Optimal Derivatives of Noisy Simulations*

Jorge Moré and Stefan Wild

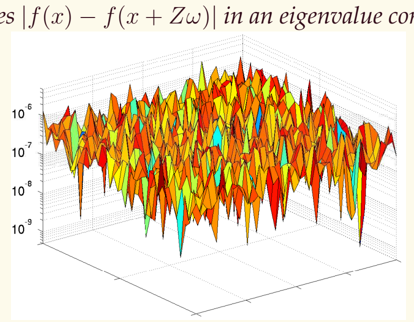
Mathematics and Computer Science Division, Argonne National Laboratory

*Research supported by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

Computational Noise in Simulations

Finite precision + finite processes lead to noise in simulations throughout computational science & engineering.

Differences $|f(x) - f(x + Z\omega)|$ in an eigenvalue computation



Computational noise can result from

- Iteratively solving systems of PDEs
- Adaptively computing integrals
- Discretizations/meshes, estimating eigenvalues

…and other fundamental calculations.

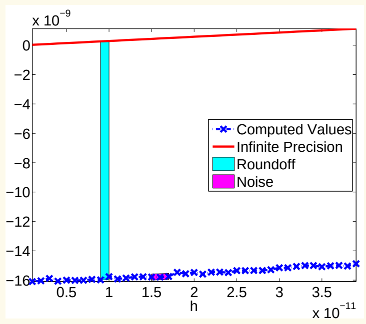
Computational noise is not:

◇ truncation error

$R_{n+1}(x) = f(x) - \sum_{i=0}^{n} P_i(x)$

◇ roundoff error

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



The Noise Level ε_f and Differences

Univariate stochastic model, $f(t) = f_s(t) + \varepsilon(t)$, $t \in \mathcal{I}$,

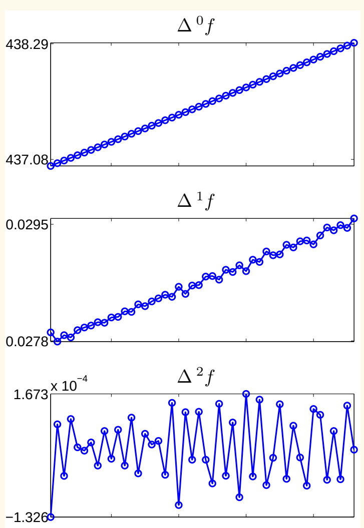
f_s smooth, deterministic

ε iid on $\mathcal{I}$. →no distribution assumptions←

Definition. The noise level of f is $\varepsilon_f = (\mathrm{Var}\{\varepsilon(t)\})^{1/2}$.

If f_s is k-times differentiable,

$$\begin{aligned} \Delta^k f(t) &= \Delta^{k-1} f(t+h) - \Delta^{k-1} f(t) \\ &= \mathbf{f_s^{(k)}}(\xi_k)\mathbf{h^k} + \boldsymbol{\Delta^k \varepsilon(t)}, \qquad \xi_k \in (t, t+kh). \end{aligned}$$



To quantify noise: Make h small enough and k large enough to remove the smooth component.

Theorem [4]:
If $\{\varepsilon(t + ih) : i = 0, \ldots, m\}$ iid

1. $\gamma_k \mathrm{E}\left\{[\Delta^k \varepsilon(t)]^2\right\} = \varepsilon_f^2$, $\qquad \gamma_k = \frac{(k!)^2}{(2k)!}$.
2. If f_s is continuous at t, $\lim_{h \to 0} \gamma_k \mathrm{E}\left\{\left[\Delta^k f(t)\right]^2\right\} = \varepsilon_f^2$.

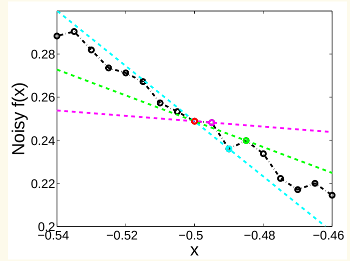
`ECNoise` [4] uses $\sigma_k = \left(\overline{\gamma_k[\Delta^k f(t_i)]^2}\right)^{1/2}$ to estimate ε_f

- Devices for choosing k and verifying h is small enough
- Empirically validated for deterministic f
- Requires only a few (6-8) f evaluations.

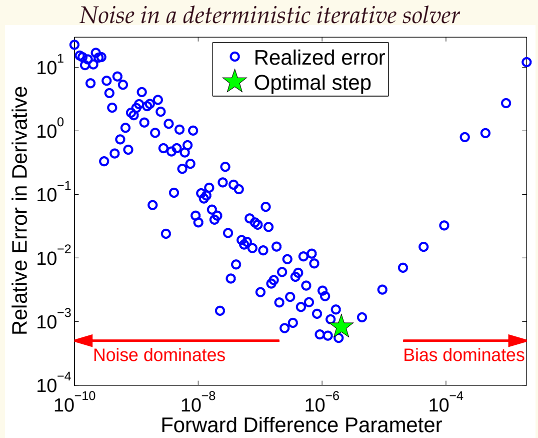
Noisy Forward Differences

A practical problem with differences: How to choose h?

$$f'(t_0) \approx \frac{f(t_0 + h) - f(t_0)}{h}$$



Noise in a deterministic iterative solver



Optimal Forward Difference Step Size

Assume

- $f(t) = f_s(t) + \epsilon$ on $\mathcal{I} = \{t_0 + h : 0 \le h \le h_0\}$
- f_s twice differentiable, $\mu_L \le |f_s''| \le \mu_M$ on $\mathcal{I}$

and minimize the mean-squared-error

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

Theorem [5]:

$$\frac{1}{4}\mu_L^2 h^2 + 2\frac{\varepsilon_f^2}{h^2} \le \min_{0 \le h \le h_0} \mathrm{E}\{\mathcal{E}(h)\} \le \frac{1}{4}\mu_M^2 h^2 + 2\frac{\varepsilon_f^2}{h^2}$$

For h_0 sufficiently large:

1. Upper bound minimized by $h_M = 8^{1/4}\left(\frac{\varepsilon_f}{\mu_M}\right)^{1/2}$.
2. When $\mu_L > 0$, h_M is near-optimal:

$$\mathrm{E}\{\mathcal{E}(h_M)\} = \sqrt{2}\mu_M \varepsilon_f \le \left(\frac{\mu_M}{\mu_L}\right) \min_{0 \le h \le h_0} \mathrm{E}\{\mathcal{E}(h)\}.$$

Other Forward Difference Step Sizes

1. Square-root of machine precision: $h = \sqrt{\epsilon_{\mathrm{mach}}}$.
2. Based on a uniform bound on roundoff error, $|f(t) - f_\infty(t)| \le \varepsilon_A$, [2] minimizes an upper bound on l_1 error to obtain $h_A = 2\left(\frac{\varepsilon_A}{\mu_M}\right)^{1/2}$.

→ Requires estimate of ε_A and $h_A \le h_0$.

Ex.- $f(t) = (x_0 + tp)^T A_n^{-2}(x_0 + tp)$, with

- A_n = (ill-conditioned) $n \times n$ Hilbert matrix
- A_n^{-1} computed by Gaussian elimination.

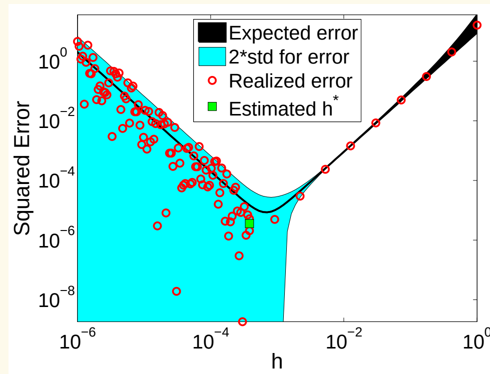
n	h_M	h_A	est(h_M)	est(h_A)	est($\sqrt{\epsilon_{\mathrm{mach}}}$)
5	2.4e-06	6.9e-06	20.4350	20.4350	20.4357
6	4.2e-06	3.4e-05	28.8913	28.8915	28.8940
7	1.2e-05	1.8e-04	37.2351	37.2362	37.2558
8	4.0e-05	1.0e-03	43.9805	43.9869	44.2150
9	9.3e-05	5.3e-03	56.1004	56.1427	51.6127
10	2.0e-04	2.9e-02	67.1231	67.3678	60.1290
11	7.8e-04	1.6e-01	80.3738	81.8401	279.6018

Roundoff errors are systematic

⇒ obtain more digits of f_∞' using h_M.

Stochastic Results

Estimate $f_s'(t) = E\{t^3 + 10^{-6}U_{[-2\sqrt{3}, 2\sqrt{3}]}\}'$ at $t = 1$.



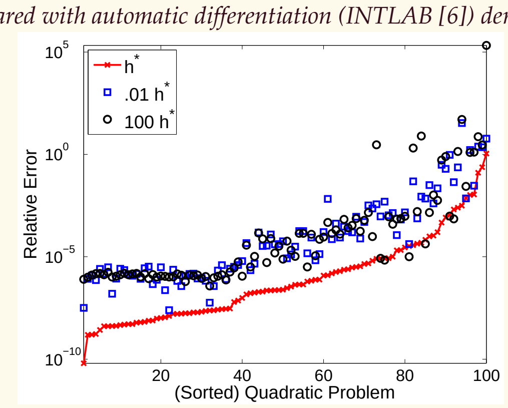
- Expected error & uncertainty region predicted by theory
- h_m falls near minimum of $\mathrm{E}\{\mathcal{E}(h)\}$.

Deterministic Results

Ex.- Quadratic Functions and BiCGSTAB

$\phi(t) = \|y_\tau(t)\|^2$, where $Ay_\tau = x_0 + tp$ is solved with BiCGSTAB with tolerance $\tau = 10^{-3}$ and A is a UF matrix [1].

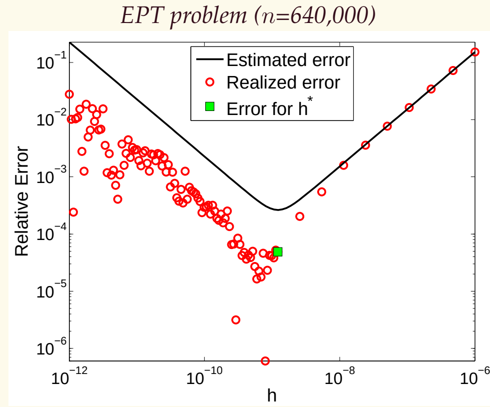
Compared with automatic differentiation (INTLAB [6]) derivative



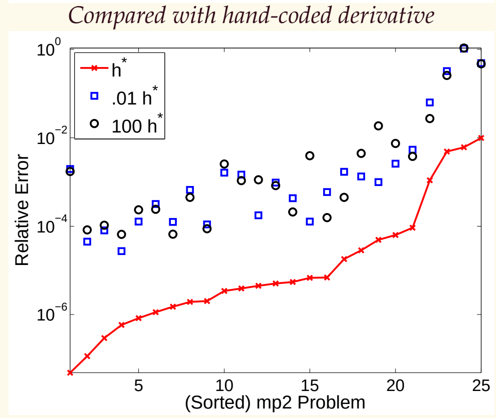
- Exhibits behavior similar to stochastic differences
- h_M obtains 2 more correct digits than $10^{\pm 2} h_M$.

Ex.- Highly Nonlinear `MINPACK-2` Problems from [3]

EPT problem (n=640,000)



Compared with hand-coded derivative



- Accurate estimates obtained even when f'' not constant
- Relatively insensitive to misestimation of $|f''|$ and ε_f.

Extension: Central Differences

First derivatives

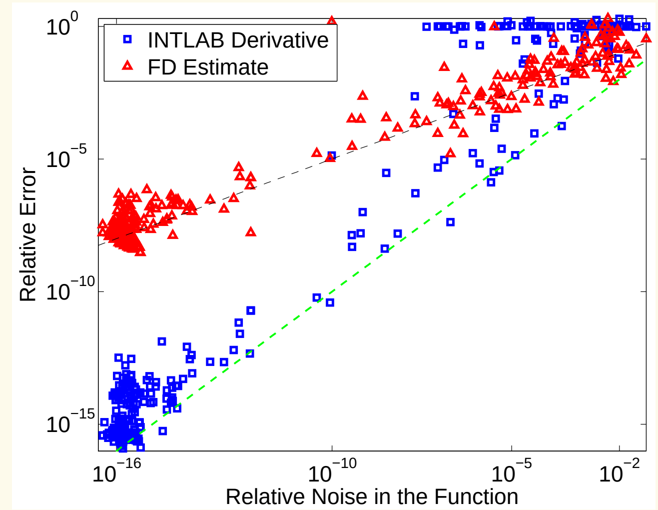
- $|h_M| = \gamma_5 \left(\frac{\varepsilon_f}{\mu_M}\right)^{1/3}$, $\qquad \gamma_5 = 3^{1/3}$, $\quad |f_s^{(3)}| \in [\mu_L, \mu_M]$
- $\mathrm{E}\{\mathcal{E}_c(h_M)\} \le \left(\frac{\mu_M}{\mu_L}\right)^{2/3} \min_{|h| \le h_0} \mathrm{E}\{\mathcal{E}_c(h)\}$.

Second derivatives

- $|h_M| = \gamma_7 \left(\frac{\varepsilon_f}{\mu_M}\right)^{1/4}$, $\qquad \gamma_7 = 2^{5/8}\, 3^{1/8}$, $\quad |f_s^{(4)}| \in [\mu_L, \mu_M]$
- $\mathrm{E}\{\mathcal{E}_2(h_M)\} \le \left(\frac{\mu_M}{\mu_L}\right) \min_{|h| \le h_0} \mathrm{E}\{\mathcal{E}_2(h)\}$
- gives rough estimate of $|f_s''|$ for forward difference h.

Differences vs. Derivatives

BiCGSTAB quadratics $\phi(t) = \|y_\tau(t)\|^2$ with more tolerances τ.



- Derivatives are noisier than functions
- Different tolerances yield very different noise levels
- Computed derivative can be much noisier than finite difference estimate, especially when f is noisy.

Summary

A few extra function evaluations can give better derivatives!

- Computational noise complicates analysis of simulation-based functions
- Stochastic theory yields near-optimal step sizes
- Requires only coarse estimates of noise and $|f''|$
- Works on deterministic functions in practice
- More robust than h_A & $\sqrt{\epsilon_{\mathrm{mach}}}$
- Extends to higher-order difference schemes and derivatives
- Finite differences can produce reliable derivatives for very noisy functions.

References

[1] T.A. Davis and Y. Hu. The University of Florida Sparse Matrix Collection. *ACM Trans. on Mathematical Software*, 38(1), 2011. To appear.

[2] P.E. Gill, W. Murray, M.A. Saunders, and M.H. Wright. Computing forward-difference intervals for numerical optimization. *SIAM J. Scientific and Statistical Computing*, 4:310–321, 1983.

[3] C.-J. Lin and J.J. Moré. Newton's method for large bound-constrained optimization problems. *SIAM J. Optimization*, 9:1100–1127, 1999.

[4] J.J. Moré and S.M. Wild. Estimating computational noise. *SIAM J. Scientific Computing*, 33(3):1292–1314, 2011.

[5] J.J. Moré and S.M. Wild. Estimating derivatives of noisy simulations. *ACM Trans. on Mathematical Software*, 38(3), 2011. To appear.

[6] S.M. Rump. INTLAB - INTerval LABoratory. In Tibor Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Dordrecht, 1999.