Recent Advances on Optimization
Toulouse, France
July 24-26, 2013

Do You Trust Your Algorithms?

Uncertainty and Sensitivity in Complex Systems

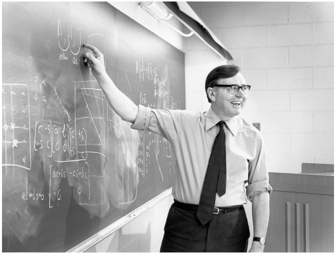
Jorge Moré

Joint work with Stefan Wild

Mathematics and Computer Science Division
Argonne National Laboratory

In the Beginning

The development of automatic digital computers has made it possible to carry out computations involving a very large number of arithmetic operations and this has stimulated a study of the cumulative effect of rounding errors.



J. H. Wilkinson, Rounding Errors in Algebraic Processes, 1963

Rounding Errors and Stability

The output of an algorithm $\mathcal{A}$ is defined by $f : \mathbb{R}^n \mapsto \mathbb{R}$.

- $f_\infty(x)$: Output when $\mathcal{A}$ is executed in infinite precision
- $f(x)$: Output when $\mathcal{A}$ is executed in working precision

Rounding errors are measured by $|f_\infty(x) - f(x)|$

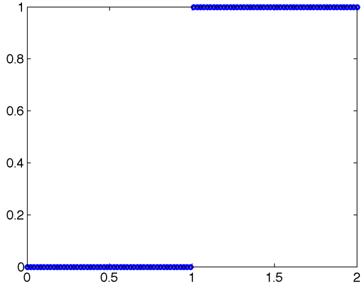
For backward-stable computations,

$$f_\infty(x + \delta x) = f(x)$$

for *small* perturbations δx.

Rounding Errors: A Cautionary Example

```
f = x;
for k = 1:L
  f = sqrt(f);
end
for k = 1:L
  f = f^2;
end
f = f^2;
```



Plot of f for $L \geq 60$

⚠ This algorithm is not backward stable.

W. Kahan, Interval arithmetic options in the proposed IEEE . . . standard, 1980

HP-15C Advanced Functions Handbook, 1982

Mindless Assessment of Roundoff

Repeat the computation but . . .

- ◇ in higher precision
- ◇ with a different rounding mode
- ◇ with random rounding
- ◇ use slightly different inputs
- ◇ use interval arithmetic

How futile are mindless assessments of roundoff in floating-point computation?
W. Kahan, 2006. Work in progress, 56 pages.

CADNA: A library for estimating round-off error propagation,
F. Jézéquél and J-M. Chesneaux, Computer Physics Communications, 2008.

Uncertainty and Computational Noise

The uncertainty in f is an estimate of

$$|f(x + \delta x) - f(x)|$$

for a *small* perturbation δx.

⚠ If the computed f is backward-stable, then the uncertainty is

$$|f_\infty(x + \delta x_r) - f(x)|$$

for a small δx_r. This is an estimate of the rounding errors.

J. Moré and S. Wild, Estimating computational noise, 2011.

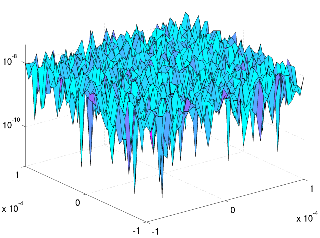
Argonne NATIONAL LABORATORY

Research Issues

- What is a noisy function f?
- Determine the noise (uncertainty) in f with a few evaluations
- Reliably approximate a derivative of f
- How do you optimize f?

Argonne NATIONAL LABORATORY

Computational Noise ~ Uncertainty



Leading causes of noise

◇ 10^X flops

◇ Iterative calculations

◇ Adaptive algorithms

◇ Mixed precision

Definition. The *noise level* of the computed f in a region Ω is

$$\varepsilon_f = \mathrm{E}\left\{ \tfrac{1}{2}\Big(f(\boldsymbol{x}_2) - f(\boldsymbol{x}_1) \Big)^2 \right\}^{1/2},$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are $\mathbf{iid}$ random vectors with range in Ω.

Argonne NATIONAL LABORATORY

Two Theorems

Let $\boldsymbol{x}$ be a random vector with range in Ω.

Theorem 1.

$$\varepsilon_f = \mathrm{E}\left\{ |f(\boldsymbol{x}) - \mu|^2 \right\}^{1/2}, \qquad \mu = \mathrm{E}\left\{ f(\boldsymbol{x}) \right\}$$

Theorem 2.

$$\varepsilon_f \le \mathrm{E}\left\{ |f(\boldsymbol{x}) - f_\infty(\boldsymbol{x})|^2 \right\}^{1/2} + \max_{x_1, x_2 \in \Omega} \| f(x_2) - f(x_1) \|$$

⚠ $\varepsilon_R = \mathrm{E}\left\{ |f(\boldsymbol{x}) - f_\infty(\boldsymbol{x})|^2 \right\}^{1/2}$ measures rounding errors.

The Noise Level ε_f and Uncertainty

Let μ be the expected value of $f(\boldsymbol{x})$.

Chebyshev inequality

$$\mathcal{P}\Big\{|f(\boldsymbol{x})-\mu| > \gamma\varepsilon_f\Big\} \le \frac{1}{\gamma^2}$$

Cauchy-Schwartz inequality

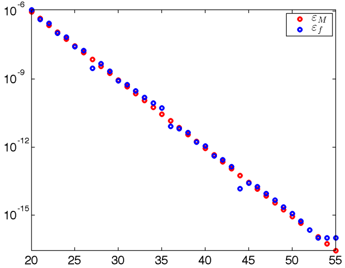
$$\mathrm{E}\left\{|f(\boldsymbol{x})-\mu|\right\} \le \varepsilon_f$$

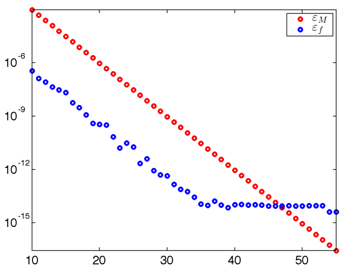
Two Claims

- The noise level ε_f is a measure of the uncertainty of f
- We can determine ε_f in a few function evaluations

The Noise Level ε_f: Basic Examples

Experiments with truncating $x \in \mathbb{R}^n$ to t bits: chop(x, t)



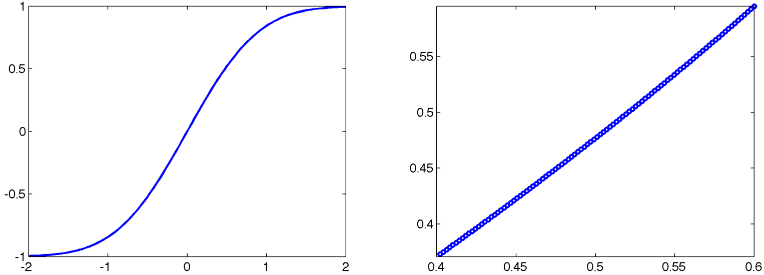


$x \mapsto \|\text{chop}(x, t)\|^2$ with $n = 1$ (left) and $n = 10^6$ (right)

Argonne NATIONAL LABORATORY

Case Study: Nonlinear Solvers

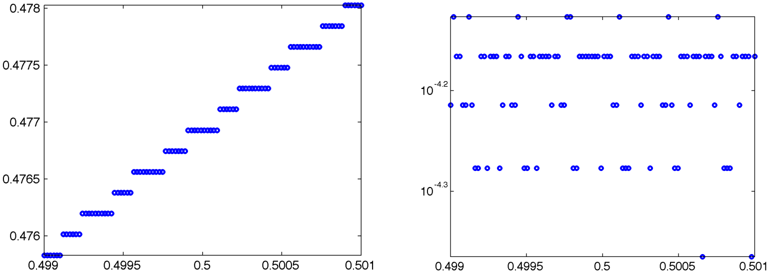
Compute f by solving $\operatorname{erf}[f(t)] = t$ with tolerance τ



erf (left) and the inverse f computed with bisection (right)

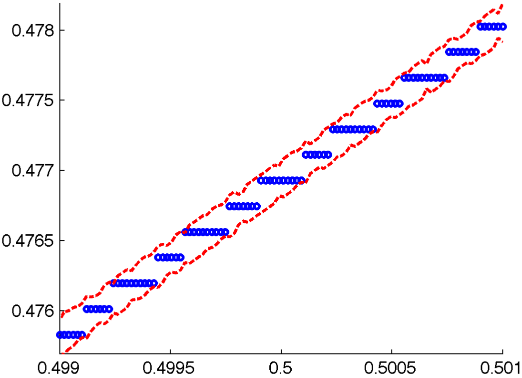
Bisection tolerance: $\tau = 10^{-4}$

Case Study: Nonlinear Solvers



Computed f with bisection (left) and the absolute noise ε_f (right)

Case Study: Nonlinear Solvers



Computed f (blue) and $f_s \pm 2\varepsilon_f$ (red)

f_s is the linear l_2 fit to the f values

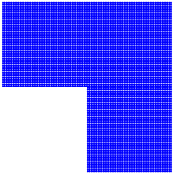
Case Study: Eigenvalue Solvers

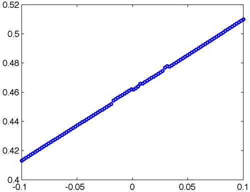
Define $f : \mathbb{R} \mapsto \mathbb{R}$ by

$$f(t) = \sum_{i=1}^{p} \lambda_i \Big(A + \text{diag}\,(x_b + tp) \Big)$$

where $\lambda_i(\cdot)$ is the i-th eigenvalue computed with tolerance τ.

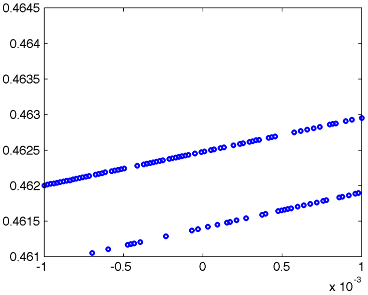
Vectors x_b and p are random in $[0, 1]$

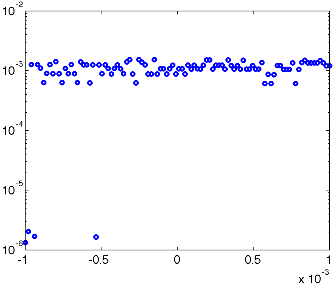




A is the Laplacian on an L-shaped region, $n = 7203$

Case Study: Eigenvalue Solvers

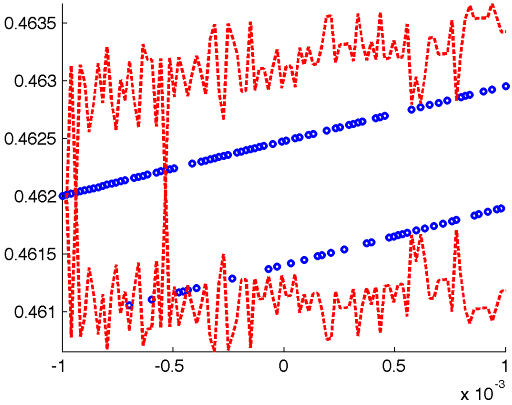




Computed f with eigs (left) and the relative noise ε_f (right)

Tolerance $\tau = 10^{-3}$

Case Study: Eigenvalue Solvers



The computed f (blue) and $f_s \pm 2\varepsilon_f$ (red)

f_s is the linear l_2 fit to the f values

The Noise Level: Assumptions and Analysis

We assume that the computed function $f : \mathbb{R}^n \mapsto \mathbb{R}$ satisfies

$$f[\boldsymbol{x}(t)] = f_s(t) + \boldsymbol{\varepsilon}(t), \qquad t \in [0,1]$$

where $f_s : \mathbb{R} \mapsto \mathbb{R}$ is smooth and $\boldsymbol{\varepsilon} : \mathbb{R} \mapsto \mathbb{R}$ is the **iid** noise.

This model accounts for

- ◇ Changes in computer, software libraries, operating system, ...
- ◇ Code changes and reformulations
- ◇ Asynchronous, highly-concurrent algorithms
- ◇ Stochastic methods
- ◇ Variable/adaptive precision methods

Argonne NATIONAL LABORATORY

ECnoise: Computing the Noise Level

◇ Construct the k-th order differences of f

$$\Delta^{k+1} f(t) = \Delta^k f(t+h) - \Delta^k f(t).$$

1.74e+03	4.92e-04	-1.98e-06	4.02e-06	-6.95e-06	9.74e-06	-1.03e-05	8.40e-06
1.74e+03	4.90e-04	2.04e-06	-2.93e-06	2.79e-06	-5.11e-07	-1.85e-06	
1.74e+03	4.92e-04	-8.92e-07	-1.39e-07	2.28e-06	-2.36e-06		
1.74e+03	4.91e-04	-1.03e-06	2.14e-06	-7.83e-08			
1.74e+03	4.90e-04	1.11e-06	2.07e-06				
1.74e+03	4.91e-04	3.18e-06					
1.74e+03	4.94e-04						
1.74e+03							

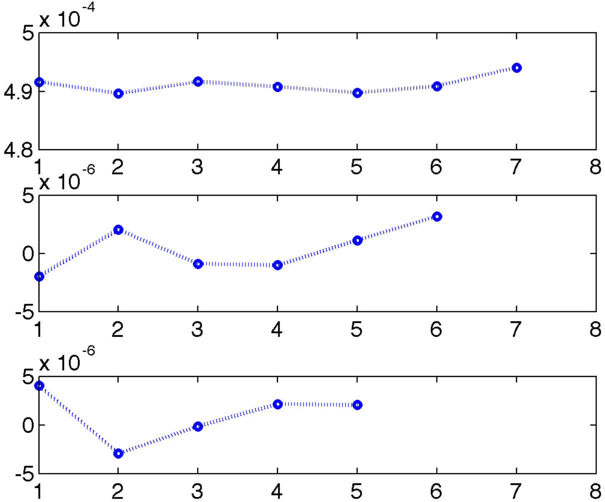
◇ Estimate the noise level from

$$\lim_{h \to 0} \gamma_k \, \mathbb{E} \left\{ \left[\Delta^k f(t) \right]^2 \right\} = \varepsilon_f^2, \qquad \gamma_k = \frac{(k!)^2}{(2k)!}.$$

R. W. Hamming, Introduction to Applied Numerical Analysis, 1971

J. Moré and S. Wild, Estimating computational noise, 2011.

ECnoise: Noise Levels for $\Delta f, \Delta^2 f, \ldots$



Argonne NATIONAL LABORATORY

The Simplest Simulations: Krylov Solvers

Define $f_\tau : \mathbb{R}^n \mapsto \mathbb{R}$ as the iterative solution of a Krylov solver,

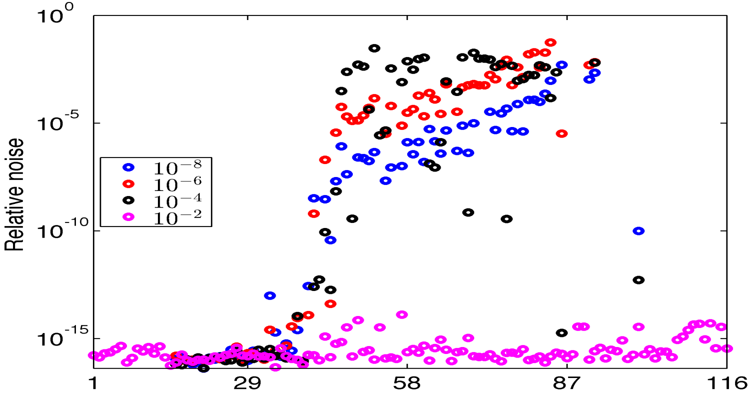
$$f_\tau(x) = \|y_\tau(x)\|^2, \qquad Ay_\tau(x) = b(x),$$

with relative residual error τ. We use $b(x) = x$.

⚠ $y_\tau : \mathbb{R}^n \mapsto \mathbb{R}^n$ is continuously differentiable for almost all τ

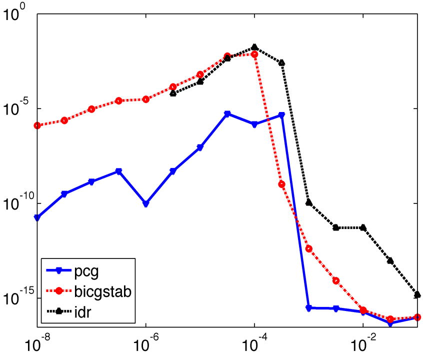
- ◇ UF symmetric positive definite matrices (116) with $n \leq 10^4$
- ◇ Scaling: $A \leftarrow D^{-1/2} A D^{-1/2}, \quad D = \text{diag}(a_{i,i})$
- ◇ Solvers: bicgstab (similar results for pcg, minres, gmres, ...)
- ◇ Tolerances: $\tau \in [10^{-8}, 10^{-1}]$

What is the Noise Level of Krylov Simulations?



Distribution of ε_f for f_τ (bicgstab)

New Phenomena: Noise Level Transitions



ε_f as a function of tolerance τ

Computing Derivatives of Noisy Simulations

Research Issues

- ◇ What is the noise level of the derivative?
- ◇ Is the noise level of the derivative higher than ε_f?

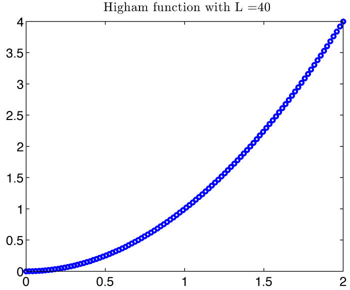
Algorithmic (Automatic) differentiation in MATLAB

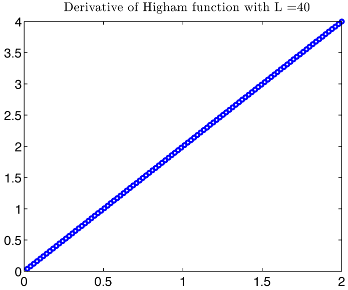
- ◇ IntLab (Siegfried Rump, Hamburg)
- ◇ AdiMat (Andre Vehreschild, Aachen)

In our numerical results we use IntLab.

J. Moré and S. Wild, Estimating derivatives of noisy simulations, 2012.

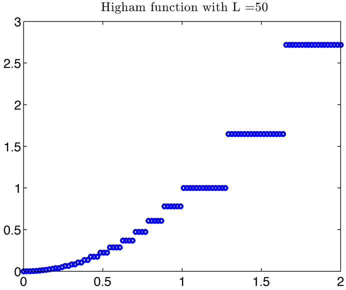
Uncertainty in f': Cautionary Examples

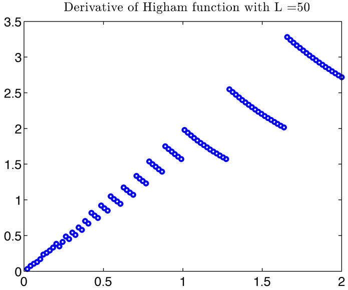




Higham function f with $L = 40$ (left) and f' (right)

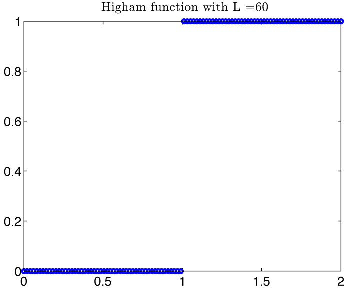
Uncertainty in f': Cautionary Examples

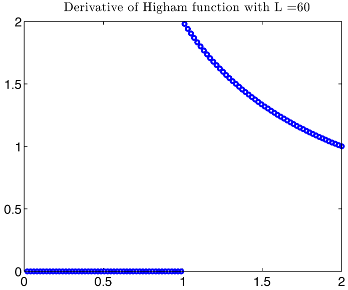




Higham function f with $L = 50$ (left) and f' (right)

Uncertainty in f': Cautionary Examples





Higham function f with $L = 60$ (left) and f' (right)

Measuring Uncertainty in f'

$$\text{re}(f') = \text{re}\Big\{ f'(x_0; p), f'(x_0; (1+\varepsilon)p) \Big\}, \qquad \varepsilon = \varepsilon_M .$$

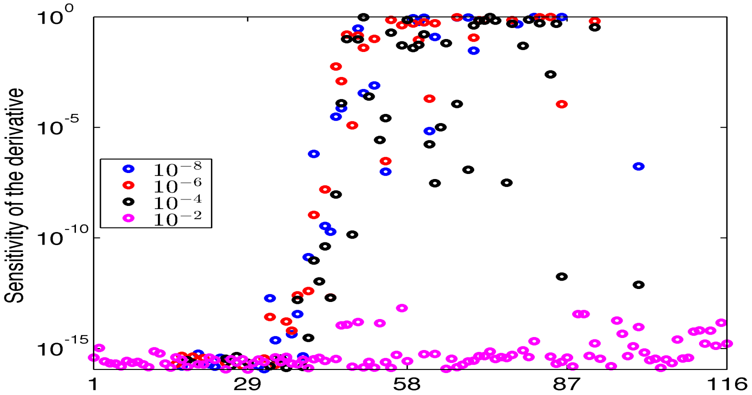
⚠ With infinite precision $\text{re}(f') = \varepsilon$ for any $\varepsilon > 0$

The function $\text{re}\,(\cdot,\cdot)$ is the relative error metric (Ziv [1982])

$$\text{re}\,(\alpha, \beta) = \frac{|\alpha - \beta|}{\max(|\alpha|, |\beta|)} .$$

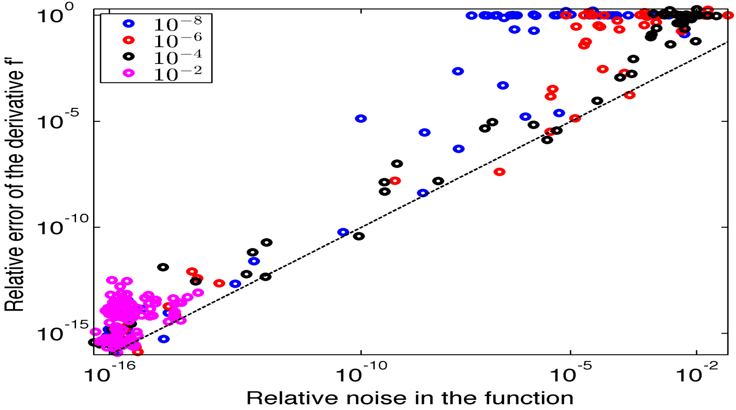
- ◇ $\text{re}\,(\alpha, \beta) \in (\frac{1}{2}, 1]$ if and only if $\max(|\alpha|, |\beta|) \geq 2 \min(|\alpha|, |\beta|)$
- ◇ $\text{re}\,(\alpha, \beta) \in (1, 2]$ if and only if α and β have opposite signs.

Can You Trust Derivatives?



Distribution of re(f'_τ) for f_τ (bicgstab)

Derivatives have Uncertainty re$(f'_\tau) \gg \varepsilon_f$



Distribution of $(\varepsilon_f, \text{re}(f'_\tau))$ for f_τ (bicgstab)

Dashed line is (t, t)

Argonne NATIONAL LABORATORY

Further Reading

S. Wild, Estimating Computational Noise in Numerical Simulations
www.mcs.anl.gov/~wild/cnoise

- J. Moré and S. Wild, *Estimating Computational Noise*, SIAM Journal on Scientific Computing, 33 (2011).
- J. Moré and S. Wild, *Estimating Derivatives of Noisy Simulations*, ACM Trans. Mathematical Software, 38 (2012).
- J. Moré and S. Wild, *Do You Trust Derivatives or Differences?*, Mathematics and Computer Science Division, Preprint ANL/MCS-P2067-0312, April 2012