SUPPLEMENT II
Example of estimation of the set of parameters for the linear function under conditions of uncertainty

Estimation of the set of parameters for a linear function is carried out. The function is a measuring dependence of some weighting unit

$$y(x,a,b) = ax + b .$$

Here, x is the argument, input weight; a is the coefficient of inclination; b is a constant parameter.

The model of the error contains only the absolute component. The probability distribution law or any other statistical characteristics of the error are absent. The sample is homogeneous, i.e., the maximal value of the error is the same for all measurements. The true value of the coefficient of inclination is 1.0 kg/kg (kilograms). The true value of the constant parameter is 0.1 kg. The constraint on the maximal value of the error is $\varepsilon_{\max} = 0.05$ kg. The step of the input values of the argument is 0.1 kg. The value of the argument is known exactly.

The following consistent sample of the length $N = 8$ is given (the errors are obtained by the standard random number generator from the interval $[-\varepsilon_{\max}, \varepsilon_{\max}]$):

$\{x_i\}$ = {0; 0.1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7} kg,

$\{y_i\}$ = {0,055; 0,245; 0,260; 0,360; 0,545; 0,562; 0,651; 0,760} kg,

$\{y_i^*\}$ = {0,100; 0,200; 0,300; 0,400; 0,500; 0,600; 0,700; 0,800} kg,

where $\{y_i^*\}$ are the true values of the function (for illustration).

The bounds of the uncertainty sets $H_i = [h_{\text{H}i},\ h_{\text{B}i}]$, $i = 1,8$ are calculated

[0.005, 0.105], [0,195, 0,295], [0.210, 0.310], [0.310, 0.410],
[0.495, 0.595], [0.510, 0.612], [0.601, 0.701], [0.710, 0.810] kg.

The partial informational sets are calculated for sample pairs of measurements with numbers $\{1, i\}$, $i = 2,\ldots,8$ only. Such sets are parallelograms. The initial *PIS* G_{12} is presented in Fig.SII.1 in the dash line. Its vertices in the plane $a \times b$ "incline–constant" have the values of coordinates {1,900; 0,105}, {2,900; 0,005}, {1,900; 0,005}, {0,900; 0,105}, the physical dimentsons are kg/kg and kg, respectively.

Since the input sample is consistent, the resultant informational set exists (is not empty). Its vertices have the coordinates {0,994; 0,105}, {1,013; 0,094}, {0,999; 0,095}, {0,975; 0,105}, kg/kg and kg, respectively.

The estimation, formally calculated by the standard Least Squares Mean method (as it used to do in the GOST and in practice), has the coordinates $\overline{a}$ = 0,9595 kg/kg and $\overline{b}$ = 0,0940 kg.

In the plane of parameters (Fig.SII.1a), the initial *PIS* (in dash line) is shown for comparison with the resultant *IS* (in solid lines), the true value of parameters is marked with the cross, the LSQM estimation is shown by a circle.

Comparison of the *IS* and the initial *PIS* (both in the general view Fig.SII.1a and in the zoomed fragment Fig.SII.1b) shows significant decreasing of the *IS* sizes. Note the important fact that the ***LSQM estimation is not consistent with the given sample*** since it lies out of the *IS*.

By the *IS*, the unconditional intervals ofparameters are found as the *IS* projections onto the corresponding axes I_a = [0.975, 1.013] kg/kg, I_b = [0.094, 0.105] kg.

Since the probabilistic characteristics of the errors are unknown, the two-dimensional probability distribution low of the coefficient of inclination a and the constant b is unknown also. So, it is

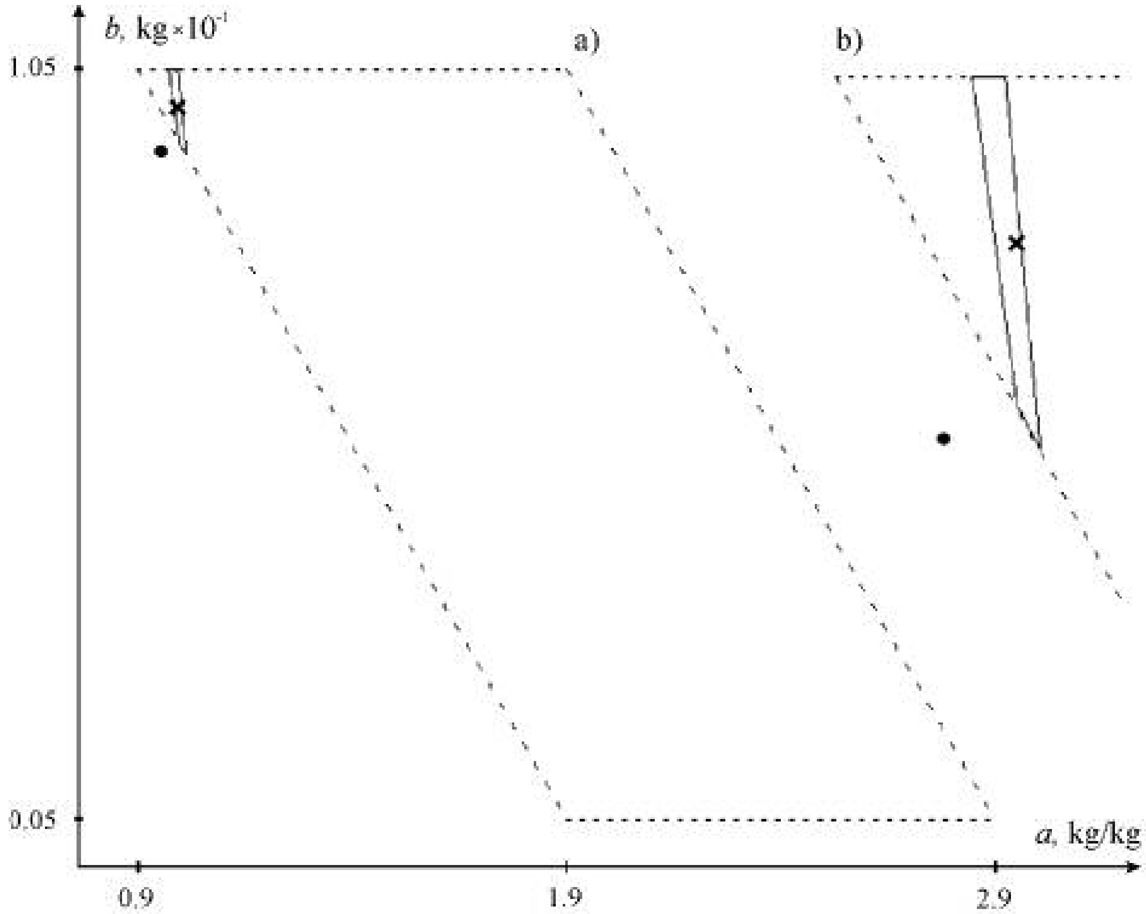

Figure SII.1: The informational set of parameters (solid boundaries); one of the initial partial informational sets (dashes); a) general view; b) zoomed fragment; the true value marked by the cross; the LSQM value is shown by the black circle

possible to calculate formally some unconditional estimations of the square means σ_a and σ_b of these parameters and the standard intervals $\pm 2\sigma_a$ and $\pm 2\sigma_b$ relatively the LSQM mean values $\overline{a} =$ 0,9595 kg/kg and $\overline{b}$ =0,0940 kg. But such intervals are not justified and in practice are rather rough, significantly larger than the unconditional intervals I_a and I_b, constructed by the resultant informational set.

Results of estimations in the plane "input weight–dependence" are shown in Fig.SII.2. The crosses mark the measurements, the vertical lines show the uncertainty sets of the measurements. The lines in rare dashes mark the initial rough boundaries of possible dependence values, the solid lines shows the boundaries of the resultant tube constructed by the informational set. The tube bounds at the given values of the argument have the following values

$\{x_i\}$ = {0; 0.1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7} kg,

$\{T_{\text{lo}}(x_i\}$ = {0,937; 0,195; 0,295; 0,395; 0,495; 0,592; 0,690; 0,787} kg,

$\{T_{\text{up}}(x_i\}$ = {0.105; 0,204; 0,303; 0,403; 0,502; 0,602; 0,701; 0,802} kg,

and the width of the resultant tube at the given values of the argument is

$\{T_{\text{up}}(x_i) - F_{\text{lo}}(x_i)\}$ = {0.0112; 0,0094; 0,0088; 0,0082; 0,0076; 0,0095; 0,0114; 0,0151} kg.

Figure SII.2 shows that the domain of admissible values of the dependence is significantly smaller of the initial one. The dashed straight line corresponds to the LSQM point $\overline{a}$,$\overline{b}$. It is seen

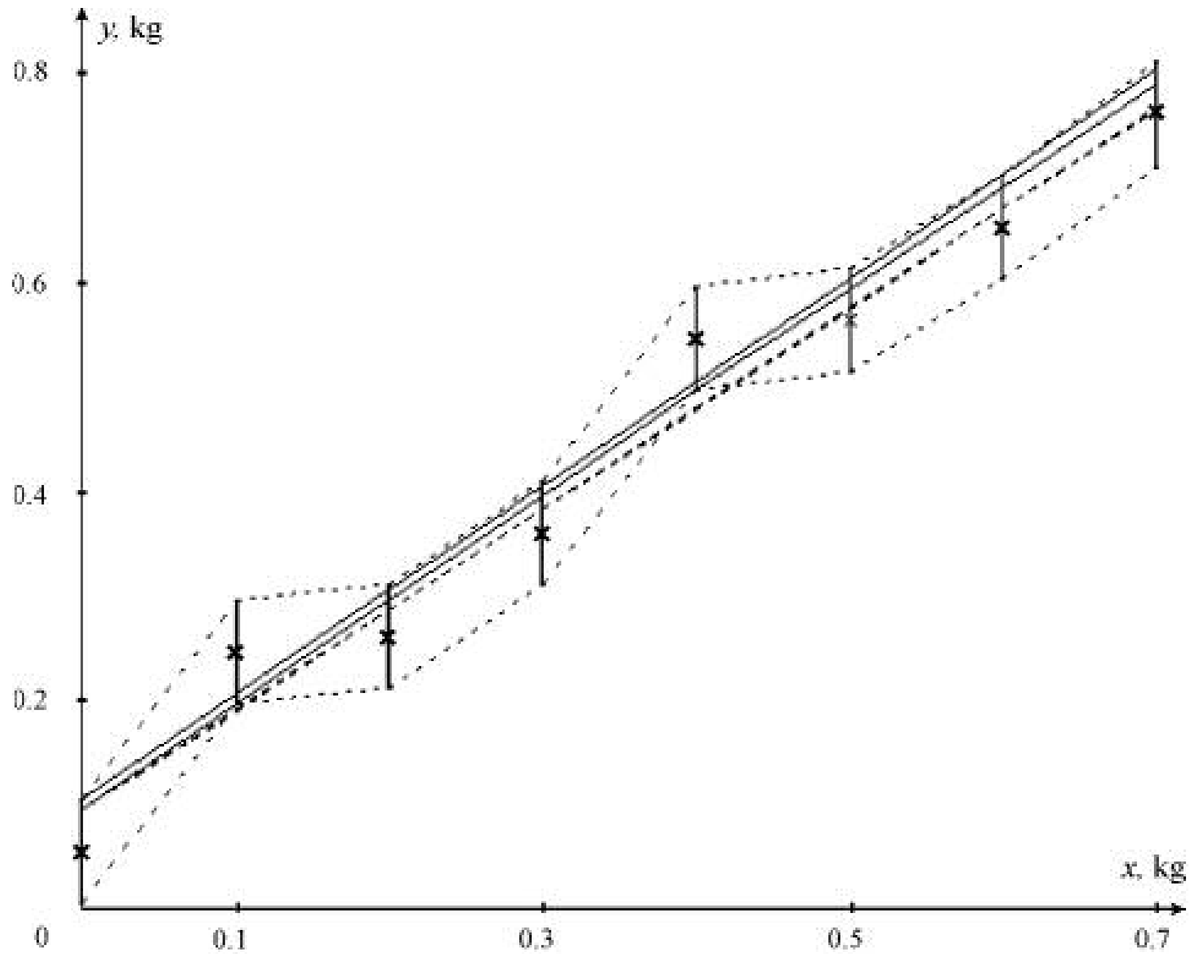

Figure SII.2: The input sample (crosses), the uncertainty intervals of measurements (vertical segments), the rough boundaries (sparse dashes), the tube of admissible function values (solid boundaries), the LSQM line (dashes)

that this line is ***inconsistent*** with the given sample, since it goes out of the uncertainty intervals of two measurement:

No.2 $y(x_2,\bar{a},\bar{b}) = 0.1899 < h_{lo2} =$ 0,1950 kg, No.5 $y(x_5,\bar{a},\bar{b}) = 0.4778 < h_{lo5} =$ 0,4950 kg.

The section of the resultant tube with the maximal width corresponds to the argument value $x_8 =$ 0,7 kg, and the central value of the section is $y_c = 0.7945$ kg.

Here, the maximal deviation of the admissible dependence from the central value is

$d_{\max} = (0.8020 - 0.7870)/2 = 0.0075$ kg.

The minimal limit level of the error realized in the input sample has the lower estimation

$\varepsilon^*_{\max} = 0.0405$ kg,

and coordinates of the limit point informational set are

$a^* =$ 0.9797 kg/ kg, $b^* =$ 0.1040 kg.

Thus, the discussed Example of estimation of the admissible set of parameters for a linear function parameters shows that for a short sample under essential level of the errors really appeared in the sample, the approach on the basis of the information sets allows to perform more sophisticated analysis and gives the better results (in the accuracy) than the standard procedures on the basis of the classic statistical methods.