$$

In contrast to the data resolution matrix $\mathbf{H}$ for the original system, the data resolution matrix $\bar{\mathbf{H}}$ for the reduced system indicates the impact of the observations only on the remaining parameters $\mathbf{x}_1$.

Due to the computation and inversion of normal equation matrices in equations (3.2) and (3.4) the derivation of $\bar{\mathbf{H}}$ should rather be based on the singular value decomposition of the design matrix of the partitioned system (3.1): With the singular value decomposition of $\mathbf{A}_2 = \mathbf{U}_2 \mathbf{S}_2 \mathbf{V}_2$ the projector $\mathbf{P}_{\mathbf{A}_2}^{\perp}$ can be computed via

$$\mathbf{P}_{\mathbf{A}_2}^{\perp} = \mathbf{I} - \mathbf{U}_{2_{u_2}} \mathbf{U}'_{2_{u_2}} \tag{3.5}$$

with $\mathbf{U}_{2_{u_2}}$ indicating the first u_2 columns of $\mathbf{U}_2$. Since $\bar{\mathbf{H}}$ is the projector onto the subspace formed by the columns of $\bar{\mathbf{A}}_1 = \mathbf{P}_{\mathbf{A}_2}^{\perp} \mathbf{A}_1$ it can also be derived by using the singular value decomposition of $\bar{\mathbf{A}}_1 = \bar{\mathbf{U}}_1 \bar{\mathbf{S}}_1 \bar{\mathbf{V}}_1$:

$$\bar{\mathbf{H}} = \bar{\mathbf{U}}_{1_{u_1}} \bar{\mathbf{U}}'_{1_{u_1}} \tag{3.6}$$

with $\bar{\mathbf{U}}_{1_{u_1}}$ consisting of the first u_1 columns of $\bar{\mathbf{U}}_1$.

Determination of cluster impact on parameter subsets

For the determination of the impact of a cluster of observations on single (or groups of) parameters, equation (3.3) is used to project onto one-dimensional (or multi-dimensional) subspaces of the data space (by effectively reducing $u - v$ parameters and leaving v parameters of the original functional model). Although in the following investigations $v = 1$, this concept can also be used for groups of parameters (i.e., for $2 \leq v \leq n - 1$).
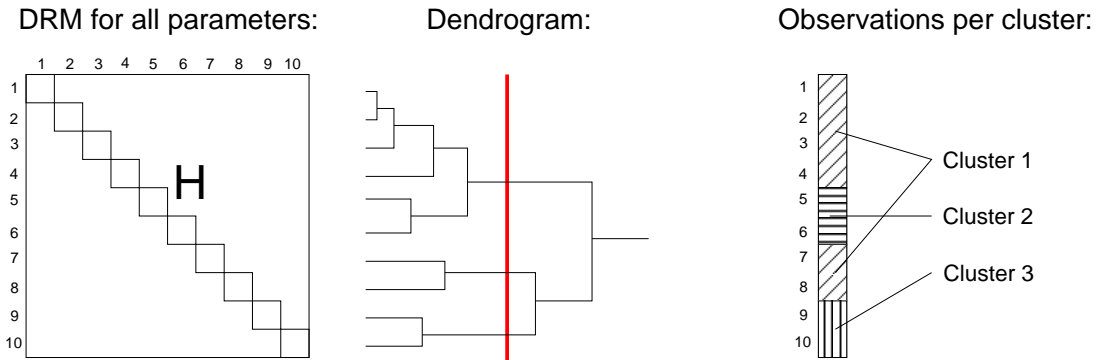
After performing the cluster analysis of the data resolution matrix $\mathbf{H}$ of the original system and after the determination of observation clusters by cutting the dendrogram at a reasonable height, the original system is gradually reduced (i.e., $v = 1$). Using the data resolution matrix $\bar{\mathbf{H}}$ of the reduced system, for each cluster the impact factors $\bar{h}_{ii}$ for those observations which belong to the current cluster are averaged by summing up the impact factors $\bar{h}_{ii}$ cluster-wise and dividing the sum by the number of elements (observations) in the respective cluster. This yields an 'average cluster impact factor $\bar{h}_{Cluster\ i}$ of cluster i on the parameter (group) $\mathbf{x}_1$' which is independent of the size (number of members) of the current cluster. Figure 3.6 provides a graphical visualisation of these steps.

3.4   Interpretation of Cluster Analysis results

The size (i.e., the number of observations in a cluster) and the 'average cluster impact factor $\bar{h}_{Cluster\ i}$ of cluster i on the parameter (group) $\mathbf{x}_1$' can be used to interpret the results of a cluster analysis and to formulate recommendations whether the cluster size should be enlarged or reduced. As shown in e.g. HOAGLIN and WELSCH 1978 the average impact of a single observation is u/n. Similarly, a group of observations (or a cluster) can be considered as of 'medium' (or average) importance if its average cluster impact factor $\bar{h}_{Cluster\ i}$ is close to u/n (e.g. $\pm$ 50% of $\bar{h}$). In the same way, further classifications can be performed by comparing the average cluster impact factor $\bar{h}_{Cluster\ i}$ with u/n.

The cluster size can be expressed as percentage of the number of observations $n_{Cluster}$ in the current cluster with respect to the total number of observations n. Table 3.2 shows an interpretation scheme which contains the cluster size as the first criterion (upper row) and the impact factor relative to the mean impact factor (i.e., u/n) as the second criterion (left column).

1. Cluster Analysis (generation of observation groups):

DRM for all parameters: Dendrogram: Observations per cluster:

2. Determination of cluster impact:

Parameter 1:   Parameter 2:   ···   Parameter u:

$\bar{h}_{Cluster\ 1,}$ $\bar{h}_{Cluster\ 2,}$ $\bar{h}_{Cluster\ 3,}$ $\bar{h}_{Cluster\ 1,}$ $\bar{h}_{Cluster\ 2,}$ $\bar{h}_{Cluster\ 3,}$ ··· $\bar{h}_{Cluster\ 1,}$ $\bar{h}_{Cluster\ 2,}$ $\bar{h}_{Cluster\ 3,}$
Parameter 1 Parameter 1 Parameter 1   Parameter 2 Parameter 2 Parameter 2   Parameter u Parameter u Parameter u
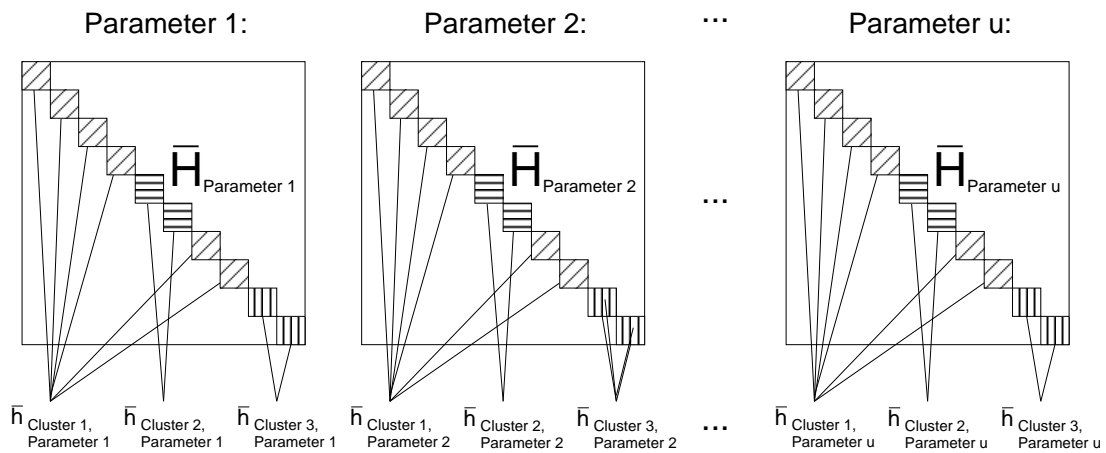
Figure 3.6: Flowchart for determination of cluster impact $\bar{h}_{Cluster\ i}$ of cluster i on parameter subset $\mathbf{x}_1$

Cluster impact factor \ Cluster Size	$n_{Cluster} = n$	Large 99% – 66%	Medium 65% – 33%	Small 32% – 1%	$n_{Cluster} = 1$
Very important ($\bar{h} > 200\% \frac{u}{n}$)	all	— — —	O	+++	++
Important ($\bar{h} > 150\% \frac{u}{n}$)	observations	O	+	++	+
Medium ($50\% \frac{u}{n} < \bar{h} < 150\% \frac{u}{n}$)	clustered	–	O	+	O
Low importance ($\bar{h} < 50\% \frac{u}{n}$)	into one	— —	–	O	–
Unimportant ($\bar{h} < 1\% \frac{u}{n}$)	cluster	— — —	— —	–	Observation negligible

Table 3.2: Rules for interpretation of Cluster Analysis results: ++ indicates that cluster size must be significantly enlarged (i.e., that observations should be controlled by appropriate (independent) observations), − − − and −− denote significant decrease, + and − indicate only minor changes and for O the cluster size is appropriate.

As shown in the first row of table 3.2 the size of a very important cluster can be slightly reduced if it contains a large number of observations. On the other hand, a small but very important cluster should be enlarged, i.e., that the small number of observations in this cluster should be controlled by appropriate (independent) observations. If all observations have been clustered into one cluster ($n_{Cluster} = n$), no reasonable conclusion can be drawn (see first column). On the other hand, if the dendrogram cut has been performed before the first clustering step, every cluster contains only one observation ($n_{Cluster} = 1$). The last column thus represents a solely interpretation of the impact factors of each observation (with the first case indicating an extreme high-leverage observation which should be controlled/supported by several independent observations).

In general, a cluster with a large impact factor and with only a small number of members should be significantly enlarged ('++') while the size of a large cluster with a low impact factor must be significantly reduced ('−−'). The symbols ('+') and ('−') indicate recommendations for only minor size variations. Some cases (e.g. all observations clustered into one cluster or each observation is a single cluster) need special treatment. The percentage values in table 3.2 have been derived from experience and thus only indicate interpretation guidelines. The actual decision for size variations depends on the purpose of the experiment and also depends on the costs for performing other observations.

3.5   Regression diagnostics tool-flowchart

The singular value decomposition of the design matrix of an adjustment problem (and hence the data resolution matrix derived from the singular value decomposition) and cluster analysis methods form the two main components of the *regression diagnostics tool* developed in this thesis. As mentioned in the introduction, in geodetic literature the term 'regression' is often used for the determination of the parameters of adjusting straight lines or other polynomials. Here, however, the term 'regression' describes the general procedure of parameter estimation in linear models (see also e.g. COOK and WEISBERG 1982, or DODGE and JURECKOVÁ 2000).

Based on the interpretation guidelines described above, the general procedure of a singular value decomposition- and cluster analysis-based *regression diagnostics tool* can be formulated. As shown in figure 3.7 on the facing page the general procedure starts with the definition of the adjustment problem to be solved. This also contains a reasonable parametrisation of the functional model relating observations and unknown parameters. After performing suitable measurements, the design matrix $\mathbf{A}$ can be set up and can be decomposed by singular value decomposition.

Since the results of the entire procedure should only be interpreted in case of a full rank adjustment problem, the singular values σ_i (or the condition number) have to be used for the detection of rank deficiencies or a weak condition. Optionally, the correlations and thus the separability of the parameters should be analysed by using the correlation matrix (COR) or the model resolution matrix (MRM). If there is an (almost) rank deficiency it is recommended to re-formulate the adjustment problem before proceeding with the regression diagnostics procedure.

If all parameters can be estimated (separately), the data resolution matrix $\mathbf{H}$ can be computed. At first, its main diagonal elements (i.e., the impact factors h_{ii}) should be checked for the existence of high-leverage observations. Depending on the purpose of the adjustment and depending on the consequences of high-leverage observations, appropriate steps might be necessary (such as elimination of observations or addition of new observations).

The next step consists of the analysis of the impact co-factors h_{ij} and the generation of groups of observations by cluster analysis of the data resolution matrix $\mathbf{H}$. After cutting the dendrogram at a reasonable height the 'average cluster impact factors $\bar{h}_{Cluster\ i}$' can be computed.

For each parameter (usually starting with the most important parameter) the impact of each cluster can be assessed using the interpretation guidelines listed in table 3.2 on the previous page. Depending on the need of performing changes of the observation structure, the entire process needs to be repeated (starting from the generation of the design matrix $\mathbf{A}$). Otherwise the procedure is completed.

Regression diagnostics tool – flowchart

Definition of adjustment problem
– choice of functional model
– definition of parameters of interest

Generation / Measurement of (appropriate) observations

Set up of design matrix A

Computation of Singular Value Decomposition of A

Investigation of singular values σ :
– check for rank deficiency
– check of condition number
– ...

(Optional) investigation of correlations between parameters (COR and MRM)

All parameters estimable and separable? — No

Yes

Computation of Data Resolution Matrix H

Analysis of impact factors h_{ii}:
– existence of high–leverage observations $h_{ii} > 2\frac{u}{n}$?

Analysis of impact co–factors h_{ij}:
– **Cluster Analysis** (and generation of dendrogram)
– **Dendrogram cut**
– **Determination of**
"average cluster impact factor $\bar{h}_{\text{Cluster i}}$ of cluster i on parameter x"

Loop over all parameters (from most to least important parameter):
– Comparison of cluster impact factor and cluster size
(see interpretation guidelines)

| Cluster size small compared to cluster importance? | Cluster of medium size and of medium importance? | Cluster size large compared to cluster importance? |

Add independent observations | **No changes necessary** | **Remove some observations**

Repeat analysis | | **Repeat analysis**
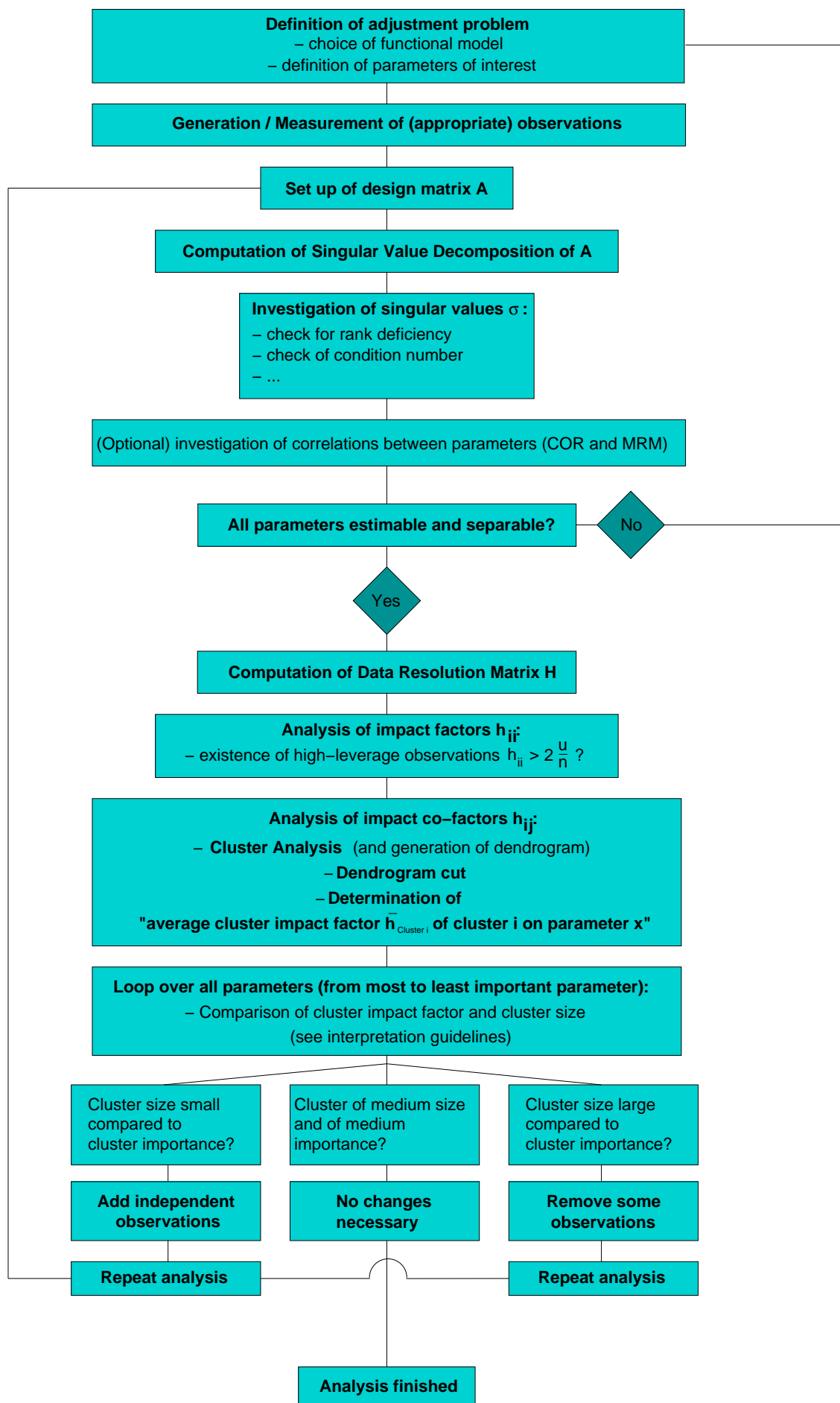
Analysis finished

Figure 3.7: Regression diagnostics flowchart

4. Geodetic Very Long Baseline Interferometry

4.1 Introduction

The principle of Very Long Baseline Interferometry (VLBI) has been developed in the 1970ies and was at first mainly used for the investigation of astronomical and astrophysical phenomena (see e.g. COHEN and SHAFFER 1971). This principle is based on a classical interferometer in the visible spectrum which has been invented as early as 1890 by Michelson (MICHELSON 1890). While the two receiving devices of a classical radio interferometer are connected, this is not the case for a long baseline interferometer. Here, the distances between the receivers can be up to 12.000 km (see figure 4.1). At both stations, the signals of an extra-galactic radio source are received and provided with time marks generated by highly precise atomic clocks (usually hydrogen-masers) before they are stored digitally on tapes or discs (CAMPBELL 2004).
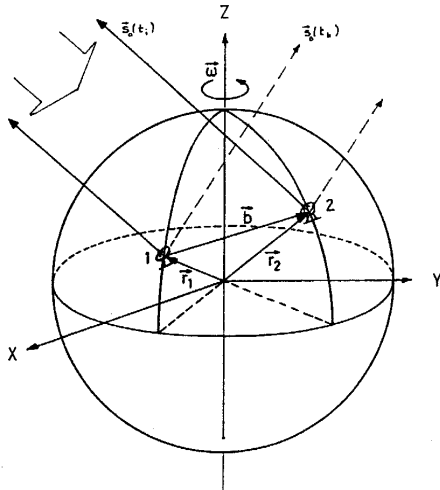

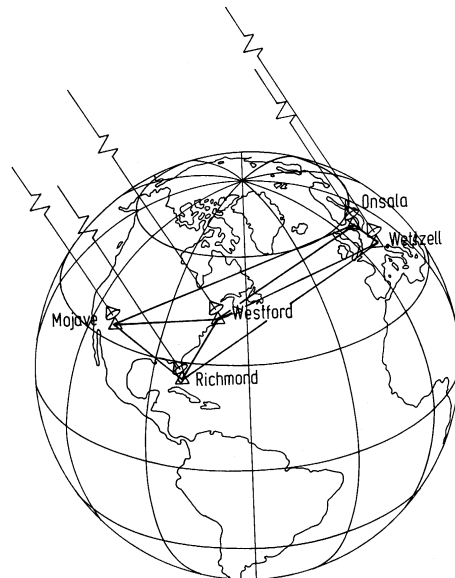Figure 4.1: Single-baseline-interferometer (CAMPBELL 2004)

Figure 4.2: VLBI network with five stations (CAMPBELL 2004)

The data is sent to specially designed computers (correlators) and brought to coherency. Within the correlation process the difference τ of the arrival times of the signal at both stations is determined and represents the primary geodetic observable. It is often simply called 'delay' and can nowadays be determined with an accuracy of approximately 20-30 picoseconds (= 6-10 [mm]) (see e.g. SOVERS et al. 1998).

Soon after the first use for astronomical purposes the geodetic use of the VLBI-principle was recognized (e.g. SHAPIRO 1974, MA 1978, or CAMPBELL and WITTE 1978). In addition, besides the baseline vector other parameters such as e.g. the rotation of the earth (i.e., polar motion x_p, y_p and $\Delta UT1$ as well as nutation $d\psi, d\epsilon$), atmospheric behaviour, tidal effects, etc. can be determined and are included in the functional model.

By performing common observations of the same radio source by different stations, global observation networks can be formed. These networks can be used to connect regional geodetic reference systems (and can therefore be used for the generation of global reference systems) as well as for a more precise determination of earth orientation parameters (compared to observations on single baselines) (see figure 4.2).

Compared to other space geodetic techniques (such as GPS, SLR/LLR or DORIS) VLBI has the advantage of having a direct connection to the quasi-inertial system of the radio sources which enables analysts to determine earth orientation parameters with a long time stability and free of any hypothesis. Hence VLBI is the only technique (except for optical astronomical techniques) which connects the sky-fixed reference system (CRF) directly to the earth-fixed system (TRF) via the earth orientation parameters (EOP).

The basic principle of VLBI has been described by many authors. For more details see e.g. CAMPBELL 1979, SCHUH 1987, NOTHNAGEL 1991, or TAKAHASHI ET AL. 2000.

4.2 Basic models in VLBI data analysis

4.2.1 The functional model of VLBI

The *geometrical time delay* of a non-rotating plane or spatial radio interferometer whose two stations are connected by the baseline vector $\mathbf{b} = \mathbf{r_2} - \mathbf{r_1}$ (with $\mathbf{r}_i$ being the geocentric vectors of the observation sites, respectively) can be mathematically described by the scalar product

$$\tau_{geom} = \tau_2 - \tau_1 = -\frac{1}{c} \cdot \mathbf{b} \cdot \mathbf{k}, \tag{4.1}$$

where c denotes the velocity of the radio signal (i.e., the velocity of light), $\mathbf{k}$ denotes a unit vector in the direction of the radio source and τ_1 and τ_2 denote the arrival times of the radio signal at the two stations respectively (see e.g. NOTHNAGEL 1991).

Further generalisation of model (4.1) leads to a spatial, kinematic interferometer. Since the baseline vector $\mathbf{b}$ is defined in an earth-fixed reference system while the vector $\mathbf{k}$ in direction of the radio source is defined in a sky-bound reference system, one of these reference systems has to be transformed into the other one. For a better physical interpretation, the three rotations necessary for this transformation are usually decomposed into five individual rotations which are represented by four rotation matrices $\mathbf{W}, \mathbf{S}, \mathbf{N}$ and $\mathbf{P}$. These matrices correspond to polar motion (wobble, x_p and y_p), earth rotation (spin, $dUT1$), nutation ($d\psi, d\epsilon$) and precession (z, ξ_A, Θ_A) respectively (see e.g. ROBERTSON 1975, MA 1978, NOTHNAGEL 1991, or SEEBER 2003). Equation (4.1) thus becomes:

$$\begin{aligned} \tau_{geom} &= -\frac{1}{c} \cdot \mathbf{b} \cdot \mathbf{R} \cdot \mathbf{k} \\ &= -\frac{1}{c} \cdot \mathbf{b} \cdot \mathbf{W} \cdot \mathbf{S} \cdot \mathbf{N} \cdot \mathbf{P} \cdot \mathbf{k} \end{aligned} \tag{4.2}$$

The matrices $\mathbf{W}, \mathbf{S}, \mathbf{N}$ and $\mathbf{P}$ are usually expressed by means of Eulerian rotation angles around the respective rotation axes. A more detailed description can be found in e.g. NOTHNAGEL 1991, or SOVERS et al. 1998. NOTHNAGEL 1991 also lists the coordinate systems associated with the different rotations.

Since equation (4.2) only describes the *geometrical* delay, a more sophisticated model has to be used to model *real* VLBI observations which are affected by various effects on their way through interstellar space, the Solar System, and the Earth's atmosphere. Therefore further terms accounting for e.g. the changing behaviour of station clocks, the delay caused by atmospheric influences, tidal or loading effects, etc. have to be added. Hence the basic geometrical has to be extended to

$$\begin{aligned} \tau_{obs} &= -\frac{1}{c} \cdot \mathbf{b} \cdot \mathbf{W} \cdot \mathbf{S} \cdot \mathbf{N} \cdot \mathbf{P} \cdot \mathbf{k} \\ &\quad + \tau_{j-abb.} + \tau_{t-abb.} + \tau_{Rel.} + \tau_{Tid.} + \tau_{Load.} \\ &\quad + \tau_{Ion.} + \tau_{Instr.} + \tau_{Atm_h} + \tau_{Clock} + \tau_{Atm_w} \\ &\quad + \dots \end{aligned} \tag{4.3}$$

where the following terms are used:

$\tau_{j-abb.}$	annual abberation because of the motion of the earth around the solar system barycenter
$\tau_{t-abb.}$	diurnal abberation because of the rotation of the earth
$\tau_{Rel.}$	relativistic effects
$\tau_{Tid.}$	deformation of the earth because of tides and because of changes of the angular momentum due to ocean tides
$\tau_{Load.}$	deformation of the earth because of loading effects e.g. due to ocean tides and atmospheric pressure changes
$\tau_{Ion.}$	ionospheric correction
$\tau_{Instr.}$	instrumental corrections
τ_{Atm_h}	atmospheric refraction (hydrostatic part)
τ_{Atm_w}	atmospheric refraction (wet part)
τ_{Clock}	relative clock offset and clock rate (and additional terms)

A more explicit formulation of model (4.3) reads:

$$
\begin{aligned}
\tau_{obs} \;=\; & -\frac{1}{c} \cdot \begin{pmatrix} x_A - x_B \\ y_A - y_B \\ z_A - z_B \end{pmatrix} \cdot \mathbf{R}(x_p, y_p, dUT1, d\psi, d\epsilon, z, \xi_A, \Theta_A) \cdot \begin{pmatrix} \cos\delta \cdot \cos h(t) \\ \cos\delta \cdot \sin h(t) \\ \sin\delta \end{pmatrix} \\
& + \tau_{j-abb.} + \tau_{t-abb.} + \tau_{Rel.} + \tau_{Tid.} + \tau_{Load.} \\
& + \tau_{Ion.} + \tau_{Instr.} + \tau_{Atm_h} + \tau_{Clock} + \tau_{Atm_w} \\
& + \ldots
\end{aligned}
\tag{4.4}
$$

with x_i, y_i, z_i being the geocentric coordinates of the particular station, $\mathbf{R}$ being the rotation matrix between the celestial and the terrestrial reference system (and thus dependent on polar motion x_p, y_p, earth rotation $\Delta UT1$, nutation $d\psi, d\epsilon$ and precession z, ξ_A, Θ_A) and $h(t)$ and δ being the hour angle and declination of the observed radio source.

The way of parametrisation usually depends on the target parameters to be investigated and depends on the number of stations participating as well as on the duration of the session. As described in chapter 2 also in VLBI data analysis different mathematical models (possibly with different kinds of parametrisations) may describe the observations equally well. Although official recommendations exist (MCCARTY and PETIT 2003), the choice of a particular model and the choice of a particular parametrisation is quite arbitrary and may vary from analyst to analyst. In particular the following physical models are subject to these arbitrary choices.

Physical models

Clock behaviour and piecewise linear modelling

One of the largest constituents of the signal delays is caused by the differences in the behaviour of the station frequency standards. After choosing an arbitrary clock as the reference clock for the entire observing network the remaining clocks show both a constant difference (= clock offset) and a linear (= clock trend) or an even higher rate of change relative to the reference clock. Thus, an appropriate clock with a presumedly high frequency stability should be chosen as the reference standard for the entire network.

From an algebraic point of view it is of no consequence to the least-squares solution which station clock is chosen as the reference one (SCHUH 1987). Since the clock parameters (offset, trend, etc.) also 'absorb' physical effects with a similar signature (as e.g. instrumental effects and relativistic effects of higher order) special attention should be paid to this type of parameter.

In order to describe the clock behaviour in a mathematical way, usually a simple polynomial approach is chosen, as e.g.

$$\tau_{Clock} = CL_0 + CL_1 \cdot t + CL_2 \cdot (t - t_0)^2 + \dots. \tag{4.5}$$

In practical VLBI data analysis usually up to three parameters (offset CL_0, trend CL_1 and squared term CL_2) are used. In addition to a simple polynomial further so-called *piece-wise linear parameters* are used to account for higher variations of the frequency standards. For piece-wise linear modelling a linear behaviour of the effect to be modelled is assumed for certain intervals. One of the different kinds of parametrisation is based on determining new clock rates for each interval by estimating $\Delta\tau_{Clockrate_i}$ (see e.g. SCHUH 1987, or TESMER 2004):

$$\Delta\tau_{Clock}(t_i) = \Delta\tau_{Clockoffset} + \Delta\tau_{Clockrate_1}(t_1 - t_0) + \Delta\tau_{Clockrate_2}(t_2 - t_1) + \dots + \Delta\tau_{Clockrate_n}(t_i - t_{n-1}) \tag{4.6}$$

In order to avoid numerical problems (as e.g. rank deficiencies) and to stabilize the parameter estimation process, constraints (or pseudo-observations) have to be included in intervals with only a small number of observations. Usually this type of pseudo-observations constrains the particular rate segment to zero and allows for a certain variation by assigning an appropriate formal error.

Piece-wise linear modelling is also used when describing other effects such as e.g. atmospheric behaviour, atmosphere gradients or sub-daily earth rotation variations.

Atmospheric refraction and atmospheric mapping functions

On their way to the radio telescopes, radio signals have to pass the atmosphere, i.e., the electrically charged part (ionosphere) and the electrically neutral part (troposphere) of the atmospheric layers up to a height of approximately 50 kilometres (HOFMANN-WELLENHOF, B. et al. 2003). Depending on the state of these layers the signals are distorted. The impact of the ionosphere can be eliminated almost completely by performing dual frequency measurements. The signal path delay caused by the atmosphere (as well as the ionospheric delay) is called *refraction* and can change between approximately 2.3 m in zenith direction (approx. 8 ns) and almost 25 m at elevations of 5° (NOTHNAGEL 2000).

Since the tropospheric signal path delay depends on the path length of the signals through the atmosphere it therefore also depends on the elevation ϵ of the radio source observed. These elevation dependencies are usually expressed as so-called *mapping functions* which describe the relations of the signal path delay in zenith direction ('zenith delay') $\delta\rho_{trp}^0$ and the signal path delay in source direction $\delta\rho_{trp}(\epsilon)$. The most simple form of such a mapping function reads:

$$m_{trp}(\epsilon) = \frac{1}{\sin(\epsilon)}. \tag{4.7}$$

Hence the signal path delay in source direction is

$$\delta\rho_{trp}(\epsilon) = m_{trp}(\epsilon) \cdot \delta\rho_{trp}^0. \tag{4.8}$$

Due to the fact that water vapor only occurs up to heights of approximately 15 km (while the entire atmosphere reaches heights of more than 50 km) it has been found appropriate to use different mapping functions for the hydrostatic and the wet part of the atmosphere. Hence the total atmospheric signal path delay in source direction can be modelled as:

$$\delta\rho_{trp}(\epsilon) = m_{trp,d}(\epsilon) \cdot \delta\rho_{trp,d}^0 + m_{trp,w}(\epsilon) \cdot \delta\rho_{trp,w}^0. \tag{4.9}$$

The simple mapping function $m_{trp}(\epsilon) = \dfrac{1}{\sin(\epsilon)}$ has been found to be inaccurate for highly precise applications (even at elevation angles of 20°). During the last decades more precise mapping functions have been

developed (as e.g. CfA-, MTT-, Neill- or the Vienna Mapping functions) some of them taking into account the atmospheric situation above the observation site or making use of numerical weather models. Details about mapping functions can be found in e.g. NOTHNAGEL 2000, or BOEHM 2004.

As for the modelling of clock behaviour also piecewise linear modelling approaches are used to describe the atmospheric behaviour. Depending on the analyst the interval length for piecewise linear modelling of the atmosphere is usually set to values between 2 hours and 20 minutes.

4.2.2   Partial derivatives and design matrix

For the adjustment of VLBI observations by using the Gauss-Markoff-model (see chapter 2) the partial derivatives of the particular functional model chosen by the analyst with respect to the unknown parameters have to be computed and arranged in the design matrix. In general, the design matrix contains many more parameters than discussed in the following paragraph (see e.g. NOTHNAGEL 1991).

The dimension of the design matrix equals *(number of observed delays + number of constraints) × (number of unknowns)*. When choosing the order of unknowns as *station coordinates, clock coefficients, atmosphere coefficients* the structure of the design matrix may be as follows (example for a 3-station network):

$$
A = \begin{vmatrix}
Stations_{kl} & Clocks_{kl} & Atmosphere_{kl} \\
Stations_{kl} & Clocks_{kl} & Atmosphere_{kl} \\
Stations_{kl} & Clocks_{kl} & Atmosphere_{kl} \\
Stations_{kl} & Clocks_{kl} & Atmosphere_{kl} \\
\dots & \dots & \dots
\end{vmatrix}
\tag{4.10}
$$

with

$$
\text{Stations} = \begin{vmatrix}
\frac{\delta\tau_i}{\delta x_a} & \frac{\delta\tau_i}{\delta y_a} & \frac{\delta\tau_i}{\delta z_a} & \frac{\delta\tau_i}{\delta x_b} & \frac{\delta\tau_i}{\delta y_b} & \frac{\delta\tau_i}{\delta z_b} & 0 & 0 & 0 \\
\frac{\delta\tau_j}{\delta x_a} & \frac{\delta\tau_j}{\delta y_a} & \frac{\delta\tau_j}{\delta z_a} & 0 & 0 & 0 & \frac{\delta\tau_j}{\delta x_c} & \frac{\delta\tau_j}{\delta y_c} & \frac{\delta\tau_j}{\delta z_c} \\
0 & 0 & 0 & \frac{\delta\tau_k}{\delta x_b} & \frac{\delta\tau_k}{\delta y_b} & \frac{\delta\tau_k}{\delta z_b} & \frac{\delta\tau_k}{\delta x_c} & \frac{\delta\tau_k}{\delta y_c} & \frac{\delta\tau_k}{\delta z_c} \\
\frac{\delta\tau_l}{\delta x_a} & \frac{\delta\tau_l}{\delta y_a} & \frac{\delta\tau_l}{\delta z_a} & \frac{\delta\tau_l}{\delta x_b} & \frac{\delta\tau_l}{\delta y_b} & \frac{\delta\tau_l}{\delta z_b} & 0 & 0 & 0 \\
\frac{\delta\tau_m}{\delta x_a} & \frac{\delta\tau_m}{\delta y_a} & \frac{\delta\tau_m}{\delta z_a} & 0 & 0 & 0 & \frac{\delta\tau_m}{\delta x_c} & \frac{\delta\tau_m}{\delta y_c} & \frac{\delta\tau_m}{\delta z_c} \\
0 & 0 & 0 & \frac{\delta\tau_n}{\delta x_b} & \frac{\delta\tau_n}{\delta y_b} & \frac{\delta\tau_n}{\delta z_b} & \frac{\delta\tau_n}{\delta x_c} & \frac{\delta\tau_n}{\delta y_c} & \frac{\delta\tau_n}{\delta z_c} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots
\end{vmatrix}
$$

$$
\text{Clocks} = \begin{vmatrix}
\frac{\delta\tau_i}{\delta CL0_a} & \frac{\delta\tau_i}{\delta CL1_a} & \frac{\delta\tau_i}{\delta CL2_a} & \frac{\delta\tau_i}{\delta CL0_b} & \frac{\delta\tau_i}{\delta CL1_b} & \frac{\delta\tau_i}{\delta CL2_b} & 0 & 0 & 0 \\
\frac{\delta\tau_j}{\delta CL0_a} & \frac{\delta\tau_j}{\delta CL1_a} & \frac{\delta\tau_j}{\delta CL2_a} & 0 & 0 & 0 & \frac{\delta\tau_j}{\delta CL0_c} & \frac{\delta\tau_j}{\delta CL1_c} & \frac{\delta\tau_j}{\delta CL2_c} \\
0 & 0 & 0 & \frac{\delta\tau_k}{\delta CL0_b} & \frac{\delta\tau_k}{\delta CL1_b} & \frac{\delta\tau_k}{\delta CL2_b} & \frac{\delta\tau_k}{\delta CL0_c} & \frac{\delta\tau_k}{\delta CL1_c} & \frac{\delta\tau_k}{\delta CL2_c} \\
\frac{\delta\tau_l}{\delta CL0_a} & \frac{\delta\tau_l}{\delta CL1_a} & \frac{\delta\tau_l}{\delta CL2_a} & \frac{\delta\tau_l}{\delta CL0_b} & \frac{\delta\tau_l}{\delta CL1_b} & \frac{\delta\tau_l}{\delta CL2_b} & 0 & 0 & 0 \\
\frac{\delta\tau_m}{\delta CL0_a} & \frac{\delta\tau_m}{\delta CL1_a} & \frac{\delta\tau_m}{\delta CL2_a} & 0 & 0 & 0 & \frac{\delta\tau_m}{\delta CL0_c} & \frac{\delta\tau_m}{\delta CL1_c} & \frac{\delta\tau_m}{\delta CL2_c} \\
0 & 0 & 0 & \frac{\delta\tau_n}{\delta CL0_b} & \frac{\delta\tau_n}{\delta CL1_b} & \frac{\delta\tau_n}{\delta CL2_b} & \frac{\delta\tau_n}{\delta CL0_c} & \frac{\delta\tau_n}{\delta CL1_c} & \frac{\delta\tau_n}{\delta CL2_c} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots
\end{vmatrix}
$$

$$
\text{Atmosphere parameters} = \begin{vmatrix}
\frac{\delta\tau_i}{\delta AT_a} & \frac{\delta\tau_i}{\delta AT_b} & 0 \\
\frac{\delta\tau_j}{\delta AT_a} & 0 & \frac{\delta\tau_j}{\delta AT_c} \\
0 & \frac{\delta\tau_k}{\delta AT_b} & \frac{\delta\tau_k}{\delta AT_c} \\
\frac{\delta\tau_l}{\delta AT_a} & \frac{\delta\tau_l}{\delta AT_b} & 0 \\
\frac{\delta\tau_m}{\delta AT_a} & 0 & \frac{\delta\tau_m}{\delta AT_c} \\
0 & \frac{\delta\tau_n}{\delta AT_b} & \frac{\delta\tau_n}{\delta AT_c} \\
\dots & \dots & \dots
\end{vmatrix}
$$

Some of the partial derivatives of the observation equation (4.3) with respect to the unknown parameters read (NOTHNAGEL 1991):

$$\frac{\partial \tau_{obs}}{\partial X_a} = \frac{1}{c} \cdot \cos(h(t)) \cdot \cos(\delta) \tag{4.11}$$

$$\frac{\partial \tau_{obs}}{\partial Y_a} = \frac{1}{c} \cdot \sin(h(t)) \cdot \cos(\delta) \tag{4.12}$$

$$\frac{\partial \tau_{obs}}{\partial Z_a} = \frac{1}{c} \cdot \sin(\delta) \tag{4.13}$$

$$\frac{\partial \tau_{obs}}{\partial X_b} = -\frac{1}{c} \cdot \cos(h(t)) \cdot \cos(\delta) \tag{4.14}$$

$$\frac{\partial \tau_{obs}}{\partial Y_b} = -\frac{1}{c} \cdot \sin(h(t)) \cdot \cos(\delta) \tag{4.15}$$

$$\frac{\partial \tau_{obs}}{\partial Z_b} = -\frac{1}{c} \cdot \sin(\delta) \tag{4.16}$$

$$\frac{\partial \tau_{obs}}{\partial CL0_a} = 1 \tag{4.17}$$

$$\frac{\partial \tau_{obs}}{\partial CL1_a} = t - t_0 \tag{4.18}$$

$$\frac{\partial \tau_{obs}}{\partial CL2_a} = (t - t_0)^2 \tag{4.19}$$

$$\frac{\partial \tau_{obs}}{\partial CL0_b} = -1 \tag{4.20}$$

$$\frac{\partial \tau_{obs}}{\partial CL1_b} = -(t - t_0) \tag{4.21}$$

$$\frac{\partial \tau_{obs}}{\partial CL2_b} = -(t - t_0)^2 \tag{4.22}$$

$$\frac{\partial \tau_{obs}}{\partial AT_a} = \frac{1}{c} \cdot m(\epsilon) \tag{4.23}$$

$$\frac{\partial \tau_{obs}}{\partial AT_b} = -\frac{1}{c} \cdot m(\epsilon) \tag{4.24}$$

$$\frac{\partial \tau_{obs}}{\partial x_{p0}} = -\frac{1}{c} \cdot \big(b_x \cdot \sin(\delta) - b_z \cdot \cos(\delta) \cdot \cos(h(t))\big) \tag{4.25}$$

$$\frac{\partial \tau_{obs}}{\partial x_{p1}} = -\frac{1}{c} \cdot \big(b_x \cdot \sin(\delta) - b_z \cdot \cos(\delta) \cdot \cos(h(t))\big) \cdot (t - t_0) \tag{4.26}$$

$$\frac{\partial \tau_{obs}}{\partial y_{p0}} = -\frac{1}{c} \cdot \big(b_y \cdot \sin(\delta) - b_z \cdot \cos(\delta) \cdot \sin(h(t))\big) \tag{4.27}$$

$$\frac{\partial \tau_{obs}}{\partial y_{p1}} = -\frac{1}{c} \cdot \big(b_y \cdot \sin(\delta) - b_z \cdot \cos(\delta) \cdot \sin(h(t))\big) \cdot (t - t_0) \tag{4.28}$$

$$\frac{\partial \tau_{obs}}{\partial \Delta UT1_0} = \frac{1}{c} \cdot \Omega \cdot \cos(\delta) \cdot \big(b_x \cdot \sin(h(t)) - b_y \cdot \cos(h(t))\big) \tag{4.29}$$

$$\frac{\partial \tau_{obs}}{\partial \Delta UT1_1} = \frac{1}{c} \cdot \Omega \cdot \cos(\delta) \cdot \big(b_x \cdot \sin(h(t)) - b_y \cdot \cos(h(t))\big) \cdot (t - t_0) \tag{4.30}$$

$$\frac{\partial \tau_{obs}}{\partial \alpha} = -\frac{1}{c} \cdot \cos(\delta) \cdot \big(b_x \cdot \sin(h(t)) - b_y \cdot \cos(h(t))\big) \tag{4.31}$$

$$\frac{\partial \tau_{obs}}{\partial \delta} = -\frac{1}{c} \cdot \big(\sin(\delta) \cdot (-b_x \cdot \cos(h(t)) - b_y \cdot \sin(h(t))) + b_z \cdot \cos(\delta)\big) \tag{4.32}$$

with

$h(t)$	Greenwich hour angle of the radio source
δ	declination of the radio source
c	velocity of light
$t - t_0$	time passed since the beginning of the session
ϵ	elevation of the radio telescope
Ω	conversion factor from universal time to sidereal time (≈ 1)
$m(\epsilon)$	atmospheric mapping function.

4.2.2.1   Sensitivity of observation equations / partial derivatives

Due to the relative nature of the VLBI technique the observations (i.e., arrival time *differences*) are not sensitive to some parameters such as absolute geocentric station coordinates or the origin of right ascension (SOVERS et al. 1998). This information must be supplied by other techniques.

The information content of VLBI observables heavily depends on the configuration of the experiment. Thus, the observation schedule (i.e., station location, source position and the actual orientation of the source with respect to the baseline) plays a crucial role in determining the types and precision of parameters that can be estimated. Due to the fact that only the scalar product of baseline vector and vector in source direction is observed, some observation geometries impose limitations on the estimability (and separability) of certain parameters (so-called critical (baseline) configuration, cf. SOVERS et al. 1998, or TAKAHASHI 1994).

In order to assess the parameters that can be estimated the partial derivatives of the observation equation of a functional model with respect to its parameters have to be analysed. In geodetic adjustment computations this is in particular used for the analysis of error propagation. On the other hand these investigations also show the sensitivity of an observation concerning a certain parameter (see e.g. NIEMEIER 2002, WALTER 1973, or LUNDQVIST 1984).

In the case of VLBI, investigations of the magnitudes and the variations of the partial derivatives of equation (4.3) with respect to the parameters to be determined have been performed very early by e.g. WALTER 1973, DERMANIS and MUELLER 1978, or MA 1978. WALTER 1973 describes in detail that 'the chances to separate the various effects are widened if the partial derivatives differ in amplitude and phase'. Conversely, this means that similarities of partial derivatives indicate a low separability (or a high correlation) between the parameters involved. Algebraically this is shown by at least two similar columns of the design matrix leading to numerical problems due to (an almost) rank deficiency of the design matrix.

Partial derivatives with respect to earth orientation parameters

Investigations of the magnitudes and the variations of the partial derivatives $\partial \tau_{obs} / \partial x_{p0}$ (eq. 4.25), $\partial \tau_{obs} / \partial y_{p0}$ (eq. 4.27) and $\partial \tau_{obs} / \partial \Delta UT1_0$ (eq. 4.29) show the sensitivities of VLBI observations with respect to the observation geometry (i.e., concerning baseline orientation and source position relative to the observing baseline).

The partial derivative $\partial \tau_{obs} / \partial x_{p0}$ (eq. 4.25) shows that observations on baselines parallel to the y-axis (i.e., $b_x = b_z = 0$, with b_i indicating the respective baseline component) are not capable for the determination of the x_p-component (which is a rotation around the y-axis), while observations on baselines parallel to the x-axis (i.e., $b_y = b_z = 0$) are not suitable for the determination of the y_p-component of polar motion (which is a rotation around the x-axis), see eq. 4.27. Additionally, equations 4.25 and 4.27 show the capability of baselines with a long north-south-extension (i.e., with large b_z-components) for the determination of polar motion. The $\partial \tau_{obs} / \partial \Delta UT1_0$-partial derivative (eq. 4.29) reveals the importance of observations on baselines with a large east-west-extension for the determination of earth rotation (i.e., for $\Delta UT1_0$). This agrees with the fact that in general VLBI observations are not sensitive to an effect which is perpendicular to the orientation of that baseline (see MORITZ 1987, or FISCHER 2006).

Concerning the source position the three partial derivatives show that especially radio sources with low (absolute) declinations are necessary for the determination and separation of x_{p0}, y_{p0} and $\Delta UT1_0$. Tabular 4.1 gives an overview of criteria for optimal EOP determination depending on baseline orientation and source position (BROUWER 1985).

	Baseline parallel to Equatorial plane		Baseline parallel to rotation axis	
	Equatorial sources	Polar sources	Equatorial sources	Polar sources
x_p, y_p	NO	YES	YES	NO
$\Delta UT1$	YES	NO	NO	NO

Table 4.1: Parameter estimability as a function of baseline orientation and source position (BROUWER 1985)

Partial derivatives with respect to height, clock and atmosphere parameters

The analysis of the partial derivatives of equation (4.3) with respect to the topocentric station height, clock offset and atmosphere offset (see equations 4.17 to 4.24) leads to a very common problem in space geodetic positioning: the separability of vertical station motion, clock offset and atmospheric variations. Due to the similarities of the partial derivatives at high elevation angles the separability of the three effects decreases with higher elevations (see also VENNEBUSCH 2002).

Figure 4.3 shows the impact of variations in station height, clock offset and atmospheric behaviour on the observations (for the case of GPS-observations, but it can be used for VLBI as well). Since a physical effect can only be determined from observations which are affected by the particular effect, this also reveals the kind of observations necessary for the determination and separation of the three effects: Observations with low elevations are especially needed for the separation of station height and atmospheric behaviour. Solely observations in zenith direction are not suitable for the separation of the three effects (see ROTHACHER 2003). NOTHNAGEL 1991 demonstrated empirically that observations to sources located in those parts of the common visible area of the celestial sphere that both stations can point at with very low elevations (so-called 'elevation cusps', for an example see sources S5 and S6 in figure 5.7 on page 81) are of particular importance for the determination of clock offsets.

Unfortunately, using observations with low elevations (which are necessary for the separability of the three effects), errors in the atmospheric models propagate into the remaining parameters and degrade the entire solution (BOEHM 2004).

$\delta\rho_h(z) = \cos(z)\cdot\delta\rho_h(0) \; ; \; \delta\rho_h(0) = \delta h$ $\qquad$ $\delta\rho_{clk}(z) = \delta\rho_{clk}(0) = c\cdot\delta t_R$ $\qquad$ $\delta\rho_{trp}(z) = mf(z)\cdot\delta\rho_{trp}(0) \sim 1/\cos(z)\cdot\delta\rho_{trp}(0)$
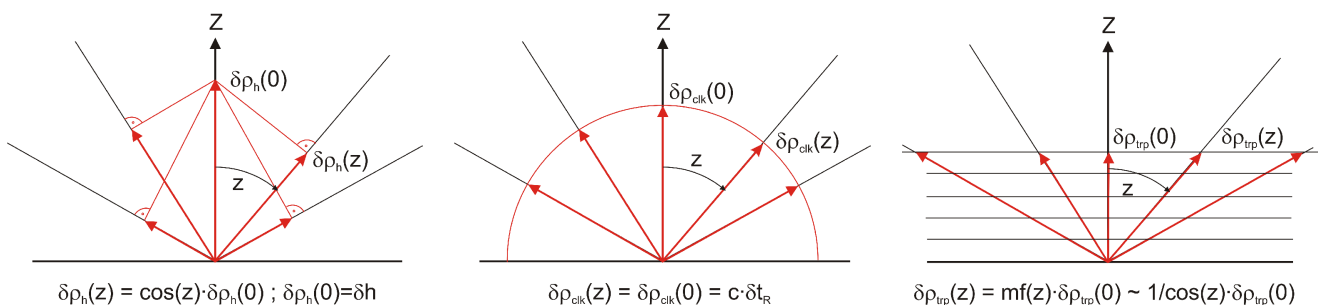
Figure 4.3: Impact of station height variations, clock offset and tropospheric delay on (GPS-)observations (ROTHACHER 2003)

4.2.2.2 Other partial derivatives

The similarities of the partial derivatives (4.29) and (4.31) show the lack of capability for the separation of $\Delta UT1$ and right ascension α of a radio source. Since these equations are equal (except for the factor $\Omega \approx 1$ and the sign) no separation is possible between variations in right ascension of a radio source and variations in earth rotation. A similar effect occurs in GPS with the ascending node and $\Delta UT1$ (ROTHACHER 2003).

Another example for critical (baseline) configurations and thus for inseparable parameters is given by baselines parallel to the equatorial plane observing sources with different right ascensions but identical declinations. With this configuration no separation between the Δz-component of the baseline and the clock offset between both stations is possible since variations in one of these parameters have the same impact on the observables (BROUWER 1985). This will be of importance in chapter 5 (case 6).

4.3   Parameter estimation in VLBI data analysis

The determination of parameters in equation 4.3 on page 57 is a typical adjustment problem and is often performed by least-squares estimation methods. Thus, the apriori model is refined by estimating model parameter corrections which best fit the data.

For the estimation of parameters from VLBI observations various different approaches exist, such as e.g.

- classical least-squares based on the solution of normal equations,

- Kalman filter and square root information filter approaches (e.g. ANDERSEN 2000) or

- collocation approaches (e.g. TITOV 2002).

For the investigations in the next chapter, the design matrix as used for the computation of the normal equations is being used. In practice, the dimension of the design matrix of a typical VLBI session is about 2000×200. For the common adjustment of several sessions (so-called global solution) these dimensions increase drastically.

4.4   Estimability limitations

Due to the limitations imposed by only observing the scalar product of the baseline and signal propagation vectors, unambiguous separation of parameters is only possible with sufficient spatially and temporally distributed observations, i.e., by avoiding observations performed only in *critical (baseline) configurations*. In practice, VLBI observation schedules are mostly generated by optimizing 'sky coverage', i.e., by observing sources in as many different positions in the commonly visible part of the celestial sphere as possible (STEUFMEHL 1994).

In addition to the limitations of the information content of VLBI observations due to the observing geometry, the *separation of physical parameters* is further complicated by linear combinations of a subset of parameters which may produce identical variations in the observables similar to other (linear combinations of) parameters of the model (so-called degeneracies). All such potential degeneracies must be identified and accounted for in the parameter estimation procedures (SOVERS et al. 1998).

Both critical (baseline) configurations and separation problems will be of importance in the next chapter.

5. Design analyses of plane and spatial interferometers

5.1 Introduction

The theoretical background derived in chapters 1 to 4 can now be used to investigate the design of interferometers, i.e., to analyse the type of parameters that can be estimated after measuring arrival time differences of signals emitted by extra-terrestrial radio sources and to analyse which parameters are affected by certain groups of observations. These methods can be used to achieve a deeper understanding of the impact of single (or groups of) observations on the adjustment process and can thus be used to optimise the observation schedule by neglecting observations (or observation groups) with small impact factors or by supporting/controlling observations with high impact factors by appropriate (independent) observations.

The main intention of this chapter is to show the suitability of the regression diagnostics tool developed in the previous chapters by testing its agreement with existing VLBI knowledge (such as e.g. found by the analysis of the functional model of VLBI in chapter 4). In some cases, however, new knowledge can be obtained which can only be found by investigating the entire observation schedule (and not just the partial derivatives of single observations). At first, the regression tool based on singular value decomposition and cluster analysis will be applied to a plane, static interferometer (or 2D-interferometer). In a second step, the investigations will be generalised and applied to spatial, rotating interferometers (i.e., to 3D-interferometer and thus to the VLBI principle).

5.2 VLBI observation schedule analysis software $qtSVD$

The author of this thesis developed $qtSVD$, a software package mainly designed for singular value decomposition-based analyses of VLBI observation schedules. The software is written in object-oriented C++ and uses the graphical user interface (GUI)-library Qt[1] which enables platform independent software development with user-friendly and mask-oriented dialogs. For matrix computations the *GNU Scientific Library (GSL)* has been used. For screenshots of $qtSVD$, see figures 5.2 and 5.3.

$qtSVD$ can also be used for investigations of other adjustment problems such as polynomial regression, plane interferometers or arbitrary design matrices (in MATLAB format). The software performs:

- visualisation of design matrices (for an example, see figure 5.2)

- singular value decomposition of design matrices

- computation of data resolution matrices, model resolution matrices, cofactor and correlation matrices

- cluster analysis of data resolution matrices

- visualisation of cluster analysis dendrograms

- computation of the 'average cluster impact factor $\bar{h}_{Cluster\ i}$ of cluster i on the parameter (group) $\mathbf{x}_1$'

For the analysis of VLBI observation schedules $qtSVD$ also

- reads observation schedules from NGS-files (i.e., ASCII files containing the VLBI observations of individual sessions),

---

[1]available at http://www.trolltech.com

- generates the design matrix after choosing an appropriate parametrisation (the computation of the design matrix is based on the VLBI data analysis software OCCAM (see Titov et al. 2004)),

- provides a three-dimensional visualisation of the network and the observation geometry with user-defined viewport settings (for an example, see figure 5.3).

The following investigations have been carried out with _qtSVD_, as well as the generation of all plots and tables.

5.3   Plane static interferometer

As shown in figure 5.1 the most simple interferometer consists of a single fixed baseline surrounded by fixed 'radio sources' considered to emit plane waves of radio signals. Due to the close relation to visible parts of the sky of a spatial interferometer (with a maximum of only one half of the celestial sphere being visible from both stations of the baseline) here only a half-circle is considered. The baseline is considered to be parallel to the x-axis of a plane, cartesian coordinate system where the origin agrees with the mid point of the baseline. At both ends of the baseline, signal receivers equipped with atomic clocks are assumed.
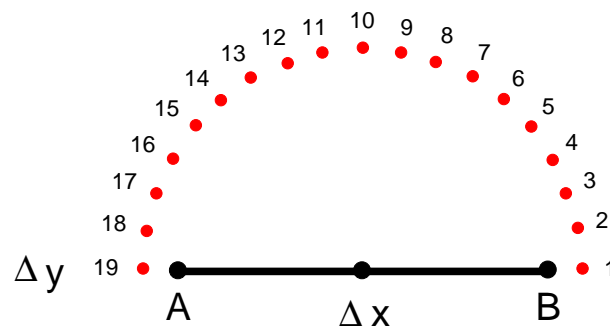
Figure 5.1: Plane static interferometer

Functional model

Observations τ_i may be obtained by measuring the difference of the arrival times of a 'radio signal' at both ends of the baseline. Estimable parameters are:

- the Δx component of the baseline,

- the Δy component of the baseline and

- the clock offset CL_0 of one clock with respect to the other clock.

The functional model of a plane, static interferometer thus reads:

$$\tau = \mathbf{b} \cdot \mathbf{k} + CL_0 = \left[\begin{array}{c} \Delta x \\ \Delta y \end{array} \right] \cdot \left[\begin{array}{c} k_1 \\ k_2 \end{array} \right] + CL_0 = \Delta x \cdot k_1 + \Delta y \cdot k_2 + CL_0 \qquad (5.1)$$

with $\mathbf{b}$ being the unit vector of the baseline and $\mathbf{k}$ being a unit vector in source direction. Note that the source vector $\mathbf{k}$ contains the coordinates of the unit vector into the direction of the respective source.
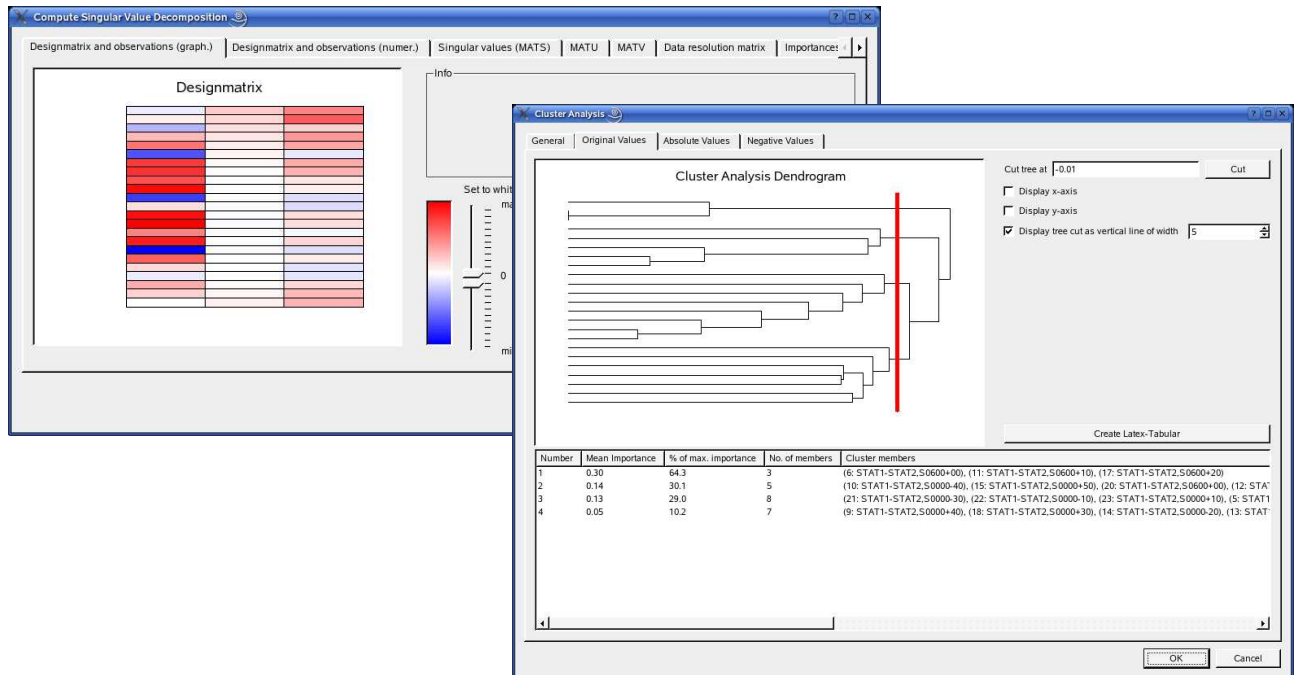
Figure 5.2: Screenshot of *qtSVD* (Matrix visualisation and cluster analysis modules)
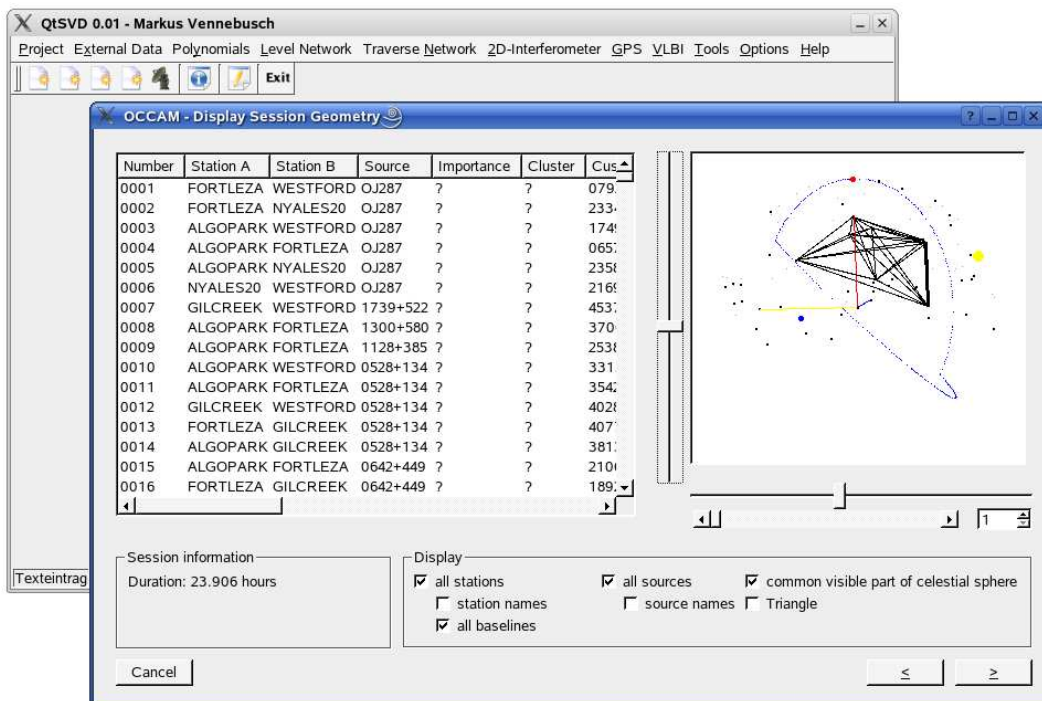

Figure 5.3: Screenshot of *qtSVD* (VLBI session visualisation module)

5.3.1   Investigation of parameter estimability

In the following sections the estimability of three parameter types (and combinations thereof) in the plane interferometer shown in fig. 5.1 are investigated. The three parameters Δx baseline component, Δy baseline component and clock offset CL_0 at station B can be combined in the following way:

1. Δx baseline component only (corresponds to estimating a change of the length of the baseline),

2. Δy baseline component only (corresponds to estimating a change of the height of one station),

3. clock offset CL_0 at station B only (corresponds to estimating the radius of a circle around station B),

4. Δx and Δy baseline component,

5. Δx baseline component and clock offset CL_0 at station B,

6. Δy baseline component and clock offset CL_0 at station B and

7. Δx and Δy baseline component and clock offset CL_0 at station B.

For each of the seven cases an identical set of 19 'observations' is obtained by simulating measurements of the arrival time differences of the signals from each source. Each source is observed only once in a counter-clockwise sense, starting at source 1 (see fig. 5.1). Since only the design of the experiment is of interest, no observation vector is present. Thus, for the following investigations only the design matrix $\mathbf{A}$ is used.

Figure 5.4 shows the components $\mathbf{S}, \mathbf{V}, \mathbf{U}$ of the singular value decomposition of the design matrix $\mathbf{A}$ as well as the data resolution matrix (DRM) and the model resolution matrix (MRM), the impact factors of the observations, the cluster analysis of the data resolution matrix and the correlation matrix $\mathbf{COR}$ of the estimated parameters for each of the seven investigated cases (see tabular 5.1).

	Element	Description
General	$\mathbf{S}$	Singular values, indicating presence of rank deficiencies
Model space (Parameters)	$\mathbf{V}$	Right singular vectors, indicating the amount of impact on the adjusted parameters
	$\mathbf{MRM}$	Model resolution matrix, indicates relations of parameters in case of a rank deficiency
	$\mathbf{COR}$	Correlation matrix, indicating correlations between parameters
Data space (Observations)	$\mathbf{U}$	Left singular vectors, indicating impact of observations on parameters (or linear combination thereof)
	h_{ii}	Impact factors, indicating the importance of individual observations
	$\mathbf{DRM}$	Data resolution matrix, main-diagonal elements h_{ii} indicating impact factors of observations, off-diagonal elements h_{ij} indicating impact co-factors and thus a jointly influential or counteracting effect of observations
	Dendro-gram	Result of cluster analysis of the data resolution matrix, indicating influential subsets of observations

Table 5.1: Singular value decomposition- and Cluster Analysis-based analysis elements
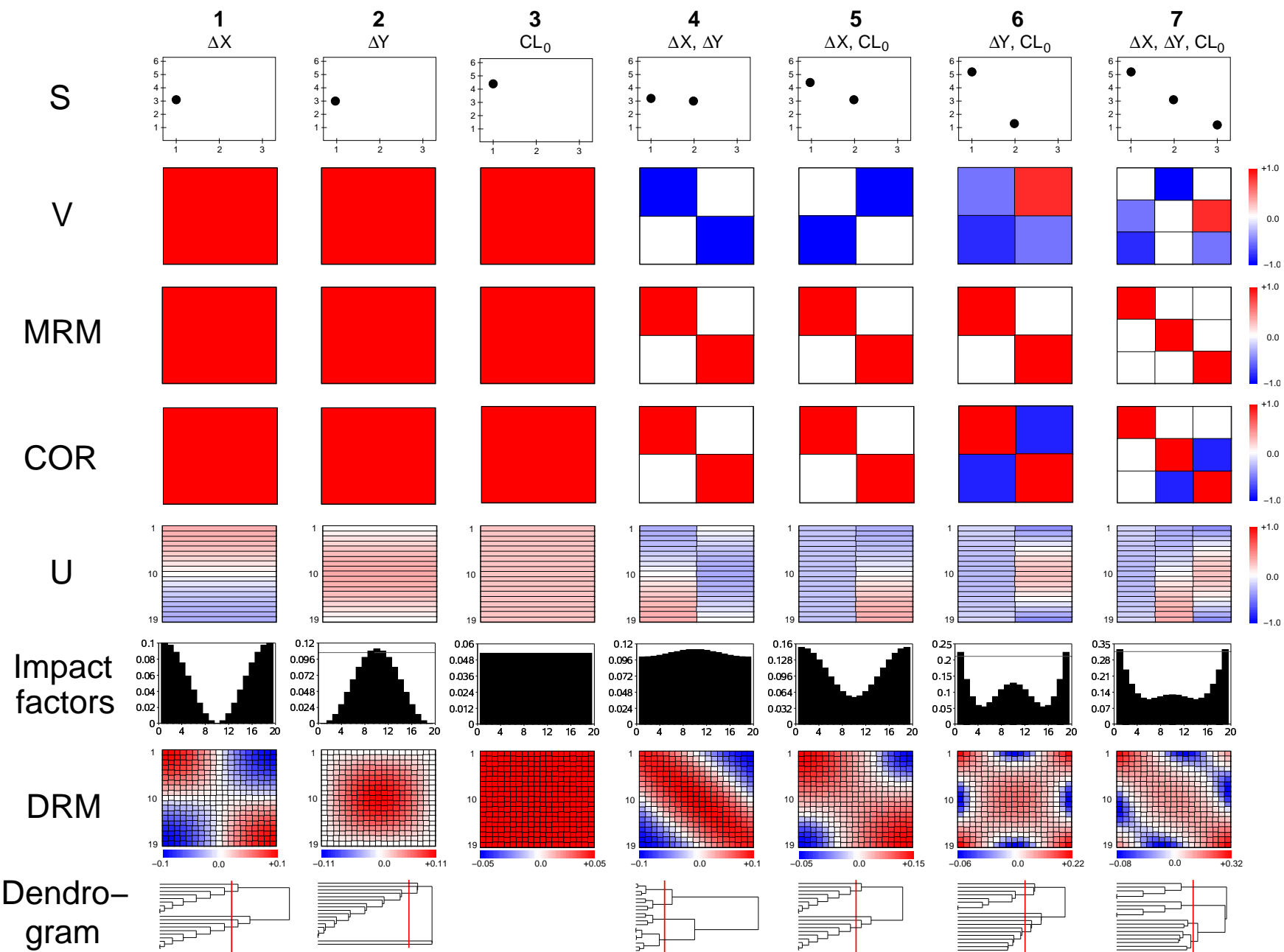
Figure 5.4: Analysis components of a plane interferometer. Red lines in impact factor diagrams indicate a recommended threshold for 'high-leverage observations' at $2 \cdot \frac{u}{n}$ (with u = number of unknowns, n = number of observations).

Interpretation guidelines

For a correct interpretation of the elements listed in table 5.1 and displayed in figures 5.4, 5.8, 5.9 and 5.16, eq. 2.47 on page 39 should be reviewed. This formula describes the least-squares solution of a linear system $\mathbf{Ax} = \mathbf{y}$ by using the singular value decomposition of $\mathbf{A}$:

$$\hat{\mathbf{x}} = \sum_{i=1}^{r} \frac{1}{\sigma_i} \mathbf{v}_i \left(\mathbf{u}'_i \cdot \mathbf{y} \right) = \mathbf{V}_r \mathbf{S}_r^{-1} \mathbf{U}'_r \mathbf{y}. \tag{5.2}$$

Eq. 5.2 shows that very small singular values σ_i (or singular values which even equal zero) lead to unreasonable contributions to the least squares solution, caused by (near) rank deficiencies.

After verifying the absence of rank deficiencies (i.e., then $r = u$), each of the r summands of eq. 5.2 needs to be related to a certain parameter by identifying the largest element(s) of each right singular vector $\mathbf{v}_i$ (for $i = 1, \ldots, r$). As shown below, only in very few cases a single element of a $\mathbf{v}_i$-vector is significantly larger than the remaining elements. In many cases, however, several elements are of similar size and thus show that the corresponding summand affects more than just one parameter.

In any case, for each summand the corresponding left singular vector $\mathbf{u}_i$ shows the impact each observation has on the parameters identified by analysing the corresponding right singular vector $\mathbf{v}_i$. This can even be performed without a real observation vector.

Since the singular values σ_i are usually sorted in ascending order, the right singular vector $\mathbf{v}_1$ of the first summand (which is computed with the largest singular value, σ_1) reveals those parameters (or a linear combination thereof) which is best determined. On the other hand, the last summand (which is computed with the smallest singular value, σ_r) displays those parameters (or parameter linear combinations) which are worst determined.

Information content of plane static interferometer observations

For the following investigations, the information content of individual observations performed in a plane static interferometer will be important. Figure 5.5 shows three observation geometries with very low elevations, medium elevations and very high elevations, respectively.

The Δx component of this baseline can be interpreted as a variation of the baseline length. The Δy component might be regarded as a height variation of station B. The circle around station B displays the offset CL_0 of the clock at station B with respect to the reference clock at station A. More detailed explanations will be given below.
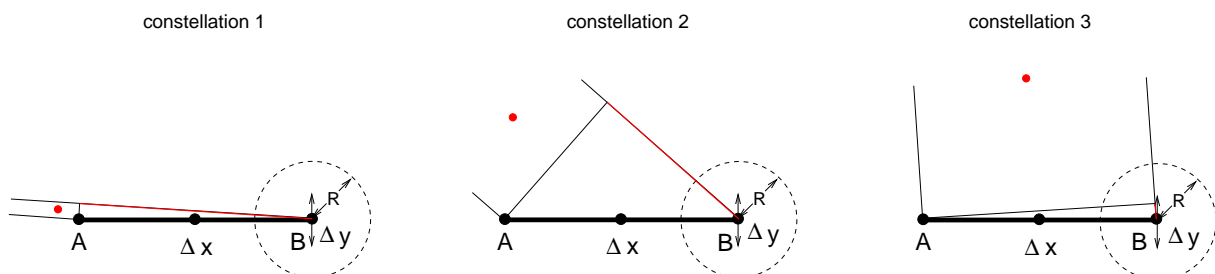

Figure 5.5: Information content of plane static interferometer observations. The Δx component displays a variation of the baseline length. The Δy component visualises a height variation of station B and the circle around station B displays the clock offset CL_0 of station B with respect to the reference clock at station A.

Case 1: Estimation of Δx baseline component only

- **General / Singular values:**
  As mentioned above, every singular value decomposition-based analysis of an adjustment problem should start with the investigation of the singular values σ_i, indicating a potential rank deficiency of the design matrix. Here, the only singular value is $\sigma_1 = 3.16 > 0$ and thus no rank deficiency is present.

- **Model space / Right singular vectors (Matrix V):**
  Since in this case only one parameter is estimated the **V**-matrix equals a 1×1-identity matrix and thus does not provide any useful information.

- **Model space / Model resolution matrix (MRM):**
  Since the model resolution matrix **MRM** is computed by the right singular vectors it also does not provide any useful information for this case.

- **Model space / Correlation matrix (COR):**
  Similar to the two items above, the correlation matrix does not give any useful information for this particular case.

- **Data space / Left singular vectors (Matrix U):**
  The **U**-matrix (i.e., the left singular vector of the design matrix) already indicates a significant difference in the impact of the individual observations. Considering the absolute values of the elements of the left singular vector, a decreasing impact of observations 1 to 9 can be seen. On the other hand, observations 11 to 19 show an increasing impact. Observation 10 is of negligible impact for the estimation of the Δx baseline component.

- **Data space / Impact factors:**
  The main-diagonal elements of the Data Resolution Matrix indicate the overall impact of each observation on the parameters to be estimated. As already seen in the left singular vector, for the determination of the Δx baseline component, the first and the last observations mainly affect the estimation process. Observation 10 could have been omitted completely since it does not provide any information.

- **Data space / Data resolution matrix (DRM):**
  A jointly influential or a counter-acting effect of observations (see section 3.2 on page 50) can be detected by row-wise investigating the Data Resolution Matrix (red = jointly influential effect, blue = counter-acting effect): A jointly influential effect and thus a common degree of information can be seen for observations 1 to 9 and observations 11 to 19, respectively. For observations 1 to 9 a decreasing significant counter-acting effect to observations 11 to 19 can be seen (and vice versa), since observations to sources with an angular difference of approximately 180° 'pull' the baseline component into opposite directions.

  For the determination of the Δx baseline component, the 10th observation does not have any impact at all (see 10th row of DRM). Nor does the 10th observation affect the remaining observations (see 10th column of DRM). Thus, this observation could have been omitted. Geometrically, this becomes obvious from figure 5.5 on the preceding page (constellation 3): For observations performed with high elevations, a small variation in the Δx baseline component hardly affects the arrival time difference of the signal. Hence, these kind of observations are not sensitive to the Δx baseline component and are thus not suited for the determination of the Δx baseline component. On the other hand, constellation 1 in figure 5.5 shows that the first and the last few observations are needed for the Δx baseline component.

  This means, that in order to obtain important observations for the determination of the Δx baseline component, sources located in direction of the baseline vector have to be used. Observations to sources orthogonal to the baseline are of no use for this parameter. Obviously, this depends on the orientation of the reference system. In general, observations to sources orthogonal to the baseline cannot be used for the determination of the baseline length.

- **Data space / Cluster analysis:**
  Cluster analysis of the data resolution matrix yields a dendrogram with two branches indicating two different groups of observations (see figure 5.6 for the enlarged dendrogram). Observations 1 and 2 and

Cluster Analysis Results - Case 1				
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter [-]: Δx
1	1	0.00	9	0.003
2	1	0.00	10	0.000
3	1	0.00	11	0.003
4	8	0.06	12, 13, 14, 15, 16, 17, 18, 19	**0.062**
5	8	0.06	8, 7, 6, 5, 4, 3, 1, 2	**0.062**

Table 5.2: Cluster analysis results for case 1 (Estimation of Δx baseline component only)

Cluster Analysis Dendrogram
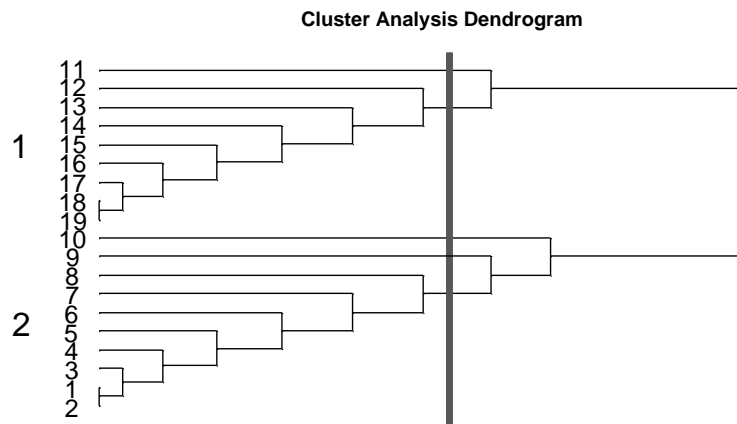
Figure 5.6: Enlarged dendrogram for case 1 (estimation of Δx baseline component)

observations 19 and 18 have been clustered at first. The last single observation which has been clustered is observation 10, showing again that this observation is least 'similar' to the remaining observations.

In order to get reasonable clusters, the dendrogram has been cut to form five clusters. Table 5.2 shows the results of cluster analysis as well as the impact of each cluster on the Δx baseline component.

Case 2: Estimation of Δy baseline component only

- **General / Singular values:**
  The only singular value is $\sigma_1 = 3.00 > 0$ and thus no rank deficiency is present.

- **Model space / Right singular vectors (Matrix V):** See case 1.

- **Model space / Model resolution matrix (MRM):** See case 1.

- **Model space / Correlation matrix (COR):** See case 1.

- **Data space / Left singular vectors (Matrix U):**
  Compared to case 1, the left singular vector (matrix **U**) has an opposite structure: Here, the first and the last observations only have a small impact, while the middle observations have a large impact on the determination of the Δy baseline component. The largest impact is produced by the 10th observation.

- **Data space / Impact factors:**
  Consequently, the same effect can be seen in the impact factors.

Cluster Analysis Results - Case 2				
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter: Δy
1	1	0.00	1	0.000
2	1	0.00	2	0.003
3	1	0.00	18	0.003
4	1	0.00	19	0.000
5	15	0.07	17, 3, 4, 16, 5, 15, 6, 14, 13, 7, 8, 12, 9, 10, 11	**0.066**

Table 5.3: Cluster analysis results for case 2 (Estimation of Δy baseline component only)

- **Data space / Data resolution matrix (DRM):**
  The data resolution matrix shows a distinct agglomeration of observations. This indicates that the middle observations possess a similar information content (decreasing to the first and the last observation, respectively) and that there are no counter-acting observations present.

  The most important observations for the determination of the Δy baseline component have thus to be performed to sources lying orthogonal to the baseline (see also constellation 3 in figure 5.5 on page 70). The least important (and thus negligible) observations are performed to sources lying in direction of the baseline vector. Again, this depends on the orientation of the coordinate system. In general, observations to sources orthogonal to a baseline can only be used to determine a position variation orthogonal to the baseline, since these observations are most affected by this effect.

- **Data space / Cluster analysis:**
  Cluster analysis of the data resolution matrix shows a sequential clustering of the most similar observations. Observations 1 and 19 have been clustered at last. Table 5.3 shows the results as well as the impact on the Δy baseline component.

Case 3: Estimation of the clock offset CL_0 at station B only

- **General / Singular values:**
  The only singular value is $\sigma_1 = 4.36 > 0$ and thus no rank deficiency is present.

- **Model space / Right singular vectors (Matrix V):** See case 1.

- **Model space / Model resolution matrix (MRM):** See case 1.

- **Model space / Correlation matrix (COR):** See case 1.

- **Data space / Left singular vectors (Matrix U):**
  In this case the **U** matrix consists of a column vector with identical elements ($u_i = 0.23$). This shows that each observation has the same impact on the parameter estimation process and thus to the determination of the clock offset. This is obviously caused by the fact that the design matrix consists of a vector of constants and is thus independent of the observation geometry.

- **Data space / Impact factors:**
  Consequently, the impact factors show the same effect, i.e., each observation is equally important for the determination of the clock offset.

- **Data space / Data resolution matrix (DRM):**
  For this case, every element of the DRM equals 0.053 and thus for the determination of the clock offset no classification of observations can be performed. Each observation is equally important and no counter-acting observations are present.

- **Data space / Cluster analysis:**
  Due to the homogeneous structure of the DRM cluster analysis cannot be performed, i.e., no clusters can be formed.

Excursus: A similar effect can be seen when investigating the data resolution matrix of a polynomial adjustment with just one parameter (i.e., computation of the arithmetic mean). In this case every observation has the same impact factor. This is due to the fact that the design matrix consists of a column with a constant value.

Case 4: Estimation of Δx and Δy baseline component

- **General / Singular values:**
  The singular values $\sigma_1 = 3.16$ and $\sigma_1 = 3.00$ show that both parameters can be estimated with approximately the same accuracy.

- **Model space / Right singular vectors (Matrix V):**
  The **V**-matrix indicates that both parameters can be determined independently.

- **Model space / Model resolution matrix (MRM):**
  Since no rank deficiency is present the MRM equals an identity matrix.

- **Model space / Correlation matrix (COR):**
  Similarly, the diagonal structure of the model resolution matrix also shows that both parameters can be determined independently.

- **Data space / Left singular vectors (Matrix U):**
  Since the largest element of the first right singular vector is in the first row (and thus affects the first parameter), the first singular value and the first left singular vector (see matrix **U**) are used to analyse the impact of the observations on the determination of the Δx baseline component: As in the first case, the (absolute) values of the first left singular vector show that the first and the last observations exert the largest impact on the determination of the Δx baseline component.

  On the other hand, the second right singular vector affects the second parameter. Thus, the second singular value and the second left singular vector influence the determination of the Δy baseline component: Here, the (absolute) values of the second left singular vector show the same behaviour as in case 2.

- **Data space / Impact factors:**
  For the determination of both baseline components the middle observations (i.e., observations 8 to 12) are of slightly higher importance. This is caused by the fact that these observations have to determine the Δy baseline component alone and do not posses any 'counter-parts' on the other side of the baseline and are thus less controlled.

  Numerically, the impact factors of this case (and only of this case!) consist of a superposition of the two first elementary cases, because $\mathrm{DRM}_{Case\ 4} = \mathrm{DRM}_{Case\ 1} + \mathrm{DRM}_{Case\ 2}$.

- **Data space / Data resolution matrix (DRM):**
  Since this data resolution matrix is the sum of the data resolution matrix of case 1 and the data resolution matrix of case 2, for the first and the last observation (rows 1 and 19) identical patterns as in case 1 can be seen. For the middle observation (row 10) the pattern is identical to the pattern of observation 10 in case 2.

  For the remaining observations a significant jointly influential effect of neighboring observations can be seen in the distinct main-diagonal structure. Counter-acting effects mainly occur for observations to sources with an opposing 'counter part', i.e., mainly for the first and last few observations. Again, geometrical interpretations are aided by figure 5.5 on page 70.

Cluster Analysis Results - Case 4					
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter: Δx	Δy
1	4	0.11	12, 13, 14, 15	0.034	**0.073**
2	4	0.10	16, 17, 18, 19	**0.090**	0.011
3	4	0.10	3, 4, 1, 2	**0.090**	0.011
4	7	0.11	10, 11, 8, 9, 5, 6, 7	0.020	**0.088**

Table 5.4: Cluster analysis results for case 4 (Estimation of Δx and Δy baseline component)

- **Data space / Cluster analysis:**
  Figure 5.4 and table 5.4 show the results of the cluster analysis. The four clusters show a symmetric classification of observations 1 to 4 and 16 to 19 in clusters 3 and 2, respectively. These clusters mostly affect the Δx baseline component. Clusters 1 and 4 show a slightly unsymmetric classification of four and seven observations, respectively. Members of these clusters mainly affect the Δy baseline component.

  This case also reveals the difficulty of choosing an appropriate height for the dendrogram cut. Depending on the (subjective) user decision, more or less symmetric clusters can be generated. The actual number of clusters thus depends on the purpose of the experiment and on the problem to be analysed.

Case 5: Estimation of the Δx baseline component and the clock offset CL_0 at station B

- **General / Singular values:**
  The two singular values ($\sigma_1 = 4.35$ and $\sigma_2 = 3.16$) indicate the absence of rank-deficiencies.

- **Model space / Right singular vectors (Matrix V):**
  The **V** matrix consists of a column-permutated identity matrix and thus every parameter can be determined separately.

- **Model space / Model resolution matrix (MRM):**
  Due to the absence of rank deficiencies the model resolution matrix equals an identity matrix.

- **Model space / Correlation matrix (COR):**
  The correlation matrix is an identity matrix and thus indicates that every parameter can be determined separately.

- **Data space / Left singular vectors (Matrix U):**
  Since numerically this case consists of a superposition of cases 1 and 3, the **U** matrix has a similar structure as the individual **U** matrices of cases 1 and 3. As shown in the right singular vectors (matrix **V**) the first left singular vector affects the clock parameter. Again, every observation exerts the same impact on this parameter.

  On the other hand, the (absolute values of the) second left singular vector show a similar structure as the left singular vector of case 1. Due to the second right singular vector, the Δx baseline component is thus mainly affected by the first and the last observations.

- **Data space / Impact factors:**
  Similar to case 1, the first and last few observations are of main importance. In contrast to case 1, however, the middle observations also have a significant impact on the regression results since they are necessary for the clock offset determination. This is caused by the fact that the data resolution matrix of this case is the sum of the data resolution matrix of case 1 and the data resolution matrix of case 3. Thus, the impact factors of this case consist of the impact factors of case 1 shifted by a constant value.

Cluster Analysis Results - Case 5					
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter:	
				Δx	CL_0
1	1	0.06	9	0.003	0.053
2	1	0.05	10	0.000	0.053
3	1	0.06	11	0.003	0.053
4	8	0.11	12, 13, 14, 15, 16, 17, 18, 19	**0.062**	0.053
5	8	0.11	8, 7, 6, 5, 4, 3, 1, 2	**0.062**	0.053

Table 5.5: Cluster analysis results for case 5 (Estimation of Δx baseline component and the clock offset CL_0 at station B)

- **Data space / Data resolution matrix (DRM):**
  As for all previous cases with the Δx baseline component being an unknown parameter, the first and the last observations are concurrent observations. The middle observations show a similar, almost homogeneous pattern as in case 3. Thus, these observations possess a similar information content and have no counter-acting observations. Again, this is due to the superposition of the cases 1 and 3.

- **Data space / Cluster analysis:**
  The cluster analysis of the data resolution matrix shows a similar structure as in case 1: Observations 1 to 9 are clustered into the first cluster. Observations 10 to 19 are clustered into the second cluster. As in case 1, observations 1 and 2 and observations 18 and 19 are clustered at first; observation 10 at last.

  Cutting the dendrogram at a reasonable height to form five clusters yields that the first (cluster 5) and the last (cluster 4) few observations are both necessary for the determination of the Δx baseline component and the clock offset. The middle observations (clusters 1 to 3) mainly affect the determination of the clock offset only since these observations are not sensitive to variations in the Δx baseline component (see table 5.5).

Case 6: Estimation of the Δy baseline component and the clock offset CL_0 at station B

For this case, interpretations are complicated because some matrices are not of diagonal structure. This is caused by the fact that some observations are conducted close to critical configurations for the determination of both the Δy baseline component and the clock offset CL_0 (see table 5.6 on page 78). For observations to sources orthogonal to this baseline, a change in the Δy baseline component and a change in the clock offset CL_0 have the same effect (see constellation 1 in table 5.6). Furthermore, due to the lack of supporting diametral observations (i.e., observations to sources below the baseline), both effects cannot be completely separated (cf. BROUWER 1985). As constellations 2 to 3a/b show, separability is improved (i.e., the correlation coefficient decreases) if observations performed with low elevations are included. The best separability is given for constellations containing diametral observations (see constellation 3b).

Obviously, this only holds for the current baseline and source geometry. A rotation of the baseline or the entire celestial sphere yields different results and thus different relations between the parameters to be determined.

- **General / Singular values:**
  Although both singular values ($\sigma_1 = 5.15$ and $\sigma_2 = 1.23$) indicate the absence of rank-deficiencies they also show that both parameters can only be determined with different accuracies. From the second right singular vector the first parameter (Δy baseline component) is identified as the parameter which is determined weaker than the clock offset CL_0.

- **Model space / Right singular vectors (Matrix V):**
  The weak separability of the two effects can also be seen in the **V**-matrix: Since no distinct diagonal structure can be seen, only the impact of observations on linear combinations of these parameters can be investigated.

- **Model space / Model resolution matrix (MRM):**
  As in the previous cases, the model resolution matrix does not reveal any relations between the unknown parameters, since no rank deficiency is present.

- **Model space / Correlation matrix (COR):**
  The correlation matrix shows a significant correlation of the Δy baseline component and the clock offset CL_0 of -0.87.

- **Data space / Left singular vectors (Matrix U):**
  Due to the weak separability of the two parameters, the **U**-matrix cannot be used for clear statements and must not be over-interpreted. Just a higher general impact of the first, middle and last observations can be detected. Observations 5 and 15 are of minor impact especially for (1.) the Δy baseline component and (2.) the clock offset CL_0.

- **Data space / Impact factors:**
  Due to the structure of the **U**-matrix the impact factors indicate the highest impact for the first and last observations as well as for the middle observations. The lowest impact factors are given for observations with a $45°$ angle between baseline vector and vector in source direction. An interpretation is given below.

- **Data space / Data resolution matrix (DRM):**
  The analysis of the data resolution matrix is complicated by the weak separation problem and must not be over-interpreted. Numerically (and in contrast to the previous cases), this data resolution matrix and thus the impact factors are not a superposition of the data resolution matrices of any of the three elementary cases 1 to 3.

- **Data space / Cluster analysis:**
  Cluster analysis classifies the observations into two main groups with the upper cluster containing the first and the last observations which are of importance for both the Δy baseline component and the clock offset. The second cluster containing the remaining observations is of major importance for the Δy baseline component only.

  The dendrogram cut displayed in figure 5.4 shows a classification into eight clusters. Again, the first and the last observations (cluster 7) are of major importance for both parameters (see table 5.7) while the observations to sources with a $45°$ angle between baseline vector and vector in source direction (clusters 1-3 and 4-6) have the smallest mean impact factor. The middle observations (cluster 8) mostly affect the Δy baseline component. Again, these interpretations are complicated by the weak separability of both parameters.

Since observations to sources orthogonal to this baseline cannot be used for the separation of both parameters, the remaining observations have to be used instead. Although theoretically the first and the last observations are not suitable for the solely determination of the Δy baseline component (see case 2), cluster analysis still detects these observations as useful for the determination of both parameters. This is due to the lack of suitable alternative observations. In connection with the determination of the clock offset, these observations are still the most useful ones among all available observations. In any case, the separability of both effects (parameters) is weak (see table 5.6).

In general (i.e., also for the three-dimensional case), separation of effects is complicated if both effects act in the same direction. It is further complicated if there are no diametral (supporting) observations available. If possible, for real experiments a re-parametrisation would be appropriate.

Case 6 - Separability of Δy and CL_0			
	Condition number:	correlation ($\rho_{\Delta y, CL_0}$):	cofactor matrix (parameter order: $\Delta y, CL_0$):
Constellation 1			
	277.5	-0.99	$\begin{bmatrix} 6550.29 & -6484.21 \\ -6484.21 & 6419.12 \end{bmatrix}$
Constellation 2			
	12.9	-0.98	$\begin{bmatrix} 28.14 & -17.90 \\ -17.90 & 11.72 \end{bmatrix}$
Constellation 3a			
	7.4	-0.77	$\begin{bmatrix} 17.13 & -2.94 \\ -2.94 & 0.84 \end{bmatrix}$
Constellation 3b			
	12.3	-0.57	$\begin{bmatrix} 49.94 & -2.89 \\ -2.89 & 0.50 \end{bmatrix}$

Table 5.6: Separability of Δy and CL_0 in case 6. For each constellation, three different observations have been performed to each of the three sources displayed in the respective diagram. For constellations 1, 2 and 3a, the observation 'triangle' is only displayed for observations to the second (or middle) source, respectively. For constellation 3b, the third observation is visualised. Vertical arrows indicate a height change (Δy) of station B. R indicates the radius of a circle around station B and thus visualises the clock offset CL_0 of the clock at station B with respect to the reference clock at station A.

Cluster Analysis Results - Case 6					
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter: Δy	CL_0
1	1	0.06	4	0.005	0.030
2	1	0.05	5	0.001	0.008
3	1	0.07	6	0.013	0.000
4	1	0.07	14	0.013	0.000
5	1	0.05	15	0.001	0.008
6	1	0.06	16	0.005	0.030
7	6	0.15	17, 3, 18, 2, 1, 19	**0.096**	**0.143**
8	7	0.11	13, 7, 12, 8, 11, 9, 10	**0.055**	0.009

Table 5.7: Cluster analysis results for case 6 (Estimation of Δy baseline component and the clock offset CL_0 at station B)

Case 7: Estimation of the Δx baseline component, Δy baseline component and the clock offset CL_0 at station B

The determination of the three parameters Δx baseline component, Δy baseline component and the clock offset CL_0 depicts a superposition of the cases 1 and 6 (not of cases 1, 2 and 3!). Thus, this experiment design again contains observations which are not suitable for the determination of any parameter (cf. case 6).

- **General / Singular values:**
  The three singular values $\sigma_1 = 5.15, \sigma_2 = 3.16$ and $\sigma_3 = 1.24$ show the absence of rank deficiencies. Again, however, a significant decline is visible and shows that the parameters cannot be determined with equal accuracy.

- **Model space / Right singular vectors (Matrix V):**
  The **V**-matrix is not of diagonal structure. Only the first parameter ($=\Delta x$ baseline component) is not affected by any other parameter (see first row of **V**). The remaining parameters (Δy baseline component and clock offset CL_0) can only be analysed together, since these parameters are both affected by the first and the third right singular vector.

- **Model space / Model resolution matrix (MRM):**
  Since no rank deficiency is present, the model resolution matrix consists of an identity matrix.

- **Model space / Correlation matrix (COR):**
  The correlation matrix shows the same correlation (-0.87) between the Δy baseline component and the clock offset CL_0 as in case 6. The Δx baseline component is not correlated with any other parameter and can thus be determined separately.

- **Data space / Left singular vectors (Matrix U):**
  The first and the third left singular vector agree with the left singular vectors of case 6. The second singular vector is the same as the left singular vector of case 1.

- **Data space / Impact factors:**
  The first and the last observations are of high importance while the middle observations nearly have the same (medium) impact factors. Numerically, the impact factors are the sum of the impact factors of cases 1 and 6.

- **Data space / Data resolution matrix (DRM):**
  Since this data resolution matrix consists of the sum of the data resolution matrix of case 1 and the data resolution matrix of case 6 (containing critical configurations) it must not be over-interpreted. Almost every observation possesses a counter-acting observation and thus complicates the geometric interpretation.

Cluster Analysis Results - Case 7						
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter: Δx	Δy	CL_0
1	5	0.20	5, 4, 3, 1, 2	**0.084**	**0.059**	**0.093**
2	5	0.20	15, 16, 17, 18, 19	**0.084**	**0.059**	**0.093**
3	3	0.11	14, 12, 13	0.026	0.033	0.004
4	6	0.12	8, 11, 9, 10, 6, 7	0.014	0.052	0.009

Table 5.8: Cluster analysis results for case 7 (Estimation of Δx baseline component, Δy baseline component and the clock offset CL_0 at station B)

- **Data space / Cluster analysis:**
  The cluster analysis shown in figure 5.4 shows a classification into three main groups consisting of the first five observations, the last five observations and the remaining observations, respectively.

  For a finer differentiation, the dendrogram has been cut to form four clusters (see table 5.8). The first and the second cluster consist of the first and the last observations and possess the largest mean importance for the entire parameter set. In addition the Δy baseline component is affected by the fourth cluster which mainly consists of the observations to sources orthogonal to the baseline.

  As in case 6, the first and the last observations are also necessary for the determination of the Δy baseline component. Again, this is caused by the additional estimation of the clock offset and the lack of suitable supporting observations.

5.3.2   Conclusions from plane static interferometer investigations

Based on the analyses performed so far, it could be shown that the regression diagnostics tool based on singular value decomposition and cluster analysis yields plausible and geometrically comprehensible results. It is possible to detect groups of observations with different impact on the parameters of interest. Using the terms of chapter 3: Both jointly influential and counter-acting groups of observations can be identified. Observation groups with large mean impact factors significantly affect the estimation process and thus should be controlled by appropriate (independent) observations. On the other hand, observation groups with small impact factors or which affect parameters of minor interest could be reduced or even neglected.

For some cases, however, the difficulties for performing a proper dendrogram cut became obvious: Sometimes, the appropriate height for a dendrogram cut does not agree with a large similarity difference. Instead, the dendrogram cut has to be performed in such a way that an appropriate number of clusters is generated. This again shows the subjective (and thus ambiguous) part of the regression diagnostics procedure.

Furthermore, the regression diagnostics tool developed in the first chapters can be used to detect degeneracies or critical (baseline) configurations (such as in case 6). In these cases (such as case 6, where the Δy baseline component and the clock offset CL_0 at station B could hardly be separated), singular value decomposition reveals indeterminable parameters (or indeterminable linear combinations thereof). The interpretation of the singular values will be of even more importance in the following investigations (for the estimation of the x_p-parameter).

The main purpose of the analyses performed so far has been to show the suitability of the regression diagnostics tool for simple interferometers. In the following, more complex cases (i.e., both three-dimensional or spatial and rotating or kinematic interferometers) will be treated.

5.4 Spatial kinematic interferometer

The generalisation of a plane static interferometer first leads to a spatial static interferometer and then to a spatial kinematic interferometer. The latter equals the VLBI principle if the rotation axis of the interferometer coincides with the z-axis of a geocentric, earth-fixed terrestrial coordinate system. An example for a single-baseline interferometer is shown in figure 5.7.

As described in chapter 4, observations performed in a single-baseline spatial kinematic interferometer (or within a single-baseline VLBI observation session) can also be used to estimated (at least) the following parameters:

- Terrestrial reference system:
    - site positions x_B, y_B, z_B of stations B (station A has to be kept fixed as a reference station)
- Earth orientation parameters:
    - polar motion x_p, y_p
    - earth rotation $dUT1$
- Auxiliary parameters:
    - clock offset CL_0 and clock rate CL_1 of one clock with respect to the other clock (reference clock)
    - atmospheric zenith path delays AT_A at station A and AT_B at station B.

and combinations thereof. Depending on the baseline length, baseline orientation and on the observation geometry some of these parameters may not be estimable.

In order to obtain reasonable results, all parameters have to be transformed to the same unit. For the following investigations every parameter has been scaled to the unit 'meter' or 'meter/day' for the clock rate parameter.
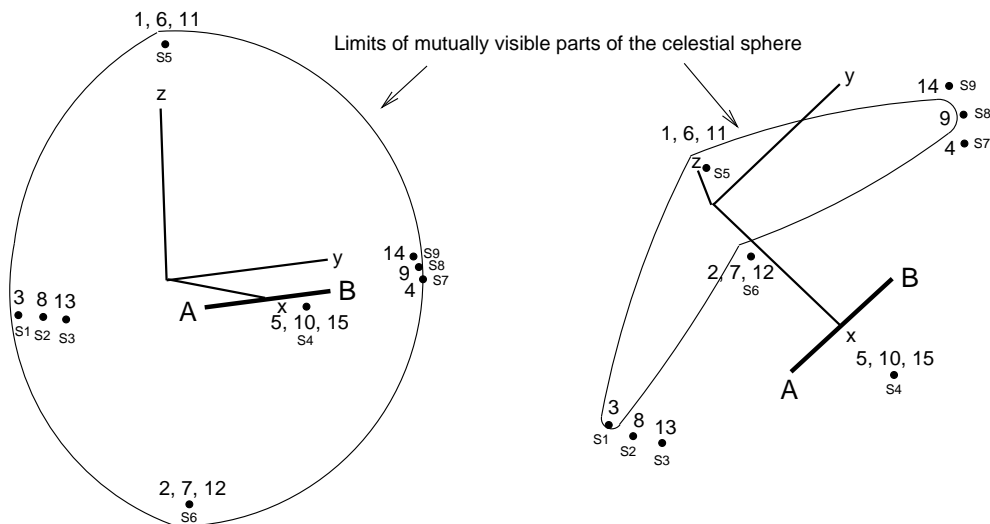
Figure 5.7: Spatial interferometer geometry as used for investigations (left: equatorial view, right: polar view), 'Sx' indicates the source name, numbers indicate the observation number.

5.4.1   Estimability investigations for basic parameters

Network geometry and observation schedule

For the following investigations of parameter estimability in spatial kinematic interferometers, a fictitious interferometer as shown in figure 5.7 is used. The two stations form an equatorial baseline with a length of 5000 km, parallel to the y-axis and located on the tip of the x-axis of a geocentric coordinate system. Nine artificial radio sources are located both at the celestial equator (declination: 0°) and near to the celestial poles (declination: 85° and -85°). A list of these sources is shown in table 5.9[2].

As shown in table 5.10 the observation schedule consists of 15 observations with a duration of four minutes each. The entire observation duration of one hour is divided into three groups of five observations, respectively. The observations of each group are performed in a similar sequence:

- observation to a polar source with high declination (source 5 (S0300+85)), i.e., close to the horizon of both stations (source is located in the northern 'elevation cusp')

- observation to a polar source with low declination (source 6 (S0300-85)), also close to the horizon of both stations (source is located in the southern 'elevation cusp')

- observation to an equatorial source with a high right ascension (either source 1 (S2230+00), source 2 (S2255+00) or source 3 (S2330+00), depending on hour angle and visibility) and thus close to the horizon of station B

- observation to an equatorial source with a low right ascension (either source 7 (S0700+00), source 8 (S0730+00) or source 9 (S0755+00), depending on hour angle and visibility) and thus close to the horizon of station A

- observation to an equatorial source (source 4 or S0300+00) which is close to the zenith of both stations.

This scheme is repeated three times, yielding fifteen observations in total. Due to the rotation of the interferometer some equatorial sources set or rise during the observation period so that different equatorial sources close to the horizon of one of the stations have to be observed.

Estimability analysis

In order to understand the complex situations of parameter estimation within real VLBI-observation sessions (as treated in section 5.4.3 on page 99), some basic parameter sets have to investigated first:

8. Estimation of site positions x_B, y_B, z_B of station B only

9. Estimation of clock offset CL_0 of station B only

10. Estimation of clock rate CL_1 of station B only

11. Estimation of atmospheric zenith path delay AT_A at station A only

12. Estimation of atmospheric zenith path delay AT_B at station B only

13. Estimation of polar motion y_p only

14. Estimation of earth rotation $dUT1$ only

As for the plane interferometer investigations, the singular value decomposition-based analysis components as displayed in figure 5.8 on the facing page will be analysed.

---

[2]The second to fifth character of the source name indicate the right ascension (in [hh min]), the sixth character indicates the sign of the declination and the seventh and eighth character indicate the declination of the source (in degrees).
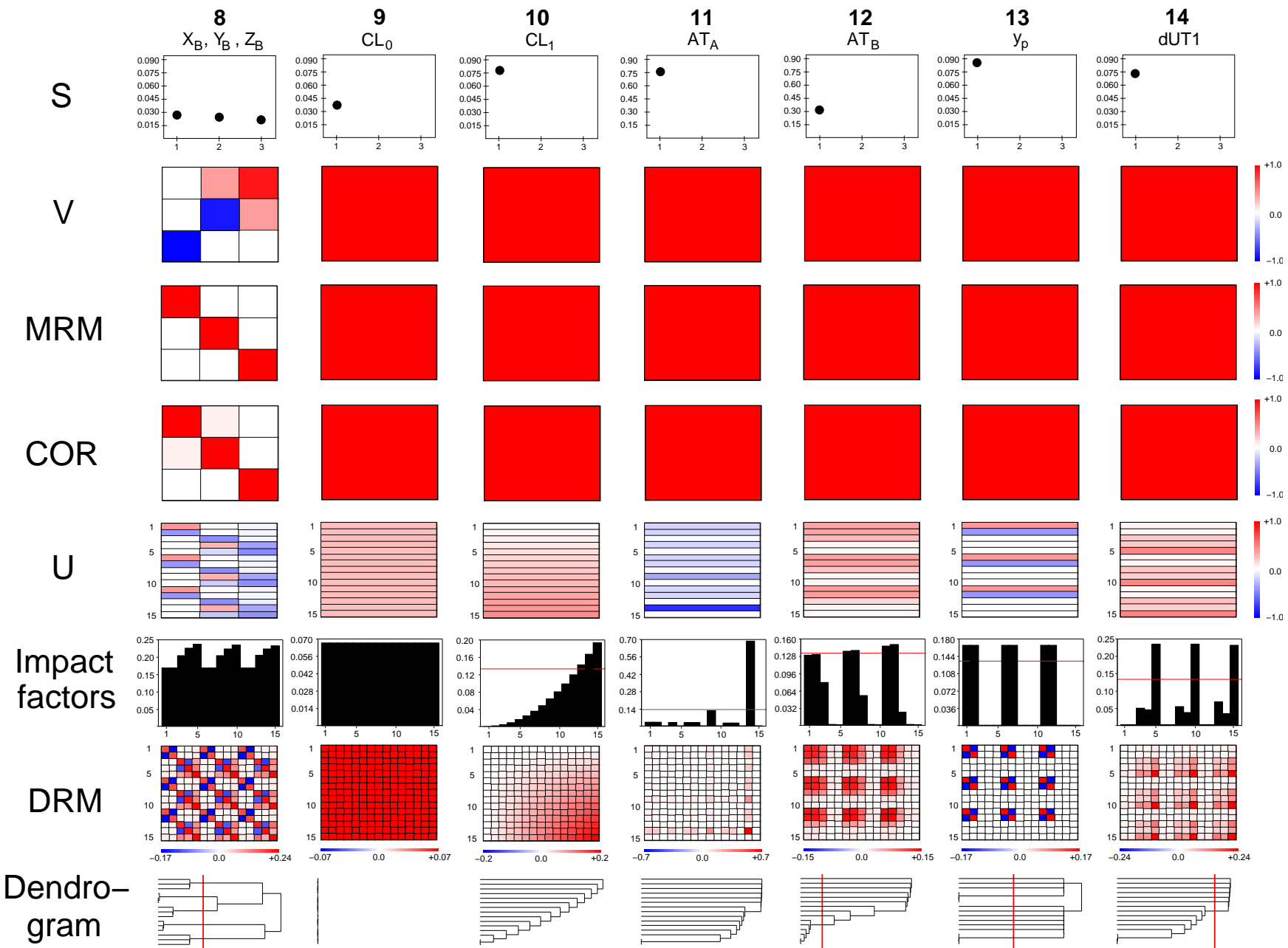
Figure 5.8: Analysis components of a spatial kinematic interferometer (basic parameters), red lines in impact factor diagrams indicate a recommended threshold for 'high-leverage observations' at $2 \cdot \frac{u}{n}$ (with u = number of unknowns, n = number of observations).

Case 8: Site positions x_B, y_B, z_B of station B only

Using the fifteen observations described above, for the determination of the site positions x_B, y_B, z_B of station B, the singular values show that every component can be determined with nearly the same quality. According to the model resolution matrix and to the correlation matrix the three parameters can be well separated.

The left singular vectors (Matrix $\mathbf{U}$) show distinct differences in the impact of each observation onto the three parameters: The first left singular vector $\mathbf{u}_1$ shows (together with the first right singular vector $\mathbf{v}_1$) that the six observations to the polar sources (sources 5 (S0300+85) and 6 (S0300-85)) are almost solely responsible for the determination of z_B. According to the second left singular vector $\mathbf{u}_2$ and the second right singular vector $\mathbf{v}_2$, observations to sources with high right ascensions (source 1 to source 3 (S2230+00, S2255+00, S2330+00)) and to sources 7 and 8 (S0700+00 and S0730+00) are of main importance for the y_B parameter (and to a small amount for x_B). Observations to sources in zenith direction (source 4 (S0300+00)) are (together with observations to sources 7 and 8 and for this baseline geometry) mainly necessary for x_B (see third left singular vector $\mathbf{u}_3$ and third right singular vector $\mathbf{v}_3$).

The impact factors and the elements of the data resolution matrix show significant impact of every observation. The impact factors increase for observations with decreasing availibility of supporting 'counterpart observations', i.e., observations to source 4 (S0300+00) are of main importance since no diametrally opposite observation is possible. The off-diagonal elements of the data resolution matrix (impact co-factors) also show a distinct pattern of three groups with five observations each. Row-wise investigation of the data resolution matrix reflects the supporting character of observations to identical or similar sources.

The cluster analysis of the data resolution matrix also shows the generation of five clusters with each of them containing three observations. Table 5.11 shows the mean impact factors of each cluster onto each parameter together with modification recommendations (as listed in table 3.2 on page 53).

These results can only partly be generalised to other interferometer geometries: Due to the rotation of the interferometer, observations to polar sources are always of main importance for the determination of z coordinates. Statements about the impact of other sources also depend on the baseline orientation with respect to the terrestrial reference system. As a rule of thumb, observations to sources in x axis direction or in y axis direction also account for the x or y component, respectively.

Case 9: Clock offset CL_0 of station B only

As for the clock offset determination in a plane static interferometer (see page 73) also in a spatial kinematic interferometer every observation is of equal importance. Thus, all impact factors and impact co-factors are equal and no cluster analysis can be performed.

As for the plane interferometer this is caused by the fact that the design matrix for this case consists of a vector of ones (see partial derivatives $\frac{\partial \tau_{obs}}{\partial CL0_a}$ and $\frac{\partial \tau_{obs}}{\partial CL0_b}$ in equations (4.17) and (4.20) on page 61) and thus resembles the determination of the arithmetic mean for arbitrarily spaced observations.

The determination of the clock offset of the clock at station B with respect to the reference clock at station A can be interpreted geometrically: As shown in the middle of figure 4.3 on page 63 the clock offset can be visualised as a circle (or sphere) around the station whose clock offset has to be determined. The radius of this circle (or sphere) corresponds to the metric value of the clock offset (i.e., multiplied by the velocity of light).

Code	Source	right ascension [hh min sec]	declination [^{o}]
S1	S2230+00	22 30 0.0	0.0
S2	S2255+00	22 55 0.0	0.0
S3	S2330+00	23 30 0.0	0.0
S4	S0300+00	03 00 0.0	0.0
S5	S0300+85	03 00 0.0	85.0
S6	S0300−85	03 00 0.0	-85.0
S7	S0700+00	07 00 0.0	0.0
S8	S0730+00	07 30 0.0	0.0
S9	S0755+00	07 55 0.0	0.0

Table 5.9: Source list for spatial interferometer investigations

No.	Source	Start of observation [yyyy.doy.hh:mm:ss]	Az. A [^{o}]	Az. B [^{o}]	El. A [^{o}]	El. B [^{o}]
1	S0300+85	2000.265.02:30:00	2	359	4	5
2	S0300−85	2000.265.02:34:00	178	181	4	5
3	S2230+00	2000.265.02:38:00	270	270	49	7
4	S0700+00	2000.265.02:42:00	90	90	5	47
5	S0300+00	2000.265.02:46:00	90	270	66	71
6	S0300+85	2000.265.02:50:00	2	358	5	5
7	S0300−85	2000.265.02:54:00	178	182	5	5
8	S2255+00	2000.265.02:58:00	270	270	50	7
9	S0730+00	2000.265.03:02:00	90	90	1	44
10	S0300+00	2000.265.03:06:00	90	270	70	67
11	S0300+85	2000.265.03:10:00	2	358	5	5
12	S0300−85	2000.265.03:14:00	179	182	5	5
13	S2330+00	2000.265.03:18:00	270	270	54	12
14	S0755+00	2000.265.03:22:00	90	90	0.8	44
15	S0300+00	2000.265.03:26:00	90	270	76	61

Table 5.10: Observation list for spatial interferometer investigations

Cluster Analysis Results - Case 8						
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter: X	Y	Z
1	3	0.23	15, 5, 10	**0.234** ⇒ o	0.001 ⇒ -	0.000 ⇒ -
2	3	0.22	14, 4, 9	0.050 ⇒ o	**0.185** ⇒ o	0.000 ⇒ -
3	3	0.20	13, 3, 8	0.047 ⇒ o	**0.147** ⇒ o	0.000 ⇒ -
4	3	0.17	12, 2, 7	0.002 ⇒ -	0.000 ⇒ -	**0.167** ⇒ o
5	3	0.17	11, 1, 6	0.002 ⇒ -	0.000 ⇒ -	**0.167** ⇒ o

Table 5.11: Cluster analysis results for case 8 (Site positions x_B, y_B, z_B of station B only)

Case 10: Clock rate CL_1 of station B only

In a similar way the clock rate CL_1 can be interpreted as a linear increase or decrease of the radius of a circle (or sphere) around the station whose clock parameter has to be determined.

As shown in figure 5.8 the impact of observations on the clock rate CL_1 is correlated with the time of the observation. This is obviously caused by the structure of the partial derivatives $\frac{\partial \tau_{obs}}{\partial CL1_a}$ and $\frac{\partial \tau_{obs}}{\partial CL1_b}$ (see equations (4.18) or (4.21) on page 61) and resembles the determination of the slope of a regression line through equally spaced observations (i.e., without offset determination).

The data resolution matrix for this case shows the increasing supporting nature of the observations with increasing observation duration. Cluster analysis of this data resolution matrix first clusters observations n and $n-1$, in a second step observations $n, n-1, n-2$, etc. and finally clusters all observations $n, n-1, n-2, \ldots, 1$. For this case, a dendrogram cut is not reasonable.

Case 11: Atmospheric zenith path delay AT_A at station A only

For the determination of the atmospheric zenith path delay AT_A at station A the left singular vector $\mathbf{u}_1$ shows that some observations are of no relevance (e.g. observations 3, 5, 8, 10, 13 and 15). These observations have negligible impact factors. According to table 5.10 on the previous page these observations are performed with high elevations at station A. On the other hand, observations 9 and 14 (to the equatorial sources S0730+00 and S0755+00) are performed with very low elevations (0.8^o and 1.0^o at station A) and thus possess high or very high impact factors. In particular, observation 14 has been observed with an elevation of only $0.86\ [^o]$ and thus possesses an extraordinary high impact on AT_A. A slight decrease in the right ascension of source S0755+00 increases the elevation of this observation and leads to a distinct decrease of this observation (not shown here).

Even in the case of a slight decrease of the right ascension of source S0755+00 the data resolution looks like the data resolution matrix shown in figure 5.10 on the preceding page: Only observations to sources S0700+00, S0730+00 and S0755+00 are of very high importance for AT_A. Cluster analysis of the corresponding data resolution matrix first clusters these observations with a large similarity distance to the next clustering step (see bottom of dendrogram). Due to this extraordinary structure of the data resolution matrix and of the dendrogram a dendrogram cut is not reasonable.

Case 12: Atmospheric zenith path delay AT_B at station B only

In a similar way, for the determination of the atmospheric zenith path delay AT_B at station B observations performed with very low elevations at station B are of main importance. Since observations 1 and 2, 6 and 7 and 11 and 12 possess similar elevations (of $\approx 5[^o]$) they can all be classified as 'high leverage observations' for AT_B (see red line in the corresponding impact factor diagram in figure 5.8 on page 83).

Contrary to the previous case, the data resolution matrix clearly shows the supporting nature of similar observations. As indicated by the dendrogram of the cluster analysis of this data resolution matrix (also see table 5.12 on page 88) observations to polar sources (with elevations below $\approx 10\ [^o]$) are clustered first.

Although one would expect similarities to case 11, this is not the case due to the significant differences in the elevation angles which are the driving factors of the partial derivatives $\frac{\partial \tau_{obs}}{\partial AT}$.

Case 13: Polar motion y_p only

Since the baseline of this interferometer is parallel to the y-axis of the terrestrial reference system this interferometer is insensitive to variations in the x_p component of the polar motion. This parameter can thus not be determined (i.e., the design matrix only contains zeros, the only singular value equals zero and thus indicates that x_p cannot be determined).

The y_p component of polar motion, however, can be determined as the singular value (of ≈ 0.08) reveals. The left singular vector $\mathbf{u}_1$ and thus the impact factors show that only observations to polar sources are needed for the determination of y_p. The remaining observations (to equatorial sources) could have been omitted if only y_p is of interest. The data resolution matrix shows that every observation to the same polar source acts as a supporting observation with equal impact (red squares in the data resolution matrix). Observations to south polar sources contribute to y_p with the same amount (but in opposite direction) as observations to north polar sources (as indicated by blue squares in the data resolution matrix).

After cutting the dendrogram at a reasonable height only two clusters with more than one observation remain (see table 5.13 on the next page). These clusters consist of observations 1, 6 and 11 and observations 2, 7 and 12, respectively. The former cluster contains the observations to the north polar source S0300+85, the latter cluster contains the observations to the south polar source S0300−85.

Case 14: Earth rotation $dUT1$ only

The last basic parameter to be investigated is $dUT1$, i.e., the phase of the rotation of the interferometer. For this parameter observations to polar sources are of no relevance. This is also indicated by the components of the left singular vector $\mathbf{u}_1$ and thus in the impact factors and impact co-factors. In general, only observations to equatorial sources can be used for the determination of the rotational phase. This agrees with the results described in chapter 4. Observations to equatorial sources orthogonal to the baseline (i.e., to source S0300+00 and only for this interferometer geometry) which can be observed with high elevations only are of particular importance and thus exceed the recommended threshold for high-leverage observations (as indicated by the red line in the impact factor diagram of case 14 in figure 5.8 on page 83).

Again, the data resolution matrix shows the supporting nature of observations to source S0300+00 and to the remaining equatorial sources. The cluster analysis of this data resolution matrix yields a dendrogram which first shows the clustering of the three observations to source S0300+00 (see the bottom of the dendrogram). With a large similarity distance, the remaining observations (to equatorial sources) are clustered next. At last the polar observations are clustered. The results are summarized in table 5.14 on the next page.

Cluster Analysis Results - Case 12				
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter: AT_B
1	1	0.08	3	$0.079 \Rightarrow$ o
2	1	0.00	4	$0.002 \Rightarrow$ -
3	1	0.00	5	$0.001 \Rightarrow$ -
4	1	0.06	8	$0.055 \Rightarrow$ o
5	1	0.00	9	$0.002 \Rightarrow$ -
6	1	0.00	10	$0.001 \Rightarrow$ -
7	1	0.02	13	$0.025 \Rightarrow$ -
8	1	0.00	14	$0.002 \Rightarrow$ -
9	1	0.00	15	$0.001 \Rightarrow$ -
10	6	0.14	1, 2, 6, 7, 11, 12	**0.139** $\Rightarrow$ o

Table 5.12: Cluster analysis results for case 12 (Atmospheric zenith path delay AT_B at station B only)

Cluster Analysis Results - Case 13				
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter: y_p
1	1	0.00	3	$0.000 \Rightarrow$ —
2	1	0.00	4	$0.000 \Rightarrow$ —
3	1	0.00	5	$0.000 \Rightarrow$ —
4	1	0.00	8	$0.000 \Rightarrow$ —
5	1	0.00	9	$0.000 \Rightarrow$ —
6	1	0.00	10	$0.000 \Rightarrow$ —
7	1	0.00	13	$0.000 \Rightarrow$ —
8	1	0.00	14	$0.000 \Rightarrow$ —
9	1	0.00	15	$0.000 \Rightarrow$ —
10	3	0.17	1, 6, 11	**0.167** $\Rightarrow$ +
11	3	0.17	2, 7, 12	**0.167** $\Rightarrow$ +

Table 5.13: Cluster analysis results for case 13 (Polar motion y_p only)

Cluster Analysis Results - Case 14				
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter: $dUT1$
1	1	0.00	1	$0.002 \Rightarrow$ -
2	1	0.00	2	$0.002 \Rightarrow$ -
3	1	0.00	6	$0.002 \Rightarrow$ -
4	1	0.00	7	$0.002 \Rightarrow$ -
5	1	0.00	11	$0.002 \Rightarrow$ -
6	1	0.00	12	$0.002 \Rightarrow$ -
7	9	0.11	14, 9, 4, 3, 8, 13, 15, 5, 10	**0.110** $\Rightarrow$ +

Table 5.14: Cluster analysis results for case 14 (Earth rotation $dUT1$ only)

5.4.2 Estimability investigations of composed parameter sets

The functional models for data analyses of real VLBI sessions always have to include more than the individual basic parameters discussed so far. Besides the main geophysical parameters of interest (such as site positions or earth orientation parameters) at least the clock offsets of all atomic clocks with respect to one reference clock have to be included. For a single-baseline session the following cases are of interest:

15. Estimation of site coordinates x_B, y_B, z_B and clock offset CL_0

16. Estimation of clock offset CL_0 and earth rotation $dUT1$

17. Estimation of clock offset CL_0, clock rate CL_1 and earth rotation $dUT1$

18. Estimation of atmospheric zenith path delay AT_A, clock parameters CL_0, CL_1 and earth rotation $dUT1$

19. Estimation of clock parameters CL_0, CL_1, atmospheric zenith path delay AT_B and earth rotation $dUT1$

20. Estimation of atmospheric zenith path delay AT_A, clock parameters CL_0, CL_1, atmospheric zenith path delay AT_B and earth rotation $dUT1$

Cases 17 and 20 show the most common parametrisations for single-baseline sessions such as INTENSIVE sessions (see e.g. FISCHER 2006). The remaining cases are either needed for a deeper insight or since they are components of cases 17 and 20. As for the basic parameters, cases 15 to 20 will be analysed by investigating the analysis components displayed in figure 5.9 on the following page.

Case 15: Site coordinates x_B, y_B, z_B and clock offset CL_0

Using the same observations as in cases 8 to 14, site positions x_B, y_B, z_B of station B and a constant bias CL_0 of the frequency standard of station B with respect to the clock at station A can be estimated. As shown by the singular values in figure 5.9 every parameter can be determined well. The **V**-matrix, however, already shows that the first parameter (x_B) and the last parameter (CL_0) are together affected by the same observations: Hence, the observations with large components in the first left singular vector $\mathbf{u}_1$ and in the last left singular vector $\mathbf{u}_4$ (mainly observations to source S0300+00 and to the polar sources) both affect x_B and CL_0.

The correlation matrix also shows a large positive correlation (of 0.78) between x_B and CL_0. This can again be interpreted geometrically: With this network geometry and this baseline orientation no separation between a variation in x_B and a variation in CL_0 (when considered as the radius of a sphere around station B) is possible due to the lack of supporting observations for the zenith observations to source S0300+00. Supporting observations would be conducted to the diametrally opposite source S1800+00 which is not visible during this session.

The most important observations are conducted to polar sources S0300+85 and S0300-85. As shown in table 5.15 the dendrogram cut reveals that these observations are of main importance for both z_B and CL_0. This confirms the results given in NOTHNAGEL 1991 where observations to sources in elevation cusps have been recognized as important for the clock offset determination. The data resolution matrix and the cluster analysis results show the generation of five groups with three observations each. These groups consist of observations to either the same source or to neighbouring sources which obviously have the same information content.

As for case 8, zenith observations to source S0300+00 are of main importance for x_B and observations to the remaining equatorial sources are of importance for y_B.
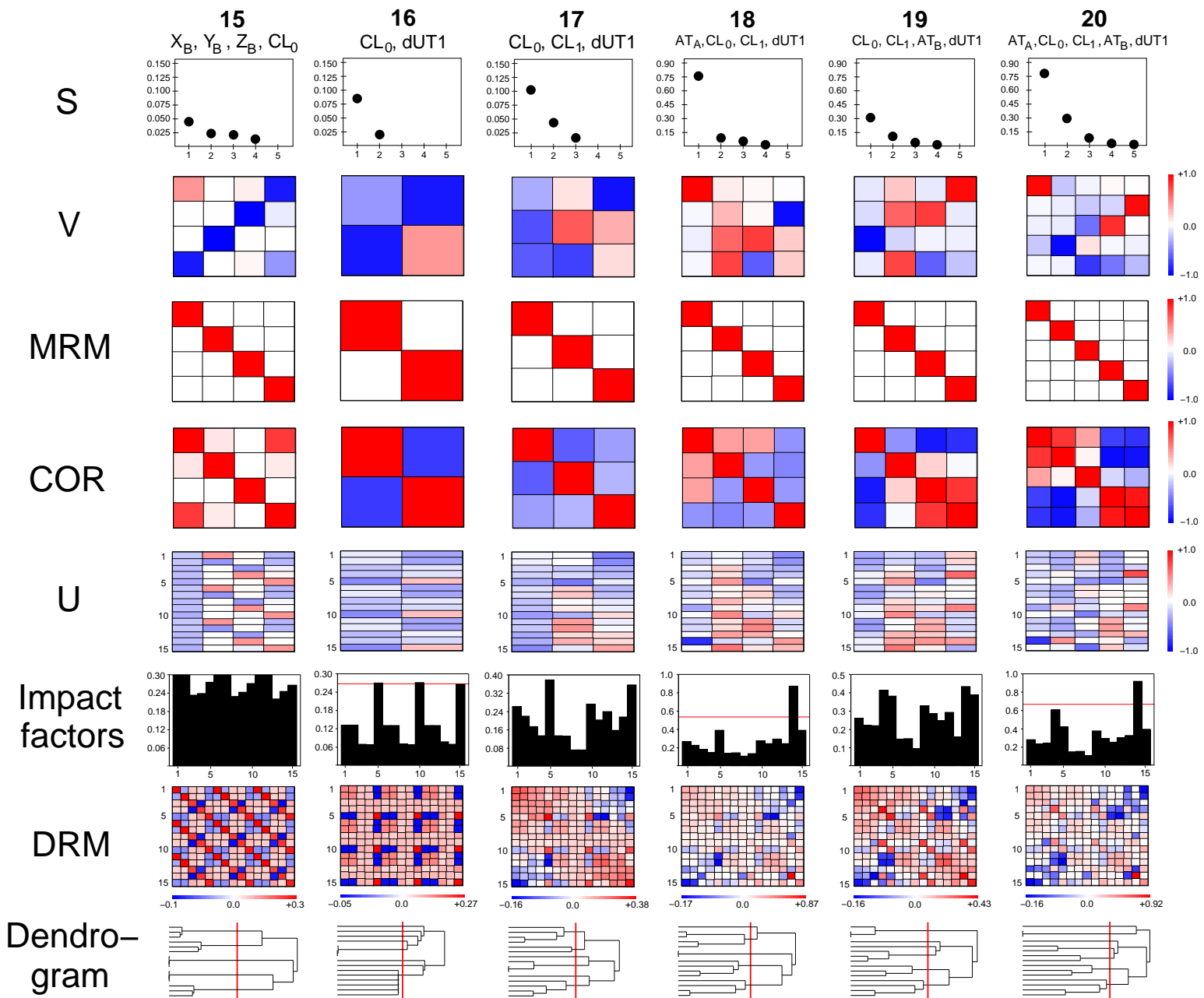
Figure 5.9: Analysis components of a spatial kinematic interferometer (composed parameters), red lines in impact factor diagrams indicate a recommended threshold for 'high-leverage observations' at $2 \cdot \frac{u}{n}$ (with u = number of unknowns, n = number of observations).

Case 16: Clock offset CL_0 and earth rotation $dUT1$

A common determination of both the rotation of the interferometer (i.e., $dUT1$) and the constant clock offset CL_0 is complicated by the high correlation of both parameters (correlation coefficient ≈ -0.77). In addition, the right singular vectors do not allow a clear assignment of observations to only one parameter. Nevertheless, high impact and a similar information content of observations to source S0300+00 and to the polar sources S0300+85 and S0300-85 is shown in the data resolution matrix. After the dendrogram cut the cluster analysis results (see table 5.16 on the next page) show that observations to source S0300+00 are still of main importance for $dUT1$ and observations to polar sources (and thus observations to sources in elevation cusps) are of main importance for the clock offset CL_0.

Case 17: Clock offset CL_0, clock rate CL_1 and earth rotation $dUT1$

The separability of the three parameters CL_0, CL_1 and $dUT1$ is also complicated by high correlations of up to -0.63 (between clock offset CL_0 and clock rate CL_1). Again, a clear assignment of observations to parameters is difficult due to the non-diagonal structure of the **V**-matrix.

For this case and for the following cases, no clear (or regular) pattern in the data resolution matrices and thus in the impact factors can be seen. A human interpretation is complicated and so the strengths of the automatic, cluster analysis-based analysis methods become obvious.

Although the dendrogram of the cluster analysis of the data resolution matrix of this case shows the presence of three main groups, for a better separation six clusters have been formed. Table 5.17 on the following page shows that for this parametrisation the two observations 8 and 9 to equatorial sources could have been neglected. On the other hand, observations 5, 10 and 15 (to source S0300+00) are still of main importance for the determination of the rotation of the interferometer (i.e., for $dUT1$).

For the clock offset determination observations 1 to 4 are most important. As table 5.10 on page 85 shows, these observations are performed into each direction of the celestial sphere and thus 'scan' the sky in diametral directions. Geometrically, for the radius determination of a sphere around station B observations into diametral directions are needed.

For the clock rate determination the same observation constellation has to be repeated at the end of the session. As table 5.17 shows, the last observations performed into each direction of the celestial sphere are needed for the determination of radius variations of the sphere around station B.

Case 18: Atmospheric zenith path delay AT_A, clock parameters CL_0, CL_1 and $dUT1$

After including the atmospheric zenith path delay AT_A at station A a distinct difference in the singular values can be seen. In connection with the first right singular vector, this means that AT_A can be determined with the lowest variance. In addition, the separability of all parameters has been improved since the absolute values of all correlations are below 0.5.

As the right singular vectors $\mathbf{v}_1$ and $\mathbf{v}_4$ show, only for the atmospheric zenith path delay AT_A and for the clock offset CL_0 clear relations between observations and parameters can be found: The elements of the first left singular vector $\mathbf{u}_1$ (which mainly affect AT_A) show a clear correlation with the elevations of the observations at station A (see table 5.10 on page 85). The smaller the elevations, the larger the elements of $\mathbf{u}_1$ and thus the larger the impact on AT_A.

For the clock offset determination observations 1 to 4 are of main importance. Again, these are observations in every direction of the celestial sphere.

As in the previous case, the data resolution matrix and thus the impact factors do not show a clear pattern. Due to its low elevation at station A observation 14 has again a very high impact on AT_A (cf. case 11) as well as observations 5, 10, 13 and 15, which are of main importance for the determination of $dUT1$ (cf. case 14). This can also be seen in the cluster analysis results showed in table 5.18 on the next page.

	Cluster Analysis Results - Case 15						
Cluster:	No. of members:	Mean impact factor:	Members:	Impact on parameter:			
				X	Y	Z	CL_0
1	3	0.30	6, 1, 11	$0.064 \Rightarrow$ o	$0.001 \Rightarrow$ -	$\mathbf{0.167} \Rightarrow$ o	$\mathbf{0.129} \Rightarrow$ o
2	3	0.30	2, 7, 12	$0.064 \Rightarrow$ o	$0.001 \Rightarrow$ -	$\mathbf{0.167} \Rightarrow$ o	$\mathbf{0.129} \Rightarrow$ o
3	3	0.27	15, 5, 10	$\mathbf{0.202} \Rightarrow$ o	$0.002 \Rightarrow$ -	$0.000 \Rightarrow$ -	$0.035 \Rightarrow$ o
4	3	0.24	4, 9, 14	$0.002 \Rightarrow$ -	$\mathbf{0.174} \Rightarrow$ o	$0.000 \Rightarrow$ -	$0.018 \Rightarrow$ o
5	3	0.23	13, 3, 8	$0.001 \Rightarrow$ -	$\mathbf{0.155} \Rightarrow$ o	$0.000 \Rightarrow$ -	$0.022 \Rightarrow$ o

Table 5.15: Cluster analysis results for case 15 (Site coordinates x_B, y_B, z_B and clock offset CL_0)

	Cluster Analysis Results - Case 16				
Cluster:	No. of members:	Mean impact factor:	Members (observations):	Impact on parameter:	
				CL_0	$dUT1$
1	1	0.07	3	$0.018 \Rightarrow$ -	$0.001 \Rightarrow$ -
2	1	0.07	4	$0.022 \Rightarrow$ -	$0.000 \Rightarrow$ —
3	1	0.07	8	$0.015 \Rightarrow$ -	$0.003 \Rightarrow$ -
4	1	0.07	9	$0.029 \Rightarrow$ -	$0.000 \Rightarrow$ —
5	1	0.08	13	$0.008 \Rightarrow$ -	$0.009 \Rightarrow$ -
6	1	0.07	14	$0.034 \Rightarrow$ -	$0.001 \Rightarrow$ —
7	3	0.27	15, 5, 10	$0.035 \Rightarrow$ o	$\mathbf{0.202} \Rightarrow$ ++
8	6	0.13	7, 6, 11, 12, 1, 2	$\mathbf{0.128} \Rightarrow$ o	$0.063 \Rightarrow$ -

Table 5.16: Cluster analysis results for case 16 (Clock offset CL_0 and earth rotation $dUT1$)

	Cluster Analysis Results - Case 17					
Cluster:	No. of members:	Mean impact factor:	Members:	Impact on parameter:		
				CL_0	CL_1	$dUT1$
1	1	0.07	8	$0.011 \Rightarrow$ -	$0.000 \Rightarrow$ —	$0.003 \Rightarrow$ -
2	1	0.07	9	$0.008 \Rightarrow$ -	$0.004 \Rightarrow$ -	$0.001 \Rightarrow$ —
3	3	0.33	15, 5, 10	$0.044 \Rightarrow$ o	$0.066 \Rightarrow$ o	$\mathbf{0.194} \Rightarrow$ o
4	4	0.20	4, 3, 1, 2	$\mathbf{0.161} \Rightarrow$ o	$0.097 \Rightarrow$ o	$0.017 \Rightarrow$ o
5	4	0.20	13, 14, 11, 12	$0.010 \Rightarrow$ o	$\mathbf{0.102} \Rightarrow$ o	$0.059 \Rightarrow$ o
6	2	0.13	6, 7	$0.082 \Rightarrow$ o	$0.001 \Rightarrow$ -	$0.056 \Rightarrow$ o

Table 5.17: Cluster analysis results for case 17 (Clock offset CL_0, clock rate CL_1 and $dUT1$)

	Cluster Analysis Results - Case 18						
Cluster:	No. of members:	Mean impact:	Members:	Impact on parameter:			
				AT_A	CL_0	CL_1	$dUT1$
1	1	0.10	8	$0.034 \Rightarrow$ -	$0.028 \Rightarrow$ -	$0.003 \Rightarrow$ -	$0.001 \Rightarrow$ —
2	2	0.50	9, 14	$\mathbf{0.357} \Rightarrow$ o	$\mathbf{0.080} \Rightarrow$ o	$0.002 \Rightarrow$ -	$0.026 \Rightarrow$ o
3	4	0.20	4, 3, 1, 2	$0.006 \Rightarrow$ o	$\mathbf{0.128} \Rightarrow$ o	$\mathbf{0.092} \Rightarrow$ o	$0.012 \Rightarrow$ o
4	4	0.32	5, 10, 13, 15	$0.031 \Rightarrow$ o	$0.018 \Rightarrow$ o	$\mathbf{0.092} \Rightarrow$ o	$\mathbf{0.118} \Rightarrow$ o
5	4	0.20	6, 7, 11, 12	$0.027 \Rightarrow$ o	$0.057 \Rightarrow$ o	$0.065 \Rightarrow$ o	$0.107 \Rightarrow$ o

Table 5.18: Cluster analysis results for case 18 (Atmospheric zenith path delay AT_A, clock parameters CL_0, CL_1 and $dUT1$)

Cluster Analysis Results - Case 19							
Cluster:	No. of members:	Mean impact:	Members:	Impact on parameter:			
				CL_0	CL_1	AT_B	$dUT1$
1	1	0.09	8	$0.009 \Rightarrow$ -	$0.000 \Rightarrow$ —	$0.023 \Rightarrow$ -	$0.024 \Rightarrow$ -
2	1	0.16	13	$0.006 \Rightarrow$ -	$0.081 \Rightarrow$ -	$0.001 \Rightarrow$ —	$0.001 \Rightarrow$ —
3	3	0.39	4, 9, 14	$\mathbf{0.249} \Rightarrow$ o	$0.069 \Rightarrow$ o	$\mathbf{0.251} \Rightarrow$ o	$\mathbf{0.174} \Rightarrow$ o
4	3	0.35	5, 10, 15	$0.039 \Rightarrow$ o	$0.068 \Rightarrow$ o	$0.016 \Rightarrow$ o	$\mathbf{0.126} \Rightarrow$ o
5	3	0.23	3, 1, 2	$0.022 \Rightarrow$ o	$\mathbf{0.094} \Rightarrow$ o	$0.016 \Rightarrow$ o	$0.023 \Rightarrow$ o
6	4	0.21	6, 7, 11, 12	$0.014 \Rightarrow$ o	$0.056 \Rightarrow$ o	$0.032 \Rightarrow$ o	$0.001 \Rightarrow$ -

Table 5.19: Cluster analysis results for case 19 (Clock parameters CL_0, CL_1, atmospheric zenith path delay AT_B and $dUT1$)

Case 19: Clock parameters CL_0, CL_1, atmospheric zenith path delay AT_B and $dUT1$

Compared to case 18, estimating the atmospheric zenith path delay AT_B at station B instead of AT_A at station A yields very different relations between observation groups and affected parameters: Although every parameter can be estimated (i.e., all singular values > 0) the absolute values of the correlation coefficients between some parameters are above 0.8. Only the clock rate CL_1 can be well separated from the remaining parameters. Especially high correlations between clock offset CL_0, AT_B and $dUT1$ and between AT_B and $dUT1$ complicate the interpretation of the cluster analysis results. Thus, some clusters commonly affect several parameters (see e.g. cluster 3 and cluster 4 in table 5.19).

In addition, none of the impact factors exceeds the 'high-leverage threshold' (of $2 \cdot \dfrac{u}{n}$). However, as the mean impact of each cluster in table 5.19 clearly show, the most important observations for this parametrisation are those performed to sources S0700+00, S0730+00, S0755+00 and S0300+00.

Case 20: Atmospheric zenith path delay AT_A, clock parameters CL_0, CL_1, atmospheric zenith path delay AT_B and $dUT1$

The parameters AT_A, CL_0, CL_1, AT_B and $dUT1$ depict the most common parametrisation for single-baseline VLBI networks with a large east-west extension. As the singular values and the first two right singular vectors show, the two atmospheric zenith path delays AT_A and AT_B are best determined. The parameter with the lowest accuracy (or highest variance) is again the clock offset CL_0.

Again, the separability of the five parameters is weak. Except for the correlations of the clock rate CL_1 with the remaining parameters all other (absolute values of the) correlation coefficients are above 0.8. This also complicates the interpretation of the cluster analysis results.

The two largest elements of the **U**-matrix belong to observations 4 and 14. Thus, the impact factors for these observations show the large importance of these observations. This is again caused by the fact that equatorial sources are needed for most of these parameters and by the low elevation of observation 14.

Cluster analysis of the data resolution matrix shows that observation 8 and observations 6, 7, 11 and 12 are candidates for negligible observations. Among the most important observations are again observations 4 and 14. These observations are mainly responsible for the determination of AT_A, CL_0, AT_B and $dUT1$ (see table 5.20 on the next page).

In contrast to cases 15 and 16 for the clock offset determination observations to equatorial sources are mainly needed. In this case, observations to sources in elevation cusps are not of high importance for the determination of CL_0.

Cluster Analysis Results - Case 20								
Cl.:	No. of mbrs.:	Mean impact:	Members:	Impact on parameter:				
				AT_A	CL_0	CL_1	AT_B	$dUT1$
1	1	0.10	8	$0.012 \Rightarrow$ -	$0.001 \Rightarrow$ —	$0.003 \Rightarrow$ —	$0.000 \Rightarrow$ —	$0.000 \Rightarrow$ —
2	1	0.92	14	$\mathbf{0.484} \Rightarrow$ o	$0.103 \Rightarrow$ -	$0.000 \Rightarrow$ —	$0.048 \Rightarrow$ -	$0.082 \Rightarrow$ -
3	2	0.49	4, 9	$0.118 \Rightarrow$ o	$\mathbf{0.361} \Rightarrow$ o	$0.015 \Rightarrow$ o	$\mathbf{0.349} \Rightarrow$ o	$\mathbf{0.250} \Rightarrow$ o
4	2	0.35	5, 10	$0.021 \Rightarrow$ o	$0.027 \Rightarrow$ o	$0.065 \Rightarrow$ o	$0.026 \Rightarrow$ o	$0.105 \Rightarrow$ o
5	2	0.35	13, 15	$0.083 \Rightarrow$ o	$0.046 \Rightarrow$ o	$\mathbf{0.138} \Rightarrow$ o	$0.042 \Rightarrow$ o	$0.059 \Rightarrow$ o
6	3	0.25	3, 1, 2	$0.016 \Rightarrow$ o	$0.008 \Rightarrow$ o	$\mathbf{0.110} \Rightarrow$ o	$0.029 \Rightarrow$ o	$0.026 \Rightarrow$ o
7	4	0.21	6, 7, 11, 12	$0.003 \Rightarrow$ -	$0.001 \Rightarrow$ -	$0.058 \Rightarrow$ o	$0.008 \Rightarrow$ o	$0.003 \Rightarrow$ -

Table 5.20: Cluster analysis results for case 20 (Atmospheric zenith path delay AT_A, clock parameters CL_0, CL_1, atmospheric zenith path delay AT_B and $dUT1$)

5.4.2.1   Effect of omitting observations on the cofactors of the estimated parameters

As shown in section 2.4.3 on page 42, the impact factors play a crucial role in the determination of the *increase of uncertainty* and thus on the cofactors of the estimated parameters. In order to present some practical applications, for cases 14 and 20 up to five observations (both important and less important observations) will be omitted and the changes in the cofactors of the respective parameters (with respect to the cofactors obtained by using all observations) are analysed.

Increase of uncertainty for case 14 (estimation of $dUT1$ only)

As shown in the impact factor plot (also see figure 5.8 on page 83), for the determination of $dUT1$, observations 5, 10 and 15 (i.e., observations to the equatorial source S4, which is orthogonal to the current baseline) and observations 13 and 8 are of main importance. On the other hand, observations 1, 2 and 6, 7 and 11, 12 (to polar sources S5 or S6) could be neglected, if only $dUT1$ is of interest.

Figure 5.10 on the facing page shows the effect of successively omitting the five most important observations and the five least important observations on the cofactor of $dUT1$. Although neglecting an observation always leads to an increase of the corresponding cofactor, it can be seen that neglecting observations with large impact factors has a significantly higher impact on the cofactor of $dUT1$ than the omission of observations with low impact factors.

Increase of uncertainty for case 20 (estimation of AT_A, CL_0, CL_1, AT_B and $dUT1$)

Omitting the five most important observations and the five least important observations of case 20 shows a similar effect (see figure 5.12 on page 96): For every parameter, large cofactor changes can be seen if the first most important observations have been neglected. Similar to case 14, omitting observations with low impact factors only has a small impact on the cofactors of the estimated parameters.

For some parameters (such as CL_1) the effect of neglecting 33% of the most important observations is almost equal to the effect of neglecting 33% of the least important observations (see modification numbers 9 and 10). This also depicts the general sensitivity of some parameters to changes in the observation structure.

In general, both case 14 and case 20 show that observations with large impact factors have to be obtained carefully since the (purely geometric) effect of the observations on the cofactors would be amplified if large formal errors are present. Thus, for the *increase of uncertainty* both the geometry of the experiment design and the observational error have to be taken into account (see also FÖRSTNER 1992).

Obviously, omitting an observation changes the impact factors of the remaining observations and thus also changes the entire impact co-factor situation. Figure 5.11 on the facing page shows the effect of omitting observation 14 (of case 20). Consequently, after omitting observations the regression diagnostics tool has to be applied to the new situation, i.e., to the modified design matrix.

Increase of uncertainty (case 14)

Impact factors

Modification number:	Observations omitted:
1	5
2	1
3	5, 10
4	1, 6
5	5, 10, 15
6	1, 6, 11
7	5, 10, 15, 13
8	1, 6, 11, 12
9	5, 10, 15, 13, 8
10	1, 6, 11, 12, 7
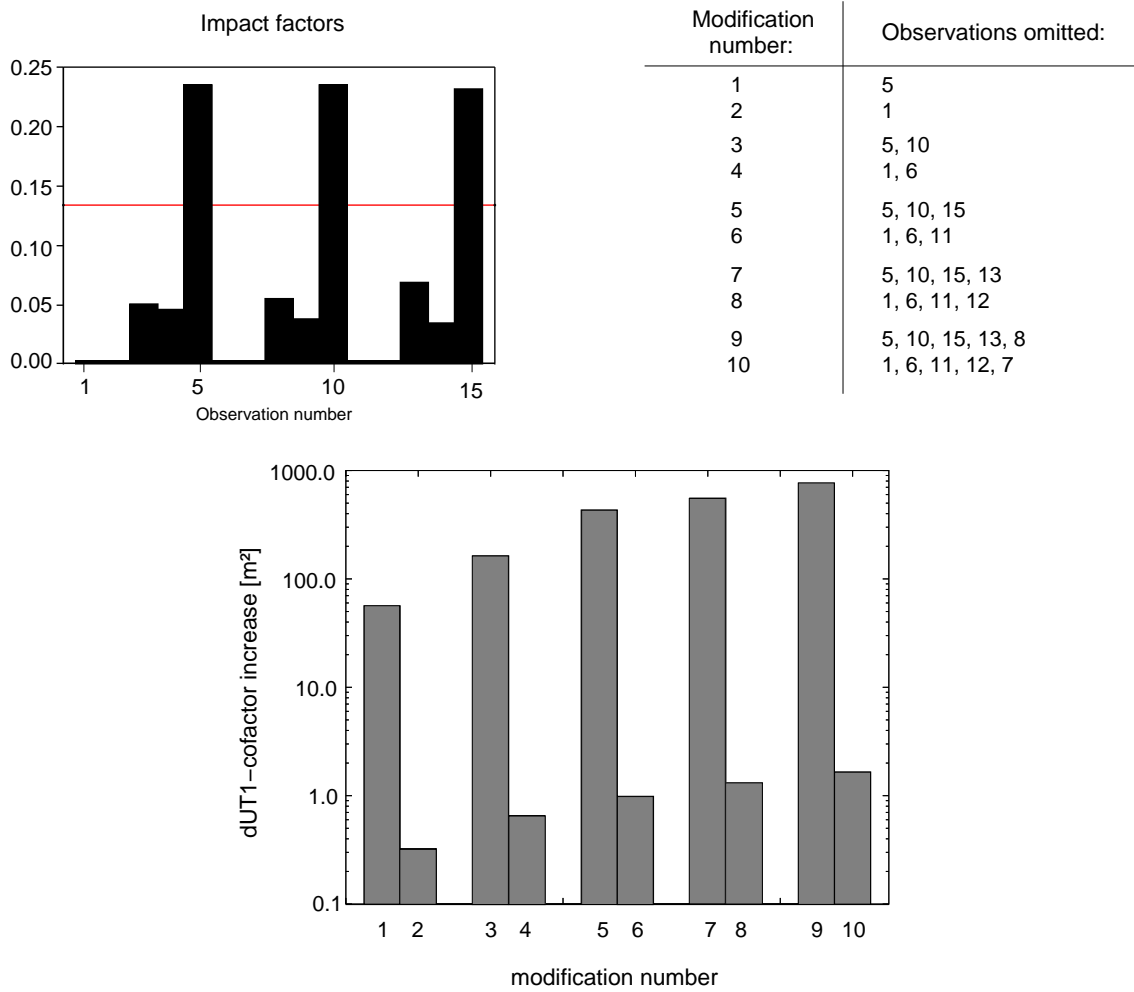
Figure 5.10: Increase of uncertainty after omission of observations: cofactor increase with respect to the cofactors computed by using all observations (for case 14, logarithmic scale for cofactor plot).

Impact factors (all observations)

Impact factors (observation 14 omitted)
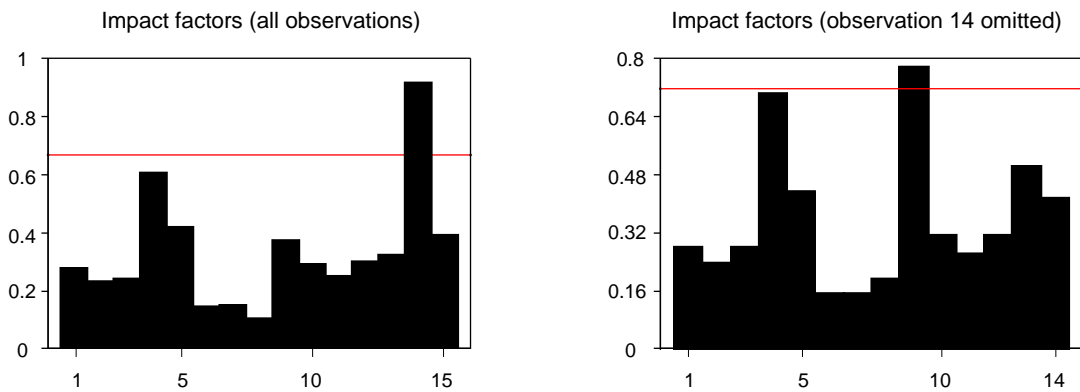
Figure 5.11: Impact factor changes after omission of observation 14 (of case 20)

Increase of uncertainty (case 20)
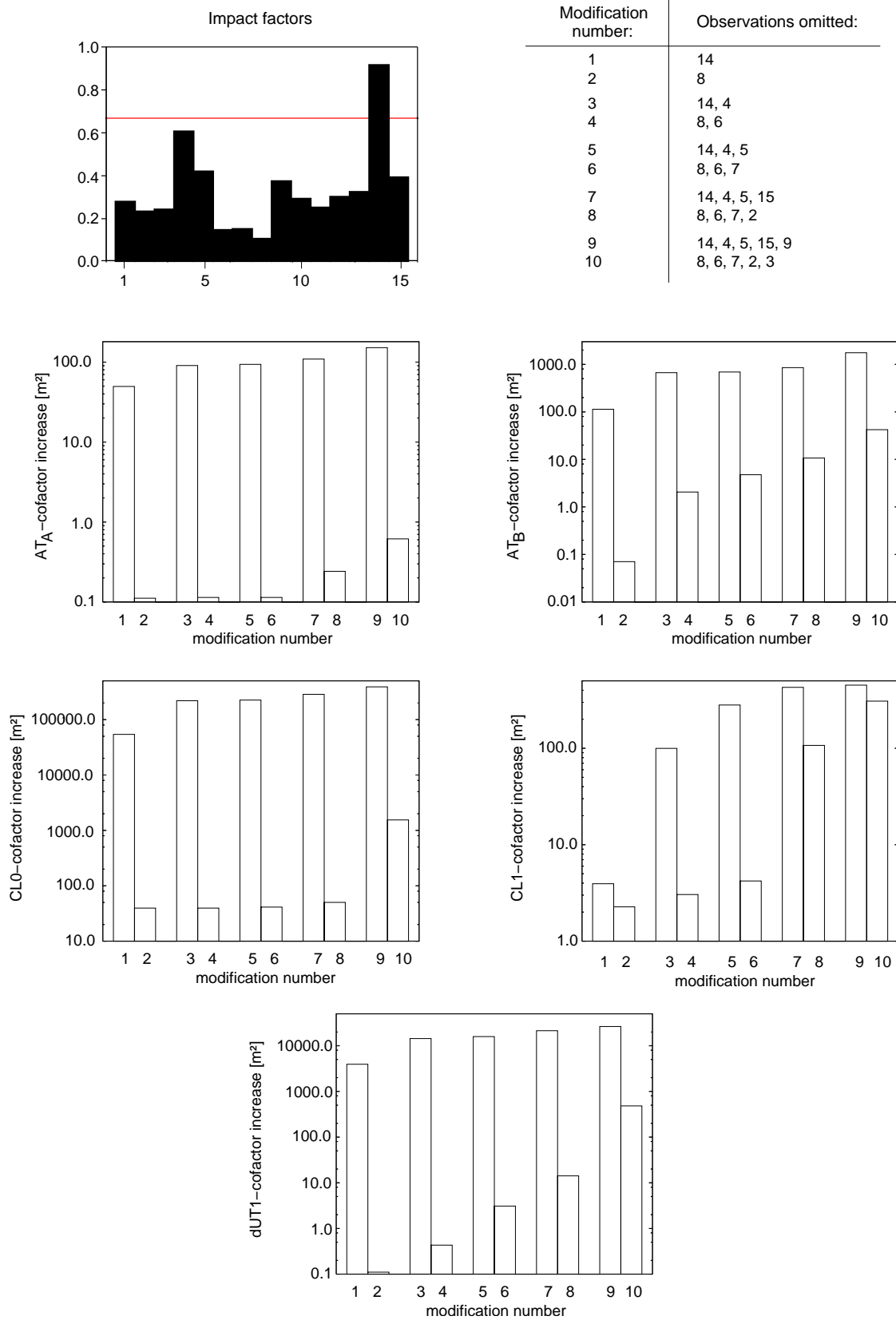
Figure 5.12: Increase of uncertainty after omission of observations: cofactor increase with respect to the cofactors computed by using all observations (for case 20, logarithmic scale for cofactor plots).

5.4.2.2 Effects of reordering observations on impact factors

For short observation durations (as considered here), reordering of observations does not significantly change the impact factors. If the basic structure of five observations of the original observation schedule (which is repeated three times, see table 5.10 on page 85) is changed to

- Source S1 (or S2 or S3, depending on hour angle and visibility),

- Source S7 (or S8 or S9, depending on hour angle and visibility),

- Source S4,

- Source S5,

- Source S6,

almost identical impact factors can be seen (for an example see figure 5.13). This is obviously caused by the fact that the topocentric observation geometry (i.e., azimuth and elevation of the radio telescope when pointing at a source) change only slightly. Of course, this changes, the longer the session duration and the larger the differences between the observation times of the exchanged observations.
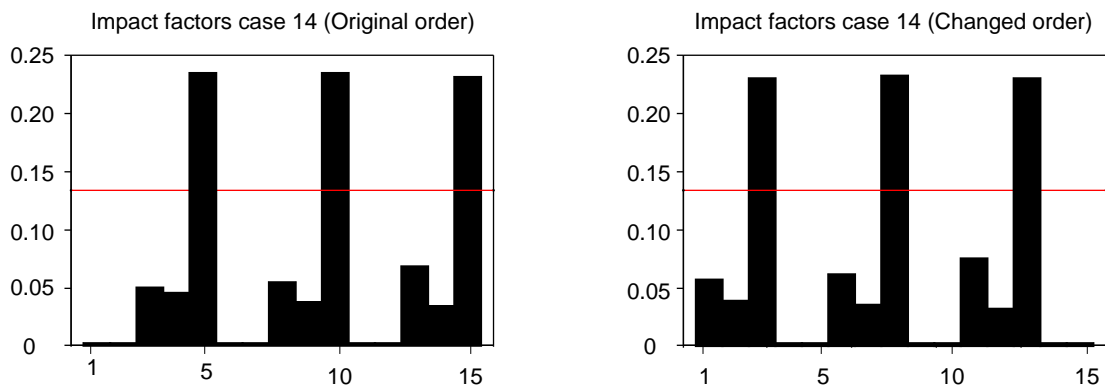
Figure 5.13: Impact factor changes after reordering of observations (for case 14)

5.4.2.3 Conclusions from spatial kinematic interferometer investigations

The main intention of the previous sections was to extend the results obtained from the analyses of plane static interferometers to spatial kinematic interferometers. As for the plane interferometer, applying the regression diagnostics tool to the determination of basic parameters (i.e., site positions x, y, z, clock parameters CL_0 and CL_1, atmospheric zenith path delays AT_A and AT_B, polar motion x_p and y_p or earth rotation $dUT1$) yields geometrically comprehensible results which agree with (or supplement) the theoretical considerations in chapter 4:

- For the determination of site positions x, y, z, observations to sources lying approximately in the direction of the axis of the coordinate system are needed.

- For the sole determination of the clock offset (CL_0), every observation is of equal importance. For the clock rate (CL_1), observations at the end of the observing session are most important.

- As expected (see chapter 4), observations performed with low elevations are needed in particular for the determination of atmospheric zenith path delays AT_A and AT_B.

- In agreement with the analysis of the partial derivatives in chapter 4, observations to polar sources are mainly needed for the determination of polar motion x_p and y_p. Observations to equatorial sources are of no relevance for these parameters.

- On the other hand, for the sole determination of earth rotation $dUT1$, observations to equatorial sources lying orthogonal to the baseline (at the time of the observation) are needed. Observations to polar sources can be omitted. This also agrees with the theoretical considerations of chapter 4.

For composed parameter sets (and thus for more complex and geometrically less comprehensible observation configurations), the strengths of the regression diagnostics tool become even more obvious. In these cases, the sole investigation of partial derivatives does not reveal the overall effects of all available observations on the entire parameter set. Here, the regression diagnostics tool based on singular value decomposition and cluster analysis provides a more detailed insight into the adjustment problem. For the current spatial kinematic interferometer it could be shown that

- for certain parametrisations (see e.g. case 15: x, y, z and CL_0 and case 16: CL_0 and $dUT1$) groups of observations to sources in elevation cusps are needed for the determination and separation of the clock offset CL_0 (see NOTHNAGEL 1991).

- for the common estimation of clock offset CL_0, clock rate CL_1 and earth rotation $dUT1$ (case 17) different observations are needed: Here, the group of observations 1 to 4 (i.e., observations into every direction of the mutually visible part of the celestial sphere) is responsible for the clock offset determination. The same observation constellation is needed at the end of the observing session in order to determine the clock rate parameter. For $dUT1$, still observations to equatorial sources are of main importance.

- for more complex parametrisations and for data resolution matrices without an obvious regular pattern, the regression diagnostics tool still recognizes groups of observations. The interpretation, however, is complicated by the unavoidable increase of complexity in the relations between the parameters involved.

Furthermore, it could be shown that the impact factors also express the *increase of uncertainty*, i.e., the effect of omitting observations on the cofactors of the estimated parameters: The higher the impact factor of an observation i, the higher the effect of omitting the ith observation on the formal error(s) of the estimated parameter(s).

In summary, it could be shown that the regression diagnostics tool developed in the first chapters yields both plausible and (geometrically) comprehensible results. In addition to the verification of knowledge based on the analysis of partial derivatives of single observations (as performed in chapter 4), new findings arose from investigating the entire design matrix, i.e., by analysing the geometry of the entire observing session.

After investigating artificial interferometers, the regression diagnostics tool will now be applied to a real, single-baseline observing session.

5.4.3 Estimability investigations for a real, single-baseline VLBI session

As a final application example a real single-baseline VLBI session is being investigated. Therefore, an arbitrary INTENSIVE2-session (IVS-code: K05072, from march 13, 2005) has been chosen. The two stations involved are Wettzell (Germany) and Tsukuba (Japan) which form a baseline with a length of 8445 km. Due to the long east-west-extension of this baseline, this session type is especially sensitive for the determination of earth rotation variations. During the session duration of approximately one hour, 29 observations (to 16 sources) have been generated (see table 5.22 on page 102). The observation schedule for this session has been generated at the Geodetic Institute of the University of Bonn using the schedule generation software SKED and with the optimisation criterion of maximum sky coverage (FISCHER 2006). As for all real VLBI sessions, the sources are not distributed homogeneously and thus the observations do not show such a regular pattern as e.g. in the previous application example (see table 5.21 on the following page[3]). Both the network geometry and the source distribution are shown in figure 5.14.
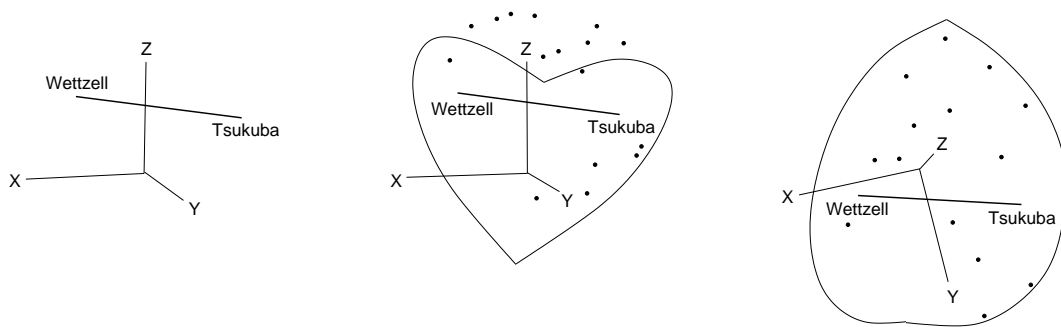
Figure 5.14: INTENSIVE2-session geometry (left: network geometry (near equatorial view), middle: network geometry and source distribution (near equatorial view), right: network geometry and source distribution (polar view))

As for the fictitious interferometer in the previous application example, a variety of parameters (and parameter sets) can be estimated. In the following, these parameter sets will be analysed:

21. Estimation of clock offset CL_0 of station Wettzell with respect to station Tsukuba only

22. Estimation of clock rate CL_1 of station Wettzell with respect to station Tsukuba only

23. Estimation of clock offset CL_0 and clock rate CL_1 of station Wettzell with respect to station Tsukuba

24. Estimation of atmospheric zenith path delay $AT_{Wettzell}$ at station Wettzell only

25. Estimation of atmospheric zenith path delay $AT_{Tsukuba}$ at station Tsukuba only

26. Estimation of earth rotation $dUT1$ only

27. Estimation of atmospheric zenith path delay $AT_{Wettzell}$, clock parameters CL_0, CL_1, atmospheric zenith path delay $AT_{Tsukuba}$ and earth rotation $dUT1$

Since some of these cases have already been treated in the previous application examples, some interpretations can be kept brief. Case 27 depicts the typical parametrisation as used in routine data analysis of single-baseline VLBI sessions for earth rotation determination and thus forms the most important case. The analysis components are shown in figure 5.16 on page 105.

---

[3]The first four characters of the source name indicate the (approximate) right ascension (in [hh min]), the fifth character indicates the sign of the declination and the sixth and seventh character indicate the (approximate) declination of the source (in degrees).

No.	Source	right ascension [hh min sec]	declination [^{o}]
1	0059+581	1 2 45.7	58.24
2	0106+013	1 8 38.7	1.35
3	0119+115	1 21 41.5	11.49
4	0133+476	1 36 58.5	47.51
5	0229+131	2 31 45.8	13.22
6	0235+164	2 38 38.9	16.36
7	0602+673	6 7 52.6	67.20
8	0804+499	8 8 39.6	49.50
9	0955+476	9 58 19.6	47.25
10	1044+719	10 48 27.6	71.43
11	1128+385	11 30 53.2	38.15
12	1300+580	13 2 52.4	57.48
13	1357+769	13 57 55.3	76.43
14	1803+784	18 0 45.6	78.28
15	1807+698	18 6 50.6	69.49
16	2037+511	20 38 37.0	51.19

Table 5.21: Source list for real interferometer investigations

Case 21: Estimation of clock offset CL_0 of station Wettzell with respect to station Tsukuba only

As for the clock offset determination in a plane interferometer and a spatial, kinematic interferometer (cases 3 and 9) for the solely determination of the clock offset CL_0 of station Wettzell with respect to station Tsukuba every observation is of the same importance and no dendrogram can be formed.

Case 22: Estimation of clock rate CL_1 of station Wettzell with respect to station Tsukuba only

In a similar way, for the solely determination of the clock rate CL_1 of station Wettzell with respect to the clock at Tsukuba the time of the observation is of crucial importance for the impact the particular observation has on the parameter estimation process. Again, this resembles the impact of observations on the determination of the slope of a regression line (without estimating an axis offset).

The data resolution matrix shows the increasing impact for each observation as well as the supporting nature of the last observations. Due to this structure, the dendrogram shows a sequential clustering of (at first) the last observations up to the first observation, which is clustered at last (see also case 10).

Case 23: Estimation of clock offset CL_0 and clock rate CL_1 of station Wettzell with respect to station Tsukuba

The common estimation of the clock offset CL_0 and the clock rate CL_1 of one station with respect to a reference clock has not been treated so far. Here, the reference clock is the clock at Tsukuba station.

The analysis of these parameters is complicated by the high correlation coefficient (of -0.86). Since the **V**-matrix does not show a clear diagonal structure, a unique relation between observations and parameters is difficult. The impact factors, however, show a very clear and almost symmetrical increase of the importances of the first and the last observations, while the middle observations are of mean importance. This exactly resembles the situation when estimating the two parameters of an adjusting straight line (regression line with the axis offset estimated at the epoch of the first observation): the first observations are of main importance

for both the axis offset and the slope determination, while the last observations are mainly responsible for the slope determination. The same results can be seen after cutting the dendrogram to form two groups of observations (see table 5.23 on the next page): Here, the first observations are both responsible for the clock offset and the clock rate. The last observations are almost solely necessary for the clock rate determination.

Case 24: Estimation of atmospheric zenith path delay $AT_{Wettzell}$ at station Wettzell only

Similar to the atmospheric zenith path delay determination in the artificial interferometer in the previous example, observations performed with low elevations at Wettzell station are of main importance for the estimation of $AT_{Wettzell}$. As shown in table 5.22 on the following page, observations with large impact factors (such as observations 2, 4, 13, 14, 18, 20 and 23) are observed with very low elevations (below 11 [^{o}], at Wettzell). These are also observations which have been clustered at first. Due to the absence of distinct clusters a dendrogram cut is not reasonable.

Case 25: Estimation of atmospheric zenith path delay $AT_{Tsukuba}$ at station Tsukuba only

As shown in table 5.22 on the next page, in general, observations at Wettzell station have been observed with lower elevations than at Tsukuba station. Consequently, for the determination of the atmospheric zenith path delay $AT_{Tsukuba}$ at Tsukuba, less important observations are available than for the determination of $AT_{Wettzell}$ in case 24. But again, a clear relation between the size of the impact factors and the elevations at Tsukuba can be seen.

Also in this case, the cluster analysis of the data resolution matrix shows that observations 2, 16, 8 and 24 are grouped at first. The remaining observations are clustered in a sequential order without forming different clusters. Therefore, a dendrogram cut is not reasonable.

Case 26: Estimation of earth rotation $dUT1$ only

As mentioned in case 14 of the artificial spatial interferometer, equatorial sources (i.e., sources with low declinations) lying almost orthogonal to the baseline are of main importance for the determination of the earth rotation $dUT1$. This can also be seen in the impact factors of this case in figure 5.16 on page 105: Observations with large impact factors close to the threshold for very important observations have been performed to sources 0106+013 (observation 21), 0119+115 (observations 11, 19 and 27) or 0229+131 (observation 28) which all possess declinations below 15 [^{o}].

The data resolution matrix for this case shows the presence of several supporting observations mainly in the second half of the session. The dendrogram of the cluster analysis of this matrix shows that the most important observations (observation 21 to source 0106+013, observation 27 to source 0119+115, observation 19 to source 0119+115 and observation 11 also to source 0119+115) are grouped at first. The dendrogram also shows the presence of mainly two clusters with 17 and 12 observations, respectively. As shown in table 5.24 on page 103 the first cluster is of main importance for the determination of $dUT1$. The first cluster consists of the above mentioned observations to equatorial or near-equatorial sources. The second cluster only contains observations to sources located on the northern part of the mutually visible part of the celestial sphere and which are candidates for negligible observations (see figure 5.14 on page 99).

Case 27: Estimation of atmospheric zenith path delay $AT_{Tsukuba}$, clock parameters CL_0, CL_1, atmospheric zenith path delay $AT_{Wettzell}$ and earth rotation $dUT1$

The estimation of the atmospheric zenith path delay $AT_{Tsukuba}$ at Tsukuba, the clock parameters CL_0 and CL_1 of the clock behaviour at Wettzell, the atmospheric zenith path delay $AT_{Wettzell}$ at Wettzell and of the earth rotation $dUT1$ represents a realistic parametrisation of an INTENSIVE2-session. Since this case

No.	Source	Start of observation [MJD] ([min.])	Azimuth Tsukuba [^{o}]	Azimuth Wettzell [^{o}]	Elevation Tsukuba [^{o}]	Elevation Wettzell [^{o}]
1	0059+581	53442.314 (0.0)	321.3	47.2	51.73	45.11
2	1128+385	53442.316 (2.8)	49.2	319.3	9.59	10.14
3	1357+769	53442.317 (4.3)	8.6	340.2	24.97	47.76
4	0955+476	53442.319 (7.2)	50.5	339.4	29.32	10.28
5	1044+719	53442.320 (8.6)	22.6	345.7	32.13	34.68
6	1807+698	53442.321 (10.0)	351.2	338.1	17.51	65.51
7	0602+673	53442.322 (11.5)	17.8	12.0	55.63	28.35
8	2037+511	53442.324 (14.4)	326.7	64.3	12.98	83.60
9	1803+784	53442.325 (15.8)	355.6	349.7	25.32	58.61
10	0133+476	53442.327 (18.7)	304.4	57.8	55.68	37.89
11	0119+115	53442.328 (20.1)	254.6	91.2	40.18	16.29
12	0059+581	53442.329 (21.6)	320.7	48.5	48.80	47.88
13	0955+476	53442.331 (24.4)	52.8	342.8	32.13	9.31
14	0804+499	53442.332 (25.9)	53.7	1.2	50.62	8.98
15	1807+698	53442.334 (28.8)	353.7	335.0	16.99	64.29
16	2037+511	53442.335 (30.2)	328.3	51.9	11.17	85.94
17	0133+476	53442.337 (33.1)	304.2	59.1	53.27	39.89
18	0235+164	53442.338 (34.5)	245.8	75.8	55.00	9.71
19	0119+115	53442.339 (36.0)	257.7	94.3	37.02	18.94
20	0229+131	53442.341 (38.8)	244.9	80.8	50.96	8.97
21	0106+013	53442.342 (40.3)	250.5	104.6	27.34	13.89
22	1044+719	53442.344 (43.2)	22.8	348.2	34.76	33.33
23	0955+476	53442.345 (44.6)	53.7	345.3	35.40	8.37
24	1300+580	53442.346 (46.0)	29.5	326.1	16.38	30.38
25	0059+581	53442.348 (48.9)	320.2	50.5	45.32	51.19
26	0133+476	53442.349 (50.4)	304.6	61.3	50.21	42.48
27	0119+115	53442.351 (53.3)	260.2	97.8	33.78	21.60
28	0229+131	53442.352 (54.7)	248.1	83.4	47.94	11.62
29	0235+164	53442.353 (56.2)	250.9	80.4	50.92	13.21

Table 5.22: Observation list for real interferometer investigations (INTENSIVE2-session K05072, 13-3-2005), MJD = modified julian date, ([min.]) indicates the minutes passed since the first observation.

Cluster Analysis Results - Case 23					
Cluster:	No. of members:	Mean impact:	Members (observations):	Impact on parameter: CL_0	CL_1
1	14	0.07	14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 1, 2	**0.062** $\Rightarrow$ o	**0.036** $\Rightarrow$ o
2	15	0.07	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29	0.009 $\Rightarrow$ -	**0.033** $\Rightarrow$ -

Table 5.23: Cluster analysis results for case 23 (Estimation of clock offset CL_0 and clock rate CL_1 of station Wettzell with respect to station Tsukuba)

Cluster Analysis Results - Case 26				
Cluster:	No. of members:	Mean impact:	Members (observations):	Impact on parameter: $dUT1$
1	17	0.05	6, 16, 8, 1, 12, 25, 10, 17, 26, 18, 20, 29, 28, 11, 19, 21, 27	**0.047** $\Rightarrow$ o
2	12	0.02	9, 7, 15, 3, 22, 5, 14, 23, 24, 13, 2, 4	0.016 $\Rightarrow$ -

Table 5.24: Cluster analysis results for case 26 (Estimation of earth rotation $dUT1$ only)

resembles case 20 of an artificial interferometer, many results of this case are similar to the results of case 20. The differences in the results are mainly due to the complexity of this case (i.e., different baseline orientation and inhomogeneous source distribution).

Also for this real interferometer, the three best determined parameters are the atmospheric zenith path delay $AT_{Wettzell}$ at Wettzell, the earth rotation parameter $dUT1$ and the atmospheric zenith path delay $AT_{Tsukuba}$ at Tsukuba. Again, the weakest determined parameter is the clock offset CL_0 (cf. case 20). As in case 20 a strong correlation (of 0.75) between the first atmosphere parameter (here: $AT_{Tsukuba}$) and the clock offset CL_0 exists. The remaining correlations, however, improved and are much lower than the corresponding correlations in case 20 (the absolute values of all remaining correlation coefficients are below 0.5).

In agreement with case 20, the most important observations are those performed with low elevations (as e.g. observations 2, 16, 23, and 24). The geometry of the three most important observations is shown in figure 5.15 on the next page. The cluster analysis of the data resolution matrix shows that these observations belong to the three most important clusters 1 to 3, which are mainly responsible for $dUT1$, $AT_{Tsukuba}$ and $AT_{Wettzell}$ (see table 5.25). Due to their large impact onto the estimation process, these observations should be supplied (or controlled) by appropriate (independent) observations.

The remaining clusters are of importance for either $dUT1$ or the clock offset CL_0. Cluster 4 contains observations to equatorial sources, cluster 5 contains observations to the middle of the mutually visible part of the celestial sphere. As for case 20, for the determination of the clock offset CL_0, observations to sources in every part of the celestial sphere are needed.

From these results it can be concluded that cluster 5 contains observations which are candidates for observations that can be omitted, since this cluster mainly affects the auxiliary parameter CL_0. On the other hand, observations of cluster 1 and cluster 2 are important for the main parameter $dUT1$ and should thus be controlled by further, independent observations.

Cluster Analysis Results - Case 27								
Cl.:	No. of mbrs.:	Mean impact:	Members:	Impact on parameter:				
				AT_{TS}	CL_0	CL_1	AT_{WZ}	$dUT1$
1	4	0.21	4, 14, 13, 23	0.015	0.010	0.024	**0.047**	**0.052**
2	4	0.20	22, 24, 25, 26	0.015	0.010	**0.094**	**0.051**	**0.066**
3	5	0.25	15, 6, 2, 8, 16	**0.133**	0.037	0.016	0.036	0.016
4	8	0.14	27, 28, 29, 19, 11, 21, 18, 20	0.004	0.011	0.023	0.035	**0.044**
5	8	0.12	17, 9, 7, 3, 5, 12, 1, 10	0.024	**0.081**	0.033	0.018	0.012

Table 5.25: Cluster analysis results for case 27 (Estimation of atmospheric zenith path delay $AT_{Tsukuba}$, clock parameters CL_0, CL_1, atmospheric zenith path delay $AT_{Wettzell}$ and earth rotation $dUT1$)

Effects of modifications of cluster 5 on the cofactors of the parameters

In the following, it is investigated how much a modification of the observations of one cluster affects the cofactors of the parameters to be estimated: Since cluster 5 (containing observations 17, 9, 7, 3, 5, 12, 1 and 10) mainly affects the clock offset CL_0 (while the earth rotation parameter $dUT1$ is least affected), it is assumed that replacing the eight observations of cluster 5 with the observations of cluster 1 and cluster 2 results in an increase of the cofactor of CL_0 (i.e., a degradation of this parameter's accuracy) and a decrease of the cofactor of $dUT1$ (i.e., an improvement of the accuracy of this parameter)[4].

Table 5.26 shows that replacing cluster 5 by clusters 1 and 2 indeed mainly increases the cofactor of CL_0. This agrees with the *increase of uncertainty*-investigations of cases 14 and 20 performed on page 94. The cofactor of $dUT1$, however, is only slightly affected (i.e., the accuracy of $dUT1$ is only slightly improved). This again shows that the *increase of uncertainty* mainly quantifies the effect of *omitting* observations on the cofactors of the parameters.

The modifications described above only slightly affect the singular values and the correlations between the parameters. Thus, a visualisation is not reasonable.

	Cofactor change				
	AT_{TS} [m^2]	CL_0 [m^2]	CL_1 [m^2/day^2]	AT_{WZ} [m^2]	$dUT1$ [m^2]
Original observations	288.1	4384.8	449.5	97.4	46.6
Cluster 5 replaced	322.3	7690.7	556.4	106.7	45.6

Table 5.26: Case 27: Effects of modifications of cluster 5 on the cofactors of the parameters

Observation 2 (to source 1128+385)   Observation 16 (to source 2037+511)   Observation 24 (to source 1300+580)
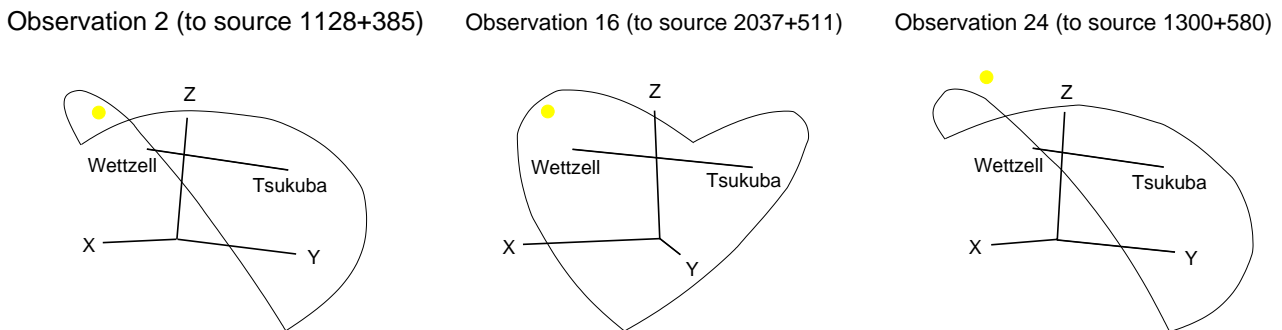
Figure 5.15: Geometry of the three most important observations of INTENSIVE2-session K05072.

---

[4]Obviously, due to the rotation of the earth, after the replacement of the observations of cluster 5 with the observations of clusters 1 and 2, cluster 5 consists of similar (but not identical) observations as contained in clusters 1 and 2.
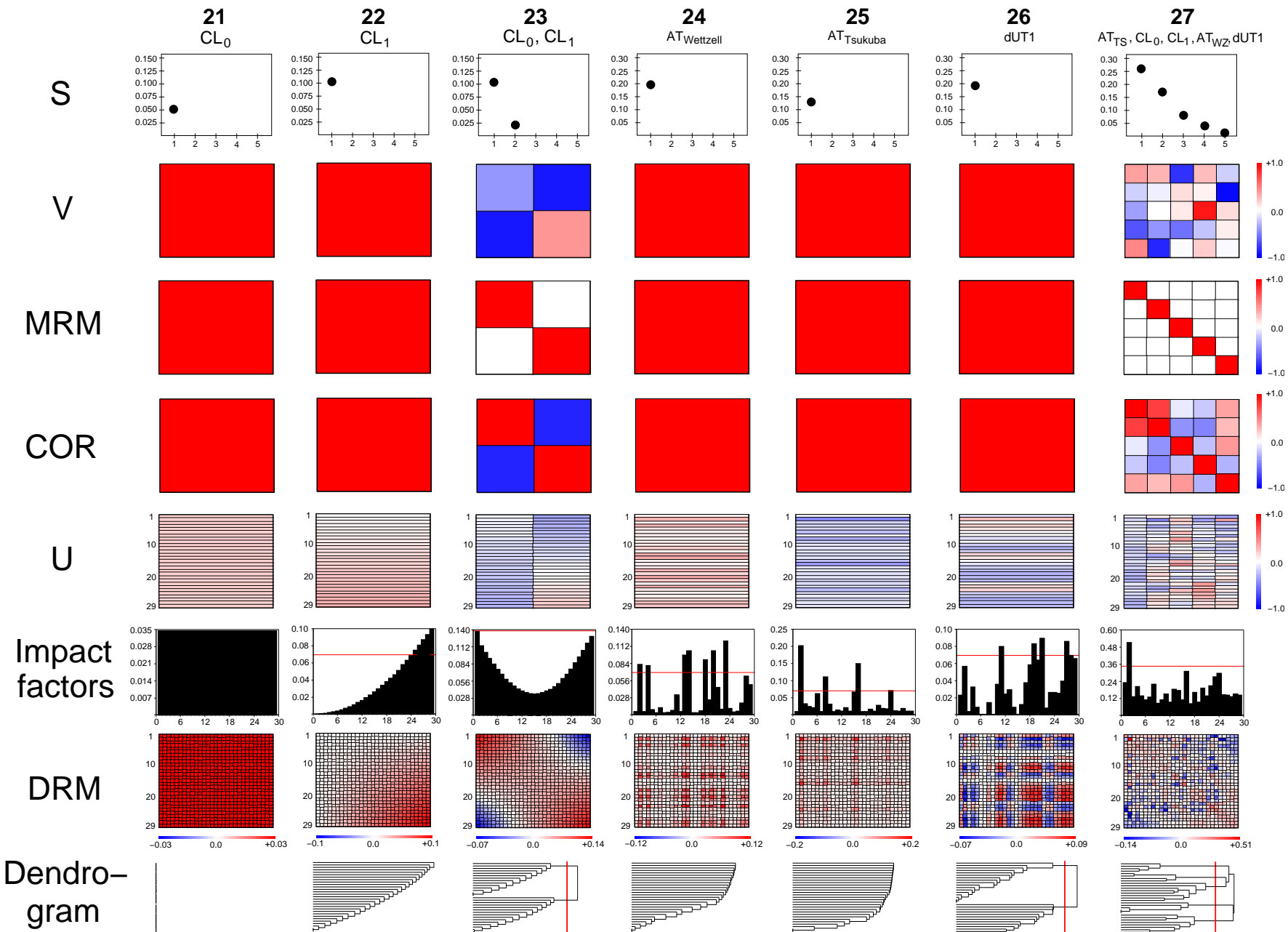
Figure 5.16: Analysis components of a real, single-baseline interferometer; red lines in impact factor diagrams indicate a recommended threshold for 'high-leverage observations' at $2 \cdot \frac{u}{n}$ (with u = number of unknowns, n = number of observations).

5.4.3.1   Conclusions from real, single-baseline VLBI session investigations

For a real VLBI session, interpretations are complicated by the skew orientation of the baseline (i.e., with respect to a geocentric, cartesian reference system), by the inhomogeneous source distribution and the irregular observation order. Nevertheless, most of the conclusions which can be drawn from investigating a real, single-baseline VLBI session resemble those obtained from the investigations of an artificial spatial kinematic interferometer. Especially for the most realistic parametrisation (case 27), the need for first investigating less complex cases became obvious. Only by considering the results of cases 8 to 20, cases 21 to 27 can be understood.

As mentioned in the previous conclusions, the more complex the functional model, the more difficulties arise due to the unavoidable increase of the correlations between the parameters to be estimated. Nevertheless, it could be shown that the regression diagnostics tool is also suited for real interferometers with complex observation structure. In most cases, groups of observations as well as their impact on the parameters to be estimated could be detected. Final conclusions will be given in the next chapter.

6. Summary, Conclusions & Outlook

Summary

The objective of this thesis is the development of a regression diagnostics tool, which can be used to improve the robustness and reliability of VLBI solutions. Therefore, the regression diagnostics tool must be able to detect the influence of single or groups of observations as well as their impact on the estimated parameters. In the language of parameter estimation in linear models: A tool for the detection of groups of high-leverage (and thus influential) observations and their impact on linear combinations of parameters (and thus on the parameters itself) is needed. Hence, such a regression diagnostics tool depicts an extension of the common investigations of the partial derivatives of a functional model with respect to the parameters to be estimated. While investigating partial derivates only provides the sensitivity of single observations on certain parameters (and thus the impact of single observations on the adjustment process), the regression diagnostics tool developed in this thesis analyses the entire design or the entire geometry of an experiment by taking into account the entire design matrix of an adjustment problem. Although the regression diagnostics tool developed in this thesis can be applied to any adjustment problem, in this thesis only applications to geodetic VLBI are presented.

In order to develop the regression diagnostics tool, at first some algebraic background needs to be summarised. Therefore, chapter 1 provides the basics of euclidean vector spaces, projections onto subspaces as well as the geometric aspects of the least-squares approach. In this context, the singular value decomposition of a design matrix is of fundamental importance since it provides new bases for the four subspaces of a matrix and is used to compute a so-called data resolution matrix. This matrix contains the so-called impact factors and impact co-factors which are used in the subsequent chapters for assessing the influence of observations. In chapter 2 the close relationship of the algebraic background of chapter 1 (i.e., the so-called vector space approach) with geodetic adjustment theory (or the theory of linear models) is presented. Furthermore, the close relationship of impact factors and redundancy numbers is given as well as geometrical interpretations of impact factors and impact co-factors. Since impact co-factors represent a common information content of observations, they can be used to detect groups of similar observations. In order to identify groups of similar observations, so-called cluster analysis methods are applied to the elements of the data resolution matrix. These methods are described in chapter 3. Furthermore, this chapter provides methods for measuring the impact of groups of observations onto individual parameters. The computation of the data resolution matrix after performing the singular value decomposition of the associated design matrix and the application of cluster analysis methods to the elements of the data resolution matrix form the main steps of the regression diagnostics tool developed in this thesis.

After a short review of the VLBI principle in chapter 4 the regression diagnostics tool is applied to plane and spatial interferometers. Therefore, chapter 5 describes the *qtSVD* software, which has been developed by the author of this thesis to

- set up the design matrix of a VLBI session,

- perform the singular value decomposition of the design matrix and to compute the data resolution matrix,

- to apply cluster analysis algorithms to the data resolution matrix, to visualize the cluster analysis dendrogram and to detect groups of similar observations and

- to compute the impact of each group of observations on each parameter to be estimated.

In order to show the capabilities of the regression diagnostics tool, it is applied to a plane, static interferometer and to spatial, kinematic interferometers. The latter ones are divided into an artificial spatial interferometer and a real VLBI session.

Conclusions

Based on both theoretical considerations and practical applications it could be shown that

- the regression diagnostics tool yields plausible and (geometrically) comprehensible results.

- the regression diagnostics tool can be used to detect groups of jointly influential and counter-acting observations.

- the regression diagnostics tool determines the impact of each group of observations onto the parameters to be estimated.

- the regression diagnostics tool can be used to detect degeneracies (or critical (baseline) configurations) and thus parameters which cannot be estimated or separated (cf. case 6).

- a technical realisation (i.e., a software implementation) of the regression diagnostics is possible and can be used for several (geodetic) adjustment problems.

The main benefit of the regression diagnostics tool developed in this thesis is thus the ability to detect weak parts of the design of a (VLBI-)experiment. The weak parts (such as inappropriate observation groups or indeterminable parameters) can then be improved or further investigated by the analyst.

The results obtained from applying the regression diagnostics tool to plane and spatial interferometers agree with (or even extend) existing VLBI analysis strategies. In addition, the regression diagnostics tool provides the *increase of uncertainty* due to the omission of observations. It thus shows which observations should be controlled (or supplied) by appropriate (independent) observations. In other words: It could be shown, that the regression diagnostics tool developed in this thesis *is* able to detect weak parts of the design of (not only VLBI-) experiments. In general, the strengths of the regression diagnostics tool become obvious for complex observation geometries and thus for experiments with data resolution matrices which do not possess a regular pattern.

Difficulties may arise in the only subjective part of the regression diagnostics procedure, i.e., the dendrogram cut. Depending on the form of the dendrogram a reasonable height for a dendrogram cut might not agree with a large similarity difference. In these cases, the analyst has to make an appropriate decision.

Outlook

In future, the regression diagnostics tool developed in this thesis needs to be applied to other single-baseline interferometers, to larger VLBI networks and for real VLBI session scheduling. In order to extend the pure analysis functionality of *qtSVD*, either the ability for *schedule improvement suggestions* should be added to *qtSVD* or the methods developed in this thesis should be implemented in existing VLBI scheduling software, such as SKED. Furthermore, the regression diagnostics tool should be applied to other geodetic adjustment problems such as geodetic networks or other geodetic space techniques.

In addition to the application of data space investigations (as performed in this thesis), the regression diagnostics tool could also be used for model space analyses and could thus be used to improve the estimability and separability of geodetic and geophysical parameters.

List of Figures

List of Tables

References

ACKERMANN, FRITZ (1981) *Grundlagen und Verfahren zur Erkennung grober Datenfehler.* in: Vorträge des Lehrgangs Numerische Photogrammetrie (IV) an der Universität Stuttgart, Heft 7, Institut für Photogrammetrie, Stuttgart.

ÁDÁM, JÓZSEF (1982) *A detailed study of the duality relation for the least squares adjustment in euclidean spaces.* Bulletin Géodésique Vol. 56, pp. 180-195.

ANDERSEN, PER HELGE (2000) *Multi-level arc combination with stochastic parameters.* Journal of Geodesy, 2000, 74, pp. 531-551.

BELSLEY, D.A., E. KUH and R.E. WELSCH (1980) *Regression Diagnostics: Identifying influential data and sources of collinearity.* John Wiley & Sons, New York.

BLANK, S.J., KRIKORIAN, N. and SPRING, D. (1989) *A geometrically inspired proof of the singular value decomposition.* American Mathematical Monthly, Vol. 96, Issue 3, p. 238-239.

BOEHM, JOHANNES (2004) *Troposphärische Laufzeitverzögerung in der VLBI.* Geowissenschaftliche Mitteilungen, Heft Nr. 68, Schriftenreihe der Studienrichtung Vermessungswesen und Geoinformation, Technische Universität Wien, ISSN 1811-8380.

BROUWER, F.J.J. (1985) *On the Principles, Assumptions and Methods of Geodetic Very Long Baseline Interferometry.* Netherlands Geodetic Commission, Vol. 7, Number 4, Delft.

CAMPBELL, JAMES (1979) *Die Radiointerferometrie auf langen Basen als geodätisches Messprinzip hoher Genauigkeit.* DGK Reihe C, Heft 254, Verlag des Instituts für Angewandte Geodäsie, Frankfurt am Main.

CAMPBELL, JAMES (2004) *VLBI for Geodesy and Geodynamics.* in: The Role of VLBI in Astrophysics, Astrometry and Geodesy, F. Mantovani and A. Kus (eds.), pp. 359-381, Kluwer Academic Publishers.

CAMPBELL, JAMES and BERTHOLD WITTE (1978) *Grundlage und geodätische Anwendung der Very Long Baseline Interferometry (VLBI).* Zeitschrift für Vermessungswesen, 103, S. 10-20.

CASPARY, WILHELM and KLAUS WICHMANN (1994) *Lineare Modelle.* R. Oldenbourg Verlag, München Wien, ISBN 3-486-22910-9.

COHEN, M.H. and D.B. SHAFFER (1971) *Positions of radio sources from long baseline interferometry.* Astron. Journ. Vol. 76, pp. 91-101.

COOK, R. DENNIS and SANFORD WEISBERG (1982) *Residuals and Influence in Regression.* Chapman and Hall, New York London, ISBN 0-412-24280-0.

DEHLERT, GARY W. (2000) *A first course in Design and Analysis of Experiments.* W.H. Freeman and Company, New York, ISBN 0-7167-3510-5.

DERMANIS, ATHANASIOS (1977) *Geodetic linear estimation techniques and the norm choice problem.* manuscripta geodaetica, Vol. 2, pp. 15-97.

DERMANIS, ATHANASIOS and ERIK GRAFAREND (1981) *Estimability analysis of geodetic, astrometric and geodynamical quantities in very long baseline interferometry.* Geophys. J. R. Astr. Soc.

DERMANIS, ATHANASIOS and IVAN MUELLER (1978) *Earth rotation and network geometry optimization for very long baseline interferometers.* Bulletin Géodésique, Vol. 52, No. 2, pp. 131-158.

DERMANIS, ATHANASIOS and RAINER RUMMEL (2000) *Data analysis methods in geodesy.* in: Geomatic Methods for the Analysis of Data in the Earth Sciences, Athanasios Dermanis and Armin Grün and Fernando Sanso (Eds.), Springer-Verlag Berlin Heidelberg New York, ISBN 3-540-67476-4.

DODGE, YADOLAH and JANA JURECKOVÁ (2000) *Adaptive Regression.* Springer, New York, ISBN 0-387-98965-X.

DUDA, RICHARD O., PETER E. HART and DAVID G. STORK (2000) *Pattern Classification.* John Wiley & Sons Inc., New York.

EEG, JØRGEN (1986) *On the Adjustment of Observations in the Presense of Blunders.* Geodætisk Institut, Technical Report No. 1, København.

EVEN-TZUR, GILAD (2006) *Datum definition and its influence on the reliability of geodetic networks.* Zeitschrift für Vermessungswesen 2/2006, p. 87-95.

FISCHER, DOROTHEE (2006) *Quality Aspects of Short Duration VLBI Observations for UT1 Determinations.* Ph.D. thesis, Geodätisches Institut der Rheinischen Friedrich-Wilhelms-Universität Bonn.

FÖRSTNER, WOLFGANG (1987) *Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision.* Computer Vision, Graphics, and Image Processing 40, p. 273-310.

FÖRSTNER, WOLFGANG (1992) *Uncertain geometric relationships and their use for object location in digital images.* Institut für Photogrammetrie, Universität Bonn, Tutorial.

GOLUB, GENE (1965) *Numerical methods for solving linear least squares problems.* Numerische Mathematik 7, pp. 206-216.

GOLUB, GENE and W. KAHAN (1965) *Calculating the singular values and pseudo-inverse of a matrix.* SIAM J. Numer. Anal., 2 (1965), pp. 205-224.

GOLUB, GENE and C. REINSCH (1970) *Singular value decomposition and least squares solutions.* Numerische Mathematik 14, pp. 403-420.

GRAMLICH, G. and H. WERNER (2000) *Numerische Mathematik mit MATLAB.* dpunkt.verlag, Heidelberg, ISBN 3-932588-55-X.

GRAY, J. BRIAN and ROBERT F. LING (1984) *K-Clustering as a detection tool for influential subsets in regression.* Technometrics, Vol. 26, No. 4, p. 305-318.

HOAGLIN, DAVID C. and ROY E. WELSCH (1978) *The hat matrix in regression and ANOVA.* The American Statistican, Vol. 32, No. 1, p. 17-22.

HOFMANN-WELLENHOF, B., LEGAT, K. and WIESER, M. (2003) *Navigation. Principles of Positioning and Guidance.* Springer Wien New York, ISBN 3-211-00828-4.

JACKSON, J. EDWARD (2003) *A User's Guide To Principal Components.* John Wiley & Sons, New York.

KALMAN, DAN (1996) *A singularly valuable decomposition: the SVD of a matrix.* College Mathematics Journal, Vol. 27, No. 1 (Jan., 1996), pp. 2-23.

KOCH, KARL RUDOLF (1999) *Parameter Estimation and Hypothesis Testing in Linear Models.* Springer-Verlag Berlin Heidelberg, New York.

KSHIRSAGAR, ANANT (1983) *A Course in Linear Models.* Marcel Dekker, Inc., New York.

LAWSON, CHARLES and RICHARD HANSON (1995) *Solving Least Squares Problems.* Society for Industrial and Applications Mathematics (SIAM), Philadelphia, ISBN 0-89871-3560.

LAY, DAVID C. (2003) *Linear Algebra and Its Applications.* Addison-Wesley.

LEICK, ALFRED (1990) *GPS Satellite Surveying.* John Wiley & Sons Inc., New York.

LUNDQVIST, GÖRAN (1984) *Radio Interferometry as a Probe of Tectonic Plate Motion.* Pd.D. thesis, Chalmers University of Technology, Göteborg.

MA, C. ET AL. (1990) *Measurement of horizontal motions in Alaska using very long baseline interferometry.* J. Geophys. Res., 95(B13), pp. 21991-22011.

MA, CHOPO (1978) *Very Long Baseline Interferometry Applied to Polar Motion, Relativity and Geodesy.* NASA Technical Memorandum 79582, University of Maryland, Maryland.

MCCARTY and PETIT (2003) *IERS Conventions 2003.* IERS Technical Note 32, Observatoire de Paris.

MEISSL, PETER (1982) *Least Squares Adjustment - A modern approach.* Mitteilungen der geodätischen Institute der TU Graz, Folge 43, Graz.

MENKE, WILLIAM (1984) *Geophysical Data Analysis: Discrete Inverse Theory.* Academic Press, Inc., Orlando, ISBN 0-12-490920-5.

MEYER, CARL (2000) *Matrix Analysis and Applied Linear Algebra.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, ISBN 0-89871-454-0.

MICHELSON, A.A. (1890) *On the application of interference methods to astronomical measurements.* Philos. Magazine, 5th Series.

MORITZ, H., I. MUELLER (1987) *Earth Rotation - Theorie and Observation.* Ungar Publishing Company, New York.

NIEMEIER, WOLFGANG (2002) *Ausgleichungsrechnung.* deGruyter Berlin New York, ISBN 3-11-014080-2.

NOTHNAGEL, AXEL (1991) *Radiointerferometrische Beobachtungen zur Bestimmung der Polbewegung unter Benutzung langer Nord-Süd-Basislinien.* DGK Reihe C, Heft 368, Verlag des Instituts für Angewandte Geodäsie, Frankfurt am Main.

NOTHNAGEL, AXEL (2000) *Der Einfluss des Wasserdampfes auf die modernen raumgestützten Messverfahren.* Mitteilungen des Bundesamtes für Kartographie und Geodäsie, Band 16, Verlag des Bundesamtes für Kartographie und Geodäsie, Frankfurt am Main, ISBN 3-88648-100-X.

PARKER, ROBERT L. (1994) *Geophysical Inverse Theory.* Princeton University Press, Princeton.

PRESS, WILLIAM H., BRIAN P. FLANNERY, SAUL A. TEUKOSLY and WILLIAM T. VETTERLING (1986) *Numerical Recipes - The Art of Scientific Computing.* Cambridge University Press, Cambridge.

RIPLEY, B.D. (1996) *Pattern Recognition and Neuronal Networks.* Cambridge University Press, Cambridge.

ROBERTSON, DOUGLAS SCOTT (1975) *Geodetic and Astrometric Measurements with Very-Long-Baseline Interferometry.* Department of Earth and Planetary Sciences, Massachusetts Institute of Technology Cambridge, Massachusetts.

ROMESBURG, H.C. (2004) *Cluster Analysis for Researchers.* Lulu Press, North Carolina, online available at http://www.lulu.com.

ROTHACHER, MARKUS (2003) *Erdmessung und Satellitengeodäsie 2 (Lecture material).* Institut für Astronomische und Physikalische Geodäsie, Technische Universität München, München.

SAVILLE, DAVID and GRAHAM WOOD (1997) *Statistical Methods: The Geometric Approach.* Springer, New York, ISBN 3-540-97517-9.

SCALES, JOHN A., MARTIN L. SMITH and SVEN TREITEL (2001) *Introductory Geophysical Inverse Theory.* Samizdat Press, online available at ftp://samizdat.mines.edu.

SCHUH, HARALD (1987) *Die Radiointerferometrie auf langen Basen zur Bestimmung von Punktverschiebungen und Erdrotationsparametern.* DGK Reihe C, Heft 328, Verlag der Bayerischen Akademie der Wissenschaften, München.

SCHUH, H. ET AL. (2006) *IVS-WG3 Report on Data Analysis.* IVS Memorandum 2006-006v01, available from ftp://ivscc.gsfc.nasa.gov/pub/memos/ivs-2006-006v01.pdf.

SCHWARZ, HANS RUDOLF (1997) *Numerische Mathematik.* B. G. Teubner, Stuttgart.

SEARLE, SHAYLE R. (1982) *Matrix Algebra useful for statistics.* John Wiley & Sons, New York.

SEEBER, GÜNTER (2003) *Satellite Geodesy.* de Gruyter, Berlin, New York.

SHAPIRO, I.I. ET AL. (1974) *Transcontinental baseline and the rotation of the Earth measured by radio interferometry.* Science, Vol. 186, pp. 920-922.

SNIEDER, ROEL and JEANNOT TRAMPERT (2000) *Linear and nonlinear inverse problems.* in: Geomatic Methods for the Analysis of Data in the Earth Sciences, Athanasios Dermanis and Armin Grün and Fernando Sanso (Eds.), Springer-Verlag Berlin Heidelberg New York, ISBN 3-540-67476-4.

SOVERS, OJARS J., JOHN L. FANSELOW and CHRISTOPHER S. JACOBS (1998) *Astrometry and geodesy with radio interferometry: experiments, models, results.* Reviews of Modern Physics, Vol. 70, No. 4.

STEUFMEHL, H. (1994) *Optimierung von Beobachtungsplänen in der Langbasisinterferometrie (VLBI).* DGK, Reihe C, Heft Nr. 406, Verlag des Instituts für Angewandte Geodäsie, Frankfurt am Main.

STEWART, G. W. (1993) *On the early history of the singular value decomposition.* SIAM Review, 35 (4), Philadelphia, pp. 551-566, ISSN 0036-1445.

STRANG, GILBERT (2003) *Lineare Algebra.* Springer-Verlag Berlin Heidelberg New York, ISBN 3-54043949-8.

STRANG, GILBERT and KAI BORRE (1997) *Linear Algebra, Geodesy and GPS.* Wellesley-Cambridge Press, Wellesley, ISBN 0-96140-8863.

TAKAHASHI, YUKIO (1994) *Estimation of Errors in VLBI Data and Position Determination Error.* Journal of the Geodetic Society of Japan, Vol. 40, No. 4, pp. 309-331, Tokyo.

TAKAHASHI ET AL. (2000) *Wave Summit Course: Very Long Baseline Interferometers.* Ohmsha, Ltd., Tokyo.

TESMER, VOLKER (2004) *Das stochastische Modell bei der VLBI-Auswertung.* DGK Reihe C, Heft 573, Verlag der Bayerischen Akademie der Wissenschaften, München, ISBN 3-76965012-3.

TEUNISSEN, P.J.G. (1985) *The Geometry of Geodetic Inverse Linear Mapping And Non-Linear Adjustment.* Netherlands Geodetic Commission, Vol. 8, Number 1, Delft.

TEUNISSEN, P.J.G. (2003) *Adjustment theory, an introduction.* Vereniging voor Studie- en Studentenbelangen te Delft (VSSD), Delft, ISBN 90-407-1974-8.

TITOV, OLEG (2002) *Global analysis of VLBI data by least-squares collocation method.* EGS XXVII General Assembly, Nice, 21-26 April 2002.

TITOV, OLEG, VOLKER TESMER and JOHANNES BOEHM (2004) *OCCAM v.6.0 software for VLBI data analysis.* In: Vandenberg, N., K. Baver (Eds.): IVS 2004 General Meeting Proceedings, NASA/CP-2004-212255, 311-314.

TOUTENBURG, HELGE (2003) *Lineare Modelle.* Physica-Verlag, Heidelberg.

TREFETHEN, L. N., D. BAU (1997) *Numerical Linear Algebra.* Society for Industrial and Applications Mathematics (SIAM), Philadelphia, ISBN 0-89871-3617.

VANICEK, P. and E. KRAKIWSKY (1986) *Geodesy: The Concepts.* Elsevier Science Publishers B.V., Amsterdam.

VENNEBUSCH, MARKUS (2002) *Untersuchungen zu Korrelationen zwischen Parametern in der VLBI-Auswertung.* unpublished diploma thesis, Rheinische Friedrich-Wilhelms-Universität zu Bonn.

WALTER, H.G. (1973) *Astrometrical Applications of Long Baseline Interferometry.* Bulletin Groupe de Recherches de Geodesie Spatiale (GRGS) No. 10, Toulouse.

Acknowledgements