



**HAL**  
open science

## A Priori Error Analysis and Spring Arithmetic

Gilles Chabert, Luc Jaulin

► **To cite this version:**

Gilles Chabert, Luc Jaulin. A Priori Error Analysis and Spring Arithmetic. *SIAM Journal on Scientific Computing*, 2009, 31 (3), pp.2214-2230. 10.1137/070696982 . hal-00428952

**HAL Id: hal-00428952**

**<https://hal.science/hal-00428952>**

Submitted on 30 Oct 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## A PRIORI ERROR ANALYSIS AND SPRING ARITHMETIC\*

GILLES CHABERT<sup>†</sup> AND LUC JAULIN<sup>‡</sup>

**Abstract.** Error analysis is defined by the following concern: Bounding the output variation of a (nonlinear) function with respect to a given variation of the input variables. This paper investigates this issue in the framework of interval analysis. The classical way of analyzing the error is to linearize the function around the point corresponding to the actual input, but this method is local and not reliable. Both drawbacks can be easily circumvented by a combined use of interval arithmetic and domain splitting. However, because of the underlying linearization, a standard interval algorithm leads to a pessimistic bound, and even simply fails (i.e., returns an infinite error) in case of singularity. We propose an original nonlinear approach where intervals are used in a more sophisticated way through the so-called *springs*. This new structure allows one to represent an (infinite) set of intervals constrained by their midpoints and their radius. The output error is then calculated with a spring arithmetic in the same way as the image of a function is calculated with interval arithmetic. Our method is illustrated on two examples, including an application of geopositioning.

**Key words.** error analysis, interval arithmetic, global optimization

**AMS subject classification.** 65G40

**DOI.** 10.1137/070696982

**1. Introduction.** In this paper, we consider an equation  $y = f(x)$  where  $x$  is a vector of *uncertain* input parameters and  $y$  a vector of outputs. For the sake of generality, *input* and *output* refers here to the mathematical meaning: “input” is used to designate a quantity  $x$  that can be fixed whereas “output” is the quantity  $y$  we want to determine from  $x$  by evaluating  $y = f(x)$ .

Note that from the physical standpoint, these terms may not match. In the context of parameter estimation [4], they would even be assigned in the other way around since the system outputs would correspond to the measured data and the system inputs to the sought parameters.

For the sake of simplicity, we shall assume all through this introduction that  $f$  is a function from  $\mathbb{R}$  to  $\mathbb{R}$ ; i.e.,  $x$  and  $y$  are not vectors.

We consider the situation where the (bounded) *uncertain* input  $x$  can only be fixed up to a given precision vector  $\delta_x$ , say by a measure. Hence, if  $x_m$  is a measure of the real value  $x_r$ , then

$$|x_m - x_r| \leq \delta_x.$$

One fundamental issue is to estimate how this uncertainty impacts the computed output, i.e., the distance between the actual output  $y_r$  (satisfying  $y_r = f(x_r)$ ) and the computed output  $y_m$  (satisfying  $y_m = f(x_m)$ ). This distance is called the *output error*.

The purpose of this paper is to compute an a priori bound of the output error, i.e., before the measure. We focus especially on the reliability and the accuracy of this bound including the case of large input errors.

---

\*Received by the editors July 12, 2007; accepted for publication (in revised form) December 16, 2008; published electronically May 7, 2009.

<http://www.siam.org/journals/sisc/31-3/69698.html>

<sup>†</sup>Ecole des Mines de Nantes LINA CNRS UMR 6241, 4 rue Alfred Kastler 44300 Nantes, France (gilles.chabert@emn.fr).

<sup>‡</sup>ENSIETA, 2 rue François Verny 29806 Brest Cedex 9, France (luc.jaulin@ensieta.fr).

In the case of a nonexplicit model  $f$  (e.g., a numerical program), the problem is usually referred to as *sensitivity analysis* [2] and tackled by statistical methods. In this context, the uncertainty is usually not bounded but described by a distribution of probability. We do not consider such models here. Our mapping  $f$  is a standard formally well-identified function, i.e., a composition of *elementary* functions (exp,  $\sqrt{\cdot}$ , cos, etc.) and usual operators (+, −, ×, etc.).

We shall address this issue in the framework of interval analysis. Next subsection introduces notations related to intervals and an overview of our contribution will be given in the two subsequent subsections.

**1.1. Interval-related notations.** Intervals will either be represented by the infimum-supremum convention

$$[a, b] = \{x \in \mathbb{R}, a \leq x \leq b\}$$

or by the midpoint-radius convention (see, e.g., [13, 9]):

$$\langle m, r \rangle = \{x \in \mathbb{R}, |x - m| \leq r\}.$$

In any case, symbols associated to intervals will be surrounded by brackets. If  $[x]$  is the interval  $[a, b]$  (or  $\langle m, r \rangle$ ), the following characteristics are standard:

---

|               |                      |   |              |
|---------------|----------------------|---|--------------|
| (lower bound) | $[x]^- :=$           | $a$   | $(= m - r),$ |
| (upper bound) | $[x]^+ :=$           | $b$   | $(= m + r),$ |
| (midpoint)    | $\text{mid } [x] :=$ | $(a + b)/2$   | $(= m),$     |
| (radius)      | $\text{rad } [x] :=$ | $(b - a)/2$   | $(= r),$     |
| (magnitude)   | $\text{mag } [x] :=$ | $\max\{ [x]^-- ,  [x]^+ \}$                                 |              |
| (mignitude)   | $\text{mig } [x] :=$ | $\min\{ [x]^-- ,  [x]^+ \}$ if $0 \notin [x]$ , 0 otherwise |              |

The set of intervals is denoted by  $\mathbb{IR}$  and a vector of intervals is often simply called a *box*. A degenerated interval  $[a, a]$  is identified to the real number  $a$ .

We assume the reader to be familiar with interval arithmetic [10, 1, 11, 3].

Given  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , an *interval extension*  $[f]$  of  $f$  is a mapping from  $\mathbb{IR}^n$  to  $\mathbb{IR}^m$  such that

$$\begin{cases} \forall x \in \mathbb{R}^n & f(x) = [f](x) \text{ and} \\ \forall [x] \in \mathbb{IR}^n & f([x]) \subseteq [f]([x]), \end{cases}$$

where  $f([x])$  denotes the set-theoretical image of  $[x]$  by  $f$ .

Moreover, for any set  $\Sigma$ , we will denote  $\square \Sigma$  the smallest box enclosing  $\Sigma$ .

**1.2. A posteriori error analysis and interval arithmetic.** Let us fix  $x_m$ . The output error  $\delta_y(x_m)$  obeys the following definition.

$$(1) \quad \delta_y(x_m) := \sup\{|f(x_m) - f(x_r)|, |x_r - x_m| \leq \delta_x\}.$$

Since the result of a classical linearization is not guaranteed (see section 2.1), one can resort to interval analysis.

In section 2.2, we shall describe two interval methods. The first one is an interval variant of the classical linearization. The second one (called the *nonlinear method*) leads to the following formula: If  $[f]$  is an interval extension of  $f$  (see section 1.1), then

$$(2) \quad \delta_y(x_m) \leq 2 \max\{\text{rad } [f]([x_m - \delta_x, x_m]), \text{rad } [f]([x_m, x_m + \delta_x])\}.$$

Hence, the nonlinear method requires only to enclose the range of a function and interval arithmetic is well-suited for such a purpose.

**1.3. A priori error analysis and spring arithmetic.** In the previous section, the error analysis is made when  $x_m$  is known, i.e., *after* the measures. In a large variety of situations, computing an a priori bound for  $|y_m - y_r|$ , i.e., *before* the measure, is more relevant. Formally, if we denote by  $[x_m]$  the set of all possible measures, we look up now for

$$(3) \quad \hat{\delta}_y := \sup\{\delta_y(x_m), x_m \in [x_m]\}.$$

An upper bound of  $\hat{\delta}_y$  can still be obtained with the interval linear method. However, the result is often not satisfactory (see section 2.2) and one could rather try to extend the nonlinear method. As explained in section 2.3, the *local* bound given by (2) leads to the following *global* formula:

$$(4) \quad \hat{\delta}_y \leq 2 \times \sup\left\{ \begin{array}{l} \text{rad } [f]([x]), \quad \text{mid } [x] \in [x_m] + \langle 0, \delta_x/2 \rangle \\ \text{rad } [x] = \delta_x/2 \end{array} \right\}.$$

Relation (4) can be viewed as a global optimization problem over a set of intervals. To our knowledge, no method exists so far for this kind of problem. A broad outline of our approach is now given.

Since the scope of our method is wider than the problem of computing the bound given by (4), let us first describe a more general situation. On the one hand, the condition  $\text{rad } [x] = \delta_x/2$  can be replaced by  $\text{rad } [x] \in [r]$  where  $[r]$  is an interval. Hence, we deal now with an *uncertain* interval  $[x]$  whose midpoint and radius both belong to intervals. The key idea is to collapse both uncertainties into the same entity called *spring*<sup>1</sup> (see section 3) and denoted by  $\langle [m], [r] \rangle$ , with  $[m] := [x_m] + [-\delta_x/2, \delta_x/2]$ . This spring represents the set of all intervals  $[x]$  satisfying  $\text{mid } [x] \in [m]$  and  $\text{rad } [x] \in [r]$ . Such construction is directly inspired from Kulpa's *diagrammatic approach* of intervals [7] that we shall briefly introduce further. Hence,

$$(5) \quad \hat{\delta}_y \leq 2 \times \sup\{\text{rad } [f]([x]), [x] \in \langle [m], [r] \rangle\}.$$

On the other hand, instead of maximizing the radius of  $[f]$  over this spring, we can consider the set of all intervals described by  $[f]$ , i.e.,  $[f](\langle [m], [r] \rangle)$ . However, since this set can have a complicated shape, we actually look for the smallest spring  $\langle Y \rangle$  enclosing it. The point is that the largest radius in  $\langle Y \rangle$  coincides with the largest radius in  $[f](\langle [m], [r] \rangle)$ ; i.e., the bound given by (4) matches

$$\sup\{\text{rad } [y], [y] \in \langle Y \rangle\}.$$

Now, our method consists in applying a spring arithmetic (see section 3.1) to enclose into a spring the range of  $[f]$  as interval arithmetic allows one to enclose into an interval the range of  $f$ . If  $F$  denotes a *spring extension* of  $[f]$  (see section 3.2), then

$$(6) \quad \langle Y \rangle \subseteq F(\langle [m], [r] \rangle),$$

with an equality if  $\langle [m], [r] \rangle$  is degenerated, i.e., a single interval. Furthermore, if  $F$  is *convergent* (see section 3.3), the enclosure (6) can be made as precise as desired by splitting the input spring.

<sup>1</sup>Since a spring is somehow an iron wire with a variable amplitude, it can be identified to an interval with a variable radius.

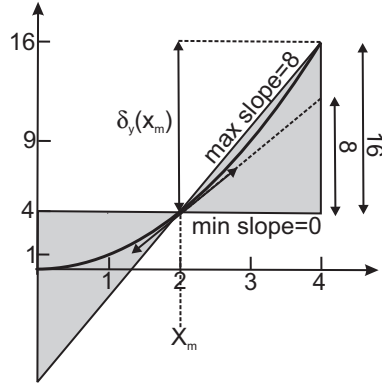


FIG. 1. *Effects of a linearization.*

**2. Error analysis.** First of all, all the previous formulae carry over to a function  $f$  from  $\mathbb{R}^n$  to  $\mathbb{R}^p$  by interpreting absolute values, suprema, inequalities, radii, and midpoints componentwise. In particular, each component of  $f$  is considered independently and we are dealing with vectors  $\delta_y(x_m)$  and  $\hat{\delta}_y$  in  $\mathbb{R}^p$ . More precisely, we are interested in a componentwise safe and accurate upper bound for  $\hat{\delta}_y$ . In addition, the notation  $\langle m, r \rangle$  with vectors  $m$  and  $r$  (e.g.,  $\langle 0, \delta_x \rangle$ ) must also be understood componentwise in the sequel.

We introduce below standard interval-free and interval-based approaches for bounding  $\hat{\delta}_y$ .

**2.1. A standard interval-free approach.** It is common thought that estimating the output error amounts to a simple linearization. Indeed, provided that  $f$  is differentiable and  $\delta_x$  sufficiently small,

$$(7) \quad \delta_y(x_m) \sim |J(x_m)| \cdot \delta_x,$$

where  $J$  denotes the Jacobian matrix of  $f$  (the absolute value is interpreted entrywise). But since approximating  $f$  by a linear mapping is valid only around  $x_m$ , this approximation is not guaranteed. Consider the example of Figure 1. We have  $f(x) = x^2$ ,  $x_m = 2$ , and  $\delta_x = 2$ . Then (7) provides  $|y_m - y_r| \sim 8$  whereas  $x_r = 4$  implies  $|y_m - y_r| = 12$ .

Furthermore, the value of  $\delta_y(x_m)$  can only be approximated with (7) when  $x_m$  is known. This method does not apply for an a priori error analysis.

**2.2. A linear interval-based approach.** Let us first consider the a posteriori error analysis.

Interval arithmetic allows one to make a “rigorous linearization” of  $f$ , providing a reliable bound of  $\delta_y(x_m)$ . Indeed, let us denote by  $[x]$  the box  $\langle x_m, \delta_x \rangle$ . The “hull variant” [12] of the mean value theorem gives

$$(8) \quad \exists J \in \square J([x]) \quad f(x_r) - f(x_m) = J(x_r - x_m).$$

A similar formula can be obtained with *interval slopes* (see, e.g., [11, 14]). The following bound is then derived from the previous formula:

$$(9) \quad \delta_y(x_m) \leq (\text{mag } \square J([x])) \cdot \delta_x,$$

where  $\text{mag}$  is interpreted entrywise. For any interval extension  $[J]$  of the Jacobian matrix (e.g., obtained by automatic differentiation), we therefore have

$$(10) \quad \delta_y(x_m) \leq (\text{mag } [J]([x])) \cdot \delta_x.$$

However, there may be an important lack of accuracy and there are two fundamental reasons for that.

The first reason is related to the use of intervals and the problem can be easily bypassed. On the contrary, the second one needs a deep change in the strategy.

First, substituting  $[J]([x])$  for  $\square J([x])$  may introduce an overestimation. This overestimation is usually related to the multi-incidence of the variables in the expression of  $J$ . This overestimation can however be arbitrarily reduced by splitting the domains as soon as the underlying interval extension of  $J$  is *convergent* (see, e.g., [10] or [11]). Hence, if one split  $[x]$  into a paving  $[x_1], \dots, [x_k]$ , then

$$\delta_y(x_m) \leq (\text{mag } \square_{1 \leq i \leq k} [J]([x_i])) \cdot \delta_x$$

is likely to yield a sharper bound than (10).

The second reason is inherent to the linearization. Relation (9) is often pessimistic because the function is somehow assimilated to a linear mapping with the largest possible slopes. Still in the example of Figure 1, (9) gives  $\delta_y(x_m) \leq (\text{mag } 2 \times [0, 4]) \cdot \delta_x$ , i.e.,  $\delta_y(x_m) \leq 16$ , while in the worst case the variation of  $f$  equals  $4^2 - 2^2 = 12$ .

This loss of accuracy can be arbitrarily large and gets magnified in the multi-variable case. In presence of a singularity, such as  $f(x) = \sqrt{x}$  with  $0 \in [x]$ ,  $f$  is assimilated to a vertical line (the derivative being  $+\infty$ ) and the interval method simply fails, whatever the interval extension is. The same problem arises with interval slopes when the expansion point is chosen near the singularity.

Nevertheless, (10) can straightforwardly be extended to get a bound of the a priori error  $\hat{\delta}_y$ . By considering the box  $[x_m] + \langle 0, \delta_x \rangle$  inside which all  $x_m$  and  $x_r$  belong, we have

$$\hat{\delta}_y \leq (\text{mag } [J]([x_m] + \langle 0, \delta_x \rangle)) \cdot \delta_x.$$

**2.3. A nonlinear interval-based approach.** We propose now a different approach that can be qualified as *nonlinear*. As before, let us first focus on the a posteriori error analysis.

Assume  $n = 1$  (monovariate case). Since both  $f(x_r)$  and  $f(x_m)$  belong to either  $f([x_m - \delta_x, x_m])$  or  $f([x_m, x_m + \delta_x])$ , the distance  $|f(x_r) - f(x_m)|$  is necessarily smaller than the greatest diameter (i.e., twice the greatest radius) of these two intervals (see Figure 2):

$$(11) \quad \delta_y(x_m) \leq 2 \max\{\text{rad } \square f([x_m - \delta_x, x_m]), \text{rad } \square f([x_m, x_m + \delta_x])\}.$$

The generalization for an arbitrary  $n$  is straightforward: the supremum has to be calculated among a set of  $2^n$  boxes obtained by a componentwise combination of intervals of the previous form. Let us call  $\mathcal{S}(x_m, \delta_x)$  this set of boxes. We have

$$(12) \quad \delta_y(x_m) \leq 2 \max_{[x] \in \mathcal{S}(x_m, \delta_x)} \text{rad } \square f([x]).$$

In practice, given an interval extension  $[f]$ , the last inequality implies by inclusion isotonicity:

$$(13) \quad \delta_y(x_m) \leq 2 \max_{[x] \in \mathcal{S}(x_m, \delta_x)} \text{rad } [f]([x]).$$

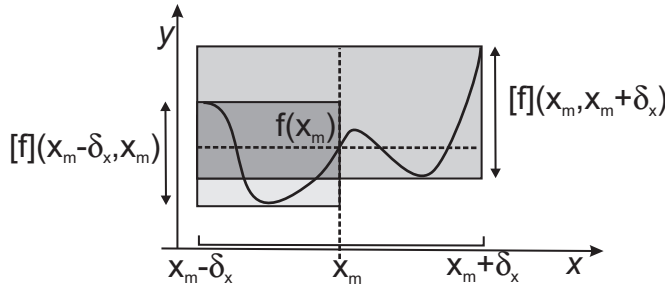


FIG. 2. A posteriori output error with interval enclosures.

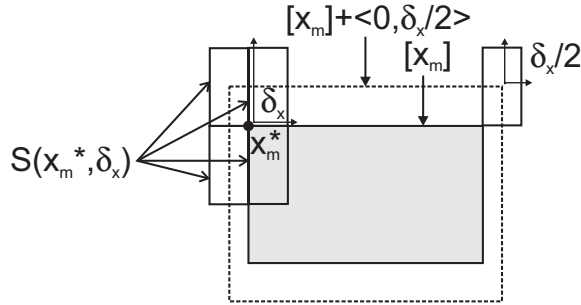


FIG. 3. The set of boxes involved in Formula (14). In two dimensions, the range of  $[f]$  has to be calculated for each  $x_m \in [x_m]$  on four rectangles (this is illustrated with a particular point  $x_m^*$ ). This set is also described by the constraints  $\text{mid } [x] \in [x_m] + \langle 0, \delta_x/2 \rangle$  and  $\text{rad } [x] = \delta_x/2$ .

Let us turn now to the a priori error analysis. Since  $x_m$  ranges over  $[x_m]$ , the overall bound  $\hat{\delta}_y$  satisfies:

$$(14) \quad \hat{\delta}_y \leq \sup_{x_m \in [x_m]} \max_{[x] \in \mathcal{S}(x_m, \delta_x)} \text{rad } [f]([x]).$$

Now, all  $[x]$  in (14) satisfy  $\text{mid } [x] \in [x_m] + \langle 0, \delta_x/2 \rangle$  (the box  $[x_m]$  “enlarged” on each dimension by  $\pm\delta_x/2$ ) and  $\text{rad } [x] = \delta_x/2$ . This is illustrated in Figure 3. Hence,

$$(15) \quad \hat{\delta}_y \leq 2 \times \sup\{\text{rad } [f]([x]), \text{mid } [x] \in [x_m] + \langle 0, \delta_x/2 \rangle \text{ and } \text{rad } [x] = \delta_x/2\}.$$

Therefore, computing an a priori bound  $\hat{\delta}_y$  with (15) requires the ability to maximize the radius produced by  $[f]$  among a set of boxes constrained by their midpoint and their radius. A good way to represent the search space (i.e., the set of all boxes under consideration) is by using springs.

Moreover, in the introduction, the problem has been generalized into the problem of computing a spring enclosure of the range of  $[f]$  over a spring.

Our method for enclosing the range of  $[f]$  is directly inspired by the *natural interval extension*. Let us remind how the latter works. An enclosure of the range of a function  $f$  is obtained with the following induction:

- For each elementary function or operation ( $\exp(x)$ ,  $\sqrt{x}$ ,  $x + y$ , etc.) the range is computed with the interval counterpart ( $\exp[x]$ ,  $\sqrt{[x]}$ ,  $[x] + [y]$ , etc.).
- The range of the compound function  $f$  is built by composing the range of the subexpressions.

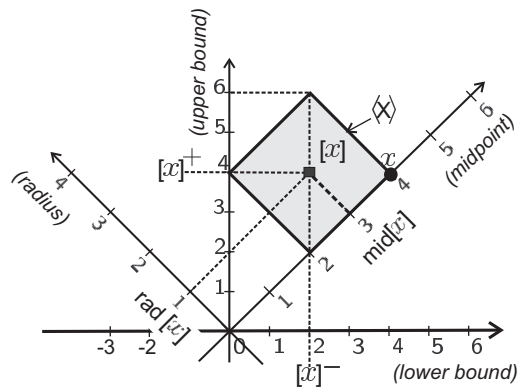


FIG. 4. A diagrammatic representation of springs. Every interval can be represented in a plane by a vector formed with its endpoints. In this plane, the two diagonals define the midpoint-radius frame. Hence, points on the first diagonal (such as  $x = [4, 4]$ ) correspond to degenerated intervals and points on the second one to intervals centered on zero. A spring is a rectangle in the midpoint-radius frame.

- Furthermore, the natural extension of  $f$  is *convergent*; i.e., the overestimation tends to zero with the size of the input box. This means that the overall accuracy can be made arbitrarily high by splitting the domain, as we already mentioned in section 2.2 with the Jacobian matrix.

Let us now get back to the range of the inclusion function  $[f]$ . The exact same induction principle can be used:

- A spring arithmetic is defined to compute the range of addition, subtraction, etc. with interval operands.
- A *natural spring extension* of  $[f]^2$  is then obtained similarly as we have just explained for intervals.
- Convergence comes also with a similar meaning.

In this way, the a priori output error can be calculated by combining the natural spring extension (of an interval extension  $[f]$ ) with splitting.

**3. Springs.** A *spring* is a pair of two intervals  $\langle [m], [r] \rangle$ . This pair represents the set of all intervals  $\langle m, r \rangle$  such that

$$m \in [m] \quad \text{and} \quad r \in [r].$$

A spring can be graphically represented with a rectangle rotated by 45 in the plane—called *diagram*—where an interval  $[a, b]$  is identified to the point  $(a, b)$ . This is shown in Figure 4.

Different interval sets (not including springs but very analogical to) have been thoroughly studied by Kulpa (see, e.g., [5, 7]) who has been developing since the 90's a methodology for proving interval relations based on diagrams. This *diagrammatic* approach has turned to be powerful for relations involving logical connectors and (in)equality symbols between endpoints. These relations are, in particular, at the core

<sup>2</sup>Of course, the expression  $[f]([x])$  must be a composition of elementary functions, arithmetic operators, and (in addition) interval operators such as midpoint, radius, etc. In particular, the natural spring extension cannot be applied to the mean value extension [11] if the Jacobian matrix results from a black box algorithm.



of temporal logic. It has also helped to prove new properties on interval arithmetics [6, 8].

Our approach allows one to deal with relations involving nonlinear expressions and radii such as (5).

Since the purpose here is to define “intervals of intervals,” one could legitimately wonder if a definition based on the infimum-supremum representation (rather than the midpoint-radius one) would not be more suitable.

Indeed, back to regular intervals, the appropriateness of one representation in comparison with the other has been a matter of study.<sup>3</sup> However, the question of which representation to choose grows out of the fact that they are mathematically equivalent, both characterizing the exact same object (an interval). The difference lies in the computational cost of the arithmetic built on top of these. Now, an “interval of intervals” represents a completely distinct object depending on whether the endpoints or the midpoints/radii vary. In the diagram of Figure 4, it is either a plain rectangle or a “lozenge,” respectively. In the context of a priori error analysis, it is clear that only the latter option makes sense. The potential efficiency of an arithmetic with the former option would clearly be counterbalanced by the wrapping effect when switching back and forth from one structure to the other. At all events, calculation speed is prior importance in this paper, the focus being rather on accuracy.

We shall use capital letters with angle brackets (e.g.,  $\langle X \rangle$ ) to denote a spring. The set of all springs will be represented by the following symbol:  $\langle \mathbb{IR} \rangle$ . Let  $\langle X \rangle = ([d], [r])$  be a spring. The following definitions come naturally:

$$\begin{aligned} \text{mid } \langle X \rangle &= [m] \\ \text{rad } \langle X \rangle &= [r] \end{aligned}$$

so that

$$[x] \in \langle X \rangle \iff (\text{mid } [x] \in \text{mid } \langle X \rangle) \wedge (\text{rad } [x] \in \text{rad } \langle X \rangle).$$

Applying the definition of inclusion in terms of membership leads to

$$\langle X \rangle \subseteq \langle Y \rangle \iff (\text{mid } \langle X \rangle \subseteq \text{mid } \langle Y \rangle) \wedge (\text{rad } \langle X \rangle \subseteq \text{rad } \langle Y \rangle).$$

The magnitude of a spring  $\langle [m], [r] \rangle$  can be defined as the set of the magnitudes of all  $[x] \in \langle X \rangle$ . We have

$$(16) \quad \text{mag } \langle [m], [r] \rangle = [( \text{mig } [m] ) + [r]^- , ( \text{mag } [m] ) + [r]^+].$$

*Remark.*  $\text{mag } \langle [m], [r] \rangle$  is not equal to  $\{ |x|, x \in ([m] + [-1, 1] \times [r]) \}$ . Indeed, take  $\langle X \rangle := \langle [0, 3], [1, 2] \rangle$ . The interval  $[x]$  in  $\langle X \rangle$  with the smallest magnitude is  $[-1, 1]$  (since the radius of  $[x]$  must be greater or equal to 1). Therefore the lower bound of  $\text{mag } \langle X \rangle$  is 1, and not 0. This is illustrated in Figure 5.

A number of properties of springs could be exhibited. We shall, however, consider only here what is needed for the aim of error analysis, i.e., a spring arithmetic.

---

<sup>3</sup>Basically, it turns out that the midpoint-radius representation is less efficient than the endpoint-based one if the interval multiplication (and division) has to match the power set counterpart. However, by slightly loosing the result of interval multiplication, the midpoint-radius representation can be made faster and more compliant with parallel computers (see [15]). The representations also have different properties in regard to floating-point issues (not considered in this paper).

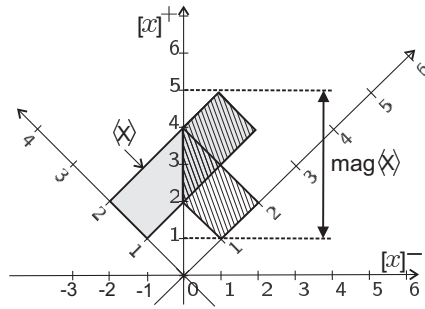


FIG. 5. *Magnitude of a spring. The hatched area represents the set  $\mathcal{Y}$  of intervals with lower bound  $[x]^-$  and upper bound  $[x]^+$  as  $[x]$  describes the spring  $\langle X \rangle := \langle [0, 3], [1, 2] \rangle$ . Since  $\text{mag } [x] = \max\{|[x]^-, |[x]^+|\}$  and since  $\mathcal{Y}$  is above the diagonal, the set of all magnitudes coincides with  $\{[x]^+, [x] \in \mathcal{Y}\}$ , i.e.,  $[1, 5]$ .*

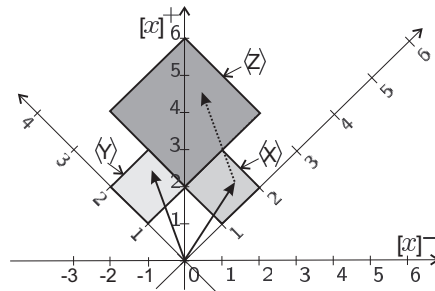


FIG. 6. *Addition of springs. Since the addition of two intervals matches the addition of two vectors in the diagram and since rectangles are preserved by addition, the (power set) sum of two springs (here,  $\langle X \rangle := \langle [1, 2], [0, 1] \rangle$  and  $\langle Y \rangle := \langle [0, 1], [1, 2] \rangle$ ) coincide with a spring (here,  $\langle Z \rangle := \langle [1, 3], [1, 3] \rangle$ ).*

**3.1. Spring arithmetic.** A spring arithmetic can be derived from the interval arithmetic. For any binary operator  $\circ$  We define

$$\langle X \rangle \circ \langle Y \rangle = \diamond\{[z] \in \mathbb{IR} \mid \exists [x] \in \langle X \rangle, \exists [y] \in \langle Y \rangle, [z] = [x] \circ [y]\},$$

where  $\diamond$  stands for the smallest spring (according to the inclusion above) containing the set of intervals in argument.

An explicit formula for addition and subtraction of springs is derived very intuitively from their interval counterparts (see Figure 6):

PROPOSITION 1 (addition and subtraction). *Let  $\langle X \rangle = ([m_x], [r_x])$  and  $\langle Y \rangle = ([m_y], [r_y])$  be two springs.*

$$\begin{aligned} \langle X \rangle + \langle Y \rangle &= ([m_x] + [m_y], [r_x] + [r_y]) \\ \langle X \rangle - \langle Y \rangle &= ([m_x] - [m_y], [r_x] + [r_y]). \end{aligned}$$

*Proof.* For every  $[x] \in \langle X \rangle$  and  $[y] \in \langle Y \rangle$  we have  $\text{mid}([x] \pm [y]) = (\text{mid } [x]) \pm (\text{mid } [y]) \in [m_x] \pm [m_y]$  and similarly  $\text{rad}([x] \pm [y]) = (\text{rad } [x]) + (\text{rad } [y]) \in [r_x] + [r_y]$ . The converse inclusion holds since the interval addition or subtraction  $[m_x] \pm [m_y]$  and  $[r_x] + [r_y]$  makes no overestimation.  $\square$

Per contra, the extension of multiplication (and division) to springs cannot be obtained so easily because of the combined effect of radii and midpoints such operations

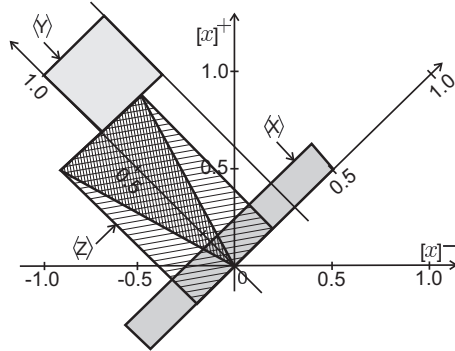


FIG. 7. *Multiplication of springs. The (power set) multiplication of the springs  $\langle X \rangle := \langle [-0.5, 0.5], [0, 0.2] \rangle$  and  $\langle Y \rangle := \langle [0, 0.3], [0.7, 1.0] \rangle$  is not a spring but the set of intervals with a triangle shape. The smallest spring enclosing this set is  $\langle Z \rangle := \langle [-0.21, 0.21], [0, 0.7] \rangle$ . It is defined as the (spring) product of  $\langle X \rangle$  and  $\langle Y \rangle$ .*

involve. This explains why the next proposition is considerably more complicated than the previous one. Note that the spring multiplication introduces a wrapping effect, contrary to interval arithmetic (where the result of each operation coincide with the exact range). This is illustrated in Figure 7. We will skip the division.

PROPOSITION 2 (multiplication). *Let  $\langle X \rangle = ([m_x], [r_x])$  and  $\langle Y \rangle = ([m_y], [r_y])$  be two springs. Define  $x_1, \dots, x_4$  and  $y_1, \dots, y_4$  whose values are endpoints of  $[x_m]$  and  $[y_m]$ , respectively (or, in addition, 0 in the case of  $x_1$  and  $y_1$ ), such that*

$$\begin{aligned} |x_1| &= \text{mig } [m_x], & |y_1| &= \text{mig } [m_y], \\ |x_2| &= \text{mag } [m_x], & |y_2| &= \text{mag } [m_y], \\ x_3 y_3 &= \min\{[m_x]^- [m_y]^-, [m_x]^- [m_y]^+, [m_x]^+ [m_y]^-, [m_x]^+ [m_y]^+\}, \\ x_4 y_4 &= \max\{[m_x]^- [m_y]^-, [m_x]^- [m_y]^+, [m_x]^+ [m_y]^-, [m_x]^+ [m_y]^+\}. \end{aligned}$$

Define also

$$\begin{aligned} [x_1] &:= \langle x_1, [r_x]^- \rangle & [y_1] &:= \langle y_1, [r_y]^- \rangle, \\ [x_2] &:= \langle x_2, [r_x]^+ \rangle & [y_2] &:= \langle y_2, [r_y]^+ \rangle, \\ [x_3] &:= \langle x_3, [r_x]^+ \rangle & [y_3] &:= \langle y_3, [r_y]^+ \rangle && \text{if } x_3 y_3 \leq 0, \\ [x_3] &:= \langle x_3, [r_x]^- \rangle & [y_3] &:= \langle y_3, [r_y]^- \rangle && \text{otherwise.} \\ [x_4] &:= \langle x_4, [r_x]^- \rangle & [y_4] &:= \langle y_4, [r_y]^- \rangle && \text{if } x_4 y_4 \leq 0, \\ [x_4] &:= \langle x_4, [r_x]^+ \rangle & [y_4] &:= \langle y_4, [r_y]^+ \rangle && \text{otherwise.} \\ [z_1] &:= [x_1] \times [y_1] & [z_2] &:= [x_2] \times [y_2], \\ [z_3] &:= [x_3] \times [y_3] & [z_4] &:= [x_4] \times [y_4]. \end{aligned}$$

Then

$$\langle X \rangle \times \langle Y \rangle = \langle [\text{mid } [z_3], \text{mid } [z_4]], [\text{rad } [z_1], \text{rad } [z_2]] \rangle.$$

*Proof.* Our proof relies on the following formulas of interval multiplication that can be all found in [11, p. 23].

- (a)  $\text{mid } ([x] \times [y]) = (\text{mid } [x])(\text{mid } [y]) + \text{sign}((\text{mid } [x])(\text{mid } [y])) \times \inf((\text{rad } [x])|\text{mid } [y]|, (\text{rad } [y])|\text{mid } [x]|, (\text{rad } [x])(\text{rad } [y]))$
- (b)  $\text{rad } ([x] \times [y]) \geq (\text{mag } [x])(\text{rad } [y])$
- (c)  $\text{rad } ([x] \times [y]) \leq (\text{mag } [x])(\text{rad } [y]) + |\text{mid } [y]|(\text{rad } [x])$
- (d)  $\text{rad } ([x] \times [y]) = |\text{mid } [x]|(\text{rad } [y]) + |\text{mid } [y]|(\text{rad } [x])$  if  $0 \notin [x]$  and  $0 \notin [y]$
- (e)  $\text{rad } ([x] \times [y]) = (\text{mag } [x])(\text{rad } [y])$  or  $(\text{mag } [y])(\text{rad } [x])$  if  $0 \in [x]$  or  $0 \in [y]$

**Lower bound for the radius.** Assume first  $\text{mig } [m_x] > [r_x]^-$  and  $\text{mig } [m_y] > [r_y]^-$ . Then 0 is neither in  $[x_1]$  nor in  $[y_1]$  and so it is for every box  $[x] \times [y] \in \langle X \rangle \times \langle Y \rangle$  satisfying  $\text{rad } [x] = [r_x]^-$  and  $\text{rad } [y] = [r_y]^-$ . Using (d), we have

$$\text{rad } ([x] \times [y]) = |\text{mid } [x]|(\text{rad } [y]) + |\text{mid } [y]|(\text{rad } [x]).$$

Hence,

$$\text{rad } ([x] \times [y]) \geq (\text{mig } [m_x])[r_y]^- + (\text{mig } [m_y])[r_x]^-$$

i.e.,

$$\text{rad } ([x] \times [y]) \geq \text{rad } [z_1].$$

Assume now that either  $\text{mig } [m_x] \leq [r_x]^-$  or  $\text{mig } [m_y] \leq [r_y]^-$ . Then, either  $0 \in [x_1]$  or  $0 \in [y_1]$ . Hence, by (e), either  $\text{rad } [z_1] = (\text{mag } [x_1])(\text{rad } [y_1])$  or  $\text{rad } [z_1] = (\text{mag } [y_1])(\text{rad } [x_1])$ . In the first case, since

$$\text{rad } ([x] \times [y]) \geq (\text{mag } [x])(\text{rad } [y])$$

by (b), and since  $[x_1]$  is the interval of  $\langle X \rangle$  with the smallest magnitude and radius then  $\text{rad } ([x] \times [y]) \geq \text{rad } [z_1]$ . The second case symmetrically leads to the same inequality. Finally, if we denote by  $\langle Z \rangle = \langle [m_z], [r_z] \rangle$  the result of  $\langle X \rangle \times \langle Y \rangle$ , then  $[r_z]^- = \text{rad } [z_1]$ .

**Upper bound for the radius.** For every box  $([x] \times [y]) \in \langle X \rangle \times \langle Y \rangle$  satisfying  $\text{rad } [x] = [r_x]^+$  and  $\text{rad } [y] = [r_y]^+$ , by (c) and (16) we have:

$$\text{rad } ([x] \times [y]) \leq ((\text{mag } [m_x]) + [r_x]^+)[r_y]^+ + (\text{mag } [m_y])[r_x]^+$$

Put  $x_2^1 := x_2 - [r_x]^+$ ,  $x_2^2 := x_2 + [r_x]^+$ ,  $y_2^1 := y_2 - [r_y]^+$ , and  $y_2^2 := y_2 + [r_y]^+$ . Since  $\text{rad } ([x_2][y_2]) = \text{rad } ((-x_2)[y_2]) = \text{rad } ([x_2](-y_2))$  we can assume  $x_2 \geq 0$  and  $y_2 \geq 0$  (i.e.,  $x_2^2 y_2^2 - x_2^1 y_2^1 \geq 0$ ) when computing  $\text{rad } [z_2]$ . Then  $0.5 \times (x_2^2 y_2^2 - x_2^1 y_2^1) \in \text{rad } [z_2]$ . Next,

$$\begin{aligned} 0.5 \times (x_2^2 y_2^2 - x_2^1 y_2^1) &= 0.5 \times ((x_2 + [r_x]^+)(y_2 + [r_y]^+) - (x_2 - [r_x]^+)(y_2 - [r_y]^+)) \\ &= (x_2[r_y]^+ + y_2[r_x]^+ + [r_x]^+[r_y]^+) \\ &= (\text{mag } [m_x])[r_y]^+ + (\text{mag } [m_y])[r_x]^+ + [r_x]^+[r_y]^+ \\ &\geq \text{rad } ([x] \times [y]), \end{aligned}$$

which means that  $[r_z]^+ = \text{rad } [z_2]$ .

**Lower and upper bound for the midpoint.** Assume first that  $x_4 y_4 \geq 0$ . Notice that in this case,  $\forall (x, y) \in \text{mid } [x] \times \text{mid } [y]$  such that  $\text{sign } (xy) = 1$ , we have  $|x_4| \geq |x|$  and  $|y_4| \geq |y|$ . Consider now a box  $[x] \times [y] \in \langle X \rangle \times \langle Y \rangle$ . If  $\text{sign } (\text{mid } [x] \times \text{mid } [y]) = +1$ , then, by using (a) with the previous remark,

$$\begin{aligned} x \text{mid } ([x] \times [y]) &\leq (\text{mid } [x])(\text{mid } [y]) \\ &\quad + \inf((\text{rad } [x])|\text{mid } [y]|, (\text{rad } [y])|\text{mid } [x]|, (\text{rad } [x])(\text{rad } [y])) \\ &\leq x_4 y_4 + \inf([r_x]^+|y_4|, [r_y]^+|x_4|, [r_x]^+[r_y]^+) \\ &\leq \text{mid } [z_4] \end{aligned}$$

If  $\text{sign } (\text{mid } [x] \times \text{mid } [y]) = -1$ , then  $\text{mid } ([x] \times [y]) \leq (\text{mid } [x])(\text{mid } [y]) \leq x_4 y_4 \leq \text{rad } [z_4]$ .

Assume now that  $x_4 y_4 \leq 0$ . Then  $\text{sign}(\text{mid}[x] \times \text{mid}[y])$  can only be  $-1$ . We have

$$\begin{aligned} \text{mid}([x] \times [y]) &\leq (\text{mid}[x])(\text{mid}[y]) \\ &\quad - \inf((\text{rad}[x])|\text{mid}[y]|, (\text{rad}[y])|\text{mid}[x]|, (\text{rad}[x])(\text{rad}[y])) \\ &\leq x_4 y_4 + \inf([r_x]^- |y_4|, [r_y]^- |x_4|, [r_x]^- [r_y]^-) \\ &\leq \text{mid}[z_4]. \end{aligned}$$

Hence,  $[m_z]^+ = \text{mid}[z_4]$ . The lower bound for the midpoint is obtained similarly.  $\square$

**3.2. Elementary functions with spring argument.** The definition of elementary functions with spring argument follows the same principle.<sup>4</sup>

DEFINITION 1 (elementary function with spring argument). *Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be an elementary function. The homonym function  $f : \langle \mathbb{IR} \rangle \rightarrow \langle \mathbb{IR} \rangle$  with a spring argument satisfies*

$$\forall \langle X \rangle \in \langle \mathbb{IR} \rangle \quad f(\langle X \rangle) \supseteq \diamond\{[y] \in \mathbb{IR} \mid \exists [x] \in \langle X \rangle, [y] = \square f([x])\}.$$

All the elementary functions (sqr, sqrt, cos, exp, etc.) can be built with spring arguments by considering their well-known properties of variation (monotonicity, symmetry, periodicity). We first illustrate our purpose by considering a convex and increasing function (e.g., the exponential function). Next, we will give the formula for cosine (skipping the proof for the sake of concision).

PROPOSITION 3. *Let  $f : \mathbb{R} \mapsto \mathbb{R}$  such that for all  $x \in \mathbb{R}$ ,  $f'(x) \geq 0$  and  $f''(x) \geq 0$ . For all spring  $\langle X \rangle := \langle [m], [r] \rangle$ , we have*

$$f(\langle X \rangle) = \langle [\text{mid}[y_1], \text{mid}[y_2]], [\text{rad}[y_1], \text{rad}[y_2]] \rangle,$$

where

$$\begin{aligned} [y_1] &= f(\langle [m]^-, [r]^- \rangle), \\ [y_2] &= f(\langle [m]^+, [r]^+ \rangle). \end{aligned}$$

*Proof.* Let  $\langle X \rangle = \langle [m], [r] \rangle$  be a spring. Take  $m \in [m]$  and  $r \in [r]$ . First,  $f$  is increasing so for any  $x$  and  $y$  in  $\langle m, r \rangle$ ,  $|f(y) - f(x)| \leq f(m+r) - f(m-r)$ . Now, since  $f'$  is increasing, we have

$$f(m+r) - f(m-r) = \int_{m-r}^{m+r} f'(x) dx \leq \int_{m-r}^{m+r} f'(x + ([m]^+ - m)) dx.$$

By changing the bounds of integration and then using the positivity of  $f'$ ,

$$f(m+r) - f(m-r) \leq \int_{[m]^+ - r}^{[m]^+ + r} f'(x) dx \leq \int_{[m]^+ - [r]^+}^{[m]^+ + [r]^+} f'(x) dx.$$

Hence,

$$\forall x, y \in \langle m, r \rangle \quad |f(y) - f(x)| \leq f([m]^+ + [r]^+) - f([m]^+ - [r]^+),$$

---

<sup>4</sup>To avoid any confusion, we shall not call such function *spring extensions* since a spring extension is related to interval (not real) functions.

i.e.,

$$\text{rad } f(\langle m, r \rangle) \leq \text{rad } f(\langle [m]^+, [r]^+ \rangle) = \text{rad } [y_2].$$

Similarly,

$$\begin{aligned} [m]^+ \geq m &\implies [m]^+ - m \geq 0 \implies [m]^+ - m \geq m - [m]^+ \\ &\implies [m]^+ - m + ([r]^+ - r) \geq m - [m]^+ + ([r]^+ - r) \\ &\implies ([m]^+ + [r]^+) - (m + r) \geq (m - r) - ([m]^+ - [r]^+) \\ &\implies \int_{m+r}^{[m]^+ + [r]^+} f'(x) dx \geq \int_{[m]^+ - [r]^+}^{m-r} f'(x) dx \quad (\text{since } f' \text{ is increasing}) \\ &\implies f([m]^+ + [r]^+) - f(m + r) \geq f(m - r) - f([m]^+ - [r]^+) \\ &\implies 0.5(f([m]^+ + [r]^+) + f([m]^+ - [r]^+)) \geq 0.5(f(m - r) + f(m + r)) \\ &\implies \text{mid } f(\langle [m]^+, [r]^+ \rangle) \geq \text{mid } f(\langle m, r \rangle). \end{aligned}$$

The lower bound for the midpoint and the radius comes with a very similar reasoning.  $\square$

A piecewise analysis inspired by the previous proof allows one to build a spring variant of all elementary functions. We provide here the formula of the cosine function, under the form of algorithm. This formula is, however, only valid when the upper bound of the radius does not exceed  $\pi/2$ .

```

begin function cosine( $\langle [m], [r] \rangle$ ) returns spring
   $[m] \leftarrow [\text{mig } [m], \text{mag } [m]]$ 
2
  if ( $\text{rad } [m] \geq \pi$ )  $[m] \leftarrow [0, \pi]$ 
  else  $[m] \leftarrow [m] - (I([m]^- / 2\pi) \times 2\pi)$ 
5
  if ( $[m]^+ \geq 2\pi$ )  $[m] \leftarrow [0, \max([m]^+ - 2\pi, 2\pi - [m]^-)]$ 
7
  if ( $[m]^- \geq \pi$ )  $[m] \leftarrow 2\pi - [m]$ 
  else if ( $[m]^+ \geq \pi$ )  $[m] \leftarrow [\min(2\pi - [m]^+, [m]^-), \pi]$ 
10
  if ( $[m]^+ \leq \pi/2$ )  $[x_1] \leftarrow \langle [m]^+, [r]^+ \rangle$ 
  else  $[x_1] \leftarrow \langle [m]^+, [r]^- \rangle$ 
  if ( $[m]^- \leq \pi/2$ )  $[x_2] \leftarrow \langle [m]^-, [r]^- \rangle$ 
  else  $[x_2] \leftarrow \langle [m]^-, [r]^+ \rangle$ 
   $[x_3] \leftarrow \langle \min([m]^-, \pi - [m]^+), [r]^- \rangle$ 
  if ( $[m]^- \leq \pi/2$ ) and ( $[m]^+ \geq \pi/2$ )  $[x_4] \leftarrow \langle \pi/2, [r]^+ \rangle$ 
  else if ( $[m]^- \geq \pi/2$ )  $[x_4] \leftarrow \langle [m]^-, [r]^+ \rangle$ 
  else  $[x_4] \leftarrow \langle [m]^+, [r]^+ \rangle$ 
  return  $\langle [\text{mid } \cos([x_1]), \text{mid } \cos([x_2])], [\text{rad } \cos([x_3]), \text{rad } \cos([x_4])] \rangle$ 
end function

```

Lines 1 to 10 reduce  $[m]$  inside  $[0, \pi]$  using periodicity and symmetry. More precisely:

- at line 2  $\longrightarrow [m] \subseteq [0, +\infty)$ ,
- at line 5  $\longrightarrow [m] \subseteq [0, 4\pi]$ ,
- at line 7  $\longrightarrow [m] \subseteq [0, 2\pi]$ ,
- at line 10  $\longrightarrow [m] \subseteq [0, \pi]$ .

**3.3. Spring extension of an interval function.**

DEFINITION 2. Let  $[f] : \mathbb{IR} \rightarrow \mathbb{IR}$  be an interval function. The mapping  $F : \langle \mathbb{IR} \rangle \rightarrow \langle \mathbb{IR} \rangle$  is a spring extension of  $[f]$  if

$$\begin{cases} \forall [x] \in \mathbb{IR} & F([x]) = [f]([x]) \\ \forall \langle X \rangle \in \langle \mathbb{IR} \rangle & \diamond [f]([x]) \subseteq F(\langle X \rangle) \end{cases}$$

The definition of a spring extension is generalized to the multivariable case in the same way as interval extensions, and the *minimal spring extension* of  $f$ , denoted by  $f^\diamond$ , is defined simply by inserting in the previous definition an equality sign in place of  $\subseteq$ .

The natural spring extension  $F$  is inclusion isotone with respect to intervals as the natural interval arithmetic is with respect to real numbers. With an extension of the Hausdorff distance to springs (seen as couples of intervals), one can even prove that this extension is convergent, i.e., that the distance between  $F(\langle X \rangle)$  and  $F(\langle Y \rangle)$  is in “big o” of the distance between  $\langle X \rangle$  and  $\langle Y \rangle$ . Back in the context of a priori error analysis, we can resort to a split-and-eval strategy to improve the accuracy of the result: If  $[m] = [x_m] + \langle 0, \delta_x/2 \rangle$  is split into  $[m_1], \dots, [m_k]$ , then

$$\hat{\delta}_y \leq \max_{1 \leq i \leq k} \{ \text{mag rad } F(\langle [m_i], \delta_x/2 \rangle) \} \leq \text{mag rad } F(\langle [m], \delta_x/2 \rangle).$$

However, the convergence to  $\hat{\delta}_y$  is guaranteed when  $k$  tends to infinity only if the interval extension  $[f]$  is minimal. The accuracy of our method is inevitably conditioned by the accuracy of the underlying interval extension.

**4. Applications.** Let us now compare the nonlinear “spring approach” with the linear approach described in section 2.2. Of course, a simple function such as  $x \mapsto \sqrt{x}$  around 0 would show the defects of a linearization and make our approach better. To offer a more convincing comparison, we have chosen unfavorable conditions on purpose for the spring approach: The functions below are very smooth, almost flat (not involving exponentials nor polynomials, etc.).

**4.1. A typical example.** The problem under interest is related to geopositioning accuracy.

The earth is assimilated to a sphere of radius  $R := 6366.2km$  with an associated frame  $(m, i, j, k)$ , where  $m$  is the center of the sphere,  $k$  and  $i$  vectors pointing, respectively, towards the north pole and the Greenwich meridian (see Figure 8). A point on the surface of the earth is usually localized (say, by a GPS) with spherical coordinates: The *longitude*  $\alpha$  and the *latitude*  $\beta$ .

Assume now that the GPS provides angles with uncertainties bounded by  $(\delta_\alpha, \delta_\beta)$ . The question is as follows. Given the coordinates  $(\alpha, \beta)$  of an object  $p$  returned by a GPS, what is the worst-case error made by calculating the Cartesian coordinates  $(x, y, z)$  of  $p$  in a local frame  $(m_0, i_0, j_0, k_0)$  where  $m_0 = (\alpha_0, \beta_0)$  is another (fixed) point on the surface and  $(i_0, j_0, k_0)$  are vectors pointing, respectively, towards the north pole, the east and the center<sup>5</sup> of the sphere?

By applying standard transformations, one finds out that the vector  $(x, y, z)$  of an object with longitude  $\alpha$  and latitude  $\beta$  in the frame  $(m_0, i_0, j_0, k_0)$  matches  $f(\alpha, \beta)$ ,

---

<sup>5</sup>As it is generally the convention in a submarine context.

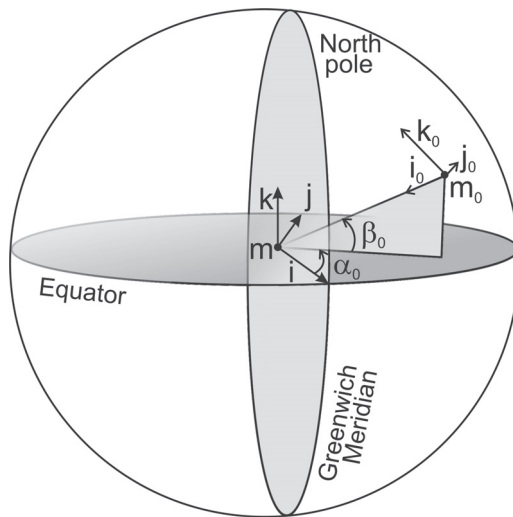


FIG. 8. Localization with longitude and latitude.

with

$$f(\alpha, \beta) := \begin{pmatrix} \sin \beta \cos \beta_0 - \cos \alpha \cos \beta \cos \alpha_0 \sin \beta_0 + \cos \beta \sin \alpha \sin \alpha_0 \sin \beta_0 \\ -\cos \alpha \cos \beta \sin \alpha_0 - \cos \beta \sin \alpha \cos \alpha_0 \\ -\sin \beta \sin \beta_0 - \cos \alpha \cos \beta \cos \alpha_0 \cos \beta_0 + \cos \beta \sin \alpha \cos \beta_0 \sin \alpha_0 \end{pmatrix}.$$

We have detailed in this paper how to compute the addition, subtraction, multiplication, and cosine of springs which are precisely the operations involved in the expression of  $h$  (the extension of the sine function to springs is easily derived from the cosine thanks to the relation  $\sin(x) = \cos(\pi/2 - x)$ ). Next figure shows the results we have obtained with  $\alpha_0 = 40^\circ$  and  $\beta_0 = 50^\circ$  (in degrees). The domains for the midpoints of  $\alpha$  and  $\beta$  are, respectively,  $[-\pi, \pi]$  and  $[-\pi/2, +\pi/2]$ , which means that the result is valid for a point  $p$  lying anywhere on the sphere. The input error bound is  $\delta_\alpha = \delta_\beta = 4.10^{-7}$ . Finally, the output error bound was computed with different values of the splitting precision  $w$ , from  $2^0$  down to  $2^{-9}$ . The best bound we found (i.e., with  $w = 2^{-9}$ ) with spring arithmetic is

$$\hat{\delta}_x \leq 4.084m, \hat{\delta}_y \leq 3.599m, \hat{\delta}_z \leq 3.944m.$$

This result was computed in less than 2 minutes on a standard laptop. It compares advantageously to the result obtained with the linear approach (see Figure 9). The best bound we found with the linear approach is

$$\hat{\delta}_x \leq 4.507m, \hat{\delta}_y \leq 5.101m, \hat{\delta}_z \leq 4.192m,$$

with similar running time. We can see that on the  $y$  coordinate, the bound exceeds by 1.4 meter the one we have got with springs.

**4.2. A dynamical example.** The second example aims at showing the potential of the spring arithmetic in the context of dynamical systems where a priori error analysis is also a key problem (directly connected to, e.g., experiment planning in automation). Consider the following model of a discrete-time dynamical system:

$$\begin{cases} x_0 &= u_0 \\ x_{k+1} &= f(x_k, u_k), \end{cases}$$



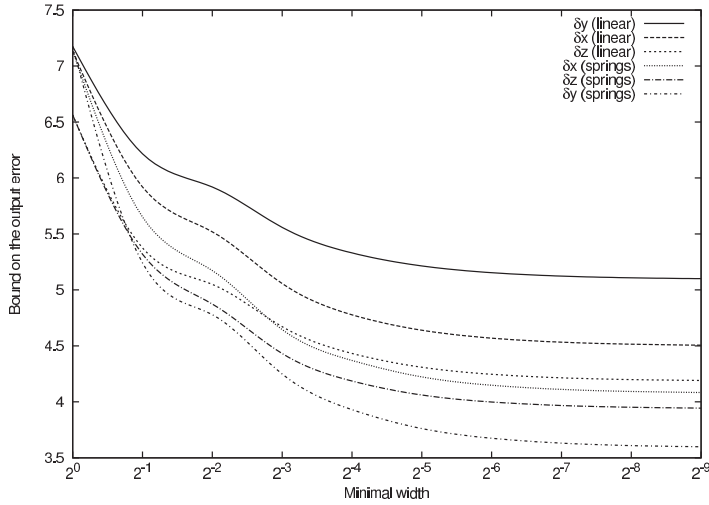


FIG. 9. Interpolated output error with both approaches. Each curve represents the output error obtained for a coordinate with respect to the minimal width  $w$ . In the case of springs, this width corresponds to the minimal diameter of the midpoint of  $[\alpha]$  (or  $[\beta]$ ). In the linear method, this width is simply the diameter of  $[\alpha]$  (or  $[\beta]$ ). Since values chosen for  $w$  decrease exponentially, the limit values can be read from the plot with a strong confidence (i.e., adding smaller values for  $w$  would be nearly useless).

where at the  $k^{th}$  time step,  $x_k$  is the system state and  $u_k$  an input. Assume that all inputs  $u_0, \dots, u_k$  share the same domain and the same uncertainty, i.e., belong to the same spring  $\langle U \rangle$ . The question is to compute  $\delta_{x_k}$ , i.e., to observe how varies the uncertainty on  $x_k$  as  $k$  grows.

First, derivating  $x_k$  as an expression of  $u_0, \dots, u_k$  is unreasonable for large values of  $k$  (such derivation would lead, in particular, to an expression with  $k$  occurrences of  $u_0$ ). The only way to linearize the problem seems to process iteratively as follows:

$$\begin{aligned} [x_0] &:= [u] \\ \delta_{x_0} &:= \delta_u \\ [x_{k+1}] &:= f([x_k], [u]) \\ \delta_{x_{k+1}} &:= \left| \frac{\partial f}{\partial x}([x_k], [u]) \right| \times \delta_{x_k} + \left| \frac{\partial f}{\partial u}([x_k], [u]) \right| \times \delta_u \end{aligned}$$

Note that bisection is not possible in this problem since computing  $\delta_{x_k}$  would generate  $2^k$  boxes. Alternatively, the spring approach can be implemented with a straightforward iteration:

$$\begin{aligned} \langle x_0 \rangle &:= \langle U \rangle \\ \langle x_{k+1} \rangle &:= f(\langle x_k \rangle \times \langle U \rangle), \end{aligned}$$

and the error bound  $\delta_{x_k}$  at the  $k^{th}$  step is nothing but  $(\text{mag rad}(\langle x_k \rangle))$ .

We now compare both approaches with  $f(x, u) = \cos(x \times u)$  and  $\langle u \rangle = \langle [-1, 1], 0.1 \rangle$ . Table 1 gives the results which speak for themselves.

Because of the recursion, the overestimation produced by the linearizations gets accumulated (with however an asymptotic effect due to a contractive partial derivative with respect to  $x$ ). The computed bound becomes quickly insignificant. This accumulation explains why a sharper method (such as springs) is even more crucial in a dynamical context.

TABLE 1

| $k$ | $\delta_{x_k}$ (with springs) | $\delta_{x_k}$ (with linearizations) |
|-----|-------------------------------|--------------------------------------|
| 1   | 0.154338                      | 0.176488                             |
| 2   | 0.213634                      | 0.262137                             |
| 4   | 0.303511                      | 0.428413                             |
| 8   | 0.400928                      | 0.741786                             |
| 16  | 0.451354                      | 1.29854                              |
| 32  | 0.458032                      | 2.17861                              |
| 64  | 0.458118                      | 3.28505                              |
| 128 | 0.458118                      | 4.18118                              |
| 256 | 0.458118                      | 4.5029                               |
| 512 | 0.458118                      | 4.53019                              |

**5. Conclusion.** For decades, interval analysis has turned out to be the right framework to deal with uncertainties. However, when the uncertainty is itself subject to uncertainty, i.e., when the midpoint and the radius of an interval may vary, the standard arithmetic has to be replaced by the so-called *spring arithmetic*. To our knowledge, springs had not been studied before and this paper contains a first study. Of course, the spring arithmetic has to be completed and many further algebraic properties should be investigated. However, putting the stress on the application side is probably more useful as a first development than an exhaustive abstract theory. We have shown that this new arithmetic allows one to perform a rigorous a priori error analysis as easily as computing an enclosure of the function range with interval arithmetic. We tried in this paper to lay the foundation stone of a new arithmetic and, with no doubt, a lot of work still has to be done.

## REFERENCES

- [1] G. ALEFELD AND J. HERZBERGER, *Introduction to Interval Computations*, Academic Press, New York, 1983.
- [2] D. G. CACUCI, *Sensitivity and Uncertainty Analysis, Theory*, Chapman & Hall, New York, 2003.
- [3] L. JAULIN, M. KIEFFER, O. DIDRIT, AND E. WALTER, *Applied Interval Analysis*, Springer, New York, 2001.
- [4] L. JAULIN AND E. WALTER, *Guaranteed bounded-error parameter estimation for nonlinear models with uncertain experimental factors*, *Automatica*, 35 (1999), pp. 849–856.
- [5] Z. KULPA, *Diagrammatic representation of interval space in proving theorems about interval relations*, *Reliable Comput.*, 3 (1997), pp. 209–217.
- [6] Z. KULPA, *Diagrammatic representation for interval arithmetic*, *Linear Algebra Appl.*, 324 (2001), pp. 55–80.
- [7] Z. KULPA, *A diagrammatic approach to investigate interval relations*, *J. Visual Languages Computing*, 17 (2006), pp. 466–502.
- [8] Z. KULPA AND S. MARKOV, *On the inclusion properties of interval multiplication: A diagrammatic study*, *BIT Numer. Math.*, 4 (2003), pp. 791–810.
- [9] S. MARKOV, *Computation of Algebraic Solutions of Interval Systems via Systems of Coordinates*, Kluwer/Plenum, New York, 2001, pp. 103–114.
- [10] R. MOORE, *Interval Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [11] A. NEUMAIER, *Interval Methods for Systems of Equations*, Cambridge University Press, Cambridge, 1990.
- [12] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
- [13] L. PETKOVIC AND M. PETKOVIC, *Inequalities in circular arithmetic: A survey*, *Math. Appl.*, 430 (1998), pp. 325–340.
- [14] D. RATZ, *An optimized interval slope arithmetic and its application*, technical report, Institut für Angewandte Mathematik, Universität Karlsruhe, 1996.
- [15] S. M. RUMP, *Fast and parallel interval arithmetic*, *BIT Numer. Math.*, 39 (1999), pp. 534–554.