Argonne NATIONAL LABORATORY

Doing Moore with Less – Inexactness for Supercomputing

3rd International Workshop on Post Moore's Era Supercomputing

Sven Leyffer, Stefan Wild, Mike Fagan, Marc Snir,
Krishna Palem, Kazutomo Yoshii, and Hal Finkel

Argonne National Laboratory, Rice University, and University of Illinois at Urbana-Champaign

November, 11 2018

Outline

Newton's Method: Workhorse for DOE Applications

Discretization of PDEs leads to large, sparse, nonlinear systems $F(x) = 0$

Algorithm 1: Newton's Method

Given x^0, choose tol $\epsilon > 0$, and set $k = 0$

repeat

Linear Solve: Bulk of Computational Work:

$$\text{Solve} \quad \nabla F^k s = -F^k, \quad \text{where } F^k = F(x^k) \quad \ldots \text{ for } \quad s = s^k$$

Line-search for α along s^k to reduce $\|F(x^k + \alpha s^k)\|$

Update $x^{k+1} = x^k + \alpha s^k$, and set $k = k + 1$

until $\|F(x^k)\| \leq \epsilon$

Goal

Exploit inexactness and precision hierarchy in solution

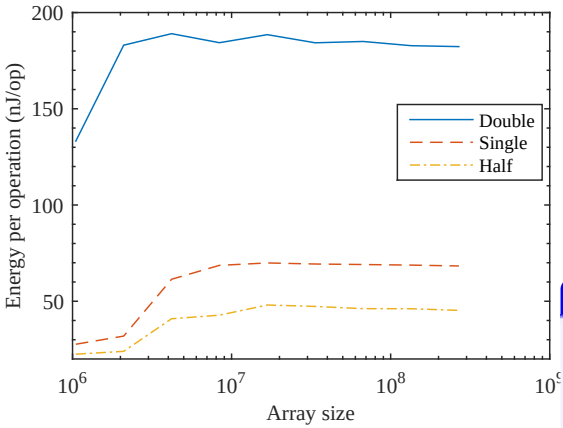
Newton's Method: Workhorse for DOE Applications

- Generic for many DOE apps: simulation, design, control
- Linear systems $\nabla F^k s = -F^k$ large and sparse
  $\Rightarrow$ apply preconditioned (iterative) Krylov method
- Exploitation of inexact Krylov solvers well understood:
  (Dembo, Eisenstat, and Steihaug, 1982)

Goal: Exploit Precision Hierarchy in Solution

- Develop strategies for precision control
- Perform hardware measurements to investigate effect
- Build models to predict effect on different hardware
- Explore power/precision trade-offs on commercial off-the-shelf (COTS) processors

Motivation: Energy per Operation for Fused-Multiply-Add



Algorithm 2: Fused-Multiply-Add

for $i = 1,\ i < N,\ i+= $ *NELTSV256*
do
  Load three 256-bit vectors: a, b, c
  Fused Multiply-Add: $a \leftarrow a \odot b + c$
  Store
end

Observations

- Double → Single: Factor 2.67
  - Larger savings: fewer cache misses
- Single → Half: Factor 1.49
  - Fewer savings: no HW support

Outline

Inexact Newton's Method

Nonlinear system of equations (e.g. discretized PDE) $F(x) = 0$

Algorithm 3: Inexact Newton's Method

Given x^0, choose tol $\epsilon > 0$, and set $k = 0$

repeat

Evaluate $F^k = F(x^k)$ and $\nabla F^k = F(x^k)$

Approximately solve the following system for s

$$\nabla F^k s = -F^k \quad \text{such that} \quad \frac{\|\nabla F(x^k)s + F(x^k)\|}{\|F(x^k)\|} \leq \eta_k,$$

... where $0 \leq \eta_k < 1$ is a sequence of tolerances $\eta_k \searrow 0$

Line-search for α along s^k to reduce $\|F(x^k + \alpha s^k)\|$

Update $x^{k+1} = x^k + \alpha s^k$, and set $k = k + 1$

until $\|F(x^k)\| \leq \epsilon$

Inner iterations performed using BI-CGSTAB ...

Inner Iterations: BI-CGSTAB

Algorithm 4: *BI-CGSTAB: Approx. solve* $\nabla F(x^k)s = -F(x^k)$

Set $r^0 = -F(x^k)$; $q^0 = r^0$, $s^0 = v^0 = p^0 = 0$; $i = 0$

while $\|r^i\| > \eta_k \|F(x^k)\|$ **do**

- $i \leftarrow i + 1$ and compute $\rho_i = \mathrm{VecVec}(q^0, r^{i-1})$
- **if** $\rho_i = 0$ **then** BI-CGSTAB method fails
- $\beta_i = \frac{\rho_i}{\rho_{i-1}} \frac{\alpha_{i-1}}{\omega_{i-1}}$
- $p^i = r^{i-1} + \beta_i (p^{i-1} - \omega_{i-1} v^{i-1})$
- $v^i = \mathrm{MatVec}(\nabla F(x^k), p^i)$
- $\alpha_i = \frac{\rho_i}{\mathrm{VecVec}(q^0, v^i)}$
- $u^i = r^{i-1} - \alpha_i v^i$ **if** $\|u^i\| = 0$ **then** $s^i = s^{i-1} + \alpha_i p^i$; **exit**
- $t^i = \mathrm{MatVec}(\nabla F(x^k), u^i)$
- $\omega_i = \frac{\mathrm{VecVec}(t^i, u^i)}{\mathrm{VecVec}(t^i, t^i)}$
- $s^i = s^{i-1} + \alpha_i p^i + \omega_i u^i$
- $r^i = u^i - \omega_i t^i$; compute $\|r^i\|$

end

MatVec$(\cdot, \cdot)$ and VecVec$(\cdot, \cdot)$ are fused-add-multiplies

Generic Test Problems

1D Laplace Problem

$$
\begin{aligned}
F_1(x) &= b_1 + 4x_1 - x_2 \\
F_i(x) &= b_i - x_{i-1} + 4x_i - x_{i+1}, \qquad i = 2, \ldots, n-1 \\
F_n(x) &= b_n - x_{n-1} + 4x_n
\end{aligned}
$$

where $b_1 = 1.0$, $b_i = -2.0$ for $i = 2, \ldots, n-1$, and $b_n = 4.0$ **well conditioned**

Chained Rosenbrock Problem

$$
\begin{aligned}
F_1(x) &= 2a(x_1 - 1) - 400x_i \left(x_2 - x_1^2\right) \\
F_i(x) &= 200 \left(x_i - x_{i-1}^2\right) + 2a(x_i - 1) - 400x_i \left(x_{i+1} - x_i^2\right), \qquad i = 2, \ldots, n-1 \\
F_n(x) &= 200 \left(x_n - x_{n-1}^2\right)
\end{aligned}
$$

where $a = 1$ **badly conditioned**

Convergence and Conditioning: Laplace, $N = 10^6$

OptTol	Double Iters	Single Iters	Half Iters	Remarks
1E-07	9 14			
1E-06	8 10	8 10		
1E-04	7 8	7 8	12 18	LS error
1E-02	6 6	6 6	6 6	

Observations

- Tridiagonal Jacobian gives good data locality
- Comparable outer/inner iterations for moderate OptTol
- Low precision gives line-search errors (LS error)

Convergence and Conditioning: Rosenbrock, $N = 10^5$

OptTol	Double Iters	Single Iters	Half Iters
1E-07	32 81	LS-error	NaN
1E-05	31 76	36 103	NaN
1E-03	30 72	29 69	NaN

RelErr	Double Iters	Single Iters	Hal's Iters
1E-07	29 68	28 65	LS-error
1E-05	6 6	6 6	6 6
1E-03	5 5	5 5	5 5

Half precision: overflow in residual (norm) computation & ill-conditioning!

Half Precision for Rosenbrock

Use non-standard half precision

Hal's C++ precision classes

	exponent	mantissa
IEEE	5	10
half	7	8

Convergence and Conditioning: Rosenbrock, $N = 10^5$

OptTol	Double Iters	Single Iters	Half Iters
1E-07	32 81	LS-error	NaN
1E-05	31 76	36 103	NaN
1E-03	30 72	29 69	NaN

RelErr	Double Iters	Single Iters	Hal's Iters
1E-07	29 68	28 65	LS-error
1E-05	6 6	6 6	6 6
1E-03	5 5	5 5	5 5

Half precision: overflow in residual (norm) computation & ill-conditioning!

Half Precision for Rosenbrock

Use non-standard half precision

Hal's C++ precision classes

	exponent	mantissa
IEEE	5	10
half	7	8

Beyond Simplistic Use of Lower Precision

Explore alternative (adaptive) use of low precision

Outline

Improving Solutions by Reinvesting Energy

So far, compared half, single, double costs ... use energy savings to improve solution

Reinvestment of Energy

Baseline Run inexact Newton in double precision to OptTol

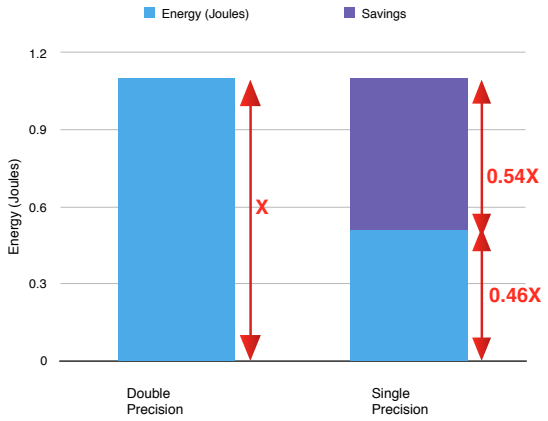
Reinvestment Run inexact Newton in single-then-double to OptTol

... measure improvement factor in energy savings!

Experimental set-up

1. DELL T1700 with Intel core i7 4770 and 1GB DDR3 RAM
2. Measure energy using RAPL hardware counter
3. SIMD (vector) arithmetic in single/double precision

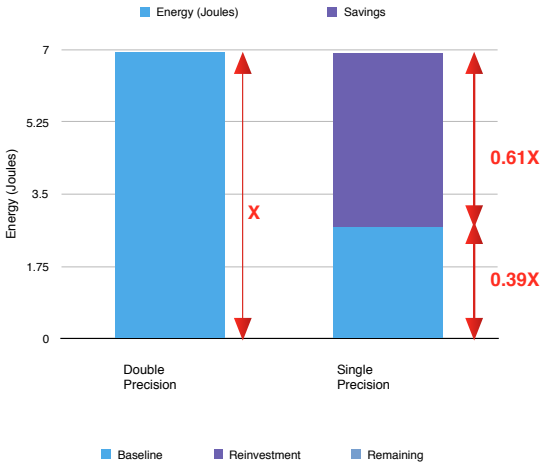
Reinvestment Gain for Laplace: $\epsilon = 10^{-6}$



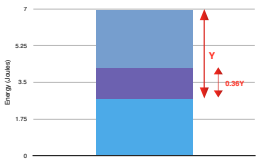
Class	Joules
Double	1.10
Reinvest	0.84
Single	0.51

Reinvestment only uses fraction of saved energy!

Reinvestment Gain for Rosenbrock: $\epsilon = 10^{-2}$



Class	Joules
Double	6.92
Reinvest	4.20
Single	2.69



Reinvestment only uses fraction of saved energy!

Outline

Energy Model for Linear Convergence

Assume linear convergence with rate $\lambda \Rightarrow$ error satisfies $\epsilon_{k+1} \leq \epsilon_k/\lambda$

- Two precisions, $p_1 < p_2$ for desired accuracy ϵ with $\epsilon \geq 2^{-p_2+s_2}$
  $\Rightarrow$ energy of precision p_2 to attain accuracy ϵ is

$$E_{\text{base}} = E(p_2)\frac{\log \frac{1}{\epsilon}}{\log \lambda}$$

  where $E(p)$ is energy for single Newton iteration

- Hybrid: k_1 iterations at precision p_1 and k_2 iterations at precision p_2
  $\Rightarrow$ energy of hybrid scheme is

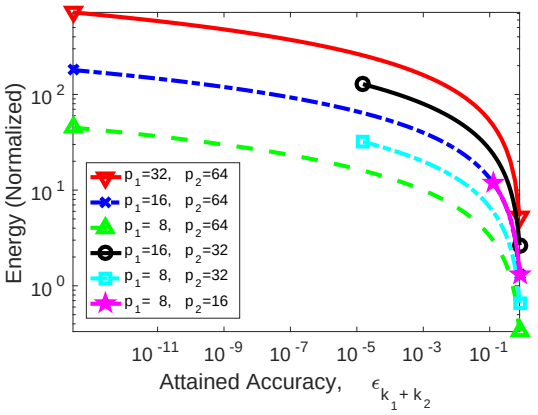
$$E_{\text{hybrid}} = E(p_2)\left(k_1\frac{E(p_1)}{E(p_2)} + k_2\right).$$

To get $E_{\text{hybrid}} \leq E_{\text{base}}$ we need

$$k_2 \leq \frac{\log \frac{1}{\epsilon}}{\log \lambda} - k_1\frac{E(p_1)}{E(p_2)}. \Leftrightarrow \epsilon_{\text{hybrid}} \leq \lambda^{-k_1-k_2} \leq \lambda^{-k_1\left(1-\frac{E(p_1)}{E(p_2)}\right)-\frac{\log \frac{1}{\epsilon}}{\log \lambda}}.$$

$\Rightarrow$ choose number of iterations, k_1, at lower precision as large as possible (obviously)

Normalized Energy for Linear Convergence



Optimal value $k_1 = \frac{1}{\log \lambda} \min \left\{ p_1 - s_1, \log \frac{1}{\epsilon} \right\}$ and corresponding

$$k_2 = \frac{1}{\log \lambda} \min \left\{ p_2 - s_2, \max \left\{ 0, \log \frac{1}{\epsilon} - k_1 \log \lambda \right\} \right\}$$

Conclusions and Outlook

Exploiting energy savings from adaptive precision

- Newton-Krylov solver ... workhorse of DOE applications
- Greater energy savings due to cache effects
- Inexact Newton could benefit from flexible half-precision
- Reinvestment gives double precision results at lower cost
- Reinvestment model points to (obvious) optimal strategy

Outlook

- Energy management critical for sustained growth in HPC
- Future likely to be heterogeneous: CPU/GPU/quantum
  ... DOE investing in quantum & neuromorphic computing
- *I don't see a way to replicate Moore's law with inexactness ...*
  ... need quantum, neuromorphic, ... combined effect