Inexact Newton Methods, Newton-Krylov Methods, and Extensions for Large-Scale Underdetermined Systems

Homer Walker
DOE Office of Advanced Scientific Computing
and
Worcester Polytechnic Instititute

Includes joint work with Roger Pawlowski (SNL), John Shadid (SNL), and Joseph Simonis (WPI/Boeing).

Supported in part by the DOE ASC program and Office of Science MICS program, Sandia National Laboratories CSRI, and the DOE-funded University of Utah C-SAFE.

Classical Newton's Method

Problem: $F(u) = 0, \quad F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ continuously differentiable.

Newton's Method

Given an initial u.

Iterate:

Solve $F'(u)s = -F(u)$.

Update $u \leftarrow u + s$.

Guiding application: *discretized nonlinear PDEs*.

Typically . . .

- quadratic, mesh-independent *local* convergence $\Rightarrow$ *globalize*,
- n is *very* large, $F'(u)$ is sparse and may be infeasible to evaluate/store $\Rightarrow$ *Krylov subspace method*.

Globalizations of Newton's Method

We can't *guarantee* convergence to a solution . . .

. . . but we can make it more likely.

Globalizations of Newton's Method

We can't *guarantee* convergence to a solution . . .

. . . but we can make it more likely.

Idea: Repeat as necessary . . .

- *Test* a step for acceptable progress.
- If unacceptable, *modify* it and test again.

Globalizations of Newton's Method

We can't *guarantee* convergence to a solution . . .

. . . but we can make it more likely.

Idea: Repeat as necessary . . .

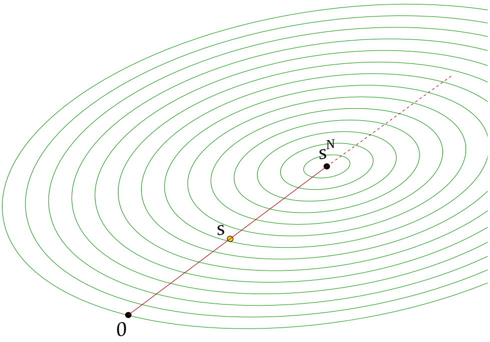
- *Test* a step for acceptable progress.
- If unacceptable, *modify* it and test again.

Major approaches:

- *Backtracking* (linesearch, damping).
- *Trust region*.

Backtracking (Linesearch, Damping) Globalization

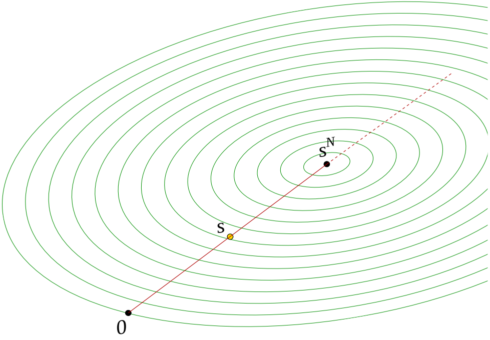
- $s \longleftarrow \theta s^N$ for an appropriate $\theta > 0$.
  - s^N is a *descent direction* for $\|F\|$ at x
  - $\Rightarrow$ s is acceptable for sufficiently small $\theta > 0$.



Red: feasible s
Green: level curves of $\|F(u) + F'(u)s\|$

Backtracking (Linesearch, Damping) Globalization

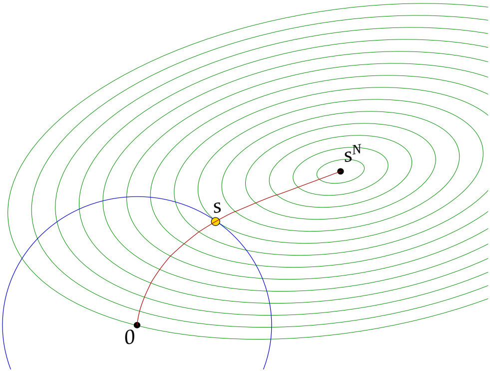
- $s \longleftarrow \theta s^N$ for an appropriate $\theta > 0$.
  - s^N is a *descent direction* for $\|F\|$ at x
  - $\Rightarrow s$ is acceptable for sufficiently small $\theta > 0$.
- s^N may be only a "weak" descent direction if $F'(u)$ is ill-conditioned.



Red: feasible s
Green: level curves of $\|F(u) + F'(u)s\|$

Trust-Region Globalization

- $s = \arg\min_{\|w\| \leq \delta} \|F(u) + F'(u)\,w\|$.

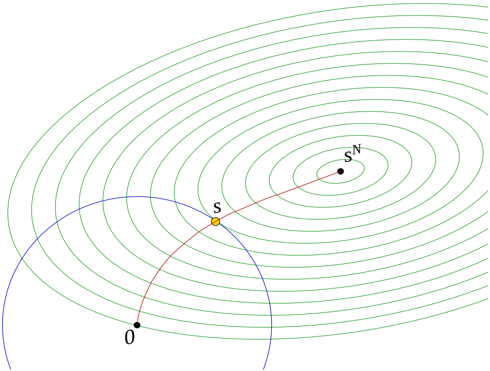


Red: feasible s
Green: level curves of $\|F(u) + F'(u)s\|$
Blue: trust region boundary

Trust-Region Globalization

- $s = \arg\min_{\|w\| \le \delta} \|F(u) + F'(u)\,w\|$.

- Computing s accurately may be problematic.

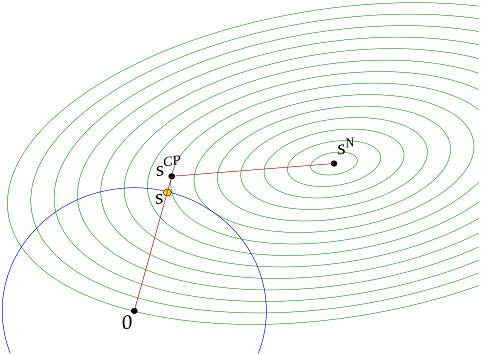


Red: feasible s
Green: level curves of $\|F(u) + F'(u)s\|$
Blue: trust region boundary

The Dogleg Step

Define

- *Cauchy point* $s^{CP} \equiv \arg\min_{0 \le \lambda < \infty} \|F(u) - F'(u)\,\lambda \nabla f(u)\|, \; f(u) \equiv \frac{1}{2}\|F(u)\|^2$
- *dogleg curve* $\Gamma^{DL}\colon\; 0 \to s^{CP} \to s^N$
- *dogleg step* $s = \arg\min_{\|w\| \le \delta,\, w \in \Gamma^{DL}} \|F(u) + F'(u)\, w\|$



Newton–Krylov Methods

Use a *Krylov subspace method* to approximately solve $F'(u)\, s = -F(u)$.

For solving $Ax = b$ . . .

Krylov Subspace Method

Given x_0, determine . . .

$$x_k = x_0 + z_k,$$

$$z_k \in \mathcal{K}_k \equiv \mathrm{span}\,\{r_0, Ar_0, \ldots, A^{k-1} r_0\},$$

A few examples . . .

CG/CR, GMRES, BCG, CGS, QMR, TFQMR (QMRCGS), QMR-squared, BiCGSTAB, BiCGSTAB2, BiCGSTAB(ℓ), QMRCGSTAB, Arnoldi (FOM/IOM), GMRESR, GCR, GMBACK, MINRES, SYMMLQ, ORTHODIR, ORTHOMIN, ORTHORES, Axelsson, SYMMBK, CGNR, CGNE, LSQR,. . . .

Newton–Krylov Methods

Special appeal of Krylov subspace methods:

- Most require only products of $F'(u)$ with vectors $\Longrightarrow$ *"matrix-free"* implementations.
- They have desirable *optimality properties*.
  - GMRES and other "minimum residual" methods minimize the linear residual norm $\|F(u) + F'(u)\, s\|$ (the *linear model norm*) over each $\mathcal{K}_k$.
  - For *optimization*, say $\min_{u \in R^n} f(u)$, $f : \boldsymbol{R}^n \rightarrow \boldsymbol{R}^1$,
    - CG minimizes the *local quadratic model* $q(s) \equiv f(u) + \nabla f(u)^T s + \frac{1}{2}\, s^T \nabla^2 f(u) s$ over each $\mathcal{K}_k$.
    - The first CG step is the steepest descent step.

Inexact Newton Methods

Inexact Newton methods (Dembo–Eisenstat–Steihaug 1982) provide a framework for analysis and implementation.

Inexact Newton Method

Given an initial u.

Iterate:

Find **some** $\eta \in [0, 1)$ and s that satisfy

$$\|F(u) + F'(u)\, s\| \leq \eta \|F(u)\|.$$

Update $u \leftarrow u + s$.

Inexact Newton Methods

Regard *Newton–Krylov methods as a special case* . . .

- Choose $\eta \in [0, 1)$.
- Apply the Krylov solver to $F'(u)\, s = -F(u)$ until

$$\|F(u) + F'(u)\, s\| \leq \eta \|F(u)\|.$$

Inexact Newton Methods

Regard *Newton–Krylov methods as a special case* ...

- Choose $\eta \in [0, 1)$.
- Apply the Krylov solver to $F'(u)\, s = -F(u)$ until

$$\|F(u) + F'(u)\, s\| \leq \eta \|F(u)\|.$$

The issue of when to stop the linear iterations becomes the issue of choosing the "forcing term" η.

Local Convergence of Inexact Newton Methods

Dembo–Eisenstat–Steihaug (1982): *Local convergence is controlled by the forcing terms.*

Theorem: *Suppose $F(u_*) = 0$ and $F'(u_*)$ is invertible. If $\{u_k\}$ is an inexact Newton sequence with u_0 sufficiently near u_*, then*

- $\eta_k \le \eta_{\max} < 1 \quad \Longrightarrow \quad u_k \to u_*$ *linearly in the norm* $\|w\|_{F'(u_*)} \equiv \|F'(u_*)\, w\|$,
- $\eta_k \to 0 \quad \Longrightarrow \quad u_k \to u_*$ *superlinearly.*

If also F' is Hölder continuous with exponent p at u_*, then

- $\eta_k = O(\|F(u_k)\|^p) \quad \Longrightarrow \quad u_k \to u_*$ *with q-order $1 + p$.*

More on forcing terms later . . .

Globalizations of Newton–Krylov Methods

Present a subset of results in

R. P. Pawlowski, J. P. Simonis, J. N. Shadid, HW, *Globalization techniques for Newton–Krylov methods and applications to the fully coupled solution of the Navier–Stokes equations*, SIREV, 48 (2006), 700–721.

- Describe two representative Newton–Krylov globalizations:
  - a backtracking method,
  - a dogleg trust-region method.
- Outline their theoretical support and discuss a few implementational details.
- Report on numerical experiments.

The Backtracking Method

The backtracking method (Eisenstat-HW 1994) is . . .

Inexact Newton Backtracking (INB) Method

Given an initial u and $\eta_{\max} \in [0, 1)$, $t \in (0, 1)$, and $0 < \theta_{\min} < \theta_{\max} < 1$.

Iterate:

Choose initial $\eta \in [0, \eta_{\max}]$ and s such that

$$\|F(u) + F'(u)\, s\| \leq \eta \|F(u)\|.$$

While $\|F(u+s)\| > [1 - t(1-\eta)]\|F(u)\|$, do:

Choose $\theta \in [\theta_{\min}, \theta_{\max}]$.

Update $s \leftarrow \theta s$ and $\eta \leftarrow 1 - \theta(1-\eta)$.

Update $u \leftarrow u + s$.

Global Convergence Theorem

Theorem: *If $\{u_k\}$ produced by the INB method has a limit point u_* such that $F'(u_*)$ is nonsingular, then $F(u_*) = 0$ and $u_k \to u_*$. Furthermore, the initial s_k and η_k are accepted for all sufficiently large k.*

Possibilities:

- $\|u_k\| \to \infty$.
- $\{u_k\}$ has limit points, and F' is singular at each one.
- $\{u_k\}$ converges to u_* such that $F(u_*) = 0$, $F'(u_*)$ is nonsingular, and *asymptotic convergence is determined by the initial η_k's.*

The Dogleg Method

Inexact Newton Dogleg (INDL) Method

Given an initial u and $\eta_{\max} \in [0,1)$, $t \in (0,1)$, $0 < \theta_{\min} < \theta_{\max} < 1$, and $0 < \delta_{\min} \le \delta$.

Iterate:

Choose $\eta \in [0, \eta_{\max}]$ and s^{IN} such that

$$\|F(u) + F'(u)\, s^{IN}\| \le \eta \|F(u)\|.$$

Evaluate s^{CP} and determine $s \in \Gamma^{DL}$: $0 \to s^{CP} \to s^{IN}$.

While *ared* $< t \cdot$ *pred* do:

Choose $\theta \in [\theta_{\min}, \theta_{\max}]$.

Update $\delta \leftarrow \max\{\theta\delta, \delta_{\min}\}$.

Redetermine $s \in \Gamma^{DL}$.

Update $u \leftarrow u + s$ and update δ.

Dogleg Details

- Sufficient decrease is based on the inexact Newton condition and

$$ared \equiv \|F(u)\| - \|F(u+s)\| \qquad \text{(actual reduction)}$$

$$pred \equiv \|F(u)\| - \|F(u) + F'(u)s\| \qquad \text{("predicted" reduction)}$$

- Update δ a la Dennis–Schnabel (1983).

- Determine $s \in \Gamma^{DL}$ by the "standard strategy":
  - If $\|s^{IN}\| \leq \delta$, then $s = s^{IN}$;
  - else, if $\|s^{CP}\| \geq \delta$, then $s = (\delta/\|s^{CP}\|)\, s^{CP}$;
  - else, $s = (1-\tau)s^{CP} + \tau s^{IN}$, where $\tau \in (0,1)$ is uniquely determined so that $\|s\| = \delta$.

- Alternative dogleg strategies and refinements are given in R. P. Pawlowski, J. P. Simonis, HW, J. N. Shadid, *Inexact Newton dogleg methods*, SINUM, 46 (2007-2008), 2112-2132.

Global Convergence Theorem

Recall: u is a *stationary point* of $\|F\| \iff \|F(u)\| \le \|F(u) + F'(u)\, s\|$ for all s.

Theorem: *If u_* is a limit point of $\{u_k\}$ produced by the INDL method, then u_* is a stationary point of $\|F\|$. If additionally $F'(u_*)$ is nonsingular, then $F(u_*) = 0$ and $u_k \longrightarrow u_*$; furthermore, $s_k = s_k^{IN}$ for all sufficiently large k.*

Possibilities:

- $\|u_k\| \longrightarrow \infty$.
- $\{u_k\}$ has limit points, and each is a stationary point of F.
- $\{u_k\}$ converges to u_* such that $F(u_*) = 0$, $F'(u_*)$ is nonsingular, and *asymptotic convergence is determined by the initial η_k's.*

Choosing $\theta \in [\theta_{\min}, \theta_{\max}]$

Two typical procedures were used in the numerical experiments (see Dennis–Schnabel (1983)).

- Choose θ to *minimize a quadratic* $p(t)$ that satisfies $p(0) = \frac{1}{2}\|F(u)\|^2$, $p(1) = \frac{1}{2}\|F(u+s)\|^2$, and $p'(0) = \frac{d}{dt}\frac{1}{2}\|F(u+ts)\|^2\Big|_{t=0}$.
- Choose θ to minimize
  - a *quadratic* on the first reduction,
  - a *cubic* on subsequent reductions.

Choosing the Forcing Terms

Two choices were used in the numerical experiments.

- **Small constant forcing terms**: $\eta_k = 10^{-4}$ for each k
  $\Rightarrow$ fast local linear convergence.

- **Adaptive forcing terms**: "Choice 1" from (Eisenstat–HW 1996)

$$\eta_k = \min\left\{ \frac{\left| \|F(u_k)\| - \|F(u_{k-1}) + F'(u_{k-1})\, s_{k-1}\| \right|}{\|F(u_{k-1})\|}, \eta_{\max} \right\}.$$

Theorem: *Suppose $F(u_*) = 0$ and $F'(u_*)$ is invertible. Let $\{u_k\}$ be an inexact Newton sequence with each η_k given as above. If u_0 is sufficiently near u_*, then $u_k \rightarrow u_*$ with*

$$\|u_{k+1} - u_*\| \leq \beta \|u_k - u_*\| \cdot \|u_{k-1} - u_*\|, \qquad k = 1, 2, \ldots$$

for a constant β independent of k.

Numerical Experiments

- Test problems: Three benchmark flow problems in 2D and 3D . . .
  - lid-driven cavity,
  - thermal convection,
  - backward-facing step.
- PDEs: Low Mach number Navier–Stokes equations with heat transport as appropriate.
- Discretization: Pressure stabilized streamline upwind Petrov–Galerkin FEM.
- Algorithms and software: Newton–GMRES implementations in the Sandia *NOX* nonlinear solver suite, with GMRES and domain-based (overlapping Schwarz) ILU preconditioners from the Sandia *Aztec* package. The simulation driver was the Sandia *MPSalsa* parallel reacting flow code.
- Problem sizes: 25,263 to 1,042,236 unknowns.
- Machines: 8 CPUs on a 16-node, 32-CPU IBM Linux cluster; 100 CPUs on Sandia's 256-node, 512-CPU Institutional Cluster.

Robustness

2D and 3D Thermal Convection	$\mathrm{Ra} =$	$10^3, 10^4, 10^5, 10^6$
2D and 3D Backward Facing Step	$\mathrm{Re} =$	100, 200, . . . , 700, 750, 800
2D Lid Driven Cavity	$\mathrm{Re} =$	1000, 2000, . . . , 10, 000
3D Lid Driven Cavity	$\mathrm{Re} =$	100, 200, . . . , 1000

Total numbers of failures:

Method	Forcing Term	2D Problems		3D Problems		All Problems	
Backtracking, Quadratic Only	Adaptive	0	10	0	0	0	10
	10^{-4}	10		0		10	
Dogleg	Adaptive	0	10	0	0	0	10
	10^{-4}	10		0		10	
No Globalization	Adaptive	15	33	4	14	19	47
	10^{-4}	18		10		28	

Efficiency

2D Thermal Convection	$\mathrm{Ra} =$	10^3, 10^4, 10^5
3D Thermal Convection	$\mathrm{Ra} =$	10^3, 10^4, 10^5, 10^6
2D and 3D Backward Facing Step	$\mathrm{Re} =$	100, 200, ..., 700
2D and 3D Lid Driven Cavity	$\mathrm{Re} =$	100, 200, ..., 1000

Method	Forcing Term	Inexact Newton Steps	Backtracks per INS	GMRES Iterations per INS	Normalized Time
Backtracking, Quadratic Only	Adaptive	16.0	0.13	62.2	0.77
	10^{-4}	9.23	0.18	163	1.0 (REF)
Dogleg	Adaptive	17.0	NA	85.3	0.83
	10^{-4}	10.7	NA	168	1.01

⟵ Geometric Means ⟶

Observations

- These globalizations have good theoretical support and are effective on these test problems, especially with adaptive forcing terms.

Observations

- These globalizations have good theoretical support and are effective on these test problems, especially with adaptive forcing terms.

- Causes of failure in our experiments:
  - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 10/10 dogleg failures.
  - Globalization failure: 7/33 backtracking/linesearch failures.

Observations

- These globalizations have good theoretical support and are effective on these test problems, especially with adaptive forcing terms.
- Causes of failure in our experiments:
  - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 10/10 dogleg failures.
  - Globalization failure: 7/33 backtracking/linesearch failures.
- Backtracking with quadratic minimization and adaptive forcing terms seems to be a clear first choice for implementation.

Observations

- These globalizations have good theoretical support and are effective on these test problems, especially with adaptive forcing terms.
- Causes of failure in our experiments:
    - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 10/10 dogleg failures.
    - Globalization failure: 7/33 backtracking/linesearch failures.
- Backtracking with quadratic minimization and adaptive forcing terms seems to be a clear first choice for implementation.
- No globalization or choice of forcing terms is always best.

Observations

- These globalizations have good theoretical support and are effective on these test problems, especially with adaptive forcing terms.
- Causes of failure in our experiments:
  - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 10/10 dogleg failures.
  - Globalization failure: 7/33 backtracking/linesearch failures.
- Backtracking with quadratic minimization and adaptive forcing terms seems to be a clear first choice for implementation.
- No globalization or choice of forcing terms is always best.
- *Many* factors contribute to success: problem formulation, discretization, preconditioning, variable scaling, accuracy, . . .

Observations

- These globalizations have good theoretical support and are effective on these test problems, especially with adaptive forcing terms.
- Causes of failure in our experiments:
  - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 10/10 dogleg failures.
  - Globalization failure: 7/33 backtracking/linesearch failures.
- Backtracking with quadratic minimization and adaptive forcing terms seems to be a clear first choice for implementation.
- No globalization or choice of forcing terms is always best.
- *Many* factors contribute to success: problem formulation, discretization, preconditioning, variable scaling, accuracy, . . .
- For more, see the SIREV and SINUM papers.

The Underdetermined System Problem

Problem: Given $F : I\!R^m \rightarrow I\!R^n$ with $m > n$, find u_* such that $F(u_*) = 0$.

Assume F is continuously differentiable throughout.

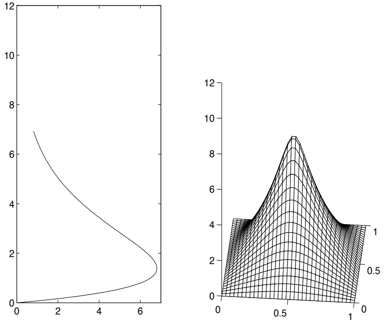
Examples:

- Parameter-dependent problems with unknown parameters.
- Time-dependent problems with periodic solutions.
- Nonlinear eigenvalue problems.

The Bratu (Gelfand) Problem

In 2D, this is

$$\begin{aligned} \Delta u + \lambda e^u &= 0 \text{ in } \mathcal{D} \equiv [0,1] \times [0,1], \\ u &= 0 \text{ on } \partial\mathcal{D}. \end{aligned}$$



Left: $\|u\|$ *vs.* λ. *Right: solution at final* λ *value.*

The Model Algorithm

Extend Newton's method with ...

Algorithm NU: Newton's Method (Underdetermined)

Given u_0.

For $k = 0, 1, \ldots$

Find $s_k \in \mathbb{R}^m$ such that

$$F'(u_k)s_k = -F(u_k), \quad s_k \perp \mathcal{N}(F'(u_k)).$$

Set $u_{k+1} = u_k + s_k$.

Appeal:

- This *pseudo-inverse* characterization of s_k is optimally conditioned.
- The algorithm has local convergence (up to quadratic) like that of Newton's method (HW–Watson 1990, Levin–Ben Israel 2001).

An Inexact Newton Extension

Extend inexact Newton methods with . . .

Algorithm INU: Inexact Newton Method (Underdetermined)

Given u_0.

For $k = 0, 1, \ldots$

Find $\eta_k \in [0, 1)$ and $s_k \in I\!R^m$ such that

$$\|F(u_k) + F'(u_k)s_k\| \leq \eta_k \|F(u_k)\|, \quad s_k \perp \mathcal{N}(F'(u_k)).$$

Set $u_{k+1} = u_k + s_k$.

Local Convergence Analysis

Hypothesis: The following hold in an open, convex $\Omega \subseteq \boldsymbol{R}^m$:

- F' is full-rank in Ω.
- There are $\gamma \geq 0$ and $p \in (0, 1]$ such that $\|F'(\tilde{u}) - F'(u)\| \leq \gamma \|\tilde{u} - u\|^p$ for all u, $\tilde{u} \in \Omega$.
- There is a μ such that $\|F'(u)^+\| \leq \mu$ for all $u \in \Omega$.

For $\rho > 0$, set $\Omega_\rho \equiv \{u \in \Omega : \ \|\tilde{u} - u\| \leq \rho \Rightarrow \tilde{u} \in \Omega\}$.

Local Convergence Analysis

Hypothesis: The following hold in an open, convex $\Omega \subseteq \boldsymbol{R}^m$:

- F' is full-rank in Ω.
- There are $\gamma \geq 0$ and $p \in (0, 1]$ such that $\|F'(\tilde{u}) - F'(u)\| \leq \gamma \|\tilde{u} - u\|^p$ for all u, $\tilde{u} \in \Omega$.
- There is a μ such that $\|F'(u)^+\| \leq \mu$ for all $u \in \Omega$.

For $\rho > 0$, set $\Omega_\rho \equiv \{u \in \Omega : \ \|\tilde{u} - u\| \leq \rho \Rightarrow \tilde{u} \in \Omega\}$.

Theorem: *Suppose that this hypothesis holds and that $\rho > 0$ is given. Assume that $\eta_k \leq \eta_{\max} < 1$ for all k. Then there exists an $\epsilon > 0$ depending only on γ, p, μ, ρ, and $\eta_{\max}$ such that if $u_0 \in \Omega_\rho$ and $\|F(u_0)\| \leq \epsilon$, then the iterates $\{u_k\}$ determined by Algorithm INU are well-defined and converge to $u_* \in \Omega$ such that $F(u_*) = 0$. Moreover, if $u_k \neq u_*$ for all k, then*

$$\limsup_{k \to \infty} \frac{\|F'(u_*)(u_{k+1} - u_*)\|}{\|F'(u_*)(u_k - u_*)\|} \leq \eta_{\max}. \qquad (\star)$$

Additionally, if $\eta_k \to 0$, then the convergence is q-superlinear, and if $\eta_k = O(\|F(u_k)\|^p)$, then the convergence is of q-order $1 + p$.

Local Convergence Analysis

Hypothesis: The following hold in an open, convex $\Omega \subseteq \boldsymbol{R}^m$:

- F' is full-rank in Ω.
- There are $\gamma \geq 0$ and $p \in (0, 1]$ such that $\|F'(\tilde{u}) - F'(u)\| \leq \gamma \|\tilde{u} - u\|^p$ for all u, $\tilde{u} \in \Omega$.
- There is a μ such that $\|F'(u)^+\| \leq \mu$ for all $u \in \Omega$.

For $\rho > 0$, set $\Omega_\rho \equiv \{u \in \Omega : \ \|\tilde{u} - u\| \leq \rho \Rightarrow \tilde{u} \in \Omega\}$.

Theorem: *Suppose that this hypothesis holds and that $\rho > 0$ is given. Assume that $\eta_k \leq \eta_{\max} < 1$ for all k. Then there exists an $\epsilon > 0$ depending only on γ, p, μ, ρ, and $\eta_{\max}$ such that if $u_0 \in \Omega_\rho$ and $\|F(u_0)\| \leq \epsilon$, then the iterates $\{u_k\}$ determined by Algorithm INU are well-defined and converge to $u_* \in \Omega$ such that $F(u_*) = 0$. Moreover, if $u_k \neq u_*$ for all k, then*

$$\limsup_{k \to \infty} \frac{\|F'(u_*)(u_{k+1} - u_*)\|}{\|F'(u_*)(u_k - u_*)\|} \leq \eta_{\max}. \qquad (\star)$$

Additionally, if $\eta_k \to 0$, then the convergence is q-superlinear, and if $\eta_k = O(\|F(u_k)\|^p)$, then the convergence is of q-order $1 + p$.

Remark: One can show that $\|F'(u_*)(u_k - u_*)\| \geq C\|u_k - u_*\|$ for all large k. Then it follows from $(\star)$ that $u_k \to u_*$ r-linearly.

A Backtracking Method

Extend the INB method with ...

Algorithm INBU:

Given u_0 and $t \in (0,1)$, $\eta_{\max} \in [0,1)$, and $0 < \theta_{\min} < \theta_{\max} < 1$.

For $k = 0$ step 1 until ∞ do:

Find initial $\eta_k \in [0, \eta_{\max}]$ and s_k such that

$$\|F(u_k) + F'(u_k)s_k\| \leq \eta_k \|F(u_k)\|, \quad s_k \perp \mathcal{N}(F'(u_k)).$$

Evaluate $F(u_k + s_k)$.

While $\|F(u_k + s_k)\| > [1 - t(1 - \eta_k)]\, \|F(u_k)\|$, do

Choose $\theta \in [\theta_{\min}, \theta_{\max}]$.

Update $s_k \leftarrow \theta s_k$ and $\eta_k \leftarrow 1 - \theta(1 - \eta_k)$.

Evaluate $F(u_k + s_k)$.

Set $u_{k+1} = u_k + s_k$.

Global Convergence Theorem

Theorem: *Suppose that* $\{u_k\}$ *is generated by Algorithm INBU. If* u_* *is a limit point of* $\{u_k\}$ *such that* $F'(u_*)$ *is full-rank, then* $F(u_*) = 0$ *and* $u_k \to u_*$*. Furthermore, the initial* η_k *and* u_k *are accepted without modification in the while-loop for all sufficiently large* k*.*

Possibilities:

- $\|u_k\| \to \infty$.
- $\{u_k\}$ has limit points, and F' is rank-deficient at each.
- $\{u_k\}$ converges to u_* such that $F(u_*) = 0$, $F'(u_*)$ is full-rank, and asymptotic convergence is determined by the initial η_k's.

Note: By taking $\eta_{\max} = 0$ in Algorithm INBU, we obtain a bactracking extension of Algorithm NU, to which this theorem applies.

Solving for s_k

Extend the technique in (SISC, 2000) for *adapting Krylov subspace methods*.

Set $\ell = m - n$. Let $\{v_1, \ldots, v_\ell\}$ be an orthonormal basis of $\mathcal{N}(F'(u_k))$.

- For $i = 1, \ldots, \ell$,
  Obtain a Householder P_i such that $P_i \ldots P_1 v_i = e_{n-i+1} \in \mathbf{R}^m$.
- Set $Q = P_1 \ldots P_\ell \begin{pmatrix} I_n \\ 0 \end{pmatrix} \in \mathbf{R}^{m \times n}$.
- Apply the Krylov subspace method to approximately solve

$$F'(u_k) Q \tilde{s}_k = -F(u_k)$$

- Set $s_k = Q \tilde{s}_k \in \mathbf{R}^m$.

Cost:

- $O(\ell^2 m)$ flops and $O(\ell m)$ storage for $P_1, \ldots, P_\ell$.
- $O(\ell m)$ flops for each Q-product.

Obtaining an Orthonormal Basis of $\mathcal{N}(F'(u_k))$

For $i = 1, \ldots, \ell$,

- Obtain an initial v_i orthogonalized against $v_1, \ldots, v_{i-1}$ and normalized.
- Obtain Δv_i such that $F'(u_k)(v_i + \Delta v_i) = 0$ and $\Delta v_i \perp v_1, \ldots, v_i$. (Take $P_{i+1} = \ldots = P_\ell = I_m$ in forming Q.)
- Update $v_i \leftarrow (v_i + \Delta v_i)/\|v_i + \Delta v_i\|$.

Cost: $O(\ell^2 m)$ flops plus ℓ solves.

Implementation details:

- MATLAB code.
- Parameters: $t = 10^{-4}$, $\eta_{\max} = .9$, $[\theta_{\min}, \theta_{\max}] = [.1, .9]$.
- Krylov solver: Restarted GMRES applied as just outlined.
- Forcing terms: "Choice 1" from (Eisenstat–W 1996), with $\eta_0 = \eta_{\max} = .9$.
- Backtracking: $\theta \in [\theta_{\min}, \theta_{\max}]$ chosen to minimize an interpolating quadratic.

Test Problems

PDEs on $\mathcal{D} = [0,1] \times [0,1]$.

2D Bratu Problem: $\Delta u + \lambda e^u = 0$ in $\mathcal{D}$, $u = 0$ on $\partial\mathcal{D}$

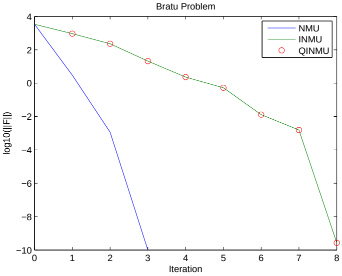
- Unknowns u, λ; $u_0 = 2\sin(\pi u)\sin(\pi y)$, $\lambda_0 = 7.0$.
- Centered differences, 50×50 grid $\Rightarrow n = 2500$, $m = 2501$.
- GMRES(20), up to 3 restarts, Poisson-solver right preconditioning.

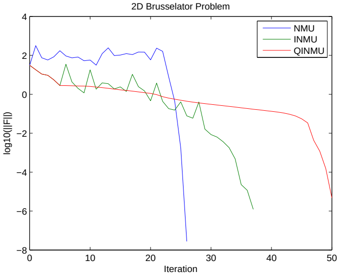
2D Brusselator Problem:

$$\begin{aligned}
\partial u/\partial t &= \alpha\Delta u + 1 + u^2 v - 4.4u \text{ in } \mathcal{D} \\
\partial v/\partial t &= \alpha\Delta v + 1 + 3.4u - u^2 v \text{ in } \mathcal{D} \\
\partial u/\partial n &= \partial v/\partial n = 0 \text{ on } \partial\mathcal{D}
\end{aligned}$$

- $\alpha = .002 \Rightarrow$ periodic solution.
- Unknowns u, v, T (period); $u_0 = 0.5 + y$, $v_0 = 1 + 5x$, $T = 7.5$.
- Centered differences, 21×21 grid $\Rightarrow n = 882$, $m = 883$.
- GMRES(50), up to 10 restarts, Poisson-solver right preconditioning.

Test Results: Bratu and Brusselator Problems





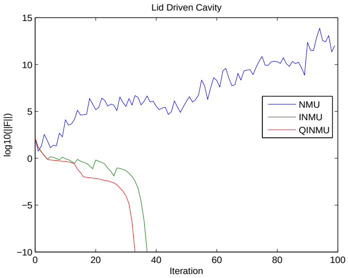
- (Inexact) Newton iterations vs. $\log_{10} \|F\|$.
- NMU = Algorithm NU.
- INMU = Algorithm INU.
- QINMU = Algorithm INBU

Note: On the Brusselator problem, the Algorithm NU iterates converged to the trivial solution (with zero period).

Test Results: Lid-Driven Cavity Problem

2D Lid-Driven Cavity Problem: $\frac{1}{\mathrm{Re}}\Delta^2 u - (u_y \Delta u_x + u_x \Delta u_y) = 0$ in $\mathcal{D}$, with $u = 0$ on $\partial\mathcal{D}$, $u_n = 0$ on the sides and bottom, and $u_n = 1$ on the top.

- Unknowns u, Re; $u_0 = 0$, $\mathrm{Re}_0 = 1000$.
- Centered differences, 40×40 staggered grid $\Rightarrow n = 1600$, $m = 1601$.
- GMRES(50), up to 10 restarts, biharmonic-solver right preconditioning.



- (Inexact) Newton iterations vs. $\log_{10} \|F\|$.
- NMU = (exact) Newton's method.
- INMU = Algorithm INU.
- QINMU = Algorithm INBU

Summary

We have:

- extended inexact Newton methods to underdetermined systems;
- provided local and global convergence results;
- reported results of limited numerical experiments.

Still needed:

- extensions to trust-region methods for underdetermined systems;
- *much* more testing.