Managing computational errors: approximation errors and roundoff errors

Sources of errors

modelling errors

problem → math. model → approximation → program → computed solution

data

data errors

approximation errors

roundoff errors

roundoff errors

Typical problems

find a polynomial P
that approximates f
on $[a,b]$
goal: evaluate the elementary function f

approximation errors

estimate or bound
$\|f(x) - P(x)\|_{[a,b],\infty}$

roundoff errors

estimate or bound $\|f(x) - \widehat{\hat{P}(\hat{x})}\|_{[a,b],\infty}$
or $\|P(x) - \widehat{\hat{P}(\hat{x})}\|_{[a,b],\infty}$
(since $\|f(x) - \widehat{\hat{P}(\hat{x})}\|_{[a,b],\infty} \leq \|f(x) - P(x)\|_{[a,b],\infty} + \|P(x) - \widehat{\hat{P}(\hat{x})}\|_{[a,b],\infty}$)
where $\hat{x}$: machine-representable number,
$\hat{P}$: polynomial with machine-representable coeff,
and $\widehat{\hat{P}(\hat{x})}$: machine evaluation

Estimates, bounds, applications

Interval arithmetic and arbitrary precision (MPFI)

returns guaranteed
enclosures of the results

to refine the enclosures:
contracting iterations
- Newton
- Taylor expansions
- retro−propagation
- interval bisection

arbitrary precision
- to contract intervals
  beyond the usual precision
- automatic adaptation
  of the computing precision

GAPPA

automatic proof generation
of arithmetic properties

in particular,
bounds errors due to
floating−point arithmetic

tools:
- properties on the fp operations
- interval arithmetic to get bounds
  - on the values
  - on the rounding errors
- interval bisection
- expression rewriting

returns a formal proof (Coq)

MEPLib

machine−efficient polynomials lib.

input: function *f*, domain *[a,b]*

output: "best" polyns s.t.
- approximation error is small
- coeff exactly representable
  in the target format
- other constraints on the coeff

criteria for "best" polyns:
- approximation error
- evaluation error
- speed of evaluation
- memory size for the coeff
- energy consumption

approximation error
- given by Maple
- given by exhaustive tests

CRlibm

elementary functions
- argument reduction before
- and reconstruction after
- the polynomial evaluation
- sometimes errorless
- and sometimes not

in double precision
- use of
- – IEEE−754 rounding
- – exact IEEE−754 operations
- – perf−oriented eval. scheme
- – Gappa−assisted error bound

with correct rounding
- need up to 158 bits of precision
- double−double & triple−double
- rounding error of each op.

HOTBM

Higher−Order Table−Based Methods

elementary functions
in hardware
with fixed−point formats

approximation error
- polynomial P given by a minimax

method error
- evaluation scheme
  (table−multiply−and−add)
- choice of the bitwidth
  of the datapath
  (choice of the coeff of $\hat{P}$)

roundoff error
- fill the tables

exhaustive tests
- to check the total error