Globalization Techniques for Newton–Krylov Methods

Homer Walker
Mathematical Sciences Department
Worcester Polytechnic Institute
January 30, 2004

Joint work with Roger Pawlowski (SNL), J. N. Shadid (SNL), J. P. Simonis (WPI).
Outline

• Introduce Newton’s method, globalizations, Newton–Krylov methods.

• Consider three representative globalizations.
  ▶ Describe the globalizations and their theoretical support.
  ▶ Report on extensive experiments with these methods applied to the steady-state Navier–Stokes equations on massively parallel machines.
**Nonlinear problem:** \( F(x) = 0, \quad F : \mathbb{R}^n \rightarrow \mathbb{R}^n. \)

Assume that \( F \) is continuously differentiable throughout.

**Newton’s Method:**

Given an initial \( x \).

Iterate:

- Solve \( F'(x)s = -F(x) \).
- Update \( x \leftarrow x + s \).
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.

- $s \leftarrow \theta s^N$ for an appropriate $\theta$.

- $s^N$ may be a “weak” descent direction if $F'(x)$ is ill-conditioned.

Green ellipses are level curves of $\|F(x) + F'(x)s\|$. 
**Trust region globalization.**

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

- Can’t be computed exactly.
The dogleg step.

- $\Gamma^{DL}: 0 \rightarrow s^{CP} \rightarrow s^{N}$.

- $s = \arg\min_{\|w\| \leq \delta, w \in \Gamma^{DL}} \|F(x) + F'(x)w\|$.
**Newton–Krylov methods.**

Use a *Krylov subspace method* to approximately solve $F'(x) s = -F(x)$. 
**Newton–Krylov methods.**

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

---

**Krylov Subspace Method:** For $Ax = b$ ...

Given $x_0$, set $r_0 = b - Ax_0$ and determine ...

$x_k = x_0 + z_k,$

$z_k \in \mathcal{K}_k \equiv \text{span} \{ r_0, Ar_0, \ldots, A^{k-1}r_0 \},$
Newton–Krylov methods.

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

---

**Krylov Subspace Method:** For $Ax = b$ ...

Given $x_0$, set $r_0 = b - Ax_0$ and determine ... 

$x_k = x_0 + z_k$, 

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

---

A few examples: CG/CR, GMRES, BCG, CGS, BiCGSTAB, QMR, TFQMR, MINRES, SYMMLQ, GCR, CGNR, CGNE, ...
Krylov subspace methods have *special appeal* for solving \( F'(x) s = -F(x) \).

They require only products of \( F'(x) \) (and, in some cases, \( F'(x)^T \) as well) with vectors.

\[ \Rightarrow \text{“matrix-free” implementations through . . .} \]

\( \rightarrow \) finite-difference approximation, e.g.,

\[ F'(x) v \approx \frac{1}{h} \left[ F(x + hv) - F(x) \right] \]

\( \rightarrow \) exact evaluation by *automatic differentiation*. 

They have desirable *optimality properties*.

- GMRES and other “minimum residual” methods minimize over $\mathcal{K}_k$ the **linear residual norm**

$$\| F(x) + F'(x) s \| = \text{linear model norm}$$
They have desirable *optimality properties*.

- GMRES and other “minimum residual” methods minimize over $\mathcal{K}_k$ the **linear residual norm** $= \|F(x) + F'(x)s\| = \text{linear model norm}$

- For *optimization*, say $\min_{x \in \mathbb{R}^n} f(x), \quad f : \mathbb{R}^n \to \mathbb{R}^1$,

  - The $k$th CG step minimizes over $\mathcal{K}_k$ the **local quadratic model** $= f(x) + \nabla f(x)^T s + \frac{1}{2} s^T \nabla^2 f(x)s$

  - The first CG step is the steepest descent step.
**Inexact Newton methods** (Dembo–Eisenstat–Steihaug 1982) provide a framework for analysis and implementation.

**Inexact Newton Method:**
Given an initial $x$.
Iterate:
Find **some** $\eta \in [0, 1)$ and $s$ that satisfy
\[
\|F(x) + F'(x) s\| \leq \eta \|F(x)\|.
\]
Update $x \leftarrow x + s$. 
Regard *Newton–Krylov methods as a special case* . . .

- Choose $\eta \in [0, 1)$.

- Apply the iterative linear solver to $F'(x) s = -F(x)$ until

  $$
  \|F(x) + F'(x) s\| \leq \eta \|F(x)\|.
  $$
Regard Newton–Krylov methods as a special case . . .

• Choose $\eta \in [0,1)$.

• Apply the iterative linear solver to $F'(x) s = -F(x)$ until

$$\|F(x) + F'(x) s\| \leq \eta \|F(x)\|.$$
Dembo–Eisenstat–Steihaug (1982): \textit{Local convergence is controlled by the forcing terms.}

\textbf{Theorem:} Suppose \( F(x_*) = 0 \) and \( F'(x_*) \) is invertible. If \( \{x_k\} \) is an inexact Newton sequence with \( x_0 \) sufficiently near \( x_* \), then

- \( \eta_k \leq \eta_{\text{max}} < 1 \implies x_k \to x_* \text{ linearly in norm} \)
  \[ \|w\|_{F'(x_*)} \equiv \|F'(x_*) w\|, \]

- \( \eta_k \to 0 \implies x_k \to x_* \text{ superlinearly}, \)

If also \( F' \) is Lipschitz continuous at \( x_* \), then

- \( \eta_k = O(\|F(x_k)\|) \implies x_k \to x_* \text{ quadratically.} \)
Efficiency and robustness may be improved by adaptive forcing terms (Eisenstat–W 1996).

“Choice 1”: \( \eta_k = \min \{ \eta_{\text{max}}, \tilde{\eta}_k \} \), where \( \eta_{\text{max}} \in [0, 1) \) and

\[
\tilde{\eta}_k = \frac{\|F(x_k)\| - \|F(x_{k-1}) + F'(x_{k-1}) s_{k-1}\|}{\|F(x_{k-1})\|}
\]  \hfill (*)

**Theorem:** Suppose \( F(x_*) = 0 \) and \( F'(x_*) \) is invertible. Let \( \{x_k\} \) be an inexact Newton sequence with each \( \eta_k \) given by \((*)\). If \( x_0 \) is sufficiently near \( x_* \), then \( x_k \rightarrow x_* \) with

\[\|x_{k+1} - x_*\| \leq \beta \|x_k - x_*\| \cdot \|x_{k-1} - x_*\|, \quad k = 1, 2, \ldots\]

for a constant \( \beta \) independent of \( k \).
Globalizations of Newton–Krylov methods.

• Describe three representative Newton–Krylov globalizations:
  — a backtracking method,
  — a linesearch method,
  — a dogleg method.

• Outline their theoretical support.

• Report on numerical experiments.
Globalizations of Newton–Krylov methods.

- Describe three representative Newton–Krylov globalizations:
  - a backtracking method,
  - a linesearch method,
  - a dogleg method.

- Outline their theoretical support.

- Report on numerical experiments.

Note: The methods and convergence results are outlined for general inexact Newton methods.
The **backtracking method** (Eisenstat-W 1994) is . . .

---

**Inexact Newton Backtracking (INB) Method.**

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

**Iterate:**

- **Choose initial** $\eta \in [0, \eta_{\text{max}}]$ and $s$ such that  
  \[ \| F(x) + F'(x) s \| \leq \eta \| F(x) \|. \]

- While $\| F(x + s) \| > [1 - t(1 - \eta)] \| F(x) \|$, do:
  - **Choose** $\theta \in [\theta_{\text{min}}, \theta_{\text{max}}]$.
  - **Update** $s \leftarrow \theta s$ and $\eta \leftarrow 1 - \theta(1 - \eta)$.

- **Update** $x \leftarrow x + s$. 

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

**Possibilities:**

- \( \|x_k\| \to \infty \).
- \( \{x_k\} \) has limit points, and \( F' \) is singular at each one.
- \( \{x_k\} \) converges to \( x_* \) such that \( F(x_*) = 0 \), \( F'(x_*) \) is nonsingular, and asymptotic convergence is determined by the initial \( \eta_k \)’s.
Choosing $\theta \in [\theta_{\text{min}}, \theta_{\text{max}}]$. We considered two procedures:

- Choose $\theta$ to minimize a quadratic $p(t)$ that satisfies
  \[
  p(0) = \frac{1}{2} \| F(x_k) \|^2, \quad p(1) = \frac{1}{2} \| F(x_k + s_k) \|^2, \quad \text{and}
  \]
  \[
  p'(0) = \frac{d}{dt} \left. \frac{1}{2} \| F(x_k + ts_k) \|^2 \right|_{t=0}.
  \]

- Choose $\theta$ to minimize
  
  - a quadratic on the first reduction,
  
  - a cubic on subsequent reductions.
The **linesearch method** (Moré–Thuente 1984) is ...

**Inexact Newton Moré–Thuente Linesearch (INMTL) Method.**

Given an initial $x$ and $\eta_{\text{max}} \in [0, 1)$.

**Iterate:**

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

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

Apply the **Moré–Thuente linesearch** to determine a final $s$.

**Update** $x \leftarrow x + s$. 
The Moré–Thuente linesearch.

With $\phi(\lambda) \equiv \frac{1}{2} \|F(x + \lambda s)\|^2$, the linesearch finds (with high likelihood) a $\lambda \geq 0$ satisfying the strong Wolfe conditions:

$$\phi(\lambda) \leq \phi(0) + \alpha \phi'(0) \lambda$$  \hspace{1cm} \text{(sufficient decrease condition)}

$$|\phi'(\lambda)| \leq \beta |\phi'(0)|$$  \hspace{1cm} \text{ (“curvature” condition)}
**Theorem:** Suppose that $x_0$ is given and $F$ is Lipschitz continuously differentiable on $\mathcal{L}(x_0) \equiv \{ x : \|F(x)\| \leq \|F(x_0)\| \}$. Assume that $\{x_k\}$ is produced by the INMTL method such that, for each $k$, the $\lambda$ determined by the Moré–Thuente line search satisfies the strong Wolfe conditions. If $\{x_k\}$ has a subsequence $\{x_{k_j}\}$ such that $F'(x_{k_j})$ is nonsingular for each $j$ and $\{\|F'(x_{k_j})^{-1}\|\}$ is bounded, then $F(x_k) \to 0$. If $\{x_k\}$ has a limit point $x_*$ such that $F'(x_*)$ is nonsingular, then $F(x_*) = 0$ and $x_k \to x_*$. 
Theorem: Suppose that \( x_0 \) is given and \( F \) is Lipschitz continuously differentiable on \( \mathcal{L}(x_0) \equiv \{ x : \| F(x) \| \leq \| F(x_0) \| \} \). Assume that \( \{ x_k \} \) is produced by the INMTL method such that, for each \( k \), the \( \lambda \) determined by the Moré–Thuente linesearch satisfies the strong Wolfe conditions. If \( \{ x_k \} \) has a subsequence \( \{ x_{k_j} \} \) such that \( F'(x_{k_j}) \) is nonsingular for each \( j \) and \( \{ \| F'(x_{k_j})^{-1} \| \} \) is bounded, then \( F(x_k) \to 0 \). If \( \{ x_k \} \) has a limit point \( x_* \) such that \( F'(x_*) \) is nonsingular, then \( F(x_*) = 0 \) and \( x_k \to x_* \).

Lemma: Suppose that \( \{ x_k \} \) produced by the INMTL method converges to \( x_* \) such that \( F(x_*) = 0 \) and \( F'(x_*) \) is nonsingular. Then the Wolfe \( \alpha \)-condition holds with \( \lambda = 1 \) for all sufficiently large \( k \) if \( \alpha < \frac{1 - \lim_{k \to \infty} \eta_k}{2} \) and only if \( \alpha < \frac{1}{2(1 - \liminf_{k \to \infty} \eta_k)} \). Additionally, the Wolfe curvature condition holds for all sufficiently large \( k \) if \( \beta > \frac{\lim_{k \to \infty} \eta_k (1 + \lim_{k \to \infty} \eta_k)}{1 - \lim_{k \to \infty} \eta_k} \).
For the **dogleg method**, make the straightforward extension:

- $\|F(x) + F'(x) s^{IN}\| \leq \eta \|F(x)\|$ 
- $\Gamma^{DL}: 0 \rightarrow s^{CP} \rightarrow s^{IN}$. 
The \textbf{dogleg method} is \ldots

\begin{boxedminipage}{\textwidth}
\textbf{Inexact Newton Dogleg (INDL) Method}

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

Iterate:

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

$$\|\mathbf{F}(x) + \mathbf{F}'(x)s^{IN}\| \leq \eta \|\mathbf{F}(x)\|.$$

Evaluate $s^{CP}$ and determine $s \in \Gamma_{DL}$.

While $\text{ared} < t \cdot \text{pred}$ do:

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

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

Redetermine $s \in \Gamma_{DL}$.

Update $x \leftarrow x + s$ and update $\delta$.
\end{boxedminipage}
• Sufficient decrease is based on the inexact Newton condition and

\[ \text{ared} \equiv \|F(x)\| - \|F(x + s)\| \quad \text{(actual reduction)} \]

\[ \text{pred} \equiv \|F(x)\| - \|F(x) + F'(x)s\| \quad \text{("predicted" reduction)} \]

• Determine \( s \in \Gamma^{DL} \) by the "standard strategy":

\( \triangleright \) If \( \|s^{IN}\| \leq \delta \), then \( s = s^{IN} \);

\( \triangleright \) else, if \( \|s^{CP}\| \geq \bar{\delta} \), then \( s = (\bar{\delta}/\|s^{CP}\|) s^{CP} \);

\( \triangleright \) else, \( s = (1 - \tau)s^{CP} + \tau s^{IN} \), where \( \tau \in (0, 1) \) is uniquely determined so that \( \|s\| = \delta \).
Recall: $x$ is a **stationary point** of $\|F\|$ $\iff$ $\|F(x)\| \leq \|F(x) + F'(x)\ s\|$ for all $s$.

**Theorem:** Assume $F$ is continuously differentiable. If $x_*$ is a limit point of $\{x_k\}$, then $x_*$ is a stationary point of $\|F\|$. If additionally $F'(x_*)$ is nonsingular, then $F(x_*) = 0$ and $x_k \to x_*$; furthermore, $s_k = s_k^{IN}$ for all sufficiently large $k$. 
Numerical experiments.

- **Goal**: To compare the effectiveness of these globalizations.
- **Test problems**: Three benchmark flow problems in 2D and 3D.
- **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 methods with these globalizations were implemented in the Sandia NOX nonlinear solver suite. The GMRES routine and domain-based (overlapping Schwarz) ILU preconditioners were 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.
The test problems (2D versions).

Thermal Convection Problem

Lid Driven Cavity Problem

Backward-Facing Step Problem
A **robustness study.** The table shows *numbers of failures.*

<table>
<thead>
<tr>
<th>Method</th>
<th>Forcing Term</th>
<th>2D Problems</th>
<th>3D Problems</th>
<th>All Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backtracking, Quadratic Only</td>
<td>Ch. 1</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>10^{-4}</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Backtracking, Quadratic/Cubic</td>
<td>Ch. 1</td>
<td>1</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>10^{-4}</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Moré–Thuente Linesearch</td>
<td>Ch. 1</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>10^{-4}</td>
<td>9</td>
<td>2</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Dogleg</td>
<td>Ch. 1</td>
<td>6</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>10^{-4}</td>
<td>10</td>
<td>0</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Full Step</td>
<td>Ch. 1</td>
<td>15</td>
<td>33</td>
<td>4</td>
</tr>
<tr>
<td>10^{-4}</td>
<td>18</td>
<td>10</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>
An **efficiency study.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Forcing Term</th>
<th>Inexact Newton Steps</th>
<th>Backtracks per INS</th>
<th>GMRES Iterations per INS</th>
<th>Normalized Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backtracking, Quadratic/Cubic</td>
<td>Ch. 1</td>
<td>16.11</td>
<td>0.14</td>
<td>55.04</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>9.23</td>
<td>0.21</td>
<td>146.94</td>
<td>1.01</td>
</tr>
<tr>
<td>Backtracking, Quadratic Only</td>
<td>Ch. 1</td>
<td>16.17</td>
<td>0.14</td>
<td>55.29</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>9.25</td>
<td>0.18</td>
<td>146.99</td>
<td>1.0 (REF)</td>
</tr>
<tr>
<td>Moré–Thuente Linesearch</td>
<td>Ch. 1</td>
<td>14.99</td>
<td>0.16</td>
<td>59.60</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>8.56</td>
<td>0.17</td>
<td>146.62</td>
<td>0.93</td>
</tr>
<tr>
<td>Dogleg</td>
<td>Ch. 1</td>
<td>16.18</td>
<td>NA</td>
<td>65.28</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>10.40</td>
<td>NA</td>
<td>151.91</td>
<td>1.13</td>
</tr>
</tbody>
</table>
Concluding observations.

- These globalizations have good theoretical support and are practically effective, especially with adaptive forcing terms.
**Concluding observations.**

- These globalizations have good theoretical support and are practically effective, especially with adaptive forcing terms.

- Causes of failure in our experiments:
  - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 20/20 dogleg failures.
  - Globalization failure: 7/33 backtracking/linesearch failures.
**Concluding observations.**

- These globalizations have good theoretical support and are practically effective, especially with adaptive forcing terms.

- Causes of failure in our experiments:
  - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 20/20 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.
Concluding observations.

- These globalizations have good theoretical support and are practically effective, especially with adaptive forcing terms.

- Causes of failure in our experiments:
  - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 20/20 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.
  - ... but ... 

- No globalization or choice of forcing terms is always best.
Concluding observations.

- These globalizations have good theoretical support and are practically effective, especially with adaptive forcing terms.

- Causes of failure in our experiments:
  - Fatal near-stagnation: 26/33 backtracking/linesearch failures; 20/20 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.

  ... but ...

- No globalization or choice of forcing terms is always best.

- Many factors contribute to success: problem formulation, discretization, preconditioning, variable scaling, accuracy, ...