West Coast Optimization Meeting
Seattle, Washington, May 13-15, 2016

Computational Noise: Uncertainty and Sensitivity

Jorge Moré

Joint work with Stefan Wild

Mathematics and Computer Science Division
Argonne National Laboratory

Connections

- ◇ PhD, University of Illinois 1983
- ◇ Summer faculty position at Argonne, 1985.
- ◇ University of Washington, 1985 –

Publications

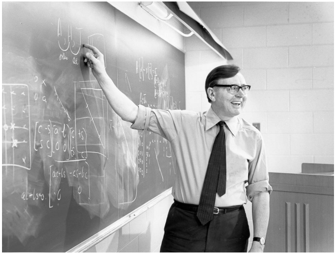
- ◇ On the identification of active constraints, 1988
- ◇ Convergence properties of trust region methods . . . 1990
- ◇ Exposing contraints, 1994

Computational noise → High Precision → J. Borwein → Jim Burke

High-precision computation: Mathematical physics and dynamics, Bailey, Barrio, Borwein, Applied Mathematics and Computation, 2012.

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 finite precision

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

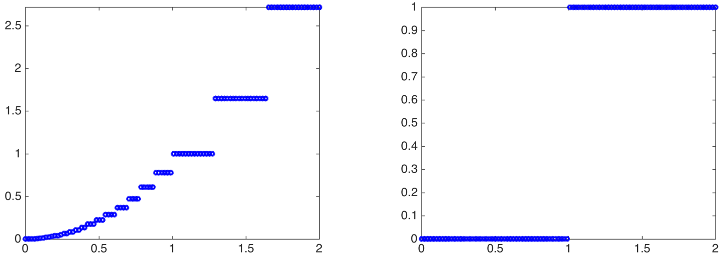
For backward-stable computations,

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

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 $L = 50$ (left) and $L \geq 55$ (right)

N. Higham, Accuracy and Stability of Numerical Algorithms, 2002

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

Assessments of Roundoff

Repeat the computation but ...

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

Monte Carlo arithmetic: How to gamble with floating point and win,
D. S. Parker, B. Pierce and P. R. Eggert, Computing in Science Engineering, 2000

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.

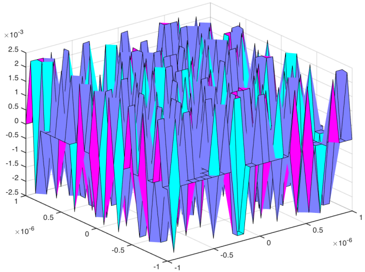
⚠ If the computed f is backward-stable, then the uncertainty is

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

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

J. Moré and S. Wild, Estimating computational noise, SIAM Journal on Scientific Computing, 2011.

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?

Computational Noise $\sim$ Uncertainty

Definition. The *noise level* of 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_0 and all permissible perturbations of x_0

Leading causes of noise

- 10^X flops
- Iterative calculations
- Adaptive algorithms
- Mixed precision

Two Theorems

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

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

where μ is the expected value of $f(\boldsymbol{x})$.

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

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

where μ is the (weighted) average of the function values.

The function $x \mapsto f\left[\mathsf{chop}(x, t)\right]$ is a step function.

The Noise Level ε_f and Uncertainty

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

Chebyshev inequality

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

Cauchy-Schwartz inequality

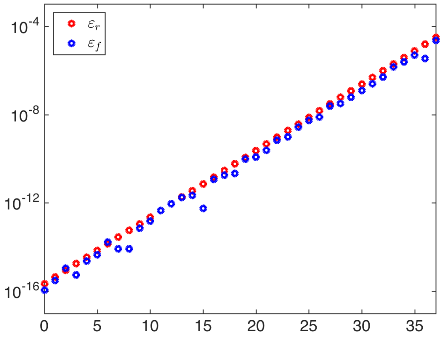
$$\mathrm{E}\left\{|f(\boldsymbol{x}) - \mu|\right\} \le \varepsilon_f$$

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

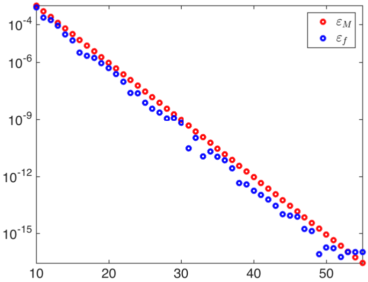
Case Study: The Higham Function

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



$$\varepsilon_r(k) = 2^{(-k+52)}$$

Case Study: Mixed Precision Quadratics



$x \mapsto \|\text{chop}(x,t)\|^2$ where $\text{chop}(x,t)$ truncates $x \in \mathbb{R}^n$ to t bits

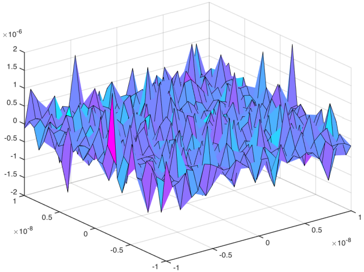
$n = 4$ (left) and $n = 10^4$ (right)

Argonne NATIONAL LABORATORY

Case Study: Eigenvalue Solvers

$$f(x) = \sum_{i=1}^{p} \lambda_i \Big(A + \text{diag}\,(x) \Big), \quad p = 5$$

where $\lambda_i(\cdot)$ is the i-th smallest eigenvalue in magnitude.



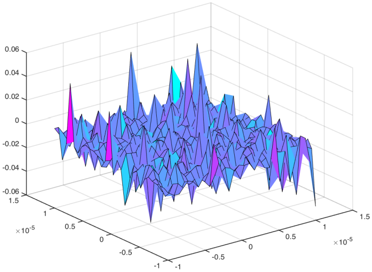
$\varepsilon_f = 3.5 \; 10^{-7}$

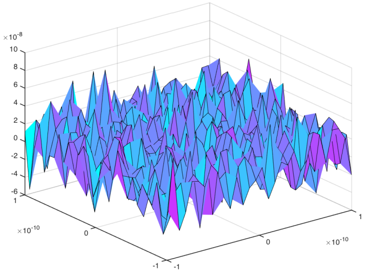
A is the sparse Laplacian on an L-shaped region

eigs with $\tau = 10^{-3}$.

Case Study: Linear (Krylov) Solvers

$$f(x) = \|A^{-1}x\|^2$$





bicgstab, $\varepsilon_f = 7.8\ 10^{-3}$

pcg , $\varepsilon_f = 1.6\ 10^{-8}$

Sparse matrix A from the University of Florida collection, $\tau = 10^{-3}$

index = 38 (left) , index = 35 (right)

Analysis

A realistic model of a finite-precision function is defined by

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

where $f_s : \mathbb{R} \mapsto \mathbb{R}$ is smooth ($f_s = f_\infty$ is an option)

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

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

- Construct the k-th order differences of f

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

- Note that ε_f can be determined from $\Delta^k \varepsilon(t)$

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

- Estimate the noise level of f from

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

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

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

ECnoise: Difference Tables

1.56e+03	-6.92e+00	1.32e+00	3.40e+01	-1.21e+02	2.93e+02	-6.12e+02
1.56e+03	-5.61e+00	3.53e+01	-8.70e+01	1.72e+02	-3.19e+02	
1.55e+03	2.97e+01	-5.17e+01	8.53e+01	-1.46e+02		
1.58e+03	-2.20e+01	3.36e+01	-6.11e+01			
1.56e+03	1.16e+01	-2.75e+01				
1.57e+03	-1.59e+01					
1.56e+03						
Noise levels						
	1.24e+01	1.39e+01	1.57e+01	1.77e+01	1.93e+01	2.01e+01

bicgstab, index = 38, $\varepsilon_f = 7.8\ 10^{-3}$

5.29e+02	7.53e-06	-5.74e-06	-8.38e-06	3.41e-05	-4.74e-05	-7.65e-06
5.29e+02	1.80e-06	-1.41e-05	2.58e-05	-1.33e-05	-5.51e-05	
5.29e+02	-1.23e-05	1.16e-05	1.25e-05	-6.84e-05		
5.29e+02	-6.82e-07	2.41e-05	-5.59e-05			
5.29e+02	2.34e-05	-3.18e-05				
5.29e+02	-8.38e-06					
5.29e+02						
Noise levels						
	8.32e-06	8.08e-06	7.08e-06	5.35e-06	3.24e-06	2.52e-07

pcg, index = 35, $\varepsilon_f = 1.6\ 10^{-8}$

Argonne NATIONAL LABORATORY

Krylov Solvers: Distribution of $\varepsilon_f(\tau)$

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

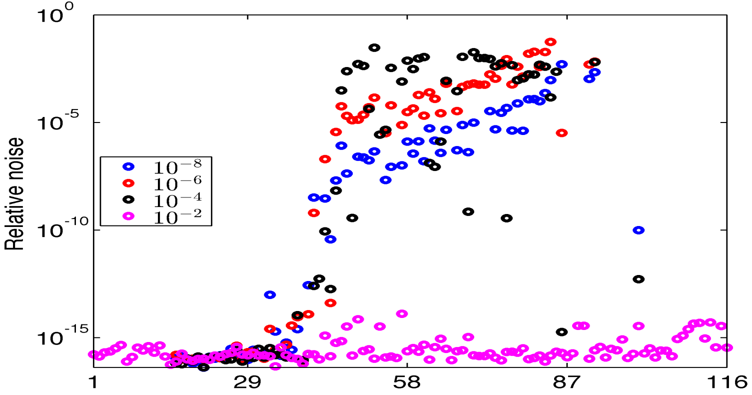
$$f_\tau(x) = \|y_\tau(x)\|^2, \qquad \|Ay_\tau(x) - b\| \le \tau\|b\|,$$

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: $A \leftarrow D^{-1/2} A D^{-1/2}$, $D = \text{diag}(a_{i,i})$
- ◇ Solvers: bicgstab (similar results for pcg, minres, gmres, . . .)
- ◇ Tolerances: $\tau \in [10^{-8}, 10^{-1}]$

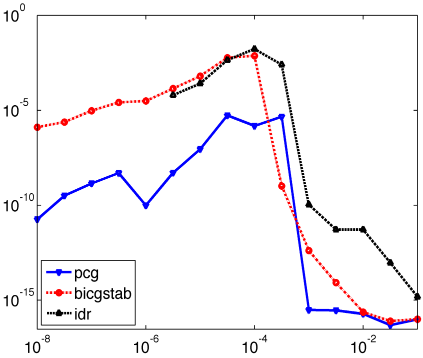
What is the Noise Level of Krylov Solvers?



Distribution of ε_f for f_τ (bicgstab)

Argonne NATIONAL LABORATORY

Noise Level Transitions



ε_f as a function of tolerance τ

Impact of Noise on Derivatives

We measure the uncertainty of f' with

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

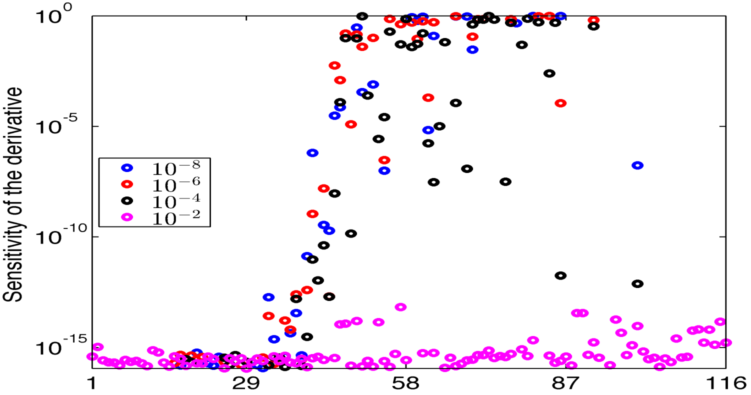
- ◇ With infinite precision $\text{re}(f') = \varepsilon/(1+\varepsilon)$ for any $\varepsilon > 0$
- ◇ We expect $\text{re}(f')$ to be small in floating point arithmetic.

Two AD algorithms (forward mode) were used to compute f':

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

IntLab was used in the numerical results.

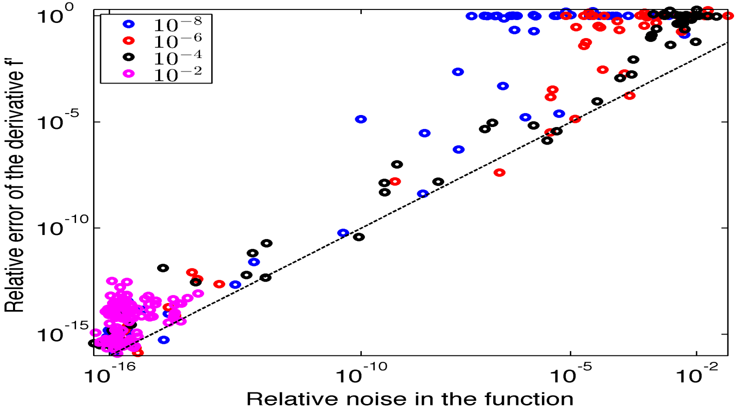
Can You Trust Derivatives?



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

Argonne NATIONAL LABORATORY

$\text{re}(f'_\tau) \gg \varepsilon_f$



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

Optimal Difference Estimate of the Derivative

Define the RMS expected error in the derivative by

$$\mathcal{E}(h) = \mathrm{E}\left\{ \left(\frac{f(t_0+h) - f(t_0)}{h} - f_s'(t_0) \right)^2 \right\}^{1/2}.$$

Theorem. If $\mu_L = \min |f''|$, $\mu_M = \max |f''|$, and

$$h^* = \gamma_2 \left(\frac{\varepsilon_f}{\mu} \right)^{1/2}, \qquad \gamma_2 = 8^{1/4} \approx 1.68,$$

where μ is an estimate of $|f''(t_0)|$ in $[\mu_L, \mu_M]$, then

$$\mathcal{E}(h^*) \le (\gamma_1 \mu_M \, \varepsilon_f)^{1/2} \le \left(\frac{\mu_M}{\mu_L} \right)^{1/2} \min_{0 < h \le h_0} \mathcal{E}(h)$$

Optimal Difference Parameter

The estimate μ of $|f''(t_0)|$ is obtained from

$$\mu = \frac{f(t_0 - h) - 2f(t_0) + f(t_0 + h)}{h^2}$$

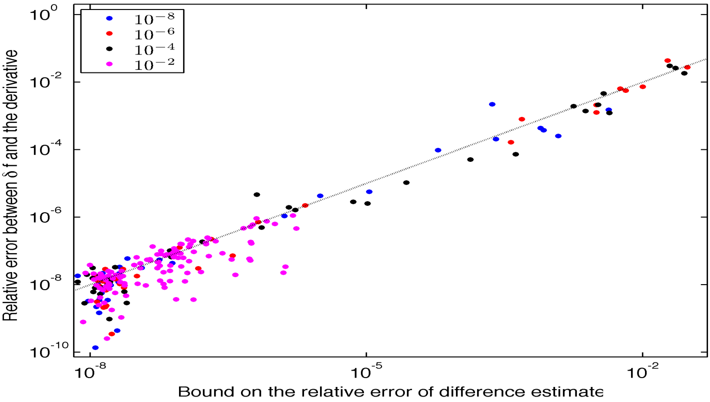
where h is of order $\varepsilon_f^{1/4}$. Set

$$\delta f(t_0) = \frac{f(t_0 + h^*) - f(t_0)}{h^*}$$

Claim: If $\text{re}(\delta f, f')$ is the relative error between δf and f', then

$$\text{re}(\delta f, f') \sim \text{re}(\delta f) \equiv \frac{\mathcal{E}(h^*)}{|f'(t_0)|}$$

re(δf) $\sim$ re($\delta f, f'$)



Distribution of (re(δf_τ), re($\delta f_\tau, f'_\tau$)) 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), 1292-1314.
- J. Moré and S. Wild, *Estimating Derivatives of Noisy Simulations*, ACM Trans. Mathematical Software, 38 (2012), 19:1–19:21
- J. Moré and S. Wild, *Do You Trust Derivatives or Differences?*, Journal of Computational Physics, 273 (2014), 268 – 277.