15 Jan. 2018, 14h15

ÉCOLE NORMALE SUPÉRIEURE DE LYON

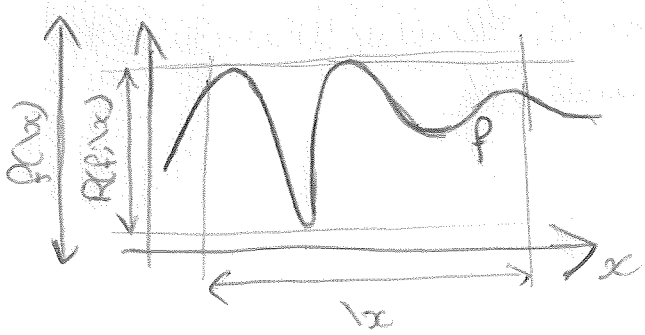
VALIDATED NUMERICS #1 by W. T.

(Mathematics) ∩ (Scientific Computing)

Goals: Computational proofs aimed at continuous problems.

- We will use "comp. repr." sets (intervals) to explore the continuum of $\mathbb{R}$.
- Operations on intervals are described by interval analysis.
- Gives us the ability to enclose the range of a function: $R(f;\mathbb{x}) = \{f(x); x \in \mathbb{x}\}$

$R(f;\mathbb{x}) \subseteq f(\mathbb{x})$ ← enclosure, computable.



Applications: evaluation, equation solving, global optimization, quadrature, ode/bvp/pde(s)

Strategy: Given a "purely" mathematical problem, break it up into a mathematical part and a finite, numerical part.

Example: compute $S = \sum_{k=1}^{\infty} 1/k^2$ to 12 correctly rounded digits. Then we write:

$$S = \underbrace{\sum_{k=1}^{N} 1/k^2}_{S_n} + \underbrace{\sum_{k=n+1}^{\infty} 1/k^2}_{S_n^*} \quad \text{where } S_n^* \in \left[\frac{1}{N+1}, \frac{1}{N}\right]$$

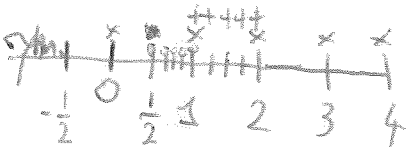
Any real number x can be expressed in base β as $x = \sigma \sum_{k=-\infty}^{n} b_k \beta^k = \sigma(b_n b_{n-1} \dots b_0 b_{-1} \dots)_\beta$. This expression is not unique. First, we use scientific notation $x = \sigma(\underbrace{b_0 b_1 b_2 \dots}_{\text{mantissa}})_\beta \beta^e$. There more props:

① $\forall i$, we have $0 \le b_i \le \beta - 1$; ② For $x \neq 0$, $1 \le b_0 \le \beta - 1$.

③ For ∞ many i, $0 \le b_i \le \beta - 2$; These props. gives a unique repr. of x.

More conditions ④ $\check{e} \le e \le \hat{e}$; ⑤ $m = (b_0 b_1 \dots b_{p-1})$ where p is precision.

Properties ①-⑤ gives $IF = IF(\beta, p, \check{e}, \hat{e})$, a finite set of "floating point numbers"

$IF(2,3,-1,+1)$: 

There is a lot of space around zero.

③' for $e > \check{e}$ $1 \le b_0 < \beta - 1$. When $e = \check{e}$, b_0 is allowed to be zero (subnormal number)

With subnormal numbers, $x = y \Leftrightarrow x \ominus y = 0$ $\quad x, y \in IF$

Problem: IF is not closed under arithmetic operations: for $x, y \in IF$, $x * y \overset{?}{\in} IF$, $* \in \{+, -, \times, /\}$

Rounding: operation $\circ : \mathbb{R} \to IF$ should satisfy: (R1) $x \in IF \Rightarrow \circ(x) = x$; (R2) $x, y \in \mathbb{R}, x \le y \Rightarrow \circ(x) \le \circ(y)$

Examples: "Truncate" (Round towards 0), "Round up" (Round towards $+\infty$), "Round down" (Round towards $-\infty$), Round towards nearest even.

Theorem If x and y are normal fp numbers with $x * y \neq 0$, then

$$\left| \frac{x * y - x \circledast y}{x * y} \right| < \varepsilon_M$$

where ε_M is machine epsilon, with $\varepsilon_M = \beta^{-(p-1)}$.

There is no information when two operations are done successively. Because the operands of the 2nd operation are not fp numbers *a priori*.