Anderson accelerated Newton's method near singular points: theory and implementation

Applied and Computational Mathematics Seminar
University of North Texas

DMS 2011519

Matt Dallas

Dept. of Mathematics
University of Dallas

Joint work with:
Sara Pollock (U FL) and
Leo Rebholz (Clemson)

April 11,2025

UNIVERSITY of DALLAS | MATHEMATICS DEPARTMENT

Introduction

Newton's Method

Anderson Acceleration (AA)

Adaptive Safeguarding

Numerical Examples

Conclusion

Introduction

We're interested in efficiently solving

$$f(x) = 0$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is nonlinear, C^3, and $f(x^*) = 0 \implies f'(x^*)$ is singular. Any such solution x^* is called a **singular point**.

Why nonlinear?

- Nonlinear Integral Equations

Chandrasekhar H-equation

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

- Nonlinear Partial Differential Equations
  - The Wikipedia page titled "List of nonlinear partial differential equations" lists **103 PDEs**.
  - Many of these have an entire Wikipedia page of their own. For example:

Incompressible Navier-Stokes

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} - \Delta \mathbf{u} + \nabla p - \mathbf{g} = 0$$

$$\nabla \cdot \mathbf{u} = 0$$

Minimal Surface Equation

$$\nabla \cdot \left(Du / \sqrt{1 + |Du|^2} \right) = 0$$

Why $\mathbb{R}^n$?

Nonlinear Equation $\xrightarrow{\textit{Discretize}}$ Nonlinear function on $\mathbb{R}^n$

Finite elements, *finite difference, collocation, quadrature, etc. .*

Remark: one may analyze $f(x) = 0$ where f is a function between Banach spaces by taking $f'(x)$ to be the Fréchet derivative. A suitable generalization for the singular problem is to assume $f'(x)$ is a Fredholm operator of index zero [DKK83].

Why Singular Problems?

Figure: A real life bifurcation post earthquake. **Bifurcation point are necessarily singular points**. Image taken from December 2020 AMS Notices [BP20].

A standard nonlinear solver: Newton's method

$$x_{k+1} = x_k - f'(x_k)^{-1} f(x_k) \tag{1}$$

Local quadratic convergence when $f'(x^*)$ is nonsingular [Ort].

Local linear convergence when $f'(x^*)$ is singular [DKK83].

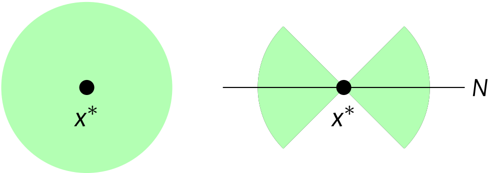

Figure: **Left**: Domain of convergence for Newton's method when $f'(x^*)$ is **nonsingular**. **Right**: Example domain of convergence when $f'(x^*)$ is **singular**, and $N = \text{null}\, f'(x^*)$.

Improving Newton's method for singular problems

Richardson extrapolation and overrelaxtion (Griewank, SIAM Review, 1985 [Gri85]) can achieve arbitrarily fast linear convergence and superlinear convergence respectively if the order of the root is know or approximately known (bordering is also useful).

An alternative to Newton's method for singular problems is the Levenberg-Marquardt (LM) method, which achieves local quadratic convergence [KYF04] if the local error bound holds:

$$\operatorname{dist}\left(x_k, f^{-1}(0)\right) \leq C\|f(x_k)\|.$$

In the absence of the local error bound, one may assume f is 2-regular at x^* along a direction $v \in N$, i.e., the linear map

$$\Phi(\cdot) := f'(x^*)(\cdot) + P_N f''(x^*)(v, \cdot)$$

is nonsingular, in which case LM converges linearly in a starlike domain around x^*.

What about Anderson?

Anderson Acceleration accelerates convergence of linearly converging fixed-point iterations [PR21], e.g., Newton's method applied to a singular problem, by improving the rate of convergence.

Can enlarge domain of convergence (observed, for example, by Pollock and Schwartz in [PS20]).

For singular problems in particular, no knowledge of the order of the root required, it's relatively cheap, and is theoretically supported under the assumption of 2-regularity.

It has been shown by Izmailov, Kurennoy, and Solodov in [IKS18] that under modest conditions[1]

$$\text{2-regular} \implies \textbf{no}\text{ local error bound}$$

Thus Anderson can recover superlinear convergence when competitors like LM cannot.

[1] The solution set is Clarke regular, and f is *strictly differentiable* at solution x^* with respect to solution set.

Anderson Acceleration (AA)

There are several forms Anderson acceleration can take, and different authors express different preferences. Each form is derived from the same starting point.

Consider the problem of computing a fixed point x^* of the function $g : \mathbb{R}^n \to \mathbb{R}^n$ with the fixed point iteration $x_{k+1} = g(x_k)$. Define $w_{k+1} = g(x_k) - x_k$ as the residual at step k.

Anderson Acceleration (AA)

Select x_0, and set $x_1 = g(x_0)$. Set an algorithmic depth $m \geq 0$, $w_1 = g(x_0) - x_0$, and $k = 1$.
Until convergence, do

1. Set $m_k = \min\{k, m\}$.
2. Take $\alpha_{k-m_k}^{(k+1)}, \ldots, \alpha_k^{(k+1)}$ that solve the

$$\min_{\sum_{j=k-m_k}^{k} \alpha_j = 1} \left\| \sum_{j=k-m_k}^{k} \alpha_j w_{j+1} \right\| \tag{2}$$

3. With damping factor $0 < \beta_k \leq 1$, set

$$x_{k+1}^{AA} = \sum_{j=k-m_k}^{k} \alpha_j^{(k+1)} x_j + \beta_k \sum_{j=k-m_k}^{k} \alpha_j^{(k+1)} w_{j+1}. \tag{3}$$

4. Set $k = k + 1$.

Remark: The norm $\|\cdot\|$ in step 2 is frequently taken to be the 2-norm, and in this talk $\|\cdot\|$ is always the 2-norm, but the 1-norm and and ∞-norm have also been studied [TK15].

Anderson Acceleration (AA)

In what follows, we will take $\beta_k = 1$ for all k. Since $\alpha_k^{(k+1)} = 1 - \sum_{j=k-m_k}^{k-1} \alpha_j^{(k+1)}$, the optimization subproblem is equivalent to the unconstrained problem

$$\min_{\alpha \in \mathbb{R}^{m_k}} \left\| w_{k+1} - \sum_{j=k-m_k}^{k-1} \alpha_j (w_{k+1} - w_{j+1}) \right\| \tag{4}$$

Similarly, the Anderson iterate x_{k+1} can be written as

$$\begin{aligned} x_{k+1}^{AA} &= x_k + w_{k+1} - \sum_{j=k-m_k}^{k-1} \alpha_j^{(k+1)} [x_k + w_{k+1} - (x_j + w_{j+1})] \\ &= g(x_k) - \sum_{j=k-m_k}^{k-1} \alpha_j^{(k+1)} [g(x_k) - g(x_j)] \end{aligned} \tag{5}$$

Equation (5) is one form of the Anderson update seen in the literature.

Anderson Acceleration (AA)

Another common form is obtained from (5) by the following.

$$
\begin{aligned}
x_{k+1}^{AA} &= g(x_k) - \sum_{j=k-m_k}^{k-1} \alpha_j^{(k+1)} [g(x_k) - g(x_j)] \\
&= g(x_k) - \sum_{j=k-m_k}^{k-1} \alpha_j^{(k+1)} \sum_{n=j+1}^{k} [g(x_n) - g(x_{n-1})] \\
&= g(x_k) - \sum_{n=k-m_k+1}^{k} \gamma_n^{(k+1)} [g(x_n) - g(x_{n-1})], \qquad (6)
\end{aligned}
$$

where $\gamma_n^{(k+1)} = \sum_{j=k-m_k}^{n-1} \alpha_j^{(k+1)}$. Equation (6) is equivalent to a third form seen in the literature, but it is typically written in terms of the residual. This is shown in the next slide.

Anderson Acceleration (AA)

Using $w_{k+1} = g(x_k) - x_k$, (6) becomes

$$x_{k+1} = x_k + w_{k+1} - \sum_{n=k-m_k+1}^{k} \gamma_n^{(k+1)} \left[x_n - x_{n-1} + w_{n+1} - w_n \right].$$

Define the $n \times m_k$ matrices

$$E_k = \Big((x_k - x_{k-1}) \cdots (x_{k-m_k+1} - x_{k-m_k}) \Big) \qquad F_k = \Big((w_{k+1} - w_k) \cdots (w_{k-m_k+2} - w_{k-m_k+1}) \Big),$$

and $\gamma^{(k+1)} = \left(\gamma_k^{(k+1)}, \ldots, \gamma_{k-m_k+1}^{(k+1)} \right)^T$ where $\gamma^{(k+1)} = \operatorname{argmin}_{\gamma \in \mathbb{R}^{m_k}} \| w_{k+1} - F_k \gamma \|$. Then the $(k+1)st$ Anderson update may be written as

$$x_{k+1}^{AA} = x_k + w_{k+1} - (E_k + F_k)\gamma^{(k+1)}. \tag{7}$$

AA milestones

(1965) Introduced by D.G. Anderson.

(1980) The closely related method, DIIS or Pulay Mixing, is introduced by Peter Pulay in *Convergence acceleration of iterative sequences. The case of SCF iteration*.

(2009) Fang and Saad prove that AA is a type of multisecant method in *Two classes of multisecant methods for nonlinear acceleration*.

(2011) Walker and Ni show that, for linear problems, AA is equivalent to the well-known GMRES method in *Anderson Acceleration for Fixed-Point Iterations*.

(2015) Toth and Kelley provide first convergence proof of Anderson for contractive operators.

(2020) Evans, Pollock, Rebholz, and Xiao prove that Anderson improves rate of convergence for linearly convergent fixed-point iterations.

Open prior to 2023: why is Anderson effective when applied to **singular problems**?
Convergence? Answers in next section.

AA as a quasi-Newton method

Fang and Saad's 2009 paper [FS08] demonstrates the x_{k+1}^{AA} can be viewed as a Type-II[2] Broyden (multisecant) method. That is,

$$x_{k+1}^{AA} = x_k - G_k w_{k+1}, \tag{8}$$

where $G_k F_k = E_k$. Noting that $\gamma^{(k+1)} = (F_k^T F_k)^{-1} F_k^T w_{k+1}$ in (6), it follows that

$$G_k = -I + (E_k + F_k)(F_k^T F_k)^{-1} F_k^T, \tag{9}$$

and G_k minimizes $\|G + I\|_F$ over all $n \times m_k$ matrices for which $GF_k = E_k$ [FS08, Gri12]. Here $\|\cdot\|_F$ is the Frobenius norm.

[2] Type-II multisecant methods approximate the inverse Jacobian directly.

AA as nonlinear GMRES

Walker and Ni's 2011 paper [WN11] showed that if *AA* is applied to an affine function $g(x) = Ax + b$ with $m_k = k$ for all k, then it is equivalent to GMRES in the sense that

$$\sum_{i=0}^{k} \alpha_k^{(k+1)} x_i^{AA} = x_k^{GMRES}, \quad \text{and} \tag{10}$$

$$x_{k+1}^{AA} = g(x_k^{GMRES}) \tag{11}$$

Here, the $\alpha_i^{(k+1)}$ terms are those obtained from the optimization step of the AA algorithm.

Thus, GMRES and AA are equivalent in the sense that the GMRES iterates are easily obtained from the AA iterates and vice versa.

AA theory for nonsingular problems

- If $\|g(x) - g(y)\| \leq \kappa_g \|x - y\|$, $\|g'(x)z - g'(y)z\| \leq \hat{\kappa}_g \|x - y\| \|z\|$, $\|w_{k+1} - w_k\| \geq \sigma \|x_k - x_{k-1}\|$, and $m = 1$, then the residual w_{k+1} satisfies [PR21, Theorem 4.3]

$$\|w_{k+1}\| \leq \theta_k \kappa_g \|w_k\| + \text{h.o.t} \quad (12)$$

  - $\theta_k := \|w_{k+1} - F_k \gamma^{(k+1)}\| / \|w_{k+1}\| \leq 1$ is the **optimization gain**.
  - AA works because $\theta_k \leq 1$, and is frequently much less than one in practice. Thus AA decreasing the Lipschitz constant κ_g.
  - To control the condition number of $F_k^T F_k$, one may use filtering [PR23].
  - $\|w_{k+1} - w_k\| \geq \sigma \|x_k - x_{k-1}\|$ holds if, for example, $f'(x^*)$ is nonsingular, and is essentially a regularity condition on $w_{k+1} = w(x) := -f'(x)^{-1} f(x)$. Such a condition makes sense given that AA can be viewed as a quasi-Newton method applied to $w(x)$.

The singular case

Consider again the problem computing the solution x^* to the nonlinear system of equations $f(x) = 0$, and we remove the assumption that $f'(x^*)$ is nonsingular, then standard Newton theory tells us that $x_{k+1}^{Newt} = x_k - f'(x_k)^{-1} f(x_k)$ will converge linearly to x^* from x_0 in a **starlike domain** about x_0.

To be precise, if we let P_N denote the projection onto the null space N of $f'(x^*)$, and P_R the projection onto the range R, then in the best case[3] we have [Gri80, DKK83]

$$\|P_R(x_{k+1}^{Newt} - x^*)\| \leq C_1 \|x_k^{Newt} - x^*\|^2 \tag{13}$$

$$\|P_N(x_{k+1}^{Newt} - x^*)\| \leq \frac{1}{2} \|P_N(x_k^{Newt} - x^*)\| + C_2 \|P_R(x_k^{Newt} - x^*)\| + C_3 \|x_k^{Newt} - x^*\|^2 \tag{14}$$

Newton's method is a linearly convergent fixed point iteration $\implies$ Anderson should help.

Q: How does Anderson improve convergence for singular problems? Does it converge?

[3] $f'(x^*)$ is two regular along N and $\dim N = 1$.

Acceleration & Convergence for singular problems

1. (**D.**,Pollock, 2023) Acceleration is due to the **optimization gain.** Near N and x^*, we have

Newton-Anderson

$$\|P_N e_{k+1}\| \leq \kappa\theta_{k+1}\|P_N e_k\|, \kappa < 1$$

Newton

$$\|P_N e_{k+1}^{\text{Newt}}\| \leq \kappa\|P_N e_k^{\text{Newt}}\|$$

2. (**D.**,Pollock, Rebholz, 2024) For x_0 sufficiently close to N and x^*, and $x_1 = x_0 + w_1$, adaptive γ-safeguarded Newton-Anderson (γ NAA($\hat{r}$)) remains well-defined and converges to x^* with

$$\|P_R e_{k+1}\| \leq c_4 \max\{|1 - \lambda_{k+1}\gamma_{k+1}| \, \|e_k\|^2, |\lambda_{k+1}\gamma_{k+1}| \, \|e_{k-1}\|^2\}$$
$$\|P_N e_{k+1}\| < \kappa\theta_{k+1}^{\lambda}\|P_N e_k\|.$$

Adaptive γ-Safeguarded Newton-Anderson

Algorithm Newton-Anderson(1)

Choose $x_0 \in \mathbb{R}^n$. Set $w_1 = -f'(x_0)^{-1}f(x_0)$, and $x_1 = x_0 + w_1$.
for k=1,2,... **do**
  $w_{k+1} \leftarrow -f'(x_k)^{-1}f(x_k)$
  $\gamma_{k+1} \leftarrow (w_{k+1} - w_k)^T w_{k+1} / \|w_{k+1} - w_k\|_2^2$
  $x_{k+1} \leftarrow x_k + w_{k+1} - \gamma_{k+1}(x_k - x_{k-1} + w_{k+1} - w_k)$
end for

Algorithm Adaptive γ-Safeguarded Newton-Anderson

Choose $x_0 \in \mathbb{R}^n$ and $\hat{r} \in (0,1)$. Set $w_1 = -f'(x_0)^{-1}f(x_0)$ and $x_1 = x_0 + w_1$
for k=1,2,... **do**
  $w_{k+1} \leftarrow -f'(x_k)^{-1}f(x_k)$
  $\gamma_{k+1} \leftarrow (w_{k+1} - w_k)^T w_{k+1} / \|w_{k+1} - w_k\|_2^2$
  $\eta_{k+1} \leftarrow \|w_{k+1}\| / \|w_k\|$
  $r_{k+1} \leftarrow \min\{\eta_{k+1}, \hat{r}\}$
  $\beta_{k+1} \leftarrow r_{k+1}\eta_{k+1}$
  $\lambda^a \leftarrow 1$
  if $\gamma_{k+1} = 0$ **or** $\gamma_{k+1} \geq 1$ **then**
    $\lambda^a \leftarrow 0$
  else if $|\gamma_{k+1}| / |1 - \gamma_{k+1}| > \beta_{k+1}$ **then**
    $\lambda^a \leftarrow \dfrac{\beta_{k+1}}{\gamma_{k+1}\left(\beta_{k+1} + \text{sign}(\gamma_{k+1})\right)}$
  end if
  $x_{k+1} \leftarrow x_k + w_{k+1} - \lambda^a \gamma_{k+1}(x_k - x_{k-1} + w_{k+1} - w_k)$
end for

Adaptive γ-safeguarding

In our first paper on Newton-Anderson for singular problems [DP23], we used an earlier version of γ-safeguarding that was not adaptive. A parameter r is set at the start of the solve and remained fixed.

It was found that, when applying Newton-Anderson to with γ-safeguarding to *nonsingular problems*, both it and standard Newton-Anderson underperformed compared to standard Newton. Reducing r closer to zero improved this, but standard Newton was still best.

Moreover...

Adaptive γ-safeguarding

It was recently shown by Rebholz and Xiao in [RX23] that Anderson reduces order of convergence of superlinearly convergent fixed-point iterations:

$$\text{order} \mapsto \frac{\text{order} + 1}{2} \quad \text{(with depth } m = 1\text{)}.$$

So while Anderson is helpful in the preasymptotic regime, it is not ideal (locally!) when problem is nonsingular.

Idea: develop an adaptive safeguarding scheme that will

1. preserve convergence results for singular problems; and
2. automatically detect nonsingular problems and respond by "turning off" Anderson to recover local quadratic convergence; and

Adaptive γ-safeguarding

In non-adaptive γ-safeguarding from D., Pollock, 2023 [DP23], we have one tunable parameter $r \in (0,1)$ that we fix at the start of the solve.

$$r \approx 0 \implies \text{Newton-like step} \quad \text{and} \quad r \approx 1 \implies \text{Newton-Anderson-like step}$$

Adaptive γ-safeguarding: replace r with r_{k+1} that may change at each iteration. A satisfactory choice of r_{k+1} should satisfy three criteria.

1. $r_{k+1} \approx 0$ if $\|P_N e_k\| / \|P_N e_{k-1}\| \approx 0$;
2. $r_{k+1} \approx 1$ if $\|P_N e_k\| / \|P_N e_{k-1}\| \approx 1$; and
3. $\lim_{k\to\infty} r_{k+1} = 0$ if $F'(x^*)$ is nonsingular.

$$r_{k+1} = \frac{\|w_{k+1}\|}{\|w_k\|} \text{ meets all three criteria since } \frac{\|w_{k+1}\|}{\|w_k\|} \approx \frac{\|P_N e_k\|}{\|P_N e_{k-1}\|} \text{ near } N \text{ and } x^*.$$

Take $r_{k+1} = \min\{\|w_{k+1}\|/\|w_k\|, \hat{r}\}$ to preserve local convergence, where $\hat{r} \in (0,1)$.

Adaptive γ-Safeguarded Newton-Anderson

Algorithm Newton-Anderson(1)

Choose $x_0 \in \mathbb{R}^n$. Set $w_1 = -f'(x_0)^{-1} f(x_0)$, and $x_1 = x_0 + w_1$.
for k=1,2,... **do**
  $w_{k+1} \leftarrow -f'(x_k)^{-1} f(x_k)$
  $\gamma_{k+1} \leftarrow (w_{k+1} - w_k)^T w_{k+1} / \|w_{k+1} - w_k\|_2^2$
  $x_{k+1} \leftarrow x_k + w_{k+1} - \gamma_{k+1}(x_k - x_{k-1} + w_{k+1} - w_k)$
end for

Algorithm Adaptive γ-Safeguarded Newton-Anderson

Choose $x_0 \in \mathbb{R}^n$ and $\hat{r} \in (0,1)$. Set $w_1 = -f'(x_0)^{-1} f(x_0)$ and $x_1 = x_0 + w_1$
for k=1,2,... **do**
  $w_{k+1} \leftarrow -f'(x_k)^{-1} f(x_k)$
  $\gamma_{k+1} \leftarrow (w_{k+1} - w_k)^T w_{k+1} / \|w_{k+1} - w_k\|_2^2$
  $\eta_{k+1} \leftarrow \|w_{k+1}\| / \|w_k\|$
  $r_{k+1} \leftarrow \min\{\eta_{k+1}, \hat{r}\}$
  $\beta_{k+1} \leftarrow r_{k+1} \eta_{k+1}$
  $\lambda^a \leftarrow 1$
  if $\gamma_{k+1} = 0$ **or** $\gamma_{k+1} \geq 1$ **then**
    $\lambda^a \leftarrow 0$
  else if $|\gamma_{k+1}| / |1 - \gamma_{k+1}| > \beta_{k+1}$ **then**
    $\lambda^a \leftarrow \dfrac{\beta_{k+1}}{\gamma_{k+1} \left(\beta_{k+1} + \operatorname{sign}(\gamma_{k+1})\right)}$
  end if
  $x_{k+1} \leftarrow x_k + w_{k+1} - \lambda^a \gamma_{k+1} (x_k - x_{k-1} + w_{k+1} - w_k)$
end for

γ-Safeguarding's effect on an AA step

At step k in adaptive γ-safeguarding, the parameter $r_{k+1} = \min\{\|w_{k+1}\|/\|w_k\|, \hat{r}\}$, where $\hat{r} \in (0, 1)$ is set at the start of the solve.

Now r_{k+1}, not the fixed parameter r from non-adaptive safeguarding, determines how strictly the AA iterates are scaled to a Newton iterate should the condition in line 11 be met. In particular, when $\|w_{k+1}\|/\|w_k\| < \hat{r}$, $r_{k+1} = \|w_{k+1}\|/\|w_k\|$. Therefore,

- if the residual is decreased significantly from the previous step, we take a more Newton-like step.
- Otherwise, seeking acceleration, we take an AA-like step.

Remark: These conditions superficially resemble those of methods such as pseudo-transient continuation [KK98] and the Eisentat-Walker choice 2 for the forcing term in inexact Newton methods [EW95].

γ-Safeguarding's effect on an AA step

We can be a bit more precise. Letting λ^a_{k+1} be the quantity returned in line 12 of the adaptive safeguarding algorithm, we have [DPR24, Lemma 3.2]

$$|\lambda^a_{k+1}\gamma_{k+1}| \leq \frac{r_{k+1}\|w_{k+1}\|/\|w_k\|}{1-\|w_{k+1}\|/\|w_k\|} \tag{15}$$

when $\lambda^a_{k+1} = 1$ (the condition in line 11 was not met), and

$$|\lambda^a_{k+1}\gamma_{k+1}| \leq \frac{r_{k+1}\|w_{k+1}\|/\|w_k\|}{1+\text{sign}(\gamma_{k+1})\|w_{k+1}\|/\|w_k\|} \tag{16}$$

when $\lambda^a_{k+1} < 1$. Recall that the x^{NA}_{k+1} step with γ-safeguarding may be written as

$$x^{NA}_{k+1} = x^{Newt}_{k+1} - \lambda^a_{k+1}\gamma_{k+1}\left(x^{Newt}_{k+1} - x^{Newt}_k\right). \tag{17}$$

γ-Safeguarding's effect on an AA step

From equations (15), (16), and (17), we can obtain the general bound

$$\|x_{k+1}^{NA} - x_{k+1}^{Newt}\| \leq C \left(\frac{r_{k+1}\|w_{k+1}\|/\|w_k\|}{1 - \|w_{k+1}\|/\|w_k\|} \right) \max\{\|e_{k+1}^{Newt}\|, \|e_k^{Newt}\|\} \tag{18}$$

if $\lambda_{k+1}^a = 1$ and

$$\|x_{k+1}^{NA} - x_{k+1}^{Newt}\| \leq C \left(\frac{r_{k+1}\|w_{k+1}\|/\|w_k\|}{1 + \text{sign}(\gamma_{k+1})\|w_{k+1}\|/\|w_k\|} \right) \max\{\|e_{k+1}^{Newt}\|, \|e_k^{Newt}\|\} \tag{19}$$

if $\lambda_{k+1}^a < 1$. Here C is a constant that depends only on f.

Asymptotic behavior in nonsingular case

If $f'(x^*)$ is **nonsingular**, then letting $e_{k+1} := x_{k+1} - x^*$ denote the error at the $(k+1)$st step of adaptive γ-safeguarded NA, we have (Corollary 3.4, **D.**, Pollock, Rebholz, 2024)

$$\|e_{k+1}\| \leq \left(\frac{C_1}{1-\hat{r}^2} + C_2 \right) \|e_k\|^2. \tag{20}$$

when $\|w_{k+1}\|/\|w_k\| < \hat{r}$ and $\|e_k\| < \|e_{k-1}\|$.

- We thus have a flexible algorithm with a tunable parameter $\hat{r} \in (0,1)$ that accelerates convergence for singular problems, and recovers quadratic convergence for nonsingular problems.

Numerics

In the following slides, we apply Newton (Newt), Newton-Anderson (NA), and Newton-Anderson with adaptive γ-safeguarding and parameter $\hat{r}$ (γNAA($\hat{r}$)) to two parameter-dependent incompressible flow problems. The goal is to demonstrate two general strategies one can use when implementing γ-safeguarding, and also show that

1. Newton-Anderson with adaptive γ-safeguarding performs competitively compared to standard Newton and Newton-Anderson near bifrucation points, which are necessarily singular points;
2. adaptive γ-safeguarding automatically detects nonsingular problems and recovers quadratic convergence; and
3. that with the right choice of $\hat{r}$, Newton-Anderson with adaptive γ-safeguarding can converge when both Newton-Anderson and Newton fail to converge.

Two general strategies

1. **Asymptotic safeguarding**
   - Adaptive γ-safeguarding is not applied until the residual is less than some tolerance τ. In [DPR24], we choose $\tau = 10^{-1}$.
2. **Preasymptotic safeguarding**
   - Adaptive γ-safeguarding is active during the entire solve.

Incompressible Channel Flow

$$\begin{cases} -\mu\Delta\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla p = \mathbf{0} \\ \nabla\cdot\mathbf{u} = \mathbf{0} \end{cases} \qquad \begin{aligned} \mathbf{u} &= \mathbf{u}_{\text{in}}, \quad \Gamma_{\text{in}} \\ \mathbf{u} &= 0, \quad \Gamma_{\text{wall}} \\ -p\mathbf{n} + (\mu\nabla\mathbf{u})\mathbf{n} &= 0, \quad \Gamma_{\text{out}} \end{aligned} \tag{21}$$

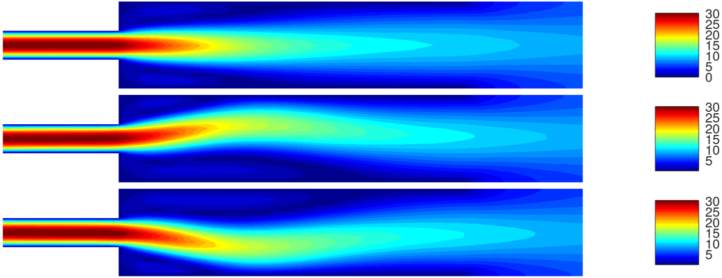

Figure: Stable Solutions to channel flow problem for different $\mu \in (0.9, 1)$.

$\mathcal{P}_2 - \mathcal{P}_1$ Taylor-Hood elements. 12, 734 velocity dof and 1,672 pressure dof.

Asymptotic safeguarding: incompressible channel flow

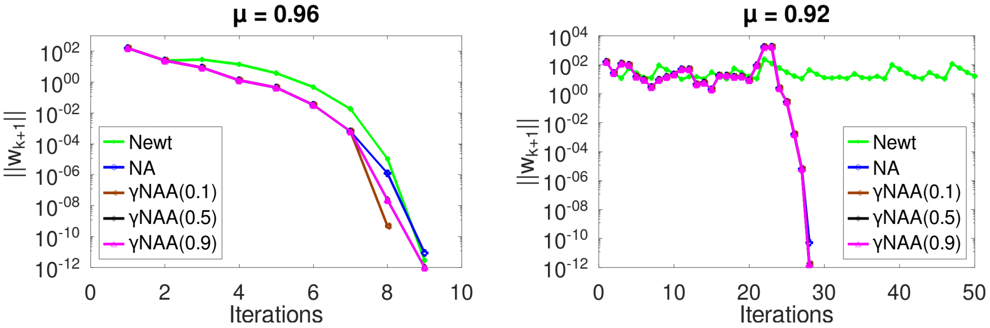

Figure: Residual history comparing Newton, Newton-Anderson, and Newton-Anderson with adaptive γ-safeguarding applied once $\|w_{k+1}\| < 10^{-1}$.

Preasymptotic safeguarding: incompressible channel flow

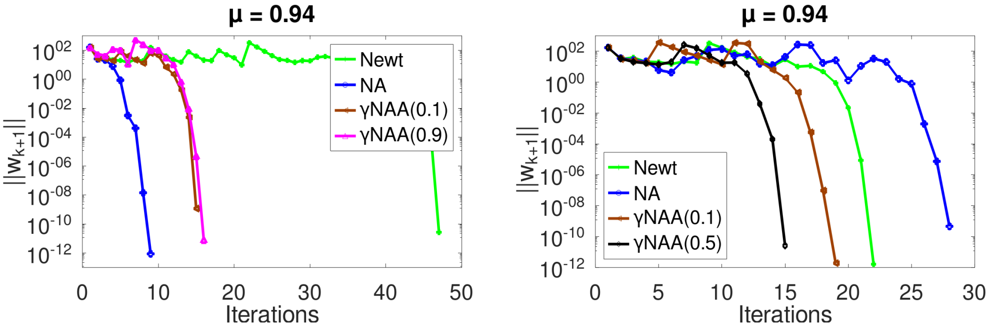

Figure: Left: Residual history for Newton, Newton-Anderson, and γ-safeguarded Newton-Anderson applied to an incompressible Navier-Stokes equations. Right: Residual history starting from different starting point from that of the figure on the left.

Rayleigh-Bénard convection

$$\begin{cases} -\mu\Delta\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla p - \mathrm{Ri}\, T\mathbf{e}_y = \mathbf{0} \\ \nabla\cdot\mathbf{u} = 0 \\ -\kappa\Delta T + \mathbf{u}\cdot\nabla T = 0 \end{cases} \qquad \begin{aligned} T = 1,\ \mathbf{x} \in \Gamma_1 &:= \{1\}\times(0,1), \\ T = 0,\ \mathbf{x} \in \Gamma_2 &:= \{0\}\times(0,1), \\ \nabla T\cdot\mathbf{n} = 0,\ \mathbf{x} \in \Gamma_3 &:= (0,1)\times\{0,1\}, \\ \mathbf{u} = \mathbf{0},\ \mathbf{x} \in \partial\Omega &= \Gamma_1\cup\Gamma_2\cup\Gamma_3. \end{aligned} \tag{22}$$

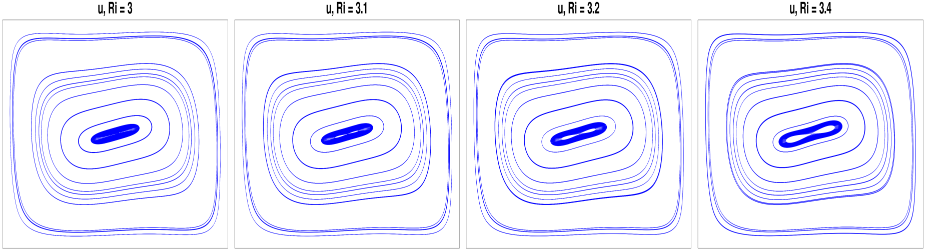

Figure: Velocity streamlines showing transition from one eddy to two eddies.

$\mathcal{P}_2 - \mathcal{P}_1^0$ Scott-Vogelius elements. 7,258 velocity dof and 5,346 pressure dof.

Preasymptotic safeguarding: Rayleigh-Bénard Convection

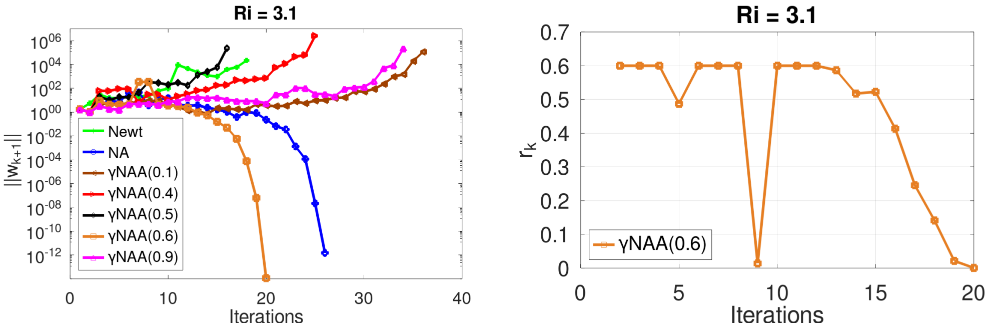

Figure: Sample results for preasymptotic safeguarding. Left: Residual history. Right: r_k history.

Right plot shows γ NAA($\hat{r}$) is working as intended since $r_k := \min\left\{\frac{\|w_{k+1}\|}{\|w_k\|}, \hat{r}\right\} \to 0$.

Preasymptotic safeguarding: Rayleigh-Bénard Convection

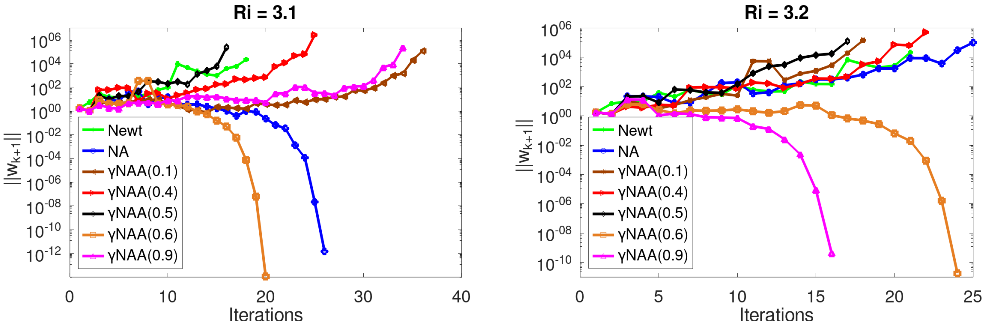

Figure: Sample results for preasymptotic safeguarding. Left: Residual history for Ri $= 3.1$. Right: Residual history for Ri $= 3.2$.

γ NAA($\hat{r}$) performs competitively with Newton-Anderson, and can recover convergence when both Newton and Newton-Anderson diverge, but *is still be sensitive to* $\hat{r}$.

Summary of Results

1. Acceleration and Convergence:[4]
   - Anderson Acceleration improves rate of convergence of Newton's method in singular case; and
   - γ-safeguarded Newton-Anderson converges locally.
2. Adaptive Safeguarding:[5]
   - adaptive γ-safeguarded Newton-Anderson effectively solves nonlinear PDEs near singular points; and
   - automatically detects nonsingular problems and recovers quadratic convergence.

Book on the way: Sara Pollock and Leo Rebholz, *Anderson Acceleration for Numerical PDEs*, in production now with SIAM.

---

[4] **D.**, S. Pollock, Newton-Anderson at Singular Points, IJNAM, 2023

[5] **D.**, S. Pollock, L. G. Rebholz, Analysis of an Adaptive Safeguarded Newton-Anderson Algorithm with Applications to Fluid Problems, ACSE, 2024

Ongoing Work

- Active projects
  1. Extending results to algorithmic depth $m > 1$ and dim $N > 1$.
  2. Developing theory for Anderson applied to perturbed Newton methods.
  3. Is γ-safeguarding useful as a general fixed point iteration safeguard?
- Semi-active or future projects
  1. Applications of γ-safeguarded Newton-Anderson to more complex problems. Is it robust and practical?
  2. Is there a way to tweak parameter $\hat{r}$ in the preasymptotic regime that leads to some kind of global convergence result?

References I

D. G. Anderson.
Iterative procedures for nonlinear integral equations.
J. Assoc. Comput. Mach., 12(4):547–560, 1965.

Renato G. Bettiol and Paolo Piccione.
Instability and bifurcation.
Notices of the AMS, December:1679 – 1691, 2020.

D. W. Decker, H. B. Keller, and C. T. Kelley.
Convergence rates for Newton's method at singular points.
SIAM J. Numer. Anal., 20(2):296–314, 1983.

M. Dallas and S. Pollock.
Newton-Anderson at singular points.
Int. J. Numer. Anal. and Mod., 20(5):667–692, 2023.

References II

M. Dallas, S. Pollock, and L. G. Rebholz.
Analysis of an adaptive safeguarded Newton-Anderson algorithm of depth one with applications to fluid problems.
Advances in Computational Science and Engineering, 2(3):246–270, 2024.

C. Evans, S. Pollock, L. Rebholz, and M. Xiao.
A proof that Anderson acceleration improves the convergence rate in linearly converging fixed point methods (but not in those converging quadratically).
SIAM J. Numer. Anal., 58(1):788–810, 2020.

Stanely C. Eisenstat and Homer F. Walker.
Choosing the forcing terms in an inexact newton method.
SIAM J. Sci. Comput., 1995.

H. Fang and Y. Saad.
Two classes of multisecant methods for nonlinear acceleration.
Numer. Linear Algebra Appl., 16(3):197–221, 2008.

References III

A.O. Griewank.
Starlike domains of convergence for Newton's method at singularities.
Numer. Math., 35:95–111, 1980.

A.O. Griewank.
On solving nonlinear equations with simple singularities or nearly singular solutions.
SIAM Review, 27:537–563, 1985.

A Griewank.
Broyden updating, the good and the bad!
Documenta Mathematica, pages 301–315, 2012.

A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov.
Critical solutions of nonlinear equations: local attraction for Newton-type methods.
Math. Progam., 167:355–379, 2018.

References IV

C. T. Kelley and David E. Keyes.
Convergence analysis of pseudo-transient continuation.
SIAM Journal on Numerical Analysis, 35(2):508–523, 1998.

Christian Kanzow, Nobuo Yamashita, and Masao Fukushima.
Levenberg–marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints.
J. Comput. Appl. Math, 172:375–397, 2004.

J.M. Ortega.
The Newton-Kantorovich theorem.
The American Mathematical Monthly, 75(6):658–660.

Sara Pollock and Leo G. Rebholz.
Anderson acceleration for contractive and noncontractive operators.
IMA J. Numer. Anal., 41(4):2841–2872, 2021.

References V

Sara Pollock and Leo Rebholz.
Filtering for Anderson acceleration.
SIAM J. Sci. Comput., 45(4):A1571–A1590, 2023.

S. Pollock and H. Schwartz.
Benchmarking results for the Newton–Anderson method.
Results Appl. Math., 8:100095, 2020.

Convergence acceleration of iterative sequences. the case of scf iteration.
Chemical Physics Letters, 73(2):393–398, 1980.

Leo G. Rebholz and Mengying Xiao.
The effect of Anderson acceleration on superlinear and sublinear convergence.
Journal of Scientific Computing, 96(2), 2023.

References VI

A. Toth and C. T. Kelley.
Convergence analysis for Anderson acceleration.
SIAM J. Numer. Anal., 53(2):805–819, 2015.

H. F. Walker and P. Ni.
Anderson acceleration for fixed-point iterations.
SIAM J. Numer. Anal., 49(4):1715–1735, 2011.

Thank you!