Convergence of Composed Nonlinear Iterations

Matthew Knepley

Computational and Applied Mathematics
Rice University

SPPEXA Symposium 2016
Leibniz Supercomputing Centre
München, Freistaat Bayern January 25, 2016

Left Nonlinear Preconditioning

- Nonlinearly preconditioned inexact Newton algorithms, Cai and D. E. Keyes, SISC, 2002.
- A parallel nonlinear additive Schwarz preconditioned inexact Newton algorithm for incompressible Navier-Stokes equations, Hwang, Cai, J. Comp. Phys., 2005.
- Field-Split Preconditioned Inexact Newton Algorithms, Liu, Keyes, SISC, 2015.

Right Nonlinear Preconditioning

- A parallel two-level domain decomposition based one-shot method for shape optimization problems, Chen, Cai, IJNME, 2014.
- Nonlinearly preconditioned optimization on Grassman manifolds for computing approximate Tucker tensor decompositions, De Sterck, Howse, SISC, 2015.
- Nonlinear FETI-DP and BDDC Methods, Klawonn, Lanser, Rheinbach, SISC, 2014.

Algorithmic Formalism

- Composing Scalable Nonlinear Algebraic Solvers, Brune, Knepley, Smith, Tu, SIAM Review, 2015.

Type	Sym	Statement	Abbreviation
Additive	$+$	$\mathbf{x} + \alpha(\mathcal{M}(\mathcal{F}, \mathbf{x}, \mathbf{b}) - \mathbf{x})$ $+ \beta(\mathcal{N}(\mathcal{F}, \mathbf{x}, \mathbf{b}) - \mathbf{x})$	$\mathcal{M} + \mathcal{N}$
Multiplicative	$*$	$\mathcal{M}(\mathcal{F}, \mathcal{N}(\mathcal{F}, \mathbf{x}, \mathbf{b}), \mathbf{b})$	$\mathcal{M} * \mathcal{N}$
Left Prec.	$-_L$	$\mathcal{M}(\mathbf{x} - \mathcal{N}(\mathcal{F}, \mathbf{x},), \mathbf{x}, \mathbf{b})$	$\mathcal{M} -_L \mathcal{N}$
Right Prec.	$-_R$	$\mathcal{M}(\mathcal{F}(\mathcal{N}(\mathcal{F}, \mathbf{x}, \mathbf{b})), \mathbf{x}, \mathbf{b})$	$\mathcal{M} -_R \mathcal{N}$
Inner Lin. Inv.	$\backslash$	$\mathbf{y} = \vec{J}(\mathbf{x})^{-1}\vec{r}(\mathbf{x}) = \mathrm{K}(\vec{J}(\mathbf{x}), \mathbf{y}_0, \mathbf{b})$	$\mathcal{N} \backslash \mathrm{K}$

How Helpful is Theory?

Consider Linear Multigrid,

- Local Fourier Analysis (LFA)
  - Multi-level adaptive solutions to boundary-value problems, Brandt, Math. Comp., 1977.
- Idealized Relaxation (IR)
  Idealized Coarse-Grid Correction (ICG)
  - On Quantitative Analysis Methods for Multigrid Solutions, Diskin, Thomas, Mineck, SISC, 2005.

How Helpful is Theory?

How about Nonlinear Multigrid?

- Full Approximation Scheme (FAS)
  - Convergence of the multigrid full approximation scheme for a class of elliptic mildly nonlinear boundary value problems,
    Reusken, Num. Math., 1987.
  - Analysis only for Picard
- Overbroad conclusions based on experiments
  - Nonlinear Multigrid Methods for Second Order Differential Operators with Nonlinear Diffusion Coefficient,
    Brabazona, Hubbard, Jimack, Comp. Math. App., 2014.
- People feel helpless when it fails or stagnates

How Helpful is Theory?

How about Newton's Method?

- We have an asymptotic theory
  - *On Newton's Method for Functional Equations*, Kantorovich, Dokl. Akad. Nauk SSSR, 1948.
- We need a non-asymptotic theory
  - The Rate of Convergence of Newton's Process, Ptak, Num. Math., 1976.
- People feel helpless when it fails or stagnates

How Helpful is Theory?

How about Nonlinear Preconditioning?

- Some guidance
  - *Nonlinear Preconditioning Techniques for Full-Space Lagrange-Newton Solution of PDE-Constrained Optimzation Problems*,
    Yang, Hwang, Cai, SISC, to appear.
- Left preconditioning (Newton $-_L$ NASM) handles local nonlinearities
- Right preconditioning (Nonlinear Elimination) handles nonlinear global coupling

Outline

Rate of Convergence

What should be a Rate of Convergence? [Ptak, 1977]:

1. It should relate quantities which may be measured or estimated during the actual process
2. It should describe accurately in particular the initial stage of the process, not only its asymptotic behavior . . .

$$\|x_{n+1} - x^*\| \leq c\|x_n - x^*\|^q$$

Rate of Convergence

What should be a Rate of Convergence? [Ptak, 1977]:

1. It should relate quantities which may be measured or estimated during the actual process
2. It should describe accurately in particular the initial stage of the process, not only its asymptotic behavior . . .

$$\|x_{n+1} - x_n\| \leq c\|x_n - x_{n-1}\|^q$$

Rate of Convergence

What should be a Rate of Convergence? [Ptak, 1977]:

1. It should relate quantities which may be measured or estimated during the actual process
2. It should describe accurately in particular the initial stage of the process, not only its asymptotic behavior ...

$$\|x_{n+1} - x_n\| \leq \omega(\|x_n - x_{n-1}\|)$$

where we have for all $r \in (0, R]$

$$\sigma(r) = \sum_{n=0}^{\infty} \omega^{(n)}(r) < \infty$$

Nondiscrete Induction

Define an approximate set $Z(r)$, where $x^* \in Z(0)$ implies $f(x^*) = 0$.

Nondiscrete Induction

Define an approximate set $Z(r)$, where $x^* \in Z(0)$ implies $f(x^*) = 0$.

For Newton's method, we use

$$Z(r) = \left\{ x \middle| \|f'(x)^{-1} f(x)\| \leq r, d(f'(x)) \geq h(r), \|x - x_0\| \leq g(r) \right\},$$

where

$$d(A) = \inf_{\|x\| \geq 1} \|Ax\|,$$

and $h(r)$ and $g(r)$ are positive functions.

Nondiscrete Induction

Define an approximate set $Z(r)$, where $x^* \in Z(0)$ implies $f(x^*) = 0$.

For $r \in (0, R]$,

$$Z(r) \subset U(Z(\omega(r)), r)$$

implies

$$Z(r) \subset U(Z(0), \sigma(r)).$$

Nondiscrete Induction

For the fixed point iteration

$$x_{n+1} = G x_n,$$

if I have

$$x_0 \in Z(r_0)$$

and for $x \in Z(r)$,

$$\begin{aligned} \|Gx - x\| &\leq r \\ Gx &\in Z(\omega(r)) \end{aligned}$$

then

Nondiscrete Induction

For the fixed point iteration

$$x_{n+1} = G x_n,$$

if I have

$$x_0 \in Z(r_0)$$

and for $x \in Z(r)$,

$$\|Gx - x\| \le r$$
$$Gx \in Z(\omega(r))$$

then

$$x^* \in Z(0)$$
$$x_n \in Z(\omega^{(n)}(r_0))$$

Nondiscrete Induction

For the fixed point iteration

$$x_{n+1} = G x_n,$$

if I have

$$x_0 \in Z(r_0)$$

and for $x \in Z(r)$,

$$\begin{aligned} \|Gx - x\| &\leq r \\ Gx &\in Z(\omega(r)) \end{aligned}$$

then

$$\begin{aligned} \|x_{n+1} - x_n\| &\leq \omega^{(n)}(r_0) \\ \|x_n - x^*\| &\leq \sigma(\omega^{(n)}(r_0)) \end{aligned}$$

Nondiscrete Induction

For the fixed point iteration

$$x_{n+1} = G x_n,$$

if I have

$$x_0 \in Z(r_0)$$

and for $x \in Z(r)$,

$$\begin{aligned} \|Gx - x\| &\leq r \\ Gx &\in Z(\omega(r)) \end{aligned}$$

then

$$\begin{aligned} \|x_n - x^*\| &\leq \sigma(\omega(\|x_n - x_{n-1}\|)) \\ &= \sigma(\|x_n - x_{n-1}\|) - \|x_n - x_{n-1}\| \end{aligned}$$

Newton's Method

$$\omega_{\mathcal{N}}(r) = cr^2$$

Newton's Method

$$\omega_{\mathcal{N}}(r) = \frac{r^2}{2\sqrt{r^2 + a^2}}$$
$$\sigma_{\mathcal{N}}(r) = r + \sqrt{r^2 + a^2} - a$$

where

$$a = \frac{1}{k_0}\sqrt{1 - 2k_0 r_0},$$

k_0 is the (scaled) Lipschitz constant for f', and
r_0 is the (scaled) initial residual.

Newton's Method

$$\omega_{\mathcal{N}}(r) = \frac{r^2}{2\sqrt{r^2 + a^2}}$$
$$\sigma_{\mathcal{N}}(r) = r + \sqrt{r^2 + a^2} - a$$

This estimate is *tight* in that the bounds hold with equality for some function f,

$$f(x) = x^2 - a^2$$

using initial guess

$$x_0 = \frac{1}{k_0}.$$

Also, if equality is attained for some n_0, this holds for all $n \geq n_0$.

Newton's Method

$$\omega_{\mathcal{N}}(r) = \frac{r^2}{2\sqrt{r^2 + a^2}}$$
$$\sigma_{\mathcal{N}}(r) = r + \sqrt{r^2 + a^2} - a$$

If $r \gg a$, meaning we have an inaccurate guess,

$$\omega_{\mathcal{N}}(r) \approx \frac{1}{2} r,$$

whereas if $r \ll a$, meaning we are close to the solution,

$$\omega_{\mathcal{N}}(r) \approx \frac{1}{2a} r^2.$$

Left vs. Right

Left:

$$\mathcal{F}(x) \Longrightarrow x - \mathcal{N}(\mathcal{F}, x, b)$$

Right:

$$x \Longrightarrow y = \mathcal{N}(\mathcal{F}, x, b)$$

Heisenberg vs. Schrödinger Picture

Left vs. Right

Left:

$$\mathcal{F}(x) \Longrightarrow x - \mathcal{N}(\mathcal{F}, x, b)$$

Right:

$$x \Longrightarrow y = \mathcal{N}(\mathcal{F}, x, b)$$

Heisenberg vs. Schrödinger Picture

$\mathcal{M} -_R \mathcal{N}$

We start with $x \in Z(r)$, apply $\mathcal{N}$ so that

$$y \in Z(\omega_{\mathcal{N}}(r)),$$

and then apply $\mathcal{M}$ so that

$$x' \in Z(\omega_{\mathcal{M}}(\omega_{\mathcal{N}}(r))).$$

Thus we have

$$\omega_{\mathcal{M} -_R \mathcal{N}} = \omega_{\mathcal{M}} \circ \omega_{\mathcal{N}}$$

Non-Abelian

$\mathcal{N} -_R$ NRICH

$$
\begin{aligned}
\omega_{\mathcal{N}} \circ \omega_{\text{NRICH}} &= \frac{1}{2}\frac{r^2}{\sqrt{r^2 + a^2}} \circ cr, \\
&= \frac{1}{2}\frac{c^2 r^2}{\sqrt{c^2 r^2 + a^2}}, \\
&= \frac{1}{2}\frac{c r^2}{\sqrt{r^2 + (a/c)^2}}, \\
&= \frac{1}{2} c \frac{r^2}{\sqrt{r^2 + \tilde{a}^2}},
\end{aligned}
$$

Non-Abelian

$$\mathcal{N} -_R \text{NRICH: } \frac{1}{2}c\frac{r^2}{\sqrt{r^2+\tilde{a}^2}}$$

NRICH $-_R \mathcal{N}$

$$\begin{aligned}
\omega_{\text{NRICH}} \circ \omega_{\mathcal{N}} &= cr \circ \frac{1}{2}\frac{r^2}{\sqrt{r^2+a^2}}, \\
&= \frac{1}{2}c\frac{r^2}{\sqrt{r^2+a^2}}, \\
&= \frac{1}{2}c\frac{r^2}{\sqrt{r^2+a^2}}.
\end{aligned}$$

Non-Abelian

$$\mathcal{N} -_R \text{NRICH}: \tfrac{1}{2}c\frac{r^2}{\sqrt{r^2+\tilde{a}^2}}$$

$$\text{NRICH} -_R \mathcal{N}: \tfrac{1}{2}c\frac{r^2}{\sqrt{r^2+a^2}}$$

The first method also changes the onset of second order convergence.

Outline

Non-Abelian

$\mathcal{N} -_R$ NRICH

$$
\begin{aligned}
\omega_{\mathcal{N}} \circ \omega_{\mathrm{NRICH}} &= \frac{1}{2} \frac{r^2}{\sqrt{r^2 + a^2}} \circ cr, \\
&= \frac{1}{2} \frac{c^2 r^2}{\sqrt{c^2 r^2 + a^2}}, \\
&= \frac{1}{2} \frac{c r^2}{\sqrt{r^2 + (a/c)^2}}, \\
&= \frac{1}{2} c \frac{r^2}{\sqrt{r^2 + \tilde{a}^2}},
\end{aligned}
$$

Non-Abelian

$$\mathcal{N} -_R \text{NRICH: } \frac{1}{2}c\frac{r^2}{\sqrt{r^2+\tilde{a}^2}}$$

NRICH $-_R \mathcal{N}$

$$\begin{aligned}
\omega_{\text{NRICH}} \circ \omega_{\mathcal{N}} &= cr \circ \frac{1}{2}\frac{r^2}{\sqrt{r^2+a^2}}, \\
&= \frac{1}{2}c\frac{r^2}{\sqrt{r^2+a^2}}, \\
&= \frac{1}{2}c\frac{r^2}{\sqrt{r^2+a^2}}.
\end{aligned}$$

Non-Abelian

$$\mathcal{N} -_R \text{NRICH}: \frac{1}{2}c\frac{r^2}{\sqrt{r^2+\tilde{a}^2}}$$

$$\text{NRICH} -_R \mathcal{N}: \frac{1}{2}c\frac{r^2}{\sqrt{r^2+a^2}}$$

The first method also changes the onset of second order convergence.

Composed Rates of Convergence

Theorem

If ω_1 *and* ω_2 *are convex rates of convergence, then* $\omega = \omega_1 \circ \omega_2$ *is a rate of convergence.*

Composed Rates of Convergence

Theorem

If ω_1 and ω_2 are convex rates of convergence, then $\omega = \omega_1 \circ \omega_2$ is a rate of convergence.

First we show that

$$\omega(s) \leq \frac{s}{r}\omega(r),$$

which means that convex rates of convergence are non-decreasing.

This implies that compositions of convex rates of convergence are also convex and non-decreasing.

Composed Rates of Convergence

Theorem

If ω_1 and ω_2 are convex rates of convergence, then $\omega = \omega_1 \circ \omega_2$ is a rate of convergence.

Then we show that

$$\omega(r) < r \qquad \forall r \in (0, R)$$

by contradiction.

Composed Rates of Convergence

Theorem

If ω_1 and ω_2 are convex rates of convergence, then $\omega = \omega_1 \circ \omega_2$ is a rate of convergence.

This is enough to show that

$$\omega_1(\omega_2(r)) < \omega_1(r),$$

and in fact

$$(\omega_1 \circ \omega_2)^{(n)}(r) < \omega_1^{(n)}(r).$$

Multidimensional Induction Theorem

Preconditions

Theorem

Let

- p *(1 for our case) and* m *(2 for our case) be two positive integers,*
- X *be a complete metric space and* $D \subset X^p$*,*
- $G : D \to X^p$ *and* $F : D \to X^{p+1}$ *be defined by* $Fu = (u, Gu)$*,*
- $F_k = P_k F$*,* $-p + 1 \leq k \leq m$*, the components of* F*,*
- $P = P_m$*,*
- $Z(r) \subset D$ *for each* $r \in T^p$*,*
- ω *be a rate of convergence of type* (p, m) *on* T*,*
- $u_0 \in D$ *and* $r_0 \in T^p$*.*

Multidimensional Induction Theorem

Theorem

If the following conditions hold

$$
\begin{aligned}
u_0 &\in Z(r_0), \\
PFZ(r) &\subset Z(\tilde{\omega}(r)), \\
\|F_k u - F_{k+1} u\| &\leq \omega_k(r),
\end{aligned}
$$

for all $r \in T^p$*,* $u \in Z(r)$*, and* $k = 0, \ldots, m-1$*, then*

1. u_0 *is admissible, and* $\exists x^* \in X$ *such that* $(P_k u_n)_{n \geq 0} \to x^*$*,*
2. *and the following relations hold for* $n > 1$*,*

$$
\begin{aligned}
P u_n &\in Z(\tilde{\omega}(r_0)), \\
\|P_k u_n - P_{k+1} u_n\| &\leq \omega_k^{(n)}(r_0), \qquad 0 \leq k \leq m-1, \\
\|P_k u_n - x^*\| &\leq \sigma_k(\tilde{\omega}(r_0)) \qquad 0 \leq k \leq m;
\end{aligned}
$$

Multidimensional Induction Theorem

Theorem

If the following conditions hold

$$
\begin{aligned}
u_0 &\in Z(r_0), \\
PFZ(r) &\subset Z(\tilde{\omega}(r)), \\
\|F_k u - F_{k+1} u\| &\leq \omega_k(r),
\end{aligned}
$$

for all $r \in T^p$, $u \in Z(r)$, *and* $k = 0, \ldots, m-1$, *then*

1. u_0 *is admissible, and* $\exists x^* \in X$ *such that* $(P_k u_n)_{n \geq 0} \to x^*$,
2. *and the following relations hold for* $n > 1$,

$$
\|P_k u_n - x^*\| \leq \sigma_k(r_n), \qquad 0 \leq k \leq m.
$$

where $r_n \in T^p$ *and* $Pu_{n-1} \in Z(r_n)$.

Multidimensional Induction Theorem

Theorem

If the following conditions hold

$$
\begin{aligned}
u_0 &\in Z(r_0), \\
PFZ(r) &\subset Z(\tilde{\omega}(r)), \\
\|F_k u - F_{k+1} u\| &\leq \omega_k(r),
\end{aligned}
$$

for all $r \in T^p$*,* $u \in Z(r)$*, and* $k = 0, \ldots, m-1$*, then*

1. u_0 *is admissible, and* $\exists x^* \in X$ *such that* $(P_k u_n)_{n \geq 0} \to x^*$*,*
2. *and the following relations hold for* $n > 1$*,*

$$
\begin{aligned}
P u_n &\in Z(\tilde{\omega}(r_0)), \\
\|P_k u_n - P_{k+1} u_n\| &\leq \omega_k^{(n)}(r_0), \qquad 0 \leq k \leq m-1, \\
\|P_k u_n - x^*\| &\leq \sigma_k(\tilde{\omega}(r_0)) \qquad 0 \leq k \leq m;
\end{aligned}
$$

Multidimensional Induction Theorem

Theorem

If the following conditions hold

$$
\begin{aligned}
u_0 &\in Z(r_0), \\
PFZ(r) &\subset Z(\omega \circ \psi(r)), \\
\|F_0 u - F_1 u\| &\le r, \\
\|F_1 u - F_2 u\| &\le \psi(r),
\end{aligned}
$$

for all $r \in T^p$, $u \in Z(r)$, and $k = 0, \ldots, m-1$, then

1. *u_0 is admissible, and $\exists x^* \in X$ such that $(P_k u_n)_{n \ge 0} \to x^*$,*
2. *and the following relations hold for $n > 1$,*

$$
\begin{aligned}
Pu_n &\in Z(\tilde\omega(r_0)), \\
\|P_k u_n - P_{k+1} u_n\| &\le \omega^{(n)}_k(r_0), \qquad 0 \le k \le m-1, \\
\|P_k u_n - x^*\| &\le \sigma_k(\tilde\omega(r_0)) \qquad 0 \le k \le m;
\end{aligned}
$$

Composed Newton Methods

Theorem

Suppose that we have two nonlinear solvers

- $\mathcal{M}$, Z_1, ω,
- $\mathcal{N}$, Z_0, ψ,

and consider $\mathcal{M} -_R \mathcal{N}$*, meaning a single step of* $\mathcal{N}$ *for each step of* $\mathcal{M}$*.*

Concretely, take $\mathcal{M}$ *to be the Newton iteration, and* $\mathcal{N}$ *the Chord method. Then the assumptions of the theorem above are satisfied using* $Z = Z_1$ *and*

$$\omega(r) = \{\psi(r), \omega \circ \psi(r)\},$$

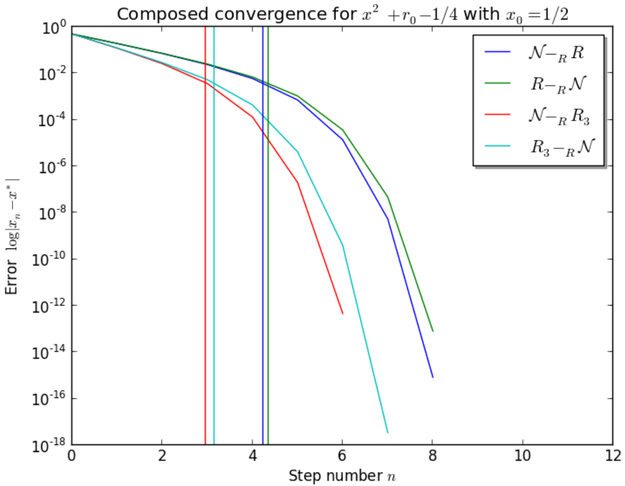
giving us the existence of a solution, and both a priori and a posteriori bounds on the error.

Example

$$f(x) = x^2 + (0.0894427)^2$$

n	$\|x_{n+1} - x_n\|$	$\|x_{n+1} - x_n\| - w^{(n)}(r_0)$	$\|x_n - x^*\| - s(w^{(n)}(r_0))$
0	1.9990e+00	$< 10^{-16}$	$< 10^{-16}$
1	9.9850e-01	$< 10^{-16}$	$< 10^{-16}$
2	4.9726e-01	$< 10^{-16}$	$< 10^{-16}$
3	2.4470e-01	$< 10^{-16}$	$< 10^{-16}$
4	1.1492e-01	$< 10^{-16}$	$< 10^{-16}$
5	4.5342e-02	$< 10^{-16}$	$< 10^{-16}$
6	1.0251e-02	$< 10^{-16}$	$< 10^{-16}$
7	5.8360e-04	$< 10^{-16}$	$< 10^{-16}$
8	1.9039e-06	$< 10^{-16}$	$< 10^{-16}$
9	2.0264e-11	$< 10^{-16}$	$< 10^{-16}$
10	0.0000e+00	$< 10^{-16}$	$< 10^{-16}$

Example



Matrix iterations also 1D scalar once you diagonalize

Pták's nondiscrete induction and its application to matrix iterations, Liesen, IMA J. Num. Anal., 2014.

Conclusions

Nonlinear Preconditioning
is a powerful technique,
but we need more
theoretical guidance,
algorithmic structure,
and rules of thumb.

Conclusions

Nonlinear Preconditioning
is a powerful technique,
but we need more
theoretical guidance,
algorithmic structure,
and rules of thumb.

Conclusions

Nonlinear Preconditioning
is a powerful technique,
but we need more
theoretical guidance,
algorithmic structure,
and rules of thumb.

Conclusions

Nonlinear Preconditioning
is a powerful technique,
but we need more
theoretical guidance,
algorithmic structure,
and rules of thumb.