#### 4ECM Prize lecture, June 29 2004

# Validated numerics

and the art of dividing by zero

Warwick Tucker Department of Mathematics Uppsala University warwick@math.uu.se

June 23, 2004

# Validated numerics

and the art of dividing by zero

Warwick Tucker Department of Mathematics Uppsala University warwick@math.uu.se



### What is validated numerics?

- set-valued mathematics;
- intervals replace real numbers.

### Why use validated numerics?

- provides rigorous error bounds;
- models uncertainty;
- may produce faster numerical methods.

Early work: R. C. Young (1931), M. Warmus (1956), T. Sunaga (1958), R. E. Moore (1959) *Interval Analysis* (1966).

### Intervals

We will adopt the short-hand notation

 $[a] = [\underline{a}, \overline{a}] = \{ x \in \mathbb{R} \colon \underline{a} \le x \le \overline{a} \},\$ 

and let  ${\rm I\!R}$  denote the set of all compact intervals of the real line:

 $\mathbb{R} = \{ [a] : \underline{a}, \overline{a} \in \mathbb{R}; \underline{a} \leq \overline{a} \}.$ 

We allow for *thin* intervals with  $\underline{a} = \overline{a}$ .

**Example:**  $[1,\pi] \in \mathbb{R}$ , but not [2,1] or  $[1,\infty]$ .

As sets, intervals inherit relations such as

 $= \subseteq \subset \neq \not \subset \dots$ Furthermore, we can define the operations $[a] \sqcup [b] = [\min\{\underline{a}, \underline{b}\}, \max\{\overline{a}, \overline{b}\}],$ 

 $[a] \cap [b] = \begin{cases} \emptyset & : \text{ if } \overline{a} < \underline{b} \text{ or } \overline{b} < \underline{a}, \\ [\max\{\underline{a}, \underline{b}\}, \min\{\overline{a}, \overline{b}\}] & : \text{ otherwise.} \end{cases}$ 

## **Useful functions**

Functions from  $\mathbb{R}$  to  $\mathbb{R}$ :

 $\operatorname{rad}([a]) = \frac{1}{2}(\bar{a} - \underline{a}); \qquad \operatorname{mid}([a]) = \frac{1}{2}(\bar{a} + \underline{a}),$  $\operatorname{mig}([a]) = \begin{cases} 0 & : \text{ if } 0 \in [a], \\ \min\{|\underline{a}|, |\overline{a}|\} & : \text{ otherwise;} \end{cases}$  $\max\{|\underline{a}|, |\overline{a}|\}.$ 

Functions from  $\mathbb{R}$  to  $\mathbb{R}$ :

 $abs([a]) = \{|a| : a \in [a]\} = [mig([a]), mag([a])].$ 

### $\mathbb{R}$ as a metric space:

We can turn  $\mathbb{R}$  into a metric space by equipping it with the Hausdorff distance:

 $d([a], [b]) = \max\{|\underline{a} - \underline{b}|, |\overline{a} - \overline{b}|\}.$ 

Using the metric, we can define the notion of a convergent sequence of intervals:

$$\lim_{k \to \infty} [a_k] = [a] \quad \Leftrightarrow \quad \lim_{k \to \infty} d([a_k], [a]) = 0.$$

### Arithmetic over $\mathbb{R}$ :

**Definition:** If  $\star$  is one of the operators  $+, -, \times, \div$ , and if  $[a], [b] \in \mathbb{R}$ , then

$$[a] \star [b] = \{a \star b \colon a \in [a], b \in [b]\},\$$

except that  $[a] \div [b]$  is undefined if  $0 \in [b]$ .

Uncountable many cases to consider!

Continuity, monotonicity, and compactness  $\Rightarrow$ 

$$[a] + [b] = [\underline{a} + \underline{b}, \overline{a} + \overline{b}]$$
  

$$[a] - [b] = [\underline{a} - \overline{b}, \overline{a} - \underline{b}]$$
  

$$[a] \times [b] = [\min\{\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}, \max\{\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}]$$
  

$$[a] \div [b] = [a] \times [1/\overline{b}, 1/\underline{b}], \quad \text{if } 0 \notin [b].$$

On a computer we use *directed rounding*:

 $[a] + [b] = [\bigtriangledown (\underline{a} \oplus \underline{b}), \triangle (\overline{a} \oplus \overline{b})].$ 

We then have  $[a] \star [b] \supseteq \{a \star b \colon a \in [a], b \in [b]\}.$ 

### **Properties of interval arithmetic**

(1) IA is associative and commutative.

(2) IA is *not* distributive: [-1,1]([-1,0] + [3,4]) = [-1,1][2,4] = [-4,4], [-1,1][-1,0] + [-1,1][3,4] = [-1,1] + [-4,4] = [-5,5].We do, however, always have

 $[a]([b] + [c]) \subseteq [a][b] + [a][c].$ 

(3) IA has no multiplicative or additive inverse:

 $0 \in [a] - [a];$   $1 \in [a] \div [a].$ 

(4) IA is *inclusion monotonic*, i.e., if  $[a] \subseteq [a']$ , and  $[b] \subseteq [b']$ , then

 $[a] \star [b] \subseteq [a'] \star [b'],$ 

where we demand that  $0 \notin [b']$  for division.

### **Interval extensions**

One of the the main goals is to enclose the range of a function f:

$$R(f; D) = \{f(x) \colon x \in D\}.$$

This is achieved by constructing an *interval extension*  $F \colon \mathbb{R} \to \mathbb{R}$  of the real-valued function  $f \colon \mathbb{R} \to \mathbb{R}$ .

Monotone functions are easy!

$e^{[x]}$	=	$[e^{\underline{x}},e^{\overline{x}}]$	
$\sqrt{[x]}$	=	$[\sqrt{\underline{x}},\sqrt{\overline{x}}]$	if $0 \leq \underline{x}$
$\log[x]$	=	$[\log \underline{x},\log \overline{x}]$	if $0 < \underline{x}$
$\arctan[x]$	=	[arctan $\underline{x},$ arctan $\overline{x}$ ] .	

Piecewise monotone functions are also OK!

 $[x]^{n} = \begin{cases} [\underline{x}^{n}, \overline{x}^{n}] & : \text{ if } n \in \mathbb{Z}^{+} \text{ is odd,} \\ [\text{mig}([x])^{n}, \text{mag}([x])^{n}] & : \text{ if } n \in \mathbb{Z}^{+} \text{ is even,} \\ [1,1] & : \text{ if } n = 0, \\ [1/\overline{x}, 1/\underline{x}]^{-n} & : \text{ if } n \in \mathbb{Z}^{-}; \ 0 \notin [x]. \end{cases}$ 

### **Standard/elementary functions**

We define the class of *standard* functions to be the set

 $\mathfrak{S} = \{e^x, \log x, x^a, \operatorname{abs} x, \sin x, \cos x, \tan x, \dots\}$ 

 $\ldots$ , arccos x, arctan x, sinh x, cosh x, tanh x}.

For any  $f \in \mathfrak{S}$ , we can construct a *sharp* interval extension F, i.e.,

$$f \in \mathfrak{S} \Rightarrow R(f; [x]) = F([x]).$$

Building new functions is easy...

We use finite combinations of constants, elements of  $\mathfrak{S}$ ,  $\{+, -, \times, \div\}$ , and  $\circ$  to build the *elementary* functions  $\mathfrak{E}$ . Interval versions of  $\mathfrak{S}$  and  $\{+, -, \times, \div\}$  provide the corresponding interval extensions.

... but we may now overestimate the range.

If  $f(x) = \frac{x}{1+x^2}$ , then  $F([x]) = \frac{[x]}{1+[x]^2}$ . For the interval [x] = [1, 2], we have

$$R(f; [1, 2]) = \left[\frac{2}{5}, \frac{1}{2}\right] \subseteq \left[\frac{1}{5}, 1\right] = F([1, 2]).$$

Looks are important!

$$f_1(x) = 1 - x^2 = (1 - x)(1 + x) = f_2(x),$$

but

$$F_1([x]) = 1 - [x]^2 \neq (1 - [x])(1 + [x]) = F_2([x]),$$
  
since

$$F_1([-1,1]) = 1 - [-1,1]^2 = [1,1] - [0,1] = [0,1],$$
  

$$F_2([-1,1]) = (1 - [-1,1])(1 + [-1,1])$$
  

$$= [0,2] \times [0,2] = [0,4].$$

Different representations – different functions.

### **Interval enclosures**

**Theorem 1:** If  $f \in \mathfrak{E}$ , and F([x]) is well-defined, then

$$R(f; [x]) \subseteq F([x]).$$

How tight is the enclosure?

**Theorem 2:** If  $f \in \mathfrak{E}$ ,  $[x] = [x_1] \cup \cdots \cup [x_k]$ , and F([x]) is well-defined, then

$$R(f; [x]) \subseteq \bigcup_{i=1}^{k} F([x_i]) \subseteq F([x]).$$

If f is Lipschitz on [x] there is a  $K \ge 0$  s.t.

$$\operatorname{rad}\left(\bigcup_{i=1}^{k} F([x_i])\right) - \operatorname{rad}\left(R(f; [x])\right) \leq K \max_{i} \operatorname{rad}\left([x_i]\right).$$

I.A. (almost) gives us access to R(f; [x]).

### **Computer-aided proofs**

An important consequence of Theorem 1 is

$$y \notin F([x]) \Rightarrow y \notin R(f; [x]).$$

**Exercise:** Let  $f(x) = (\sin x - x^2 + 1) \cos x$ . Prove that  $f(x) \neq 0$  for  $x \in [0, \frac{1}{2}]$ .

Solution: Define  $F([x]) = (\sin [x] - [x]^2 + 1) \cos [x]$ . Then, by Theorem 1, we have

$$R(f; [0, \frac{1}{2}]) \subseteq F([0, \frac{1}{2}]) = \dots$$
$$\dots = [\frac{3}{4} \cos \frac{1}{2}, 1 + \sin \frac{1}{2}] \subseteq [0.65818, 1.4795].$$



### **Graph enclosures**

Exercise: Draw the graph of the function  $f(x) = \cos^3 x + \sin x$ over the interval [ 5.5]

over the interval [-5, 5].

Solution: Define  $F([x]) = \cos^3 [x] + \sin [x]$ , and bisect the domain into smaller pieces until

 $\max_{i} \operatorname{rad} \left( F([x_i]) \right) \leq \operatorname{TOL}.$ 



### "Impossible" cases too

Exercise: Draw the graph of the function  $f_a(x) = x^2 - \frac{3}{10}e^{-(a(x-\frac{1}{2}))^2}$ for a = 200 over the interval [-1, 1].



Even when *a* is huge, the I.A.-methods *cannot* miss the sharp bend! Conventional methods *must* miss it.

### **Root enclosures**

Exercise: Find all roots of the function

$$f(x) = \sin x(x - \cos x)$$

over the domain [-10, 10].

Solution: Define  $F([x]) = \sin [x]([x] - \cos [x])$ , bisect the domain, and throw away all pieces that satisfy  $0 \notin F([x_i])$ .



# Implicit curves ( $\mathbb{R}^2$ )

Exercise: Draw the locus defined by  $f(x,y) = \sin(\cos x^2 + 10 \sin y^2) - y \cos x = 0;$ restricted to the domain  $[-5,5] \times [-5,5].$ 

MATLAB produces the following picture:



The locus |f(x, y)| = 0, however, is empty!?!

### Implicit curves...

The validated enclosure in both cases is



Number of solution boxes: 19718 Maximal box diameter : 0.0195312 #-----#

### Implicit curves...

A sharper look at the previous picture...



No zeroes can exist outside the boxes:

 $0 \notin F([x^{(i)}]) \Rightarrow 0 \notin R(f; [x^{(i)}]).$ 

### Optimization

Exercise: Find the minimum of the function

$$f_a(x) = x^2 - \frac{3}{10}e^{-(a(x-\frac{1}{2}))^2}$$

for a = 10000 over the interval  $\left[-\frac{3}{2}, \frac{3}{2}\right]$ .

Solution: Adaptive branch and bound.

#	#
Domain :	[-1.5,1.5]
Tolerance :	9.09495e-13 (2^-40)
Function calls :	3810
Global minimizer :	[4.9999998318185e-01,4.9999998348404e-01]
Global minimum :	-5.000000833367428e-02 +- 3.412409245e-13
Non-rigorous estimate:	3.113063833543e-09
#	#

### Prove that the minimum is negative:

#	#
Domain :	[-1.5,1.5]
Tolerance :	0.03125 (2^-5)
Function calls :	67
Global minimizer :	[4.9996948242187e-01,5.0001525878907e-01]
Global minimum :	-4.914849425348847e-02 +- 8.820223934e-04
Non-rigorous estimate:	3.71e-04
#	#

### Newton's method in ${\ensuremath{\mathbb R}}$

Assume that  $f \in C^1([x], \mathbb{R})$ , and that  $x^* \in [x]$ is a root of f. Also assume that  $0 \notin F'([x])$ , and define

$$N_f([x]) = \operatorname{mid}([x]) - \frac{f(\operatorname{mid}([x]))}{F'([x])}$$

**Theorem 3:** Assume that  $N_f([x])$  exists. Then

 $N_f([x]) \cap [x] = \emptyset \Rightarrow f$  has no roots in [x];

 $N_f([x]) \subseteq [x] \Rightarrow f$  has a unique root in [x].



### Newton's method in ${\ensuremath{\mathbb R}}$

Set 
$$[x^{(0)}] = [x]$$
, and consider the sequence  
 $[x^{(k+1)}] = N_f([x^{(k)}]) \cap [x^{(k)}], \quad k \in \mathbb{N}.$ 

Bonus: The iterates behave very regularly.

**Theorem 4:** If  $N_f([x^{(0)}])$  exists, and  $[x^{(0)}]$  contains a root  $x^*$  of f, then so do all  $[x^{(k)}]$ ,  $k \in \mathbb{N}$ . Furthermore, the intervals  $[x^{(k)}]$  form a nested sequence converging to  $x^*$ .

Major drawback: The Newton operator  $N_f$  is undefined when  $0 \in F'([x])$ .

Fix 1: Use bisection to single out subintervals  $[\tilde{x}]$  on which  $0 \notin F'([\tilde{x}])$ . These are sent to the interval Newton method. Also keep all small  $[\tilde{x}]$  with  $0 \in F'([\tilde{x}])$  and  $0 \in F([\tilde{x}])$ .

#### Newton's method in $\mathbb{R}$

Example: Consider the function

$$f_a(x) = x^2 - \frac{3}{10}e^{-(a(x-\frac{1}{2}))^2}$$

with a = 10000 over the interval  $\left[\frac{1}{4}, 1\right]$ . Recall that this function has a *very* sharp bend.

```
#-----#
Searching the domain [0.25,1] with TOL = 0.00390625 (2^-8).
After the adaptive bisection, we have:
    1 subdomain where f may have roots.
There may be roots within
    1: [0.496094,0.501953]
#------##
```

#### Decrease the tolerance...

### **Extended interval arithmetic**

Fix 2: Extend the interval arithmetic to allow for division by zero.

Example:  $[1,2] \div [-3,5] = ???$  $[1,2] \div [-3,5] = [1,2] \div ([-3,0] \cup [0,5])$  $= ([1,2] \div [-3,0]) \cup ([1,2] \div [0,5])$ 

Now we *define* the division as follows:

 $[1,2] \div [-3,0] \stackrel{\text{def}}{=} \lim_{\varepsilon \to 0^-} [1,2] \div [-3,\varepsilon] = [-\infty,-\frac{1}{3}]$  and

$$[1,2] \div [0,5] \stackrel{\text{def}}{=} \lim_{\varepsilon \to 0^+} [1,2] \div [\varepsilon,5] = [\frac{1}{5},+\infty].$$

Answer:  $[1,2] \div [-3,5] = [-\infty, -\frac{1}{3}] \cup [\frac{1}{5}, +\infty].$ 

### The extended interval Newton Method

When computing the Newton iterates, the unbounded intervals are intersected back to compact domains by the scheme



**Example:** Again the function  $f_{10000}(x)$ .

#		
π		<del>п</del>
Domain	:	[2.500000e-01,1.000000e+00]
Tolerance	:	9.536743e-07 (2^-20)
Unique root within	:	[+0.49995676466028721,+0.49995838064162701]
Unique root within	:	[+0.50004262913481123,+0.50004274553256134]
Function calls	:	34
#		#

#### The Lorenz equations

Introduced in 1963 by Edward Lorenz.

$$\dot{x}_{1} = -\sigma x_{1} + \sigma x_{2}$$
$$\dot{x}_{2} = \rho x_{1} - x_{2} - x_{1} x_{3}$$
$$\dot{x}_{3} = -\beta x_{3} + x_{1} x_{2},$$

Classical parameters:  $\sigma = 10$ ,  $\beta = 8/3$ ,  $\varrho = 28$ .

Symmetry:  $S(x_1, x_2, x_3) = (-x_1, -x_2, x_3).$ 

Three fixed points: the origin 0 and

$$C^{\pm} = (\pm \sqrt{\beta(\varrho - 1)}, \pm \sqrt{\beta(\varrho - 1)}, \varrho - 1).$$

Stability: The origin is a saddle point with

$$0<-\lambda_3<\lambda_1<-\lambda_2.$$

 $C^{\pm}$  are unstable spirals.

## Solutions of the Lorenz equations



### The Lorenz equations...

Thus, the stable manifold of the origin  $W^{s}(0)$  is two-dimensional, and the unstable manifold of the origin  $W^{u}(0)$  is one-dimensional.

Constant divergence:

$$\frac{\partial \dot{x}_1}{\partial x_1} + \frac{\partial \dot{x}_2}{\partial x_2} + \frac{\partial \dot{x}_3}{\partial x_3} = -(\sigma + \beta + 1),$$

The volume of a solid at time t can be expressed as

$$V(t) = V(0)e^{-(\sigma+\beta+1)t} \approx V(0)e^{-13.7t},$$

for the classical parameter values.

Absorbing region:  $\ensuremath{\mathcal{U}}$  containing the origin.

Maximal invariant set:

$$\mathcal{A} = \bigcap_{t \ge 0} \varphi(\mathcal{U}, t).$$

 $\mathcal{A}$  must have zero volume, and  $W^u(0) \subseteq \mathcal{A}$ .

### Lorenz observed:

- [1] An attracting invariant set  ${\cal A}$
- [2] Sensitive dependence on i.c.
- [3] Fractal structure of  ${\cal A}$
- [4] Robustness of  ${\cal A}$



He observed a strange attractor!

# A geometric model:

Introduced by Guckenheimer and Williams (1979)



Return map:  $R: \Sigma \setminus \Gamma \to \Sigma$ .

The return plane  $\Sigma$  is foliated by stable leaves.

Projecting along these stable leaves induces a 1-d singular function:

$$f\colon [-1,1]\to [-1,1]$$



The function  $f: [-1, 1] \rightarrow [-1, 1]$  satisfies:

[1] 
$$f(-x) = -f(x);$$
  
[2]  $\lim_{x\to 0} f'(x) = +\infty;$   
[3]  $f''(x) < 0$  on  $(0, 1];$   
[4]  $f'(x) > \sqrt{2};$ 

[1] - [4]  $\Rightarrow$  f is topologically transitive on [-1, 1].

What is a strange attractor?

We need to prove:

- (1) There exists a compact  $N \subset \Sigma$ , such that  $R(N \setminus \Gamma) \subset N$ .
- (2) On N, there exists a cone field  $\mathfrak{C}$  such that for all  $x \in N$ ,

 $DR(x) \cdot \mathfrak{C}(x) \subset \mathfrak{C}(R(x)).$ 

(3) There exists C > 0 and  $\lambda > 1$  such that for all  $v \in \mathfrak{C}(x)$ ,  $x \in N$ , we have

 $|DR^n(x)v| \ge C\lambda^n |v|, \qquad n \ge 0.$ 

Open conditions - Perfect for I.A.!

Computing R and DR requires solving intervalvalued differential equations. This is still an open field of research. How do we use these results?

(1) proves the existence of an attracting set. This *could* be a single periodic orbit.

(2)+(3) rule out the possibility of just observing a stable periodic orbit.

Strong enough expansion  $\Rightarrow$  topological transitivity.

*R* area contracting + expansion in  $\mathfrak{C}(x) \Rightarrow$  stable foliation.

**Theorem:** For the classical parameter values, the Lorenz equations support a robust strange attractor  $\mathcal{A}$  – the Lorenz attractor!

### The wrapping effect

**Example:** Solve the ODE  $(\dot{x}_1, \dot{x}_2) = (x_2, -x_1)$ .



Exponential growth of the solution set!

Fix: The Lorenz equations are strongly volume contracting. An adaptive bisection scheme dampens the overestimation of R and DR.

Analytic (non-numerical) techniques are needed near the fixed point of the flow.

### Other articles using I.A.

D. Gabai, G. R. Mayerhoff, and N. Thurston, *Homotopy hyperbolic 3-manifolds are hyperbolic*. Annals of Mathematics, **157**, 335–431, 2003.

J. Hass, M. Hutchings, and R. Schlafly, *The Double Bubble Conjecture*. Electr. Research Announcements of the Amer. Math. Soc., **1**, 98–102, 1995.

T. C. Hales, *Some algorithms arising in the proof of the Kepler conjecture*. Discrete and computational geometry, **25**, 489–507, Algorithms Combin., 2003.

K. Makino and M. Berz, *Taylor Models and Other Validated Functional Inclusion Methods*. Int. J. of Pure and Appl. Math., **4**, 379–456, 2003.

I. Mitrea, W. Tucker, *Some Counterexamples for the Spectral Radius Conjecture*. Differential and Integral Equations, **16:12**, 1409–1439, 2003.

N. S. Nedialkov, K. R. Jackson and G. F. Corliss, Validated Solutions of Initial Value Problems for Ordinary Differential Equations. J. Applied Math. and Comp., **105**, 21–68, 1999.

P. Zgliczynski, *Attracting fixed points for the Kuramoto-Sivashinsky equation*, SIAM J. Applied Dynamical Systems, **1:2**, 215–235, 2002.

#### **References:**

G. Alefeld and J. Herzberger, *Introduction to Interval Computations*. Academic Press, New York, 1983.

W. Kahan, *IEEE Standard 754 for Binary Floating-Point Arithmetic*. lecture notes, 1996. Available from http://www.cs.berkeley.edu/~wkahan/

E. N. Lorenz, *Deterministic Non-periodic Flow*. J. Atmos. Sci. **20** 130–141, 1963.

R. E. Moore, *Interval Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

T. Sunaga, *Theory of interval algebra and its application to numerical analysis*. In: RAAG Memoirs, Ggujutsu Bunken Fukuy-kai. Tokyo, **2** 29–46, 1958.

W. Tucker, A Rigorous ODE Solver and Smale's 14th Problem, Found. Comput. Math. **2:1**, 53–117, 2002.

M. Warmus, *Calculus of Approximations*. Bulletin de l'Academie Polonaise de Sciences, **4:5** 253–257, 1956.

R. C. Young, *The algebra of multi-valued quantities*. Mathematische Annalen, **104** 260–290, 1931.

#### Interval Computations Web Page

http://www.cs.utep.edu/interval-comp