

Interval Methods for Data Fitting under Uncertainty: A Probabilistic Treatment

Vladik Kreinovich

University of Texas at El Paso,
El Paso, USA

`vladik@utep.edu`

Sergey P. Shary*

Institute of Computational Technologies
and Novosibirsk State University,
Novosibirsk, Russia

`shary@ict.nsc.ru`

Abstract

How to estimate parameters from observations subject to errors and uncertainty? Very often, the measurement errors are random quantities that can be adequately described by the probability theory. When we know that the measurement errors are normally distributed with zero mean, then the (asymptotically optimal) Maximum Likelihood Method leads to the popular least squares estimates. In many situations, however, we do not know the shape of the error distribution, we only know that the measurement errors are located on a certain interval. Then the maximum entropy approach leads to a uniform distribution on this interval, and the Maximum Likelihood Method results in the so-called minimax estimates. We analyse specificity and drawbacks of the minimax estimation under essential interval uncertainty in data and discuss possible ways to solve the difficulties. Finally, we show that, for the linear functional dependency, the minimax estimates motivated by the Maximum Likelihood Method coincide with those produced by the Maximum Consistency Method that originate from interval analysis.

Mathematics Subject Classification: 62J02, 62J05, 65G40

Keywords: data fitting, regression, interval uncertainty, maximum likelihood method, maximum consistency method.

*Corresponding author.

1 Introduction

The paper is devoted to data analysis under uncertainty, when we do not know exact values of our measurements, observations, etc., but instead we have at our disposal some alternative information. These can be bounds of possible errors of the measured quantities, both lower and upper, which is equivalent to determining interval results of measurements. On the other hand, these can be probabilistic characteristics of the errors corrupting the measured quantities, and then we can use well-developed methods of probability theory to process our data.

Data analysis methods based on probability theory are very elaborate and popular, they have been applied to processing the results of measurements and observations for about two centuries. Interval methods came into being in the middle of the XX century, and nowadays they only start winning the favor of engineers and practitioners. The purpose of our paper is to show that the results and conclusions drawn by these different approaches are in good conformity with each other, so that any one of them can be used to justify and substantiate (or even to verify) the other.

Interval methods in data analysis originated from the pioneering works by L. V. Kantorovich¹ [6] and F. C. Schweppe [18]. For the last decades, they have been developed deeply and extensively by many researchers throughout the world.

Broad research area and a great variety of applications resulted in great terminological diversity: doing similar (and even the same) things, people speak of “guaranteed parameter estimation”, “set-membership estimation”, “bounded-error approach”, “interval data fitting”, “interval regression”, etc. To get an insight into the current state of knowledge in this field, the reader can consult, e.g., [4, 8] and the literature cited there.

Our paper appears in the special issue of the journal *Reliable Computing* devoted to Ramon E. Moore and his scientific heritage. The topic of the paper has a direct relationship to the activity of Ramon Moore who suggested, in [10], an elegant way to present and describe the solution sets in nonlinear parameter estimation.

Formulation of the problem. In many situations, we know the general form of the functional dependency between the quantities $x = (x_1, \dots, x_k)$ and y , i.e., we know that

$$y = f(x, p), \tag{1}$$

where $p = (p_1, \dots, p_l)$ is a parameter. Based on actual values of the variables x and y , we have to find the values of p_1, \dots, p_l that correspond to a specific function f from the parametric family (1). This problem is referred to as “parameter estimation problem”, “data fitting problem”, “regression problem”, “curve fitting”, etc.

¹Leonid Kantorovich was one of the founders, along with George Dantzig, of the linear programming, as well as Nobel Prize winner in economic sciences.

To find the parameters p , we repeatedly measure the corresponding values of x and y . As a result, for each measurement $i = 1, 2, \dots, m$, we get the corresponding values $x^{(i)} = (x_1^{(i)}, \dots, x_k^{(i)})$ and $y^{(i)}$. Our task is to determine such $p^* = (p_1^*, \dots, p_l^*)$ that the function $y = f(x, p^*)$ “best fits” the measurement data set

$$\begin{aligned} & x^{(1)}, \quad y^{(1)}, \\ & x^{(2)}, \quad y^{(2)}, \\ & \dots, \quad \dots, \\ & x^{(m)}, \quad y^{(m)} \end{aligned} \quad (2)$$

(see Fig. 1). At this point, we have to explain the sense in which we understand the “best fitting”.

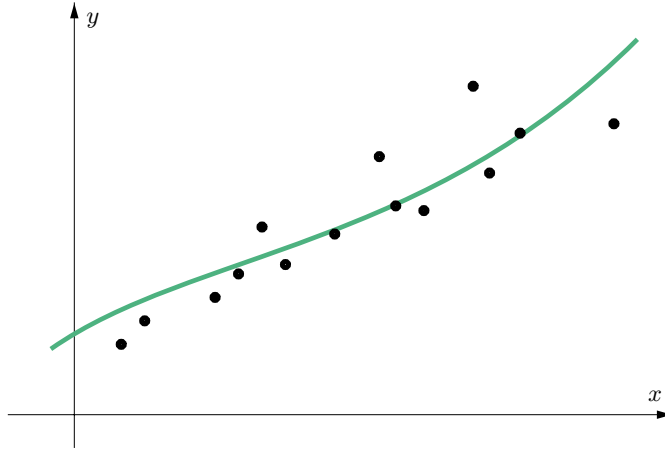


Figure 1: In data fitting problem, we have to construct a line that best fits measurement data.

Ideally, the “best fit” line should go through all the measurement points of the set (2). This happens when the measurements are so accurate that we can safely ignore any errors and assume that the measured values of the quantities x_1, x_2, \dots, x_k and y are exact. In this case, we get a system of equations

$$\begin{cases} f(x^{(1)}, p) = y^{(1)}, \\ f(x^{(2)}, p) = y^{(2)}, \\ \vdots \quad \quad \quad \vdots \\ f(x^{(m)}, p) = y^{(m)}, \end{cases} \quad (3)$$

with l unknowns p_1, \dots, p_l . Having solved the system, we obtain the desired values of the parameters that correspond to the data analyzed. In general, to find l unknowns, it is sufficient to have l different equations. Thus, in the ideal case of absolutely accurate measurements and providing that f adequately

describes the actual functional dependency under study, we would be able to determine the values of all l parameters p_i after $m = l$ measurements.

However, measurements are not precise. They can even include so-called outliers, measurement results that largely deviate from the rest of data set, which may be caused by experimental errors, etc. The function f may not perfectly reflect the functional dependency that exists between x and y in practice. Usually, f is considered as a “main part” (to within neglected terms) of the actual function between x and y . And so on. In such situations, we cannot expect that the equalities $y^{(i)} = f(x^{(i)}, p)$, $i = 1, 2, \dots, m$, are exactly satisfied for some p . Then the “best fit” line should be a best approximation, in a prescribed sense, of the data set (2) (this is the situation described at Fig. 1).

In such situations, the equation system (3) cannot be solved exactly, and we have to find its solution in some generalized sense. The latter is especially true since we often try to make as much measurements as possible: every measurement provides us with additional information about f . Hence, the equation system (3) can be overdetermined.

We are going to consider the simplified situation with the measurement errors when the measured values $x_j^{(i)}$ are exact, but the measured values $\tilde{y}^{(i)}$ are, in general, different from the actual values $y^{(i)} = f(x^{(i)}, p)$.

2 Probabilistic approaches

In this section, we discuss how to estimate parameters when our measurements are not exact, but we have certain probabilistic information about measurement errors. The material we present below is quite standard, but we survey it to make our paper self-sufficient, as well as to prepare and motivate further conclusions. Notice that even the fact that probabilistic approaches are applicable to processing specific data is not trivial and should be substantiated on its own.

Typical probabilistic prerequisites. In some cases, we can be sure that the probability theory is an adequate tool for describing the measurement errors. Moreover, the probability distributions of the measurement errors $\Delta y^{(i)} \stackrel{\text{def}}{=} \tilde{y}^{(i)} - y^{(i)}$ or at least the shape of the corresponding probability distributions are known. Let us consider this kind of situation.

First of all, from the repeated measurements, we often know, with good accuracy, the mean value of the measurement error. In such situations, we can correct the measurement results by subtracting this bias (mean value). So, we can suppose that the mean value of the measurement error is equal to 0.

Frequently, when all major measurement disturbances are eliminated, the remaining errors are formed from many small error sources. Under reasonable conditions, the probability distribution of the sum of N small independent random variables is known to tend to the normal distribution when N increases; this fact is called the *Central Limit Theorem* [17, 27]. Thus, when the measurement error comes from the joint effect of a large number of small independent

components, we can safely assume that the resulting probability distribution of the overall measurement error is normal. For a general normal Gaussian distribution with zero mean, the probability density as a function of the error Δy has the form

$$\rho(\Delta y) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(\Delta y)^2}{2\sigma^2}\right) \quad (4)$$

for an appropriate value σ (called *standard deviation*).

We would remind the following empirical fact: in practice, about 60% of the measuring instruments have a normally distributed measurement error; see, e.g., [14, 15].

How to estimate the parameters in the case of normal error distribution. It is well known that an asymptotically optimal way to determine the parameters of a distribution from the sample values is the *Maximum Likelihood Method*. According to it, we select the values of the unknown parameters so that the probability of the given sample of values (in the case of discrete probability distributions) or probability density at the given sample of values (in the case of continuous probability distributions) are the largest possible; see, e.g., [7, 9].

Measurement errors corresponding to different measurements are usually assumed to be independent. The probability of several independent events occurring together is equal to the product of the corresponding probabilities. Thus, for the normal distribution (4), the above implies that we have to select the unknowns p and σ that maximize the product

$$\mathcal{L} = \prod_{i=1}^m \rho(\Delta y^{(i)}) = \prod_{i=1}^m \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(\Delta y^{(i)})^2}{2\sigma^2}\right).$$

Substituting the expression $\Delta y^{(i)} = \tilde{y}^{(i)} - f(x^{(i)}, p)$ into this formula, we conclude that p and σ should maximize the product

$$\mathcal{L} = \prod_{i=1}^m \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(\tilde{y}^{(i)} - f(x^{(i)}, p))^2}{2\sigma^2}\right).$$

The constant coefficient $1/(\sigma\sqrt{2\pi})$ is the same for all i , so we have

$$\begin{aligned} \mathcal{L} &= \frac{1}{(\sigma\sqrt{2\pi})^m} \cdot \prod_{i=1}^m \exp\left(-\frac{(\tilde{y}^{(i)} - f(x^{(i)}, p))^2}{2\sigma^2}\right) \\ &= \frac{1}{(\sigma\sqrt{2\pi})^m} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^m (\tilde{y}^{(i)} - f(x^{(i)}, p))^2\right). \end{aligned}$$

From monotonicity of the function \exp , it follows that maximizing \mathcal{L} is equivalent to minimizing the sum

$$\sum_{i=1}^m (\tilde{y}^{(i)} - f(x^{(i)}, p))^2. \quad (5)$$

We thus arrive at the well-known *least squares method* independently proposed by A.-M. Legendre and K.F. Gauss (see, e.g., [3, 7]): the values of p should be selected so that the sum of the squares of the approximation errors $\Delta y^{(i)} = \tilde{y}^{(i)} - f(x^{(i)}, p)$ is the smallest possible. Using the least squares method is an asymptotically optimal way to estimate the parameters from the observations when the measurement errors are normally distributed with zero mean [3].

However, the above approach and the least squares method are not universal. As we have mentioned, for about 40% of measuring instruments, the measurement errors are not normally distributed. Moreover, often, we do not have enough measurements to determine the actual shape of the corresponding probability distribution (this is known as “small sample size” problem). In many such situations, the only information we know is that the possible values of the measurement error are located within some bounds $\underline{\Delta}$ and $\overline{\Delta}$, but we do not know the probability of different values within the interval $[\underline{\Delta}, \overline{\Delta}]$. What shall we do in this case?

How to estimate the parameters when the error distribution is not known. Situations when we do not know the exact probabilistic distribution of the measurement errors are ubiquitous. Then, several different distributions can be compatible with our knowledge. Some of these distributions are more informative, some are less. It is reasonable to select, among all the distributions, the one with the least possible amount of information, i.e., the one which does not add anything to our knowledge that the random variable is located on a given interval.

If a random variable has the probability density function $\rho(x)$, the amount of information it bears is usually described by the *entropy*

$$S = - \int_{-\infty}^{\infty} \rho(x) \ln \rho(x) dx, \quad (6)$$

see, e.g., [1, 5, 12]. In fact, the entropy gives a measure of how chaotic the probabilistic distribution is. So, the more entropy of a distribution, the less informative it is. Then the above idea means that, among all the probability distributions $\rho(x)$ compatible with our knowledge, we should select the one for which entropy (6) attains the largest possible value. This consideration is known as the *maximum entropy approach*.

Let us apply the maximum entropy approach to the case when the only information about the probability distribution is that it is located on the interval $[\underline{\Delta}, \overline{\Delta}]$, i.e., that $\rho(x) = 0$ for $x \notin [\underline{\Delta}, \overline{\Delta}]$. Thus, we need to maximize the entropy

$$S = - \int_{-\infty}^{+\infty} \rho(x) \ln \rho(x) dx = - \int_{\underline{\Delta}}^{\overline{\Delta}} \rho(x) \ln \rho(x) dx. \quad (7)$$

Additionally, expression (7) should be maximized under the constraint that the

overall probability is 1, i. e.,

$$\int_{\underline{\Delta}}^{\overline{\Delta}} \rho(x) dx = 1. \quad (8)$$

The method of Lagrange multipliers reduces this constraint optimization problem to an unconstrained problem of maximizing the function

$$-\int_{\underline{\Delta}}^{\overline{\Delta}} \rho(x) \ln \rho(x) dx + \lambda \cdot \left(\int_{\underline{\Delta}}^{\overline{\Delta}} \rho(x) dx - 1 \right),$$

where λ is a Lagrange multiplier. Taking a (variational) derivative over $\rho(x)$ and equating this derivative to 0, we conclude that $-\ln \rho(x) - 1 + \lambda = 0$, hence $\ln \rho(x) = \lambda - 1 = \text{const}$ and thus, $\rho(x) = \text{const}$. The value of this constant can be determined from equality (8) that the overall probability should be equal to 1. Therefore, we have a *uniform* distribution with the probability density

$$\rho(x) = \frac{1}{\overline{\Delta} - \underline{\Delta}} = \text{const}. \quad (9)$$

The conclusion that we should select a uniform distribution is in good accordance with common sense. Indeed, since we have no reason to believe that some values from the interval $[\underline{\Delta}, \overline{\Delta}]$ are more probable and some values are less probable, it makes sense to select a distribution in which all these values are equally probable, i. e., a uniform distribution on this interval.²

As we have mentioned, it is usually reasonable to assume that the main part of the systematic error (bias) has been eliminated and thus, the mean value for the measurement error is 0. For a uniform distribution, the mean value is equal to the midpoint of the corresponding interval, and the above requirement takes the form

$$\frac{\underline{\Delta} + \overline{\Delta}}{2} = 0.$$

So, if we denote $\Delta \stackrel{\text{def}}{=} \overline{\Delta}$, then $\underline{\Delta} = -\Delta$, and the interval should be $[-\Delta, \Delta]$ for a certain value Δ .

How to estimate the parameters in the case of uniform error distribution. In this situation, we have $l + 1$ unknowns: l parameters $p = (p_1, \dots, p_l)$ and the parameter Δ that characterizes the measurement uncertainty. To find the parameters, we are going to use, similar to the case of normal error distribution, the Maximum Likelihood Method, i. e., we maximize the product

$$\mathcal{L} = \prod_{i=1}^m \rho(\Delta y^{(i)}).$$

²The mathematical fact that the uniform distribution provides the maximum entropy is widely known, being a “mathematical folklore”, and its proof can be found in many sources, e. g., in [12].

The probability distribution $\rho(x)$ is located on the interval $[-\Delta, \Delta]$. Thus, if the absolute value $|\Delta y^{(i)}|$ of one of the measurement errors $\Delta y^{(i)}$ exceeds Δ , then the corresponding factor $\rho(\Delta y^{(i)})$ is equal to 0 and therefore, the entire product \mathcal{L} is equal to 0.

To find the desired maximum of \mathcal{L} , we have to consider only the values p and Δ for which

$$|\Delta y^{(i)}| \leq \Delta \quad (10)$$

for all $i = 1, 2, \dots, m$, i.e., for which $\Delta \geq \max_i |\Delta y^{(i)}|$. For such values, $\rho(\Delta y^{(i)}) = 1/(2\Delta)$ according to (9), and the product \mathcal{L} takes the form

$$\mathcal{L} = \frac{1}{(2\Delta)^m}.$$

This expression is the largest if and only if the corresponding value Δ is the smallest.

For fixed p , the only restriction on Δ is that

$$\Delta \geq \max_{1 \leq i \leq m} |\Delta y^{(i)}| = \max_{1 \leq i \leq m} |\tilde{y}^{(i)} - f(x^{(i)}, p)|.$$

Thus, for each tuple p , the smallest possible Δ is the smallest value that satisfies this inequality, i.e., it is the value

$$\Delta = \max_{1 \leq i \leq m} |\tilde{y}^{(i)} - f(x^{(i)}, p)|. \quad (11)$$

We want to select the parameters p for which the probability \mathcal{L} is the largest, i.e., equivalently, for which the value Δ described by formula (11) is the smallest. Thus, in situations where the shape of the probability distribution of the measurement error is not known, we should select the parameters p for which expression (11) is the smallest possible, i.e., for which

$$\min_p \max_{1 \leq i \leq m} |\tilde{y}^{(i)} - f(x^{(i)}, p)| \quad (12)$$

is attained.

The above result, expressed by formulas (11)–(12), is not quite trivial, and it makes sense to comment it. In principle, estimating the parameters $p = (p_1, p_2, \dots, p_l)$ of the unknown dependency $y = f(x, p)$ amounts to computing a “pseudo-solution” to the equation system (3). To do that, we construct the so-called *defect vector* which is made up of the differences $(\tilde{y}^{(i)} - f(x^{(i)}, p))$, $i = 1, 2, \dots, m$, between the left-hand and right-hand sides of system (3), and then make it as small as possible. In fact, this means that we take a norm of the defect vector, as a general measure of how large it is, and then minimize this norm. If the minimum of the norm of the defect equals zero, we get an exact solution to system (3). Otherwise, a pseudo-solution is obtained that corresponds to approximation of the data set by the best-fit line (the situation depicted at Fig. 1).

A norm of the defect vector can be constructed from the defects of separate components $(\tilde{y}^{(i)} - f(x^{(i)}, p))$, $i = 1, 2, \dots, m$, in various ways. In particular,

taking the Euclidean norm (2-norm) of the defect results in (5) and corresponds to the least squares method. Our reasoning above shows that only the max-norm (Chebyshev norm) has a clear probabilistic interpretation when we do not know of any information about the probability density functions of the error. Formula (11) expresses the essence of popular “minimax estimation” (see, e.g., [4, 7, 28]), and it is often used without any probabilistic context.

To conclude, it is worth noting that the experience of applying the Maximum Likelihood Method for joint considering interval and probabilistic approaches in the data fitting problems is not new. Zhilin in [29] used the maximum likelihood estimation in experimental study of the correlation between different types of estimates. The conclusion made in [29] on the base of extensive numerical simulation is that the non-probabilistic interval-based estimates are very good and even the best ones for uniform or nearly uniform probabilistic distributions.

3 Specificity of interval approaches

In interval approaches to the data fitting problem, the measurements errors are supposed to be bounded, and these bounds have a deterministic character now. If we introduce the intervals

$$\mathbf{y}^{(i)} \stackrel{\text{def}}{=} [\tilde{y}^{(i)} - \Delta, \tilde{y}^{(i)} + \Delta], \quad i = 1, 2, \dots, m,$$

then the inequalities (10)

$$|\Delta y^{(i)}| = |\tilde{y}^{(i)} - f(x^{(i)}, p)| \leq \Delta$$

can be equivalently rewritten as

$$f(x^{(i)}, p) \in \mathbf{y}^{(i)} \quad i = 1, 2, \dots, m. \quad (13)$$

In this case, the sense of the parameter estimation is depicted at Fig. 2: we have to construct a line from a given parametric family that “best fits” the interval data in a prescribed sense. Notice that such “best fit line” may go through each segment representing data uncertainty or, alternatively, may not do that. In the former case, when the constructed line intersects all the uncertainty intervals $(x_i, \mathbf{y}^{(i)})$, $i = 1, 2, \dots, m$, we will speak that the parameters of the line are *consistent* with the interval data.

Overall, instead of the equation system (3), we arrive at an analogous system of interval equations

$$\begin{cases} f(x^{(1)}, p) = \mathbf{y}^{(1)}, \\ f(x^{(2)}, p) = \mathbf{y}^{(2)}, \\ \vdots \quad \quad \quad \vdots \\ f(x^{(m)}, p) = \mathbf{y}^{(m)}, \end{cases} \quad (14)$$

that we have to “solve” with respect to p for parameter estimation.

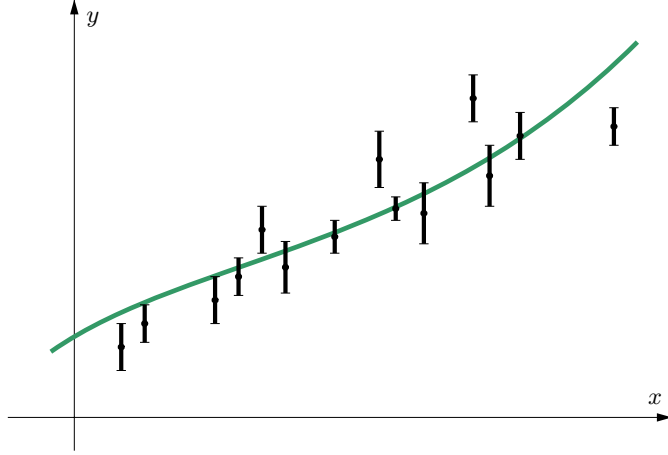


Figure 2: Data fitting problem for interval measurement data:
we have to construct a line that best fits the data.

In modern interval analysis, there exist several concepts of “solutions” and “solution sets” to interval systems of equations, and the most relevant to our task is the so-called “united solution set”. It is defined as the set of all solutions to the usual point systems of the same form with the parameters taken from the prescribed intervals. For (14), the united solution set is formally determined as

$$\Xi = \{ p \in \mathbb{R}^n \mid (\exists y^{(1)} \in \mathbf{y}^{(1)}) \dots (\exists y^{(m)} \in \mathbf{y}^{(m)}) \\ (f(x^{(1)}, p) = y^{(1)}) \dots (f(x^{(m)}, p) = y^{(m)}) \}, \quad (15)$$

being made up of solutions to the usual (point) systems of the form (3) with $y^{(1)}, y^{(2)}, \dots, y^{(m)}$ from the intervals $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(m)}$. If the solution set Ξ is not empty, it describes the set of parameters consistent with the measurement data set

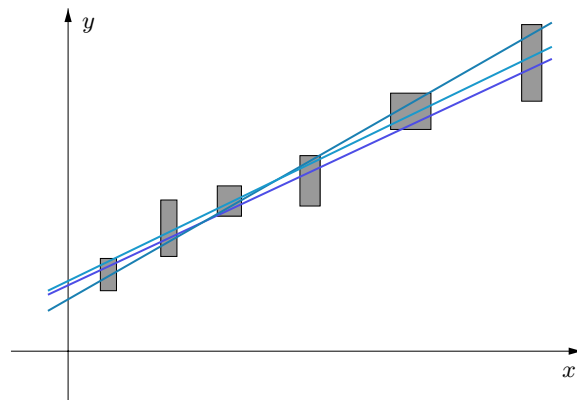
$$\begin{aligned} x^{(1)}, & \quad \mathbf{y}^{(1)}, \\ x^{(2)}, & \quad \mathbf{y}^{(2)}, \\ \dots, & \quad \dots, \\ x^{(m)}, & \quad \mathbf{y}^{(m)}, \end{aligned} \quad (16)$$

and each $p \in \Xi$ can be taken as a solution to the data fitting problem, i. e., as an estimate of the parameters. In general, set (15) may not coincide with the set of solutions to the data fitting problem, and this is why we are going to use, with respect to (15), the term *feasible parameter set* popular in the set-membership estimation theory.

The Demidenko paradox. The case of nonempty feasible parameter set Ξ is the most favorable in interval data fitting, but we do not have to reject the opposite case when the set Ξ is empty. Then there do not exist p strictly

consistent with the data set (16), but this in no way means that the data fitting problem is not solvable. This only means that its solution has the other status. For better understanding our reasoning, it is worth reminding that for usual non-interval data fitting problem, which is the limit case of the interval problem statement, the consistency between the data and parameters in the sense of our definition is mostly unachieved, being a very rare and exceptional event.

Bad accuracy and wide uncertainty intervals enable one to construct many models consistent with the data.



For better accuracy and narrow uncertainty intervals, there might not exist a model consistent with the data.

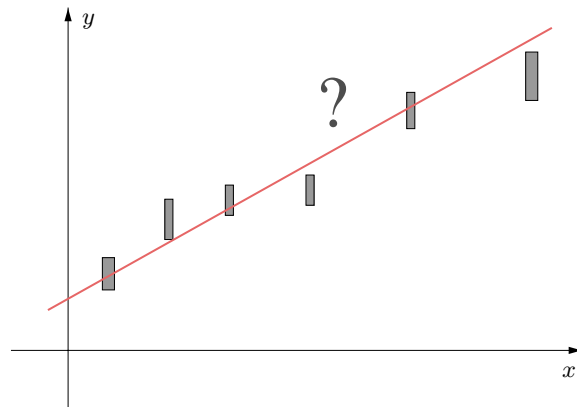


Figure 3: The essence of the Demidenko paradox.

Moreover, if possible emptiness of the feasible parameter set is not taken into account, we can come to a paradox that was first noticed by E. Demidenko in the note [2]. Its essence can be described in short by the phrase “the better, the

worse". Specifically, the more accurate are our measurements and more narrow the uncertainty intervals $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(m)}$, the more likely that the feasible parameter set Ξ is empty and the data fitting problem looks "unsolvable". Conversely, if we organize crude measurement and our errors are large, the intervals $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(m)}$ are wide, but this enlarges the feasible parameter set Ξ , and it is easier to take an estimate of the parameters from it.

In fact, the Demidenko paradox is based on the assumptions that

- (i) the intervals $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(m)}$ represent rigorous bounds on the actual values of our measurements and falling outside them is impossible,
- (ii) the function f exactly represents the functional dependency between x and y .

Any one or both of the above points can be violated in real life problems. For example, we can consider the intervals $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(m)}$ as "soft" bounds on the respective measurements: these bounds are not obligatory for the actual values of $y^{(1)}, y^{(2)}, \dots, y^{(m)}$, but rather serve as estimates of their ranges. Our data can have outliers that spoil the ideal measurements picture. The function f may be an approximation to the actual dependency between x and y taken (e.g., from physical, chemical, economic, etc., reasons) merely as a dominant trend. And so on.

Therefore, the emptiness of the feasible parameter set should not be perceived as a signal for terminating the solution of the data fitting problem. We have to accept the possibility for the feasible parameter set to be empty, the possibility that there are no parameters strictly consistent with the processed data. Even more, our approach to the data fitting and parameter estimation for interval data should successfully cope with this "double-layer character" of the problem, providing a uniform treatment of the two situations when the feasible parameter set is either empty or nonempty. This is a specific feature of the data processing under essentially interval uncertainty.

Minimax estimation for interval data. Let us turn to the minimax estimation (11)–(12) derived from the Maximum Likelihood Method in the previous section.

Inclusions (13) mean that if the upper bound on the absolute value of the measurement error is equal to Δ , then the actual value $f(x^{(i)}, p)$ of the quantity y agrees with the measurement result $\tilde{y}^{(i)}$. In these terms, expression (11) is the smallest bound Δ for which all the measurement results agree with the actual values; the selected value p is the one for which the difference between the measurement results and the actual values is the smallest possible. Overall, a natural probabilistic approach to estimating parameters, the Maximum Likelihood Method, leads to an "interval-ready" formula (12) for parameter estimation, which can be directly applied to data fitting under interval uncertainty. However, some points should be tuned to this specific case.

Under certain conditions, we may have *a priori* bounds on some parameters

p_j , i. e., the values \underline{p}_j and \bar{p}_j for which the actual (unknown) value p_j satisfies the inequality $\underline{p}_j \leq p_j \leq \bar{p}_j$. Then we should minimize expression (11) for Δ only over the tuples $p = (p_1, \dots, p_l)$ which satisfy these inequalities.

Very often, we can be interested only in some of the parameters. Without losing generality, let us assume that we have to find the values of the parameters p_1, \dots, p_ℓ for some $\ell < l$, while the parameters $p_{\ell+1}, \dots, p_l$ are not interesting for us. Then we can use the mathematical fact that

$$\begin{aligned} \min_{p_1, p_2, \dots, p_l} \max_{1 \leq i \leq m} |\tilde{y}^{(i)} - f(x^{(i)}, p)| \\ = \min_{p_1, p_2, \dots, p_\ell} \min_{p_{\ell+1}, \dots, p_l} \max_{1 \leq i \leq m} |\tilde{y}^{(i)} - f(x^{(i)}, p)|. \end{aligned}$$

In other words, minimizing expression (11) over all the tuples (p_1, \dots, p_l) is equivalent to minimizing an auxiliary expression

$$\Phi(p_1, p_2, \dots, p_\ell) \stackrel{\text{def}}{=} \min_{p_{\ell+1}, \dots, p_l} \max_{1 \leq i \leq m} |\tilde{y}^{(i)} - f(x^{(i)}, p)| \quad (17)$$

over the parameters p_1, \dots, p_ℓ . This trick may prove very helpful in situations when we want to convert some uncertain factors into parameters.

Plateau-like extremum region. Yet another specific feature of the data fitting problem under essential interval uncertainty is that its feasible parameter set Ξ may be a “real” non-singleton set made up of infinitely many points. In fact, such feasible parameter sets can have nonempty interior and nonzero measure, so that the solution to the minimization problem (17) is achieved at a whole region in \mathbb{R}^ℓ . In mathematical terms, this feature is expressed by the fact that the feasible parameter set, i. e., the united solution set to the interval equation system (14) that we have to solve for constructing the best fit line, is a solid set. For every point of such feasible parameter set, the norm of the defect $\max_i |\tilde{y}^{(i)} - f(x^{(i)}, p)|$ is precisely zero. So, the objective function (17) attains its minimum at a zero level plateau like that depicted at the right-hand side of Fig. 4. That is very disadvantageous for our minimization problem, both in itself and due to practical reasons.

First of all, checking whether a point belongs to the feasible parameter set amounts to testing unstable equality of the objective function (17) to zero. Furthermore, the points of the feasible parameter set Ξ are, in reality, different from each other with respect to the data fitting problem. In particular, it makes sense to distinguish between points of the boundary and interior of the set Ξ , since the latter are stable under data perturbations. Choosing the interior points of the feasible parameter set Ξ as parameter estimates would be more preferable, but the objective function (17) does not allow us to do that.

By and large, it is intuitively clear that the estimates taken somewhere in the “middle” of the feasible parameter set Ξ are “more robust” and even “more probable”. This is confirmed by computational experiments from [20] demonstrating that the uniform probabilistic distributions over input variables

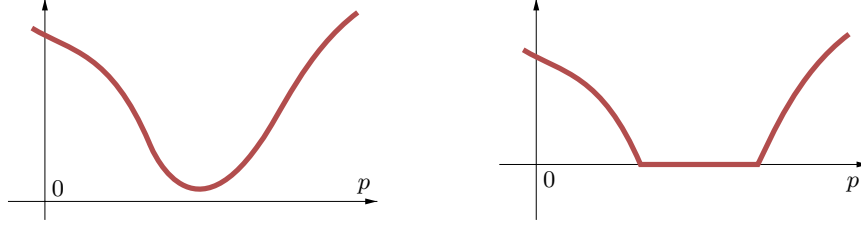


Figure 4: Various configurations of the minima in (12) and (17) depending on their signs.

of a function lead to a non-uniform distribution over its range of values, with the maximum probability density in the central area.

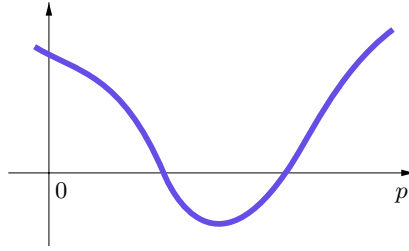


Figure 5: Desirable configuration of the negative minimum.

The situation with the zero-level plateau in the minimax estimation (Fig. 4, right) should be corrected, and some parameter estimation methods really do that during their execution. For example, in the uncertainty center method from [16, 29], the estimate is taken as the middle point of the outer box for the feasible parameter set. Ideally, it would be desirable to have the configuration of the minimum for (17) like that depicted at Fig. 5, where the objective function helps distinguishing the interior points from the feasible parameter set. In the last section, we are going to show how one can naturally improve, in the above sense, the minimax estimation for interval data fitting with the linear functional dependency.

4 Straight line fitting

As an example of applying the ideas developed in the preceding sections, let us consider the problem of solving an $m \times n$ -system of interval linear equations that arises in the interval data fitting problem for linear functional dependency. It is often called “linear regression problem”, etc.

Suppose that we are given a system of linear equations

$$\sum_{j=1}^n a_{ij} z_j - b_i = 0, \quad 1 \leq i \leq m, \quad (18)$$

under interval uncertainty, when we only know that a_{ij} and b_i belong to intervals $\mathbf{a}_{ij} = [\underline{a}_{ij}, \bar{a}_{ij}]$ and $\mathbf{b}_i = [\underline{b}_i, \bar{b}_i]$ respectively. One can think that an interval $m \times n$ -matrix $\mathbf{A} = (\mathbf{a}_{ij})$ and an interval m -vector $\mathbf{b} = (\mathbf{b}_i)$ are given that specify an interval linear system of algebraic equations $\mathbf{A}z - \mathbf{b} = 0$, which is equivalent to $\mathbf{A}z = \mathbf{b}$. We are interested in the values of the variables z_1, \dots, z_n that “best fit” equalities (18) under the uncertainty represented by the data set $\mathbf{a}_{ij}, \mathbf{b}_i, i = 1, \dots, m, j = 1, \dots, n$ (see Fig. 6).

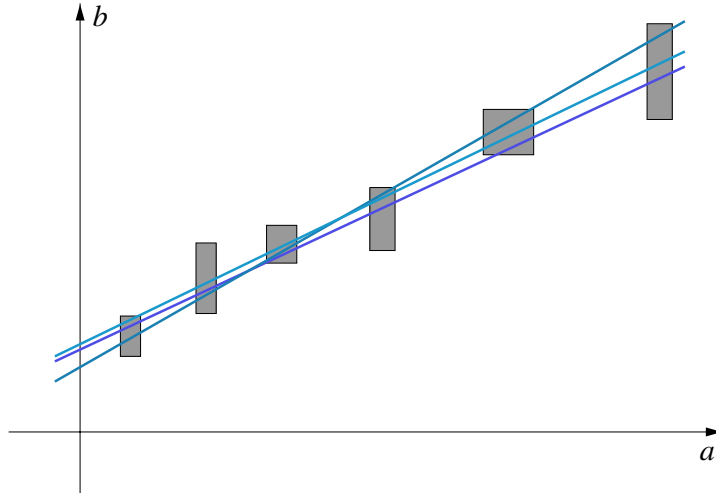


Figure 6: Straight line fitting for intervally uncertain data.

The problem is in close relationship with the general data fitting problem we have considered in the previous sections. One can see here m observations based on which we would like to determine the desired parameters z_1, z_2, \dots, z_n . As for the interval data $\mathbf{a}_{ij}, \mathbf{b}_i, i = 1, \dots, m, j = 1, \dots, n$, we will not consider them as measurement results, but represent them as additional parameters using the technique elaborated at the end of the previous section. So, in our problem statement, the number ℓ of the sought-for parameters is equal to n , and these parameters p_1, p_2, \dots, p_ℓ from formulation (17) coincide with z_1, z_2, \dots, z_n from (18). In addition to these $\ell = n$ parameters of interest, we also have the following auxiliary parameters:

- $m \cdot n$ parameters $a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n$,
for which we are given the bounds \underline{a}_{ij} and \bar{a}_{ij} ;
- m parameters $b_i, 1 \leq i \leq m$,
for which we are given the bounds \underline{b}_i and \bar{b}_i .

In total, we consider $mn + m + n$ parameters, which can be represented as a long composite vector $p = (z, A, b)$, $A = (a_{ij})$, $b = (b_i)$. As an input $x^{(i)}$ of the function that describes the i -th observation, we can simply take $x^{(i)} = i$. Overall, the model-based value $f(x, p)$ of the observed functional dependency is exactly

$$f(x^{(i)}, p) = f(i, z, A, b) = \sum_{j=1}^n a_{ij} z_j - b_i,$$

while the observed value $y^{(i)}$ is always equal to 0.

For this problem, formulation (17) originated from the maximum likelihood approach leads to selecting the values z_1, \dots, z_n for which the following expression is minimized:

$$\Phi(z) = \min_{a_{ij} \in \mathbf{a}_{ij}, b_i \in \mathbf{b}_i} \max_{1 \leq i \leq m} \left| \sum_{j=1}^n a_{ij} z_j - b_i \right|.$$

Here, each of the parameters a_{ij} , $j = 1, 2, \dots, n$, and b_i occurs only in the i -th expression under $\min_{1 \leq i \leq m}$. So, minimizing $\max_{1 \leq i \leq m}$ is equivalent to minimizing the corresponding expression for each i :

$$\Phi(z) = \max_{1 \leq i \leq m} \min_{a_{ij} \in \mathbf{a}_{ij}, b_i \in \mathbf{b}_i} \left| \sum_{j=1}^n a_{ij} z_j - b_i \right|. \quad (19)$$

Next, the inner minimum in (19) can be found explicitly. The range of values of the linear expression $\sum_{j=1}^n a_{ij} z_j - b_i$ over all $a_{ij} \in \mathbf{a}_{ij}$ and $b_i \in \mathbf{b}_i$ coincides with its natural interval extension due to the fundamental theorem of interval arithmetic [11, 13]. Therefore,

$$\min_{a_{ij} \in \mathbf{a}_{ij}, b_i \in \mathbf{b}_i} \left| \sum_{j=1}^n a_{ij} z_j - b_i \right| = \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \mathbf{b}_i \right\rangle,$$

where $\langle \cdot \rangle$ means mignitude of the interval (see [13, 25]), the smallest distance from points of the interval to zero:

$$\langle \mathbf{u} \rangle = \begin{cases} \min\{|\overline{\mathbf{u}}|, |\underline{\mathbf{u}}|\}, & \text{if } 0 \notin \mathbf{u}, \\ 0, & \text{otherwise.} \end{cases}$$

Overall, we arrive at the following optimization problem:

$$\text{find } \min_{z_1, z_2, \dots, z_n} \Phi(z_1, z_2, \dots, z_n), \quad (20)$$

where

$$\Phi(z) = \Phi(z_1, z_2, \dots, z_n) = \max_{1 \leq i \leq m} \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \mathbf{b}_i \right\rangle. \quad (21)$$

We would remind that, for interval systems of linear algebraic equations $\mathbf{A}z = \mathbf{b}$, the *united solution set* (also called just *solution set*) is defined as the set of all solutions to the point systems $Az = b$ with $A \in \mathbf{A}$ and $b \in \mathbf{b}$. In formal terms, the united solution set is

$$\Xi(\mathbf{A}, \mathbf{b}) \stackrel{\text{def}}{=} \{z \in \mathbb{R}^n \mid (\exists A \in \mathbf{A})(\exists b \in \mathbf{b})(Az = b)\}. \quad (22)$$

As before, the set $\Xi(\mathbf{A}, \mathbf{b})$ is the “feasible parameter set” for our data fitting problem under interval uncertainty, i.e., the set where the strict consistency takes place between the data and the parameters of the straight line.

If the set $\Xi(\mathbf{A}, \mathbf{b})$ is empty, then the minimum in the optimization problem (21)–(20) is strictly positive, which corresponds to the left-hand side picture at Fig. 4. A specific feature of our situation is that the objective function $\Phi(z)$ is nonsmooth (piecewise linear), and its minimum is attained at a sharp point of its graph. If the set $\Xi(\mathbf{A}, \mathbf{b})$ is nonempty, then the minimum in (20) is zero, which corresponds to the right-hand side picture at Fig. 4. The minimum is then attained at any point of the set $\Xi(\mathbf{A}, \mathbf{b})$, i.e., the united solution set to the interval linear system $\mathbf{A}z = \mathbf{b}$.

Taking into account the consideration of the preceding section, the next question arises: how can we correct the construction to improve the objective function and avoid plateau-like minima sets at the zero level in the case of nonempty feasible parameter sets? This can be done in a natural way we describe in the following section.

5 Maximum consistency method

To bring the results of the preceding section into correlation with known interval-related estimation techniques, it is necessary to transform the problem statement (20)–(21). We will need the following property of the magnitude (see [13, 25]): for any intervals \mathbf{u} and \mathbf{v} ,

$$\langle \mathbf{u} \pm \mathbf{v} \rangle \geq \langle \mathbf{u} \rangle - |\mathbf{v}|. \quad (23)$$

The equality instead of the nonstrict inequality (23) holds true, in particular, if $\langle \mathbf{u} \rangle \geq |\mathbf{v}|$ and the interval \mathbf{v} is balanced (symmetric with respect to zero), i.e., has the form $\mathbf{v} = [-\tilde{v}, \tilde{v}]$ for a nonnegative \tilde{v} . The short proof is as follows:

$$\begin{aligned} \langle \mathbf{u} \pm \mathbf{v} \rangle &= \min_{\substack{u \in \mathbf{u} \\ v \in \mathbf{v}}} |u \pm v| \geq \min_{\substack{u \in \mathbf{u} \\ v \in \mathbf{v}}} (|u| - |v|) \\ &= \min_{u \in \mathbf{u}} |u| - \max_{v \in \mathbf{v}} |v| = \langle \mathbf{u} \rangle - |\mathbf{v}|. \end{aligned}$$

If \mathbf{v} is a balanced interval, then, for any $u \in \mathbf{u}$, there exists such $v \in \mathbf{v}$ that the equality $|u \pm v| = |u| - |v|$ holds, no matter what the interval \mathbf{u} is. Hence, the equality instead of the inequality in the above relations really takes place.

Recognizing functional Uni. For each index $i = 1, 2, \dots, m$, we can evaluate subexpressions of (21) as follows:

$$\begin{aligned} \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \mathbf{b}_i \right\rangle &= \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \text{mid } \mathbf{b}_i - [-1, 1] \cdot \text{rad } \mathbf{b}_i \right\rangle \\ &\geq \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \text{mid } \mathbf{b}_i \right\rangle - |[-1, 1] \cdot \text{rad } \mathbf{b}_i| \\ &= \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \text{mid } \mathbf{b}_i \right\rangle - \text{rad } \mathbf{b}_i \end{aligned}$$

due to property (23). Notice that the equality instead of nonstrict inequality is valid for the case

$$\left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \text{mid } \mathbf{b}_i \right\rangle \geq \text{rad } \mathbf{b}_i.$$

Since the mignitude is always nonnegative, we have

$$\left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \mathbf{b}_i \right\rangle = \max \{ \psi_i(z), 0 \},$$

where

$$\psi_i(z) \stackrel{\text{def}}{=} \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \text{mid } \mathbf{b}_i \right\rangle - \text{rad } \mathbf{b}_i.$$

In summary,

$$\Phi(z) = \max_{1 \leq i \leq m} \max \{ \psi_i(z), 0 \},$$

and, reversing the order of the two maximum operations, we get

$$\Phi(z) = \max \{ \Psi(z), 0 \}, \quad (24)$$

where

$$\Psi(z) \stackrel{\text{def}}{=} \max_{1 \leq i \leq m} \psi_i(z) = \max_{1 \leq i \leq m} \left\{ \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \text{mid } \mathbf{b}_i \right\rangle - \text{rad } \mathbf{b}_i \right\}.$$

One can see that $\Psi(z)$ differs only by the opposite sign from the so-called *recognizing functional* of the united solution set $\Xi(\mathbf{A}, \mathbf{b})$ to the interval linear system $\mathbf{A}z = \mathbf{b}$, which is defined as

$$\text{Uni}(z) \stackrel{\text{def}}{=} \min_{1 \leq i \leq m} \left\{ \text{rad } \mathbf{b}_i - \left\langle \text{mid } \mathbf{b}_i - \sum_{j=1}^n \mathbf{a}_{ij} z_j \right\rangle \right\}.$$

It was proposed and studied in [21, 24]. Thus, the minimized functional $\Phi(z) = \max \{ \Psi(z), 0 \}$ is related to the functional $\text{Uni}(z)$ as

$$\Phi(z) = \max \{ -\text{Uni}(z), 0 \} = -\min \{ \text{Uni}(z), 0 \}.$$

Recognizing functional Uss. We can perform our transformations of expression (21) for Φ in another way. For each index $i = 1, 2, \dots, m$, there holds

$$\begin{aligned}
& \left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \mathbf{b}_i \right\rangle \\
&= \left\langle \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij} + [-1, 1] \cdot \text{rad } \mathbf{a}_{ij}) z_j - \text{mid } \mathbf{b}_i - [-1, 1] \cdot \text{rad } \mathbf{b}_i \right\rangle \\
&= \left\langle \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j + \sum_{j=1}^n [-1, 1] (\text{rad } \mathbf{a}_{ij}) z_j - \text{mid } \mathbf{b}_i + [-1, 1] \cdot \text{rad } \mathbf{b}_i \right\rangle \\
&\quad \text{due to distributivity of multiplication with respect} \\
&\quad \text{to addition for the common point multipliers } z_j \\
&\geq \left\langle \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j - \text{mid } \mathbf{b}_i \right\rangle - \left| \sum_{j=1}^n [-1, 1] (\text{rad } \mathbf{a}_{ij}) z_j + [-1, 1] \cdot \text{rad } \mathbf{b}_i \right| \\
&\quad \text{due to property (23) of the mignitude} \\
&= \left| \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j - \text{mid } \mathbf{b}_i \right| - \sum_{j=1}^n (\text{rad } \mathbf{a}_{ij}) |z_j| - \text{rad } \mathbf{b}_i,
\end{aligned}$$

since all the intervals $[-1, 1] (\text{rad } \mathbf{a}_{ij}) z_j$, $[-1, 1] \cdot \text{rad } \mathbf{b}_i$ are symmetric with respect to zero and $(\sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j - \text{mid } \mathbf{b}_i)$ is a point, not interval. Therefore,

$$\left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \mathbf{b}_i \right\rangle \geq \left| \text{mid } \mathbf{b}_i - \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j \right| - \sum_{j=1}^n (\text{rad } \mathbf{a}_{ij}) |z_j| - \text{rad } \mathbf{b}_i,$$

and the equality instead of nonstrict inequality is valid for the case

$$\left| \text{mid } \mathbf{b}_i - \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j \right| \geq \sum_{j=1}^n (\text{rad } \mathbf{a}_{ij}) |z_j| + \text{rad } \mathbf{b}_i,$$

Since the mignitude is always nonnegative, we have

$$\left\langle \sum_{j=1}^n \mathbf{a}_{ij} z_j - \mathbf{b}_i \right\rangle = \max \{ v_i(z), 0 \},$$

where

$$v_i(z) \stackrel{\text{def}}{=} \left| \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j - \text{mid } \mathbf{b}_i \right| - \sum_{j=1}^n (\text{rad } \mathbf{a}_{ij}) |z_j| - \text{rad } \mathbf{b}_i.$$

In summary,

$$\Phi(z) = \max_{1 \leq i \leq m} \max \{ v_i(z), 0 \},$$

and, reversing the order of the two maximum operations, we get

$$\Phi(z) = \max \{ \Upsilon(z), 0 \}, \quad (25)$$

where

$$\Upsilon(z) \stackrel{\text{def}}{=} \max_{1 \leq i \leq m} \left\{ \left| \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j - \text{mid } \mathbf{b}_i \right| - \sum_{j=1}^n (\text{rad } \mathbf{a}_{ij}) |z_j| - \text{rad } \mathbf{b}_i \right\},$$

One can see that $\Upsilon(z)$ differs only by the opposite sign from the so-called *recognizing functional* of the united parameter set $\Xi(\mathbf{A}, \mathbf{b})$ to the interval linear system $\mathbf{A}z = \mathbf{b}$, which is defined as

$$\text{Uss}(z) = \min_{1 \leq i \leq m} \left\{ \text{rad } \mathbf{b}_i + \sum_{j=1}^n (\text{rad } \mathbf{a}_{ij}) |z_j| - \left| \text{mid } \mathbf{b}_i - \sum_{j=1}^n (\text{mid } \mathbf{a}_{ij}) z_j \right| \right\}.$$

It was introduced and investigated in [22, 23, 26]. Thus, the minimized functional $\Phi(z) = \max \{ \Upsilon(z), 0 \}$ is related to the functional $\text{Uss}(z)$ as

$$\Phi(z) = \max \{ -\text{Uss}(z), 0 \} = -\min \{ \text{Uss}(z), 0 \}.$$

The sense and general properties of recognizing functionals. The recognizing functionals Uni and Uss were introduced in [21, 22, 23, 24, 26] to give a numerical measure that characterizes compatibility of interval linear equation systems.

According to the result from [24], a point $z \in \mathbb{R}^n$ belongs to the united solution set $\Xi(\mathbf{A}, \mathbf{b})$ if and only if

$$\langle \mathbf{A}z - \text{mid } \mathbf{b} \rangle \leq \text{rad } \mathbf{b},$$

where the mignitude is applied component-wise. So, the amount to which the right-hand side of the above inequality exceeds the left-hand side, i.e., the difference

$$\text{rad } \mathbf{b} - \langle \mathbf{A}z - \text{mid } \mathbf{b} \rangle, \quad (26)$$

can be taken as a numerical consistency measure of the parameter vector z and the data \mathbf{A}, \mathbf{b} . In order to construct a unified scalar characteristic, we convolve vector (26) by taking the minimum of its components. This way, the functional Uni is obtained.

The membership of a point $z \in \mathbb{R}^n$ in the united solution set $\Xi(\mathbf{A}, \mathbf{b})$ to the interval linear equation system $\mathbf{A}z = \mathbf{b}$ is equivalent to the nonnegativeness of the functional Uni in z :

$$z \in \Xi(\mathbf{A}, \mathbf{b}) \quad \Longleftrightarrow \quad \text{Uni}(z, \mathbf{A}, \mathbf{b}) \geq 0.$$

Well-known Oettli-Prager inequality may serve as a base for yet another consistency measure. Namely, a point $z \in \mathbb{R}^n$ is known (see [13]) to belong to the united solution set $\Xi(\mathbf{A}, \mathbf{b})$ if and only if

$$|\text{mid } \mathbf{A} \cdot z - \text{mid } \mathbf{b}| \leq \text{rad } \mathbf{A} \cdot |z| + \text{rad } \mathbf{b},$$

where the magnitude (absolute value) is applied componentwise. So, the amount to which the right-hand side of the above Oettli-Prager inequality exceeds the left-hand side, i. e., the difference

$$\text{rad } \mathbf{A} \cdot |z| + \text{rad } \mathbf{b} - |\text{mid } \mathbf{A} \cdot z - \text{mid } \mathbf{b}|, \quad (27)$$

can be taken as a numerical consistency measure of the parameter vector z and the data \mathbf{A}, \mathbf{b} . In order to construct a unified scalar characteristic, we convolve vector (27) by taking the minimum of its components. This way, the functional Uss is obtained.

The membership of a point $z \in \mathbb{R}^n$ in the united solution set $\Xi(\mathbf{A}, \mathbf{b})$ to the interval linear equation system $\mathbf{A}z = \mathbf{b}$ is equivalent to the nonnegativeness of the functional Uss in z :

$$z \in \Xi(\mathbf{A}, \mathbf{b}) \quad \Longleftrightarrow \quad \text{Uss}(z, \mathbf{A}, \mathbf{b}) \geq 0.$$

Below, we give a short survey of the properties of the recognizing functionals Uni and Uss as they are presented in the works [21, 22, 23, 24, 26].

The functionals Uni and Uss are concave functions of z in each orthant of the space \mathbb{R}^n . If, additionally, the matrix \mathbf{A} is such that its columns with the numbers from the set $J = \{j_1, j_2, \dots, j_r\}$, $r \leq n$, are interval, while the rest columns are entirely point (noninterval), then the functionals Uni and Uss are concave on each of the 2^r sets of the form $\{z \in \mathbb{R}^n \mid z_j \geq 0, j \in J\}$, where “ \geq ” denotes one of the relations “ \geq ” or “ \leq ”.

The functionals Uni and Uss are polyhedral, that is, their graphs are made up of finite numbers of hyperplane pieces. Additionally, the functional Uni reaches a finite maximum with respect to z over the entire space \mathbb{R}^n .

On the other hand, for interval linear systems with nonempty united solution set, the recognizing functionals can help distinguishing interior and boundary of the solution set, etc. In particular, if $\text{Uni}(z) > 0$ or $\text{Uss}(z) > 0$, then z is a point from the topological interior of the solution set, which immediately follows from the continuity of Uni and Uss. Under certain additional requirements, the converse is also true [22, 23]. Overall, the recognizing functionals prove to be very useful tool for examination of the “fine structure” of the solution sets.

Both functionals, Uss and Uni, have similar properties, and each one of them is intended for its own problems. In particular, the functional Uni takes into account the right-hand side of the system to a greater extent than the functional Uss. So, Uni may be useful in the problems where the corresponding variables are of much importance.

The recognizing functionals may be considered as a way to find a solution to a system of interval equations in situations when we have underestimated the

uncertainty with which we know the parameters. As a result, the corresponding system of interval linear equations has no solutions, i.e., its united solution set is empty. Then one can use the vector for which the functional Uni or Uss attains its largest possible value as a pseudo-solution with the minimum possible inconsistency.

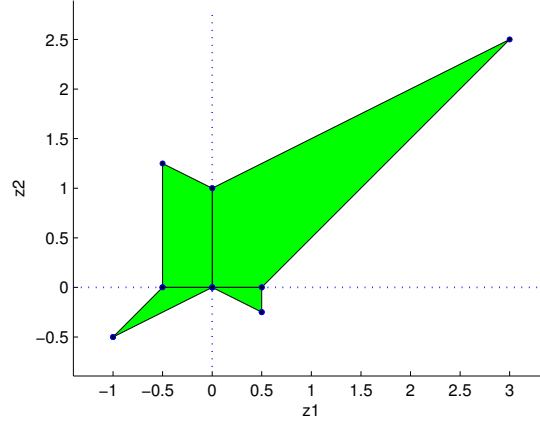


Figure 7: Solution set to the interval linear system (28).

An example of interval linear system and its recognizing functionals.

As an example, we consider the interval linear algebraic system

$$\begin{pmatrix} [2, 4] & [-2, 0] \\ [-1, 1] & [2, 4] \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} [-1, 1] \\ [0, 2] \end{pmatrix}, \quad (28)$$

and its solution set is depicted at Fig. 7.³

The pictures at Fig. 8 show graphs of the recognising functionals Uni and Uss of the solution set for system (28). The difference between the two pictures is small, but it exists: at the graph of Uss , the maximum is a “sharp peak” of the height 1.44 attained at the point $(0.1, 0.33)^\top$, while the graph of Uni has a flat maximum region at the height 1 around the argument point $(0.5, 0.5)^\top$.

For negative values, the functionals Uni and Uss totally coincide with each other, as was substantiated at the beginning of the section. But for nonnegative values, each one of them shows its own “consistency measure”, according to its origin and construction, for the data given by the interval linear system (28).

Maximum consistency method. In view of the above results, the situation when the interval system has empty solution set corresponds to the case when

³The picture has been obtained with the use of the visualization package `IntLinInc2D` [19].

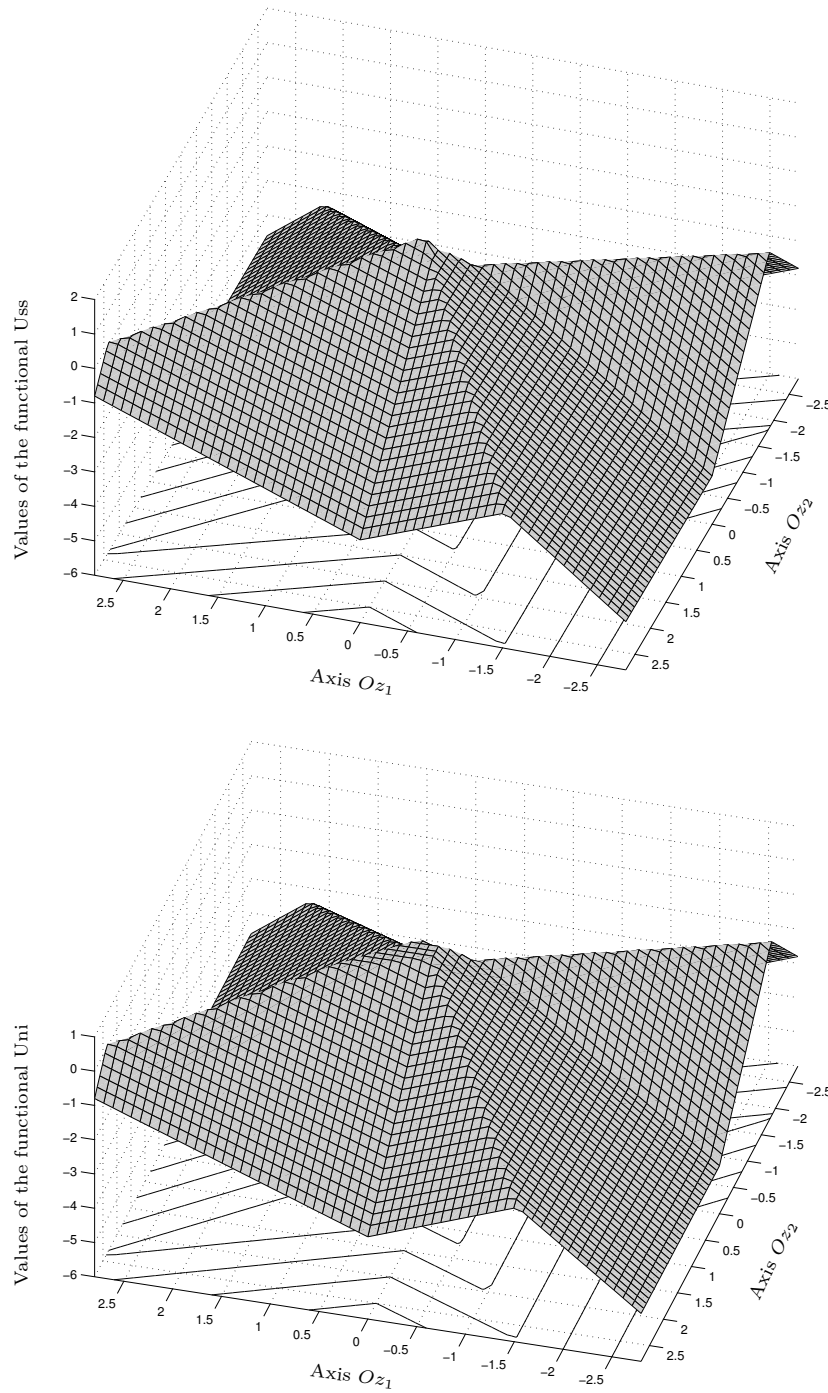


Figure 8: Graphs of the recognizing functional Uni and Uss for system (28).

$\Psi(z) > 0$ and $\Upsilon(z) > 0$ for all z , i. e., when

$$\text{Uni}(z) = -\Psi(z) < 0 \quad \text{and} \quad \text{Uss}(z) = -\Upsilon(z) < 0 \quad \text{for all } z \in \mathbb{R}^n.$$

Then we have

$$-\Phi(z) = \min\{\text{Uni}(z), 0\} = \text{Uni}(z)$$

and, at the same time,

$$-\Phi(z) = \min\{\text{Uss}(z), 0\} = \text{Uss}(z).$$

Hence, minimizing $\Phi(z)$ in our problem (20)–(21) is equivalent to unconstrained maximization of $\text{Uni}(z)$ or $\text{Uss}(z)$.

Data fitting (parameter estimation) approach based on maximization of the recognizing functional for the solution set is called the *maximum consistency method*. As an estimate of parameters, we take the argument of the recognizing functional where its maximum is attained. The term “maximum consistency” is justified by the fact that the recognizing functionals, both Uni and Uss , show, as was demonstrated earlier, the degree of consistency between the data \mathbf{A} , \mathbf{b} and parameters z_1, z_2, \dots, z_n of the regression line. If the maximum of the recognizing functional is nonnegative, then our estimate is consistent with the input data in the sense of the definition from Section 3. If the maximum of the recognizing functional is negative, no consistency can be achieved between any parameter estimate and the data, but the point we have found is the best one since it provides the minimum possible value of “inconsistency”. In the limit case of non-interval (point) data, maximization of any recognizing functional is equivalent to minimizing the Chebyshev norm (maximum-norm) of the defect of the equation system (18) (see [26]). The Maximum Consistency Method thus turns into the so-called “Chebyshev data smoothing” which is successfully applied in data processing.

The idea of the Maximum Consistency Method comes from the interval analysis, while the idea of minimizing the functional $\Phi(z)$ in (20)–(21) comes from a natural probabilistic approach, the Maximum Likelihood Method. In the case of empty feasible parameter set, the estimates produced by these two methods coincide. The Maximum Consistency Method, thereby, gets a probabilistic justification.

To improve the situation with the zero-level minimum plateau in problem (20)–(21), it makes sense to change the objective function Φ within the solution set $\Xi(\mathbf{A}, \mathbf{b})$, if it is not empty. We can take either Ψ instead of $\Phi = \max\{\Psi, 0\}$, or Υ instead of $\Phi = \max\{\Upsilon, 0\}$, thus removing the cut-off with zero in both cases. This will delete the zero plateau at the minimum of Φ for nonempty feasible parameter set. On the other hand, this idea has been already implemented in the Maximum Consistency Method, where one maximizes the recognizing functionals $\text{Uni} = -\Psi$ or $\text{Uss} = -\Upsilon$. In point of fact, the Maximum Consistency Method is a uniform numerical procedure in which an automatic switching takes place, depending on a specific situation, between the two dissimilar parts of the data fitting problem under interval uncertainty that correspond to empty and nonempty feasible parameter set.

References

- [1] B. CHOKR AND V. KREINOVICH, How far are we from complete knowledge: complexity of knowledge acquisition in Dempster-Shafer approach. In: R.R. Yager, J.Kacprzyk, and M.Pedrizzi (eds.), *Advances in the Dempster-Shafer Theory of Evidence*. Wiley, New York, 1994, pp. 555–576.
- [2] E.Z. DEMIDENKO, Comment II to the Paper of A.P. Voshchinin, A.F. Bochkov, and G.R. Sotirov “A method of data analysis under interval nonstatistical error”, *Industrial Laboratory*, vol. 56 (1990), No. 7, pp. 83–84. (in Russian) Accessible at <http://www.nsc.ru/interval/Library/Thematic/DataProcs/VosBochSoti.pdf>
- [3] P. C. HANSEN, V. PEREYRA, AND G. SCHERER, *Least Squares Data Fitting with Applications*. The John Hopkins University Press, Baltimore, Maryland, 2013.
- [4] L. JAULIN, M. KIEFFER, O. DIDRIT, AND E. WALTER, *Applied Interval Analysis*. Springer, London, 2001.
- [5] E. T. JAYNES AND G. L. BRETTHORST, *Probability Theory: the Logic of Science*. Cambridge University Press, Cambridge, UK, 2003.
- [6] L. V. KANTOROVICH, On some new approaches to numerical methods and processing observation data. *Siberian Math. Journal*, vol. 3 (1962), No. 5, pp. 701–709. (in Russian) Electronic version is accessible at <http://www.nsc.ru/interval/Introduction/Kantorovich62.pdf>
- [7] E. L. LEHMANN AND G. CASELLA, *Theory of Point Estimation*. Springer, New York, 2003.
- [8] M. MILANESE, J. NORTON, H. PIET-LAHANIER, AND E. WALTER, EDS., *Bounding Approaches to System Identification*. Plenum Press, New York, 1996.
- [9] R. B. MILLAR, *Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB*. Wiley, Chichester, UK, 2011.
- [10] R. E. MOORE, Parameter sets for bounded-error data. *Mathematics and Computers in Simulation*, vol. 34 (1992), No. 2, pp. 113–119.
- [11] R. E. MOORE, R. B. KEARFOTT, M. J. CLOUD, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
- [12] H. T. NGUYEN, V. KREINOVICH, B. WU, AND G. XIANG, *Computing Statistics under Interval and Fuzzy Uncertainty*. Springer Verlag, Berlin, Heidelberg, 2012.
- [13] A. NEUMAIER, *Interval Methods for Systems of Equations*. Cambridge University Press, Cambridge, 1990.

- [14] P. V. NOVITSKII AND I. A. ZOGRAPH, *Estimating the Measurement Errors*, Energoatomizdat, Leningrad, 1991. (in Russian)
- [15] A. I. ORLOV, How often are the observations normal? *Industrial Laboratory*, vol. 57 (1991), No. 7, pp. 770–772. (in Russian)
- [16] N. M. OSKORBIN, A. V. MAKSIMOV, AND S. I. ZHILIN, Construction and analysis of empirical dependencies by uncertainty center method, *The News of Altai State University*, vol. 5 (1998), No. 1, pp. 37–40. (in Russian)
- [17] S. G. RABINOVICH, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.
- [18] F. C. SCHWEPPE, Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Trans. Autom. Control*, AC-13 (1968), No. 1, pp. 22–28.
- [19] I. A. SHARAYA, IntLinInc2D — a software package for visualization of the solution sets to interval linear systems of relations with two unknowns. Version for MATLAB. Available at <http://www.nsc.ru/interval/sharaya/index.html#codes>
- [20] S. P. SHARY, Interval analysis or Monte-Carlo methods? *Computational Technologies*, vol. 12 (2007), No. 1, pp. 103–115. (in Russian) Accessible at <http://www.ict.nsc.ru/jct/getfile.php?id=946>
- [21] S. P. SHARY, Solvability of interval linear equations and data analysis under uncertainty. *Automation and Remote Control*, vol. 73 (2012), No. 2, pp. 310–322. DOI: 10.1134/S0005117912020099
- [22] S. P. SHARY AND I. A. SHARAYA, Recognition of solvability of interval equations and its application to data analysis. *Computational Technologies*, vol. 18 (2013), No. 3, pp. 80–109. (in Russian) Accessible at <http://www.ict.nsc.ru/jct/getfile.php?id=1551>
- [23] S. P. SHARY AND I. A. SHARAYA, On solvability recognition for interval linear systems of equations, *Optimization Letters*, published online May 6, 2015, pp. 1–14. DOI: 10.1007/s11590-015-0891-6
- [24] S. P. SHARY, New characterizations for the solution set to interval linear systems of equations. *Applied Mathematics and Computation*, vol. 265 (2015), pp. 570–573. DOI: 10.1016/j.amc.2015.05.029
- [25] S. P. SHARY, *Finite-Dimensional Interval Analysis*. Institute of Computational Technologies, Novosibirsk, 2015. (in Russian) Accessible at <http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf>
- [26] S. P. SHARY, Maximum consistency method for data fitting under interval uncertainty. *Journal of Global Optimization*, published online July 18, 2015, pp. 1–16. DOI: 10.1007/s10898-015-0340-1

- [27] D. J. SHESKIN, *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall/CRC, Boca Raton, Florida, 2004.
- [28] A. B. TSYBAKOV, *Introduction to Nonparametric Estimation*. Springer, New York, 2009.
- [29] S. I. ZHILIN, On fitting empirical data under interval error, *Reliable Computing*, vol. 11 (2005), pp. 433–442. DOI: 10.1007/s11155-005-0050-3
Accessaible at <http://www.nsc.ru/interval/Library/Thematic/DataProcs/Zhilin.pdf>