

---

# Describing the Quality of Inequality Constrained Estimates

L. Roese-Koerner, B. Devaraju, W.-D. Schuh, and N. Sneeuw

---

## Abstract

A key feature of geodetic adjustment theory is the description of stochastic properties of the estimated quantities. A variety of tools and measures have been developed to describe the quality of ordinary least-squares estimates, for example, variance-covariance information, redundancy numbers, etc. Many of these features can easily be extended to a constrained least-squares estimate with equality constraints. However, this is not true for inequality constrained estimates.

In many applications in geodesy the introduction of inequality constraints could improve the results (e.g. filter and network design or the regularization of ill-posed problems). This calls for an adequate stochastic modeling accompanying the already highly developed estimation theory in the field of inequality constrained estimation. Therefore, in this contribution, an attempt is made to develop measures for the quality of inequality constrained least-squares estimates combining Monte Carlo methods and the theory of quadratic programming. Special emphasis is placed on the derivation of confidence regions.

---

## Keywords

Confidence regions • Convex optimization • Inequality constrained least-squares • Monte Carlo method • Stochastic modeling

---

## 1 Introduction and Motivation

In many applications in geodesy some bounds or restrictions on the parameters are known in advance. Truncating the parameter space by formulating this knowledge as inequality constraints often helps to improve the results. One can think of the estimation of non-negative variance components for

example or constraints on the power spectral density in the design of decorrelation filters (Roese-Koerner et al. 2012a).

However, besides the process of actually determining a solution of a problem, it is also important to give a measure of its accuracy. In presence of inequality constraints, it is no longer possible to project the uncertainty of the observations to the parameters by applying the law of error propagation. This is, because there is no analytical relationship between observations and parameters. Even if a variance-covariance (VCV) matrix could be obtained, it would not yield a realistic error description, as one has to deal with truncated probability density functions (PDFs).

Up to now, there have been different approaches for a quality description in presence of inequality constraints. For example Liew (1976) proposed to first identify all constraints, which are exactly satisfied (by solving the problem). Afterwards, these constraints are treated as equality constraints, all other constraints are discarded and the VCV

---

L. Roese-Koerner (✉) • W.-D. Schuh  
Institute of Geodesy and Geoinformation, University of Bonn, Bonn,  
Germany  
e-mail: [lutz.roese-koerner@geod.uni-bonn.de](mailto:lutz.roese-koerner@geod.uni-bonn.de);  
[schuh@geod.uni-bonn.de](mailto:schuh@geod.uni-bonn.de)

B. Devaraju • N. Sneeuw  
Institute of Geodesy, University of Stuttgart, Stuttgart, Germany  
e-mail: [devaraju@gis.uni-stuttgart.de](mailto:devaraju@gis.uni-stuttgart.de); [sneeuw@gis.uni-stuttgart.de](mailto:sneeuw@gis.uni-stuttgart.de)

matrix of the equality constrained problem is computed. A disadvantage of this method is, that inactive constraints (e.g. constraints, which do not hold with equality) are neglected and do not constrain the confidence region.

Geweke (1986) and Zhu et al. (2005) treated inequalities as prior information in a Bayesian sense and truncated the probability density functions. To obtain a valid PDF, the function has to be scaled, which could be thought of as distributing the probability mass in the infeasible region over the whole function, which might be not realistic. Furthermore, the numerical evaluation of the PDF is computationally expensive in the multivariate case.

In this paper, we aim at giving a quality description for Inequality Constrained Least-Squares (ICLS) problems using Monte Carlo methods. In contrast to the idea of scaling, we want to obtain a PDF of the estimated parameters, which is identical to the PDF of an Ordinary Least-Squares (OLS) estimate inside the feasible region and where all probability mass in the infeasible region is projected onto its boundaries.

The paper is organized as follows. In Sect. 2 we define the ICLS problem and provide references on several solvers. Our proposed method is described in Sect. 3, as is the derivation of confidence regions and a brief description of a sensitivity analysis for the constraints. In Sect. 4 a case study is carried out to illustrate the application of our approach. The insights that have been gained are summarized in Sect. 5.

## 2 Background

First, the well-known linear OLS estimation model is extended to a linear ICLS estimation. We assume a deterministic model of the form

$$\mathbf{y} + \mathbf{v} = \mathbf{A}\mathbf{x}, \quad (1)$$

with vector of observations  $\mathbf{y}$ , vector of residuals  $\mathbf{v}$ ,  $(n \times m)$  design matrix  $\mathbf{A}$  and vector of unknown parameters  $\mathbf{x}$ .  $n$  is the number of observations,  $m$  the number of parameters. The design matrix is assumed to have full rank  $m$  and all quantities are assumed to be real valued. The (possibly fully populated) VCV matrix  $\mathbf{Q}$  of the observations is assumed to be known. The aim is to minimize the quadratic form

$$\Phi(\mathbf{x}) = \mathbf{v}^T \mathbf{Q}^{-1} \mathbf{v}. \quad (2)$$

Clearly, this aim can be achieved, by applying the usual OLS estimator

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}^{-1} \mathbf{y}. \quad (3)$$

Throughout this paper, symbols marked with a hat refer to unconstrained quantities whereas tildes refer to quantities of a constrained solution.

We now introduce  $p$  inequality constraints in a matrix vector notation:

$$\mathbf{B}^T \mathbf{x} \leq \mathbf{b}. \quad (4)$$

$\mathbf{B}$  is a  $(m \times p)$  matrix of constraints and  $\mathbf{b}$  the corresponding right-hand side. As it is not known in advance which constraints will lead to changes in the parameters (i.d. are exactly satisfied or “active”) the usual techniques of equality constrained estimation can not be applied. However, many algorithms have been developed to solve such an ICLS problem of the form

$$\text{minimize } \Phi(\mathbf{x}) = \mathbf{v}^T \mathbf{Q}^{-1} \mathbf{v} \quad (5a)$$

$$\text{subject to } \mathbf{B}^T \mathbf{x} \leq \mathbf{b} \quad (5b)$$

(which is often referred to as Quadratic Program (QP) as we want to minimize a quadratic objective function  $\Phi(\mathbf{x})$  with respect to some linear constraints). Most of the solvers can be subdivided into two classes: Simplex Methods and Interior Point Methods.

In each iteration of a Simplex method a search direction is computed and projected onto a subset of the constraints. If at least one constraint is active in the solution, the optimal point will be at the boundary of the feasible set (the set where all constraints are satisfied). Therefore, one follows the borderline of the feasible set until the optimal solution is reached. If it is not on the boundary, in the last iteration the projection is neglected, resulting in a step into the interior of the feasible set. Examples for solvers of this type are the Active Set Method (cf. Gill et al. 1981, pp. 167–173) or Dantzig’s Simplex Algorithm for Quadratic Programming (Dantzig 1998, pp. 490–498).

Interior Point Methods on the other hand, substitute the original—possibly hard to solve—problem by a sequence of easier to solve ones. Then a so called “central path” through the interior of the feasible region is followed until the optimal solution is reached. Examples are the Logarithmic Barrier method or primal-dual methods (cf. Boyd and Vandenberghe 2004, pp. 568–571 respectively pp. 609–613).

Other approaches also include the idea of aggregating all inequality constraints into one complex equality constraint (Peng et al. 2006) or transforming (5) into a Linear Complementarity Problem (cf. Koch 2006, pp. 24–25), which can be solved using Lemke’s Algorithm (cf. Fritsch 1985).

As we want to focus on the quality description, we will not pursue the process of actually solving an ICLS problem but refer to the above mentioned authors. All results within this paper were computed using the Active Set Method.

### 3 MC-QP Method

As a VCV matrix is no longer representative in the inequality constrained case, we want to directly propagate the probability density of the observations. Especially in cases, where either no analytical solution is known or it would computationally be very expensive to obtain, Monte Carlo (MC) methods often are used. For the ICLS problem, MC integration seems to be perfectly suited as the probability distribution of the observations is assumed to be known but there is no analytical relationship between observations and parameters.

So the idea is to draw  $M$  samples of the observations, solve the resulting QPs and use them to obtain an empirical probability density function of the estimated parameters.

#### 3.1 Propagation of the Probability Density

Assuming that we have a fully populated VCV matrix  $\mathbf{Q}$  of the observations, we want to draw  $M$  samples  $\mathbf{s}_{\mathcal{Y}}^{(i)}$  of the observations, which follow the normal distribution

$$\mathcal{Y} \sim N(E\{\mathcal{Y}\}, \mathbf{Q}). \quad (6)$$

$E\{\mathcal{Y}\}$  denotes the expectation operator of the random variable  $\mathcal{Y}$ , the random counterpart of the deterministic variable  $\mathbf{y}$ . As estimator of  $E\{\mathcal{Y}\}$  we use an unconstrained OLS estimate

$$\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}} \quad (7a)$$

$$= \mathbf{A}(\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{y}. \quad (7b)$$

The process of sampling from the above described distribution can be done as follows (cf. Koch 2007, p. 197): First  $\mathbf{Q}$  is factorized into the product of two upper right triangular matrices  $\mathbf{R}$  using the Cholesky factorization

$$\mathbf{Q} = \mathbf{R}^T\mathbf{R}. \quad (8)$$

Afterwards,  $M$  standard normal distributed samples  $\mathbf{s}_{\mathcal{E}}^{(i)}$  are drawn from the distribution

$$\mathcal{E} \sim N(\mathbf{0}, \mathbf{I}), \quad (9)$$

with identity matrix  $\mathbf{I}$  and transformed to

$$\mathbf{s}_{\mathcal{Y}}^{(i)} = \hat{\mathbf{y}} + \mathbf{R}^T\mathbf{s}_{\mathcal{E}}^{(i)}, \quad i = 1, \dots, M. \quad (10)$$

These samples are used as input for the quadratic program (5), which is solved using the Active Set Method, producing  $M$  solution  $\mathbf{s}_{\tilde{\mathcal{X}}}^{(i)}$ . Hence, we can achieve an

empirical joint density function of the parameters  $\tilde{\mathcal{X}}$  by computing and normalizing the histogram of the solutions.

One can think of this approach as computing  $M$  different instances of the problem, resulting in  $M$  different objective functions with  $M$  different minima. However, as the constraints remain unchanged, all minima inside the feasible region coincide with the solution of an OLS estimation. All solutions outside the feasible region on the contrary are projected to the closest point (in the metric, given through the objective function) that fulfills all constraints. As all these points will be at the boundary of the feasible set, this will lead to the accumulation of probabilities, described in Sect. 1.

The task of solving  $M$  different QPs is computational demanding. However, as with the solution of the original problem a good initial solution is known, QP algorithms will converge fast.

#### 3.2 Confidence Regions (HPD Regions)

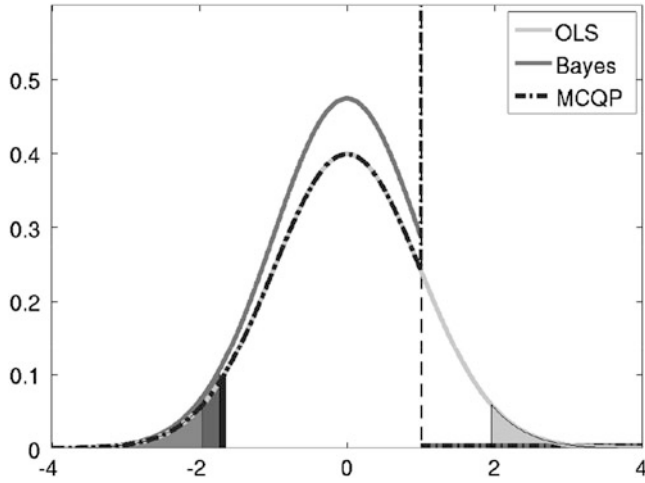
As with the MC-QP approach we not only obtain a point estimate but a whole empirical PDF, we can easily compute confidence intervals. In Chen and Shao (1999) and the Supplement 1 to the ‘‘Guide to the expression of uncertainty in measurement’’ (GUM, ISO 2008, pp. 5 and 30) the confidence interval of a Monte Carlo estimate (called highest probability density (HPD) region) is defined as the shortest interval, which contains  $1 - \alpha$  percent of the data (with  $1 - \alpha$  being the level of significance). This definition extends naturally to the  $n$ -dimensional case:

The  $(1 - \alpha)$ -confidence region  $\Omega$  of an  $n$ -dimensional problem is defined as the smallest region containing  $1 - \alpha$  percent of the data

$$P\{\mathbf{x}|\Omega\} = 1 - \alpha. \quad (11)$$

This region can be computed by simply sorting the values of the  $n$ -dimensional histogram described in Sect. 3.1 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. One has to be aware that this region is not necessarily connected, due to the accumulation of probability mass at the boundary of the feasible region (Chen and Shao 1999, p. 84).

Figure 1 illustrates such confidence intervals for a one dimensional example for  $M \rightarrow \infty$ . The PDF of the OLS estimate with  $E\{\mathcal{X}\} = 0$  is plotted in light gray. The  $\alpha$  percent, which are 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.



**Fig. 1** Probability density functions and confidence intervals of different estimates for a one dimensional problem: 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

The PDF of the MC-QP estimate 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  $\alpha$  percent not included in the confidence interval are at the left tail of the PDF (depending on the constraint). So the confidence interval is bounded on one side by the constraint and on the other side by the  $\alpha$  percent that are “most unlikely”. As can be seen in Fig. 1, 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.

### 3.3 Influence of the Constraints

So far, we investigated the distribution of the estimated parameters and their confidence region. However, it might be also of interest, to determine the influence of the constraints onto the parameters. This can be done either on a global level, determining if the overall change in the result due to the constraints is significant or at a local level, investigating the individual influence of each constraint on each parameter. Due to the limited space, we will only very briefly discuss

the different options for such a sensitivity analysis and provide some references for further reading. A more detailed discussion could for example be found in Roese-Koerner et al. (2012b).

On a global scale, one can perform a hypothesis testing. Here, the sum of squared residuals of the unconstrained OLS estimate is compared with the sum of squared changes through the constraints (Koch 1981). Another global measure is the ratio of the probability mass inside and outside the feasible region (measured by checking in each Monte Carlo iteration if at least one constraint is active or not).

To analyse the local influence, the Lagrangian

$$L(\mathbf{x}, \mathbf{k}) = \Phi(\mathbf{x}) + \mathbf{k}^T (\mathbf{B}^T \mathbf{x} - \mathbf{b}) \quad (12)$$

of the ICLS problem (5) is needed. It is the sum of the original objective function  $\Phi(\mathbf{x})$  and the rearranged constraints multiplied with the Lagrange multipliers  $\mathbf{k}$ . The Lagrange multipliers of the optimal solution can be determined and give a measure for the “activeness” of a constraint. This can be used to quantify the influence of each constraint on each parameter (cf. Boyd and Vandenberghe 2004, p. 252).

## 4 Case Study

In this case study, the MC-QP method is applied to derive stochastic information of some quantities of a very simple ICLS problem: the estimation of a line of best fit with a constrained minimal intercept. We have intentionally chosen a simple problem to focus on the methodological aspects of the MC-QP method and on the comparison of different confidence regions.

Assume the following uncorrelated and normal distributed observations  $\mathbf{y}$  to be measured at the supporting points  $\mathbf{t}$ :

$$\mathbf{y} = [-4.0 \ 0.0 \ 1.5 \ 3.0 \ 2.0]^T, \quad \mathbf{t} = [1 \ 2 \ 3 \ 4 \ 5]^T.$$

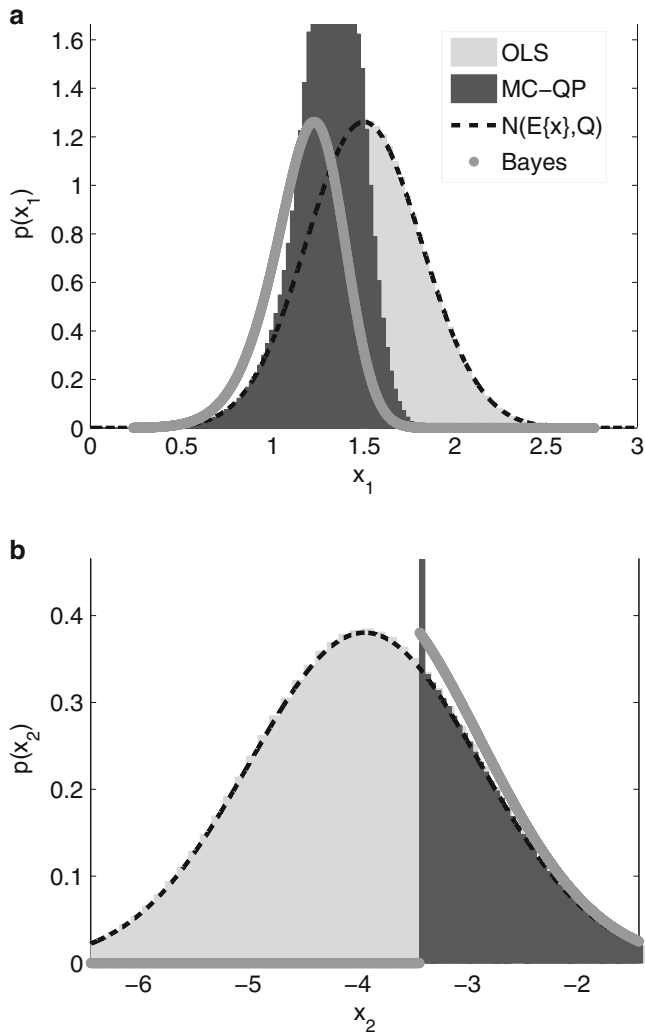
The deterministic model reads

$$y_i + v_i = x_1 t_i + x_2. \quad (13)$$

The parameter space should be constrained, so that only intercepts of at least  $-3.5$  are accepted:

$$x_2 \geq -3.5 \iff \underbrace{\begin{bmatrix} 0 & -1 \end{bmatrix}}_{\mathbf{B}^T} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} \leq \underbrace{[3.5]}_{\mathbf{b}}. \quad (14)$$

Therefore, we have an ICLS problem in the form of (5) and can apply the MC-QP method. The unconstrained OLS solution reads



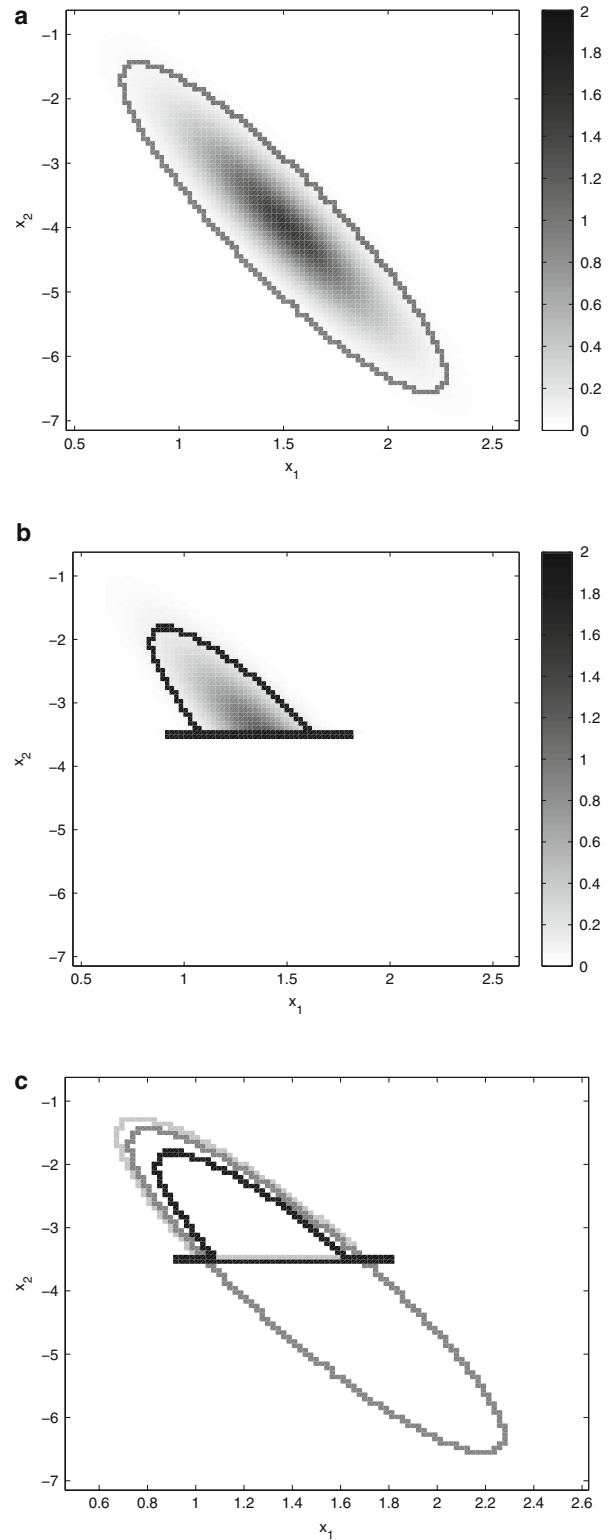
**Fig. 2** Empirical marginal densities of the parameters after  $M = 10,000,000$  Monte Carlo iterations. The PDF of the Monte Carlo estimates are plotted in *light gray* (OLS) and *dark gray* (ICLS), the *dashed line* is the analytical PDF of the OLS estimate and the PDF of the Bayesian approach is plotted in *gray*. (a) Estimates of the marginal density of  $x_1$ , (b) estimates of the marginal density of  $x_2$

$$\hat{\mathbf{x}} = \begin{bmatrix} 1.5 \\ -4.0 \end{bmatrix} \quad (15)$$

and the ICLS solution, which was obtained using the Active Set Method, reads

$$\tilde{\mathbf{x}} = \begin{bmatrix} 1.3636 \\ -3.5000 \end{bmatrix}. \quad (16)$$

Comparison of the marginal densities of parameter  $x_2$ , which are illustrated in Fig. 2b, shows, that the MC-QP estimate (dark gray bars) is nearly identical to the OLS estimate (light gray bars) inside the feasible region. All probability mass



**Fig. 3** Empirical joint densities and confidence regions of the parameters after  $M = 10,000,000$  Monte Carlo iterations. (a) OLS: joint PDF and confidence region, (b) MC-QP: joint PDF and confidence region, (c) confidence regions of Bayesian (*light gray*), OLS (*darker gray*) and MC-QP approach (*black*)

left of the constraint is projected onto the boundary of the feasible set, resulting in a peak at  $x_2 = -3.5$ . The Bayesian estimate (gray curve) is a scaled version of the analytical PDF of the OLS estimate (dashed curve) inside the feasible region.

The more “peaky” form and the shift of the maximum of the MC-QP estimate of parameter  $x_1$  (Fig. 2a) results from correlations between the parameters. The shift is even stronger in the Bayesian approach. Figure 3a illustrates the joint densities of the OLS estimate and the corresponding 95 % confidence region. The joint PDF of the MC-QP estimate is shown in Fig. 3b and is identical to the OLS estimate inside the feasible region. Here, the accumulation on the boundary can be seen as well. In Fig. 3c the confidence regions of the different estimates are compared. As discussed in Sect. 3.2, the confidence region of the MC-QP estimate (black) becomes smallest due to the accumulation of the probability mass on the boundary. On the contrary, applying the Bayesian method (light gray) leads to a bigger confidence region due to the scaling of the PDF. In this case study, the confidence region of the Bayes estimate is even bigger than the one of the OLS estimate (darker gray) because a huge part of the PDF is truncated.

## 5 Summary and Outlook

The proposed MC-QP method allows a stochastic description of ICLS estimates but it is computationally expensive to apply. It was shown that the introduction of inequality constraints within this framework leads to smaller confidence regions.

The possibilities of a sensitivity analysis (which were only mentioned briefly here) as well as the determination of the influence of constraints on correlations between parameters are to be addressed in future work.

## References

- 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, Princeton
- 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/Heidelberg/New York/Tokyo, pp 169–184
- Geweke J (1986) Exact Inference in the inequality constrained normal linear regression model. *J Appl Econ* 1(2):127–41
- Gill P, Murray W, Wright M (1981) Practical optimization. Academic Press, London
- ISO (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, Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, Geneva
- Koch A (2006) Semantische Integration von zweidimensionalen GIS-Daten und Digitalen Geländemodellen. Ph.D. thesis, Universität Hannover
- Koch KR (1981) Hypothesis testing with inequalities. *Bollettino di geodesia e scienze affini* 2:136–144
- 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, Krasbutter I, Schuh WD (2012a) A constrained quadratic programming technique for data-adaptive design of decorrelation filters. In: Sneeuw N, Novak P, Crespi M, Sansò F, Sideris MG (eds) VII Hotine-Marussi Symposium on Mathematical Geodesy, Rome, 2009, Springer, International Association of Geodesy Symposia, vol 137, pp 165–170. DOI 10.1007/978-3-642-22078-4\_25
- Roese-Koerner L, Devaraju B, Sneeuw N, Schuh WD (2012b) A stochastic framework for inequality constrained estimation. *J Geodesy* 86(11):1005–1018. Doi:10.1007/s00190-012-0560-9
- 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