Convergence of the EDIIS Algorithm

C. T. Kelley
NC State University
tim_kelley@ncsu.edu
Supported by NSF, DOE(CASL/ORNL), ARO

CityU, May 2018

Outline

Collaborators

- My NCSU Students: Alex Toth, Austin Ellis
- Multiphysics coupling
    - ORNL: Steven Hamilton, Stuart Slattery, Kevin Clarno, Mark Berrill, Tom Evans, Jean-Luc Fattebert
    - Sandia: Roger Pawlowski, Alex Toth
- Electronic Structure Computations at NCSU
    - Jerry Bernholc, Emil Briggs, Miro Hodak, Elena Jakubikova, Wenchang Lu
- Hong Kong Polytechnic: Xiaojun Chen

Anderson Acceleration Algorithm

Solve fixed point problems

$$\mathbf{u} = \mathbf{G}(u)$$

faster than Picard iteration

$$\mathbf{u}_{k+1} = \mathbf{G}(\mathbf{u}_k).$$

Motivation (Anderson 1965) SCF iteration in electronic structure computations.

Why not Newton?

Newton's method

$$\mathbf{u}_{k+1} = \mathbf{u}_k - (I - \mathbf{G}'(\mathbf{u}_k))^{-1}(\mathbf{u}_k - \mathbf{G}(u_k))$$

- converges faster,
- does not require that $\mathbf{G}$ be a contraction,
- needs $\mathbf{G}'(\mathbf{u})$ or $\mathbf{G}'(\mathbf{u})\mathbf{w}$.

Sometimes you will not have $\mathbf{G}'$.

Electronic Structure Computations

Nonlinear eignevalue problem: Kohn-Sham equations

$$\mathbf{H}_{ks}[\psi_i] = -\frac{1}{2}\nabla^2\psi_i + V(\rho)\psi_i = \lambda_i\psi_i \quad i = 1, \dots, N$$

where the charge density is

$$\rho = \sum_{i=1}^{N} \|\psi_i\|^2.$$

Write this as

$$\mathbf{H}(\rho)\Psi = \Lambda\Psi$$

Self-Consistent Field iteration (SCF)

Given ρ

- Solve the linear eigenvalue problem

$$\mathbf{H}(\rho)\Psi = \Lambda\Psi$$

  for the N eigenvalues/vectors you want.
- Update the charge density via

$$\rho \leftarrow \sum_{i=1}^{N} \|\psi_i\|^2.$$

- Terminate if change in ρ is sufficiently small.

SCF as a fixed-point iteration

SCF is a fixed point iteration

$$\rho \leftarrow G(\rho)$$

Not clear how to differentiate G

- termination criteria in eigen-solver
- multiplicities of eigenvalues not know at the start

Bad news: you really have a fixed point problem in Ψ!

Multiphysics Coupling

Given several simulators: $\{S_j\}_{j=1}^{N_S}$

- The simulators depend on a partition $\{X_j\}_{j=1}^{N_S}$ of the primary variables
- S_i computes X_i as a function of $Z_i = \{X_j\}_{j \neq i}$
- The maps S_i could contain
  - Black-box solvers
  - Legacy codes
  - Table lookups
  - Internal stochastics
- Jacobian information very hard to get.

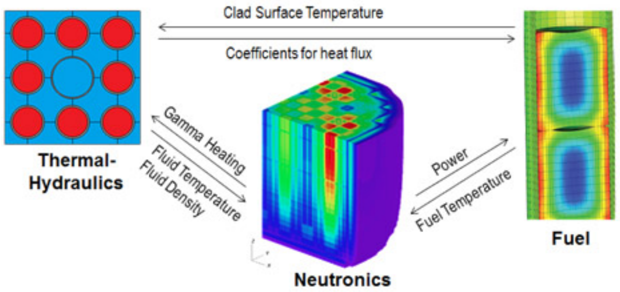
Iteration to self-consistency

Chose one X_i to **expose**. Then

- for $j = 1 : N_S, j \neq i$
  $X_j = S_j(Z_j)$
- $X_i \leftarrow S_i(Z_i)$

This is a fixed point problem

Example: $N_S = 3$


Clad Surface Temperature
Coefficients for heat flux
Thermal-
Hydraulics
Gamma Heating
Fluid Temperature
Fluid Density
Neutronics
Power
Fuel Temperature
Fuel

Anderson Acceleration

$\texttt{anderson}(\mathbf{u}_0, \mathbf{G}, m)$

$\mathbf{u}_1 = \mathbf{G}(\mathbf{u}_0)$; $\mathbf{F}_0 = \mathbf{G}(\mathbf{u}_0) - \mathbf{u}_0$

for $k = 1, \dots$ **do**

$m_k \le \min(m, k)$

$\mathbf{F}_k = \mathbf{G}(\mathbf{u}_k) - \mathbf{u}_k$

Minimize $\| \sum_{j=0}^{m_k} \alpha_j^k \mathbf{F}_{k-m_k+j} \|$ subject to $\sum_{j=0}^{m_k} \alpha_j^k = 1$.

$\mathbf{u}_{k+1} = \sum_{j=0}^{m_k} \alpha_j^k \mathbf{G}(\mathbf{u}_{k-m_k+j})$

end for

Other names for Anderson

- Pulay mixing (Pulay 1980)
- Direct iteration on the iterative subspace (DIIS) Rohwedder/Scheneider 2011
- Nonlinear GMRES (Washio 1997)

Terminology

- m, depth. We refer to Anderson(m).
  Anderson(0) is Picard.
- $\mathbf{F}(\mathbf{u}) = \mathbf{G}(\mathbf{u}) - \mathbf{u}$, residual
- $\{\alpha_j^k\}$, coefficients
  Minimize $\| \sum_{j=0}^{m_k} \alpha_j^k \mathbf{F}_{k-m_k+j} \|$ subject to $\sum_{j=0}^{m_k} \alpha_j^k = 1$.
  is the optimization problem.
- $\| \cdot \|$, ℓ^2, ℓ^1, or ℓ^∞

Solving the Optimization Problem

Solve the linear least squares problem:

$$\min \left\| \mathbf{F}_{m_k} - \sum_{j=0}^{m_k-1} \alpha_j^k (\mathbf{F}_{k-m_k+j} - \mathbf{F}_k) \right\|_2^2,$$

for $\{\alpha_j^k\}_{j=0}^{m_k-1}$ and then

$$\alpha_{m_k}^k = 1 - \sum_{j=0}^{m_k-1} \alpha_j^k.$$

More or less what's in the codes.

Details

- Many codes (RMG, for example) solve the normal equations. Not clear how bad that is.
- Using QR would be better. More on this later.
- LP solve for $\| \cdot \|_1$ and $\| \cdot \|_\infty$.
  That's bad for our customers.

Convergence Theory

- Most older work assumed unlimited storage or very special cases.
  - For unlimited storage, Anderson looks like a Krylov method and it is equivalent to GMRES (Walker-Ni 2011). See also (Potra 2012).
  - Anderson is also equivalent to a multi-secant quasi-Newton method (Fang-Saad + many others).
- In practice $m \leq 5$ most of the time and 5 is generous.
- The first general convergence results for the method as implemented in practice are ours.
- Convergence results have been local.

Convergence Results: Toth-K 2015

Critical idea: prove acceleration instead of convergence.

- Assume $\mathbf{G}$ is a contraction, constant c.
  Objective: do no worse than Picard
- Local nonlinear theory; $\|\mathbf{e}_0\|$ is small.
- Better results for $\| \cdot \|_2$.

Linear Problems, Toth, K 2015

Here

$$\mathbf{G}(\mathbf{u}) = \mathbf{M}\mathbf{u} + \mathbf{b}, \ \|\mathbf{M}\| \le c < 1, \text{ and } \mathbf{F}(\mathbf{u}) = \mathbf{b} - (\mathbf{I} - \mathbf{M})\mathbf{u}.$$

Theorem: $\|\mathbf{F}(\mathbf{u}_{k+1})\| \le c\|\mathbf{F}(\mathbf{u}_k)\|$

Proof: I

Since $\sum \alpha_j = 1$, the new residual is

$$
\begin{aligned}
\mathbf{F}(\mathbf{u}_{k+1}) &= b - (I - \mathbf{M})\mathbf{u}_{k+1} \\
&= \sum_{j=0}^{m_k} \alpha_j \left[b - (I - \mathbf{M})(b + \mathbf{M}\mathbf{u}_{k-m_k+j}) \right] \\
&= \sum_{j=0}^{m_k} \alpha_j \mathbf{M} \left[b - (I - \mathbf{M})\mathbf{u}_{k-m_k+j} \right] \\
&= \mathbf{M} \sum_{j=0}^{m_k} \alpha_j \mathbf{F}(\mathbf{u}_{k-m_k+j})
\end{aligned}
$$

Take norms to get ...

Proof: II

$$\|\mathbf{F}(\mathbf{u}_{k+1})\| \le c \left\| \sum_{j=0}^{m_k} \alpha_j \mathbf{F}(\mathbf{u}_{k-m_k+j}) \right\|$$

Optimality implies that

$$\left\| \sum_{j=0}^{m_k} \alpha_j \mathbf{F}(\mathbf{u}_{k-m_k+j}) \right\| \le \|\mathbf{F}(\mathbf{u}_k)\|.$$

That's it.
Use Taylor for the nonlinear case, which means local convergence.

Assumptions: $m = 1$

- There is $\mathbf{u}^* \in R^N$ such that $\mathbf{F}(\mathbf{u}^*) = \mathbf{G}(\mathbf{u}^*) - \mathbf{u}^* = 0$.
- $\|\mathbf{G}(\mathbf{u}) - \mathbf{G}(\mathbf{v})\| \le c\|u - v\|$ for $\mathbf{u}, \mathbf{v}$ near $\mathbf{u}^*$.
- $\mathbf{G}$ is continuously differentiable near $\mathbf{u}^*$

G *has a fixed point and is a smooth contraction in a neighborhood of that fixed point.*

Convergence for Anderson(1) with ℓ^2 optimization

Anderson(1) converges and

$$\limsup_{k \to \infty} \frac{\|\mathbf{F}(\mathbf{u}_{k+1})\|_2}{\|\mathbf{F}(\mathbf{u}_k)\|_2} \le c.$$

Very special case:

- Optimization problem is trivial.
- No iteration history to keep track of.

On the other hand . . .

Assumptions: $m > 1$, any norm

- The assumptions for $m = 1$ hold.
- There is M_α such that for all $k \ge 0$

$$\sum_{j=1}^{m_k} |\alpha_j| \le M_\alpha.$$

Do this by

- Hoping for the best.
- Reduce m_k until it happens.
- Reduce m_k for conditioning(?)

Convergence for Anderson(m), any norm.

Toth-K, Chen-K

If u_0 is sufficiently close to u^* then the Anderson iteration converges to u^* r-linearly with r-factor no greater than $\hat{c}$. In fact

$$\limsup_{k\to\infty} \left(\frac{\|\mathbf{F}(\mathbf{u}_k)\|}{\|\mathbf{F}(\mathbf{u}_0)\|} \right)^{1/k} \leq c. \tag{1}$$

Anderson acceleration is not an insane thing to do.

Comments

- The local part is serious and is a problem in the chemistry codes.
- No guarantee the convergence is monotone. See this in practice.
- The conditioning of the least squares problem can be poor. But that has only a small(???) effect on the results.
- The results do not completely reflect practice in that...
  - Theory seems to be sharp for some problems. But . . . convergence can sometimes be very fast. Why?
  - Convergence can depend on physics. The mathematics does not yet reflect that.
  - There are many variations in the chemistry/physics literature, which are not well understood.

EDIIS: Kudin, Scuseria, Cancès 2002

EDIIS (Energy DIIS) globalizes Anderson by constraining $\alpha_j^k \geq 0$. The optimization problem is

$$\text{Minimize } \left\| \mathbf{F}_k - \sum_{j=0}^{m_k-1} \alpha_j^k (\mathbf{F}_{k-m_k+j} - \mathbf{F}_k) \right\|_2^2 \equiv \| \mathbf{A}\alpha^k - \mathbf{F}_k \|_2^2.$$

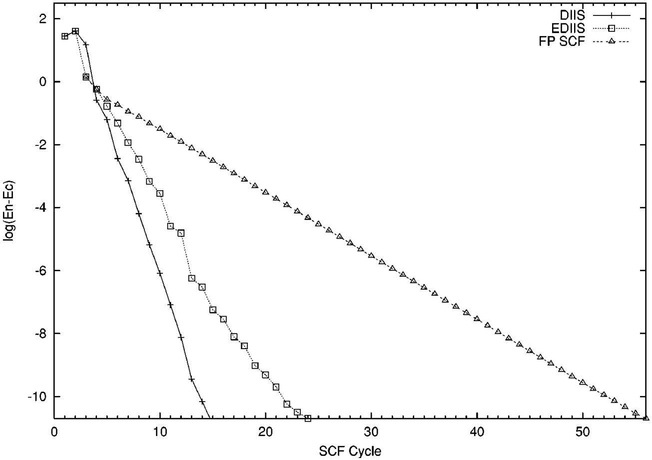
subject to

$$\sum_{j=0}^{m_k-1} \alpha_j^k \leq 1, \alpha_j^k \geq 0.$$

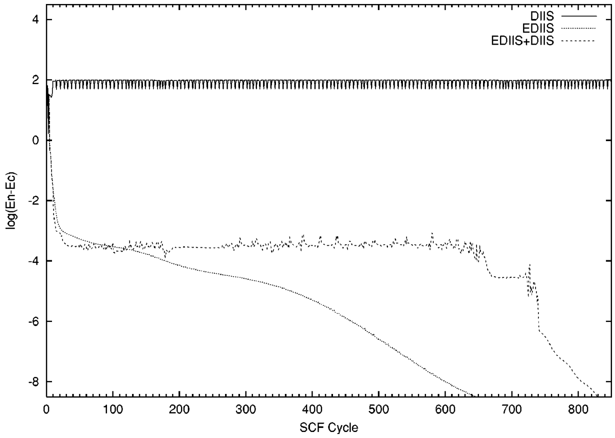
Solving the optimization problem

- Solve as a QP and we'd have to compute $\mathbf{A}^T\mathbf{A}$.
  $\mathbf{A}$ is often very ill-contitioned.
- We used QR before which exposed the ill-contitioning less badly.
- The Golub-Saunders active set method (1969!) does that.
- You're looking for the minimum in a smaller set, can that hurt?

Easy problem from Kudin et al



Hard problem from Kudin et al



Example from Radiative Transfer

Chandrasekhar H-equation

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

$\omega \in [0,1]$ is a physical parameter.
$F'(H^*)$ is singular when $\omega = 1$.

$$\rho(G'(H^*)) \leq 1 - \sqrt{1-\omega} < 1$$

Numerical Experiments

- Discretize with 500 point composite midpoint rule.
- Compare Newton-GMRES with Anderson(m).
- Terminate when $\|F(u_k)\|_2/\|F(u_0)\|_2 \le 10^{-8}$
- $\omega = .5, .99, 1.0$
- $0 \le m \le 6$
- ℓ^1, ℓ^2, ℓ^∞ optimizations
- Tabulate
  - κ_{max}: max condition number of least squares problems
  - S_{max}: max absolute sum of coefficients

Newton-GMRES vs Anderson(0)

Function evaluations:

	Newton-GMRES			Fixed Point		
ω	.5	.99	1.0	.5	.99	1.0
Fs	12	18	49	11	75	23970

Anderson(m)

		ℓ^1 Optimization			ℓ^2 Optimization			ℓ^∞ Optimization		
ω	m	Fs	κ_{max}	S_{max}	Fs	κ_{max}	S_{max}	Fs	κ_{max}	S_{max}
0.50	1	7	1.00e+00	1.4	7	1.00e+00	1.4	7	1.00e+00	1.5
0.99	1	11	1.00e+00	3.5	11	1.00e+00	4.0	10	1.00e+00	10.1
1.00	1	21	1.00e+00	3.0	21	1.00e+00	3.0	19	1.00e+00	4.8
0.50	2	6	1.36e+03	1.4	6	2.90e+03	1.4	6	2.24e+04	1.4
0.99	2	10	1.19e+04	5.2	10	9.81e+03	5.4	10	4.34e+02	5.9
1.00	2	18	1.02e+05	43.0	16	2.90e+03	14.3	34	5.90e+05	70.0
0.50	3	6	7.86e+05	1.4	6	6.19e+05	1.4	6	5.91e+05	1.4
0.99	3	10	6.51e+05	5.2	10	2.17e+06	5.4	11	1.69e+06	5.9
1.00	3	22	1.10e+08	18.4	17	2.99e+06	23.4	51	9.55e+07	66.7

Anderson(m)

		ℓ^1 Optimization			ℓ^2 Optimization			ℓ^∞ Optimization		
ω	m	Fs	κ_{max}	S_{max}	Fs	κ_{max}	S_{max}	Fs	κ_{max}	S_{max}
0.50	4	7	2.64e+09	1.5	6	9.63e+08	1.4	6	9.61e+08	1.4
0.99	4	11	1.85e+09	5.2	11	6.39e+08	5.4	11	1.61e+09	5.9
1.00	4	23	2.32e+08	12.7	21	6.25e+08	6.6	35	1.38e+09	49.0
0.50	5	7	1.80e+13	1.4	6	2.46e+10	1.4	6	2.48e+10	1.4
0.99	5	11	3.07e+10	5.2	12	1.64e+11	5.4	13	3.27e+11	5.9
1.00	5	21	2.56e+09	21.8	27	1.06e+10	14.8	32	4.30e+09	190.8
0.50	6	7	2.65e+14	1.4	6	2.46e+10	1.4	6	2.48e+10	1.4
0.99	6	12	4.63e+11	5.2	12	1.49e+12	5.4	12	2.27e+11	5.9
1.00	6	31	2.61e+10	45.8	35	1.44e+11	180.5	29	3.51e+10	225.7

Observations

- For $m > 0$, Anderson(m) is much better than Picard
- Anderson(m) does better than Newton GMRES
- There is little benefit in $m \geq 3$
- ℓ^∞ optimization seems to be a poor idea
- ℓ^1 optimization appears fine, but the cost is not worth it

Convergence of EDIIS: Chen-K 2017

If $\mathbf{G}$ is a contraction in convex Ω then

$$\|\mathbf{e}_k - \mathbf{u}^*\| \leq c^{k/(m+1)} \|\mathbf{e}_0 - \mathbf{u}^*\|$$

and the convergence is the same as the local theory when near $\mathbf{u}^*$ i.e.

$$\limsup_{k \to \infty} \left(\frac{\|\mathbf{F}(\mathbf{u}_k)\|}{\|\mathbf{F}(\mathbf{u}_0)\|} \right)^{1/k} \leq c.$$

Similar to global results for Newton's method.
Reflects practice reported by Kudin et al.

Example from Radiative Transfer

Chandrasekhar H-equation

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

$\omega \in [0,1]$ is a physical parameter.
$F'(H^*)$ is singular when $\omega = 1$.

$$\rho(G'(H^*)) \le 1 - \sqrt{1-\omega} < 1$$

Numerical Experiments

- Discretize with 500 point composite midpoint rule.
- Compare EDIIS/Anderson/Optimization problem methods
  - Matlab `lsqlin` active set (Golub-Saunders 1969)
  - Matlab `lsqlin` interior point (Coleman-Li 1994)
- Terminate when $\|F(u_k)\|_2/\|F(u_0)\|_2 \le 10^{-8}$
- $\omega = .5$

Table and Figure

- Tabulate
  - Computed convergence rate at terminal iteration k

$$\left(\frac{\|\mathbf{F}(\mathbf{h}_k)\|}{\|\mathbf{F}(\mathbf{h}_0)\|} \right)^{1/k}$$

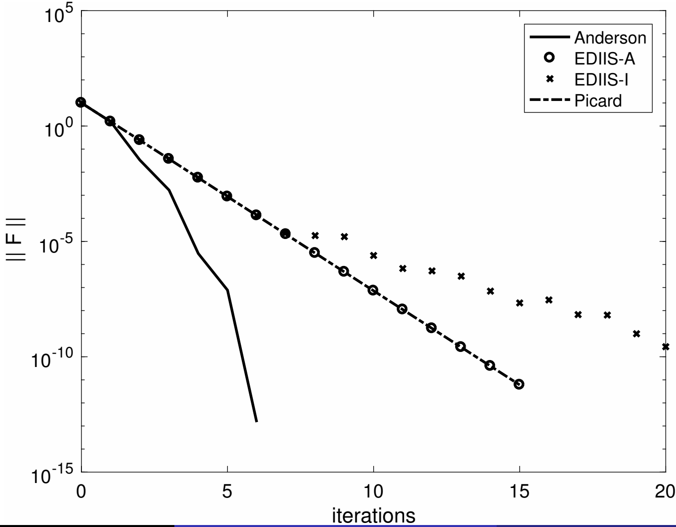
  - Spectral radius $\rho(\mathbf{G}'(H^*))$
- Plot: residiual histories

Convergence Rates

Anderson	Picard	EDIIS-A	EDIIS-I	$\rho(\mathcal{G}(H^*))$
1.06e-02	1.72e-01	1.72e-01	3.14e-01	2.93e-01

- Why is Picard so good? There's theory.
- Why is Anderson so good? There's no theory.

Residual Histories



Assumptions: $m = 1$

- There is $u^* \in R^N$ such that $\mathbf{F}(\mathbf{u}^*) = \mathbf{G}(\mathbf{u}^*) - \mathbf{u}^* = 0$.
- $\|\mathbf{G}(\mathbf{u}) - \mathbf{G}(v)\| \le c\|\mathbf{u} - \mathbf{v}\|$ for $\mathbf{u}, \mathbf{v}$ near $\mathbf{u}^*$.

Words: **G** has a fixed point and is a contraction.
We can do prove something without assuming differntiability ...

Convergence for Anderson(1) with ℓ^2 optimization

Let c be small enough so that

$$\hat{c} \equiv \frac{3c - c^2}{1 - c} < 1.$$

Let $c < \hat{c} < 1$ Anderson(1) converges and

$$\|\mathbf{F}(\mathbf{u}_{k+1})\|_2 \leq \hat{c}\|\mathbf{F}(\mathbf{u}_k)\|_2$$

Proof I

If $m = 1$ then

$$\mathbf{u}_{k+1} = (1 - \alpha^k)\mathbf{G}(\mathbf{u}_k) + \alpha^k \mathbf{G}(\mathbf{u}_{k-1}),$$

where

$$\alpha^k = \frac{\mathbf{F}(\mathbf{u}_k)^T(\mathbf{F}(\mathbf{u}_k) - \mathbf{F}(\mathbf{u}_{k-1}))}{\|\mathbf{F}(\mathbf{u}_k) - \mathbf{F}(\mathbf{u}_{k-1})\|^2}$$

Proof II

Write

$$\mathbf{F}(\mathbf{u}_{k+1}) = \mathbf{G}(\mathbf{u}_{k+1}) - u_{k+1} = A_k + B_k,$$

where

$$A_k = \mathbf{G}(\mathbf{u}_{k+1}) - \mathbf{G}((1-\alpha^k)u_k + \alpha^k u_{k-1})$$

and

$$B_k = \mathbf{G}((1-\alpha^k)u_k + \alpha^k u_{k-1}) - u_{k+1}.$$

We will estimate A_k and B_k separately to prove the estimate.

Proof III: Estimation of $\|A_k\|$

$$
\begin{aligned}
\|A_k\| &= \|\mathbf{G}(\mathbf{u}_{k+1}) - \mathbf{G}((1-\alpha^k)u_k + \alpha^k u_{k-1})\| \\
&\le c\|u_{k+1} - (1-\alpha^k)u_k - \alpha^k u_{k-1}\| \\
&= c\|(1-\alpha^k)(\mathbf{G}(\mathbf{u}_k) - u_k) - \alpha^k(\mathbf{G}(\mathbf{u}_{k-1}) - u_{k-1})\| \\
&= c\|(1-\alpha^k)\mathbf{F}(\mathbf{u}_k) - \alpha^k\mathbf{F}(\mathbf{u}_{k-1})\| \le c\|\mathbf{F}(\mathbf{u}_k)\|,
\end{aligned}
$$

The last inequality follows from optimality of the coefficients.

Proof IV: Estimation of $\|B_k\|$

To begin

$$\begin{aligned} B_k &= \mathbf{G}((1-\alpha^k)u_k + \alpha^k u_{k-1}) - (1-\alpha^k)\mathbf{G}(\mathbf{u}_k) - \alpha^k \mathbf{G}(\mathbf{u}_{k-1}) \\ &= \mathbf{G}(\mathbf{u}_k + \alpha^k \delta_k) - \mathbf{G}(\mathbf{u}_k) + \alpha^k(\mathbf{G}(\mathbf{u}_k) - \mathbf{G}(\mathbf{u}_{k-1})) \end{aligned}$$

Using contractivity

$$\|B_k\| \le 2c|\alpha^k| \, \|\delta_k\|.$$

Next, estimate the product $|\alpha^k|\|\delta_k\|$.

Proof VI: Estimation of $\|B_k\|$

The difference in residuals is

$$\mathbf{F}(\mathbf{u}_k) - \mathbf{F}(\mathbf{u}_{k-1}) = \mathbf{G}(\mathbf{u}_k) - \mathbf{G}(\mathbf{u}_{k-1}) + \delta_k.$$

Using contractivity $\|\mathbf{G}(\mathbf{u}_k) - \mathbf{G}(\mathbf{u}_{k-1})\| \leq c\|\delta_k\|$ we obtain

$$\|\mathbf{F}(\mathbf{u}_k) - \mathbf{F}(\mathbf{u}_{k-1})\| \geq (1 - c)\|\delta_k\|.$$

Hence

$$\|\delta_k\| \leq \|\mathbf{F}(\mathbf{u}_k) - \mathbf{F}(\mathbf{u}_{k-1})\|/(1 - c).$$

Proof VII: Final result

Finally, use the formula for α^k to obtain

$$|\alpha^k| \|\delta_k\| \le \frac{\|\mathbf{F}(\mathbf{u}_k)\|}{\|\mathbf{F}(\mathbf{u}_k) - \mathbf{F}(\mathbf{u}_{k-1})\|} \|\delta_k\| \le \frac{\|\mathbf{F}(\mathbf{u}_k)\|}{1-c}.$$

So

$$\begin{aligned} \|\mathbf{F}(\mathbf{u}_{k+1})\| & \le c\|\mathbf{F}(\mathbf{u}_k)\| + \tfrac{2c\|\mathbf{F}(\mathbf{u}_k)\|}{1-c} \\ & = \tfrac{3c-c^2}{1-c}\|\mathbf{F}(\mathbf{u}_k)\| = \hat{c}\|\mathbf{F}(\mathbf{u}_k)\|. \end{aligned}$$

Smooth Case

Assume that $\mathbf{G}'$ is Lipschitz continuous. Then if $\|e_0\|$ is sufficiently small Anderson(1) converges and

$$\limsup_{k\to\infty} \frac{\|\mathbf{F}(\mathbf{u}_{k+1})\|_2}{\|\mathbf{F}(\mathbf{u}_k)\|_2} \le c.$$

Proof I: Exploiting smoothness

The only difference is the estimate for B_k. Using the differentiability assumption

$$
\begin{aligned}
B_k &= \mathbf{G}((1-\alpha^k)u_k + \alpha^k u_{k-1}) - (1-\alpha^k)\mathbf{G}(\mathbf{u}_k) - \alpha^k \mathbf{G}(\mathbf{u}_{k-1}) \\
&= \mathbf{G}(\mathbf{u}_k + \alpha^k \delta_k) - \mathbf{G}(\mathbf{u}_k) + \alpha^k(\mathbf{G}(\mathbf{u}_k) - \mathbf{G}(\mathbf{u}_{k-1})) \\
&= \int_0^1 \mathbf{G}'(\mathbf{u}_k + t\alpha^k \delta_k)\alpha^k \delta_k \, dt - \alpha^k \int_0^1 \mathbf{G}'(\mathbf{u}_k + t\delta_k)\delta_k \, dt \\
&= \alpha^k \int_0^1 \left[\mathbf{G}'(\mathbf{u}_k + t\alpha^k \delta_k) - \mathbf{G}'(\mathbf{u}_k + t\delta_k)\right] \delta_k \, dt.
\end{aligned}
$$

Proof II: Lipschitz continuity of $\mathbf{G}'$

So, if γ is the Lipschitz constant of $\mathbf{G}'$,

$$\|B_k\| \le \gamma|\alpha^k||(1-\alpha^k)|\|\delta_k\|^2/2.$$

By definition,

$$|\alpha^k||1-\alpha^k| \le \frac{\|\mathbf{F}(\mathbf{u}_k)\|\|\mathbf{F}(\mathbf{u}_{k-1})\|}{\|\mathbf{F}(\mathbf{u}_k)-\mathbf{F}(\mathbf{u}_{k-1})\|^2}.$$

Contractivity implies that

$$\|\mathbf{F}(\mathbf{u}_k)-\mathbf{F}(\mathbf{u}_{k-1})\| \ge (1-c)\|\delta_k\|$$

So ...

Proof III: Final estimate

$$
\begin{aligned}
\|\mathbf{F}(\mathbf{u}_{k+1})\| &\le \|A_k\| + \|B_k\| \\
&\le c\|\mathbf{F}(\mathbf{u}_k)\| + \frac{\gamma\|\mathbf{F}(\mathbf{u}_k)\|\|\mathbf{F}(\mathbf{u}_{k-1})\|}{2(1-c)^2} \\
&= \|\mathbf{F}(\mathbf{u}_k)\|(c + O(\|\mathbf{F}(\mathbf{u}_{k-1})\|))
\end{aligned}
$$

proving the result if e_0 is sufficiently small.

Can we use semi-smoothness to do this?

What do we need to get ...

$$
\begin{aligned}
B_k &= \mathbf{G}((1-\alpha^k)u_k + \alpha^k u_{k-1}) - (1-\alpha^k)\mathbf{G}(\mathbf{u}_k) - \alpha^k \mathbf{G}(\mathbf{u}_{k-1}) \\
&= \mathbf{G}(\mathbf{u}_k + \alpha^k \delta_k) - \mathbf{G}(\mathbf{u}_k) + \alpha^k(\mathbf{G}(\mathbf{u}_k) - \mathbf{G}(\mathbf{u}_{k-1})) \\
&= o(\|\mathbf{F}(\mathbf{u}_k)\|)?
\end{aligned}
$$

Continuity of $\mathbf{G}'$ is enough.

References

- D. G. Anderson,
  Iterative Procedures for Nonlinear Integral Equations, Journal of the ACM, 12 (1965), pp. 547–560.
- P. Pulay,
  Convergence acceleration of iterative sequences. The case of SCF iteration. Chemical Physics Letters, 73 (1980), pp. 393–398.
- K. N. Kudin, G. E. Scuseria, and E. Cancès,
  A black-box self-consistent field convergence algorithm: One step closer, Journal of Chemical Physics, 116 (2002), pp. 8255–8261,

References

- A. Toth and C. T. Kelley, Convergence analysis for Anderson acceleration, SIAM J. Numer. Anal., 53 (2015), pp. 805 – 819.
- A. Toth, J. A. Ellis, T. Evans, S. Hamilton, C. T. Kelley, R. Pawlowski, and S. Slattery, Local improvement results for Anderson acceleration with inaccurate function evaluations, 2016. To appear in SISC.
- S. Hamilton, M. Berrill, K. Clarno, R. Pawlowski, A. Toth, C. T. Kelley, T. Evans, and B. Philip, An assessment of coupling algorithms for nuclear reactor core physics simulations, Journal of Computational Physics, 311 (2016), pp. 241–257.
- X. Chen, C. T. Kelley, and players to be named, Analysis and Implementation of EDIIS, in progress.

Summary

- Proofs use derivatives
  - Can semi-smooth analysis do the job?
- EDIIS has a harder optimization problem