# ON DEVELOPING RUSSIAN–ISO STANDARD (GUIDE) TO PROCESSING MEASUREMENT INFORMATION UNDER CONDITIONS OF UNCERTAINTY OF MEASURING ERRORS AND SMALL NUMBER OF MEARUREMENTS (ON THE BASIS OF METHODS OF THE INTERVAL ANALYSIS)[1]

S.I. Kumkov

*Institute of Mathematics and Mechanics, UrB of RAS, S.Kovalevskaya str. 16*
*Ekaterinburg GSP-384, 620219, RUSSIA* *E-mail:* kumkov@imm.uran.ru

**Abstract**

The paper deals with application of methods of optimal observation and the applied interval analysis to estimation of the informational set of parameters of a describing polynomial, i.e., the set of parameters consistent with the given sample of measurements. Estimation is implemented under conditions of uncertainty of probabilistic characteristics of measurement errors. Only geometric constrains are imposed on the errors. Linearity of the describing polynomial with respect to the parameters to be estimated allows to concretize effectively procedures of the interval analysis, to use fast procedures of the convex analysis, and to obtain exact description of a polyhedral of admissible values of parameters. Theoretical validity and simplicity of algorithms and corresponding software allow to recommend the suggested approach to elaboration of an official Standard (Guide) to processing corrupted information under conditions of uncertainty and small number of measurements in the sample.

**Key words:** standard, guide to estimation, sample of measurements, measurement errors, uncertainty conditions, measurement uncertainty set, describing polynomial, parameters, partial informational set of parameters. informational set, tube of dependencies, inconsistent sample, consistent subsample

## 1. Introduction

Elaboration of the suggested Standard (Guide) is stipulated by necessity of development and extension of methods for processing measurements corrupted by errors. Especial need appears in the cases when application of standard methods of the mathematical statistics is hampered because of uncertainty or complete absence of information about probabilistic characteristics of measuring errors and under short length (small number of measurements) of a sample to be processed. Under such conditions the maximal difficulty is in computation of the admissible regions of actual values of constants, regions of actual values of parameters of describing dependencies, and regions of their actual values (the tube).

The main application of the suggested Guide is procession of measurements in destructing tests, inter-laboratory tests, round comparisons, and under difficulties for tests repeating caused by some technological or economical reasons. The Guide is elaborated in development and extension of the Russian Standard [1], the Guide [2], and the European Standard [3].

## 2. The main terms and definitions

In comparison to [4], the following new basic notions and definitions are introduced.

**Describing dependency.** It is a function of some given type $y = f(x, P)$ that depends on the argument $x \in [x_{\min}, x_{\max}]$ values on the interval of its variation and on values of vector $P$ of the parameters. It is necessary to estimate the set of admissible values of parameters and the region of admissible values of the dependence (the enhanced "tube" of the dependence values).

**Model of measurement error.** It is some known structure of the measuring error. Here, the minimal data about the model are used. The model of general type can contain both a relative $\delta_i$ and absolute $\varepsilon_i$ components of the error

$$y_i = y(x_i, P) = y_i^*(x_i, P)(1 + \delta_i) + \varepsilon_i, \ \ |\delta_i| \le \delta_{\max}, \ \ |\varepsilon_i| \le \varepsilon_{\max},$$

where $i$ is the measurement number; $x_i$ is the argument value at the $i$th measurement; $y_i$ is the corrupted measurement; $y_i^*(x_i, P)$ is the unknown true value of the dependency; $\delta_{\max}$, $\varepsilon_{\max}$ are maximal values of bounds on the measurement errors. If the data about the error structure are absent, then the errors are constrained by some equivalent summary value $\varepsilon_{\max}$.

**Measurement uncertainty set $H$ of a single measurement (MUS).** It is a set (region) of the argument and function values consistent with the error bounds in this measurement. In the case of

[1]The work was supported by the Russian Foundation for Basic Research (projects nos. 06–01–00414, 04–01–96099) and by Corporation "New Informational Technologies in Aviation" (contract no. 53/2005).

the model with the absolute error components of the maximal values $\Delta x_{\max}$ and $\Delta y_{\max}$ and a uniform sample

$H_i = [x_i - \Delta x_{\max}, x_i + \Delta x_{\max}] \times [y_i - \Delta y_{\max}, y_i + \Delta y_{\max}]$.

In the general case of the model MUS of each measurement is constructed by special formulas [7]. In the case when the argument is exactly measured, MUS of this measurement is the **uncertainty interval**

$H_i = [y_i - \Delta y_{\max}, y_i + \Delta y_{\max}]$.

**Actual (admissible) dependency.** It is a function of the given type that passes through MUS's of all its measurements.

**Actual (admissible) value of the parameters vector.** The value that defines an actual dependency.

**A priory set $IS_0$ of possible values of the parameters.** A region of values of the parameters vector, whose bounds have been approximately estimated *a priori* theoretically or defined by results of some previous measurements.

**Partial informational set $PIS_{i,j,k}$ of actual values of parameters.** It is a set (region) of values of the parameters consistent with some group of $i, j, k$ measurements of the given sample.

**Informational set $IS(P)$ of the dependency parameters.** It is a set (region) of values of the parameters consistent with all measurements of the given sample and the given type of the dependency.

**Tube $F(x)$ of actual values of he dependency.** It is a set (region) of values of the dependency consistent with the given sample of measurements, given type of the dependency, and the informational set of the parameters. The tube is described by the following totality of intervals:

$F(x) = \{[F_{i,\min},\, F_{i,\max}],\, i = 1, N\},\;\; F_{i,\min} = \min_{P \in IS(P)}\{y(x_i, P)\},\;\; F_{i,\max} = \max_{P \in IS(P)}\{y(x_i, P)\}$.

**Consistent sample.** The given sample is consistent if its informational set is not empty. The equivalent definition: the sample is consistent if there exists at least one actual dependency.

In the most simple case of the dependence $y = C = \text{Const}$, i.e., a physical constant, the following definitions are simplified.

**Measurement of uncertainty set $H$ of a single measurement.** It is a set (region) of possible values of a single measurement that depends on the constraints of error of a measuring unit or a way of measuring. The constraints are defined on the basis of some preliminary research in correspondence with the error model. In the case of only absolute component in the error model, the **measurement uncertainty interval** of each measurement is described $H_i = [y_i - \Delta y_{\max}, y_i + \Delta y_{\max}]$.

**Interval $I(C)$ actual values of the physical constant.** It is an interval of values of the measurement that are consistent with the given sample of measurements. The interval is bounded from below by the maximal lower frontier of the MUS of single measurements; and from above this interval is bounded by the minimal upper frontier of the MUS of single measurements

$I(C) = \bigcap_i H_i = [h_{\min}, h_{\max}],\;\; i = 1, N,\;\; h_{\min} = \max_i\{y_i - \Delta y_{\max}\},\;\; h_{\max} = \min_i\{y_i + \Delta y_{\max}\}$.

### 3. Construction of informational sets of parameters

The approach is based on methods of the optimal observation under conditions of uncertainty [5] and methods of the interval analysis [6–8], when only geometric constraints are imposed on the measurement error. Any probabilistic characteristics of the errors are unknown.

As the first variant of the Standard (Guide) under consideration, the most simple and practically interesting case is suggested when the process under investigation is described by the first (or second) order polynomial of a independent argument and the right-hand side that are linear with respect to the unknown coefficients (parameters). Usage of the mentioned type of the dependency on parameters has allowed to apply fast procedures from the convex analysis for construction of the sets of admissible values of parameters, i.e., the informational sets. The Guide can be used also for estimation of dependency parameters in the cases when, by changing of variables, the dependency can be reduced to the mentioned linear type with respect to the parameters [9–11].

Let us illustrate the main procedures of the Guide on an example of the linear dependency $y = Ax + B$, whose parameters have to be estimated on the basis of the given sample of corrupted measurements $\{y_i, x_i\}$, $i = 1, N$ with the length $N$ at the values $x_i$ of the argument. The argument values (abscissa) are assumed to be ordered by increasing $x_i < x_j$ for all $i < j$. For simplicity of description assume also that the error model contains only the absolute component $\Delta y_{\max}$. The following totality of MUS's is put in correspondence to the given sample:

$H_i = [h_{i\min}, h_{i\max}],\;\; h_{i\min} = y_i - \Delta y_{\max},\;\; h_{i\max} = y_i + \Delta y_{\max}],\;\; i = 1, N$.

**Estimation of unconditional intervals of the parameters.** Similarly to [1] let us estimate the unconditional (independent) intervals $I_A$ and $I_B$ that are useful in practice. For each pair of MUS's $H_i$, $H_j$ of measurements $i = 1, N-1; j = i+1, N$ the bounds of the partial intervals of the parameter $A$ are computed

$A_{i,j} = [a_{i,j,\min}, a_{i,j,\max}],\;\; a_{i,j,\min} = (h_{j\min} - h_{i\max})/(x_j - x_i),\;\; a_{i,j,\max} = (h_{j\max} - h_{i\min})/(x_j - x_i)$.

The bounds of the partial intervals of the parameter $B$ are computed

$B_{i,j} = [b_{i,j,\min}, b_{i,j,\max}],\;\; b_{i,j,\min} = h_{i\min} - a_{i,j,\max} x_i,\;\; b_{i,j,\max} = h_{i\max} - a_{i,\min} x_i$.

The unconditional intervals for the parameters are found by intersection of the partial intervals of pairs of the measurements $i = \overline{1, N-1}$, $j = \overline{i+1, N}$,

$I_A = \bigcap_{i,j} A_{i,j} = [a_{\min}, a_{\max}]$, $a_{\min} = \max\{a_{i,j,\min}\}$, $a_{\max} = \min\{a_{i,j,\max}\}$,

$I_B = \bigcap_{i,j} B_{i,j} = [b_{\min}, b_{\max}]$, $b_{\min} = \max\{b_{i,j,\min}\}$, $b_{\max} = \min\{b_{i,j,\max}\}$.

These unconditional intervals give the box-estimate from above of the region of possible values for the parameters to be estimated. Such an estimate even being useful in practice for primary approximate calculations could be rather rough.

**Exact estimate of the two-dimensional informational set $IS(A, B)$ of the parameters.** This set is an analogue of the confident two-dimensional region [1] of the parameters values under given level of confident probability. For each pair of the MUS's $H_i$ and $H_j$ of measurements $i = \overline{1, N-1}$; $j = \overline{i+1, N}$, the partial two-dimensional informational sets are computed

$PIS_{i,j}(A, B) = PIS_{i,j}$.

in the plane $a \times b$ of the parameters to be estimated. Since the right-hand side of the considered dependency is linear in these parameters and the argument values are strictly ordered, each set $PIS_{i,j}$ is a closed convex quadrangle, namely, a parallelogram. Such a parallelogram is given by four vertices and linear edges between them. Here, the interval analysis procedures for explicit construction of two-dimensional sets of parameters are evidently concretized by application of the function inverse to the given describing one [6,7,9–11]. The vertices (ordered clockwise, for instance, from the upper-right one) are computed for each parallelogram as follows:

$a^1_{i,j} = (h_{j,\max} - h_{i,\max})/(x_j - x_i)$, $b^1_{i,j} = h_{i,\max} - a^1_{i,j} x_i$,

$a^2_{i,j} = (h_{j,\max} - h_{i,\min})/(x_j - x_i)$, $b^2_{i,j} = h_{i,\min} - a^2_{i,j} x_i$,

$a^3_{i,j} = (h_{j,\min} - h_{i,\min})/(x_j - x_i)$, $b^3_{i,j} = h_{i,\min} - a^3_{i,j} x_i$,

$a^4_{i,j} = (h_{j,\min} - h_{i,\max})/(x_j - x_i)$, $b^4_{i,j} = h_{i,\max} - a^4_{i,j} x_i$.

The exact estimate of the two-dimensional informational set $IS(A, B)$ of the parameters is found by specialized fast procedure of intersection of the convex partial informational sets over all pairs of the measurements [9–11]

$IS(A, B) = \bigcap_{i,j} PIS_{i,j}$, $i = \overline{1, N-1}$, $j = \overline{i+1, N}$.

This set is a closed convex polygonal with constrained number of vertices and exact linear edges between them.

**Tube $F(x)$ of the actual (admissible) values of the estimated linear dependency.** In the suggested Guide, the tube of the actual values of the dependency is an analogue of the confident region of values of the dependence at each value of the argument. According to the above definition, the tube $F(x) = \{[F_{i,\min}, F_{i,\max}]$, $i = \overline{1, N}\}$ of the actual values of the given dependency is found

$F_{i,\min} = \min_{(A,B) \in IS(A,B)}\{Ax_i + B\}$, $F_{i,\max} = \max_{(A,B) \in IS(A,B)}\{Ax_i + B\}$.

**Evaluation of consistency of the given samples.** Application of the ideology of the interval analysis has allowed to introduce a new crucial procedure **for constructive analysis of consistency** of the given sample of corrupted measurements. Namely, the opportunity of direct detection of the consistency – inconsistency of the sample appears under the given level $\Delta y_{\max}$ of the maximal value of the error. In the process of computation of the two-dimensional informational set $IS(A, B)$ under actual value of the constraint on the error, the intersection of the partial informational sets **for a consistent sample is always non-empty**. If the sample is **inconsistent**, for example, because of presence of at least one outlier [1,2,4], the intersection is **always empty**, i.e., the informational set does not exist.

**Partition of an inconsistent sample into consistent subsamples and detection of single outliers.** In the existent Standards [1,2] that are based on methods of the mathematical statistics, detection and elimination of outliers is implemented by the standard rule of "two sigmas". But if several outliers of different signs are present in the sample or there is a large systematic shift of a group of measurements with respect to the rest part of the sample, the rule of "two sigmas" could be unapplicable or lead to fatal distortion of the result.

Usage of ideology of the interval analysis has allowed to elaborate constructive procedure for **partition of an inconsistent sample into a totality of internally consistent subsamples.** Here, the procedures of separation of complete subgraphs from a given graph [12] are applied as follows. A totality of the partial informational sets

$\{PIS_{i,j},\ i = \overline{1, N-1}; j = \overline{i+1, N}\}$,

is regarded as a graph with these sets as the vertices. Each arbitrary pair of vertices $PIS_{i,j}$ and $PIS_{k,l}$ is called connected by the edge $(i, j) - (k, l)$ if their intersection $PIS_{i,j} \bigcap PIS_{k,l}$ is non-empty. Otherwise, the vertices are not connected, and the edge is absent.

A group of vertices forms a complete connected subgraph, if they all are pairwise connected by the edges. In the terms of the partial informational sets, it means non-emptiness of their complete intersection. In other words it means that the measurements of the corresponding numbers forms a **consistent subsample.** Thus, if some consistent subsample forms a complete subgraph with a number of vertices, for example equal to $M$, then to each vertex of this subgraph at least $M$ edges comes. In contrast to this case, each single outlier leads to isolated vertices (the isolated partial informational sets), which have no connecting edges.

Algorithms of separation of complete connected subgraph are wide known. But since the number of pairwise partial informational sets increases significantly when the number of measurements of the sample increases, then to obtain fast procedures, special algorithms for construction of the partial informational sets and separation of consistent subsamples in the case of inconsistent given sample have been elaborated.

Considered Guide was applied to estimation of both polynomial dependencies with dimension of the parameters vector of three and more and nonlinear dependencies. Special hybrid algorithm has been elaborated. In it, the grid representation of a part of parameters is used, especially, for the parameters entering nonlinearly into the dependency. But for rest part of parameters linearly entering into the dependency, the described exact estimation of two-dimensional cross-sections of the multi-dimensional informational sets is carried out. The Guide was successfully applied also to the cases when, using the change of variables, it is succeeded to reduce a nonlinear dependency in parameters of the describing function to a linear one.

In the Institute of Mathematics and Mechanics (IMM) of the Ural Branch of the Russian Academy of Sciences in cooperation with the Ural Scientific-Research Institute of Metrology (UNIIM) and the Metrological Center of the Ural Branch of the Russian Academy of Sciences (CERTIMET), a preliminary version of the suggested Guide and corresponding software for computations have been elaborated. Specialized variants of the Guide and software were probated on a representative number of practical problems of procession of experimental information in thermal-physics, high-temperature electrochemistry, organic synthesis of medicaments, estimation of parameters of stability of reference specimens, in air traffic control systems, and guidance systems of space vehicles [9–11].

Preliminary materials have been prepared for cooperative work with interested Russian organizations and organizations abroad elaborating ISO-Standards (for example, in the European Community).

## 4. References


1. GOST 8.207-76. The State System of Providing the Uniformity of Measuring. Direct Measurings with Multiple Observations. Methods for Processing Results of Observations.
2. MI 2083-93. Recommendations. The State System of Providing the Uniformity of Measuring. Implicit Measurings. Computation of Results of Measurings and Estimation of Their Errors.
3. Guide to the Expression of Uncertainty in Measurement. International Organization for Standardization. ISO, ISBN 92-67-10188-9, 1992.
4. STANDARD RMG 29–99. Metrology. The Main Terms and Definitions.
5. Kurzhanskii A.B. Control and Observation under Conditions of Uncertainty. Nauka, Moscow, 1977 (in Russian).
6. Neumaier A. Interval Methods for Systems of Equations. Cambridge University Press, Cambridge, 1990.
7. Jaulin L., Kiefer M., Didrit O., Walter E. Applied Interval Analysis. Springer, 2001, London, Berlin.
8. Milanese M. and Norton J.P. (Eds.) Bounding Approaches to System Identification. Plenum Press, 1996, London, UK, New York, USA.
9. Kumkov S.I. Method of Informational Sets in Problems of Filtration, Identification, and Control. Tampere International Center for Signal Processing, Tampere, Finland, March 2001, http:/sigwww.cs.tut.fi/TICSP/PRESENTATIONS/2001.htm , pp.1-33.
10. Kumkov S.I. Informational Sets in Applied Problems of Evaluation. Proceedings of the IFAC International Workshop "Nonsmooth and Discontinuous Problems of Control and Optimization". June, 17-20, 1998. Chelyabinsk, Russia. pp. 130-134.
11. Kumkov S.I. Application of Minimax Evaluation Procedures for Expansion of Metrological Standards on Noised Measurements. Proceeding of International Conference "SIMULATION-98". University of York, UK, 30 September - 2 October 1998, IEE Publication Number 457, pp.162-167.
12. Swami M., Tkhulasiraman K. Graphs, Nets, and Algorithms. Mir, Moscow, 1983 (in Russian).