Anderson Acceleration: An Overview

Homer Walker (Emeritus)
Mathematical Sciences Department
Worcester Polytechnic Instititute
walker@wpi.edu

Slides downloadable at www.wpi.edu/∼walker.

Derived from a method of D. G. Anderson (1965).

Consider a fixed-point iteration: $x_{k+1} = g(x_k)$, $g : \mathbb{R}^n \to \mathbb{R}^n$.

Anderson Acceleration

Given x_0 and $m \geq 1$, set $x_1 = g(x_0)$.

For $k = 1, 2, \ldots$

Set $m_k = \min\{m, k\}$.

Set $F_k = \left[f(x_{k-m_k}), \ldots, f(x_k)\right]$, where $f(x) \equiv g(x) - x$.

Solve $\min_{\alpha \in \mathbb{R}^{m_k+1}} \|F_k \alpha\|_2$ s. t. $\sum_{i=0}^{m_k} \alpha_i = 1$.

Set $x_{k+1} = \sum_{i=0}^{m_k} \alpha_i g(x_{k-m_k+i})$.

Can allow a damped step: $x_{k+1} = (1 - \beta_k) \sum_{i=0}^{m_k} \alpha_i x_{k-m_k+i} + \beta_k \sum_{i=0}^{m_k} \alpha_i g(x_{k-m_k+i})$.

Suppose $g(x) = Ax + b$ for $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$.

Then $x_{k+1} = \sum_{i=0}^{m_k} \alpha_i g(x_{k-m_k+i}) = g(\sum_{i=0}^{m_k} \alpha_i x_{k-m_k+i})$, or

$$x_{k+1} = g(x_{min})$$

where $x_{min} = \sum_{i=0}^{m_k} \alpha_i x_{k-m_k+i}$ has minimal residual within the affine subspace containing $\{x_{k-m_k+i}\}_{i=0,\ldots,m_k}$.

Perspectives

- In a broad category with . . .
  - "charge-mixing" methods for electronic-structure computations;
    Pulay (1980, 1982), Kerker (1981), Eyert (1996), Thomas–Fermi (Raczkowski et al. 2001), Broyden (Kresse–Furthmüller 1996), C. Yang et al. (2008)
  - methods based on quasi-Newton updating;
    Eirola–Nevanlinna (1989), U. Yang(1995), Eyert (1996), Fang–Saad (2008), Degroote et al. (2009), Haelterman et al. (2009, 2010)
  - Krylov acceleration methods.
    Washio–Oosterlee (1997), Carlson–Miller (1998), Oosterlee–Washio (2000), Hager–Zhang (2006)
- Essentially (or nearly) the same method has been independently described several times.
  Pulay (1980, 1982), Washio–Oosterlee (1997), Carlson–Miller (1998), Oosterlee–Washio (2000), Degroote et al. (2009), Haelterman et al. (2009, 2010)
- Various names: Anderson mixing, nonlinear GMRES, Krylov acceleration, DIIS or Pulay mixing, QN-ILS.

Perspectives (cont.)

Another broad category ...

- *vector-extrapolation* methods, especially polynomial methods: (reduced-rank, minimal-polynomial, modified minimal-polynomial);
- *vector* and *topological* ε*-algorithms*.

See the book by Brezinski & Redivo-Zaglia (1991); surveys by Brezinski (2000), Jbilou–Sadok (2000), Smith–Ford–Sidi (1987), ...

Brezinski, Redivo-Zaglia, & Saad (2018) formulate a variant of reduced-rank extrapolation (RRE) that is “similar in spirit, but not quite equivalent, to AA.”

Anderson Acceleration and GMRES

Assume ...

- $g(x) = Ax + b$ for $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$.
- Anderson acceleration is not truncated, i.e., $m_k = k$ for each k.
- $(I - A)$ is nonsingular.
- (Unrestarted) GMRES is applied to $(I - A)x = b$ with initial point x_0.

W–Ni (2011)

Suppose also that, for some $k > 0$, $r_{k-1}^{GMRES} \neq 0$ and $\|r_{j-1}^{GMRES}\|_2 > \|r_j^{GMRES}\|_2$ for $0 < j < k$. Then

$$\sum_{i=0}^{k} \alpha_i x_i^{AA} = x_k^{GMRES} \text{ and } x_{k+1}^{AA} = g(x_k^{GMRES}).$$

Potra & Engler (2013) extended to allow a damped step.

Anderson Acceleration and GMRES (cont.)

Consider ...

- $Ax = b$ for nonsingular $A \in \mathbb{R}^n$.
- $A = M - N$ for nonsingular $M \in \mathbb{R}^{n\times n}$.
- Stationary iteration $x_{k+1} = g(x_k) \equiv M^{-1}Nx_k + M^{-1}b$.

Assume ...

- Anderson acceleration is not truncated, i.e., $m_k = k$ for each k.
- GMRES is applied to $M^{-1}Ax = M^{-1}b$ with initial point x_0.

Corollary

Suppose also that, for some $k > 0$, $r_{k-1}^{GMRES} \neq 0$ and $\|r_{j-1}^{GMRES}\|_2 > \|r_j^{GMRES}\|_2$ for $0 < j < k$. Then

$$\sum_{i=0}^{k} \alpha_i x_i^{AA} = x_k^{GMRES} \text{ and } x_{k+1}^{AA} = g(x_k^{GMRES}).$$

Anderson Acceleration and Multi-Secant Updating

Fang–Saad (2008)

Successive Anderson acceleration iterates are related by

$$x_{k+1} = x_k - B_k^{-1} f(x_k), \qquad (\star)$$

where $f(x) \equiv g(x) - x$ *and* B_k *is the second Broyden multi-secant update of* $-I$ *satisfying*

$$B_k(x_{i+1} - x_i) = f(x_{i+1}) - f(x_i), \qquad k - m_k \leq i \leq k-1. \qquad (\star\star)$$

- Clarifies and extends work of Eyert (1996).
- Degroote et al. (2009) and Haelterman at al. (2009, 2010) have independently developed very similar QN-LS and QN-ILS methods using multi-secant updating.

Anderson Acceleration and Multi-Secant Updating (cont.)

- Fang–Saad (2008) define an *Anderson family* of methods of the form ($\star$), with Anderson acceleration designated the *Type II* method.

- The Anderson family *Type I* method has the form ($\star$), in which B_k is the first Broyden multi-secant update of $-I$ satisfying ($\star\star$).

- For affine g, W–Ni (2011) give results relating the Type I method to the Arnoldi method (FOM) analogous to those relating AA to GMRES:

$$\sum_{i=0}^{k} \alpha_i x_i^{Type\ I} = x_k^{Arnoldi} \ \textit{and}\ x_{k+1}^{Type\ I} = g(x_k^{Arnoldi}).$$

Anderson Acceleration and Multi-Secant Updating (cont.)

Can this updating approach be used to exploit problem structure?

Anderson Acceleration and Multi-Secant Updating (cont.)

Can this updating approach be used to exploit problem structure?

Set $\mathcal{X}_k = (\Delta x_{k-m_k}, \ldots, \Delta x_{k-1})$, $\mathcal{F}_k = (\Delta f_{k-m_k}, \ldots, \Delta f_{k-1})$.

Suppose we want an update of $-I$ satisfying $B_k \mathcal{X}_k = \mathcal{F}_k$ and $B_k = B_k^T$.

Anderson Acceleration and Multi-Secant Updating (cont.)

Can this updating approach be used to exploit problem structure?

Set $\mathcal{X}_k = (\Delta x_{k-m_k}, \ldots, \Delta x_{k-1})$, $\mathcal{F}_k = (\Delta f_{k-m_k}, \ldots, \Delta f_{k-1})$.

Suppose we want an update of $-I$ satisfying $B_k \mathcal{X}_k = \mathcal{F}_k$ and $B_k = B_k^T$.

Schnabel (1983): $\exists\ B_k$ such that $B_k \mathcal{X}_k = \mathcal{F}_k$ and $B_k = B_k^T \iff \mathcal{X}_k^T \mathcal{F}_k = \mathcal{F}_k^T \mathcal{X}_k$.

Anderson Acceleration and Multi-Secant Updating (cont.)

Can this updating approach be used to exploit problem structure?

Set $\mathcal{X}_k = (\Delta x_{k-m_k}, \ldots, \Delta x_{k-1})$, $\mathcal{F}_k = (\Delta f_{k-m_k}, \ldots, \Delta f_{k-1})$.

Suppose we want an update of $-I$ satisfying $B_k \mathcal{X}_k = \mathcal{F}_k$ and $B_k = B_k^T$.

Schnabel (1983): $\exists\ B_k$ such that $B_k \mathcal{X}_k = \mathcal{F}_k$ and $B_k = B_k^T \iff \mathcal{X}_k^T \mathcal{F}_k = \mathcal{F}_k^T \mathcal{X}_k$.

Boutet et al. (2019, 2021a, 2021b) are exploring modified strategies in a more general multi-secant updating context.

Convergence

Assume ...

- $g : D \to D$ for closed $D \subseteq \mathbb{R}^n$;
- g is continuously differentiable on D;
- $\exists\ \kappa\ \in (0,1)$ such that $\|g(y) - g(x)\| \leq \kappa \|y - x\|$ for x, $y \in D$.

This implies ...

- $\exists!\ x_* \in D$ such that $x_* = g(x_*)$.
- Fixed-point iterates converge globally with $\|x_{k+1} - x_*\| \leq \kappa \|x_k - x_*\|$.
- $\|g'(x)\| \leq \kappa$ in D and $I - g'(x)$ is invertible in D.

Chen–Kelley (2019)

Suppose also that $\exists\ M$ such that $\sum_{i=0}^{m_k} |\alpha_i| \leq M$ for all k. Then the AA residuals and iterates converge locally and r-linearly to x_* with

$$\limsup_{k\to\infty} \|f(x_k)\|^{1/k} = \limsup_{k\to\infty} \|x_k - x_*\|^{1/k} \leq \kappa.$$

Convergence (cont.)

Toth–Kelley (2015) give a slightly weaker r-linear convergence result, assuming g is Lipschitz continuously differentiable. They also ...

- show that the convergence is q-linear if either (a) $\|\cdot\| = \|\cdot\|_2$ and $m_k = 1$ for all k, or (b) g is linear;
- discuss the condition $\sum_{i=0}^{m_k} |\alpha_i| \leq M$ and give strategies for enforcing it;
- discuss using norms other than $\|\cdot\|_2$ in the minimization problem for the α_i.

Chen–Kelley (2019) consider EDIIS from Kudin et al. (2002), described as differing from AA by requiring each $\alpha_i \geq 0$.

Assuming only convexity of D and contractivity of g in D, they show global r-linear convergence in D, with

$$\|x_k - x_*\| \leq \left(\kappa^{1/(m+1)}\right)^k \|x_0 - x_*\|.$$

Convergence (cont.)

Assume ...

- $g : \mathbb{R}^n \to \mathbb{R}^n$ is uniformly Lipschitz continuously differentiable;
- $\exists\ \kappa \in (0,1)$ such that $\|g(y) - g(x)\| \le \kappa \|y - x\|$ for $x, y \in \mathbb{R}^n$.

(Implications as before.)

Evans et al. (2020)

Suppose also that $\exists\ M$ and $\epsilon > 0$ such that for all $k > m$, $\sum_{i=0}^{m-1} |\alpha_i| \le M$ and $|\alpha_m| \ge \epsilon$. Then

$$\|f(x_{k+1})\| \le \theta_{k+1} \left[(1 - \beta_k) + \kappa\beta_k\right] \|f(x_k)\| + \sum_{i=0}^{m} \mathcal{O}\left(\|f(x_{k-m+i})\|^2\right), \qquad (\star\star\star)$$

where β_k = damping parameter and $\theta_{k+1} = \| \sum_{i=0}^{m} \alpha_i f(x_{k-m+i}) \| / \|f(x_k)\|$.

Note: $\theta_{k+1} \le 1$.

Convergence (cont.)

With $\beta_k = 1$, $(\star\star\star)$ gives

$$\|f(x_{k+1})\| \leq \kappa \left\|\sum_{i=0}^{m} \alpha_i f(x_{k-m+i})\right\| + h.o.t. \leq \kappa \|f(x_k)\| \; + \; h.o.t.$$

Since $f(x) = [g'(x_*) - I](x - x_*) + h.o.t$, this gives

$$\|x_{k+1} - x_*\|_* \leq (\kappa + \delta)\|x_k - x_*\|_* \; + \; h.o.t.,$$

where $\|v\|_* \equiv \|f'(x_*)v\| = \|[I - g'(x_*)]v\|$ and $\delta > 0$ is arbitrarily small.

Convergence (cont.)

With $\beta_k = 1$, $(\star\star\star)$ gives

$$\|f(x_{k+1})\| \le \kappa \left\|\sum_{i=0}^{m} \alpha_i f(x_{k-m+i})\right\| + h.o.t. \le \kappa \|f(x_k)\| \ + \ h.o.t.$$

Since $f(x) = [g'(x_*) - I](x - x_*) + h.o.t$, this gives

$$\|x_{k+1} - x_*\|_* \le (\kappa + \delta)\|x_k - x_*\|_* \ + \ h.o.t.,$$

where $\|v\|_* \equiv \|f'(x_*)v\| = \|[I - g'(x_*)]v\|$ and $\delta > 0$ is arbitrarily small.

With continuous differentiability and contractivity, we have directly

$$\|x_{k+1} - x_*\| \ = \ \left\|\sum_{i=0}^{m} \alpha_i [g(x_{k-m+i}) - g(x_*)]\right\| \le \left\|\sum_{i=0}^{m} \alpha_i g'(x_*)(x_{k-m+i} - x_*)\right\| + h.o.t.$$

$$\le \ \kappa \left\|\sum_{i=0}^{m} \alpha_i x_{k-m+i} - x_*\right\| + h.o.t.$$

If g is affine and AA is not truncated, then $\left\|x_{k+1}^{AA} - x_*\right\| \le \kappa \left\|x_k^{GMRES} - x_*\right\|$.

Convergence (cont.)

With $\beta_k = 1$, $(\star\star\star)$ gives

$$\|f(x_{k+1})\| \leq \kappa \left\|\sum_{i=0}^{m} \alpha_i f(x_{k-m+i})\right\| + h.o.t. \leq \kappa \|f(x_k)\| \ + \ h.o.t.$$

Since $f(x) = [g'(x_*) - I](x - x_*) + h.o.t$, this gives

$$\|x_{k+1} - x_*\|_* \leq (\kappa + \delta)\|x_k - x_*\|_* \ + \ h.o.t.,$$

where $\|v\|_* \equiv \|f'(x_*)v\| = \|[I - g'(x_*)]v\|$ and $\delta > 0$ is arbitrarily small.

With continuous differentiability and contractivity, we have directly

$$\begin{aligned}\|x_{k+1} - x_*\| &= \left\|\sum_{i=0}^{m} \alpha_i [g(x_{k-m+i}) - g(x_*)]\right\| \leq \left\|\sum_{i=0}^{m} \alpha_i g'(x_*)(x_{k-m+i} - x_*)\right\| + h.o.t. \\ &\leq \kappa \left\|\sum_{i=0}^{m} \alpha_i x_{k-m+i} - x_*\right\| + h.o.t.\end{aligned}$$

If g is affine and AA is not truncated, then $\left\|x_{k+1}^{AA} - x_*\right\| \leq \kappa \left\|x_k^{GMRES} - x_*\right\|$.

Caution: All $h.o.t.$ depend on x_{k-m+i} for $i = 0, \ldots, m$.

Wrapping up . . .

- Have outlined major theoretical properties of AA and its relationship to other algorithms.
- Promising areas of research (IMHO):
  - exploiting problem structure in AA,
  - sharpening local convergence results.
- Slides downloadable at `www.wpi.edu/~walker`.

Wrapping up . . .

- Have outlined major theoretical properties of AA and its relationship to other algorithms.
- Promising areas of research (IMHO):
  - exploiting problem structure in AA,
  - sharpening local convergence results.
- Slides downloadable at `www.wpi.edu/~walker`.

Thank you!

References

D. G. Anderson (1965), *Iterative Procedures for Nonlinear Integral Equations*, J. Assoc. Comput. Machinery, vol. 12, 547-560.

N. Boutet, R. Haelterman, and J. Degroote (2019), *Secant update version of quasi-Newton PSB with weighted multisecant equations*, Comput. Optim. Appl., vol. 75, 441-466.

N. Boutet, R. Haelterman, and J. Degroote (2021a) , *Secant Update generalized version of PSG: a new approach*, Comput. Optim. Appl., vol. 78, 953-982.

N. Boutet, R. Haelterman, and J. Degroote (2021b), *New approach for Secant Update generalized version of PSB*, 9th International Conference on System Modeling and Optimization, Budapest. Available at http://hdl.handle.net/1854/LU-8699524.

C. Brezinski, M. Redivo–Zaglia, and Y. Saad (2018), *Shanks sequence transfomations and Anderson acceleration*, SIAM Review, vol. 60, 646-669.

N. N. Carlson and K. Miller (1998), *Design and application of a gradient-weighted moving finite element code I: in one dimension*, SIAM J. Sci. Comput., vol. 19, 728-765.

X. Chen and C. T. Kelley (2019), *Convergence of the EDIIS algorithm for nonlinear equations*, SIAM J. Sci. Comput., vol. 41, A365-A379.

J. E. Dennis, Jr., and H. F. Walker (1981), *Convergence theorems for least-change secant update methods*, SIAM J. Numer. Anal., vol. 18, 949-987.

J. Degroote, K.-J. Bathe, D. Van Heule, and J. Vierendeels (2009), *Performance of a new partitioned procedure versus a monolithic procedure in fluid-structure interaction*, Comput. & Structures, vol. 87, 798-801.

T. Eirola and O. Nevanlinna (1989), *Accelerating with rank-one updates*, Lin. Alg. Appl., vol. 121, 511-520.

C. Evans, S. Pollock, L. G. Rebholz, and M. Xiao (2020), *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., vol. 58, 788-810.

References (cont.)

V. Eyert (1996), *A comparative study on methods for convergence acceleration of iterative vector sequences*, J. Comput. Phys., vol. 124, 271-285.

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

J.. Greenstadt (1970), *Variations on variable-metric methods*, Math. Comp., vol. 24, 1-18.

H. Haelterman, J. Degroote, D. Van Heule, & J. Vierendeels (2009), *The quasi-Newton least squares method: a new and fast secant method analyzed for linear systems*, SIAM J. Numer. Anal.,vol. 47, 2347-2368.

H. Haelterman, J. Degroote, D. Van Heule, & J. Vierendeels (2010), *On the similarities between the quasi-Newton inverse least squares method and GMRes*, SIAM J. Numer. Anal., vol. 47, 4660-4679.

W. W. Hager & H. Zhang (2006), *A survey of nonlinear conjugate gradient methods*, Pac. J. Optimization, vol. 2, 35-58.

G. P. Kerker (1981), *Efficient iteration scheme for self-consistent pseudopotential calculations*, Phys. Rev. B, vol. 23, 3082-3084.

G. Kresse and J. Furthmüller (1996), *Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set*, Phys. Rev. B, vol. 54, 169-186.

K. N. Kudin, G. E. Scuseria, and E. Cancés (2002), *A black-box self-consistent field convergence algorithm: One step closer*, J. Chem. Phys., vol. 116, 8255-8261.

P. A. Lott, H. F. Walker, C. S. Woodward, and U. M. Yang (2012), *An accelerated Picard method for nonlinear systems related to variably saturated flow*, Adv. Water Resour., vol. 38, 92-101.

C. W. Oosterlee and T. Washio (2000), *Krylov subspace acceleration of nonlinear multigrid with application to recirculating flows*, SIAM J. Sci. Comput., vol. 21, 1670-1690.

F. A. Potra and H. Engler (2013), *A characterization of the behavior of the Anderson acceleration on linear problems*, Lin. Alg. Appl., vol. 438, 1002-1011.

References (cont.)

M. J. D. Powell (1970), *A new algorithm for unconstrained optimization*, in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter, eds., Academic Press, New York.

P. Pulay (1980), *Convergence acceleration of iterative sequences. The case of SCF iteration*, Chem. Phys. Letters, vol. 73, 393-398.

P. Pulay (1982), *Improved SCF convergence*, J. Computational Chem., vol. 3, 556-560.

D. Raczkowski, A. Canning, and L. W. Wang (2001), *Thomas–Fermi charge mixing for obtaining self-consistency in density functional calculations*, Phys. Rev. B, vol. 64, 121101 (R).

R. B. Schnabel (1983), *Quasi-Newton Methods Using Multiple Secant Equations*, University of Colorado at Boulder Department of Computer Science Report CU-CS-247-83.

A. Toth and C. T. Kelley (2015), *Convergence analysis for Anderson acceleration*, SIAM J. Numer. Anal., vol 53, 805-819.

A. Toth, J. A. Ellis, T. Evans, S. Hamilton, C. T. Kelley, R. Pawlowski, and S. Slattery (2017), *Local improvement results for Anderson acceleration with inaccurate function evaluations*, SIAM J. Sci. Comput., vol. 39, 847-865.

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

T. Washio and C. W. Oosterlee (1997), *Krylov subspace acceleration for nonlinear multigrid schemes*, Electronic Trans. Numer. Anal., vol. 6, 271-290.

C. Yang, J. C. Meza, B. Lee, and L.-W. Wang (2009), *KSSOLV—a MATLAB toolbox for solving the Kohn–Sham equations*, ACM Trans. Math. Softw., vol. 36, 1-35.

U. M. Yang (1995), *A Family of Preconditioned Iterative Solvers for Sparse Linear Systems*, PhD thesis, Department of Computer Science, University of Illinois at Urbana-Champaign.