Thirty Years of Parallel Computing at Argonne
Argonne, Illinois, May 14-15, 2013

Do You Trust Your Algorithms?

Uncertainty and Sensitivity in Complex Systems

Jorge Moré

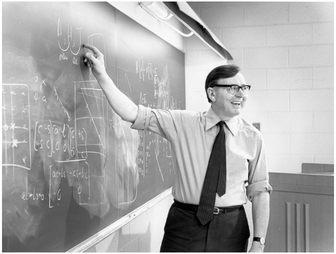
Joint work with Stefan Wild

Mathematics and Computer Science Division
Argonne National Laboratory

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 $|f_\infty(x) - f(x)|$

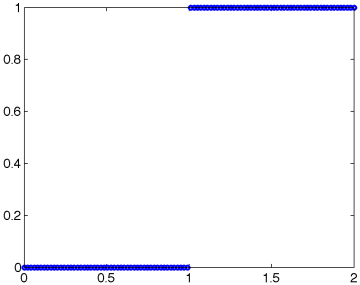
For backward-stable computations,

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

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.

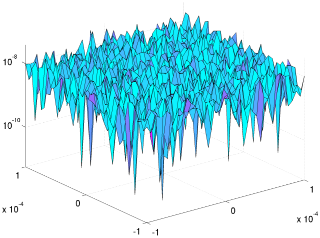
J. Moré and S. Wild, Estimating computational noise, 2011.

Argonne NATIONAL LABORATORY

Research Issues

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

Computational Noise $\sim$ 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

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

where Ω contains x and all permissible perturbations of x

Argonne NATIONAL LABORATORY

Two Theorems

Theorem 1. If $\mathcal{F}$ is the space of all $\mathbf{iid}$ $\boldsymbol{x} \mapsto \Omega$,

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

Note. The computed function f is a step function.

Theorem 2. If f is a step function with values $v_1 \ldots v_p$, then there are weights $w_k \geq 0$ with $\sum w_k = 1$ such that

$$\varepsilon_f = \left(\sum_{k=1}^{p} w_k (v_k - \mu)^2 \right)^{1/2}$$

Argonne NATIONAL LABORATORY

The Noise Level ε_f and Uncertainty

Chebyshev's inequality. If μ is the expected value of $f(\boldsymbol{x})$ then

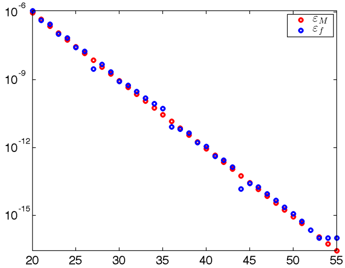
$$|f(\boldsymbol{x}) - \mu| \leq \gamma \varepsilon_f$$

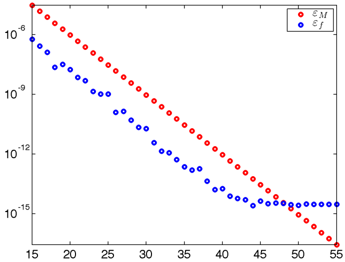
is likely to hold for $\gamma \geq 1$ of modest size.

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

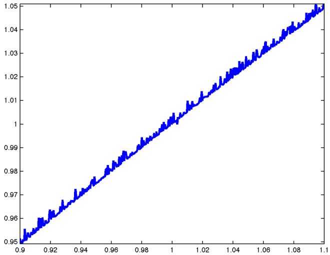


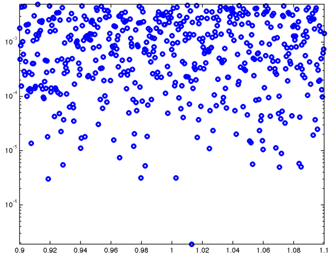


$x \mapsto \text{chop}(x,t)^2$ (left) and $x \mapsto \text{ept}[\text{chop}(x,t)]$ (right)

chop(x,t) truncates x to t bits; ept(x) evaluates an n-dimensional quadratic function

Newton's Method





Newton's method for $t \mapsto t^{1/2}$ with tolerance $\tau = 10^{-2}$

Computed function f (left) and noise (right)

$$\varepsilon_f \sim 7 \cdot 10^{-4}$$

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 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

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							

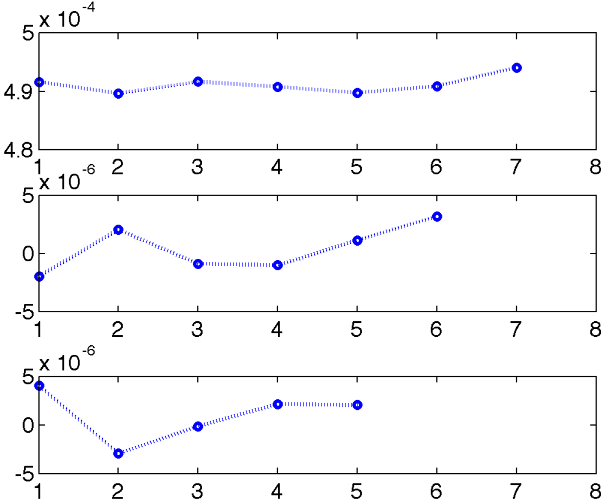
◇ Estimate the noise level from

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

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

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

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



The Simplest Simulations: Krylov Solvers

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

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

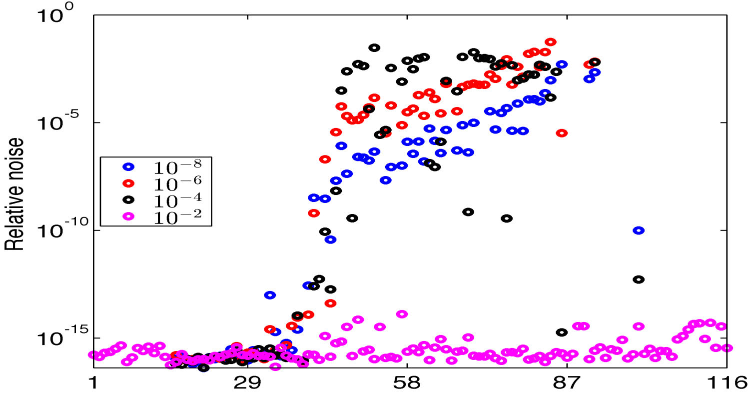
where b is a function of the input x. We use $b = x$.

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

- ◇ UF symmetric positive definite matrices (116) with $n \le 10^4$
- ◇ 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: $\tau \in [10^{-8}, 10^{-1}]$

Argonne NATIONAL LABORATORY

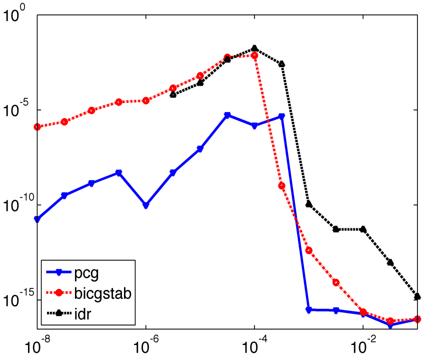
What is the Noise Level of Krylov Simulations?



Distribution of ε_f for f_τ (bicgstab)

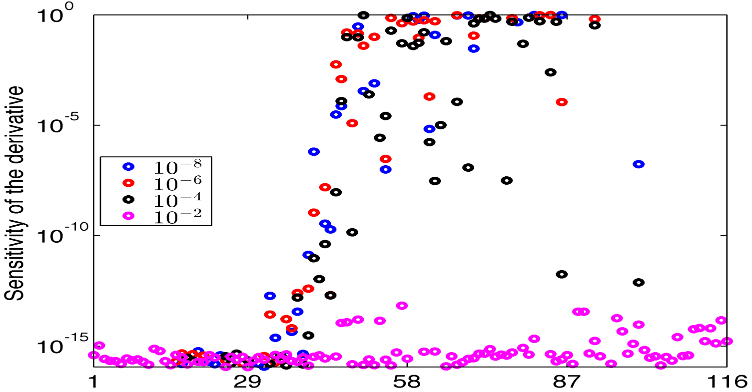
Argonne NATIONAL LABORATORY

New Phenomena: Noise Level Transitions



ε_f as a function of tolerance τ

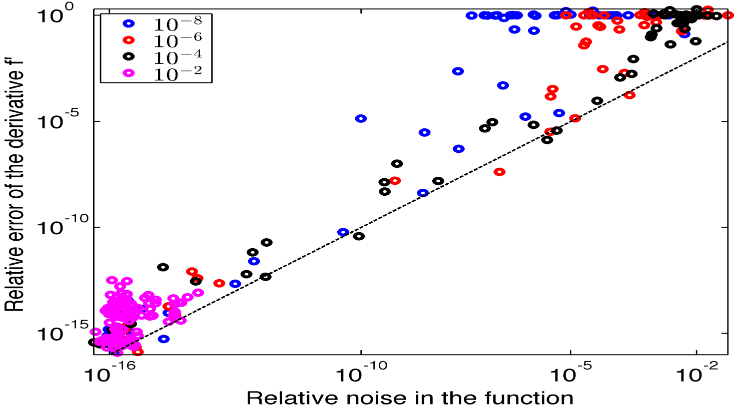
Can You Trust Derivatives?



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

Argonne NATIONAL LABORATORY

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



Distribution of $(\varepsilon_f, \mathrm{re}(f'_\tau))$ for f_τ (bicgstab)
Dashed line is (t, t)

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