DGK Deutsche Geodätische Kommission
der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 759

Lutz Rolf Roese-Koerner

Convex Optimization

for Inequality Constrained Adjustment Problems

München 2015

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck

ISSN 0065-5325 ISBN 978-3-7696-5171-3

Diese Arbeit ist gleichzeitig veröffentlicht in:
Schriftenreihe des Instituts für Geodäsie und Geoinformation
der Rheinischen Friedrich-Wilhelms Universität Bonn
ISSN 1864-1113, Nr. 50, Bonn 2015

DGK Deutsche Geodätische Kommission
der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 759

Convex Optimization for Inequality Constrained Adjustment Problems

Inaugural-Dissertation zur
Erlangung des Grades
Doktor-Ingenieur (Dr.-Ing.)
der Hohen Landwirtschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms Universität
zu Bonn

vorgelegt von

Lutz Rolf Roese-Koerner

aus Bad Neuenahr-Ahrweiler

München 2015

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck

ISSN 0065-5325 ISBN 978-3-7696-5171-3

Diese Arbeit ist gleichzeitig veröffentlicht in:
Schriftenreihe des Instituts für Geodäsie und Geoinformation
der Rheinischen Friedrich-Wilhelms Universität Bonn
ISSN 1864-1113, Nr. 50, Bonn 2015

Adresse der Deutschen Geodätischen Kommission:

Deutsche Geodätische Kommission

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 89 – 23 031 1113 ● Telefax +49 – 89 – 23 031 - 1283 / - 1100

e-mail hornik@dgfi.badw.de ● http://www.dgk.badw.de

Diese Publikation ist als pdf-Dokument veröffentlicht im Internet unter den Adressen /
This volume is published in the internet

<http://dgk.badw.de> / <http://hss.ulb.uni-bonn.de/2015/4142/4142.pdf>

Prüfungskommission

Referent: Prof. Dr. techn. Wolf-Dieter Schuh

Korreferenten: Prof. Dr. Hans-Peter Helfrich

Prof. Dr.-Ing. Nico Sneeuw

Tag der mündlichen Prüfung: 10.07..2015

ISSN 0065-5325 ISBN 978-3-7696-5171-3

Convex Optimization for Inequality Constrained Adjustment Problems

Summary

Whenever a certain function shall be minimized (e.g., a sum of squared residuals) or maximized (e.g., profit) optimization methods are applied. If in addition prior knowledge about some of the parameters can be expressed as bounds (e.g., a non-negativity bound for a density) we are dealing with an optimization problem with inequality constraints. Although, common in many economic and engineering disciplines, inequality constrained adjustment methods are rarely used in geodesy. Within this thesis methodology aspects of convex optimization methods are covered and analogies to adjustment theory are provided. Furthermore, three shortcomings are identified which are—in the opinion of the author—the main obstacles that prevent a broader use of inequality constrained adjustment theory in geodesy. First, most optimization algorithms do not provide quality information of the estimate. Second, most of the existing algorithms for the adjustment of rank-deficient systems either provide only one arbitrary particular solution or compute only an approximative solution. Third, the Gauss-Helmert model with inequality constraints was hardly treated in the literature so far. We propose solutions for all three obstacles and provide simulation studies to illustrate our approach and to show its potential for the geodetic community.
Thus, the aim of this thesis is to make accessible the powerful theory of convex optimization with inequality constraints for classic geodetic tasks.

Konvexe Optimierung für Ausgleichungsaufgaben mit Ungleichungsrestriktionen

Zusammenfassung

Methoden der konvexen Optimierung kommen immer dann zum Einsatz, wenn eine Zielfunktion minimiert oder maximiert werden soll. Prominente Beispiele sind eine Minimierung der Verbesserungsquadratsumme oder eine Maximierung des Gewinns. Oft liegen zusätzliche Vorinformationen über die Parameter vor, die als Ungleichungen ausgedrückt werden können (beispielsweise eine Nicht-Negativitätsschranke für eine Dichte). In diesem Falle erhält man ein Optimierungsproblem mit Ungleichungsrestriktionen. Ungeachtet der Tatsache, dass Methoden zur Ausgleichungsrechnung mit Ungleichungen in vielen ökonomischen und ingenieurwissenschaftlichen Disziplinen weit verbreitet sind, werden sie dennoch in der Geodäsie kaum genutzt.
In dieser Arbeit werden methodische Aspekte der konvexen Optimierung behandelt und Analogien zur Ausgleichungsrechnung aufgezeigt. Desweiteren werden drei große Defizite identifiziert, die – nach Meinung des Autors – bislang eine häufigere Anwendung von restringierten Ausgleichungstechniken in der Geodäsie verhindern. Erstens liefern die meisten Optimierungsalgorithmen ausschließlich eine Schätzung der unbekannten Parameter, jedoch keine Angabe über deren Genauigkeit. Zweitens ist die Behandlung von rangdefekten Systemen mit Ungleichungsrestriktionen nicht trivial. Bestehende Verfahren beschränken sich hier zumeist auf die Angabe einer beliebigen Partikulärlösung oder ermöglichen gar keine strenge Berechnung der Lösung. Drittens wurde das Gauß-Helmert-Modell mit Ungleichungsrestriktionen in der Literatur bisher so gut wie nicht behandelt. Lösungsmöglichkeiten für alle genannten Defizite werden in dieser Arbeit vorgeschlagen und kommen in Simulationsstudien zum Einsatz, um ihr Potential für geodätische Anwendungen aufzuzeigen.
Diese Arbeit soll somit einen Beitrag dazu leisten, die mächtige Theorie der konvexen Optimierung mit Ungleichungsrestriktionen für klassisch geodätische Aufgabenstellungen nutzbar zu machen.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 State of the Art 2

1.3 Scientific Context and Objectives of this Thesis 3

1.4 Outline 5

I Fundamentals **7**

2 Adjustment Theory **8**

2.1 Mathematical Foundations 8

2.2 Models 13

2.3 Estimators 21

3 Convex Optimization **23**

3.1 Convexity 23

3.2 Minimization with Inequality Constraints 24

3.3 Quadratic Program (QP) 26

3.4 Duality 29

3.5 Active-Set Algorithms 32

3.6 Feasibility and Phase-1 Methods 39

II Methodological Contributions **43**

4 A Stochastic Framework for ICLS **44**

4.1 State of the Art 45

4.2 Quality Description 47

4.3 Analysis Tools for Constraints 50

4.4 The Monte Carlo Quadratic Programming (MC-QP) Method 53

4.5 Example 54

5 Rank-Deficient Systems with Inequalities **59**
5.1 State of the Art . . . 60
5.2 Extending Active-Set Methods for Rank-Deficient Systems . . . 61
5.3 Rigorous Computation of a General Solution . . . 65
5.4 A Framework for the Solution of Rank-Deficient ICLS Problems . . . 71
5.5 Example . . . 72

6 The Gauss-Helmert Model with Inequalities **77**
6.1 State of the Art . . . 77
6.2 WLS Adjustment in the Inequality Constrained GHM . . . 78
6.3 Example . . . 87

III Simulation Studies **93**

7 Applications **94**
7.1 Robust Estimation with Inequality Constraints . . . 94
7.2 Stochastic Description of the Results . . . 97
7.3 Applications with Rank-Deficient Systems . . . 104
7.4 The Gauss-Helmert Model with Inequalities . . . 108

IV Conclusion and Outlook **113**

8 Conclusion and Outlook **114**
8.1 Conclusion . . . 114
8.2 Outlook . . . 115

A Appendix **i**
A.1 Transformation of the Dual Function of an ICLS Problem . . . i
A.2 Deactivating Active Constraints in the Active-Set Algorithm . . . ii
A.3 Reformulation of the Huber Loss Function . . . iv
A.4 Data of the Positive Cosine Example . . . v

Tables of Symbols **vii**

List of Abbreviations **viii**

List of Figures **viii**

List of Tables **ix**

References **xv**

1. Introduction

1.1 Motivation

Many problems in both science and society can be mathematically formulated as **optimization problems**. The term *optimization* already implies that the task is to find a solution that is *optimal*—i.e. better (or at least not worse) than any other existing solution. Thus, optimization methods always come into play if a cost function should be minimized or a profit function should be maximized. Prominent examples in geodesy include the minimization of the sum of squared residuals or the maximization of a likelihood function.

In many cases additional information on (or restrictions of) the unknown parameters exist that could be expressed as constraints. Equality constraints can be easily incorporated in the adjustment process to take advantage of as much information as possible. However, it is not always possible to express knowledge on the parameters as equalities. Possible examples are a lower bound of zero for non-negative quantities such as densities or repetition numbers, or an upper bound for the maximal feasible slope of a planned road. In these cases, inequalities are used to express the additional knowledge on the parameters which leads to an **optimization problem with inequality constraints**.

If in addition it is known that only a global minimum exists, a **convex optimization problem with inequality constraints** has to be solved.

Despite the fact that there is a rich body of work on convex optimization with inequality constraints in many engineering and economic disciplines, its application in classic geodetic adjustment tasks is rather scarce.

Within this thesis methodological aspects from the field of convex optimization with inequality constraints are described and analogies to classic adjustment theory are pointed out. Furthermore, we identified the following three shortcomings, which are—in our opinion—the main obstacles that prevent a broader use of inequality constrained adjustment theory in geodesy:

Quality description. It is a fundamental concept in geodesy that not only the value of an estimated quantity is of interest but also its accuracy. Information on the accuracy is usually provided as a variance-covariance matrix, which allows us to extract standard deviations and information on the correlation of different parameters. However, this is no longer possible in the presence of inequality constraints as their influence on the parameters cannot be described analytical. Furthermore, symmetric confidence regions are no longer sufficient to describe the accuracy as the parameter space is truncated by inequality constraints.

Rank-deficient systems. Many applications in geodesy lead to a rank-deficient system of normal equations. Examples are the second order design of a geodetic network with more weights to be estimated than entries in the criterion matrix, or the adjustment of datum-free networks. However, most state-of-the-art optimization algorithms for inequality constrained problems are either not capable of solving a singular system of linear equations or provide only one of an infinite number of solutions.

Gauss-Helmert model. Oftentimes, not only the relationship between a measured quantity (i.e. an observation) and the unknown parameters has to be modeled, but also the relationship between two or more observations themselves. In this case, it is not possible to perform an adjustment in the Gauss-Markov model, and the Gauss-Helmert model is used instead. However, it is not straightforward to perform an inequality constrained estimation in the Gauss-Helmert model.

1.2 State of the Art

A rich literature on convex optimization exists, including textbooks such as Gill et al. (1981), Fletcher (1987) or Boyd and Vandenberghe (2004).

The same holds true for a special case of convex optimization problem: the quadratic program. Here, a quadratic objective function should be minimized subject to some linear constraints. Two out of many works on this topic are Simon and Simon (2003) and Wong (2011). The former proposed a quadratic programming approach to set up an inequality constrained Kalman Filter for aircraft turbofan engine health estimation. The latter examined certain active-set methods for quadratic programming in positive (semi-)definite as well as in indefinite systems.

Not only in the geodetic community, there exist many articles on the solution of an Inequality Constrained Least-Squares (ICLS) problem as a quadratic program. Stoer (1971) for example proposed, what he defined as "a numerically stable algorithm for solving linear least-squares problems with linear equalities and inequalities as additional constraints". Klumpp (1973) formulated the problem of estimating an optimal horizontal centerline in road design as a quadratic program.
Fritsch and Schaffrin (1980) and Koch (1981) were the first to address inequality constrained least-squares problems in geodesy. While the former formulated the design of optimal FIR filters as ICLS problem, the latter examined hypothesis testing with inequality constraints. Later on Schaffrin (1981), Koch (1982, 1985) and Fritsch (1982) transformed the quadratic programming problem resulting from the first and second order design of a geodetic network into a linear complementarity problem and solved it via Lemke's algorithm. Fritsch (1983, 1985) examined further possibilities resulting from the use of ICLS for the design of FIR filters and other geodetic applications. Xu et al. (1999) proposed an ansatz to stabilize ill-conditioned LCP problems.
A more recent approach stems from Peng et al. (2006) who established a method to express many simple inequality constraints as one intricate equality constraint in a least-squares context. Koch (2006) formulated constraints for the semantical integration of two-dimensional objects and digital terrain models. Kaschenz (2006) used inequality constraints as an alternative to the Tikhonov regularization leading to a non-negative least-squares problem. She applied her proposed framework to the analysis of radio occultation data from GRACE (Gravity Recovery And Climate Experiment). Tang et al. (2012) used inequalities as smoothness constraints to improve the estimated mass changes in Antarctica from GRACE observations, which leads again to a quadratic program.

Much less literature exists on the quality description of inequality constrained estimates. The probably most cited work in this area is the frequentist approach of Liew (1976). He first identified the active constraints and used them to approximate an inequality constrained least-squares problem by an equality constrained one. Geweke (1986) on the other hand suggested a Bayesian approach, which was further developed and introduced to geodesy by Zhu et al. (2005).
However, both approaches are incomplete. While in the first ansatz, the influence of inactive constraints is neglected, the second ansatz ignores the probability mass in the infeasible region. Thus, we propose the MC-QP method in Chap. 4 which overcomes both drawbacks.

The probably most import contributions to the area of rank-deficient systems with inequality constraints are the works of Werner (1990), Werner and Yapar (1996) and Wong (2011). In the two former articles a projector theoretical approach for the rigorous computation of a general solution of ICLS problems with a possible rank defect is proposed using generalized inverses. In the latter an extension of classic active-set methods for quadratic programming is described, which enables us to compute a particular solution despite a possible rank defect.
However, the ansatz from Werner (1990) and Werner and Yapar (1996) is only suited for small-scale problems, and the approach of Wong (2011) lacks a description of the homogeneous solution. Thus, we propose a framework for rank-deficient and inequality constrained problems in Chap. 5 that is applicable to larger problems and provides a particular as well as a homogeneous solution.

To the best of our knowledge, nearly no literature on the inequality constrained Gauss-Helmert model exists. The works which come closest treat the mixed model—which can be seen as a generalization of the Gauss-Helmert model. Here, the works of Famula (1983), Kato and Hoijtink (2006), Davis et al. (2008) and Davis (2011) should be mentioned.
In Chap. 6 we describe two approaches to solve an inequality constrained Gauss-Helmert model. One that uses standard solvers and one that does not but therefore takes advantage of the special structure of the Gauss-Helmert model.

1.3 Scientific Context and Objectives of this Thesis

As implied by the title, this work is located at the transition area between **mathematical optimization** and **adjustment theory**. Both fields are concerned with the estimation of unknown parameters in such a way that an objective function is minimized or maximized. As a consequence the topics overlap to a large extent and sometimes differ merely in the terminology used.

As there exist many different definitions and as the assignment of a certain concept to one of the two fields might sometimes be arbitrary, we state in the following how the terms will be used within this thesis.

Adjustment theory. In adjustment theory (cf. Chap. 2) not only the parameter estimate but also its accuracy is of interest. Thus, we have a functional as well as a stochastic model for a specific problem as it is common in geodesy. This also includes testing theory and the propagation of variances.
Furthermore, the term adjustment theory is often connected with three classic geodetic models: the Gauss-Markov model, the adjustment with condition equations and the Gauss-Helmert model. All of these models can be combined with an estimator that defines what characterizes an optimal estimate. By far the most widely used estimator is the L_2 norm estimator which leads to the famous least-squares adjustment. Furthermore, it can be shown to be the Best Linear Unbiased Estimator (BLUE) in the aforementioned models (cf. Koch, 1999, p. 156–158 and p. 214–216, respectively).

Mathematical optimization. In mathematical optimization (cf. Chap. 3) on the other hand, mostly the estimate itself matters. Thus, we are usually dealing with deterministic quantities only.
Furthermore, the optimization theory deals with methods to reformulate a certain problem in a more convenient form and provides many different algorithms to solve it. In addition, inequality constrained estimates are usually assigned to this field. While in adjustment theory the focus lies on the problem and a way to model it mathematically, in optimization the focus is more on general algorithmic developments leading to powerful methods for many tasks.

Some words might be in order to distinguish the topic of this thesis from related work in other fields.

1. We are using frequentist instead of Bayesian inference. Thus, we assume that the unknown parameters have a fixed but unknown value and are deterministic quantities (cf. Koch, 2007, p. 34). Thus, no stochastic prior information on the parameters is used. However, in Sect. 4.2.3 we borrow the concept of Highest Posterior Density Regions, which originally stems from Bayesian statistics but can also be applied in a frequentist framework.

2. Recently, many works have been published dealing with the errors-in-variables model and thus leading to a total least-squares estimate. Some of them even include inequality constraints (De Moor, 1990, Zhang et al., 2013, Fang, 2014, Zeng et al., 2015). However, the errors-in-variables model can be seen as a special case of the Gauss-Helmert model (cf. Koch, 2014). Thus, we decided to treat the most general version of the Gauss-Helmert model in Chap. 6 instead of a special case.

3. Whenever we mention the term "inequalities" we are talking about constraints for the quantities which should be estimated. This is a big difference to the field of "censoring" (cf. Kalbfleisch and Prentice, 2002, p. 2) that deals with incomplete data. An example for an inequality in a censored data problem would be an **observation** whose quantity is only known by a lower bound as in "The house was sold for at least €500 000." As a consequence, it is not known if the house was sold for €500 000, €600 000 or €1 000 000. Within this thesis we explicitely exclude censored data.

Purpose of this Work. The purpose of this thesis it to make accessible the powerful theory of convex optimization—especially the estimation with inequality constraints—for classic geodetic tasks. Besides the extraction of certain analogies between convex optimization methods and approaches from adjustment theory, this is attempted by removing the three main obstacles identified in Sect. 1.1 and thus includes

- the extension of existing convex optimization algorithms by a stochastic framework.
- a new strategy to treat rank-deficient problems with inequalities.
- a formulation of a Gauss-Helmert model with inequality constraints as a quadratic program in standard form.

The contents of this thesis have been partly published in the following articles.

Roese-Koerner, Devaraju, Sneeuw and Schuh (2012). A stochastic framework for inequality constrained estimation. *Journal of Geodesy*, 86(11):1005–1018, 2012.

Roese-Koerner and Schuh (2014). Convex optimization under inequality constraints in rank-deficient systems. *Journal of Geodesy*, 88(5):415–426, 2014.

Roese-Koerner, Devaraju, Schuh and Sneeuw (2015). Describing the quality of inequality constrained estimates. In Kutterer, Seitz, Alkhatib, and Schmidt, editors, *Proceedings of the 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11)*, *IAG Symp. 140*, pages 15–20. Springer, Berlin, Heidelberg, 2015.

Roese-Koerner and Schuh (2015). Effects of different objective functions in inequality constrained and rank-deficient least-squares problems. In *VIII Hotine-Marussi Symposium on Mathematical Geodesy*, *IAG Symp. 142*. Springer, Berlin, Heidelberg, 2015 (accepted).

Halsig, Roese-Koerner, Artz, Nothnagel and Schuh (2015). Improved parameter estimation of zenith wet delay using an inequality constrained least squares method. In *IAG Scientific Assembly, Potsdam 2013*, *IAG Symp. 143*. Springer, Berlin, Heidelberg, 2015 (accepted)

1.4 Outline

This thesis is subdivided in four parts: fundamentals, methodological contributions, simulation studies, and conclusion and outlook.

The first part consists of two chapters and is devoted to fundamental mathematical principles. The basics of adjustment theory are described in Chap. 2. This includes mathematical concepts, such as the quadratic form, optimality conditions for unconstrained problems, systems of linear equations as well as different models and estimators.

Existing methods in convex optimization are covered in Chap. 3. Here, the term convexity is defined and the minimization with inequality constraints is explained. Furthermore, a special optimization problem—the quadratic program—is introduced alongside the concept of duality. The last sections of the chapter are devoted to active-set methods as one possible way to solve an inequality constrained problem, and to methods for finding a feasible solution, respectively.

Our own contributions are summarized in Part Two, which consists of three chapters. A stochastic framework for the quality description of inequality constrained estimates is proposed in Chap. 4. It combines the actual quality description with some analytic methods for constraints. This includes two global measures (an hypothesis test and a measure for the probability mass in the infeasible region) and one local measure: A sensitivity analysis that allows us to examine the influence of each constraint on each parameter. The chapter concludes with a simple example of the proposed framework.

Rank-deficient systems with inequalities are the topic of Chap. 5. Here an extension for active-set methods is described which is required to compute a particular solution of a singular system. Furthermore, the rigorous computation of a general solution is described and a framework for the solution of rank-deficient ICLS problems is proposed. The application of the framework is illustrated with a simple two-dimensional example with a one-dimensional manifold.

The Gauss-Helmert model with inequalities is treated in Chap. 6. We show how to reformulate the inequality constrained Gauss-Helmert model as a quadratic program in standard form. Furthermore, the special structure of the Gauss-Helmert model is exploited leading to a tailor-made transformation of the Gauss-Helmert model. In addition, the Karush-Kuhn-Tucker optimality conditions for the tailor-made approach are derived and a simple example is provided.

Six different applications of inequality constrained adjustment form the third part of this thesis—the simulation studies in Chap. 7. Here, the Huber estimator is reformulated as a quadratic program in standard form, as an example for robust estimation with inequality constraints. Furthermore, the stochastic description mentioned above is illustrated again at two close-to-reality examples: The estimation of a positive definite covariance function and a VLBI adjustment with constraints on the tropospheric delay. Afterwards, the second order design of a geodetic network and an engineering example with strict welding tolerances are portrayed. Both tasks lead to rank-deficient systems. In the last simulation study, the Gauss-Helmert model with inequality constraints is used for the optimal design of a vertical road profile.

The major findings of this thesis are summarized in the last part: conclusion and outlook (Chap. 8). Furthermore, some thoughts for further works are developed.

Part I

Fundamentals

2. Adjustment Theory

In this chapter, some basic principles of adjustment theory are reviewed and the mathematical concepts applied are introduced. The distinction between models (e.g., the Gauß-Markov model) and estimators (e.g., the L_2 norm estimator) is discussed. Special emphasis is put on rank-deficient systems.

Due to the overlap of adjustment theory and optimization mentioned in the introduction, some topics (such as optimality conditions) could have been assigned to either of both chapters. These topics are mentioned as soon as they are necessary for understanding the following concepts.

2.1 Mathematical Foundations

In order to facilitate the understanding of this thesis and to counteract possible ambiguities, some basic mathematical concepts are reviewed and necessary expressions are defined. This includes quadratic forms and convexity as well as minimization problems and the Gauss-Jordan algorithm, which will be important when dealing with rank-deficient systems.

2.1.1 Quadratic Forms and the Definiteness of a Matrix

The expression

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} - \boldsymbol{c}^T\boldsymbol{x} + c, \tag{2.1}$$

is called a *quadratic form*. $\boldsymbol{C}$ is a symmetric $m \times m$ matrix, $\boldsymbol{c}$ and $\boldsymbol{x}$ are $m \times 1$ vectors and c is a scalar quantity. Depending on the definiteness of $\boldsymbol{C}$, four different cases can be distinguished (cf. Shewchuk, 1994).

$\boldsymbol{C}$ is positive definite: For every vector $\boldsymbol{x} \neq \boldsymbol{0}$

$$\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} > 0 \tag{2.2}$$

holds true. All eigenvalues of a positive definite matrix are positive. A quadratic form with a positive definite matrix $\boldsymbol{C}$ is depicted in Fig. 2.1(a).

$\boldsymbol{C}$ is negative definite: For every vector $\boldsymbol{x} \neq \boldsymbol{0}$

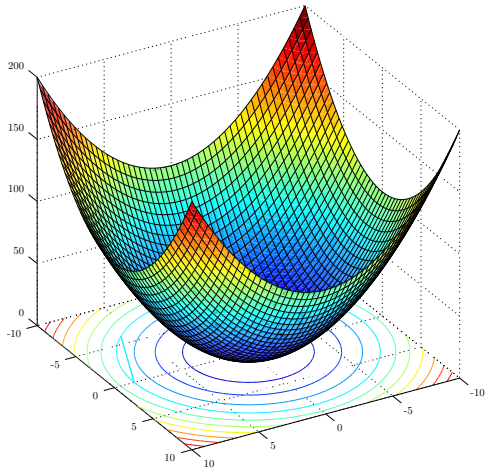
$$\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} < 0 \tag{2.3}$$

holds true. All eigenvalues of a negative definite matrix are negative. A quadratic form with a negative definite matrix $\boldsymbol{C}$ is depicted in Fig. 2.1(b).

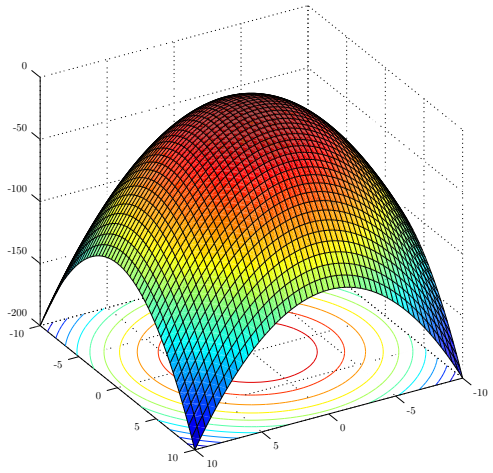
$\boldsymbol{C}$ is positive semi-definite: All eigenvalues of a positive semi-definite matrix are non-negative (i.e., positive or zero). If at least one eigenvalue is zero, $\boldsymbol{C}$ has a rank defect and is called singular. A quadratic form with a positive semi-definite matrix $\boldsymbol{C}$ with a rank defect of one is depicted in Fig. 2.1(c).

$\boldsymbol{C}$ is indefinite: An indefinite matrix has positive as well as negative eigenvalues. As a consequence, the corresponding quadratic form has a saddle point (cf. Fig. 2.1(d)).

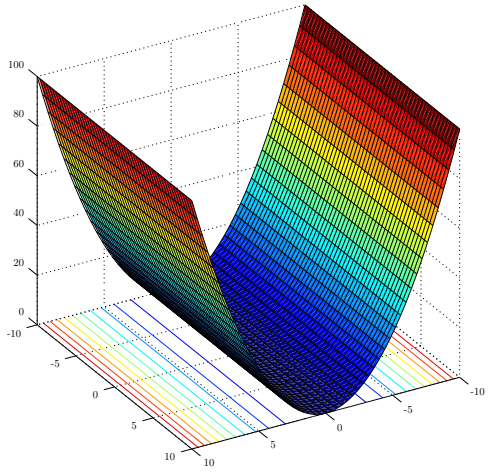
Within this work, especially positive (semi-)definite matrices will be used.



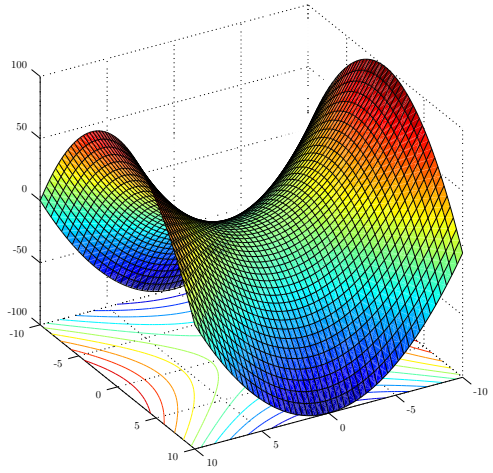
(a) Positive definite matrix
(Eigenvalues: $\lambda_1 = \lambda_2 = 1$)



(b) Negative definite matrix
(Eigenvalues: $\lambda_1 = \lambda_2 = -1$)



(c) Positive semi-definite matrix
(Eigenvalues: $\lambda_1 = 0, \lambda_2 = 1$)



(d) Indefinite Matrix
(Eigenvalues: $\lambda_1 = -1, \lambda_2 = 1$)

Figure 2.1: Quadratic forms in the bivariate case for matrices with a different type of definiteness. Modified and extended from Shewchuk (1994).

2.1.2 Optimality Conditions for Unconstrained Problems

In order to find an optimal solution an extremum problem has to be solved. Depending on the task, optimization can either mean to minimize or to maximize an objective function. As it is always possible to transform a maximization problem into a minimization problem by multiplying the objective function by minus one, only minimization problems will be addressed. In the following, it is assumed that the objective function is a quadratic form and thus twice differentiable.

The *necessary condition* for a minimum of a function is that its gradient vanishes in the critical point $\widetilde{\boldsymbol{x}}$. For the quadratic form (2.1), this leads to

$$\nabla f(\boldsymbol{x})|_{\boldsymbol{x}=\widetilde{\boldsymbol{x}}} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x})|_{\boldsymbol{x}=\widetilde{\boldsymbol{x}}} := \left[\begin{array}{c} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_m} \end{array}\right]\Bigg|_{\boldsymbol{x}=\widetilde{\boldsymbol{x}}} = \boldsymbol{C}\widetilde{\boldsymbol{x}} - \boldsymbol{c} = \boldsymbol{0}, \tag{2.4}$$

with ∇ being the gradient operator. In case of a positive (semi-)definite Hessian matrix, this condition is not only necessary, but also *sufficient*. This can easily be verified by examining (2.1). Its Hessian reads

$$\nabla\left(\nabla f(\boldsymbol{x})\right) = \boldsymbol{C}^T = \boldsymbol{C}. \tag{2.5}$$

If $\widetilde{\boldsymbol{x}}$ is a minimum, there is no point $\boldsymbol{x}$ with a lower value of the objective function. An application of Taylor's formula yields

$$f(\boldsymbol{x}) = f(\widetilde{\boldsymbol{x}}) + \underbrace{\nabla f(\boldsymbol{x})|_{\boldsymbol{x}=\widetilde{\boldsymbol{x}}}}_{=\boldsymbol{0}}(\boldsymbol{x}-\widetilde{\boldsymbol{x}}) + \underbrace{\frac{1}{2}(\boldsymbol{x}-\widetilde{\boldsymbol{x}})^T\boldsymbol{C}(\boldsymbol{x}-\widetilde{\boldsymbol{x}})}_{\geq\boldsymbol{0},\text{ if }\boldsymbol{C}\text{ is pos. semi-def.}}. \tag{2.6}$$

If $\boldsymbol{C}$ is positive (semi-)definite, there is no point $\boldsymbol{x}$ for which (2.6) has a lower value than for $\widetilde{\boldsymbol{x}}$. Thus, $\widetilde{\boldsymbol{x}}$ is the desired minimizer.

If $\boldsymbol{C}$ is positive definite, $\boldsymbol{x}$ is the unique minimizer of (2.1). If $\boldsymbol{C}$ is positive semi-definite, there is a manifold of solutions (cf. Sect. 2.1.3.2).

2.1.3 Systems of Linear Equations (SLE)

Estimating optimal parameters usually includes solving a system of linear equations (SLE). This can be done in many different ways. In the following, the Gauss-Jordan algorithm is described as one possible method for solving a system of linear equations. Furthermore, the solution of rank-deficient systems is explained.

2.1.3.1 Gauss-Jordan Algorithm

The Gauss-Jordan algorithm is a method to solve a system of linear equations with either one right-hand side $\boldsymbol{n}$, i.e.

$$\boldsymbol{N}_{[m\times m]}\boldsymbol{x}_{[m\times 1]} = \boldsymbol{n}_{[m\times 1]},$$

or multiple right-hand sides $\boldsymbol{n}_i$, i.e.

$$\boldsymbol{N}_{[m\times m]}\boldsymbol{X}_{[m\times b]} = \begin{bmatrix}\boldsymbol{n}_1 & \boldsymbol{n}_2 & \dots & \boldsymbol{n}_b\end{bmatrix}_{[m\times b]}.$$

Its most basic form is shown in Alg. 1. For a practical implementation, a version with pivoting (cf. Press et al., 2007, p. 41–46) is highly recommended to gain numerical stability. "MATLAB notation" is used to denote the access to specify certain parts or components of a matrix or vector. $\boldsymbol{G}(i,:)$ represents the i-th row of the Gauss-Jordan matrix and $\boldsymbol{G}(:,j)$ its j-th column. If $\boldsymbol{N}$ is of full rank m and the right-hand sides $\boldsymbol{n}_i$ are contained in its column space, the result of the Gauss-Jordan algorithm can be written as

$$\boldsymbol{G} = \begin{bmatrix}\boldsymbol{I}_{[m\times m]} & \boldsymbol{X}_{[m\times b]}\end{bmatrix} \tag{2.7}$$

and thus contains the desired solution $\boldsymbol{X}$ of the linear system (cf. line 16 of Alg. 1). The algorithm can be used to compute the inverse of a matrix, too (cf. Press et al., 2007, p. 41–46). To obtain for example the inverse $\boldsymbol{N}^{-1}$, the columns of the identity matrix $\boldsymbol{I}$ have to be used as right-hand sides, yielding

$$\boldsymbol{N}\boldsymbol{N}^{-1} = \boldsymbol{I}. \tag{2.8}$$

This will be helpful when dealing with rank-deficient linear systems as described in the next section.

Algorithm 1: Basic Gauss-Jordan algorithm (without pivoting)

// Reduces and solves the system of linear equations $\boldsymbol{AX} = \boldsymbol{B}$

Data:

$\boldsymbol{A}_{[m\times m]}, \boldsymbol{B}_{[m\times b]}$	...	Square Matrix $\boldsymbol{A}$ and b corresponding right-hand sides comprised in $\boldsymbol{B}$
$i_{\max}$	...	Maximal number of reduction steps ($i_{\max} \leq m-1$)
ϵ	...	Defintion of numeric zero

Result:

$\boldsymbol{X}_{[m\times b]}$	...	Solution of the system of linear equations

$\boldsymbol{G} = [\boldsymbol{A}|\boldsymbol{B}]$; *// Gauss-Jordan matrix: expand* $\boldsymbol{A}$ *with the right-hand sides* $\boldsymbol{B}$
$\boldsymbol{G}(1,:) = \boldsymbol{G}(1,:)/\boldsymbol{G}(1,1)$;
// Loop over all rows
for $(i = 1 : i_{\max})$ **do**
  for $(j = i+1 : m)$ **do**
    $\boldsymbol{G}(j,:) = \boldsymbol{G}(j,:) - \boldsymbol{G}(j,i) * \boldsymbol{G}(i,:)$;
  end
  if $\boldsymbol{G}(i+1,i+1) < \epsilon$ **then**
    break;
  end
  $\boldsymbol{G}(i+1,:) = \boldsymbol{G}(i+1,:)/\boldsymbol{G}(i+1,i+1)$;
  for $(k = 1 : i)$ **do**
    $\boldsymbol{G}(k,:) = \boldsymbol{G}(k,:) - \boldsymbol{G}(k,i+1) * \boldsymbol{G}(i+1,:)$;
  end
end
$\boldsymbol{X} = \boldsymbol{G}(:, m+1 : m+b)$; *// extract solution* $\boldsymbol{X}$ *from* $\boldsymbol{G}$ *(cf.* (2.7)*)*
return $\boldsymbol{X}$

2.1.3.2 Systems with Rank-Deficiencies

If the column rank r of an $m \times m$ matrix $\boldsymbol{N}$ is less than m, i.e.

$$\mathrm{Rg}(\boldsymbol{N}) = r < m, \tag{2.9}$$

then there are $d = m - r$ linearly dependent columns. The matrix has a rank defect of d and the linear system

$$\boldsymbol{N}\boldsymbol{x} = \boldsymbol{n} \tag{2.10a}$$

is underdetermined. Without loss of generality it is assumed that $\boldsymbol{N}$ can be rearranged and partitioned in a way that the $r \times r$ matrix $\boldsymbol{N}_{11}$ is of full rank

$$\begin{bmatrix} \underset{[r\times r]}{\boldsymbol{N}_{11}} & \underset{[r\times d]}{\boldsymbol{N}_{12}} \\ \underset{[d\times r]}{\boldsymbol{N}_{12}} & \underset{[d\times d]}{\boldsymbol{N}_{22}} \end{bmatrix} \begin{bmatrix} \underset{[r\times 1]}{\boldsymbol{x}_1} \\ \underset{[d\times 1]}{\boldsymbol{x}_2} \end{bmatrix} = \begin{bmatrix} \underset{[r\times 1]}{\boldsymbol{n}_1} \\ \underset{[d\times 1]}{\boldsymbol{n}_2} \end{bmatrix}. \tag{2.10b}$$

As (2.10) is an underdetermined system of linear equations, there is not one unique solution but a manifold. Therefore, the general solution

$$\widetilde{\boldsymbol{x}}(\boldsymbol{\lambda}) = \boldsymbol{x}_{\mathrm{P}} + \boldsymbol{x}_{\mathrm{hom}}(\boldsymbol{\lambda}) \tag{2.11}$$

consists of a particular solution $\boldsymbol{x}_{\mathrm{P}}$ and a solution $\boldsymbol{x}_{\mathrm{hom}}(\boldsymbol{\lambda})$ of the homogeneous system

$$\boldsymbol{N}\boldsymbol{x}_{\mathrm{hom}}(\boldsymbol{\lambda}) = \boldsymbol{0}. \tag{2.12}$$

$\boldsymbol{x}_{\mathrm{hom}}(\boldsymbol{\lambda})$ depends on d free parameters λ_i, which can be chosen arbitrarily.

In the following, two equivalent ways to determine a particular and a homogeneous solution of (2.10) are described. The first involves the theory of generalized inverses, while the latter uses the Gauss-Jordan algorithm.

Solution with Generalized Inverses. The homogeneous solution of (2.12) is given by

$$\boldsymbol{x}_{\mathrm{hom}}(\boldsymbol{\lambda}) = \boldsymbol{X}_{\mathrm{hom}}\ \boldsymbol{\lambda}. \tag{2.13}$$

The columns of the $m\ \times\ d$ matrix $\boldsymbol{X}_{\mathrm{hom}}$ form a basis for the nullspace of $\boldsymbol{N}$ yielding

$$\boldsymbol{N}\boldsymbol{X}_{\mathrm{hom}} = \boldsymbol{0}. \tag{2.14}$$

According to Koch (1999, p. 59) the matrix

$$\boldsymbol{X}_{\mathrm{hom}} = \begin{bmatrix} -\boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12} \\ \boldsymbol{I} \end{bmatrix} \tag{2.15}$$

fulfills (2.14) and is therefore a feasible choice. A particular solution of (2.10) can be obtained by applying the symmetric reflexive generalized inverse

$$\boldsymbol{N}_{\mathrm{RS}}^{-} = \begin{bmatrix} \boldsymbol{N}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{2.16}$$

as described e.g., in Koch (1999, p. 57), resulting in

$$\boldsymbol{x}_{\mathrm{P}} = \boldsymbol{N}_{\mathrm{RS}}^{-}\ \boldsymbol{n} = \begin{bmatrix} \boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1 \\ \boldsymbol{0} \end{bmatrix}. \tag{2.17}$$

In addition, $\boldsymbol{X}_{\mathrm{hom}}$ can be used to compute the Moore-Penrose inverse (also known as pseudo inverse)

$$\boldsymbol{N}^{+} = \left(\boldsymbol{N} + \boldsymbol{X}_{\mathrm{hom}}\boldsymbol{X}_{\mathrm{hom}}^{T}\right)^{-1} - \boldsymbol{X}_{\mathrm{hom}}\left(\boldsymbol{X}_{\mathrm{hom}}^{T}\boldsymbol{X}_{\mathrm{hom}}\boldsymbol{X}_{\mathrm{hom}}^{T}\boldsymbol{X}_{\mathrm{hom}}\right)^{-1}\boldsymbol{X}_{\mathrm{hom}}^{T} \tag{2.18}$$

(cf. Koch, 1999, p. 61).

Solution with the Gauss-Jordan Algorithm. An equivalent way to obtain the particular and the homogeneous solution mentioned above can be developed using the Gauss-Jordan algorithm. Applying Alg. 1 to (2.10) in a symbolic way yields

$$\begin{array}{cc|c|l}
\boldsymbol{N}_{11} & \boldsymbol{N}_{12} & \boldsymbol{n}_1 & |\boldsymbol{N}_{11}^{-1}\cdot(I) \\
\boldsymbol{N}_{21} & \boldsymbol{N}_{22} & \boldsymbol{n}_2 & \\
\hline
\boldsymbol{I} & \boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12} & \boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1 & \\
\boldsymbol{N}_{21} & \boldsymbol{N}_{22} & \boldsymbol{n}_2 & |(II) - \boldsymbol{N}_{21}\cdot(I) \\
\hline
\boldsymbol{I} & \boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12} & \boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1 & \\
\boldsymbol{0} & \boldsymbol{N}_{22} - \boldsymbol{N}_{21}\boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12} & \boldsymbol{n}_2 - \boldsymbol{N}_{21}\boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1. &
\end{array} \tag{2.19}$$

As $\boldsymbol{N}$ has a rank defect of d and thus d linearly dependent columns the Schur form

$$\boldsymbol{N}_{22} - \boldsymbol{N}_{21}\boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12} \tag{2.20}$$

in the last line of (2.19) is equal to zero. If $\boldsymbol{n}$ is an element of the column space of $\boldsymbol{N}$, the same holds true for

$$\boldsymbol{n}_2 - \boldsymbol{N}_{21}\boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1, \tag{2.21}$$

resulting in

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1 \\ \boldsymbol{0} \end{bmatrix}. \tag{2.22a}$$

Reformulating yields

$$\boldsymbol{x}_1 + \boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12}\boldsymbol{x}_2 = \boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1, \tag{2.22b}$$

thus

$$\boldsymbol{x}_1 = \boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1 - \boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12}\boldsymbol{x}_2. \tag{2.22c}$$

With the choice $\boldsymbol{x}_2 := \boldsymbol{\lambda}$, the general solution

$$\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{N}_{11}^{-1}\boldsymbol{n}_1 \\ \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{x}_\mathrm{P}} + \underbrace{\begin{bmatrix} -\boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12} \\ \boldsymbol{I} \end{bmatrix}}_{\boldsymbol{X}_\mathrm{hom}} \boldsymbol{\lambda} \tag{2.23}$$

follows. The particular and the homogeneous solution are clearly the same as those in (2.17) and (2.15). Throughout this thesis, the Gauss-Jordan algorithm will be used to solve underdetermined systems of linear equations.

2.2 Models

In this section, the three fundamental models of adjustment theory are introduced. The *Gauss-Markov Model* (GMM), the *adjustment with condition equations* and the *Gauss-Helmert Model* (GHM). It is well known that the first two models can be seen as special cases of the latter (cf. Niemeier, 2002, p. 152). Nonetheless, due to their importance in adjustment theory, all models are introduced separately and connections are pointed out.

Note that the term *model* (e.g., GMM, GHM, etc...) refers to the functional and stochastic model of an adjustment problem, only. In principle, this does not include an *estimator* (e.g., least-squares estimation, least absolute deviations, etc...), which is a statement about the objective function to minimize (e.g., the sum of squared residuals). However—if not mentioned otherwise—the least-squares estimator is used in all models.

2.2.1 Gauss-Markov Model (GMM)

In the following, the GMM is reviewed in its unconstrained and its equality constrained form. Furthermore, the well-known *weighted least-squares* (WLS) estimate in the GMM is introduced.

2.2.1.1 Unconstrained Gauss-Markov Model

Let the functional model, i.e., the relationship between the n observations ℓ_i and the m unknown parameters x_j, be defined as

$$\begin{aligned} l_1 + v_1 &= f(x_1, x_2, \ldots, x_m) \\ l_2 + v_2 &= f(x_1, x_2, \ldots, x_m) \\ &\vdots \\ l_n + v_n &= f(x_1, x_2, \ldots, x_m). \end{aligned}$$

The n residuals v_i are unknown as well. If this relationship is linear, the observation equations can be written in matrix vector form

$$\boldsymbol{\ell} + \boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}. \tag{2.24}$$

$\boldsymbol{\ell}$ is the $n \times 1$ vector of observations and $\boldsymbol{x}$ the $m \times 1$ vector of parameters. The $n \times m$ design matrix $\boldsymbol{A}$ is assumed to be of full rank m (if not mentioned otherwise) and the n residuals are comprised in

$$\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{\ell}. \tag{2.25}$$

The random vector $\boldsymbol{\mathcal{L}}$ is here assumed to be normally distributed with known variance-covariance (VCV) matrix $\boldsymbol{\Sigma}$, i.e.

$$\boldsymbol{\mathcal{L}} \sim N(\boldsymbol{A}\boldsymbol{\xi}, \boldsymbol{\Sigma}). \tag{2.26}$$

$\boldsymbol{\xi}$ is the true parameter vector. The observations comprised in $\boldsymbol{\ell}$ are assumed to be a realization of the random vector $\boldsymbol{\mathcal{L}}$. In case of a non-linear relationship, the observation equations have to be linearized first (cf. Koch, 1999, p. 155–156) using a Taylor approximation of degree one

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) \approx \underbrace{f(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}_0}}_{\boldsymbol{\ell}_0} + \underbrace{\nabla_{\boldsymbol{x}} f(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}_0}}_{\boldsymbol{A}} \Delta\boldsymbol{x}, \tag{2.27}$$

resulting in

$$\boldsymbol{\ell} - \boldsymbol{\ell}_0 + \boldsymbol{v} = \Delta\boldsymbol{\ell} + \boldsymbol{v} = \boldsymbol{A}\Delta\boldsymbol{x}. \tag{2.28}$$

As a consequence, the adjustment has to be carried out using an iterative scheme (also known as Gauss-Newton approach, cf. Nocedal and Wright, 1999, p. 259–262). In each iteration only the increment $\Delta\boldsymbol{x}$ is computed and the vector of parameters is updated using

$$\boldsymbol{x} = \boldsymbol{x}_0 + \Delta\boldsymbol{x}. \tag{2.29}$$

In the following we will assume that the observation equations are linear. Minimizing the weighted sum of squared residuals

$$\begin{aligned} \Phi(\boldsymbol{x}) &= \boldsymbol{v}(\boldsymbol{x})^T \boldsymbol{\Sigma}^{-1} \boldsymbol{v}(\boldsymbol{x}) && \text{(2.30a)} \\ &= (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{\ell})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{\ell}) && \text{(2.30b)} \\ &= \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} && \text{(2.30c)} \\ &= \boldsymbol{x}^T \boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n}^T \boldsymbol{x} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell}, && \text{(2.30d)} \end{aligned}$$

with the substitutions

$$\begin{aligned} \boldsymbol{N} &= \boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{A}, && \text{(2.31a)} \\ \boldsymbol{n} &= \boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell}, && \text{(2.31b)} \end{aligned}$$

leads to the *Weighted Least-Squares (WLS) adjustment in the Gauss-Markov model*:

Weighted Least-Squares (WLS) adjustment in the Gauss-Markov model	
objective function:	$\Phi(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n}^T\boldsymbol{x} + \boldsymbol{\ell}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} \ldots \text{Min}$
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m$.

(2.32)

Closed formulas for the solution can be derived by setting the gradient of the objective function with respect to $\boldsymbol{x}$ equal to zero (cf. Sect. 2.1.2)

$$\nabla_{\boldsymbol{x}}\Phi(\boldsymbol{x}) = 2\boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n} \stackrel{!}{=} \boldsymbol{0}, \tag{2.33}$$

yielding the normal equation system

$$\boldsymbol{N}\widetilde{\boldsymbol{x}} = \boldsymbol{n} \tag{2.34}$$

and the least-squares estimator in the GMM (cf. Koch, 1999, p. 158)

$$\widetilde{\boldsymbol{x}} = \boldsymbol{N}^{-1}\boldsymbol{n} \tag{2.35a}$$

$$= \left(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell}. \tag{2.35b}$$

The VCV matrix of the estimated parameters is given by

$$\boldsymbol{\Sigma}\{\widetilde{\boldsymbol{\mathcal{X}}}\} = \boldsymbol{N}^{-1} = \left(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}. \tag{2.36}$$

The (estimated) residuals read

$$\boldsymbol{v} = \boldsymbol{A}\widetilde{\boldsymbol{x}} - \boldsymbol{\ell} \tag{2.37}$$

and the adjusted observations are given by

$$\widetilde{\boldsymbol{\ell}} = \boldsymbol{\ell} + \boldsymbol{v}. \tag{2.38}$$

2.2.1.2 Equality Constrained Gauss-Markov Model

If there exist $\bar{p}$ linear equality constraints

$$\bar{\boldsymbol{B}}^T\boldsymbol{x} = \bar{\boldsymbol{b}}, \tag{2.39}$$

which have to be strictly fulfilled, (2.32) has to be extended to the *Equality Constrained (weighted) Least-Squares (ECLS) adjustment in the Gauss-Markov model*:

Equality constrained least-squares (ECLS) adjustment in the GMM	
objective function:	$\Phi(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n}^T\boldsymbol{x} + \boldsymbol{\ell}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} \ldots \text{Min}$
constraints:	$\bar{\boldsymbol{B}}^T\boldsymbol{x} = \bar{\boldsymbol{b}}$
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m$.

(2.40)

$\bar{\boldsymbol{B}}$ is the $m \times \bar{p}$ matrix of equality constraints and the $\bar{p} \times 1$ vector $\bar{\boldsymbol{b}}$ is the corresponding right-hand side. Here and in the following, all quantities originating from equality constrained problems are marked by a bar to distinguish them from inequality constraints. The Lagrangian

$$L(\boldsymbol{x}, \bar{\boldsymbol{k}}) = \Phi(\boldsymbol{x}) + 2\bar{\boldsymbol{k}}^T(\bar{\boldsymbol{B}}^T\boldsymbol{x} - \bar{\boldsymbol{b}}) \tag{2.41a}$$

$$= \boldsymbol{x}^T\boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n}^T\boldsymbol{x} + \boldsymbol{\ell}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} + 2\bar{\boldsymbol{k}}^T(\bar{\boldsymbol{B}}^T\boldsymbol{x} - \bar{\boldsymbol{b}}), \tag{2.41b}$$

can be used to compute necessary and sufficient optimality conditions of the constrained optimization problem (2.40) (cf. Koch, 1999, p. 171–172).

The $\bar{p} \times 1$ vector $\bar{\boldsymbol{k}}$ contains the Lagrange multipliers. We will see later in Sect. 4.3.3 that—together with the matrix of constraints $\bar{\boldsymbol{B}}$ and the normal equation matrix $\boldsymbol{N}$—they can be used as a measure for the distortion through the associated constraint (cf. Lehmann and Neitzel, 2013). However, one should be careful as the **absolute value** of the Lagrange multipliers does not matter as they are dimensionless. Therefore, often a multiple $\alpha\bar{\boldsymbol{k}}$ is used instead of the Lagrange multipliers $\bar{\boldsymbol{k}}$ themselves, in order to obtain convenient equations. Throughout this thesis, we will use this whenever it seems appropriate. Sometimes, even **negative** scaling factors α are applied to the Lagrange multipliers when dealing with **equality constraints**. However, this should not be done with Lagrange multipliers that are linked to inequality constraints as their sign is important for the interpretation (cf. Sect. 3.5.1.6).

The gradients of the Lagrangian (2.41a) read

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\bar{\boldsymbol{k}}) = 2\boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n} + 2\bar{\boldsymbol{B}}\bar{\boldsymbol{k}} \tag{2.42}$$

$$\nabla_{\bar{\boldsymbol{k}}} L(\boldsymbol{x},\bar{\boldsymbol{k}}) = 2(\bar{\boldsymbol{B}}^T\boldsymbol{x} - \bar{\boldsymbol{b}}). \tag{2.43}$$

The extended normal equations

$$\begin{bmatrix} \boldsymbol{N} & \bar{\boldsymbol{B}} \\ \bar{\boldsymbol{B}}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \bar{\boldsymbol{k}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{n} \\ \bar{\boldsymbol{b}} \end{bmatrix} \tag{2.44}$$

result from setting the rearranged gradients equal to zero. Thus, the estimated parameters $\widetilde{\boldsymbol{x}}$ can be computed by solving (2.44) and the values of the residuals and adjusted observations can be obtained by evaluating (2.37) and (2.38), respectively. The same holds true for the VCV matrix of the estimated parameters which can be extracted from the inverse of the extended normal equation matrix.

2.2.2 Adjustment with Condition Equations

Instead of a relationship between observations and parameters, it can be more convenient to establish r conditions between two or more observations

$$g_1(\ell_1, \ell_2, \ldots, \ell_n) \stackrel{!}{=} 0 \tag{2.45a}$$

$$\vdots$$

$$g_r(\ell_1, \ell_2, \ldots, \ell_n) \stackrel{!}{=} 0. \tag{2.45b}$$

This ansatz is especially well suited for problems with a small redundancy $r = n - m$. If the conditions are linear, these condition equations can easily be expressed in matrix vector notation

$$\bar{\boldsymbol{B}}^T(\boldsymbol{\ell} + \boldsymbol{v}) - \bar{\boldsymbol{b}} = \boldsymbol{0}, \tag{2.46a}$$

leading to

$$\bar{\boldsymbol{B}}^T\boldsymbol{v} + \bar{\boldsymbol{B}}^T\boldsymbol{\ell} - \bar{\boldsymbol{b}} = \boldsymbol{0} \tag{2.46b}$$

and thus

$$\bar{\boldsymbol{B}}^T\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{0}, \tag{2.46c}$$

with the substitution

$$\boldsymbol{w} := \bar{\boldsymbol{B}}^T \boldsymbol{\ell} - \bar{\boldsymbol{b}}. \tag{2.47}$$

If the condition equations (2.45) are non-linear, they have to be linearized first, using a Taylor approximation of degree one

$$\begin{aligned}
\boldsymbol{g}(\widetilde{\boldsymbol{\ell}}) &= \boldsymbol{g}(\boldsymbol{\ell}_0 + \Delta\boldsymbol{\ell} + \boldsymbol{v}) && \text{(2.48a)}\\
&= \underbrace{\boldsymbol{g}(\boldsymbol{\ell})|_{\boldsymbol{\ell}=\boldsymbol{\ell}_0}}_{\boldsymbol{w}_0} + \underbrace{\nabla_{\boldsymbol{\ell}}\, \boldsymbol{g}(\boldsymbol{\ell})|_{\boldsymbol{\ell}=\boldsymbol{\ell}_0}}_{\bar{\boldsymbol{B}}^T}(\Delta\boldsymbol{\ell} + \boldsymbol{v}) && \text{(2.48b)}\\
&= \boldsymbol{w}_0 + \bar{\boldsymbol{B}}^T(\Delta\boldsymbol{\ell} + \boldsymbol{v}) && \text{(2.48c)}\\
&= \underbrace{\boldsymbol{w}_0 + \bar{\boldsymbol{B}}^T\Delta\boldsymbol{\ell}}_{\boldsymbol{w}} + \bar{\boldsymbol{B}}^T\boldsymbol{v} && \text{(2.48d)}\\
&= \bar{\boldsymbol{B}}^T\boldsymbol{v} + \boldsymbol{w} = 0 && \text{(2.48e)}
\end{aligned}$$

and the adjustment has to be carried out using an iterative scheme.

Minimizing the sum of squared residuals in this model yields the *adjustment with condition equations* (cf. Koch, 1999, p. 220–221):

ADJUSTMENT WITH CONDITION EQUATIONS

objective function:	$\Phi(\boldsymbol{v}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} \ldots \text{Min}$
constraints:	$\bar{\boldsymbol{B}}^T\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{0}$
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n$.

(2.49)

The normal equations read

$$\left[\bar{\boldsymbol{B}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}}\right]\boldsymbol{k} = -\boldsymbol{w}, \tag{2.50}$$

and the optimal values for the Lagrange multipliers $\bar{\boldsymbol{k}}$, the residuals $\boldsymbol{v}$ and the adjusted observations $\widetilde{\boldsymbol{\ell}}$ can be obtained via

$$\begin{aligned}
\widetilde{\bar{\boldsymbol{k}}} &= -(\bar{\boldsymbol{B}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}})^{-1}\boldsymbol{w}, && \text{(2.51)}\\
\boldsymbol{v} &= \boldsymbol{\Sigma}\bar{\boldsymbol{B}}\,\widetilde{\bar{\boldsymbol{k}}}, && \text{(2.52)}\\
\widetilde{\boldsymbol{\ell}} &= \boldsymbol{\ell} + \boldsymbol{v}. && \text{(2.53)}
\end{aligned}$$

A proof using duality theory is given in Sect. 3.4.1. This proof is an alternative to the classical ones and spans a bridge between optimization and geodetic adjustment theory. The VCV matrix of the adjusted observations reads

$$\boldsymbol{\Sigma}\{\widetilde{\boldsymbol{\mathcal{L}}}\} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\bar{\boldsymbol{B}}(\bar{\boldsymbol{B}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}})^{-1}\bar{\boldsymbol{B}}^T\boldsymbol{\Sigma}. \tag{2.54}$$

2.2.3 Gauss-Helmert Model

The Gauss-Helmert Model (GHM, Helmert, 1872, p. 40–41) can be described as a combination of the Gauss-Markov model (cf. Sect. 2.2.1) and the adjustment with condition equations (cf. Sect. 2.2.2).

2.2.3.1 Unconstrained Gauss-Helmert Model (GHM)

Given

$$\begin{aligned} \boldsymbol{\ell} &\dots n \times 1 \\ \boldsymbol{\Sigma} = \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{L}}\} &\dots n \times n \end{aligned}$$

we assume b possibly non-linear functional relationships between one or more observation(s) and up to m parameters

$$\boldsymbol{g}(\boldsymbol{\ell}, \boldsymbol{x}) \stackrel{!}{=} \boldsymbol{0}, \tag{2.55}$$

which will be introduced as equality constraints. To linearize the functional relationship (2.55), one has to split up the quantities

$$\widetilde{\boldsymbol{x}} = \boldsymbol{x}_0 + \Delta\boldsymbol{x} \tag{2.56}$$

$$\widetilde{\boldsymbol{\ell}} = \boldsymbol{\ell} + \boldsymbol{v} = \underbrace{\boldsymbol{\ell} - \boldsymbol{\ell}_0}_{\Delta\boldsymbol{\ell}} + \boldsymbol{\ell}_0 + \boldsymbol{v} = \boldsymbol{\ell}_0 + \Delta\boldsymbol{\ell} + \boldsymbol{v} \tag{2.57}$$

and choose a Taylor point related to both parameters **and** computed observations (cf. Lenzmann and Lenzmann, 2004, Koch, 2014)

$$\boldsymbol{z}_0 = \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{\ell}_0 \end{bmatrix}. \tag{2.58}$$

Linearization using a Taylor series and the Taylor point mentioned above yields

$$\begin{aligned} \boldsymbol{g}(\widetilde{\boldsymbol{\ell}}, \widetilde{\boldsymbol{x}}) &= \boldsymbol{g}(\boldsymbol{\ell}_0 + \Delta\boldsymbol{\ell} + \boldsymbol{v}, \boldsymbol{x}_0 + \Delta\boldsymbol{x}) \\ &= \underbrace{\boldsymbol{g}(\boldsymbol{\ell}, \boldsymbol{x})|_{\boldsymbol{\ell}=\boldsymbol{\ell}_0, \boldsymbol{x}=\boldsymbol{x}_0}}_{\boldsymbol{w}_0} + \underbrace{\nabla_{\boldsymbol{\ell}}\, \boldsymbol{g}(\boldsymbol{\ell}, \boldsymbol{x})|_{\boldsymbol{\ell}=\boldsymbol{\ell}_0, \boldsymbol{x}=\boldsymbol{x}_0}}_{\overline{\boldsymbol{B}}^T_{\mathrm{GHM}}} (\Delta\boldsymbol{\ell} + \boldsymbol{v}) + \underbrace{\nabla_{\boldsymbol{x}}\, \boldsymbol{g}(\boldsymbol{\ell}, \boldsymbol{x})|_{\boldsymbol{\ell}=\boldsymbol{\ell}_0, \boldsymbol{x}=\boldsymbol{x}_0}}_{\boldsymbol{A}} \Delta\boldsymbol{x} \\ &= \boldsymbol{w}_0 + \overline{\boldsymbol{B}}^T_{\mathrm{GHM}} (\Delta\boldsymbol{\ell} + \boldsymbol{v}) + \boldsymbol{A}\Delta\boldsymbol{x} \\ &= \underbrace{\boldsymbol{w}_0 + \overline{\boldsymbol{B}}^T_{\mathrm{GHM}} \Delta\boldsymbol{\ell}}_{\boldsymbol{w}} + \overline{\boldsymbol{B}}^T_{\mathrm{GHM}} \boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} \\ &= \boldsymbol{w} + \overline{\boldsymbol{B}}^T_{\mathrm{GHM}} \boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} = 0. \end{aligned} \tag{2.59}$$

With the least-squares objective function

$$\boldsymbol{v}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{v} \dots \min,$$

the *weighted least-squares adjustment in the Gauss-Helmert model* can be stated:

Weighted least-squares adjustment in the Gauss-Helmert model	
objective function:	$\Phi(\boldsymbol{v}) = \boldsymbol{v}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{v} \dots \mathrm{Min}$
constraints:	$\overline{\boldsymbol{B}}^T_{\mathrm{GHM}} \boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = 0$
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n, \Delta\boldsymbol{x} \in \mathbb{R}^m.$

(2.60)

The corresponding Lagrangian reads

$$L(\boldsymbol{v}, \Delta\boldsymbol{x}, \bar{\boldsymbol{k}}_{\mathrm{GHM}}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} - 2\bar{\boldsymbol{k}}_{\mathrm{GHM}}^T(\bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w}). \tag{2.61}$$

Once again setting the gradients to zero yields the first order optimality conditions

$$\nabla_{\boldsymbol{v}} L(\boldsymbol{v}, \Delta\boldsymbol{x}, \bar{\boldsymbol{k}}_{\mathrm{GHM}}) = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{v} - 2\bar{\boldsymbol{B}}_{\mathrm{GHM}}\bar{\boldsymbol{k}}_{\mathrm{GHM}} \stackrel{!}{=} \boldsymbol{0} \tag{2.62}$$

$$\nabla_{\Delta\boldsymbol{x}} L(\boldsymbol{v}, \Delta\boldsymbol{x}, \bar{\boldsymbol{k}}_{\mathrm{GHM}}) = -2\boldsymbol{A}^T\bar{\boldsymbol{k}}_{\mathrm{GHM}} \stackrel{!}{=} \boldsymbol{0} \tag{2.63}$$

$$\nabla_{\bar{\boldsymbol{k}}_{\mathrm{GHM}}} L(\boldsymbol{v}, \Delta\boldsymbol{x}, \bar{\boldsymbol{k}}_{\mathrm{GHM}}) = -2(\bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w}) \stackrel{!}{=} \boldsymbol{0}, \tag{2.64}$$

leading to

$$\boldsymbol{v} = \boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\mathrm{GHM}}\bar{\boldsymbol{k}}_{\mathrm{GHM}} \tag{2.65}$$

$$\boldsymbol{A}^T\bar{\boldsymbol{k}}_{\mathrm{GHM}} = \boldsymbol{0} \tag{2.66}$$

$$\bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0}. \tag{2.67}$$

Inserting (2.65) in (2.67) yields

$$\bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\mathrm{GHM}}\bar{\boldsymbol{k}}_{\mathrm{GHM}} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0}. \tag{2.68}$$

In matrix vector notation, the normal equations (2.68) and (2.66) read

$$\begin{bmatrix} \bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\mathrm{GHM}} & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{k}}_{\mathrm{GHM}} \\ \Delta\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{0} \end{bmatrix}. \tag{2.69}$$

Solving this linear system yields an estimate $\widetilde{\Delta\boldsymbol{x}}$. $\widetilde{\boldsymbol{x}}, \boldsymbol{v}$ and $\widetilde{\boldsymbol{\ell}}$ can be computed using (2.56), (2.37) and (2.38), respectively. As this is a linearized form, iterations according to the Gauss-Newton method (cf. Nocedal and Wright, 1999, p. 259–262) may be necessary. It can be shown that a unique solution exists, if the design matrix is of full rank

$$\mathrm{Rg}(\boldsymbol{A}) = m \tag{2.70}$$

and the rank of the composed matrix $\begin{bmatrix} \bar{\boldsymbol{B}}_{\mathrm{GHM}}^T & \boldsymbol{A} \end{bmatrix}$ is equal to the number of constraints

$$\mathrm{Rg}(\begin{bmatrix} \bar{\boldsymbol{B}}_{\mathrm{GHM}}^T & \boldsymbol{A} \end{bmatrix}) = p. \tag{2.71}$$

As in the equality constrained GMM, the VCV matrix of the estimated parameters can be extracted from the inverse of the normal equation matrix (cf. (2.69)).

2.2.3.2 Equality Constrained Gauss-Helmert Model (ECGHM)

Now we assume that there exist $\bar{p}$ linear equality constraints concerning the parameters

$$\bar{\boldsymbol{B}}^T\boldsymbol{x} = \bar{\boldsymbol{b}}^* \tag{2.72a}$$

which have to be strictly fulfilled. Splitting up the parameter vector yields

$$\bar{\boldsymbol{B}}^T(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) = \bar{\boldsymbol{b}}^*, \tag{2.72b}$$

leading to

$$\bar{\boldsymbol{B}}^T\Delta\boldsymbol{x} = \bar{\boldsymbol{b}}^* - \bar{\boldsymbol{B}}^T\boldsymbol{x}_0 =: \bar{\boldsymbol{b}}. \tag{2.72c}$$

Thus, problem (2.60) has to be extended to the *Equality Constrained (weighted) least-squares adjustment in the Gauss-Helmert Model* (ECGHM):

Equality Constrained (weighted) LS adjustment in the GHM (ECGHM)		
objective function:	$\Phi(\boldsymbol{v}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} \ldots \text{Min}$	(2.73)
constraints:	$\bar{\boldsymbol{B}}_{\text{GHM}}^T\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = 0$ $\bar{\boldsymbol{B}}^T\Delta\boldsymbol{x} = \bar{\boldsymbol{b}}$	
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n, \Delta\boldsymbol{x} \in \mathbb{R}^m.$	

As in the unconstrained case the Lagrangian

$$L(\boldsymbol{v}, \Delta\boldsymbol{x}, \bar{\boldsymbol{k}}_{\text{GHM}}, \bar{\boldsymbol{k}}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} - 2\bar{\boldsymbol{k}}_{\text{GHM}}^T(\bar{\boldsymbol{B}}_{\text{GHM}}^T\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w}) - 2\bar{\boldsymbol{k}}^T(\bar{\boldsymbol{B}}^T\Delta\boldsymbol{x} - \bar{\boldsymbol{b}}) \tag{2.74}$$

can be used to compute a solution. Again the multiples and signs of the Lagrange multipliers of the inherent constraint $\bar{\boldsymbol{k}}_{\text{GHM}}$ and the additional equality constraint $\bar{\boldsymbol{k}}$ are chosen in a way to make the following computations as convenient as possible. Setting the gradients to zero yields

$$\nabla_{\boldsymbol{v}} L(\boldsymbol{v}, \bar{\boldsymbol{k}}_{\text{GHM}}) = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{v} - 2\bar{\boldsymbol{B}}_{\text{GHM}}\bar{\boldsymbol{k}}_{\text{GHM}} \stackrel{!}{=} \boldsymbol{0} \tag{2.75}$$

$$\nabla_{\Delta\boldsymbol{x}} L(\bar{\boldsymbol{k}}_{\text{GHM}}, \bar{\boldsymbol{k}}) = -2\boldsymbol{A}^T\bar{\boldsymbol{k}}_{\text{GHM}} - \bar{\boldsymbol{B}}\bar{\boldsymbol{k}} \stackrel{!}{=} \boldsymbol{0} \tag{2.76}$$

$$\nabla_{\bar{\boldsymbol{k}}_{\text{GHM}}} L(\boldsymbol{v}, \Delta\boldsymbol{x}) = -2(\bar{\boldsymbol{B}}_{\text{GHM}}^T\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w}) \stackrel{!}{=} \boldsymbol{0} \tag{2.77}$$

$$\nabla_{\bar{\boldsymbol{k}}} L(\Delta\boldsymbol{x}) = -\left(\bar{\boldsymbol{B}}^T\Delta\boldsymbol{x} - \bar{\boldsymbol{b}}\right) \stackrel{!}{=} \boldsymbol{0}, \tag{2.78}$$

leading to

$$\boldsymbol{v} = \boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\text{GHM}}\bar{\boldsymbol{k}}_{\text{GHM}} \tag{2.79}$$

$$\boldsymbol{A}^T\bar{\boldsymbol{k}}_{\text{GHM}} + \bar{\boldsymbol{B}}\bar{\boldsymbol{k}} = \boldsymbol{0} \tag{2.80}$$

$$\bar{\boldsymbol{B}}_{\text{GHM}}^T\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0} \tag{2.81}$$

$$\bar{\boldsymbol{B}}^T\Delta\boldsymbol{x} = \bar{\boldsymbol{b}}. \tag{2.82}$$

It should be noted that (2.79) and (2.81) are identical to the equivalent formulations (2.65) and (2.67) in the unconstrained GHM. Inserting (2.79) in (2.81) yields

$$\bar{\boldsymbol{B}}_{\text{GHM}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\text{GHM}}\bar{\boldsymbol{k}}_{\text{GHM}} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0}. \tag{2.83}$$

The equations (2.83), (2.80) and (2.82) can be combined to the system

$$\begin{bmatrix} \bar{\boldsymbol{B}}_{\text{GHM}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\text{GHM}} & \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{A}^T & \boldsymbol{0} & \bar{\boldsymbol{B}} \\ \boldsymbol{0} & \bar{\boldsymbol{B}}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{k}}_{\text{GHM}} \\ \Delta\boldsymbol{x} \\ \bar{\boldsymbol{k}} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{0} \\ \bar{\boldsymbol{b}} \end{bmatrix}. \tag{2.84}$$

Solving this system of linear equations yields estimates for the parameters $\widetilde{\Delta\boldsymbol{x}}$ (and $\widetilde{\boldsymbol{x}}$) and the Lagrange multipliers $\bar{\boldsymbol{k}}_{\text{GHM}}$ and $\bar{\boldsymbol{k}}$ corresponding to the inherent constraints and the additional equality constraints, respectively. As this is a linearized form, iterations according to the Gauss-Newton method (cf. Nocedal and Wright, 1999, p. 259–262) may be necessary. Again, the VCV matrix of the estimated parameters can be extracted from the inverse of the extended normal equation matrix (cf. (2.84)).

2.3 Estimators

One of the main targets in adjustment theory is the estimation of optimal parameters. However, there are different options to define optimality. Each of these requires a certain objective function $\Phi(\boldsymbol{v})$—which is a function of the residuals $\boldsymbol{v}$—to be minimal. Depending on that choice, different estimators can be distinguished.

In the following, it is assumed that all observations are independently and identically distributed. However, possible correlations and differing weights can be accounted for by introducing a metric $\boldsymbol{P}$ of the vector space $\boldsymbol{v}$ is defined in.

L_2 Norm Estimator. Hitherto, only the L_2 norm estimator was described. Its objective function reads

$$\Phi_{L_2}(\boldsymbol{v}) = ||\boldsymbol{v}||_2^2 = \boldsymbol{v}^T\boldsymbol{v} = v_1^2 + v_2^2 + ... + v_n^2 \tag{2.85}$$

and is convex. The L_2 norm estimator is the one most commonly used in adjustment theory and also known as *ordinary least-squares estimator*. In the linear(ized) GMM, it leads to the unbiased estimator with minimal variance (cf. Jäger et al., 2005, p. 105). However, it is not a robust estimator as its influence function (not shown here) is unbounded (cf. Peracchi, 2001, p. 508). Thus, the estimator is sensitive to outliers, in the sense of "an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data" (Barnett and Lewis, 1994, p. 7).

L_1 Norm Estimator. The objective function of the L_1 norm estimator (cf. Koch, 1999, p. 262) reads

$$\Phi_{L_1}(\boldsymbol{v}) = ||\boldsymbol{v}||_1 = |v_1| + |v_2| + \ldots + |v_n| \tag{2.86}$$

and is convex. It does not yield an estimate with minimal variance for normally distributed observations as the L_2 norm estimator does. However, this *least absolute deviations estimator* is a *robust estimator* and thus less sensitive concerning outliers.

Huber Estimator. The loss function of the Huber estimator

$$\rho_{\text{Huber}}(v) = \begin{cases} \frac{1}{2}v^2, & |v| \le k \\ k|v| - \frac{1}{2}k^2, & |v| > k \end{cases} \tag{2.87}$$

can be seen as a mixture of the L_1 and L_2 norm estimators. k is a tuning parameter, which is often set to 1.5 or 2.0, depending on the assumed portion of outliers in the data (cf. Koch, 1999, p. 260). The corresponding objective function is defined as the sum of the loss functions of all observations

$$\Phi(\boldsymbol{v}) = \sum_{i=1}^{n} \rho(v_i). \tag{2.88}$$

The Huber estimator is a robust estimator and has a convex objective function (cf. Boyd and Vandenberghe, 2004, p. 299).

Hampel Estimator. Due to its loss function

$$\rho_{\text{Hampel}}(v) = \begin{cases} \frac{1}{2}v^2, & |v| < k_1 \\ k_1 |v| - \frac{1}{2}k_1^2, & k_1 \leq |v| \leq k_2 \\ k_1 k_2 - \frac{1}{2}k_1^2 + \frac{1}{2}k_1(k_3 - k_2)\left[1 - \left(\frac{k_3 - |v|}{k_3 - k_2}\right)^2\right], & k_2 \leq |v| \leq k_3 \\ k_1 k_2 - \frac{1}{2}k_1^2 + \frac{1}{2}k_1(k_3 - k_2), & |v| > k_3 \end{cases}, \tag{2.89}$$

the Hampel estimator is a robust estimator, too. However, its loss function is not convex (cf. Suykens et al., 2003, p. 163). Standard values for the tuning parameters are $k_1 = 2, k_2 = 4, k_3 = 8$ (cf. Jäger et al., 2005, p. 119).

L_∞ Norm Estimator. The L_∞ norm objective function

$$\Phi_{L_\infty}(\boldsymbol{v}) = ||\boldsymbol{v}||_\infty = \lim_{p \to \infty} \left(|v_1|^p + |v_2|^p + \ldots + |v_n|^p\right)^{1/p} = \max |v_i| \tag{2.90}$$

depends only on the largest absolute value $v_{\max}$. Therefore, it is very sensitive to outliers. This estimator is also called *Chebyshev* or *Min-Max estimator* (cf. Jäger et al., 2005, p. 127) as the maximal residual is minimized. It has a convex objective function (cf. Boyd and Vandenberghe, 2004, p. 634–635 together with p. 72).

3. Convex Optimization

According to Rockafellar (1993), "the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity." As within this thesis, the focus is on *convex optimization*, some basic principles of convex optimization theory are reviewed in this chapter.

Special emphasis is put on inequality constrained optimization and a taxonomy of different optimization problems is established. *Quadratic programs* and *active-set algorithms* to solve them are introduced. In the optimization community the term *program* is used as a synonym for *optimization problem*. Furthermore, the concepts of *duality* and *feasibility* are explained.

3.1 Convexity

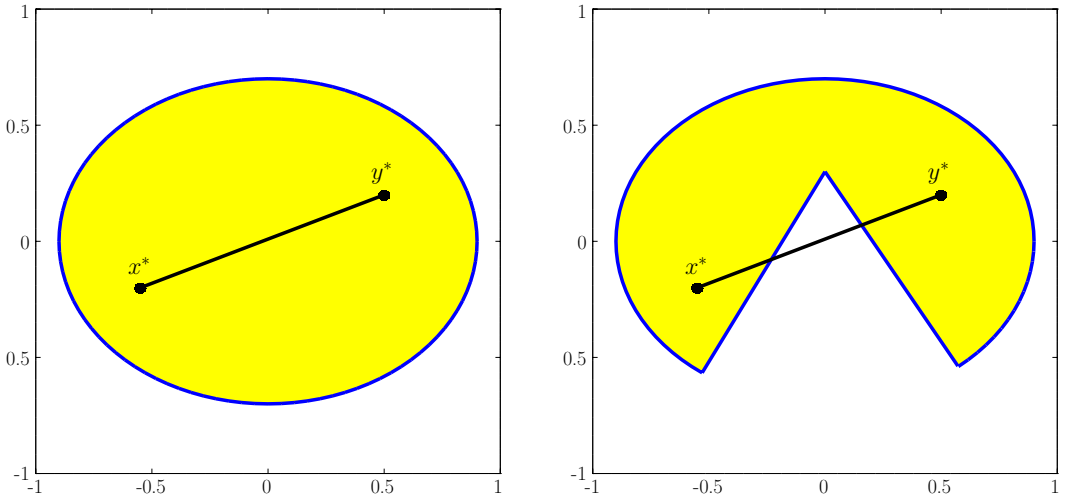
According to Boyd and Vandenberghe (2004, p. 137), *convex optimization* is defined as the task of minimizing a "convex objective function over a convex set". Therefore, the terms *convex function* and *convex set* are defined in the following.

3.1.1 Convex Set

A set $C \subseteq \mathbb{R}^m$ is convex if and only if

$$\alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{y} \in C \tag{3.1}$$

holds for any $\alpha \in (0, 1)$ and any $\boldsymbol{x}, \boldsymbol{y} \in C$ (cf. Boyd and Vandenberghe, 2004, p. 23). Geometrically, this means that the line segment between any two points $\boldsymbol{x}$ and $\boldsymbol{y}$ in C lies in C. This is visualized for the bivariate case in Fig. 3.1.



(a) Convex set (b) Non-convex set

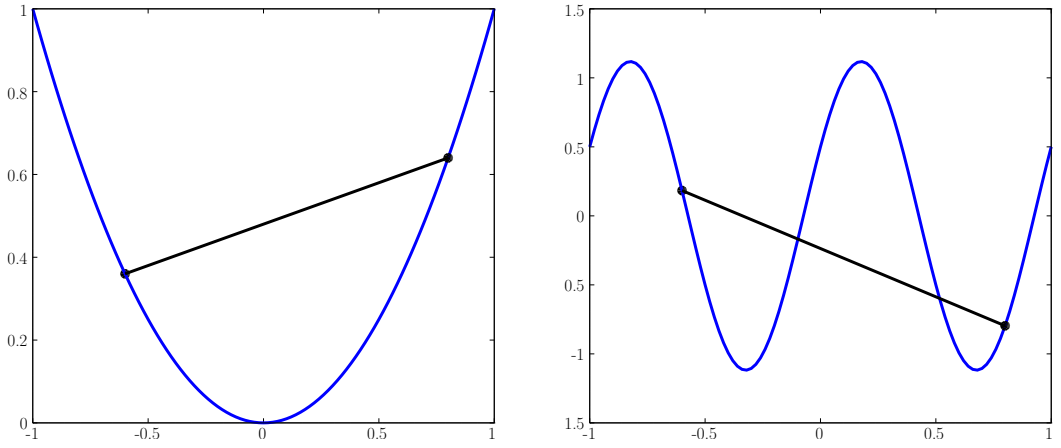
Figure 3.1: In a convex set *(a)*, the line of sight between any two points of the set lies completely within the set. This is not the case for a non-convex set *(b)*.

3.1.2 Convex Function

A function $f : C \to \mathbb{R}$ that maps the non-empty convex set $C \subseteq \mathbb{R}^m$ to $\mathbb{R}$, is convex if and only if

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}), \quad \alpha \in (0,1), \quad \forall\, \boldsymbol{x}, \boldsymbol{y} \in C \tag{3.2}$$

holds (cf. Boyd and Vandenberghe, 2004, p. 67). Geometrically, this is equivalent to the statement that the line segment between any two points $\boldsymbol{x}$ and $\boldsymbol{y}$ in C is never below the graph of the function. For the univariate case, this is visualized in Fig. 3.2. It should be noted that—due to the above definition—a linear function (e.g., a line or a plane) is a special case of a convex function, too.



(a) Convex function (b) Non-convex function

Figure 3.2: A convex function *(a)* has only one minimum. This need not be the case for a non-convex function *(b)*.

An essential property for optimization is that every convex function has exactly one minimum value. However, it is possible that there are many adjacent points with minimal value of the objective function. Therefore, in convex optimization, techniques to find a local minimum are sufficient as any local minimum found is the global minimum of the function. Furthermore, it should be noted that every quadratic form (cf. Sect. 2.1.1) with a symmetric and positive (semi-)definite matrix is a convex function (cf. Boyd and Vandenberghe, 2004, p. 71).

3.2 Minimization with Inequality Constraints

In the following, the focus is on inequality constrained estimation problems of the form

Minimization with inequality constraints	
objective function:	$\Phi(\boldsymbol{x}) \ldots \text{Min}$
constraints:	$\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$ $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m$.

(3.3)

$\Phi(\boldsymbol{x})$ is the objective function, $\boldsymbol{g}(\boldsymbol{x})$ are possibly non-linear inequality constraints and $\boldsymbol{h}(\boldsymbol{x})$ are linear or non-linear equality constraints. The vector $\boldsymbol{x}$ contains the optimization variables. Without

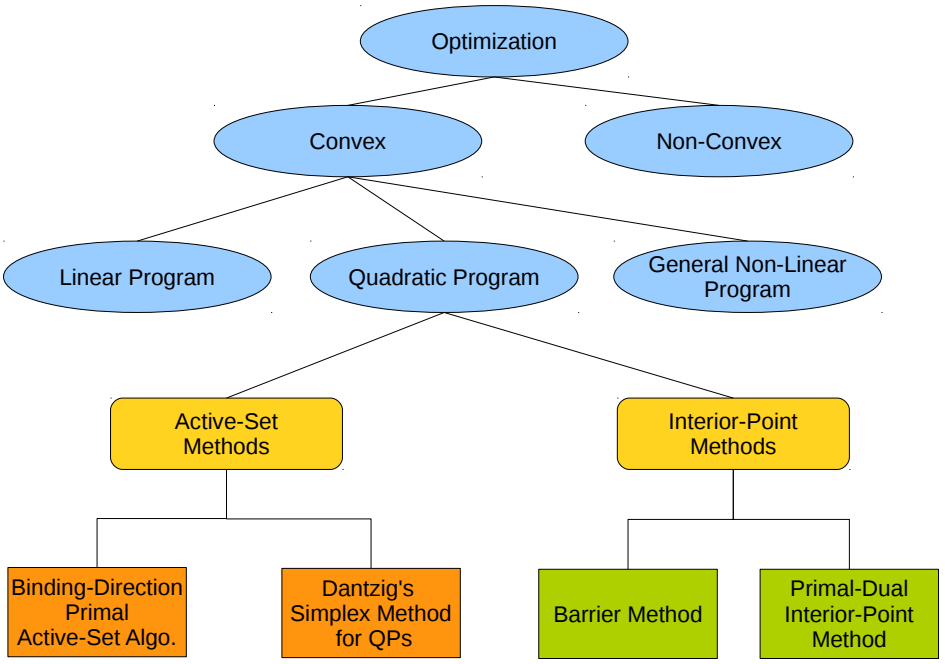


Figure 3.3: Taxonomy of optimization problems. Problem categories are indicated by blue ellipses. The two main classes of solvers are shown as yellow boxes. Some existing algorithms are depicted as orange (active-set methods) and green (interior-point methods) rectangles, respectively.

loss of generality only less than or equal to constraints will be treated within this thesis. This is sufficient as any greater than or equal to constraint can be easily transformed into a less than or equal to constraint.

Although the program (3.3) contains both types of constraints, problems of this type will be referred to as "inequality constrained problems" and not as "inequality and equality constrained problems". This abbreviation of notation is legitimated by the fact that if both types of constraints appear, the inequalities are the much more challenging ones. Furthermore, it is easy to incorporate equality constraints in almost any algorithm for inequality constrained estimation.

Maybe the biggest difference between unconstrained (or equality constrained) optimization and inequality constrained optimization is that for the latter, it is not known beforehand which constraints will influence the result. Equality constraints in general influence the result. However, this is not the case for inequalities. Due to this fact, there exist only iterative algorithms to solve inequality constrained problems. In general, such a problem is much harder to solve than an equality or unconstrained one.

To position the current work within the context of optimization, Fig. 3.3 shows a taxonomy of inequality constrained optimization problems.

The blue ellipses represent different problem categories. The general *optimization* problem (3.3) can be subdivided into *convex* and *non-convex* ones. As stated above, in the following only convex optimization problems are treated. Three sub-cases of convex optimization problems: *linear programs*, *quadratic programs* and *general non-linear programs* are depicted in order of ascending complexity. We will focus on Quadratic Programs (QP, cf. Sect. 3.3) only, which include Inequality Constrained Least-Squares (ICLS) adjustment as a prominent example.

There are two main classes of solvers for QPs: exact *active-set methods* and numerical *interior-point methods*, which are depicted as yellow boxes in Fig. 3.3. Within this thesis, the focus lies on active-set methods as those allow a warm-start (i.e., providing an initial solution), which will be beneficial in combination with the Monte Carlo approach used in Chap. 4.

Some existing algorithms are shown as orange (active-set methods) and green (interior-point methods) rectangles, respectively. A *binding-direction primal active-set method* is described in Sect. 3.5.1. *Dantzig's Simplex Method for Quadratic Programs* is described in Dantzig (1998, p. 490–497), and a *barrier method* as well as a *primal-dual interior-point method* can be found for example in Boyd and Vandenberghe (2004, p. 568–571 and p. 609–613, respectively).

The process of solving an inequality constrained optimization problem can be subdivided into two phases. While *phase 1* is concerned with feasibility, i.e., finding a point that fulfills all constraints, *phase 2* deals with optimality, i.e., determining the point that minimizes the objective function (cf. Wong, 2011, p. 1). Within this thesis, we will focus on the second phase. However, some approaches to obtain a feasible solution are discussed in Sect. 3.6.

3.3 Quadratic Program (QP)

An optimization problem

Quadratic Program (QP)	
objective function:	$\Phi_{\mathrm{QP}}(\boldsymbol{x}) = \gamma_1 \boldsymbol{x}^T \boldsymbol{C} \boldsymbol{x} + \gamma_2 \boldsymbol{c}^T \boldsymbol{x} \ \ldots \mathrm{Min}$
constraints:	$\boldsymbol{B}^T \boldsymbol{x} \leq \boldsymbol{b}$ $\overline{\boldsymbol{B}}^T \boldsymbol{x} = \bar{\boldsymbol{b}}$
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m$

(3.4)

is called a *Quadratic Program* (QP) in standard form (cf. Fletcher, 1987, p. 229). It consists of a quadratic and convex objective function Φ_{QP}, p linear inequality constraints and $\bar{p}$ linear equality constraints. γ_1 and γ_2 are scalar quantities, allowing to weight the quadratic and the linear term in the objective function. If not mentioned otherwise we will use $\gamma_1 = 0.5$ und $\gamma_2 = 1$ throughout this thesis. This is for reasons of convenience only. $\boldsymbol{C}$ is a symmetric and positive (semi-)definite $m \times m$ matrix (cf. Sect. 3.1.2) and $\boldsymbol{c}$ is a $m \times 1$ vector. In Chap. 5 we will see that the solution is unique if $\boldsymbol{C}$ is positive definite. In the positive semi-definite case, there will be a manifold of solutions (if it is not resolved through the constraints).

$\boldsymbol{B}$ and $\overline{\boldsymbol{B}}$ are the $m \times p$ and $m \times \bar{p}$ matrices of inequality and equality constraints, respectively. The $p \times 1$ vector $\boldsymbol{b}$ and the $\bar{p} \times 1$ vector $\bar{\boldsymbol{b}}$ are the corresponding right-hand sides. As in the last section, quantities corresponding to equalities are marked with a bar. As only linear constraints are allowed, the feasible set—i.e., the region, in which all constraints are satisfied—is always a convex set (cf. Sect. 3.1.1). Therefore, we are again minimizing a convex function over a convex set. As a result, there is only one minimum. If not stated otherwise, the optimization variable $\boldsymbol{x}$ is allowed to obtain positive or negative values (which is a difference to the standard form of a linear program).

It is often beneficial to transform an optimization problem into such a standard form, as there exists a large variety of algorithms to solve problems in that particular form (such as those mentioned in Sect. 3.2).

3.3.1 Karush-Kuhn-Tucker (KKT) Optimality Conditions

In this section, necessary and sufficient optimality conditions for a QP are derived. They are often referred to as *Karush-Kuhn-Tucker (KKT) conditions*. As in Sect. 2.2.1.2, the necessary conditions are obtained by minimizing the Lagrangian

$$L(\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}) = \Phi_{\mathrm{QP}}(\boldsymbol{x}) + \boldsymbol{k}^T(\boldsymbol{B}^T\boldsymbol{x} - \boldsymbol{b}) - \bar{\boldsymbol{k}}^T(\bar{\boldsymbol{B}}^T\boldsymbol{x} - \bar{\boldsymbol{b}}) \tag{3.5a}$$

$$= \gamma_1 \boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} + \gamma_2 \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{k}^T(\boldsymbol{B}^T\boldsymbol{x} - \boldsymbol{b}) - \bar{\boldsymbol{k}}^T(\bar{\boldsymbol{B}}^T\boldsymbol{x} - \bar{\boldsymbol{b}}). \tag{3.5b}$$

The derivatives with respect to the optimization variable $\boldsymbol{x}$ and with respect to both Lagrange multiplier vectors $\boldsymbol{k}$ and $\bar{\boldsymbol{k}}$ yield the gradients

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}) = 2\gamma_1 \boldsymbol{C}\boldsymbol{x} + \gamma_2 \boldsymbol{c} + \boldsymbol{B}\boldsymbol{k} - \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} \tag{3.6a}$$

$$\nabla_{\boldsymbol{k}} L(\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}) = \boldsymbol{B}^T\boldsymbol{x} - \boldsymbol{b} \tag{3.6b}$$

$$\nabla_{\bar{\boldsymbol{k}}} L(\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}) = -(\bar{\boldsymbol{B}}^T\boldsymbol{x} - \bar{\boldsymbol{b}}). \tag{3.6c}$$

Setting these gradients (less than or) equal to zero yields the first three KKT conditions

$$2\gamma_1 \boldsymbol{C}\boldsymbol{x} + \gamma_2 \boldsymbol{c} + \boldsymbol{B}\boldsymbol{k} - \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} \overset{!}{=} \boldsymbol{0} \tag{3.7a}$$

$$\boldsymbol{B}^T\boldsymbol{x} \overset{!}{\leq} \boldsymbol{b} \tag{3.7b}$$

$$\bar{\boldsymbol{B}}^T\boldsymbol{x} \overset{!}{=} \bar{\boldsymbol{b}}. \tag{3.7c}$$

The first equation states the relationship between the parameters $\boldsymbol{x}$ and the Lagrange multipliers. The second and the third one assure that all constraints are fulfilled. However, special care has to be taken of the inequality constraints. As they are "permeable" in one direction an additional sign constraint for the Lagrange multipliers has to be introduced

$$k_i \overset{!}{\geq} 0, \quad \forall i = 1, 2, \ldots, p \tag{3.7d}$$

(cf. Boyd and Vandenberghe, 2004, p. 244). Furthermore, the condition of complementary slackness (cf. Boyd and Vandenberghe, 2004, p. 243), states that all products

$$k_i \ (\boldsymbol{B}(:,i)^T\boldsymbol{x} - b_i) \overset{!}{=} 0, \quad \forall i = 1, 2, \ldots, p \tag{3.7e}$$

of the Lagrange multiplier and the corresponding inequality constraint have to be equal to zero. We will pick up this idea in Sect. 3.5.1 again, as it allows to separate constraints that influence the result from those that do not. Together, the five equations of (3.7) form the KKT conditions and are the necessary optimality conditions of a QP.

If there exists a solution of the convex QP (3.4)—i.e., the constraints are not contradictory, meaning that the feasible region is not empty—the necessary conditions become also sufficient. This is known as *Slater's condition* or *Slater's constraint qualification* (cf. Boyd and Vandenberghe, 2004, p. 226–227). Consequently, any point $\boldsymbol{x}$ that satisfies the KKT conditions (3.7) is a solution of the QP (3.4) and is called *KKT point*.

3.3.2 Inequality Constrained Least-Squares Adjustment as Quadratic Program

The majority of problems treated within this thesis are *Inequality Constrained (weighted) Least-Squares* (ICLS) problems in the GMM:

Inequality constrained least-squares (ICLS) adjustment in the GMM		
objective function:	$\Phi(\boldsymbol{x}) = \boldsymbol{v}(\boldsymbol{x})^T \boldsymbol{\Sigma}^{-1} \boldsymbol{v}(\boldsymbol{x}) \ldots \text{Min}$	
constraints:	$\boldsymbol{B}^T \boldsymbol{x} \leq \boldsymbol{b}$ $\bar{\boldsymbol{B}}^T \boldsymbol{x} = \bar{\boldsymbol{b}}$	(3.8)
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m$.	

As stated above this problem is easier to solve if it can be expressed as a QP in standard form. It can easily be seen that the constraints as well as the optimization variable already are conform with the notation of the standard form (3.4). In this section, it is shown that the objective function can easily be transformed into standard form, too. Using the well known transformations from (2.30), the objective function reads

$$\Phi(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{N} \boldsymbol{x} - 2\boldsymbol{n}^T \boldsymbol{x} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell}. \tag{3.9}$$

Neglecting the last part (which is constant and thus irrelevant for minimization) and using the substitutions

$$\boldsymbol{C} := 2\boldsymbol{N}, \tag{3.10a}$$
$$\boldsymbol{c} := -2\boldsymbol{n}, \tag{3.10b}$$

the objective function (3.9) reads

$$\Phi(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{x}^T \boldsymbol{C} \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x}. \tag{3.11}$$

Thus, the ICLS problem can be expressed as a QP in standard form with $\gamma_1 = \frac{1}{2}$ and $\gamma_2 = 1$, and a great variety of QP algorithms (e.g., those mentioned in Sect. 3.5) can be used to solve it.

Note that the substitutions of $\boldsymbol{C}$ and $\boldsymbol{c}$ were performed this way, to obtain exactly those numbers for the two scaling factors γ_1 and γ_2. This is done for reasons of convenience only, as these values are used very often in the optimization literature. However, the choice is completely arbitrary.

It should be mentioned that there are sometimes deviating opinions on what is called a "linear problem" in the optimization and the adjustment community. In the adjustment community, a problem with *linear observation equations*

$$\boldsymbol{\ell} + \boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}$$

is called a linear problem (cf. Sect. 2.2.1). However, minimizing the sum of squared residuals, leads to the **quadratic objective function** (3.9) and thus to a quadratic program.

In the optimization community, a linear problem often refers to a *Linear Program* (LP), i.e., to the problem of minimizing a **linear objective function** subject to some linear constraints. Within this thesis, the term "linear problem" will be used as done in the adjustment community. Thus, it refers to a problem with linear observation equations.

3.4 Duality

The basic idea of the *duality principle* is that for every constrained optimization problem (called the *primal problem*) there exists an equivalent representation, called the *(Lagrange) dual problem*. Both representations are linked via the Lagrangian (3.5a). The objective function of the dual problem is defined as the minimum of the Lagrangian with respect to $\boldsymbol{x}$

$$\Psi(\boldsymbol{k},\bar{\boldsymbol{k}}) = \min\left\{L(\boldsymbol{x},\boldsymbol{k},\bar{\boldsymbol{k}}) : \boldsymbol{x} \in \mathbb{R}^m\right\} \tag{3.12}$$

and is called the *Lagrange dual function* or simply *dual function* (cf. Boyd and Vandenberghe, 2004, p. 216). While the optimization variables of the primal problem are the parameters $\boldsymbol{x}$, the Lagrange multipliers $\boldsymbol{k}$ and $\bar{\boldsymbol{k}}$ are the optimization variables of the dual problem—the so called *dual variables*. If Slater's condition holds (which is always the case for a QP, cf. Sect. 3.3.1), it can be shown that by solving the dual problem

Lagrange dual problem	
objective function:	$\Psi(\boldsymbol{k},\bar{\boldsymbol{k}}) \ldots \mathrm{Max}$
constraints:	$\boldsymbol{k} \geq \boldsymbol{0}$
optim. variable:	$\boldsymbol{k} \in \mathbb{R}^p,\ \bar{\boldsymbol{k}} \in \mathbb{R}^{\bar{p}}$

(3.13)

optimal values for the primal variables can be obtained (cf. Boyd and Vandenberghe, 2004, p. 248), too. Sometimes this is more convenient than solving the primal problem. It should be emphasized that the only constraints in the dual formulation are that the Lagrange multipliers linked with inequality constraints have to be nonnegative. The Lagrange multipliers of the equality constraints are not sign-constrained.

Furthermore, it can be beneficial to explicitly compute the values of the Lagrange multipliers as they are "a quantitative measure of how active a constraint is" (cf. Boyd and Vandenberghe, 2004, p. 252). Along this line of thought, Lehmann and Neitzel (2013) proposed an hypothesis test for the compatibility of constraints using the Lagrange multipliers.

In the following section, the duality principle is used to derive the well-known formulas of an adjustment with condition equations in an alternative way.

3.4.1 Duality in Least-Squares Adjustment

For the primal problem of the adjustment with condition equations (cf. Sect. 2.2.2)

Primal problem (adjustment with condition equations)	
objective function:	$\Phi(\boldsymbol{v}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} \ldots \mathrm{Min}$
constraints:	$\bar{\boldsymbol{B}}^T\boldsymbol{v} + \bar{\boldsymbol{w}} = \boldsymbol{0}$
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n,$

(3.14)

the Lagrange function reads

$$L(\boldsymbol{v},\bar{\boldsymbol{k}}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} - 2\bar{\boldsymbol{k}}^T(\overline{\boldsymbol{B}}^T\boldsymbol{v} + \overline{\boldsymbol{w}}), \tag{3.15}$$

yielding the dual function

$$\Psi(\bar{\boldsymbol{k}}) = \min\left\{L(\boldsymbol{v},\bar{\boldsymbol{k}}) : \boldsymbol{v} \in \mathbb{R}^n\right\} \ldots \text{Max}. \tag{3.16}$$

In order to minimize (3.15) with respect to $\boldsymbol{v}$ its derivative is set equal to zero

$$\nabla_{\boldsymbol{v}} L(\boldsymbol{v},\bar{\boldsymbol{k}}) = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{v} - 2\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} \stackrel{!}{=} \boldsymbol{0}, \tag{3.17a}$$

yielding

$$\boldsymbol{v} = \boldsymbol{\Sigma}\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}}. \tag{3.17b}$$

Inserting (3.17b) in (3.15) yields

$$\Psi(\bar{\boldsymbol{k}}) = \bar{\boldsymbol{k}}^T\overline{\boldsymbol{B}}^T\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} - 2\bar{\boldsymbol{k}}^T(\overline{\boldsymbol{B}}^T\boldsymbol{\Sigma}\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} + \overline{\boldsymbol{w}}) \tag{3.18a}$$

$$= -\bar{\boldsymbol{k}}^T\overline{\boldsymbol{B}}^T\boldsymbol{\Sigma}\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} - 2\bar{\boldsymbol{k}}^T\overline{\boldsymbol{w}}. \tag{3.18b}$$

This results in the dual problem

DUAL PROBLEM (ADJUSTMENT WITH CONDITION EQUATIONS)

objective function:	$\Psi(\bar{\boldsymbol{k}}) = -\bar{\boldsymbol{k}}^T\overline{\boldsymbol{B}}^T\boldsymbol{\Sigma}\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} - 2\bar{\boldsymbol{k}}^T\overline{\boldsymbol{w}} \ldots \text{Max}$
constraints:	–
optim. variable:	$\bar{\boldsymbol{k}} \in \mathbb{R}^{\bar{p}}$.

(3.19)

As there are no inequality constraints, no sign restrictions concerning the Lagrange multipliers $\bar{\boldsymbol{k}}$ are required. Therefore, using the duality principle, the constrained estimation problem (3.14) was transformed into an unconstrained one.

To maximize $\Psi(\bar{\boldsymbol{k}})$, the derivative with respect to $\bar{\boldsymbol{k}}$ is computed and set equal to zero

$$\nabla_{\bar{\boldsymbol{k}}}\Psi(\bar{\boldsymbol{k}}) = -2\overline{\boldsymbol{B}}^T\boldsymbol{\Sigma}\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} - 2\overline{\boldsymbol{w}} \stackrel{!}{=} \boldsymbol{0}, \tag{3.20a}$$

yielding

$$\widetilde{\bar{\boldsymbol{k}}} = -(\overline{\boldsymbol{B}}^T\boldsymbol{\Sigma}\overline{\boldsymbol{B}})^{-1}\overline{\boldsymbol{w}}. \tag{3.20b}$$

Inserting (3.20b) in (3.17b) yields the desired primal variables—the residuals $\boldsymbol{v}$. Equations (3.20b) and (3.17b) are exactly those used in the classic geodetic approach of an adjustment with condition equations (cf. Sect. 2.2.2). Thus, the adjustment with condition equations can be seen as an example for solving the primal problem via its dual formulation and therefore for the impact of the duality principle in geodesy.

However, the duality principle offers many more concepts that may be beneficial for the process of solving an optimization problem. One of them—the so called duality gap—is introduced in the next section.

3.4.2 The Duality Gap

The difference between the primal and dual objective function

$$d(\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}) = \Phi(\boldsymbol{x}) - \Psi(\boldsymbol{k}, \bar{\boldsymbol{k}}) \tag{3.21}$$

is called *duality gap* and can be evaluated for any triplet $\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}$. If Slater's condition holds (which is always the case for a QP, cf. Sect. 3.3.1), the duality gap vanishes in the point of the optimal solution (cf. Fig. 3.4). This is equivalent to the statement that in the KKT point (cf. Sect. 3.3.1),

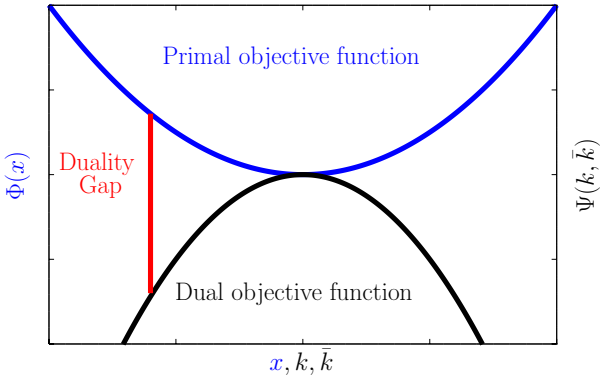


Figure 3.4: Schematic of the duality gap (*red line segment*) of a univariate problem with equality and inequality constraints. Strong duality holds. Thus, the values of the primal (*blue graph*) and the dual objective function (*black graph*) are identical in the point of the optimal solution.

the values of the primal (*blue graph*) and dual objective function (*black graph*) are identical

$$\Phi(\widetilde{\boldsymbol{x}}) = \Psi(\widetilde{\boldsymbol{k}}, \widetilde{\bar{\boldsymbol{k}}}). \tag{3.22}$$

This is called *strong duality* (cf. Boyd and Vandenberghe, 2004, p. 226). Summarized, strong duality enables us to test whether the optimization algorithm has already converged (i.e., the duality gap is zero) or how close we are to the optimal solution (the smaller the duality gap, the closer is the KKT point).
It should be mentioned that there exist non-convex optimization problems with a duality gap of zero (cf. Geiger and Kanzow, 2002, p. 326). Thus, it is not possible to use strong duality to derive a statement about the convexity of a problem.

In the following section, the dual problem of an ICLS adjustment in the GMM is derived exemplarily.

3.4.2.1 Dual Formulation of an Inequality Constrained Least-Squares Problem

As stated in Sect. 3.3.2, the primal objective function of the ICLS problem (3.8) reads

$$\begin{aligned}\Phi(\boldsymbol{x}) &= \boldsymbol{v}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{v} \\ &= \boldsymbol{x}^T \boldsymbol{N} \boldsymbol{x} - 2\boldsymbol{n}^T \boldsymbol{x} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} \ldots \text{Min}\end{aligned}$$

and the corresponding Lagrangian is given by

$$L(\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}) = \boldsymbol{x}^T \boldsymbol{N} \boldsymbol{x} - 2\boldsymbol{n}^T \boldsymbol{x} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} + 2\boldsymbol{k}^T (\boldsymbol{B}^T \boldsymbol{x} - \boldsymbol{b}) - 2\bar{\boldsymbol{k}}^T (\bar{\boldsymbol{B}}^T \boldsymbol{x} - \bar{\boldsymbol{b}}). \tag{3.23}$$

Note that again multiples of the Lagrange multipliers are used to obtain simpler derivatives. To obtain Lagrange's dual function, the gradient with respect to $\boldsymbol{x}$ is computed

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}) = 2\boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n} + 2\boldsymbol{B}\boldsymbol{k} - 2\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}}. \tag{3.24}$$

Setting it equal to zero

$$2\boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n} + 2\boldsymbol{B}\boldsymbol{k} - 2\overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} \stackrel{!}{=} \boldsymbol{0} \tag{3.25a}$$

yields the desired minimal value of the primal variable

$$\widetilde{\boldsymbol{x}} = \boldsymbol{N}^{-1}(\boldsymbol{n} - \boldsymbol{B}\boldsymbol{k} + \overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}}). \tag{3.25b}$$

Inserting (3.25b) in (3.23) yields the objective function of the dual problem

$$\begin{aligned}
\Psi(\boldsymbol{k}, \bar{\boldsymbol{k}}) =& \min\left\{ L(\boldsymbol{x}, \boldsymbol{k}, \bar{\boldsymbol{k}}) : \boldsymbol{x} \in \mathbb{R}^m \right\} \\
=& \widetilde{\boldsymbol{x}}^T \boldsymbol{N} \widetilde{\boldsymbol{x}} - 2\boldsymbol{n}^T \widetilde{\boldsymbol{x}} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} + 2\boldsymbol{k}^T (\boldsymbol{B}^T \widetilde{\boldsymbol{x}} - \boldsymbol{b}) - 2\bar{\boldsymbol{k}}^T (\overline{\boldsymbol{B}}^T \widetilde{\boldsymbol{x}} - \bar{\boldsymbol{b}}) && (3.26a) \\
=& -\boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} - \bar{\boldsymbol{k}}^T \overline{\boldsymbol{B}}^T \boldsymbol{N}^{-1} \overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} + 2\boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} \\
& + 2\left(\boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{B} - \boldsymbol{b}^T\right) \boldsymbol{k} - 2\left(\boldsymbol{n}^T \boldsymbol{N}^{-1} \overline{\boldsymbol{B}} - \bar{\boldsymbol{b}}^T\right) \bar{\boldsymbol{k}} && (3.26b) \\
& - \boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{n} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell}.
\end{aligned}$$

The complete transformation is stated in Appendix A.1. Applying (3.26), the ICLS dual problem reads

ICLS Dual Problem

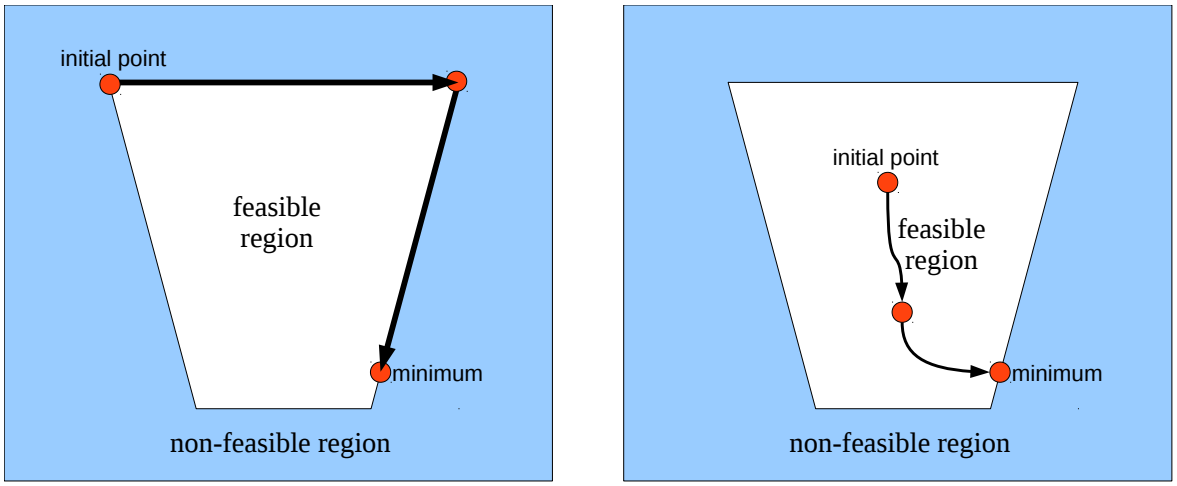
objective function:	$-\boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} - \bar{\boldsymbol{k}}^T \overline{\boldsymbol{B}}^T \boldsymbol{N}^{-1} \overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}} + 2\boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \overline{\boldsymbol{B}}\,\bar{\boldsymbol{k}}$ $+ 2\left(\boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{B} - \boldsymbol{b}^T\right) \boldsymbol{k} - 2\left(\boldsymbol{n}^T \boldsymbol{N}^{-1} \overline{\boldsymbol{B}} - \bar{\boldsymbol{b}}^T\right) \bar{\boldsymbol{k}}$ $- \boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{n} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} \ldots \mathrm{Max}$
constraints:	$\boldsymbol{k} \geq \boldsymbol{0}$
optim. variable:	$\boldsymbol{k} \in \mathbb{R}^p,\ \bar{\boldsymbol{k}} \in \mathbb{R}^{\bar{p}}$.

(3.27)

As problem (3.27) can be easily transformed into a QP in standard form (cf. Appendix A.1), any QP algorithm can be used to obtain a solution. In the next section, some algorithms to solve either the primal or the dual problem are explained.

3.5 Active-Set Algorithms

There exists a big variety of algorithms for solving inequality constrained convex optimization problems of type (3.3). Most of them can be subdivided into two main classes: active-set and interior-point methods (cf. Fig. 3.3). While algorithms of the first type are usually tailor-made for quadratic and linear programs, the latter are applicable to a wide range of problem types. As within this thesis the focus is on QPs, both classes would be suited. However, many active-set methods allow a warm-start (i.e., allow to specify an initial solution). As this is beneficial for the Monte Carlo methods presented in Chap. 4, we will focus on active-set methods.



(a) Active-set approach

(b) Interior-point approach

Figure 3.5: Basic ideas of *(a)* active-set and *(b)* interior-point approaches. While algorithms of the first type follow the boundary of the feasible set, the latter follow a central path through the interior of the feasible region.

Furthermore, in active-set algorithms, the constraints enter in an exact way, while interior-point methods use relaxed constraints, which are tightened in each iteration. Compared with active-set approaches, interior-point methods usually need less, but more expensive, iterations to solve an optimization problem (cf. Gould, 2003).

It is also possible to transform a QP into a linear complementarity problem (LCP, cf. Koch, 2006, p. 24–25) and solve it e.g., via *Lemke's Algorithm* (cf. Fritsch, 1985) or into a *Least-Distance Program* (LDP) and solve it e.g., via the LDP algorithm described in Lawson and Hanson (1974, p. 165). More recent approaches include the aggregation of all inequality constraints into one equality constraint with a high complexity (Peng et al., 2006).

As mentioned above, within this thesis, only active-set methods are used. For a comparison of different active-set, LCP and LDP algorithms see e.g., Roese-Koerner (2009).

The main idea of active-set methods is to follow the boundary of the *feasible region* (i.e., the region in the parameter space, where all constraints are satisfied; *white region* in Fig. 3.5(a)) in an iterative approach until the optimal solution (*red circle* in the lower right) is reached. This is done by extracting the constraints that hold as equality constraints (called *active set*) in the point of the current solution and therefore solve a sequence of equality constrained subproblems (cf. Wong, 2011, p. 2). If at least one constraint is active, the point of the optimal solution will always be at the boundary of the feasible region. For a brief overview of different types of active-set methods see e.g., Nocedal and Wright (1999, Chap. 16) or Wong (2011).

In the following section, a basic version of the algorithm is stated, which allows to solve strictly convex problems. An extension to the case with a singular matrix $\boldsymbol{C}$ in the objective function is discussed in Sect. 5.2.

3.5.1 Binding-Direction Primal Active-Set Algorithm

The method described in the following is a *binding-direction primal active-set method*, which is a combined version of the algorithms described in Gill et al. (1981), Best (1984) and Wölle (1988). For reasons of readability we will refer to it only as "the" active-set method. The method relies on two concepts for different types of constraints and directions, which will be explained first.

3.5.1.1 Active, Inactive and Violated Constraints

An inequality constraint

$$b_{j1}x_1 + b_{j2}x_2 + \ldots + b_{jm}x_m \leq b_j \tag{3.28}$$

is called *active* if it holds as the equality constraint

$$b_{j1}x_1 + b_{j2}x_2 + \ldots + b_{jm}x_m = b_j. \tag{3.29}$$

This is equivalent to the statement that $\boldsymbol{x}$ is on the boundary of the feasible set. A constraint is *inactive* if the strict inequality

$$b_{j1}x_1 + b_{j2}x_2 + \ldots + b_{jm}x_m < b_j \tag{3.30}$$

holds. In this case there is a "buffer" between $\boldsymbol{x}$ and the constraint. If

$$b_{j1}x_1 + b_{j2}x_2 + \ldots + b_{jm}x_m > b_j, \tag{3.31}$$

the constraint is *violated*, which must not happen throughout the iterations of the algorithm described in the next sections. For each feasible point $\boldsymbol{x}$ a *working set*

$$\boldsymbol{W}^T\boldsymbol{x} = \boldsymbol{w}, \tag{3.32}$$

can be assembled, in which all active inequality constraints (3.29) and all equality constraints are comprised in the $m \times p_w$ matrix $\boldsymbol{W}$ and the $p_w \times 1$ vector $\boldsymbol{w}$. All p_v inactive inequality constraints are combined to

$$\boldsymbol{V}^T\boldsymbol{x} < \boldsymbol{v}, \tag{3.33}$$

with the $m \times p_v$ matrix $\boldsymbol{V}$ and the $p_v \times 1$ vector $\boldsymbol{v}$. Taken together, both sets yield the (possibly permuted) original set of constraints

$$\underbrace{\begin{bmatrix}\boldsymbol{W}^T \\ \boldsymbol{V}^T\end{bmatrix}}_{\boldsymbol{B}^T}\boldsymbol{x} \leq \underbrace{\begin{bmatrix}\boldsymbol{w} \\ \boldsymbol{v}\end{bmatrix}}_{\boldsymbol{b}}. \tag{3.34}$$

Each point $\boldsymbol{x}^{(i)}$ has its specific set of active and inactive constraints. Thus, it would be consequent to write

$$\boldsymbol{W}^{(i)^T}\boldsymbol{x}^{(i)} = \boldsymbol{w}^{(i)} \tag{3.35}$$

$$\boldsymbol{V}^{(i)^T}\boldsymbol{x}^{(i)} < \boldsymbol{v}^{(i)}. \tag{3.36}$$

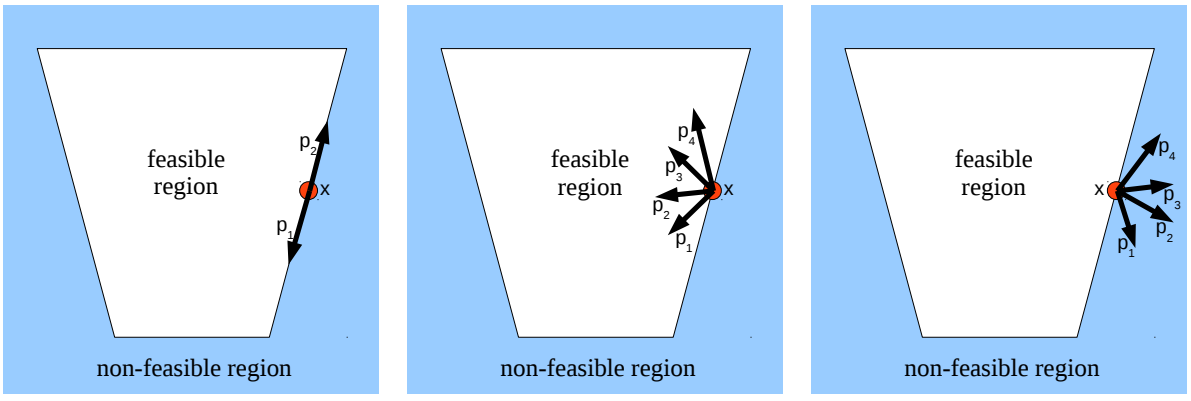
However, we decided to drop the iteration index $^{(i)}$ of the sets in the following whenever it seems appropriate to keep the formulas tidy. Thus, one should keep in mind that these sets change at each iteration.

3.5.1.2 Binding, Non-binding and Infeasible Directions

Within the algorithms, the parameter vector will be updated repeatedly using a *search direction* $\boldsymbol{p}^{(i)}$ and a *step length* $q^{(i)}$

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + q^{(i)}\boldsymbol{p}^{(i)}, \qquad q^{(i)} > 0. \tag{3.37}$$

Therefore, it is crucial to be able to determine if a step in a potential search direction $\boldsymbol{p}^{(i)}$ can violate a constraint. For inactive constraints a step in any direction is possible, if $q^{(i)}$ is chosen small enough. Thus, it is sufficient to examine solely the active constraints as they can be violated through an update in the "wrong" direction (cf. Fig. 3.6(c)). A direction is called *binding* concerning



(a) Binding directions (b) Non-binding directions (c) Infeasible directions

Figure 3.6: Binding, non-binding and infeasible directions for an active constraint in a bivariate example. $\boldsymbol{p}_i$ are the vectors of directions starting from point $\boldsymbol{x}$. Figure modified from Roese-Koerner (2009, p. 22).

all active constraints, if

$$\boldsymbol{W}^T\boldsymbol{p}^{(i)} = \boldsymbol{0} \tag{3.38}$$

holds. This is equivalent to the statement that $\boldsymbol{p}^{(i)}$ is in the nullspace of $\boldsymbol{W}$ and all constraints active in $\boldsymbol{x}^{(i)}$ will stay active after the next iteration (cf. Fig. 3.6(a) and (3.29)), due to

$$\begin{aligned}
\boldsymbol{W}^T\boldsymbol{x}^{(i+1)} &= \boldsymbol{W}^T\left(\boldsymbol{x}^{(i)} + q^{(i)}\boldsymbol{p}^{(i)}\right) && \text{(3.39a)}\\
&= \boldsymbol{W}^T\boldsymbol{x}^{(i)} + q^{(i)}\underbrace{\boldsymbol{W}^T\boldsymbol{p}^{(i)}}_{=\boldsymbol{0}} && \text{(3.39b)}\\
&= \boldsymbol{W}^T\boldsymbol{x}^{(i)} && \text{(3.39c)}\\
&= \boldsymbol{w}.
\end{aligned}$$

This property will be essential for the active-set algorithm. A direction is called *non-binding* concerning the constraint j if

$$\boldsymbol{W}(:,j)^T\boldsymbol{p}^{(i)} < 0 \tag{3.40}$$

holds. As a consequence, the constraint j will become inactive in the point $\boldsymbol{x}^{(i+1)}$ (cf. Fig. 3.6(b) and (3.30))

$$\begin{aligned}
\boldsymbol{W}^T\boldsymbol{x}^{(i+1)} &= \boldsymbol{W}^T\left(\boldsymbol{x}^{(i)} + q^{(i)}\boldsymbol{p}^{(i)}\right) && \text{(3.41a)}\\
&= \underbrace{\boldsymbol{W}^T\boldsymbol{x}^{(i)}}_{=\boldsymbol{w}_j} + \underbrace{q^{(i)}}_{>0}\underbrace{\boldsymbol{W}^T\boldsymbol{p}^{(i)}}_{<0} && \text{(3.41b)}\\
&< \boldsymbol{w}_j.
\end{aligned}$$

Algorithm 2: Binding-Direction Primal Active-Set Algorithm

// Iterative algorithm to solve a quadratic program in standard form (3.4)

Data:

$\boldsymbol{C}_{[m\times m]}, \boldsymbol{c}_{[m\times 1]}$	...	Matrix and vector with the coefficients of the objective function
$\boldsymbol{B}_{[m\times p]}, \boldsymbol{b}_{[p\times 1]}$	...	Matrix and corresponding right-hand side of the inequality constraints
$\bar{\boldsymbol{B}}_{[m\times \bar{p}]}, \bar{\boldsymbol{b}}_{[\bar{p}\times 1]}$	...	Matrix and corresponding right-hand side of the equality constraints
$\boldsymbol{x}^{(0)}_{[m\times 1]}$	...	feasible initial solution vector

Result:

$\boldsymbol{x}_{[m\times 1]}$	...	Vector containing the solution of the QP
$\boldsymbol{k}_{w[p_w\times 1]}$	...	Vector containing the Lagrange multipliers of the active constraints

1 $[\boldsymbol{W}, \boldsymbol{w}, \boldsymbol{V}, \boldsymbol{v}] = \text{findActiveConstraints}(\boldsymbol{B}, \boldsymbol{b}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{b}}, \boldsymbol{x})$ *// cf. Sect. 3.5.1.1*
2 **for** $i = 1 : i_{\max}$ **do**
3 $\quad [\boldsymbol{p}, \boldsymbol{k}_w] = \text{computeSearchDirection}(\boldsymbol{C}, \boldsymbol{c}, \boldsymbol{W}, \boldsymbol{x})$ *// cf. Sect. 3.5.1.4*
4 $\quad q = \text{computeStepLength}(\boldsymbol{V}, \boldsymbol{v}, \boldsymbol{x}, \boldsymbol{p})$ *// cf. Sect. 3.5.1.5*
5 $\quad \boldsymbol{x} = \boldsymbol{x} + q\boldsymbol{p}$
6 $\quad [\boldsymbol{W}, \boldsymbol{w}, \boldsymbol{V}, \boldsymbol{v}] = \text{updateActiveSet}(\boldsymbol{W}, \boldsymbol{w}, \boldsymbol{V}, \boldsymbol{v}, \boldsymbol{x}, q)$ *// cf. Sect. 3.5.1.6*
7 $\quad$ **if** $\min(\boldsymbol{k}_w \geq 0)$ **then**
8 $\quad\quad$ break
9 $\quad$ **end**
10 **end**
11 **return** $\boldsymbol{x}, \boldsymbol{k}_w$

For an *infeasible* direction concerning the constraint j

$$\boldsymbol{W}(:, j)^T \boldsymbol{p}^{(i)} > 0 \tag{3.42}$$

holds. The feasible region would be left through a step of an arbitrary step size in this direction (cf. Fig. 3.6(c) and (3.31))

$$\begin{aligned} \boldsymbol{W}^T \boldsymbol{x}^{(i+1)} &= \boldsymbol{W}^T \left(\boldsymbol{x}^{(i)} + q^{(i)} \boldsymbol{p}^{(i)} \right) && \text{(3.43a)} \\ &= \underbrace{\boldsymbol{W}^T \boldsymbol{x}^{(i)}}_{=\boldsymbol{w}_j} + \underbrace{q^{(i)}}_{>0} \underbrace{\boldsymbol{W}^T \boldsymbol{p}^{(i)}}_{>0} && \text{(3.43b)} \\ &> \boldsymbol{w}_j. \end{aligned}$$

This must not happen throughout the algorithm.

3.5.1.3 Outline of the Binding-Direction Primal Active-Set Method

Having the concepts described in the last two paragraphs at hand an outline of the Binding-Direction Primal Active-Set algorithm can be stated, which is given in Alg. 2.

Starting at a feasible point $\boldsymbol{x}^{(0)}$ an initial active set can be obtained by evaluating the constraints at $\boldsymbol{x}^{(0)}$ (line 1 of Alg. 2). It has to be taken care of that the initial matrix of active constraints $\boldsymbol{W}$ is

of full column rank (cf. Wong, 2011, p. 10). Therefore, it might be necessary to eliminate dependent columns from $\boldsymbol{W}$ and the corresponding values in $\boldsymbol{w}$.

Afterwards a search direction $\boldsymbol{p}$, the Lagrange multipliers of the active constraints $\boldsymbol{k}_w$ (line 3) and a step length q (line 4) are computed, and an update of the solution is performed (line 5). Subsequently, a decision has to be made, which constraints stay in the active set and which are dropped (line 6). The last four steps are repeated iteratively until all Lagrange multipliers associated with inequalities are non-negative. Starting with the search direction all steps are explained in more detail in the following.

3.5.1.4 Search Direction $\boldsymbol{p}$

The search direction $\boldsymbol{p}^{(i)}$ shall be computed in a way that all active constraints

$$\boldsymbol{W}^{(i)^T}\boldsymbol{x}^{(i)} = \boldsymbol{w}^{(i)}$$

are kept active after the update

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \boldsymbol{p}^{(i)} \tag{3.44}$$

and such that the value of the objective function of problem (3.4) becomes minimal in the current subspace (i.e., the nullspace of the matrix of active constraints). This ensures that we move along the boundary of the feasible region towards the minimum. We have intentionally omitted the step length q here, as we want to derive a search direction which directly points to the minimum of the current subspace without the need for a "scaling". The step length will come into play again in the next paragraph, as it is essential for not violating constraints inactive in the point of the current solution.

Therefore, the following subproblem has to be solved. It should be noted that in this case $\boldsymbol{p}$ is the optimization variable and $\boldsymbol{x}$ is fixed (iteration indices were omitted)

Search direction Subproblem	
objective function:	$\Phi(\boldsymbol{p}) = \frac{1}{2}(\boldsymbol{x}+\boldsymbol{p})^T\boldsymbol{C}(\boldsymbol{x}+\boldsymbol{p}) + \boldsymbol{c}^T(\boldsymbol{x}+\boldsymbol{p}) \ldots \mathrm{Min}$
constraints:	$\boldsymbol{W}^T\boldsymbol{p} = \boldsymbol{0}$
optim. variable:	$\boldsymbol{p} \in \mathbb{R}^m$.

(3.45)

The search direction can either be obtained by computing a direction to the unconstrained minimum and using projectors to map it to the nullspace of the matrix of active constraints $\boldsymbol{W}$ (cf. Wölle, 1988, p. 39) or by an approach using the Lagrangian of the problem (cf. Best, 1984). We will use the latter as it is numerically more stable and can be extended to allow the treatment of a positive semi-definite matrix $\boldsymbol{C}$ (cf. Chap. 5). The Lagrangian of problem (3.45) reads

$$L(\boldsymbol{p}, \boldsymbol{k}_w) = \frac{1}{2}(\boldsymbol{x}+\boldsymbol{p})^T\boldsymbol{C}(\boldsymbol{x}+\boldsymbol{p}) + \boldsymbol{c}^T(\boldsymbol{x}+\boldsymbol{p}) + \boldsymbol{k}_w^T(\boldsymbol{W}^T\boldsymbol{p}) \tag{3.46a}$$

$$= \frac{1}{2}\boldsymbol{p}^T\boldsymbol{C}\boldsymbol{p} + \boldsymbol{x}^T\boldsymbol{C}\boldsymbol{p} + \frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{p} + \boldsymbol{k}_w^T\boldsymbol{W}^T\boldsymbol{p}. \tag{3.46b}$$

$\boldsymbol{k}_w$ are the Lagrange multipliers linked with the active constraints. The gradients of (3.46) with respect to $\boldsymbol{p}$ and $\boldsymbol{k}_w$ read

$$\nabla_{\boldsymbol{p}} L(\boldsymbol{p}, \boldsymbol{k}_w) = \boldsymbol{C}\boldsymbol{p} + \boldsymbol{C}\boldsymbol{x} + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{k}_w$$

$$= \boldsymbol{C}\boldsymbol{p} + \boldsymbol{g} + \boldsymbol{W}\boldsymbol{k}_w, \tag{3.47}$$

$$\nabla_{\boldsymbol{k}_w} L(\boldsymbol{p}, \boldsymbol{k}_w) = \boldsymbol{W}^T\boldsymbol{p}. \tag{3.48}$$

The gradient

$$\boldsymbol{g} = \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{c}, \tag{3.49}$$

was used as substitution. Setting the derivatives equal to zero results in

$$\begin{bmatrix} \boldsymbol{C} & \boldsymbol{W} \\ \boldsymbol{W}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{k}_w \end{bmatrix} = \begin{bmatrix} -\boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix}. \tag{3.50}$$

Therefore, by solving the equation system above, it is possible to compute a search direction that will not violate any active constraint. As a byproduct, the Lagrange multipliers of all active constraints are obtained, which will be used later on to determine which active constraints should be deactivated. However, no statement about inactive constraints has been made, yet. This will be done in the next section, when dealing with the step length q.

3.5.1.5 Step Length q

The search direction was computed in a way that no active constraint will be violated after the parameter update. A maximal feasible step length $q \in (0, 1]$ is determined within this paragraph, to ensure that also no inactive constraint will be violated. As a first step, those constraints shall be identified that could possibly be violated through a step in direction $\boldsymbol{p}$.

According to Gill et al. (1981, p. 169), those inactive constraints for which

$$\boldsymbol{V}(:, j)^T \boldsymbol{p}^{(i)} \leq 0 \tag{3.51}$$

holds, cannot be violated through a step in direction $\boldsymbol{p}$. This can be verified by examining the product

$$\begin{aligned} \boldsymbol{V}(:, j)^T \boldsymbol{x}^{(i+1)} &= \boldsymbol{V}(:, j)^T \left(\boldsymbol{x}^{(i)} + q^{(i)} \boldsymbol{p}^{(i)} \right) && \text{(3.52a)} \\ &= \underbrace{\boldsymbol{V}(:, j)^T \boldsymbol{x}^{(i)}}_{<v_j} + \underbrace{q^{(i)}}_{<0} \underbrace{\boldsymbol{V}(:, j)^T \boldsymbol{p}^{(i)}}_{\geq 0} && \text{(3.52b)} \\ &< v_j. \end{aligned}$$

This is equivalent to the statement that the constraint j will stay inactive in the next iteration step. Therefore, it could be concluded that only constraints for which

$$\boldsymbol{V}(:, j)^T \boldsymbol{p}^{(i)} > 0 \tag{3.53}$$

holds, can become active or violated in the next step. Those constraints shall be called *potentially violated constraints*. If

$$\boldsymbol{V}(:, j)^T \boldsymbol{x}^{(i+1)} = \boldsymbol{V}(:, j)^T \left(\boldsymbol{x}^{(i)} + q_j^{(i)} \boldsymbol{p}^{(i)} \right) = v_j \tag{3.54a}$$

holds for a certain q_j, the constraint j will become active in the next iteration step. Reformulating (3.54a) with respect to q_j can be used to determine the distance to the constraint j in the direction $\boldsymbol{p}$

$$q_j^{(i)} = \frac{v_j - \boldsymbol{V}(:, j)^T \boldsymbol{x}^{(i)}}{\boldsymbol{V}(:, j)^T \boldsymbol{p}^{(i)}}. \tag{3.54b}$$

Therefore, the maximal feasible step length q_{max} is restrained by the constraint j that would first be violated through a step in direction $\boldsymbol{p}$. As described above, the best achievable step length would be one, as in this case, the new point $\boldsymbol{x}^{(i+1)}$ will be the minimum of the current subspace. As a consequence, the optimal step length is defined as

$$q = \min \left(q_j^{(i)}, 1 \right), \qquad \forall \, \{ j \mid \boldsymbol{V}_j^T \boldsymbol{p}^{(i)} > 0 \}. \tag{3.55}$$

3.5.1.6 Update of the Active Set

So far it was described how an optimal solution in the current subspace can be obtained. Within this paragraph, a strategy to update the set of active constraints—i.e., drop constraints from the set or include new ones—is explained. As this determines the subspace and therefore the sequence of equality constrained subproblems to be solved, it can be seen as the "core" of the algorithm. Two cases have to be distinguished, depending on the chosen step length q.

If q is less than 1, the hitherto inactive constraint j prevents that a "full step" can be taken. In this case, the constraint j is removed from the inactive set $\boldsymbol{V}, \boldsymbol{v}$ and included in the active set $\boldsymbol{W}, \boldsymbol{w}$. Therefore, in the next iteration step a new equality constrained optimization problem has to be solved.

If $q = 1$ holds, due to its design, the point $\boldsymbol{x}^{(i+1)}$ is the minimum of the current subspace (cf. problem (3.45)). This means it is the point which has the minimal value of the objective function that keeps all active constraints of iteration step (i) also active after step $(i + 1)$. As a consequence, it is not possible to further minimize the objective function without removing a constraint from the active set. Therefore, it is mandatory to identify those active constraints that prevent a decrease of the objective function. It can be shown that all inequality constraints which are associated with a negative Lagrange multiplier are the ones to deactivate (a proof is provided in Appendix A.2). It should be noted that equality constraints will never be removed from the active set. If there are no negative Lagrange multipliers, it is not possible to further reduce the value of the objective function. Thus, $\boldsymbol{x}^{(i+1)}$ is the optimal solution and the algorithm terminates.

The combination of all steps described above yields the binding-direction primal active-set approach outlined in Sect. 3.5.1.3 and Alg. 2.

3.6 Feasibility and Phase-1 Methods

The topics *feasibility* and *phase-1 methods* are closely related. While the former is concerned with the question whether there is a feasible solution of an optimization problem, the latter deals with the determination of such a point that fulfills all constraints and can therefore be used as initial value for the second phase: the minimization of the objective function.

3.6.1 Feasibility

Whenever the task is to examine if the constraints are consistent, a *feasibility problem* (cf. Boyd and Vandenberghe, 2004, p. 129) has to be solved

Feasibility problem	
objective function:	–
constraints:	$\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$ $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m$.

(3.56)

As the aim is to determine if a feasible point exists or if the constraints are inconsistent, no objective function is stated. As the transformation of the feasibility problem (3.56) into a QP in standard form is not straightforward, some approaches are outlined in the next section.

3.6.2 Finding an Initial Point

As many phase-2 algorithms require a (primal) feasible initial point (e.g, the binding-direction active-set method described in Sect. 3.5.1), many different approaches to solve the feasibility problem (3.56) exist. One of these phase-1 methods is outlined in the next section and references to other approaches are provided afterwards. In the following we restrict ourselves to the case of linear constraints.

3.6.2.1 The Sum-of-Infeasibilities Ansatz

In the *sum-of-infeasibilities ansatz* (cf. Boyd and Vandenberghe, 2004, p. 580) the feasibility problem (3.56) is extended by one auxiliary variable s_i for each inequality constraint. Those can be combined to a $p \times 1$ vector $\boldsymbol{s}$, yielding

PHASE-1: SUM OF INFEASIBILITIES	
objective function:	$\mathbf{1}^T \boldsymbol{s} \ldots \text{Min}$
constraints:	$\boldsymbol{B}^T \boldsymbol{x} - \boldsymbol{b} \leq \boldsymbol{s}$ $\bar{\boldsymbol{B}}^T \boldsymbol{x} = \bar{\boldsymbol{b}}$
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{s} \in \mathbb{R}^p.$

(3.57)

$\mathbf{1}$ is a $p \times 1$ vector containing only Ones. Therefore, the objective function can be seen as a summation of the existing "infeasibilities" of the inequality constraints. If there exists a point $\boldsymbol{x}^{(0)}$ for which all s_i are non-positive, the problem is feasible and $\boldsymbol{x}^{(0)}$ can be used as initial value. If no such point exists, the constraints are inconsistent.

In contrast to problem (3.56), the problem (3.57) is a LP in standard form (cf. Sect. 3.3.2). Therefore it can be solved by any LP solver (e.g., by Dantzig's Simplex Method, Dantzig, 1998, p. 94–108) or by any QP solver which is able to deal with a positive semi-definite (or non-existing) matrix $\boldsymbol{C}$ in the objective function (cf. Chapter 5).

Furthermore, it is trivial to find a feasible starting point for problem (3.57) or to prove that there exists none: First, a solution of the usually underdetermined linear equation system of the equality constraints

$$\bar{\boldsymbol{B}}^T \boldsymbol{x} = \bar{\boldsymbol{b}}$$

has to be found. This can for example be done by using the Gauss-Jordan algorithm described in Sect. 2.1.3.1. If the equation system has no solution, the equality constraints are inconsistent and it is not possible to fulfill all constraints. Subsequently, initial values for the auxiliary variables are computed by evaluating

$$s_i^{(0)} = \max\Big(\boldsymbol{B}(:,i)^T \boldsymbol{x}^{(0)} - b_i, 0\Big). \tag{3.58}$$

If all $s_i^{(0)}$ are non-positive, $\boldsymbol{x}^{(0)}$ is a feasible initial solution for the original optimization problem and the second phase can be started. Otherwise, the initial values $\boldsymbol{s}^{(0)}$ and $\boldsymbol{x}^{(0)}$ are used to solve program (3.57) with a suitable solver. The computation can be aborted, as soon as all s_i take a non-positive value. If optimal values for $\boldsymbol{s}$ and $\boldsymbol{x}$ are found and there is still a positive s_i, the constraints are inconsistent and no feasible solution exists. In this case, the positive entries of $\boldsymbol{s}$ are connected to those constraints that could not be satisfied.

3.6.2.2 Alternatives

Instead of using the ansatz described above, it is also possible to use standard LP, QP or LCP solvers that need no initial solution (such as Lemke's Algorithm, cf. Fritsch, 1985) and abort the iterations as soon as a feasible point is encountered. The advantage is that the resulting initial point is likely to be closer to the optimal solution than in approaches that do not take the objective function into account. The drawback is that these algorithms usually converge slower to a feasible solution as they are not tailor-made for finding an initial point. As a consequence, using the sum-of-infeasibilities approach is usually more preferable than using another solver to determine a feasible solution.

When dealing with phase-1 problems, sometimes the *Big-M-method* is mentioned. At first glance, it looks very similar to the sum-of-infeasibilities approach mentioned above, as auxiliary variable are used in a very similar way. However, the original objective function is used, too, and the auxiliary variables are weighted in the objective function using a weight M with a big value. As a consequence it suffers from numerical difficulties (especially for larger problems) so that Gill et al. (1981, p. 199) conclude that it "is not reliable in practice".

Part II

Methodological Contributions

4. A Stochastic Framework for ICLS

This chapter is a modified and extended version of the articles *A stochastic framework for inequality constrained estimation* (Roese-Koerner et al., 2012) and *Describing the quality of inequality constrained estimates* (Roese-Koerner et al., 2015).

Quality description is one of the key features of geodetic inference. In Ordinary or Equality Constrained Least-Squares (OLS and ECLS, respectively) adjustment, a wide range of tools to quantify the accuracy of an estimate is available. As there is an analytical solution, it is possible to propagate the accuracy of the observations (given in the form of a covariance matrix) to the parameters. This is no longer possible in the inequality constrained case, as the parameters are not linked analytically with the observations. Therefore, iterative solvers are used (either active-set or interior-point methods, cf. Sect. 3.2), and the law of error propagation (cf. Koch, 1999, p. 100) can no longer be applied. The idea of a symmetric standard deviation, in the sense of an interval around estimated quantities, is no longer sufficient to describe the accuracy, dealing with a parameter space truncated by inequality constraints, which can destroy symmetry.

Some analytical methods have been developed for describing the quality of inequality constrained estimates. However, these methods either ignore the probability mass in the infeasible region (cf. Geweke, 1986, Zhu et al., 2005) or the influence of inactive constraints (cf. Liew, 1976) and therefore yield only approximate results.

Nevertheless, it is crucial to have a measure for the accuracy of an estimate. In contrast to standard deviations, confidence regions can be given in truncated parameter spaces, as they can be adapted to the constraints. In order to construct confidence regions, the Probability Density Function (PDF) of the estimates must be known. Due to the above mentioned difficulties in applying analytical techniques for quality description of ICLS estimates, the approach proposed in this chapter is a Monte Carlo technique. This has the advantage that no intricate analytical relationship has to be evaluated and that it is well suited for parallel computing.

We use a binding-direction primal active-set method (cf. Sect. 3.5.1) to solve M instances of the original problem with randomized observation vectors to compute an empirical PDF of the estimated parameters. Knowing this discrete approximation of the PDF, Highest Posterior Density (HPD) regions are computed and several ideas of what constitutes a *best* estimator in this case are discussed.

In addition, the influence of the constraints on the estimate is examined. Therefore, a sensitivity analysis is described, as a local measure for the influence of each constraint on each parameter. Furthermore, two global measures are provided: an hypothesis testing is outlined to determine if data and constraints are consistent, and it is shown how the ratio of probability mass inside the feasible region can be computed.

The purpose of this chapter is to describe a frequentist framework for the stochastic description of ICLS estimates (cf. problem (3.8)), which overcomes many of the drawbacks of the existing methods. We subdivide the task of defining a stochastic framework for ICLS estimates into the problem of quality description (Sect. 4.2) and the problem of measuring the influence of the constraints (Sect. 4.3). Afterwards, both parts are combined to the aforementioned stochastic framework and referred to as *Monte Carlo Quadratic Programming method* (MC-QP, Sect. 4.4). The case of an inequality constrained estimate of a line of best fit serves as proof of concept (Sect. 4.5).

4.1 State of the Art

All existing methods to find the solution of an inequality constrained estimation problem are based on iterative procedures (cf. Sect. 3.2). Therefore—and due to the presence of inequality constraints itself—it is difficult to describe the quality of the estimate. In the following sections two existing approaches to give a measure for the accuracy of those estimates are described: The active constraint approach, which uses frequentist statistics, and a Bayesian method. Both methods as well as their strengths and weaknesses are outlined in the next sections.

4.1.1 Active Constraint Approach

The idea of Liew (1976) is to reduce an inequality constraint problem to an equality constraint one. As active constraints hold as equality constraints, only those constraints active in the optimal solution are taken into account and are treated as equalities. This approach consists of four steps:

1. As it is not known beforehand which of the constraints will be active for the optimal solution, in a first step the ICLS problem is solved (for example via the active-set method described in Sect. 3.5.1).

2. Afterwards, the Lagrange multipliers are used to identify the active constraints $\boldsymbol{W}, \boldsymbol{w}$ (cf. Sect. 3.5.1.1). This is possible, as the rule of complementary slackness (3.7e) states that only Lagrange multipliers of active constraints are different from zero (cf. Sect. 3.3.1). We will later use the reverse conclusion that the larger the value of a Lagrange multiplier is, the stronger is its influence on the solution.

3. Having identified those constraints which hold as equalities, and therefore have an associated Lagrange multiplier with a positive value, a standard equality constrained least-squares estimate (cf. Sect. 2.2.1.2) can be carried out, discarding all inactive constraints

ECLS SUBPROBLEM	
objective function:	$\Phi(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{N}\boldsymbol{x} - 2\boldsymbol{n}^T\boldsymbol{x} + \boldsymbol{\ell}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} \ldots \text{Min}$
constraints:	$\boldsymbol{W}^T\boldsymbol{x} = \boldsymbol{w}$
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m$.

(4.1)

4. A variance-covariance (VCV) matrix of the equality constrained problem is computed.

One severe shortcoming of this method is that—due to the restriction on the first two moments—one only gets a symmetric PDF of the ICLS problem rather than the one shown in Fig. 4.1, where the ICLS PDF is the region bounded by the constraints. Though the inactive constraints do not contribute to the estimation process, it will be shown in Sect. 4.5 that they do matter in estimating a PDF. Furthermore, the treatment of active constraints as equalities may lead to overoptimistic variances. One can easily imagine worst case scenarios where it even leads to a (highly unrealistic) variance of zero, for example the univariate case with one active constraint. As only one realization of the random variable $\boldsymbol{\mathcal{L}}$ is used to compute the set of active constraints (and the VCV matrix), the approach is not robust to changes in the active set due to (small) changes in the observations.

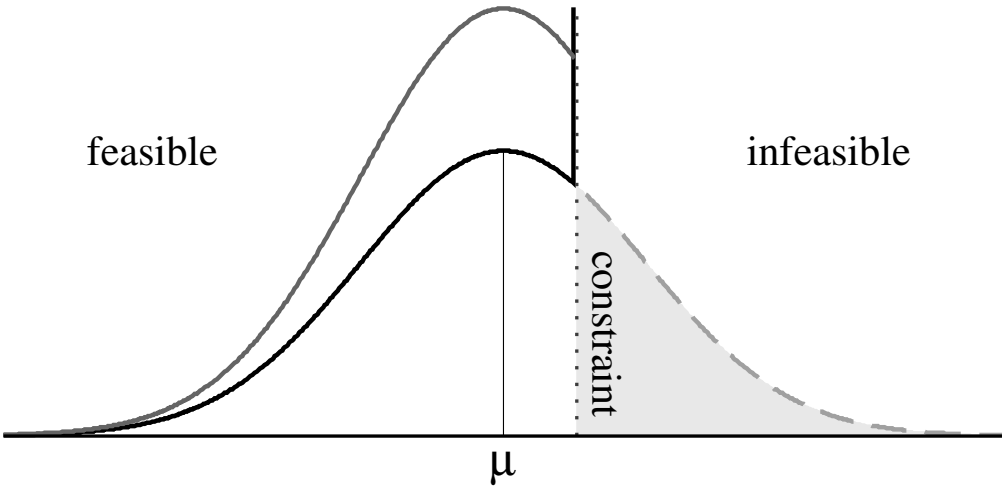


Figure 4.1: Effect of a single inequality constraint (*dotted line*) on the PDF of a parameter with expectation value μ (univariate case). As the probability mass has to be conserved, it is necessary to decide how to handle the part of the probability mass of the unconstrained estimate (*dashed line*) that is truncated by the constraint (*gray shaded area*). This could either be done by scaling the whole function (approach of Zhu et al., 2005, *gray line*) or by accumulating that probability mass at the boundary of the feasible set (MC-QP approach, *black line*). Figure taken from Roese-Koerner et al. (2012).

4.1.2 Bayesian Approach

Besides the frequentist approaches, the problem of ICLS estimation can also be tackled using Bayesian statistics. As a consequence, the unknown parameters are treated as random vectors (in contrast to the frequentist approach, cf. Koch, 2007, p. 34). Thus, the inequality constraints are converted into prior information on the parameters. Geweke (1986) suggested this approach, which was further developed and introduced to geodesy by Zhu et al. (2005).

According to Bayes's theorem

$$f(\boldsymbol{x}|\boldsymbol{\ell}) \quad = \quad \frac{f(\boldsymbol{\ell}|\boldsymbol{x})f(\boldsymbol{x})}{f(\boldsymbol{\ell})} \tag{4.2}$$

the posterior probability density $f(\boldsymbol{x}|\boldsymbol{\ell})$ of the parameters, given a set of observations, can be described as the product of the likelihood function

$$f(\boldsymbol{\ell}|\boldsymbol{x}) = f(\boldsymbol{v}) \quad \propto \quad \exp\left\{-\frac{1}{2\sigma_0^2}\boldsymbol{v}^T\,\boldsymbol{\Sigma}^{-1}\,\boldsymbol{v}\right\} \tag{4.3}$$

(depending on the residuals $\boldsymbol{v}$ and the VCV matrix of the observations $\boldsymbol{\Sigma}$) and the prior distribution $f(\boldsymbol{x})$ of the parameters (cf. Koch, 2007, p. 89). Usually a uniform distribution of the form

$$f(\boldsymbol{x}) \quad \propto \quad \begin{cases} 1 & , \text{ if } \boldsymbol{B}^T\boldsymbol{x} \leq \boldsymbol{b} \\ 0 & , \text{ otherwise} \end{cases} \tag{4.4}$$

is used. In statistics, the normalization term $f(\boldsymbol{\ell})$ is often neglected, resulting in

$$f(\boldsymbol{x}|\boldsymbol{\ell}) \quad \propto \quad f(\boldsymbol{\ell}|\boldsymbol{x})f(\boldsymbol{x}). \tag{4.5}$$

The idea with this method is that we assume that the observations $\boldsymbol{\ell}$ and the residuals $\boldsymbol{v}$ *a priori* follow the normal probability density function $f(\boldsymbol{\ell})$ and $f(\boldsymbol{\ell}|\boldsymbol{x}) = f(\boldsymbol{v})$, respectively, and further

we assume that the possible solutions $\boldsymbol{x}$ *a priori* follow a uniform density function within the given constrained bounds. Thus, inserting these probability densities in (4.5) provides us with an *a posteriori* density function as follows

$$f(\boldsymbol{x}|\boldsymbol{\ell}) \quad \propto \quad \begin{cases} \exp\left\{-\frac{1}{2\sigma_0^2}\boldsymbol{v}^T\,\boldsymbol{\Sigma}^{-1}\,\boldsymbol{v}\right\} & , \quad \text{if} \quad \boldsymbol{B}^T\boldsymbol{x} \leq \boldsymbol{b} \\ 0 & , \quad \text{otherwise} \end{cases} . \tag{4.6}$$

The posterior density is a truncated (piece-wise continuous) version of the non-informative prior distribution of the observations, where the truncation points are determined by the constraints. If the truncated posterior distribution has to be treated like a probability density function, then the area under the function must be equal to one. This is not the case for the truncated function. One way of dealing with this condition is to normalize the posterior density with a constant factor. However, this results in a scaling of the whole PDF (see *gray line* in Fig. 4.1) meaning that the constraints influence not only the boundary regions but the entire PDF.

An advantage of this approach is that it gives an analytical expression for the stochastic description of the ICLS estimate. However, the numerical evaluation of the PDF is computationally very expensive in the multivariate case. Furthermore, distributing the probability mass that is outside the feasible region over the whole PDF (i.e., scaling), might not be the most realistic treatment. We suggest that a more proper treatment would be to move the probability mass outside the feasible region only as far as needed to satisfy the constraints (see *black line* in Fig. 4.1). Therefore, in the next section we propose a (frequentist) Monte Carlo approach, which follows the idea described above.

One could also think of a modification of the Bayesian method (for example using Dirac delta functions to model the singularities at the boundaries of the feasible set as is done in Albertella et al., 2006). However, the modification of the Bayesian approach will not be pursued here as it is not the intention of this thesis.

4.2 Quality Description

Along the line of thought described at the end of the last section we develop a method to compute a (possibly multivariate) PDF of the estimated parameters with the property that all the probability mass in the infeasible region is projected to the nearest spot in the feasible region due to the metric of our objective function. In the absence of analytical expressions, we use a Monte Carlo method to compute an empirical PDF of the parameters, which also allows to derive confidence regions.

4.2.1 Deriving the Posterior PDF of the Parameters

The general idea is to generate M samples of the observations, according to their distribution (which is assumed to be known). All M realizations of the observations are seen as independent problems and solved via an optimization method (e.g., an active-set method) resulting in M realizations of the estimated parameters. If M is chosen large enough, the histogram of the parameters will be an adequate approximation of the PDF of the estimated parameters. In the following, all steps will be described in detail.

For the quality description, it is essential that the design matrix $\boldsymbol{A}$ and the VCV matrix $\boldsymbol{\Sigma}$ are of full column rank. Therefore, it is not possible to combine the stochastic framework described in this chapter with the ansatz to handle rank-deficient problems provided in Chap. 5.

First we compute the ICLS solution $\widetilde{\boldsymbol{x}}$ and the WLS solution of the unconstrained problem (2.32)

$$\widehat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell}, \tag{4.7}$$

which will be used to determine the expectation value of the observations

$$E\{\boldsymbol{\mathcal{L}}\} = \widehat{\boldsymbol{\ell}} = E\{\boldsymbol{A}\boldsymbol{\mathcal{X}}\} = \boldsymbol{A}\widehat{\boldsymbol{x}}. \tag{4.8}$$

In the following, the adjusted observations $\widehat{\boldsymbol{\ell}}$ are interpreted as approximation of the true observations $\boldsymbol{\ell}$ and thus treated as a specific realization of a random vector. Hence, the vector $\widehat{\boldsymbol{\ell}}$ is a constant.
Henceforth in this chapter, unconstrained quantities are marked with a hat to distinguish them from quantities of an adjustment with constraints, which are indicated by a tilde. Assuming the most general case in which the VCV matrix $\boldsymbol{\Sigma}$ of observation vector $\boldsymbol{\mathcal{L}}$ is fully populated, a Monte Carlo simulation for correlated data can be carried out using the Cholesky factorization (cf. Alkhatib and Schuh, 2007). Therefore, the positive definite VCV matrix $\boldsymbol{\Sigma}$ of the observations is decomposed into the product of a lower triangular matrix $\boldsymbol{R}^T$ and an upper triangular matrix $\boldsymbol{R}$:

$$\boldsymbol{\Sigma} = \boldsymbol{R}^T\boldsymbol{R}. \tag{4.9}$$

This is expedient to model the full stochastic information of the observations. Afterwards, M independent samples $\boldsymbol{s}_{\boldsymbol{e}}^{(i)}$ are generated using the standard normal distribution $\boldsymbol{\mathcal{E}} \sim N(\boldsymbol{0}, \boldsymbol{I})$. The superscript (i) denotes the number of the sample $(i = 1, 2, ...M)$. Now the vector $\boldsymbol{s}_{\boldsymbol{e}}^{(i)}$ is transformed to

$$\boldsymbol{s}_{\Delta\boldsymbol{\ell}}^{(i)} = \boldsymbol{R}^T\boldsymbol{s}_{\boldsymbol{e}}^{(i)}.$$

All vectors generated in that manner are realizations of the random vector $\Delta\boldsymbol{\mathcal{L}} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ representing the colored noise of the observations. Adding the noise vectors to the estimated observations $\widehat{\boldsymbol{\ell}}$ (which are treated as constant quantities) we get M realizations of the observation vector

$$\boldsymbol{\ell}^{(i)} = \widehat{\boldsymbol{\ell}} + \boldsymbol{s}_{\Delta\boldsymbol{\ell}}^{(i)}. \tag{4.10}$$

For each of these M realizations of the observations, we compute a sample $\boldsymbol{s}_{\widetilde{\boldsymbol{x}}}^{(i)}$ of the estimated parameters $\widetilde{\boldsymbol{x}}$ using the binding-direction primal active-set method to solve the ICLS problem (cf. Sect. 3.5.1). Usually, when performing a Monte Carlo simulation to determine the accuracy of an estimate, an empirical VCV matrix is estimated from the parameters

$$\boldsymbol{\Sigma}\{\widetilde{\boldsymbol{\mathcal{X}}}\} = E\{(\widetilde{\boldsymbol{\mathcal{X}}} - E\{\widetilde{\boldsymbol{\mathcal{X}}}\})(\widetilde{\boldsymbol{\mathcal{X}}} - E\{\widetilde{\boldsymbol{\mathcal{X}}}\})^T\}. \tag{4.11}$$

However, as mentioned before, this second central moment would not contain the full stochastic information in the inequality constrained case because we have to deal with truncated PDFs. Therefore, it is more conducive to compute an m-dimensional histogram of the parameters. This histogram can be seen as a discrete approximation of the joint PDF of the parameters. Approximations of the marginal densities can be computed the same way, adding up the particular rows of the hyper matrix of the histogram. The quality of approximation of the continuous PDF depends directly on M (cf. Alkhatib and Schuh, 2007), which therefore has to be chosen in a way that allows a satisfactory approximation while keeping the computation time at an acceptable level. In each Monte Carlo iteration a new optimization problem has to be solved. However, as the solution of the original ICLS problem can be used as feasible initial value for the parameters, convergence of the active-set method is usually quite fast.

It should be emphasized again that we compute a histogram as discrete approximation of a continuous PDF here. As a consequence, all probability mass in the infeasible region concentrates at the

"last bin(s)" in the feasible region. In a continuous PDF this probability mass would concentrate exactly on the boundary of the feasible region and thus have an extent of zero. This would lead to problems with the claim that the integral of a density has to be one. In the discrete histogram on the other hand all bins have an extent larger than zero. Thus, it is possible to quantify the probability mass concentrated at a single point and the probability mass of all bins sums up to one.

4.2.2 Optimal ICLS Solution

Having obtained an approximation of the posterior density, one can think of at least four different possibilities to define an optimal point estimate (cf. Zhu et al., 2005): The mean, the median, the mode and the solution that minimizes the original ICLS problem. As the introduction of inequality constraints might lead to multi-modal distributions (due to the accumulation at the boundaries), mean and mode are in general improper.

Empirical studies have shown that the median of the PDF and the point that minimizes the original problem often are very similar but not necessarily the same. Henceforth in this thesis, we will use the term *solution* to refer to the solution of the original problem as it is more convenient to compute. As this is either the WLS estimate or its projection onto the boundary of the feasible set, it is *best* in the sense that it is the solution with the smallest sum of squared residuals of all feasible points.

4.2.3 Highest Posterior Density (HPD) Regions

Different definitions of confidence regions have been developed. One concept, which is perfectly suited to be combined with Monte Carlo methods is called *highest posterior density* (HPD) region and originates from Bayesian statistics (cf. Koch, 1999, p. 71). It gives a quality measure of the estimate in form of a region Ω containing a certain percentage (e.g., 95 percent) of the samples

$$P(\boldsymbol{x} \in \Omega | \boldsymbol{\ell}) = 1 - \alpha. \tag{4.12}$$

$1-\alpha$ is the level of significance. According to Chen and Shao (1999) a region is called a HPD region if it is the smallest possible region with the property that every point inside has a higher density than every point outside the region.

Benefits of HPD regions are that they do not rely on (asymptotic) normality assumptions, and are able to describe also multimodal densities. However, one has to be aware that they are computationally expensive to obtain, and may not be connected in the multimodal case (Chen and Shao, 1999, p. 84). As stated in GUM Supplement 1 (Joint Committee for Guides in Metrology, 2008, p. 30), HPD intervals of a unimodal and one-dimensional problem can be computed by sorting the results of the Monte Carlo study and discarding the smallest and largest $\frac{\alpha}{2}$ percent.
For the multivariate and/or multimodal case, this region can be computed in a similar fashion. However, we have to use the values of the m-dimensional histogram described above instead of the estimates themselves. They are sorted by value, starting with the largest one. Afterwards, the cumulative sum is computed until $1-\alpha$ percentage is reached. All bins of the histogram that added up to $1-\alpha$ percentage form the confidence region. To allow for a meaningful sorting the chosen number M of Monte Carlo iterations has to be large enough. This is especially true for high dimensional problems.

In contrast to the traditional approach working with the first two moments of the PDF, in the HPD approach no assumptions about the geometry of the confidence regions are needed. This is a necessary feature as we will have to deal with ellipses truncated by constraints and possibly extended along the boundary of the feasible set.

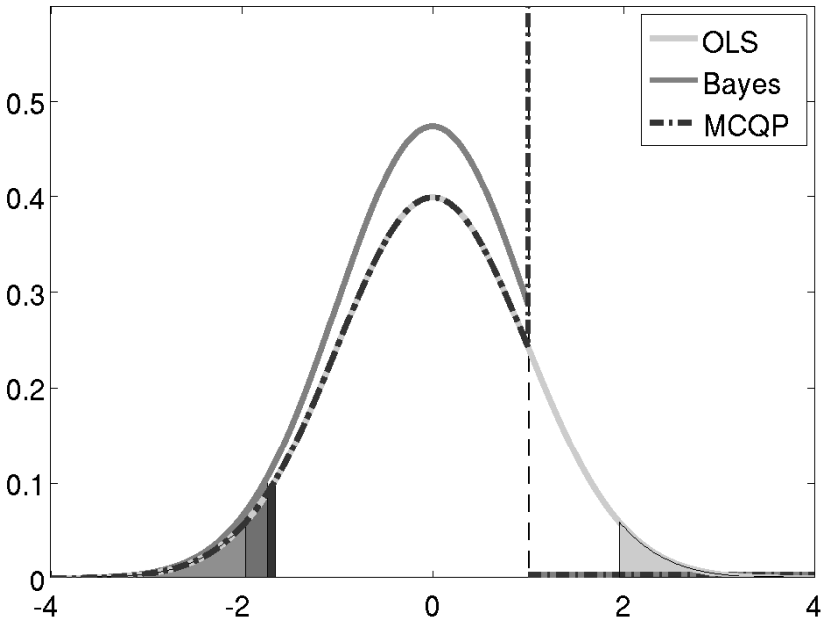


Figure 4.2: Approximations of the probability density functions and confidence intervals of different estimates for a one-dimensional problem and $M \to \infty$: OLS (*light gray*), Bayesian ICLS approach (*gray*) and MC-QP method (*dark gray, dash-dotted*). The *dashed line* represents the inequality constraint $x \leq 1$ and the minimal value of the objective function is reached for $x = 0$. The parts not contained in the confidence intervals are the *shaded regions* below the curves. In contrast to an OLS estimate the confidence region of an ICLS estimate can be asymmetric, which is why the OLS PDF has symmetric shaded regions, while those of Bayes and MC-QP are one-sided. Figure taken from Roese-Koerner et al. (2015).

Figure 4.2 illustrates such confidence intervals for a one-dimensional example for $M \to \infty$. The PDF of the OLS estimate with $E\{\mathcal{X}\} = 0$ is plotted in *light gray*. The α percent not included in the confidence interval (*shaded areas*) are symmetrically distributed at both tails of the distribution ($\frac{\alpha}{2}$ at each side). This symmetry is destroyed when introducing the inequality constraint $x \leq 1$ and performing an ICLS estimate.

The PDF of the MC-QP estimate (cf. Sect. 4.4) is indicated by the *dash-dotted dark gray line*, which coincides with the OLS estimate in the feasible region and accumulates all the probability mass in the infeasible region at its boundary. Thus, as the confidence interval contains the values which are most likely, the whole α percent not included in the confidence interval are at the left tail of the PDF (depending on the constraint). As a consequence, the confidence interval is bounded on one side by the constraint and on the other side by the α percent that are "most unlikely". As can been seen in Fig. 4.2, this results in a smaller confidence interval.

However, this is not true for the Bayesian estimate (*gray curve*). The symmetry is destroyed here as well, but the scaling of the PDF leads to a shift of the beginning of the $(1 - \alpha)$-percent-interval and therefore to a bigger interval compared to the MC-QP method.

4.3 Analysis Tools for Constraints

Besides the actual quality description, it might also be reasonable to measure the influence of the constraints on the solution. We will present tools to perform such an analysis: Two global measures to investigate if the data support the constraints in general, and a local measure to determine the influence of each constraint.

4.3.1 Global Measure I: Testing the Plausibility of the Constraints

In order to test if the introduction of inequality constraints results in a significant change in the parameters, a testing procedure according to Wald (1943) is applied, which was first used in the inequality constrained case by Koch (1981). Usually this test is used in the equality constrained case. In order to apply the Wald test for inequality constraints, the ICLS problem is solved, all p_w active constraints

$$\boldsymbol{W}^T\boldsymbol{x} = \boldsymbol{w}$$

are treated as equality constraints and all inactive constraints are neglected as in the approach of Liew (1976). Then the estimation is carried out as a two-step approach.
In a first step a WLS adjustment is done:

$$\boldsymbol{\ell} + \boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}, \quad \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{L}}\} = \boldsymbol{\Sigma} \tag{4.13}$$

$$\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\} = (\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1} \tag{4.14}$$

$$\widehat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} = \boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} \tag{4.15}$$

$$\widehat{\boldsymbol{v}} = \boldsymbol{A}\widehat{\boldsymbol{x}} - \boldsymbol{\ell}. \tag{4.16}$$

Afterwards, an adjustment of the equality constraints is carried out

$$\boldsymbol{W}^T(\widehat{\boldsymbol{x}} + \boldsymbol{r}) = \boldsymbol{w} \tag{4.17}$$

$$\boldsymbol{r} = -\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}\boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}\boldsymbol{W})^{-1}(\boldsymbol{W}^T\widehat{\boldsymbol{x}} - \boldsymbol{w}) \tag{4.18}$$

$$\widetilde{\boldsymbol{x}} = \widehat{\boldsymbol{x}} + \boldsymbol{r}. \tag{4.19}$$

Now, the estimated *a posteriori* variance factor of the first step

$$\widehat{s}_1^2 = \frac{\widehat{\boldsymbol{v}}^T\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{v}}}{n-m} \tag{4.20}$$

and of the second step

$$\widehat{s}_2^2 = \frac{\boldsymbol{r}^T\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}^{-1}\boldsymbol{r}}{p_w} \tag{4.21}$$

are computed and the actual hypothesis testing is carried out via the Wald test:

Test hypothesis H_0 vs. hypothesis H_A:
H_0: The changes through the equalities are **not** significant.
H_A: The changes through the equalities are significant.

As $\widehat{\mathcal{S}}_1^2$ and $\widehat{\mathcal{S}}_2^2$ are independent (c.f. Koch, 1999, p. 273–274), the test statistic

$$\mathcal{T} = \frac{\widehat{\mathcal{S}_2^2}}{\widehat{\mathcal{S}_1^2}} \sim F_{p_w,n-m} \tag{4.22}$$

follows the Fisher distribution, as it is the ratio of two independent, χ^2 distributed quantities, which are reduced by their degrees of freedom ($n-m$ and p_w, respectively). Now, a level of significance $1-\alpha$ is chosen (e.g., 95 percent) and a one-sided test is carried out. If the test statistic of the Wald test is less than or equal to the critical value of the Fisher distribution, then the changes due to the constraints are not significant. This is equivalent to the statement that the constraints only lead to negligible changes in the parameters. If the test statistic is greater than the critical value, then the constraints are very *strong* and will change the result significantly. In this case, each constraint can be tested separately, if required.

Similar to Liew (1976), we can conclude from this test solely if the equality constraint problem is supported by the data or not. We can draw no conclusions what will happen if other inequality constraints become active, because in that case a different hypothesis testing would be carried out, due to the changes in the test statistic and in the degrees of freedom. However, for the set of constraints active in the actual solution, the hypothesis testing does allow for an interpretation as stated above.

4.3.2 Global Measure II: Probability Mass in the Infeasible Region

Another global measure of the change in the result due to the introduction of constraints is the ratio d of estimates in which at least one constraint is active (ICLS estimates) compared to the total number of estimates (= number of Monte Carlo samples). If at least one constraint is active, then the unconstrained WLS solution will be in the infeasible region. Therefore, d is an unbiased estimator of the probability mass outside the feasible region. If it is close to one, then in nearly every sample of the Monte Carlo study the optimal solution is projected onto the boundary of the feasible set. If d is close to zero, then the constraints have solely a very small influence on the estimation process.

4.3.3 Local Measure: Sensitivity Analysis

In order to determine the influence of each constraint on each parameter, the derivative with respect to the parameters $\boldsymbol{x}$ of the Lagrangian (3.23) of the ICLS is revisited. As stated in equation (3.25b), setting it equal to zero and resolving for $\boldsymbol{x}$ yields

$$\widetilde{\boldsymbol{x}} = \boldsymbol{N}^{-1}(\boldsymbol{n} - \boldsymbol{B}\widetilde{\boldsymbol{k}} + \overline{\boldsymbol{B}}\,\widetilde{\overline{\boldsymbol{k}}}).$$

The Lagrange multipliers $\widetilde{\boldsymbol{k}}$ and $\widetilde{\overline{\boldsymbol{k}}}$ are marked with a tilde, too, to emphasize that the adjusted quantities are used (i.e., after the algorithm has converged). The equation stated above can be rewritten as

$$\begin{aligned}
\widetilde{\boldsymbol{x}} &= \boldsymbol{N}^{-1}(\boldsymbol{n} - \boldsymbol{B}\widetilde{\boldsymbol{k}} + \overline{\boldsymbol{B}}\,\widetilde{\overline{\boldsymbol{k}}}) && \text{(4.23a)}\\
&= \boldsymbol{N}^{-1}\boldsymbol{n} - \boldsymbol{N}^{-1}\boldsymbol{B}\widetilde{\boldsymbol{k}} + \boldsymbol{N}^{-1}\overline{\boldsymbol{B}}\,\widetilde{\overline{\boldsymbol{k}}} && \text{(4.23b)}\\
&= \widehat{\boldsymbol{x}} \underbrace{-\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}\boldsymbol{B}\widetilde{\boldsymbol{k}} + \boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}\overline{\boldsymbol{B}}\,\widetilde{\overline{\boldsymbol{k}}}}_{\Delta\boldsymbol{x}} = \widehat{\boldsymbol{x}} + \Delta\boldsymbol{x}. && \text{(4.23c)}
\end{aligned}$$

With this explicit relation between the unconstrained solution $\widehat{\boldsymbol{x}}$, the constrained solution $\widetilde{\boldsymbol{x}}$ and the Lagrange multipliers $\widetilde{\boldsymbol{k}}$ and $\widetilde{\overline{\boldsymbol{k}}}$ it is possible to perform a sensitivity analysis.
The perturbation of $\widehat{\boldsymbol{x}}$ consists of three parts:

1. influence of the Lagrange multipliers $\widetilde{\boldsymbol{k}}$ and $\widetilde{\overline{\boldsymbol{k}}}$,
2. influence of the design $\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}$,
3. influence of the matrix $\boldsymbol{B}$ of constraints.

As a rule of thumb one can say that the larger the value of the Lagrange multiplier is, the larger is the perturbation by the related active constraint. If there are no correlations between the parameters and each constraint contains only one parameter (called independent constraints), then

one constraint only influences one parameter. If there are correlations between the parameters, then the constraints will also have an influence on all the correlated parameters. The individual influence of each constraint on the parameters can be determined (only Lagrange multipliers of active constraints have values different from zero) by evaluating (4.23a). It should be kept in mind that equality constraints are always active.

According to Boyd and Vandenberghe (2004, p. 252) the Lagrange multipliers can be interpreted as a measure for the *activeness* of a constraint. If the Lagrange multiplier $\widetilde{k}_i$ is zero, there will be no change in the sum of squared residuals $\Phi(\boldsymbol{x})$ through the ith constraint. For small values of $\widetilde{k}_i$, there will be an effect on $\Phi(\boldsymbol{x})$ which is small, whereas for large $\widetilde{k}_i$ even small changes to the constraints can result in great changes in $\Phi(\boldsymbol{x})$. Needless to say that this is just a rule of thumb, as in (4.23a) there still are the influences of the VCV matrix $\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}$ and the constraint matrices $\boldsymbol{B}$ and $\overline{\boldsymbol{B}}$.

4.4 The Monte Carlo Quadratic Programming (MC-QP) Method

In order to gain as much information about the quality of the estimate, all tools described above are now combined into a framework for the stochastic description of ICLS estimates. In this Monte Carlo Quadratic Programming (MC-QP) method, we are no longer restricted to computing solely the solution of an ICLS estimator, but can also determine some of its statistical properties. The proposed stochastic framework is summarized as pseudo code in Alg. 3.

Algorithm 3: A Stochastic framework for ICLS estimates: The MC-QP method (modified from Roese-Koerner et al. (2012)).

```
[x̂, ℓ̂] = solveWlsProblem(A, ℓ, Σ)                                   // cf. Sect. 2.2.1.1
[x̃, ℓ̃, k̃] = solveIclsProblem(A, ℓ, Σ, B, b, B̄, b̄, x̂)               // cf. Sect. 3.5.1
hypothesisTesting(x̂, A, Σ, ℓ̂, B, b, B̄, b̄)                          // cf. Sect. 4.3.1
for i = 1 : M do
    generate sample s_L^(i) of L ~ N(A X̂, Σ)                        // cf. Sect. 4.2.1
    [s_x̃^(i), s_ℓ̃, s_k̃] = solveIclsProblem(s_L^(i), A, Σ, B, b, B̄, b̄, x̃)   // cf. Sect. 3.5.1
end
f = computeEmpiricalPdf(s_x̃)                                        // cf. Sect. 4.2.1
deriveConfidenceRegions(f)                                          // cf. Sect. 4.2.3
d = computeProbabilityMassInInfeasibleRegion(s_k̃)                   // cf. Sect. 4.3.2
sensitivityAnalysis(A, Σ, B, B̄, k̃, k̃̃)                              // cf. Sect. 4.3.3
```

First the WLS and ICLS solutions are determined and an hypothesis testing is carried out to determine if the constraints are plausible. If the null hypothesis is discarded (meaning that the data do not support the constraints), one can decide whether to compute the solution with or without constraints (depending on the problem).

Afterwards a Monte Carlo simulation is carried out, the empirical probability density function of the parameters $\widetilde{\boldsymbol{\mathcal{X}}}$ is computed, and their confidence regions are derived. The influence of each constraint on each parameter can be determined in a sensitivity analysis using the Lagrange multipliers $\widetilde{\boldsymbol{k}}$ and $\widetilde{\widetilde{\boldsymbol{k}}}$. As an overall measure for the influence of the constraints the percentage d of the probability mass outside the feasible region and on its boundary can be computed.

It should be emphasized again that is approach is only valid for problems in which the design matrix $\boldsymbol{A}$ as well as the VCV matrix $\boldsymbol{\Sigma}$ are of full column rank.

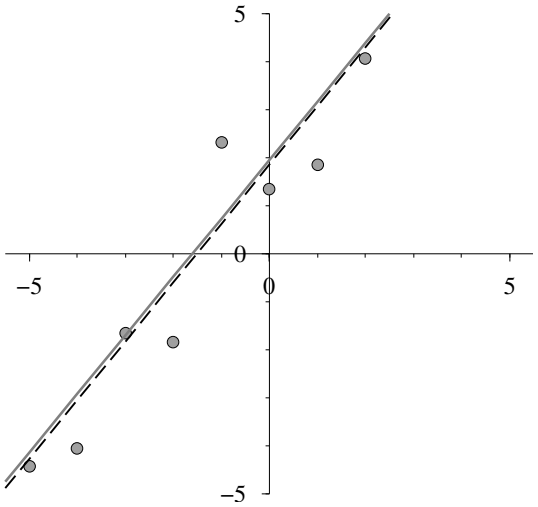


Figure 4.3: Illustration of some of the observations (*gray dots*), the WLS line-fit (*gray line*) and the ICLS line-fit (*dashed black line*). Figure taken from Roese-Koerner et al. (2012).

4.5 Example

In this section, two bivariate examples will be shown to elucidate the features of the MC-QP method. In both cases the slope (x_1) and the intercept (x_2) of a line of best fit (cf. Fig. 4.3) are to be estimated. Stochastic description based on the Bayesian method of Zhu et al. (2005) will also be shown for comparison. We restricted ourselves to inequality constraints only, as this is the most interesting case. However, adding additional equality constraints is straightforward.

In the application part (Chap. 7), two more examples are provided in which the proposed stochastic framework proves useful. In Sect. 7.2.1 the task ist to estimate a positive definite covariance function and in Sect. 7.2.2 we apply constraints on the tropospheric delay in a VLBI setting. The examples focus mainly on the sensitivity analysis capabilities to illustrate the full potential of the MC-QP approach.

4.5.1 Line of Best Fit with Independent Constraints

The observations along the line are generated by taking an arbitrary slope $x_1 = 1.3$ and intercept $x_2 = 1.5$ and to these true values white-noise is added in form of the vector $\boldsymbol{\mathcal{E}} \sim N(\mathbf{0}, \boldsymbol{I})$:

$$\ell_i = t_i\, x_1 + x_2 + e_i \tag{4.24a}$$

$$\boldsymbol{\ell} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}, \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{4.24b}$$

To this line-fitting problem inequality constraints are added, which are given as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 1.30 \\ 1.85 \end{bmatrix}. \tag{4.24c}$$

Therefore, the constraint matrix becomes $\boldsymbol{B} = \boldsymbol{I}$. The identity matrix implies that the constraint applied to one parameter is functionally independent of the one applied to the other parameter. Such constraints will be referred to as *independent constraints*. In such cases, the application of constraints is equivalent to applying constraints in a univariate case, provided the parameters are uncorrelated.

Table 4.1: Numerical results from the line-fitting problem with independent constraints. The table on the left shows the WLS ($\widehat{\boldsymbol{x}}$) and ICLS ($\widetilde{\boldsymbol{x}}$) estimates. The table on the right shows the sensitivity analysis performed with Lagrange multipliers ($\widetilde{\boldsymbol{k}}$), which indicate that there is only one active constraint ($\widetilde{k}_2$) and it contributes to the perturbations ($\Delta\boldsymbol{x}$) in the ICLS estimates. $d = 67\,\%$ is the percentage of probability mass in the infeasible region. The result of the Wald test shows that the test statistic 0.17 is smaller than the critical value of 4.38, meaning that the data supports the constraints (cf. Sect. 4.3.1).

	$\widehat{\boldsymbol{x}}$	$\widetilde{\boldsymbol{x}}$
x_1	1.218	1.224
x_2	1.953	1.850

	$\widetilde{k}_1 = 0$	$\widetilde{k}_2 = 3.474$
Δx_1	0	0.007
Δx_2	0	−0.103

ICLS estimates : $d = 67\,\%$, Wald test: $0.17 < 4.38$

Now, the stochastic framework outlined in Sect. 4.4 is applied to this ICLS problem. The numerical results show that the WLS estimate of the slope already satisfies the constraint but that of the intercept does not. After applying the ICLS estimation both estimates have been changed (cf. Tab. 4.1). Sensitivity analysis of the parameters using the Lagrange multipliers ($\widetilde{\boldsymbol{k}}$) shows that the constraint on the slope is inactive ($\widetilde{k}_1 = 0$), while the constraint on the intercept is active ($\widetilde{k}_2 = 3.474$). However, the active constraint on the intercept contributes to both the changes in slope and intercept, and that is due to the negative correlation introduced by the design matrix $\boldsymbol{A}$. Analytically, this can be explained using (4.23a), where if $\boldsymbol{B} = \boldsymbol{I}$ and only inequality constraints exist, then

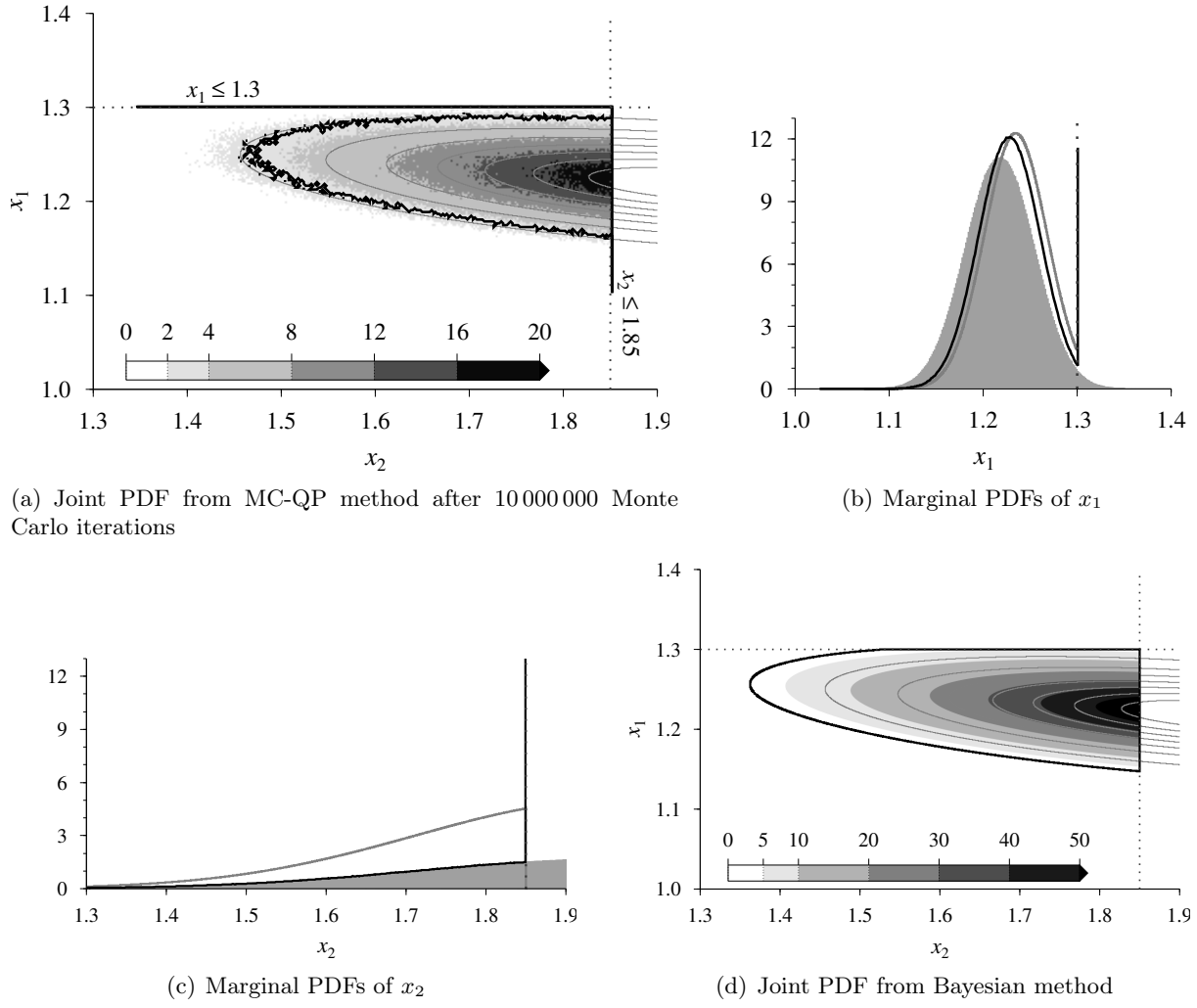
$$\Delta\boldsymbol{x} = -\frac{1}{2}\,\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}\,\widetilde{\boldsymbol{k}}. \tag{4.25}$$

In (4.25), the VCV matrix $\boldsymbol{\Sigma}\{\widehat{\boldsymbol{\mathcal{X}}}\}$ describes the correlations between the parameters. If the parameters are correlated in the independent constraints case, they directly affect the perturbations $\Delta\boldsymbol{x}$.

It is valid to question the utility of the inactive constraints in ICLS problems as they do not directly contribute to the estimation process. However, the inactive constraints in addition to the active constraints define the parameter space, and therefore the probability space of the estimates. Therefore, while those constraints that are known to be inactive in the optimal solution can be neglected within the estimation process, they might be essential for describing the quality of the parameter estimates.

The quality of the estimates from WLS, MC-QP and the Bayesian methods are shown in Fig. 4.4. The peak of the ICLS PDF of slope (x_1) is slightly shifted from that of the WLS PDF. This is an interesting case, since the constraints that were applied were independent of each other. The reason for this shift in the peak is the negative correlation between slope and intercept of the line: if the slope increases the intercept has to decrease and vice-versa. A similar shift has not taken place in the PDF of the intercept, whose WLS estimate has not satisfied the constraints. This is due to the spread of the probability densities of the intercept, and hence the slope introduced by the correlation does not affect the accumulation of the densities. Therefore, the inactively constrained estimates will undergo a shift in their values based on the correlation between the parameters and the size of the changes in the values of the actively constrained estimates. In this context, it should be mentioned that even though the constrained and unconstrained estimates seem to be very similar in this example (cf. Fig. 4.3), their respective probability densities are entirely different (cf. Fig. 4.4).

In general, whether a frequentist or a Bayesian approach is followed, one should arrive at the same result for the confidence regions and the PDF curves (neglecting the roughness which is a consequence of the Monte Carlo sampling). However, in the example problem (4.24), the two approaches differ drastically (cf. Fig. 4.4). The drastic difference is mainly due to the way in which the boundary



(a) Joint PDF from MC-QP method after 10 000 000 Monte Carlo iterations

(b) Marginal PDFs of x_1

(c) Marginal PDFs of x_2

(d) Joint PDF from Bayesian method

Figure 4.4: Probability density functions (joint and marginal) from the line-fitting problem with independent inequality constraints. The *contours* in joint PDFs and *gray bars* in the marginal PDFs are from WLS estimates. In the marginal PDFs *black lines* indicate MC-QP method and *gray lines* indicate Bayesian approach. The *dotted lines* indicate the truncation by the constraints. The HPD region in the joint PDFs is marked as a *black contour line*. The accumulation (MC-QP) and scaling (Bayesian) difference is clearly evident both in the joint and marginal PDFs. Figure taken from Roese-Koerner et al. (2012).

of the truncated parameter space is treated in the Bayesian method (scaling or accumulating). This difference ends up in the different sizes of the HPD regions: a compact HPD region for the MC-QP approach and wider one for the Bayesian approach. A modification of the Bayesian method for a similar treatment of the boundary conditions should provide the same quality description as the MC-QP method (cf. Sect. 4.1.2).

4.5.2 Line of Best Fit with Dependent Constraints

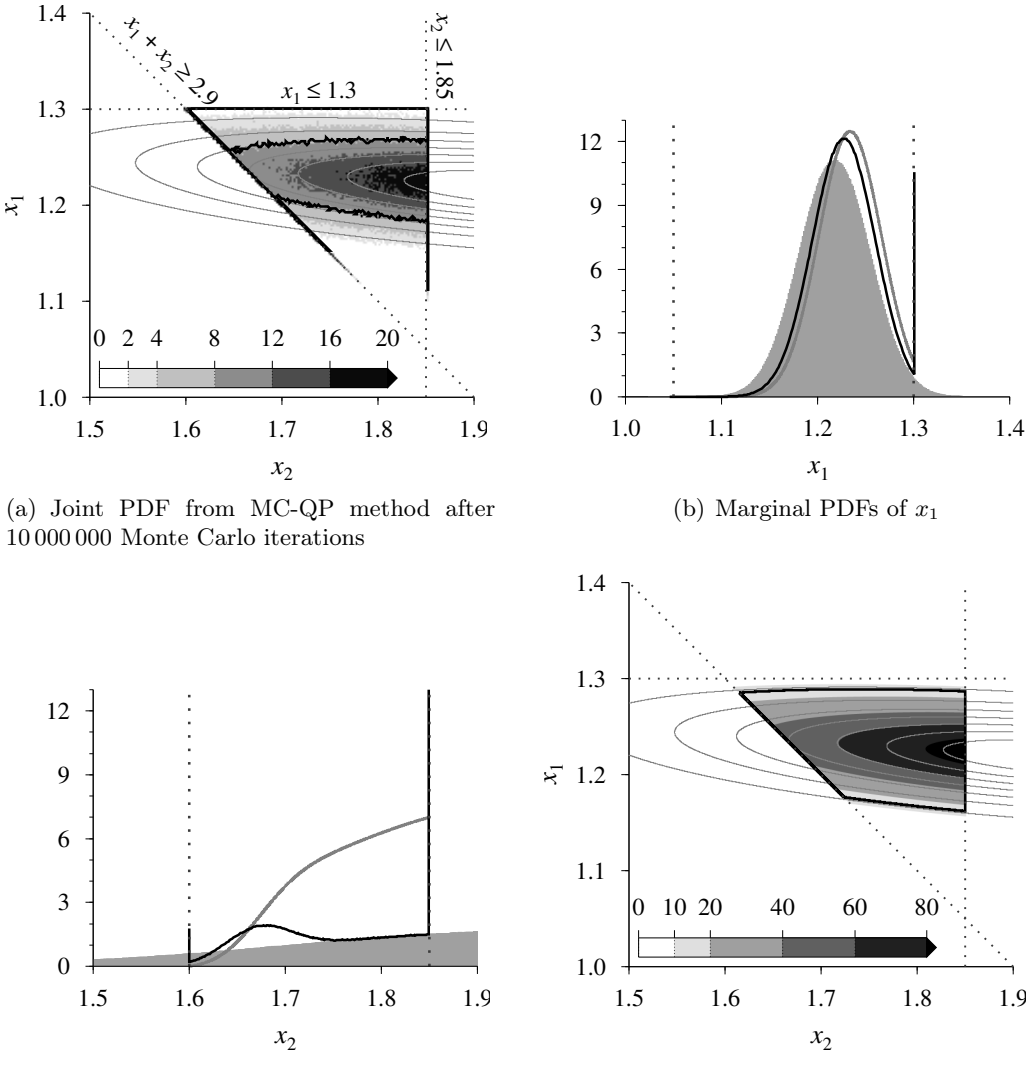
An additional constraint is added to the line-fitting problem defined in (4.24) such that it relates both the parameters, as follows

$$\begin{bmatrix} x_1 \\ x_2 \\ -x_1 - x_2 \end{bmatrix} \leq \begin{bmatrix} 1.30 \\ 1.85 \\ 2.90 \end{bmatrix}. \tag{4.26}$$

Table 4.2: Results from the line-fitting problem with dependent constraints. The arrangement of the tables is the same as that of Tab. 4.1. Despite the ICLS estimates being identical to those in the independent constraints case, the probability mass in the infeasible region, indicated by d, is significantly different. This is a clear indication of the influence of inactive constraints on the statistical properties of the estimate.

	$\widehat{\mathbf{x}}$	$\widetilde{\boldsymbol{x}}$		$\widetilde{k}_1 = 0$	$\widetilde{k}_2 = 3.474$	$\widetilde{k}_3 = 0$
x_1	1.218	1.224	Δx_1	0	0.007	0
x_2	1.953	1.850	Δx_2	0	−0.103	0

ICLS estimates : $d = 79\,\%$, Wald test: $0.17 < 4.38$



(a) Joint PDF from MC-QP method after 10 000 000 Monte Carlo iterations

(b) Marginal PDFs of x_1

(c) Marginal PDFs of x_2

(d) Joint PDF from Bayesian method

Figure 4.5: Probability density functions (joint and marginal) from the line-fitting problem with dependent constraints. From the joint PDFs it is clear that this is a bounded constraint problem as the densities are confined to the triangular region formed by the constraints. Due to the inclined plane $x_1 + x_2 \geq 3$, there is complex accumulation of the probability densities taking place in the marginal PDF of x_2. Figure taken from Roese-Koerner et al. (2012).

Now the constraint matrix $\boldsymbol{B}$ is not an identity matrix anymore due to the constraints being dependent on both the parameters. Such constraints will be called *dependent constraints*. The stochastic framework is applied to the ICLS problem subject to the constraints of (4.26). The third constraint that was added is an inactive constraint, and hence the ICLS estimated parameters have the same values as in the problem with independent constraints. This is further confirmed by the Lagrange multipliers of the constraints (cf. Tab. 4.2).

Though the estimates of the ICLS problem with independent constraints and dependent constraints are equivalent, their joint PDFs are completely different. The striking difference is seen in the Marginal Density Function (MDF) of the intercept x_2 (cf. Fig. 4.5). While the MDF from the independent constraints is only affected at the boundary between the feasible and infeasible region, the MDF from the dependent constraints is affected on either side of the boundary. This is clearly due to the constraint $(x_1 + x_2 \geq 3)$ cutting diagonally across the joint density function. The addition of the third constraint, although inactive, is felt most in the HPD region: The region is far more compact for the dependent constraints than in the independent constraints case.

5. Rank-Deficient Systems with Inequalities

This chapter is a modified and extended version of the articles *Convex optimization under inequality constraints in rank-deficient systems* (Roese-Koerner and Schuh, 2014) and *Effects of different objective functions in inequality constrained and rank-deficient least-squares problems* (Roese-Koerner and Schuh, 2015).

Solving a rank-deficient normal equation system results in not one unique but a manifold of solutions. In the unconstrained case, a rigorous general solution can easily be computed using the theory of generalized inverses (Koch, 1999, p. 48–59). This also includes the computation of a unique solution via the Moore-Penrose inverse. However, this is no longer possible in the presence of inequality constraints. Most state-of-the-art optimization algorithms—e.g., the active-set method described in Sect. 3.5.1 and in Gill et al. (1981, p. 167–173) or the interior-point methods described in Boyd and Vandenberghe (2004, p. 568–571 and p. 609–613, respectively)—are either not able to deal with an objective function with a rank-deficient matrix or yield only one of the infinite number of particular solutions.

Despite the highly sophisticated estimation theory for unconstrained rank-deficient systems and well defined but inequality constrained systems (cf. Sect. 3.2), only few publications (cf. Sect. 5.1) have been dedicated to singular optimization problems with inequality constraints.
However, in geodesy, these problems occur on many occasions. Examples are the second order design of a geodetic network with more weights to be estimated than entries in the criterion matrix, the adjustment of datum-free networks or a spline approximation with data gaps and additional information on the function behavior.

In this chapter a line of thought similar to that of Werner and Yapar (1996) will be followed (cf. Sect. 5.1). Accordingly, we propose a two step approach for a rigorous computation of the general solution. In a first step, the constraints are neglected and a general solution of the unconstrained problem is computed. Subsequently, the constraints are taken into account and a second optimization problem is solved, depending on whether there is an intersection of manifold and feasible region or not. To keep the equations tidy, the focus of this chapter is on problems with inequality constraints only. Possible equality constraints have to be removed before (cf. Boyd and Vandenberghe, 2004, p. 142–143).

The proposed approach has four major advantages. First, it allows for a description of the manifold including the constraints as these will be reformulated in terms of a basis for the nullspace of the design matrix. Second, an additional optimization step is performed in this nullspace to provide a particular solution, which fulfills certain optimality criteria and allows us to aim for desirable properties such as sparsity. Third, it is not restricted to small scale problems. Fourth, the proposed method detects if the dimension of the manifold of the unconstrained problem is reduced through the introduction of inequality constraints (up to the case of a unique solution).

The current state of the art of inequality constrained rank-deficient systems is reviewed in Sect. 5.1. An extension of the binding-direction primal active-set method (cf. Sect. 3.5.1), which facilitates the computation of a particular solution despite a possible rank defect, is provided in Sect. 5.2. The ideas and computations for the rigorous computation of general ICLS solutions are described in Sect. 5.3 and combined to a framework in Sect. 5.4. The simple synthetic example in Sect. 5.5 illustrates the application of the framework and serves as proof of concept. Two close-to-reality examples—the second order design of a geodetic network and an application with strict welding tolerances—are provided in Sect. 7.3.1 in the applications chapter.

5.1 State of the Art

Barrodale and Roberts (1978) presented a modification of the standard simplex method for linear programming, which is able to handle rank-deficient problems. However, extending it to a QP is not straightforward.

Schaffrin (1981) treated the special case that in addition to the linear inequality constraints all parameters are restricted to be non-negative. He developed a method to compute a particular solution through the introduction of slack variables. However, his approach is limited to problems with only non-negative parameters.

Best (1984) identified equivalence relations between different quadratic programming algorithms. This includes an extension of a generic active-set method for singular equation systems. However, no homogeneous solution is provided.

O'Leary and Rust (1986) developed an approach for computing confidence regions for ill-posed weighted non-negative least-squares problems. The inequality $\boldsymbol{x} \geq \mathbf{0}$ is used to truncate the solution space and eliminate the non-uniqueness of the solution. However, their approach cannot easily be extended to general linear constraints as it aims to completely resolve the manifold, which is not always possible.

Fletcher and Johnson (1997) proposed a nullspace method for ill-conditioned QPs with solely equality constraints: The aim is to compute the nullspace of the matrix $\overline{\boldsymbol{B}}^T$ of equality constraints. This allows to reformulate a problem with equality constraints as a problem without constraints. Under certain conditions, it is now possible to compute a solution even if the coefficient matrix or the matrix of constraints is ill-conditioned. This could be applied in an active-set approach to solve an inequality constrained problem as a sequence of equality constrained ones. However, as the focus of their contribution is on ill-conditioned problems and not on rank-deficient ones, the computation of a general solution is not discussed.

Dantzig's Simplex Method for Quadratic Programming (Dantzig, 1998, p. 490–498) allows for the computation of a particular solution in case of a rank-deficient design matrix. However, no statements about the homogeneous solution are given.

Xu et al. (1999) analyzed the stability of ill-conditioned Linear Complementarity Problems (LCP) in geodesy, which could be used to solve an ICLS problem. In case of an unstable LCP due to an ill-conditioned LCP matrix (a case often encountered when processing GPS data), they proposed a regularization of the LCP matrix.

Among many others, Geiger and Kanzow (2002, p. 362–365) described a Tikhonov regularization for ill-conditioned convex optimization problems.

In the projector theoretical approach of Werner (1990) and Werner and Yapar (1996) a rigorous computation of the general solution of ICLS problems with possibly rank-deficient matrices $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ is performed using generalized inverses. First, an ordinary least-squares solution is computed, then the update to the ICLS solutions is computed in an iterative approach. As the ICLS solution is obtained by testing arbitrary subsets of constraints, this approach is mostly suited for small scale problems with few constraints.

Wong (2011) described extensions of different types of active-set methods for rank-deficient systems. Similar to the approach of Dantzig mentioned above, only a particular solution is computed and no statements about the homogeneous solution are provided. Nonetheless, we will revisit her ideas in the next section, because it enables us to find a particular solution that fulfills all constraints.

5.2 Extending Active-Set Methods for Rank-Deficient Systems

When dealing with a rank-deficient system, the critical part of active-set methods (e.g., of the binding-direction primal active-set method described in Sect. 3.5.1) is the computation of the search direction $\boldsymbol{p}$ and of the step length q. This is, because the system of equations (5.1) might be inconsistent, which makes it hard to determine a value for $\boldsymbol{p}$ (cf. Sect. 5.2.1.3). Furthermore, q might not be uniquely determined if the point $\boldsymbol{x}^{(i)}$ is part of the solution manifold of the current subspace.

Best (1984) and Wong (2011), for instance, proposed some techniques to overcome these difficulties. In the following sections, the ideas of Wong (2011) are explained (Sect. 5.2.1) and a modified version of the binding-direction primal active-set method using the Gauß-Jordan algorithm is provided (Sect. 5.2.2), which enables us to compute an optimal—but not necessarily unique—solution despite a rank defect. In our proposed framework this method will be used in case the manifold is resolved through the inequality constraints (cf. Sect. 5.3.4).

5.2.1 Modifications for Rank-Deficient Systems

5.2.1.1 Using a Subspace Minimizer as Initial Point

As pointed out in Sect 3.5.1.6, in each iteration of the algorithm either the point $\boldsymbol{x}^{(i)}$ is a minimum of the subspace defined by problem (3.45)—in this case a constraint can be removed from the active set—or the subspace is modified by adding a new constraint to the active set. In the latter case $\boldsymbol{x}^{(i)}$ is not a minimum of the current subspace. For the alternative computation of the search direction described in Sect. 5.2.1.2, it is essential that the initial point is a minimum of the current subspace and that all active constraints are linearly independent (cf. Wong, 2011, p. 19–20). Therefore, before starting the actual algorithm a subspace minimizer has to be computed.

One option to compute such a minimizer and the corresponding active set is to start the "original" active-set method (Alg. 2) at a feasible, but not necessarily optimal point and proceed as long as new constraints become active in each iteration. For the computation of the search direction $\boldsymbol{p}$ the Gauss-Jordan algorithm can be used to determine a particular solution of the system of equations (5.1). Determining a step length is not an issue here, as each step is "blocked" by a constraint. There exist much more sophisticated and efficient methods to compute a subspace minimizer and the corresponding active set. However, as they are not in the focus of the current work, the reader is referred to Wong (2011, p. 57–62).

5.2.1.2 Determination of the Search Direction p for Rank-Deficient Systems

As stated in Sect. 3.5.1.4, the desired search direction $\boldsymbol{p}$ should have the property that all active constraints

$$\boldsymbol{W}^T\boldsymbol{x} = \boldsymbol{w}$$

are kept active after the update step

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + q\boldsymbol{p}.$$

To determine such a search direction $\boldsymbol{p}$, the Search Direction Subproblem

Search direction Subproblem	
objective function:	$\Phi(\boldsymbol{p}) = \frac{1}{2}(\boldsymbol{x}+\boldsymbol{p})^T\boldsymbol{C}(\boldsymbol{x}+\boldsymbol{p}) + \boldsymbol{c}^T(\boldsymbol{x}+\boldsymbol{p}) \ldots \mathrm{Min}$
constraints:	$\boldsymbol{W}^T\boldsymbol{p} = \boldsymbol{0}$
optim. variable:	$\boldsymbol{p} \in \mathbb{R}^m$

(cf. problem (3.45)) has to be solved. As described in Sect. 3.5.1.4, this subproblem directly leads to the system of linear equations

$$\underbrace{\begin{bmatrix} \boldsymbol{C} & \boldsymbol{W} \\ \boldsymbol{W}^T & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{H}} \underbrace{\begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{k}_w \end{bmatrix}}_{\boldsymbol{z}} = \begin{bmatrix} -\boldsymbol{g} \\ \boldsymbol{0} \end{bmatrix} \tag{5.1}$$

(identical to (3.50)). $\boldsymbol{k}_w$ are the Lagrange multipliers associated with the active constraints. In some cases this system is singular or even inconsistent. In case of a singular but consistent system, a particular solution can be computed via the Gauss-Jordan algorithm and used as search direction.

An inconsistent system can for example appear if some entries of the main diagonal of $\boldsymbol{C}$ are zero but the corresponding elements in the gradient $\boldsymbol{g}$ are not. Wong (2011, p. 13) states that in this case the search direction subproblem (3.45) is unbounded below and proposes to determine a "search direction of zero curvature" instead. Her approach is stated here in a slightly modified version. The search direction $\boldsymbol{p}$ should have the following three properties:

1. The search direction should not be ascending. Thus, the scalar product with the gradient has to be non-positive

$$\boldsymbol{g}^T\boldsymbol{p} \leq 0. \tag{5.2a}$$

2. The curvature of the objective function in search direction $\boldsymbol{p}$ should be zero

$$\boldsymbol{p}^T\boldsymbol{C}\boldsymbol{p} = 0. \tag{5.2b}$$

3. All active constraints should stay active after the update

$$\boldsymbol{W}^T\boldsymbol{p} = \boldsymbol{0}. \tag{5.2c}$$

To compute such a search direction, a homogeneous solution $\boldsymbol{Z}_{\mathrm{hom}}$ of (5.1) is computed using the Gauss-Jordan algorithm (Wong, 2011, p. 13). If the rank defect is larger than one, $\boldsymbol{Z}_{\mathrm{hom}}$ is a matrix and each of its columns fulfills

$$\boldsymbol{H}\boldsymbol{Z}_{\mathrm{hom}}(:,i) = \boldsymbol{0}. \tag{5.3}$$

Therefore, an arbitrary column j of the homogeneous solution can be selected to compute the search direction

$$\boldsymbol{z}_{\mathrm{hom},j} := \begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{k}_w \end{bmatrix} = \boldsymbol{Z}_{\mathrm{hom}}(:,j). \tag{5.4}$$

Then the scalar product $\boldsymbol{g}^T\boldsymbol{p}$ is evaluated. If it is positive, the vector $\boldsymbol{z}_{\mathrm{hom},j}$ is multiplied with -1. As a homogeneous solution can be arbitrarily scaled, the vector is still a basis of the nullspace of $\boldsymbol{H}$. Thus, a step in this "mirrored" direction does not influence the set of active constraints. Therefore,

the basic properties of the algorithm are not violated and a viable and binding search direction is found (Wong, 2011, p. 13). However, it is essential for this computation that the initial solution is a subspace minimizer (cf. Sect. 5.2.1.1) to prevent cycling of the algorithm.

Furthermore, as the homogeneous solution is not unique, it is preferable to compute the values of the Lagrange multipliers $\boldsymbol{k}_w$ of the active constraints independently. For example, by solving (Wong, 2011, p. 10)

$$\boldsymbol{W}\boldsymbol{k}_w = -\boldsymbol{g}. \tag{5.5}$$

As a consequence, in the function `computeSearchDirectionRankDefect` in line 4 of Alg. 4, three different cases have to be distinguished

1. If (5.1) is consistent and $\boldsymbol{H}$ is well-defined, the unique solution of the system is determined.
2. If (5.1) is consistent but $\boldsymbol{H}$ has a rank defect, a particular solution is computed.
3. If (5.1) is inconsistent, one column of the homogeneous solution is selected.

5.2.1.3 Determination and Handling of the Step Length q for Rank-Deficient Systems

In the non-singular case, due to the construction of the search direction $\boldsymbol{p}$, the optimal step length is $q = 1$ and it was only necessary to check, if a step of full length would lead to the violation of an inactive constraint (cf. Sect. 3.5.1.5). However, this is not sufficient for singular systems, as $\boldsymbol{p}$ might be a line of zero curvature. As a consequence, there might be no optimal step length and different cases have to be distinguished depending on the curvature of the objective function in search direction (cf. Wong, 2011, 13–14) and the presence of blocking constraints.

The following four cases have to be distinguished:

1. $\boldsymbol{p}$ is a line of positive curvature and no inactive constraint is blocking a step of full length. In this case $\boldsymbol{p}$ is the projected Newton direction and thus correctly scaled by design (cf. Sect. 3.5.1.4). This yields $q = 1$ and the new solution will be a minimizer of the current subspace.
2. $\boldsymbol{p}$ is a line of positive curvature and there is a blocking constraint. As stated in (3.54b) the maximal feasible step length

   $$q = \min\left(q_j^{(i)}\right) = \min\left(\frac{v_j - \boldsymbol{V}(:,j)^T\boldsymbol{x}^{(i)}}{\boldsymbol{V}(:,j)^T\boldsymbol{p}^{(i)}}\right) \tag{5.6}$$

   is chosen and a new constraint will become active after the update step.
3. $\boldsymbol{p}$ is a line of zero curvature and there is no inactive constraint in search direction. In this case, no other constraint will become active, no matter how the step length is chosen. Thus, the objective function is unbounded below and the algorithm terminates without a solution (Wong, 2011, p. 15).
4. $\boldsymbol{p}$ is a line of zero curvature and there is a blocking constraint. As a constraint prevents that the problem is unbounded below, the maximal feasible step length (5.6) is computed and a new constraint will become active after the update step.

This distinction of cases is formalized in the function `computeStepLengthRankDefect` in line 6 of Alg. 4.

Algorithm 4: Binding-Direction Primal Active-Set Algorithm for Rank-Deficient Systems

// Iterative algorithm to solve a (possibly rank-deficient) quadratic program in standard form (3.4)

Data:

$\boldsymbol{C}_{[m\times m]}, \boldsymbol{c}_{[m\times 1]}$	...	Matrix and vector with the coefficients of the objective function
$\boldsymbol{B}_{[m\times p]}, \boldsymbol{b}_{[p\times 1]}$	...	Matrix and corresponding right-hand side of the inequality constraints
$\overline{\boldsymbol{B}}_{[m\times \bar{p}]}, \overline{\boldsymbol{b}}_{[\bar{p}\times 1]}$	...	Matrix and corresponding right-hand side of the equality constraints
$\boldsymbol{x}^{(0)}_{[m\times 1]}$	...	feasible initial solution vector

Result:

$\boldsymbol{x}_{[m\times 1]}$	...	Vector containing the solution of the QP
$\boldsymbol{k}_{w[p_w\times 1]}$	...	Vector containing the Lagrange multipliers of the active constraints

```
1  [W, w, V, v] = findActiveConstraints(B, b, B̄, b̄, x)
2  [x, W, w, V, v] = computeSubspaceMinimizer(C, c, W, w, V, v, x)        // cf. Sect. 5.2.1.1
3  for i = 1 : i_max do
4  |   [p, k_w] = computeSearchDirectionRankDefect(C, c, W, x)           // cf. Sect. 5.2.1.2
5  |   if i > 1 then
6  |   |   q = computeStepLengthRankDefect(H, V, v, x, p)               // cf. Sect. 5.2.1.3
7  |   |   x = x + qp
8  |   else
9  |   |   q = 1
10 |   end
11 |   [W, w, V, v] = updateActiveSet(W, w, V, v, x, q)                 // cf. Sect. 3.5.1.6
12 |   if min(k_w ≥ 0) then
13 |   |   break
14 |   end
15 end
16 return x, k_w
```

5.2.2 Binding-Direction Primal Active-Set Method for Rank-Deficient Systems

The ideas described above are now combined to a modified version of the binding-direction primal active-set method described in Sect. 3.5.1.3. This algorithm is capable of finding a particular solution of a QP in standard form (3.4) despite a possible rank-deficient matrix $\boldsymbol{C}$ and stated in Alg. 4.

Starting at a feasible point $\boldsymbol{x}^{(0)}$ an initial active set can be obtained by evaluating the constraints at $\boldsymbol{x}^{(0)}$ (line 1 of Alg. 4). Afterwards a solution $\boldsymbol{x}$ that minimizes the current subspace is computed and a corresponding active set which includes only linearly independent constraints is obtained (line 2 and Sect. 5.2.1.1). Subsequently, a search direction $\boldsymbol{p}$ and the Lagrange multipliers of active constraints $\boldsymbol{k}_w$ are computed (line 4 and Sect. 5.2.1.2). The next steps are the computation of a step length q (line 6 and Sect. 5.2.1.3) and of an update of the parameters (line 7). As the algorithm starts with a minimizer of the current subspace, these two steps are omitted in the first iteration as it is not possible to reduce the value of the objective function without deactivating a constraint. Subsequently, as in the original version of the algorithm, a decision has to be made, which constraints

stay in the active set and which are dropped (line 11). These steps are repeated iteratively until all Lagrange multipliers associated with inequalities are non-negative.

It should be noted that in the general framework this algorithm is solely used if manifold and feasible region are disjunct (cf. Sect. 5.3.1.2).

5.3 Rigorous Computation of a General Solution

Before describing the actual framework, it is instructive to examine how the introduction of inequalities can change the original problem and especially the manifold.

5.3.1 Changes Through Inequality Constraints

We will distinguish two main cases depending on whether there is an intersection of the feasible region and the manifold of solutions (*case 1*) or not (*case 2*). As it is crucial for understanding the theory our proposed framework is built on, the differences to the unconstrained case (cf. Sect. 2.1.3.2) will be described in some detail.

Figure 5.1(a) shows the *isolines* of the objective function of the unconstrained, bivariate optimization problem stated in Sect. 5.5.1.1. One particular solution is computed using the Moore-Penrose inverse (2.18) and depicted as *orange cross*. The one-dimensional manifold of solutions is shown as a *dashed black line*.

5.3.1.1 Case 1: Intersection of Manifold and Feasible Region

This case can be subdivided into three sub-cases. Let *case 1a* be defined as a problem with only one or more inactive inequality constraints that are parallel to the solution manifold. As the unconstrained solution already is in the feasible region, the parallel constraint(s) do not influence the solution at all.

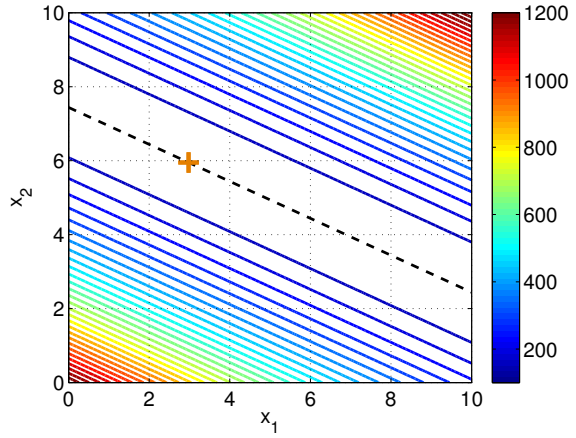
Case 1b, another possible influence of constraints, is shown in Fig. 5.1(b). Here, the constraint is not parallel to the manifold and therefore restricts it. The manifold is still one-dimensional. However, due to the constraint, it is no longer a line but a half-line, since it is now bounded in one direction but still unbounded in the other one. Introduction of more constraints can further restrict the manifold to a line segment.

The introduction of constraints can also lead to a decrease of the dimension of the manifold (*case 1c*). This is always the case for equality constraints which are not parallel to the manifold of solutions. However, the same can be true for inequality constraints. An easy to imagine example is the case where one equality constraint is expressed by two inequality constraints.

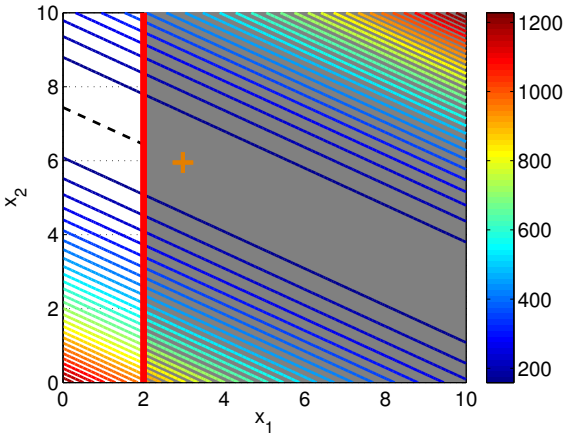
5.3.1.2 Case 2: Manifold and Feasible Region are Disjunct

In some cases, the introduction of inequality constraints enables us to determine a unique solution of a rank-deficient optimization problem (*case 2a*, *green cross* in Fig. 5.1(c)). This is always the case, if the solution of the constrained problem is not contained in the original manifold and if there are no active parallel constraints.

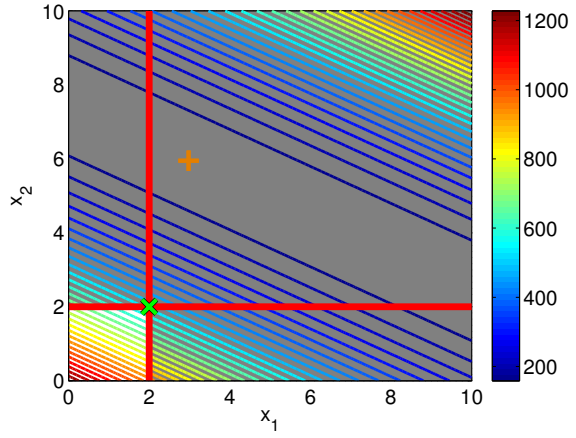
In *case 2b*, the influence of a constraint that is parallel to the manifold of solutions is further examined (cf. Fig. 5.1(d)). It is obvious that the active inequality constraint (*red line*) shifts the one-dimensional manifold but does not constrain it further. Therefore, there is still a manifold of the same dimension as in the original problem.



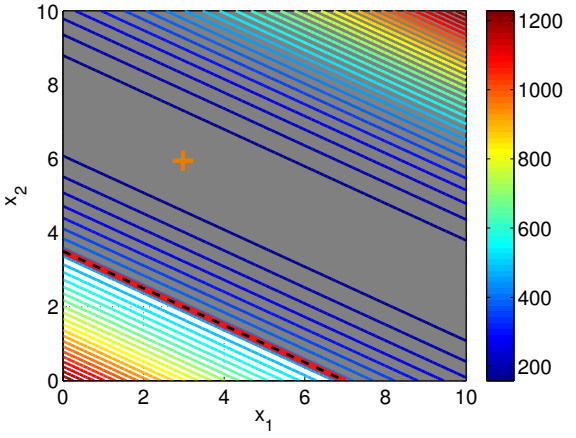
(a) Isolines of the objective function



(b) *Case 1b*: Manifold and feasible region intersect and the manifold is restricted by the constraint



(c) *Case 2a*: Manifold and feasible region are disjunct, unique solution



(d) *Case 2b*: Manifold and feasible region are disjunct, but there is still a manifold of solutions due to the active parallel constraint

Figure 5.1: *Isolines* of the objective function of the bivariate convex optimization problem, which is stated in Sect. 5.5.1. The one-dimensional manifold (i.e., a straight line) is plotted as *dashed black line*. A particular solution of the unconstrained case was computed via the Moore-Penrose inverse and is shown as *orange cross*. The infeasible region is *shaded* and inequalities are plotted as *red lines*. In (c) the unique ICLS solution is depicted as *green cross*. Figure taken from Roese-Koerner and Schuh (2014).

5.3.1.3 General Remarks and Outline of the Framework

It is important to notice that the direction of the manifold cannot be changed through the introduction of inequality constraints. More specifically, a translation (*case 2b*), a general restriction (*case 1b*) or a dimension reduction (*case 1c* and *2a*) of the manifold are possible, but never a rotation. This leaves us in the comfortable situation that it is possible to determine the homogeneous solution of an ICLS problem by determining the homogeneous solution of the corresponding unconstrained WLS problem and reformulate the constraints in relation to this manifold. Therefore, our framework consists of the following major parts that will be explained in detail in the next sections:

To compute a general solution of an ICLS problem (3.8), we compute a general solution of the unconstrained WLS problem and perform a change of variables to reformulate the constraints in terms of the free variables of the homogeneous solution. Next, we determine if there is an intersection between the manifold of solutions and the feasible region. In case of an intersection, we determine the shortest solution vector in the nullspace of the design matrix with respect to the inequality constraints and reformulate the homogeneous solution and the inequalities accordingly. If there is no intersection, we use the modified active-set method described in Sect. 5.2.2 to compute a particular solution and determine the uniqueness of the solution by checking for active parallel constraints.

It should be noted that any QP solver that is capable of handling singular systems can be used to determine this particular solution. In earlier publications (Roese-Koerner and Schuh, 2014, 2015) Dantzig's Simplex Algorithm for QPs was used for this purpose. However, we found that the modified active-set method is less memory consuming than Dantzig's algorithm (at least in its basic formulation) and therefore replaced the latter.

5.3.2 Transformation of Parameters

In order to solve the ICLS problem (3.8), as a first step, we compute a general solution of the unconstrained WLS problem (2.32) as described in Sect. 2.1.3.2

$$\widetilde{\boldsymbol{x}}^{\mathrm{WLS}}(\boldsymbol{\lambda}) = \boldsymbol{x}_{\mathrm{P}}^{\mathrm{WLS}} + \boldsymbol{x}_{\mathrm{hom}}(\boldsymbol{\lambda}) = \boldsymbol{x}_{\mathrm{P}}^{\mathrm{WLS}} + \boldsymbol{X}_{\mathrm{hom}}\,\boldsymbol{\lambda}. \tag{5.7}$$

Afterwards, we perform a change of variables and insert (5.7) and (2.13) into the inequality constraints

$$\boldsymbol{B}^T\boldsymbol{x} \leq \boldsymbol{b}$$

to reformulate them in terms of the free variables $\boldsymbol{\lambda}$ at the point of the particular solution

$$\boldsymbol{B}^T\left(\boldsymbol{x}_{\mathrm{P}}^{\mathrm{WLS}} + \boldsymbol{X}_{\mathrm{hom}}\,\boldsymbol{\lambda}\right) \leq \boldsymbol{b}, \tag{5.8a}$$

leading to

$$\boldsymbol{B}^T\boldsymbol{x}_{\mathrm{P}}^{\mathrm{WLS}} + \boldsymbol{B}^T\boldsymbol{X}_{\mathrm{hom}}\,\boldsymbol{\lambda} \leq \boldsymbol{b}. \tag{5.8b}$$

With the substitutions

$$\boldsymbol{B}_{\boldsymbol{\lambda}}^T := \boldsymbol{B}^T\boldsymbol{X}_{\mathrm{hom}}, \tag{5.9}$$

$$\boldsymbol{b}_{\boldsymbol{\lambda}} := \boldsymbol{b} - \boldsymbol{B}^T\boldsymbol{x}_{\mathrm{P}}^{\mathrm{WLS}}, \tag{5.10}$$

(5.8b) reads

$$\boldsymbol{B}_{\boldsymbol{\lambda}}^T\boldsymbol{\lambda} \leq \boldsymbol{b}_{\boldsymbol{\lambda}}, \tag{5.11}$$

being the desired formulation of the inequality constraints. If the extended matrix $[\boldsymbol{B}_{\boldsymbol{\lambda}}^T|\boldsymbol{b}_{\boldsymbol{\lambda}}]$ has rows that contain only zeros, these rows can be deleted from the equation system, as they belong to inactive inequality constraints which are parallel to the solution manifold but do not shift the optimal solution (cf. *case 1a* in Sect. 5.3.1.1).

For some applications, it is possible to directly state a basis $\boldsymbol{X}_{\text{hom}}$ for the nullspace of $\boldsymbol{A}$. This offers the advantage that the free parameters $\boldsymbol{\lambda}$ are interpretable. For example in the adjustment of a geodetic direction and distance network, the free parameters represent translation, rotation and scaling of the network and an explicite basis for the nullspace is known (cf. Meissl, 1969).

For an horizontal network, where there is a known basis for the nullspace of the design matrix, Xu (1997) worked out constraints (such as a positive scaling parameter) for the free parameters to obtain a *geodetically meaningful* solution. These constraints can easily be included in our framework by using the known basis of the nullspace as homogeneous solution $\boldsymbol{X}_{\text{hom}}$ and expanding the constraints $\boldsymbol{B}_{\boldsymbol{\lambda}}^T$ and $\boldsymbol{b}_{\boldsymbol{\lambda}}$ to the free parameters.

Next, we examine if the constraints (5.11) form a feasible set or if they are inconsistent. This can be done by formulating and solving a feasibility problem (cf. Boyd and Vandenberghe, 2004, p. 579–580). If the constraints form a feasible set, there is at least one set of free parameters λ_i that fulfills all constraints. This is equivalent to the statement that there is an intersection between the manifold of solutions and the feasible region of the original ICLS problem which is described in Sect. 5.3.3. If the constraints are inconsistent, the manifold and the feasible region are disjoint and we will proceed as described in Sect. 5.3.4.

5.3.3 Case 1: Intersection of Manifold and Feasible Region

If there is an intersection of manifold and feasible region, we aim for the determination of a unique particular solution $\widetilde{\boldsymbol{x}}_{\text{P}}$ that fulfills certain predefined optimality criteria. Therefore, a second optimization problem is introduced (e.g., minimizing the length of the solution vector with respect to the norm L_{P}). As this minimization takes place in the nullspace of the design matrix $\boldsymbol{A}$, the value of the objective function of the original problem (3.8) does not change

Nullspace optimization problem	
objective function:	$\Phi_{\text{NS}}\left(\boldsymbol{x}_{\text{P}}^{\text{ICLS}}(\boldsymbol{\lambda})\right) \ldots \text{Min}$
constraints:	$\boldsymbol{B}_{\boldsymbol{\lambda}}^T\boldsymbol{\lambda} \leq \boldsymbol{b}_{\boldsymbol{\lambda}}$
optim. variable:	$\boldsymbol{\lambda} \in \mathbb{R}^d$.

(5.12)

The minimization yields optimal free parameters $\widetilde{\boldsymbol{\lambda}}$, whose insertion in (5.7) results in

$$\widetilde{\boldsymbol{x}}_{\text{P}}^{\text{ICLS}} = \boldsymbol{x}_{\text{P}}^{\text{WLS}} + \boldsymbol{X}_{\text{hom}}\widetilde{\boldsymbol{\lambda}}, \tag{5.13}$$

which is a unique particular solution that fulfills all constraints. As the reformulated constraints depend on the chosen particular solution, we have to adapt them to the new particular solution using (5.8). Now, we can combine our results to a rigorous general solution of the ICLS problem (3.8)

$$\widetilde{\boldsymbol{x}}^{\text{ICLS}}(\boldsymbol{\lambda}) = \widetilde{\boldsymbol{x}}_{\text{P}}^{\text{ICLS}} + \boldsymbol{x}_{\text{hom}}(\boldsymbol{\lambda}), \tag{5.14a}$$

$$\boldsymbol{B}_{\boldsymbol{\lambda}}^T\boldsymbol{\lambda} \leq \boldsymbol{b}_{\boldsymbol{\lambda}}. \tag{5.14b}$$

In the following, the influence of the use of different norms and functional relationships of the objective function Φ_{NS} of the nullspace optimization problem (5.12) is discussed.

5.3.3.1 Nullspace Optimization

As stated above, the whole minimization takes place in the nullspace of the design matrix and the value of the original objective function (3.8) will not change. Nonetheless, choosing a suitable second objective function for a particular problem can be helpful to achieve properties such as sparsity of the parameter vector. If this second objective function is chosen in a way that the nullspace optimization problem becomes a QP, it can for example be solved using the active-set method described in Sect. 3.5.1. In case the problem is an LP or a general non-linear problem, specialized solvers (e.g., CVX, see Grant and Boyd, 2014) have to be used.

Two different parts of the objective function can be distinguished, which will be examined in some detail: the *functional relationship* and the *norm*, with respect to which the minimization (or maximization) shall be performed.

Functional Relationship. The functional relationship strongly depends on the application. However, there are two main concepts which are applicable to a large variety of problems.

First, the length of the parameter vector $\boldsymbol{x}$ can be minimized. This can for example be beneficial if not absolute coordinates but coordinate differences are estimated, which should be close to the initial coordinates. In a Bayesian interpretation this could be seen as introduction of isotropic prior information on the parameters.
More sophisticated approaches include a weighted minimization of the length of the parameter vector. For example if prior knowledge about the magnitude of the parameters is given. Prominent examples are Kaula's rule of thumb in gravity field estimation or the demand for a decay of the amplitudes of higher frequencies in signal processing to achieve a square integrable function. A mathematical example for the minimization of the length of the parameter vector is provided in Sect. 5.5.1.2.

The other main concept is to maximize the distance of the solution to the constraints

$$||\boldsymbol{B}^T\boldsymbol{x}(\boldsymbol{\lambda}) - \boldsymbol{b}|| \ldots \max, \tag{5.15a}$$

thus

$$||\boldsymbol{B}^T\boldsymbol{x}_\mathrm{P} + \boldsymbol{B}^T\boldsymbol{X}_\mathrm{hom}\boldsymbol{\lambda} - \boldsymbol{b}|| \ldots \max. \tag{5.15b}$$

This could be beneficial, if the constraints constitute a kind of outermost threshold. In Sect. 7.3.2 an engineering application including welding tolerances is described, in which this type of *functional relationship* is applied. One of the main advantages of the use of inequality constraints is that they do not influence the result, if they are not active. Therefore, it it usually not possible to provide a buffer to the boundary of the feasible set without losing estimation quality (shown by an increased value of the original objective function). However, due to the optimization in the nullspace of $\boldsymbol{A}$, we are in the unique position to apply such a buffer without deteriorating the estimate.

Norms. In the following, we will point out some general aspects of different norms and their influence on the most classical functional relationship: the length of the solution vector. L_p norms used in adjustment theory include the L_1 norm, the L_2 norm and the L_∞ norm (cf. Sect. 2.3 and Jäger et al., 2005, Boyd and Vandenberghe, 2004, p. 125–128 and p. 635, respectively).

Most often, the length of a vector is minimized with respect to the L_2 norm (also known as Euclidean norm), which results in the nullspace objective function

$$\Phi_\mathrm{NS}^{L2}\left(\boldsymbol{x}_\mathrm{P}^\mathrm{ICLS}(\boldsymbol{\lambda})\right) = ||\boldsymbol{x}_\mathrm{P}^\mathrm{ICLS}(\boldsymbol{\lambda})||_2 \tag{5.16a}$$

$$= \left(\boldsymbol{x}_\mathrm{P}^\mathrm{ICLS}(\boldsymbol{\lambda})\right)^T \boldsymbol{x}_\mathrm{P}^\mathrm{ICLS}(\boldsymbol{\lambda}). \tag{5.16b}$$

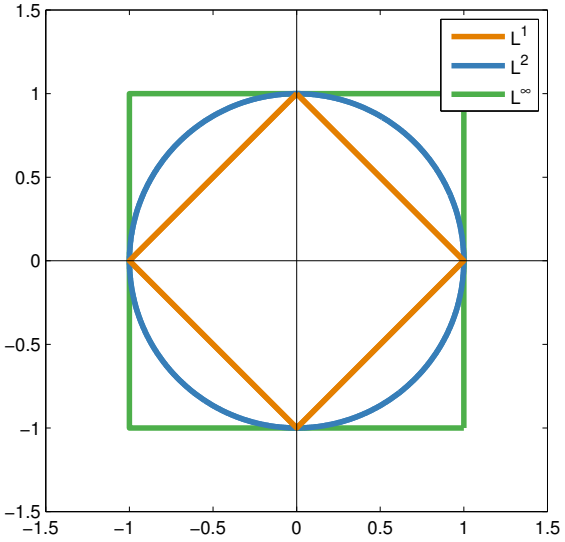


Figure 5.2: Two-dimensional *unit spheres* of L_1 (*orange*), L_2 (*blue*) and L_∞ norm (*green*). The L_1 diamond evolves from the summation of x and y value, while for the L_∞ square, only the value of the largest quantity is decisive. Figure taken from Roese-Koerner and Schuh (2015).

Inserting (5.7) yields

$$\Phi_{\mathrm{NS}}^{L2}(\boldsymbol{\lambda}) = \left(\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}} + \boldsymbol{X}_{\mathrm{hom}}\,\boldsymbol{\lambda}\right)^T \left(\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}} + \boldsymbol{X}_{\mathrm{hom}}\,\boldsymbol{\lambda}\right) \tag{5.16c}$$

$$= \boldsymbol{\lambda}^T \boldsymbol{X}_{\mathrm{hom}}^T \boldsymbol{X}_{\mathrm{hom}} \boldsymbol{\lambda} + 2\left(\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}}\right)^T \boldsymbol{X}_{\mathrm{hom}} \boldsymbol{\lambda} + \left(\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}}\right)^T \boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}} \tag{5.16d}$$

and the following quadratic program:

L_2 NORM NULLSPACE OPTIMIZATION PROBLEM

objective function:	$\Phi_{\mathrm{NS}}^{L2}(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^T \boldsymbol{X}_{\mathrm{hom}}^T \boldsymbol{X}_{\mathrm{hom}} \boldsymbol{\lambda} + 2\left(\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}}\right)^T \boldsymbol{X}_{\mathrm{hom}} \boldsymbol{\lambda} \ldots \mathrm{Min}$
constraints:	$\boldsymbol{B}_{\boldsymbol{\lambda}}^T \boldsymbol{\lambda} \leq \boldsymbol{b}_{\boldsymbol{\lambda}}$
optim. variable:	$\boldsymbol{\lambda} \in \mathbb{R}^d$.

(5.17)

The last term of objective function (5.16) is constant and was neglected as it does not influence the minimization result. In case of solely inactive constraints, the resulting solution will be equivalent to the one obtained via the Moore-Penrose inverse in the unconstrained case. If the constraints prevent that the optimal unconstrained solution is reached, our particular solution will be the one with shortest length among all solutions that minimize the objective function and fulfill the constraints. The procedure could therefore be seen as a kind of pseudoinverse that takes inequality constraints into account.

Using the L_2 norm is a quite natural choice which can easily be visualized geometrically. However, as the influence of one element on the norm decreases if its absolute value becomes smaller, a minimization with respect to the L_2 norm will seldom result in a sparse vector. This can be verified by examining the *blue* L_2 norm unit sphere depicted in Fig. 5.2.

The naive choice to achieve maximal sparsity of a vector would be a minimization with respect to the L_0 norm (e.g., the number of nonzero elements). However, a minimization with respect to the L_0 norm is a combinatorical problem and computationally very demanding. Therefore, e.g., Candes

et al. (2006) approximated the L_0-minimization problem by the L_1-minimization problem in the context of compressed sensing. They showed that a minimization with respect to the L_1 norm in most cases yields sparse results, too. This is used in many compressed sensing algorithms. Minimization with respect to the L_1 norm is a convex optimization problem and can be formulated as a linear program (cf. Dantzig, 1998, p. 60–62). As can be seen in Fig. 5.2, the corners of the *orange* L_1 norm unit sphere coincide with the coordinate grid. This is equivalent to the statement that one parameter is zero there, yielding a sparse solution.

Application of the L_∞ norm (also known as Chebyshev norm) results in a parameter vector with minimal maximal value that is usually not sparse. See Fig. 5.2 for the corresponding unit sphere. The L_∞ norm is often applied, when trying to maximize the distance to the constraints, in order to maximize the minimal buffer to the boundary. However, it should be noted that L_∞ norm minimization is an intricate task as there is often no analytical solution (cf. Grant et al., 2006). Within this thesis we used the CVX software (Grant and Boyd, 2014) whenever an L_∞ norm minimization is performed.

5.3.4 Case 2: Manifold and Feasible Region are Disjunct

If there is no intersection between the feasible region and the manifold of solutions, there either is a unique solution or at least one active constraint is parallel to the solution manifold meaning the solution is still non-unique (cf. Sect. 5.3.1.2). To compute a particular solution of problem (3.8) the modified binding-direction primal active-set method discussed in Sect. 5.2.2 can be used, resulting in an arbitrary optimal solution $\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}}$. To decide whether the solution is unique, we check for active parallel constraints. If there is at least one, there exists a manifold of solutions, which is a shifted version of the original one. In this case, we proceed as described in Sect. 5.3.3 using $\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}}$ instead of $\boldsymbol{x}_{\mathrm{P}}^{\mathrm{WLS}}$.

If there is a unique solution, it means that the introduction of constraints has resolved the manifold yielding

$$\widetilde{\boldsymbol{x}}^{\mathrm{ICLS}} = \boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}} \tag{5.18}$$

as the final result.

5.4 A Framework for the Solution of Rank-Deficient ICLS Problems

Using the tools described, we devised a framework for the computation of a rigorous general solution of rank-deficient ICLS problems, which is shown in Alg. 5.

The first three lines correspond to the transformation of parameters and the test for intersection of feasible region and solution manifold (described in Sect. 5.3.2). If there is no intersection (*case 2*, *feasible* = *false*) a particular solution is computed as described in Sect. 5.3.4 and a test for parallel constraints is performed (lines 4–12). If no active constraint is parallel to the manifold, the computed particular solution is unique and the algorithm terminates (line 8).

If there is an intersection of the sets (*case 1*, *feasible* = *true*) or an active parallel constraint, we proceed as described in Sect. 5.3.3 and solve a second optimization problem (lines 13–17).

In lines 2, 11 and 17 a transformation of the constraints is computed. This is necessary, as the constraints with respect to the Lagrange multipliers k_i depend on the chosen particular solution. Fortunately, this transformation is a computationally cheap operation.

Algorithm 5: A framework to estimate the general solution of ICLS problems with rank defect

1 $[\boldsymbol{x}_\mathrm{P}, \boldsymbol{x}_\mathrm{hom}] = \mathrm{GeneralWlsSolution}(\boldsymbol{A}, \boldsymbol{\ell}, \boldsymbol{\Sigma})$ // *cf. Sect. 5.3.2*
2 $[\boldsymbol{B}_{\boldsymbol{\lambda}}, \boldsymbol{b}_{\boldsymbol{\lambda}}] = \mathrm{transformConstraints}(\boldsymbol{B}, \boldsymbol{b}, \boldsymbol{x}_\mathrm{P}, \boldsymbol{x}_\mathrm{hom})$ // *cf. Sect. 5.3.2*
3 $feasible = \mathrm{testFeasibility}(\boldsymbol{B}_{\boldsymbol{\lambda}}, \boldsymbol{b}_{\boldsymbol{\lambda}})$
4 **if** $feasible == false$ **then**
5 | *// Case 2*
6 | $[\boldsymbol{x}_\mathrm{P}, \tilde{\boldsymbol{k}}] = \mathrm{computeParticularSolution}(\boldsymbol{A}, \boldsymbol{\ell}, \boldsymbol{\Sigma}, \boldsymbol{B}, \boldsymbol{b})$ // *cf. Sect. 5.3.4*
7 | **if** $activeParallelConstraint == false$ **then**
8 | | $\widetilde{\boldsymbol{x}}_\mathrm{P}^\mathrm{ICLS} = \boldsymbol{x}_\mathrm{P}, \boldsymbol{x}_\mathrm{hom} = \emptyset, \boldsymbol{B}_{\boldsymbol{\lambda}} = \emptyset, \boldsymbol{b}_{\boldsymbol{\lambda}} = \emptyset$
9 | | **return** $(\widetilde{\boldsymbol{x}}_\mathrm{P}^\mathrm{ICLS}, \boldsymbol{x}_\mathrm{hom}, \boldsymbol{B}_{\boldsymbol{\lambda}}, \boldsymbol{b}_{\boldsymbol{\lambda}})$
10 | **else**
11 | | $[\boldsymbol{B}_{\boldsymbol{\lambda}}, \boldsymbol{b}_{\boldsymbol{\lambda}}] = \mathrm{transformConstraints}(\boldsymbol{B}, \boldsymbol{b}, \boldsymbol{x}_\mathrm{P}, \boldsymbol{x}_\mathrm{hom})$ // *cf. Sect. 5.3.2*
12 | **end**
13 **end**
14 *// Case 1 or at least one active parallel constraint*
15 $\widetilde{\boldsymbol{\lambda}} = \mathrm{solveNullspaceOptimizationProblem}(\boldsymbol{x}_\mathrm{P}, \boldsymbol{x}_\mathrm{hom}, \boldsymbol{B}_{\boldsymbol{\lambda}}, \boldsymbol{b}_{\boldsymbol{\lambda}})$ // *cf. Sect. 5.3.3*
16 $\widetilde{\boldsymbol{x}}_\mathrm{P}^\mathrm{ICLS} = \boldsymbol{x}_\mathrm{P} + \boldsymbol{x}_\mathrm{hom}(\widetilde{\boldsymbol{\lambda}})$
17 $[\boldsymbol{B}_{\boldsymbol{\lambda}}, \boldsymbol{b}_{\boldsymbol{\lambda}}] = \mathrm{transformConstraints}(\boldsymbol{B}, \boldsymbol{b}, \widetilde{\boldsymbol{x}}_\mathrm{P}^\mathrm{ICLS}, \boldsymbol{x}_\mathrm{hom})$ // *cf. Sect. 5.3.2*
18 **return** $(\widetilde{\boldsymbol{x}}_\mathrm{P}^\mathrm{ICLS}, \boldsymbol{x}_\mathrm{hom}, \boldsymbol{B}_{\boldsymbol{\lambda}}, \boldsymbol{b}_{\boldsymbol{\lambda}})$

One may ask the question: Why not directly compute a particular ICLS solution as performed in line 6 of the framework? We intentionally decided to first compute an unconstrained general solution and to check for set intersection in order to achieve optimal runtime behavior. This is because solving the original quadratic program is the most expensive operation performed within this framework. If feasible set and manifold intersect, we can avoid solving the original problem directly. Instead a general unconstrained solution is computed and an optimization problem with respect to the free parameters λ_i is solved. Computationally this is a lot cheaper, as the dimension reduces from an m-dimensional problem to a d-dimensional problem. The number m of parameters is usually much greater than the dimension d of the rank defect. Hence, computations become faster.

It should be noted that it is not possible to apply the MC-QP method described in Chap. 4 to a rank-deficient problem. This is for two reasons: First, the computation of **a** particular solution via the active-set method is somewhat arbitrary, which does not go well with an algorithm using a Monte Carlo approach. Second, the concept of perturbing the observations as input for each Monte Carlo step may not effect the Nullspace optimization procedure and can therefore lead to an overoptimistic stochastic description.

5.5 Example

The presented framework for the rigorous computation of a general solution of an ICLS problem was applied to a small two-dimensional synthetic example. We chose this setup because of its simplicity which allows to concentrate on details of the presented framework and exploit the fact that it is possible to explicitly draw the objective function of a two-dimensional problem.

Two more applications—the task of estimating optimal repetition numbers in an underdetermined system, and a setup with strict welding tolerances—are used to underline the potential of the framework for classical geodetic applications. Both examples can be found in the applications chapter (Sect. 7.3.1 and Sect. 7.3.2, respectively).

5.5.1 2D Synthetic Example with 1D Manifold

The two summands of the weighted sum

$$\ell_i + v_i = x_1 + 2x_2 \tag{5.19}$$

shall be estimated in a Gauss-Markov model. We assume each observation to follow a normal distribution with a standard deviation of one and state that all observations are independent and identically distributed. Therefore, the VCV matrix $\boldsymbol{\Sigma}$ is an identity matrix. It should be noted that this is done for reasons of simplicity only, as the proposed framework is able to handle the case of correlated observations, too.

The resulting two-dimensional system clearly has a rank defect of one, as it is solely possible to estimate one summand depending on the other one

$$\boldsymbol{\ell} + \boldsymbol{v} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{A}\boldsymbol{x}. \tag{5.20}$$

Obviously, both columns of the design matrix $\boldsymbol{A}$ are linearly dependent. Given the following synthetic observations

$$\boldsymbol{\ell}^T = \begin{bmatrix} 23.2 & 16.4 & 12.9 & 8.2 & 13.7 \end{bmatrix}, \tag{5.21}$$

we will demonstrate the computation of the unconstrained OLS solution in Sect. 5.5.1.1, as this is identical to the first steps of our framework. Next, we will introduce two different sets of constraints to cover *case 1* where manifold and feasible region intersect (Sect. 5.5.1.2) as well as *case 2*, in which there is no such intersection (Sect. 5.5.1.3). The isolines of the objective function are shown in Fig. 5.1(a).

5.5.1.1 Unconstrained OLS Solution

The elements of the normal equations read

$$\boldsymbol{N} = \boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix}, \qquad \boldsymbol{n} = \boldsymbol{A}^T\boldsymbol{\ell} = \begin{bmatrix} 74.40 \\ 148.80 \end{bmatrix}. \tag{5.22}$$

Applying the Gauss-Jordan algorithm to solve

$$\boldsymbol{N}\boldsymbol{x} = \boldsymbol{n} \tag{5.23}$$

yields

$$\begin{array}{cc|c|l} 5 & 10 & 74.40 & |:5 \\ 10 & 20 & 148.80 & |-2(I) \\ \hline 1 & 2 & 14.88 & \\ 0 & 0 & 0. & \end{array} \tag{5.24}$$

Thus, inserting (2.17) and (2.15) in (2.11) the general solution reads

$$\widetilde{\boldsymbol{x}}^{\mathrm{OLS}}(\lambda) = \underbrace{\begin{bmatrix} 14.88 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_{\mathrm{P}}^{\mathrm{OLS}}} + \underbrace{\begin{bmatrix} -2 \\ 1 \end{bmatrix}}_{\boldsymbol{X}_{\mathrm{hom}}} \lambda = \boldsymbol{x}_{\mathrm{P}}^{\mathrm{OLS}} + \boldsymbol{X}_{\mathrm{hom}}\lambda. \tag{5.25}$$

As expected, there is no unique optimal solution but a manifold, which is expressed by an arbitrary particular solution $\boldsymbol{x}_{\mathrm{P}}^{\mathrm{OLS}}$ and a solution $\boldsymbol{X}_{\mathrm{hom}}$ of the homogeneous system. The manifold is represented by the dashed black line in Fig. 5.1(a) for the arbitrarily chosen interval $2.44 \leq \lambda \leq 7.44$.

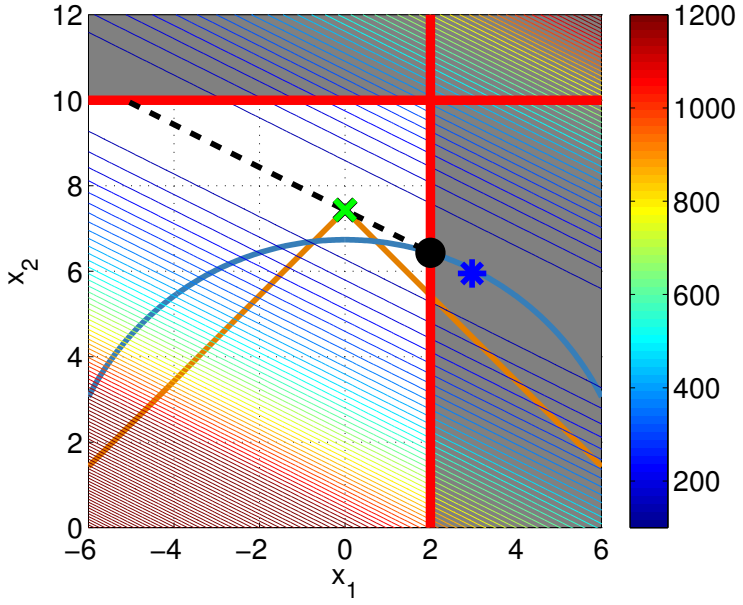


Figure 5.3: Contour lines of the objective function of case 1. *Red lines* represent constraints, the infeasible region is *shaded*. The *dashed black line* indicates the manifold of solutions. The *green cross* is the L_1 norm solution, the *black circle* the L_2 solution. For comparison the solution using a pseudoinverse (*blue star*) is shown, too. Appropriately scaled unit spheres are depicted in *orange* (L_1) and *light blue* (L_2). Figure modified from Roese-Koerner and Schuh (2015).

5.5.1.2 Case 1: Intersection of Manifold and Feasible Region

Introduction of the constraints

$$x_1 \leq 2, \quad x_2 \leq 10 \tag{5.26a}$$

and thus

$$\boldsymbol{B}^T\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{x} \leq \begin{bmatrix} 2 \\ 10 \end{bmatrix} = \boldsymbol{b} \tag{5.26b}$$

restricts the manifold as can be seen in Fig. 5.3. Transforming the constraints in the point $\boldsymbol{x}_{\mathrm{P}}^{\mathrm{OLS}}$ of the particular solution with respect to the free parameter λ according to (5.8) yields

$$-2\lambda \leq -12.88, \quad \lambda \leq 7, \tag{5.27a}$$

leading to

$$\boldsymbol{B}_\lambda^T\lambda = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \lambda \leq \begin{bmatrix} -12.88 \\ 7 \end{bmatrix} = \boldsymbol{b}_\lambda. \tag{5.27b}$$

Now, a feasibility problem is solved to determine if a feasible solution exists or if the constraints are contradictory. For this trivial example one can easily find a solution that fulfills all constraints, e.g., $\lambda = 7$. Thus, as there is an intersection of manifold and feasible set (cf. Fig. 5.3) a second optimization problem in the nullspace of the design matrix has to be solved. We chose to examine the different effects of minimizing the length of the parameter vector with respect to the L_1 or L_2 norm:

Synthetic 2D example: Nullspace optimization	
objective function:	$\|\|\boldsymbol{x}_{\mathrm{P}}^{\mathrm{OLS}} + \boldsymbol{X}_{\mathrm{hom}}\lambda\|\|_p \ldots \mathrm{Min}$
constraints:	$\boldsymbol{B}_{\lambda}^{T}\lambda = \begin{bmatrix} -2 \\ 1 \end{bmatrix}\lambda \leq \begin{bmatrix} -12.88 \\ 10 \end{bmatrix} = \boldsymbol{b}_{\lambda}$
optim. variable:	$\lambda \in \mathbb{R}$.

The resulting value $\widetilde{\lambda}$ is used to compute the shortest solution vector that fulfills all constraints. Depending on the chosen norm this results in either

$$\widetilde{\boldsymbol{x}}_{\mathrm{P}}^{\mathrm{ICLS},L_2} = \boldsymbol{x}_{\mathrm{P}}^{\mathrm{OLS}} + \boldsymbol{X}_{\mathrm{hom}}\widetilde{\lambda}^{L_2} = \begin{bmatrix} 2.00 \\ 6.44 \end{bmatrix} \qquad (\textit{black circle} \text{ in Fig. 5.3}) \tag{5.28}$$

or

$$\widetilde{\boldsymbol{x}}_{\mathrm{P}}^{\mathrm{ICLS},L_1} = \boldsymbol{x}_{\mathrm{P}}^{\mathrm{OLS}} + \boldsymbol{X}_{\mathrm{hom}}\widetilde{\lambda}^{L_1} = \begin{bmatrix} 0.00 \\ 7.43 \end{bmatrix} \qquad (\textit{green cross} \text{ in Fig. 5.3}), \tag{5.29}$$

respectively. As expected, utilization of the L_1 norm in the nullspace optimization step leads to a sparse solution without an increase in the sum of squared residuals of the original problem. Considering the changed particular solution, we transform the constraints again and achieve as final result

$$\widetilde{\boldsymbol{x}}^{L_2} = \widetilde{\boldsymbol{x}}_{\mathrm{P}}^{\mathrm{ICLS},L_2} + \boldsymbol{x}_{\mathrm{hom}}^{\mathrm{OLS}}(\lambda) = \begin{bmatrix} 2.00 \\ 6.44 \end{bmatrix} + \lambda \begin{bmatrix} -2 \\ 1 \end{bmatrix} \tag{5.30a}$$

$$\text{subject to} \begin{bmatrix} -2 \\ 1 \end{bmatrix}\lambda \leq \begin{bmatrix} 0 \\ 0.56 \end{bmatrix}, \tag{5.30b}$$

and

$$\widetilde{\boldsymbol{x}}^{L_1} = \widetilde{\boldsymbol{x}}_{\mathrm{P}}^{\mathrm{ICLS},L_1} + \boldsymbol{x}_{\mathrm{hom}}^{\mathrm{OLS}}(\lambda) = \begin{bmatrix} 0.00 \\ 7.43 \end{bmatrix} + \lambda \begin{bmatrix} -2 \\ 1 \end{bmatrix} \tag{5.31a}$$

$$\text{subject to} \begin{bmatrix} -2 \\ 1 \end{bmatrix}\lambda \leq \begin{bmatrix} 2 \\ 2.57 \end{bmatrix}, \tag{5.31b}$$

respectively. Due to the introduction of inequality constraints the manifold is no longer a line, but a line segment. The constraint $x_1 \leq 2$ prevents that the ICLS solution is identical to the solution obtained via the pseudoinverse. This can also be seen in the final solution of the L_2 case (5.30), as there is one value on the right hand side of the transformed constraints that is zero—meaning that the constraint is active. As the absolute value of the second transformed constraint is small, it can be concluded that the manifold is only a small line segment which is also depicted in Fig. 5.3.

5.5.1.3 Case 2: Manifold and Feasible Region are Disjunct

This section deals with the same synthetic example as described in Sect. 5.5.1.1, but the second constraint is narrowed to demonstrate the case, in which the introduction of inequality constraints resolves the manifold and yields a unique solution. Let the constraints be

$$x_1 \leq 2, \quad x_2 \leq 2, \tag{5.32a}$$

leading to

$$\boldsymbol{B}^{T}\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\boldsymbol{x} \leq \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \boldsymbol{b}. \tag{5.32b}$$

This setup is depicted in Fig. 5.1(c). The constraint transformation yields

$$-2\lambda \leq -12.88, \quad \lambda \leq 2 \tag{5.33a}$$

and thus

$$\boldsymbol{B}_\lambda^T \lambda = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \lambda \leq \begin{bmatrix} -12.88 \\ 2 \end{bmatrix} = \boldsymbol{b}_\lambda. \quad \lightning \tag{5.33b}$$

These new constraints (5.33b) are contradictory as—due to constraint 2—the maximal feasible value of λ is 2, which is not enough to satisfy the first inequality. Therefore, one particular solution $\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}}$ of the original problem and the corresponding Lagrange multipliers $\boldsymbol{k}$ are computed using the binding-direction primal active-set method for rank-deficient systems described in Sect. 5.2.2, resulting in

$$\boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}} = \begin{bmatrix} 2.00 \\ 2.00 \end{bmatrix}, \quad \boldsymbol{k} = \begin{bmatrix} 88.8 \\ 177.60 \end{bmatrix}. \tag{5.34}$$

As there is no active constraint that is parallel to the manifold of solutions, the introduction of inequality constraints resolves the manifold and the computed particular solution

$$\widetilde{\boldsymbol{x}}^{\mathrm{ICLS}} = \boldsymbol{x}_{\mathrm{P}}^{\mathrm{ICLS}} = \begin{bmatrix} 2.00 \\ 2.00 \end{bmatrix} \tag{5.35}$$

is unique. This can be geometrically verified, considering Fig. 5.1(c), where $\widetilde{\boldsymbol{x}}^{\mathrm{ICLS}}$ is shown as *green cross*.

6. The Gauss-Helmert Model with Inequalities

In many applications it is crucial to model the relationship between two or more **observations** explicitly. If in addition some **parameters** of the system should be estimated, the Gauss-Helmert Model (GHM) is the model of choice to perform an adjustment. The GHM may be viewed as a special case of the *mixed model* (Koch, 1999, p. 213).

In this chapter an extension of the *weighted least-squares adjustment in the Gauss-Helmert model* described in Sect. 2.2.3 to the inequality constrained case is developed. As equality constraints are inherent to the GHM, we are dealing with a problem with both equality and inequality constraints.

Despite the fact that the GHM is often used in geodesy, not much literature (cf. Sect. 6.1) is devoted to the (weighted) least-squares adjustment in the *Inequality Constrained Gauss-Helmert Model* (ICGHM).

It should be noted that we slightly abuse the abbreviations here: For ICLS (the acronym for an *Inequality Constrained Least-Squares adjustment*, the (least-squares) **estimator** was eponymous and it is usually assumed that the Gauss-Markov model is used.
On the other hand, for ICGHM the name resulted from the **model** and it is assumed that the least-squares estimator is used (if not mentioned otherwise). However, as ICLS is a well established term, we have chosen to sacrifice some consistency here, for the sake of an easier distinction between both approaches.

The reformulation of the ICGHM as a quadratic program in standard form and the appendant KKT conditions are developed in Sect. 6.2.1. A less general but more sophisticated transformation is described in Sect. 6.2.2 as is an adaption of the binding-direction primal active-set algorithm to the peculiarities of this transformation. Again, the chapter concludes with a simple example, which serves as proof of concept (Sect. 6.3).

6.1 State of the Art

Famula (1983) described the use of inequality constraints for the estimation of non-negative variance components in the mixed model for a problem in animal science. However, besides the claim for non-negative variance components no further constraints are applied.

Kato and Hoijtink (2006) proposed a Bayesian approach for estimation and model selection in the linear mixed model with inequality constraints. They apply the Gibbs Sampler to problems with order constraints and sample from a truncated normal distribution. However, no (KKT) optimality conditions are provided and the focus is mostly on the sampling strategy.

Davis et al. (2008) and Davis (2011) explained a method for maximum likelihood estimation in what they called “general linear mixed models under linear inequality constraints”. The term *general* in this context means that no assumptions on the distribution of the observations are required. An active-set method is utilized in their approach. However, the (KKT) optimality conditions are only mentioned very briefly and the focus is put on hypothesis testing and the aspect of missing data.

6.2 WLS Adjustment in the Inequality Constrained GHM

The unconstrained Gauss-Helmert model was stated in Sect. 2.2.3 and reads

Weighted least-squares adjustment in the Gauss-Helmert model	
objective function:	$\Phi(\boldsymbol{v}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} \ldots \text{Min}$
constraints:	$\bar{\boldsymbol{B}}_{\text{GHM}}^T\,\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = 0$
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n, \Delta\boldsymbol{x} \in \mathbb{R}^m.$

Now, in addition to the inherent equality constraint, some equality constraints

$$\bar{\boldsymbol{B}}_{\text{eq}}^T \begin{bmatrix} \widetilde{\boldsymbol{\ell}} \\ \boldsymbol{x} \end{bmatrix} = \bar{\boldsymbol{b}}_{\text{eq}}^* \tag{6.1}$$

and some inequality constraints

$$\boldsymbol{B}^T \begin{bmatrix} \widetilde{\boldsymbol{\ell}} \\ \boldsymbol{x} \end{bmatrix} \leq \boldsymbol{b}^* \tag{6.2}$$

shall be introduced. As the optimization variables represent solely *changes* of the original variables

$$\widetilde{\boldsymbol{\ell}} = \boldsymbol{\ell} + \boldsymbol{v}, \tag{6.3}$$

$$\boldsymbol{x} = \boldsymbol{x}^0 + \Delta\boldsymbol{x}, \tag{6.4}$$

the constraints (6.1) and (6.2) have to be reformulated, yielding

$$\bar{\boldsymbol{B}}_{\text{eq}}^T \left(\begin{bmatrix} \boldsymbol{\ell} \\ \boldsymbol{x}^0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} \right) = \bar{\boldsymbol{b}}_{\text{eq}}^*, \tag{6.5a}$$

which implies

$$\bar{\boldsymbol{B}}_{\text{eq}}^T \begin{bmatrix} \boldsymbol{\ell} \\ \boldsymbol{x}^0 \end{bmatrix} + \bar{\boldsymbol{B}}_{\text{eq}}^T \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} = \bar{\boldsymbol{b}}_{\text{eq}}^*. \tag{6.5b}$$

Rearranging the terms yields

$$\bar{\boldsymbol{B}}_{\text{eq}}^T \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} = \bar{\boldsymbol{b}}_{\text{eq}}^* - \bar{\boldsymbol{B}}_{\text{eq}}^T \begin{bmatrix} \boldsymbol{\ell} \\ \boldsymbol{x}^0 \end{bmatrix} =: \bar{\boldsymbol{b}}_{\text{eq}}. \tag{6.5c}$$

An analogous reformulation of the inequality constraints leads to

$$\boldsymbol{B}^T \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} \leq \boldsymbol{b}^* - \boldsymbol{B}^T \begin{bmatrix} \boldsymbol{\ell} \\ \boldsymbol{x}^0 \end{bmatrix} =: \boldsymbol{b} \tag{6.6}$$

and thus to the

WLS adjustment in the Inequality Constrained GHM (ICGHM)	
objective function:	$\boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} \ldots \text{Min}$
constraints:	$\bar{\boldsymbol{B}}_{\text{GHM}}^T\,\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0}$
	$\bar{\boldsymbol{B}}_{\text{eq}}^T \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} = \bar{\boldsymbol{b}}_{\text{eq}}$
	$\boldsymbol{B}^T \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} \leq \boldsymbol{b}$
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n, \Delta\boldsymbol{x} \in \mathbb{R}^m.$

(6.7)

As the GHM can be seen as a combination of the adjustment with condition equations (cf. Sect. 2.2.2) and the adjustment in the GMM (cf. Sect. 2.2.1) the residuals $\boldsymbol{v}$ as well as the parameters $\Delta\boldsymbol{x}$ are the optimization variables. It is therefore possible to directly constrain the parameters and the residuals (and thus the adjusted observations). Therefore, it will prove convenient to partition the matrix $\boldsymbol{B}$ of inequality constraints in four submatrices, yielding

$$\boldsymbol{B}^T \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_{\boldsymbol{v}}^T & \boldsymbol{B}_{\boldsymbol{v}\Delta\boldsymbol{x}}^T \\ \boldsymbol{B}_{\Delta\boldsymbol{x}\boldsymbol{v}}^T & \boldsymbol{B}_{\Delta\boldsymbol{x}}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} \leq \boldsymbol{b}. \tag{6.8}$$

It should be noted that in most cases the involved matrices $\boldsymbol{B}_{\boldsymbol{v}}$ and $\boldsymbol{B}_{\Delta\boldsymbol{x}}$ are not symmetric. Furthermore

$$\boldsymbol{B}_{\boldsymbol{v}\Delta\boldsymbol{x}} \neq \boldsymbol{B}_{\Delta\boldsymbol{x}\boldsymbol{v}} \tag{6.9}$$

holds. If solely constraints are introduced that concern $\Delta\boldsymbol{x}$ **or** $\boldsymbol{v}$, $\boldsymbol{B}$ becomes a (usually non-symmetric) block-diagonal matrix. However, we chose to state the constraints in their most general form which includes also constraints that affect both $\Delta\boldsymbol{x}$ **and** $\boldsymbol{v}$.

One drawback of the use of reduced parameters $\Delta\boldsymbol{x}$ as optimization variables is that iterations and—due to the dependency from the constraints on the current solution—a recomputation of the constraints may be necessary if the initial point is far away from the optimal solution. However, this is not a result of the introduction of inequalities but true for the unconstrained case, too. In the unconstrained case, the Gauss-Newton method (cf. Nocedal and Wright, 1999, p. 259–262) is used as iteration scheme. Its extension to the constrained case is known as Lagrange-Newton method or Sequential Quadratic Programming (SQP, cf. Fletcher, 1987, p. 304–317). It is a generalization of the Quadratic Programming theory, which allows to solve a general non-linear optimization problem as a sequence of QPs.

Two comments concerning the ideas developed in the last chapters (namely the stochastic description and the treatment of rank-deficient systems) and their usefulness for the ICGHM might be appropriate. First, it is indeed possible (and often worthwhile) to combine the stochastic description treated in Chap. 4 with the ICGHM. However, we chose to focus on the development of the ICGHM framework without the stochastic description part in order to keep the formulas and algorithms tidy.
Unfortunately, the combination of the methods for rank-deficient systems described in Chap. 5 with the ICGHM is not that easy, due the non-linearities inherent to the GHM and the resulting need for Lagrange-Newton iterations. As a consequence, the computation of one particular solution is possible, but the computation of a homogeneous solution with respect to the constraints might prove difficult and is not pursued here.

6.2.1 Formulating the ICGHM as Standard Quadratic Program

In order to formulate the ICGHM problem (6.7) as a quadratic program in standard form (cf. Sect. 3.3):

Quadratic Program in Standard Form	
objective function:	$\Phi_{\mathrm{QP}}(\boldsymbol{z}) = \gamma_1 \boldsymbol{z}^T \boldsymbol{C} \boldsymbol{z} + \gamma_2 \boldsymbol{c}^T \boldsymbol{z} \ \ldots \mathrm{Min}$
constraints:	$\boldsymbol{B}^T \boldsymbol{z} \leq \boldsymbol{b}$
	$\bar{\boldsymbol{B}}^T \boldsymbol{z} = \bar{\boldsymbol{b}}$
optim. variable:	$\boldsymbol{z} \in \mathbb{R}^{m+n}$,

the optimization variable, the objective function and the constraints have to be adapted.

Optimization Variable. The parameters $\Delta\boldsymbol{x}$ and $\boldsymbol{v}$ are combined to the vector

$$\boldsymbol{z} := \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix}. \tag{6.10}$$

Sometimes $\Delta\boldsymbol{z}$ is used instead of $\boldsymbol{z}$ to explicitly denote that it contains the residuals and the reduced parameters.

Objective Function. Consequently, the objective function reads

$$\underbrace{\begin{bmatrix} \boldsymbol{v}^T & \Delta\boldsymbol{x}^T \end{bmatrix}}_{\boldsymbol{z}^T} \underbrace{\begin{bmatrix} \boldsymbol{\Sigma}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_m \end{bmatrix}}_{\boldsymbol{C}} \underbrace{\begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix}}_{\boldsymbol{z}} \ldots \text{Min.} \tag{6.11}$$

As the VCV matrix $\boldsymbol{\Sigma}$ is assumed to be positive definite, $\boldsymbol{C}$ is a positive semi-definite matrix, which has direct consequences for the solution algorithm (cf. Sect. 6.2.2). Without loss of generality the other terms of the objective function can be set as follows:

$$\boldsymbol{c} := \boldsymbol{0}, \qquad \gamma_1 := 1, \qquad \gamma_2 := 1. \tag{6.12}$$

Constraints. The inequality constraints are already almost in standard form, only substitution (6.10) has to be applied, yielding

$$\begin{bmatrix} \boldsymbol{B}_{\boldsymbol{v}}^T & \boldsymbol{B}_{\boldsymbol{v}\Delta\boldsymbol{x}}^T \\ \boldsymbol{B}_{\Delta\boldsymbol{x}\boldsymbol{v}}^T & \boldsymbol{B}_{\Delta\boldsymbol{x}}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} = \boldsymbol{B}^T \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} = \boldsymbol{B}^T \boldsymbol{z} \leq \boldsymbol{b}. \tag{6.13}$$

Reformulation of the inherent equality constraints is also straightforward as

$$\bar{\boldsymbol{B}}_{\text{GHM}}^T \boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0} \tag{6.14a}$$

implies

$$\begin{bmatrix} \bar{\boldsymbol{B}}_{\text{GHM}}^T & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} = -\boldsymbol{w}, \tag{6.14b}$$

thus

$$\begin{bmatrix} \bar{\boldsymbol{B}}_{\text{GHM}}^T & \boldsymbol{A} \end{bmatrix} \boldsymbol{z} = -\boldsymbol{w}. \tag{6.14c}$$

Combination with the additional equalities (6.5) yields

$$\underbrace{\begin{bmatrix} \begin{bmatrix} \bar{\boldsymbol{B}}_{\text{GHM}}^T & \boldsymbol{A} \end{bmatrix} \\ \bar{\boldsymbol{B}}_{\text{eq}}^T \end{bmatrix}}_{\bar{\boldsymbol{B}}^T} \boldsymbol{z} = \underbrace{\begin{bmatrix} -\boldsymbol{w} \\ \bar{\boldsymbol{b}}_{\text{eq}} \end{bmatrix}}_{\bar{\boldsymbol{b}}}. \tag{6.15}$$

6.2.1.1 The ICGHM with Independent Constraints

In the last section, the ICGHM was reformulated as a QP in standard form. Thus, the usual KKT optimality conditions of a standard QP (which are derived in Sect. 3.3.1) apply. However, in this section, the special case of independent equality and inequality constraints for parameters and adjusted observations

$$\begin{bmatrix} \bar{\boldsymbol{B}}_{\boldsymbol{v}}^T & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{B}}_{\Delta\boldsymbol{x}}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{b}}_{\boldsymbol{v}} \\ \bar{\boldsymbol{b}}_{\Delta\boldsymbol{x}} \end{bmatrix}, \tag{6.16}$$

$$\begin{bmatrix} \boldsymbol{B}_{\boldsymbol{v}}^T & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_{\Delta\boldsymbol{x}}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} \leq \begin{bmatrix} \boldsymbol{b}_{\boldsymbol{v}} \\ \boldsymbol{b}_{\Delta\boldsymbol{x}} \end{bmatrix} \tag{6.17}$$

is examined, because it allows for a direct interpretation of the Lagrange multipliers involved. Furthermore, inequality constraints that apply to both parameters **and** adjusted observations are rather scarce. In order to keep the formulas tidy, the optimization variables $\boldsymbol{v}$ and $\Delta\boldsymbol{x}$ are not combined to one vector (as was done in the last section) but kept separately. Thus, the problem for which we will derive the KKT conditions reads

INEQUALITY CONSTRAINED GHM WITH INDEPENDENT CONSTRAINTS

objective function:	$\boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} \ldots \text{Min}$
constraints:	$\bar{\boldsymbol{B}}_{\text{GHM}}^T\,\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0}$
	$\bar{\boldsymbol{B}}_{\boldsymbol{v}}^T\boldsymbol{v} = \bar{\boldsymbol{b}}_{\boldsymbol{v}}$
	$\bar{\boldsymbol{B}}_{\Delta\boldsymbol{x}}^T\Delta\boldsymbol{x} = \bar{\boldsymbol{b}}_{\Delta\boldsymbol{x}}$
	$\boldsymbol{B}_{\boldsymbol{v}}^T\boldsymbol{v} \leq \boldsymbol{b}_{\boldsymbol{v}}$
	$\boldsymbol{B}_{\Delta\boldsymbol{x}}^T\Delta\boldsymbol{x} \leq \boldsymbol{b}_{\Delta\boldsymbol{x}}$
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n, \Delta\boldsymbol{x} \in \mathbb{R}^m.$

(6.18)

6.2.1.2 KKT Conditions of the Standard Transformation

The Lagrangian corresponding to problem (6.18) reads

$$\begin{aligned} L(\boldsymbol{v},\Delta\boldsymbol{x},\bar{\boldsymbol{k}}_{\text{GHM}},\bar{\boldsymbol{k}}_{\boldsymbol{v}},\bar{\boldsymbol{k}}_{\Delta\boldsymbol{x}},\boldsymbol{k}_{\boldsymbol{v}},\boldsymbol{k}_{\Delta\boldsymbol{x}}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} &- 2\bar{\boldsymbol{k}}_{\text{GHM}}^T(\bar{\boldsymbol{B}}_{\text{GHM}}^T\,\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w}) \\ &- 2\bar{\boldsymbol{k}}_{\boldsymbol{v}}^T(\bar{\boldsymbol{B}}_{\boldsymbol{v}}^T\boldsymbol{v} - \bar{\boldsymbol{b}}_{\boldsymbol{v}}) - 2\bar{\boldsymbol{k}}_{\Delta\boldsymbol{x}}^T(\bar{\boldsymbol{B}}_{\Delta\boldsymbol{x}}^T\Delta\boldsymbol{x} - \bar{\boldsymbol{b}}_{\Delta\boldsymbol{x}}) \\ &+ 2\boldsymbol{k}_{\boldsymbol{v}}^T(\boldsymbol{B}_{\boldsymbol{v}}^T\boldsymbol{v} - \boldsymbol{b}_{\boldsymbol{v}}) + 2\boldsymbol{k}_{\Delta\boldsymbol{x}}^T(\boldsymbol{B}_{\Delta\boldsymbol{x}}^T\Delta\boldsymbol{x} - \boldsymbol{b}_{\Delta\boldsymbol{x}}). \end{aligned} \tag{6.19}$$

$\bar{\boldsymbol{k}}_{\text{GHM}}$ denotes the Lagrange multipliers corresponding to the GHM inherent equality constraint (6.14), $\bar{\boldsymbol{k}}_{\boldsymbol{v}}$ and $\bar{\boldsymbol{k}}_{\Delta\boldsymbol{x}}$ those linked with the additional equality constraints (6.16) and $\boldsymbol{k}_{\boldsymbol{v}}$ and $\boldsymbol{k}_{\Delta\boldsymbol{x}}$ those linked with the inequality constraints (6.17). Now the gradients with respect to the optimization variables can be computed and set equal to zero

$$\nabla_{\boldsymbol{v}} L(\boldsymbol{v},\bar{\boldsymbol{k}}_{\text{GHM}},\bar{\boldsymbol{k}}_{\boldsymbol{v}},\boldsymbol{k}_{\boldsymbol{v}}) = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{v} - 2\bar{\boldsymbol{B}}_{\text{GHM}}\bar{\boldsymbol{k}}_{\text{GHM}} - 2\bar{\boldsymbol{B}}_{\boldsymbol{v}}\bar{\boldsymbol{k}}_{\boldsymbol{v}} + 2\boldsymbol{B}_{\boldsymbol{v}}\boldsymbol{k}_{\boldsymbol{v}} \stackrel{!}{=} \boldsymbol{0} \tag{6.20}$$

$$\Rightarrow \boldsymbol{v} = \boldsymbol{Q}(\bar{\boldsymbol{B}}_{\text{GHM}}\bar{\boldsymbol{k}}_{\text{GHM}} + \bar{\boldsymbol{B}}_{\boldsymbol{v}}\bar{\boldsymbol{k}}_{\boldsymbol{v}} - \boldsymbol{B}_{\boldsymbol{v}}\boldsymbol{k}_{\boldsymbol{v}}) \tag{6.21}$$

$$\nabla_{\Delta\boldsymbol{x}} L(\bar{\boldsymbol{k}}_{\text{GHM}},\bar{\boldsymbol{k}}_{\Delta\boldsymbol{x}},\boldsymbol{k}_{\Delta\boldsymbol{x}}) = -2\boldsymbol{A}^T\bar{\boldsymbol{k}}_{\text{GHM}} - 2\bar{\boldsymbol{B}}_{\Delta\boldsymbol{x}}\bar{\boldsymbol{k}}_{\Delta\boldsymbol{x}} + 2\boldsymbol{B}_{\Delta\boldsymbol{x}}\boldsymbol{k}_{\Delta\boldsymbol{x}} \stackrel{!}{=} \boldsymbol{0}. \tag{6.22}$$

The gradients with respect to the Lagrange multipliers follow to be

$$\nabla_{\bar{\boldsymbol{k}}_{\mathrm{GHM}}} L(\boldsymbol{v}, \Delta\boldsymbol{x}) = \bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\, \boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} \stackrel{!}{=} \boldsymbol{0} \tag{6.23}$$

$$\nabla_{\bar{\boldsymbol{k}}_{\boldsymbol{v}}} L(\boldsymbol{v}) = \bar{\boldsymbol{B}}_{\boldsymbol{v}}^T \boldsymbol{v} - \bar{\boldsymbol{b}}_{\boldsymbol{v}} \stackrel{!}{=} \boldsymbol{0} \tag{6.24}$$

$$\nabla_{\bar{\boldsymbol{k}}_{\Delta\boldsymbol{x}}} L(\Delta\boldsymbol{x}) = \bar{\boldsymbol{B}}_{\boldsymbol{x}}^T \Delta\boldsymbol{x} - \bar{\boldsymbol{b}}\Delta\boldsymbol{x} \stackrel{!}{=} \boldsymbol{0} \tag{6.25}$$

$$\nabla_{\boldsymbol{k}_{\boldsymbol{v}}} L(\boldsymbol{v}) = \boldsymbol{B}_{\boldsymbol{v}}^T \boldsymbol{v} - \boldsymbol{b}_{\boldsymbol{v}} \stackrel{!}{\leq} \boldsymbol{0} \tag{6.26}$$

$$\nabla_{\boldsymbol{k}_{\Delta\boldsymbol{x}}} L(\Delta\boldsymbol{x}) = \boldsymbol{B}_{\Delta\boldsymbol{x}}^T \Delta\boldsymbol{x} - \boldsymbol{b}\Delta\boldsymbol{x} \stackrel{!}{\leq} \boldsymbol{0}, \tag{6.27}$$

which are nothing else than the original constraints. The conditions of complementary slackness (cf. Sect. 3.3.1) read

$$k_{\boldsymbol{v}_i}(\boldsymbol{B}_{\boldsymbol{v}}(:,i)^T \boldsymbol{v} - \boldsymbol{b}_{\boldsymbol{v}_i}) \stackrel{!}{=} 0, \qquad \forall\ i = 1, 2, \ldots p_{\boldsymbol{v}} \tag{6.28}$$

$$k_{\Delta\boldsymbol{x}_i}(\boldsymbol{B}_{\Delta\boldsymbol{x}}(:,j)^T \Delta\boldsymbol{x} - \boldsymbol{b}_{\Delta\boldsymbol{x}_j}) \stackrel{!}{=} 0, \qquad \forall\ j = 1, 2, \ldots p_{\Delta\boldsymbol{x}}. \tag{6.29}$$

All Lagrange multipliers linked with inequality constraints have to be non-negative (cf. Sect. 3.3.1)

$$k_{\boldsymbol{v}_i} \stackrel{!}{\geq} 0, \qquad \forall\ i = 1, 2, \ldots p_{\boldsymbol{v}} \tag{6.30}$$

$$k_{\Delta\boldsymbol{x}_j} \stackrel{!}{\geq} 0, \qquad \forall\ j = 1, 2, \ldots p_{\Delta\boldsymbol{x}}. \tag{6.31}$$

Thus, it is directly possible to examine the influence of a constraint on the objective function by interpreting the associated Lagrange multiplier. However, one has to be careful as Lagrange multipliers are dimensionless quantities. Thoughts on the computation of some measures that are easier to interpret can be found in Sect. 4.3.

As described above, it is possible to transform an ICGHM problem into a QP in standard form. However, this may lead to a large "overhead" of the problem, which could lead to numerical difficulties and an increase in run time. Under certain circumstances, a more elegant and efficient way to solve an ICGHM problem can be found, which is explained in the next section.

6.2.2 Tailor-Made Transformation of the ICGHM

As stated above, every ICGHM can be reformulated as a QP in standard form. However, this has some drawbacks:

1. As can be seen in (6.11) the matrix $\boldsymbol{C}$ in the objective function will always be positive semi-definite and have many rows and columns containing only zeros. If dense matrix computations are used, this will generate a large computational "overhead" and an increase in run-time and memory requirements.
2. As $\Delta\boldsymbol{x}$ and $\boldsymbol{v}$ are both used as optimization variables (cf. (6.10)), in every Lagrange-Newton iteration feasible initial values have to be provided. Especially for $\boldsymbol{v}$ this might be difficult in some cases (due to the presence of both equality and inequality constraints).

Therefore, in this section, a more sophisticated transformation of the ICGHM is introduced, which should overcome the disadvantages mentioned above. However, it requires a tailor-made computation of the search direction of the active-set method described in Sect. 3.5. Thus, it will not be possible to

solve this formulation of the ICGHM with a standard solver. Furthermore, it is solely possible here to incorporate constraints to the parameters $\boldsymbol{x}$ and not to the adjusted observations $\widetilde{\boldsymbol{\ell}}$. However, the dimension of the problem will be much smaller than in the "standard formulation" and the system can be solved more efficiently.

We will follow a similar line of thought as in the equality constrained GHM (ECGHM, cf. Sect. 2.2.3.2). This is possible because with active-set methods, an inequality constrained problem is solved as a sequence of equality constrained ones (using in each iteration only the active constraints, cf. Sect. 3.5.1). This can be exploited to reduce the size of the system.

6.2.2.1 The ICGHM with Constraints Concerning Solely the Parameters $\Delta\boldsymbol{x}$

As stated above, only constraints concerning the parameters $\Delta\boldsymbol{x}$ shall be introduced in this section

$$\bar{\boldsymbol{B}}_{\Delta\boldsymbol{x}}^T \Delta\boldsymbol{x} = \bar{\boldsymbol{b}}_{\boldsymbol{x}}, \tag{6.32}$$

$$\boldsymbol{B}_{\Delta\boldsymbol{x}}^T \Delta\boldsymbol{x} \leq \boldsymbol{b}_{\boldsymbol{x}}. \tag{6.33}$$

Thus, the resulting problem reads

ICGHM with Constraints Concerning Solely the Parameters $\Delta\boldsymbol{x}$

objective function:	$\boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} \ldots \text{Min}$
constraints:	$\bar{\boldsymbol{B}}_{\text{GHM}}^T\, \boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0}$
	$\bar{\boldsymbol{B}}_{\Delta\boldsymbol{x}}^T \Delta\boldsymbol{x} = \bar{\boldsymbol{b}}_{\Delta\boldsymbol{x}}$
	$\boldsymbol{B}_{\Delta\boldsymbol{x}}^T \Delta\boldsymbol{x} \leq \boldsymbol{b}_{\Delta\boldsymbol{x}}$
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n, \Delta\boldsymbol{x} \in \mathbb{R}^m.$

(6.34)

As hinted in the last section, we will exploit the fact that in active-set methods the inequality constrained problem is solved as a sequence of equality constrained ones. Therefore, in each iteration the active constraints

$$\boldsymbol{B}_{\Delta\boldsymbol{x},\text{active}} \Delta\boldsymbol{x} = \boldsymbol{b}_{\Delta\boldsymbol{x},\text{active}}$$

are identified and a working set

$$\overline{\boldsymbol{W}} := \left[\bar{\boldsymbol{B}}_{\Delta\boldsymbol{x}} \quad \boldsymbol{B}_{\Delta\boldsymbol{x},\text{active}}\right] \tag{6.35}$$

$$\overline{\boldsymbol{w}} := \begin{bmatrix} \bar{\boldsymbol{b}}_{\Delta\boldsymbol{x}} \\ \boldsymbol{b}_{\Delta\boldsymbol{x},\text{active}} \end{bmatrix} \tag{6.36}$$

is derived, which contains all equality constraints as well as the active inequalities (cf. Sect. 3.5.1.1). In contrast to Sect. 3.5.1.1 we mark the working set with a bar, too. This is for two reasons. First, we would like to emphasize that these constraints are treated as equalities. Second, to avoid confusion with the vector of misclosures $\boldsymbol{w}$. The problem which has to be solved in each iteration of the active-set method reads

Tailor-Made Subproblem: ICGHM Including Only Active Constraints

objective function:	$\boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} \ldots \text{Min}$
constraints:	$\bar{\boldsymbol{B}}_{\text{GHM}}^T\, \boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0}$
	$\overline{\boldsymbol{W}}^T \Delta\boldsymbol{x} = \overline{\boldsymbol{w}}$
optim. variable:	$\boldsymbol{v} \in \mathbb{R}^n, \Delta\boldsymbol{x} \in \mathbb{R}^m.$

(6.37)

It should be noted that the working set changes in every active-set iteration. Thus, a tailor-made adaption of a binding-direction primal active-set algorithm is provided in Sect. 6.2.2.3. For the sake of an easier interpretation we chose not to include the optimization variables $\boldsymbol{v}$ and $\Delta\boldsymbol{x}$ in one vector but to keep them separately. The same holds true for the inherent and the additional constraints.

6.2.2.2 KKT Conditions of the Tailor-Made Transformation

As problem (6.37) is an equality constrained problem, the Lagrangian

$$L(\boldsymbol{v}, \Delta\boldsymbol{x}, \bar{\boldsymbol{k}}_{\mathrm{GHM}}, \bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}}) = \boldsymbol{v}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{v} - 2\bar{\boldsymbol{k}}_{\mathrm{GHM}}^T(\bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\boldsymbol{v} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w}) - 2\bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}}^T(\overline{\boldsymbol{W}}^T\Delta\boldsymbol{x} - \overline{\boldsymbol{w}}) \tag{6.38}$$

is identical to (2.74)—the Lagrangian of ECGHM problem (2.73)—but with $\overline{\boldsymbol{W}}$ and $\overline{\boldsymbol{w}}$ instead of $\bar{\boldsymbol{B}}$ and $\bar{\boldsymbol{b}}$, respectively. The same holds true for the KKT conditions resulting from setting the gradients equal to zero and the substitution (2.79) of $\boldsymbol{v}$

$$\bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\mathrm{GHM}}\bar{\boldsymbol{k}}_{\mathrm{GHM}} + \boldsymbol{A}\Delta\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{0}, \tag{6.39}$$

$$\boldsymbol{A}^T\bar{\boldsymbol{k}}_{\mathrm{GHM}} + \overline{\boldsymbol{W}}\bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}} = \boldsymbol{0}, \tag{6.40}$$

$$\overline{\boldsymbol{W}}^T\Delta\boldsymbol{x} = \overline{\boldsymbol{w}}, \tag{6.41}$$

as well for the resulting KKT system

$$\begin{bmatrix} \bar{\boldsymbol{B}}_{\mathrm{GHM}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\mathrm{GHM}} & \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{A}^T & \boldsymbol{0} & \overline{\boldsymbol{W}} \\ \boldsymbol{0} & \overline{\boldsymbol{W}}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{k}}_{\mathrm{GHM}} \\ \Delta\boldsymbol{x} \\ \bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{0} \\ \overline{\boldsymbol{w}} \end{bmatrix}. \tag{6.42}$$

6.2.2.3 A Tailor-Made Active-Set Method for the Gauss-Helmert Model

As we have substituted $\boldsymbol{v}$ in order to reduce the size of the resulting KKT system (6.42), some small changes to the active-set method described in Sect. 3.5.1 are required as now only a part of the original parameter vector

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{v} \\ \Delta\boldsymbol{x} \end{bmatrix} \tag{6.43}$$

is contained in the equation system. More specifically, the computation of the search direction $\boldsymbol{p}$ has to be modified.

Therefore, we have to recall that if no inactive constraints prevent that a step of full length is taken (cf. Sect. 3.5.1.5)

$$\Delta\boldsymbol{x}^{(i+1)} = \Delta\boldsymbol{x}^{(i)} + \boldsymbol{p}^{(i)} \tag{6.44}$$

holds. As a consequence, it is possible to compute the search direction by solving (6.42) and computing the difference

$$\boldsymbol{p}^{(i)} = \Delta\boldsymbol{x}^{(i+1)} - \Delta\boldsymbol{x}^{(i)}. \tag{6.45}$$

The resulting modified active-set method is shown in Alg. 6. One peculiarity should be mentioned. In (6.38) we chose to introduce a negative weight for the Lagrange multiplier $\bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}}$ of the active constraints in order to obtain a symmetric matrix in the KKT system (6.42). As a consequence,

Algorithm 6: Binding-Direction Primal Active-Set Algorithm Tailored for the GHM

Data:

$\boldsymbol{B}_{[m\times p]}, \boldsymbol{b}_{[p\times 1]}$	...	Matrix and corresponding right-hand side of the inequality constraints
$\bar{\boldsymbol{B}}_{[m\times \bar{p}]}, \bar{\boldsymbol{b}}_{[\bar{p}\times 1]}$	...	Matrix and corresponding right-hand side of the equality constraints
$\bar{\boldsymbol{B}}_{\mathrm{GHM}[n\times b]}, \boldsymbol{w}_{[b\times 1]}$	...	Matrix with the GHM inherent equality constraints and misclosure
$\boldsymbol{\Sigma}_{[n\times n]}$	...	VCV matrix of the observations
$\Delta\boldsymbol{x}^{(0)}_{[m\times 1]}$	...	Feasible initial (reduced) solution vector

Result:

$\Delta\boldsymbol{x}_{[m\times 1]}$	...	Vector containing the (reduced) solution of the QP
$\bar{\boldsymbol{k}}_{\mathrm{GHM}[b\times 1]}$	...	Vector containing the Lagrange multipliers of the GHM inherent constraints
$\bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}[p_{\bar{w}}\times 1]}$	...	Vector containing the Lagrange multipliers of active additional constraints

$[\overline{\boldsymbol{W}}, \overline{\boldsymbol{w}}, \boldsymbol{V}, \boldsymbol{v}] = \mathrm{findActiveConstraints}(\boldsymbol{B}, \boldsymbol{b}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{b}}, \Delta\boldsymbol{x})$ *// cf. Sect. 3.5.1.1*

for $i = 1 : i_{\max}$ **do**

  $\Delta\boldsymbol{x}_{\mathrm{old}} = \Delta\boldsymbol{x}$

  solve $\left(\begin{bmatrix} \bar{\boldsymbol{B}}^T_{\mathrm{GHM}} \boldsymbol{\Sigma} \bar{\boldsymbol{B}}_{\mathrm{GHM}} & \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{A}^T & \boldsymbol{0} & \overline{\boldsymbol{W}} \\ \boldsymbol{0} & \overline{\boldsymbol{W}}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{k}}_{\mathrm{GHM}} \\ \Delta\boldsymbol{x} \\ \bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{0} \\ \overline{\boldsymbol{w}} \end{bmatrix} \right)$ *// cf. Sect. 6.2.1.2*

  $\boldsymbol{p} = \Delta\boldsymbol{x} - \Delta\boldsymbol{x}_{\mathrm{old}}$

  $q = \mathrm{computeStepLength}(\boldsymbol{V}, \boldsymbol{v}, \Delta\boldsymbol{x}, \boldsymbol{p})$ *// cf. Sect. 3.5.1.5*

  $\Delta\boldsymbol{x} = \Delta\boldsymbol{x} + q\boldsymbol{p}$

  $[\overline{\boldsymbol{W}}, \overline{\boldsymbol{w}}, \boldsymbol{V}, \boldsymbol{v}] = \mathrm{updateActiveSet}(\overline{\boldsymbol{W}}, \overline{\boldsymbol{w}}, \boldsymbol{V}, \boldsymbol{v}, \Delta\boldsymbol{x}, q)$ *// cf. Sect. 3.5.1.6*

  if $\max(\bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}} \leq 0)$ **then**

    break

  end

end

return $\Delta\boldsymbol{x}, \bar{\boldsymbol{k}}_{GHM}, \boldsymbol{k}_{\bar{w}}$

the abortion criterion of the algorithm has to be adapted: Now, all $\bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}}$ should be less or equal to zero—and not greater or equal to zero as in Alg. 2 and Alg. 4.

It should be emphasized that this algorithm is to be seen as a kind of "inner loop" as all matrices of constraints depend on the values of the current solution $\boldsymbol{x}^{(j)}$ and the adjusted observations. Therefore, they have to be recomputed every time Alg. 6 terminates. The "outer loop" is formalized in Alg. 7.

One comment concerning the runtime of Alg. 6 might be in order: The computationally most expensive operation in this algorithm is the solution of the system of linear equations (6.42). As there are only small changes in this system in every iteration, it might be worthwhile to examine its structure. In every iteration only the quantities $\overline{\boldsymbol{W}}, \overline{\boldsymbol{w}}$ and $\boldsymbol{w}$ change and the upper left block of the KKT matrix

$$\begin{bmatrix} \bar{\boldsymbol{B}}^T_{\mathrm{GHM}} \boldsymbol{\Sigma} \bar{\boldsymbol{B}}_{\mathrm{GHM}} & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{0} \end{bmatrix} \tag{6.46}$$

Algorithm 7: WLS estimate in the Inequality Constrained Gauss-Helmert Model (ICGHM)

Data:

$\boldsymbol{\ell}_{[n\times 1]}$	...	Vector of observations
$\boldsymbol{\Sigma}_{[n\times n]}$	...	VCV matrix of the observations
$\boldsymbol{x}^{(0)}_{[m\times 1]}$	...	Feasible initial solution vector

Result:

$\boldsymbol{x}_{[m\times 1]}$	...	Vector containing the solution of the QP
$\bar{\boldsymbol{k}}_{\mathrm{GHM}[b\times 1]}$	...	Vector containing the Lagrange multipliers of the GHM inherent constraints
$\bar{\boldsymbol{k}}_{\bar{\boldsymbol{w}}[p_{\bar{w}}\times 1]}$	...	Vector containing the Lagrange multipliers of active additional constraints

$\boldsymbol{\ell}_0 = \boldsymbol{\ell}$
for $j = 1 : j_{\max}$ **do**
  $\Delta\boldsymbol{\ell} = \boldsymbol{\ell} - \boldsymbol{\ell}_0$
  $\Delta\boldsymbol{x} = \mathbf{0}$ *// ensure feasibility of the starting point*
  $\boldsymbol{A} = \mathrm{assembleDesignmatrix}(\boldsymbol{\ell}_0, \boldsymbol{x}_0)$
  $[\boldsymbol{B}, \boldsymbol{b}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{B}}_{\mathrm{GHM}}] = \mathrm{assembleConstraints}(\boldsymbol{\ell}_0, \boldsymbol{x}_0)$ *// cf. (6.5), (6.6) and (6.15)*
  $\boldsymbol{w}_0 = \mathrm{computeMisclosure}(\boldsymbol{\ell}_0, \boldsymbol{x}_0)$
  $\boldsymbol{w} = \bar{\boldsymbol{B}}^T_{\mathrm{GHM}}\Delta\boldsymbol{\ell} + \boldsymbol{w}_0$
  $[\Delta\boldsymbol{x}, \bar{\boldsymbol{k}}_{\mathrm{GHM}}, \bar{\boldsymbol{k}}_{\bar{\boldsymbol{w}}}] = \mathrm{ActiveSetAlgoForGhm}(\boldsymbol{B}, \boldsymbol{b}, \bar{\boldsymbol{B}}, \bar{\boldsymbol{b}}, \bar{\boldsymbol{B}}_{\mathrm{GHM}}, \boldsymbol{w}, \boldsymbol{\Sigma}, \Delta\boldsymbol{x})$ *// cf. Alg. 6*
  $\boldsymbol{x}_0 = \boldsymbol{x}_0 + \Delta\boldsymbol{x}$
  $\boldsymbol{v} = \boldsymbol{\Sigma}\bar{\boldsymbol{B}}_{\mathrm{GHM}}\bar{\boldsymbol{k}}_{\mathrm{GHM}}$ *// cf. Sect. 2.2.3.1*
  $\boldsymbol{\ell}_0 = \boldsymbol{\ell} + \boldsymbol{v}$
  if $\max|\Delta\boldsymbol{x}| < \epsilon$ **then**
    break
  end
end
return $\boldsymbol{x}_0, \bar{\boldsymbol{k}}_{GHM}, \boldsymbol{k}_{\bar{w}}$

stays constant. Furthermore, interim values for the Lagrange multipliers $\boldsymbol{k}_{\mathrm{GHM}}$ of the GHM inherent constraint are not required for further computations. Thus, there is a huge potential to optimize the solution strategy.

One could e.g., use an LU decomposition and precompute the reduction of the constant block (6.46) or use the Gauss-Jordan algorithm (cf. Sect. 2.1.3.1) for this task, which leads to the Schur form. This precomputation only has to be computed once and can then be used in every iteration yielding a system that is already partly reduced. Furthermore, it is not necessary to completely reduce the system, as the values of the $\boldsymbol{k}_{\mathrm{GHM}}$ only have to be computed explicitly when the algorithm terminates. As the focus of this chapter is on the methodologic development, these thoughts on improvements of the run-time behavior are mentioned only briefly and not explained in detail.

6.2.3 Comparison of Both ICGHM Reformulations

A comparison of the two ICGHM reformulations presented in Sect. 6.2.1 and Sect. 6.2.2, respectively is provided in Table 6.1. In summary—due to the faster and more stable computation—it

Table 6.1: Advantages and drawbacks of two possible representations of the ICGHM: The formulation as a standard quadratic program described in Sect. 6.2.1 and the formulation via a tailor-made approach described in Sect. 6.2.2.

ICGHM ...	**... as a standard QP**	**... as a tailor-made approach**
Constraints possible for	$\boldsymbol{x}$ and $\widetilde{\boldsymbol{l}}$	only $\boldsymbol{x}$
Required feasible initial values	$\Delta\boldsymbol{x}$ and $\boldsymbol{v}$ (difficult)	only $\boldsymbol{x}$
Use of standard solvers	possible	not possible
Numerics	relatively unstable	stable
Speed and memory requirem.	large SLEs, slow	smaller SLEs, faster

is recommendable to use the tailor-made approach, whenever possible. The requirements are that there exist solely constraints for the parameters $\Delta\boldsymbol{x}$ and—as a result of the tailored computation of the search direction—that the use of a standard solver is not mandatory.

6.3 Example

In order to obtain results which can be validated, an example for the least-squares adjustment in the GHM from Lenzmann and Lenzmann (2004) is used as proof of concept and one additional inequality constraint is introduced. As the introduced constraint affects solely a parameter x, the solution is computed using the tailor-made approach developed in Sect. 6.2.2.

Problem Description

A parabola through the origin of the coordinate system shall be estimated. The functional model reads

$$y_i = x t_i^2. \tag{6.47}$$

The y_i as well as the t_i are observed and assumed to be independent and normally distributed, the unknown parameter x is to be estimated. The observations read

$$t_1 = 2.5, \quad t_2 = 4.0 \tag{6.48}$$

$$y_1 = 4.8, \quad y_2 = 5.0 \tag{6.49}$$

and are combined to the vector of observations

$$\boldsymbol{\ell} = \begin{bmatrix} t_1 \\ t_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 4.0 \\ 4.8 \\ 5.0 \end{bmatrix}. \tag{6.50}$$

It should be noted that the residuals in both the y_i and the t_i will be larger than usual in most geodetic setups. Lenzmann and Lenzmann (2004) intentionally designed the example this way, to point out the differences between a strict and an approximative linearization scheme.

Following (2.55) we reformulate the functional relationship into

$$g_i(\boldsymbol{\ell}, x) = y_i - x t_i^2 \stackrel{!}{=} 0. \tag{6.51}$$

The redundancy is one, as the functional relationship consists of two equations and one parameter shall be estimated.

Initialization $(i = 0)$

We use

$$\boldsymbol{z}_0^{(0)} = \begin{bmatrix} \widetilde{\boldsymbol{\ell}}_0^{(0)} \\ x_0^{(0)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\ell} \\ 0.2 \end{bmatrix} \tag{6.52}$$

as initial values. The index in brackets is the number of the current iteration. Subsequently, the following quantities can be derived:

The non-linearized misclosure

$$\boldsymbol{w}_0 = \boldsymbol{g}(\boldsymbol{\ell}, x)|_{\boldsymbol{\ell}=\boldsymbol{\ell}_0, x=x_0} = \begin{bmatrix} y_1 - x_0^{(0)} t_1^2 \\ y_2 - x_0^{(0)} t_2^2 \end{bmatrix} = \begin{bmatrix} 3.55 \\ 1.80 \end{bmatrix}, \tag{6.53}$$

the matrix

$$\overline{\boldsymbol{B}}_{\mathrm{GHM}}^T = \nabla_{\boldsymbol{\ell}}\, \boldsymbol{g}(\boldsymbol{\ell}, x)|_{\boldsymbol{\ell}=\boldsymbol{\ell}_0, x=x_0} = \begin{bmatrix} \frac{\partial g_1}{\partial t_1} & \frac{\partial g_1}{\partial t_2} & \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} \\ \frac{\partial g_2}{\partial t_1} & \frac{\partial g_2}{\partial t_2} & \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} \end{bmatrix} \tag{6.54a}$$

$$= \begin{bmatrix} -2x_0^{(0)} t_1 & 0 & 1 & 0 \\ 0 & -2x_0^{(0)} t_2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1.6 & 0 & 1 \end{bmatrix} \tag{6.54b}$$

containing the GHM inherent constraint and the design matrix

$$\boldsymbol{A} = \nabla_x\, \boldsymbol{g}(\boldsymbol{\ell}, x)|_{\boldsymbol{\ell}=\boldsymbol{\ell}_0, x=x_0} = \begin{bmatrix} \frac{\partial g_1}{\partial x} \\ \frac{\partial g_2}{\partial x} \end{bmatrix} = \begin{bmatrix} -t_1^2 \\ -t_2^2 \end{bmatrix} = \begin{bmatrix} -6.25 \\ -16.00 \end{bmatrix}. \tag{6.55}$$

As in the first iteration the observations $\boldsymbol{\ell}$ are identical to the computed observations $\boldsymbol{\ell}_0$, (2.59) yields that the linearized and the non-linearized misclosures are the same

$$\boldsymbol{w} = \boldsymbol{w}_0. \tag{6.56}$$

Strictly speaking, the quantities $\boldsymbol{w}_0, \boldsymbol{w}, \overline{\boldsymbol{B}}_{\mathrm{GHM}}, \boldsymbol{A}$ and all matrices and right-hand-sides of constraints would need an iteration index, too. However, as their values change only in every **outer** iteration and as we intend to focus on the **inner** iterations (i.e., the adjusted active-set method), these indices were neglected. However, the matrices and vectors of the active and inactive sets of constraints $\overline{\boldsymbol{W}}^{(i)}, \overline{\boldsymbol{w}}^{(i)}, \boldsymbol{V}^{(i)}$ and $\boldsymbol{v}^{(i)}$ have an iteration index as their values and dimensions change during the inner iterations. Lenzmann and Lenzmann (2004) showed that the minimization of the sum of squared residuals in the unconstrained GHM yields the solution

$$\widetilde{x} = 0.46. \tag{6.57}$$

Therefore, we chose to introduce the constraint

$$x^{(0)} = x_0^{(0)} + \Delta x^{(0)} \leq 0.4 \tag{6.58}$$

which implies

$$\underbrace{\begin{bmatrix} 1 \end{bmatrix}}_{\boldsymbol{B}^T} (x_0^{(0)} + \Delta x^{(0)}) \leq \underbrace{\begin{bmatrix} 0.4 \end{bmatrix}}_{\boldsymbol{b}^{(0)*}} \tag{6.59}$$

and thus

$$\boldsymbol{B}^T \Delta x^{(0)} \leq \boldsymbol{b}^* - \boldsymbol{B}^T x_0^{(0)} = \underbrace{0.2}_{\boldsymbol{b}}. \tag{6.60}$$

As we expect the reduced parameters to be small, in the initialization step

$$\Delta x^{(0)} = 0 \tag{6.61}$$

is assumed. As shown in line 1 of Alg. 6 as a first step the active set is computed. As the constraint is not active in $x_0^{(0)}$ the components of the active set $\overline{\boldsymbol{W}}^{(0)}$ and $\overline{\boldsymbol{w}}^{(0)}$ are empty and the set of inactive constraints reads

$$\boldsymbol{V}^{(0)} = \begin{bmatrix} 1 \end{bmatrix}, \quad \boldsymbol{v}^{(0)} = \begin{bmatrix} 0.2 \end{bmatrix}. \tag{6.62}$$

First Iteration ($i = 1$)

As there are no active constraints, the KKT system (6.42) can be reduced to the form

$$\begin{bmatrix} \overline{\boldsymbol{B}}_{\mathrm{GHM}}^T \boldsymbol{Q} \overline{\boldsymbol{B}}_{\mathrm{GHM}} & \boldsymbol{A}^{(0)} \\ \boldsymbol{A}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{k}}_{\mathrm{GHM}}^{(1)} \\ \Delta x^{(1)} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{w}^{(0)} \\ \boldsymbol{0} \end{bmatrix} \tag{6.63a}$$

yielding

$$\begin{bmatrix} 2.00 & 0.00 & -6.25 \\ 0.00 & 3.56 & -16.00 \\ -6.25 & -16.00 & 0.00 \end{bmatrix} \begin{bmatrix} \bar{k}_{1,\mathrm{GHM}}^{(1)} \\ \bar{k}_{2,\mathrm{GHM}}^{(1)} \\ \Delta x^{(1)} \end{bmatrix} = \begin{bmatrix} -3.55 \\ -1.80 \\ 0.00 \end{bmatrix}. \tag{6.63b}$$

No standard deviations of the observations are given in Lenzmann and Lenzmann (2004), but they are assumed to be independent and identically distributed. Therefore, instead of the VCV matrix $\boldsymbol{\Sigma}$ the cofactor matrix $\boldsymbol{Q} = \boldsymbol{I}$ was used. Solving (6.63) yields

$$\begin{bmatrix} \bar{k}_{1,\mathrm{GHM}}^{(1)} \\ \bar{k}_{2,\mathrm{GHM}}^{(1)} \\ \Delta x^{(1)} \end{bmatrix} = \begin{bmatrix} -1.12 \\ 0.44 \\ 0.21 \end{bmatrix}. \tag{6.64}$$

Computing the search direction p as stated in line 5 yields

$$p^{(1)} = \Delta x^{(1)} - \Delta x^{(0)} = 0.21. \tag{6.65}$$

According to (3.54b) and line 7 this results in a maximal feasible step length of

$$q^{(1)} = \frac{\boldsymbol{v}^{(0)} - \boldsymbol{V}^{(0)T} \Delta x^{(1)}}{\boldsymbol{V}^{(0)T} p^{(1)}} = \frac{0.20 - 0}{0.21} = 0.95 \tag{6.66}$$

and thus in the update (line 7)

$$\Delta x^{(1)} = \Delta x^{(0)} + q^{(1)} p^{(1)} = 0 + 0.95 \cdot 0.21 = 0.2. \tag{6.67}$$

An update of the active set (line 8) yields empty matrices $\boldsymbol{V}^{(1)}$ and $\boldsymbol{v}^{(1)}$ and

$$\overline{\boldsymbol{W}}^{(1)} = \begin{bmatrix} 1 \end{bmatrix}, \quad \overline{\boldsymbol{w}}^{(1)} = \begin{bmatrix} 0.2 \end{bmatrix}. \tag{6.68}$$

As the Lagrange multipliers of the active constraints $\bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}}$ are not recomputed at this step (cf. Alg. 6), the algorithm will not terminate, but start a new iteration.

Second Iteration ($i = 2$)

This time, the KKT system (6.42) contains also the matrix of active constraints and therefore reads

$$\begin{bmatrix} \bar{\boldsymbol{B}}_{\mathrm{GHM}}^T \boldsymbol{Q} \bar{\boldsymbol{B}}_{\mathrm{GHM}} & \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{A}^T & \boldsymbol{0} & \overline{\boldsymbol{W}}^{(1)} \\ \boldsymbol{0} & \overline{\boldsymbol{W}}^{(1)T} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{k}}_{\mathrm{GHM}}^{(2)} \\ \Delta x^{(2)} \\ \bar{\boldsymbol{k}}_{\overline{\boldsymbol{w}}}^{(2)} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{w} \\ \boldsymbol{0} \\ \overline{\boldsymbol{w}}^{(1)} \end{bmatrix}, \tag{6.69a}$$

leading to

$$\begin{bmatrix} 2.00 & 0.00 & -6.25 & 0.00 \\ 0.00 & 3.56 & -16.00 & 0.00 \\ -6.25 & -16.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 & 0.00 \end{bmatrix} \begin{bmatrix} \bar{k}_{1,\mathrm{GHM}}^{(2)} \\ \bar{k}_{2,\mathrm{GHM}}^{(2)} \\ \Delta x^{(2)} \\ k_{\overline{\boldsymbol{w}}}^{(2)} \end{bmatrix} = \begin{bmatrix} -3.55 \\ -1.80 \\ 0.00 \\ 0.20 \end{bmatrix}. \tag{6.69b}$$

Solving (6.69) yields

$$\begin{bmatrix} \bar{k}_{1,\mathrm{GHM}}^{(2)} \\ \bar{k}_{2,\mathrm{GHM}}^{(2)} \\ \Delta x^{(2)} \\ k_{\overline{\boldsymbol{w}}}^{(2)} \end{bmatrix} = \begin{bmatrix} -1.15 \\ 0.39 \\ 0.20 \\ -0.90 \end{bmatrix}. \tag{6.70}$$

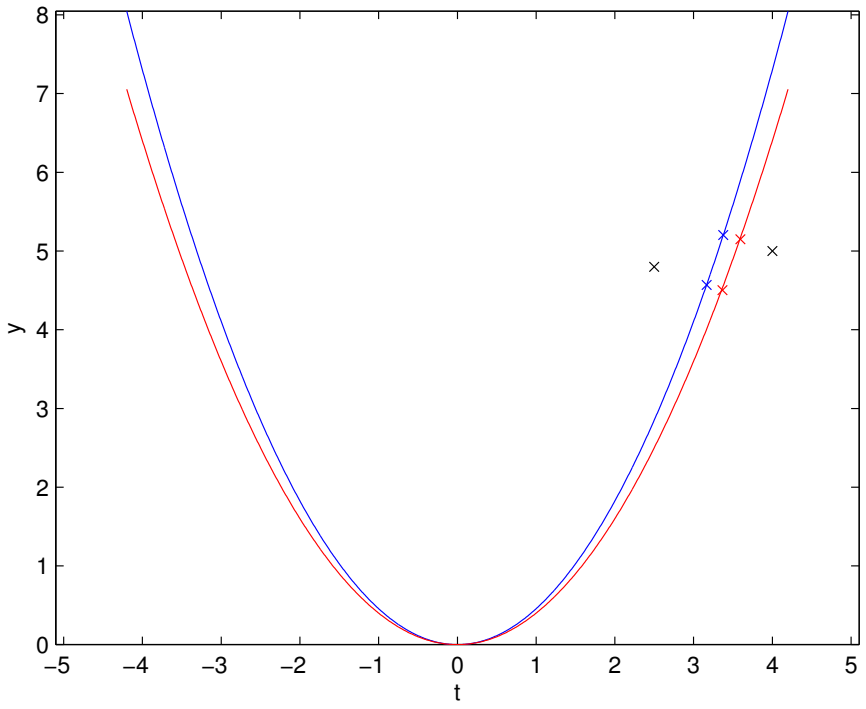


Figure 6.1: Parabola example: Observations (*black crosses*) and parabolas and adjusted observations estimated in the unconstrained GHM (*blue curve* and *crosses*, respectively) and the inequality constrained GHM (*red curve* and *crosses*, respectively). It is clearly visible that the model (a parabola without translational term) was chosen in a way that large residuals result.

Computing the search direction p as stated in line 5 yields

$$p^{(2)} = \Delta x^{(2)} - \Delta x^{(1)} = 0. \tag{6.71}$$

As a consequence, no step in any direction would be made in this iteration. As furthermore the Lagrange multiplier of the only active inequality constraint $k_{\overline{\boldsymbol{w}}}^{(2)}$ is negative, the algorithm terminates (line 9–11 of Alg. 6) with the result

$$\widetilde{x} = x_0 + \Delta x = 0.2 + 0.2 = 0.4. \tag{6.72}$$

Now the quantities $\boldsymbol{w}_0, \overline{\boldsymbol{B}}_{\mathrm{GHM}}, \boldsymbol{A}, \boldsymbol{w}$ and all matrices and right-hand-sides of constraints which depend on $\widetilde{x}$ or $\widetilde{\boldsymbol{\ell}}$ are recomputed and the modified active-set method is started again (cf. Alg. 7). However, the algorithm terminates after one iteration as the initial values are already optimal. Hence, they are the solution of the original problem. The residuals read

$$\boldsymbol{v} = \boldsymbol{Q}\overline{\boldsymbol{B}}_{\mathrm{GHM}}\bar{\boldsymbol{k}}_{\mathrm{GHM}} = \begin{bmatrix} 0.87 \\ -0.40 \\ -0.30 \\ 0.15 \end{bmatrix} \tag{6.73}$$

and thus

$$\widetilde{\boldsymbol{\ell}} = \boldsymbol{\ell} + \boldsymbol{v} = \begin{bmatrix} 3.37 \\ 3.60 \\ 4.50 \\ 5.15 \end{bmatrix} \tag{6.74}$$

follows. The observations (*black crosses*) as well as the parabolas and adjusted observations estimated in the unconstrained GHM (*blue curve* and *crosses*, respectively) and the inequality constrained GHM (*red curve* and *crosses*, respectively) are depicted in Fig. 6.1.

Part III

Simulation Studies

7. Applications

In this chapter a variety of applications is described, in which the methodology developed in the last chapters proves useful. This could either mean that an alternative to classic approaches is shown (e.g., the Huber estimator with inequality constraints in Sect. 7.1.1) or that the proposed solution provides more features than existing inequality constrained approaches (e.g., the stochastic description of the Very Long Baseline Interferometry (VLBI) estimates in Sect. 7.2.2). In case we have already published the example only a brief description is given and the reference to the original article is provided. The aim is primarily to show that inequalities are relevant to numerous geodetic applications.

The Huber estimator with inequality constraints was developed by Mangasarian and Musicant (2000). In Sect. 7.1 their approach is reviewed and introduced to geodesy as an example for the robust estimation with inequality constraints. Two applications for the stochastic framework provided in Chap. 4 are described in Sect. 7.2: The estimation of the coefficients of a positive definite covariance matrix and of the tropospheric zenith wet delay from VLBI observations.
Two examples for problems with rank-deficient systems (cf. Chap. 5) are given in Sect. 7.3: The second order design of a geodetic network and an engineering problem in which strict welding tolerances must not be exceeded.
An example for the use of an ICGHM (cf. Chap. 6) is provided in Sect. 7.4: Here, the vertical height gradient of a planned street is to be estimated. As there are constraints (e.g., concerning the maximal slope), inequalities come into play.

7.1 Robust Estimation with Inequality Constraints

In the following section, an ansatz from Mangasarian and Musicant (2000) is shown that allows to reformulate the robust Huber estimator (cf. Sect. 2.3) as a QP in standard form. This has two advantages:

1. Usually, the Iteratively Reweighted Least-Squares (IRLS) procedure is used to compute a solution with the Huber estimator. However, Neitzel (2004, p. 64) showed (for the L_1 case) that this approach can sometimes lead to a wrong solution. If the Huber estimator is formulated as a QP, any standard QP solver can be used instead of the IRLS approach.
2. It is possible to introduce equality and inequality constraints.

Linear Programs are probably the most prominent among all robust inequality constrained estimators. However, as this thesis focuses on QPs instead of LPs, they will not be treated here. There exists a rich literature on linear programming. Good starting points for the interested reader are the fundamental works of Dantzig (1998) und Dantzig and Thapa (2003) for the Simplex approach for LPs or e.g. Boyd and Vandenberghe (2004) for interior-point methods in LPs and general convex programs.

7.1.1 Huber Estimator

In contrast to e.g., the Hampel estimator (cf. Sect. 2.3), the loss function of the Huber estimator

$$\rho_{\text{Huber}}(v) = \begin{cases} \frac{1}{2}v^2, & |v| \leq k \\ k|v| - \frac{1}{2}k^2, & |v| > k \end{cases} \tag{7.1}$$

(identical to (2.87)) is convex. However, this composite function in only once continuously differentiable. As the objective function of a QP has to be twice differentiable (cf. Sect. 2.1.2), it is not straightforward to reformulate the Huber estimator as a QP in standard form. We will follow the line of thought of Mangasarian and Musicant (2000) (and later Jabr (2005)) who reformulate the loss function of the Huber estimator as a convex and twice differentiable function.

7.1.1.1 Expressing the Huber Estimator as Standard QP

Mangasarian and Musicant (2000) stated that the function

$$\rho(v) = \min\left\{ f(v,y) = \frac{1}{2}y^2 + k|v-y| : y \in \mathbb{R} \right\} \tag{7.2}$$

is equivalent to the loss function (7.1) of the Huber estimator. y is an additional optimization variable. A proof of the statement is provided in Appendix A.3. The objective function (cf. Sect. 2.3) corresponding to loss function (7.2) reads

$$\min\left\{ \phi(\boldsymbol{v},\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{y} + k||\boldsymbol{v}-\boldsymbol{y}||_1 : \boldsymbol{v} \in \mathbb{R}^n, \boldsymbol{y} \in \mathbb{R}^n \right\}. \tag{7.3}$$

Expressing the residuals $\boldsymbol{v}$ as a function of the parameters $\boldsymbol{x}$

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{\ell}$$

leads to

$$\min\left\{ \phi(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{y} + k||\boldsymbol{A}\boldsymbol{x}-\boldsymbol{\ell}-\boldsymbol{y}||_1 : \boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{y} \in \mathbb{R}^n \right\}. \tag{7.4}$$

In a next step, the L_1 norm part of the composite objective function (7.4) shall be expressed via side conditions. Thus, an additional variable $t \in \mathbb{R}^n$ is introduced and a transformation of parameters is performed, yielding the QP

HUBER ESTIMATOR AS QUADRATIC PROGRAM

objective function:	$\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y} + k\ \mathbf{1}^T\boldsymbol{t} \ldots \text{Min}$
constraints:	$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{\ell} - \boldsymbol{y} \le \boldsymbol{t}$ $-\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\ell} + \boldsymbol{y} \le \boldsymbol{t}$
optim. variable:	$\boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{y} \in \mathbb{R}^n, \boldsymbol{t} \in \mathbb{R}^n.$

(7.5)

With the following substitutions

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{t} \end{bmatrix}, \qquad \boldsymbol{B}^T = \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{I} & -\boldsymbol{I} \\ -\boldsymbol{A} & \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{\ell} \\ -\boldsymbol{\ell} \end{bmatrix} \tag{7.6}$$

the Huber estimator can be expressed as a QP in standard form

HUBER ESTIMATOR AS QP IN STANDARD FORM

objective function:	$\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y} + k\ \mathbf{1}^T\boldsymbol{t} \ldots \text{Min}$
constraints:	$\boldsymbol{B}^T\boldsymbol{z} \le \boldsymbol{b}$
optim. variable:	$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{x}^T & \boldsymbol{y}^T & \boldsymbol{t}^T \end{bmatrix}^T \in \mathbb{R}^{m+2n}.$

(7.7)

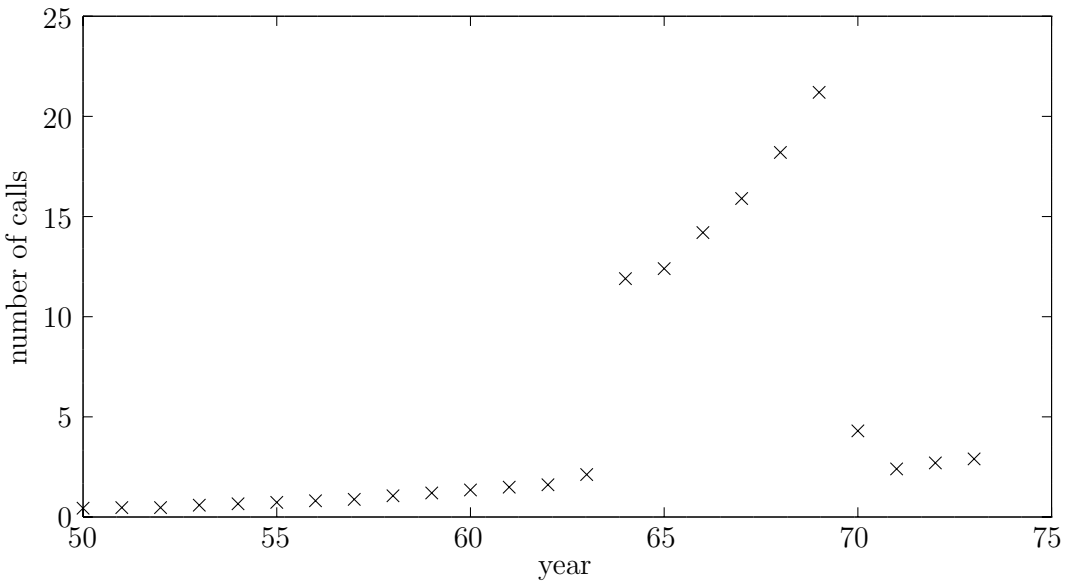


Figure 7.1: Number of international phone calls (*black crosses*) from Belgium.

As a consequence, it is not only possible to use any QP solver for the estimation with the Huber estimator, but also to apply additional equality and inequality constraints. Hitherto it is assumed that the observations are independent and identically distributed. This is because in its current state, the algorithm is only capable of dealing with uncorrelated observations and a decorrelation in presence of outliers is not straightforward.

In the following section the reformulated Huber estimator is applied to a real data example.

7.1.1.2 Data

We will use a classic data set from robust statistics, provided e.g., by Rousseeuw and Leroy (2003, p. 25–26). Plotted in Fig. 7.1 are the observed total number (in tens of millions) ℓ_i of international phone calls from Belgium versus the year t_i. Clearly, the observations from the years 1964 to 1969 do not match the rest of the data. Rousseeuw and Leroy (2003, p. 25–26) stated that the reason is that in these years another recording system was used that counts the total number of *minutes* instead of the total number of *calls*. However, if we want to estimate how the call characteristic develops over time, we must either eliminate these values or use an estimator that is not too much influenced by outliers—e.g., the Huber estimator described above.

We adjust a straight line to the data in a robust way yielding the functional model

$$\ell_i + v_i = x_1 t_i + x_2. \tag{7.8}$$

The observables $\ell_1 \dots \ell_n$ are assumed to be independent and identically distributed.

7.1.1.3 Results

Figure 7.2 shows the resulting straight lines from applying the L_2 norm estimator (*blue line*), the L_1 norm estimator (*green line*) and the Huber estimator with $k = 1.5$ (*orange line*). The *red line* shows the result of the Huber estimator with the additional constraint that the intercept is not allowed to be less than -6. The additional constraint was introduced to illustrate the capability of the reformulated Huber estimator to incorporate additional side conditions.

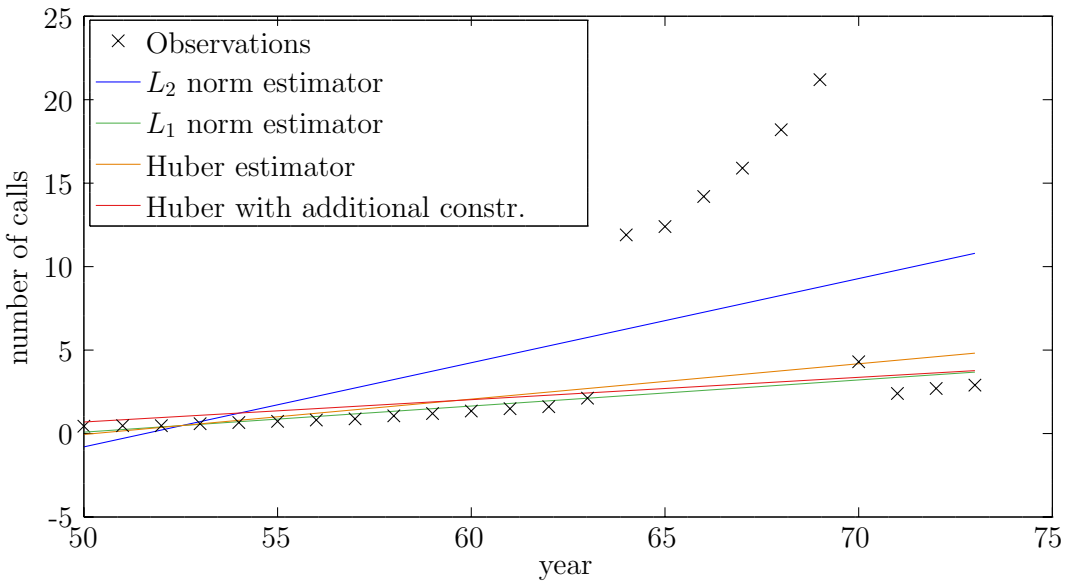


Figure 7.2: Number of international phone calls (*black crosses*) from Belgium and the lines of best fit, estimated with the L_2 norm estimator (*blue line*), the L_1 norm estimator (*green line*) and the Huber estimator (*orange line*). The *red line* shows the result of the Huber estimator with the additional constraint that the intercept is not allowed to be less than -6.

7.2 Stochastic Description of the Results

7.2.1 Estimation of a Positive Definite Covariance Function

This section contains a modified version of the application described in the article *A stochastic framework for inequality constrained estimation* (Roese-Koerner et al., 2012).

The 2D example with dependent and independent constraints described in Sect. 4.5 showed the utility of the stochastic framework developed here. In order to demonstrate the usefulness of the stochastic framework in a more realistic scenario, a multi-dimensional example is chosen. Also, in the multivariate case the importance and benefits of sensitivity analysis becomes more explicit.

7.2.1.1 Data

The multi-dimensional example that will be used here is that of *fitting a positive cosine expansion to a set of observations* (cf. Figure 7.3). The maximum degree of the cosine expansion is $m = 9$, and since it is a positive cosine expansion, the coefficients of the expansion must be positive, which will be guaranteed by the independent inequality constraints enforced on the parameters. Therefore, the observation equation of our model for the i-th observation and the parametric constraints read

$$\ell_i + v_i = \frac{x_0}{2} + x_1 \cos 1\omega t_i + \ldots + x_m \cos m\omega t_i, \tag{7.9a}$$

$$x_j \geq 0, \quad j \in [0\ m], \tag{7.9b}$$

with the period $T = 1$ and the angular frequency $\omega = \frac{2\pi}{T}$. The $n = 50$ supporting points t_i are equally spaced in the interval $[0\ 0.5)$. The $m + 1 = 10$ unknown parameters are denoted $x_0 \ldots x_m$. In matrix form the above equations read

$$\boldsymbol{\ell} + \boldsymbol{v} = \boldsymbol{A}\boldsymbol{x}, \tag{7.9c}$$

$$\boldsymbol{x} \geq 0. \tag{7.9d}$$

Again, the true observations were corrupted by a vector of white noise: $\boldsymbol{\mathcal{E}} \sim N(\mathbf{0}, \boldsymbol{I})$. There will be small correlations between some of the estimated parameters as we have not sampled a complete period. Supporting points t_i and the original observations ℓ_i are listed in Tab. A.1 in Appendix A.4.

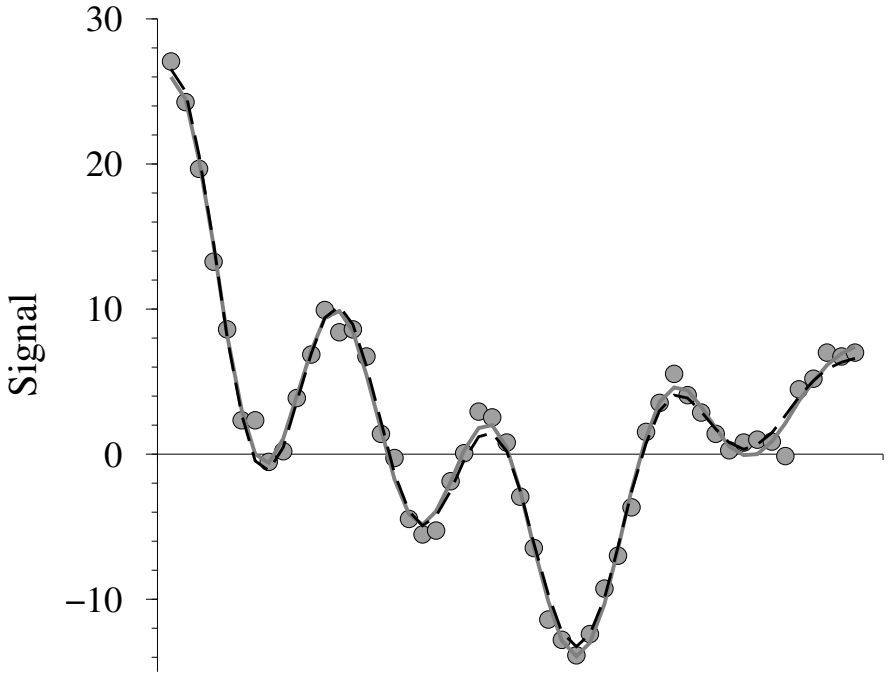


Figure 7.3: Observations (*gray dots*) and the WLS (*gray line*) and ICLS (*black line*) fits to the observations of the positive cosine function estimation problem. Figure taken from Roese-Koerner et al. (2012).

7.2.1.2 Results

The results from the unconstrained estimation show that there are three parameters (x_3, x_4 and x_7) that have negative values. As a consequence of the introduction of the constraints and the application of the stochastic framework described in Alg. 3, these three parameters take the value of zero in the ICLS estimate (cf. Table 7.1, left-hand-side table). The estimates of these parameters have been moved to the boundary not just by the respective independent constraints but also by the other active constraints. Again, this is due to the correlation between the parameters, as encountered in the 2D case (cf. (4.25)).

Furthermore, the largest Lagrange multiplier is that of $\widetilde{k}_7$, and the corresponding absolute change Δx_7 to the parameter x_7 is the largest as well, which is again explained by (4.25). Figure 7.4 shows the range of joint and marginal PDFs that can be expected from a multi-dimensional estimation problem. At least one constraint was active in every estimate. Thus, the probability mass in the infeasible area (cf. Sect. 4.3.2) is $d = 100\,\%$. However, the Wald test (cf. Sect.4.3.1) yields a test statistic of 1.98, which is less than the critical value of 2.08. This is much closer to the critical value in comparison to the previous examples in Sect. 4.5. However, the data still supports the constraints.

Although the method is based on the empirical Monte Carlo approach, substantial quality information can already be obtained from the quadratic programming algorithms themselves. For example, the Lagrange multipliers can be used to carry out a sensitivity analysis, which clearly demarcates the active and inactive constraints and their respective contributions to the perturbation of all the parameters.

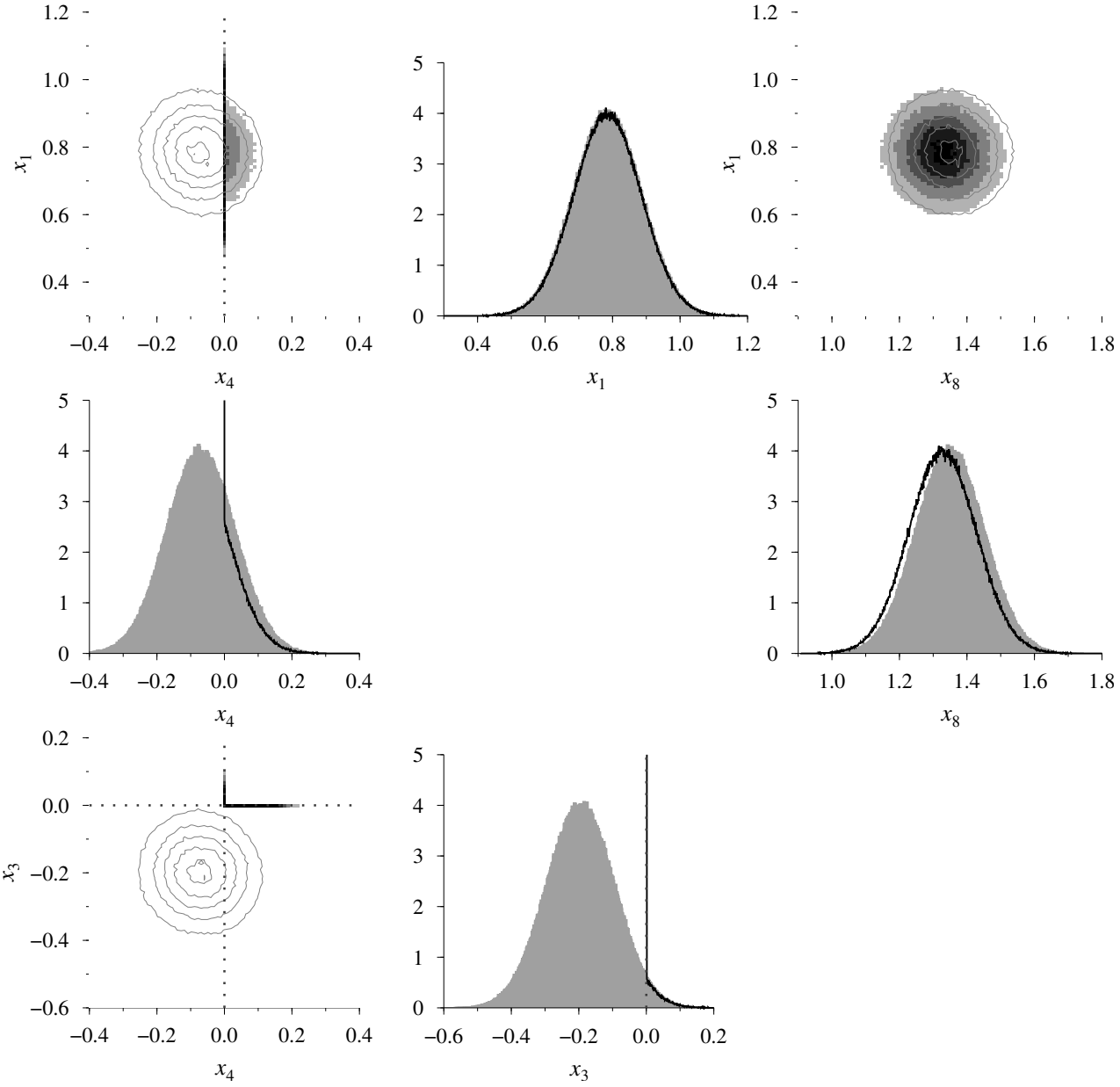


Figure 7.4: Probability density functions (joint and marginal) of the positive cosine function estimation problem after 1 000 000 Monte Carlo iterations. Three different pairs of joint PDFs are shown: x_3 & x_4, x_4 & x_1 and x_1 & x_8. The *contour lines* are from the WLS estimate and the *dotted lines* are the constraints. These pairs show the range of joint PDFs that can be expected from a multi-dimensional estimation problem. While some joint PDFs are entirely in the feasible region, some of them are either partly or entirely in the infeasible region. Due to the negligible correlation between the parameters, all the WLS joint PDFs are nearly circular. Figure taken from Roese-Koerner et al. (2012).

Table 7.1: Results from the ICLS estimation and sensitivity analysis for fitting a positive cosine function. The table on the left shows the estimates from WLS ($\hat{\boldsymbol{x}}$) and ICLS ($\tilde{\boldsymbol{x}}$), and the table on the right shows the sensitivity analysis based on the Lagrange multipliers ($\widetilde{\boldsymbol{k}}$). Sensitivity analysis is only shown for the active constraints as the contributions from inactive constraints ($\widetilde{k}_0$, $\widetilde{k}_1$, $\widetilde{k}_2$, $\widetilde{k}_5$, $\widetilde{k}_6$, $\widetilde{k}_8$ and $\widetilde{k}_9$) are always zero. In all estimates at least one constraint is active ($d = 100\,\%$). Nonetheless, the Wald test yields a test statistic of 1.98, which is less than the critical value of 2.08. Table taken from Roese-Koerner et al. (2012).

	$\hat{\boldsymbol{x}}$	$\tilde{\boldsymbol{x}}$
x_0	2.318	2.274
x_1	0.785	0.785
x_2	2.095	2.073
x_3	−0.195	0
x_4	−0.070	0
x_5	0.637	0.637
x_6	0.443	0.421
x_7	−0.362	0
x_8	1.350	1.328
x_9	1.039	1.039

	$\widetilde{k}_3 = 38.959$	$\widetilde{k}_4 = 18.534$	$\widetilde{k}_7 = 72.252$	$\Delta\boldsymbol{x}$
Δx_0	−0.016	0.002	−0.030	−0.044
Δx_1	0.001	−0.004	0.003	0
Δx_2	−0.008	0.001	−0.015	−0.022
Δx_3	**0.196**	−0.004	0.003	0.195
Δx_4	−0.008	**0.093**	−0.015	0.070
Δx_5	0.001	−0.004	0.003	0
Δx_6	−0.008	0.001	−0.015	−0.022
Δx_7	0.002	−0.004	**0.364**	0.362
Δx_8	−0.008	0.001	−0.015	−0.022
Δx_9	0.001	−0.004	0.003	0

ICLS estimates : $d = 100\,\%$, Wald test: $1.98 < 2.08$

7.2.2 Constraints on the Tropospheric Delay in a VLBI Setting

This example was already published in the article *Improved Parameter Estimation of Zenith Wet Delay Using an Inequality Constrained Least Squares Method* (Halsig et al., 2015).

It is crucial for space geodetic techniques such as Global Navigation Satellite Systems (GNSS) or VLBI, to model the path of the signal through the atmosphere as precisely as possible. Due to refractivity variations in the neutral atmosphere, the signal is delayed. It is of particular importance to account for the influence of the lower part of the atmosphere—the troposphere—which otherwise would be a main error source. This total Slant Troposphere Delay (STD) can be subdivided into two parts: a hydrostatic component and a wet component. These components can be further split up into the product of a specific mapping function $mf_h(e)$ and $mf_w(e)$, respectively, and the Zenith Hydrostatic Delay (ZHD) and the Zenith Wet Delay (ZWD), respectively, yielding

$$\text{STD} = mf_h(e)\ \text{ZHD} + mf_w(e)\ \text{ZWD}. \tag{7.10}$$

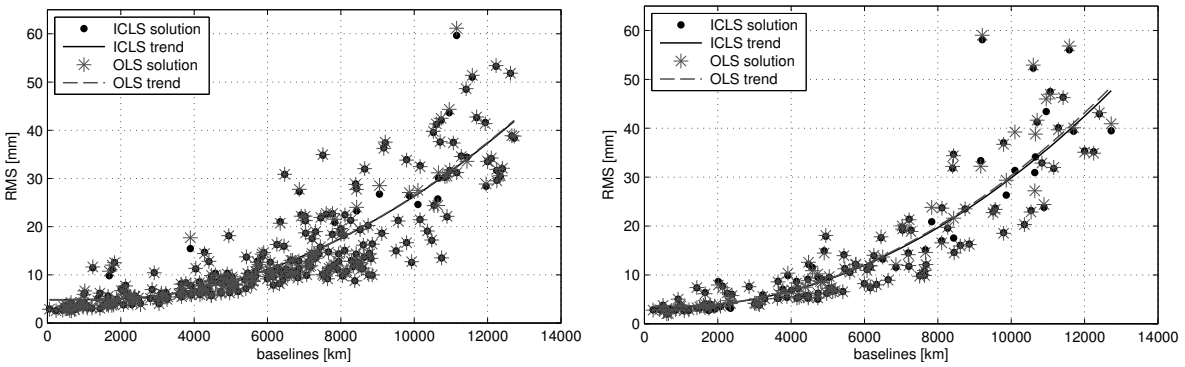
Both mapping functions depend on the elevation angle e. While the ZHD is usually modeled using its relationship to surface air pressure, the ZWD is often estimated in the VLBI analysis. This is because it strongly depends on the water vapor content in the atmosphere, which is subject to high variations. It sometimes occurs in the adjustment process that negative ZWD values are estimated. From a physical point of view those negative values are not possible as that would imply negative water vapor values, too. This phenomenon occurs most often at stations in cold regions. Here, one would expect small positive ZWD values as there is not much water vapor at temperatures below 0 °C. If we assume that these implausible ZWD values occur because unmodeled non-tropospheric effects are absorbed by the ZWDs (Halsig et al., 2015), it should be possible to improve the estimation results by introducing non-negativity constraints to the ZWDs. As there are correlations between ZWDs and the target parameters of a VLBI adjustment (e.g., terrestrial positions of the stations or Earth orientation parameters) those will be affected, too.

7.2.2.1 Data

We use 2333 VLBI databases from 1993 to 2010 provided by the International VLBI Service for Geodesy and Astrometry (IVS). A classic VLBI parameterization was used: The positions of the radio sources were fixed and terrestrial positions, Earth orientation parameters, clock behavior and tropospheric delays are estimated. More details on the parameterization and the models used are provided in Halsig et al. (2015). The only inequality constraints that were introduced are non-negativity restrictions for the ZWDs.

7.2.2.2 Results

Global Results. In 454 of the 2333 databases at least one constraint was active, meaning that the occurrence of at least one negative ZWD value was prevented. As a measure for the improvement through the inequalities, the baseline repeatabilities were computed. Thus, a linear trend was fitted to all repeated measurements of one baseline, as a linear behavior of the station coordinates in time is expected. The resulting Root Mean Square (RMS) of the computation for each baseline that is measured at least 30 times is shown in Fig. 7.5(a). Results from the solution with inequality



(a) Repeatabilities for all baselines that are measured at least 30 times.

(b) Repeatabilities for baselines from sessions with at least one active constraint.

Figure 7.5: Baseline repeatabilities of an OLS (*gray asterisks*) and an ICLS (*black dots*) estimate. Figures taken from Halsig et al. (2015).

constraints (ICLS) are shown as *black dots*. Results from an unconstrained estimate (OLS, *gray asterisks*) are shown for comparison. A quadratic polynomial was fitted to the data and is represented as a *solid line* (ICLS) and a *dashed line* (OLS), respectively. As only about 20 % of the sessions were affected by the constraints, we show in Fig. 7.5(b) the results for the baselines of sessions with at least one active constraint. Here, it can be seen that the trend of the ICLS estimates increases slightly less than that of the OLS estimates.

For a better comparison the differences between the OLS and the ICLS solution are depicted in Fig. 7.6. While the RMS of 9 % of the baselines could be improved through the non-negativity constraints for at least 1 mm (*black bars*), 90 % remain unchanged (*light gray bars*) and 1 % worsened (*dark gray bars*).

Station Specific Results. Negative ZWD values were often estimated in cold regions as the true value would be close to zero. Thus, we chose the station in Ny Ålesund (Spitzsbergen, Norway, 17.01.02) to demonstrate the effect of the new constraints. The OLS (*gray asterisks*) and ICLS (*black dots*) ZWD estimates and their corresponding HPD intervals (cf. Sect. 4.2.3) are depicted in

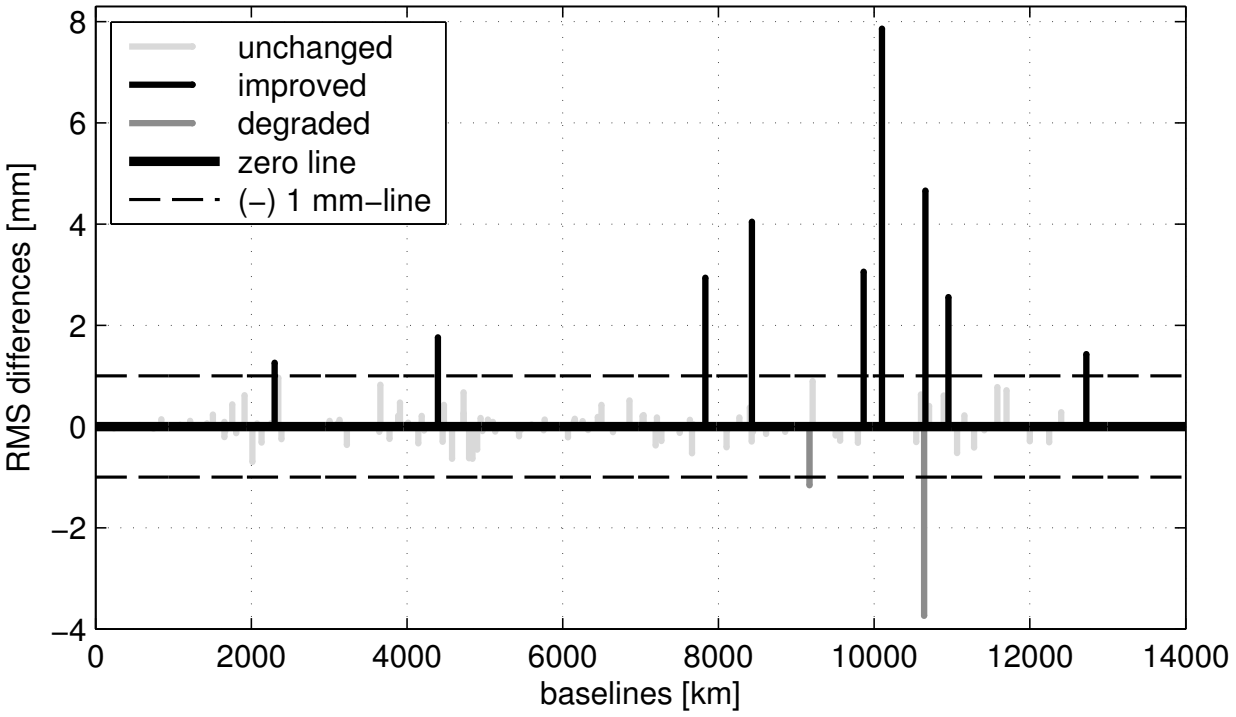


Figure 7.6: Differences between the OLS and the ICLS solution. Figure taken from Halsig et al. (2015).

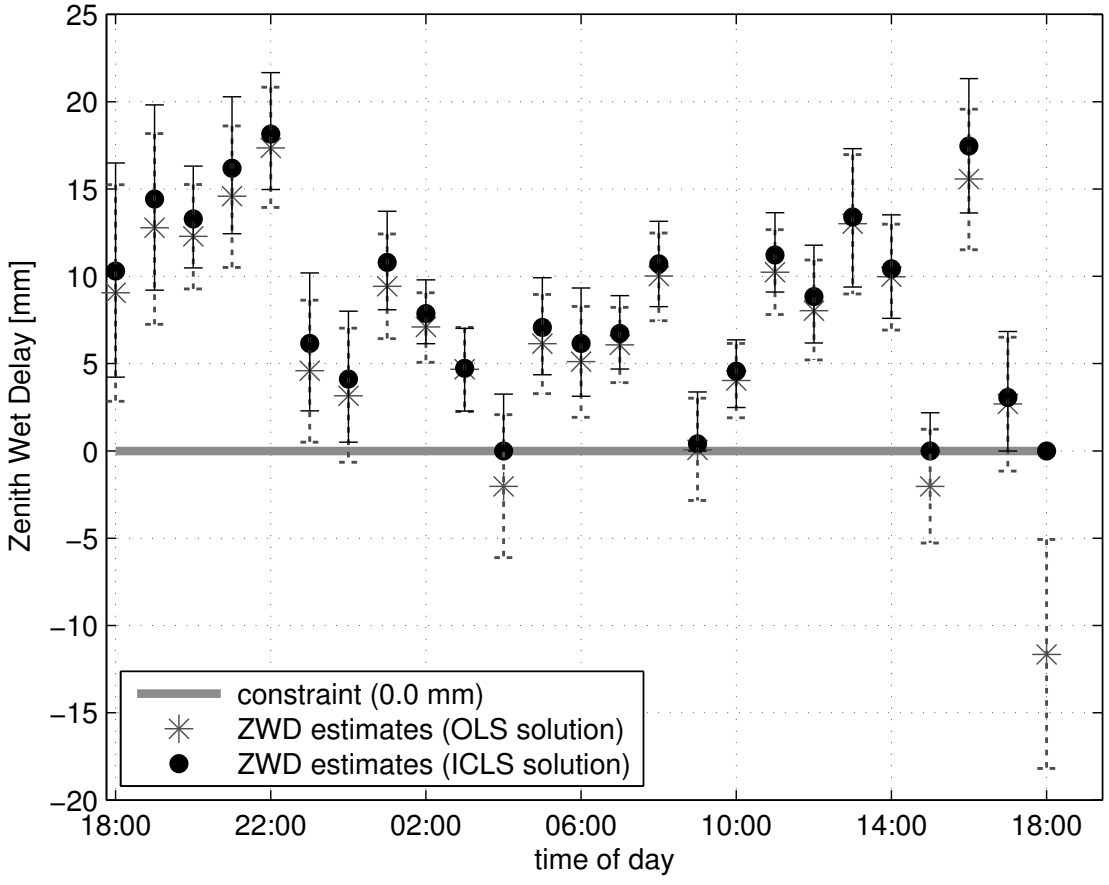
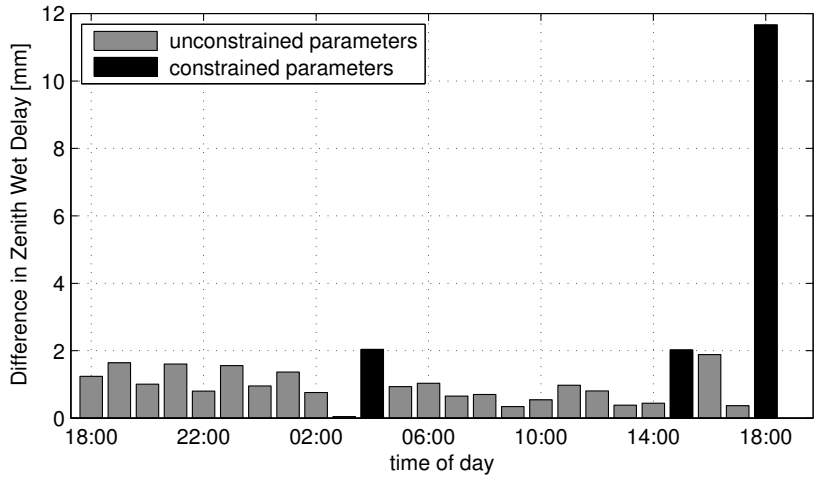
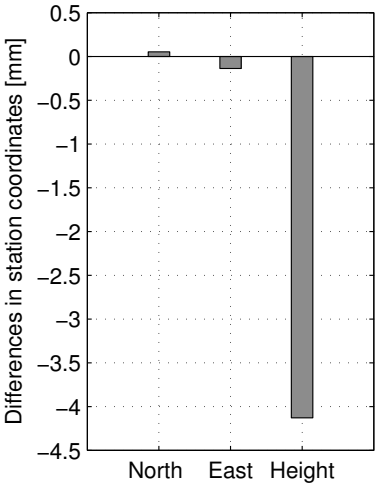


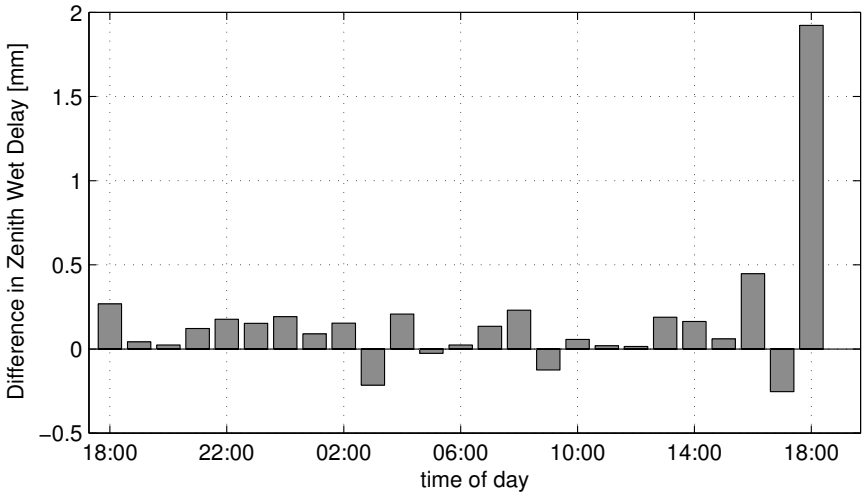
Figure 7.7: OLS (*gray asterisks*) and ICLS (*black dots*) ZWD estimates from Ny Ålesund and their corresponding HPD intervals. Figure taken from Halsig et al. (2015).



(a) Ny Ålesund: ZWD differences. *Black bars* indicate active constraints.



(b) Ny Ålesund: Coordinate differences due to constraints on the ZWDs.



(c) Wettzell: ZWD differences due to constraints in Ny Ålesund.

Figure 7.8: Differences between the OLS and the ICLS estimates. Note the different scales of the bar plots. Figures taken from Halsig et al. (2015).

Fig. 7.7. Due to the positive correlations between the ZWDs, not only the values of the negative ZWDs change, but also the values of all other ZWDs increase. The difference of both estimates is depicted in Fig. 7.8(a). A *black bar* indicates that the corresponding non-negativity constraint becomes active.

The same influence can be seen in the station coordinates of Ny Ålesund shown in Fig. 7.8(b). Here, the largest influence was measured in the vertical station component. It should be noted that there is an effect on the parameters of the other stations, too. However, an evaluation of this effect through a sensitivity analysis (cf. Sect. 4.3.3) shows that it is much smaller than the effect on the "constrained station" itself. For example, the influence of the constraints on the ZWDs at Ny Ålesund on the ZWDs estimated at the station in Wettzell (Bavarian Forest, Germany) is depicted in Fig. 7.8(c).

Due to the results shown above, it seems promising to use inequality constraints to model physical boundaries in the VLBI adjustment. However, the constraints have to be carefully applied and—up to now—are not used in operational VLBI analysis. This is because it is not always possible to ade-

quately model the ZHD component (e.g. because of incomplete pressure data). In an unconstrained estimate, the influence of mis-modeled ZHDs is absorbed in the corresponding ZWDs. This would not necessarily be possible for the ICLS estimate. Thus, the inequalities should only be applied, if it is guaranteed that the ZHDs are modeled with sufficient precision.

7.3 Applications with Rank-Deficient Systems

7.3.1 Second Order Design (SOD) of a Geodetic Network

This section contains a modified version of the application described in the article *Convex optimization under inequality constraints in rank-deficient systems* (Roese-Koerner and Schuh, 2014).

The design of a geodetic network is a classic optimization task in geodesy, which received a lot of attention in the past (cf. Grafarend and Sansò, 1985) and still is a topic of ongoing research (cf. Dalmolin and Oliveira, 2011, Doma and Sedeek, 2014). We focused on the SOD of a direction and distance network as there it is most likely for a rank defect to appear. However, the same methodology can be applied for the SOD of a GNSS network, too. See e.g., Yetkin et al. (2011), Doma (2013) or Mehrabi and Voosoghi (2014) for approaches to determine an optimal set of GPS baselines in an horizontal network.

7.3.1.1 Problem Description

Aim of the SOD of an existing geodetic direction and/or distance network is to determine optimal observation weights p_i in order to develop an observation plan that fulfills some predefined optimality criteria. Usually, one tries to design a network in a way that the estimated coordinates are of a similar accuracy and have only small correlations. Variances and covariances can be described via the cofactor matrix $\boldsymbol{Q}\{\widetilde{\boldsymbol{\mathcal{X}}}\}$ of the estimated parameters. Therefore, a link to the matrix of observation weights $\boldsymbol{P}$ can be established as

$$\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A} = \boldsymbol{Q}\{\widetilde{\boldsymbol{\mathcal{X}}}\}^{-1}. \tag{7.11}$$

$\boldsymbol{Q}\{\widetilde{\boldsymbol{\mathcal{X}}}\}$ can be replaced by a *target cofactor matrix*. This criterion matrix $\boldsymbol{Q}_{\boldsymbol{x}}$ contains the specification of an optimal cofactor matrix (e.g., uncorrelatedness and similar variances) which serve as observations. Classic choices for the criterion matrix are either an identity matrix or a matrix of Taylor-Karman type (cf. Grafarend and Schaffrin, 1979). Using the Khatri-Rao product $\odot$ (7.11) can be rewritten in matrix vector notation

$$(\boldsymbol{A}^T \odot \boldsymbol{A}^T)\boldsymbol{p} \stackrel{!}{=} \boldsymbol{q}_{\text{inv}}. \tag{7.12}$$

The vectors $\boldsymbol{p}$ and $\boldsymbol{q}_{\text{inv}}$ contain the entries of $\boldsymbol{P}$ and $\boldsymbol{Q}_{\boldsymbol{x}}^{-1}$, respectively. Using the substitutions

$$\boldsymbol{M} := \boldsymbol{A}^T \odot \boldsymbol{A}^T, \tag{7.13}$$

the L_2-norm objective function can be formulated as

$$\Phi(\boldsymbol{p}) = \left((\boldsymbol{A}^T \odot \boldsymbol{A}^T)\boldsymbol{p} - \boldsymbol{q}_{\text{inv}}\right)^T \left((\boldsymbol{A}^T \odot \boldsymbol{A}^T)\boldsymbol{p} - \boldsymbol{q}_{\text{inv}}\right) \tag{7.14a}$$

$$= \left(\boldsymbol{M}\boldsymbol{p} - \boldsymbol{q}_{\text{inv}}\right)^T \left(\boldsymbol{M}\boldsymbol{p} - \boldsymbol{q}_{\text{inv}}\right). \tag{7.14b}$$

$\boldsymbol{Q}_{\boldsymbol{x}}$ is a symmetric $m \times m$ matrix and all $\frac{m(m+1)}{2}$ elements of its upper right triangle can be used as observations. If the number n of weights p_i that should be determined is less or equal to this length

of $\boldsymbol{q}_{\text{inv}}$, the weights can be estimated using a standard GMM. However, in practice it often occurs that there are more weights to be estimated than entries of the criterion matrix given. Especially, if we have to deal with small networks or with larger networks with many fixed coordinates. In these cases, the system is underdetermined resulting in a rank-deficient normal equation matrix.

There is a direct relationship between repetition number n_i and the corresponding weight

$$p_i = n_i \frac{\sigma_0^2}{\sigma_{\ell_i}^2}. \tag{7.15}$$

σ_0^2 is the variance factor and $\sigma_{\ell_i}^2$ the variance of observation ℓ_i. As negative or huge repetition numbers cannot be realized, box constraints are applied to the weights to ensure that they are nonnegative

$$p_i \geq 0 \tag{7.16}$$

and do not exceed a certain maximal repetition number

$$p_i \leq n_{\max} \frac{\sigma_0^2}{\sigma_{\ell_i}^2}. \tag{7.17}$$

In matrix-vector notation, the constraints read

$$-\boldsymbol{I}\boldsymbol{p} \leq \boldsymbol{0}, \tag{7.18a}$$

$$\operatorname{diag}(\boldsymbol{\Sigma})\,\boldsymbol{p} \leq n_{\max}\sigma_0 \boldsymbol{1}_n. \tag{7.18b}$$

The operator $\operatorname{diag}(\boldsymbol{\Sigma})$ extracts all diagonal elements of the matrix $\boldsymbol{\Sigma}$ and preserves its original dimension. $\boldsymbol{1}_n$ is a vector of length n that contains only Ones. The corresponding optimization problem reads

Example: Second Order Design	
objective function:	$(\boldsymbol{M}\boldsymbol{p} - \boldsymbol{q}_{\text{inv}})^T (\boldsymbol{M}\boldsymbol{p} - \boldsymbol{q}_{\text{inv}}) \ldots \text{Min}$
constraints:	$\begin{bmatrix} -\boldsymbol{I} \\ \operatorname{diag}(\boldsymbol{\Sigma}) \end{bmatrix} \boldsymbol{p} \leq \begin{bmatrix} \boldsymbol{0} \\ n_{\max}\sigma_0 \boldsymbol{1}_n \end{bmatrix}$
optim. variable:	$\boldsymbol{p} \in \mathbb{R}^n$.

Naturally, computation of individual weights for each observation does not yield a final result of the network optimization process. This is, because it is not viable in practice to measure some directions from one standpoint more often than others. Therefore, the estimation of individual weights usually presents the first step of a three step approach. In a second step, measurements with little impact are identified and eliminated. Finally, in the third step, group weights—e.g., for all observations from one standpoint—are estimated (cf. Müller, 1985). However, we focused on the first step only, because it is most likely for a rank defect to appear there.

7.3.1.2 Data

We have applied the framework described in Sect. 5.4 to determine optimal weights for a horizontal network located in the "Messdorfer Feld" in Bonn. The network consists of three fixed datum points (*black triangles*) and eight new points (*black dots*), whose coordinates are to be estimated (cf. Fig. 7.9). All points are located within sight distance from each other, so that directions and

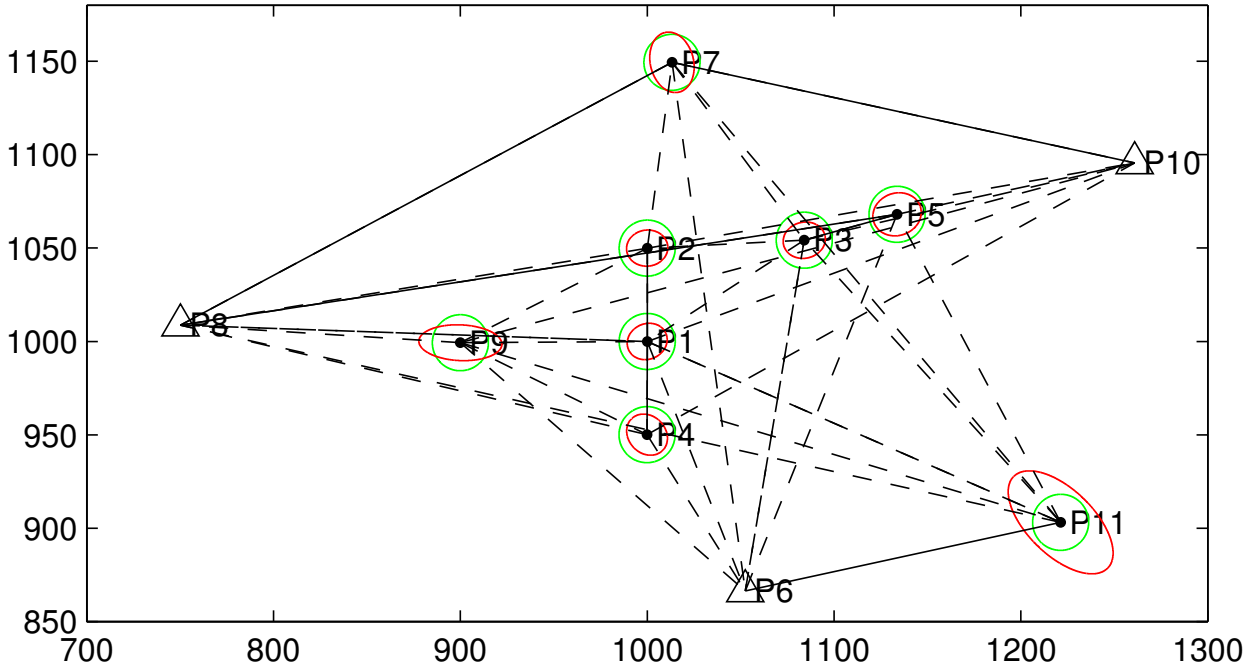


Figure 7.9: SOD of a geodetic network with 3 fixed points (*black triangles*) and 8 new points (*black dots*). Distance measurements are shown as *solid lines* and direction measurements as *dashed lines*. *Green ellipses* depict the optimal error ellipses which are approximated by the *red* ones. Figure taken from Roese-Koerner and Schuh (2014).

distances between all pairs of points could be measured theoretically. A criterion matrix of Taylor-Karman type (cf. Grafarend and Schaffrin, 1979) is chosen, resulting in the target error ellipses plotted in *green*. The dimensions of the network lead to $\frac{16(16+1)}{2} = 136$ entries of the criterion matrix, which serve as observations, and to 162 weights to be estimated, resulting from the 162 possible direction or distance measurements. A tachymeter with an accuracy of

$$\sigma_{\text{dir}} = 0.4\,\text{mgon}, \tag{7.19}$$

$$\sigma_{\text{dist}} = 1\,\text{mm} + 1\,\text{ppm} \tag{7.20}$$

shall be used. σ_{dir} is the assumed accuracy of a direction measurement and σ_{dist} the assumed accuracy of a distance measurement.

7.3.1.3 Results

The network shown in Fig. 7.9 has been designed using the quantities stated above, assuming an arbitrarily chosen maximum repetition number of 50 and applying the presented framework. Since this approach approximates the inverse of the criterion matrix instead of the criterion matrix itself, a factor was computed and used to rescale $\boldsymbol{p}$ to prevent over-optimization (proposed in Müller, 1985). The *red error ellipses* indicate values of the resulting cofactor matrix $\boldsymbol{Q}\{\widetilde{\boldsymbol{\mathcal{X}}}\}$.

As a result of the optimization procedure a total of 146 measurements should be performed (118 direction measurements, *dashed lines*, and 28 distance measurements, *solid lines*). No measurement should be repeated more than 5 times. Furthermore, the introduction of inequalities resolves the rank defect resulting in a unique solution. As expected, the resulting error ellipses of points in the center of the network are more *circle-like* than those of the points on the borders.

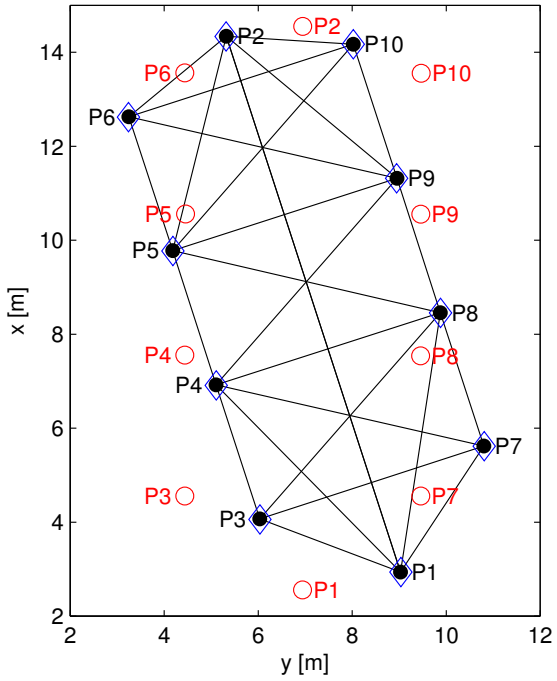


Figure 7.10: Distance measurements (*black lines*) are performed between points $P1$ to $P10$ (*black dots*). A particular OLS solution (*blue diamonds*) and the ICLS solution (*red circles*) with maximal minimal distance to the constraints are shown. Figure taken from Roese-Koerner and Schuh (2015).

7.3.2 Strict Welding Tolerances in a Rank-Deficient Network

This section contains a modified version of the application described in the article *Effects of different objective functions in inequality constrained and rank-deficient least-squares problems* (Roese-Koerner and Schuh, 2015).

7.3.2.1 Problem Description

This case study is based on an engineering problem. We assume that some prefabricated building material shall be fitted between other elements so that the parts can be welded together. In order to make welding possible, tolerances have to be fulfilled strictly.

Figure 7.10 depicts the test case. 26 distance measurements (*black lines*) are performed between the ten points $P1$ to $P10$ (*black dots*). Their 20 coordinates are the parameters to be estimated in a GMM. As no datum is defined, estimating absolute values of the coordinates is a rank-deficient problem.

7.3.2.2 Data

Points $P3$ to $P6$ are located at the left-hand side of the gap the new part is supposed to fill, and the points $P7$ to $P10$ are located on its right-hand side. $P1$ and $P2$ are external points to stabilize the network. It shall be determined if the new part fits between both lines of points. This can be achieved by setting up the 16 linear constraints

$$y_{7,8,9,10} - y_{3,4,5,6} \leq 5.03\,\mathrm{m} \tag{7.21}$$

and the 16 linear constraints

$$y_{7,8,9,10} - y_{3,4,5,6} \geq 5.00\,\mathrm{m}, \tag{7.22}$$

resulting in a rank-deficient ICLS problem in form of (3.8). While the first constraints guarantee that the new part is not allowed to be wider than 5.03 m, the latter ensure that it is not smaller than 5.00 m (otherwise the gap would be too large for welding). The constraints force the estimated points to align almost parallel to the x axis (cf. *red circles* in Fig. 7.10). If more than two of the 32 constraints mentioned above are active, the new part will not meet the tolerances. Incompatible elements can be detected via an analysis of the Lagrange multipliers (cf. Sect. 4.3.3).

If the manifold is not resolved through the introduction of constraints, a nullspace optimization has to be performed. This can be used to maximize the minimal distance to the constraints

$$\Phi_{\mathrm{NS}}^{L\infty} = ||\boldsymbol{B}^T\boldsymbol{x}(\boldsymbol{\lambda}) - \boldsymbol{b}||_\infty \ldots \max. \tag{7.23}$$

Using the Chebyshev norm is always beneficial if tolerances instead of standard deviations are given.

7.3.2.3 Results

The optimization problem was solved using the CVX software (Grant and Boyd, 2014). Results are shown in Fig. 7.10. In the chosen scenario, no constraint is active. Therefore, the new part will fit in the gap and welding is possible.

Figure 7.11 shows the welding boundary for the existing parts (*gray area*), the new part (*black area*) and the "gaps" at its left-hand (Fig. 7.11(a)) and right-hand (Fig. 7.11(b)) side. Note the different scales in x and y direction and the breach in the y axis. Adjusted coordinates of the ICLS estimate with $\Phi_{\mathrm{NS}}^{L\infty}$ (*red circles*) are compared with those of an ICLS estimate with

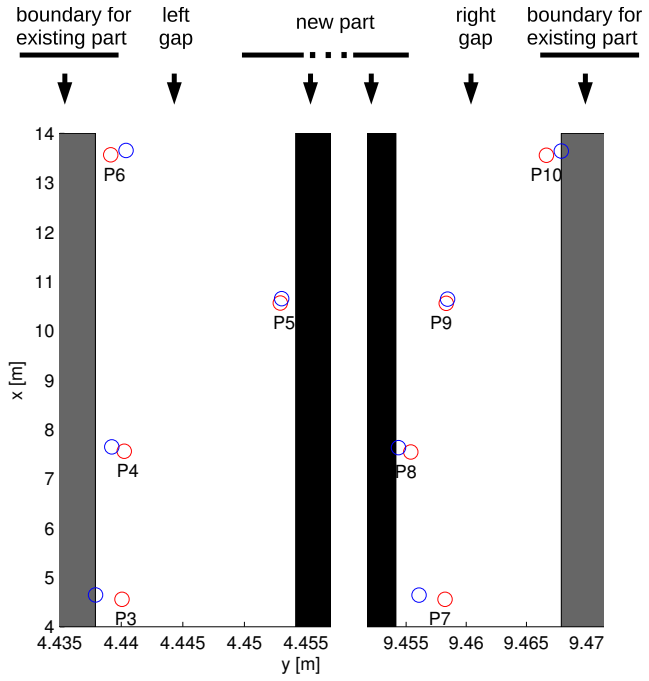
$$\Phi_{\mathrm{NS}}^{L2} = ||\boldsymbol{x}||_2 \ldots \min \tag{7.24}$$

as the nullspace objective function (*blue circles*).

While both estimates provide a decision *if* the new part will fit, only the adjustment with maximal minimal distance to the constraints allows us to determine *how well* the new part will fit. This can be seen in Fig. 7.11, where for this estimate the minimal distance to the constraints is at least 2.5 mm at each side (namely for the points 5, 6, 8 and 10). In contrast, the blue points 3 and 10 are exactly on the boundary. So there is clearly a benefit in choosing a suited objective function for the nullspace optimization.

7.4 The Gauss-Helmert Model with Inequalities

In the following section the design of the profile (i.e. the vertical aspect) of a road is examined. We will see that uncertainties exist in the horizontal as well as in the vertical data, and that explicit relationships between the observations have to be handled. Thus, the Gauss-Helmert model proves useful. During the design, driving dynamics should be considered which could be done by adding additional inequality constraints concerning the maximal feasible slope or its change. Some references for the geometric design of railroads and roads and the role of the ICGHM therein are provided in the next section. However, one could easily imagine applications in other areas that call for the use of the ICGHM.
In cartographic generalization for example, the position of labeling elements (e.g., a text box with the name of a city) should be optimized in a way that it is close to the corresponding map element, but does not hide other important information. This could be parameterized using the ICGHM.
Another example would be the design of switches in rail stations. In order to optimize the maximum speed, the switch should be designed in a way that only minimal differences in curvature appear.



(a) Gap on the left side.

(b) Gap on the right side.

Figure 7.11: Boundary for the existing parts (*gray area*), the new part (*black area*) and the "gaps" at the left-hand (a) and right-hand (b) side of the new part. The axes are scaled differently and there is a breach in the y axis, so most of the new part is not shown. Adjusted coordinates of the above mentioned ICLS estimate with maximal minimal distance to the constraints (*red circles*) are compared with those of an ICLS estimate with Φ_{NS}^{L2} as nullspace objective function (*blue circles*). Figures taken from Roese-Koerner and Schuh (2015)

7.4.1 Optimal Design of a Vertical Road Profile

Although a few three dimensional approaches exist (cf. Kang et al., 2012, Kühn, 2013), up to now the design process of a road is usually split up in three different steps: The design of horizontal alignment, vertical profile und cross section (Vitkiene and Puodziukas, 2014).

First, the horizontal alignment (i.e. the "route of the road") is designed (Koch, 1976). This is mainly due to the fact that many external forces such as land use, environment, preservation of historical and cultural values (Vitkiene and Puodziukas, 2014) limit the freedom of the designer in this step. Afterwards the vertical profile is designed. The main aspects of consideration here are the fit to the topography, safety of the road users and driving dynamic considerations such as average speed and the minimization of sudden slope changes ("ruptures"). As a last step, the cross section is designed according to the estimated number of vehicles that will use this road.

In this section we will focus on the second design step: defining the vertical road profile. Thus, we assume that the first design step is already completed and the horizontal alignment of the road is known. This enables us to sample discrete height values along the route of the road (cf. Fig. 7.12) from a Digital Elevation Model (DEM).

Especially in mountainous regions, the earthwork cost can account for up to 80 % of the total building cost of the road (Kang et al., 2012). Thus, the task is to find the road profile that is optimal in the sense that the differences between the road and the surface of the Earth are minimal. As we assume uncertainties in the vertical as well as in the horizontal component, deviations in

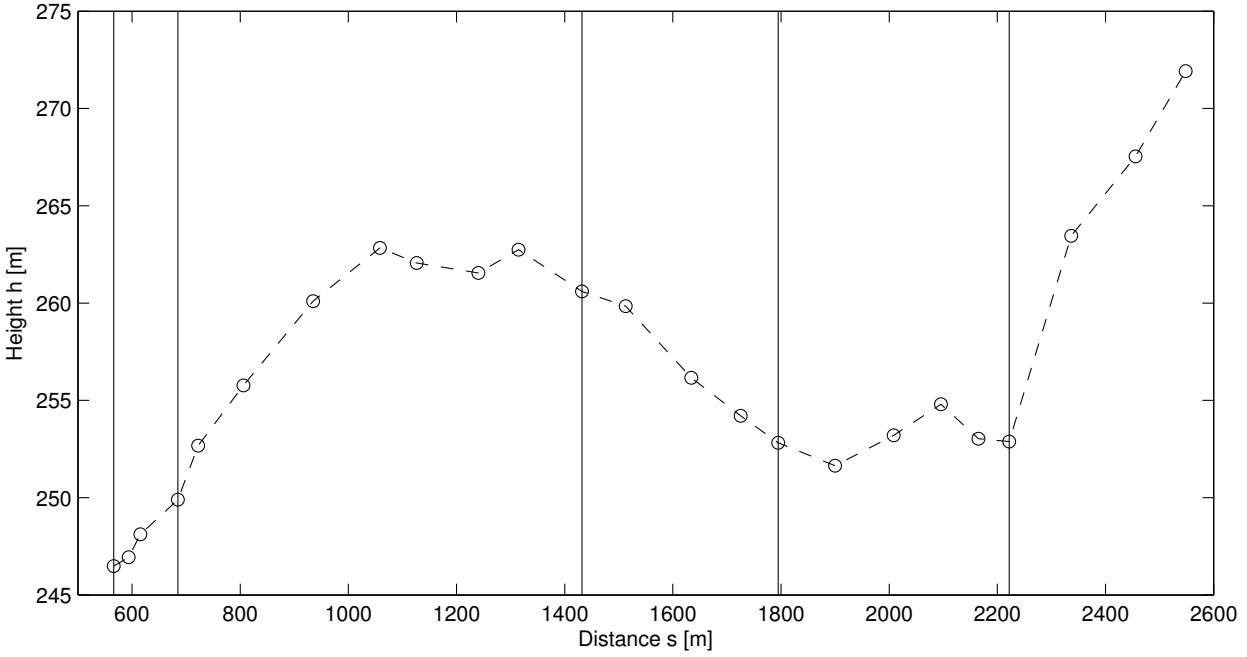


Figure 7.12: Heights h along the route of the road. The transition points of the horizontal road design are shown as vertical lines. Figure modified from Pietzsch (1989, p. 169).

both quantities should be minimized. The German road design standard "Richtlinien für die Anlage von Landstraßen" (RAL, 2012, p. 39) mentions two standard elements that should be used to parameterize the vertical component of a road:

1. Straight lines
2. Arcs to round crests and sags at the transition point of two straight lines (cf. Sect. 7.4.1.1).

The road can then be seen as a piecewise assembled function of these standard elements. Furthermore, several constraints are mentioned in the RAL (2012, p. 20), which have to be met. Their values depend on the average travel speed the road shall be designed for. In this simplified example, we will only introduce two of them and choose to design a road for an average speed of 110 km/h ("Kraftfahrstraße, Entwurfsklasse 1" RAL, 2012, p. 20). Thus, the maximal feasible slope is 4.5 %, the minimal feasible radius of a crest is 8 km, and the minimal feasible radius of a sag is 4 km.

In many textbooks the design is not performed as an adjustment but by simply taking into account the height at as many points as needed to assemble a uniquely defined system (e.g., Pietzsch, 1989, p. 169–170). However, with the ICGHM it is possible to address this task as an adjustment problem resulting in an optimal surface fit.

7.4.1.1 Elements of a Vertical Road Profile

According to RAL (2012, p. 40) the three-dimensional course of the street is essential for driving safety and driving dynamics. To design a road that is as safe and comfortable as possible, the transition points of the vertical design elements should be identical or close to identical to the transition points of the horizontal design. In Fig. 7.12 the transition points of the horizontal road design are shown as vertical lines. Several exceptions and additional design rules are stated in RAL

(2012, p. 40–46). As the idea of this section is simply to show that the ICGHM can be a useful tool for road design and not to design an actual road, we will not pursue them further.

As mentioned in the last section, according to RAL (2012, p.39) the following two design elements should be used to define the vertical profile:

Straight Line. A straight line can for example be parameterized as

$$h_i = h_A + \frac{a}{100}(s_i - s_A). \tag{7.25}$$

The height h_i is given in meter and measured at distance s_i from the beginning of the road. The point $A(s_A, h_A)$ is the starting point of the line segment. a is the slope of the road given in percentage (as it is usual in road design).

Convex Transition Between Gradients. In order to round out the transition points of two straight lines, a segment of a circle would seem appropriate. However, as the minimal radius of such a circle is constrained to be at least several thousand meters (depending on the planned average speed), the RAL (2012, p. 94) states that a polynomial of degree two

$$h_i = h_A + \frac{a_1}{100}(s_i - s_A) + \frac{(s_i - s_A)^2}{2r} \tag{7.26}$$

is a sufficient approximation of a circle. a_1 is the slope of the line on the left side of the polynomial. r is the approximated vertical curve radius (computed as radius of curvature in the peak of the polynomial). In case of a crest—i.e. the straight line on the left has a positive slope a_1 and the straight line on the right has a negative slope a_2—the radius r is positive. In case of a sag—i.e. the straight line on the left has a negative slope a_1 and the straight line on the right has a positive slope a_2—the radius r is negative.

7.4.1.2 Data

The vertical profile of the road should be designed in a way that it is as close as possible to the topography depicted in Fig. 7.12 (modified from Pietzsch, 1989, p. 169–170). As it is assumed that measurements are given at all break points (*black circles*), the points are connected by a *dashed line* to give a rough impression of the terrain. Due to the shape of the landscape we decided to use the following element sequence:

straight line 1 – crest – straight line 2 – sag – straight line 3.

The transition points of the vertical elements should be close to those of the horizontal elements, cf. Sect. 7.4.1.1. As we aim to design a road for an average speed of 110 km/h, according to RAL (2012, p. 20), the following constraints have to be met:

Maximal feasible slope: $-4.5\,\% \leq a_i \leq 4.5\,\%$,
Minimal feasible radius of a crest: $r_c \geq 8\,\text{km}$,
Minimal feasible radius of a sag: $r_s \geq 4\,\text{km}$.

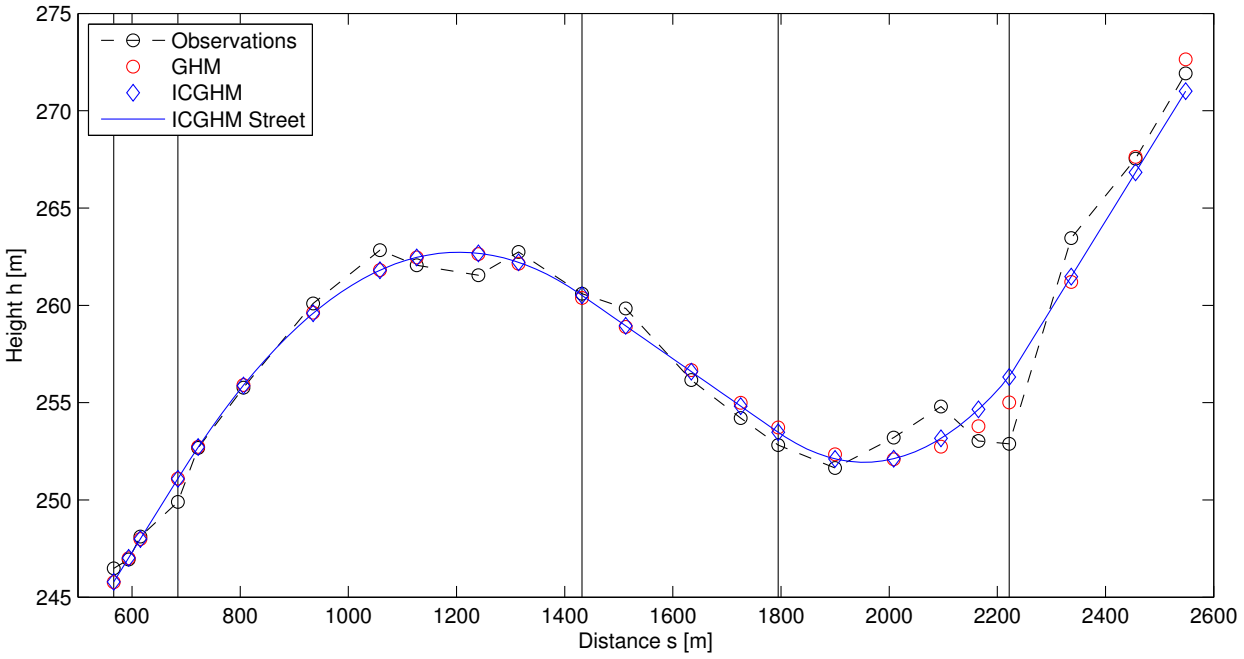


Figure 7.13: Optimal vertical gradient of the road. The adjusted observations from the ICGHM adjustment are shown as *blue diamonds* and the corresponding vertical profile of the road is depicted as *blue line*. The transition points of the horizontal road design are shown as *vertical lines*. The adjusted observations from an unconstrained GHM are shown for comparison (*red circles*).

7.4.1.3 Results

With the data shown in the last sections two adjustments were performed: An unconstrained GHM estimate and an ICGHM estimate, in which the constraints mentioned in Sect. 7.4.1.2 were incorporated. The observed heights h_i and distances s_i are assumed to be uncorrelated and given with a variance of $\sigma_{\mathcal{H}_i} = 0.1\,\mathrm{m}$ and $\sigma_{\mathcal{S}_i} = 1\,\mathrm{m}$, respectively. The results of the unconstrained GHM estimate read

$$a_1^{\mathrm{GHM}} = 4.50\,\%,\ a_2^{\mathrm{GHM}} = -1.83\,\%,\ a_3^{\mathrm{GHM}} = 5.40\,\%,\ r_c^{\mathrm{GHM}} = -11\,\mathrm{km},\ r_s^{\mathrm{GHM}} = 10\,\mathrm{km}$$

and are depicted in Fig. 7.13 as *red circles*. Clearly the slope a_3^{GHM} violates the constraint of a maximal slope of 4.5 %. The results of the ICGHM estimate read

$$a_1^{\mathrm{ICGHM}} = 4.47\,\%,\ a_2^{\mathrm{ICGHM}} = -1.94\,\%,\ a_3^{\mathrm{ICGHM}} = 4.50\,\%,\ r_c^{\mathrm{ICGHM}} = -12\,\mathrm{km},\ r_s^{\mathrm{ICGHM}} = 8\,\mathrm{km}$$

and are depicted in Fig. 7.13 as *blue diamonds*. The corresponding vertical profile of the road is depicted as a *blue line*. These are the results with the smallest value of the objective function that fulfill all design constraints.

Obviously for the design of an actual road, this result would only be a first step. In a next step an examination of the line of sight of the drivers would be performed as well as a test for certain design deficits mentioned in RAL (2012, p. 41–46). This could lead to the redesign of the horizontal and vertical alignments. Furthermore, the designer has to check, if the fit to the topography is sufficient enough or if the constraints should be relaxed—which is only possible in exceptional and justified cases (RAL, 2012, p. 39). A sensitivity analysis as described in Sect. 4.3.3 could be helpful here, to identify those constraints that distort the result most. In addition, one can think of several further constraints such as fixed points in the horizontal and vertical profiles (e.g. fixed transition points to other road segments, tunnel or bridges) that could easily be incorporated using the proposed approach.

Part IV

Conclusion and Outlook

8. Conclusion and Outlook

8.1 Conclusion

Within this thesis concepts from the field of convex optimization were derived and analogies to adjustment theory were pointed out. Furthermore, three major obstacles to the use of inequality constrained estimation approaches in geodesy were identified, possible solutions were proposed, and simulation studies were carried out to verify the approaches.

A **stochastic framework for inequality constrained estimates** was developed in Chap. 4. The MC-QP method allows for the computation of the empirical PDF of a constrained estimate using a Monte Carlo approach. Furthermore, possibly truncated confidence regions were computed. As a result of the projection of the probability mass in the infeasible region onto the boundary of the feasible set, the resulting confidence regions are usually more compact than those from the unconstrained estimate or from the Bayesian method of Zhu et al. (2005).
Thus, the major drawbacks of existing methods for the quality description of inequality constrained estimates are overcome: Our proposed framework is more robust to changes in the set of active constraints than the active-constraints approach of Liew (1976) and yields a more realistic PDF than the Bayesian method of Zhu et al. (2005) as also the information from the probability mass in the infeasible region is used. Furthermore, the empirical PDF inside the feasible region is identical to that of the unconstrained estimate and changes solely appear at its boundary.
Besides the actual quality description the framework provides us with the opportunity to analyse the constraints. On a global scale we suggested to examine if the data support the constraints by using a Wald test and provide a measure for the probability mass in the infeasible region which could be helpful to evaluate the influence of the constraints in general. On a local scale a sensitivity analysis can be performed using the Lagrange multipliers to describe the individual influence of each constraint on each parameter.
One limitation of the MC-QP method is that applying a Monte Carlo method remains a computationally demanding task.

A **framework to treat rank-deficient least-squares problems with inequality constraints** was proposed in Chap. 5. This includes an extension of the binding-direction primal active-set method (along the line of thought of Wong, 2011) that allows us to compute one particular solution despite a possible rank defect.
Furthermore, the rigorous computation of a general solution was discussed. If the solution manifold and the feasible region intersect, the proposed framework enables us to obtain a unique particular solution with certain predefined optimality criteria. Possible choices include a solution vector that has the shortest length of a all vectors in the solution manifold, or a solution in which the minimal distance to the constraints is maximized. While the former can be seen as a kind of Moore-Penrose inverse that takes the inequalities into account, the latter can be helpful to obtain a buffer to the constraints. As this optimization is performed in the nullspace of the design matrix, the value of the original objective function will not change. Thus, a solution with certain specific characteristics can be derived without sacrificing optimization quality.
The direct computation of a particular solution of the original ICLS problem is avoided (if possible) to reduce the computational complexity. Instead, the original problem is split up into three subparts: An unconstrained adjustment in the original parameter space, and a feasibility and an optimization problem, both in a vector space of lower dimensions. As a consequence, the proposed framework is not limited to small-scale problems as the projector theoretical approach of Werner and Yapar (1996).

In addition, a homogeneous solution can be computed and the constraints are reformulated with respect to the free parameters, which allows us to examine the influence of the constraints on the homogeneous solution.

The **Gauss-Helmert model with inequality constraints** was treated in Chap. 6. A straightforward way to reformulate the ICGHM as a quadratic program in standard form was provided. However, we found that this produces a large computational overhead and therefore proposed a tailor-made reformulation which exploits the special structure of the GHM. As a consequence of the reformulation it is no longer possible to use standard QP solvers. Consequently, we derived the KKT optimality conditions for the modified problem and adapted the binding-direction primal active-set method to this structure.

In summary, the findings of this thesis make it much more reasonable (and convenient!) to apply methods from the field of convex optimization to classic geodetic tasks. With the proposed stochastic framework, we no longer lack a quality description of inequality constrained estimates, which is essential for geodetic applications. Furthermore, the framework for rank-deficient ICLS problems and the introduction of the ICGHM allow for the formulation of more sophisticated models, which was not possible before. Thus, we are no longer limited to use solely standard models. In addition, the reformulation of the Huber estimator as a quadratic program in standard form (cf. Mangasarian and Musicant, 2000) was introduced to geodesy, opening up new possibilities for robust estimation with inequality constraints.

8.2 Outlook

Naturally, the removal of the three major obstacles identified in this thesis leads to further questions.

As discussed in Chap. 4 it is not possible to represent the complete stochastic information of an inequality constrained estimate in form of a VCV matrix, because the description of the first two moments of the PDF is no longer sufficient. Thus, it is difficult to describe the correlation between different parameters. In particular, as we have illustrated that constraints can also add a dependency between otherwise uncorrelated parameters, which could be interpreted as a kind of correlation, too. As a consequence, it would be worthwhile to further investigate how to describe the natural correlations as well as those introduced by constraints. Another promising field of research might be the attempt to describe the probability mass accumulated at the borderline of the feasible region analytically.

In Chap. 5 we focused on singular systems with constraints only. Here, it would be interesting to examine the case of ill-posed problems with inequalities, too. This goes along with the problem to determine the size of the rank defect, as this states the dimension of the subspace in which the nullspace optimization has to be performed. Furthermore, we explicitely excluded the application of the stochastic framework to a rank-deficient system. However, possibilities to combine both frameworks should be addressed in the future.

In the description of the tailor-made ICGHM approach in Sect. 6.2.2.3 we outlined the possibility to speed up the solution of the KKT system via a suited decomposition (e.g., an LU factorization) and the precomputation of the reduction of a constant block. This should be examined further in future work. The same holds true for the Huber estimator with inequality constraints and correlated observations (cf. Sect. 7.1.1.1).

Despite the examples mentioned before, the introduction of inequality constraints combined with the key findings of this thesis can be beneficial for many applications in many geodetic disciplines. Besides thrilling new applications that could become possible such as applying non-negativity constraints for the runoff estimated from GRACE in satellite geodesy, even approaches already using inequalities can benefit from the proposed stochastic framework. This ranges from photogrammetry (e.g. constraints on the minimal distances between cameras in a camera placement problem, see Alsadik et al., 2012) over navigation (e.g., a Kalman filter with inequality constraints, see Simon and Simon, 2003) to geoinformation sciences (e.g. constraints in the process of data aggregation for a digital terrain model, see Koch, 2006).

Altogether, we hope that this thesis contributes its share in convincing many—not only geodetic—scientists that the benefits arising from the use of inequality constrained adjustment theory outweigh the additional effort.

A. Appendix

A.1 Transformation of the Dual Function of an ICLS Problem

In the following, all transformation steps to obtain the dual objective function (3.26) of an ICLS problem are provided. Inserting (3.25b) in (3.23) yields

$$
\begin{aligned}
\Psi(\boldsymbol{k},\bar{\boldsymbol{k}}) &= \min\{L(\boldsymbol{x},\boldsymbol{k},\bar{\boldsymbol{k}}) : \boldsymbol{x} \in \mathbb{R}^m\} \\
&= \widetilde{\boldsymbol{x}}^T \boldsymbol{N} \widetilde{\boldsymbol{x}} - 2\boldsymbol{n}^T \widetilde{\boldsymbol{x}} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} + 2\boldsymbol{k}^T (\boldsymbol{B}^T \widetilde{\boldsymbol{x}} - \boldsymbol{b}) - 2\bar{\boldsymbol{k}}^T (\bar{\boldsymbol{B}}^T \widetilde{\boldsymbol{x}} - \bar{\boldsymbol{b}}) && \text{(A.1a)} \\
&= \left(\boldsymbol{n}^T - \boldsymbol{k}^T \boldsymbol{B}^T + \bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T\right) \boldsymbol{N}^{-1} \boldsymbol{N} \widetilde{\boldsymbol{x}} - 2\boldsymbol{n}^T \widetilde{\boldsymbol{x}} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} && \text{(A.1b)} \\
&\quad + 2\boldsymbol{k}^T (\boldsymbol{B}^T \widetilde{\boldsymbol{x}} - \boldsymbol{b}) - 2\bar{\boldsymbol{k}}^T (\bar{\boldsymbol{B}}^T \widetilde{\boldsymbol{x}} - \bar{\boldsymbol{b}}) \\
&= \boldsymbol{n}^T \widetilde{\boldsymbol{x}} - \boldsymbol{k}^T \boldsymbol{B}^T \widetilde{\boldsymbol{x}} + \bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T \widetilde{\boldsymbol{x}} - 2\boldsymbol{n}^T \widetilde{\boldsymbol{x}} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} && \text{(A.1c)} \\
&\quad + 2\boldsymbol{k}^T \boldsymbol{B}^T \widetilde{\boldsymbol{x}} - 2\boldsymbol{k}^T \boldsymbol{b} - 2\bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T \widetilde{\boldsymbol{x}} + 2\bar{\boldsymbol{k}}^T \bar{\boldsymbol{b}} \\
&= -\boldsymbol{n}^T \widetilde{\boldsymbol{x}} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} + \boldsymbol{k}^T \boldsymbol{B}^T \widetilde{\boldsymbol{x}} - 2\boldsymbol{k}^T \boldsymbol{b} - \bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T \widetilde{\boldsymbol{x}} + 2\bar{\boldsymbol{k}}^T \bar{\boldsymbol{b}} && \text{(A.1d)} \\
&= -\boldsymbol{n}^T \boldsymbol{N}^{-1} (\boldsymbol{n} - \boldsymbol{B}\boldsymbol{k} + \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}}) + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} \\
&\quad + \boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} (\boldsymbol{n} - \boldsymbol{B}\boldsymbol{k} + \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}}) - 2\boldsymbol{k}^T \boldsymbol{b} && \text{(A.1e)} \\
&\quad - \bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T \boldsymbol{N}^{-1} (\boldsymbol{n} - \boldsymbol{B}\boldsymbol{k} + \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}}) + 2\bar{\boldsymbol{k}}^T \bar{\boldsymbol{b}} \\
&= -\boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{n} + \boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} - \boldsymbol{n}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} \\
&\quad + \boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} - \boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} + \bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} - 2\boldsymbol{b}^T \boldsymbol{k} && \text{(A.1f)} \\
&\quad - \boldsymbol{n}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} + \boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} - \bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} + 2\bar{\boldsymbol{b}}^T \bar{\boldsymbol{k}} \\
&= -\boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} - \bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} + 2\boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} && \text{(A.1g)} \\
&\quad + 2\boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} - 2\boldsymbol{n}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} - 2\boldsymbol{b}^T \boldsymbol{k} + 2\bar{\boldsymbol{b}}^T \bar{\boldsymbol{k}} - \boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{n} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell} \\
&= -\boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{B}\boldsymbol{k} - \bar{\boldsymbol{k}}^T \bar{\boldsymbol{B}}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} + 2\boldsymbol{k}^T \boldsymbol{B}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}}\,\bar{\boldsymbol{k}} && \text{(A.1h)} \\
&\quad + 2\left(\boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{B} - \boldsymbol{b}^T\right) \boldsymbol{k} - 2\left(\boldsymbol{n}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}} - \bar{\boldsymbol{b}}^T\right) \bar{\boldsymbol{k}} - \boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{n} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell}.
\end{aligned}
$$

Throughout the transformation, the symmetry of $\boldsymbol{N}$ was used and the fact that scalar quantities can be transposed without altering the result. Stating the equation in matrix-vector notation yields

$$
\begin{aligned}
\Psi(\boldsymbol{k},\bar{\boldsymbol{k}}) = &\begin{bmatrix} \boldsymbol{k}^T & \bar{\boldsymbol{k}}^T \end{bmatrix} \underbrace{\begin{bmatrix} -\boldsymbol{B}^T \boldsymbol{N}^{-1} \boldsymbol{B} & \boldsymbol{B}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}} \\ \bar{\boldsymbol{B}}^T \boldsymbol{N}^{-1} \boldsymbol{B} & -\bar{\boldsymbol{B}}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}} \end{bmatrix}}_{=:\boldsymbol{M}} \begin{bmatrix} \boldsymbol{k} \\ \bar{\boldsymbol{k}} \end{bmatrix} \\
&+ 2\begin{bmatrix} \boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{B} - \boldsymbol{b}^T & -\boldsymbol{n}^T \boldsymbol{N}^{-1} \bar{\boldsymbol{B}} + \bar{\boldsymbol{b}}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{k} \\ \bar{\boldsymbol{k}} \end{bmatrix} - \boldsymbol{n}^T \boldsymbol{N}^{-1} \boldsymbol{n} + \boldsymbol{\ell}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\ell}.
\end{aligned} \tag{A.2}
$$

As $\boldsymbol{M}$ is a symmetric and negative (semi-)definite matrix, $\Psi(\boldsymbol{k},\bar{\boldsymbol{k}})$ is a concave function. As a consequence, $-\Psi(\boldsymbol{k},\bar{\boldsymbol{k}})$ is a convex function. Therefore, the corresponding optimization problem (3.27) can be solved with any standard QP algorithm (cf. Sect. 3.5).

A.2 Deactivating Active Constraints in the Active-Set Algorithm

In Sect. 3.5.1.6 it was stated that all inequality constraints associated with negative Lagrange multipliers prevent a decrease of the objective function of problem (3.4) if $\boldsymbol{x}^{(i)}$ is the minimum of the current subspace. In the following, the proof of this statement is provided.

If $\boldsymbol{x}^{(i)}$ is the minimum of the current subspace it is not possible to further decrease the value of the objective function by a step in a binding direction (i.e., a direction that keeps all active constraints active, cf. Sect. 3.5.1.2). Therefore, we have to identify those non-binding directions that could lead to a decrease of the objective function.

Proposition: *A sufficiently small step in a direction $\boldsymbol{p}_j$ for which*

$$\boldsymbol{g}^T\boldsymbol{p}_j < 0 \tag{A.3}$$

holds, will decrease the value of the objective function.

Proof: The quadratic objective function of problem (3.4) can be completely described by a Taylor expansion up to degree two

$$\Phi(\boldsymbol{x}^{(i)}+q\boldsymbol{p}) = \Phi(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}^{(i)}} + q\nabla_{\boldsymbol{x}}\Phi(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}^{(i)}}\boldsymbol{p} + \frac{1}{2}q^2\boldsymbol{p}^T\nabla_{\boldsymbol{x}}\left(\nabla_{\boldsymbol{x}}\Phi(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}^{(i)}}\right)\boldsymbol{p} \tag{A.4a}$$

$$= \Phi(\boldsymbol{x}^{(i)}) + q\left(\boldsymbol{C}\boldsymbol{x}^{(i)}+\boldsymbol{c}\right)^T\boldsymbol{p} + \frac{1}{2}q^2\boldsymbol{p}^T\boldsymbol{C}\boldsymbol{p}. \tag{A.4b}$$

If (A.3) holds, then there exists a small positive scalar q for which the value of the objective function decreases. Inserting the gradient

$$\boldsymbol{g} = \boldsymbol{C}\boldsymbol{x}^{(i)} + \boldsymbol{c},$$

the objective function reads

$$\Phi(\boldsymbol{x}^{(i)}+q\boldsymbol{p}) = \Phi(\boldsymbol{x}^{(i)}) + \underbrace{q}_{>0}\underbrace{\boldsymbol{g}^T\boldsymbol{p}}_{<0} + \underbrace{\frac{1}{2}q^2\boldsymbol{p}^T\boldsymbol{C}\boldsymbol{p}}_{>0} \tag{A.5}$$

$$< \Phi(\boldsymbol{x}^{(i)}), \quad \text{for small } q.$$

□

The proof that a step in a direction $\boldsymbol{p}_j$ for which

$$\boldsymbol{g}^T\boldsymbol{p}_j \geq 0 \tag{A.6}$$

holds, can only increase the value of the objective function is obtained in an analogous way. Therefore, $\boldsymbol{x}^{(i)}$ can only be the point with smallest objective function value in the feasible region, if (A.6) holds for all non-binding directions.

In a next step, a connection between the Lagrange multipliers of active constraints and the relationship (A.3) shall be established. This will be helpful to identify those active constraints which have to be deactivated. As $\boldsymbol{x}^{(i)}$ is the minimum of the current subspace, the search direction $\boldsymbol{p}$ computed in that point is zero. Therefore, the derivative (3.47) of the Lagrangian with respect to $\boldsymbol{p}$ reduces to

$$\boldsymbol{g} + \boldsymbol{W}\boldsymbol{k}_w = \boldsymbol{0} \tag{A.7a}$$

and thus

$$\boldsymbol{g} = -\boldsymbol{W}\boldsymbol{k}_w \tag{A.7b}$$
$$= -\boldsymbol{W}(:,1)\boldsymbol{k}_w(1) - \boldsymbol{W}(:,2)\boldsymbol{k}_w(2) - \ldots - \boldsymbol{W}(:,p_w)\boldsymbol{k}_w(p_w). \tag{A.7c}$$

This is plausible as—so far—a minimization in the nullspace of the active set $\boldsymbol{W}$ was performed. Therefore, the gradient $\boldsymbol{g}$ in the point of the current solution has no parts in the nullspace but will lie completely in the range space of $\boldsymbol{W}$. As a consequence, $\boldsymbol{g}$ can be expressed as a linear combination of the columns of $\boldsymbol{W}$ which form a basis of its nullspace.

Inserting (A.7a) in (A.3) yields the desired connection between Lagrange multipliers and constraints

$$\boldsymbol{g}^T\boldsymbol{p}_j = -\boldsymbol{k}_w^T\boldsymbol{W}^T\boldsymbol{p}_j \tag{A.8a}$$
$$= -\boldsymbol{k}_w(1)\boldsymbol{W}(:,1)^T\boldsymbol{p}_j - \boldsymbol{k}_w(2)\boldsymbol{W}(:,2)^T\boldsymbol{p}_j - \ldots - \boldsymbol{k}_w(p_w)\boldsymbol{W}(:,p_w)^T\boldsymbol{p}_j. \tag{A.8b}$$

Proposition: $\boldsymbol{x}^{(i)}$ *is the point with minimal value of the objective function within the feasible region, if and only if all Lagrange multipliers connected with inequality constraints have a non-negative value.*

Proof: As proved in Sect. 3.5.1.2

$$\boldsymbol{W}(:,k)^T\boldsymbol{p}_j \leq 0, \quad \forall\ k = 1,2,...,p_w$$

holds for all feasible directions. As the columns of $\boldsymbol{W}$ are linearly independent by construction (cf. Gill et al., 1981, p. 201), no matter if there exist linear dependent constraints in general, it is possible to construct a direction $\boldsymbol{p}_j$ for which

$$\boldsymbol{W}(:,k)^T\boldsymbol{p}_j < 0 \tag{A.9a}$$

and

$$\boldsymbol{W}(:,i)^T\boldsymbol{p}_j = 0,\ i = 1,2,\ldots,k-1,k+1,\ldots p_w \tag{A.9b}$$

hold. Inserting (A.9) in (A.8b) yields

$$\boldsymbol{g}^T\boldsymbol{p}_j = -\boldsymbol{k}_w(k)\underbrace{\boldsymbol{W}(:,k)^T\boldsymbol{p}_j}_{<0}. \tag{A.10}$$

Consequently, the value of the objective function would decrease through a step in direction $\boldsymbol{p}_j$ if a negative Lagrange Multiplier associated with inequality constraint k exists. In this case $\boldsymbol{x}^{(i)}$ cannot be the minimum and the proposition is proved. □

Therefore, it is possible to use the sign of the Lagrange multipliers to identify the inequality constraint(s) that should be deactivated next (namely those that are linked with a negative Lagrange multiplier). If more than one constraint has a negative Lagrange multiplier, usually the one with the Lagrange multiplier that has the largest absolute value is dropped. However, one should keep in mind that equality constraints are always active and should never be dropped from the active set. Therefore, they should not enter the testing procedure described above.

A.3 Reformulation of the Huber Loss Function

The following reformulation of (7.1) was proposed by Mangasarian and Musicant (2000).

Proposition: *The function*

$$\rho(v) = \min\left\{ f(v,y) = \frac{1}{2}y^2 + k|v-y| : y \in \mathbb{R} \right\} \tag{A.11}$$

is equivalent to the loss function

$$\rho_{\text{Huber}}(v) = \begin{cases} \frac{1}{2}v^2, & |v| \le k \\ k|v| - \frac{1}{2}k^2, & |v| > k \end{cases} \tag{A.12}$$

of the classic Huber estimator (cf. (7.1)*).*

Proof: The derivative of (A.11) with respect to the scalar y reads

$$\frac{\partial f(v,y)}{\partial y} = y - k \cdot \operatorname{sign}(v-y). \tag{A.13}$$

As it is not defined for $v = y$ the subderivative is computed instead (cf. Rockafellar and Wets, 2009, p. 299), yielding the following distinction of cases

$$\frac{\partial f(v,y)}{\partial y} = \begin{cases} y - k, & v > y \\ y + \lambda k,\ \lambda \in [-1,1] & v = y \\ y + k, & v < y \end{cases}. \tag{A.14}$$

$f(v,y)$ becomes minimal if the subderivative is equal to zero, leading to

$$y = \begin{cases} k, & v > y = k \\ v = -\lambda k,\ \lambda \in [-1,1] & v = y = -\lambda k \\ -k, & v < y = -k \end{cases}. \tag{A.15}$$

Thus, the value of y is identical to the value of the influence function

$$\Psi_{\text{H}}(v) = \begin{cases} k, & v > k \\ v, & |v| \le k \\ -k, & v < -k \end{cases} \tag{A.16}$$

of the standard Huber estimator. Inserting (A.15) in (A.11) leads to

$$\rho(v) = \begin{cases} \frac{1}{2}k^2 + k(v-k) = kv - \frac{1}{2}k^2, & v > k \\ \frac{1}{2}v^2, & |v| \le k \\ \frac{1}{2}k^2 + k(-v-k) = -kv - \frac{1}{2}k^2, & v < -k \end{cases}. \tag{A.17}$$

which is identical to (A.12). Thus, the proof is complete. □

A.4 Data of the Positive Cosine Example

In Tab. A.1, the supporting points t_i as well as the observations ℓ_i of the example of estimating a positive definite covariance function in Sect. 7.2.1 are provided. The MATLAB random number generator was used to generate 1 000 000 samples of the observations (cf. Sect. 7.2.1).

Table A.1: Supporting points t_i and observations ℓ_i from the example of estimating a positive definite covariance function in Sect. 7.2.1.

i	t_i	ℓ_i	i	t_i	ℓ_i
1	0.0000	13.6661	26	0.2500	-1.7469
2	0.0100	11.6741	27	0.2600	-3.6271
3	0.0200	9.2258	28	0.2700	-6.9844
4	0.0300	5.8736	29	0.2800	-7.2926
5	0.0400	4.4700	30	0.2900	-7.9395
6	0.0500	0.9921	31	0.3000	-6.8677
7	0.0600	2.7769	32	0.3100	-4.9126
8	0.0700	0.4992	33	0.3200	-4.2674
9	0.0800	0.6154	34	0.3300	-2.6859
10	0.0900	2.9608	35	0.3400	1.0446
11	0.1000	4.3965	36	0.3500	2.1318
12	0.1100	6.3031	37	0.3600	3.7316
13	0.1200	4.3356	38	0.3700	2.4828
14	0.1300	4.9459	39	0.3800	1.8105
15	0.1400	4.2990	40	0.3900	0.9967
16	0.1500	0.5812	41	0.4000	0.3656
17	0.1600	0.4734	42	0.4100	1.1483
18	0.1700	-2.6287	43	0.4200	1.2223
19	0.1800	-3.3337	44	0.4300	0.6707
20	0.1900	-3.3755	45	0.4400	-0.8721
21	0.2000	-0.7667	46	0.4500	3.1633
22	0.2100	0.1865	47	0.4600	3.3745
23	0.2200	2.3498	48	0.4700	4.8484
24	0.2300	1.8760	49	0.4800	4.3550
25	0.2400	0.7216	50	0.4900	4.4772

Tables of Symbols

Scalars, Vectors and Matrices

a, α	Scalar
$\boldsymbol{a}$	Vector
$\widetilde{\boldsymbol{a}}$	Estimated vector
$\boldsymbol{\mathcal{A}}$	Random vector
$\boldsymbol{\alpha}$	True vector
$\boldsymbol{A}$	Matrix
$\bar{\boldsymbol{B}}, \bar{\boldsymbol{b}}$	Matrix and vector of equality constraints
$\boldsymbol{B}, \boldsymbol{b}$	Matrix and vector of inequality constraints
$\boldsymbol{I}_m$	Identity matrix of size $[m \times m]$
$\boldsymbol{A}(i,j)$	Element in row i and column j of matrix $\boldsymbol{A}$
$\boldsymbol{A}(i,:)$	Row i of matrix $\boldsymbol{A}$
$\boldsymbol{A}(:,j)$	Column j of matrix $\boldsymbol{A}$
$\boldsymbol{a}(i)$	Element in row i of vector $\boldsymbol{a}$
$\mathbf{1}$	Vector containing only Ones

Mathematical Symbols and Operators

$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}^T \quad \frac{\partial f(\boldsymbol{x})}{\partial x_2}^T \quad \dots \quad \frac{\partial f(\boldsymbol{x})}{\partial x_m}^T\right]^T$	Gradient of $f(\boldsymbol{x})$ with respect to $\boldsymbol{x}$
$\boldsymbol{x} \in C$	$\boldsymbol{x}$ is an element of the set C
$\forall$	For all
$\odot$	Khatri-Rao product
$\text{diag}(\boldsymbol{N})$	Extracts all diagonal elements of matrix $\boldsymbol{N}$

List of Abbreviations

BLUE	Best Linear Unbiased Estimator
DEM	Digital Elevation Model
ECGHM	Equality Constrained Gauss-Helmert Model
ECLS	Equality Constrained Least-Squares
GHM	Gauss-Helmert Model
GMM	Gauss-Markov Model
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GRACE	Gravity Recovery And Climate Experiment
HPD	Highest Posterior Density
ICGHM	Inequality Constrained Gauss-Helmert Model
ICLS	Inequality Constrained Least-Squares
IRLS	Iteratively Reweighted Least-Squares
IVS	International VLBI Service for Geodesy and Astrometry
KKT	Karush-Kuhn-Tucker
LCP	Linear Complementarity Problem
LDP	Least-Distance Program
LP	Linear Program
MC-QP method	Monte Carlo Quadratic Programming method
MDF	Marginal Density Function
OLS	Ordinary Least-Squares
PDF	Probability Density Function
QP	Quadratic Program
RAL	Richtlinien für die Anlage von Landstraßen (German road design standard)
RMS	Root Mean Square
SLE	System of Linear Equations
SOD	Second Order Design
SQP	Sequential Quadratic Programming
STD	Slant Troposphere Delay
VCV matrix	Variance CoVariance matrix
VLBI	Very Long Baseline Interferometry
WLS	Weighted Least-Squares
ZHD	Zenith Hydrostatic Delay
ZWD	Zenith Wet Delay

List of Figures

2.1 Quadratic forms for matrices with a different type of definiteness. 9

3.1 Convex and non-convex set. 23
3.2 Convex and non-convex function. 24
3.3 Taxonomy of optimization problems. 25
3.4 Duality gap. 31
3.5 Basic ideas of active-set and interior-point approaches. 33
3.6 Binding, non-binding and infeasible directions for an active constraint. 35

4.1 Effect of a single inequality constraint on the PDF of a parameter. 46
4.2 Probability density functions and confidence intervals of different estimates. 50
4.3 WLS and ICLS estimates for a line-fit example. 54
4.4 Joint and marginal PDFs from the line-fitting problem with independent constraints. 56
4.5 Joint and marginal PDFs from the line-fitting problem with dependent constraints. . 57

5.1 Isolines of the objective function of a bivariate convex optimization problem. 66
5.2 Two-dimensional *unit spheres* of L_1, L_2 and L_∞ norm. 70
5.3 Contour lines of the objective function of a rank-deficient problem. 74

6.1 Parabola through the origin. 90

7.1 Number of international phone calls from Belgium. 96
7.2 Straight lines fitted to the number of international phone calls from Belgium. 97
7.3 WLS and ICLS fits to the observations of a positive cosine function. 98
7.4 Joint and marginal PDFs of the positive cosine function estimation problem. 99
7.5 Baseline repeatabilities of an OLS and an ICLS estimate. 101
7.6 Differences between the OLS and the ICLS solution in the baseline estimation. . . . 102
7.7 OLS and ICLS ZWD estimates from Ny Ålesund and their corresponding HPD intervals. 102
7.8 Differences between the OLS and the ICLS estimates for Ny Ålesund. 103
7.9 SOD of a geodetic network with 3 fixed points and 8 new points. 106
7.10 Geodetic network of the welding tolerances example. 107
7.11 Boundary for the existing parts, the new part and the "gaps". 109
7.12 Heights h along the route of the road. 110
7.13 Optimal vertical gradient of the road. 112

List of Tables

4.1 Results from the line-fitting problem with independent constraints. 55

4.2 Results from the line-fitting problem with dependent constraints. 57

6.1 Advantages and drawbacks of two possible representations of the ICGHM. 87

7.1 Results and sensitivity analysis from the positive cosine function example. 100

A.1 Data from the example of estimating a positive definite covariance function. v

References

Albertella, A., N. Cazzaniga, M. Crespi, L. Luzietti, F. Sacerdote and F. Sansò. Deformations detection by a Bayesian approach: Prior information representation and testing criteria definition. In Sansò, F., and A. J. Gill, editors, *Geodetic Deformation Monitoring: From Geophysical to Engineering Roles*, IAG Symp. 131, pages 30–37. Springer-Verlag. Berlin, Heidelberg, 2006.

Alkhatib, H., and W.-D. Schuh. Integration of the Monte Carlo covariance estimation strategy into tailored solution procedures for large-scale least squares problems. *Journal of Geodesy*, 81:53–66, 2007. doi:10.1007/s00190-006-0034-z.

Alsadik, B. S., M. Gerke and G. Vosselman. Optimal camera network design for 3D modeling of cultural heritage. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 7–12, 2012. doi:10.5194/isprsannals-I-3-7-2012.

Barnett, V., and T. Lewis. *Outliers in Statistical Data, 3rd Ed.* John Wiley and Sons, Chichester, 1994.

Barrodale, I., and F. D. K. Roberts. An efficient algorithm for discrete l1 linear approximation with linear constraints. *SIAM Journal on Numerical Analysis*, 15(3):603–611, 1978.

Best, M. J. Equivalence of some quadratic programming algorithms. *Mathematical Programming*, 30(1):71–87, 1984. URL `http://www.springerlink.com/content/515n1162r2811v84/`.

Boyd, S., and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. URL `http://www.stanford.edu/~boyd/cvxbook/`.

Candes, E., J. Romberg and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52 (2):489–509, 2006. doi:10.1109/TIT.2005.862083.

Chen, M.-H., and Q.-M. Shao. Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92, 1999. URL `http://www.jstor.org/stable/1390921`.

Dalmolin, Q., and R. Oliveira. Inverse eigenvalue problem applied to weight optimisation in a geodetic network. *Survey Review*, 43(320):187–198, 2011. doi:10.1179/003962611X12894696205028.

Dantzig, G. *Linear Programming and Extensions*. Princeton University Press, New Jersey, USA, 1998.

Dantzig, G., and M. Thapa. *Linear Programming 2: Theory and Extensions*. Springer-Verlag. Berlin, 2003.

Davis, K., C. Park and S. Sinha. Constrained inference in generalized linear and mixed models. In *Proceedings of the Survey Methods Section, SSC Annual Meeting*, Ottawa, 2008.

Davis, K. *Constrained Statistical Inference in Generalized Linear, and Mixed Models with Incomplete Data*. PhD thesis, Carleton University, Ottawa, Canada, 2011.

De Moor, B. Total linear least squares with inequality constraints. Esat-sista report 1190-2, Department of Electrical Engineering, Katholieke Universiteit Leuven, 1990.

Doma, M. Particle swarm optimization in comparison with classical optimization for GPS network design. *Journal of Geodetic Science*, 3(4):250–257, 2013.

Doma, M., and A. Sedeek. Comparison of PSO, GAs and analytical techniques in second-order design of deformation monitoring networks. *Journal of Applied Geodesy*, 8(1):21–30, 2014.

Famula, T. Inequality constrained estimation of genetic parameters. *Theoretical and Applied Genetics*, 66:291–296, 1983.

Fang, X. On non-combinatorial weighted total least squares with inequality constraints. *Journal of Geodesy*, 88(8):805–816, 2014. doi:10.1007/s00190-014-0723-y.

Fletcher, R. *Practical Methods of Optimization.* Wiley-Interscience, New York, USA, 2nd edition, 1987.

Fletcher, R., and T. Johnson. On the stability of null-space methods for KKT systems. *SIAM J. Matrix Anal. Appl*, 18:938–958, 1997.

Fritsch, D. Second order design of geodetic networks: Problems and examples. In *Proceedings of the International Symposium on Geodetic Networks and Computations*, 1982.

Fritsch, D. Optimal design of two-dimensional FIR-filters. In *Proceedings IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP 83)*, pages 383–386, 1983.

Fritsch, D. Some additional informations on the capacity of the linear complementarity algorithm. In Grafarend, E., and F. Sansò, editors, *Optimization and Design of Geodetic Networks*, pages 169–184. Springer. Berlin, Heidelberg, New York, Tokyo, 1985.

Fritsch, D., and B. Schaffrin. Optimal FIR-filter design subject to inequality constraints by means of the complementarity algorithm. In *Proceedings of first european signal processing conference*, Lausanne, Switzerland, 1980.

Geiger, C., and C. Kanzow. *Theorie und Numerik restringierter Optimierungsaufgaben.* Springer-Verlag. Berlin, Heidelberg, 2002.

Geweke, J. Exact inference in the inequality constrained normal linear regression model. *Journal of Applied Econometrics*, 1(2):127–41, 1986. URL `http://ideas.repec.org/a/jae/japmet/v1y1986i2p127-41.html`.

Gill, P., W. Murray and M. Wright. *Practical Optimization.* Academic Press, London, 1981.

Gould, N. Some reflections on the current state of active-set and interior-point methods for constrained optimization. *SIAG/Optimization Views-and-News*, 14(1):2–7, 2003.

Grafarend, E., and F. Sansò. *Optimization and Design of Geodetic Networks.* Springer-Verlag. Berlin, Heidelberg, New York, Tokyo, 1985.

Grafarend, E., and B. Schaffrin. Kriterion-Matrizen I – zweidimensionale homogene und isotrope geodätische Netze. *Zeitschrift für Vermessungswesen*, 4:133–149, 1979.

Grant, M., and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.

Grant, M., S. Boyd and Y. Ye. *Global Optimization: From Theory to Implementation*, chapter Disciplined Convex Programming, pages 155–210. Springer US, 2006.

Halsig, S., L. Roese-Koerner, T. Artz, A. Nothnagel and W.-D. Schuh. Improved parameter estimation of zenith wet delay using an inequality constrained least squares method. In *IAG Scientific Assembly, Potsdam 2013*, IAG Symp. 143. Springer. Berlin, Heidelberg, 2015. accepted.

Helmert, F. *Die Ausgleichungsrechnung nach der Methode der Kleinsten Quadrate.* B.G. Teubner Verlag, Leipzig, 1872.

Jabr, R. Power system Huber M-estimation with equality and inequality constraints. *Electric Power Systems Research*, 74(2):239–246, 2005. doi:10.1016/j.epsr.2004.11.002.

Jäger, R., T. Müller, H. Saler and R. Schwäble. *Klassische und robuste Ausgleichungsverfahren.* Wichmann-Verlag, Heidelberg, 2005.

Joint Committee for Guides in Metrology. Evaluation of measurement data—Supplement 1 to the "Guide to the expression of uncertainty in measurement"—Propagation of distributions using a Monte Carlo method, 2008. URL `http://www.bipm.org/en/publications/guides/gum.html`. JCGM 101:2008.

Kalbfleisch, J. D., and R. L. Prentice. *The Statistical Analysis of Failure Time Data.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., 2nd edition, 2002. doi:10.1002/9781118032985.ch1.

Kang, M.-W., M. K. Jha and P. Schonfeld. Applicability of highway alignment optimization models. *Transportation Research Part C: Emerging Technologies*, 21(1):257–286, 2012. doi:10.1016/j.trc.2011.09.006.

Kaschenz, J. *Regularisierung unter Berücksichtigung von Residuentoleranzen.* PhD thesis, Technische Universität Berlin, 2006.

Kato, B. S., and H. Hoijtink. A Bayesian approach to inequality constrained linear mixed models: estimation and model selection. *Statistical Modelling*, 6(3):231–249, 2006. doi:10.1191/1471082X06st119oa.

Klumpp, R. *Achsberechnungen mit Hilfe der Ausgleichungs- und Optimierungsrechnung.* PhD thesis, University of Stuttgart, 1973.

Koch, A. *Semantische Integration von zweidimensionalen GIS-Daten und Digitalen Geländemodellen.* PhD thesis, University of Hannover, 2006. URL `http://www.baufachinformation.de/literatur.jsp?dis=2007089016528`. DGK series C, No. 601.

Koch, K. R. Dynamische Optimierung am Beispiel der Straßentrassierung. *Vermessungswesen und Raumordnung*, 38:281–290, 1976.

Koch, K. R. Hypothesis testing with inequalities. *Bollettino di geodesia e scienze affini*, 2:136–144, 1981.

Koch, K. R. Optimization of the configuration of geodetic networks. In *Proceedings of the international symposium on geodetic networks and computations*, pages 82–89, 1982. DGK series B, No. 258/III.

Koch, K. R. First order design: Optimization of the configuration of a network by introducing small position changes. In Grafarend, E., and F. Sansò, editors, *Optimization and Design of Geodetic Networks*, pages 56–73, Berlin - Heidelberg - New York - Tokyo, 1985. Springer.

Koch, K. R. *Parameter Estimation and Hypothesis Testing in Linear Models.* Springer–Verlag, Berlin, Heidelberg, New York, 1999.

Koch, K. R. *Introduction to Bayesian Statistics*. Springer-Verlag. Berlin, Heidelberg, New York, 2007. 2nd edition.

Koch, K. R. Robust estimations for the nonlinear Gauss Helmert model by the expectation maximization algorithm. *Journal of Geodesy*, 88(3):263–271, 2014. doi:10.1007/s00190-013-0681-9.

Kühn, W. Vernetzte Bearbeitung. Neue Methodik für den Entwurf von Straßen. *Deutsches Ingenieurblatt*, 20(1/2):14–19, 2013.

Lawson, C., and R. Hanson. *Solving Least Squares Problems*. Series in Automatic Computation. Prentice-Hall, London, 1974.

Lehmann, R., and F. Neitzel. Testing the compatibility of constraints for parameters of a geodetic adjustment model. *Journal of Geodesy*, 87(6):555–566, 2013. doi:10.1007/s00190-013-0627-2.

Lenzmann, L., and E. Lenzmann. Strenge Auswertung des nichtlinearen Gauß-Helmert-Modells. *allgemeine vermessungs-nachrichten*, 111(2):68–73, 2004.

Liew, C. K. Inequality constrained least-squares estimation. *Journal of the American Statistical Association*, 71(355):746–751, 9 1976. URL `http://www.jstor.org/stable/2285614`.

Mangasarian, O. L., and D. R. Musicant. Robust linear and support vector regression. *IEEE Transaction on pattern analysis and machine intelligence*, 22(9):950–955, 2000.

Mehrabi, H., and B. Voosoghi. Optimal observation planning of local GPS networks: assessing an analytical method. *Journal of Geodetic Science*, 4:87–97, 2014.

Meissl, P. Zusammenfassung und Ausbau der Inneren Fehlertheorie eines Punkthaufens. In Rinner, K., K. Killian and P. Meissl, editors, *Beiträge zur Theorie der geodätischen Netze*, pages 8–21. Verlag der Bayerischen Akademie der Wissenschaften, Munich, 1969. DGK series A, No. 61.

Müller, H. Second-order design of combined linear-angular geodetic networks. *Journal of Geodesy*, 59(4):316–331, 1985. doi:10.1007/BF02521066.

Neitzel, F. *Identifizierung konsistenter Datengruppen am Beispiel der Kongruenzuntersuchung geodätischer Netze*. PhD thesis, TU Munich, 2004.

Niemeier, W. *Ausgleichungsrechnung*. Walter de Gruyter, Berlin, 2002.

Nocedal, J., and S. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, 1st edition, 1999.

O'Leary, D. P., and B. W. Rust. Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems. *SIAM Journal on Scientific and Statistical Computing*, 7:473–489, 1986.

Peng, J., H. Zhang, S. Shong and C. Guo. An aggregate constraint method for inequality-constrained least squares problems. *Journal of Geodesy*, 79:705–713, 2006. doi:10.1007/s00190-006-0026-z.

Peracchi, F. *Econometrics*. John Wiley & Sons, Chichester, New York, Weinheim, Brisbane, Singapore, Toronto, 2001.

Pietzsch, W. *Straßenplanung*. Werner-Verlag, Düsseldorf, 1989.

Press, W., S. Teukolsky, W. Vetterling and B. Flannery. *Numerical Recipes*. Cambridge University Press, 2007.

RAL. Richtlinien für die Anlage von Landstraßen. Technical report, Forschungsgesellschaft für Straßen- und Verkehrswesen, 2012.

Rockafellar, R. Lagrange multipliers and optimality. *SIAM Rev.*, 35(2):183–238, 1993. doi:10.1137/1035044.

Rockafellar, R., and R. Wets. *Variational Analysis.* Fundamental Principles of Mathematical Sciences. Springer-Verlag. Berlin, 2009.

Roese-Koerner, L. Quadratische Programmierung mit Ungleichungen als Restriktionen. Master's thesis, University of Bonn, 2009. URL `http://www.igg.uni-bonn.de/tg/uploads/tx_ikgpublication/roese-koerner_09.pdf`.

Roese-Koerner, L., and W.-D. Schuh. Convex optimization under inequality constraints in rank-deficient systems. *Journal of Geodesy*, 88(5):415–426, 2014. doi:10.1007/s00190-014-0692-1.

Roese-Koerner, L., and W.-D. Schuh. Effects of different objective functions in inequality constrained and rank-deficient least-squares problems. In *VIII Hotine-Marussi Symposium on Mathematical Geodesy*, IAG Symp. 142. Springer. Berlin, Heidelberg, 2015. accepted.

Roese-Koerner, L., B. Devaraju, N. Sneeuw and W.-D. Schuh. A stochastic framework for inequality constrained estimation. *Journal of Geodesy*, 86(11):1005–1018, 2012. doi:10.1007/s00190-012-0560-9.

Roese-Koerner, L., B. Devaraju, W.-D. Schuh and N. Sneeuw. Describing the quality of inequality constrained estimates. In Kutterer, H., F. Seitz, H. Alkhatib and M. Schmidt, editors, *Proceeding of the 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11)*, IAG Symp. 140, pages 15–20. Springer-Verlag, 2015. doi:10.1007/978-3-319-10828-5_3.

Rousseeuw, P., and A. Leroy. *Robust Regression and Outlier Detection.* John Wiley and Sons, New Jersey, 2003.

Schaffrin, B. Ausgleichung mit Bedingungs-Ungleichungen. *allgemeine vermessungs-nachrichten*, 88. Jg.:227–238, 1981.

Shewchuk, J. R. An introduction to the conjugate gradient method without the agonizing pain. Technical report, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, 1994. URL `http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf`.

Simon, D., and D. L. Simon. Aircraft turbofan engine health estimation using constrained Kalman filtering. Technical report, National Aeronautics and Space Administration Cleveland OH Glenn Research Center, 2003. URL `http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA417944`.

Stoer, J. On the numerical solution of constrained least-squares problems. *SIAM Journal on Numerical Analysis*, 8(2):382–411, 1971. URL `http://www.jstor.org/stable/2949486`.

Suykens, J., G. Horvath and S. Basu, editors. *Advances in Learning Theory: Methods, Models and Applications*, volume 190 of *NATO Science.* IOS Press, 2003.

Tang, J., H. Cheng and L. Liu. Using nonlinear programming to correct leakage and estimate mass change from GRACE observation and its application to antarctica. *Journal of Geophysical Research*, 117:B11410, 2012.

Vitkiene, J., and V. Puodziukas. Design of road based on the applicable design methods used in Western European countries. In *The 9th International Conference "Environmental Engineering", Vilnius, Lithuania, 2014*, page enviro.2014.175. Vilnius Gediminas Technical University Press Technika, 2014. doi:10.3846/enviro.2014.175.

Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3):426–482, 1943. URL `http://www.jstor.org/stable/1990256`.

Werner, H. J. On inequality constrained generalized least-squares estimation. *Linear Algebra and its Applications*, 127:379–392, 1990. doi:10.1016/0024-3795(90)90351-C.

Werner, H. J., and C. Yapar. On inequality constrained generalized least squares selections in the general possibly singular Gauss-Markov model: A projector theoretical approach. *Linear Algebra and its Applications*, 237/238:359–393, 1996. doi:10.1016/0024-3795(94)00357-2.

Wölle, N. Graphischer Ausgleich durch Quadratische Programmierung mit linearen Gleichungen und Ungleichungen. Master's thesis, TU Graz, 1988.

Wong, E. *Active-Set Methods for Quadratic Programming*. PhD thesis, University of California, San Diego, 2011.

Xu, P. A general solution in geodetic nonlinear rank-deficient models. *Bollettino di geodesia e scienze affini*, 1:1–25, 1997.

Xu, P., E. Cannon and G. Lachapelle. Stabilizing ill-conditioned linear complementarity problems. *Journal of Geodesy*, 73:204–213, 1999. doi:10.1007/s001900050237.

Yetkin, M., C. Inal and C. Yigit. The optimal design of baseline configuration in GPS networks by using the particle swarm optimisation algorithm. *Survey Review*, 43(323):700–712, 2011. doi:10.1179/003962611X13117748892597.

Zeng, W., J. Liu and Y. Yao. On partial errors-in-variables models with inequality constraints of parameters and variables. *Journal of Geodesy*, 89(2):111–119, 2015. doi:10.1007/s00190-014-0775-z.

Zhang, S., X. Tong and K. Zhang. A solution to EIV model with inequality constraints and its geodetic applications. *Journal of Geodesy*, 87(1):23–28, 2013. doi:10.1007/s00190-012-0575-2.

Zhu, J., R. Santerre and X.-W. Chang. A Bayesian method for linear, inequality-constrained adjustment and its application to GPS positioning. *Journal of Geodesy*, 78(9):528–534, 2005. doi:10.1007/s00190-004-0425-y.

Acknowledgments

Writing this thesis would not have been possible without the following people.

First of all, I would like to thank my Ph.D. supervisor Wolf-Dieter Schuh. Thank you for your guidance, all the long discussions and for providing me with the freedom to find my own line of research. In addition, I want to express my gratitude to Hans-Peter Helfrich and Nico Sneeuw for both agreeing to review my thesis and for your valuable input.

Parts of this thesis have already been published before. Thus, I would like to thank the following colleagues for an always fruitful collaboration: Balaji Devaraju, Nico Sneeuw, Sebastian Halsig, Thomas Artz and Axel Nothnagel.

Moreover, I would like to thank Boris Kargoll, Karl-Rudolf Koch and Marisa Röder-Sorge for proof-reading my thesis.

Thank you, Judith Schall, for all the discussions on optimization and so many other (work- and non-work-related) topics in our reading group, in bars and during cigarette breaks without a cigarette.

This thesis would hardly exist without Jan Martin Brockmann. You are not only the best roommate one could possibly ask for, but also a very open minded-colleague and became a dear friend over the years. Thank you for all your input and the proofreading.

For the SOD of a geodetic network, I used a MATLAB software written by Maike Schumacher. I hope you are not to disappointed that I have decided against drawing the error ellipses in pink. The network data from the Messdorfer Feld war kindly provided by Florian Schölderle and Heiner Kuhlmann from the department of Geodesy from the University of Bonn.

I always liked being part of the department of Theoretical Geodesy in Bonn very much. Together with the APMG group it was a great working environment and makes it easy to enjoy work.

All computations were performed using MATLAB, Octave and CVX.

Furthermore, my parents deserve a huge “Thank you”. You supported me right from the beginning in everything I did and never stopped doing so.

Finally, I am very grateful for my wife Beate and my sons Lars and Nils. Beate, without your love and your support (and your patience), this thesis would not exist! It is really encouraging to know that there is someone that has your back no matter what.