J Geod
DOI 10.1007/s00190-012-0560-9

ORIGINAL ARTICLE

A stochastic framework for inequality constrained estimation

Lutz Roese-Koerner · Balaji Devaraju · Nico Sneeuw · Wolf-Dieter Schuh

Received: 15 November 2011 / Accepted: 24 March 2012

Abstract Quality description is one of the key features of geodetic inference. This is even more true if additional information about the parameters is available that could improve the accuracy of the estimate. However, if such additional information is provided in the form of inequality constraints, most of the standard tools of quality description (variance propagation, confidence ellipses, etc.) cannot be applied, as there is no analytical relationship between parameters and observations. 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 or the influence of inactive constraints and therefore yield only approximate results. In this article, a frequentist framework for quality description of inequality constrained least-squares estimates is developed, based on the Monte Carlo method. The quality is described in terms of highest probability density regions. Beyond this accuracy estimate, the proposed method allows to determine the influence and contribution of each constraint on each parameter using Lagrange multipliers. Plausibility of the constraints is checked by hypothesis testing and estimating the probability mass in the infeasible region. As more probability mass concentrates in less space, applying the proposed method results in smaller confidence regions compared to the unconstrained ordinary least-squares solution. The method is applied to describe the quality of estimates in the problem of approximating a time series with positive definite functions.

Keywords Stochastic framework · Inequality constrained least-squares · Monte Carlo method · Quality description · Sensitivity analysis · Lagrange multipliers

L. Roese-Koerner (✉) · W.-D. Schuh
Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany
e-mail: roese-koerner@geod.uni-bonn.de

W.-D. Schuh
e-mail: schuh@geod.uni-bonn.de

B. Devaraju · N. Sneeuw
Institute of Geodesy, University of Stuttgart, Stuttgart, Germany
e-mail: devaraju@gis.uni-stuttgart.de

N. Sneeuw
e-mail: sneeuw@gis.uni-stuttgart.de

1 Introduction

Prior knowledge about properties of the parameters is very common in many geodetic applications. This information (e.g. the knowledge that parameters lie within a certain interval) is often expressed by inequality constraints, which have to be strictly fulfilled. An example of prior knowledge of this type are constraints due to limited resources or a maximum feasible slope parameter of a surface of best fit.

Although well known in mathematics for some time, inequality constrained least-squares (ICLS) adjustment did not appear in geodetic literature before the late seventies and early eighties. Schaffrin et al. (1980), for example, transformed the task of a second order design of geodetic networks into a linear complementarity problem (LCP) and solved it with the Dantzig–Cottle algorithm to ensure positive weights. Schaffrin (1981) pointed out the potential of the same approach for the problem of estimating a (quasi-)optimal finite impulse response (FIR) filter with a constraint on the maximum approximation error, and for the problem of estimating non-negative variance components. Not only the second, but also the first order design could be improved by inequalities as was shown by Koch (1982, 1985). He introduced inequality constraints on the maximum feasible

Published online: 11 April 2012

distance for a coordinate shift as well as a constraint on the overall accuracy.

More recent work also include the semantic integration of data from a geographical information system and digital terrain models using object specific constraints (Koch 2006) as well as the inclusion of (im)possible directions of landslide movement as constraints in the determination of landslide (Song et al. 2010).

Despite the highly developed estimation theory dealing with inequality constrained problems, these techniques are far from being standard tools in geodesy. This might be due to the fact, that up to now, the existing approaches to describe the quality of inequality constrained estimates (Liew 1976; Geweke 1986; Zhu et al. 2005), have several drawbacks, e.g., a suboptimal treatment of the truncated probability mass or overoptimistic results.

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 project 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 simplex or interior point methods, cf. Sect. 2.2), and the law of error propagation 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.

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 we propose in this paper is a Monte Carlo technique. This has the advantage, that no complex analytical relationship has to be evaluated and that it is easy to parallelize.

We use the active-set method (cf. Gill et al. 1981, p. 167–173) 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, confidence regions are computed and several ideas of what constitutes a *best* estimator in this case are discussed.

Besides knowing the distribution of the parameters, it might be interesting to examine the influence of the constraints on the estimate. Therefore, a sensitivity analysis is described, which allows to quantify the influence of each constraint on each parameter. Furthermore, an hypothesis testing is outlined to determine if the data support the constraints, and it is shown how the ratio of probability mass inside the feasible region can be computed.

As quality description is a key feature of geodetic adjustment theory, the purpose of this paper is to describe a framework for the stochastic description of ICLS estimates, which overcomes many of the drawbacks of the existing methods.

The present article is organized as follows: in Sect. 2, basic principles of inequality constraint adjustment and existing methods of quality description are reviewed. In Sect. 3, we introduce a new method for the quality description of ICLS problems and provide some analysis tools to measure the contribution of the constraints. Results of such a constrained estimate are presented in Sect. 4, where the methodology is applied to the problem of fitting a time series using a cosine transformation with only non-negative coefficients. This is a task, which appears for example, when a covariance function of a time series should be estimated, because the fulfillment of the non-negativity constraints ensures that the resulting function is positive (semi-)definite. Finally, in Sect. 5, the major findings are summarized, drawbacks and advantages of the proposed method are reviewed and some future challenges in quality description are pointed out.

2 Background

2.1 Inequality constrained least-squares estimation (ICLS)

Every problem in which a convex objective function Φ should be minimized with respect to some constraints forming a convex set is called a convex optimization problem (Boyd and Vandenberghe 2004, p. 137). The (linear) ICLS problem is a special case of this type of problem as a quadratic function (the sum of squared residuals) is to be minimized subject to linear inequality constraints.

The deterministic model of a standard Gauss–Markov model (GMM) is given as

$$\mathbf{y} + \mathbf{v} = \mathbf{A}\mathbf{x}. \tag{1}$$

$\mathbf{y}$ contains the n observations, $\mathbf{v}$ is the vector of residuals, $\mathbf{A}$ the design matrix and the m unknown parameters are contained in $\mathbf{x}$. The random vector of observations $\boldsymbol{\mathcal{Y}}$ is supposed to be normally distributed with known variance–covariance (VCV) matrix

$$\boldsymbol{\mathcal{Y}} \sim N(\mathbf{A}\boldsymbol{\xi}, \mathbf{Q}).$$

$\boldsymbol{\xi}$ is the vector of true parameters. In a linear ICLS problem, we try to minimize the (weighted) sum of squared residuals subject to linear constraints

$$\text{minimize } \Phi(\mathbf{x}) = \mathbf{v}^{\mathrm{T}}\mathbf{P}\mathbf{v} \tag{2a}$$

$$\text{subject to } \mathbf{B}^{\mathrm{T}}\mathbf{x} \leq \mathbf{b}. \tag{2b}$$

P denotes the weight matrix, **B** the matrix containing p constraints and **b** the right hand side of the constraints. The representation for the linear inequality constraints chosen here can be used in all ICLS problems without loss of generality, because greater than or equal to constraints can be transformed to less than or equal to constraints. This also greatly simplifies the way the constraints are handled in the estimation process.

Many algorithms use the Lagrangian of an ICLS problem, which is computed by multiplying the rearranged constraints with the so called Lagrange multiplier vector **k** and adding the product to the objective function $\Phi(\mathbf{x})$

$$L(\mathbf{x}, \mathbf{k}) = \Phi(\mathbf{x}) + \mathbf{k}^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{x} - \mathbf{b}) \tag{3a}$$

$$= \mathbf{x}^{\mathrm{T}}\mathbf{N}\mathbf{x} - 2\mathbf{x}^{\mathrm{T}}\mathbf{n} + \mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{k}^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{x} - \mathbf{b}), \tag{3b}$$

with normal equation matrix **N** and right hand side of the normal equations **n**. The Lagrangian allows us to integrate the constraints into the objective function. This is possible, because for each constraint a new parameter—the Lagrange multiplier k_i linked with the ith constraint—is introduced. Derivation of (3b) leads to the Karush–Kuhn–Tucker conditions (cf. Boyd and Vandenberghe 2004, p.243). Especially one of them—*the condition of complementary slackness*—will be explained and used later in the active constraint approach in Sect. 2.3.1.

2.2 Solving ICLS problems

In the taxonomy of optimization problems, Eq. (2) is called a quadratic program (QP), as a quadratic objective function is to be minimized subject to linear inequality constraints (Boyd and Vandenberghe 2004, p.152). Several solvers for this type of problem exist, which can be subdivided into two classes (despite some exceptions): simplex methods and interior point methods. While the former are special solvers for quadratic and linear programs (LP), the latter are applicable to a wide range of convex optimization problems.

Simplex methods like the active-set method (Gill et al. 1981, p.167–173) or Dantzig's simplex method for quadratic programming (Dantzig 1998, p. 490–498) subdivide the set of constraints (here symbolized by the matrix of constraints **B** and the righthand side **b**) into an active part $\mathbf{B}_{\mathrm{a}}$, $\mathbf{b}_{\mathrm{a}}$ and an inactive part $\mathbf{B}_{\mathrm{i}}$, $\mathbf{b}_{\mathrm{i}}$. A constraint is called active (or binding) if it is exactly satisfied, and therefore, holds as equality constraint. It is called inactive, if it is fulfilled as strict inequality. Therefore, (2b) can be written as:

$$\mathbf{B}_{\mathrm{a}}^{\mathrm{T}}\mathbf{x} = \mathbf{b}_{\mathrm{a}} \quad \text{and} \quad \mathbf{B}_{\mathrm{i}}^{\mathrm{T}}\mathbf{x} < \mathbf{b}_{\mathrm{i}}. \tag{4}$$

The constraints that have been identified as active constraints are then used to follow the boundary of the feasible region

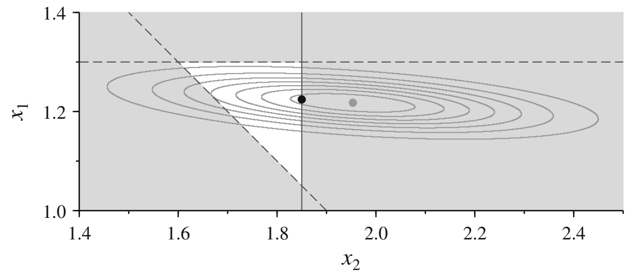


Fig. 1 Isolines of the objective function (*black ellipses*) and constraints (*straight lines*) of an ICLS problem. The optimal OLS solution (*gray dot*) is in the infeasible region (*gray shaded area*). The ICLS solution (*black dot*) is the projection of the unconstrained solution onto the boundary of the feasible set. One constraint is active in the optimal solution (*solid line*) and two are inactive (*dashed lines*)

(i.e. the region in the parameter space, where all constraints are satisfied, cf. the unshaded area in Fig. 1), until the optimal solution (black dot) is reached. The optimal solution is the point with smallest value of the objective function, that fulfills all constraints. This point will always lie at the boundary of the feasible region if at least one constraint is active (solid line). The solution of a quadratic program is therefore the projection of the solution of the unconstrained problem (gray dot) onto the feasible set, due to the metric of the problem (illustrated by the contour lines, representing isolines of the objective function).

Interior point methods like the logarithmic barrier method or primal–dual methods (Boyd and Vandenberghe 2004, p. 568–571 and p. 609–613) use a different approach: starting at a feasible point far away from the constraints, the solution follows a central path through the interior of the feasible region towards the optimum. This is done by splitting the complex original problem either into a sequence of unconstrained problems, e.g. by punishing the violation of constraints by using a penalty function (barrier methods), or into a sequence of simpler inequality constrained problem, by relaxing some conditions of the constraints (primal–dual methods).

It is also possible to transform the QP into a Linear Complementarity Problem (LCP, Koch 2006, p. 24–25) and solve it e.g. with Lemke's algorithm (cf. Fritsch 1985). More recent approaches include, for example, the aggregation of all simple inequality constraints into one complex equality constraint (Peng et al. 2006).

It is important to mention two facts: first is, that all methods theoretically give the same result, and second that if no constraint is active in the solution, the result will be identical to the OLS estimate. That is because inactive constraints (dotted lines in Fig. 1) do not influence the parameter estimation. However, as we will discuss later, they do influence its statistical properties.

2.3 Quality description in ICLS estimates

All methods described in the last section are iterative solvers. Therefore (and due to the presence of inequality constraints), it is difficult to describe the quality of the estimate. So far, two different approaches to give a measure for the accuracy of those estimates have been developed: The active constraint approach, which uses frequentist statistics and a Bayesian method.

2.3.1 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.

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 (e.g. via Lemke's algorithm, using the transformation to a LCP). Afterwards, the Lagrange multipliers are used to identify the active constraints $\mathbf{B}_a$, $\mathbf{b}_a$. This is possible, as there is the rule of complementary slackness (cf. Gill et al. 1990, p. 302)

$$k_j(\mathbf{B}^T\mathbf{x}-\mathbf{b})_j \stackrel{!}{=} 0, \quad \forall j = 1, \ldots p, \tag{5}$$

stating that only Lagrange multipliers of active constraints are different from zero. 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. 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 can be carried out, discarding all inactive constraints

$$\text{minimize } \Phi(\mathbf{x}) = \mathbf{v}^T\mathbf{P}\mathbf{v} \tag{6a}$$

$$\text{subject to } \mathbf{B}_a^T\mathbf{x} = \mathbf{b}_a. \tag{6b}$$

This allows to compute the variance–covariance matrix of the equality constrained problem.

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. 2, 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 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. Also, the approach is not robust to changes in the set of active constraints due to (small) changes in the observations.

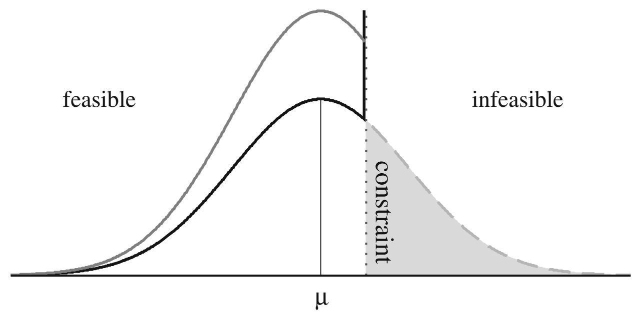


Fig. 2 Effect of a single inequality constraint (*dotted line*) on the PDF of a parameter with expectation value μ (univariate case). As the probability mass has to be conserved, it is necessary to decide how to handle the part of the probability mass of the unconstrained estimate (*dashed line*) that 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*)

2.3.2 Bayesian approach

Besides the frequentist approaches, the problem of ICLS estimation can also be tackled using Bayesian statistics. Here 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:

$$f(\mathbf{x}|\mathbf{y}) = \frac{f(\mathbf{y}|\mathbf{x})\,f(\mathbf{x})}{f(\mathbf{y})} \tag{7}$$

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

$$f(\mathbf{y}|\mathbf{x}) = f(\mathbf{v}) \propto \exp\left\{-\frac{1}{2\sigma_0^2}\mathbf{v}^T\,\mathbf{Q}^{-1}\,\mathbf{v}\right\} \tag{8}$$

depending on the residuals $\mathbf{v}$, the variance–covariance matrix of the observations $\mathbf{Q}$ and the prior distribution $f(\mathbf{x})$ of the parameters (cf. Koch 2007, p.89). Usually a uniform distribution of the form

$$f(\mathbf{x}) = \begin{cases} \frac{1}{s}, & \text{if } \mathbf{B}^T\mathbf{x} \leq \mathbf{b} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

is used. The constant s ensures, that the integral over the whole space is equal to one. In statistics, the normalization term $f(\mathbf{y})$ is often neglected, resulting in

$$f(\mathbf{x}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{x})\,f(\mathbf{x}). \tag{10}$$

The idea with this method is that we assume that the observations $\mathbf{y}$ and the residuals $\mathbf{v}$ a priori follow the normal probability density function $f(\mathbf{y})$ and $f(\mathbf{y}|\mathbf{x}) = f(\mathbf{v})$, respectively, and further we assume that the possible solutions $\mathbf{x}$ a priori follow a uniform density function within the given constrained bounds. Thus, inserting these probability densities in (10) provides us with an a posteriori density function as follows

$$f(\mathbf{x}|\mathbf{y}) \propto \begin{cases} \exp\left\{-\frac{1}{2\sigma_0^2}\mathbf{v}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{v}\right\}\frac{1}{s}, & \text{if } \mathbf{B}^{\mathrm{T}}\mathbf{x} \leq \mathbf{b} \\ 0, & \text{otherwise} \end{cases}. \quad (11)$$

The posterior density is a truncated (piece-wise continuous) version of the 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. 2) meaning that the constraints influence not only the boundary regions but the complete 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. 2). 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 (e.g. 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 the current contribution.

3 A stochastic framework for ICLS estimates

We subdivide the task of defining a stochastic framework for ICLS estimates into the problem of quality description and the problem of measuring the influence of the constraints. Afterwards, both parts are combined to the aforementioned stochastic framework.

3.1 Quality description

Along the line of thought at the end of Sect. 2.3.2 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.

3.1.1 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. the 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:

First we compute the ICLS solution $\tilde{\mathbf{x}}$ and the OLS solution of the unconstrained problem (1)

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{y}, \quad (12)$$

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

$$E\{\boldsymbol{\mathcal{Y}}\} = \hat{\mathbf{y}} = E\{\mathbf{A}\boldsymbol{\mathcal{X}}\} = \mathbf{A}\hat{\mathbf{x}}. \quad (13)$$

Henceforth in this paper, 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 variance–covariance matrix $\mathbf{Q}$ of observation vector $\boldsymbol{\mathcal{Y}}$ 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 variance–covariance matrix $\mathbf{Q}$ of the observations is decomposed into the product of two upper triangular matrices $\mathbf{R}$:

$$\mathbf{Q} = \mathbf{R}^{\mathrm{T}}\mathbf{R}. \quad (14)$$

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

$$\mathbf{s}_{\Delta\mathbf{y}}^{(i)} = \mathbf{R}^{\mathrm{T}}\mathbf{s}_{\mathbf{e}}^{(i)}.$$

All vectors generated in that manner are realizations of the random vector $\Delta\boldsymbol{\mathcal{Y}} \sim N(\mathbf{0}, \mathbf{Q})$ representing the colored noise of the observations. Adding the noise vectors to the estimated observations $\hat{\mathbf{y}}$ we get M realizations of the observation vector

$$\mathbf{y}^{(i)} = \hat{\mathbf{y}} + \mathbf{s}_{\Delta\mathbf{y}}^{(i)}. \quad (15)$$

For each of these M realizations of the observations, we compute a sample $\mathbf{s}_{\tilde{\mathbf{x}}}^{(i)}$ of the estimated parameters $\tilde{\mathbf{x}}$ using the active-set method to solve the ICLS problem. Usually, when performing a Monte Carlo simulation to determine the accuracy of an estimate, an empirical variance–covariance matrix is estimated from the parameters

$$\boldsymbol{\Sigma}\{\tilde{\boldsymbol{\mathcal{X}}}\} = E\{(\tilde{\boldsymbol{\mathcal{X}}} - E\{\tilde{\boldsymbol{\mathcal{X}}}\})(\tilde{\boldsymbol{\mathcal{X}}} - E\{\tilde{\boldsymbol{\mathcal{X}}}\})^{\mathrm{T}}\}. \tag{16}$$

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 initial value for the parameters, convergence of the active-set method is usually quite fast.

3.1.2 Optimal ICLS solution

Having obtained 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 paper, 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 OLS estimate or its projection onto the boundary of the feasible set (cf. Sect. 2.2), it is *best* in the sense that it is the solution with the smallest sum of squared residuals of all feasible points.

3.1.3 Confidence regions (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. It gives a quality measure of the estimate in form of a region Ω containing a certain percentage (e.g. 95 %) of the samples

$$P(\mathbf{x} \in \Omega|\mathbf{y}) = 1 - \alpha. \tag{17}$$

$1-\alpha$ 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 one dimensional problem can be computed by sorting the results of the Monte Carlo study and discarding the smallest and biggest $\frac{\alpha}{2}$ percent. This definition is easily extended to the multivariate case (cf. Roese-Koerner et al. 2011).

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.

3.2 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.

3.2.1 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_{a} active constraints

$$\mathbf{B}_{\mathrm{a}}^{\mathrm{T}}\mathbf{x} = \mathbf{b}_{\mathrm{a}}$$

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 an OLS adjustment is done:

$$\mathbf{y} + \mathbf{v} = \mathbf{A}\mathbf{x}, \quad \mathbf{Q}\{\boldsymbol{\mathcal{Y}}\} = \mathbf{Q} \tag{18}$$

$$\mathbf{Q}\{\hat{\boldsymbol{\mathcal{X}}}\} = (\mathbf{A}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{A})^{-1} \tag{19}$$

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{y} = \mathbf{Q}\{\hat{\boldsymbol{\mathcal{X}}}\}\mathbf{A}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{y} \tag{20}$$

$$\hat{\mathbf{v}} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{y}. \tag{21}$$

Afterwards, an adjustment of the equality constraints is carried out.

$$\mathbf{B}_\mathrm{a}^\mathrm{T}(\hat{\mathbf{x}}+\mathbf{r}) = \mathbf{b}_\mathrm{a} \tag{22}$$

$$\mathbf{r} = -\mathbf{Q}\{\hat{\boldsymbol{\mathcal{X}}}\}\mathbf{B}_\mathrm{a}(\mathbf{B}_\mathrm{a}^\mathrm{T}\mathbf{Q}\{\hat{\boldsymbol{\mathcal{X}}}\}\mathbf{B}_\mathrm{a})^{-1}(\mathbf{B}_\mathrm{a}^\mathrm{T}\hat{\mathbf{x}}-\mathbf{b}_\mathrm{a}) \tag{23}$$

$$\tilde{\mathbf{x}} = \hat{\mathbf{x}}+\mathbf{r}. \tag{24}$$

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

$$\hat{s}_1^2 = \frac{\hat{\mathbf{v}}^\mathrm{T}\mathbf{Q}^{-1}\hat{\mathbf{v}}}{n-m} \tag{25}$$

and of the second step

$$\hat{s}_2^2 = \frac{\mathbf{r}^\mathrm{T}\mathbf{Q}\{\hat{\boldsymbol{\mathcal{X}}}\}^{-1}\mathbf{r}}{p_\mathrm{a}} \tag{26}$$

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

$$\mathcal{T} = \frac{\hat{\mathcal{S}}_2^2}{\hat{\mathcal{S}}_1^2} \sim F_{p_\mathrm{a},n-m} \tag{27}$$

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_a, respectively). Now, a level of significance $1-\alpha$ is chosen (e.g. 95 %) 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 small 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 interpretation as stated above.

3.2.2 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 iterations). If at least one constraint is active, then the unconstrained OLS 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.

3.2.3 Sensitivity analysis

In order to determine the influence of each constraint on each parameter, we set the derivative with respect to the parameters $\mathbf{x}$ of the Lagrangian (3b) of the ICLS problem equal to zero

$$\frac{\partial L}{\partial \mathbf{x}^\mathrm{T}} = 2\mathbf{N}\mathbf{x} - 2\mathbf{n} + \mathbf{B}\mathbf{k} \overset{!}{=} 0. \tag{28}$$

Resolving for $\mathbf{x}$, yields

$$\tilde{\mathbf{x}} = \hat{\mathbf{x}} \underbrace{-\frac{1}{2}\mathbf{Q}\{\tilde{\boldsymbol{\mathcal{X}}}\}\mathbf{B}\tilde{\mathbf{k}}}_{\Delta\mathbf{x}} = \hat{\mathbf{x}} + \Delta\mathbf{x}. \tag{29}$$

With this explicit relation between the unconstrained solution $\hat{\mathbf{x}}$, the constrained solution $\tilde{\mathbf{x}}$ and the Lagrange multipliers $\tilde{\mathbf{k}}$ it is possible to do a sensitivity analysis. The perturbation of $\hat{\mathbf{x}}$ consists of three parts:

1. influence of the Lagrange multipliers $\tilde{\mathbf{k}}$,
2. influence of the design $\mathbf{Q}\{\tilde{\boldsymbol{\mathcal{X}}}\}$,
3. influence of the matrix $\mathbf{B}$ 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 (29).

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 $\tilde{k}_i$ is zero, there will be no change in the sum of squared residuals $\Phi(\mathbf{x})$ through the ith constraint. For small values of $\tilde{k}_i$, there will be an effect on $\Phi(\mathbf{x})$, which will be small, whereas for large $\tilde{k}_i$ even small changes to the constraints can result in great changes in $\Phi(\mathbf{x})$. Needless to say, that this is just a rule of thumb as in (29) there still are the influences of the VCV matrix $\mathbf{Q}\{\tilde{\boldsymbol{\mathcal{X}}}\}$ and the constraint matrix $\mathbf{B}$.

3.3 Monte Carlo quadratic programming 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 compute 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 Algorithm 1.

Algorithm 1: A Stochastic framework for ICLS estimates: The MC-QP method.

$[\hat{\mathbf{x}}, \hat{\mathbf{y}}]$ = solveOlsProblem($\mathbf{A}, \mathbf{y}, \mathbf{Q}$)
$[\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{k}}]$ = solveIclsProblem($\mathbf{A}, \mathbf{y}, \mathbf{Q}, \mathbf{B}, \mathbf{b}, \hat{\mathbf{x}}$)
hypothesisTesting($\hat{\mathbf{x}}, \mathbf{A}, \mathbf{Q}, \hat{\mathbf{y}}, \mathbf{B}, \mathbf{b}$)
for $i = 1 : M$ **do**
  generate sample $\mathbf{s}_{\mathcal{Y}}^{(i)}$ of $\mathcal{Y} \sim N(\mathbf{A}\widehat{\mathcal{X}},\ \mathbf{Q})$
  $[\mathbf{s}_{\tilde{\mathbf{x}}}^{(i)}, \mathbf{s}_{\tilde{\mathbf{y}}}, \mathbf{s}_{\tilde{\mathbf{k}}}]$ = solveIclsProblem($\mathbf{s}_{\mathcal{Y}}^{(i)}, \mathbf{A}, \mathbf{Q}, \mathbf{B}, \mathbf{b}, \tilde{\mathbf{x}}$)
end
$\mathbf{f}$ = computeEmpiricalPdf($\mathbf{s}_{\tilde{\mathbf{x}}}$)
deriveConfidenceRegions($\mathbf{f}$)
sensitivityAnalysis($\mathbf{A}, \mathbf{Q}, \mathbf{B}, \tilde{\mathbf{k}}$)
d = computeProbabilityMassInInfeasibleRegion($\mathbf{s}_{\tilde{\mathbf{k}}}$)

First the OLS and ICLS solutions are determined and a 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 $\tilde{\mathcal{X}}$ 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 $\tilde{\mathbf{k}}$. 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.

4 Application of MC-QP method

In this section, three examples will be shown to elucidate the framework for stochastic description of the ICLS estimates. The first two examples are bivariate examples, which are simple problems to explicitly show the features of the MC-QP method. The third example is a multivariate example, which is designed to show mainly the sensitivity analysis capabilities of the MC-QP approach. Further, most of the real-world problems are multivariate and hence, this example will show the full capability of the approach. Stochastic description based on the Bayesian method of Zhu et al. (2005) will also be shown for comparison.

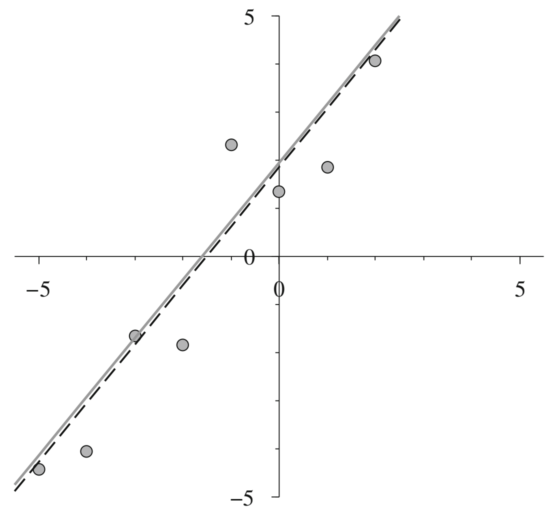


Fig. 3 Illustration of some of the observations (*gray dots*), the OLS line-fit (*gray line*) and the ICLS line-fit (*dashed black line*)

4.1 Line of best fit

The first example that will be illustrated is that of a *line of best fit*, where the parameters that need to be estimated are the *slope* (x_1) and *intercept* (x_2) of the line (Fig. 3).

4.1.1 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 $\mathcal{E} \sim N(\mathbf{0}, \mathbf{I})$:

$$y_i = t_i\, x_1 + x_2 + e_i \tag{30a}$$

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{e}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{30b}$$

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

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 1.30 \\ 1.85 \end{bmatrix}. \tag{30c}$$

Therefore, the constraint matrix becomes $\mathbf{B} = \mathbf{I}$. 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.

Now, the stochastic framework outlined in Sect. 3.3 is applied to this ICLS problem. The numerical results show that the OLS estimate of the slope already satisfies the

Table 1 Numerical results from the line-fitting problem with independent constraints)

	$\hat{\mathbf{x}}$	$\tilde{\mathbf{x}}$
x_1	1.218	1.224
x_2	1.953	1.850

	$\tilde{k}_1 = 0$	$\tilde{k}_2 = 3.474$
Δx_1	0	0.007
Δx_2	0	-0.103

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

The table on the left shows the estimates OLS ($\hat{\mathbf{x}}$) and ICLS ($\tilde{\mathbf{x}}$). The table on the right shows the sensitivity analysis performed with Lagrange multipliers ($\tilde{\mathbf{k}}$), which indicate that there is only one active constraint ($\tilde{k}_2$) and it contributes to the perturbations ($\Delta\mathbf{x}$) 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. 3.2.1)

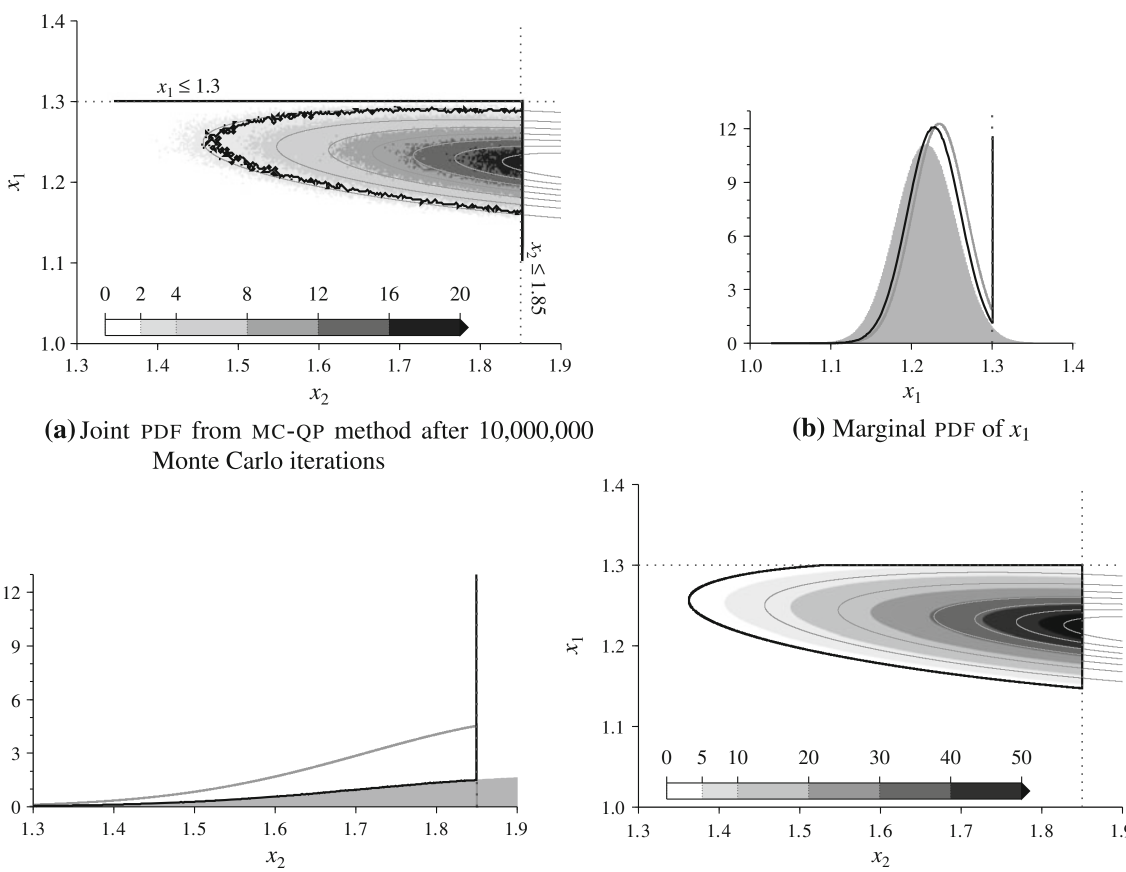
constraint but that of the intercept does not. After applying the ICLS estimation both estimates have been changed (cf. Table 1). Sensitivity analysis of the parameters using the Lagrange multipliers ($\tilde{\mathbf{k}}$) shows that the constraint on the slope is *inactive* ($\tilde{k}_1 = 0$), while the constraint on the intercept is *active* ($\tilde{k}_2 = 3.474$). 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 $\mathbf{A}$. Analytically, this can be explained using (29), where if $\mathbf{B} = \mathbf{I}$ then

$$\Delta\mathbf{x} = -\frac{1}{2}\mathbf{Q}\{\widehat{\mathcal{X}}\}\tilde{\mathbf{k}}. \tag{31}$$

In (31), the covariance matrix $\mathbf{Q}\{\widehat{\mathcal{X}}\}$, describes the correlations between the parameters. If the parameters are correlated in the independent constraints case, they directly affect the perturbations $\Delta\mathbf{x}$.

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 the inactive constraints can be neglected within the estimation process, if they are already known, they are essential for describing the quality of the parameter estimates.

The quality of the estimates from OLS, MC-QP and the Bayesian methods are shown in Fig. 4. The peak of the ICLS



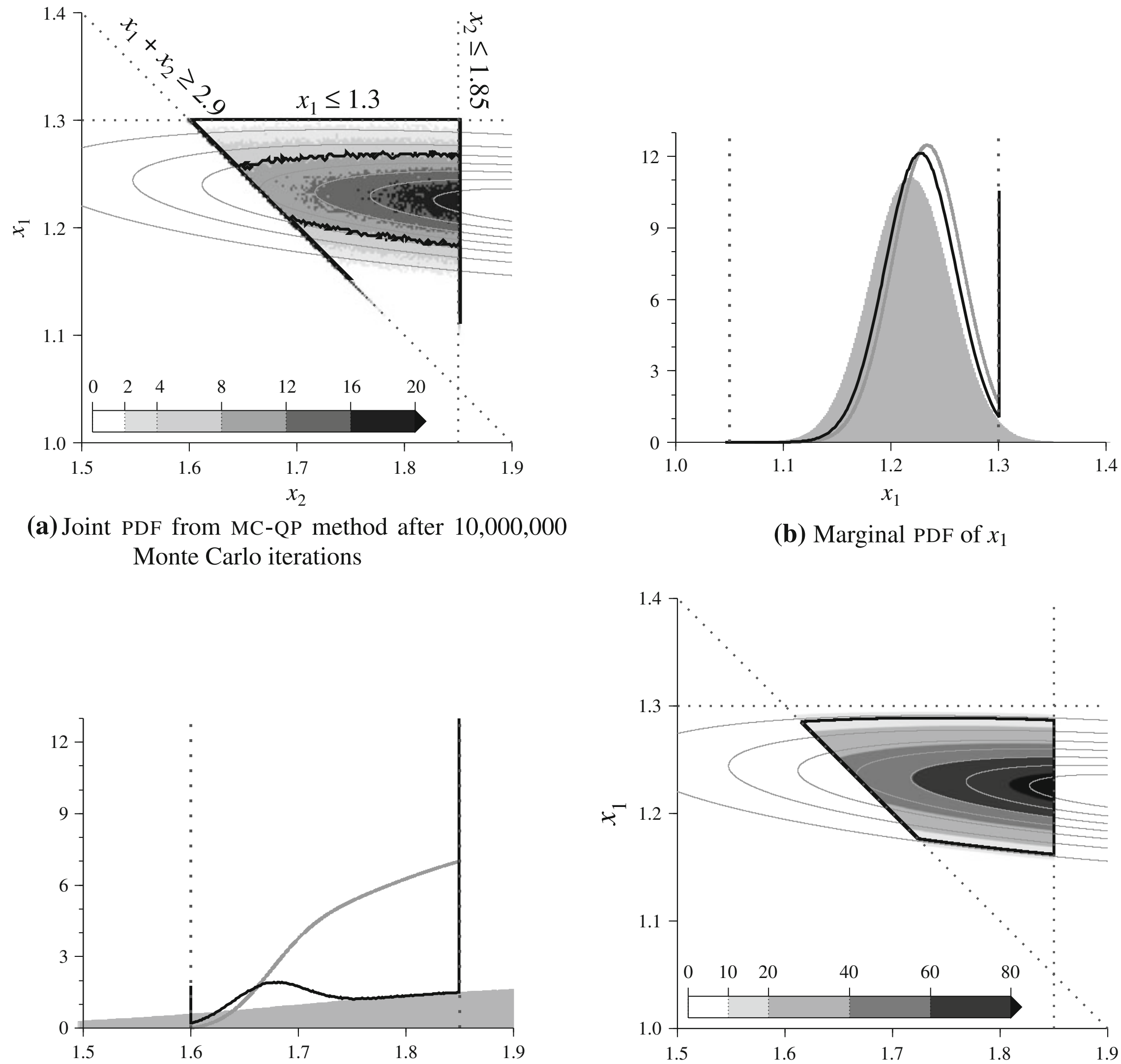
(a) Joint PDF from MC-QP method after 10,000,000 Monte Carlo iterations

(b) Marginal PDF of x_1

(c) Marginal PDF of x_2

(d) Joint PDF from Bayesian method

Fig. 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 OLS 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



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

(b) Marginal PDF of x_1

(c) Marginal PDF of x_2

(d) Joint PDF from Bayesian method

Fig. 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

Table 2 Results from the line-fitting problem with dependent constraints

	$\hat{\mathbf{x}}$	$\tilde{\mathbf{x}}$		$\tilde{k}_1 = 0$	$\tilde{k}_2 = 3.474$	$\tilde{k}_3 = 0$
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: $0.17 < 4.38$

The arrangement of the tables is the same as that of Table 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

PDF of slope (x_1) is slightly shifted from that of the OLS 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 the slope and intercept of the line: if the slope increases the intercept has to decrease and vice-versa. On further scrutiny, a similar shift has not taken place in the PDF of the intercept, whose OLS 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

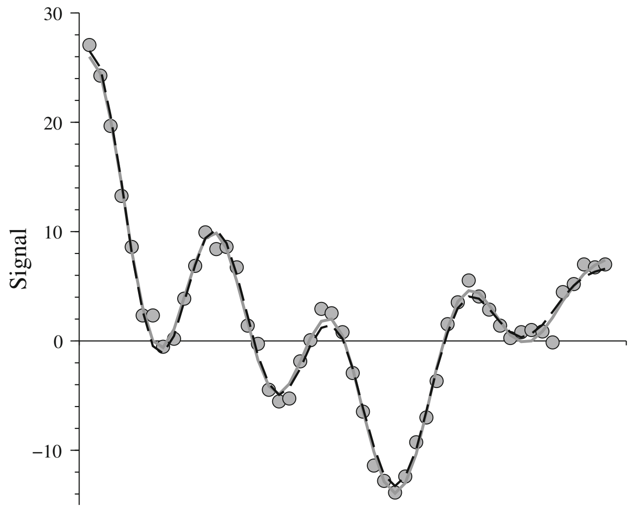


Fig. 6 Observations (*gray dots*), and the OLS (*gray line*) and ICLS (*black line*) fits to the observations of the positive cosine function estimation problem

though the constrained and unconstrained estimates seem to be very similar in this example (cf. Fig. 3), their respective probability densities are entirely different (cf. Fig. 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 (30), the two approaches differ drastically (cf. Fig. 4). The drastic difference is mainly due to the way in which the boundary 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. 2.3.2).

4.1.2 Dependent constraints

An additional constraint is added to the line-fitting problem defined in (30) such that it relates both the parameters, as follows

$$\begin{bmatrix} x_1 \\ x_2 \\ -x_1 - x_2 \end{bmatrix} \leq \begin{bmatrix} 1.30 \\ 1.85 \\ 2.90 \end{bmatrix}. \tag{32}$$

Now the constraint matrix $\mathbf{B}$ 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 (32). 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. Table 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. 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:

Table 3 Results from the ICLS estimation and sensitivity analysis for fitting a positive cosine function

	$\hat{\mathbf{x}}$	$\tilde{\mathbf{x}}$
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

	$\tilde{k}_3 = 38.959$	$\tilde{k}_4 = 18.534$	$\tilde{k}_7 = 72.252$	$\Delta\mathbf{x}$
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: $1.98 < 2.08$

The table on the left shows the estimates from OLS ($\hat{\mathbf{x}}$) and ICLS ($\tilde{\mathbf{x}}$), and the table on the right shows the sensitivity analysis based on the Lagrange multipliers ($\tilde{\mathbf{k}}$). Sensitivity analysis is only shown for the active constraints as the contributions from inactive constraints ($\tilde{k}_0$, $\tilde{k}_1$, $\tilde{k}_2$, $\tilde{k}_5$, $\tilde{k}_6$, $\tilde{k}_8$ and $\tilde{k}_9$) are always zero. Bold values indicate the influence of a constraint on the corresponding parameter. 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. This is a lot closer to the critical value compared to the previous examples

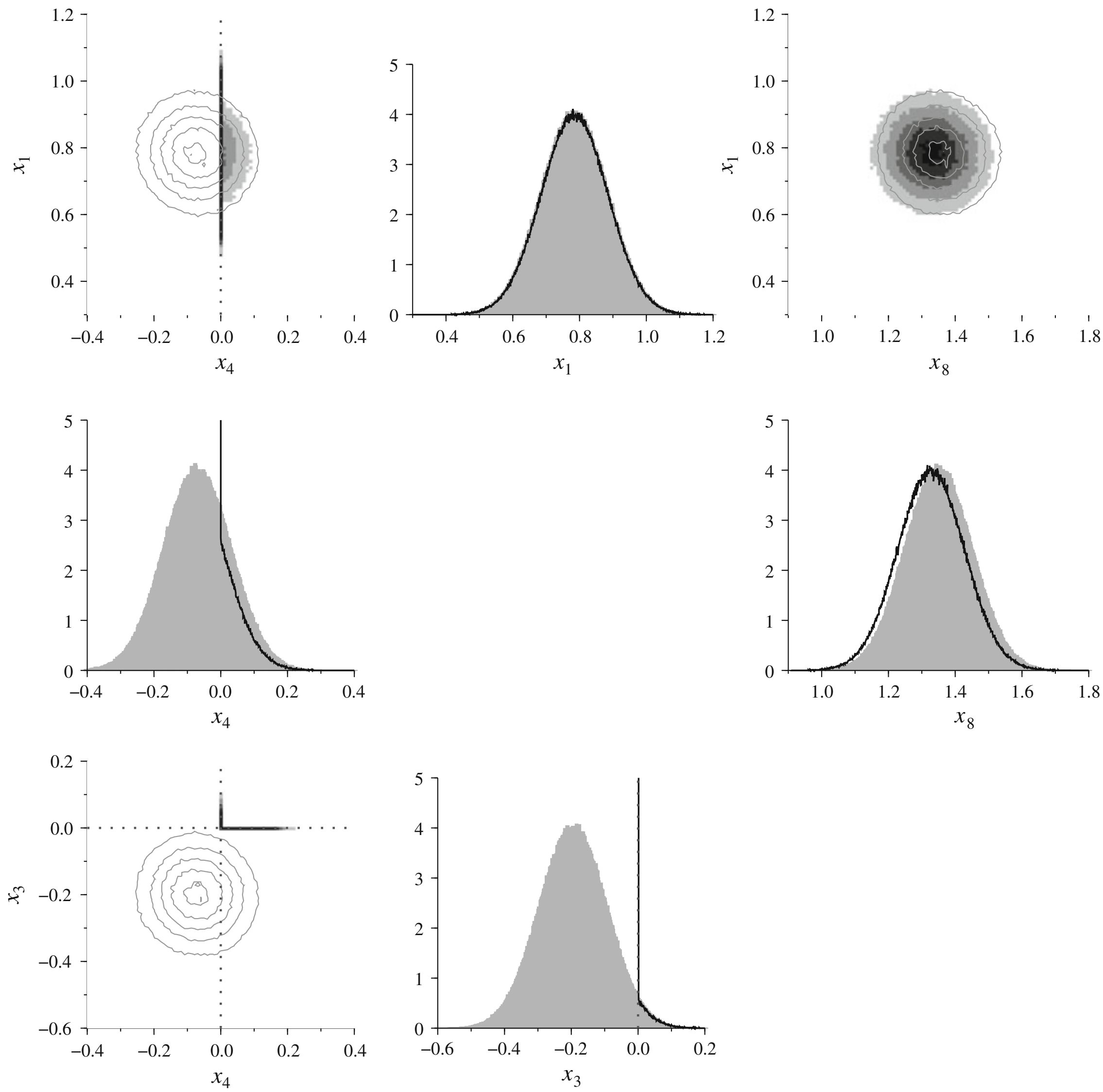


Fig. 7 Probability density functions (joint and marginal) of the positive cosine function estimation problem after 1,000,000 Monte Carlo iterations. Three different pairs of joint PDFs are shown: x_3 & x_4, x_4 & x_1 and x_1 & x_8. The *contour lines* are from the OLS estimate and the *dotted lines* 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 OLS joint PDFs are near circular

The region is far more compact for the dependent constraints than in the independent constraints case.

4.2 Estimation of a positive definite covariance function

The 2-D example with dependent and independent constraints showed the utility of the stochastic framework developed here. In order to demonstrate the utility 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.

The multi-dimensional example that will be used here is that of *fitting a positive cosine expansion to a set of observations* (cf. Fig. 6). Again, the true observations were corrupted by a vector of white noise: $\boldsymbol{\mathcal{E}} \sim N(\mathbf{0}, \mathbf{I})$. Despite the fact that we are dealing with an equally sampled positive cosine function, we have small correlations between some of the parameters as we have not sampled a complete period. 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 the independent inequality constraints enforced on the parameters. Therefore, the observation equation of our model for an arbitrary ith observation and the parametric constraints read as

$$y_i + v_i = \frac{x_0}{2} + x_1 \cos 1\omega t_i + \cdots + x_m \cos m\omega t_i, \quad \text{(33a)}$$
$$x_j \geq 0, \quad j \in [0\ m], \quad \text{(33b)}$$

with the angular frequency $\omega = \frac{2\pi}{T}$, period T, n equally spaced supporting points $t_i \in [0\ 0.5)$ and $m + 1$ unknown parameters $x_0 \ldots x_m$. In matrix form the above equations will read as

$$\mathbf{y} + \mathbf{v} = \mathbf{A}\,\mathbf{x}, \quad \text{(33c)}$$
$$\mathbf{x} \geq 0. \quad \text{(33d)}$$

After applying the stochastic framework to estimate the parameters and the quality of the estimates, the results from OLS show that there are three parameters (x_3, x_4 and x_7) that have negative values, and hence the constraints on these three parameters have become active (cf. Table 3, 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 was in the 2-D case [cf. (31)]. Further, the largest Lagrange multiplier is that of $\tilde{k}_7$ and the corresponding absolute change Δx_7 to the parameter x_7 is the largest as well, which is again explained by (31). Figure 7 shows the range of joint and marginal PDFs that can be expected from a multi-dimensional estimation problem.

Although the method is based on the empirical Monte Carlo approach, a lot of quality information can already be obtained from the quadratic programming algorithms themselves. For example, the Lagrange multipliers can be used to carry out sensitivity analysis, which clearly demarcates the active and inactive constraints and their respective contributions to the perturbation of all the parameters.

5 Summary and conclusions

A framework for the stochastic description of ICLS estimates has been developed. It has been shown that with the proposed MC-QP approach an empirical PDF of the constrained estimate can be determined. The resulting truncated confidence regions are usually smaller (more compact) than the ones of an OLS estimate or the ones computed with the Bayesian method of Zhu et al. (2005). Furthermore, we have given two global measures for the influence of the inequality constraints: the result of a hypothesis testing and the percentage of probability mass outside the feasible region. In addition, it was discussed how the local influence of each constraint on each parameter can be determined using Lagrange multipliers.

The main drawbacks of the existing methods are overcome by this new approach. The concept of projecting infeasible solutions on the boundary of the feasible set instead of neglecting them, leads to a PDF which is—in the opinion of the authors—more realistic than the one from the Bayesian method as all available information (even the part in the infeasible region) is used.

Our method is also more robust to changes in the active constraints (despite the hypothesis testing) than Liew's active constraint approach and also takes into account the influence of inactive constraints on the statistical properties of the estimate. However, carrying out a Monte Carlo approach remains a computationally expensive task.

It was shown, that no disturbance through the constraints take place inside the feasible region (i.e. the PDF in the feasible region is identical to the one of an OLS estimate). All changes due to constraints take place at the boundary of this set. If the optimal estimate lies within that feasible region, then there will be no influence of the constraints on the solution vector itself, but there still will be an influence on its statistical properties. That is because prior knowledge given in the form of inequality constraints can be used to find a smaller confidence region of the desired parameters (due to the concentration of the probability mass in the infeasible region at the boundaries of the feasible set).

The major disadvantage of ICLS problems is the inability to represent the complete stochastic information in form of a variance–covariance matrix as the description of the first two moments of the PDF is no longer sufficient. Therefore, it also difficult to describe correlations between the parameters of the ICLS problem. As can be seen for example in Fig. 5a the introduction of inequality constraints does not change the orientation of the truncated confidence (hyper-)ellipsoids. Therefore, it can be assumed that the correlations in the ICLS case are similar to the ones in the OLS case. However, we have seen in Sect. 4.1.2 that an additional dependency between the parameters can be introduced by the constraints, which might as well be interpreted as a kind of correlation. The determination of correlation will be further investigated in future work.

Acknowledgments The authors are deeply indebted to Jan Martin Brockmann and Professor Karl-Rudolf Koch for valuable discussion and support. All figures were created with the Generic Mapping Toolbox (GMT). The help of three anonymous reviewers is greatly acknowledged.

References

Albertella A, Cazzaniga N, Crespi M, Luzietti L, Sacerdote F, Sansò F (2006) Deformations detection by a Bayesian approach: prior information representation and testing criteria definition. In: Sansò F, Gill AJ (eds) IAG symposium on geodetic deformation monitoring: from geophysical to engineering roles, Springer, Jaén, Spain, pp 30–37

Alkhatib H, Schuh WD (2007) Integration of the Monte Carlo covariance estimation strategy into tailored solution procedures for large-scale least squares problems. J Geodesy 81: 53–66. doi:10.1007/s00190-006-0034-z

Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge

Chen MH, Shao QM (1999) Monte Carlo estimation of Bayesian credible and HPD intervals. J Comput Graph Stat 8(1):69–92

Dantzig G (1998) Linear programming and extensions. Princeton University Press, New Jersey

Fritsch D (1985) Some additional informations on the capacity of the linear complementarity algorithm. In: Grafarend E, Sansò F (eds) Optimization and design of geodetic networks. Springer, Berlin, pp 169–184

Geweke J (1986) Exact inference in the inequality constrained normal linear regression model. J Appl Econ 1(2):127–141

Gill P, Murray W, Wright M (1981) Practical optimization. Academic Press, London

Gill P, Murray W, Wright M (1990) Numerical linear algebra and optimization, vol 1. Addison-Wesley, Reading

Joint Committee for Guides in Metrology (2008) Evaluation of measurement data—supplement 1 to the "Guide to the expression of uncertainty in measurement"—propagation of distributions using a Monte Carlo method. JCGM 101:2008

Koch A (2006) Semantische Integration von zweidimensionalen GIS-Daten und Digitalen Geländemodellen. PhD thesis, University of Hannover, DGK series C, no. 601

Koch KR (1981) Hypothesis testing with inequalities. Bollettino di geodesia e scienze affini 2:136–144

Koch KR (1982) Optimization of the configuration of geodetic networks. In: Proceedings of the international symposium on geodetic networks and computations, München, DGK series B, no. 3

Koch KR (1985) First order design: Optimization of the configuration of a network by introducing small position changes. In: Grafarend E, Sansò F (eds) Optimization and design of geodetic networks. Springer, Berlin, pp 56–73

Koch KR (2007) Parameter estimation and hypothesis testing in linear models. Springer, Berlin

Koch KR (2007) Introduction to Bayesian statistics, 2nd edn. Springer, Berlin

Liew CK (1976) Inequality constrained least-squares estimation. J Am Stat Assoc 71(355):746–751

Peng J, Zhang H, Shong S, Guo C (2006) An aggregate constraint method for inequality-constrained least squares problems. J Geodesy 79: 705–713. doi:10.1007/s00190-006-0026-z

Roese-Koerner L, Devaraju B, Schuh WD, Sneeuw N (2011) Describing the quality of inequality constrained estimates. In: Proceedings of the 1st international workshop on the quality of geodetic observation and monitoring systems, Springer, Garching/München (submitted)

Schaffrin B (1981) Ausgleichung mit Bedingungs-Ungleichungen. Allgemeine Vermessungs-Nachrichten 88. Jg.:227–238

Schaffrin B, Krumm F, Fritsch D (1980) Positiv-diagonale Genauigkeitsoptimierung von Realnetzen über den Komplementaritäts-Algorithmus. In: Conzett R, Matthias H, Schmid H (eds) Ingenieurvermessung 80—Beiträge zum VIII. Internationalen Kurs für Ingenieurvermessung, Ferd. Dümmlers Verlag, ETH Zürich, vol 1, pp A15–A15/19

Song Y, Zhu J, Li Z (2010) The least-squares estimation of adjustment model constrained by some non-negative parameters. Surv Rev 42(315): 62–71. doi:10.1179/003962610X12572516251367

Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc 54(3):426–482

Zhu J, Santerre R, Chang XW (2005) A Bayesian method for linear, inequality-constrained adjustment and its application to GPS positioning. J Geodesy 78: 528–534. doi:10.1007/s00190-004-0425-y