Newton's Method in Mixed Precision

C. T. Kelley
NC State University
tim_kelley@ncsu.edu
Supported by NSF, ARO

PolyU, November 13 2020

Outline

1. Nonlinear Equations and Backward Error
   - Newton's Method
   - Inexact function and Jacobian
2. Linear Solver Woes
   - This Talk's Problem
   - The Backward Error Bites You
   - Probabilistic Rounding Analysis
3. Example. You figure it out.
4. Codes
5. Summary

Nonlinear Equations

Objective: solve

$$F(x) = 0$$

where

$$F = (f_1, f_2, \ldots, f_N)^T.$$

Newton's method is

$$x_+ = x_c - F'(x_c)^{-1}F(x_c).$$

Jacobian:

$$(F')_{ij} = \partial f_i / \partial x_j$$

Local Convergence to distinguished root x^*

Standard assumptions for local convergence:
There is $x^* \in D$ such that

- $F(x^*) = 0$,
- $F'(x^*)$ is nonsingular, and
- $F'(x)$ is Lipschitz continuous with Lipschitz constant γ, i. e.

$$\|F'(x) - F'(y)\| \leq \gamma \|x - y\|,$$

for all $x, y \in D$.

Rules for talking about Newton's method

- x^* is the solution in SA
  which may not be the one you want
- $e = x - x^*$ is the error
- Convergence theorems in terms of change from
  - current iteration x_c to
  - next iteration x_+

Famous local convergence theorem

Assume that the standard assumptions hold, $x_c \in D$, and that

$$\|e_c\| \le \frac{1}{2\|F'(x^*)^{-1}\|\gamma}.$$

Then

$$\|F'(x^*)^{-1}\|/2 \le \|F'(x_c)^{-1}\| \le 2\|F'(x^*)^{-1}\|.$$

Moreover, if e_+ is the Newton iterate from x_c then

$$\|e_+\| \le \gamma \|F'(x^*)^{-1}\| \|e_c\|^2 \le \|e_c\|/2.$$

For the entire iteration ...

Corollary: Assume that the standard assumptions hold, $x_0 \in D$, and that

$$\|e_0\| \leq \frac{1}{2\|F'(x^*)^{-1}\|\gamma}.$$

Then the

- Newton iteration exists (i. e. $F'(x_n)$ is nonsingular for all n),
- converges to x^*, and
- the convergence is q-quadratic

$$\|e_{n+1}\| = O(\|e_n\|^2)$$

What does this mean?

In an ideal world where

- precision is infinite,
- derivatives are analytic,
- linear solvers are exact,

Newton's method works great with good initial data.
But . . .

. . . you'll be doing it wrong.

In practice, you get

$$x_+ = x_c - J_c^{-1}(F(x_c) + E_c)$$

where

- $J_c \approx F'(x_c)$ (maybe badly)
- E_c is the (usually small) error in F

A less famous theorem

Same assumptions as for Newton plus

$$\| J_c - F'(x_c) \| \le \frac{1}{4 \| F'(x^*)^{-1} \|}.$$

Then J_c is nonsingular and x_+ satisfies

$$\| e_+ \| = O\left(\| e_c \|^2 + \| J_c - F'(x_c) \| \| e_c \| + \| E_c \| \right).$$

Local Improvement Theorem

Same assumptions as for Newton and, for all n,

$$\|\mathrm{J}_n - \mathrm{F}'(\mathrm{x}_n)\| \leq \frac{1}{4\|F'(x^*)^{-1}\|}.$$

and

$$\|\mathrm{E}_n\| \leq \epsilon_F.$$

Then

$$\|\mathrm{e}_{n+1}\| = O(\|\mathrm{e}_n\|^2 + \|\mathrm{J}_n - \mathrm{F}'(\mathrm{x}_n)\|\|\mathrm{e}_n\| + \epsilon_F).$$

The theorem does not predict convergence, rather stagnation.

Examples

- $\epsilon_F = 0$, $J_n = F(x_n)$: Newton
- $\epsilon_F > 0$, floating point error: Newton in practice
- $\epsilon_F > 0$, J_n finite difference Jacobian, step h
  - Use optimal $h = \sqrt{\epsilon_F}$ and
  - $\|e_{n+1}\| = O(\|e_n\|^2 + h\|e_n\| + \epsilon_F)$
  - Same behavior as Newton until stagnation.
- $\epsilon_F > 0$, $J_n = F'(x_0)$, chord method

Example: J_n forward difference approximation

With a difference increment of h

$$\|J_n - F'(x_n)\| = O(h)$$

where the prefactor in the O term depends on

- $\kappa(F')$
- γ: Lip constant of F'

Stagnation in action: Residual histories

$$f(x) = x - \tan(x); x_0 = 4.5$$

Indistinguishable!

Analytic	Finite Difference
1.37e-01	1.37e-01
4.13e-03	4.13e-03
3.98e-06	3.98e-06
3.69e-12	5.60e-12
8.88e-16	8.88e-16
8.88e-16	8.88e-16
8.88e-16	8.88e-16

Implementation: ignore ϵ_F

Initialize x_0, $n = 0$, termination criteria
while Not happy **do**
  Evaluate $F(x_n)$; terminate?
  Evaluate $J_n \approx F'(x_n)$
  Solve $J_n s = -F(x_n)$
  $x_{n+1} = x_n + s$
end while

Genius Idea!

- Store J in reduced precision.
- Solve in reduced precision.
  - Cut $O(N^2)$ storage by factor of 2 (single)
  - Cut $O(N^3)$ work by factor of 2 (single)
- How can you lose? Why isn't this in all the books?

The case in this talk

- ϵ_F floating point double precision roundoff
- $J_c = J_N + \Delta_{be}$ where
- Δ_{be} is the backward error
- Solver is double, single, or half precision LU
  - J_N is the nominal approximation you give the linear solver $F'(x_c)$ in double or finite-difference approximation
  - The solver returns the solution of $(J_N + \Delta_{be})s = -F(x_c) - E_c$

So the less famous theorem says . . .

$$\|e_{n+1}\| = O\left(\|e_n\|^2 + (\|J_{Nn} - F'(x_n)\| + \|\Delta_{be}\|)\|e_n\| + \epsilon_F \right).$$

The Jacobian you think you have is harmless

- Analytic Jacobian: $\|J_{Nn} - F'(x_n)\| = O(\epsilon_F)$
- Difference Jacobian: $\|J_{Nn} - F'(x_n)\| = O(\epsilon_F^{1/2})$
- But what about the backward error?
- Large backward error $\rightarrow$ slow nonlinear convergence. Can we see this numerically?

What is that backward error?

Let's look at some famous linear algebra books . . .

- J. W. DEMMEL, Applied Numerical Linear Algebra, SIAM, Philadelphia, 1997.
- NICHOLAS J. HIGHAM, Accuracy and Stability of Numerical Algorithms, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1996.

and read up on this.

What your professors told you is . . .

If you're solving $Ax = b$ and the solver shows up with

$$(A + \delta A)x = b$$

then (Demmel 97) page 49 says $\|\delta A\|_1 \le 3 g_{PP} N^3 \epsilon_S \|A\|_1$, where

- g_{PP} is the growth factor and
- ϵ_S is the unit roundoff in the precision of the solver.

Growth factor? We don't need a growth factor!

- Worst case bound 2^{N-1}. Bad but completely artificial.
- (Higham 96, p 178-8) reports on a few cases where g_{PP} is a problem. But also quotes Wilkinson who said that problematic growth factors are "extremely uncommon".

So in the spirit of optimism, we will ignore g_{PP}.

What does this mean?

Suppose $g_{PP} = 1$, you are still in trouble if N is large.
$N^3 \epsilon_S = O(1)$ if

- (double): $\epsilon_S = 10^{-16}$, $N \approx 2 \times 10^5$
- (single): $\epsilon_S = 10^{-8}$, $N \approx 5 \times 10^2$
- (half): $\epsilon_S = 10^{-4}$, $N \approx 22$

FAKE NEWS!

These results are clearly silly. What's up?

Details

Page 175-177: Componentwise backward error (ignore permutation matrix)

$$|\delta A| \le 2\gamma_N |\hat{L}||\hat{U}|$$

where $\hat{L}\hat{U} = A + \delta A$ and

$$\gamma_N = \frac{N\epsilon_S}{1 - N\epsilon_S}$$

Did the N^3 go away?

Nope!
The growth factor part is

$$|\hat{U}_{ij}| \le \hat{g}_{PP} \max_{kl} |A_{kl}|$$

So

- $|\hat{L}_{ij}| \le 1$ implies (worst case) $\|\hat{L}\|_1 \le N$
- $\|\hat{U}\|_1 \le \hat{g}_{PP} N \|\mathrm{A}\|_1$ also worse case

More N^3

- Bottom line:

$$\|\Delta_{be}\|_1 \le 2N^2 \gamma_N \hat{g}_{PP} \|A\|_1.$$

- The N^3 is from

$$N^2 \gamma_N = \frac{N^3 \epsilon_S}{1 - N \epsilon_S}$$

But these estimates are the worst case.
Are we doomed?

Nope!

Why should $|\mathrm{L}|$ have an entire row or column of 1s?
In many cases $|\hat{\mathrm{L}}||\hat{\mathrm{U}}| \le C|A|$

- A symmetric
- Totally positive A (so $L_{ij} \ge 0$ and $U_{ij} \ge 0$)

So, in the perfect world where

- $|\hat{\mathrm{L}}||\hat{\mathrm{U}}| \le C|A|$ and
- $g_{PP} = O(1)$,

$$\|\mathrm{J}_N - \Delta_{be}\|_\infty = O(N\epsilon_S)?$$

Probably even better ...

- N. J. Higham and T. Mary, A new approach to probabilistic rounding error analysis, Tech. Report 2018.33, Manchester Institute for Mathematical Sciences, School of Mathematics, The University of Manchester, 2018.
- I. C. F. Ipsen and H. Zhou, Probabilistic error analysis for inner products, 2019.

Big assumption: rounding errors are independent

Some people do not believe this.

Higham-Mary results: Lots of notation

Define

$$\tilde{\gamma}(\lambda) = \exp\left(\lambda\sqrt{N}\epsilon_S + \frac{N\epsilon_S^2}{1-\epsilon_S}\right) - 1$$

$$P(\lambda) = 1 - 2\exp\left(-\frac{\lambda^2(1-\epsilon_S)^2}{2}\right)$$

and

$$Q(\lambda, N) = 1 - N(1 - P(\lambda))$$

Limiting cases

- $N\epsilon_S$ small $\rightarrow \tilde{\gamma}(\lambda) \approx \lambda\sqrt{N}\epsilon_S$
- ϵ_S small, λ large $\rightarrow P(\lambda) \approx 1$
- N large and λ large and curated $\rightarrow Q(\lambda, N^3) \approx 1$ independently of N

At last, a theorem!

Theorem:
Use Gaussian elimination for $\mathrm{A}x = b$. The the computed LU factors $\hat{\mathrm{L}}$ and $\hat{\mathrm{U}}$ satisfy

$$\mathrm{A} + \delta \mathrm{A} = \hat{\mathrm{L}}\hat{\mathrm{U}} \text{ and } |\delta \mathrm{A}| \le (3\tilde{\gamma}(\lambda) + \tilde{\gamma}(\lambda)^2)|\hat{\mathrm{L}}||\hat{\mathrm{U}}|$$

with probability at least $Q(\lambda, N^3/3 + 3N^2/2 + 7N/6)$.
Wait! What? Is this good?

Goodness of results

Remember, we get to pick λ to make things look good.

- $N\epsilon_S$ small so $(3\tilde{\gamma}(\lambda) + \tilde{\gamma}(\lambda)^2) = O(\epsilon_S \sqrt{N})$
    - Much better than $O(N)$
- Grow $\lambda \approx \sqrt{\log(N)}$ and $Q(\lambda, N^3/3 + 3N^2/2 + 7N/6) \approx 1$

So you can use $\sqrt{N}$ with confidence(?)

What should we observe if $\sqrt{N}$ is the right thing?

- Trouble (slow nonlinear convergence) when $\sqrt{N}\epsilon_S \geq .1$
  - Double: $N \approx 10^{30}$. Not on my computer.
  - Single: $N \approx 10^{14}$. Not on my computer.
  - Half: $N \approx 10^{6}$. Maybe if we push it.
- Expectation: Single just as good as double.
- Expect to see deterioration with N for half.

Chandrasekhar H-equation

Midpoint rule discretization

$$\mathcal{F}(H)(\mu) = H(\mu) - \left(1 - \frac{c}{2}\int_0^1 \frac{\mu H(\mu)}{\mu + \nu}\, d\nu\right)^{-1} = 0.$$

- Defined on $C[0, 1]$
- $\mathcal{F}'$ nonsingular for $0 \le c < 1$.
  Simple fold singularity at $c = 1$.
- Any sensible discretization inherits the singularity structure.

Discrete Problem

$$F(u)_i \equiv u_i - \left(1 - \frac{c}{2N} \sum_{j=1}^{N} \frac{u_j \mu_i}{\mu_j + \mu_i}\right)^{-1} = 0.$$

Midpoint rule says

$$\frac{c}{2N} \sum_{j=1}^{N} \frac{u_j \mu_i}{\mu_j + \mu_i} = \frac{c(i - 1/2)}{2N} \sum_{j=1}^{N} \frac{u_j}{i + j - 1}.$$

so can evaluate F in $O(N \log(N))$ work with FFT.

Analytic Jacobian

Define M by

$$\mathrm{M}(\mathrm{u})_i = \frac{c(i-1/2)}{2N} \sum_{j=1}^{N} \frac{u_j}{i+j-1}$$

and compute the Jacobian analytically as

$$\mathrm{F}'(\mathrm{u}) = \mathrm{I} - \mathrm{diag}(\mathrm{G}(\mathrm{u}))^2 \mathrm{M}$$

where

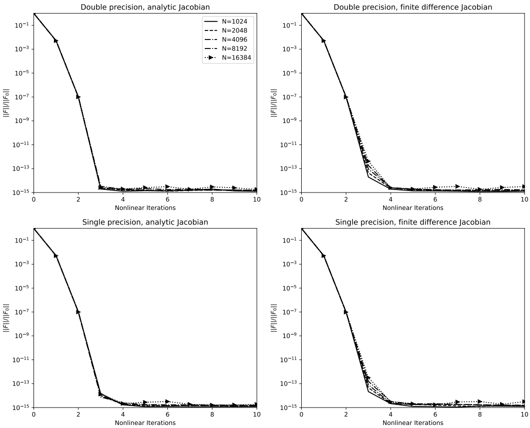
$$\mathrm{G}(\mathrm{u})_i = \left(1 - \frac{c}{2N} \sum_{j=1}^{N} \frac{u_j \mu_i}{\mu_j + \mu_i} \right)^{-1}.$$

Takes $O(N^2)$ work.

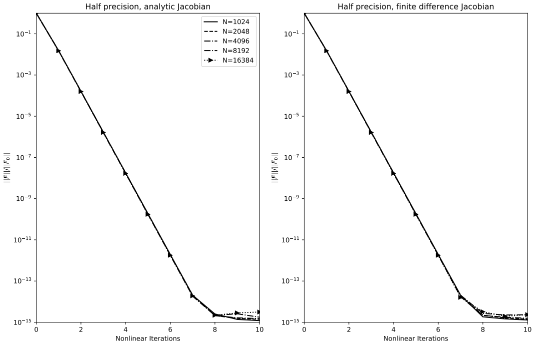
Experiments

- $c = .5, .99, 1.0$ (no theory for $c = 1.0$)
- Analytic and forward difference Jacobians
  Theory predicts single as good as double
- Double, single, and half precision factor/solve
- Everything else in double
- $N = 2^p$, $p = 10, \dots, 14$, $2^{14} = 16384$
  Larger N took far too long in half.

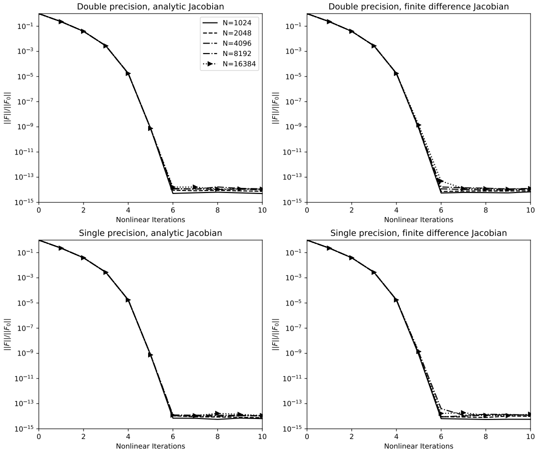
$c = .5$, double and single



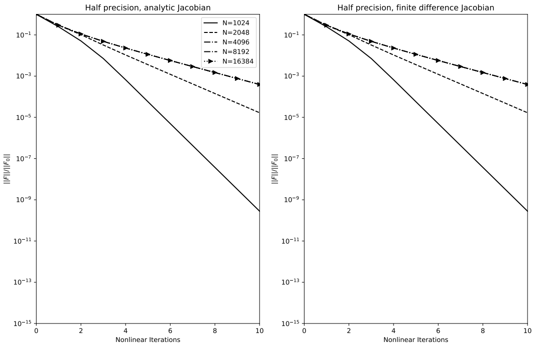
$c = .5$, half, not quadratic looking



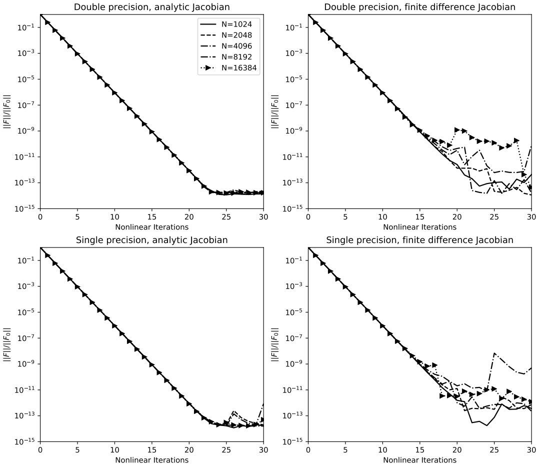
$c = .99$, double and single



$c = .99$, half, Wait! What?



$c = 1.0$, double and single, theory not from this talk



What's up with $c = 1$?

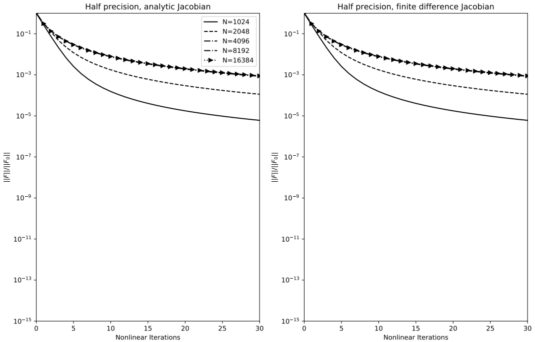
It's like $f(x) = x^2 = 0$.

- $x^* = 0$
- $f'(x) = 2x$ so $f'(x^*) = 0$. Singular!
- Newton: $x_+ = x_c - x_c^2/(2x_c) = x_c/2$ if $x_c \neq 0$
  Not quadratic!
- And why does the difference Jacobian go south?

$$f'(x) = 0 \text{ implies } (f(x+h) - f(x))/h = O(h)$$

so you're not entitled to much.

$c = 1.0$, half, DOOM! Some theory out there



What? Is that converging at all?

Back to $x^2 = 0$.

- Chord method: $x_+ = x_c - f'(x_0)^{-1} f(x_c)$
- $x_0 = 1$
- $x_+ = x_c - x_c^2/2 = x_c(1 - x_c/2)$
- Then (exercise for faculty)

$$\lim_{n \to \infty} \frac{x_n}{2/n} = 1.$$

- Sublinear convergence, sad!

Reproduciblity

- Codes in Julia (no joke!)
  - Julia makes managing reproducitlity easy.
  - You can use plain vanilla Jupyter notebooks.
- Results in the paper
  `https://github.com/ctkelley/MPResults`
  - Solver + H-equation in Julia
  - Story in Notebooks
    pdf works all the time; note book via html works sometimes

New book under contract

Solving Nonlinear Equations with Iterative Methods: Solvers and Examples in Julia
SIAM: Publication sometime in 2022

Three parts

- Print book: sequel to FA1:
  C. T. Kelley, Solving Nonlinear Equations with Newton's Method, number 1 in Fundamentals of Algorithms, SIAM, Philadelphia, 2003.
- IJulia (aka Jupyter) notebook at
  https://github.com/ctkelley/NotebookSIAMFANL
- Julia package with solvers+test problems+examples
  https://github.com/ctkelley/SIAMFANLEquations.jl

Warning!

- Under development and changing constantly
  - As the Julia people say "breaking changes" are possible
- Not formally registered yet
  - Once registered I'll have stable branch for the package/notebook
  - For now, the master branch is your best bet

Summary

- Low quality linear solvers are just fine
  - Single precision $\to$ same nonlinear results
  - Half precision $\to$ not great
  - The precision for you is 32!
  - $c = 1.0$ is different
- Software out there.
- Book in progress.