*Journal of Global Optimization* **28:** 121–137, 2004.



# A Rigorous Lower Bound for the Optimal Value of Convex Optimization Problems

CHRISTIAN JANSSON
*Institute of Computer Science III, Technical University Hamburg-Harburg, Schwarzenbergstraße 95, 21071 Hamburg, Germany. (e-mail: jansson@tu-harburg.de)*

(Received 23 December; accepted in revised form 28 April 2003)

**Abstract.** In this paper, we consider the computation of a rigorous lower error bound for the optimal value of convex optimization problems. A discussion of large-scale problems, degenerate problems, and quadratic programming problems is included. It is allowed that parameters, which define the convex constraints and the convex objective functions, may be uncertain and may vary between given lower and upper bounds. The error bound is verified for the family of convex optimization problems which correspond to these uncertainties. It can be used to perform a rigorous sensitivity analysis in convex programming, provided the width of the uncertainties is not too large. Branch and bound algorithms can be made reliable by using such rigorous lower bounds.

**Mathematics Subject Classifications.** 90C25, 90C06, 65G30, 49Q12

**Key words:** Convex programming, Convex relaxations, Global optimization, Interval arithmetic, Large-scale problems, Quadratic programming, Rigorous error bounds, Sensitivity analysis

## 1. Introduction

Convex optimization plays an important role in practice. Firstly, many applications can be modelled as linear or convex problems. Secondly, convex optimization is essentially applied in global optimization, where so-called convex relaxations are solved sequentially within branch and bound frameworks. In order to discard subproblems containing no global optimal points, a lower bound of the optimal value for the convex relaxation is required. There are several books which give substantial attention to relaxation techniques for mixed integer nonlinear programming problems and are recommended to readers. These include Floudas [3], Tawaralani and Sahinidis [15], and the Encyclopedia of Optimization [4].

On a computer rounding errors occur which may effect the computed approximation of an optimal solution. This effect depends on the algorithm and on the problem. Especially ill-conditioned problems may influence the computed approximation drastically, yielding a non-backward-stable approximation; that is, the approximation is not the exact solution of a slightly perturbed problem, because data dependencies, due to the coefficients and the functions which define the convex problem, may allow only special perturbations on a small nonlinear manifold of the input space.

Ill-conditioning is not a rare phenomenon in practice, even in the linear case. In a recent study of Ordóñez and Freund [12] it is stated that 72% of the lp-instances in the NETLIB Linear Programming Library [8] (which contains many industrial problems) are ill-conditioned. After applying CPLEX 7.1 presolve (a preprocessing heuristic for linear programming problems) 19% maintain the property of being ill-conditioned. Hence, the computation of rigorous error bounds may be useful in practice.

For linear mixed integer programming problems with exact input data it is shown in a recent preprint of Neumaier and Shcherbina [11] how, the use of a careful pre- and postprocessing together with interval arithmetic, can guarantee safe forward error bounds for the solution. In [6] results for rigorously solving linear programming problems with uncertain input data and unbounded variables are contained. For the case of exact input data and bounded variables some results of Section 3 in Neumaier and Shcherbina [11] and Section 6 in [6] coincide. Examples, where commercial lp-solvers failed, are presented in both preprints.

Our major goal is to show that similar results can be obtained in the nonlinear convex case. It turns out that a rigorous lower bound for the optimal value of a convex programming problem can be computed by postprocessing only the output of a nonlinear solver. An engagement into the code of the nonlinear solver is not necessary. It is shown that the error bound is very sharp provided that the computed approximations are close to a Karush-Kuhn-Tucker point. A consequence is that also non state of the art solvers (which may produce more frequently wrong approximations) can be used in a safe manner by judging the output with the error bound. This postprocessing algorithm needs in most applications only a fraction of the computational work, which is required by the nonlinear solver. Moreover, the algorithm is also applicable for sparse problems, degenerate problems, and problems with uncertain input data.

The paper is organized as follows. Section 2 contains notation and elementary definitions of interval arithmetic. In Section 3 the basic theory is presented, and in the following section, as a special case, quadratic programming problems are considered. Section 5 presents algorithms for computing a rigorous lower bound, and in Section 6 an illustrative example is studied. In Section 7, some remarks are given how infeasibility of convex optimization problems can be rigorously verified. Finally, some conclusions are given.

## 2. Notation

Throughout this paper we use the following notation. $\mathbf{R}$, $\mathbf{R}^n$, $\mathbf{R}^n_+$, and $\mathbf{R}^{m\times n}$ denote the sets of real numbers, real vectors, real nonnegative vectors, and real $m \times n$

matrices, respectively. Comparisons $\leqslant$, absolute value $|.|$, min, max, inf and sup are used entrywise for vectors and matrices.

The coefficients of a real $m\times n$ matrix $A$ are denoted by $A_{ij}$, its columns by $A_{:j}$, its rows by $A_{i:}$, and its transpose by $A^T$. For subsets $I, J$ of indices $A_{:J}$ is the submatrix of $A$ with columns $A_{:j}$ where $j\in J$, $A_{I:}$ is the submatrix of $A$ with rows $A_{i:}$ where $i\in I$, and $A_{IJ}$ is the submatrix of $A$ with coefficients $A_{ij}$ where $i\in I$ and $j\in J$.

We require only some elementary definitions about interval arithmetic which are described here. There is a number of textbooks on interval arithmetic and self-validating methods which can be highly recommended to readers. These include Alefeld and Herzberger [1], Moore [7], and Neumaier [9,10].

If $\mathbf{V}$ is one of the spaces $\mathbf{R}$, $\mathbf{R}^n$, $\mathbf{R}^{m\times n}$, and $\underline{v}, \overline{v}\in\mathbf{V}$, then the box

$$\mathbf{v} := [\underline{v}, \overline{v}] := \{v\in\mathbf{V} : \underline{v}\leqslant v\leqslant\overline{v}\} \tag{1}$$

is called an *interval quantity* in $\mathbf{IV}$ with *lower bound* $\underline{v}$ and *upper bound* $\overline{v}$. In particular, $\mathbf{IR}$, $\mathbf{IR}^n$, and $\mathbf{IR}^{m\times n}$ denote the set of real intervals $\mathbf{a} = [\underline{a}, \overline{a}]$, the set of real interval vectors $\mathbf{x} = [\underline{x}, \overline{x}]$, and the set of real interval matrices $\mathbf{A} = [\underline{A}, \overline{A}]$, respectively. The real operations $A\circ B$ with $\circ\in\{+,-,\cdot,/\}$ between real numbers, real vectors and real matrices can be generalized to *interval operations*. The result $\mathbf{A}\circ\mathbf{B}$ of an interval operation is defined as the interval hull of all possible real results, that is

$$\mathbf{A}\circ\mathbf{B} := \cap\{\mathbf{C}\in\mathbf{IV} : A\circ B\in\mathbf{C} \quad \text{for all} \quad A\in\mathbf{A}, B\in\mathbf{B}\}. \tag{2}$$

All interval operations can be easily executed by working appropriately with the lower and upper bounds of the interval quantities. For example, in the simple case of addition, we obtain

$$\mathbf{A}+\mathbf{B} = [\underline{A}+\underline{B}, \overline{A}+\overline{B}]. \tag{3}$$

Interval multiplications and divisions require a distinction of cases. For interval quantities $\mathbf{A}, \mathbf{B}\in\mathbf{IV}$ we define

$$\mathrm{mid}\mathbf{A} := (\underline{A}+\overline{A})/2 \quad \text{as the } \textit{midpoint}, \tag{4}$$

$$\mathrm{rad}\mathbf{A} := (\overline{A}-\underline{A})/2 \quad \text{as the } \textit{radius}, \tag{5}$$

$$|\mathbf{A}| := \sup\{|A| : A\in\mathbf{A}\} \quad \text{as the } \textit{absolute value}, \tag{6}$$

$$\mathbf{A}^+ := \max\{0, \overline{A}\}, \tag{7}$$

$$\mathbf{A}^- := \min\{0, \underline{A}\}. \tag{8}$$

Moreover, the comparison in $\mathbf{IV}$ is defined by

$$\mathbf{A}\leqslant\mathbf{B} \quad \text{iff} \quad \overline{A}\leqslant\underline{B},$$

and other relations are defined analogously. Real quantities $v$ are embedded in the interval quantities by identifying $v = \mathbf{v} = [v, v]$.

In interval arithmetic linear systems of equations with inexact input data are treated by working with an interval matrix $\mathbf{A} \in \mathbf{IR}^{n\times n}$ and a right hand side $\mathbf{b} \in \mathbf{IR}^n$. Frequently, the aim is to compute an interval vector $\mathbf{x} \in \mathbf{IR}^n$ containing the *solution set*

$$\Sigma(\mathbf{A},\mathbf{b}) := \{x \in \mathbf{R}^n : Ax = b \text{ for some } A \in \mathbf{A},\, b \in \mathbf{b}\} \tag{9}$$

This is an NP-hard problem, but there are several methods that compute enclosures $\mathbf{x}$ with $O(n^3)$ operations for certain types of interval matrices. A precise description of such methods, required assumptions, and approximation properties can be found, for example, in Neumaier [9]. Roughly speaking, it turns out that for interval matrices with $\|I - R\mathbf{A}\| < 1$ ($R$ is an approximate inverse of the midpoint mid $\mathbf{A}$) there are several methods which compute an enclosure $\mathbf{x}$, and the radius rad$\mathbf{x}$ decreases linearly with decreasing radii rad$\mathbf{A}$ and rad$\mathbf{b}$. For the computation of enclosures in the case of large-scale linear systems the reader is referred to Rump [14].

## 3. Rigorous Lower Bound of the Optimal Value

We consider the convex optimization problem

$$\begin{aligned} &\min f(x)\\ &\text{s.t. } G(x) \leqslant 0\\ &\quad\;\; Hx = h\\ &\quad\;\; \underline{x} \leqslant x \leqslant \overline{x} \end{aligned} \tag{10}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $G : \mathbf{R}^n \to \mathbf{R}^m$, $H \in \mathbf{R}^{l\times n}$, $h \in \mathbf{R}^l$, and $f$ and all components $G_i$ of $G$ are convex functions. The simple bounds $\underline{x}_j < \overline{x}_j$ may also be infinite. We partition the set of indices $\{1, \ldots, n\}$ into the four sets $J^{\pm\infty}$, $J^{+\infty}$, $J^{-\infty}$, and $J$ denoting the indices where both simple bounds are infinite, the simple upper bounds are infinite, the simple lower bounds are infinite, and both simple bounds are finite, respectively.

The optimal value is denoted by $f^*$, and optimal points are denoted by $x^*$. We assume that for the functions $f$ and $G$ the gradients, or in the nonsmooth case subgradients, are available, which we denote by $\nabla f(x)$ and $\nabla G(x) = (\nabla G_1(x), \ldots, \nabla G_m(x))$. The convexity implies that for each $\tilde{x} \in \mathbf{R}^n$ the inequalities

$$f(x) \geqslant f(\tilde{x}) + \nabla f(\tilde{x})^T (x - \tilde{x}) \quad \text{for } x \in \mathbf{R}^n, \tag{11}$$

and

$$G(x) \geqslant G(\tilde{x}) + \nabla G(\tilde{x})^T (x - \tilde{x}) \quad \text{for } x \in \mathbf{R}^n, \tag{12}$$

are satisfied.

The following lemma provides a (possibly infinite) lower bound of the optimal value for problem (10).

LEMMA 1. *Let $\tilde{x} \in \mathbf{R}^n$, $\tilde{y} \in \mathbf{R}^m_+$, $\tilde{z} \in \mathbf{R}^l$, and define the defect*

$$d := \nabla f(\tilde{x}) + \nabla G(\tilde{x}) \cdot \tilde{y} - H^T \tilde{z}. \tag{13}$$

*Then*

$$\begin{aligned} \underline{f}^* := {} & f(\tilde{x}) + (G(\tilde{x}) - \nabla G(\tilde{x})^T \tilde{x})^T \tilde{y} + h^T \tilde{z} \\ & + \underline{x}^T d^+ + \overline{x}^T d^- - \nabla f(\tilde{x})^T \tilde{x} \end{aligned} \tag{14}$$

*is a lower bound of $f^*$, that is $\underline{f}^* \leqslant f^*$.*

*Proof.* The inequalities (11) and (12) provide affine lower bound functions of $f$ and $G$, and by replacing in (10) the convex functions $f$ and $G_i$ for $i = 1, \ldots, m$ by their affine lower bound functions, we obtain a linear programming problem

$$\begin{aligned} \min\ & f(\tilde{x}) + \nabla f(\tilde{x})^T (x - \tilde{x}) \\ \text{s.t. } & G(\tilde{x}) + \nabla G(\tilde{x})^T (x - \tilde{x}) \leqslant 0 \\ & Hx = h \\ & \underline{x} \leqslant x \leqslant \overline{x}, \end{aligned} \tag{15}$$

which has the property that each feasible solution for problem (10) is feasible for (15), and the optimal value of (15) is a lower bound for the optimal value of (10).

This linear program can be written in the form

$$\begin{aligned} \min\ & \nabla f(\tilde{x})^T x + (f(\tilde{x}) - \nabla f(\tilde{x})^T \tilde{x}) \\ \text{s.t. } & \nabla G(\tilde{x})^T x \leqslant -G(\tilde{x}) + \nabla G(\tilde{x})^T \tilde{x} \\ & Hx = h \\ & \underline{x} \leqslant x \leqslant \overline{x}, \end{aligned}$$

The corresponding dual problem is

$$\begin{aligned} \max\ & (G(\tilde{x}) - \nabla G(\tilde{x})^T \tilde{x})^T y + h^T z + \\ & \underline{x}^T u - \overline{x}^T v + f(\tilde{x}) - \nabla f(\tilde{x})^T \tilde{x} \\ \text{s.t. } & -\nabla G(\tilde{x}) y + H^T z + u - v = \nabla f(\tilde{x}) \\ & y \geqslant 0,\ u \geqslant 0,\ v \geqslant 0. \end{aligned} \tag{16}$$

If we define

$$\tilde{u} := d^+, \quad \tilde{v} := -d^-,$$

then $\tilde{u} \geqslant 0$, $\tilde{v} \geqslant 0$, and from $\tilde{y} \geqslant 0$ we obtain immediately that $\tilde{y}, \tilde{z}, \tilde{u}, \tilde{v}$ are feasible for the dual problem with the objective value $\underline{f}^*$, which is by duality theory less than or equal $f^*$. □

Notice that this lower bound may also be infinite, because the infinity of the simple bounds may yield infinite terms $\underline{x}^T d^+$ or $\overline{x}^T d^-$. Especially, it follows that infinite terms are avoided, if $d_j \leqslant 0$ for $j \in J^{-\infty}$, $d_j \geqslant 0$ for $J^{+\infty}$ and $d_j = 0$ for $j \in J^{\pm\infty}$. Hence, the quality of this lower bound depends not only on the quality of the approximation $\tilde{x}$, but also on the defect $d$, and therefore on the quality of $\tilde{y}$ and $\tilde{z}$.

Frequently, in applications optimization problems are defined by functions which depend on uncertain parameters. We model these uncertainties by intervals; that is, we assume that the functions depend on parameters $p$, and $p \in \mathbf{p} \in \mathrm{I}\mathbf{R}^k$. This yields a family of optimization problems

$$\begin{aligned} &\min f(x;p) \\ &\text{s.t. } G(x;p) \leqslant 0 \\ &\quad H(p)x = h(p) \\ &\quad \underline{x} \leqslant x \leqslant \overline{x}, \end{aligned} \tag{17}$$

depending on $p \in \mathbf{p}$, where we assume:

(a) For every $p \in \mathbf{p}$ the functions $f(x;p)$ and $G(x;p)$ are convex.
(b) For each fixed $\tilde{x} \in [\underline{x}, \overline{x}]$ enclosures

$$\begin{aligned} &\{f(\tilde{x};p) : p \in \mathbf{p}\} \subseteq \mathbf{f} \in \mathrm{I}\mathbf{R}, \\ &\{G(\tilde{x};p) : p \in \mathbf{p}\} \subseteq \mathbf{G} \in \mathrm{I}\mathbf{R}^m, \\ &\{\nabla f(\tilde{x};p) : p \in \mathbf{p}\} \subseteq \nabla\mathbf{f} \in \mathrm{I}\mathbf{R}^n, \\ &\{\nabla G(\tilde{x};p) : p \in \mathbf{p}\} \subseteq \nabla\mathbf{G} \in \mathrm{I}\mathbf{R}^{n\times m} \end{aligned} \tag{18}$$

can be computed.
(c) The coefficients of the linear equation are allowed to vary within intervals; that is $H(p) \in \mathbf{H} \in \mathrm{I}\mathbf{R}^{l\times n}$ and $h(p) \in \mathbf{h} \in \mathrm{I}\mathbf{R}^l$.

The enclosures (18) can be calculated on a computer by means of interval arithmetic and automatic differentiation (see Griewank [5], and Rall and Corliss [13]). We mention that in the backward mode the computational costs for computing the gradient of $G$ is at most five times the cost for one function evaluation of $G$. This property holds true independently of the dimension $n$.

The following theorem provides the theoretical basis for algorithms computing a rigorous lower bound of the optimal values for the family (17) of convex optimization problems.

THEOREM 1. *Let* $\tilde{x} \in \mathbf{R}^n$, $\mathbf{y} \in \mathrm{I}\mathbf{R}^m$ *with* $\mathbf{y} \geqslant 0$, $\mathbf{z} \in \mathrm{I}\mathbf{R}^l$, *and let*

$$\mathbf{d} := \nabla\mathbf{f} + \nabla\mathbf{G} \cdot \mathbf{y} - \mathbf{H}^T \cdot \mathbf{z}. \tag{19}$$

*Suppose further that*

(*i*) $\mathbf{d}_{J-\infty} \leqslant 0$, $\mathbf{d}_{J+\infty} \geqslant 0$, *and*
(*ii*) *for every* $\nabla f \in \nabla\mathbf{f}$, $\nabla G \in \nabla\mathbf{G}$, *and* $H \in \mathbf{H}$ *there exists* $y \in \mathbf{y}$ *and* $z \in \mathbf{z}$ *such that the equations*

$$(\nabla G \cdot y - H^T \cdot z + \nabla f)_{J\pm\infty} = 0 \tag{20}$$

*are fulfilled.*

*Then*

$$\underline{f}^* := \min\big\{\mathbf{f}+(\mathbf{G}-\nabla\mathbf{G}^T\tilde{x})^T\cdot\mathbf{y}+\mathbf{h}^T\mathbf{z} \\ \underline{x}^T_{JuJ^{+\infty}}\mathbf{d}^+_{JuJ^{+\infty}}+\overline{x}^T_{JuJ^{-\infty}}\mathbf{d}^-_{JuJ^{-\infty}}-\nabla\mathbf{f}^T\tilde{x}\big\} \tag{21}$$

*is a finite lower bound of the optimal value for all optimization problems of the family (17), that is*

$$\underline{f}^* \leqslant f^*(p) \quad \textit{for } p\in\mathbf{p}. \tag{22}$$

*Proof.* Let $p\in\mathbf{p}$ be chosen fixed. Because of assumption (ii) and the inclusion conditions (18), it follows that there exist $y(p)\in\mathbf{y}$ and $z(p)\in\mathbf{z}$ such that the equations

$$(\nabla G(\tilde{x};p)\cdot y(p)-H(p)^T\cdot z(p))_j+(\nabla f(\tilde{x};p))_j=0 \tag{23}$$

are fulfilled for all $j\in J^{\pm\infty}$.

Hence, the components of the defect

$$d(p):=\nabla f(\tilde{x};p)+\nabla G(\tilde{x};p)\cdot y(p)-H(p)^T\cdot z(p) \tag{24}$$

are equal to zero for $j\in J^{\pm\infty}$. Lemma 1 yields

$$f^*(p) \geqslant f(\tilde{x};p)+(G(\tilde{x};p)-\nabla G(\tilde{x};p)^T\tilde{x})^T y(p)+h(p)^T z(p) \\ +\underline{x}^T d^+(p)+\overline{x}^T d^-(p)-\nabla f(\tilde{x};p)^T\tilde{x}. \tag{25}$$

Now the terms $\underline{x}_j d_j^+(p)$ and $\overline{x}_j d_j^-(p)$ are equal zero for $j\in J^{\pm\infty}$, and from assumption (i) it follows that

$$d_j(p)\in\mathbf{d}_j\leqslant 0 \quad \text{for } j\in J^{-\infty}$$

and

$$d_j(p)\in\mathbf{d}_j\geqslant 0 \quad \text{for } j\in J^{+\infty}.$$

Hence, the products $\underline{x}_j d_j^+(p)$ and $\overline{x}_j d_j^-(p)$ vanish for $j\in J^{-\infty}$ and $j\in J^{+\infty}$, respectively. Therefore,

$$\underline{x}^T d^+(p)=\underline{x}^T_{JuJ^{+\infty}} d^+_{JuJ^{+\infty}} \quad \text{and} \quad \overline{x}^T d^-(p)=\overline{x}^T_{JuJ^{-\infty}} d^-_{JuJ^{-\infty}}$$

and these scalar products are finite. Therefore, the right hand side of (25) is finite. The inclusion properties of interval arithmetic yield that for each $p\in\mathbf{p}$ the right hand side of (25) is contained in the interval quantity

$$\mathbf{f}+(\mathbf{G}-\nabla\mathbf{G}^T\tilde{x})^T\cdot\mathbf{y}+\mathbf{h}^T\mathbf{z}+\underline{x}_J^T\mathbf{d}_J^+ +\overline{x}_J^T\mathbf{d}_J^- -\nabla\mathbf{f}^T\tilde{x}.$$

Hence, $\underline{f}^*$ is a finite lower bound of the optimal values for this family of optimization problems. □

The lower bound (21) is computable for large-scale linear and convex optimization problems, because the computation of the enclosures (18) exploits the sparse structures of functions just as the interval matrix-vector operations for computing **d** and $\underline{f}^*$. Degenerate problems (10) (that is, more than $n$ constraints are active) also can be rigorously bounded from below, provided the local solver has computed sufficiently good approximations.

The output of a local solver applied to the convex optimization problem (10) consists of an approximate optimal solution $\tilde{x} \in \mathbf{R}^n$, approximate Lagrange multipliers $\tilde{y} \in \mathbf{R}^m$ for the inequalities $G(x) \leqslant 0$, multipliers $\tilde{z} \in \mathbf{R}^l$ for the equations $Hx = h$, and multipliers $\tilde{u}$ and $\tilde{v}$ for the simple bound constraints $\underline{x} \leqslant x$ and $x \leqslant \overline{x}$, respectively. The following theorem shows that, roughly spoken, the lower bound (21) is sharp, provided all calculations are executed exactly and an Karush-Kuhn-Tucker point is known.

THEOREM 2. *Assume that the vector $p$ of parameters is known exactly, $\tilde{x}$, $\mathbf{y} := \tilde{y}$, $\mathbf{z} := \tilde{z}$, $\tilde{u}$ and $\tilde{v}$ satisfy the Karush-Kuhn-Tucker conditions of* (10) *exactly, and all computations in Theorem* 1 *are executed exactly, then $\underline{f}^* = f^*$.*

*Proof.* The Karush-Kuhn-Tucker conditions for problem (10) are

$$\nabla f(\tilde{x}) + \nabla G(\tilde{x}) \cdot \tilde{y} - \tilde{u} + \tilde{v} - H^T \tilde{z} = 0, \tag{26}$$

$$\tilde{y} \geqslant 0,\ \tilde{u} \geqslant 0,\ \tilde{v} \geqslant 0, \tag{27}$$

$$G(\tilde{x})^T \cdot \tilde{y} = 0, \tag{28}$$

$$(\underline{x}_j - \tilde{x}_j)\tilde{u}_j = 0 \text{ for all } j \text{ with finite } \underline{x}_j, \tag{29}$$

$$(\tilde{x}_j - \overline{x}_j)\tilde{v}_j = 0 \text{ for all } j \text{ with finite } \overline{x}_j. \tag{30}$$

Assuming exact computations, we obtain from (19)

$$d = \nabla f(\tilde{x}) + \nabla G(\tilde{x}) \cdot \tilde{y} - H^T \cdot \tilde{z}.$$

Since $\tilde{x}_j$ cannot be equal to $\underline{x}_j$ and to $\overline{x}_j$, for each index $j$ one of the Lagrange parameter $\tilde{u}_j$ or $\tilde{v}_j$ must be equal to zero. If

$$\tilde{v}_j > 0 \text{ then } \tilde{u}_j = 0,\ \tilde{x}_j = \overline{x}_j, \text{ and } \tilde{v}_j = -d_j > 0, \tag{31}$$

or if

$$\tilde{u}_j > 0 \text{ then } \tilde{v}_j = 0,\ \tilde{x}_j = \underline{x}_j, \text{ and } \tilde{u}_j = d_j > 0. \tag{32}$$

Hence,

$$\underline{x}_J^T d_J^+ = \tilde{x}_J^T \tilde{u}_J, \text{ and } \overline{x}_J^T d_J^- = -\tilde{x}_J^T \tilde{v}_J. \tag{33}$$

Using equation (26) for evaluating $-\nabla f(\tilde{x})^T \tilde{x}$, a short computation shows that

$$\underline{f}^* = f(\tilde{x}) + \underline{x}^T d^+ + \overline{x}^T d^- - \tilde{u}^T \tilde{x} + \tilde{v}^T \tilde{x}.$$

In the case $d_j=0$ equation (26) yields $-\tilde{u}_j+\tilde{v}_j=0$, implying $(-\tilde{u}_j+\tilde{v}_j)\tilde{x}_j=0$. Hence,

$$\underline{f}^*=f(\tilde{x})+\underline{x}_J^T d_J^+ +\overline{x}_J^T d_J^- -\tilde{u}_J^T\tilde{x}_J+\tilde{v}_J^T\tilde{x}_J=f(\tilde{x}).$$

□

## 4. Quadratic Programming

As a special case, we consider the standard quadratic programming problem

$$\begin{aligned} &\min\ c^T x+\tfrac{1}{2}x^T Qx \\ &\text{s.t.}\ \ Hx=h \\ &\qquad 0\leqslant x\leqslant \overline{x}, \end{aligned} \tag{34}$$

where the input data are known exactly; that is, $c\in\mathbf{R}^n$, $h\in\mathbf{R}^l$, $H$ is a $l\times n$ matrix, $Q$ is a positive semidefinite $n\times n$ matrix, and the simple bounds $\underline{x}=0,\overline{x}\in\mathbf{R}^n$ are finite. The following corollary is a consequence of Theorem 1.

COROLLARY 1. *Let $\tilde{x}\in\mathbf{R}^n$, $\tilde{z}\in\mathbf{R}^l$, and*

$$d:=c+Q\tilde{x}-H^T\tilde{z}. \tag{35}$$

*Then the optimal value $f^*$ of the quadratic programming problem* (34) *is bounded from below by the finite value*

$$f^*\geqslant \underline{f}^*:=h^T\tilde{z}-\frac{1}{2}\tilde{x}^T Q\tilde{x}+\overline{x}^T d^-. \tag{36}$$

*Proof.* Since there are no inequalities and the simple bounds are finite, all assumptions of Theorem 1 are satisfied. Because all input data are real, the interval operations coincide with the real operations, and from formula (21) we obtain

$$\begin{aligned} f^* &\geqslant c^T\tilde{x}+\tfrac{1}{2}\tilde{x}^T Q\tilde{x}+h^T\tilde{z}+\overline{x}^T d^- -(c+Q\tilde{x})^T\tilde{x} \\ &= h^T\tilde{z}-\tfrac{1}{2}\tilde{x}^T Q\tilde{x}+\overline{x}^T d^-. \end{aligned}$$

Since $\overline{x}\in\mathbf{R}^n$ is finite, $\underline{f}^*$ is finite. □

In order to compute a rigorous lower bound for the quadratic programming problem (34), we assume that a quadratic programming routine has computed an approximate optimal solution $\tilde{x}$ with Lagrange multiplies $\tilde{z}$. Then, using only the rounding modes on a computer, the following small program suffice:

ALGORITHM 1.

> **roundup**;
> $\alpha = H^T \tilde{z}$;
> $\gamma = \frac{1}{2}\tilde{x}^T (Q\tilde{x})$
> **rounddown**;
> $d = c + Q\tilde{x} - \alpha$;
> $\beta = \overline{x}^T d^-$;
> $\underline{f}^* = h^T \tilde{z} + \beta - \gamma$;

The fact that only rounding modes are necessary was first observed for the standard linear programming problem with exact input data and finite simple bounds by Neumaier and Shcherbina [11]. We see that this observation holds also true for standard quadratic programming problems.

If the input is uncertain, but can be bounded by interval quantities $c \in \mathbf{c}$, $h \in \mathbf{h}$, $H \in \mathbf{H}$, and $Q \in \mathbf{Q}$, we obtain

$$\underline{f}^* = \min(\mathbf{h}^T \tilde{z} - \frac{1}{2}\tilde{x}^T \mathbf{Q}\tilde{x} + \overline{x}^T \mathbf{d}^-), \tag{37}$$

where

$$\mathbf{d} := \mathbf{c} + \mathbf{Q}\tilde{x} - \mathbf{H}^T \tilde{z}.$$

## 5. Algorithms

Next, we describe two algorithms for computing a rigorous lower bound for the optimal value of convex optimization problems which are based on the previous analysis.

In this section it is assumed that a local nonlinear solver has computed an approximation $\tilde{x}$ of an optimal point $x^*$, and Lagrange parameter $\tilde{y}$ and $\tilde{z}$ corresponding to the inequalities $G(x; \text{mid}\mathbf{p}) \leqslant 0$ and the equations $H(\text{mid}\mathbf{p}) \cdot x = h(\text{mid}\mathbf{p})$, respectively. No assumptions are made about the quality of the approximations. Moreover, we assume that the interval quantities $\mathbf{f}$, $\mathbf{G}$, $\nabla\mathbf{f}$, and $\nabla\mathbf{G}$ are computed rigorously.

Firstly, we discuss the case where all simple bounds are finite. Then the conditions (i) and (ii) of Theorem 1 are automatically satisfied, and the following small algorithm suffice:

ALGORITHM 2.

1. *Set* $\mathbf{y} := \tilde{y}^+$, *and* $\mathbf{z} := \tilde{z}$.
2. *Compute* $\mathbf{d} := \nabla\mathbf{f} + \nabla\mathbf{G} \cdot \mathbf{y} - \mathbf{H}^T \cdot \mathbf{z}$.
3. *Compute* $\underline{f}^* := \min\{\mathbf{f} + (\mathbf{G} + \nabla\mathbf{G}^T \tilde{x})^T \mathbf{y} + \mathbf{h}^T \mathbf{z} + \underline{x}_J^T \mathbf{d}_J^+ + \overline{x}_J^T \mathbf{d}_J^- - \nabla\mathbf{f}^T \tilde{x}\}$.

Theorem 1 requires that $\mathbf{y} \geqslant 0$. Hence, if the local solver has computed untruly negative components $\tilde{y}_i$, we set this component equal to zero in the first step. Then all assumptions of Theorem 1 are satisfied, and $\underline{f}^*$ is a rigorous lower bound of the optimal value. Notice that there occurs no overestimation during the computation of $\mathbf{d}$, since the radii of $\mathbf{y}$ and $\mathbf{z}$ are equal to zero.

The computation of $\underline{f}^*$ requires $O((m+l)\cdot n)$ interval operations, and additionally the operations for computing the interval quantities (18). Hence, we obtain a very cheap postprocessing algorithm.

For large simple bounds and $\text{rad}\mathbf{p} > 0$ the lower bound $\underline{f}^*$ may be poor, because the terms $\underline{x}_J^T \mathbf{d}_J^+$ or $\overline{x}_J^T \mathbf{d}_J^-$ may be large. Then it is useful to treat large simple bound components in the same way as infinite bounds (see below).

Secondly, we discuss the case where $J^{\pm\infty}$ is empty, but the index sets $J^{-\infty}$ and/or $J^{+\infty}$ are nonempty. Then, additionally, condition (i) in Theorem 1 must be satisfied. We investigate first this condition for an index $j \in J^{-\infty}$ yielding the inequality

$$\mathbf{d}_j := (\nabla \mathbf{f})_j + (\nabla \mathbf{G} \cdot \mathbf{y})_j - (\mathbf{H}^T \cdot \mathbf{z})_j \leqslant 0. \tag{38}$$

This inequality may be not true for the choice $\mathbf{y} := \tilde{y}^+$ and $\mathbf{z} := \tilde{z}^+$.

The reason is that good approximations $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$ for the midpoint problem of (17) approximate very well the optimal solution of the linearized midpoint problem (16). Hence, if $|\underline{x}_j|$ and $|\overline{x}_j|$ are large, then maximizing the term $\underline{x}^T u - \overline{x}^T v$ in the objective function of (16) leads in many cases to $u_j = v_j = 0$, and for the constraint in (16) it follows that for $p = \text{mid}\mathbf{p}$

$$-(\nabla G(\tilde{x}; p) \cdot \tilde{y})_j + (H(p)^T \cdot \tilde{z})_j = (\nabla f(\tilde{x}; p))_j.$$

Therefore, the defect satisfies

$$d_j(p) := (\nabla f(\tilde{x}; p))_j + (\nabla G(\tilde{x}; p) \cdot \tilde{y})_j - (H(p)^T \cdot \tilde{z})_j = 0,$$

and $0 = d_j(p) \in \mathbf{d}_j$, violating frequently the inequality $\mathbf{d}_j \leqslant 0$.

In order to enforce the inequality (38), the idea is to compute approximations $\tilde{y}$ and $\tilde{z}$ of a perturbed problem in a way such that

$$d_j(p) \leqslant -\varepsilon_j, \quad \varepsilon_j > 0 \tag{39}$$

is fulfilled. This is equivalent to

$$-(\nabla G(\tilde{x}; p) \cdot \tilde{y})_j + (H(p)^T \cdot \tilde{z})_j \leqslant (\nabla f(\tilde{x}; p))_j + \varepsilon_j. \tag{40}$$

Then, for the midpoint problem, the inequality (39) is satisfied, and, in order that for all problems in $\mathbf{p}$ this inequality holds true, we have to choose $\varepsilon_j$ large enough.

Our suggestion is

$$\begin{aligned} \varepsilon_j = 2\{&\text{rad}\nabla \mathbf{f})_j + \text{rad}(\nabla \mathbf{G} \cdot \tilde{y})_j + \text{rad}(\mathbf{H}^T \cdot \tilde{z})_j \\ &+ \varepsilon^*(|\nabla f(\tilde{x}; p)_j| + (|\nabla G(\tilde{x}; p)| \cdot |\tilde{y}|)_j + (|H(p)^T| \cdot |\tilde{z}|)_j\}. \end{aligned} \tag{41}$$

This definition takes into consideration the interval input data $\mathbf{p}$ and the accuracy $\varepsilon^*$ of the nonlinear solver, which is applied to the midpoint problem $p = \text{mid}\,\mathbf{p}$. For an index $j \in J^{+\infty}$ similar arguments lead to

$$d_j(p) \geqslant \varepsilon_j, \quad \varepsilon_j > 0, \tag{42}$$

which is equivalent to

$$-(\nabla G(\tilde{x};p)\cdot\tilde{y})_j + (H(p)^T\cdot\tilde{z})_j \leqslant \nabla f(\tilde{x};p)_j - \varepsilon_j. \tag{43}$$

Now combining (40) and (43) results in computing componentwise a vector $c(\varepsilon)$, where

$$c_j(\varepsilon) := \begin{cases} (\nabla f(\tilde{x};p))_j + \varepsilon_j, & \text{if } j \in J^{-\infty} \\ (\nabla f(\tilde{x},p))_j - \varepsilon_j, & \text{if } j \in J^{+\infty} \\ (\nabla f(\tilde{x};p))_j & \text{otherwise,} \end{cases} \tag{44}$$

and $\varepsilon_j$ are defined by formula (41). Then with an lp-solver we compute for the perturbed linearized dual problem

$$\begin{array}{ll} \max & (G(\tilde{x};p) - \nabla G(\tilde{x};p)^T\cdot\tilde{x})^T\cdot y + h(\hat{\mathbf{p}})^T\cdot z + \\ & \underline{x}^T u - \overline{x}^T v \\ \text{s.t.} & -\nabla G(\tilde{x};p)\cdot y + H(p)^T z + u - v = c(\varepsilon) \\ & y \geqslant 0,\ u \geqslant 0,\ v \geqslant 0. \end{array} \tag{45}$$

approximations $\tilde{y}$, $\tilde{z}$, $\tilde{u}$ and $\tilde{v}$. As before, we set $\mathbf{y} := \tilde{y}^+$ and $\mathbf{z} := \tilde{z}$, and test the conditions (i) of Theorem 1. If the conditions are not fulfilled, then the process is repeated by doubling the parameters $\varepsilon_j$ and updating once more $c$ as in (44).

Summarizing, we obtain the following algorithm.

ALGORITHM 3.

1. *Compute $\varepsilon_j$ by formula (41) for $j \in J^{-\infty} \cup J^{+\infty}$.*
2. *Compute $c(\varepsilon)$ by formula (44).*
3. *Compute with an lp-solver approximations $\tilde{y}$ and $\tilde{z}$ for the optimal solution of the perturbed problem (45). If the lp-solver gives a warning and cannot find approximations, then STOP: No rigorous lower bound.*
4. *Set $\mathbf{y} := \tilde{y}^+$, and $\mathbf{z} := \tilde{z}$.*
5. *Compute $\mathbf{d} := \nabla\mathbf{f} + \nabla\mathbf{G}\cdot\mathbf{y} - \mathbf{H}^T\cdot\mathbf{z}$.*
6. *If condition (i) of Theorem 1 is satisfied, then STOP: Lower bound is $\underline{f}^* = \min\{\mathbf{f} + (\mathbf{G} - \nabla\mathbf{G}^T\tilde{x})^T\cdot\mathbf{y} + \mathbf{h}^T\mathbf{z} + \underline{x}_J^T\mathbf{d}_J^+ + \overline{x}_J^T\mathbf{d}_J^- - \nabla\mathbf{f}^T\tilde{x}\}$.*
7. *Set $\varepsilon_j = 2\varepsilon_j$ for $j \in J^{-\infty} \cup J^{+\infty}$, and goto step 2.*

It follows that in each iteration (i.e. one execution of step 2 to step 7) the parameters $\varepsilon_j$ are doubled, yielding in few iterations (in many cases only between one and three) either a rigorous lower bound in step 6 or a STOP in step 3. Notice that

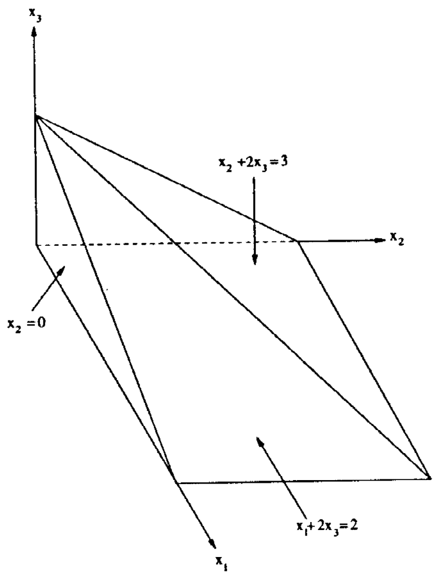


*Figure 1.* The set of feasible solutions for the degenerate problem.

the main computational costs for this bound are required by solving the perturbed linearized dual problem (45) in step 3. However, these costs are small compared to the expenses of the nonlinear solver for computing approximate optimal solutions of the convex problem (10). Moreover, lp-solvers are much more robust than nonlinear solvers.

If a lower simple bound $|\underline{x}_j|$ is very large, then the inequality $\underline{x}_j \leqslant x_j$ can be treated as one of the inequalities $G(x) \leqslant 0$, and the index $j$ can be put into $J^{-\infty}$. This equivalent transformation has the advantage that the term $\underline{x}_j \mathbf{d}_j^+$ vanishes, perhaps leading to a reduction of the overestimation. Analogously, large upper simple bounds may be handled.

For variables $x_j$ with $j \in J^{\pm\infty}$ the linear interval system (20) must be solved. A detailed description for solving such systems can be found in [6]. For avoiding the solution of this interval system a seemingly good modification would be to describe free variables as the difference of two non-negative variables. But transformation leads to an ill-posed linear programming problem which contains in each neighbourhood of the input data problems with empty feasible domain. This would imply the lower bound minus infinity.[1]

## 6. Example

For the purpose of illustration of the previous analysis, we consider the convex optimization problem

[1]I wish to thank Arnold Neumaier for this hint.

$$\begin{aligned}
\min\quad & p_1x_1^2 + p_2x_2^2 + (p_3x_3-2)^2 \\
\text{s.t.}\quad & p_4x_1 + 2p_5x_3 - 2p_6 \leqslant 0 \\
& p_7x_2 + 2p_8x_3 - 2p_9 \leqslant 0 \\
& 0 \leqslant x_1, x_2, x_3 \leqslant 2
\end{aligned} \tag{46}$$

The set of feasible points is illustrated in Figure 1 in the case $p_i = 1$ for $i = 1, \ldots, 9$.

Obviously, in this case the optimal point is $x^* = (0,0,1)^T$ with optimal value $f^* = 1$. The point $x^*$ is the intersection of four facets in the three-dimensional space implying that $x^*$ is degenerate. This is not a simple degeneracy caused by redundant constraints, since deleting one of the above constraints changes the set of feasible solutions.

Writing down the Karush-Kuhn-Tucker conditions, and observing that the constraints $x_i \leqslant 2$ for $i = 1,2,3$ and $x_3 \geqslant 0$ are not active for $x^*$, it follows that the corresponding Lagrange multipliers are zero, and we obtain the conditions

$$\begin{aligned}
2x_1^* + y_1^* - y_3^* &= 0 \\
2x_2^* + y_2^* - y_4^* &= 0 \\
2(x_3^*-2) + 2y_1^* + 2y_2^* &= 0
\end{aligned} \tag{47}$$

Using $x^* = (0,0,1)$ yields

$$y_1^* = y_3^*,\ y_2^* = y_4^*,\ y_1^* + y_2^* = 1 \tag{48}$$

which has, due to the degeneracy, no unique solution.

To simplify matters for the following illustration of Algorithm 2, we assume that the local nonlinear solver has calculated the exact optimal solution and multipliers

$$\tilde{x} = x^* = (0,0,1)^T,\ \tilde{y} = (y_1^*, y_2^*)^T = \left(\frac{1}{2}, \frac{1}{2}\right)^T \tag{49}$$

In order to illustrate the effect of interval input data, every parameter $p_i$ is allowed to vary in the interval $[1-r, 1+r]$ independently from the other parameters $p_j$ with $i \neq j$; that is, we look on interval parameter $\mathbf{p}_j$. For the radius we assume $0 \leqslant r < 1$.

In the first step of Algorithm 2 we set $\mathbf{y} := \tilde{y}^+ = (1/2, 1/2)^T$. For the second step, the enclosures (18) must be calculated:

$$\begin{aligned}
\{f(\tilde{x}; p) : p \in \mathbf{p}\} &\subseteq \mathbf{p}_1\tilde{x}_1 + \mathbf{p}_2\tilde{x}_2 + (\mathbf{p}_3\tilde{x}_3 - 2)^2 \\
&= ([1-r, 1+r]\cdot 1 - 2)^2 \\
&= [1-2r+r^2, 1+2r+r^2] =: \mathbf{f},
\end{aligned}$$

$$\{G(\tilde{x};p) : p \in \mathbf{p}\} \subseteq \begin{pmatrix} 2[1-r,1+r]\cdot 1-2[1-r,1+r] \\ 2[1-r,1+r]\cdot 1-2[1-r,1+r] \end{pmatrix}$$
$$= \begin{pmatrix} [-4r,4r] \\ [-4r,4r] \end{pmatrix} =: \mathbf{G},$$

$$\{\nabla f(\tilde{x};p) : p \in \mathbf{p}\} \subseteq \begin{pmatrix} 2\mathbf{p}_1 \cdot \tilde{x}_1 \\ 2\mathbf{p}_2 \cdot \tilde{x}_2 \\ 2\mathbf{p}_3(\mathbf{p}_3 \cdot \tilde{x}_3 - 2) \end{pmatrix}$$
$$= \begin{pmatrix} 0 \\ 0 \\ [-2-4r-2r^2, -2+4r-2r^2] \end{pmatrix} =: \nabla\mathbf{f},$$
$$\{\nabla G(\tilde{x};p) : p \in \mathbf{p}\} \subseteq \left( \begin{pmatrix} [1-r,1+r] \\ 0 \\ [2-2r,2+2r] \end{pmatrix}, \begin{pmatrix} 0 \\ [1-r,1+r] \\ [2-2r,2+2r] \end{pmatrix} \right)$$
$$=: \nabla\mathbf{G}.$$

Hence, we obtain

$$\mathbf{d} = \begin{pmatrix} 0 \\ 0 \\ [-2-4r-2r^2, -2+4r-2r^2] \end{pmatrix} + \nabla\mathbf{G} \cdot \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$
$$= \begin{pmatrix} [\frac{1}{2}-\frac{r}{2}, \frac{1}{2}+\frac{r}{2}] \\ [\frac{1}{2}-\frac{r}{2}, \frac{1}{2}+\frac{r}{2}] \\ [-6r-2r^2, 6r-2r^2] \end{pmatrix}.$$

Hence, $\mathbf{d}^- = (0,0,[-6r-2r^2,0])^T$, and in step 3 of Algorithm 2

$$\underline{f}^* = \min\{\mathbf{f} + (\mathbf{G} - \nabla\mathbf{G}^T \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix})^T \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} - \nabla\mathbf{f}^T \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}^T \cdot \mathbf{d}^-\}$$

yields $\underline{f}^* = 1 - 24r - r^2$.

## 7. Rigorous Infeasibility Test

In branch and bound algorithms a subproblem is discarded if the local nonlinear solver detects infeasibility (see for example Borchers and Mitchell [2]). It is not a rare phenomenon that especially nonlinear solver find no feasible solutions, although the subproblem contains such ones. A consequence is that the global minimum solutions may be cut off.

The first possibility for verifying rigorously that a convex subproblem (relaxation) contains no feasible solution, is to proceed as in the two-phase-method:

nonnegative artificial variables are introduced as customary, and then the sum of the artificial variables is minimized. This auxiliary problem remains convex. If the optimal value is greater zero, then the problem is infeasible. A rigorous lower bound of this optimal value can be computed with the previous algorithms. If this rigorous lower bound is greater zero then we have obtained a safe certificate of infeasibility. Two-phase nonlinear solvers provide an approximate solution and appropriate Lagrange multipliers, which may be directly used for computing a rigorous lower bound. An additional call for the auxiliary problem is not required in this case.

Another approach for verifying infeasibility for linear programs is described in Neumaier and Shcherbina [11]. It is based on the observation that the dual of an infeasible problem is unbounded or infeasible, and typically lp-solvers compute a ray exposing this. This information can be used for a certificate of infeasibility. One way to extend this method to the convex case is to use the linear relaxation (15).

## 8. Conclusions

We have described algorithms for computing a lower bound for the optimal value of convex optimization problems. The input data of these problems may be uncertain. This lower bound (i) is rigorously valid, (ii) it uses only approximate solutions of the nonlinear solver, (iii) nothing is assumed about the quality of the approximations, (iv) it requires, at least for finite simple bounds, only a fraction of the computational work for solving the convex problem, and (v) it can also be used to verify infeasibility. We think that many branch and bound algorithms can be made completely rigorous by using such postprocessing tools, whereupon the computational time increases only by a small amount in most cases.

## References


1. Alefeld, G. and Herzberger, J., (1983) *Introduction to Interval Computations*. Academic Press, New York.
2. Borchers, B. and Mitchell, J.E., (1994). An improved branch and bound algorithm for mixed integer nonlinear Programs. *Computers Ops Res.*, 21(4), 359–367.
3. Floudas, C.A., (2000). *Deterministic Global Optimization – Theory, Methods and Applications*, volume 37 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Dordrecht, Boston, London.
4. Floudas, C.A. and Pardalos, P.M., (2001). *Encyclopedia of Optimization*. Kluwer Academic Publishers.
5. Griewank, A., (1989). *On Automatic Differentiation*, volume 88 of *Mathematical Programming*. Kluwer Academic Publishers, Boston MA.
6. Jansson, C., 2002. Rigorous lower and upper bounds in linear programming. Technical Report 02.1, Forschungsschwerpunkt Informations- und Kommunikationstechnik, TU Hamburg-Harburg. http://www.ti3.tu-harburg.de/paper/jansson/verification.ps.
7. Moore, R.E., (1979). *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, PA.
8. NETLIB Linear Programming Library. http://www.netlib.org/lp.

9. Neumaier, A., (1990). Interval methods for systems of equations. In: *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.
10. Neumaier, A., (2001). *Introduction to Numerical Analysis*. Cambridge University Press, Cambridge.
11. Neumaier, A. and Shcherbina, O. Safe bounds in linear and mixed-integer programming. `http://www.mat.univie.ac.at/~neum/papers.html`.
12. Ordóñez, F. and Freund, R.M., (2002). Computational experience and the explanatory value of condition numbers for linear optimization. Technical Report MIT Operations Research Center Working Paper OR361-02, MIT. submitted for publication to SIAM J. Optimization. http://web.mit.edu/rfreund/www/CVfreund.pdf.
13. Rall, L.B. and Corliss, G.F., (1999). Automatic differentiation: Point and interval AD. In Pardalos, P.M. and Floudas, C.A. (eds.), *Encyclopedia of Optimization*. Kluwer, Dordrecht.
14. Rump, S.M., (1993). Validated solution of large linear systems. In: Albrecht, R., Alefeld, G. and Stetter, H.J., (eds), *Validation numerics: Theory and Applications*, volume 9 of *Computing Supplementum*, Springer, Berlin, pp. 191–212.
15. Tawaralani, M. and Sahinidis, N.V., (2002). *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming*. Kluwer Academic Publishers.