Numerical verification
for systems of linear and nonlinear equations

Milan Hladík

Faculty of Mathematics and Physics,
Charles University in Prague, Czech Republic
http://kam.mff.cuni.cz/~hladik/

Programy a algoritmy numerické matematiky – PANM 19
Hejnice, 24. – 29. června 2018

Motivation: Numerical errors

Example (Rump, 1988)

Consider the expression

$$f = 333.75b^6 + a^2(11a^2b^2 - b^6 - 121b^4 - 2) + 5.5b^8 + \frac{a}{2b},$$

with

$$a = 77617, \quad b = 33096.$$

Calculations from 80s gave

$$\begin{array}{rl} \text{single precision} & f \approx 1.172603\ldots \\ \text{double precision} & f \approx 1.1726039400531\ldots \\ \text{extended precision} & f \approx 1.172603940053178\ldots \\ \text{the true value} & f = -0.827386\ldots \end{array}$$

Motivation: Computer-assisted proofs

Kepler conjecture

What is the densest packing of balls? (Kepler, 1611)

That one how the oranges are stacked in a shop.

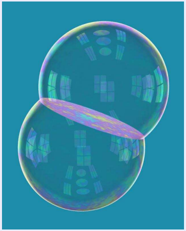
The conjecture was proved by T.C. Hales (2005).

Double bubble problem

What is the minimal surface of two given volumes?

Two pieces of spheres meeting at an angle of 120°.

Hass and Schlafly (2000) proved the equally sized case.
Hutchings et al. (2002) proved the general case.



Verification: Introduction

Can we obtain rigorous numerical results by using floating-point arithmetic?

Yes, by extending to interval arithmetic.

Example

$$\frac{10}{3} \in [3.33333333333333333, \ 3.33333333333333334],$$
$$\sqrt{2} \in [1.41421356237309504\mathrm{88}, \ 1.41421356237309504\mathrm{89}].$$

Interval computations

Notation

An interval matrix

$$\boldsymbol{A} := [\underline{A}, \overline{A}] = \{A \in \mathbb{R}^{m\times n} \mid \underline{A} \le A \le \overline{A}\}.$$

The center and radius matrices

$$A^c := \frac{1}{2}(\overline{A} + \underline{A}), \quad A^\Delta := \frac{1}{2}(\overline{A} - \underline{A}).$$

The set of all $m \times n$ interval matrices: $\mathbb{IR}^{m\times n}$.

Main problem

Let $f\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ and $\boldsymbol{x} \in \mathbb{IR}^n$. Determine the image

$$f(\boldsymbol{x}) = \{f(x) : x \in \boldsymbol{x}\}.$$

Monotone functions

If $f\colon \boldsymbol{x} \to \mathbb{R}$ is non-decreasing, then $f(\boldsymbol{x}) = [f(\underline{x}), f(\overline{x})]$.

(Similarly for piece-wise monotone functions.)

Interval arithmetic

Interval arithmetic (incl. rounding, IEEE standard)

$$
\begin{aligned}
\boldsymbol{a} + \boldsymbol{b} &= [\underline{a} + \underline{b}, \overline{a} + \overline{b}], \\
\boldsymbol{a} - \boldsymbol{b} &= [\underline{a} - \overline{b}, \overline{a} - \underline{b}], \\
\boldsymbol{a} \cdot \boldsymbol{b} &= [\min(\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}), \max(\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b})], \\
\boldsymbol{a} / \boldsymbol{b} &= [\min(\underline{a}/\underline{b}, \underline{a}/\overline{b}, \overline{a}/\underline{b}, \overline{a}/\overline{b}), \max(\underline{a}/\underline{b}, \underline{a}/\overline{b}, \overline{a}/\underline{b}, \overline{a}/\overline{b})], \quad 0 \notin \boldsymbol{b}.
\end{aligned}
$$

Theorem (Basic properties of interval arithmetic)

- *Interval addition and multiplication is commutative and associative.*
- *It is not distributive in general, but sub-distributive instead,*

$$\forall \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{IR} : \boldsymbol{a}(\boldsymbol{b} + \boldsymbol{c}) \subseteq \boldsymbol{a}\boldsymbol{b} + \boldsymbol{a}\boldsymbol{c}.$$

Example ($\boldsymbol{a} = [1, 2]$, $\boldsymbol{b} = 1$, $\boldsymbol{c} = -1$)

$$
\begin{aligned}
\boldsymbol{a}(\boldsymbol{b} + \boldsymbol{c}) &= [1, 2] \cdot (1 - 1) = [1, 2] \cdot 0 = 0, \\
\boldsymbol{a}\boldsymbol{b} + \boldsymbol{a}\boldsymbol{c} &= [1, 2] \cdot 1 + [1, 2] \cdot (-1) = [1, 2] - [1, 2] = [-1, 1].
\end{aligned}
$$

Direct usage of interval arithmetic: No, please

Why not to replace all operations by the interval operations from the very beginning?

Example (Amplification factor for the interval Gaussian elimination)

n	20	50	100	170
amplification	10^2	10^5	10^{10}	10^{16}

Advice

Postpone interval computation to the very end.

Verification

Verification

Compute a solution by floating-point arithmetic, and then to verify that the result is correct or determine rigorous distance to a true solution.

Typically, we can prove uniqueness ($=$ the problem is well posed).
Therefore, we can verify only robust properties!
Verifying singularity of a matrix thus cannot be performed!

Verification paradigm

- *every* computation on a computer should be done in a verified way
- we want not much extra computational cost

Verification for nonlinear equations

Verification method for one root of a function $f\colon \mathbb{R}^n \to \mathbb{R}^n$.

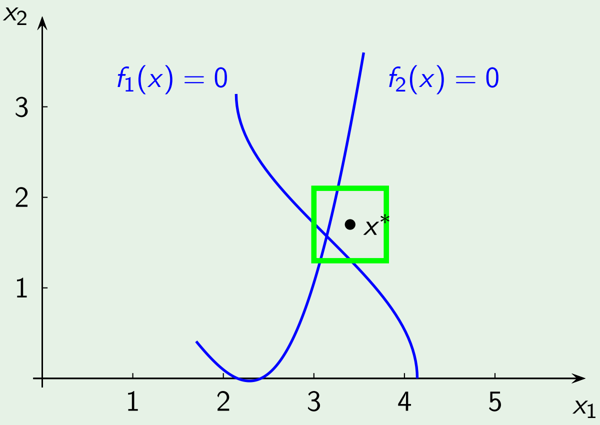
Problem statement

- Given $x^* \in \mathbb{R}^n$ a numerically computed ($=$ approximate) solution of the system $f(x) = 0$,
- find a small interval $0 \in \boldsymbol{y} \in \mathbb{IR}^n$ such that the true solution lies in $x^* + \boldsymbol{y}$.

Illustration of verification

Example

Illustration of the verification of x^* to be a solution of $f(x) = 0$.



Ingredients

Brouwer fixed-point theorem

Let U be a convex compact set in $\mathbb{R}^n$ and $g\colon U \to U$ a continuous function. Then there is a fixed point, i.e., $\exists x \in U : g(x) = x$.

Observation

Finding a root of $f(x)$ is equivalent to finding a fixed-point of the function $g(y) \equiv y - C \cdot f(x^* + y)$, where C is any nonsingular matrix of order n.

Perron theory of nonnegative matrices

- If $|A| \leq B$, then $\rho(A) \leq \rho(B)$.
  ($\leq$ is meant entrywise and $\rho(\cdot)$ is the spectral radius)
- If $A \geq 0$, $x > 0$ and $Ax < \alpha x$, then $\rho(A) < \alpha$.

Lemma

If $\boldsymbol{z} + \boldsymbol{R}\boldsymbol{y} \subseteq \mathrm{int}\, \boldsymbol{y}$, then $\rho(R) < 1$ for every $R \in \boldsymbol{R}$.

Proof. $|R|y^\Delta < y^\Delta$, whence by Perron theory $\rho(R) < 1$. $\square$

Cooking

Theorem

Suppose $0 \in \boldsymbol{y}$. *Now if*

$$-C \cdot f(x^*) + (I - C \cdot \nabla f(x^* + \boldsymbol{y})) \cdot \boldsymbol{y} \subseteq int\, \boldsymbol{y},$$

then:

- C *and every matrix in* $\nabla f(x^* + \boldsymbol{y})$ *are nonsingular, and*
- *there is a unique root of* $f(x)$ *in* $x^* + \boldsymbol{y}$.

Proof.

By the mean value theorem,

$$f(x^* + y) \in f(x^*) + \nabla f(x^* + \boldsymbol{y})y.$$

By the assumptions, the function

$$g(y) = y - C \cdot f(x^* + y) \in -C \cdot f(x^*) + (I - C \cdot \nabla f(x^* + \boldsymbol{y}))\boldsymbol{y} \subseteq int\, \boldsymbol{y}$$

has a fixed point, which shows "existence".

By Lemma, C and $\nabla f(x^* + \boldsymbol{y})$ are nonsingular, implying "uniqueness". □

Eating

Implementation

- take $C \approx \nabla f(x^*)^{-1}$ (numerically computed inverse),
- take $\boldsymbol{y} := C \cdot f(x^*)$ and repeat inflation

$$\boldsymbol{y} := \Big(- C \cdot f(x^*) + (I - C \cdot \nabla f(x^* + \boldsymbol{y})) \cdot \boldsymbol{y} \Big) \cdot [0.9, 1.1] + 10^{-20}[-1, 1]$$

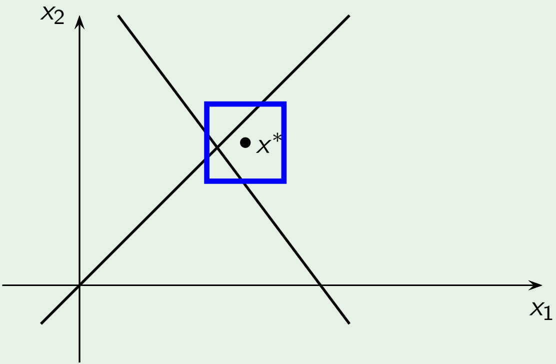
until the assumption of Theorem are satisfied.

Verification of a linear system of equations

Problem formulation

Given a real system $Ax = b$ and x^* approximate solution, find $\boldsymbol{y} \in \mathbb{IR}^n$ such that $A^{-1}b \in x^* + \boldsymbol{y}$.

Example



Verification of a linear system of equations

Given the system $Ax = b$ and an approximate solution x^*.

Theorem

Suppose $0 \in \boldsymbol{y}$. *Now if*

$$C(b - Ax^*) + (I - CA)\boldsymbol{y} \subseteq int\, \boldsymbol{y},$$

then:

- *C and A are nonsingular,*
- *there is a unique solution of $Ax = b$ in $x^* + \boldsymbol{y}$.*

Proof.

Use the previous result with $f(x) = Ax - b$. □

Implementation

- take $C \approx A^{-1}$ (numerically computed inverse),

Verification of a linear system of equations

ε-inflation method (Caprani and Madsen, 1978, Rump, 1980)

Repeat inflating $\boldsymbol{y} := [0.9, 1.1]\boldsymbol{x} + 10^{-20}[-1, 1]$ and updating

$$\boldsymbol{x} := C(b - Ax^*) + (I - CA)\boldsymbol{y}$$

until $\boldsymbol{x} \subseteq int\, \boldsymbol{y}$.

Then, $\Sigma \subseteq x^* + \boldsymbol{x}$.

Results

- Verification is theoretically 9–12 times slower than solving the original problem, practically only about 7 times slower (for random instances of dimension 100 to 2000).

Verification of a linear system of equations

Example

Let A be the Hilbert matrix of size 10 (i.e., $a_{ij} = \frac{1}{i+j-1}$), and $b := Ae$.

Then $Ax = b$ has the solution $x = e = (1, \dots, 1)^T$.

Approximate solution by Matlab:

0.999999999235452
1.000000065575364
0.999998607887449
1.000012638750021
0.999939734980300
1.000165704992114
0.999727989024899
1.000263042205847
0.999861803020249
1.000030414871015

Enclosing interval by ε-inflation method (2 iterations):

[0.99999973843401, 1.00000026238575]
[0.99999843048508, 1.00000149895660]
[0.99997745481481, 1.00002404324710]
[0.99978166603900, 1.00020478046370]
[0.99902374408278, 1.00104070076742]
[0.99714060702796, 1.00268292103727]
[0.99559932282378, 1.00468935360003]
[0.99546972629357, 1.00425202249136]
[0.99776781605377, 1.00237789028988]
[0.99947719419921, 1.00049082925529]

Overestimation factor about 20; compare $\kappa(A) \approx 1.6 \cdot 10^{13}$.

Verification of a linear system of equations

Challenge

- verification for large systems
  (one cannot use preconditioning by the inverse matrix)

Verification of other problems

- linear algebraic problems (eigenvalues, rank, decompositions,...)
- optimization (linear, semidefinite programming,...)
- infinite-dimensional problems (ODE,...)

References

S.M. Rump.
Verification methods: Rigorous results using floating-point arithmetic.
Acta Numerica, 19:187–449, 2010.

Software

Matlab/Octave libraries

- *Interval* for Octave (by O. Heimlich),
  interval arithmetic and elementary functions
  `https://wiki.octave.org/Interval_package`
- *Intlab* (by S.M. Rump),
  interval arithmetic and elementary functions
  `http://www.ti3.tu-harburg.de/~rump/intlab/`
  - *Versoft* (by J. Rohn), verification software
  - *Lime* (by M. Hladík, J. Horáček et al.), under development

Other languages libraries

- *Int4Sci Toolbox* (by Coprin team, INRIA),
  A Scilab Interface for Interval Analysis
  `http://www-sop.inria.fr/coprin/logiciels/Int4Sci/`
- *C++ libraries*: C-XSC, PROFIL/BIAS, BOOST interval, FILIB++,...
- *many others*: for Fortran, Pascal, Julia, Maple, Python,...

When no verification is used...

The Patriot Missile failure, Gulf War, Feb. 25, 1991

- Small rounding error of binary representation of $\frac{1}{10}$ expanded to $0.34\,s$ during 100 hours.
- As a consequence, the battery failed to intercept an incoming Iraqi Scud missile, which killed 28 soldiers.

The sinking of the Sleipner A offshore platform Norway, Aug. 13, 1991

- Inaccurate finite element approximation of the linear elastic model – the shear stresses were underestimated by 47%.
- The structure sprang a leak and needed to be sunk under a controlled operation.