An Adaptive Safeguarded Newton-Anderson Algorithm for Solving Nonlinear Problems Near Singular Points

SIAM TX-LA
Baylor University, Waco, TX

DMS 2011519

Matt Dallas

Dept. of Mathematics
University of Dallas

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

March 19, 2025

UNIVERSITY of DALLAS | MATHEMATICS DEPARTMENT

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.

Notation: P_R and P_N denote orthogonal projections onto the range and null space of $F'(x^*)$, $e_k := x_k - x^*$, and $\|\cdot\| = \|\cdot\|_2$.

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].

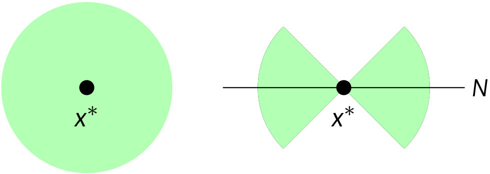

Figure: **Left**: Domain of convergence for Newton's method when $F'(x^*)$ is **nonsingular**. **Right**: Example domain of convergence when $F'(x^*)$ is **singular**, and $N = \text{null}\, F'(x^*)$.

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^*.

Anderson Acceleration (AA)

(1965) First 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?

Why Anderson?

Anderson Acceleration can achieve superlinear convergence when applied to linearly convergent fixed-point iterations [PR21], e.g., Newton's method applied to a singular problem.

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

No knowledge of the order of the root required; it's computationally cheap; and, for singular problems, is theoretically supported under the assumption of 2-regularity.

It has been shown by Izmailov, Kurennoy, and Solodov in [IKS18] that

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

Thus Anderson can recover superlinear convergence when LM cannot.

Anderson Acceleration

Suppose we seek a fixed point of g, and we have computed $m+1$ iterates $\{x_k, x_{k-1}, \ldots, x_{k-m}\}$ where $x_i = g(x_{i-1})$. Let $w_{k+1} = g(x_k) - x_k$.

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

and $\gamma_{k+1} = \operatorname{argmin}_{\gamma \in \mathbb{R}^m} \|w_{k+1} - F_k \gamma\|_2$. Then

$$x_{k+1}^{AA} = x_k + \beta w_{k+1} - (E_k + \beta F_k)\gamma_{k+1}. \tag{2}$$

Here $\beta \in (0,1]$ is a damping parameter. The results in this talk apply to $\beta = 1$ and $m = 1$.

We define the **optimization gain** [PR21] as

$$\theta_{k+1} := \frac{\|w_{k+1} - F_k \gamma_{k+1}\|}{\|w_{k+1}\|}. \tag{3}$$

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

Acceleration & Convergence

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

$$\|P_R e_{k+1}\| \leq c_4 \max\{|1 - \lambda_{k+1}\gamma_{k+1}| \, \|e_k\|^2, |\lambda_{k+1}\gamma_{k+1}| \, \|e_{k-1}\|^2\}$$

$$\|P_N e_{k+1}\| < \kappa\theta_{k+1}^{\lambda}\|P_N e_k\|.$$

Anderson slows superlinear convergence

Anderson improves the rate of convergence of linearly convergent fixed-point iterations.

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

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

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

Goal: develop an adaptive safeguarding scheme that will

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

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 $\hat{r} \in (0,1)$.

Asymptotic Behavior of γNAA($\hat{r}$): Nonsingular problems

If $F'(x^*)$ is **nonsingular**, then letting $e_{k+1} := x_{k+1} - x^*$ denote the error at the $(k+1)$st γ NAA($\hat{r}$) step, we have (Corollary 3.4, **D.**, Pollock, Rebholz, 2024)

$$\|e_{k+1}\| \leq \left(\frac{C_1}{1 - \hat{r}^2} + C_2 \right) \|e_k\|^2. \tag{4}$$

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

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

Example from Incompressible Channel Flow

$$\begin{cases} -\mu\Delta\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla p = \mathbf{0} \\ \nabla\cdot\mathbf{u} = \mathbf{0} \end{cases} \qquad \begin{aligned} \mathbf{u} &= \mathbf{u}_{\text{in}}, \quad \Gamma_{\text{in}} \\ \mathbf{u} &= 0, \quad \Gamma_{\text{wall}} \\ -p\mathbf{n} + (\mu\nabla\mathbf{u})\mathbf{n} &= 0, \quad \Gamma_{\text{out}} \end{aligned} \tag{5}$$

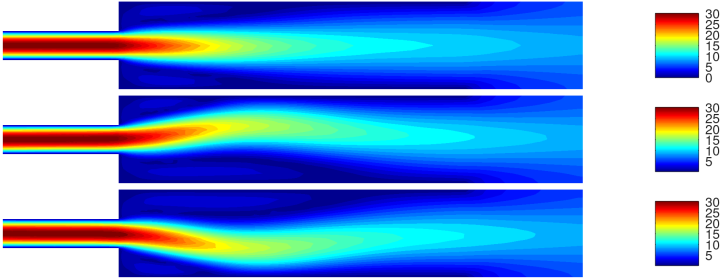

Figure: Stable Solutions to channel flow problem for different μ.

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

Example from Incompressible Channel Flow

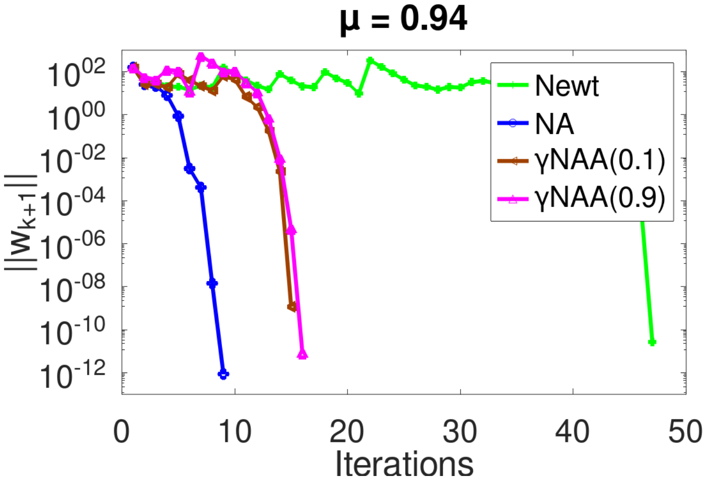

Figure: Residual history for Newton, Newton-Anderson, and γ-safeguarded Newton-Anderson applied to an incompressible Navier-Stokes equation.

Rayleigh-Bénard Convection

$$
\begin{cases} -\mu\Delta\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla p - \mathrm{Ri}\, T\mathbf{e}_y = \mathbf{0} \\ \nabla\cdot\mathbf{u} = 0 \\ -\kappa\Delta T + \mathbf{u}\cdot\nabla T = 0 \end{cases}
\qquad
\begin{aligned} T &= 1,\ \mathbf{x}\in\Gamma_1 := \{1\}\times(0,1), \\ T &= 0,\ \mathbf{x}\in\Gamma_2 := \{0\}\times(0,1), \\ \nabla T\cdot\mathbf{n} &= 0,\ \mathbf{x}\in\Gamma_3 := (0,1)\times\{0,1\}, \\ \mathbf{u} &= \mathbf{0},\ \mathbf{x}\in\partial\Omega = \Gamma_1\cup\Gamma_2\cup\Gamma_3. \end{aligned} \tag{6}
$$

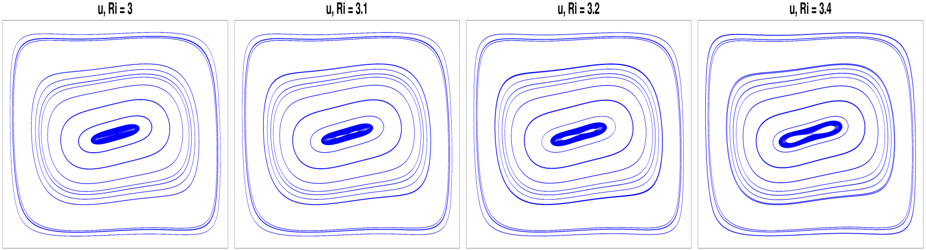

Figure: Velocity streamlines showing transition from one eddy to two eddies.

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

Numerical Results: Rayleigh-Bénard Convection

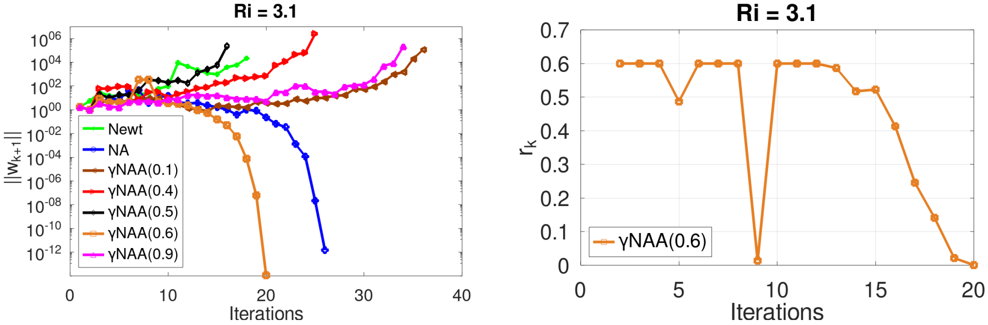

Figure: Sample results for preasymptotic safeguarding. Left: Residual history. Right: r_k history.

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

Numerical Results: Rayleigh-Bénard Convection

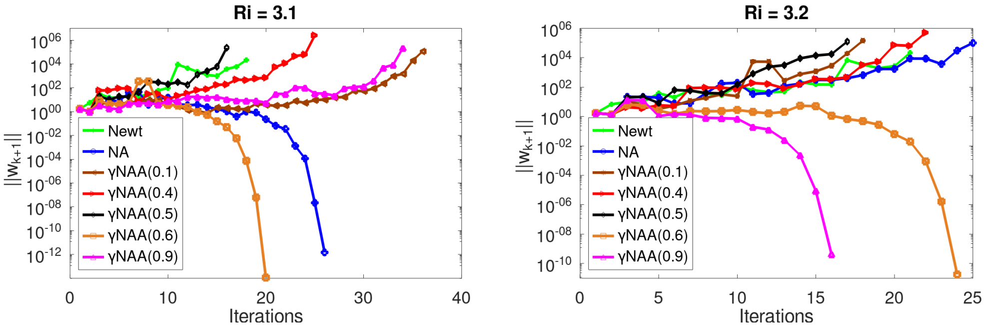

Figure: Sample results for preasymptotic safeguarding. Left: Residual history for $\mathrm{Ri} = 3.1$. Right: Residual history for $\mathrm{Ri} = 3.2$.

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

---

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

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

References I

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

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.

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.

References II

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.

H. Fang and Y. Saad.

Two classes of multisecant methods for nonlinear acceleration.

Numer. Linear Algebra Appl., 16(3):197–221, 2008.

A.O. Griewank.

On solving nonlinear equations with simple singularities or nearly singular solutions.

SIAM Review, 27:537–563, 1985.

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 III

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.

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

References IV

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.

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!

Figure: D., Pollock, IJNAM, 2023

Figure: D., Pollock, Rebholz, ACSE, 2024

DMS 2011519