SIAM J. NUMER. ANAL.
Vol. 53, No. 2, pp. 805–819

CONVERGENCE ANALYSIS FOR ANDERSON ACCELERATION[*]

ALEX TOTH[†] AND C. T. KELLEY[†]

Abstract. Anderson(m) is a method for acceleration of fixed point iteration which stores $m+1$ prior evaluations of the fixed point map and computes the new iteration as a linear combination of those evaluations. Anderson(0) is fixed point iteration. In this paper we show that Anderson(m) is locally r-linearly convergent if the fixed point map is a contraction and the coefficients in the linear combination remain bounded. Without assumptions on the coefficients, we prove q-linear convergence of Anderson(1) and, in the case of linear problems, Anderson(m). We observe that the optimization problem for the coefficients can be formulated and solved in nonstandard ways and report on numerical experiments which illustrate the ideas.

Key words. nonlinear equations, Anderson acceleration, local convergence

AMS subject classification. 65H10

DOI. 10.1137/130919398

1. Introduction.

Anderson acceleration (also known as Anderson mixing) [1] is essentially the same as Pulay mixing (also known as direct inversion on the iterative subspace, or DIIS) [22, 28, 29, 30, 32] and the nonlinear GMRES method [4, 23, 25, 34]. The method has been widely used to accelerate the self-consistent field (SCF) iteration in electronic structure computations. Recent papers [12, 27, 30, 31, 33] show that the method is related to multisecant quasi-Newton methods or, in the case of linear problems, GMRES. None of these results leads to a convergence proof, even in the linear case, unless the available storage is large enough to allow GMRES to take a number of iterations equal to the dimension of the problem. Anderson acceleration does not require the computation or approximation of Jacobians or Jacobian-vector products, and this can be an advantage over Newton-like methods.

Anderson acceleration is a method for solving fixed point problems

$$u = G(u), \tag{1.1}$$

where $u \in R^N$ and $G : R^N \to R^N$. In this paper we prove that Anderson iteration converges when G is a contraction, which implies that the conventional fixed point iteration

$$u_{k+1} = G(u_k) \tag{1.2}$$

also converges. In the linear case we show that the convergence rate is no worse than that of fixed point iteration. In the nonlinear case we we show that, in a certain sense, the convergence speed can be made as close to the speed of fixed point iteration as

[*] Received by the editors May 2, 2013; accepted for publication (in revised form) January 22, 2015; published electronically March 24, 2015. This work was partially supported by the Consortium for Advanced Simulation of Light Water Reactors (www.casl.gov), an Energy Innovation Hub (http://www.energy.gov/hubs) for Modeling and Simulation of Nuclear Reactors under U.S. Department of Energy contract DE-AC05-00OR22725, National Science Foundation grants DMS-1406349, CDI-0941253, OCI-0749320, and SI2-SSE-1339844, and Army Research Office grants W911NF-11-1-0367 and W911NF-07-1-0112.

http://www.siam.org/journals/sinum/53-2/91939.html

[†] Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 (artoth@ncsu.edu, Tim_Kelley@ncsu.edu).

one likes, provided the initial iterate is sufficiently near a solution and the coefficients $\{\alpha_j^k\}$ (defined below) in the iteration remain bounded.

The Anderson iteration maintains a history of residuals

$$F(u) = G(u) - u$$

of depth at most $m + 1$, where m is an algorithmic parameter. When m is important, we will call the iteration Anderson(m). Anderson(0) is fixed point iteration by definition.

A formal algorithmic description follows.

```
anderson($u_0, G, m$)
  $u_1 = G(u_0)$; $F_0 = G(u_0) - u_0$
  for $k = 1, \dots$ do
    $m_k = \min(m, k)$
    $F_k = G(u_k) - u_k$
    Minimize $\| \sum_{j=0}^{m_k} \alpha_j^k F_{k-m_k+j} \|$ subject to
    $\sum_{j=0}^{m_k} \alpha_j^k = 1$.
    $u_{k+1} = (1 - \beta_k) \sum_{j=0}^{m_k} \alpha_j^k u_{k-m_k+j} + \beta_k \sum_{j=0}^{m_k} \alpha_j^k G(u_{k-m_k+j})$
  end for
```

One could use any norm in the minimization step, as we will see in section 2. Typically one uses the ℓ^2 norm, so the minimization problem can be formulated as a linear least squares problem [33] and solved easily. In this approach we solve the unconstrained problem

$$\min \left\| F(u_k) + \sum_{j=0}^{m_k-1} \alpha_j^k (F(u_{k-m_k+j}) - F(u_k)) \right\|^2 \tag{1.3}$$

for $\{\alpha_j^k\}_{j=0}^{m_k-1}$. Then we recover $\alpha_{m_k}^k$ by

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

This formulation of the linear least squares problem is not optimal for implementation, a point we discuss in more detail in section 3. Our analysis uses (1.3) because it explicitly displays the coefficients $\{\alpha_j^k\}$.

If one uses the ℓ^1 or ℓ^∞ norms, the optimization problem can be expressed as a linear program [13], for which there are many efficient solvers.

The mixing parameters $\{\beta_k\}$ are generally determined heuristically. For this discussion we will set $\beta_k \equiv 1$. If G is linear, i.e., $G(u) \equiv Mu + b$ (see section 2.1), then one can show [33] that $u_{k+1} = G(u_k^{GMRES})$ for $k \le m$, where u_k^{GMRES} is the kth GMRES iteration for the linear system $(I - M)u = b$, provided $I - M$ is nonsingular, and the GMRES residual norms are strictly decreasing. In particular, if $m > N$, G is linear, $I - M$ is nonsingular, and the residuals strictly decrease, then the iteration will converge to the solution. However, even in the linear case [27] the Anderson iterations can stagnate at an incorrect result. In the nonlinear case [12, 30] one can show that, for $m \le N$, Anderson iteration is a type of multisecant method, but one which is not covered by any previous convergence theory.

In most applications m is small. $m = 1$ or $m = 2$ is common for large electronic structure computations [2]. In these cases (large N and small m) there are no convergence results even in the linear case.

In this paper we prove three convergence results for Anderson acceleration when applied to contractive mappings. We prove that the residuals converge q-linearly for linear problems in section 2.1 and, in section 2.3, under certain conditions when $m = 1$. In the general nonlinear case we prove local r-linear convergence in section 2.2 under the assumption that the coefficients $\{\alpha_j^k\}$ remain bounded. All our results remain valid in an infinite-dimensional Banach space setting.

2. Convergence.

In this section we prove three convergence theorems. In the linear case the convergence of the residuals is q-linear. Recall that a sequence $\{w_k\}$ converges q-linearly with q-factor $c \in [0, 1)$ to w^* if

$$\|w_{k+1} - w^*\| \leq c\|w_k - w^*\|$$

for all $k \geq 0$. In the nonlinear case the convergence of the residuals is r-linear. This means that there is $\hat{c} \in (0, 1)$ and $M > 0$ such that

$$\|w_k - w^*\| \leq M\hat{c}^k\|w_0 - w^*\|.$$

In both cases, the convergence of u_k to the solution is r-linear.

Finally, in Theorem 2.4 we prove q-linear convergence of the Anderson(1) residuals to zero when the optimization is done in the ℓ^2 norm. Note that [9, 16] if the residuals, either linear or nonlinear, converge q-linearly to zero, then the errors converge q-linearly to zero in the norm $\|\cdot\|_*$, which is defined by

$$\|w\|_* = \|F'(u^*)w\|.$$

All of the proofs depend on the fact that if $\{\alpha_j\}_{j=0}^{m_k}$ is the solution of the least squares problem at iteration k, then by definition

$$\left\|\sum_{j=0}^{m_k} \alpha_j F(u_{k-m_k+j})\right\| \leq \|F(u_k)\|. \tag{2.1}$$

2.1. Linear problems.

In this section M is a linear operator with $\|M\| = c < 1$. We consider the linear fixed point problem

$$u = G(u) \equiv Mu + b.$$

The residual in this case is

$$F(u) = G(u) - u = b - (I - M)u.$$

THEOREM 2.1. *If $\|M\| = c < 1$, then the Anderson iteration converges to $u^* = (I - M)^{-1}b$ and the residuals converge q-linearly to zero with q-factor c.*

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

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

Hence, by (2.1)

$$\|F(u_{k+1})\| \leq c\|F(u_k)\|,$$

as asserted. □

If we set $e = u - u^*$, then $F(u) = -(I - M)e$. So q-linear convergence of residuals implies that

$$(1-c)\|e_k\| \leq \|F(u_k)\| \leq c^k\|F(u_0)\| \leq c^k(1+c)\|e_0\|$$

and hence

$$\|e_k\| \leq \left(\frac{1+c}{1-c}\right) c^k\|e_0\|,$$

which is r-linear convergence with r-factor c.

2.2. Nonlinear problems and local r-linear convergence. In this section we prove a local r-linear convergence result. Our result applies to any iteration of the form

$$u_{k+1} = \sum_{j=0}^{m_k} \alpha_j^k G(u_{k-m_k+j}) \tag{2.2}$$

for a fixed m and $m_k = \min(m, k)$ if the coefficients α_j^k satisfy Assumption 2.1.

ASSUMPTION 2.1.

1. $\|\sum_{j=0}^{m_k} \alpha_j F(u_{k-m_k+j})\| \leq \|F(u_k)\|$ *(which is* (2.1)*) holds.*
2. $\sum_{j=0}^{m_k} \alpha_j = 1$.
3. *There is* M_α *such that for all* $k \geq 0$ $\sum_{j=1}^{m_k} |\alpha_j| \leq M_\alpha$.

The first two parts of Assumption 2.1 are trivially satisfied by Anderson acceleration. The boundedness requirement in the third part would follow, for example, from uniform well-conditioning of the ℓ^2 nonlinear least squares problem (1.3), which, as [33] observes, is not guaranteed. In fact, as we show by example in section 3, the least squares problem can become highly ill-conditioned while the coefficients still remain bounded. We have not seen a case in our testing where the coefficients become large, but we are not able to prove that they remain bounded, hence the assumption. A method to modify m_k in response to ill-conditioning was proposed in [33]. We propose to address the boundedness of the coefficients directly.

One can modify Anderson acceleration to enforce boundedness of the coefficients by, for example,

- restarting the iteration when the coefficients exceed a threshold;
- imposing a bound constraint on the linear least squares problem and solving that problem with the method of [6]; or
- minimizing in ℓ^1 or ℓ^∞, adding the bound as a constraint, and formulating the resulting problem as a linear program.

The first of these is by far the simplest and, based on our experience, is unlikely to change the iteration at all.

The assumptions we make on the nonlinearity G and the solution u^*, Assumption 2.2, imply the usual standard assumptions [10, 16, 26] for local convergence of Newton's method. As is standard we will let F' denote the Jacobian of F and $e = u - u^*$.

ASSUMPTION 2.2.

- *There is $u^* \in R^N$ such that $F(u^*) = G(u^*) - u^* = 0$.*
- *G is Lipschitz continuously differentiable in the ball $\mathcal{B}(\hat\rho) = \{u \,|\, \|e\| \leq \hat\rho\}$ for some $\hat\rho > 0$.*
- *There is $c \in (0, 1)$ such that for all $u, v \in \mathcal{B}(\hat\rho)$, $\|G(u) - G(v)\| \leq c\|u - v\|$.*

The last of these assumptions implies that $\|G'(u)\| \leq c < 1$ for all $u \in \mathcal{B}(\hat\rho)$, and hence $F'(u^*)$ is nonsingular. We will let $G^* = G'(u^*)$. We will need a special case of a result (Lemma 4.3.1) from [16]. We will denote the Lipschitz constant of F' in $\mathcal{B}(\rho)$ by γ.

LEMMA 2.2. *For $\rho \leq \hat\rho$ sufficiently small and all $u \in \mathcal{B}(\rho)$,*

$$\|F(u) - F'(u^*)e\| \leq \frac{\gamma}{2}\|e\|^2 \tag{2.3}$$

and

$$\|e\|(1 - c) \leq \|F(u)\| \leq (1 + c)\|e\|. \tag{2.4}$$

THEOREM 2.3. *Let Assumption 2.2 hold and let $c < \hat c < 1$. Then if u_0 is sufficiently close to u^*, the Anderson iteration converges to u^* r-linearly with r-factor no greater than $\hat c$. In fact*

$$\|F(u_k)\| \leq \hat c^k \|F(u_0)\| \tag{2.5}$$

and

$$\|e_k\| \leq \frac{(1 + c)}{1 - c} \hat c^k \|e_0\|. \tag{2.6}$$

Proof. We let $u_0 \in \mathcal{B}(\rho)$ and assume that $\rho \leq \hat\rho$, where $\hat\rho$ is from the statement of Lemma 2.2. We will prove (2.5). Inequality (2.6) will follow from (2.5) and Lemma 2.2.

Reduce ρ if needed so that $\rho < 2(1 - c)/\gamma$ and

$$\frac{\left(\frac{c}{\hat c} + \left(\frac{M_\alpha \gamma\rho}{2(1-c)}\right)\hat c^{-m-1}\right)}{\left(1 - \frac{\gamma\rho}{2(1-c)}\right)} \leq 1. \tag{2.7}$$

Now reduce $\|e_0\|$ further so that

$$\left(\frac{M_\alpha(c + \gamma\rho/2)}{1 - c}\right)\hat c^{-m}\|F(u_0)\| \leq \left(\frac{M_\alpha(1 + c)(c + \gamma\rho/2)}{1 - c}\right)\hat c^{-m}\|e_0\| \leq \rho. \tag{2.8}$$

We will proceed by induction. Assume for all $0 \leq k \leq K$ that

$$\|F(u_k)\| \leq \hat c^k \|F(u_0)\|, \tag{2.9}$$

which clearly holds for $K = 0$.

Equations (2.8) and (2.9) imply that $\|e_k\| \leq \rho$ for $1 \leq k \leq K$. Hence, by (2.3)

$$F(u_k) = F'(u^*)e_k + \Delta_k,$$

where

$$\|\Delta_k\| \leq \frac{\gamma}{2}\|e_k\|^2. \tag{2.10}$$

This implies that

$$G(u_k) = u^* + G^* e_k + \Delta_k. \tag{2.11}$$

By (2.11) and the fact that $\sum \alpha_j^K = 1$,

$$\begin{aligned} u_{K+1} &= u^* + \sum_{j=0}^{m_K} \alpha_j^K (G^* e_{K-m_K+j} + \Delta_{K-m_K+j}) \\ &= u^* + \sum_{j=0}^{m_K} (\alpha_j^K G^* e_{K-m_K+j}) + \bar{\Delta}_K, \end{aligned} \tag{2.12}$$

where

$$\bar{\Delta}_K = \sum_{j=0}^{m_K} \alpha_j^K \Delta_{K-m_K+j}.$$

Our next task is to estimate $\bar{\Delta}_K$. Formulas (2.10) and (2.12) imply that

$$\|\bar{\Delta}_K\| = \sum_{j=0}^{m_K} |\alpha_j^K| \gamma \|e_{K-m_K+j}\|^2/2. \tag{2.13}$$

Lemma 2.2, the induction hypothesis, and the fact that

$$K - m_K + j = K - \min(m, K) + j \geq K - m$$

imply that

$$\begin{aligned} \|e_{K-m_K+j}\|^2 &\leq \|e_{K-m_K+j}\| \left(\frac{1}{1-c}\right) \|F(u_{K-m_K+j})\| \leq \left(\frac{\rho}{1-c}\right) \|F(u_{K-m_K+j})\| \\ &\leq \left(\frac{\rho}{1-c}\right) \hat{c}^{(K-m_K+j)} \|F(u_0)\| \leq \left(\frac{\rho}{1-c}\right) \hat{c}^{K-m} \|F(u_0)\|. \end{aligned} \tag{2.14}$$

So, since $\sum |\alpha_j^K| \leq M_\alpha$, we have

$$\|\bar{\Delta}_K\| \leq \left(\frac{M_\alpha \gamma \rho}{2(1-c)}\right) \hat{c}^{K-m} \|F(u_0)\| \leq \left(\frac{M_\alpha \gamma \rho}{2(1-c)}\right) \hat{c}^{-m} \|F(u_0)\|. \tag{2.15}$$

Write (2.12) as

$$e_{K+1} = \sum_{j=0}^{m_K} (\alpha_j^K G^* e_{K-m_K+j}) + \bar{\Delta}_K.$$

The induction hypothesis implies that, for $0 \leq j \leq m_k$,

$$\begin{aligned} \|e_{K-m_K+j}\| &\leq \left(\frac{1}{1-c}\right) \|F(u_{K-m_K+j})\| \\ &\leq \left(\frac{1}{1-c}\right) \hat{c}^{K-m_K+j} \|F(u_0)\| \leq \left(\frac{1}{1-c}\right) \hat{c}^{-m} \|F(u_0)\|, \end{aligned}$$

and hence

$$\left\|\sum_{j=0}^{m_K}\alpha_j^K G^* e_{K-m_K+j}\right\| \le \left(\frac{M_\alpha c}{1-c}\right)\hat{c}^{-m}\|F(u_0)\|. \tag{2.16}$$

Combining (2.12), (2.15), and (2.16) yields

$$\|e_{K+1}\| \le \|F(u_0)\|\left(\frac{M_\alpha(c+\gamma\rho/2)}{1-c}\right)\hat{c}^{-m} \le \rho$$

by (2.8).

Since $\|e_{K+1}\| \le \rho \le \hat{\rho}$, we may apply (2.11) with $k = K+1$ to obtain

$$F(u_{K+1}) = (G^* - I)e_{K+1} + \Delta_{K+1},$$

where, by Lemma 2.2,

$$\|\Delta_{K+1}\| \le \frac{\gamma}{2}\|e_{K+1}\|^2.$$

So, since G^* and $G^* - I$ commute,

$$\begin{aligned} F(u_{K+1}) &= G^*\sum_{j=0}^{m_K}\alpha_j^K(G^*-I)e_{K-m_K+j} + (G^*-I)\bar{\Delta}_K + \Delta_{K+1} \\ &= G^*\sum_{j=0}^{m_K}\alpha_j^K F(u_{K-m_K+j}) - \alpha_j^K\Delta_{K-m_K+j} + (G^*-I)\bar{\Delta}_K + \Delta_{K+1} \\ &= G^*\sum_{j=0}^{m_K}\alpha_j^K F(u_{K-m_K+j}) - \bar{\Delta}_K + \Delta_{K+1}. \end{aligned} \tag{2.17}$$

Combine (2.17) with (2.4) to obtain

$$\|\Delta_{K+1}\| \le \left(\frac{\gamma}{2(1-c)}\right)\rho\|F(u_{K+1})\|. \tag{2.18}$$

The induction hypothesis, (2.1), (2.15), (2.17), and (2.18) imply that

$$\begin{aligned} \|F(u_{K+1})\|\left(1-\frac{\gamma\rho}{2(1-c)}\right) &\le \|F(u_{K+1})\| - \|\Delta_{K+1}\| \\ &\le c\left\|\sum_{j=0}^{m_K}\alpha_j^K F(u_{K-m_K+j})\right\| + \|\bar{\Delta}_K\| \\ &\le c\|F(u_K)\| + \|\bar{\Delta}_K\| \\ &= \left(\frac{c}{\hat{c}} + \left(\frac{M_\alpha\gamma\rho}{2(1-c)}\right)\hat{c}^{-m-1}\right)\hat{c}^{K+1}\|F(u_0)\|. \end{aligned} \tag{2.19}$$

Therefore

$$\|F(u_{K+1})\| \le \frac{\left(\frac{c}{\hat{c}} + \left(\frac{M_\alpha\gamma\rho}{2(1-c)}\right)\hat{c}^{-m-1}\right)}{\left(1-\frac{\gamma\rho}{2(1-c)}\right)}\hat{c}^{K+1}\|F(u_0)\| \le \hat{c}^{K+1}\|F(u_0)\|$$

since

$$\frac{\left(\frac{c}{\hat{c}} + \left(\frac{M_\alpha \gamma \rho}{2(1-c)}\right)\hat{c}^{-m-1}\right)}{\left(1 - \frac{\gamma\rho}{2(1-c)}\right)} \leq 1$$

by (2.7). This completes the proof. □

2.3. Convergence for Anderson(1). In this section we prove convergence for Anderson(1) with the ℓ^2 norm. We will assume that Assumption 2.2 holds. We show directly that the coefficients are bounded if c is sufficiently small and prove q-linear convergence of the residuals in that case. The analysis here is quite different from the Anderson(m) case in the previous section, depending heavily on both $m = 1$ and the fact that the optimization problem for the coefficients is a linear least squares problem.

We will express the iteration as

$$u_{k+1} = (1 - \alpha^k)G(u_k) + \alpha^k G(u_{k-1}) \tag{2.20}$$

and note that

$$\alpha^k = \frac{F(u_k)^T(F(u_k) - F(u_{k-1}))}{\|F(u_k) - F(u_{k-1})\|^2}. \tag{2.21}$$

THEOREM 2.4. *Assume that Assumption* 2.2 *holds, that* $u_0 \in \mathcal{B}(\hat{\rho})$*, and that* c *is small enough so that*

$$\hat{c} \equiv \frac{3c - c^2}{1 - c} < 1. \tag{2.22}$$

Then the Anderson(1) *residuals with* ℓ^2 *optimization residuals converge q-linearly with q-factor* $\hat{c}$.

Proof. We induct on k. Assume that

$$\|F(u_k)\| \leq \hat{c}\|F(u_{k-1})\| \tag{2.23}$$

for all $0 \leq k \leq K$. Inequality (2.23) is trivially true for $K = 1$, since Anderson(0) is successive substitution and $c < \hat{c}$.

Now define

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

and

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

Clearly

$$F(u_{K+1}) = G(u_{K+1}) - u_{K+1} = A_K + B_K. \tag{2.24}$$

We will obtain an estimate of $F(u_{K+1})$ by estimating A_K and B_K separately.

By definition of the Anderson iteration (2.20) and contractivity of G,

$$\begin{aligned}\|A_K\| &= \|G(u_{K+1}) - G((1-\alpha^K)u_K + \alpha^K u_{K-1})\| \\ &\leq c\|u_{K+1} - (1-\alpha^K)u_K - \alpha^K u_{K-1}\| \\ &= c\|(1-\alpha^K)(G(u_K) - u_K) - \alpha^K(G(u_{K-1}) - u_{K-1})\| \\ &= c\|(1-\alpha^K)F(u_K) - \alpha^K F(u_{K-1})\| \leq c\|F(u_K)\|,\end{aligned} \tag{2.25}$$

where the last inequality follows from the optimality property of the coefficients.

Now let

$$\delta_K = u_{K-1} - u_K.$$

To estimate B_K we note that

$$\begin{aligned}B_K &= G((1-\alpha^K)u_K + \alpha^K u_{K-1}) - (1-\alpha^K)G(u_K) - \alpha^K G(u_{K-1}) \\ &= G(u_K + \alpha^K\delta_K) - G(u_K) + \alpha^K(G(u_K) - G(u_{K-1})) \\ &= \int_0^1 G'(u_K + t\alpha^K\delta_K)\alpha^K\delta_K\,dt - \alpha^K\int_0^1 G'(u_K + t\delta_K)\delta_K\,dt \\ &= \alpha^K\int_0^1 \left[G'(u_K + t\alpha^K\delta_K) - G'(u_K + t\delta_K)\right]\delta_K\,dt.\end{aligned} \tag{2.26}$$

This leads to the estimate

$$\|B_K\| \leq |\alpha^K|\|\delta_K\|\int_0^1 \|G'(u_K + t\alpha^K\delta_K) - G'(u_K + t\delta_K)\|\,dt \leq 2c|\alpha^K|\|\delta_K\|. \tag{2.27}$$

The next step is to estimate α^K. The difference in residuals is

$$\begin{aligned}F(u_K) - F(u_{K-1}) = G(u_K) - G(u_{K-1}) + \delta_K &= \delta_K - \int_0^1 G'(u_{K-1} - t\delta_K)\delta_K\,dt \\ &= \left(I - \int_0^1 G'(u_{K-1} - t\delta_K)\,dt\right)\delta_K.\end{aligned}$$

Since $\|G'(u)\| \leq c$ for all $u \in \mathcal{B}(\rho)$ we have

$$\|\delta_K\| \leq \|F(u_K) - F(u_{K-1})\|/(1-c). \tag{2.28}$$

Combine (2.28) and (2.21) to obtain

$$|\alpha^K|\|\delta_K\| \leq \frac{\|F(u_K)\|}{\|F(u_K) - F(u_{K-1})\|}\|\delta_K\| \leq \frac{\|F(u_K)\|}{1-c}. \tag{2.29}$$

Now, we combine (2.25), (2.27), and (2.29) to obtain

$$\frac{\|F(u_{K+1})\|}{\|F(u_K)\|} \leq c + \frac{2c}{1-c} = \hat{c}. \tag{2.30}$$

Our assumption that $\hat{c} < 1$ completes the proof. □

Note that q-linear convergence of the residuals implies that the coefficients are bounded because

$$|\alpha^K| \le \frac{\|F(u_K)\|}{\|F(u_K) - F(u_{K-1})\|} \le \frac{\hat{c}\|F(u_{K-1})\|}{\|F(u_{K-1})\|(1-\hat{c})} \le \frac{\hat{c}}{1-\hat{c}}.$$

For sufficiently good initial data, one can use the proof of Theorem 2.4 to prove q-linear convergence for all $c \in [0,1)$.

Corollary 2.5. *Assume that Assumption* 2.2 *holds and that $\hat{c} \in (c,1)$. Then if $\|e_0\|$ is sufficiently small, the Anderson*(1) *residuals with ℓ^2 optimization converge q-linearly with q-factor no larger than $\hat{c}$. Moreover*

$$\limsup_{k\to\infty} \frac{\|F(u_{k+1})\|}{\|F(u_k)\|} \le c. \tag{2.31}$$

Proof. We use the standard assumptions, (2.21), and (2.26) to estimate B_K. Note that $1+\gamma$ is an upper bound for the Lipschitz constant of G'. Equations (2.21) and (2.26) imply that

$$\|B_K\| \le \frac{(1+\gamma)|\alpha^K||1-\alpha^K|\|\delta_K\|^2}{2} \le \frac{(1+\gamma)\|F(u_K)\|\|F(u_{K-1})\|\|\delta_K\|^2}{2\|F(u_K) - F(u_{K-1})\|^2}.$$

Equation (2.28) implies that

$$\frac{\|F(u_K) - F(u_{K-1})\|}{\|\delta_K\|} \ge (1-c),$$

and hence

$$\|B_K\| \le \frac{(1+\gamma)\|F(u_{K-1})\|\|F(u_K)\|}{2(1-c)^2}.$$

Therefore, we can use (2.24) and (2.25) to obtain

$$\frac{\|F(u_{K+1})\|}{\|F(u_K)\|} \le c + \frac{(1+\gamma)\|F(u_{K-1})\|}{2(1-c)^2}. \tag{2.32}$$

The right side of (2.32) is $\le \hat{c}$ for all $K \ge 1$ if $\|e_0\|$ is sufficiently small. In that case, the residuals converge to zero, and (2.31) follows. □

3. Numerical experiments.

In this section we report on some simple numerical experiments. We implement Anderson acceleration using the approach from [12, 21, 33], which is equivalent to (1.3), but organizes the computation to make the coefficient matrix easy to update by adding a single column and deleting another. This makes it possible to use fast methods to update the QR factorization [14] to solve the sequence of linear least squares problems if one does the optimization in the ℓ^2 norm. According to [33] and our own experience, this form has modestly better conditioning properties.

For the kth iteration we solve

$$\min_{\theta \in R^{m_k}} \left\| F(u_k) - \sum_{j=0}^{m_k-1} \theta_j (F(u_{k-m_k+j+1}) - F(u_{k-m_k+j})) \right\| \tag{3.1}$$

to obtain a vector $\theta^k \in R^{m_k}$. Then the next iteration is

$$u_{k+1} = G(u_k) - \sum_{j=0}^{m_k-1} \theta_j^k (G(u_{k-m_k+j+1}) - G(u_{k-m_k+j})). \tag{3.2}$$

In terms of (1.3),

$$\alpha_0 = \theta_0, \alpha_j = \theta_j - \theta_{j-1} \text{ for } 1 \leq j \leq m_k - 1 \text{ and } \alpha_{m_k} = 1 - \theta_{m_k-1}.$$

3.1. Conditioning. Suppose $G : R^2 \to R^2$,

$$G(u) = \begin{pmatrix} g(u_1, u_2) \\ g(u_1, u_2) \end{pmatrix},$$

and

$$u_0 = \begin{pmatrix} w \\ w \end{pmatrix}.$$

Then Anderson(2), using the ℓ^2 norm, may fail because the linear least squares problem will be rank-deficient. One could, of course, take the minimum norm solution with no ill effects.

Now consider a perturbation of such a problem. The linear least squares problem will be ill-conditioned, but the size of the coefficients for Anderson(2) may remain bounded. We illustrate this with a simple example.

Let

$$G(u) = \begin{pmatrix} \cos((u_1 + u_2)/2) \\ \cos((u_1 + u_2)/2) + 10^{-8} \sin(u_1^2) \end{pmatrix}.$$

We applied Anderson acceleration to this problem with an initial iterate $u_0 = (1, 1)^T$. In Table 1 we tabulate the residual norms, the condition number of the coefficient matrix for the optimization problem for the coefficients, and the ℓ^1 norm of the coefficients. We terminate the iteration when the residual norm falls below 10^{-10}. As one can see, the condition number becomes very large with little effect on the coefficient norm. We will see a similar effect for the more interesting problem in section 3.2.

TABLE 1
Iteration statistics for Anderson(2).

k	Residual norm	Condition number	Coefficient norm
0	6.501e-01		
1	4.487e-01	1.000e+00	1.000e+00
2	2.615e-02	2.016e+10	4.617e+00
3	7.254e-02	1.378e+09	2.157e+00
4	1.531e-04	3.613e+10	1.184e+00
5	1.185e-05	2.549e+11	1.000e+00
6	1.825e-08	3.677e+10	1.002e+00
7	1.048e-13	1.574e+11	1.092e+00

3.2. Using the ℓ^1 and ℓ^∞ norms. In this section we compare the use of Anderson acceleration with the ℓ^1, ℓ^2, and ℓ^∞ norms with a fixed point iteration (Anderson(0)) and a Newton–Krylov iteration. We use the `nsoli.m` MATLAB code from [16, 17] for the Newton–Krylov code and solve the linear programs for the ℓ^1 and ℓ^∞ optimization problems for the coefficients with the CVX MATLAB software [7, 15]. We used the SeDuMi solver and set the `precision` in cvx to `high`. The ℓ^1 and ℓ^∞ optimizations are significantly more costly than the ℓ^2 optimization. While the iteration with the ℓ^1 optimization seems to perform well, one would need a special-purpose linear programming code tuned for this application to make that approach practical.

As an example we take the composite midpoint rule discretization of the Chandrasekhar H-equation [3, 5]

$$H(\mu) = G(H) \equiv \left(1 - \frac{\omega}{2}\int_0^1 \frac{\mu}{\mu+\nu} H(\nu)\, d\nu\right)^{-1}. \tag{3.3}$$

In (3.3) $\omega \in [0,1]$ is a parameter and one seeks a solution $H^* \in C[0,1]$. The solution $H^*(\mu) \geq 1$ satisfies

$$\|H^*\|_\infty \leq \min\left(3, \frac{1}{\sqrt{1-\omega}}\right)$$

and is an increasing function of μ and ω. This fact and a monotonicity argument can be used to show that if $\epsilon > 0$ is sufficiently small and u and v are in $\mathcal{B}(\epsilon)$,

$$\|G(u) - G(v)\| \leq \frac{(1+\epsilon)^2 \|H^*\|_\infty^2 \omega}{2} \|u - v\|$$

for any L^p norm. This inequality also carries over to the discrete problems. In particular, G is a local contraction for $\omega = .5$, which is one of our test cases.

It is known [18, 24], both for the continuous problem and its midpoint rule discretization, that if $\omega < 1$,

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

where ρ denotes spectral radius. Hence the local convergence theory in sections 2.2 and 2.3 applies for some choice of norm. It is also known that if the initial iterate is nonnegative and componentwise smaller than H^*, then fixed point iteration will converge with a q-factor no larger than ω. For $\omega = 1$ the nonlinear equation is singular, but both fixed point iteration and Newton-GMRES will converge from our choice of initial iterate [18, 19, 20].

For integral equation problems such as this one, where preconditioning is not necessary, Newton-GMRES will be faster than a conventional approach in which one constructs, stores, and factors a Jacobian matrix. Newton-GMRES will converge to truncation error at a cost proportional to the square of the size of the spatial grid, whereas the cost of even one factorization will be cubic. However, for this problem G' is a compact operator, which implies that the coefficient matrix for the optimization problem for the coefficients could become ill-conditioned as the iteration converges, as we see in the tables below.

We report on computational results with an $N = 500$ point composite midpoint rule discretization. We compare Anderson(m) for $m = 0, \dots, 6$ and the ℓ^1, ℓ^2, and ℓ^∞ norms with a Newton-GMRES iteration. The Newton-GMRES iteration used a reduction factor of .1 for the linear residual with each Newton step. Other approaches to the linear solver, such as the one from [11], performed similarly. We terminate the nonlinear iterations when $\|F(u_k)\|_2/\|F(u_0)\|_2 \leq 10^{-8}$.

We consider values of $\omega = .5, .99, 1.0$, with the last value being one for which F' is singular at the solution and $\rho(G(H^*)) = 1$. When $\omega = 1$, Newton's method will be linearly convergent [8, 20]. One can see this in Table 2 by the increase in the number of function calls for $\omega = 1$. The initial iteration was $(1, 1, \dots, 1)^T$ for all cases. This is a good initial iterate for $\omega = .5$ and a marginal one for the other two cases.

In the tables we use function calls as a measure of cost. This is an imperfect metric, and the comparisons should be viewed as qualitative. The costs of the ℓ^1 and ℓ^∞ optimizations are more than solving the ℓ^2 least squares problem, and we have ignored the orthogonalization cost within Newton-GMRES. For large problems where the evaluation of G is costly, the cost of the linear program solve should be relatively low, especially for small values of m.

We begin with the base cases of Newton-GMRES and fixed point iteration. In Table 2 we tabulate, for each value of ω, the number of calls to the function needed for termination. Note that each Newton-GMRES iteration needs one call for the residual evaluation and one for each of the finite-difference Jacobian-vector products.

Table 2
Function evaluations for Newton-GMRES and fixed point iteration.

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

In Table 3 we tabulate $\rho(G'(H^*))$ and the ℓ^p norms of $G'(H^*)$ as functions of ω and p. These are the computed values using the output of the Newton-GMRES iteration. Because of the singularity of $F'(H^*)$, the iteration for $\omega = 1$ has roughly four digits of accuracy [8], which accounts for the difference in the spectral radius for $\omega = 1$ from the true value of 1.0.

Table 3
Norms and spectral radius of Jacobians.

ω	$\|G'(H^*)\|_1$	$\|G'(H^*)\|_2$	$\|G'(H^*)\|_\infty$	$\rho(G'(H^*))$
0.5	3.422e-01	1.994e-01	2.712e-01	1.528e-01
0.99	1.714e+00	1.182e+00	2.095e+00	7.959e-01
1.00	2.137e+00	1.541e+00	2.926e+00	9.999e-01

In Table 4, we tabulate ω, m, the cost in function evaluations for convergence, the maximum (over the entire iteration) κ_{max} of the condition number of the matrix in the optimization problem, and S_{max}, the maximum of the sum of the absolute values of the coefficients $\{\alpha_j^k\}$. In all cases Anderson iteration is competitive with Newton-GMRES, and is significantly better for all values of m when the optimization uses the ℓ^1 and ℓ^2 norms. Note that κ_{max} becomes very large as m increases, but that the performance of the algorithm does not degrade in a significant way with increasing m. In fact, for this problem $m = 3$ (ℓ^1) and $m = 4$ (ℓ^2) have the fewest calls to the fixed point map, even though κ_{max} is quite large.

Table 4
Anderson iteration for H-equation.

		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
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

4. Conclusions.

We prove local r-linear convergence of Anderson iteration when applied to contractive mappings under the assumption that the coefficients are bounded. We prove q-linear convergence of the residuals for linear problems. For the special case of Anderson(1), we prove q-linear convergence of the residuals. Numerical results illustrate the ideas.

Acknowledgments. The authors are grateful to Steve Wright, who pointed them to the CVX code, and to Jerry Bernholc, Emil Briggs, Kai Fan, Miro Hodak, Wenchang Lu, Homer Walker, and Carol Woodward for several enlightening conversations about Anderson acceleration. We also appreciate the efforts of two very knowledgeable and thoughtful referees.

REFERENCES

[1] D. G. Anderson, *Iterative procedures for nonlinear integral equations*, J. Assoc. Comput. Mach., 12 (1965), pp. 547–560.

[2] J. Bernholc, *private communication*, 2012.

[3] I. W. Busbridge, *The Mathematics of Radiative Transfer*, Cambridge Tracts 50, Cambridge University Press, Cambridge, UK, 1960.

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

[5] S. Chandrasekhar, *Radiative Transfer*, Dover, New York, 1960.

[6] T. F. Coleman and Y. Li, *A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables*, SIAM J. Optim., 6 (1996), pp. 1040–1058.

[7] CVX Research, Inc., *CVX: MATLAB Software for Disciplined Convex Programming*, version 2.0, http://cvxr.com/cvx, 2012.

[8] D. W. Decker and C. T. Kelley, *Newton's method at singular points.* I, SIAM J. Numer. Anal., 17 (1980), pp. 66–70.

[9] R. S. Dembo, S. C. Eisenstat, and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.

[10] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics Appl. Math. 16, SIAM, Philadelphia, 1996.

[11] S. C. Eisenstat and H. F. Walker, *Globally convergent inexact Newton methods*, SIAM J. Optim., 4 (1994), pp. 393–422.

[12] H.-R. Fang and Y. Saad, *Two classes of multisecant methods for nonlinear acceleration*, Numer. Linear Algebra Appl., 16 (2009), pp. 197–221.

[13] M. Ferris, O. Mangasarian, and S. Wright, *Linear Programming with MATLAB*, SIAM, Philadelphia, 2007.

[14] G. H. Golub and C. G. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996.

[15] M. Grant and S. Boyd, *Graph implementations for nonsmooth convex programs*, in Recent Advances in Learning and Control, V. Blondel, S. Boyd, and H. Kimura, eds., Lecture Notes in Control and Inform. Sci. 371, Springer-Verlag, London, 2008, pp. 95–110.

[16] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, Frontiers Appl. Math. 16, SIAM, Philadelphia, 1995.

[17] C. T. Kelley, *Solving Nonlinear Equations with Newton's Method*, Fundam. Algorithms 1, SIAM, Philadelphia, 2003.

[18] C. T. Kelley and T. W. Mullikin, *Solution by iteration of H-equations in multigroup neutron transport*, J. Math. Phys., 19 (1978), pp. 500–501.

[19] C. T. Kelley and Z. Q. Xue, *Inexact Newton methods for singular problems*, Optim. Methods Softw., 2 (1993), pp. 249–267.

[20] C. T. Kelley and Z. Q. Xue, *GMRES and integral operators*, SIAM J. Sci. Comput., 17 (1996), pp. 217–226.

[21] G. Kresse and J. Furthmüller, *Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set*, Comput. Materials Sci., 6 (1996), pp. 15–50.

[22] L. Lin and C. Yang, *Elliptic preconditioner for accelerating the self-consistent field iteration in Kohn–Sham density functional theory*, SIAM J. Sci. Comput., 35 (2013), pp. S277–S298.

[23] K. Miller, *Nonlinear Krylov and moving nodes in the method of lines*, J. Comput. Appl. Math., 183 (2005), pp. 275–287.

[24] T. W. Mullikin, *Some probability distributions for neutron transport in a half space*, J. Appl. Probab., 5 (1968), pp. 357–374.

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

[26] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[27] F. A. Potra and H. Engler, *A characterization of the behavior of the Anderson acceleration on linear problems*, Linear Algebra Appl., 438 (2013), pp. 1002–1011.

[28] P. Pulay, *Convergence acceleration of iterative sequences. The case of SCF iteration*, Chem. Phys. Lett., 73 (1980), pp. 393–398.

[29] P. Pulay, *Improved SCF convergence acceleration*, J. Comput. Chem., 3 (1982), pp. 556–560.

[30] T. Rohwedder and R. Schneider, *An analysis for the DIIS acceleration method used in quantum chemistry calculations*, J. Math. Chem., 49 (2011), pp. 1889–1914.

[31] Y. Saad, J. R. Chelikowsky, and S. M. Shontz, *Numerical methods for electronic structure calculations of materials*, SIAM Rev., 52 (2010), pp. 3–54.

[32] R. Schneider, T. Rohwedder, A. Neelov, and J. Blauert, *Direct minimization for calculating invariant subspaces in density functional computations of the electronic structure*, J. Comput. Math., 27 (2008), pp. 360–387.

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

[34] T. Washio and C. W. Oosterlee, *Krylov subspace acceleration for nonlinear multigrid schemes*, Electron. Trans. Numer. Anal., 6 (1997), pp. 271–290.