Maximum Consistency Method for Data Fitting under Interval Uncertainty

Sergey P. Shary
Institute of Computational Technologies,
Novosibirsk, Russia

For the linear regression model $b = a_1x_1 + a_2x_2 + \cdots + a_nx_n$, we consider the problem of data fitting under interval uncertainty. Let an interval $m{\times}n$-matrix $\boldsymbol{A} = (\boldsymbol{a}_{ij})$ and an interval m-vector $\boldsymbol{b} = (\boldsymbol{b}_i)$ represent the input data and output responses of the model respectively, such that $a_1 \in \boldsymbol{a}_{i1}$, $a_2 \in \boldsymbol{a}_{i2}$, … , $a_n \in \boldsymbol{a}_{in}$, $b \in \boldsymbol{b}_i$ in the i-th experiment, $i = 1, 2, \ldots, m$. It is necessary to find the coefficients $x_1, x_2, \ldots, x_n$ that best fit the above linear relation for the data given.

A family of values of the parameters $x_1, x_2, \ldots, x_n$ is called *consistent* with the interval data ($\boldsymbol{a}_{i1}$, $\boldsymbol{a}_{i2}$, … , $\boldsymbol{a}_{in}$), $\boldsymbol{b}_i$, $i = 1, 2, \ldots, m$, if, for every index i, there exist such point representatives $a_{i1} \in \boldsymbol{a}_{i1}$, $a_{i2} \in \boldsymbol{a}_{i2}$, … , $a_{in} \in \boldsymbol{a}_{in}$, $b_i \in \boldsymbol{b}_i$ that $a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i$. The set of all the parameter values consistent with the data given form a *parameter uncertainty set*. As an estimate of the parameters, it makes sense to take a point from the parameter uncertainty set providing that it is nonempty. Otherwise, if the parameter uncertainty set is empty, then the estimate should be a point where maximal “consistency” (in a prescribed sense) with the data is achieved.

The parameter uncertainty set is nothing but the solution set $\Xi(\boldsymbol{A}, \boldsymbol{b})$ to the interval system of linear equations $\boldsymbol{A}x = \boldsymbol{b}$ defined in interval analysis: $\Xi(\boldsymbol{A}, \boldsymbol{b}) = \{\, x \mid Ax = b$ for some A from $\boldsymbol{A}$ and b from $\boldsymbol{b}\,\}$. For the above data fitting problem, we propose, as the consistency measure, the values of the *recognizing functional* of the solution set $\Xi(\boldsymbol{A}, \boldsymbol{b})$, which is defined as

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

where “mid” and “rad” mean the midpoint and radius of an interval. The functional Uss “recognizes” the points of $\Xi(\boldsymbol{A}, \boldsymbol{b})$ by the sign of its values: $x \in \Xi(\boldsymbol{A}, \boldsymbol{b})$ if and only if $\mathrm{Uss}\,(x, \boldsymbol{A}, \boldsymbol{b}) \ge 0$. Additionally, Uss has reasonably good properties as a function of x and $\boldsymbol{A}$, $\boldsymbol{b}$.

As an estimate of the parameters in the data fitting problem, we take the value of $x = (x_1, x_2, \ldots, x_n)$ that provides maximum of the recognizing functional Uss (Maximum Consistency Method). Then,

- if the parameter uncertainty set is nonempty, we get a point from it,
- if the parameter uncertainty set is empty, we get a point that still has maximum possible consistency (determined by the functional Uss) with the data given.

In our work, we discuss properties of the recognizing functional Uss, interpretation and features of the estimates obtained by the Maximum Consistency Method as well as correlation with the other approaches to data fitting under interval uncertainty.

References

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.

S.P. Shary and I.A. Sharaya, Recognizing solvability of interval equations and its application to data analysis, *Computational Technologies*, vol. 18 (2013), No. 3, pp. 80-109. (in Russian)