This is Technical Report No. 020811 of Department of Econometrics, University of Economics, Prague. It serves as pre-publication medium for papers before they are ready for final publication.

On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data

Michal Černý[1] Jaromír Antoch[2] and Milan Hladík[3]

Abstract. We consider linear regression models where both input data (the values of independent variables) and output data (the observations of the dependent variable) are affected by loss of information caused by uncertainty, indeterminacy, rounding or censoring. Some or all of the crisp data, which are unavailable, are replaced by known intervals. We introduce a possibilistic generalization of the least squares estimator, so called OLS-set for the interval model. Investigation of the OLS set allows us to quantify whether the replacement of crisp values by interval values can have a significant effect on the usual OLS estimator; or, in other words, whether the loss of information caused by replacement of crisp values by intervals can be considered as serious. We show that in the model with both interval input and output data in general there is no computationally feasible way to describe the OLS set reasonably (assuming $P \neq NP$). Nevertheless, we show and compare various approximation methods that could be useful in particular cases. We also focus on restricted versions of the general interval linear regression model. We show that in the crisp input – interval output model, the OLS set is a convex polyhedron of a special structure. We present both exact and approximate methods for description of the OLS set, in particular interval enclosures and ellipsoidal enclosures. We also present a meta-algorithm, called Reduction and Reconstruction Recursion, which can be used for computation of vertex and facet description of the OLS set. We discuss special cases of the regression model, e.g. a model with repeated observations and a model where we want to estimate a single regression parameter. We illustrate the approaches by examples.

Keywords. Interval data; uncertain data; censored data; possibilistic regression; computational complexity; zonotopes

[1]Department of Econometrics, Faculty of Computer Science and Statistics, University of Economics, Prague; Winston Churchill Square 4, 13067 Prague, Czech Republic; cernym@vse.cz

[2]Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague; Sokolovská 83, 18675 Prague, Czech Republic; antoch@karlin.mff.cuni.cz

[3]Department of Applied Mathematics, Faculty of Mathematics and Physics, Charles University, Prague; Malostranské náměstí 25, 11000 Prague, Czech Republic; hladik@kam.mff.cuni.cz

1 Introduction

Consider the linear regression model

$$y = X\beta + \varepsilon, \tag{1}$$

where y denotes the vector of observations of the dependent variable, X denotes the design matrix of the regression model, β denotes the vector of unknown regression parameters and ε is the vector of disturbances. We do not make any special assumptions on ε; we just assume that for estimation of β a linear estimator can be used, i.e. an estimator of the form

$$\widehat{\beta} = Qy, \tag{2}$$

where Q is a matrix. In particular we shall concentrate on the Ordinary Least Squares (OLS) estimator, which corresponds to the choice $Q = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ in (2). Nevertheless, the theory is also applicable for other linear estimators, such as the Generalized Least Squares (GLS) estimator, which corresponds to the choice $Q = (X^{\mathrm{T}}\Omega^{-1}X)^{-1}\Omega^{-1}X^{\mathrm{T}}$ in (2), where Ω is either known or estimated covariance matrix of ε. Other examples include estimation methods which, at the beginning, exclude outliers and then apply OLS or GLS. These estimators are often used in analysis of contaminated data.

Throughout the paper, the symbol n stands for the number of observations and the symbol p stands for the number of regression parameters, as it is usual in statistics.

We shall treat X and y as constants representing observed values of the independent and dependent variables, respectively. Then the tuple (X, y) is called *data* for the regression model.

1.1 Interval data in the linear regression model

We shall study the situation when the data (X, y) cannot be observed directly. Instead of y_i and X_{ij}, only intervals of the form $[\underline{y}_i, \overline{y}_i]$ and $[\underline{X}_{ij}, \overline{X}_{ij}]$ are available, where we are guaranteed that for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, p\}$,

$$y_i \in [\underline{y}_i, \overline{y}_i] \quad \text{and} \quad X_{ij} \in [\underline{X}_{ij}, \overline{X}_{ij}],$$

where y_i denotes the i-th element of y and X_{ij} denotes the (i, j)-th element of X.

The replacement of crisp values by intervals is henceforth referred to as "censoring". In some literature, this process is also called "trimming", "uncertaintification" or "intervalization".

1.2 Motivation

Inclusion of interval data in linear regression models is suitable for modeling variety of real-world problems. For example:

- The data (X, y) have been interval-censored. This is often the case of medical, epidemiologic or demographic data — only interval-censored data are published while the exact individual values are kept secret.

- Data are rounded. If we store data using data types of restricted precision, then instead of exact values we are only guaranteed that the true value is in an interval of width 2^{-d} where d is the number of bits of the data type for representation of the non-integer part. For example, if we store data as integers, then we know only the interval $[\tilde{y} - 0.5, \tilde{y} + 0.5]$ instead of the exact value y, where $\tilde{y}$ is y rounded to the nearest integer. This application is important in the theory of reliable computing.

- The data are uncertain or unstable. For that reason it might be inappropriate to describe them in terms of fixed values (X, y) only.

- Categorial data may be sometimes interpreted as interval data; for example, credit rating grades can be understood as intervals of credit spreads over the risk-free yield curve.

In econometric regression models, it is often the case that varying variables are represented by their average or median values. For example, if the exchange rate for a period of one year should be included in the regression model, usually the average rate of that year is taken. However, it might be more appropriate to regard the exchange rate as an interval inside which the variable changes.

More applications of interval data are found in econometrics [7], information science [10], ergonomics [9], optimization and operational research [12], [34], [38], [66], speeach learning [41] and in pattern recognition [35], [39].

A variety of methods for estimation of regression parameters in a regression with interval data has been developed; they are studied in statistics ([8], [19], [33], [37], [40], [45], [50], [71]), where also robust regression methods have been proposed ([29], [46]), in fuzzy theory ([21], [26], [27], [67], [68], [69]) as well as in computer science ([11], [28], [31]). An algebraic treatment of least squares methods for interval data has been considered in [5] and [15].

1.3 Interval numbers, vectors and matrices

If two real matrices X_1, X_2 are of the same dimension, the relation $X_1 \leq X_2$ is understood componentwise.

Definition 1. *(a) Let $\underline{a}$ and $\overline{a}$ be scalars such that $-\infty < \underline{a} \leq \overline{a} < \infty$. The **interval number $\boldsymbol{a}$** is the closed interval $[\underline{a}, \overline{a}]$.*

*(b) Let $\underline{X} \leq \overline{X}$ be two $M \times N$ real matrices. The **interval matrix $\boldsymbol{X} = [\underline{X}, \overline{X}]$** is the set*

$$\{X \in \mathbb{IR}^{M \times N} : \underline{X} \leq X \leq \overline{X}\}.$$

*The **interval vector $\boldsymbol{y} = [\underline{y}, \overline{y}]$** is a special case of the interval matrix with one column.*

Interval numbers, vectors and matrices are typeset in boldface.

Arithmetic operations + and × with interval numbers $\boldsymbol{a} = [\underline{a}, \overline{a}]$ and $\boldsymbol{b} = [\underline{b}, \overline{b}]$ are defined in a natural way (see [1]):

$$\begin{aligned} \boldsymbol{a} + \boldsymbol{b} &= [\underline{a} + \underline{b}, \overline{a} + \overline{b}], \\ \boldsymbol{a} \cdot \boldsymbol{b} &= [\min\{\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}, \max\{\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}]. \end{aligned} \tag{3}$$

From the definition, the following lemma is obvious:

Lemma 2. *A finite sequence of sums and products of interval numbers is a bounded set.* □

1.4 The possibilistic approach to linear regression models with interval data

To recall: we are in the situation that only intervals $(\boldsymbol{X}, \boldsymbol{y})$ are available instead of the exact values (X, y) such that $X \in \boldsymbol{X}$ and $y \in \boldsymbol{y}$. The replacement of crisp values (X, y) by interval values $(\boldsymbol{X}, \boldsymbol{y})$ causes some loss of information. *The main aim of this text is to quantify how the loss of information caused by replacement of* (X, y) *by* $(\boldsymbol{X}, \boldsymbol{y})$ *(henceforth referred to as "censoring")*[4] *influences the OLS estimator* $\widehat{\beta}$. In particular, our aim is to study tools which can help the analyst of data $(\boldsymbol{X}, \boldsymbol{y})$ to understand whether censoring of data can cause a "significant" or "insignificant" uncertainty about $\widehat{\beta}$.

The next definition generalizes of the notion of the estimator $\widehat{\beta}$ for the case when the crisp values (X, y) in are replaced by intervals $(\boldsymbol{X}, \boldsymbol{y})$ in (1).

Definition 3. *(a) A tuple* $(\boldsymbol{X}, \boldsymbol{y})$*, where* $\boldsymbol{X}$ *is an* $n \times p$ *interval matrix and* $\boldsymbol{y}$ *is an* $n \times 1$ *interval vector, is called* ***data of an interval regression model****, or just* ***interval model*** *for short.*

(b) The ***OLS-set*** *of the interval model* $(\boldsymbol{X}, \boldsymbol{y})$ *is defined as*

$$OLS(\boldsymbol{X}, \boldsymbol{y}) = \{\beta \in \mathbb{R}^p : (\exists X \in \boldsymbol{X})(\exists y \in \boldsymbol{y}) X^{\mathrm{T}} X \beta = X^{\mathrm{T}} y\}.$$

The motivation for the definition is straightforward. Our aim is to use OLS to obtain an estimate of the unknown vector of regression parameters β in the model (1). However, observations are censored, i.e., we only know intervals $\boldsymbol{X}$ and $\boldsymbol{y}$ that are guaranteed to contain the directly unobservable data (X, y). Then, the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ contains *all possible* values of OLS-estimates of β as X and y range over $\boldsymbol{X}$ and $\boldsymbol{y}$, respectively. We say that $OLS(\boldsymbol{X}, \boldsymbol{y})$ is a *possibilistic* version of the notion of the OLS estimator.

The set $OLS(\boldsymbol{X}, \boldsymbol{y})$ captures the loss of information caused by censoring of the data included in the regression model. For a user of such a regression model, it is essential to understand whether the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ is, in some

[4] Though, for simplicity of presentation, we speak about "censoring", it should be kept in mind that not only censoring, but also rounding, uncertainty, indeterminacy, instability or interval nature of data can be the reason for inclusion of intervals in the model.

sense, "large" or "small"; that is, whether the impact of the loss of information on the OLS estimator may be considered to be serious or not. More generally, the user of $(\boldsymbol{X}, \boldsymbol{y})$ needs a suitable description of the set $OLS(\boldsymbol{X}, \boldsymbol{y})$. When $p = 2$ or $p = 3$, then the set can be visualized in the parameter space using standard numerical methods. However, in higher dimensions visualization is quite complicated. Hence we need methods for a suitable description of the set $OLS(\boldsymbol{X}, \boldsymbol{y})$; in particular, we would like to design computationally feasible methods. In Section 2 we shall show that this task is very hard from the computational point of view.

1.5 Two interpretations of the possibilistic approach

Possibilistic interpretation. If we do not assume any distribution on $\boldsymbol{X}$ or $\boldsymbol{y}$, then the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ contains all possible values of $\widehat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$ as X ranges over $\boldsymbol{X}$ and y ranges over $\boldsymbol{y}$. We say that $OLS(\boldsymbol{X}, \boldsymbol{y})$ is a *covering* of $\widehat{\beta}$.

The boundary of the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ can be understood as the *worst-case impact* of interval censoring on the estimator. The possibilistic approach then can be characterized as a tool for analysis of the worst case. The worst-case analysis will be illustrated by an example in Section 4.3.

Probabilistic interpretation. If X and y are random variables such that the supports of the distributions of X and y are $\boldsymbol{X}$ and $\boldsymbol{y}$, respectively, then $\widehat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$ is a random variable satisfying $\Pr[\widehat{\beta} \in OLS(\boldsymbol{X}, \boldsymbol{y})] = 1$. Then the covering $OLS(\boldsymbol{X}, \boldsymbol{y})$ can be regarded as a "region of 100% certainty about $\widehat{\beta}$".

In the theory of Sections 2–6, we do not treat (X, y) as random variables over $(\boldsymbol{X}, \boldsymbol{y})$ (though it is one of possible interpretations); hence we need no assumptions about their distribution. (The case where (X, y) are considered as random variables is studied in literature, see e.g. [63].)

1.6 Special cases of the interval regression model

An interval regression model $(\boldsymbol{X} = [\underline{X}, \overline{X}], \boldsymbol{y} = [\underline{y}, \overline{y}])$ is also called *a general model* or *interval input – interval output model*. Interesting special cases are (see [26] and [30]):

(i) *crisp input – interval output model* is a model with $\underline{X} = \overline{X}$;

(ii) *interval input – crisp output model* is a model with $\underline{y} = \overline{y}$;

(iii) *crisp input – crisp output model* is a model with $\underline{X} = \overline{X}$ and $\underline{y} = \overline{y}$.

"Crisp input – crisp output" is just another name for the traditional model (1).

If $\boldsymbol{X}$ is crisp, i.e. if $\underline{X} = \overline{X} =: X$, then instead of $OLS(\boldsymbol{X}, \boldsymbol{y})$ we write $OLS(X, \boldsymbol{y})$. (And similarly in the case of $\boldsymbol{y}$ crisp.)

1.7 The structure of the paper

Sections 2 and 6 are devoted to the OLS-set of the general model, its computational properties and approximation methods. In section 3 we present a geometric characterization of the OLS-set of the crisp input – interval output model. In Section 4 we discuss methods for approximation of the OLS-set of the crisp input – interval output mode and in Section 5 we propose a general meta-algorithms for exact descriptions of the OLS-set of the crisp input – interval output model.

2 The general model

Our aim is to find a description of the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ given $\boldsymbol{X} = [\underline{X}, \overline{X}]$ and $\boldsymbol{y} = [\underline{y}, \overline{y}]$. Such a description may take a variety of forms — for example, we might try to find the smallest enclosing ellipse, to find the small enclosing box (i.e. interval vector) or to determine other characteristics of the set such as volume or diameter.

Theorem 5, which is the main result of this Section, shows that in general we cannot expect to be successful in a computationally feasible way. The point is that any reasonable description of $OLS(\boldsymbol{X}, \boldsymbol{y})$ must allow the user to decide whether the set is bounded or not. Theorem 5 says that there is no polynomial-time method for this question (assuming $\boldsymbol{P} \neq \boldsymbol{NP}$).

Before we state and prove the Theorem, we briefly review some definitions from complexity theory.

2.1 Some complexity-theoretic notions

We sketch basic definitions needed for further reading only; more details can be found in [3], [51].

The class $\boldsymbol{P}$ is the class of sets decidable in Turing deterministic polynomial time. The class $\boldsymbol{NP}$ is the class of sets decidable in Turing nondeterministic polynomial time. The class *co-*$\boldsymbol{NP}$ is the class of complements of $\boldsymbol{NP}$-sets, i.e. *co-*$\boldsymbol{NP} = \{A : \textit{co-}A \in \boldsymbol{NP}\}$, where *co-*$A$ is the complement of A. The class $\boldsymbol{PF}$ is the class of total functions computable in Turing deterministic polynomial time.

A set A is also called *problem* A. A problem A is *reducible* to problem B if there is a function $f \in \boldsymbol{PF}$ such that $(\forall x)[x \in A \longleftrightarrow f(x) \in B]$. The function f is also called *reduction* of the problem A to the problem B.

A problem C is $\boldsymbol{NP}$-hard if any problem $A \in \boldsymbol{NP}$ is reducible to C. A problem C is *co-*$\boldsymbol{NP}$-hard if any problem $A \in$ *co-*$\boldsymbol{NP}$ is reducible to C.

Recall that the most important complexity-theoretic conjecture is that $\boldsymbol{P} \neq \boldsymbol{NP}$ which is generally believed to be true. We shall need the following elementary lemma which can be found in any textbook on complexity theory, for example in [3] and [51].

Lemma 4. *(a) The problem A is **NP**-hard if and only if the problem co-A is co-**NP**-hard;*

*(b) if A is **NP**-hard and A is reducible to C, then C is **NP**-hard;*

*(c) if $\boldsymbol{P} \neq \boldsymbol{NP}$, then for any co-**NP**-hard problem C it holds $C \notin \boldsymbol{P}$.* □

The problems in $\boldsymbol{P}$ are generally considered to be computationally feasible. The proposition (c) says that, if $\boldsymbol{P} \neq \boldsymbol{NP}$, then no *co-**NP***-hard problem is computationally feasible.

2.2 The main result of Section 2

Theorem 5. *Let $\underline{X}, \overline{X}, \underline{y}, \overline{y}$ be rational and denote $\boldsymbol{X} = [\underline{X}, \overline{X}]$ and $\boldsymbol{y} = [\underline{y}, \overline{y}]$. Deciding whether the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ is bounded is a co-**NP**-hard problem.*

Proof. Let $\boldsymbol{X}$ be an $n \times p$ interval matrix. If there is $X \in \boldsymbol{X}$ with column rank $< p$, then for any y the set

$$\{\beta : X^{\mathrm{T}} X\beta = X^{\mathrm{T}} y\}$$

is an affine space of dimension at least one, and hence is unbounded.

Assume that for every $X \in \boldsymbol{X}$, the column rank of X is p. Then $(X^{\mathrm{T}}X)^{-1}$ exists for each $X \in \boldsymbol{X}$. By Cramer's Rule, we can write

$$((X^{\mathrm{T}}X)^{-1})_{ij} = \pm \frac{\det(X^{\mathrm{T}}X)^{[i,j]}}{\det(X^{\mathrm{T}}X)}$$

where $A^{[i,j]}$ results from A by deleting the j-th row and the i-th column. By continuity of $\det(\cdot)$ on the compact set $\boldsymbol{X}$, the set

$$\{\det(X^{\mathrm{T}}X) : X \in \boldsymbol{X}\}$$

is a closed interval which, by assumption, does not contain zero. It follows that the set

$$\left\{ \frac{1}{\det(X^{\mathrm{T}}X)} : X \in \boldsymbol{X} \right\}$$

is a closed interval. Let us denote the interval $[\underline{d}, \overline{d}]$. Also the set

$$\{\det(X^{\mathrm{T}}X)^{[i,j]} : X \in \boldsymbol{X}\}$$

is an interval of the form $[\underline{\delta}_{ij}, \overline{\delta}_{ij}]$. Hence we can write

$$\begin{aligned}
\widehat{\beta}_i &= \{((X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y)_i : \ X \in \boldsymbol{X}, \ y \in \boldsymbol{y}\} \\
&= \left\{ \sum_{j=1}^{p} \left[((X^{\mathrm{T}}X)^{-1})_{ij} \cdot \sum_{k=1}^{n} X_{kj} \cdot y_k \right] : X \in \boldsymbol{X}, y \in \boldsymbol{y} \right\} \\
&\subseteq \sum_{j=1}^{p} \left[\pm[\underline{d}, \overline{d}] \cdot [\underline{\delta}_{ij}, \overline{\delta}_{ij}] \cdot \sum_{k=1}^{n} [\underline{X}_{kj}, \overline{X}_{kj}] \cdot [\underline{y}_k, \overline{y}_k] \right]
\end{aligned}$$

and the last expression is a finite sequence of sums and products of intervals. By Lemma 2 it follows that it is a bounded set.

We have shown that the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ is unbounded if and only if there is an $X \in \boldsymbol{X}$ such that the column rank of X is $< p$. By Section of 3.19 of [58], the latter problem is ***NP***-hard. We have constructed a reduction from an ***NP***-hard problem to the problem $C :=$ *"is* $OLS(\boldsymbol{X}, \boldsymbol{y})$ *unbounded?"*. By the statements (a) and (b) of Lemma 4, the problem *co-*C $=$ *"is* $OLS(\boldsymbol{X}, \boldsymbol{y})$ *bounded?"* is *co-****NP***-hard. □

It follows that if we want to find a computationally feasible description of $OLS(\boldsymbol{X}, \boldsymbol{y})$ we must reformulate the problem. We can follow (at least) two ways:

(a) either to search for descriptions and/or approximations of $OLS(\boldsymbol{X}, \boldsymbol{y})$ which are guaranteed to be correct only under additional assumptions, or

(b) to consider special cases of the general model.

There is a variety of approaches to (a), see [1], [22], [24], [25], [32], [47], [58] and a comparison study [48]. Some of them will be discussed in Section 6.

In the next section we follow the way (b) and study the restriction to the crisp input – interval output model. Observe that this restriction is the only interesting restriction among (i) – (iii) (see Section 1.6). In the crisp input – crisp output model, the set $OLS(X, y)$ is trivial — it is either a single point or an affine space in the parameter space. And the restriction to the interval input – crisp output model is ruled out by the following observation.

Corollary 6 (to the proof of Theorem 5)**.** *Let* $\underline{X}, \overline{X}$ *and* y *be rational and denote* $\boldsymbol{X} = [\underline{X}, \overline{X}]$. *Deciding whether the set* $OLS(\boldsymbol{X}, y)$ *is bounded is a co-****NP****-hard problem.*

Proof. The reduction constructed in the proof of Theorem 5 remains valid also if $\boldsymbol{y}$ is crisp. □

Remark to Theorem 5. The hardness result of Theorem 5 does not guarantee decidability of the problem whether the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ is unbounded. Let us show that the problem is decidable. (This fact is not obvious at the first sight.) We have seen that the set $OLS(\boldsymbol{X}, \boldsymbol{y})$ is unbounded iff there is a matrix $X \in \boldsymbol{X} = [\underline{X}, \overline{X}]$ which does not have full column rank. Assume that $\underline{X}$ and $\overline{X}$ are rational; then it is easy to see that if there is a matrix $X \in \boldsymbol{X}$ which does not have full column rank, then there is a *rational* matrix $X' \in \boldsymbol{X}$ which does not have full column rank. This observation proves that the problem is recursively enumerable — it suffices to enumerate all rational matrices $X \in \boldsymbol{X}$, using the fact that the set of all rational matrices $X \in \boldsymbol{X}$ is countable. (Basics of Recursion Theory can be found in [49].)

The assumption that $\underline{X}$ and $\overline{X}$ are rational implies that all entries of $\underline{X}$ and $\overline{X}$ can be multiplied by the common denominator of fractions occurring in $\underline{X}$ and $\overline{X}$. It means that we can assume that $\underline{X}$ and $\overline{X}$ are integer matrices. The following holds: there is $X \in \boldsymbol{X}$ which does not have full column rank if and only if

$$(\exists X)[\underline{X} \leq X \leq \overline{X} \ \ \& \ \ \det X^{\mathrm{T}} X = 0] \tag{4}$$

where X ranges over $\mathbb{IR}^{n\times p}$. By the assumption of integrality of $\underline{X}$ and $\overline{X}$, the expression (4) is a sentence in the language of arithmetic (see [20]) — indeed, $\det X^{\mathrm{T}}X$ is a polynomial in $X_{11},\ldots,X_{np}$ and thus can be expressed only with addition and multiplication.

Recall that the theory of Real Closed Fields (*RCF*) is a theory with arithmetical language which proves

- basic properties of the theory of fields (in particular, closure under additive inversion and closure under multiplicative inversion for all nonzero elements);
- Bolzano's Intermediate Value Theorem for all polynomials of one variable (with integer coefficients).

By Tarski's Theorem [70], the theory *RCF* is complete. And $\mathbb{IR}$ is a model of *RCF*. It follows that there is an algorithmic procedure which decides whether the sentence (4) is true (in $\mathbb{IR}$) — it suffices to enumerate all proofs of *RCF* (using the fact that *RCF* is recursively axiomatizable) and wait until a proof of (4) or its negation appears. This proves decidability.

More can be found in [55], [56], [57] and [59].

3 Characterization of the set $\boldsymbol{OLS}(X,\boldsymbol{y})$ in the crisp input – interval output model

In this section we assume that X is a crisp matrix with full column rank.

The aim of this section is twofold:

- we shall show a geometric characterization of the set $OLS(X,\boldsymbol{y})$;
- we shall show that though there are natural descriptions of the set $OLS(X,\boldsymbol{y})$, in general they cannot be computed in polynomial time.

Hence, from the computational point of view, the situation is (in some sense) as disappointing as in the general case. However, the reason is quite different, as we shall see in Theorem 11.

The negative complexity result gives a good motivation for finding approximations and for inspection of special cases. We will

- show interval and ellipsoidal approximations of the set $OLS(X,\boldsymbol{y})$;
- show that natural descriptions of the set $OLS(X,\boldsymbol{y})$ are polynomial-time constructible if we restrict ourselves to a fixed p (i.e., to a class of regression models with a fixed number of parameters).

3.1 Geometric characterization of the set $\boldsymbol{OLS}(X,\boldsymbol{y})$

First we need to review some notions from geometry of convex polyhedra; for further reading see [72]. The next definition formalizes the Minkowski sum. In geometry literature the sum is usually defined more generally. The following simpler definition is sufficient for our purposes.

Definition 7. *The **Minkowski sum** of a set $A\subseteq\mathbb{IR}^k$ and a vector $g\in\mathbb{IR}^k$ is the set*

$$A\dotplus g=\{a+\lambda g:\ a\in A,\ \lambda\in[0,1]\}.$$

It is easily seen that for a convex set A, it holds

$$A \dot{+} g = conv(A \cup \{a + g : a \in A\}),$$

where *conv* denotes the convex hull.

Definition 8. *(a) The* ***zonotope*** *generated by $g_1, \ldots, g_N \in \mathbb{IR}^k$ with shift $s \in \mathbb{IR}^k$ is the set*

$$\mathcal{Z}(s; g_1, \ldots, g_N) = (\cdots((\{s\} \dot{+} g_1) \dot{+} g_2) \dot{+} \cdots \dot{+} g_N).$$

The vectors $g_1, \ldots, g_N$ are called ***generators****.*

(b) The ***dimension*** *of the zonotope $\mathcal{Z} = \mathcal{Z}(s; g_1, \ldots, g_N)$, denoted* $\dim(\mathcal{Z})$*, is the dimension of the linear space generated by $g_1, \ldots, g_N$.*

Instead of $(\cdots((\{s\}\dot{+}g_1)\dot{+}g_2)\dot{+}\cdots\dot{+}g_N)$ we shall write $\{s\}\dot{+}g_1\dot{+}g_2\dot{+}\cdots\dot{+}g_N$ only.

It is easily seen that a zonotope is a convex polyhedron; see Figure 1.

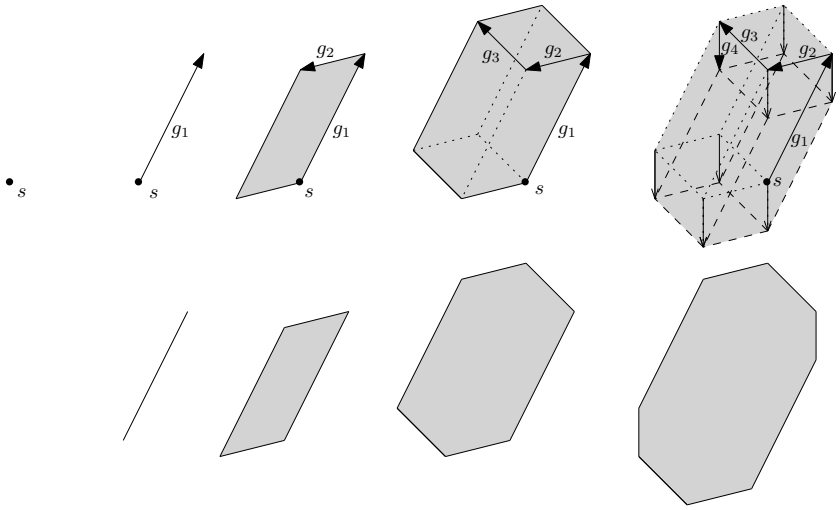

Figure 1: Sequence of zonotopes $\mathcal{Z}(s; g_1), \ldots, \mathcal{Z}(s; g_1, g_2, g_3, g_4)$ called *evolution* of a zonotope $\mathcal{Z}(s; g_1, g_2, g_3, g_4)$. In the upper part of the figure, the zonotopes are plotted with generators, while in the bottom part, the same zonotopes are plotted without generators.

The main result of this section follows.

Theorem 9. *Let $X \in \mathbb{IR}^{n \times p}$ be a matrix of full column rank and $\boldsymbol{y} = [\underline{y}, \overline{y}]$ an $n \times 1$ interval vector. Then*

$$OLS(X, \boldsymbol{y}) = \mathcal{Z}(Q\underline{y};\ Q_1(\overline{y}_1 - \underline{y}_1), \ldots, Q_n(\overline{y}_n - \underline{y}_n)),$$

where $Q = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ and Q_i is the i-th column of Q.

Proof.

$$
\begin{aligned}
&OLS(X,\boldsymbol{y})\\
&=\{Qy : y\in\boldsymbol{y}\}\\
&=\{Q\underline{y}+Q\Lambda : \Lambda\in[0,\overline{y}-\underline{y}]\}\\
&=\{Q\underline{y}+Q\Lambda : \Lambda_1\in[0,\overline{y}_1-\underline{y}_1],\ \Lambda_2\in[0,\overline{y}_2-\underline{y}_2],\ \ldots,\ \Lambda_n\in[0,\overline{y}_n-\underline{y}_n]\}\\
&=\Big\{Q\underline{y}+Q\begin{pmatrix}\Lambda_1\\0\\\vdots\\0\end{pmatrix}+Q\begin{pmatrix}0\\\Lambda_2\\\vdots\\0\end{pmatrix}+\cdots+Q\begin{pmatrix}0\\0\\\vdots\\\Lambda_n\end{pmatrix}:\\
&\qquad\qquad \Lambda_1\in[0,\overline{y}_1-\underline{y}_1],\ \Lambda_2\in[0,\overline{y}_2-\underline{y}_2],\ \ldots,\ \Lambda_n\in[0,\overline{y}_n-\underline{y}_n]\Big\}\\
&=\{Q\underline{y}+Q_1\Lambda_1+Q_2\Lambda_2+\cdots+Q_n\Lambda_n :\\
&\qquad\qquad \Lambda_1\in[0,\overline{y}_1-\underline{y}_1],\ \Lambda_2\in[0,\overline{y}_2-\underline{y}_2],\ \ldots,\ \Lambda_n\in[0,\overline{y}_n-\underline{y}_n]\}\\
&=\{Q\underline{y}+Q_1(\overline{y}_1-\underline{y}_1)\lambda_1+Q_2(\overline{y}_2-\underline{y}_2)\lambda_2+\cdots+Q_n(\overline{y}_n-\underline{y}_n)\lambda_n :\\
&\qquad\qquad \lambda_1\in[0,1],\ \lambda_2\in[0,1],\ \ldots,\ \lambda_n\in[0,1]\}\\
&=\{Q\underline{y}\}\dot{+}Q_1(\overline{y}_1-\underline{y}_1)\dot{+}Q_2(\overline{y}_2-\underline{y}_2)\dot{+}\cdots\dot{+}Q_n(\overline{y}_n-\underline{y}_n).\quad\square
\end{aligned}
\tag{5}
$$

There is a nice geometric characterization of zonotopes. Namely, a set $\mathcal{Z}\subseteq\mathbb{R}^k$ is a zonotope if and only if *there exists a number m, a matrix $Q\in\mathbb{R}^{k\times m}$ and an interval m-dimensional vector $\boldsymbol{y}$ (called m-dimensional cube) such that* $\mathcal{Z}=\{Qy : y\in\boldsymbol{y}\}$. The interesting case is $m>k$. In that case we can say that zonotopes are images of "high-dimensional" cubes in "low-dimensional" spaces under linear mappings, see Figure 2. In our setting, the set $OLS(X,\boldsymbol{y})$ is an image of the cube $\boldsymbol{y}$ under the mapping determined by the matrix $Q=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$.

In the next lemma we summarize basic properties which will be used later.

Lemma 10. *(a) Each face of the zonotope $OLS(X,\boldsymbol{y})$ is center symmetric; in particular, the set $OLS(X,\boldsymbol{y})$ itself is center symmetric.*

(b) The center of $OLS(X,\boldsymbol{y})$ is $c:=Q\underline{y}+\frac{1}{2}\sum_{i=1}^{n}Q_i(\overline{y}_i-\underline{y}_i)$.

(c) The point v is a vertex of $OLS(X,\boldsymbol{y})$ if and only if $c-2v$ is a vertex.

Proof. (a) follows from the observation that the Minkowski sum preserves center-symmetry (i.e. if A is center-symmetric, then $A\dot{+}g$ is center-symmetric). The remaining statements follow from the symmetry property. $\square$

3.2 Descriptions of the set $\boldsymbol{OLS}(X,\boldsymbol{y})$

In order the user can understand how the set $OLS(X,\boldsymbol{y})$ looks like, she/he can use any standard description applicable for convex polyhedra. In particular, three descriptions come to mind:

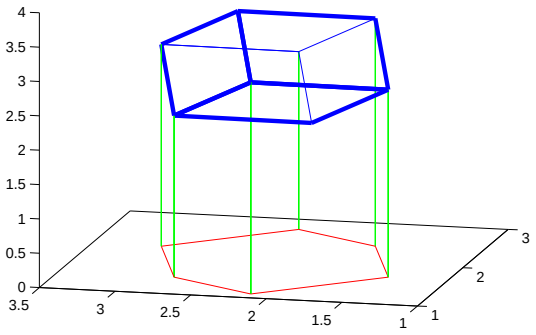

Figure 2: A zonotope as an image of a higher-dimensional cube.

(a) description of the zonotope $OLS(X, \boldsymbol{y})$ by the shift vector and the set of generators;

(b) description of the zonotope $OLS(X, \boldsymbol{y})$ by the enumeration of vertices;

(c) description of the zonotope $OLS(X, \boldsymbol{y})$ by the enumeration of facets. Each facet of a polyhedron can be identified with a supporting halfspace (or hyperplane), i.e. with an inequality of the form $a^{\mathrm{T}}x \leq \gamma$. Hence, construction of a facet description amounts to constructing a matrix A and a vector c such that $OLS(X, \boldsymbol{y}) = \{x \in \mathbb{R}^p : Ax \leq c\}$.

The description (a) has been given by the Theorem 9: we have

$$s = Q\underline{y}, \quad g_i = Q_i(\overline{y}_i - \underline{y}_i) \quad \text{for } i = 1, \ldots, n. \tag{6}$$

The descriptions (b) and (c) will be investigated in Sections 3.3, 3.4 and 3.5. Algorithms for their construction will be studied in Section 5.

3.3 A negative complexity result

It is an interesting question whether there are efficient algorithms which can construct the enumerations (b) and (c) given X, $\underline{y}$ and $\overline{y}$. We give an argument that the answer is negative. The answer follows from the simple fact that zonotopes can have too many vertices and facets. Let $V(\mathcal{Z})$ and $F(\mathcal{Z})$ denote the number of vertices and facets of a zonotope $\mathcal{Z}$, respectively.

Theorem 11 ([72]). *For a zonotope $\mathcal{Z} \subseteq \mathbb{R}^p$ with n generators it holds*

$$V(\mathcal{Z}) \leq 2\sum_{k=0}^{p-1} \binom{n-1}{k} \quad \textit{and} \quad F(\mathcal{Z}) \leq 2\binom{n}{p-1}.$$

Moreover, the bounds cannot be improved: for each n and p there are p-dimensional zonotopes $\mathcal{Z}_V$ and $\mathcal{Z}_F$ with n generators such that $V(\mathcal{Z}_V)$ and $F(\mathcal{Z}_F)$ attain the bounds. □

The numbers $V(\mathcal{Z})$ and $F(\mathcal{Z})$ cannot be bounded by a polynomial in n and p; hence, the functions enumerating vertices and facets are not in ***PF*** for the simple reason that their output cannot be bounded by a polynomial in the size of the input. By the next lemma we know that this negative result also holds for zonotopes which are OLS-sets of crisp input – interval output regression models.

Lemma 12. *For every p-dimensional zonotope $\mathcal{Z}$ with n generators, there exists a regression model $(X, \boldsymbol{y})$ with p regression parameters and n observations such that $V(OLS(X, \boldsymbol{y})) = V(\mathcal{Z})$ and $F(OLS(X, \boldsymbol{y})) = F(\mathcal{Z})$.*

Proof. Let $g_1, \ldots, g_n$ be the generators of $\mathcal{Z}$. Consider the matrix $G \in \mathbb{IR}^{p \times n}$ with columns $g_1, \ldots, g_n$ and the n-dimensional vector $\boldsymbol{z}$ with $\boldsymbol{z}_i := [0, 1]$ for $i = 1, \ldots, n$. By the same argument as in (5), the zonotope $\mathcal{Z}' = \{Gz : z \in \boldsymbol{z}\}$ is, up to a shift, the same zonotope as $\mathcal{Z}$. By the assumption that $\mathcal{Z}$ has dimension p, we know that G has full row rank, and hence GG^{T} is regular.

Consider the model $(X, \boldsymbol{y})$ with $X := G^{\mathrm{T}}$ and $\boldsymbol{y} := \boldsymbol{z}$. Then

$$OLS(X, \boldsymbol{y}) = \{(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y : y \in \boldsymbol{y}\} = \{(GG^{\mathrm{T}})^{-1}Gz : z \in \boldsymbol{z}\}.$$

The last expression shows that the zonotope $OLS(X, \boldsymbol{y})$ is an image of the zonotope $\mathcal{Z}'$ under the regular linear mapping $\xi \mapsto (GG^{\mathrm{T}})^{-1}\xi$. Regular linear mappings preserve the number of facets and vertices of polyhedra; hence $V(OLS(X, \boldsymbol{y})) = V(\mathcal{Z})$ and $F(OLS(X, \boldsymbol{y})) = F(\mathcal{Z})$. □

3.4 A positive complexity result

Theorem 11 has an interesting corollary if we treat the number p as a fixed constant (i.e. if we restrict ourselves to a class of regression models with a fixed number of regression parameters).

Corollary 13. *If p is fixed then*

$$V(OLS(X, \boldsymbol{y})) \leq O(n^{p-1}) \quad \textit{and} \quad F(OLS(X, \boldsymbol{y})) \leq O(n^{p-1}).$$

Proof. We have

$$F(OLS(X, \boldsymbol{y})) \leq 2\binom{n}{p-1} = \frac{2n(n-1)\cdots(n-p+2)}{(p-1)!} \leq 2n^{p-1} \leq O(n^{p-1}) \tag{7}$$

and

$$V(OLS(X, \boldsymbol{y})) \leq 2\sum_{k=0}^{p-1}\binom{n-1}{k} \leq 2p \cdot \max_{k \in \{0,\ldots,p-1\}}\binom{n-1}{k} \overset{(\star)}{\leq} O(n^{k_{\max}}) = O(n^{p-1}),$$

where $k_{\max}$ is the $k \in \{0, \ldots, p-1\}$ for which the maximum is attained. By well-known properties of binomial coefficients, for n large enough it holds $k_{\max} = p-1$. In the inequality $(\star)$ we used a similar estimate as in (7). □

Corollary 13 implies that the number of vertices and the number of facets is polynomial in n (provided that p is fixed). In Section 5 we show a general method for enumeration of vertices and facets running in time polynomial in *size*(*input*) + *size*(*output*). The assumption that p is fixed implies that *size*(*output*) is polynomially bounded by *size*(*input*). Thus the algorithm runs in polynomial time under that assumption.

3.5 Regression models with repeated observations

Let p be fixed. Corollary 13 can be easily strengthened to the form

$$V(OLS(X, \boldsymbol{y})) \leq O(n^{\dim(OLS(X,\boldsymbol{y}))-1}), F(OLS(X, \boldsymbol{y})) \leq O(n^{\dim(OLS(X,\boldsymbol{y}))-1}),$$

where dim refers to Definition 8(b). In statistical applications, this reduction usually does not help as it is rarely $\dim(OLS(X, \boldsymbol{y})) < p$. This could happen, for example, if the great majority of observations are crisp, i.e. if the cardinality of $\{i \in \{1, \ldots, n\} : \underline{y}_i < \overline{y}_i\}$ is smaller than p.

However, there are important special cases where another significant reduction can be reached. The reduction is based on the following easy lemma.

Lemma 14 (reduction lemma)**.** *Let $\mathcal{Z} := \mathcal{Z}(s; g_1, g_2, \ldots, g_n)$ and for some i and j, $i < j$ it holds $g_j = \alpha g_i$, where $\alpha \in \mathbb{R}$. Then*

$$\mathcal{Z} = \begin{cases} \mathcal{Z}(s; g_1, \ldots, g_{i-1}, g_i + g_j, g_{i+1}, \ldots, g_{j-1}, g_{j+1}, \ldots, g_n) & \text{if } \alpha \geq 0, \\ \mathcal{Z}(s + g_j; g_1, \ldots, g_{i-1}, g_i - g_j, g_{i+1}, \ldots, g_{j-1}, g_{j+1}, \ldots, g_n) & \text{if } \alpha < 0. \end{cases}$$

□

The generator g_j is called *redundant*. The process of removal of redundant generators may be iterated until all are removed; then we obtain a certain shift s' and a reduced set of generators $g'_1, \ldots, g'_{n'}$ with $n' \leq n$ defining the same zonotope. Later, the following notation will be useful:

$$red(s; g_1, \ldots; g_n) := (s', g'_1, \ldots, g'_{n'}). \tag{8}$$

We can reformulate Theorem 11 in the following way.

Corollary 15. *Let $\sim$ be an equivalence on rows of X: $X_i \sim X_j$ iff X_i is a multiple of X_j. Let ν be the number of equivalence classes of $\sim$. Then,*

$$V(OLS(X, \boldsymbol{y})) \leq 2 \sum_{k=0}^{p-1} \binom{\nu - 1}{k} \quad \text{and} \quad F(OLS(X, \boldsymbol{y})) \leq 2 \binom{\nu}{p-1}.$$

In particular, if p is fixed, then $V(OLS(X, \boldsymbol{y})) \leq O(\nu^{p-1})$ and $F(OLS(X, \boldsymbol{y})) \leq O(\nu^{p-1})$.

Proof. If X_i is a multiple of X_j, then $Q_i(\overline{y}_i - \underline{y}_i)$ is a multiple of $Q_j(\overline{y}_j - \underline{y}_j)$ and we may apply Lemma 14. □

If the absolute term is involved in the regression model (i.e., X contains the all-one column), then $X_i \sim X_j$ iff $X_i = X_j$. So, rather that saying that the combinatorial complexity of the zonotope $OLS(X, \boldsymbol{y})$ depends on the number of observations, it is more appropriate to say that *the complexity depends on the number of distinct design points* (i.e. distinct rows of X). Regression models with $\nu \ll n$, called *models with repeated observations*, are very important in practice.

3.6 Algorithmic properties of $\boldsymbol{OLS}(X, \boldsymbol{y})$

The following theorem summarizes some complexity-theoretic facts about the set $OLS(X, \boldsymbol{y})$, which will be useful in the next sections. In the proof we use the well-known fact that linear programming is a polynomial-time solvable problem, see [65].

Definition and Theorem 16. *Let $X, \underline{y}, \overline{y}$ and β be rational.*

*(a) We say that β is **admissible** for $(X, \boldsymbol{y})$ if $\beta \in OLS(X, \boldsymbol{y})$. The question "is β admissible?" is in $\boldsymbol{P}$.*

*(b) We say that β is **extremal** for $(X, \boldsymbol{y})$ if β is on the boundary of $OLS(X, \boldsymbol{y})$. The question "is β extremal?" is in $\boldsymbol{P}$.*

(c) The question "is β a vertex of $OLS(X, \boldsymbol{y})$?" is in $\boldsymbol{P}$.

Proof. For (a) observe that admissibility is decidable via the linear program

$$\max\ 0^{\mathrm{T}} y :\ \beta = Qy,\ \underline{y} \le y \le \overline{y}. \tag{9}$$

To prove (b) assume that the center of $OLS(X, \boldsymbol{y})$ is 0 (by Lemma 10, the zonotope $OLS(X, \boldsymbol{y})$ can be shifted). Then β is extremal iff the optimal value of the linear program

$$\max\ w :\ w\beta = Qy,\ \underline{y} \le y \le \overline{y}$$

is $w = 1$.

In the proof of (c) we may assume that all generators $g_i := Q_i(\overline{y}_i - \underline{y}_i)$ are nonzero. Observe that β is not a vertex iff there exists a generator g_i such that β can be shifted both in the direction g_i and in the direction $-g_i$. So, β is a vertex iff for each $i = 1, \ldots, n$ it holds that the linear program

$$\max\ w :\ \beta + wg_i = Qy,\ \beta - wg_i = Qz,\ \underline{y} \le y \le \overline{y},\ \underline{y} \le z \le \overline{y}$$

has the optimal value $w = 0$. □

3.7 Estimation of a linear combination of regression parameters

Assume that a nonzero vector c of parameters is given. Our aim is to estimate the linear combination $c^{\mathrm{T}}\beta$ of regression parameters.

An example of the choice of c is $c^{\mathrm{T}} = (1, 0, \ldots, 0)$. In this case we want to estimate the first regression parameter. This choice is interesting, for example, in the model $y = \beta_1 x^2 + \beta_2 x + \beta_3 + \varepsilon$; if we can measure β_1 with high precision, then we can determine whether it is zero or not. In other words, we can determine whether the model is linear in x or not.

Let us ask the question how the OLS-estimator $c^{\mathrm{T}}\widehat{\beta}$ of $c^{\mathrm{T}}\beta$ is affected by censoring.

By lemma 10 we can assume that the center of $OLS(X, \boldsymbol{y})$ is 0. Then we can solve the linear program

$$\max\ w:\ wc = Qy,\ \underline{y} \le y \le \overline{y}.$$

If w^* is the optimal value, then $\chi_c := 2w^*\|c\|$, where $\|\cdot\|$ is the L_2-norm, is the diameter of $OLS(X, \boldsymbol{y})$ in the direction c. Then χ_c measures how censoring can affect the estimator $c^{\mathrm{T}}\widehat{\beta}$ in the worst case. If the value χ_c is small, then we can say that *the loss of information caused by censoring effect is negligible.* (Often it is the case that for some choices of c the value χ_c is small while for other choices of c the value χ_c is large.) The interval $[-\chi_c, \chi_c]$ is called *covering* of $c^{\mathrm{T}}\widehat{\beta}$ or "region of 100% certainty" about $c^{\mathrm{T}}\widehat{\beta}$. In some sense, it can be seen as an analogy of the notion of the confidence interval for $c^{\mathrm{T}}\beta$ in classical (crisp) regression.

4 Approximations of the set $\boldsymbol{OLS}(X, \boldsymbol{y})$ in the crisp input − interval output model

A complete vertex description and a complete facet description of the zonotope $OLS(X, \boldsymbol{y})$ is available via the methods described in Section 5. In many cases this description is not friendly for a user — fox example, in a model with $p = 4$ regression parameters and $n = 100$ observations, the enumeration of vertices or facets can fill up a thick book. This is just one reason justifying that it is suitable to consider some user-friendly approximations of the set $OLS(X, \boldsymbol{y})$. Another reason is that the methods of Section 5 do not run in polynomial time.

Let Q denote the matrix $(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$.

4.1 Interval approximation

By (3) we obtain that for any $b \in OLS(X, \boldsymbol{y})$ and any $i = 1, \ldots, p$,

$$\underbrace{\sum_{j=1}^{n} \min\{Q_{ij}\underline{y}_j, Q_{ij}\overline{y}_j\}}_{=:\underline{b}_i} \le b_i \le \underbrace{\sum_{j=1}^{n} \max\{Q_{ij}\underline{y}_j, Q_{ij}\overline{y}_j\}}_{=:\overline{b}_i}. \tag{10}$$

Moreover, the cube $[\underline{b}, \overline{b}]$ is the smallest cube containing $OLS(X, \boldsymbol{y})$. This bound is easily computable.

4.2 Simple ellipsoidal approximation

Combinatorially complex polyhedra are often approximated with ellipses: an ellipse is a convex set which is quite flexible to approximate the shape of the polyhedron and it is sufficiently simple to be described.

An ellipse $\mathcal{E}$ is described by a center point s and a positive definite matrix E such that

$$\mathcal{E} = \{x \in \mathbb{R}^p : (x-s)^{\mathrm{T}} E^{-1} (x-s) \leq 1\}.$$

We do not know a polynomial-time algorithm for construction of the *best* (i.e., the smallest with respect to volume) circumscribing ellipse for the set $OLS(X, \boldsymbol{y})$. It is an intriguing research problem; however, we expect a hardness result on this computational problem rather than a polynomial-time algorithm.

The following ellipse $\mathcal{E} = (E, s)$ can be seen as a weaker form:

$$\begin{aligned} s &= \tfrac{1}{2} Q(\overline{y} + \underline{y}), \\ E &= Q \cdot diag\left(\tfrac{n}{4}(\overline{y}_1 - \underline{y}_1)^2, \ldots, \tfrac{n}{4}(\overline{y}_n - \underline{y}_n)^2 \right) \cdot Q^{\mathrm{T}}, \end{aligned} \tag{11}$$

where $diag(\xi_1, \ldots, \xi_n)$ denotes the diagonal matrix with diagonal entries $\xi_1, \ldots, \xi_n$. This is the ellipse which is the image of the smallest ellipse circumscribing $\boldsymbol{y}$ in $\mathbb{R}^n$ under the mapping $\xi \mapsto Q\xi$. This proves $OLS(X, \boldsymbol{y}) \subseteq \mathcal{E}$.

Another ellipsoidal approximation will be investigated in Section 4.4.

4.3 Example

Consider the regression model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \tag{12}$$

with $n = 11$ observations collected in the following table. Only interval-censored values are available to us:

$$\boldsymbol{y}_i = [\underline{y}_i, \overline{y}_i] = [\tilde{y}_i - \tfrac{1}{2}, \tilde{y}_i + \tfrac{1}{2}], \quad i = 1, \ldots, 11$$

where $\tilde{y}$ denotes the center of $\boldsymbol{y}$. This situation can occur, for example, in the case of integer rounding of measured data y.

i	1	2	3	4	5	6
x_i	−2	−1	0	1	2	3
$\underline{y}_i$	1.5	−1.5	−0.5	3.5	3.5	5.5
$\tilde{y}_i$	2	−1	0	4	4	6
$\overline{y}_i$	2.5	−0.5	0.5	4.5	4.5	6.5
i	7	8	9	10	11	
x_i	4	5	6	7	8	
$\underline{y}_i$	8.5	6.5	10.5	10.5	9.5	
$\tilde{y}_i$	9	7	11	11	10	
$\overline{y}_i$	9.5	7.5	11.5	11.5	10.5	

Using the central estimator $\tilde{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\tilde{y}$ we get

$$\tilde{\beta}_1 = 2.12, \quad \tilde{\beta}_2 = 1.2$$

and with (10) we get

$$[\underline{b}_1, \overline{b}_1] = [1.56, 2.69], \quad [\underline{b}_2, \overline{b}_2] = [1.06, 1.34].$$

We can conclude that the censoring effect couldn't have had the impact higher than ± 0.565 $[= \frac{1}{2}(2.69 - 1.56)]$ on the estimate of β_1 and the impact higher than ± 0.14 $[= \frac{1}{2}(1.34 - 1.06)]$ on the estimate of β_2.

The set (zonotope) $OLS(X, \boldsymbol{y})$, together with the enclosure B given by (10) and the ellipse (11), is plotted in Figure 3.

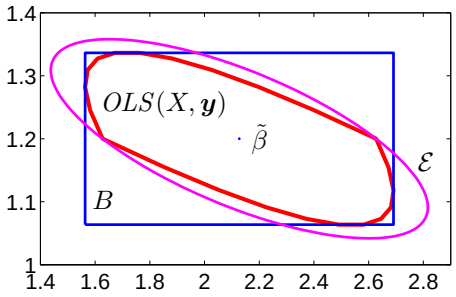

Figure 3: The set (zonotope) $OLS(X, \boldsymbol{y})$ for the regression model in the example of Section 4.3 and its approximations B and $\mathcal{E}$ given by (10) and (11), respectively.

Though the approximations of Sections 4.1 and 4.2 are quite trivial, their combination gives some nontrivial information. The enclosure B contains the point $[1.56, 1.06]$; hence, the approximation B does not rule out the case that *both regression parameters could have been affected by the maximal possible "error"* $[-0.565, -0.14]$ *in the negative direction simultaneously.* However, this case is ruled out by the fact that $[1.65, 1.06] \notin \mathcal{E}$.

4.4 Approximation of $\boldsymbol{OLS}(X, \boldsymbol{y})$ by (a form of) the Löwner-John ellipse

If $\mathcal{E}$ is an ellipse (E, s), then $\alpha \cdot \mathcal{E}$ is the ellipse (α^2, s), i.e. an ellipse blown-up by a factor α. Let $\varepsilon > 0$.

Definition 17. *Let $P \subseteq \mathbb{R}^p$ and $\varepsilon > 0$.*

(a) Any ellipse $\mathcal{E}$ satisfying $\frac{1}{p} \cdot \mathcal{E} \subseteq P \subseteq \mathcal{E}$ is called ***Löwner-John ellipse*** *for P. Any ellipse $\mathcal{E}$ satisfying $\frac{1}{p} \cdot \mathcal{E} \subseteq P \subseteq (1+\varepsilon) \cdot \mathcal{E}$ is called* ***ε-approximate Löwner-John ellipse*** *for P.*

(b) *Any ellipse* $\mathcal{E}$ *satisfying* $\frac{1}{\sqrt{p}} \cdot \mathcal{E} \subseteq P \subseteq \mathcal{E}$ *is called* ***Jordan ellipse*** *for* P. *Any ellipse* $\mathcal{E}$ *satisfying* $\frac{1}{\sqrt{p}} \cdot \mathcal{E} \subseteq P \subseteq (1+\varepsilon) \cdot \mathcal{E}$ *is called* ε***-approximate Jordan ellipse*** *for* P.

The following theorems, known as Löwner-John Theorem and Jordan Theorem, respectively, are basic results in polyhedral geometry (see [17], [65]):

(i) *any full-dimensional bounded convex set* $P \subseteq \mathbb{R}^p$ *has a Löwner-John ellipse*;

(ii) *any full-dimensional bounded centrally symmetric convex set* $P \subseteq \mathbb{R}^p$ *has a Jordan ellipse.*

Moreover, the factors $\frac{1}{p}$ and $\frac{1}{\sqrt{p}}$ in Definition 17 cannot be improved in general unless the statements (i) and (ii) are violated.

The Löwner-John and Jordan ellipses have the advantage that they provide us with *both a lower and an upper bound* on the boundary of P. When $P = OLS(X, \boldsymbol{y})$, we get an interesting approximation of $OLS(X, \boldsymbol{y})$.

If the set $OLS(X, \boldsymbol{y})$ is full-dimensional, it has a Jordan ellipse — we know that zonotopes are centrally symmetric. Of course, we would like to describe an *algorithmic method* for finding the ellipses. Unfortunately both theorems (i) and (ii) are nonconstructive.

Algorithms for construction of Löwner-John ellipses and Jordan ellipses are studied in computational geometry. In particular, the following theorem was proved by Goffin [16]; see also [6], [17], [65]. It is known as a constructive form of the Löwner-John Theorem for polyhedra.

(i) *Let* $\varepsilon > 0$ *be fixed. Given rational* A *and* b *such that the polyhedron* $P := \{x \in \mathbb{R}^p : Ax \leq b\}$ *is full-dimensional and bounded, its* ε*-approximate Löwner-John ellipse can be found in polynomial time.*

The method is known as Goffin's Algorithm and it is based on Khachiyan's Ellipsoid Method. The method can be adapted for central-symmetric polyhedra (see [6], [17]). Then, Goffin's Theorem can be stated in the following way. It is known as a constructive form of the Jordan Theorem for polyhedra.

(ii) *Let* $\varepsilon > 0$ *be fixed. Given rational* A *and* b *such that the polyhedron* $P := \{x \in \mathbb{R}^p : Ax \leq b\}$ *is full-dimensional, bounded and centrally symmetric, its* ε*-approximate Jordan ellipse can be found in polynomial time.*

At the first sight it seems that the form (ii) of Goffin's Theorem could be used for a polynomial-time construction of the Jordan ellipse for $OLS(X, \boldsymbol{y})$. But there is a serious obstacle: the algorithm requires the facet description $Ax \leq b$ of the set $OLS(X, \boldsymbol{y})$. From Section 3.3 we know that in general, computation of A and b, when generators $g_1, \ldots, g_n$ are given, is not a polynomial-time procedure.

From section 3.4 we know that it is a polynomial time procedure under the assumption that the dimension p is fixed. Indeed, the facet description can be found using the methods of Section 5. This observation together with Goffin's Theorem (ii) imply the following interesting statement:

Corollary 18. *Let $\varepsilon > 0$ and the dimension p be fixed. There is a polynomial-time algorithm which computes the ε-approximate Jordan ellipse for $OLS(X, \boldsymbol{y})$.* □

If the dimension is not fixed, we can prove only a weaker result.

Theorem 19 ([6])**.** *Let $\varepsilon > 0$ be fixed. There is a polynomial-time algorithm which computes the ε-approximate Löwner-John ellipse for $OLS(X, \boldsymbol{y})$.* □

Research problem. It is an intriguing research problem whether Theorem 19 can be reformulated with Jordan ellipse instead of Löwner-John ellipse. (Or at least with an ellipse $\mathcal{E}$ satisfying $p^{-\gamma} \cdot \mathcal{E} \subseteq OLS(X, \boldsymbol{y}) \subseteq (1+\varepsilon) \cdot \mathcal{E}$ with some $\gamma \in (\frac{1}{2}, 1)$.)

By [6], the problem is tightly interconnected with the following question. Assume that $OLS(X, \boldsymbol{y})$ is centered at zero (by Lemma 10(b), this assumption is without loss of generality). Let $K_\gamma := \{x : \|x\| \leq \gamma\}$ be a ball with radius γ. Let T be the the problem *given rational $X, \underline{y}, \overline{y}$ and $\gamma > 0$, decide whether $K_\gamma \subseteq OLS(X, \boldsymbol{y})$*. The following holds: if $T \in \overline{\boldsymbol{P}}$, in Theorem 19 the Löwner-John ellipse can be replaced by the Jordan ellipse. On the other hand, if $T \notin \boldsymbol{P}$, this fact is a serious obstacle which seems to rule out the Goffin's method from the attempts of improvement of Theorem 19.

At the moment we cannot prove $T \in \boldsymbol{P}$. We conjecture that the problem T is *co-**NP***-hard.

4.5 Approximation of volume of $\boldsymbol{OLS}(X, \boldsymbol{y})$

Volume of $OLS(X, \boldsymbol{y})$ is a natural measure of its size, i.e. a natural measure of "uncertainty" about the OLS-estimation of the regression model resulting from censoring (or uncertainty) of y. In Section 5 we shall present an algorithm for exact computation of volume of $OLS(X, \boldsymbol{y})$. However, that algorithms requires (in general) high computation resources: to our knowledge, no polynomial-time algorithm (polynomial in n, p) is known.

Theorem 16(a) combined with (10) suggests a simple procedure for Monte-Carlo approximation of volume: just generate a random point $b \in [\underline{b}, \overline{b}]$ and test its admissibility. This procedure is interesting in particular in higher dimensions.

5 The Reduction-and-Reconstruction-Recursion ("RRR") Algorithm

In this section we are still dealing with the crisp input – interval output model, i.e. $\underline{X} = \overline{X} =: X$. From the previous sections we know that in this case, the

set $OLS(X, \boldsymbol{y})$ is a zonotope. Here we introduce a meta-algorithm that can be used for answering several questions about $OLS(X, \boldsymbol{y})$. In particular, we get an algorithm for

(a) enumeration of vertices of $OLS(X, \boldsymbol{y})$,

(b) enumeration of facets of $OLS(X, \boldsymbol{y})$ (i.e. finding a description of $OLS(X, \boldsymbol{y})$ in terms of linear inequalities in $\mathbb{IR}^p$),

(c) computation of volume of $OLS(X, \boldsymbol{y})$.

Recall that an approximate algorithm for estimation of volume has been presented in Section 4.5.

Algorithms for zonotopes are discussed in [4] and [13].

Given a family $a_1, \ldots, a_k$ of vectors, $\mathcal{L}(a_1, \ldots, a_k)$ denotes the linear space generated by $a_1, \ldots, a_k$ and $\mathcal{L}(a_1, \ldots, a_k)^\perp$ denotes its orthogonal complement.

Remark. Given a sequence of generators $g_1, \ldots, g_n$ of a zonotope $\mathcal{Z} \subseteq \mathbb{IR}^p$, we can easily determine $D := \dim(\mathcal{Z})$ by evaluating the dimension of $\mathcal{L}(g_1, \ldots, g_n)$. *Volume* is understood as volume in the D-dimensional space $\mathcal{L}(g_1, \ldots, g_n)$, since $\mathcal{Z}$ is full-dimensional in that space (while it is of zero volume in $\mathbb{IR}^p$ if $D < p$).

The Reduction-and-Reconstruction-Recursion (*RRR*) Algorithm is a meta-algorithm that uses routines *BasicCase* and *Combine*. Particular choices of these routines will be specified later.

Observe that for any permutation π of $\{1, \ldots, n\}$, $\mathcal{Z}(s, g_1, \ldots, g_n) = \mathcal{Z}(s, g_{\pi(1)}, \ldots, g_{\pi(n)})$. We may assume that whenever we work with a sequence $g_1, \ldots, g_n$ of generators of a zonotope $\mathcal{Z}$, the vectors $g_1, \ldots, g_{\dim(\mathcal{Z})}$ are linearly independent. Say that the output of the procedure *red* defined by (8) meets this requirement. Let $\mathcal{P}_v(u)$ denote the orthogonal projection of the vector u into the space $\{\xi : v^{\mathrm{T}}\xi = 0\}$.

```
{1}  function RRR(s̃; g̃_1, ..., g̃_ñ)
{2}    (s, g_1, ..., g_n) := red(s̃; g̃_1, ..., g̃_ñ)
{3}    if g_1, ..., g_n are linearly independent then
{4}          OUTPUT := BasicCase(s; g_1, ..., g_n)
{5}       else
{6}          DATA_1 := RRR(s; g_1, g_2, ..., g_{n-1})
{7}          DATA_2 := RRR(P_{g_n}(s); P_{g_n}(g_1), P_{g_n}(g_2), ..., P_{g_n}(g_{n-1}))
{8}          OUTPUT := Combine(s; g_1, ..., g_n; DATA_1, DATA_2)
{9}  end.
```

The steps {6} and {7} are computed recursively.

Observe that by {2} it holds that in the collection $g_1, \ldots, g_n$ there is no pair of parallel generators. This observation is important for the correctness of Application 2.

5.1 Application 1 — Volume computation

Let us define

$$Combine(s; g_1, \ldots, g_n; \alpha_1, \alpha_2) = \alpha_1 + \alpha_2 \|g_n\|,$$

where $\|\cdot\|$ denotes the L_2-norm, and

{1} **function** *BasicCase*$(s; g_1, \ldots, g_n \in \mathbb{R}^k)$
{2} let $g_{n+1}, \ldots, g_k$ be any orthonormal basis of $\mathcal{L}(g_1, \ldots, g_n)^\perp$
{3} $OUTPUT := |\det(g_1 \ \cdots \ g_k)|$
{4} **end.**

In the basic case, the zonotope is a parallelogram and hence its volume is $|\det(g_1 \ \cdots \ g_k)|$. The vectors $g_{n+1}, \ldots, g_k$ have been appended to the sequence $g_1, \ldots, g_n$ in a way meeting the convention of the Remark from the beginning of Section 5.

By symmetry of zonotopes (see the evolution process of Figure 1), adding a new generator g_n to the zonotope $\mathcal{Z}(s; g_1, \ldots, g_{n-1})$ increases its volume α_1 by the volume of the prism with base $\mathcal{Z}(\mathcal{P}_{g_n}(s); \mathcal{P}_{g_n}(g_1), \ldots, \mathcal{P}_{g_n}(g_{n-1}))$ and height $\|g_n\|$. From the recursion, α_2 is the area of the base.

5.2 Application 2 — Enumeration of vertices

In the *Basic Case*, the zonotope is a parallelogram:

$$BasicCase(s; g_1, \ldots, g_n) := \left\{ s + \sum_{i=1}^{n} c_i g_i : c \in \{0,1\}^n \right\}.$$

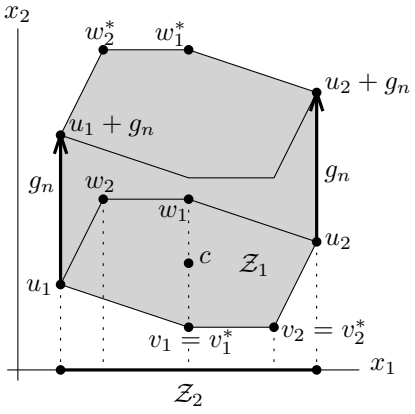

Figure 4: Enumeration of vertices. The zonotope $\mathcal{Z} := \mathcal{Z}(s; g_1, \ldots, g_n)$ is shaded. From the recursion, the vertex description $U \cup I$ of $\mathcal{Z}_1$ is available, where $U = \{u_1, u_2\}$ and $I = \{(v_1, w_1), (v_2, w_2)\}$. The center of $\mathcal{Z}_1$ is denoted c. The vertices u_1, u_2 are utmost; they can be detected with the knowledge of the vertex description of $\mathcal{Z}_2$ which is also available from recursion. The vertex description of $\mathcal{Z}$ is $u_1, u_1 + g_n, u_2, u_2 + g_n, v_1^*, w_1^*, v_2^*, w_2^*$.

We shall describe the function *Combine*$(s; g_1, \ldots, g_n; VertexSet_1, VertexSet_2)$ with reference to Figure 4. From the recursion, *VertextSet*$_1$ is the set of vertices

of the zonotope $\mathcal{Z}_1 := \mathcal{Z}(s; g_1, \ldots, g_{n-1})$ and *VertextSet*$_2$ the set of vertices of a zonotope $\mathcal{Z}_2$, which is the projection of $\mathcal{Z}_1$ on the hyperplane $\{\xi : g_n^{\mathrm{T}}\xi = 0\}$.

We say that a vertex $v \in$ *VertexSet*$_1$ is *utmost* if $\mathcal{P}_{g_n}(v) \in$ *VertexSet*$_2$. Otherwise it is *inner*. Consider the decomposition *VertexSet*$_1 = U \cup I$, where U are utmost vertices and I are inner vertices.

Now we process the utmost vertices: we set $OUTPUT_1 := \{u, u + g_n : u \in U\}$.

By center symmetry of the zonotope $\mathcal{Z}_1$, the set I of inner vertices can be seen as a set of center-symmetric pairs $(v_1, w_1), \ldots (v_\ell, w_\ell)$. For each pair (v_i, w_i), we get the transformed pair (v_i^*, w_i^*) in the way that one of the vertices is preserved and the other is shifted by the generator g_n; in Figure 4 we have, for example, $v_i^* = v_i$ and $w_i^* = w_i + g_n$ for $i = 1, 2$. Then we set $OUTPUT_2 := \{v_i^*, w_i^* : i = 1, \ldots, \ell\}$. The final output of the procedure *Combine* is $OUTPUT_1 \cup OUTPUT_2$.

5.3 Application 3 — Enumeration of facets

The procedure *BasicCase*$(s; g_1, \ldots, g_n \in \mathbb{R}^k)$ finds the facet description of the parallelogram generated by $g_1, \ldots, g_n$. Let $g_{n+1}, \ldots, g_k$ be any basis of $\mathcal{L}(g_1, \ldots, g_n)^\perp$ and let G be the matrix with columns $g_1, \ldots, g_k$. Then, the facet description is

$$\left\{ x : \begin{array}{ll} -e_i^{\mathrm{T}} G^{-1}(x - s) \le 0, \; e_i^{\mathrm{T}} G^{-1}(x - s) \le 1, & i = 1, \ldots, n; \\ -e_i^{\mathrm{T}} G^{-1}(x - s) \le 0, \; e_i^{\mathrm{T}} G^{-1}(x - s) \le 0, & i = n+1, \ldots, k \end{array} \right\},$$

where e_i is the i-th column of the unit matrix.

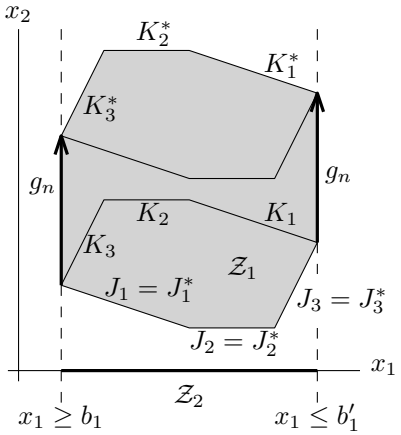

Figure 5: Enumeration of facets. The zonotope $\mathcal{Z} := \mathcal{Z}(s; g_1, \ldots, g_n)$ is shaded. From the recursion, the facet description $x_1 \ge b_1$, $x_1 \le b_1'$, $x_2 \ge 0$, $x_2 \le 0$ of $\mathcal{Z}_2$ and the facet description $(J_1, K_1), (J_2, K_2), (J_3, K_3)$ of $\mathcal{Z}_1$ is available. The facet description of $\mathcal{Z}$ is $x_1 \ge b_1$, $x_1 \le b_1'$, $(J_1^*, K_1^*), (J_2^*, K_2^*), (J_3^*, K_3^*)$.

We shall describe the function $Combine(s; g_1, \ldots, g_n \in \mathbb{IR}^k; Facets_1, Facets_2)$ with reference to Figure 5. From the recursion, $Facets_1$ is the set of facets of the zonotope $\mathcal{Z}_1 := \mathcal{Z}(s; g_1, \ldots, g_{n-1})$ and $Facets_2$ is the set of facets of the zonotope $\mathcal{Z}_2$, which is the projection of $\mathcal{Z}_1$ on the hyperplane $\{\xi : g_n^{\mathrm{T}}\xi = 0\}$. For simplicity say that g_n is parallel with e_k (which can be achieved by rotation). Then, the set $Facets_2$ contains inequalities $x_k \leq 0$ and $x_k \geq 0$. These inequalities are disregarded; the remaining ones are copied to $OUTPUT_1$.

By symmetry of $\mathcal{Z}_1$, the inequalities in $Facets_1$ can be seen as pairs $(J_i, K_i) := (a_i^{\mathrm{T}} x \geq b_i, a_i x \leq b_i')$, $i = 1, \ldots, \ell$. Each pair (J_i, K_i) is transformed to a pair (J_i^*, K_i^*) in the way that one inequality is preserved and the other is shifted by the generator g_n. In Figure 5, we have $J_i^* = J_i$ and $K_i^* = [a_i^{\mathrm{T}}(x - g_n) \leq b_i']$ for $i = 1, 2, 3$. We set $OUTPUT_2 = \{J_i^*, K_i^* : i = 1, \ldots, \ell\}$. The final output of the procedure *Combine* is $OUTPUT_1 \cup OUTPUT_2$.

6 Approximations of $\boldsymbol{OLS}(\boldsymbol{X}, \boldsymbol{y})$ in the general case

As we know from Theorem 5, in the general case $(\boldsymbol{X}, \boldsymbol{y})$ the situation is much more difficult than in the crisp input case. Therefore we cannot expect general results. However, the *co-**NP***-hardness result does not rule out that good approximations of $OLS(\boldsymbol{X}, \boldsymbol{y})$ could be found in particular cases. As the general case is very important in practice, in this section we review applicable methods for interval approximation of $OLS(\boldsymbol{X}, \boldsymbol{y})$. Those methods, which — to our knowledge — provide the tightest approximations, are illustrated by three examples. The examples should provide a basic insight showing what results we could expect from the methods. However, the main message of the examples is that — even in very simple cases — the known methods for the general case can hardly be considered as satisfactory. The examples could also serve as an impulse for further research.

For the sake of completeness, we also give a short overview of available software.

Given an interval matrix $\boldsymbol{Z} = [\underline{Z}, \overline{Z}]$, the matrix

$$Z^c = \frac{1}{2}(\underline{Z} + \overline{Z})$$

is called *center matrix* and the matrix

$$Z^\Delta = \frac{1}{2}(\overline{Z} - \underline{Z})$$

is called *radius matrix*.

Recall also that interval addition and multiplication has been defined by (3). We shall also need the division of interval numbers $\boldsymbol{a} = [\underline{a}, \overline{a}]$ and $\boldsymbol{b} = [\underline{b}, \overline{b}]$, which is defined as

$$\frac{\boldsymbol{a}}{\boldsymbol{b}} = \begin{cases} \left[\min\{\underline{a} \div \underline{b}, \underline{a} \div \overline{b}, \overline{a} \div \underline{b}, \overline{a} \div \overline{b}\}, \max\{\underline{a} \div \underline{b}, \underline{a} \div \overline{b}, \overline{a} \div \underline{b}, \overline{a} \div \overline{b}\}\right] & \text{if } 0 \notin \boldsymbol{b}, \\ \text{undefined} & \text{otherwise,} \end{cases}$$

where the relation $\div$ stands for the traditional crisp division.

A lot of effort has been devoted to finding a tight interval enclosure to $OLS(\boldsymbol{X}, \boldsymbol{y})$, that is, as small as possible box $\boldsymbol{\beta} = [\underline{\beta}, \overline{\beta}]$ satisfying $\boldsymbol{\beta} \supseteq OLS(\boldsymbol{X}, \boldsymbol{y})$. Recall that the family of linear systems describing $OLS(\boldsymbol{X}, \boldsymbol{y})$ is of the form

$$X^T X\beta = X^T y, \quad X \in \boldsymbol{X},\ y \in \boldsymbol{y}. \tag{13}$$

6.1 Relaxation

The basic approach to find an interval enclosure is by relaxation. Consider an interval system of equations

$$\boldsymbol{Z}\beta = \boldsymbol{z}, \tag{14}$$

where the results of products $\boldsymbol{Z} := \boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{z} := \boldsymbol{X}^T\boldsymbol{y}$ are calculated using interval arithmetic. The interval system (14) covers all equation systems from (13) and typically some more. This redundancy, which is called *overestimation* in interval algebra, is caused by the so called *dependency problem*, which occurs when an interval quantity appears several times in the description of the interval system (in our case it is $\boldsymbol{X}$). Dependencies are hard to deal with, and often they are relaxed. Any enclosure of the solution set of (14) serves as an enclosure of $OLS(\boldsymbol{X}, \boldsymbol{y})$, too.

It is known that determining the interval hull (i.e., the optimal enclosure, or, exact bounds of the solution set) is an ***NP***-hard problem. However, there are many methods for calculating quite sharp bounds in short time [47], [58], [60]. Some of them employ the basic form (14), however, often a preconditioning is used to improve the performance. Preconditioning means premultiplication of the system (14) by a crisp matrix; usually an inverse of Z^c (or its numerically computed approximation) is used, where Z^c is the center matrix of $\boldsymbol{Z}$. Even though it leads to further overestimation of the solution set, surprisingly the resulting enclosures are usually tighter.

First we mention three direct methods and then two iterative ones. Further, more general methods capable of dealing with dependencies are introduced, and finally other possible approaches are commented.

6.2 Interval Gaussian elimination

Implementing Gaussian elimination with interval arithmetic leads to the Interval Gaussian elimination. It works well in special cases (e.g. when the interval matrix is diagonally dominant or M-matrix), but it almost always causes high overestimation.

6.3 Hansen–Bliek–Rohn's bounds

One of the best direct methods is the Hansen–Bliek–Rohn's formula [48], [58] which works as follows. Precondition the system (14) by the inverse of Z^c and

denote the resulting interval system by $\boldsymbol{A}\beta = \boldsymbol{r}$. Suppose that $\underline{A}^{-1} \geq 0$ and denote $\beta^* := \underline{A}^{-1}(|r^c| + r^\Delta)$. Then the enclosing box $\boldsymbol{\beta} = [\underline{\beta}, \overline{\beta}]$ is of the form

$$\overline{\beta_i} := \max\left\{\beta_i^* + (\underline{A}^{-1})_{ii}(r^c - |r^c|)_i,\ \frac{1}{2(\underline{A}^{-1})_{ii} - 1}\big(\beta_i^* + (\underline{A}^{-1})_{ii}(r^c - |r^c|)_i\big)\right\},$$

$$\underline{\beta_i} := \min\big\{-\beta_i^* + (\underline{A}^{-1})_{ii}(r^c + |r^c|)_i,$$

$$\frac{1}{2(\underline{A}^{-1})_{ii} - 1}\big(-\beta_i^* + (\underline{A}^{-1})_{ii}(r^c + |r^c|)_i\big)\right\}.$$

Notice that the assumption $\underline{A}^{-1} \geq 0$ is not very restrictive. Indeed, it is frequently used as a sufficient condition for regularity of interval matrices (or its equivalent form $\rho(A^\Delta) < 1$, where $\rho(\cdot)$ stands for the spectral radius). It was proved that the Hansen–Bliek–Rohn method calculates exact bounds for the solution set of the preconditioned system $\boldsymbol{A}\beta = \boldsymbol{r}$, and so it gives a very tight enclosure to (14).

6.4 Jansson's algorithm

There is an algorithm that computes the optimal bounds. The solution set of (14) represents a finite union of convex polyhedra. When restricting on any orthant, is becomes a convex polyhedron the bounds of which can be exactly determined by solving $2p$ linear programs. Thus, the interval hull is computable by inspecting all of the 2^p orthants. This tremendeous number can be reduced by the method of Jansson [32]. The method is based on the observation that the solution set is compact and connected provided $\boldsymbol{Z}$ is regular, and it is unbounded (moreover, each topologically connected component is unbounded) in case $\boldsymbol{Z}$ is not regular. Thus, the Jansson's algorithm concentrates on just one connected component and goes through all orthants it intersects. Even though it may be exponential in the worst case, it often inspects only a fraction of all orthants.

6.5 Krawczyk's method

A popular and efficient Krawczyk's method [44], [47] is an iterative method based on the preconditioning of (14) by a point matrix Y, usually $Y \approx (Z^c)^{-1}$. Let an initial enclosure $\boldsymbol{\beta}^0$ be given. Then we iteratively calculate a nested sequence of enclosures $\boldsymbol{\beta}^k$, $k = 0, 1, \ldots$, where the iteration step is

$$\begin{array}{ll}\{1\} & \boldsymbol{\beta}' := Y\boldsymbol{z} + (I - Y\boldsymbol{Z})\boldsymbol{\beta^k} \\ \{2\} & \boldsymbol{\beta}^{k+1} := \boldsymbol{\beta}' \cap \boldsymbol{\beta}^k.\end{array}$$

Convergence and the approximation order is discussed deeply in [47].

6.6 Interval Gauss–Seidel's iteration

It is a straightforward extension of the Gauss–Seidel's iteration method for intervals [44], [47]. Once an initial enclosure $\boldsymbol{\beta}^0$ to the solution set is known, a nested sequence of enclosures $\boldsymbol{\beta}^k$, $k = 0, 1, \ldots$ is calculated. One iteration is

$\{1\}$ **for** $i = 1, \ldots, p$ **do**
$\{2\}$ $\quad \boldsymbol{\beta}'_i := \left(\boldsymbol{z}_i - \sum_{j<i} \boldsymbol{Z}_{ij} \boldsymbol{\beta}_j^{k+1} - \sum_{j>i} \boldsymbol{Z}_{ij} \boldsymbol{\beta}_j^{k} \right) / \boldsymbol{Z}_{ii}$
$\{3\}$ $\quad \boldsymbol{\beta}_i^{k+1} := \boldsymbol{\beta}'_i \cap \boldsymbol{\beta}_i^{k}$
$\{4\}$ **end do.**

For special matrices (e.g. M-matrices), convergence to the interval hull is guaranteed. Applied to a preconditioned system, Gauss–Seidel's iteration yields tighter intervals than Krawczyk's method.

6.7 Parametric approach

Due to the relaxation of (13) to (14), the resulting enclosure $\boldsymbol{\beta}$ is overestimated. To reduce the overestimation, we can consider (13) to be a nonlinear parametric interval system of equations. That is, each matrix and right-hand side entry is considered as a nonlinear function of interval parameters. Several methods for such systems are available [54], [36], however, they usage for high-dimensional problems (with respect to n and/or p) is questionable.

Since the constraint matrix in (13) is symmetric positive definite, we can consider a partial relaxation, too. That is, we relax the correlations between the interval parameters and keep only the symmetry and positive definiteness. For such interval systems, various approaches are known. The interval Cholesky's method [2] naturally extends Cholesky's method for interval data. Unfortunately, its efficiency is quite low, despite some pivot tightening improvements [14]. Since the symmetry condition is linear, we can utilize any solver for (more general) linear parametric interval systems. As numerical results in [23] indicate, this approach leads to quite tight bounds.

6.8 Software

Interval arithmetic is implemented in many programming languages. INTLAB [62] is a powerful MATLAB toolbox comprising not only the basic interval arithmetic, but also some useful interval functions. For instance, `verifylss` is a Krawczyk iteration-based function for interval linear systems of equations. VERSOFT [61] is a collection of verification and interval software written in INTLAB / MATLAB. For interval linear systems, it contains e.g. `verenclinthull`, a function for computing the Hansen–Bliek–Rohn's bounds, and `verintervalhull` for calculating the interval hull. The parametric solver by Popova [52] can handle interval systems with dependencies; the readers can experiment using the free webComputing service [53].

6.9 Other approaches

Another way to solve the problem is to rewrite it in the form

$$\begin{pmatrix} 0_p & \boldsymbol{X}^T \\ \boldsymbol{X} & I_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} 0 \\ \boldsymbol{y} \end{pmatrix}. \tag{15}$$

Again, a relaxation leads to an ordinary interval linear system of equations, and we can employ the above mentioned solvers. Because the dependence structure in (15) is simpler than in (14) (it is just a condition on symmetry), it is easily shown that the resulting interval box $\boldsymbol{\beta}$ will be tighter.

Proposition 20. *The solution set of* (15) *is contained it the solution set of* (14).

Proof. Let us consider any instance of the system (15)

$$\begin{pmatrix} 0_p & X^T \\ X' & I_n \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix},$$

where $X, X' \in \boldsymbol{X}$ and $y \in \boldsymbol{y}$. Then $\gamma = y - X'\beta$, and by substituting $X^T(y - X'\beta) = 0$. Hence $X^T X'\beta = X^T y$. Since $X^T X' \in \boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{Z}$ and $X^T y \in \boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{z}$, we have an instance of the system (14). □

The drawback of this approach is that the system may ve very large provided the number of observations is large. So it may be time-consuming to calculate sharp enough enclosures.

The constraint matrix in (15) is symmetric, so we can again utilize any method for symmetric interval systems to obtain more accurate enclosures.

An alternative approach to solve the problem is based on the adaptation of the QR factorization for interval data [15], e.g. using interval Householder's method [5], [42]. Another method is described in [43].

6.10 Examples

Example 21. *Let*

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ [0,2] & 2 & 3 & 4 & 5 \end{pmatrix}^{\mathrm{T}},$$
$$\boldsymbol{y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}^{\mathrm{T}}.$$

Obviously, the solution set consists of one point $(1, 0)$. The results of different approaches are displayed in Table 1 and illustrated in Figure 6.

In Figures 6, 7 and 8, the dashed rectangles correspond to the enclosures resulting from the system (14), whereas the solid rectangles correspond to (15). Inside the rectangles, the true solution set (computed numerically) is plotted.

Example 21 — a modified case. Now let us change the interval vector $\boldsymbol{y}$ to

$$\boldsymbol{y}' = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \end{pmatrix}^{\mathrm{T}}.$$

The results are displayed in Table 2 and in Figure 7.

Table 1: Different enclosures in Example 21.

	via system (14)	via system (15)
`verifylss`	$([-1.2858, 3.2858], [-0.7143, 0.7143])$	(1, 0)
`verenclinthull`	$([0.3043, 3.2858], [-0.7143, 0.7143])$	(1, 0)
parametric method [23]	$([0.3043, 3.2858], [-0.7143, 0.7143])$	(1, 0)
`verintervalhull`	$([0.3043, 3.2858], [-0.7143, 0.2174])$	(1, 0)

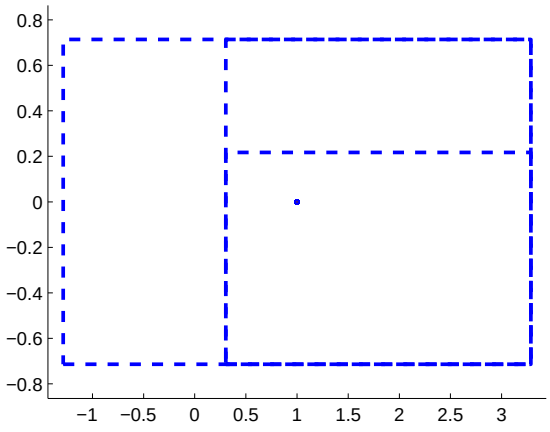

Figure 6: Example 21 (Table 1).

Example 22. *Let $\boldsymbol{X}, \boldsymbol{y}$ be given with*

$$\underline{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -2.5 & -1.5 & -0.5 & 0.5 & 1.5 & 2.5 & 3.5 & 4.5 & 5.5 & 6.5 & 7.5 \end{pmatrix}^{\mathrm{T}},$$

$$\overline{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1.5 & -0.5 & 0.5 & 1.5 & 2.5 & 3.5 & 4.5 & 5.5 & 6.5 & 7.5 & 8.5 \end{pmatrix}^{\mathrm{T}},$$

$$\underline{y} = \begin{pmatrix} 1.5 & 2.5 & 3.5 & 4.5 & 5.5 & 6.5 & 7.5 & 8.5 & 9.5 & 10.5 & 11.5 \end{pmatrix}^{\mathrm{T}},$$

$$\overline{y} = \begin{pmatrix} 2.5 & 3.5 & 4.5 & 5.5 & 6.5 & 7.5 & 8.5 & 9.5 & 10.5 & 11.5 & 12.5 \end{pmatrix}^{\mathrm{T}}.$$

The calculated enclosures are shown in Table 3 and illustrated in Figure 8.

Example 23. *Consider an example from [5]:*

$$\boldsymbol{X} = \begin{pmatrix} [0.1, 0.3] & [0.9, 1.1] \\ [8.9, 9.1] & [0.4, 0.6] \\ [0.9, 1.1] & [6.9, 7.1] \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} [0.8, 1.2] \\ [-0.2, 0.2] \\ [1.8, 2.2] \end{pmatrix}.$$

Table 2: Different enclosures in Example 21, the modified case.

	via system (14)
`verifylss`	$([-8.0520, 10.5975], [-1.9481, 3.7663])$
`verenclinthull`	$([-5.1429, 10.5975], [0.3030, 3.7663])$
parametric method [23]	$([-5.1429, 7.6884], [0.3030, 2.8572])$
`verintervalhull`	$([-5.1429, 3.1516], [0.3030, 2.8572])$
	via system (15)
`verifylss`	$([-0.2667, 2.2667], [0.6666, 1.3334])$
`verenclinthull`	$([-0.2667, 2.2667], [0.6666, 1.3334])$
parametric method [23]	$([-0.2667, 2.2667], [0.7999, 1.3334])$
`verintervalhull`	$([-0.1740, 1.7298], [0.8108, 1.3044])$

Table 3: Different enclosures in Example 22.

	via system (14)
`verifylss`	$([-11.9723, 20.0256], [-3.2903, 5.2725])$
`verenclinthull`	$([-7.5467, 19.7341], [-0.1201, 5.1945])$
parametric method [23]	$([-7.5467, 16.2139], [-0.1201, 4.2770])$
`verintervalhull`	$([-7.5467, 7.9201], [-0.1201, 4.0134])$
	via system (15)
`verifylss`	$([2.1858, 5.8142], [0.5234, 1.4766])$
`verenclinthull`	$([2.1858, 5.8142], [0.5234, 1.4766])$
parametric method [23]	$([2.2206, 5.7794], [0.6820, 1.4661])$
`verintervalhull`	$([-\infty, \infty], [-\infty, \infty])$

Table 4 displays enclosures calculated by various methods; in the last line, there is the enclosure from [5] computed by the interval Householder's method.

7 Conclusion

We have shown several properties of the set $OLS(X, \boldsymbol{y})$ in the crisp input – interval output model. We expect that some results could be improved. In particular, it might be possible to improve the Theorem 19 as discussed in Section 4.4. We also expect that further applications of the RRR metaalgorithm of Section 5 could be found, in particular in polyhedral geometry.

We also think that methods of Sections 4 and 5 are suitable for implementation in software for analysis of interval data. In particular, an implementation of a visualization tool, demonstrated in Figure 3, could be useful for users of interval data.

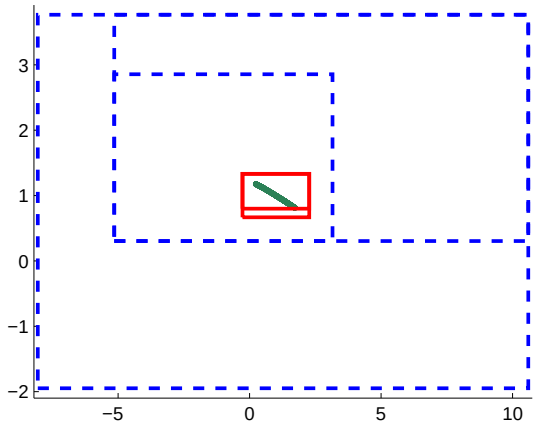

Figure 7: Example 21, the modified case (Table 2).

The Examples of Section 6, dealing with the general model, might seem disappointing. Even in very simple cases the enclosures computed by known methods are often so wide that they can hardly be useful in practical analysis of interval regression models. The main drawback is that the methods lean on relaxation as described in Section 6.1. We expect that the special form of dependence, either in the system (14) or in the system (15), could be further analyzed and utilized for improvement of the known enclosure methods. Moreover, not only interval enclosures, but also other types of enclosures (such as ellipsoidal enclosures) could be successful.

Acknowledgment

The work of the first author was supported by Grant No. F4/18/2011 of Internal Grant Agency of University of Economics, Prague. The work of the second author was supported by Grant No. 201/09/0755 of Czech Science Foundation.

References

[1] G. Alefeld, J. Herzberger, Introduction to Interval Computations, Computer Science and Applied Mathematics, Academic Press, New York, USA, 1983.

[2] G. Alefeld, G. Mayer, New criteria for the feasibility of the Cholesky method with interval data, SIAM Journal on Matrix Analysis and Applications 30 (2008) 1392–1405.

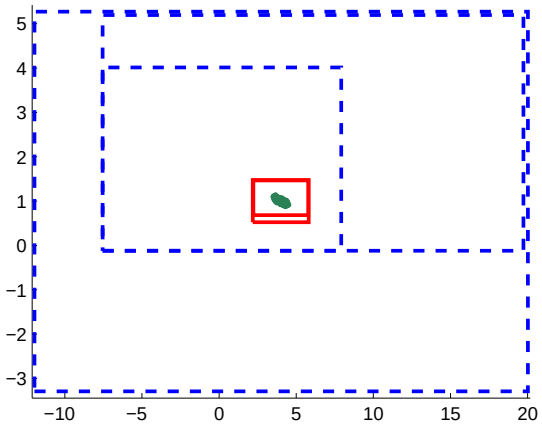

Figure 8: Example 22.

[3] S. Arora, B. Barak, Computational Complexity: A Modern Approach, Cambridge University Press, UK, 2009.

[4] D. Avis, K. Fukuda, Reverse search for enumeration, Discrete Applied Mathematics 65 (1996) 21–46.

[5] A. H. Bentbib, Solving the full rank interval least squares problem, Applied Numerical Mathematics 41 (2002) 283–294.

[6] M. Černý, Goffin's Algorithm for Zonotopes, Technical Report, University of Economics, Prague, 2011; http://nb.vse.cz/∼cernym/tr

[7] M. Černý, M. Hladík, The regression tolerance quotient in data analysis, In: M. Houda, J. Friebelová (Eds.), Procceding of Mathematical Methods in Economics 2010, University of South Bohemia, Czech Republic, 2010, pp. 98–104.

[8] M. Černý, M. Rada, A note on linear regression with interval data and linear programming, In: *Quantitative Methods in Economics: Multiple Criteria Decision Making XV,* Kluwer, Iura Edition, Slovakia, 2010, pp. 276–282.

[9] P.-T. Chang, E. S. Lee, S. A. Konz, Applying fuzzy linear regression to VDT legibility, Fuzzy Sets and Systems 80, 1996, 197–204.

[10] C. Chuang, Extended support vector interval regression networks for interval input-output data, Information Sciences 178 (2008) 871–891.

Table 4: Different enclosures in Example 23.

	via system (14)
`verifylss`	$([-0.0644, 0.0324], [0.2364, 0.3693])$
`verenclinthull`	$([-0.0643, 0.0312], [0.2418, 0.3692])$
parametric method [23]	$([-0.0643, 0.0312], [0.2418, 0.3692])$
`verintervalhull`	$([-0.0643, 0.0276], [0.2432, 0.3692])$
	via system (15)
`verifylss`	$([-0.0471, 0.0145], [0.2569, 0.3477])$
`verenclinthull`	$([-0.0470, 0.0139], [0.2587, 0.3475])$
parametric method [23]	$([-0.0469, 0.0139], [0.2587, 0.3462])$
`verintervalhull`	$([-0.0468, 0.0127], [0.2606, 0.3473])$
interval Householder method [5]	$([-0.0558, 0.0232], [0.2579, 0.3485])$

[11] J. P. Dunyak, D. Wunsch, Fuzzy regression by fuzzy number neural networks, Fuzzy Sets and Systems 112 (2000) 371–380.

[12] T. Entani, M. Inuiguchi, Group decisions in interval AHP based on interval regression analysis, In: V.-N. Huynh, Y. Nakamori, Lawry J., M. Inuiguchi (Eds.), Integrated Uncertainty Management and Applications, Advances in Soft Computing 68, Springer, Germany, 2010, pp. 269–280.

[13] J.-A. Ferrez, K. Fukuda, T. Liebling, Solving the fixed rank convex quadratic maximization in binary variables by a parallel zonotope construction algorithm, European Journal of Operational Research 166 (2005) 35–50.

[14] J. Garloff, Pivot tightening for the interval Cholesky method, Proceedings in Applied Mathematics and Mechanics (PAMM) 10 (2010) 549–550.

[15] D. M. Gay, Interval least squares—a diagnostic tool, In R. E. Moore (Ed.), Reliability in Computing: The Role of Interval Methods in Scientific Computing, Perspectives in Computing 19, Academic Press, Boston, USA, 1988, pp. 183–205.

[16] J. L. Goffin, Variable metric relaxation methods. Part II: The ellipsoid method, Mathematical Programming 30 (1984) 147–162.

[17] M. Grötschel, L. Lovász, A. Schrijver, Geometric Algorithms and Combinatorial Optimization, Springer, Germany, 1993.

[18] L. J. Guibas, A. Nguyen, L. Zhang, Zonotopes as Bounding Volumes, Proceeding SODA '03 Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Pennsylvania, 2003.

[19] P. Guo, H. Tanaka, Dual models for possibilistic regression analysis, Computational Statistics and Data Analysis 51 (2006) 253–266.

[20] P. Hájek, P. Pudlák, Metamathematics of First Order Arithmetic, Perspectives in Mathematical Logic, Springer, Germany, 1993.

[21] B. Hesmaty, A. Kandel, Fuzzy linear regression and its applications to forecasting in uncertain environment, Fuzzy Sets and Systems 15 (1985) 159–191.

[22] M. Hladík, Description of symmetric and skew-symmetric solution set, SIAM Journal on Matrix Analysis and Applications 30 (2008) 509–521.

[23] M. Hladík, Enclosures for the solution set of parametric interval linear systems, KAM-DIMATIA Series 983, Department of Applied Mathematics, Charles University, Prague, 2010, http://kam.mff.cuni.cz/∼kamserie/serie/clanky/2010/s983.ps.

[24] M. Hladík, Solution set characterization of linear interval systems with a specific dependence structure, Reliable Computing 13 (2007) 361–374.

[25] M. Hladík, Solution sets of complex linear interval systems of equations, Reliable Computing 14 (2010) 78–87.

[26] M. Hladík, M. Černý, Interval regression by tolerance analysis approach, to appear in Fuzzy Sets and Systems, 2011, DOI: 10.1016/j.fss.2011.07.009.

[27] M. Hladík, M. Černý, New approach to interval linear regression, In: R. Kasımbeyli, C. Dinçer, S. Özpeynirci, L. Sakalauskas (Eds.), 24th Mini-EURO Conference on Continuous Optimization and Information-Based Technologies in the Financial Sector MEC EurOPT 2010, Selected papers, Technika, Vilnius, Lithuania, 2010, pp. 167–171.

[28] C.-H. Huang, H.-Y. Kao, Interval regression analysis with soft-margin reduced support vector machine, Lecture Notes in Computer Science 5579, Springer, Germany, 2009, pp. 826–835.

[29] M. Inuiguchi, H. Fujita, T. Tanino, Robust interval regression analysis based on Minkowski difference, In: SICE 2002: Proceedings of the 41st SICE Annual Conference, vol. 4, Osaka, Japan, 2002, pp. 2346–2351.

[30] H. Ishibuchi, H. Tanaka, Several formulations of interval regression analysis, In: Proceedings of Sino-Japan Joint Meeting on Fuzzy Sets and Systems B2-2, Beijing, China, 1990, pp. 1–4.

[31] H. Ishibuchi, H. Tanaka, H. Okada, An architecture of neural networks with interval weights and its application to fuzzy regression analysis, Fuzzy Sets and Systems 57 (1993) 27–39.

[32] C. Jansson, Calculation of exact bounds for the solution set of linear interval systems, Linear Algebra and its Applications 251 (1997) 321–340.

[33] G. Jun-Peng, L. Wen-Hua, Regression analysis of interval data based on error theory, In: Proceedings of 2008 IEEE International Conference on Networking, Sensing and Control ICNSC, Sanya, China, 2008, pp. 552–555.

[34] M. Kaneyoshi, H. Tanaka, M. Kamei, H. Furuta, New system identification technique using fuzzy regression analysis, In: Proceedings of the First International Symposium on Uncertainty Modeling and Analysis, Baltimore, USA, 1990, pp. 528–533.

[35] H. Kashima, K. Yamasaki, A. Inokuchi, H. Saigo, Regression with interval output values, In: 19th International Conference on Pattern Recognition ICPR 2008, Tampa, USA, 2008, pp. 1–4.

[36] L. V. Kolev, Solving linear systems whose elements are nonlinear functions of intervals, Numerical Algorithms 37 (2004) 199–212.

[37] H. Lee, H. Tanaka, Fuzzy regression analysis by quadratic programming reflecting central tendency, Behaviormetrika 25 (1998) 65–80.

[38] H. Lee, H. Tanaka, Upper and lower approximation models in interval regression using regression quantile techniques, Europeran Journal of Operational Research 116 (1999) 653–666.

[39] B. Li, C. Li, J. Si, G. Abousleman, Interval least-squares filtering with applications to robust video target tracking, In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing — Proceedings, IEEE Signal Processing Society, Las Vegas, USA, 2008, pp. 3397–3400.

[40] E. de A. Lima Neto, F. de A. T. de Carvalho, Constrained linear regression models for symbolic interval-valued variables, Computational Statistics and Data Analysis 54 (2010) 333–347.

[41] P. Liu, Study on a speech learning approach based on interval support vector regression, In: Proceedings of 4th International Conference on Computer Science and Education, Nanning, China, 2009, pp. 1009–1012.

[42] T. A. Manteuffel, An interval analysis approach to rank determination in linear least squares problems, SIAM Journal on Scientific and Statistical Computing 2 (1981) 335–348.

[43] S. M. Markov, Least squares approximations under interval input data, In: C. Ullrich (Ed.), Contributions to Computer Arithmetic and Self-Validating Numerical Methods, IMACS Annals on Computing and Applied Mathematics 7, J. C. Baltzer Scientific Publishing, Basel, 1990, pp. 133–147.

[44] R. E. Moore, R. B. Kearfott, M. J. Cloud, Introduction to Interval Analysis, SIAM, Philadelphia, USA, 2009.

[45] I. Moral-Arce, J. M. Rodríguez-Póo, S. Sperlich, Low dimensional semiparametric estimation in a censored regression model, Journal of Multivariate Analysis 102, 118–129.

[46] E. Nasrabadi, S. Hashemi, Robust fuzzy regression analysis using neural networks, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 16 (2008) 579–598.

[47] A. Neumaier, Interval Methods for Systems of Equations, Cambridge University Press, UK, 1990.

[48] S. Ning, R. B. Kearfott, A comparison of some methods for solving linear interval equations, SIAM Journal on Numerical Analysis 34 (1997) 1289–1305.

[49] P. Odifreddi, Classical Recursion Theory, Elsevier, Amsterdam, Netherlands, 1992.

[50] W. Pan, R. Chappell, Computation of the NPMLE of distribution functions for interval censored and truncated data with applications to the Cox model, Computational Statistics and Data Analysis 28 (1998) 33–50.

[51] C. Papadimitriou, Computational Complexity, Addison-Wesley Longman, 1995.

[52] E. D. Popova, Parametric interval linear solver, Numerical Algorithms 37 (2004) 345–356.

[53] E. D. Popova, Webcomputing service framework, International Journal on Information Theories and Applications 13 (2006) 246–254.

[54] E. D. Popova, Solving linear systems whose input data are rational functions of interval parameters, Lecture Notes in Computer Science 4310, Springer, Germany, 2007, pp. 345–352.

[55] J. Renegar, On the computational complexity and geometry of the first-order theory of the reals. Part I: Introduction. Preliminaries. The geometry of semi-algebraic sets. The decision problem for the existential theory of the reals, Journal of Symbolic Computation 13 (1992) 255–299.

[56] J. Renegar, On the computational complexity and geometry of the first-order theory of the reals. Part II: The general decision problem. Preliminaries for quantifier elimination, Journal of Symbolic Computation 13 (1992) 301–327.

[57] J. Renegar, On the computational complexity and geometry of the first-order theory of the reals. Part III: Quantifier elimination, Journal of Symbolic Computation 13 (1992) 329–352.

[58] J. Rohn, A Handbook of Results on Interval Linear Problems, Czech Academy of Sciences, Prague, Czech Republic, 2005, http://uivtx.cs.cas.cz/∼rohn/handbook/handbook.zip.

[59] J. Rohn, Enclosing solutions of overdetermined systems of linear interval equations, Reliable Computing 2 (1996) 167–171.

[60] J. Rohn, Systems of linear interval equations, Linear Algebra and its Applications 126C (1989) 39–78.

[61] J. Rohn, VERSOFT: Verification software in MATLAB / INTLAB, version 10, 2009. http://uivtx.cs.cas.cz/∼rohn/matlab/.

[62] S. M. Rump, INTLAB — INTerval LABoratory, In: T. Csendes (Ed.) Developments in Reliable Computing, Kluwer Academic Publishers, Dordrecht, 1999, pp. 77–104, http://www.ti3. tu-harburg.de/rump/.

[63] H. Schneeweiss, J. Komlos, Probabilistic Rounding and Sheppard's Correction, Technical Report 45, Department of Statistics, University of Munich, 2008, http://epub.ub.uni-muenchen.de/8661/1/tr045.pdf.

[64] S. Schön, H. Kutterer, Using zonotopes for overestimation-free interval least-squares — some geodetic applications, Reliable Computing 11 (2005), 137–155.

[65] A. Schrijver, Theory of Linear and Integer Programming, Wiley, USA, 2000.

[66] K. Sugihara, H. Ishii, H. Tanaka, Interval priorities in AHP by interval regression analysis, Europeran Journal of Operational Research 158 (2004) 745–754.

[67] H. Tanaka, H. Lee, Fuzzy linear regression combining central tendency and possibilistic properties, In: Proceedings of the Sixth IEEE International Conference on Fuzzy Systems 1, Barcelona, Spain, 1997, pp. 63–68.

[68] H. Tanaka, H. Lee, Interval regression analysis by quadratic programming approach, IEEE Transactions on Fuzzy Systems 6 (1998) 473–481.

[69] H. Tanaka, J. Watada, Possibilistic linear systems and their application to the linear regression model, Fuzzy Sets and Systems 27 (1988) 275–289.

[70] A. Tarski, A Decision Method for Elementary Algebra and Geometry, University of California Press, Berkeley, USA, 1951.

[71] X. Zhang, J. Sun, Regression analysis of clustered interval-censored failure time data with informative cluster size, Computational Statistics and Data Analysis 54 (2010) 1817–1823.

[72] G. Ziegler, Lectures on Polytopes, Springer, Germany, 2004.