Verified error bounds for multiple roots of systems of nonlinear equations

Stef Graillat

LIP6/PEQUAN - Université Pierre et Marie Curie (Paris 6) - CNRS

Joint work with Siegfried M. Rump

CRC 2011
International Workshop on Certified and Reliable Computation
NanNing, GuangXi, China, July 17-20, 2011

General motivations: self-validating methods

Verify assumptions of mathematical theorems on the computer

- Making mathematical proofs with computers
- Getting verified results :
  → an interval enclosure of the true result
  → an approximate result with a rigorous error bound
- Possibly with proof of uniqueness
- Being fast and accurate
- Dealing with “ill-posed problems”

General motivations (cont'd)

Proofs with computers: how to do that ?

- with computer algebra systems: exact results but sometimes not efficient
- with floating-numbers: fast but often wrong results due to rounding errors

Possible solution: computing with floating-point but taking into account all the rounding errors !

Outline of the talk

1. Floating-point and interval arithmetic
2. Principle of self-validating methods
3. Multiple roots of polynomial systems
4. Numerical experiments

Outline of the talk

Floating-point numbers

Normalized floating-point numbers $\mathbb{F} \subseteq \mathbb{R}$:

$$x = \pm \underbrace{x_0.x_1 \ldots x_{M-1}}_{mantissa} \times b^e, \quad 0 \leq x_i \leq b-1, \quad x_0 \neq 0$$

b : basis, M : precision, e : exponent such that $e_{\min} \leq e \leq e_{\max}$
epsilon machine $\epsilon = b^{1-M}$

Approximation of $\mathbb{R}$ by $\mathbb{F}$ with rounding $\mathrm{fl} : \mathbb{R} \to \mathbb{F}$.
Let $x \in \mathbb{R}$ then

$$\mathrm{fl}(x) = x(1+\delta), \quad |\delta| \leq \mathbf{u}$$

Unit rounding $\mathbf{u} = \epsilon/2$ for rounding to the nearest

Standard model of floating-point arithmetic

Let $x, y \in \mathbb{F}$ and $\circ \in \{+, -, \cdot, /\}$.

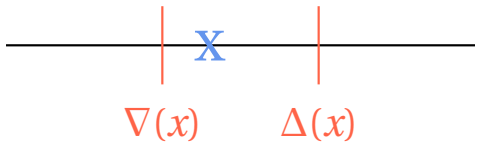
The result $x \circ y$ is not in general a floating-point number

$$\mathrm{fl}(x \circ y) = (x \circ y)(1 + \delta), \quad |\delta| \leq \mathbf{u}$$

IEEE 754 standard (1985 and 2008)
Correctly rounded : arithmetic ops $(+, -, \times, /, \sqrt{\ })$ performed as if first calculated to infinite precision, then rounded.

Type	Size	Mantissa	Exponent	Unit rounding	Interval
Single	32 bits	23+1 bits	8 bits	$\mathbf{u} = 2^{1-24} \approx 1,92 \times 10^{-7}$	$\approx 10^{\pm 38}$
Double	64 bits	52+1 bits	11 bits	$\mathbf{u} = 2^{1-53} \approx 2,22 \times 10^{-16}$	$\approx 10^{\pm 308}$

Rounding



The norm proposes 4 rounding modes:

- rounding toward $+\infty$ denoted $\Delta(x)$: return the smallest floating-point number greater or equal the exact result x
- rounding toward $-\infty$ denoted $\nabla(x)$: return the largest floating-point number less or equal the exact result x
- rounding toward 0, denoted $\mathcal{Z}(x)$: return $\Delta(x)$ for negative numbers and $\nabla(x)$ for positive numbers
- rounding to the nearest, denoted $\circ(x)$: return the nearest floating-point number of the exact result x (breaks ties by rounding to the nearest even floating-point number)

The 3 first rounding modes are called directed rounding modes.

Advantages of the standard

IEEE arithmetic is closed: every operation produces a result.
Default results:

Exception type	Default result
Invalid operation	NaN (Not a Number)
Overflow	$\pm\infty$
Divide by zero	$\pm\infty$
Underflow	subnormal numbers
Inexact	correctly rounded result

NaN is generated by operations such as $0/0$, $0\times\infty$, ∞/∞, $(+\infty)+(-\infty)$ and $\sqrt{-1}$.
Infinity symbol satisfies $\infty+\infty=\infty$, $(-1)\times\infty=-\infty$ and $(\text{finite})/\infty=0$.

Advantages of the standard

- make possible to write portable programs
- make program deterministic from one computer to another
- correctly rounded operations
- directed roundings useful for interval arithmetic

Directed roundings

Let

$$x_1 = \nabla(1/3), \quad x_2 = \Delta(1/3)$$

Then we mathematically have

$$x_1 \leq 1/3 \leq x_2 \quad \text{with} \quad x_1, x_2 \in \mathbb{F}$$

More general $a, b \in \mathbb{F}$, we have :

$$\nabla(a \circ b) \leq a \circ b \leq \Delta(a \circ b)$$

for $\circ \in \{+, -, \times, /\}$

Directed roundings (cont'd)

With INTLAB

`setround(-1)`	rounding downwards
`setround(1)`	rounding upwards
`setround(0)`	rounding to nearest

Example:

```
setround(-1)
x = 1/3
setround(1)
y = 1/3
```

Then we have the mathematical inequality

$$x \leq 1/3 \leq y$$

Interval arithmetic

- Interval arithmetic : replace numbers by intervals and compute.
- Fundamental theorem of interval arithmetic: the exact result is contained in the computed interval.
- No result is lost, the computed interval is guaranteed to contain every possible result.

Definitions

Objects

- interval of real numbers : closed connected sets of $\mathbb{R}$
  - interval for π : $[3.14159, 3.14160]$
  - data d known with absolute uncertainty of ε : $[d-\varepsilon, d+\varepsilon]$
- interval vector

$$\boldsymbol{v} = \begin{pmatrix} [1,2] \\ [2,4] \end{pmatrix}$$

- interval matrix

$$\boldsymbol{A} = \begin{pmatrix} [1,3] & [3,4] \\ [2,5] & [1,2] \end{pmatrix}$$

Representation inf-sup of intervals

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

The set of interval of $\mathbb{R}$ is denoted $\mathbb{IR}$.

Operations on intervals

Given two intervals $\boldsymbol{x}$, $\boldsymbol{y}$ and $\diamond \in \{+, -, \times, /\}$, one defines

$$\boldsymbol{x} \diamond \boldsymbol{y} = \{x \diamond y : x \in \boldsymbol{x}, y \in \boldsymbol{y}\}.$$

On can implement these operations as :

$$\begin{aligned}
\boldsymbol{x} + \boldsymbol{y} &= [\underline{x} + \underline{y}; \overline{x} + \overline{y}], \\
\boldsymbol{x} - \boldsymbol{y} &= [\underline{x} - \overline{y}; \overline{x} - \underline{y}], \\
\boldsymbol{x} \times \boldsymbol{y} &= [\min\{\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \overline{x}\overline{y}\}; \max\{\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \overline{x}\overline{y}\}], \\
\boldsymbol{x}^2 &= [\min(\underline{x}^2, \overline{x}^2),\ \max(\underline{x}^2, \overline{x}^2)] \text{ if } 0 \notin [\underline{x},\ \overline{x}], \\
&\quad\ [0, \max(\underline{x}^2, \overline{x}^2)] \text{ otherwise}, \\
1/\boldsymbol{x} &= [1/\overline{x}; 1/\underline{x}] \text{ if } 0 \notin [\underline{x},\ \overline{x}], \\
\boldsymbol{x}/\boldsymbol{y} &= \boldsymbol{x} \times 1/\boldsymbol{y} \text{ if } 0 \notin [\underline{y},\ \overline{y}], \\
\sqrt{\boldsymbol{x}} &= [\sqrt{\underline{x}},\ \sqrt{\overline{x}}] \text{ if } 0 \leq \underline{x}, \\
&\quad\ [0,\ \sqrt{\overline{x}}] \text{ otherwise.}
\end{aligned}$$

Operations on intervals

In floating-point arithmetic, if one wants validated results, one need to take into account rounding errors !

$$\begin{aligned} \boldsymbol{x}+\boldsymbol{y} &= [\nabla(\underline{x}+\underline{y}), \Delta(\overline{x}+\overline{y})] \supseteq \{x+y | x \in \boldsymbol{x}, y \in \boldsymbol{y}\} \\ \boldsymbol{x}-\boldsymbol{y} &= [\nabla(\underline{x}-\overline{y}), \Delta(\overline{x}-\underline{y})] \supseteq \{x-y | x \in \boldsymbol{x}, y \in \boldsymbol{y}\} \end{aligned}$$

where ∇ (resp. Δ) representes rounding toward $-\infty$ (resp. rounding toward $+\infty$).

Operations on intervals (cont'd)

Algebraic properties : associativity and commutativity still hold

But lost :

- the subtraction is not the inverse of addition : $\boldsymbol{x} - \boldsymbol{x} \neq [0]$
- the division is not the inverse of multiplication
- ...

Intervals and functions

Definition : an interval extension $\boldsymbol{f}$ of f must satisfy

$$\forall \boldsymbol{x}, f(\boldsymbol{x}) \subseteq \boldsymbol{f}(\boldsymbol{x}) \text{ et } \forall x, f(\{x\}) = \boldsymbol{f}(\{x\})$$

Elementary functions :

$$\begin{aligned} \exp \boldsymbol{x} &= [\exp \underline{x}, \exp \overline{x}] \\ \sin[\pi/6, 2\pi/3] &= [1/2, 1] \end{aligned}$$

Outline of the talk

Proving that a matrix is nonsingular

Theorem 1

Let A be a matrix and R another matrix such that $\|I - RA\| < 1$. Then A is nonsingular

Proof.

By contrapositive, if A is singular, there exists $x \neq 0$ such that $Ax = 0$. Then $(I - RA)x = x$ and so $\|I - RA\| \geq 1$. □

On a computer, choose for $R \approx A^{-1}$ and then compute $\|I - RA\|$ with interval arithmetic.

Proving that a matrix is nonsingular with INTLAB

Let A be a matrix of dimension n

```
R = inv(A)
C = eye(n) - R*intval(A)
nonsingular = ( norm(C,1) < 1 )
```

If nonsingular = 1, then A is nonsingular.
If nonsingular = 0, then we can say nothing

A simple approach

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\hat{x} \in \mathbb{R}^n$ unknown such that $f(\hat{x}) = 0$

Let $\tilde{x} \approx \hat{x}$ such that $f(\tilde{x}) \approx 0$

Find a bound for $\tilde{x}$: an interval X such that $\hat{x} \in X$

We have

$$f(x) = 0 \qquad \Leftrightarrow \qquad g(x) = x$$

with $g(x) := x - Rf(x)$ with $\det(R) \neq 0$.

Theorem 2 (Brouwer, 1912)

Every continuous function from a closed ball of a Euclidean space to itself has a fixed point.

A simple approach (cont'd)

By Brouwer fixed point theorem,

$$X \in \mathbb{IR}^n, \quad g(X) \subseteq X \quad \Rightarrow \quad \exists \hat{x} \in X, \; g(\hat{x}) = \hat{x} \quad \Rightarrow \quad f(\hat{x}) = 0$$

We just have to check $g(X) \subseteq X$ and prove $\det(R) \neq 0$.

But naive approach fails:

$$g(X) \subseteq X - Rf(X) \not\subseteq X$$

Bounds for the solution of nonlinear systems

Mean Value Theorem :

$$\text{if } f \in \mathscr{C}^1 \text{ then } f(x) = f(\tilde{x}) + M(x - \tilde{x}) \text{ with } M = (\frac{\partial f}{\partial x}(\xi_i))_i$$

Let $Y := X - \tilde{x}$ and

$$\begin{array}{lcll} x \in X & \Rightarrow & g(x) - \tilde{x} & = & x - \tilde{x} - Rf(x) \\ & & & = & -Rf(\tilde{x}) + (I - RM)(x - \tilde{x}) \\ & & & \in & -Rf(\tilde{x}) + (I - RM)\,Y \end{array}$$

As a consequence

$$-Rf(\tilde{x}) + (I - RM)\,Y \subseteq Y \quad \Rightarrow \quad g(X) - \tilde{x} \subseteq Y \quad \Rightarrow \quad g(X) \subseteq X$$

Bounds for the solution of nonlinear systems (cont'd)

Theorem 3 (Rump, 1983)

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ with $f = (f_1, \ldots, f_n) \in \mathscr{C}^1$, $\tilde{x} \in \mathbb{R}^n$, $X \in \mathbb{IR}^n$ with $0 \in X$ and $R \in \mathbb{R}^{n \times n}$ be given. Let $M \in \mathbb{IR}^{n \times n}$ be given such that

$$\{\nabla f_i(\zeta) : \zeta \in \tilde{x} + X\} \subseteq M_{i,:} .$$

Denote by I the $n \times n$ identity matrix and assume

$$-Rf(\tilde{x}) + (I - RM)X \subseteq int(X).$$

Then there is a unique $\hat{x} \in \tilde{x} + X$ with $f(\hat{x}) = 0$. Moreover, every matrix $\tilde{M} \in M$ is nonsingular. In particular, the Jacobian $J_f(\hat{x}) = \frac{\partial f}{\partial x}(\hat{x})$ is nonsingular.

Remark

- Note that an inclusion of the range of the gradients ∇f_i over the set $\tilde{x} + X$ needs to be computed.
- A convenient way to do this in INTLAB is by interval arithmetic and the gradient toolbox. For a given (Matlab) function `f`, for `xs` = $\tilde{x}$ and an interval vector `X`, the call

$$\texttt{M = f(gradientinit(xs + X))}$$

computes an inclusion `M`.

Outline of the talk

Verification of multiple roots

- Verification method for computing guaranteed (real or complex) error bounds for double roots of systems of nonlinear equations.
- To circumvent the principle problem of ill-posedness we prove that a slightly perturbed system of nonlinear equations has a double root.
  For example, for a given univariate function $f : \mathbb{R} \to \mathbb{R}$ we compute two intervals $X, E \subseteq \mathbb{R}$ with the property that there exists $\hat{x} \in X$ and $\hat{e} \in E$ such that $\hat{x}$ is a double root of $\bar{f}(x) := f(x) - \hat{e}$.
- If the function f has a double root, typically the interval E is a very narrow interval around zero.

Verification of multiple roots

The typical scenario in the univariate case is a function $f : \mathbb{R} \to \mathbb{R}$ with a double root $\hat{x}$, i.e. $f(\hat{x}) = f'(\hat{x}) = 0$ and $f''(\hat{x}) \neq 0$.
Consider, for example,

$$\begin{aligned} f(x) &= 18x^7 - 183x^6 + 764x^5 - 1675x^4 + 2040x^3 - 1336x^2 + 416x - 48 \\ &= (3x-1)^2(2x-3)(x-2)^4 \end{aligned}$$

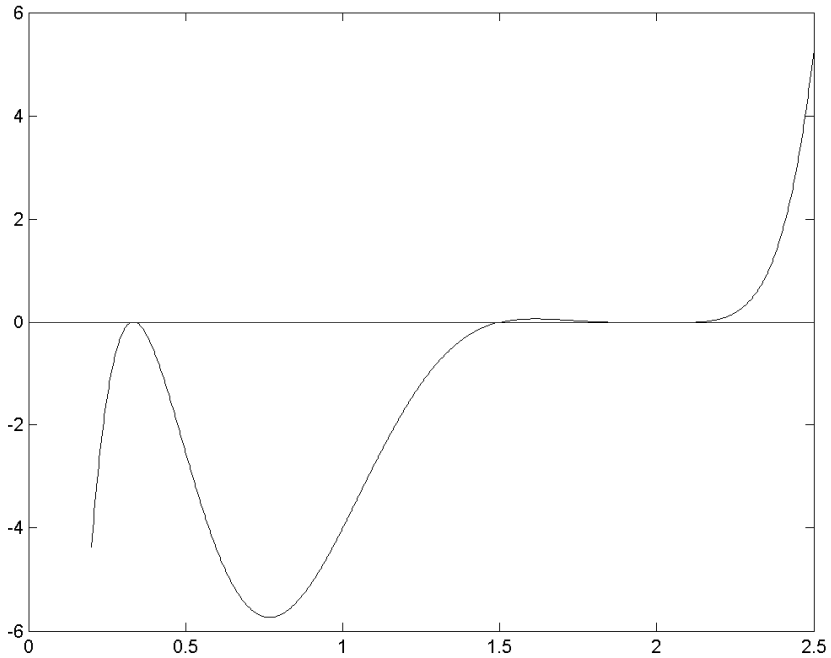


Figure: Graph of $f(x) = (3x-1)^2(2x-3)(x-2)^4$.

Verification of multiple roots

- Verification methods for multiple roots of polynomials already exist (Rump,2003). A set containing k roots of a polynomial is computed, but no information on the true multiplicity can be given.
- A hybrid algorithm based on the methods of (Rump,2003) is implemented in algorithm `verifypoly` in INTLAB. Computing inclusions `X1`, `X2` and `X3` of the simple root $x_1 = 1.5$, the double root $x_2 = 1/3$ and the quadruple root $x_3 = 2$ of f by algorithm `verifypoly` in INTLAB we obtain the following.

```
>> X1 = verifypoly(f,1.3), X2 = verifypoly(f,.3), X3 = verifypoly(f,2.1)
intval X1 =
[   1.49999999999904,   1.50000000000078]
intval X2 =
[   0.33333316656015,   0.33333343640539]
intval X3 =
[    1.99741678159164,   2.00363593397305]
```

Verification of multiple roots (cont'd)

- The accuracy of the inclusion of the double root $x_2 = 1/3$ is much less than that of the simple root $x_1 = 1.5$, and this is typical.
- If we perturb f into $\tilde{f}(x) := f(x) - \varepsilon$ for some small real constant ε and look at a perturbed root $\tilde{f}(\hat{x}+h)$ of $\tilde{f}$, then

$$0 = \tilde{f}(\hat{x}+h) = -\varepsilon + \frac{1}{2}f''(\hat{x})h^2 + \mathcal{O}(h^3)$$

  implies

$$h \sim \sqrt{2\varepsilon / f''(\hat{x})}.$$

- In general floating-point computations are afflicted with a relative error of size $\varepsilon \approx 10^{-16}$. This has the same effect as a perturbation of the given function f into $\tilde{f}$. But for double roots, we cannot expect this inclusion to be of better relative accuracy than $\sqrt{\varepsilon} \approx 10^{-8}$.

Dealing with double roots

We consider for a double root the nonlinear system $G : \mathbb{R}^2 \to \mathbb{R}$ with

$$G(x, e) = \begin{pmatrix} f(x) - e \\ f'(x) \end{pmatrix} = 0$$

in the two unknowns x and e. The Jacobian of this system is

$$J_G(x, e) = \begin{pmatrix} f'(x) & -1 \\ f''(x) & 0 \end{pmatrix},$$

so that the nonlinear system is well-conditioned for the double root $x_2 = 1/3$ of f.

Dealing with double roots (cont'd)

- Now we can apply a verification algorithm for solving general systems of nonlinear equation such as algorithm `verifynlss` in INTLAB. Indeed, applying algorithm `verifynlss` we obtain

```
>> Y2 = verifynlss(G,[.3;0])
intval Y2 =
[  3.33333333333328e-001,  3.33333333333337e-001]
[ -2.13162820728042e-014,  2.13162820728042e-014]
```

- This proves that there is a constant ε with $|\varepsilon| \leq 2.14 \cdot 10^{-14}$ such that the nonlinear equation $f(x) - \varepsilon = 0$ has a double root $\hat{x}$ with $0.333333333333328 \leq \hat{x} \leq 0.333333333333337$.

Dealing with double roots (cont'd)

- We presented the previous approach in preparation for the multivariate case;
- however, for univariate nonlinear functions we may proceed more directly.

  Suppose $X \in \mathbb{IR}$ is an inclusion of a root $\hat{x}$ of f', and use the interval evaluation of f at X to compute $E \in \mathbb{IR}$ with $f(X) \subseteq E$. In particular $f(\hat{x}) \in E$, so that there exists $\hat{e} \in E$ such that the function $g(x) := f(x) - \hat{e}$ satisfies $g(\hat{x}) = g'(\hat{x}) = 0$.
- If, moreover, the inclusion X is computed by a verification method, then $\hat{x}$ is a unique root of f' in X, and $\hat{x}$ is proved to be a double root of g.

Dealing with double roots (cont'd)

By this approach we obtain the inclusions for the double root $\hat{x}$ are of the same quality, but the inclusion for the shift is a little weaker than in `Y2`:

```
intval X =
[  3.33333333333329e-001,  3.33333333333339e-001]
intval E =
[ -3.126388037344441e-013,  2.913225216616412e-013]
```

Dealing with double roots (cont'd)

However, it is superior to expand f with respect to some point $m \in X$.
For all $x \in X$ we have $f(x) \in f(m) + f'(X)(X - m) =: E_1$, and in particular $f(\hat{x}) \in E_1$.
Here m should be close to the midpoint of X, but need not to be equal to the midpoint. In this case we obtain with

```
intval E1 =
[ -2.131628207280369e-014,  2.131628207280378e-014]
```

an inclusion of the same quality as `Y2` by solving G.

Note that we use only a univariate verification method to include a root of f', the shift E is obtained by a mere function evaluation.

The multivariate case

- Let a suitably smooth function $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\hat{x} \in \mathbb{R}^n$ be given such that $f(\hat{x}) = 0$ and the Jacobian of f at $\hat{x}$ is singular.
- A standard verification method such as `verifynlss` must fail because with an inclusion of a root the nonsingularity of the Jacobian at the root is proved as well.
- Again it is an ill-posed problem and we need some regularization technique.

The multivariate case (cont'd)

Consider the model problem

$$f(x,y) = \begin{pmatrix} f_1(x,y) \\ f_2(x,y) \end{pmatrix} = \begin{pmatrix} x^2 + (x+1)(y-1)^2 - \operatorname{asinh}((x+3)^3 + y^2)\cos(x - xy) \\ (x + 1.90871887406161 8)^2 - \sin(x)(y+1)^2 \end{pmatrix} = 0$$

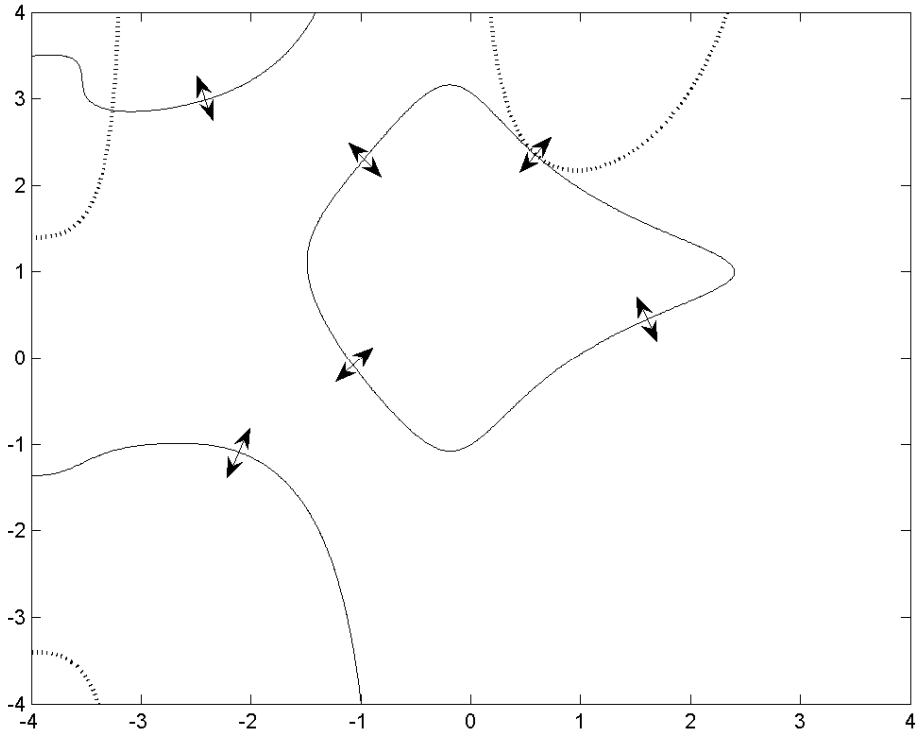


Figure: Contour lines of $f_1(x) = 0$ (solid) and $f_2(x) = 0$ (dashed)

The multivariate case (cont'd)

- As a regularization we add, similar to the univariate case, a smoothing parameter e and rewrite into

$$F(x, y, e) = \begin{pmatrix} f_1(x, y) - e \\ f_2(x, y) \\ \det J_f(x, y) \end{pmatrix} = 0 \, .$$

- The third equation forces the tangents of the zero contour lines to be parallel at the solution, whereas the first equation introduces a perturbation to f_1 so that the root becomes a double root.

This approach may work for two or three unknowns, however, an explicit formula for the determinant of the Jacobian is prohibitive for larger dimensions. Consider the following way to ensure the Jacobian to be singular.

The multivariate case (cont'd)

Let a function $f = (f_1, \ldots, f_n) : \mathbb{R}^n \to \mathbb{R}^n$ be given and let $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ be such that $f(\hat{x}) = 0$ and the Jacobian $J_f(\hat{x})$ of f at $\hat{x}$ is singular. Adding a smoothing parameter e we arrive with $g : \mathbb{R}^{n+1} \to \mathbb{R}^n$ and

$$g(x, e) = \begin{pmatrix} f_1(x) - e \\ f_2(x) \\ \cdots \\ f_n(x) \end{pmatrix} = 0$$

at n equations in $n+1$ unknowns. We force the Jacobian to be singular by

$$J_f(x)y = 0$$

for some vector y in the kernel of J_f. In order to make y unique we normalize some component of y to 1.

The multivariate case (cont'd)

Theorem 4

Let $f = (f_1, \ldots, f_n) : \mathbb{R}^n \to \mathbb{R}^n$ *with* $f \in \mathscr{C}^2$ *be given. Define* $F : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ *by*

$$F(x, e, y) = \begin{pmatrix} g(x, e) \\ J_f(x)y \end{pmatrix} = 0 \,,$$

where $x = (x_1, \ldots, x_n)$, $e \in \mathbb{R}$ *and* $y = (1, y_2, \ldots, y_n)$. *Suppose* F *suitable assumptions and yields inclusions for* $\hat{x} \in \mathbb{R}^n$, $\hat{e} \in \mathbb{R}$ *and* $\hat{y} \in \mathbb{R}^{n-1}$ *such that* $F(\hat{x}, \hat{e}, \hat{y}) = 0$. *Then* $g(\hat{x}, \hat{e}) = f(\hat{x}) - (\hat{e}, 0, \ldots, 0)^T = 0$, *and the rank of the Jacobian* $J_f(\hat{x})$ *of* f *at* $\hat{x}$ *is* $n-1$.

The multivariate case (cont'd)

The system

$$f(x_1, x_2) = \begin{pmatrix} x_1^2 - x_2^2 \\ x_1 - x_2^2 \end{pmatrix} = 0$$

yields

$$J_F(x, e, y) = \begin{pmatrix} 2x_1 & -2x_2 & -1 & 0 \\ 1 & -2x_2 & 0 & 0 \\ 0 & -2 & 0 & 1 \\ 2y & -2 & 0 & 2x_1 \end{pmatrix},$$

as the Jacobian of the augmented system, which is nonsingular for $x_1 = x_2 = 0$. Thus an inclusion is in principle possible.

The multivariate case (cont'd)

```
>> f=inline('[x(1)^2-x(2)^2;x(1)-x(2)^2]'),
   verifynlss2(f,[0.002;0.001])
f =
     Inline function:
     f(x) = [x(1)^2-x(2)^2;x(1)-x(2)^2]
intval ans =
  1.0e-323 *
[  -0.6666666666666,   0.6666666666666]
[  -1.0000000000000,   1.0000000000000]
[  -1.0000000000000,   1.0000000000000]
```

Verified multiple eigenvalues

Computing eigenvalues can be viewed as solving the nonlinear system:

$$f(x,\lambda) = \begin{pmatrix} Ax - \lambda x \\ e_k^T x - 1 \end{pmatrix} = 0\,,$$

As before we regularize the system, but now not by shifting a whole partial function but by changing an individual component a_{ij} of A:

$$g(x,\lambda,\varepsilon,y) = \begin{pmatrix} Ax - \lambda x - \varepsilon x_j e_i \\ e_k^T x - 1 \\ J_f(x,\lambda)y \end{pmatrix} = 0\,.$$

Again an inclusion is calculated. In this case, the rank of the Jacobian

$$J_f(x,\lambda) = \begin{pmatrix} A - \lambda I & -x \\ e_k^T & 0 \end{pmatrix}$$

is proved to be n and we can also prove that the eigenvalue is of geometric multiplicity one.

Outline of the talk

First example

Consider

$$f(x) = (\sin(x) - 1)(x - \alpha) \quad \text{for } \alpha := \frac{\pi}{2}(1+\varepsilon)\ .$$

The function f has a double root $\hat{x} = \pi/2$ with another simple root α of relative distance ε to $\pi/2$. Hence we expect the inclusion E of the offset e for regularization to be a narrow inclusion of zero.

ε	X	E
10^{-2}	$1.5707963267949 \pm 1.8 \cdot 10^{-14}$	$[-3.5, 1.8] \cdot 10^{-18}$
10^{-3}	$1.5707963267948 \pm 1.7 \cdot 10^{-13}$	$[-3.5, 1.8] \cdot 10^{-19}$
10^{-4}	$1.570796326795 \pm 1.6 \cdot 10^{-12}$	$[-3.5, 1.8] \cdot 10^{-20}$
10^{-5}	$1.57079632679 \pm 1.2 \cdot 10^{-10}$	$[-3.5, 1.8] \cdot 10^{-21}$
10^{-6}	$1.5707963268 \pm 1.5 \cdot 10^{-9}$	$[-3.5, 1.8] \cdot 10^{-22}$
10^{-7}	$1.570796327 \pm 1.6 \cdot 10^{-8}$	$[-3.5, 1.8] \cdot 10^{-23}$
10^{-8}	failed	

Table: Inclusions for the double root $\hat{x} = \pi/2$ and a nearby simple root α for f

Second example

Consider now

$$f(x) = (\sin(x) - 1)(x - \alpha)^2 \quad \text{for } \alpha := \frac{\pi}{2}(1 + \varepsilon) \ ,$$

so that there is a double root α near the double root $\hat{x}$. For a relative distance ε of about $\sqrt[4]{\varepsilon} \sim 10^{-4}$ the four roots behave like a quadruple root. This is confirmed by the results in the Table.

ε	X	E
10^{-2}	$1.57079632679488 \pm 1.2 \cdot 10^{-14}$	$[-2.8, 5.5] \cdot 10^{-20}$
10^{-3}	$1.5707963267948 \pm 2.4 \cdot 10^{-13}$	$[-2.8, 5.5] \cdot 10^{-22}$
10^{-4}	$1.570796326794 \pm 2.8 \cdot 10^{-12}$	$[-2.8, 5.5] \cdot 10^{-24}$
10^{-5}	failed	

Table: Inclusions for the double root $\hat{x} = \pi/2$ and a nearby double root α for f

Some systems of nonlinear equations

The first test function is

$$f(x_1, x_2) = \begin{pmatrix} e^{x_1 x_2} - \sin(x_1^2 - 2x_1 x_2) \\ x_1(x_1 - \cosh(x_2)) + x_1 \operatorname{atan}(x_2) - \alpha \end{pmatrix} = 0\,,$$

where we choose the parameter α such that the system has a nearly double root. For example, for $\alpha = 0.4$ the zero contour lines look like in Figure.

Some systems of nonlinear equations (cont'd)

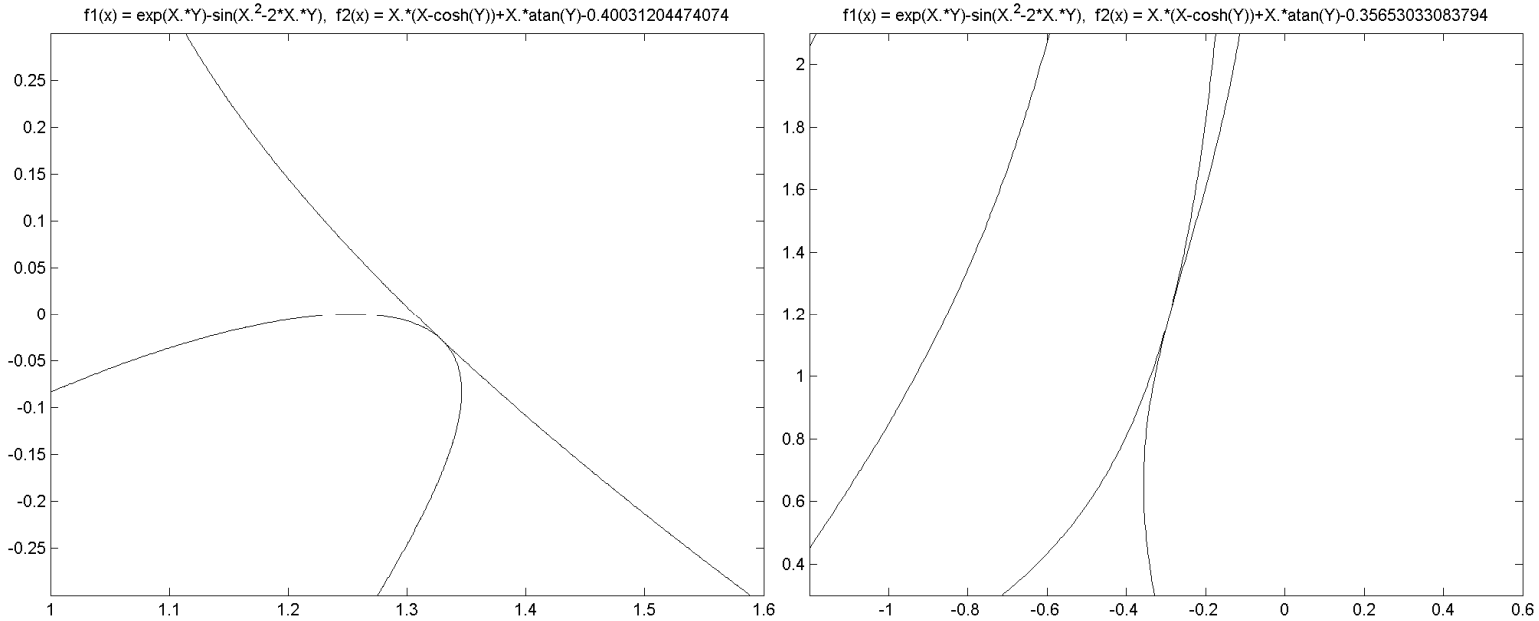


Figure: Zero contour lines of $f(x_1, x_2)$ for two different parameter values α.

X_1	X_2	X	E
$1.32889962\,1_{28}^{86}$	$1.3288995\,1_{48}^{57}$	$1.3288995683907\,1_{5}^{6}$	
$-0.0272980\,5_{59}^{67}$	$-0.0272979\,2_{88}^{98}$	$-0.027297992758\,79_{34}^{41}$	[-5.2,-5.0]$\cdot 10^{-14}$

Table: Inclusions X_1, X_2 for two single roots and X for a nearly double root for f and $\alpha = 0.4003120447407$.

Some systems of nonlinear equations (cont'd)

X_1	X_2	X	E
-0.29197330_{44}^{91}	-0.2919733_{57}^{61}	$-0.291973333127\,64_{29}^{41}$	
1.1950051_{00}^{23}	1.1950048_{53}^{69}	$1.1950049857509_{87}^{92}$	[-1.17,-0.96]$\cdot 10^{-14}$

Table: Inclusions X_1, X_2 for two single roots and X for a nearly double root for f and $\alpha = 0.35653033083794$.

Example of higher dimensions

Consider Brown's almost linear function $f : \mathbb{R}^n \to \mathbb{R}^n$ with

$$f_k(x) = x_k + \sum_{j=1}^{n} x_j - (n+1) \quad \text{for} \quad 1 \le k \le n-1 \,,$$
$$f_n(x) = \Big(\prod_{j=1}^{n} x_j\Big) - 1 - e \,,$$

where the last function is shifted by some e. One verifies that for

$$e = \left(1 - \frac{1}{n^2}\right)^{n-1} \left(1 + \frac{1}{n}\right) - 1$$

and

$$\bar{x}_k = 1 - \tfrac{1}{n^2} \quad \text{for} \quad 1 \le k \le n-1 \,,$$
$$\bar{x}_n = 1 + \tfrac{1}{n}$$

the vector $(1, \ldots, 1, -n)$ is in the kernel of the Jacobian of f.

Example of higher dimensions (cont'd)

Thus $\bar{x}$ is not a simple root of f. More precisely it is verified that there exists $\hat{x} \in X$ and $\hat{\varepsilon} \in E$ such that $f(\hat{x}) - (\hat{\varepsilon}, \ldots, 0) = 0$ and the Jacobian $J_f(\hat{x})$ of f at $\hat{x}$ is singular.

n	$X_{1\cdots n-1}$	X_n	E
10	$0.990000 \pm 1.0 \cdot 10^{-14}$	$1.100000 \pm 1 \cdot 10^{-14}$	$[-3.5, 5.8] \cdot 10^{-15}$
20	$0.997500 \pm 4.0 \cdot 10^{-14}$	$1.050000 \pm 1 \cdot 10^{-14}$	$[-1.4, 2.2] \cdot 10^{-14}$
50	$0.996000 \pm 2.1 \cdot 10^{-13}$	$1.020000 \pm 2 \cdot 10^{-14}$	$[-0.1, 1.9] \cdot 10^{-13}$
100	$0.999900 \pm 8.2 \cdot 10^{-13}$	$1.010000 \pm 2 \cdot 10^{-14}$	$[-5.4, 2.9] \cdot 10^{-13}$
200	$0.999975 \pm 3.3 \cdot 10^{-12}$	$1.005000 \pm 5 \cdot 10^{-14}$	$[-1.3, 2.0] \cdot 10^{-12}$
500	$0.999996 \pm 1.9 \cdot 10^{-11}$	$1.002000 \pm 1 \cdot 10^{-13}$	$[-0.6, 1.3] \cdot 10^{-11}$
1000	$0.999999 \pm 7.5 \cdot 10^{-11}$	$1.001000 \pm 2 \cdot 10^{-13}$	$[-1.1, 6.4] \cdot 10^{-11}$

Table: Inclusions of a double root for different dimensions.

Conclusion and future work

Conclusion:

- Efficient algorithms for computing verified and narrow error bounds with the property that a slightly perturbed system is proved to have a double root within the computed bounds
- Applied those to univariate polynomials, to multivariate polynomials and also to eigenvalue problems
- Numerical experiments have confirmed the performance of our algorithms

Future work:

- Detecting singular matrices
- Applications to approximate coprimeness

Bibliography I


Siegfried M. Rump and Stef Graillat.
Verified error bounds for multiple roots of systems of nonlinear equations.
Numer. Algorithms, 54 (2010), no. 3, 359-377.


Siegfried M. Rump.
Verification methods: Rigorous results using floating-point arithmetic.
Acta Numerica (2010), pp. 287-449.


Bo Einarsson.
Accuracy and Reliability in Scientific Computing.
Software-Environments-Tools. SIAM, Philadelphia, PA, 2005.

Bibliography II


Nicholas J. Higham.
Accuracy and stability of numerical algorithms.
Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2002.


Jean-Michel Muller et al.
Handbook of Floating-Point Arithmetic.
Birkhäuser, 2010.


R. Moore, R. Kearfott et M. Cloud.
Introduction to Interval Analysis.
SIAM, 2009.

Thank you for your attention