# Newton-Anderson at Singular Points

## SIAM Gators Seminar

Matt Dallas

Dept. of Mathematics
University of Florida

Joint work with:
Sara Pollock (UF),

September 16, 2023

DMS-{2011519}

# The Flow

Background/Motivation

↓

Singular Problems

↓

Anderson Acc

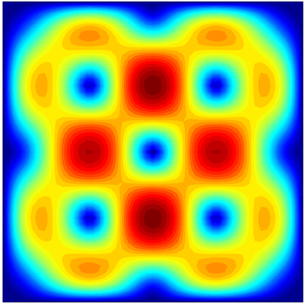
↓

Anderson Acc Newton

↓

Bifurcations

u, Ra = 8.6813e-05



orange = recent developments or ongoing work

# Introduction

- Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be nonlinear.
- Everything we'll discuss today is motivated by the problem

$$F(x) = 0$$

# Why?

- Nonlinear Integral Equations

*Chandrasekhar H-equation*

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

# Why?

- Nonlinear Integral Equations

*Chandrasekhar H-equation*

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

- 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*

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

# Why $\mathbb{R}^n$?

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

# Newton's Method

- A popular choice because of it's relative simplicity and strong local convergence results.
- The idea is to linearize $F$, and approximate a root $x^*$ by the root of the linearization.
- Suppose we have an approximate root $x^k$. Then

$$F(x) \approx F(x^k) + F'(x^k)(x - x^k),$$

and we define $x^{k+1}$ as the root of the linearization. That is,

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

# Newton's Method

- The strong local convergence property can be stated as follows:

*If $F : \mathbb{R}^n \to \mathbb{R}^n$ is $C^1$, with $F'$ Lipschitz, and $F(x^*) = 0$ with $F'(x^*)$ invertible, then there exists a neighborhood of $x^*$ such that $x^k \to x^*$ q-quadratically for any $x^0$ in this neighborhood, where $\{x_k\}$ is defined by (1).*

*Remark*: Everything stated here can be generalized to Banach spaces. The celebrated Newton-Kantorovich[1] theorem is the analogue of the above result in this more general setting.

[1] Ortega, J. M. (1968). The Newton-Kantorovich Theorem. The American Mathematical Monthly, 75(6), 658–660. https://doi.org/10.2307/2313800 [Ort68]

# Newton isn't *always* fast...

- The assumption that $F'(x^*)$ is invertible is very important. Without it, we aren't guaranteed fast local convergence.
- In fact, if $F'(x^*)$ is singular, the "best" we can hope for is[2]

$$\|P_X(x_i - x^*)\| \leq K_1 \|x_{i-1} - x^*\|^2, \quad some\ K_1 > 0,$$

$$\lim_{i\to\infty} \frac{\|P_N(x_i - x^*)\|}{\|P_N(x_{i-1} - x^*)\|} = \frac{1}{2}, \qquad i = 1, 2, \cdots.$$

- Here $X$ is range of $F'(x^*)$, and $N$ is the nullspace.

[2] Decker, D. W., Keller, H. B., & Kelley, C. T. (1983). Convergence Rates for Newton's Method at Singular Points. SIAM Journal on Numerical Analysis, 20(2), 296–314. http://www.jstor.org/stable/2157219 [DKK83]

# Newton isn't *always* fast...

The point is that when $F'(x^*)$ is singular, Newton's method will only converge linearly, and at best the convergence rate will approach $1/2$.

Recall that Newton's method can be viewed as a fixed-point method, which is linearly convergent in the singular case.

What we need is some method that accelerates linearly converging fixed point methods.

# Anderson Acceleration (AA)

(1965) Introduced by D.G. Anderson

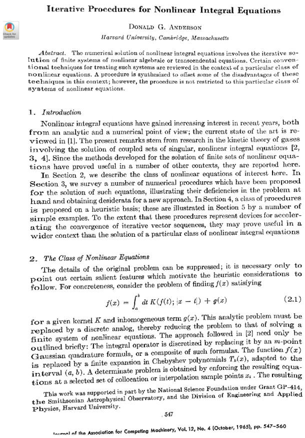
(1980) A closely related method, DIIS or Pulay Mixing, is introduced by Peter Pulay in *Convergence acceleration of iterative sequences. The case of SCF iteration*.

(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 if $F$ is linear, then AA is (in a sense) equivalent to the well-known GMRES method. *Anderson Acceleration for Fixed-Point Iterations*.

(2015) Toth and Kelley prove that AA converges in *Convergence analysis for Anderson acceleration*.

(2020) Evans, Pollock, Rebholz, and Xiao provide *A Proof That Anderson Acceleration Improves the Convergence Rate in Linearly Converging Fixed-Point Methods (But Not in Those Converging Quadratically)*.



### Iterative Procedures for Nonlinear Integral Equations

Donald G. Anderson
*Harvard University, Cambridge, Massachusetts*

*Abstract.* The numerical solution of nonlinear integral equations involves the iterative solution of finite systems of nonlinear algebraic or transcendental equations. Certain conventional techniques for treating such systems are reviewed in the context of a particular class of nonlinear equations. A procedure is synthesized to offset some of the disadvantages of these techniques in this context; however, the procedure is not restricted to this particular class of systems of nonlinear equations.

1. *Introduction*

Nonlinear integral equations have gained increasing interest in recent years, both from an analytic and a numerical point of view; the current state of the art is reviewed in [1]. The present remarks stem from research in the kinetic theory of gases involving the solution of coupled sets of singular, nonlinear integral equations [2, 3, 4]. Since the methods developed for the solution of finite sets of nonlinear equations have proved useful in a number of other contexts, they are reported here.

In Section 2, we describe the class of nonlinear equations of interest here. In Section 3, we survey a number of numerical procedures which have been proposed for the solution of such equations, illustrating their deficiencies in the problem at hand and obtaining desiderata for a new approach. In Section 4, a class of procedures is proposed on a heuristic basis; these are illustrated in Section 5 by a number of simple examples. To the extent that these procedures represent devices for accelerating the convergence of iterative vector sequences, they may prove useful in a wider context than the solution of a particular class of nonlinear integral equations

2. *The Class of Nonlinear Equations*

The details of the original problem can be suppressed; it is necessary only to point out certain salient features which motivate the heuristic considerations to follow. For concreteness, consider the problem of finding $f(x)$ satisfying

$$f(x) = \int_a^b dt\, K(f(t); |x - t|) + g(x) \quad (2.1)$$

for a given kernel $K$ and inhomogeneous term $g(x)$. This analytic problem must be replaced by a discrete analog, thereby reducing the problem to that of solving a finite system of nonlinear equations. The approach followed in [2] need only be outlined briefly: The integral operator is discretized by replacing it by an $m$-point Gaussian quadrature formula, or a composite of such formulas. The function $f(x)$ is replaced by a finite expansion in Chebyshev polynomials $T_k(x)$, adapted to the interval $(a, b)$. A determinate problem is obtained by enforcing the resulting equations at a selected set of collocation or interpolation sample points $x_i$. The resulting

This work was supported in part by the National Science Foundation under Grant GP-414, the Smithsonian Astrophysical Observatory, and the Division of Engineering and Applied Physics, Harvard University.

547

Journal of the Association for Computing Machinery, Vol. 12, No. 4 (October, 1965), pp. 547–560

Figure: Donald G. Anderson. 1965. Iterative Procedures for Nonlinear Integral Equations.

# So What is AA?

It's an extrapolation scheme that takes the previous $m$ (called the algorithmic depth) iterates, and constructs a new iterate as follows.

# So What is AA?

It's an extrapolation scheme that takes the previous $m$ (called the algorithmic depth) iterates, and constructs a new iterate as follows.

Suppose we seek a fixed poit 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 $w_{k+1} = g(x_k) - x_k$,

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

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

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

Here $\beta \in (0, 1)$.

# What is AA?

Our new results only consider depth $m = 1$ and $\beta = 1$, in which case

$$x_{k+1}^{AA} = x_k + w_{k+1} - (x_k - x_{k-1} + w_{k+1} - w_k)\gamma_{k+1}.$$

To conclude the AA section, let's go the board to see a “derivation" of AA (á la Hans De Sterk[3]).

[3] Professor of Computational and Applied Math at University of Waterloo, currently the Chair. Very nice guy.

# Anderson and Newton and Newton-Anderson

Now all the pieces are in place to study Anderson accelerated Newton's method, or Newton-Anderson (NA).

For the remainder of the talk, the only fixed-point scheme we care about is Newton's method. Therefore $g(x) = x - F'(x)^{-1}F(x)$, and $w(x) = -F'(x)^{-1}F(x)$. If $x = x_k$, we write $w(x_k) = w_{k+1}$.

# Nonsingular Newton-Anderson

- In 2021, Dr. Pollock and her collaborator Leo Rebholz[4] published *Anderson acceleration for contractive and noncontractive operators*.
- A key result in this and their (and collaborators') 2020 paper is that whether or not Anderson actually accelerates at step $k+1$ is determined by

$$\theta_{k+1} := \frac{\left\|w_{k+1} - \gamma_{k+1}(w_{k+1} - w_k)\right\|_2}{\left\|w_{k+1}\right\|_2}.$$

[4] Clemson Univeristy, also very nice guy.

# Nonsingular Newton-Anderson

- In 2021, Dr. Pollock and her collaborator Leo Rebholz[4] published *Anderson acceleration for contractive and noncontractive operators*.
- A key result in this and their (and collaborators') 2020 paper is that whether or not Anderson actually accelerates at step $k+1$ is determined by

$$\theta_{k+1} := \frac{\|w_{k+1} - \gamma_{k+1}(w_{k+1} - w_k)\|_2}{\|w_{k+1}\|_2}.$$

- Loosely speaking,

$$\theta_{k+1} << 1 \implies \text{acceleration}$$
$$\theta_{k+1} \approx 1 \implies \text{no acceleration}$$

[4] Clemson Univeristy, also very nice guy.

# Singular Newton-Anderson

- Recall that the problem $F(x) = 0$ with solution $x^*$ is *singular* if $F'(x^*)$ is singular.

# Singular Newton-Anderson

- Recall that the problem $F(x) = 0$ with solution $x^*$ is *singular* if $F'(x^*)$ is singular.
- Newton's method only converged linearly in this case.

# Singular Newton-Anderson

- Recall that the problem $F(x) = 0$ with solution $x^*$ is *singular* if $F'(x^*)$ is singular.
- Newton's method only converged linearly in this case.
- Anderson likes to accelerate linearly converging things.

# Singular Newton-Anderson

It has been observed numerically that Newton-Anderson can greatly improve convergence in singular problems.

**Table 3**
Results for degenerate problems.

| Problem | Method | Iterations ($k$) | $\|f(x_k)\|$ | $\|w_k\|$ | $q_k$ | Time (s) |
|---|---|---|---|---|---|---|
| D1. $n = 3$ | Newton | 14 | 3.991e−09 | 6.903e−05 | 1.077 | 0.000129 |
| | N. Anderson(1) | 5 | 1.656e−10 | 1.349e−05 | 1.493 | 7.152e−05 |
| | KS-acc. N. (1.0, 0.9) | 3 | 2.396e−09 | 0.0001877 | 1.993 | 4.476e−05 |
| D2. $n = 1000$ | Newton | 16 | 2.628e−09 | 0.000382 | 1.075 | 0.9698 |
| | N. Anderson(1) | 6 | 1.236e−11 | 0.001947 | 1.663 | 0.3671 |
| | N. Anderson(2) | 6 | 1.912e−11 | 2.595e−06 | 1.399 | 0.3658 |
| | KS-acc. N. (1.0, 0.9) | F | – | – | – | – |
| | KS-acc. N. (0.35, 0.1) | 4 | 3.986e−09 | 0.002449 | 1.461 | 0.4772 |
| D3. $n = 10$ | Newton | 46 | 4.339e−09 | 0.07587 | 1.057 | 0.001038 |
| | N. Anderson(1) | 17 | 7.899e−09 | 0.01347 | 1.45 | 0.0008288 |
| | N. Anderson(2) | 26 | 6.781e−11 | 0.0003507 | 1.404 | 0.002012 |
| | N. Anderson(3) | 6 | 3.964e−10 | 0.09431 | 5.87 | 0.000219 |
| | N. Anderson(4) | 5 | 7.289e−25 | 0.1056 | 23.05 | 0.0001861 |
| | KS-acc. N. (1.0, 0.9) | F | – | – | – | – |
| | KS-acc. N. (0.35, 0.1) | F | – | – | – | – |
| | KS-acc. N. (0.7, 0.3) | 20 | 1.758e−09 | 0.07474 | 1.165 | 0.0007143 |

Figure: Pollock, Schwarz, *Benchmarking results for the Newton-Anderson method*. 2020.[PS20].

# It works...but why?

Our recent work[5] answers the questions

1. What's the mechanism behind singular NA acceleration?
2. Do the NA iterates remain well-defined and converge to $x^*$?

[5] M. Dallas and S. Pollock, Newton-Anderson at Singular Points, *In press*, 2023. DOI: 10.48550/arXiv.2207.12334 To appear in The International Journal of Numerical Analysis and Modeling.

# It works...but why?

Our recent work[5] answers the questions

1. What's the mechanism behind singular NA acceleration?
2. Do the NA iterates remain well-defined and converge to $x^*$?

[5] M. Dallas and S. Pollock, Newton-Anderson at Singular Points, *In press*, 2023. DOI: 10.48550/arXiv.2207.12334 To appear in The International Journal of Numerical Analysis and Modeling.

# It works...but why?

Our recent work[5] answers the questions

1. What's the mechanism behind singular NA acceleration?
   $\longrightarrow$ Within the region of invertibility, it's actually $\theta_{k+1}$!
2. Do the NA iterates remain well-defined and converge to $x^*$?

[5] M. Dallas and S. Pollock, Newton-Anderson at Singular Points, *In press*, 2023. DOI: 10.48550/arXiv.2207.12334
To appear in The International Journal of Numerical Analysis and Modeling.

# It works...but why?

Our recent work[5] answers the questions

1. What's the mechanism behind singular NA acceleration?
   $\longrightarrow$ Within the region of invertibility, it's actually $\theta_{k+1}$!
2. Do the NA iterates remain well-defined and converge to $x^*$?
   $\longrightarrow$ Sort of. We can prove convergence not of NA itself, but of a safeguarded version which we've called $\gamma$-safeguarded NA ( $\gamma$NA($r$) ).

[5] M. Dallas and S. Pollock, Newton-Anderson at Singular Points, *In press*, 2023. DOI: 10.48550/arXiv.2207.12334
To appear in The International Journal of Numerical Analysis and Modeling.

# Regions of Invertibility

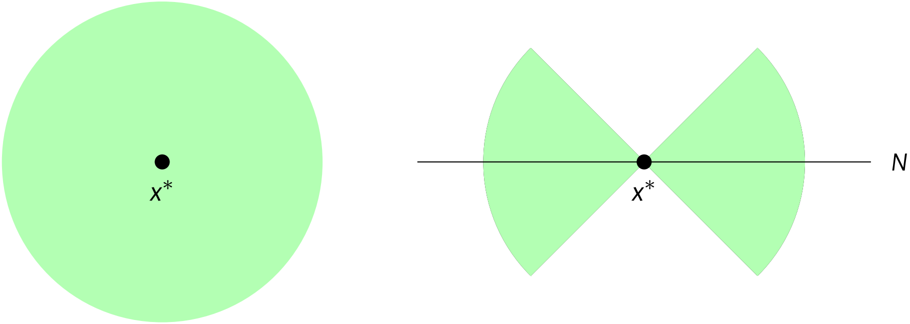



Figure: **Left**: Domain of convergence for Newton's method when $f'(x^*)$ is **nonsingular**. **Right**: Example domain of convergence when $f'(x^*)$ is **singular**. Note that these are also domains of invertibility for $f'$.

# Analysis Strategy

- The error $e_k = x_k - x^*$ may be decomposed into it's range and null space components.

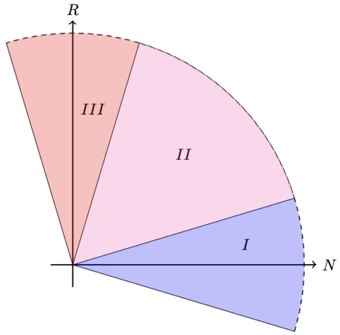



Figure: The three regions of interest in analyzing a Newton-Anderson step. The blue region is the null-dominant region, the red region is range-dominant, and the magenta region is not strongly range or null dominant. For convergence, we're most interested in region I. Note that we've assumed $x^* = 0$ for simplicity.

# Analysis Strategy

- The error $e_k = x_k - x^*$ may be decomposed into it's range and null space components.
- Therefore, the NA (depth $m = 1$) error is determined by how the dominant contributions in $e_k$ and $e_{k-1}$.
- $e_{k+1}^{NA} = \frac{1}{2}\left(P_N e_k\right)^{\alpha} + \left(T_k P_R e_k\right)^{\alpha} + q_{k-1}^{k}$.

  Where $x_k^{\alpha} = (1 - \gamma_{k+1})x_k + \gamma_{k+1}x_{k-1}$, and $q_{k-1}^{k} = \mathcal{O}(|\gamma_{k+1}|, \|e_k\|^2, \|e_{k-1}\|^2)$.

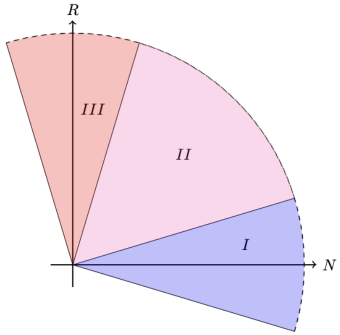



Figure: The three regions of interest in analyzing a Newton-Anderson step. The blue region is the null-dominant region, the red region is range-dominant, and the magenta region is not strongly range or null dominant. For convergence, we're most interested in region I. Note that we've assumed $x^* = 0$ for simplicity.

# Compatibility

In [DP23], we define the notion of *compatibility*. The definition essentially says that if an NA step is compatible when it behaves like a nonsingular NA step.

**Definition 4.1.** *Let* $\{x_k\}$ *be a sequence of Newton-Anderson iterates. We say that* $x_{k+1}$ *is* compatible *or a* compatible step *if there exists a moderate constant* $C > 0$ *independent of* $k$ *such that* $\|P_N e_{k+1}\| \le C\theta_{k+1}\|w_{k+1}\|$, *in which case we'll write* $P_N e_{k+1} = \mathcal{O}(\theta_{k+1}\|w_{k+1}\|)$. *Otherwise,* $(x_k, x_{k-1})$ *is an* incompatible pair, *and* $x_{k+1}$ *is* incompatible *or an* incompatible step

The point is that if $x_{k+1}$ compatible,

$$\theta_{k+1} \ \text{small} \implies \|P_N e_{k+1}\| \ \text{small}$$

# Compatibility

Thus far we have

1. a region containing $x^*$ where $F'(x)$ is invertible for all $x$ in this region, and
2. a way to quantify how well Anderson accelerates a Newton step.

How are (1) and (2) related?

# Compatibility

Paraphrasing Lemma 5.1 in [DP23], we have

*If $x_k$ and $x_{k-1}$ are close to N, then $x_{k+1}^{NA}$ is compatible.*

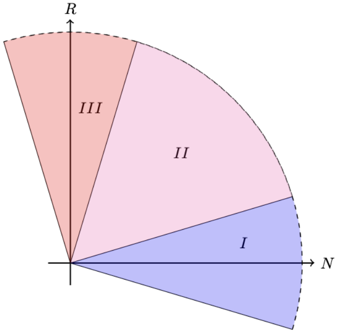
It can then be shown that

$$\|P_N e_{k+1}\| \leq \kappa \theta_{k+1}(1/2)\|P_N e_k\|, \kappa < 1.$$

Compare with the standard Newton bound near $N$ and $x^*$:

$$\|P_N e_{k+1}^{Newt}\| \leq c(1/2)\|P_N e_k^{Newt}\|,$$

with $c < 2$. Note that $\theta_{k+1} \leq 1$.



# Other Compatibility Conditions

- There are compatibility conditions for other arrangements of $x_k$ and $x_{k-1}$, but they can be stringent.

**Lemma 5.3.** *Let the assumptions of theorem 3.1 hold for $x_k$ and $x_{k-1}$, and suppose $r^e_{k+1} < 1$. Suppose $(x_k, x_{k-1})$ is a strong mixed pair. There are two cases.*

1. *(Strong NR-pair) If $E_{k+1} = L_{k+1}(P_N e_k, T_{k-1} P_R e_{k-1})$, then $(x_k, x_{k-1})$ is compatible if*
   (a) $((1-\gamma_{k+1})P_N e_k)^T(\gamma_{k+1} T_{k-1} P_R e_{k-1}) \le 0$, *and*
   (b) $(\gamma_{k+1} P_N e_{k-1})^T((1-\gamma_{k+1})T_k P_R e_k + q^k_{k-1}) \ge 0$.

   *When $(x_k, x_{k-1})$ is not compatible,*

$$\|P_N e_{k+1}\| \le C(1 + r^e_{k+1}) \max\{|1-\gamma_{k+1}| \, \|P_N e_k\|, |\gamma_{k+1}| \, \|T_{k-1} P_R e_{k-1}\|\}. \tag{31}$$

2. *(Strong RN-pair) If $E_{k+1} = L_{k+1}(T_k P_R e_k, P_N e_{k-1})$, then $(x_k, x_{k-1})$ is compatible if*
   (a) $((1-\gamma_{k+1})T_k P_R e_k)^T(\gamma_{k+1} P_N e_{k-1}) \le 0$, *and*
   (b) $((1-\gamma_{k+1} P_N e_k)^T(\gamma_{k+1} T_{k-1} P_R e_{k-1} + q^k_{k-1}) \ge 0$.

   *When $(x_k, x_{k-1})$ is not compatible,*

$$\|P_N e_{k+1}\| \le C(1 + r^e_{k+1}) \max\{|1-\gamma_{k+1}| \, \|T_k P_R e_k\|, (|\gamma_{k+1}|/2) \, \|P_N e_{k-1}\|\}. \tag{32}$$

*In both cases, $C$ denotes a constant determined by $f$.*

# Other Compatibility Conditions

The great things about pairs $x_k$ and $x_{k-1}$ near $N$ is that they are *automatically* compatible.

# Incompatible steps

Incompatible steps can still be accelerated, but there's no guarantee.

The story here is that in the worst case, $\|P_N e_{k+1}\|$ decreases very little, while $\|P_R e_{k+1}\|$ is still quadratic.

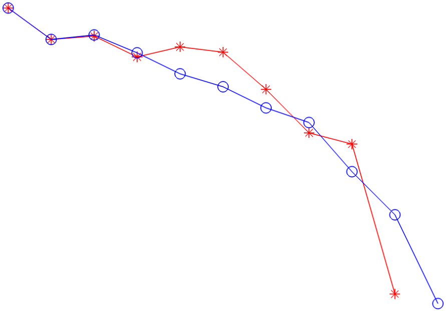
This causes the iterates to cluster around $N$, which leads to compatibility.

# $\gamma$-Safeguarding

To prove convergence, we need to ensure that the iterates remain within the region of invertibility. To achieve this, we created the $\gamma$-safeguarding algorithm.

**Algorithm 2** $\gamma$-safeguarding

Given $x_k$, $x_{k-1}$, $w_{k+1}$, $w_k$, and $\gamma_{k+1}$. Set $r \in (0,1)$ and $\lambda = 1$.
$\beta \leftarrow r\|w_{k+1}\|/\|w_k\|$
**if** $\gamma_{k+1} = 0$ **or** $\gamma_{k+1} \geq 1$ **then**
  $x_{k+1} \leftarrow x_k + w_{k+1}$
**else**
  **if** $|\gamma_{k+1}|/|1-\gamma_{k+1}| > \beta$ **then**
    **if** $\gamma_{k+1} > 0$ **and** $\beta/(\gamma_{k+1}(1+\beta)) < 1$ **then**
      $\lambda \leftarrow \beta/(\gamma_{k+1}(1+\beta))$
    **end if**
    **if** $\gamma_{k+1} < 0$ **and** $0 \leq \beta/(\gamma_{k+1}(\beta-1)) < 1$ **then**
      $\lambda \leftarrow \beta/(\gamma_{k+1}(\beta-1))$
    **end if**
  **end if**
**end if**
$\gamma_{k+1} \leftarrow \lambda\gamma_{k+1}$
$x_{k+1} \leftarrow x_k + w_{k+1} - \gamma_{k+1}(x_k - x_{k-1} + w_{k+1} - w_k)$



# Local Convergence

- We'll denote $\gamma$-safeguarded NA as $\gamma$Na($r$), where $r$ is a parameter set by the user.
- Then, paraphrasing Theorem 6.1 in [DP23], we have

*For $x_0$ sufficiently close to N and $x^*$, and $x_1 = x_0 + w_1$, $\gamma$NA(r) remains well-defined and converges to $x^*$ with*

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

# Examples

| Problem | Algorithm | Iterations | f-evals | $\|\|f(x)\|\|$ | LM/LS/PG |
|---|---|---|---|---|---|
| Eq-Combustion | Proj-Lev-Marq | 11 | 28 | 8.200e-14 | 8/3/0 |
| | N.Anderson | 35 | - | 8.510e-12 | - |
| | $\gamma$-N.Anderson(0.9) | 17 | - | 3.092e-09 | - |
| | Armijo-N.Anderson | 18 | 56 | 2.136e-10 | -/2/- |
| | $\gamma$-Armijo-N.Anderson(0.9) | 17 | 18 | 3.092e-09 | -/0/- |
| Bullard-Biegler | Proj-Lev-Marq | 13 | 26 | 6.602e-11 | 10/3/0 |
| | N.Anderson | F | - | - | - |
| | $\gamma$-N.Anderson(0.5) | 11 | - | 1.799e-12 | - |
| | Armijo-N.Anderson | 20 | 202 | 1.212e-10 | -/12/- |
| | $\gamma$-Armijo-N.Anderson(0.5) | 13 | 34 | 1.629e-11 | -/6/- |

| | | | | | |
|---|---|---|---|---|---|
| Pollock1[30] | Proj-Lev-Marq | 14 | 15 | 3.991e-09 | 14/0/0 |
| | N.Anderson | 5 | - | 1.656e-10 | - |
| | $\gamma$-N.Anderson(0.9) | 5 | - | 8.268e-10 | - |
| | Armijo-N.Anderson | 5 | 6 | 1.656e-10 | -/0/- |
| | $\gamma$-Armijo-N.Anderson(0.9) | 5 | 6 | 8.268e-10 | -/0/- |
| Dayton10[7] | Proj-Lev-Marq | F | - | - | - |
| | N.Anderson | 11 | - | 3.294e-10 | - |
| | $\gamma$-N.Anderson(0.5) | 14 | - | 5.157e-09 | - |
| | Armijo-N.Anderson | 15 | 140 | 2.153e-11 | -/4/- |
| | $\gamma$-Armijo-N.Anderson(0.5) | F | - | - | -/-/- |

Figure: **Top**: Results when applied to two **nonsingular** problems. **Bottom**: Results when applied to **singular** problems.

# Towards Adaptive Safeguarding

- $\gamma$NA($r$) is nice theoretically because it provides a convergence proof.
- Having to choose $r$, however, isn't ideal from an implementation perspective.

# Towards Adaptive Safeguarding

- $\gamma$NA($r$) is nice theoretically because it provides a convergence proof.
- Having to choose $r$, however, isn't ideal from an implementation perspective.
- To improve the situation, we propose an *adaptive* $\gamma$-safeguarded NA, which we denote by $\gamma$NA($r_k$).

# Towards Adaptive Safeguarding

- $\gamma$NA($r$) is nice theoretically because it provides a convergence proof.
- Having to choose $r$, however, isn't ideal from an implementation perspective.
- To improve the situation, we propose an *adaptive* $\gamma$-safeguarded NA, which we denote by $\gamma$NA($r_k$).
- Rather than a fixed $r \in (0, 1)$, we could take

$$r_k = \min\left(\frac{\|w_{k+1}\|}{\|w_k\|}, 0.9\right)$$

# Why $\frac{\|w_{k+1}\|}{\|w_k\|}$?

A good choice of $r_k$ should satisfy the following

1. $r_k << 1$ if $\|P_N e_{k+1}\| / \|P_N e_k\| << 1$

# Why $\frac{\|w_{k+1}\|}{\|w_k\|}$?

A good choice of $r_k$ should satisfy the following

1. $r_k << 1$ if $\|P_N e_{k+1}\| / \|P_N e_k\| << 1$
2. $r_k \approx 1$ if $\|P_N e_{k+1}\| / \|P_N e_k\| \approx 1$

# Why $\frac{\|w_{k+1}\|}{\|w_k\|}$?

A good choice of $r_k$ should satisfy the following

1. $r_k << 1$ if $\|P_N e_{k+1}\| / \|P_N e_k\| << 1$
2. $r_k \approx 1$ if $\|P_N e_{k+1}\| / \|P_N e_k\| \approx 1$
3. If $F'(x^*)$ is not singular, then we want $r_k \to 0$ as $k \to \infty$.

# Why $\frac{\|w_{k+1}\|}{\|w_k\|}$?

In other words, we want

$$r_k \approx \|P_N e_{k+1}\|/\|P_N e_k\|,$$

and in the region of invertibility,

$$\|w_{k+1}\|/\|w_k\| \approx \|P_N e_{k+1}\|/\|P_N e_k\|.$$

# Applications: Bifurcation Theory

- Consider a parameter dependent nonlinear equation:

$$F(x;\mu) = 0.$$

This could come from discretizing a parameter-dependent PDE such as

$$-\Delta \mathbf{u} + \nabla p = -\frac{1}{\eta}\left(\mathbf{u}\cdot\nabla\mathbf{u} + \mathbf{u}_t\right)$$
$$\nabla\cdot\mathbf{u} = 0.$$

# Applications: Bifurcation Theory

- Consider a parameter dependent nonlinear equation:

$$F(x;\mu) = 0.$$

This could come from discretizing a parameter-dependent PDE such as

$$-\Delta \mathbf{u} + \nabla p = -\frac{1}{\eta}(\mathbf{u}\cdot\nabla\mathbf{u} + \mathbf{u}_t)$$
$$\nabla\cdot\mathbf{u} = 0.$$

- If $(\bar{x},\bar{\mu})$ is a solution, and $F_x(\bar{x};\bar{\mu})$ is invertible, then there's a neighborhood of $(\bar{x},\bar{\mu})$ such that there's a unique solution curve $x(\mu)$ through $(\bar{x},\bar{\mu})$.

# Applications: Bifurcation Theory

- Consider a parameter dependent nonlinear equation:

$$F(x; \mu) = 0.$$

This could come from discretizing a parameter-dependent PDE such as

$$-\Delta \mathbf{u} + \nabla p = -\frac{1}{\eta}(\mathbf{u} \cdot \nabla \mathbf{u} + \mathbf{u}_t)$$
$$\nabla \cdot \mathbf{u} = 0.$$

- If $(\bar{x}, \bar{\mu})$ is a solution, and $F_x(\bar{x}; \bar{\mu})$ is invertible, then there's a neighborhood of $(\bar{x}, \bar{\mu})$ such that there's a unique solution curve $x(\mu)$ through $(\bar{x}, \bar{\mu})$.
- Bifurcations occur at $\mu^*$ when there is *not* a unique solution in a neighborhood of $\mu^*$.

# Applications: Bifurcation Theory

If there's not a unique solution near $\mu^*$, then $F_x(x^*; \mu^*)$ must be singular.

# Example: Rayleigh-Benard Convection

Models the flow of a fluid whose motion is produced by buoyancy forces. One system of equations modeling this scenario is the system of Boussinesq equations[6]:

$$
\begin{aligned}
\nabla \cdot \mathbf{u} &= 0 \\
\mathbf{u} \cdot \nabla \mathbf{u} + \nabla p - \mathrm{Pr} \nabla^2 \mathbf{u} - \mathrm{PrRa} T \hat{\mathbf{y}} &= 0 \\
\mathbf{u} \cdot \nabla T - \nabla^2 T &= 0
\end{aligned}
$$

If we fix $Pr$, then the parameter $\mu$ is Ra, the Rayleigh number.

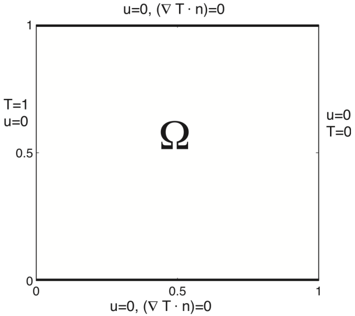
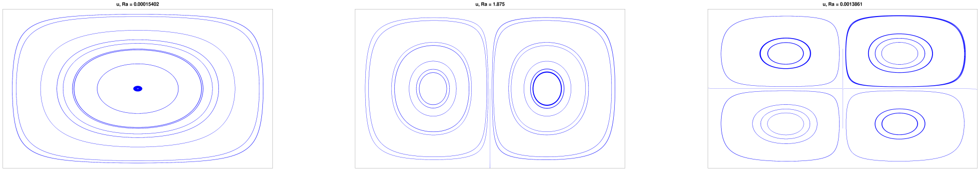



Figure: Taken from [GLRW12]

[6] See Tritton, *Physical Fluid Dynamics*, 1988.

# Rayleigh-Benard Convection




Increasing Ra $\longrightarrow$

# What we have and haven't said

- What we've shown:
  - Under certain conditions, Newton-Anderson accelerates Newton iterates in the singular case by the same mechanism seen in the nonsingular case.
  - With $\gamma$-safeguarding, Newton-Anderson converges locally, and in general faster than Newton (never worse).

# What we have and haven't said

- What we've shown:
  - Under certain conditions, Newton-Anderson accelerates Newton iterates in the singular case by the same mechanism seen in the nonsingular case.
  - With $\gamma$-safeguarding, Newton-Anderson converges locally, and in general faster than Newton (never worse).
- Open questions
  - Do these results extend to dim $N > 1$ case?

# What we have and haven't said

- What we've shown:
  - Under certain conditions, Newton-Anderson accelerates Newton iterates in the singular case by the same mechanism seen in the nonsingular case.
  - With $\gamma$-safeguarding, Newton-Anderson converges locally, and in general faster than Newton (never worse).
- Open questions
  - Do these results extend to dim $N > 1$ case?
  - What about for depth $m > 1$?

# What we have and haven't said

- What we've shown:
  - Under certain conditions, Newton-Anderson accelerates Newton iterates in the singular case by the same mechanism seen in the nonsingular case.
  - With $\gamma$-safeguarding, Newton-Anderson converges locally, and in general faster than Newton (never worse).
- Open questions
  - Do these results extend to dim $N > 1$ case?
  - What about for depth $m > 1$?
  - How small can $\theta_{k+1}$ be?

# References I

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.
*IJNAM*, In Print, 2023.

Keith J. Galvin, Alexander Linke, Leo G. Rebholz, and Nicholas E. Wilson.
Stabilizing poor mass conservation in incompressible flow problems with large irrotational forcing and application to thermal convectio n.
*Computer Methods in Applied Mechanics and Engineering*, 237-240:166–176, 2012.

J.M. Ortega.
The newton-kantorovich theorem.
*American Mathematical Monthly*, 1968.

# References II

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

# Thank you!

My support group