Maximum Consistency Method for Data Fitting under Interval Uncertainty*

Sergey P. Shary

Institute of Computational Technologies,
Novosibirsk, Russia

shary@ict.nsc.ru

Abstract

The work is devoted to application of global optimization in data fitting problem under interval uncertainty. Parameters of the linear function that best fits intervally defined data are taken as the maximum point for a special ("recognizing") functional which is shown to characterize consistency between the data and parameters. The new data fitting technique is therefore called "maximum consistency method". We investigate properties of the recognizing functional and present interpretation of the parameter estimates produced by the maximum consistency method.

Keywords: data fitting, interval uncertatinty, recognizing functional, maximum consistency, nonconvex optimization

Mathematics Subject Classification 2010: 62H99, 93B30, 90C26

1 Introduction

The subject matter of this paper is application of the global optimization methods. We consider the problem of data fitting: given empirical data, it is required to construct a functional relationship, of a fixed form, between "input" and "output" quantities that best fits the data in a prescribed sense.

More precisely, let us suppose that the quantity b is a linear function of a_1, a_2, …, a_n, i. e. such that

$$b = a_1x_1 + a_2x_2 + \ldots + a_nx_n \tag{1}$$

with some coefficients x_i, $i = 1,2,\ldots,n$. They are not known, and we have to determine them from given sets of values of both the function and its arguments. There are m such sets that we denote as

$$\begin{array}{ccccc} a_{11}, & a_{12}, & \ldots, & a_{1n}, & b_1, \\ a_{21}, & a_{22}, & \ldots, & a_{2n}, & b_2, \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1}, & a_{m1}, & \ldots, & a_{mn}, & b_m, \end{array} \tag{2}$$

*Published in *Journal of Global Optimization*, vol. 66 (2016), No. 1, pp. 111–126. DOI: 10.1007/s10898-015-0340-1.

where the first of the lower indices shows the number of the data set, or, in other words, the "observation number". We have to determine the values of x_i, $i = 1, 2, \ldots, n$, for which the function (1) approximates the data (2) with "the best accuracy".

Substituting the values from (2) to the equality (1), we arrive at the system of equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \quad\vdots \qquad\quad \vdots \qquad\quad \ddots \qquad\quad \vdots \qquad\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \end{cases} \tag{3}$$

that the sought-for coefficients x_i, $i = 1, 2, \ldots, n$, should comply with. The system (3) can be written briefly as

$$Ax = b,$$

where $A = (a_{ij})$ is an $m \times n$-matrix and $b = (b_i)$ is an m-vector made up of the values (2).

A solution to the system of linear algebraic equations (3), either common or in a generalized sense, is usually taken as an estimate of the parameters x_0, x_1, …, x_n. But the practical application of the above scheme inevitably encounters a number of difficulties that crucially complicates both the initial problem statement and its solution.

First of all, the number of observations almost always is not the same as the number of unknown parameters x_i that we have to determine. Very often, they try to get as much observations as possible, since every one of them provides us with information on the function (1). This is why the system of equations (3) is usually overdetermined. However, there exist situations when the number of observations is less than the number of unknown parameters, so that the system (3) is underdetermined.

But the main feature of the problem under study is that the measuments data (2) are always imprecise, and the system of equations (3) should be solved for inexact a_{ij} and b_i. Further analysis and our *modus operandi* in the solution of the data fitting problem depend upon the way that we describe the data imprecision (uncertainty). If we adhere to the model of errors based on the classical probability theory, this results in the traditional regression analysis (see e. g. [3, 14]). The main data fitting technique is then the least squares method and its modifications. In what follows, we are going to adopt only the terminology of the regression analysis, since it provides the most elaborated system of concepts and notions that are necessary for data fitting.

In the present paper, we consider interval model of the measurement errors when uncertainty and imprecision in the data are described by intervals of their possible values, i. e. we have at our disposal only lower and upper bounds for possible values of the measured quantities. In other words, it is supposed that the memeberships of a_{ij} and b_i in some intervals are known:

$$a_{ij} \in \boldsymbol{a}_{ij} \qquad \text{and} \qquad b_i \in \boldsymbol{b}_i, \tag{4}$$

where boldface letters denote intervals according to the informal notation standard [7].

The pioneering work on data processing under interval uncertainty was published in 1962 [6], authored by Leonid Kantorovich, inventor of linear programming. It seems like he anticipated a long-standing and mature necessity in this kind of technique, since analogous approaches had been repeatedly reinvented since the mid-60th. The first Western paper on data processing under bounded (interval) uncertainty was written by Schweppe [16], and a detailed survey of the modern state of this field can be found in the handbook [10].

2 Problem statement

What is a "solution" to the data fitting problem under interval uncertainty (4)? The following natural definition is generally accepted: *The set of parameters $x_1, x_2, \ldots, x_n$ of the linear function* (1) *is said to be consistent with the interval experimental data* $(\boldsymbol{a}_{i1}, \boldsymbol{a}_{i2}, \ldots, \boldsymbol{a}_{in}, \boldsymbol{b}_i)$, $i = 1, 2, \ldots, m$, *if, for every observation i, the measured intervals contain such representatives* $a_{i1} \in \boldsymbol{a}_{i1}$, $a_{i2} \in \boldsymbol{a}_{i2}$, $\ldots$, $a_{in} \in \boldsymbol{a}_{in}$ *and* $b_i \in \boldsymbol{b}_i$ *that*

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i.$$

The above definition is a straightforward generalization of the solution to data fitting problem for usual non-interval case, and Fig. 1 shows its main idea visually. The regression line that corresponds to the set of parameters intersects all the uncertainty boxes in the space $\mathbb{R}^{n+1}$ where the data (2) lie.

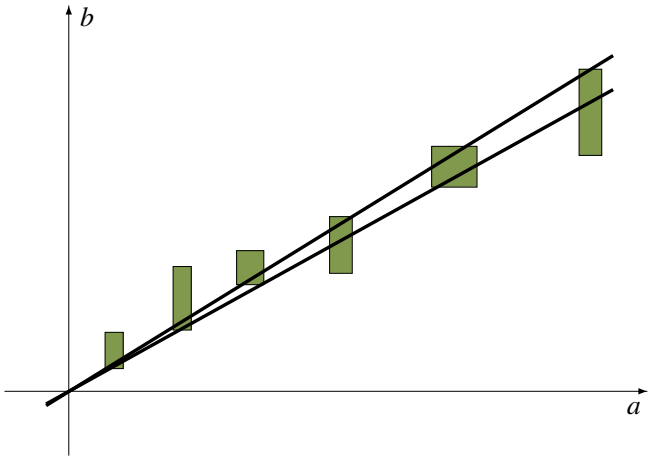

Figure 1: An illustration of consistency between data and parameters of a linear function.

Similar to the usual non-interval case, sometimes there does not exist a set of parameters consistent with the data (in the sense of the above definition). On the other hand, the data fitting problem under interval uncertainty has its own features that substantially distinguish it from the common data fitting problem for non-interval data. In the non-interval case, existence of the regression line that goes through all the data points, i.ė. existence of the parameters that exactly satisfy the system of equations (3) is, generally speaking, an delicate event that can be destroyed after arbitrarily small perturbation in the data. But for essential interval uncertainty, when the widths of the intervals are strictly greater than zero, the set of parameters consistent with the data has, as a rule, nonzero measure, and it is stable under small data perturbations. This will be clear from the further mathematical considerations (see Sections 2–3).

The fact that there does not exist a regression line going through all the uncertainty boxes, i. e. there is no parameter sets consistent with the data, in no way means that the data fitting problem is unsolvable. It only means that the consistency condition cannot be exactly satisfied by a solution to the problem. It makes sense to recall that such a situation is quite typical for the degenerate non-interval case, when we consider the traditional data fitting problem. Then we do not speak of "unsolvability" of the problem if the system of equations (3) does not have exact solutions.

Using formal mathematical language, the set of parameters consistent with the interval experimental data can be specified as

$$\Big\{ x \in \mathbb{R}^n \;\Big|\; \big(\, \exists (a_{ij}) \in (\boldsymbol{a}_{ij}) \,\big) \big(\, \exists (b_i) \in (\boldsymbol{b}_i) \,\big) \big(Ax = b \,\big) \Big\}, \tag{5}$$

where the elements a_{ij} constitute an $m \times n$-matrix $A = (a_{ij})$, while the components b_i form a vector $b = (b_i)$. In identification theory and data analysis, this set is called *parameter uncertainty set*, *set of possible values of the parameters*, *information set* [8], and so on. In interval analysis, it is a *solution set* to the interval system of linear algebraic equations $\boldsymbol{A}x = \boldsymbol{b}$, where $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ and $\boldsymbol{b} = (\boldsymbol{b}_i)$ (see [11, 12, 18]).

We are going to perform data fitting under imprecision and uncertainty according to the following general theoretical scheme:

1) a "consistency measure" is introduced between data and parameters that characterizes the accuracy to which the data are approximated by the function defined by the parameters, in other words, a "quality of data approximation";

2) as an estimate of the parameters, we take the parameter values for which the "consistency measure" attains its maximum, that is, where the maximal consistency between data and parameters is reached.

For instance, in the traditional data fitting problem for non-interval data the distance from the regression line to the data points is to be minimized (in a certain metric). Then the "consistency measure" is, in fact, this distance with the minus sign.

When solving the data fitting problem under imprecision and uncertainty, the main question is: what "consistency measure" should be used? There is a number of natural requirements that such a measure must fulfill. First of all, the "measure" should trace the quantitative difference between the cases when the information set (5) is empty or not. This can be achieved, e. g., by assigning different signs to the "measure", ether negative or positive.

For a nonempty solution set, the consistency measure should be positive (or, at least, non-negative) for the points from the solution set, where the "consistency" really takes place. For the points outside the solution set, where no "consitency" occurs, our measure should be negative.

3 Interval linear systems of equations

In this section, we consider the main object that arises in the data fitting problem under interval uncertainty when the constructed function is supposed to be linear, i. e., has the form (1). This is an interval linear algebraic system of the form

$$\begin{cases} \boldsymbol{a}_{11}x_1 + \boldsymbol{a}_{12}x_2 + \ldots + \boldsymbol{a}_{1n}x_n = \boldsymbol{b}_1, \\ \boldsymbol{a}_{21}x_1 + \boldsymbol{a}_{22}x_2 + \ldots + \boldsymbol{a}_{2n}x_n = \boldsymbol{b}_2, \\ \quad\vdots \qquad\quad \vdots \qquad\quad \ddots \qquad\quad \vdots \qquad\quad \vdots \\ \boldsymbol{a}_{m1}x_1 + \boldsymbol{a}_{m2}x_2 + \ldots + \boldsymbol{a}_{mn}x_n = \boldsymbol{b}_m, \end{cases} \tag{6}$$

or, in short,

$$\boldsymbol{A}x = \boldsymbol{b} \tag{7}$$

with an interval $m \times n$-matrix $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ and an interval m-vector $\boldsymbol{b} = (\boldsymbol{b}_i)$.

In what follows, we essentially use terminology and basic concepts of interval analysis (see e. g. [11, 12, 18]). *Interval matrix* is a rectangular table made up of interval elements. Similarly, *interval vector* is an ordered tuple of intervals disposed either horisontally (column-vector) or vertically (row-vector). Geometrical images of interval vectors are axis-aligned rectangular parallelepipeds in the standard Euclidean space $\mathbb{R}^n$. The interval vectors are also called *boxes* (see [7]). Membership of a

point in an interval vector-box is understood as set-theoretical membership, i. e., that each component of the point belongs to the corresponding interval components of the box.

An *interval linear system of equations* $\boldsymbol{A}x = \boldsymbol{b}$ is a family of point systems of linear algebraic equations (ILAS) of the same form $Ax = b$ with $A \in \boldsymbol{A}$ and $b \in \boldsymbol{b}$. The *solution set* for the interval linear system is

$$\Xi(\boldsymbol{A},\boldsymbol{b}) \;=\; \big\{\, x \in \mathbb{R}^n \mid (\exists A \in \boldsymbol{A})(\exists b \in \boldsymbol{b})(Ax = b) \,\big\},$$

i.e. the set of solutions to all usual non-interval systems $Ax = b$ whose matrices and right-hand side vectors belong to $\boldsymbol{A}$ and $\boldsymbol{b}$ respectively. The structure of the solution sets to interval linear systems has been thoroughly studied during the last decades, and details can be found, e.g., in the books [4, 12, 18].

To give a visual idea of how the solution sets to interval linear systems may look like, we consider an example of the interval linear system

$$\begin{pmatrix} 4 & [0,2] & [0,2] \\ [0,2] & 4 & [0,2] \\ [0,2] & [0,2] & 4 \end{pmatrix} x = \begin{pmatrix} [-1,1] \\ [-1,1] \\ [-1,1] \end{pmatrix}. \tag{8}$$

Its solution set is depicted in Fig. 2, and it is a non-convex polyhedron in $\mathbb{R}^3$, whose boundary "breaks", in particular, at the coordinate axes.

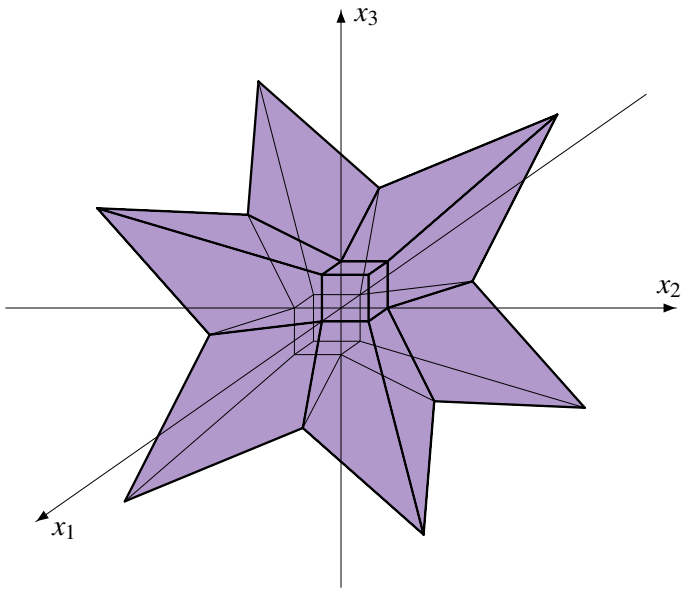

Figure 2: Solution set to the interval linear system (8).

The interval system of equations $\boldsymbol{A}x = \boldsymbol{b}$ is said to be *solvable*, if its solution set is non-empty, i. e. there holds $\Xi(\boldsymbol{A},\boldsymbol{b}) \neq \varnothing$. In the general case, checking the solvability for interval linear systems is an NP-hard problem; this fact was first established by A.V. Lakeyev et al. [9] (see also the book [4] about further results on the subject).

Analytical description of the points from the solution set to interval linear systems is given by the following result obtained by H. Beeck in [2]:

The Beeck characterization. *A point $x \in \mathbb{R}^n$ belongs to the solution set $\Xi(\boldsymbol{A},\boldsymbol{b})$ if and only if $\boldsymbol{A} \cdot x \cap \boldsymbol{b} \neq \varnothing$, i. e. the interval vectors $\boldsymbol{A} \cdot x$ and $\boldsymbol{b}$ have non-empty intersection.*

In the Beeck characterization, the product $\boldsymbol{A} \cdot x$ is understood in the sense of interval arithmetic, and we will omit the multiplication sign "$\cdot$" from now on, writing just $\boldsymbol{A}x$ instead of $\boldsymbol{A} \cdot x$. More precisely, the i-th component of the vector $\boldsymbol{A}x$ is, by definition, $\sum_{j=1}^{n} \boldsymbol{a}_{ij} x_j$, where all the operations are performed according to the rules of the classical interval arithmetic $\mathbb{IR}$ [1, 11, 12, 18]. $\mathbb{IR}$ is an algebraic system formed by closed real intervals $\boldsymbol{x} = [\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}}] \subset \mathbb{R}$ with the arithmetic operations between them defined by "representatives", i. e., as follows:

$$\boldsymbol{x} \star \boldsymbol{y} = \{\, x \star y \mid x \in \boldsymbol{x}, y \in \boldsymbol{y} \,\} \qquad \text{for} \quad \star \in \{+, -, \cdot, /\}.$$

This definition results in the following constructive formulas (see also [1, 11, 12, 18]):

$$\boldsymbol{x} + \boldsymbol{y} = \left[\underline{\boldsymbol{x}} + \underline{\boldsymbol{y}},\ \overline{\boldsymbol{x}} + \overline{\boldsymbol{y}}\right], \qquad \boldsymbol{x} - \boldsymbol{y} = \left[\underline{\boldsymbol{x}} - \overline{\boldsymbol{y}},\ \overline{\boldsymbol{x}} - \underline{\boldsymbol{y}}\right],$$
$$\boldsymbol{x} \cdot \boldsymbol{y} = \left[\min\{\underline{\boldsymbol{x}}\,\underline{\boldsymbol{y}}, \underline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}, \overline{\boldsymbol{x}}\,\underline{\boldsymbol{y}}, \overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}\},\ \max\{\underline{\boldsymbol{x}}\,\underline{\boldsymbol{y}}, \underline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}, \overline{\boldsymbol{x}}\,\underline{\boldsymbol{y}}, \overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}\}\right],$$

where underlines and overlines mean lower and upper endpoints of the respective intervals. Note that the last formula crucially simplifies if one of the operands is a degenerate interval, i. e., an exact number.

4 Recognizing functional

Let us assume that we are given the data fitting problem (1) under interval uncertainty represented by an interval matrix $\boldsymbol{A}$ and an interval vector $\boldsymbol{b}$, while the vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^\top$ determines the parameters of the linear function (1). Testing consistency between the data and parameters, equivalent to testing the Beeck characterization, amounts to examination whether the axis-aligned boxes $\boldsymbol{A}\tilde{x}$ and $\boldsymbol{b}$ intersect in the space $\mathbb{R}^n$. In doing this, we encounter the situations schematically depicted at Fig. 3.

Considering such pictures suggests the following natural idea: in the data fiting problem, we can take, as the "consistency measure" between the parameters $\tilde{x}$ and data $\boldsymbol{A}$, $\boldsymbol{b}$, a function that characterizes mutual disposition of the boxes $\boldsymbol{A}\tilde{x}$ and $\boldsymbol{b}$, the extent to which they intersect with each other or, alternatively, their deviation from each other in case of their non-intersection.

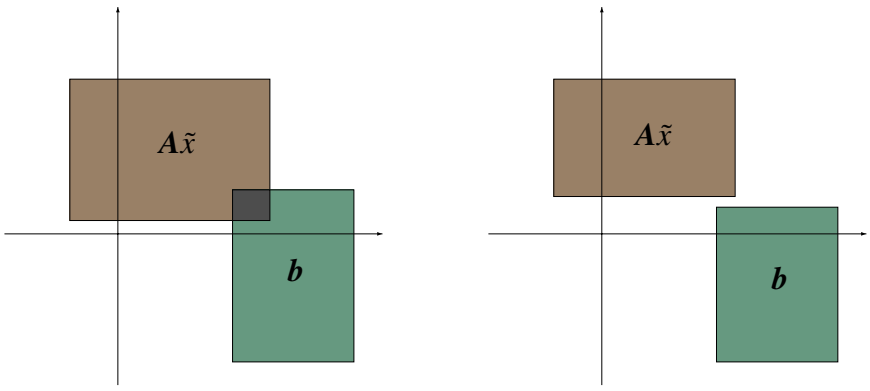

Figure 3: Varuous mutual dispositions of the boxes $\boldsymbol{A}\tilde{x}$ and $\boldsymbol{b}$ in testing the Beeck characterization.

In numerical analysis, the *residual* $A\tilde{x} - b$ (also called *defect*) between the right-hand and left-hand sides of the equation plays an important role in estimating accuracy of the approximate solution $\tilde{x}$. For interval systems of equations, an analog of the defect may be the set-theoretical difference of the

boxes $\boldsymbol{A}\tilde{x}$ and $\boldsymbol{b}$, and the size of this difference, in some reasonable sense, can serve as a "consistency measure" for the parameters $\tilde{x}$ and data $\boldsymbol{A}$, $\boldsymbol{b}$ in case of nonempty intersection $\boldsymbol{A}\tilde{x}\cap\boldsymbol{b}$. If the intersection is empty, the desired measure may be taken as a kind of distance between the boxes $\boldsymbol{A}\tilde{x}$ and $\boldsymbol{b}$.

To describe the above geometrical ideas in an analytic language, we consider first the one-dimensional case, i. e. intersection of one-dimensional intervals $\boldsymbol{a}$ and $\boldsymbol{b}$ on the real axis $\mathbb{R}$. Let us denote by $\operatorname{mid}\boldsymbol{a}$ and $\operatorname{mid}\boldsymbol{b}$ the midpoints of the intervals (half-sums of their endpoints), and let $\operatorname{rad}\boldsymbol{a}$ and $\operatorname{rad}\boldsymbol{b}$ denote their radii (half-differences of the upper and lower endpoints). It is obvious (see Fig. 4) that

$$\boldsymbol{a}\cap\boldsymbol{b}\neq\varnothing \qquad \Leftrightarrow \qquad |\operatorname{mid}\boldsymbol{b}-\operatorname{mid}\boldsymbol{a}|\leq\operatorname{rad}\boldsymbol{b}+\operatorname{rad}\boldsymbol{a}.$$

The last inequality can be written in the form

$$\operatorname{rad}\boldsymbol{b}+\operatorname{rad}\boldsymbol{a}-|\operatorname{mid}\boldsymbol{a}-\operatorname{mid}\boldsymbol{b}|\geq 0.$$

Addtitionally, in the case when the intersection of intervals $\boldsymbol{a}$ and $\boldsymbol{b}$ is empty, the absolute value of the difference $(\operatorname{rad}\boldsymbol{b}+\operatorname{rad}\boldsymbol{a}-|\operatorname{mid}\boldsymbol{a}-\operatorname{mid}\boldsymbol{b}|)$ that forms the left-hand side of the above inequality characterizes "the degree of non-intersection" for $\boldsymbol{a}$ and $\boldsymbol{b}$.

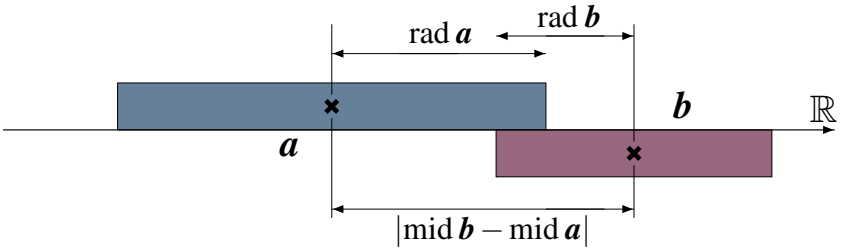

Figure 4: Intersection of intervals on the real axis (crosses mean midpoints of the intervals).

As long as the multidimensional interval vectors (boxes) are direct products of one-dimensional intervals, the following equivalence holds true:

$$\boldsymbol{A}\tilde{x}\cap\boldsymbol{b}\neq\varnothing \quad \Leftrightarrow \quad \operatorname{rad}(\boldsymbol{A}\tilde{x})_i+\operatorname{rad}\boldsymbol{b}_i-\left|\operatorname{mid}(\boldsymbol{A}\tilde{x})_i-\operatorname{mid}\boldsymbol{b}_i\right|\geq 0, \qquad i=1,2,\ldots,m. \tag{9}$$

How can we combine the results of testing the one-dimensional intersections $(\boldsymbol{A}\tilde{x})_i$ and $\boldsymbol{b}_i$ over separate components into a single expression? Generally, this can be done in several ways, but some natural requirements should be met when treating the data fitting problem. The overall expression must be negative if at least one of its one-dimensional subexpressions

$$\operatorname{rad}(\boldsymbol{A}\tilde{x})_i+\operatorname{rad}\boldsymbol{b}_i-\left|\operatorname{mid}(\boldsymbol{A}\tilde{x})_i-\operatorname{mid}\boldsymbol{b}_i\right|$$

is negative, which corresponds to empty intersection of the boxes $\boldsymbol{A}\tilde{x}$ and $\boldsymbol{b}$. The overall expression must be non-negative (positive), if all the one-dimensional subexpressions are non-negative (positive), which corresponds to nonempty intersection of the boxes $\boldsymbol{A}\tilde{x}$ and $\boldsymbol{b}$.

It is undesirable that, in the resulting expression, the one-dimensional subexpressions are tied up so that the values of one of them may substantially compensate (thus "masking") the values of the other subexpressions. The contribution of all the one-dimensional subexpressions to the overall expression should be taken in an equal manner. This is closely related to the "scalability" requirement on the resulting expression with respect to m (i. e., number of observations), when it is necessary to provide comparability of the results for different m.

Sometimes, due to mathematical reasons, it is advisable to preserve differentiability or even smoothness (continuous differentiability) of relationships to be constructed. This allows one to draw developed and well-known tools of the differential calculus for further analysis of the mathematical model. However, in our specific situation preserving smoothness of the overall expression is not necessary, since the one-dimensional subexpressions are already non-smooth, containing moduli.

For our purposes, taking minimum over the one-dimensional expressions from (9) suits us well, since it takes into account their values uniformly. To sum up, we can consider, as a "consistency measure" between the parameters $\tilde{x}$ and data $\boldsymbol{A}$, $\boldsymbol{b}$, the values of

$$\min_{1\le i\le m} \left\{ \operatorname{rad}(\boldsymbol{A}\tilde{x})_i + \operatorname{rad}\boldsymbol{b}_i - \left|\operatorname{mid}(\boldsymbol{A}\tilde{x})_i - \operatorname{mid}\boldsymbol{b}_i\right| \right\}. \tag{10}$$

To simplify the expression (10), we can notice that [1, 12, 18]

$$\operatorname{mid}(\boldsymbol{A}\tilde{x}) = (\operatorname{mid}\boldsymbol{A})\,\tilde{x} \qquad \text{and} \qquad \operatorname{rad}(\boldsymbol{A}\tilde{x}) = (\operatorname{rad}\boldsymbol{A})\,|\tilde{x}|,$$

where the operations "mid" and "rad" are applied to interval vectors and matrices in component-wise and element-wise manner. Then, instead of (10), one may write out more convenient equivalent form

$$\min_{1\le i\le m} \left\{ \operatorname{rad}\boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{rad}\boldsymbol{a}_{ij})\,|\tilde{x}_j| - \left| \operatorname{mid}\boldsymbol{b}_i - \sum_{j=1}^{n} (\operatorname{mid}\boldsymbol{a}_{ij})\,\tilde{x}_j \right| \right\}.$$

The final result of the above considerations is the following

Theorem. *Let $\boldsymbol{A}$ be an interval $m\times n$-matrix, $\boldsymbol{b}$ be an interval m-vector. Then the expression*

$$\operatorname{Uss}(x,\boldsymbol{A},\boldsymbol{b}) = \min_{1\le i\le m} \left\{ \operatorname{rad}\boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{rad}\boldsymbol{a}_{ij})\,|x_j| - \left| \operatorname{mid}\boldsymbol{b}_i - \sum_{j=1}^{n} (\operatorname{mid}\boldsymbol{a}_{ij})\,x_j \right| \right\}$$

defines such a functional $\operatorname{Uss}:\mathbb{R}^n\to\mathbb{R}$ *that the membership of a point $x\in\mathbb{R}^n$ in the solution set $\Xi(\boldsymbol{A},\boldsymbol{b})$ to the interval linear system $\boldsymbol{A}x=\boldsymbol{b}$ is equivalent to nonnegativity of the functional* Uss *in the point x. In other words, $x\in\Xi(\boldsymbol{A},\boldsymbol{b})$ if and only if* $\operatorname{Uss}(x,\boldsymbol{A},\boldsymbol{b})\ge 0$.

As a consequence of the theorem, the solution set $\Xi(\boldsymbol{A},\boldsymbol{b})$ to the interval linear system can be represented as the level set $\left\{ x\in\mathbb{R}^n \mid \operatorname{Uss}(x,\boldsymbol{A},\boldsymbol{b})\ge 0 \right\}$ of the functional Uss. It turns out that, through the sign of its values, the functional Uss "recognizes" the membership of any point in the set $\Xi(\boldsymbol{A},\boldsymbol{b})$. This is why we will call it "recognizing functional".

5 Properties of the recognizing functional

The functional Uss is obviously continuous, and even Lipschitz continuous.

Proposition 1. *The functional Uss is concave with respect to the variable x in every orthant of the space $\mathbb{R}^n$. If, in the interval matrix $\boldsymbol{A}$, l columns have nonzero widths and their numbers form the index set $J=\{j_1,j_2,\ldots,j_l\}$, $l\le n$, while the rest of the columns have zero widths (i. e., they are entirely non-interval), then the functional* $\operatorname{Uss}(x,\boldsymbol{A},\boldsymbol{b})$ *is concave with respect to x over unions of several orthants, more precisely, over each of 2^l sets of the form $\left\{ x\in\mathbb{R}^n \mid x_j \gtrless 0, j\in J \right\}$, where "$\gtrless$" means any one of the relations "$\le$" or "$\ge$".*

Proof. It is sufficient to conduct the proof for all the expressions

$$\psi_i(x) \;=\; \operatorname{rad} \boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{rad} \boldsymbol{a}_{ij})\,|x_j| - \left| \operatorname{mid} \boldsymbol{b}_i - \sum_{j=1}^{n} (\operatorname{mid} \boldsymbol{a}_{ij})\,x_j \right|,$$

$i = 1,2,\ldots,m$, since Uss is their lower envelope.

Within one orthant of $\mathbb{R}^n$, when the signs of x_j are constant, the functions

$$\operatorname{rad} \boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{rad} \boldsymbol{a}_{ij})\,|x_j| \tag{11}$$

are linear with respect to x. Additionally, the functions

$$-\left| \operatorname{mid} \boldsymbol{b}_i - \sum_{j=1}^{n} (\operatorname{mid} \boldsymbol{a}_{ij})\,x_j \right|$$

are globally concave. As the result, every $\psi_i(x)$, $i = 1,2,\ldots,m$, is also concave with respect to x within one orthant, being the sum of a linear and concave function.

The above reasoning cannot be conducted in the general case, since the expression

$$\sum_{j=1}^{n} (\operatorname{rad} \boldsymbol{a}_{ij})\,|x_j|$$

is not a linear function of x outside one orthant of the space $\mathbb{R}^n$. However, if, for some index $k \in \{1,2,\ldots,n\}$, all $\boldsymbol{a}_{ik}$ are noninterval, then $\operatorname{rad} \boldsymbol{a}_{ik} = 0$, $i = 1,2,\ldots,m$, and the expressions (11) are linear with respect to x_k.

Therefore, if the k-th column of $\boldsymbol{A}$ is noninterval, then all the functions $\psi_i(x)$ (jointly with Uss) will be concave over the sets

$$\big\{\, x \in \mathbb{R}^n \mid x_1 \gtrless 0, \ldots, x_{k-1} \gtrless 0, x_{k+1} \gtrless 0, \ldots, x_n \gtrless 0 \,\big\},$$

each of which being the union of two orthants in $\mathbb{R}^n$. Generalization of the above reasoning to the case of several noninterval columns in $\boldsymbol{A}$ is obvious. ■

To formulate our next result, we remind that, for a function $f : \mathbb{R}^n \to \mathbb{R}$, its hypograph is defined as the set $\operatorname{hyp} f = \{\, (x,t) \in \mathbb{R}^{n+1} \mid t \le f(x) \,\}$. In other words, hypograph is the graph of a function supplemented with all the points below the graph.

Proposition 2. *The functional* $\operatorname{Uss}(x,\boldsymbol{A},\boldsymbol{b})$ *is polyhedral, i. e. its graph is made up of pieces of hyperplanes, and its hypograph is a polyhedral set.*

Proof. Since $|a| = \max\{a,-a\}$, we have within every separate orthant of the space $\mathbb{R}^n$, where the signs of the components of x are constant,

$$\begin{aligned}\psi_i(x) \;=\; \operatorname{rad} \boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{rad} \boldsymbol{a}_{ij})\,\operatorname{sgn} x_j \cdot x_j \\ - \max \left\{ \operatorname{mid} \boldsymbol{b}_i - \sum_{j=1}^{n} (\operatorname{mid} \boldsymbol{a}_{ij})\,x_j \,, -\operatorname{mid} \boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{mid} \boldsymbol{a}_{ij})\,x_j \right\},\end{aligned}$$

$i = 1, 2, \ldots, m$. Therefore, within any orthant,

$$\begin{aligned}
\operatorname{Uss}(x) &= \min_{1\le i\le m} \psi_i(x) \\
&= \min_{1\le i\le m} \min \left\{ \operatorname{rad} \boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{rad} \boldsymbol{a}_{ij}) \operatorname{sgn} x_j \cdot x_j - \operatorname{mid} \boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{mid} \boldsymbol{a}_{ij}) x_j \,, \right. \\
&\qquad\qquad \left. \operatorname{rad} \boldsymbol{b}_i + \sum_{j=1}^{n} (\operatorname{rad} \boldsymbol{a}_{ij}) \operatorname{sgn} x_j \cdot x_j + \operatorname{mid} \boldsymbol{b}_i - \sum_{j=1}^{n} (\operatorname{mid} \boldsymbol{a}_{ij}) x_j \right\} \\
&= \min_{1\le i\le m} \min \left\{ -\underline{\boldsymbol{b}}_i + \sum_{j=1}^{n} \left(\operatorname{rad} \boldsymbol{a}_{ij} \operatorname{sgn} x_j + \operatorname{mid} \boldsymbol{a}_{ij}\right) x_j, \right. \\
&\qquad\qquad \left. \overline{\boldsymbol{b}}_i + \sum_{j=1}^{n} \left(\operatorname{rad} \boldsymbol{a}_{ij} \operatorname{sgn} x_j - \operatorname{mid} \boldsymbol{a}_{ij}\right) x_j \right\} \\
&= \min_{1\le i\le m} \min \left\{ -\underline{\boldsymbol{b}}_i + a_i' x \,,\, \overline{\boldsymbol{b}}_i - a_i'' x \right\}, \qquad (12)
\end{aligned}$$

where a_i' and a_i'' are vertices of the interval row vector $(\boldsymbol{a}_{i1}, \boldsymbol{a}_{i2}, \ldots, \boldsymbol{a}_{in})$, defined as

$$\begin{aligned}
\left(a_i'\right)_j &= \operatorname{rad} \boldsymbol{a}_{ij} \operatorname{sgn} x_j + \operatorname{mid} \boldsymbol{a}_{ij}, \\
\left(a_i''\right)_j &= \operatorname{rad} \boldsymbol{a}_{ij} \operatorname{sgn} x_j - \operatorname{mid} \boldsymbol{a}_{ij}, \qquad j = 1, 2, \ldots, n.
\end{aligned}$$

The expressions in curly brackets from (12), over which the minimums are taken, determine linear functions. In every orthant, $\operatorname{Uss}(x)$ is thus minimum of a finite number of linear functions, being piecewise linear in general. ■

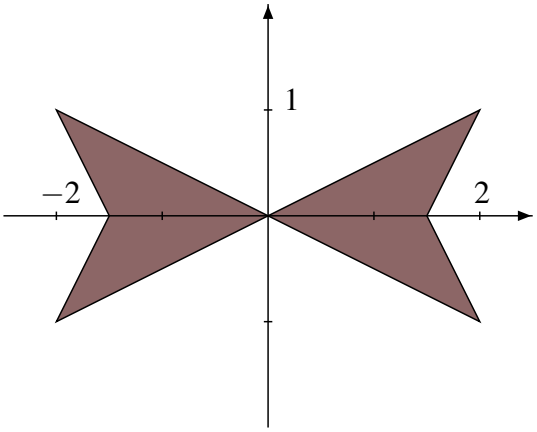

Figure 5: Solution set to the interval linear system (13).

As an example illustrating Proposition 2, we consider interval linear system of equations

$$\begin{pmatrix} [2,4] & [-1,1] \\ [-1,1] & [2,4] \end{pmatrix} x = \begin{pmatrix} [-3,3] \\ 0 \end{pmatrix}. \qquad (13)$$

Its solution set is depicted at Fig. 5, and the graph of its recognizing functional can be seen at Fig. 6.

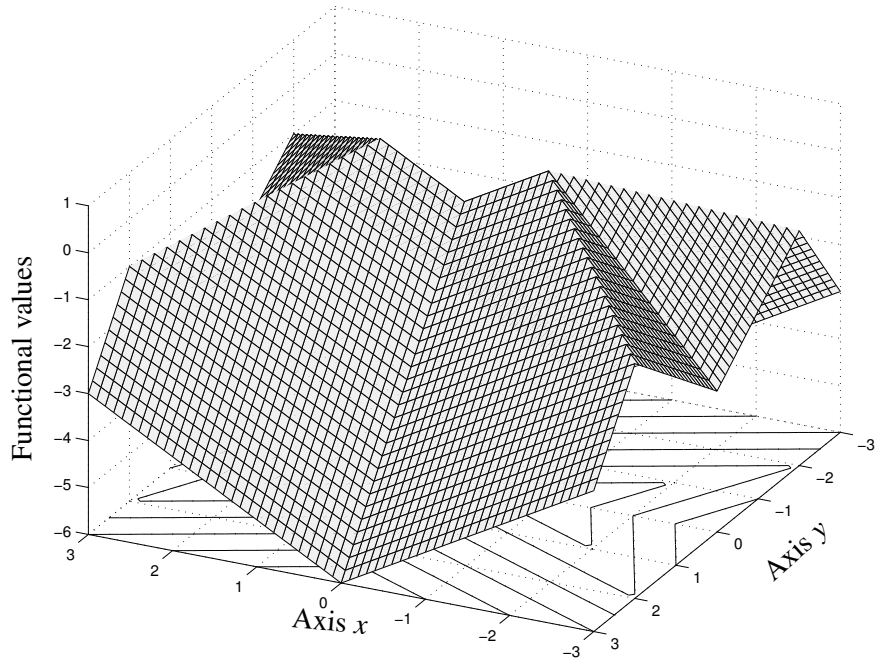

Figure 6: Graph of the recognizing functional for the system (13).

Proposition 3. *If the solution set $\Xi(\boldsymbol{A},\boldsymbol{b})$ to the interval linear system $\boldsymbol{A}x=\boldsymbol{b}$ is bounded, then the recognizing functional* $\mathrm{Uss}\,(x,\boldsymbol{A},\boldsymbol{b})$ *attains a finite maximum over the entire space $\mathbb{R}^n$.*

Proposition 4. *If* $\mathrm{Uss}\,(x,\boldsymbol{A},\boldsymbol{b})>0$*, then x is a point from topological interior* $\mathrm{int}\,\Xi(\boldsymbol{A},\boldsymbol{b})$ *of the solution set, i. e. it belongs to the solution set with some its neighbourhood.*

Proposition 5. *Let the interval linear system of equations $\boldsymbol{A}x=\boldsymbol{b}$ be such that its augmented matrix $(\boldsymbol{A},\boldsymbol{b})$ does not contain rows in which all elements have zero endpoints. Then the membership $x\in$* $\mathrm{int}\left(\Xi(\boldsymbol{A},\boldsymbol{b})\cap\mathscr{O}\right)$*, where $\mathscr{O}$ is an orthant of the space $\mathbb{R}^n$, implies the strict inequality* $\mathrm{Uss}\,(x,\boldsymbol{A},\boldsymbol{b})>0$.

We omit proofs for Propositions 3–5; the reader can find them in the works [19, 20].

6 Maximum consistency method

Turning to the data fitting problem under interval uncertainty and to the main question how to take "consistency / inconsistency measure" between the data $\boldsymbol{A}$, $\boldsymbol{b}$ and parameters x of the relationship to be constructed, we can repeatedly note that the recognizing functional Uss of the interval linear system $\boldsymbol{A}x=\boldsymbol{b}$ suits well for this purpose. Indeed, in case of nonempty information set, it possesses nonnegative and positive values for the points from this set, where the actual "consistency" really takes place. For the points outside the information set, where no "consistency" holds, the values of the functional Uss are negative. Besides, with other things being equal, the values of Uss are positive in the interior of the solution set, and they exceed the values at the boundary of the solution set, which are mostly zero.

The above considerations motivate the following approach to the solution of the data fitting problem under interval uncertainty:

As an estimate of the function parameters, we take the point that provides maximum of the recognizing functional Uss.

We will call this approach to determining the parameters of the linear function (1) that best fits the interval data as *maximum consistency method*, since the functional Uss has been shown to characterize the "consistency measure" between the parameters and data. So,

- if $\max \mathrm{Uss} \geq 0$, then the argument that delivers maximum to the functional Uss lies in a nonempty set of parameters consistent with the data;
- if $\max \mathrm{Uss} < 0$, then the set of parameters consistent with the data is empty, but the point where the maximum of Uss is attained minimizes the inconsistency.

It makes sense to note that, generally, the maximum consistency point which maximizes the recognizing functional can be non-unique (unlike, e. g., the least squares method where the solution is typically unique).

7 Properties of the maximum consistency estimates

The expression defining the recognizing functional Uss has a form that enables one to predict how the functional will change its values after this or that variation in $\boldsymbol{A}$ and $\boldsymbol{b}$. We can use such information for further correction of the interval linear system in a necessary sense. Let us consider the simplest of this kind of construction.

Note that, in the expression

$$\min_{1\leq i\leq m}\left\{\operatorname{rad}\boldsymbol{b}_i+\sum_{j=1}^{n}(\operatorname{rad}\boldsymbol{a}_{ij})\,|x_j|-\left|\operatorname{mid}\boldsymbol{b}_i-\sum_{j=1}^{n}(\operatorname{mid}\boldsymbol{a}_{ij})\,x_j\right|\right\},$$

the quantities $\operatorname{rad}\boldsymbol{b}_i$, $i=1,2,\ldots,m$, occur as addons in every subexpression standing under "min" operation. Therefore, if all $\operatorname{rad}\boldsymbol{b}_i$'s simultaneously increase or decrease by equal values, then the general minimum increases or decreases by the same value.

Uniform increase of the radii of the right-hand sides $\boldsymbol{b}_i$ by C, $C\geq 0$, is equivalent to adding the vector $C\boldsymbol{e}$ to $\boldsymbol{b}$, where $\boldsymbol{e}=\big([-1,1],\ldots,[-1,1]\big)^{\top}$. Then, for the interval system $\boldsymbol{A}x=\boldsymbol{b}+C\boldsymbol{e}$ with the widened right-hand side, there holds

$$\mathrm{Uss}\,(x,\boldsymbol{A},\boldsymbol{b}+C\boldsymbol{e}) \;=\; \mathrm{Uss}\,(x,\boldsymbol{A},\boldsymbol{b})+C,$$

and, as a consequence,

$$\max_{x\in\mathbb{R}^n}\mathrm{Uss}\,(x,\boldsymbol{A},\boldsymbol{b}+C\boldsymbol{e}) \;=\; \max_{x\in\mathbb{R}^n}\mathrm{Uss}\,(x,\boldsymbol{A},\boldsymbol{b})+C. \tag{14}$$

From (14), a practical interpretation of the maximum consistency method ensues: the value of the argument providing $\max\mathrm{Uss}$ is the first point that appears in the solution set after uniform (with respect to its midpoint) widening of the right-hand side vector.

Yet another interpretation of the maximum consistency method may be as follows: the argument of $\max\mathrm{Uss}$ gives us parameters of such a regression line that should be widened in the smallest possible amount to result in a "regression strip" intersecting all the data boxes (see Fig. 7).

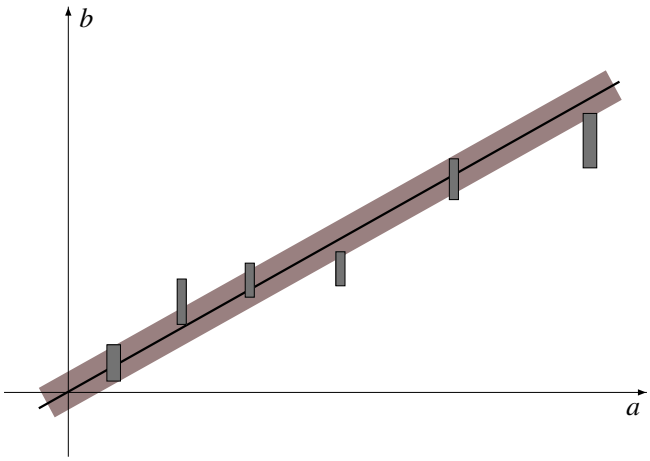

Figure 7: Regression strip (instead of thin line) as a solution to data fitting problem under interval uncertainty.

To understand relations between the maximum consistency method and the other approaches to the data fitting problem, it makes sense to consider the extreme case when the data are precise and do not have any uncertainty. What does maximum consistency method produce then?

If the input data matrix $\boldsymbol{A}$ and output data vector $\boldsymbol{b}$ are non-interval (common real), i. e. $\boldsymbol{A} = A = (a_{ij})$ and $\boldsymbol{b} = b = (b_i)$, then for all i, j

$$\operatorname{rad} \boldsymbol{a}_{ij} = 0,\ \operatorname{rad} \boldsymbol{b}_i = 0 \qquad \text{and} \qquad \operatorname{mid} \boldsymbol{a}_{ij} = a_{ij},\ \operatorname{mid} \boldsymbol{b}_i = b_i.$$

The recognizing functional of the equations system then takes the form

$$\operatorname{Uss}(x,A,b) = \min_{1\le i\le m} \left\{ -\left| b_i - \sum_{j=1}^{n} a_{ij}x_j \right| \right\} = -\max_{1\le i\le m} \left| b_i - \sum_{j=1}^{n} a_{ij}x_j \right|$$

$$= -\max_{1\le i\le m} \left| (Ax)_i - b_i \right| = -\|Ax-b\|_\infty .$$

By $\|\cdot\|_\infty$, we denote Chebyshev vector norm (∞-norm) in a finite-dimensional space $\mathbb{R}^m$, which is defined as $\|y\|_\infty = \max\limits_{1\le i\le m} |y_i|$. Then

$$\max \operatorname{Uss}(x) = \max_{x\in\mathbb{R}^n} \left(-\|Ax-b\|_\infty\right) = -\min_{x\in\mathbb{R}^n} \|Ax-b\|_\infty,$$

since $\max(-f(x)) = -\min f(x)$. Therefore, under the circumstances, maximization of the recognizing functional is equivalent to minimization of the Chebyshev norm of the discrepancy between the left-hand and right-hand sides of the equations system. Maximum consistency method then turns into Chebyshev data smoothing that has been successfully applied in data analysis for a long time (see e. g. [15]).

8 Implementation

Practical implementation of the maximum consistency method crucially depends on how efficient is the computation of $\max \operatorname{Uss}$. In the general case, this is a global optimization problem with nonsmooth and multiextrema objective function, which is quite hard to solve.

Nevetherless, to compute unconstrained maximum of Uss, we can successfully apply some presently existing techniques taking into account specificity of the functional Uss. First of all, we can try the exhaustive search over all the concavity regions of the functional Uss (in the general case, over all the orthants of $\mathbb{R}^n$) and using efficient non-smooth convex optimization methods in every one of such regions. In particular, good results have been demonstrated in the solution of this kind of problem by *r*-algorithms developed by N. Shor [21, 22] as well as separating plane method by E. Nurminski [13, 24].

The overall number of the concavity regions for the recognizing functional is determined, according to Proposition 2, by the number of essentially interval columns in the input data matrix $\boldsymbol{A}$ (put it differently, by the number of a_i's in (1) subject to interval uncertainty). If, in the data matrix $\boldsymbol{A}$, the number of interval columns is not large (say, no more than ten), then the above approach is quite practical.

One more possible approach to maximization of Uss can be based on the fact that the functional Uss belongs to the class of so-called d.c.-functions, i. e. such that can be represented as the difference of two convex functions. For optimization of such functions, a number of promising methods have been elaborated in the last years [23], which may hopefully lead to implementation of practical and efficient algorithms.

It is worth mentioning the most important particular case of the data fitting problem that corresponds to the exact definition of all the input variables a_1, a_2, ..., a_n, when interval uncertainty is present only on output b. Then the regression line (hyperplane in the general case) should go through vertical segments parallel to the b axis, not through the solid uncertainty boxes shown in Fig. 1.

As a result, instead of the interval linear system of the general form (6)–(7) we get the system of equations

$$Ax = \boldsymbol{b}$$

with a non-interval (thin) matrix $A = (a_{ij})$. Then all rad $a_{ij} = 0$, and the recognizing functional of the solution set crucially simplifies:

$$\mathrm{Uss}\,(x,\boldsymbol{A},\boldsymbol{b}) \;=\; \min_{1\le i\le m} \left\{ \mathrm{rad}\,\boldsymbol{b}_i - \left| \mathrm{mid}\,\boldsymbol{b}_i - \sum_{j=1}^{n} a_{ij}x_j \right| \right\}. \tag{15}$$

In the above expression, the "min" operation is taken over the differences between constants and concave functions with respect to x. Therefore, the recognizing functional of the form (15) is globally concave, and its graph should look similar to what is depicted at Fig. 8 rather than the multiextrema configuration of Fig. 5. The picture at Fig. 8 shows the graph of the recognizing functional for the interval linear system

$$\begin{pmatrix} 3 & -1 \\ -1 & 2 \\ 1 & 2 \\ 1 & 0 \end{pmatrix} x = \begin{pmatrix} [-2,2] \\ [0,1] \\ [-1,0] \\ [1,3] \end{pmatrix}. \tag{16}$$

For the case of non-interval matrix $\boldsymbol{A}$, maximization of the concave recognizing functional Uss can, again, rely upon developed methods of non-smooth convex optimization (e. g. those from [13, 21, 22, 24]).

9 Results and conclusions

Introduction of the recognizing functional of the solution set to interval linear algebraic systems reduces the problem of testing solvability of such systems to a convenient analytical form that enables

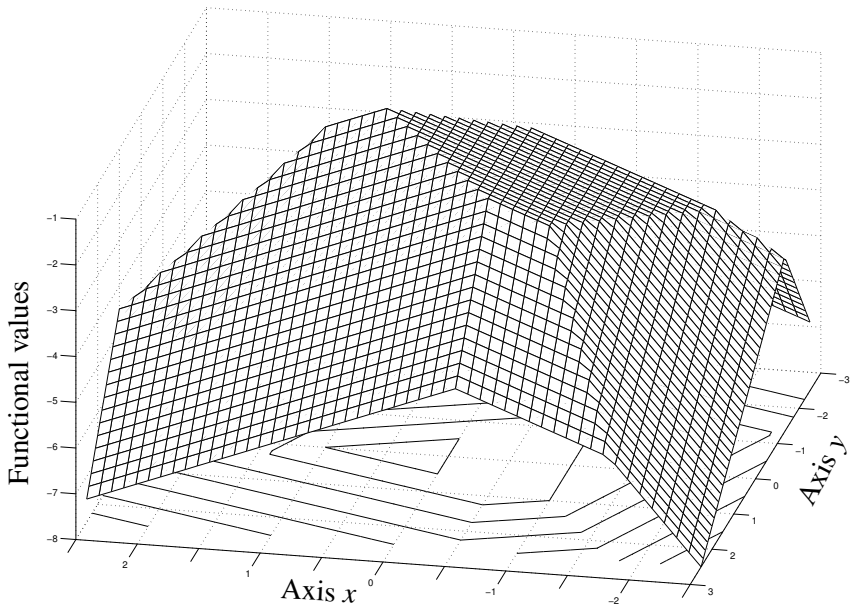

Figure 8: Graph of recognizing functional for the solution set to (16).

one correcting the initial data and the problem statement itself.

Maximization of the recognizing functional provides us with important information about solvability margin of the interval linear systems and properties of its solution set. Additionally, we can lay it down as a basis of a new technique for the solution of the data fitting problem under interval uncertainty called *maximum consistency method*. It is a good alternative to the traditional statistical approaches based on probabilistic models of observation errors.

References

[1] Alefeld, G., Herzberger, J.: Introduction to Interval Computations. Academic Press, New York (1983)

[2] Beeck, H.: Über die Struktur und Abschätzungen der Lösungsmenge von linearen Gleichungssystemen mit Intervallkoeffizienten. Computing **10**, 231–244 (1972)

[3] Draper, N. R., Smith, H.: Applied Regression Analysis, 3rd edition. Wiley-Interscience, New York (1998)

[4] Fiedler, M., Nedoma, J., Ramik, J., Rohn, J., Zimmerman, M.: Linear Optimization Problems with Inexact Data. Springer, Berlin (2006)

[5] Jaulin, L., Kieffer, M., Didrit, O., Walter, E.: Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics. Springer, London (2001)

[6] Kantorovich, L. V.: On some new approaches to numerical methods and processing observation data. Siberian Math. Journal **3** (5), 701–709 (1962) (in Russian) Electronic version is accessible at `http://www.nsc.ru/interval/Introduction/Kantorovich62.pdf`

[7] Kearfott, R. B., Nakao, M., Neumaier, A., Rump, S., Shary, S. P., van Hentenryck, P.: Standardized notation in interval analysis. Comput. Technologies **15** (1), 7–13 (2010)

[8] Kurzhanski, A. B., Vályi, I.: Ellipsoidal Calculus for Estimation and Control. Birkhäuser, Boston (1996)

[9] Lakeyev, A. V., Noskov, S. I.: A description of the set of solutions of a linear equation with intervally defined operator and right-hand side, Russian Acad. Sci. Dokl. Math. **47** (3), 518–523 (1993)

[10] Milanese, M., Norton, J., Piet-Lahanier, H., Walter, E., eds.: Bounding Approaches to System Identification. Plenum Press, New York (1996)

[11] Moore, R. E., Kearfott, R. B., Cloud, M. J.: Introduction to Interval Analysis. SIAM, Philadelphia (2009)

[12] Neumaier, A.: Interval Methods for Systems of Equations. Cambridge University Press, Cambridge (1990)

[13] Nurminski, E. A.: Separating plane algorithms for convex optimization // Mathematical Programming **76** (3), 373–391 (1997)

[14] Rao, C. R., Toutenburg, H., Shalabh, Heumann, C.: Linear Models and Generalizations. Least Squares and Alternatives. Springer, New York (2008)

[15] Remez, E. Ya.: General Computational Methods of Chebyshev Approximation. Kiev, USSR, Ukranian Academy of Sciences (1957) (in Russian) Atomic Energy Commission Translation 4491, 1–85 (1957)

[16] Schweppe, F. C.: Recursive state estimation: unknown but bounded errors and system inputs. IEEE Trans. on Autom. Control **13** (1), 22–28 (1968)

[17] Shary, S. P.: Solvability of interval linear equations and data analysis under uncertainty. Automation and Remote Control **73** (2), 310–322 (2012) doi: 10.1134/S0005117912020099

[18] Shary, S. P.: Finite-Dimensional Interval Analysis. Institute of Computational Technologies SB RAS, Novosibirsk (2013) Electronic book accessible at `http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf`

[19] Shary, S. P., Sharaya, I. A.: Recognition of solvability of interval equations and its application to data analysis. Comput. Technologies **18** (3), 80–109 (2013) (in Russian)

[20] Shary, S. P., Sharaya, I. A.: On solvability recognition for interval linear systems of equations. Optimization Letters **10** (2), 247–260 (2016) doi: 10.1007/s11590-015-0891-6

[21] Shor, N. Z.: Nondifferentiable Optimization and Polynomial Problems. Kluwer, Boston-Dordrecht-London (1998)

[22] Shor, N. Z., Stetsyuk, P. I.: Modified *r*-algorithm to find the global minimum of polynomial functions. Cybernetics and Systems Analysis **33** (4), 482–497 (1997)

[23] Strekalovsky, A. S.: On the minimization of the difference of convex functions on a feasible set. Comput. Math. and Math. Physics **43** (3), 380–390 (2003)

[24] Vorontsova, E. A.: A projective separating plane method with additional clipping for nonsmooth optimization. WSEAS Transactions on Mathematics **13**, 115–121 (2014)

[25] Zhilin, S. I.: On fitting empirical data under interval error. Reliable Computing **11** (5), 433–442 (2005)