XXIV IMEKO World Congress "Think Metrology"
August 26 – 29, 2024, Hamburg, Germany

USING KNOWN RELATION BETWEEN QUANTITIES TO MAKE MEASUREMENTS MORE ACCURATE AND MORE RELIABLE

Niklas R. Winnewisser [a,*], *Felix Mett* [a], *Michael Beer* [a], *Olga Kosheleva* [b], *Vladik Kreinovich* [c]

[a]Institute for Risk and Reliability, Leibniz University Hannover, Callinstr. 34, 30167 Hannover, Germany,
*winnewisser@irz.uni-hannover.de, beer@irz.uni-hannover.de
[b]Department of Teacher Education, University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA,
olgak@utep.edu
[c]Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA,
vladik@utep.edu

Abstract – Most of our knowledge comes, ultimately, from measurements and from processing measurement results. In this, metrology is very valuable: it teaches us how to gauge the accuracy of the measurement results and of the results of data processing, and how to calibrate the measuring instruments so as to reach the maximum accuracy. However, traditional metrology mostly concentrates on individual measurements. In practice, often, there are also relations between the current values of different quantities. For example, there is usually an known upper bound on the difference between the values of the same quantity at close moments of time or at nearby locations. It is known that taking such relation into account can lead to more accurate estimates for physical quantities. In this paper, we describe a general methodology for taking these relations into account. We also show how this methodology can help to detect faulty measuring instruments – thus increasing the reliability of the measurement results.

Keywords: relations between quantities, measurement accuracy, measurement reliability

1. FORMULATION OF THE PROBLEM

One of the main objectives of science and engineering. One of the main objectives of science and engineering is to predict the future state of the world – and to make sure that this future state is as beneficial for us as possible.

The state of the world at any given moment of time can be characterized by the corresponding values of physical quantities. Thus, what we need is to predict future values of these quantities. For example, predicting weather means predicting the future values of temperature, humidity, wind speed and direction, etc.

Measurements and resulting data processing are very important. To make the desired predictions, we need to know the relation between each future value y of the quantities of interest and current values $x_1, \ldots, x_n$ of related quantities. This relation is often described in terms of differential (and other) equations. By applying an appropriate algorithm to solve this equation, we get an algorithm $y = f(x_1, \ldots, x_n)$ for estimating y from x_i.

So, to come up with the desired estimate $\widetilde{y}$ for y, we measure the current values of the corresponding quantities $x_1 \ldots, x_n$ and use the measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$ to compute the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ for each desired future value y

Traditional metrology is very important. In general, the more accurately we measure, the more accurate our predictions. In this process, traditional metrology is very valuable: it teaches us how to gauge the accuracy of the measuring instruments and even how to calibrate the instrument so as to further increase its accuracy.

Need to take into account relation between quantities. However, traditional metrology mostly concentrates on individual measurements. In practice, often, in addition to the relation between future and present values, there are also relations between the current values of different quantities. For example, there is usually an known upper bound on the difference between the values of the same quantity at close moments of time or at nearby locations.

It is known that such relation can lead to more accurate measurement results. For example, suppose that we know that the values of the mechanical stress at two nearby locations are in intervals $[0.80, 1.00]$ and $[0.90, 1.10]$. If we know that the difference between the actual values cannot exceed 0.01, this means that the first quantity cannot be smaller than $0.90 - 0.01$, so we get a narrower interval $[0.89, 1.00]$ – and similarly, a narrower interval $[0.90, 1.01]$ for the second quantity.

In general, such relations take form of inequalities or inequalities. It is well known that any inequality or equality constraint can be described by inequalities of the type $g_j(x_1, \ldots, x_n) \le 0$:

- a general inequality constraint

$$a(x_1, \ldots, x_n) \le b(x_1, \ldots, x_n)$$

  is equivalent to

$$a(x_1, \ldots, x_n) - b(x_1, \ldots, x_n) \le 0,$$

  and

- a general equality constraint

$$a(x_1, \ldots, x_n) = b(x_1, \ldots, x_n)$$

is equivalent to two inequality constraints:

$$a(x_1, \ldots, x_n) - b(x_1, \ldots, x_n) \le 0 \text{ and}$$

$$b(x_1, \ldots, x_n) - a(x_1, \ldots, x_n) \le 0.$$

For example, the above relation of the type $|x_i - x_j| \le \delta$ means that we have two inequalities:

$$x_i - x_j \le \delta \text{ and } x_i - x_j \ge -\delta,$$

which are equivalent to

$$x_i - x_j - \delta \le 0 \text{ and } -\delta - (x_i - x_j) \le 0.$$

In this case,

$$g_1(x_1, \ldots, x_n) = x_i - x_j - \delta \text{ and}$$

$$g_2(x_1, \ldots, x_n) = -\delta - (x_i - x_j).$$

Case of indirect relations. So far, we have considered relations that directly relate the measured quantities. However, relations can also involve not-directly-measurable parameters of a model that describes the corresponding phenomenon. For example, we know that for a cantilever beam with a triangular line load linearly decreasing from q_0 at distance 0 to 0 at distance $z = \ell$, the bending x_i at location z_i is determined by the following formula:

$$x_i = c_1 \cdot \left(-\frac{1}{120\ell} z_i^5 + \frac{1}{24} z_i^4 + \frac{1}{12} z_i^3 \cdot \ell + \frac{1}{12} z_i^2 \cdot \ell^2 \right),$$

where

$$c_1 \stackrel{\text{def}}{=} \frac{q_0}{E \cdot I}.$$

We measure the values x_i, but we do not directly the value c_1. Since different measurement results are related to the same coefficient c, they are indirectly related to each other: namely, we know the exact ratios x_i/x_j of the bendings corresponding to different locations on the beam.

In general, we can have models with several not-directly-measurable parameters $c_1, \ldots, c_k$, so a general constraint can take a form

$$g_j(x_1, \ldots, x_n, c_1, \ldots, c_k) \le 0.$$

Based on the measurements, we can find estimates $\widetilde{c}_i$ for these parameters. Since these estimates are based on measurements, and measurements are not absolutely accurate, these estimates, in general, differ from the actual (unknown) values of these parameters, i.e., we have a non-zero estimation errors

$$\Delta c_i \stackrel{\text{def}}{=} \widetilde{c}_i - c_i.$$

What we do in this paper. In the paper, we analyze the general problem of using known relation between quantities to make measurements more accurate and more reliable.

Our main focus is on the – rather typical practical case – when we only know the upper bound on the measurement error. Measurement errors are usually relatively small, so terms quadratic (or higher order) with respect to measurement errors can be safely ignored. We show that under this – usual – linearization assumption, the corresponding problems can be effectively solved by known linear programming algorithms.

We also consider the case when we know the probability distributions of different measurement errors – and we know the reliability of each measuring instrument and our degree of confidence in each relation between the quantities. In this case, the problem is computationally more difficult but still solvable.

2. CASE OF INTERVAL UNCERTAINTY: A BRIEF REMINDER

Description of the situation. In this section, we consider a rather typical practical case, when for each measurement, we only know the upper bound Δ on the absolute value $|\Delta x|$ of the measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$; see, e.g., [7]. In this case, once we perform the measurement and get the measurement result $\widetilde{x}$, the only information that we now have about the actual value x of the measured quantity is that this value x is located in the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$. Because of this, this type of uncertainty is known as *interval uncertainty*; see, e.g., [2, 4–6].

The upper bound is an absolute necessity. The upper bound is the absolute minimum information that we need to know: if we do not even know the upper bound on the measurement error, this means that the actual value x can be as different from the measurement result $\widetilde{x}$ as possible. This is not a measurement, this is a wild guess.

Of course, the upper bound is not the only information we can have. Ideally, in addition to knowing this upper bound, we should also know the probabilities of different values of Δx. However, in many practical situations, the upper bound is the only information that we have about the measurement error.

In many practical situations, the upper bound is the only information we have. Actually, there are two types of such situations. One is the most common one, when we are performing routine measurements – on the factory floor, in a building, etc. In principle, we can calibrate each sensor and get the probability distribution of its measurement error, but nowadays, most sensors are very cheap – unless we are looking for very accurate ones – while calibration is very expensive. As a result, most sensors used in industry and in practice are not calibrated very accurately. For example, for a sensor that measures a body temperature, it is enough to know that it provides temperature with accuracy probably $\pm 0.2^\circ$ C, it makes no sense to come up with the exact probability distribution.

Another case when the upper bound is the only information we have is the case of state-of-the-art measurements, when the usual calibration techniques are not applicable. Indeed, the usual sensor calibration techniques assume that there is a measuring instrument which is much more accurate than the sensor that we are trying to calibrate. However, if we perform the measurements with the most accurate sensor available, then this sensor is already the most accurate, there

is no more accurate one. In this case, the best we can do is to get some theory-justified upper bound on the measurement error.

But why not assume a uniform distribution? In practice, when all we know is the upper bound on the measurement error, people often assume that the measurement error is uniformly distributed on the corresponding interval $[-\Delta, \Delta]$. At first glance, this assumption makes perfect sense: if we have no reason to assume that some values from this interval, are more probable than others, then it is reasonable to assume that all the values are equally probable, i.e., that we have a uniform distribution. This argument goes back to Laplace and is known as *Laplace Indeterminacy Principle*; see, e.g., [3].

However, one of the main purposes of metrology is to produce *guaranteed* information about the measured values of different physical quantities. From this viewpoint, as we will show on a simple example, selecting a uniform distribution can lead to misleading conclusions.

Let us consider the simplest possible relation between the desired quantity y and the easier-to-measure quantities $x_1, \ldots, x_n$:

$$y = x_1 + \ldots + x_n.$$

Let us also assume, for simplicity, that the measured values $\widetilde{x}_1, \ldots, \widetilde{x}_n$ of all these quantities are 0s, and that for each of these measurements, the only information that we have about the possible values of the measurement error $\Delta x_i = \widetilde{x}_i - x_i$ is that this value is bounded by a given positive number Δ. In this case, all we know about each actual value x_i is that this value is located in the interval $[-\Delta, \Delta]$. In this situation, what can we say about possible values of y?

Since each of the values x_i is bounded from above by Δ, the sum of n such terms cannot exceed $n \cdot \Delta$. On the other hand, it is possible that each value x_i is equal to Δ, in which case their sum is exactly $n \cdot \Delta$.

Similarly, since each of the values x_i is bounded from below by $-\Delta$, the sum of n such terms cannot be smaller than $-n \cdot \Delta$. On the other hand, it is possible that each value x_i is equal to $-\Delta$, in which case their sum is exactly $-n \cdot \Delta$. So, the range of possible values of y is the interval $[-n \cdot \Delta, n \cdot \Delta]$.

What if we assume that measurement errors are independent and uniformly distributed on the interval $[-\Delta, \Delta]$? Each of these distributions has mean 0 and variance $\Delta^2/3$. For large n, according to the Central Limit Theorem (see, e.g., [8]), the distribution of y is close to Gaussian. For this Gaussian distribution, the mean m is the sum of the means, i.e., $m = 0$, and the variance σ^2 is the sum of the variances, i.e., $\sigma^2 = n \cdot \Delta^2/3$. For Gaussian distribution, with high confidence, we usually conclude that the actual value is located in the interval

$$[m - k \cdot \sigma, m + k \cdot \sigma]:$$

- for $k = 3$, this is true with confidence 99.9%;
- for $k = 6$, this is true with confidence $1 - 10^{-8}$.

So, with high confidence, we conclude that the actual value y is located in the interval

$$\left[-\sqrt{n} \cdot \frac{\Delta}{\sqrt{3}}, \sqrt{n} \cdot \frac{\Delta}{\sqrt{3}}\right].$$

So, the upper bound resulting from the uniform-distribution assumption is proportional to $\sqrt{n}$, while the actual bound grows as n. For large n, $\sqrt{n}$ is much smaller than n. So, even in this very simple example, the uniform-distribution assumption drastically underestimates the inaccuracy of the data processing result – which explains why we will not use this assumption in our analysis.

Possibility of linearization. Measurement errors are usually relatively small, so terms quadratic (or higher order) with respect to measurement errors are much smaller than the linear terms. For example, even for a not very accurate measurement, for which the measurement error is about 10%, the square of this error is 1%, which is much smaller than 10%. For more accurate measurements, the ratio of quadratic to linear terms is even smaller. So, we can safely ignore quadratic and higher order terms in the Taylor expansion of the corresponding dependencies, and retain only linear terms.

Let us apply this idea to the error $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$ in the result of processing data caused by measurement errors. Here,

$$\Delta y = \widetilde{y} - y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(x_1, \ldots, x_n). \quad (1)$$

By definition of the measurement error Δx_i, we get $x_i = \widetilde{x}_i - \Delta x_i$. Substituting this expression into the formula (1), we conclude that

$$\Delta y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n).$$

Expanding this expression in Taylor series in terms of Δx_i and keeping only linear terms in this expansion, we conclude that

$$\Delta y = f_1 \cdot \Delta x_1 + \ldots + f_n \cdot \Delta x_n, \quad (2)$$

where we denoted

$$f_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}_{|x_1 = \widetilde{x}_1, \ldots, x_n = \widetilde{x}_n}. \quad (3)$$

Similarly, any constraint

$$g_j(x_1, \ldots, x_n, c_1, \ldots, c_k) \le 0$$

takes the form

$$g_j(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n, \widetilde{c}_1 - \Delta c_1, \ldots, \widetilde{c}_k - \Delta c_k) \le 0.$$

Expanding this expression in Taylor series in terms of Δx_i and keeping only linear terms in this expansion, we conclude that

$$\widetilde{g}_j + g_{j1} \cdot \Delta x_1 + \ldots + g_{jn} \cdot \Delta x_n + m_{j1} \cdot \Delta c_1 + \ldots + m_{jk} \cdot \Delta c_k \le 0, \quad (4)$$

where we denoted

$$\widetilde{g}_j \stackrel{\text{def}}{=} g_j(\widetilde{x}_1, \ldots, \widetilde{x}_n, \widetilde{c}_1, \ldots, \widetilde{c}_k),$$

$$g_{ji} \stackrel{\text{def}}{=} -\frac{\partial g_j}{\partial x_i}_{|x_1 = \widetilde{x}_1, \ldots, x_n = \widetilde{x}_n, c_1 = \widetilde{c}_1, \ldots, c_k = \widetilde{c}_k}, \quad (5)$$

$$m_{ji} \stackrel{\text{def}}{=} -\frac{\partial g_j}{\partial c_i}_{|x_1 = \widetilde{x}_1, \ldots, x_n = \widetilde{x}_n, c_1 = \widetilde{c}_1, \ldots, c_k = \widetilde{c}_k}.$$

Now, we are ready to describe our general methodology for taking into account relation between the variables.

3. HOW TO USE KNOWN RELATIONS BETWEEN QUANTITIES TO MAKE MEASUREMENTS MORE ACCURATE: CASE WHEN WE SIMPLY MEASURE THE QUANTITIES

Description of the problem. We measure n quantities $x_1, \ldots, x_n$ and get n measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$. For each measurement i, we know the upper bound Δ_i on the absolute values of the measurement error $\Delta x_i = \widetilde{x}_i - x_i$. In this case, we can conclude that each actual value x_i is located in the interval $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

Suppose that we also know the relations between these quantities

$$g_1(x_1, \ldots, x_n, c_1, \ldots, c_k) \le 0,$$

$$\ldots,$$

$$g_r(x_1, \ldots, x_n, c_1, \ldots, c_k) \le 0.$$

As we have mentioned in the previous section, since the measurement errors are usually small, we can describe these relations in a simpler form (4). How can we use this info to get more accurate estimates for the quantities $x_1, \ldots, x_n$?

How we can solve this problem. For each i, the smallest possible value of x_i can be obtained by solving the following constrained optimization problem:

Minimize $\widetilde{x}_i - \Delta x_i$ under the following constraints:

$$\begin{aligned} -\Delta_1 \le \Delta x_1 \le \Delta_1, \ldots, -\Delta_n \le \Delta x_n \le \Delta_n, \\ \widetilde{g}_1 + g_{11} \cdot \Delta x_1 + \ldots + g_{1n} \cdot \Delta x_n + \\ m_{11} \cdot \Delta c_1 + \ldots + m_{1k} \cdot \Delta c_k \le 0, \\ \ldots, \\ \widetilde{g}_1 + g_{r1} \cdot \Delta x_1 + \ldots + g_{jn} \cdot \Delta x_n + \\ m_{r1} \cdot \Delta c_1 + \ldots + m_{rk} \cdot \Delta c_k \le 0. \end{aligned} \tag{6}$$

To find the largest possible value of x_i, we need to maximize $\widetilde{x}_i - \Delta x_i$ under the same constraints (6).

In both problems, we optimize the value of a linear expression under linear constraints. Such problems are known as problems of *linear programming*; several efficient algorithms are known for solving these problems; see, e.g., [9].

Once we use linear programming techniques and find:

- the solution $\underline{x}_i$ to the minimization problem and
- the solution $\overline{x}_i$ to the maximization problem,

we can conclude that the range of possible values of x_i is equal to $[\underline{x}_i, \overline{x}_i]$.

Comment. A simple example that we used in Section 1 shows that this can indeed lead to a drastic improvement of accuracy.

4. HOW TO USE KNOWN RELATIONS BETWEEN QUANTITIES TO MAKE MEASUREMENTS MORE ACCURATE: CASE OF DATA PROCESSING

Description of the problem. We are interested in the value of a difficult-to-measure quantity y that is related to several easier-to-measure quantities $x_1, \ldots, x_n$ by a known relation $y = f(x_1, \ldots, x_n)$. We measure n quantities $x_1, \ldots, x_n$ and get n measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$. We plus in these measurement results into the algorithm f and get the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$. What can we conclude about the possible values of the estimation error $\Delta y = \widetilde{y} - y$?

We assume that for each measurement i, we know the upper bound Δ_i on the absolute values of the measurement error $\Delta x_i = \widetilde{x}_i - x_i$. In this case, the estimation error is determined by the formula (2). In the absence of any other information, all we can conclude about Δy is that $|\Delta y| \le \Delta$, where we denoted

$$\Delta \stackrel{\text{def}}{=} |f_1| \cdot \Delta_1 + \ldots + |f_n| \le \Delta_n.$$

Suppose now that we also know the relations between these quantities

$$g_1(x_1, \ldots, x_n, c_1, \ldots, c_k) \le 0,$$

$$\ldots,$$

$$g_r(x_1, \ldots, x_n, c_1, \ldots, c_k) \le 0.$$

As we have mentioned in the previous section, since the measurement errors are usually small, we can describe these relations in the simpler form (4). How can we use this info to get more accurate estimates for the quantities $x_1, \ldots, x_n$?

How we can solve this problem. The smallest possible value of Δy can be obtained by solving the following constrained optimization problem:

$$\text{Minimize } \widetilde{y} - (f_1 \cdot \Delta x_1 + \ldots + f_n \cdot \Delta x_n)$$

under the constraints (6).

To find the largest possible value of Δy, we need to maximize the same expression

$$\widetilde{y} - (f_1 \cdot \Delta x_1 + \ldots + f_n \cdot \Delta x_n)$$

under the same constraints (6).

Both problems are particular cases of linear programming, so we can apply efficient linear programming algorithms to solve these problems. Once we use linear programming techniques and find:

- the solution $\underline{y}$ to the minimization problem and
- the solution $\overline{y}$ to the maximization problem,

we can conclude that the range of possible values of y is equal to $[\underline{y}, \overline{y}]$.

5. HOW TO USE KNOWN RELATIONS BETWEEN QUANTITIES TO MAKE MEASUREMENTS MORE RELIABLE

Formulation of the problems. A sensor can malfunction. As a result, the value generated by this sensor will, in general, become very different from the actual value of the measured quantity. This leads to two natural questions.

The first natural question is: How can we detect that one of the sensors malfunctioned? If we can detect this, this will make the measurement results more reliable.

And the second natural question is: once we detect that one of the sensors malfunctioned, how can we determine which sensor malfunctioned?

How to detect that one of the sensors malfunctioned: analysis of the problem. Suppose that, as in our first numerical example, that we measure two related quantities x_1 and x_2 by two sensors, for each of which the upper bound on the absolute value of the measurement error is $\Delta_1 = \Delta_2 = 0.1$. Suppose that we know that the actual values x_1 and x_2 of two quantities cannot differ by more than 0.01: $|x_1 - x_2| \le 0.01$. Suppose that, as in our example, the first sensor produced the value $\widetilde{x}_1 = 0.9$, but the second sensor malfunctioned and generated the value $\widetilde{x}_2 = 2.0$ – which is far away from the actual (unknown) value x_2. In this case, we cannot find two values x_1 and x_2 for which $|x_1 - 0.9| \le 0.1$, $|x_2 - 2.0| \le 0.1$, and $|x_1 - x_2| \le 0.01$. Indeed, in this case, $x_1 \le 1.0$, $x_2 \ge 1.9$ and thus, the difference $x_2 - x_1$ is larger than or equal to $1.9 - 1.0 = 0.9$ – much larger than 0.01.

This is a typical situation: if one of the sensors malfunctions, it is highly probable that the system of constraints (6) can no longer be satisfied. So, we arrive at the following suggestion.

How to detect that one of the sensors malfunctioned: idea. If the system of inequalities (6) is inconsistent, this means that oe of the sensors malfunctioned.

Checking consistency of a system of linear inequalities can be done by the same efficient linear programming algorithms that we mentioned earlier.

How can we determine which sensor malfunctioned. If the system of inequalities (6) is no longer consistent, what we can do is try, for each i from 1 to n, to delete all inequalities involving Δx_i from the system (6) and check whether the resulting reduced system is still consistent.

- When we delete a well-functioning sensor i, the system (6) will still contain the inequalities coming from the malfunctioning sensor and thus, the system (6) will most probably still be inconsistent.
- However, when we delete the malfunctioning sensor, the remaining system will become consistent.

So, we can determine the malfunctioning sensor as the one whose deletion makes the system (6) consistent.

Comments.

- On the qualitative level, this idea is far from being new: it is used a lot. For example, if a thermometer show a temperature of 5 C in an office while in a neighboring office a thermometer shows 20C, this clearly means that one of these sensor malfunctioned. When a car flashes a red light indicating that something may be wrong, mechanics always compare with other measurement results to make sure that it is a real problem, and not a sensor malfunction.
- A similar idea can be used to detect whether the physical systems itself – whose quantities we are measuring – is malfunctioning. Usually, there are some thresholds for the corresponding quantities. For example, for a building, stresses should exceed a certain level; for a chemical plant, temperature in the reactor must lie within given bounds, etc. If some values from the interval $[\underline{y}, \overline{y}]$ that we obtain get outside these bounds, this is an indication that something needs to be done.

6. APPLICATION TO DIGITAL TWINS

Digital twins: a brief description and the main metrology-related challenges. To understand how different effects and different controls will affect a system – be it an airplane, a building, a plant – a good idea is to design an accurate simulating program. Such program are known as *digital twins*.

From the metrological viewpoint, there are two major challenges related to digital twins. The first challenge is related to the fact that we determine the parameters $c_1, \ldots, c_k$ of the corresponding model based on some preliminary measurements. The value c_i are determined based on measurements and are, thus, only known with some uncertainty: the estimated values $\widetilde{c}_i$ used in the design of the digital twin are, in general, somewhat different from the ideal best-fit values c_i; there is an estimation error $\Delta c_i = \widetilde{c}_i - c_i$. As a result, the values predicted by a digital twin are, in general, somewhat different from what we actually observe. However, the current digital twins do not provide us with any bounds on this difference. It is therefore desirable to make digital twins generate not the numerical values – as now – but rather intervals of possible values.

The second major challenges related is that the current digital twins do not learn. As we compare the predictions of the digital twin with how the actual system behaves, we get additional information that can potentially help us make the model more accurate – but in the current digital twin technology, there is no easy way to do it. How can we make a digital twin that learns? There exist effective machine learning tools like deep learning (see, e.g., [1]), but these tools require a lot of data to train, and we rarely have that much data about the actual physical system. So, what can we do?

How to make digital twins learn: main idea. The formulas describing the digital twin can be formulated as constraints. For example, if the digital twin predict the value x_i as $t_i(c_1, \ldots, c_k)$, for some algorithm t_i, this can be described as an equality $x_i = t_i(c_1, \ldots, c_k)$. Now, each measurement of the actual values of each physical quantity results in this equality constraint plus the usual constraint $|x_i - \widetilde{x}_i| \le \Delta_i$. So, to get a better constraint on the values of each of the parameters c_i, we can maximize and minimize each value $\widetilde{c}_i - \Delta c_i$ under the corresponding constraints (6). As we add more and more measurement results, the corresponding bounds $[\underline{c}_i, \overline{c}_i]$ will lead to more and more accurate description of the best-fit parameters c_i.

This will also help digital twins return intervals, and not just numbers without any estimates of their accuracy. This way, digital twins will not only predict approximate estimates – we can also use the above-described ideas to generate the bounds on possible values of $t_i(c_1, \ldots, c_k)$ when each c_i is in the corresponding interval.

How can we make this idea more practical. In the above idea, each additional measurement adds new inequalities to the system (6) of inequalities. As the number of inequalities grows, the computation time needed to solve this system of inequalities also grows, and at some point, it may exceed the computational ability of the computational system.

To make computations realistic, we can use the following natural idea:

- first, we perform a certain number M of measurements, and use the above scheme to find the better bounds

$$\underline{c}_i \le c_i = \widetilde{c}_i - \Delta c_i \le \overline{c}_i$$

on the parameters c_i;

- then, we again start measuring and comparing the measurement results with the predictions of the digital twin; once we have made M new measurements, we form a system (6) that takes into account only these new M measurements, plus uses k inequalities

$$\underline{c}_i \le c_i = \widetilde{c}_i - \Delta c_i \le \overline{c}_i$$

that we determined earlier; solving this system of inequalities leads us to more accurate bounds $\underline{c}_i$ and $\overline{c}_i$;

- then, we again perform M new measurements, etc.

In this case, the number of inequalities that we need to process remains limited.

Comment. Instead of waiting for all M measurements to complete, we can – if we have enough computational power – repeat this procedure every time we get a new measurement, but using only M latest measurement results.

7. WHAT IF WE HAVE PROBABILISTIC UNCERTAINTY?

In the previous sections, we consider the case when we only know the bounds on the measurement errors, and we do not have any information about the probabilities of different values within these bounds. In many practical situations, however, we know the corresponding probability distributions. In such situations, how can we use known relations between the measured quantities to make measurements more accurate?

One possibility is to use Monte-Carlo simulations. Namely, many times, we simulate probability distributions for all n measurement errors Δx_i, but then dismiss all simulated tuples $(\Delta x_1, \ldots, \Delta x_n)$ that do not satisfy the inequalities describing the known relations. Based on remaining tuples, we can get the histograms of the corresponding values Δx_i and/or Δy and thus, get a good understanding of the probability distributions that take into account the relations between the quantities x_i.

FUNDING STATEMENT

This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Focus Program SPP 100+ 2388, Grant Nr. 501624329, by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES), and by the AT&T Fellowship in Information Technology.

It was also supported by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts, 2016.

[2] L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2012.

[3] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

[4] B. J. Kubica, *Interval Methods for Solving Nonlinear Constraint Satisfaction, Optimization, and Similar Problems: from Inequalities Systems to Game Solutions*, Springer, Cham, Switzerland, 2019.

[5] G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.

[6] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.

[7] S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, New York, 2005.

[8] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[9] R. J. Vanderbei, *Linear Programming: Foundations and Extensions*, Springer, New York, 2014.