

From Interval Computations to Constraint-Related Set Computations: Towards Faster Estimation of Statistics and ODEs Under Interval, p-Box, and Fuzzy Uncertainty

Martine Ceberio¹, Vladik Kreinovich¹, Andrzej Pownuk², and Barnabás Bede²

¹ Department of Computer Science, University of Texas at El Paso,
El Paso, TX 79968, USA

mceberio@cs.utep.edu, vladik@utep.edu

² Department of Mathematical Sciences, University of Texas at El Paso,
El Paso, TX 79968, USA

ampownuk@utep.edu, bbede@utep.edu

Abstract. In interval computations, at each intermediate stage of the computation, we have intervals of possible values of the corresponding quantities. In our previous papers, we proposed an extension of this technique to *set computations*, where on each stage, in addition to intervals of possible values of the quantities, we also keep sets of possible values of pairs (triples, etc.). In this paper, we show that in several practical problems, such as estimating statistics (variance, correlation, etc.) and solutions to ordinary differential equations (ODEs) with given accuracy, this new formalism enables us to find estimates in feasible (polynomial) time.

1 Formulation of the Problem

Need for data processing. In many real-life situations, we are interested in the value of a physical quantity y that is difficult or impossible to measure directly. Examples of such quantities are the distance to a star and the amount of oil in a given well. Since we cannot measure y directly, a natural idea is to measure y *indirectly*. Specifically, we find some easier-to-measure quantities x_1, \dots, x_n which are related to y by a known relation $y = f(x_1, \dots, x_n)$; this relation may be a simple functional transformation, or complex algorithm (e.g., for the amount of oil, numerical solution to a partial differential equation). Then, to estimate y , we first measure or estimate the values of the quantities x_1, \dots, x_n , and then we use the results $\tilde{x}_1, \dots, \tilde{x}_n$ of these measurements (estimations) to compute an estimate \tilde{y} for y as $\tilde{y} = f(\tilde{x}_1, \dots, \tilde{x}_n)$.

Computing an estimate for y based on the results of direct measurements is called *data processing*; data processing is the main reason why computers were invented in the first place, and data processing is still one of the main uses of computers as number crunching devices.

Measurement uncertainty: from probabilities to intervals. Measurement are never 100% accurate, so in reality, the actual value x_i of i -th measured quantity can differ from the measurement result \tilde{x}_i . Because of these *measurement errors* $\Delta x_i \stackrel{\text{def}}{=} \tilde{x}_i - x_i$, the result $\tilde{y} = f(\tilde{x}_1, \dots, \tilde{x}_n)$ of data processing is, in general, different from the actual value $y = f(x_1, \dots, x_n)$ of the desired quantity y .

It is desirable to describe the error $\Delta y \stackrel{\text{def}}{=} \tilde{y} - y$ of the result of data processing. To do that, we must have some information about the errors of direct measurements.

What do we know about the errors Δx_i of direct measurements? First, the manufacturer of the measuring instrument must supply us with an upper bound Δ_i on the measurement error. If no such upper bound is supplied, this means that no accuracy is guaranteed, and the corresponding “measuring instrument” is practically useless. In this case, once we performed a measurement and got a measurement result \tilde{x}_i , we know that the actual (unknown) value x_i of the measured quantity belongs to the interval $\mathbf{x}_i = [\underline{x}_i, \bar{x}_i]$, where $\underline{x}_i = \tilde{x}_i - \Delta_i$ and $\bar{x}_i = \tilde{x}_i + \Delta_i$.

In many practical situations, we not only know the interval $[-\Delta_i, \Delta_i]$ of possible values of the measurement error; we also know the probability of different values Δx_i within this interval. This knowledge underlies the traditional engineering approach to estimating the error of indirect measurement, in which we assume that we know the probability distributions for measurement errors Δx_i .

In practice, we can determine the desired probabilities of different values of Δx_i by comparing the results of measuring with this instrument with the results of measuring the same quantity by a standard (much more accurate) measuring instrument. Since the standard measuring instrument is much more accurate than the one use, the difference between these two measurement results is practically equal to the measurement error; thus, the empirical distribution of this difference is close to the desired probability distribution for measurement error. There are two cases, however, when this determination is not done:

- First is the case of cutting-edge measurements, e.g., measurements in fundamental science. When we use the largest particle accelerator to measure the properties of elementary particles, there is no “standard” (much more accurate) located nearby that we can use for calibration: our accelerator is the best we have.
- The second case is the case of measurements in manufacturing. In principle, every sensor can be thoroughly calibrated, but sensor calibration is so costly – usually costing ten times more than the sensor itself – that manufacturers rarely do it.

In both cases, we have no information about the probabilities of Δx_i ; the only information we have is the upper bound on the measurement error.

In this case, after we performed a measurement and got a measurement result \tilde{x}_i , the only information that we have about the actual value x_i of the measured quantity is that it belongs to the interval $\mathbf{x}_i = [\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$. In such situations, the only information that we have about the (unknown) actual value of $y =$