SIAM J. NUMER. ANAL.
Vol. 57, No. 2, pp. 615–637

ANDERSON-ACCELERATED CONVERGENCE OF PICARD ITERATIONS FOR INCOMPRESSIBLE NAVIER–STOKES EQUATIONS*

SARA POLLOCK†, LEO G. REBHOLZ‡, AND MENGYING XIAO§

Abstract. We propose, analyze, and test Anderson-accelerated Picard iterations for solving the incompressible Navier–Stokes equations (NSE). Anderson acceleration has recently gained interest as a strategy to accelerate linear and nonlinear iterations, based on including an optimization step in each iteration. We extend the Anderson acceleration theory to the steady NSE setting and prove that the acceleration improves the convergence rate of the Picard iteration based on the success of the underlying optimization problem. The convergence is demonstrated in several numerical tests, with particularly marked improvement in the higher Reynolds number regime. Our tests show it can be an enabling technology in the sense that it can provide convergence when both usual Picard and Newton iterations fail.

Key words. Anderson acceleration, steady Navier–Stokes, fixed-point iteration, local convergence, global convergence

AMS subject classifications. 65N22, 65H10, 35Q30, 65N30

DOI. 10.1137/18M1206151

1. Introduction.

We consider numerical solvers for the steady incompressible Navier–Stokes equations (NSE), which are given in a domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ by

$$u \cdot \nabla u + \nabla p - \nu \Delta u = f, \tag{1.1}$$

$$\nabla \cdot u = 0, \tag{1.2}$$

$$u|_{\partial\Omega} = g, \tag{1.3}$$

where ν is the kinematic viscosity, f is a forcing, and u and p represent velocity and pressure. For simplicity of our presentation and analysis, we consider homogeneous Dirichlet boundary conditions, i.e., $g = 0$, but our theory can be extended to other common boundary conditions.

We study herein an acceleration technique applied to the Picard method for solving the steady NSE. The Picard method is commonly used for solving the steady NSE due to its stability and global convergence properties, and takes the form (suppressing a spatial discretization)

$$u_k \cdot \nabla u_{k+1} + \nabla p_{k+1} - \nu \Delta u_{k+1} = f, \tag{1.4}$$

$$\nabla \cdot u_{k+1} = 0, \tag{1.5}$$

$$u_{k+1}|_{\partial\Omega} = 0, \tag{1.6}$$

*Received by the editors August 8, 2018; accepted for publication (in revised form) February 4, 2019; published electronically March 19, 2019.

http://www.siam.org/journals/sinum/57-2/M120615.html

Funding: The work of the first author was supported in part by the National Science Foundation grants DMS 1719849 and DMS 1852876. The work of the second and third authors was supported in part by the National Science Foundation grant DMS 1522191.

†Department of Mathematics, University of Florida, Gainesville, FL 32611-8105 (s.pollock@ufl.edu).

‡Department of Mathematical Sciences, Clemson University, Clemson, SC 29634 (rebholz@rebholz@clemson.edu).

§Department of Mathematics, College of William & Mary, Williamsburg, VA 23185 (mxiao01@wm.edu).

This iteration can be written as a fixed point iteration, $u_{k+1} = G(u_k)$, with G denoting a solution operator for the Picard linearization (1.4)–(1.6).

In practice, unfortunately, the Picard iteration often converges slowly, sometimes so slowly that for all practical purposes it fails. To improve this slow convergence, we employ an acceleration strategy introduced by Anderson in 1965 [1]. In recent years, this strategy now commonly referred to as Anderson acceleration has been analyzed in the context of multisecant methods for fixed-point iterations in [6] motivated by a problem in electronic structure computations and in the context of generalized minimal residual methods in [17], where the efficacy of the method is demonstrated on a range of nonlinear problems. We further refer readers to [10, 12, 17] and the references therein for detailed discussions on both practical implementation and a history of the method and its applications. Despite its long history of use, the first convergence analysis for Anderson acceleration (in both the linear and nonlinear settings) appears in 2015 in [16], under the usual local assumptions for convergence of Newton iterations. However, this theory (which we summarize in section 2) does not prove that Anderson acceleration actually improves the convergence of a fixed-point iteration.

The main contributions of this work involve Anderson acceleration applied to the Picard iteration for the steady NSE. In this setting, we are able to prove that Anderson acceleration gives guaranteed improvement over the usual Picard iteration in a neighborhood of the fixed point. To our knowledge, this is the first proof of improved convergence for Anderson acceleration applied to a nonlinear fixed-point iteration, and thus may give insight into how a theory for general nonlinear fixed-point operators might be developed. Additionally, we show with several numerical experiments that Anderson acceleration can provide dramatic improvement in the Picard iteration, and can even be an enabling technology in the sense that it provides convergence in cases where both Picard and Newton fail. In addition to this result, we also investigate the global convergence behavior of Anderson acceleration for contractive operators. We find a relation between the gain from the optimization, bounds on the optimization coefficients, and the convergence rate of the underlying fixed-point iteration that assures the accelerated sequence converges at an improved rate, independent of the initial error.

This paper is arranged as follows. In section 2 we provide some background on Anderson acceleration and its convergence properties, and show global r-linear convergence at an improved rate based on success of the optimization problem for small enough coefficients. In section 3 we give preliminaries for the steady NSE and associated finite element spatial discretization, and provide details of properties of the solution operator of the fixed-point iteration associated with the discrete Picard linearization of the steady NSE. In section 4 we then analyze the Anderson-accelerated Picard iteration for the steady NSE. We extend the general convergence results of [10, 16] to this problem, and for the particular cases of the acceleration algorithm optimizing over either one or two additional prior residuals, prove that Anderson acceleration improves the contraction ratio of the Picard iteration. In section 5 we report on results of several numerical tests for Anderson-accelerated Picard iterations for the steady NSE, and show that it can have a dramatic positive impact.

2. Anderson acceleration. We discuss now the general Anderson acceleration algorithm and its convergence properties for contractive nonlinear operators. In later sections, we will consider the specific case of Picard iterations for the steady incompressible NSE. We start by stating the algorithm and reviewing the relevant known theory. Theorem 2.5 is a new contribution to the theory for general nonlinear con-

tractive operators. It shows that Anderson acceleration increases the convergence rate of the fixed-point iteration when the optimization coefficients satisfy certain bounds. We begin with the basic assumption of a contractive (nonlinear) operator.

Assumption 2.1. Let $G : X \to X$ be a contractive operator with contraction ratio $r < 1$, i.e.,

$$\|G(u) - G(w)\|_* \leq r\|u - w\|_* \quad \forall u, w \in X,$$

for a given space X with norm $\|\cdot\|_*$.

By standard fixed-point theory, under Assumption 2.1 there exists a unique $u^* \in X$ such that $G(u^*) = u^*$. Although in section 3 and beyond we will make specific choices for G and X, we discuss the acceleration algorithm in this form to emphasize its more general applicability.

ALGORITHM 2.2 (Anderson iteration). *The Anderson acceleration with depth m reads as follows:*
Step 0*: Choose $u_0 \in X$.*
Step 1*: Find $\tilde{u}_1 \in X$ such that $\tilde{u}_1 = G(u_0)$. Set $u_1 = \tilde{u}_1$.*
Step k: For $k + 1 = 1, 2, 3, \ldots$, set $m_k = \min\{k, m\}$.
[a] *Find $\tilde{u}_{k+1} = G(u_k)$.*
[b] *Solve the minimization problem for $\{\alpha_j^{k+1}\}_{k-m_k}^{k}$*

$$\min_{\sum_{j=k-m_k}^{k} \alpha_j^{k+1} = 1} \left\| \sum_{j=k-m_k}^{k} \alpha_j^{k+1} (\tilde{u}_{j+1} - u_j) \right\|_* .$$

[c] *Set* $u_{k+1} = \sum_{j=k-m_k}^{k} \alpha_j^{k+1} \tilde{u}_{j+1}$.

Remark 2.3. For the more general Anderson mixing algorithm, set u_{k+1} in Algorithm 2.2 by

$$u_{k+1} = \beta_{k+1} \sum_{j=k-m_k}^{k} \alpha_j^{k+1} \tilde{u}_{j+1} + (1 - \beta_{k+1}) \sum_{j=k-m_k}^{k} \alpha_j^{k+1} u_j$$

for damping parameter $0 < \beta_k \leq 1$. Here we consider the undamped case $\beta_k = 1$ for all k as under Assumption 2.1 the fixed-point iteration converges with rate $r < 1$, and $\beta_k < 1$ scales the natural convergence rate of the iteration towards unity. However, damping can be useful and is sometimes crucial for simulations in which the underlying fixed-point operator is not globally contractive. In the current context of steady Navier–Stokes under a small data assumption, the fixed-point operator is globally contractive so we do not further consider the damping in this manuscript. The effect of combining damping with acceleration is however of ongoing interest to the authors.

The convergence of Anderson acceleration is studied in [10, 16], and for general nonlinear G it is known that in a small enough neighborhood of the solution, the acceleration will not make the convergence significantly worse. To our knowledge however there is no mathematical proof that Anderson acceleration increases the convergence compared to the associated fixed-point iteration. The following result is proven in [16, Theorem 2.3], and is the best known result for (locally) contractive operators.

THEOREM 2.4 (convergence of Anderson acceleration). *Assume operator G has fixed point u^*, and satisfies the following two conditions under some norm $\|\cdot\|_*$.*

1. *G is Lipschitz continuously differentiable in a ball*

$$\mathcal{B}(\rho) = \{u \in X_h : \|u - u^*\|_* < \rho\}$$

for some $\rho > 0$.

2. *There is a $c \in (0,1)$ such that for all $u, v \in \mathcal{B}(\rho)$, $\|G(u)-G(v)\|_* \leq c\|u-v\|_*$.*

Then if $\sum_{j=1}^{m_k} |\alpha_j^k|$ is uniformly bounded for all $k > 0$, Algorithm 2.2 converges to u^* with contraction ratio no greater than $\hat{c}$, where $c < \hat{c} < 1$, provided $\|u_0 - u\|_*$ is small enough.

We improve on this result for steady NSE in section 4 where we show for the contractive operator G associated with the Picard iteration that the convergence of the residual to zero is guaranteed to be accelerated close enough to the solution. While this result depends on the particular structure of the steady NSE and cannot be immediately applied to general contractive operators, the tools we employ may give insight into how a more general result of improved convergence rate can be constructed.

Under some stronger assumptions on the coefficients α of the minimization step, we first establish a globally accelerated rate of convergence of the difference between successive iterates for general contractive operators. The idea of this analysis is to characterize the improvement in the convergence rate by the balance between the success of the optimization problem solved at each step and the magnitude of the coefficients corresponding to earlier solutions. The common link between the analysis here and in section 4 is in characterizing the improvement in convergence rate by the gain from the optimization problem. We now fix some notation used in the remainder of the article:

$$e_k \coloneqq u_k - u_{k-1}, \quad \tilde{e}_k \coloneqq \tilde{u}_k - \tilde{u}_{k-1}, \quad w_k \coloneqq G(u_k) - u_k. \tag{2.1}$$

Here, (2.1) spells out three different types of error terms used in the analysis that follows. The first, $e_k = u_k - u_{k-1}$, is the difference between accelerated iterates. Theorem 2.5 and its Corollary 2.6 address global convergence with respect to this quantity. The second, $\tilde{e}_k = \tilde{u}_k - \tilde{u}_{k-1} = G(u_{k-1}) - G(u_{k-1})$, satisfies $\|\tilde{e}_k\| \leq r\,\|e_{k-1}\|$ under Assumption 2.1. The third, $w_k = G(u_k) - u_k$, can be thought of as the residual of the fixed-point iteration, $G(u) = u$. It can also, however, be thought of as the update step between the preaccelerated $\tilde{u}_{k+1}$ and the accelerated iterate at step k, by $\tilde{u}_{k+1} = u_k + w_k$. Each of these quantities is equivalent to the error between the current iterate and the fixed point u^* as follows. Starting with the residual we have

$$u_k - u^* = (u_k - G(u_k)) + (G(u_k) - u^*) = (u_k - G(u_k)) + (G(u_k) - G(u^*)).$$

Taking norms of both sides yields $\|u_k - u^*\|_* \leq \|w_k\|_* + r\,\|u_k - u^*\|_*$ with $r < 1$ under Assumption 2.1 by which

$$\|u_k - u^*\| \leq \frac{1}{1-r}\,\|w_k\|,$$

which implies $\|u_k - u^*\|_* \to 0$ as $\|w_k\|_* \to 0$. For the reverse inequality

$$\|G(u_k) - u_k\|_* \leq \|G(u_k) - G(u^*)\|_* + \|u^* - u_k\|_* \leq (1+r)\,\|u_k - u^*\|_*.$$

In term of the difference of consecutive iterates, we also have the error $\|u_{k+1} - u^*\|_*$ goes to zero as the sequence $\|e_j\|_* \to 0$, $j = k - m_k, \ldots, k$. In other words the er-

ror in the fixed-point iteration is controlled by the m consecutive differences between iterates in the depth-m algorithm. We proceed under the additional assumption that the optimization coefficients are bounded: there is some $\alpha_M > 0$ for which $|\alpha_j^k| \leq \alpha_M,\ j = k - m_k, \dots k$, for each iteration k. Establishing $\|e_k\|_*$ converges to zero shows $\{u_k\}$ is a Cauchy sequence hence has a limit in the ambient space X; and, by the following argument that limit must be u^*. Assuming boundedness of the optimization coefficients and using Assumption 2.1 we have by Step k[c] of Algorithm 2.2 that

$$\begin{aligned} u_{k+1} - u^* &= \left(\sum_{j=k-m_k}^{k} \alpha_j^{k+1} (G(u_j) - G(u_{k+1})) \right) + (G(u_{k+1}) - u^*) \\ &= \left(\sum_{j=k-m_k}^{k} \left(\sum_{n=k-m_k}^{j} \alpha_n^{k+1} \right) (G(u_j) - G(u_{j+1})) \right) + (G(u_{k+1}) - G(u^*)). \end{aligned}$$

The technique of telescoping the sum to subtract consecutive iterates will be repeated in the next theorem used to show convergence of $\|e_k\|$ to zero. Taking norms of both sides

$$\|u_{k+1} - u^*\|_* \leq r \|u_{k+1} - u^*\|_* + r\gamma_M \sum_{j=k-m_k}^{k} \|u_{j+1} - u_j\|_* ,$$

where $\gamma_M = \max\{(m-1)\alpha_M, 1\}$ arises from maximizing the sums $\sum_{n=k-m_k}^{j} \alpha_n^{k+1}$ for $j \leq k$ under the constraint $\sum_{n=k-m_k}^{k} \alpha_n^{k+1} = 1$. This in turn implies

$$\|u_{k+1} - u^*\|_* \leq \frac{r\gamma_M}{1-r} \sum_{j=k-m_k}^{k} \|u_{j+1} - u_j\| .$$

The reverse inequality follows more easily by $\|u_{k+1} - u_k\|_* \leq \|u_{k+1} - u^*\|_* + \|u_k - u^*\|_*$.

Having established the sufficiency of convergence to zero of either quantity $\|e_k\|_*$ or $\|w_k\|_*$, we next proceed with a global convergence analysis of $\|e_k\|_*$. To aid in the analysis here and in section 4 we introduce an intermediate quantity

$$u_k^\alpha = \sum_{j=k-m_k}^{k} \alpha_j^{k+1} u_j. \tag{2.2}$$

In particular, u_k^α satisfies $\|u_{k+1} - u_k^\alpha\|_* = \theta_k \|\tilde{u}_{k+1} - u_k\|_*$, where $0 < \theta_k \leq 1$ denotes the gain of the optimization of Step k[b] by

$$\min_{\sum_{j=k-m_k}^{k} \alpha_j^{k+1} = 1} \left\| \sum_{j=k-m_k}^{k} \alpha_j^{k+1} (\tilde{u}_{j+1} - u_j) \right\|_* = \theta_k \|\tilde{u}_{k+1} - u_k\|_* . \tag{2.3}$$

As $\theta_k = 1$ corresponds to the original fixed-point iteration, it is expected that $\theta_k < 1$ for all k.

THEOREM 2.5. *Let the sequences $\{u_k\}$ and $\{\tilde{u}_k\}$ be given by Algorithm 2.2. Let G satisfy Assumption 2.1. Suppose the first m_k coefficients of each α_j^{k+1} satisfy*

$|\sum_{j=k-m_k}^{l} \alpha_j^{k+1}| \leq \eta_k$, $l = k - m_k, \ldots, k-1$, *for some* $0 < \eta_k < 1$, *for each* $k > 1$. *Define* e_k *as in* (2.1). *Then* $\|e_2\|_* \leq (\kappa\theta_1 + \eta_k) \|e_1\|_*$ *and it holds for* $2 \leq k \leq m$ *that*

$$\|e_{k+1}\|_* \leq (r\theta_k + \eta_k) \|e_k\|_* + (r\theta_k\eta_{k-1} + \eta_k) \sum_{j=1}^{k-1} \|e_j\|_* . \tag{2.4}$$

For $k > m$, $(m_{k-1} = m_k = m)$ *it holds that*

(2.5)

$$\|e_{k+1}\|_* \leq (r\theta_k + \eta_k) \|e_k\|_* + (r\theta_k\eta_{k-1} + \eta_k) \sum_{j=k-m+1}^{k-1} \|e_j\|_* + r\theta_k\eta_{k-1} \|e_{k-m}\|_* ,$$

where the sums are understood to be zero if the final index is less that the starting index.

The above theorem shows that if η is small (requiring $\{\alpha_k^{k+1}\}$ close to 1), then Algorithm 2.2 can speed up convergence. The precise relationship between r, θ, and η to assure r-linear convergence at a rate greater than r is given in the corollary that follows. This estimate also suggests one of the ways the accelerated algorithm can stall by failing to increase or even maintain the standard fixed-point convergence rate if coefficients α_j^{k+1}, $j \leq k-1$, corresponding to iterates earlier in the history are too large.

Proof. The proof makes use of the decomposition

$$\|u_{k+1} - u_k\|_* \leq \|u_{k+1} - u_k^\alpha\|_* + \|u_k^\alpha - u_k\|_* . \tag{2.6}$$

Expanding u_k as a linear combination of $G(u_j)$, $j = k - 1 - m_{k-1}, \ldots, k-1$, using the property that the coefficients of α_j^k sum to unity and telescoping the resulting difference, we have

$$\begin{aligned}
\|G(u_k) - u_k\|_* &= \left\| \sum_{j=k-1-m_{k-1}}^{k-1} \alpha_j^k (G(u_k) - G(u_j)) \right\|_* \\
&= \left\| \sum_{j=k-m_{k-1}}^{k} \left(\sum_{n=k-m_{k-1}-1}^{j-1} \alpha_n^k \right) (G(u_j) - G(u_{j-1})) \right\| \\
&\leq \|G(u_k) - G(u_{k-1})\|_* + \eta_{k-1} \sum_{j=k-m_{k-1}}^{k-1} \|G(u_j) - G(u_{j-1})\|_* \\
&\leq r \left(\|e_k\|_* + \eta_{k-1} \sum_{j=k-m_{k-1}}^{k-1} \|e_j\|_* \right), \qquad (2.7)
\end{aligned}$$

where the last inequality follows from the Lipschitz property of G. By the same reasoning as above

$$\|u_k^\alpha - u_k\|_* = \left\| \sum_{j=k-m_k+1}^{k} \left(\sum_{n=k-m_k}^{j-1} \alpha_n^{k+1} \right) e_j \right\|_* \leq \eta_k \sum_{j=k-m_k+1}^{k} \|e_j\|_* . \tag{2.8}$$

Putting (2.3), (2.7), and (2.8) together into (2.6) establishes the result. ∎

Theorem 2.5 gives an essential worst-case scenario where no cancellation between the iterates is accounted for. Nonetheless, for a given bound η we can determine sufficient optimization gain θ to ensure r-linear convergence $\|e_{k+1}\|_* \leq r^k \|e_1\|_*$, where r is the convergence rate of the underlying fixed-point iteration. A similar formula can be derived for r-linear convergence at a given rate q.

COROLLARY 2.6. *Let the sequence $\{u_k\}$ be given by Algorithm 2.2 and suppose the hypotheses of Theorem 2.5 hold true. Then r-linear convergence with factor r holds for $k \geq 1$,*

$$\|u_{k+1} - u_k\|_* \leq r^k \|u_1 - u_0\|_*, \tag{2.9}$$

if it holds that $\theta_1 < 1 - \eta_1/r$, and for $m = 1, k > 1$, $\theta_k \leq (r - \eta_k)/(r + \eta_{k-1})$, and for $m > 1$

$$\theta_k \leq \begin{cases} \left(\frac{r^k - \eta_k(1-r^k)/(1-r)}{r^k + \eta_{k-1}(r-r^k)/(1-r)}\right), & k \leq m, \\ \left(\frac{r^m - \eta_k(1-r^{m-1})/(1-r)}{r^m + \eta_{k-1}(1-r^m)/(1-r)}\right), & k > m, \end{cases} \tag{2.10}$$

and $\eta_k < r^m(1-r)/(1-r^{m-1})$.

For instance, with $r = 0.9$ and $\eta_k = \eta_{k-1} = 0.1$, we have for the $m = 1$ case $\|e_{k+1}\|_* \leq r^k \|e_1\|_*$ for $\theta_1 = 8/9$ and $\theta_k \leq (r - \eta)/(r + \eta) = 0.8, k > 1$. For $m = 2$ we require $\theta_k \leq 0.62$ for $k > 2$. The proof follows directly from the result of Theorem 2.5 by induction on k, first for $k \leq m$, then for $k > m$, and is left to the interested reader.

The relevance of this result is that it quantifies a relation between the parameters of the optimization and the contractive operator for which global convergence at a given rate will be observed. In contrast, the results in section 4 and those in [10, 16] prove an accelerated rate of convergence only once the residual is small enough. Corollary 2.6 encompasses the preasymptotic regime, describing the global convergence seen in section 5, and is consistent with results of [12] for finite difference approximations to Richards' equation in which a lack of significant dependence on choice of initial iterate is demonstrated numerically. Some examples of how the θ_k and η_k relate in practice are shown in section 5.

3. The Picard iteration for steady NSE. We next consider the steady incompressible NSE. First, we give the mathematical framework and define some notation including the Picard iteration and associated Picard solution operator. Then we prove two important properties for the solution operator in order to relate it to the developed convergence theory.

3.1. Mathematical preliminaries. We consider an open connected domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ with Lipschitz boundary $\partial\Omega$. The $L^2(\Omega)$ norm and inner product will be denoted by $\|\cdot\|$ and $(\cdot, \cdot)$, and $L^2_0(\Omega)$ denotes the zero mean subspace of $L^2(\Omega)$. Throughout this paper, it is understood by context whether a particular space is scalar or vector valued, and we do not distinguish notation.

For the natural NSE velocity and pressure spaces, we denote $X := H^1_0(\Omega)$ and $Q := L^2_0(\Omega)$. In the space X, the Poincaré inequality is known to hold: there exists $\lambda > 0$, dependent only on $|\Omega|$, such that for every $v \in X$, $\|v\| \leq \lambda \|\nabla v\|$. The dual space of X will be denoted by X' with norm $\|\cdot\|_{-1}$. We use the notation $\langle \cdot, \cdot \rangle$ to denote the dual pairing of functions in X and X'.

Define the skew-symmetric, trilinear operator $b^* : X \times X \times X \to \mathbb{R}$ by

$$b^*(u, v, w) := \frac{1}{2}(u \cdot \nabla v, w) - \frac{1}{2}(u \cdot \nabla w, v),$$

and recall from, e.g., [7] that there exists M depending only on Ω such that

$$|b^*(u,v,w)| \leq M\|\nabla u\|\|\nabla v\|\|\nabla w\| \tag{3.1}$$

for every $u, v, w \in X$.

Let τ_h be a conforming, shape-regular, and simplicial triangulation of Ω with maximum element diameter h. Denote by P_k the space of degree k globally continuous piecewise polynomials with respect to τ_h, and P_k^{disc} the space of degree k piecewise polynomials on τ_h that can be discontinuous across elements.

Throughout the paper, we consider only discrete velocity-pressure spaces $(X_h,\ Q_h) \subset (X, Q)$ that satisfy the LBB condition: there exists a constant β, independent of h, satisfying

$$\inf_{q\in Q_h} \sup_{v\in X_h} \frac{(\nabla\cdot v, q)}{\|q\|\|\nabla v\|} \geq \beta > 0.$$

Common examples of such elements include (P_2, P_1) Taylor–Hood elements, and divergence-free (P_k, P_{k-1}^{disc}) Scott–Vogelius elements on meshes with particular structure [2, 19], and see [5, 8] for other stable and divergence-free elements. We denote the discretely divergence-free velocity space by

$$V_h \coloneqq \{v \in X_h,\ (\nabla\cdot v, q) = 0\ \forall q \in Q_h\}.$$

3.2. Discrete NSEs. We can now state the discrete steady NSE problem as follows: find $(u, p) \in (X_h, Q_h)$ satisfying for all $(v, q) \in (X_h, Q_h)$,

$$b^*(u,u,v) - (p, \nabla\cdot v) + \nu(\nabla u, \nabla v) = \langle f, v\rangle, \tag{3.2}$$

$$(\nabla\cdot u, q) = 0. \tag{3.3}$$

As shown in [7, 11, 15], solutions to (3.2)–(3.3) exist and satisfy

$$\|\nabla u\| \leq \nu^{-1}\|f\|_{-1}. \tag{3.4}$$

Define the data-dependent constant $\kappa := M\nu^{-2}\|f\|_{-1}$. If the data satisfy the condition $\kappa < 1$, then the system (3.2)–(3.3) is well-posed with a unique solution pair (u, p) [7]. We will assume throughout this paper that $\kappa < 1$, and refer to this as the *small data condition*.

It will be notationally convenient to also consider the V_h formulation of (3.2)–(3.3): find $u \in V_h$ satisfying for all $v \in V_h$

$$b^*(u,u,v) + \nu(\nabla u, \nabla v) = \langle f, v\rangle. \tag{3.5}$$

The equivalence of (3.5) to (3.2)–(3.3) follows from the inf-sup condition [11].

Remark 3.1. The accuracy of the discrete solution can be improved by the use of grad-div stabilization in the discrete NSE system, i.e., by adding $\gamma(\nabla\cdot u, \nabla\cdot v)$ to the momentum equation with $\gamma > 0$ [9, 13]. To simplify the presentation, we omit this important term, as all the analysis to follow will hold if grad-div is added to the system.

The Picard iteration, stated as follows, is a common approach to solving (3.2)–(3.3).

ALGORITHM 3.2 (Picard iteration for steady NSE).
Step 1*: Choose* $u_0 \in X_h$.
Step k*: Find* $(u_k, p_k) \in (X_h, Q_h)$ *satisfying for all* $(v, q) \in (X_h, Q_h)$,

$$b^*(u_{k-1}, u_k, v) - (p_k, \nabla \cdot v) + \nu(\nabla u_k, \nabla v) = \langle f, v\rangle, \tag{3.6}$$

$$(\nabla \cdot u_k, q) = 0. \tag{3.7}$$

This algorithm converges with contraction ratio κ for any initial guess, provided $\kappa < 1$ (see [7] for a standard proof). We note that the equivalent V_h formulation of Step k of the Picard iteration can be written as find $u_k \in V_h$ satisfying for all $v \in V_h$

$$b^*(u_{k-1}, u_k, v) + \nu(\nabla u_k, \nabla v) = \langle f, v\rangle. \tag{3.8}$$

3.3. Properties of the Picard solution operator for steady NSE. In order to analyze the effect of Anderson acceleration on the steady NSE Picard iteration, we next define a solution operator for the Picard linearization of the NSE from (3.8).

DEFINITION 3.3. *Define the Picard solution operator* $G : V_h \to V_h$ *as follows. Given* $w \in V_h$, $G(w) \in V_h$ *satisfies*

$$b^*(w, G(w), v) + \nu(\nabla G(w), \nabla v) = \langle f, v\rangle \ \forall v \in V_h. \tag{3.9}$$

By this definition of G, Step k of the Picard iteration (3.8) for the steady NSE can be written simply as set $u_k = G(u_{k-1})$. The problem (3.9) is linear, and since $f \in X'$ is assumed, Lax–Milgram theory can easily be applied to show that (3.9) is well-posed and thus that the solution operator G is well-defined. By taking $v = G(w)$, the trilinear term vanishes, leaving $\nu\|\nabla G(w)\|^2 = \langle f, G(w)\rangle \le \|f\|_{-1}\|\nabla G(w)\|$, and thus we have that for any $w \in V_h$,

$$\|\nabla G(w)\| \le \nu^{-1}\|f\|_{-1}. \tag{3.10}$$

We now prove that G is Lipschitz continuously (Frechét) differentiable, and a contractive operator with contraction ratio κ.

LEMMA 3.4. *The operator* G *is Lipschitz continuously (Frechét) differentiable, and for any* $w \in V_h$ *satisfies* $\|\nabla G'(w)\| \le \kappa$.

Remark 3.5. By standard fixed-point theory, Lemma 3.4 implies convergence of the Picard algorithm, Algorithm 3.2, under the small data condition $\kappa < 1$. Moreover, the convergence is global since the result will hold for any initial guess.

Proof. For $w,\ h \in V_h$, consider equations for $G(w)$ and $G(w+h)$ defined by (3.9):

$$\begin{aligned} b^*(w, G(w), v) + \nu(\nabla G(w), \nabla v) &= \langle f, v\rangle \ \forall v \in V_h, \\ b^*(w+h, G(w+h), v) + \nu(\nabla G(w+h), \nabla v) &= \langle f, v\rangle \ \forall v \in V_h. \end{aligned}$$

Subtracting yields

$$\begin{aligned} b^*(w+h, G(w+h) - G(w), v) + b^*(h, G(w), v) \\ + \nu(\nabla(G(w+h) - G(w)), \nabla v) = 0. \end{aligned} \tag{3.11}$$

Now, setting $v = G(w+h) - G(w)$ vanishes the first nonlinear term, and produces

$$\begin{aligned} \nu\|\nabla(G(w+h) - G(w))\|^2 &\le |b^*(h, G(w), G(w+h) - G(w))| \\ &\le M\|\nabla h\|\nabla G(w)\|\|\nabla(G(w+h) - G(w))\| \\ &\le \nu^{-1}M\|f\|_{-1}\|\nabla h\|\|\nabla(G(w+h) - G(w))\|, \end{aligned}$$

thanks to (3.1) and (3.10). This reduces immediately to

$$\|\nabla(G(w+h)-G(w))\| \le \kappa\|\nabla h\|, \tag{3.12}$$

which proves G is Lipschitz continuous and contractive with contraction ratio κ.

Next we show the G is Frechét differentiable. First define for a given $w \in V_h$ an operator $A_w : V_h \to V_h$ such that for all $h \in V_h$

$$b^*(h, G(w), v) + b^*(w, A_w(h), v) + \nu(\nabla A_w(h), \nabla v) = 0 \quad \forall v \in V_h. \tag{3.13}$$

Using properties for G and b^* established above together with Lax–Milgram theory it is easily verified the this linear problem is well-posed and thus A_w is well-defined.

Subtracting (3.13) from (3.11) provides

$$\begin{aligned} &b^*(w, G(w+h)-G(w)-A_w(h), v) + \nu(\nabla(G(w+h)-G(w)-A_w(h)), \nabla v) \\ &\quad = -b^*(h, G(w+h)-G(w), v) \\ &\quad \le M\|\nabla h\|\|\nabla(G(w+h)-G(w))\|\|\nabla v\| \\ &\quad \le \kappa M\|\nabla h\|^2\|\nabla v\| \end{aligned}$$

for all $v \in V_h$ thanks to (3.1) for the first inequality and (3.12) for the second. This proves that G is Frechét differentiable at w. From (3.12) and noting $w \in V_h$ is arbitrary establishes the result. ∎

4. The Anderson-accelerated Picard iteration for NSE. In this section, we define an Anderson-accelerated Picard iteration for the steady incompressible NSE. We then provide analysis to establish local convergence of the residual at a faster rate than that of the underlying fixed-point iteration. Numerical tests to back up the theory are shown in section 5. Although the usual Picard iteration Algorithm 3.2, is stable and globally convergent under a small data condition, its convergence rate can be sufficiently slow that it may fail in practice. The goal of combining the Picard iteration with Anderson acceleration is to improve convergence properties without introducing significant extra cost.

We define the Anderson-accelerated Picard iteration for the incompressible steady NSE (AAPINSE) as Algorithm 2.2 with G given by (3.9), the solution operator for the Picard linearized NSE. We note that the optimization step of Algorithm 2.2 is negligible in computational cost compared to the linear solve associated with applying the G operator, until $m = 4$ or so, when the cost becomes of the same order of magnitude as the linear solve. For even larger m, the cost of the optimization will dominate the cost of the linear solve. Interestingly, in our tests, there is little gain in convergence speed by using $m = 4$ over $m = 3$.

Combining Theorem 2.4 with Lemma 3.4 establishes local convergence of the AAPINSE under the assumption of uniformly bounded optimization parameters and a good initial guess. We prove next for AAPINSE that the acceleration does in fact improve the convergence rate of the fixed-point iteration based on the improvement given by the optimization. We provide results below for the cases of $m = 1$ and $m = 2$. We were unable to find an easily digestible proof for general m, but expect extension to greater values of m will follow along similar lines.

THEOREM 4.1 (improved convergence of the AAPINSE residual with $m = 1$). *Suppose* $0 < |\alpha_{k-1}^k| < \bar{\alpha}$ *for some fixed* $\bar{\alpha}$. *Then on any step where* $\alpha_{k-2}^k \neq 0$, *the* $m = 1$ *Anderson-accelerated Picard iterates satisfy*

$$\|\nabla(G(u_k)-u_k)\| \le \kappa\|\nabla(G(u_{k-1})-u_{k-1})\|\left(\theta_k + C_0\,\|\nabla(G(u_{k-2})-u_{k-2})\|\right) \tag{4.1}$$

with $C_0 = \nu^{-1}M\bar{\alpha}/(1-\kappa)^2$ *and where* $0 \le \theta_k \le \theta$ *for some fixed* $\theta < 1$ *represents the improvement from the optimization at Step* k *and satisfies* (2.3).

On any step where $\alpha_{k-2}^k = 0$, meaning $u_k = G(u_{k-1})$ (the standard Picard iteration) it holds that $\theta = 1$ and $\|G(u_k) - u_k\| \le \kappa\|G(u_{k-1}) - u_{k-1}\|$. Assuming $\theta_k < \theta$ for some $\theta < 1$, Theorem (4.1) yields an improved convergence rate as k increases, based on the success of the optimization problem. Unlike Theorem 2.5, the improved convergence rate is only local; however, the assumptions on the optimization coefficients are significantly weaker.

Proof. Define $e_k, \tilde{e}_k$, and w_k by (2.1). The structure of the proof is first to establish two key inequalities that bound the error by the residual

$$\|\nabla\tilde{e}_k\| \le \kappa\|\nabla e_{k-1}\|, \tag{4.2}$$

$$\|\nabla e_k\| \le \frac{1}{1-\kappa}\|\nabla w_{k-1}\|, \tag{4.3}$$

and then to use these for the NSE-specific main result. The first inequality (4.2) follows directly from (3.12). The second follows from the decomposition $e_k = (u_k - \tilde{u}_k) + (\tilde{u}_k - u_{k-1}) = -\alpha_{k-2}^k\tilde{e}_k + w_{k-1}$. Using (4.2) we have

$$\|\nabla e_k\| \le \kappa|\alpha_{k-2}^k|\|\nabla e_{k-1}\| + \|\nabla w_{k-1}\|. \tag{4.4}$$

The first term on the right of (4.4) can be controlled by the "backwards" inequality

$$\|\nabla e_{k-1}\| \le \frac{1}{(1-\kappa)|\alpha_{k-2}^k|}\|\nabla w_{k-1}\|, \tag{4.5}$$

which follows from the closed form expression for α_{k-2}^k for $m = 1$. It is based on the contribution u_k has from $\tilde{u}_{k-1}$, and requires the assumption α_{k-2}^k is nonzero. For $m = 1$ the optimization Step k[b] of Algorithm 2.2 can be written as $\alpha_{k-2}^k = \arg\min_{\alpha\in\mathbb{R}} \|\nabla\left(w_{k-1} + \alpha\left(w_{k-2} - w_{k-1}\right)\right)\|$, from which, exploiting the Hilbert space structure,

$$\alpha_{k-2}^k\|\nabla\left(w_{k-1} - w_{k-2}\right)\|^2 = \left(\nabla w_{k-1}, \nabla\left(w_{k-1} - w_{k-2}\right)\right).$$

Applying Cauchy–Schwarz on the right reduces this to $\|\nabla(w_{k-1} - w_{k-2})\| \le \frac{1}{|\alpha_{k-2}^k|}\|\nabla w_{k-1}\|$. By the identity $w_{k-1} - w_{k-2} = \tilde{e}_k - \tilde{e}_{k-1}$ and the triangle inequality,

$$(1-\kappa)\|\nabla e_{k-1}\| \le \|\nabla e_{k-1}\| - \|\nabla\tilde{e}_k\| \le \|\nabla\left(\tilde{e}_k - e_{k-1}\right)\| \le \frac{1}{|\alpha_{k-2}^k|}\|\nabla w_{k-1}\|, \tag{4.6}$$

where the first inequality follows from (4.2). Comparing the first and last terms of (4.6) verifies (4.5), and applying (4.5) to (4.4) validates (4.3).

To establish the main result of the theorem, we make use of the two following identities which follow from Algorithm 2.2 and $u_k = \alpha_{k-1}^k\tilde{u}_k + \alpha_{k-2}^k\tilde{u}_{k-1}$:

$$\alpha_{k-1}^k\tilde{e}_k = u_k - \tilde{u}_{k-1}, \tag{4.7}$$

$$e_k + \alpha_{k-2}^k e_{k-1} = \alpha_{k-1}^k w_{k-1} + \alpha_{k-2}^k w_{k-2}. \tag{4.8}$$

From $\tilde{u}_{k+1} = G(u_k)$, and (3.9), we have for $j \ge 1$

$$\nu(\nabla\tilde{u}_{j+1}, \nabla v) + b^*(u_j, \tilde{u}_{j+1}, v) = \langle f, v\rangle \quad \text{for all } v \in V_h. \tag{4.9}$$

Adding α_{k-1}^k times (4.9) with $j = k-1$ to α_{k-2}^k times (4.9) with $j = k-2$ and applying the definition of u_k together with $\alpha_{k-1}^k + \alpha_{k-2}^k = 1$ produces the equation for u_k:

$$\nu(\nabla u_k, \nabla v) + b^*(u_{k-1}, u_k, v) - b^*(e_{k-1}, \alpha_{k-2}^k \tilde{u}_{k-1}, v) = \langle f, v\rangle. \tag{4.10}$$

Subtracting (4.10) from (4.9), with $j = k$, we obtain

$$\nu(\nabla(\tilde{u}_{k+1} - u_k), \nabla v) + b^*(u_k, \tilde{u}_{k+1} - u_k, v) + b^*(e_k, u_k, v) + \alpha_{k-2}^k b^*(e_{k-1}, \tilde{u}_{k-1}, v) = 0,$$

which by (4.7) is equivalent to

$$\nu(\nabla w_k, \nabla v) + b^*(u_k, w_k, v) + b^*(e_k + \alpha_{k-2}^k e_{k-1}, \tilde{u}_{k-1}, v) + b^*(e_k, \alpha_{k-1}^k \tilde{e}_k, v) = 0. \tag{4.11}$$

Choosing $v = w_k$ in (4.11) vanishes the second term. Applying (3.1) and (4.8) yields

$$\|\nabla w_k\| \le M\nu^{-1}\left(\|\nabla(\alpha_{k-1}^k w_{k-1} + \alpha_{k-2}^k w_{k-2})\|\|\nabla \tilde{u}_{k-1}\| + \kappa|\alpha_{k-1}^k|\|\nabla e_k\|\|\nabla e_{k-1}\|\right).$$

Finally, applying $\|\nabla \tilde{u}_{k-1}\| \le \nu^{-1}\|f\|_{-1}$ from (3.10) together with (2.3) and (4.3) we have

$$\begin{aligned}\|\nabla w_k\| &\le \kappa\theta_k\|\nabla w_{k-1}\| + \kappa\nu^{-1}M|\alpha_{k-1}^k|\|\nabla e_k\|\|\nabla e_{k-1}\| \\ &\le \kappa\|\nabla w_{k-1}\|\left(\theta + \frac{\nu^{-1}M|\alpha_{k-1}^k|}{(1-\kappa)^2}\|\nabla w_{k-2}\|\right).\end{aligned}$$

∎

Together with the contraction of the underlying fixed-point iteration, Theorem 4.1 establishes convergence of the residual to zero after the first iterate that satisfies $\|\nabla w_{k-2}\| < (1-\kappa\theta)/(\kappa C_0)$, and contraction at a faster rate than the fixed-point iteration once $\|\nabla w_{k-2}\| < (1-\theta)/C_0$. The underlying assumption that the gain from the optimization step is bounded away from unity by some fixed θ for bounded coefficients on steps for which there is a contribution to u_k from $\tilde{u}_{k-1}$ is a reasonable characterization of conditions under which the algorithm should be expected to succeed.

Next, we establish improved convergence of AAPINSE for the case $m = 2$. The proof strategy is analogous to the $m = 1$ case, but with additional technical details arising from the additional parameter in the optimization step. We provide the $m = 2$ proof as an indication that the extension to greater m would follow the same essential idea.

THEOREM 4.2 (improved convergence of the AAPINSE residual with $m = 2$). *Suppose the coefficients $|\alpha_j^{k+1}|$ are bounded, $j = k-2, k-1, k$, the coefficient corresponding to the latest fixed-point iterate satisfies $|\alpha_k^{k+1}| > \breve{\alpha} > 0$ and $\alpha_k^{k+1} > \alpha_{k-2}^{k+1}$. Then on any step where at least one of α_{k-2}^{k+1} or α_{k-1}^{k+1} is nonzero, the $m = 2$ Anderson-accelerated Picard iteration satisfies*

$$\|\nabla(G(u_{k+1}) - u_{k+1})\| \le \kappa\theta_{k+1}\|\nabla(G(u_k) - u_k)\| + \mathcal{O}(\|\nabla(G(u_{k-2}) - u_{k-2})\|^2),$$

where $0 \le \theta_{k+1} \le \theta$ for some fixed $\theta < 1$ satisfies (2.3).

The proof follows the same general strategy as the $m = 1$ case, and again establishes local convergence of the algorithm (with mild assumptions on the coefficients) after the first iterate where $\|\nabla w_{k-2}\|$ is small enough, and with an improved rate when the accelerated solution is other than the fixed-point iterate. We precede the proof with a technical lemma to establish four key inequalities which bound the difference between accelerated iterates by the latest three residuals. As this is a general result (not NSE specific), it is posed in the same notation as section 2.

LEMMA 4.3. *Let the sequences* $\{u_k\}$ *and* $\{\tilde{u}_k\}$ *be given by Algorithm* 2.2 *with* $m = 2$*, and define* $e_k, \tilde{e}_k$*, and* w_k *by* (2.1)*. Let* $G : X \to X$ *satisfy Assumption* 2.1 *with constant* $r < 1$*, where* X *is a Hilbert space with norm* $\|\cdot\|_*$ *induced by inner product* $(\,\cdot,\cdot\,)_*$*. Then the following hold for* $k > 1$*:*

$$|\alpha_k^{k+1}| \, \|e_k\|_* \leq \frac{1}{(1-r)} \left(|1 - \alpha_{k-2}^{k+1}| \, \|w_{k-1}\|_* + |\alpha_{k-2}^{k+1}| \, \|w_{k-2}\|_* \right), \tag{4.12}$$

$$|1 - \alpha_k^{k+1}| \, \|e_k\|_* \leq \frac{1}{(1-r)} \left(|1 - \alpha_k^{k+1}| \, \|w_{k-1}\|_* + (1 + |\alpha_k^{k+1}|) \, \|w_k\|_* \right), \tag{4.13}$$

$$|\alpha_{k-2}^{k+1}| \, \|e_{k-1}\|_* \leq \frac{1}{(1-r)} \left(|1 - \alpha_k^{k+1}| \, \|w_{k-1}\|_* + |\alpha_k^{k+1}| \, \|w_k\|_* \right), \tag{4.14}$$

$$|1 - \alpha_{k-2}^{k+1}| \, \|e_{k-1}\|_* \leq \frac{1}{(1-r)} \left(|1 - \alpha_{k-2}^{k+1}| \, \|w_{k-1}\|_* + (1 + |\alpha_{k-2}^{k+1}|) \, \|w_{k-2}\|_* \right). \tag{4.15}$$

Proof. Without confusion, denote α_j^{k+1} by α_j for $j = \{k-2, k-1, k\}$. First, by Assumption 2.1 and the triangle inequality we have

$$(1-r) \, \|e_n\|_* \leq \|e_n\|_* - \|\tilde{e}_{n+1}\|_* \leq \|\tilde{e}_{n+1} - e_n\|_* = \|w_n - w_{n-1}\|_* . \tag{4.16}$$

To derive (4.12) and (4.15), write the Step k[b] minimization problem of Algorithm 2.2 in the equivalent form: find (α_k, β_0) that minimize

$$\|(\alpha_k(w_k - w_{k-1}) + \beta_0(w_{k-1} - w_{k-2}) + w_{k-2})\|_*^2$$

with $\beta_0 = \alpha_k + \alpha_{k-1}$ (so from $\alpha_k + \alpha_{k-1} + \alpha_{k-2} = 1$ we have $1 - \beta_0 = \alpha_{k-2}$). Exploiting the Hilbert space structure, the critical points α_k and β_0 are the solutions of

$$\alpha_k \, \|w_k - w_{k-1}\|_*^2 = -(w_k - w_{k-1}, \beta_0 w_{k-1} + (1-\beta_0) w_{k-2})_*, \tag{4.17}$$

$$\beta_0 \, \|w_{k-1} - w_{k-2}\|_*^2 = -(w_{k-1} - w_{k-2}, \alpha_k(w_k - w_{k-1}) + w_{k-2})_*. \tag{4.18}$$

Applying Cauchy–Schwarz and triangle inequalities together to (4.17) yields

$$|\alpha_k| \, \|w_k - w_{k-1}\|_* \leq |1 - \alpha_{k-2}| \, \|w_{k-1}\|_* + |\alpha_{k-2}| \, \|w_{k-2}\|_* . \tag{4.19}$$

Applying the same estimates together with (4.19) to (4.18) yields

$$|\beta_0| \, \|w_{k-1} - w_{k-2}\|_* \leq |1 - \alpha_{k-2}| \, \|w_{k-1}\|_* + (1 + |\alpha_{k-2}|) \, \|w_{k-2}\|_* . \tag{4.20}$$

Combining (4.16) with (4.19) (respectively, (4.20)) yields (4.12) (respectively, (4.15)).

To establish (4.13) and (4.14), follow the same process with the minimization problem written in the equivalent form: find (β_1, α_{k-2}) that minimize

$$\|(w_k + \beta_1(w_{k-1} - w_k) + \alpha_{k-2}(w_{k-2} - w_{k-1})\|_*^2$$

with $\beta_1 = \alpha_{k-1} + \alpha_{k-2}$ (which implies $1 - \beta_1 = \alpha_k$). ∎

The purpose of the four estimates (4.12)–(4.15) is to bound the terms $\|\nabla e_k\|$ and $\|\nabla e_{k-1}\|$ where they appear in the following estimates by $\|\nabla w_k\|, \|\nabla w_{k-1}\|$, and $\|\nabla w_{k-2}\|$, without introducing optimization coefficients other than α_k^{k+1} in the denominator. This is important as only α_k^{k+1} is justifiably bounded away from zero. We proceed now with the proof of Theorem 4.2 applying Lemma 4.3 with $\|v\|_* = \|\nabla v\|$ and $r = \kappa$.

Proof of Theorem 4.2. Recall the solution from Step k is defined as $u_{k+1} = \alpha_k\tilde{u}_{k+1}+\alpha_{k-1}\tilde{u}_k+\alpha_{k-2}\tilde{u}_{k-1}$ with α_j^{k+1} denoted α_j for $j = \{k-2, k-1, k\}$. From the problem definition (3.8), the following equation holds for $n = \{k-2, k-1, k, k+1\}$:

$$\nu(\nabla\tilde{u}_{n+1}, \nabla v) + b^*(u_n, \tilde{u}_{n+1}, v) = \langle f, v\rangle, \tag{4.21}$$

thus as in (4.10) we have

$$\nu(\nabla u_{k+1}, \nabla v) + \sum_{j=k-2}^{k} \alpha_j b^*(u_j, \tilde{u}_{j+1}, v) = \langle f, v\rangle.$$

Subtracting the above equation from (4.21) with $n = k+1$ yields

$$\begin{aligned}\nu(\nabla(\tilde{u}_{k+2} - u_{k+1}), \nabla v) &+ b^*(u_{k+1}, \tilde{u}_{k+2} - u_{k+1}, v)\\ &+ b^*(u_{k+1}, u_{k+1}, v) - \sum_{j=k-2}^{k} \alpha_j b^*(u_j, \tilde{u}_{j+1}, v) = 0.\end{aligned} \tag{4.22}$$

Next, rewrite the last two terms on the left-hand side in terms of e_k, $\tilde{e}_k$, and u_k^α given by (2.2):

$$\begin{aligned} &b^*(u_{k+1}, u_{k+1}, v) - \sum_{j=k-2}^{k} \alpha_j b^*(u_j, \tilde{u}_{j+1}, v)\\ &\quad= b^*\left(u_{k+1} - u_k^\alpha, \tilde{u}_{k-1}, v\right) + b^*\left(u_k^\alpha, \tilde{u}_{k-1}, v\right)\\ &\qquad + b^*(u_{k+1}, u_{k+1} - \tilde{u}_{k-1}, v) - \sum_{j=k-2}^{k} \alpha_j b^*(u_j, \tilde{u}_{j+1}, v)\\ &\quad= b^*\left(u_{k+1} - u_k^\alpha, \tilde{u}_{k-1}, v\right) + b^*(u_{k+1}, u_{k+1} - \tilde{u}_{k-1}, v)\\ &\qquad - b^*(u_k, \alpha_k(\tilde{e}_{k+1} + \tilde{e}_k), v) - b^*(u_{k-1}, \alpha_{k-1}\tilde{e}_k, v).\end{aligned}$$

Now using the identity $u_{k+1} - \tilde{u}_{k-1} = \alpha_k\tilde{e}_{k+1} + (\alpha_k + \alpha_{k-1})\tilde{e}_k$, produces

$$\begin{aligned} &b^*(u_{k+1}, u_{k+1}, v) - \sum_{j=k-2}^{k} \alpha_j b^*(u_j, \tilde{u}_{j+1}, v)\\ &= b^*\left(u_{k+1} - u_k^\alpha, \tilde{u}_{k-1}, v\right) + b^*(e_{k+1}, \alpha_k\tilde{e}_{k+1} + (\alpha_k + \alpha_{k-1})\tilde{e}_k, v) + b^*(e_k, \alpha_{k-1}\tilde{e}_k, v),\end{aligned}$$

and replacing e_{k+1} by

$$e_{k+1} = (u_{k+1} - \tilde{u}_{k+1}) + (\tilde{u}_{k+1} - u_k) = -(\alpha_{k-1} + \alpha_{k-2})\tilde{e}_{k+1} - \alpha_{k-2}\tilde{e}_k + (\tilde{u}_{k+1} - u_k),$$

gives

$$\begin{aligned} &b^*(u_{k+1}, u_{k+1}, v) - \sum_{j=k-2}^{k} \alpha_j b^*(u_j, \tilde{u}_{j+1}, v)\\ &= b^*\left(u_{k+1} - u_k^\alpha, \tilde{u}_{k-1}, v\right) - b^*((\alpha_{k-1} + \alpha_{k-2})\tilde{e}_{k+1} + \alpha_{k-2}\tilde{e}_k, \alpha_k\tilde{e}_{k+1}\\ &\quad + (\alpha_k + \alpha_{k-1})\tilde{e}_k, v) + b^*(\tilde{u}_{k+1} - u_k, \alpha_k\tilde{e}_{k+1} + (\alpha_k + \alpha_{k-1})\tilde{e}_k, v) + b^*(e_k, \alpha_{k-1}\tilde{e}_k, v).\end{aligned}$$

Thus, (4.22) can be written as

$$\begin{aligned}&\nu(\nabla w_{k+1},\nabla v)+b^*(u_{k+1},w_{k+1},v)+b^*\left(u_{k+1}-u_k^\alpha,\tilde{u}_{k-1},v\right)\\&\quad-b^*((\alpha_{k-1}+\alpha_{k-2})\tilde{e}_{k+1}+\alpha_{k-2}\tilde{e}_k,\alpha_k\tilde{e}_{k+1}+(\alpha_k+\alpha_{k-1})\tilde{e}_k,v)\\&\quad+b^*(w_k,\alpha_k\tilde{e}_{k+1}+(\alpha_k+\alpha_{k-1})\tilde{e}_k,v)+b^*(e_k,\alpha_{k-1}\tilde{e}_k,v)=0.\end{aligned}\tag{4.23}$$

Next, setting $v=w_{k+1}$ in (4.23) yields

$$\begin{aligned}\nu\|\nabla w_{k+1}\|^2&=-b^*\left(u_{k+1}-u_k^\alpha,\tilde{u}_{k-1},w_{k+1}\right)\\&\quad+b^*\left((\alpha_{k-1}+\alpha_{k-2})\tilde{e}_{k+1}+\alpha_{k-2}\tilde{e}_k,\alpha_k\tilde{e}_{k+1}+(\alpha_k+\alpha_{k-1})\tilde{e}_k,w_{k+1}\right)\\&\quad-b^*(w_k,\alpha_k\tilde{e}_{k+1}+(\alpha_k+\alpha_{k-1})\tilde{e}_k,w_{k+1})-b^*(e_k,\alpha_{k-1}\tilde{e}_k,w_{k+1}),\end{aligned}\tag{4.24}$$

and we proceed to bound the right-hand side terms. For the first term

$$\begin{aligned}b^*\left(u_{k+1}-u_k^\alpha,\tilde{u}_{k-1},w_{k+1}\right)&\le M\left\|\nabla\left(u_{k+1}-u_k^\alpha\right)\right\|\|\nabla\tilde{u}_{k-1}\|\|\nabla w_{k+1}\|\\&\le\nu^{-1}M\|f\|_{-1}\theta_k\|\nabla w_k\|\|\nabla w_{k+1}\|,\end{aligned}$$

using (2.2), (2.3), and $\|\nabla\tilde{u}_{k-1}\|\le\nu^{-1}\|f\|_{-1}$. The second term of (4.24) is majorized via

$$\begin{aligned}&M\|\nabla((\alpha_{k-1}+\alpha_{k-2})\tilde{e}_{k+1}+\alpha_{k-2}\tilde{e}_k)\|\|\nabla(\alpha_k\tilde{e}_{k+1}+(\alpha_k+\alpha_{k-1})\tilde{e}_k)\|\|\nabla w_{k+1}\|\\&\quad\le M\kappa^2\|\nabla w_{k+1}\|\left(|1-\alpha_k||\alpha_k\|\nabla e_k\|^2+|1-\alpha_{k-2}||\alpha_{k-2}|\left\|\nabla e_{k-1}\right\|^2\right)\\&\qquad+M\kappa^2\left\|\nabla w_{k+1}\right\|\left(|\alpha_k||\alpha_{k-2}|+|1-\alpha_k||1-\alpha_{k-2}|\right)\|\nabla e_k\|\|\nabla e_{k-1}\|.\end{aligned}\tag{4.25}$$

Applying (4.12)–(4.15) from Lemma 4.3, (4.25) is controlled by

$$\begin{aligned}&\frac{M\kappa^2}{(1-\kappa)^2}\|\nabla w_{k+1}\|\\&\quad\times\Big(\ \big(|1-\alpha_{k-2}|\left\|\nabla w_{k-1}\right\|+|\alpha_{k-2}|\left\|\nabla w_{k-2}\right\|\big)\\&\qquad\big(|1-\alpha_k|\left\|\nabla w_{k-1}\right\|+(1+|\alpha_k|)\left\|\nabla w_k\right\|\big)+\big(|1-\alpha_k|\left\|\nabla w_{k-1}\right\|+|\alpha_k|\left\|\nabla w_k\right\|\big)\\&\qquad\big(|1-\alpha_{k-2}|\left\|\nabla w_{k-1}\right\|+(1+|\alpha_{k-2}|)\left\|\nabla w_{k-2}\right\|\big)\Big)\\&\quad+\Big(\big(|1-\alpha_{k-2}|\left\|\nabla w_{k-1}\right\|+|\alpha_{k-2}|\left\|\nabla w_{k-2}\right\|\big)\big(|1-\alpha_k|\left\|\nabla w_{k-1}\right\|+|\alpha_k|\left\|\nabla w_k\right\|\big)\\&\qquad+\big(|1-\alpha_k|\left\|\nabla w_{k-1}\right\|+(1+|\alpha_k|)\left\|\nabla w_k\right\|\big)\\&\qquad\big(|1-\alpha_{k-2}|\left\|\nabla w_{k-1}\right\|+(1+|\alpha_{k-2}|)\left\|\nabla w_{k-2}\right\|\big)\Big).\end{aligned}\tag{4.26}$$

Using (4.12) and (4.15), the third term on the right-hand side of (4.24) is bounded by

$$\begin{aligned}&M\kappa\|\nabla w_{k+1}\|\|\nabla w_k\|\left(|\alpha_k|\|\nabla e_k\|+|1-\alpha_{k-2}|\|\nabla e_{k-1}\|\right)\\&\quad\le\frac{M\kappa}{(1-\kappa)}\|\nabla w_{k+1}\|\|\nabla w_k\|\big(2|1-\alpha_{k-2}|\left\|\nabla w_{k-1}\right\|+(1+2|\alpha_{k-2}|)\left\|\nabla w_{k-2}\right\|\big).\end{aligned}$$

By the assumption $\alpha_k\ge\alpha_{k-2}$ we have

$$\alpha_{k-1}=(\alpha_{k-1}+\alpha_{k-2})-\alpha_{k-2}=(1-\alpha_k)-\alpha_{k-2}\le(1-\alpha_{k-2})-\alpha_{k-2}.$$

Using this together with (4.12), (4.14), and (4.15), the last term of (4.24) is controlled by

$$\begin{aligned} &M\kappa \left\|\nabla w_{k+1}\right\| |\alpha_{k-1}| \|\nabla e_k\| \|\nabla e_{k-1}\| \\ &\quad \leq \frac{M\kappa}{(1-\kappa)^2} \left\|\nabla w_{k+1}\right\| \frac{1}{|\alpha_k|} \big(|1-\alpha_{k-2}| \left\|\nabla w_{k-1}\right\| + |\alpha_{k-2}| \left\|\nabla w_{k-2}\right\| \big) \\ &\qquad \times \big((|1-\alpha_k| + |1-\alpha_{k-2}|) \left\|\nabla w_{k-1}\right\| + |\alpha_k| \left\|\nabla w_k\right\| + (1+|\alpha_{k-2}|) \left\|\nabla w_{k-2}\right\| \big). \end{aligned} \tag{4.27}$$

Finally, combining (4.24)–(4.27) yields

$$\begin{aligned} \left\|\nabla w_{k+1}\right\| &\leq \kappa\theta_k \left\|\nabla w_k\right\| \\ &\qquad + \frac{M\nu^{-1}\kappa}{(1-\kappa)} \Bigg(\left\|\nabla w_k\right\| \left(c_1 \left\|\nabla w_{k-1}\right\| + c_2 \left\|\nabla w_{k-2}\right\|\right) \\ &\qquad\qquad + \left(\frac{\kappa}{1-\kappa} + \frac{1}{\breve{\alpha}(1-\kappa)} \right) \times \mathcal{O}\left(\left\|\nabla w_{k-2}\right\|^2 \right) \Bigg) \\ &= \kappa\theta_k \left\|\nabla w_k\right\| + \mathcal{O}\left(\left\|\nabla w_{k-2}\right\|^2 \right), \end{aligned}$$

where all the implicitly defined constants are sums and products of the bounded $|\alpha_k|, |1-\alpha_k|, |\alpha_{k-2}|$, and $|1-\alpha_{k-2}|$. The only optimization coefficient that makes an appearance in a denominator is α_k^{k+1}. It is a reasonable assumption that this coefficient is bounded away from zero as without a contribution from the latest fixed-point iterate $\tilde{u}_{k+1}$, the new solution u_{k+1} remains spanned by the same (less one) basis vectors as u_k and should not yield an improved residual. ∎

5. Numerical experiments. Here we present numerical experiments to show the improved convergence provided by the Anderson acceleration for solving the steady NSE. As illustrated below, Anderson acceleration can provide fast convergence even when Newton and usual Picard iterations fail. Our test problems are the 2 dimensional (2D) and 3 dimensional (3D) driven cavity, at varying Reynolds numbers. All computations were done in MATLAB with the authors' codes, and fminsearch with an initial guess of $(1, 0, \ldots, 0)$ was used to solve the optimization problems.

5.1. 2D lid driven cavity. We now test AAPINSE on the 2D driven cavity, at benchmark values of $Re = 1000$, 2500, and 5000, and compare results with those of the usual Picard and Newton methods. The 2D driven cavity uses a domain $\Omega = (0,1)^2$, with no-slip boundary conditions on the sides and bottom, and a "moving lid" on the top which is implemented by enforcing the Dirichlet boundary condition $u(x,1) = \langle 1, 0 \rangle^T$. There is no forcing ($f = 0$), and the kinematic viscosity is set to be $\nu \coloneqq Re^{-1}$. We discretize with (P_2, P_1) Taylor–Hood elements on a $\frac{1}{64}$ uniform triangular mesh that provides 37,507 total degrees of freedom, and use $u_0 = 0$ as the initial guess. A sparse direct solver (backslash) was used to solve the linear systems. Our results below do not use grad-div stabilization, but we also ran the tests with grad-div stabilization using the parameter $\gamma = 1$ (see Remark 3.1), and found very similar results to those without stabilization (slightly better convergence, but essentially negligible), so we omit these results.

Plots of the velocity solutions from 4 level Anderson-accelerated Picard solvers at $Re = 1000$, 2500, and 5000 are shown in Figure 1, and these solutions match well those from recent literature [4].

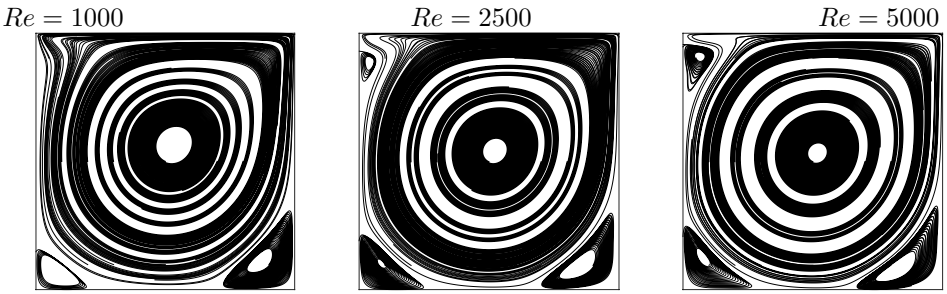

FIG. 1. *Streamline plots of the solutions from 4 level Anderson-accelerated Picard solvers at varying Re.*

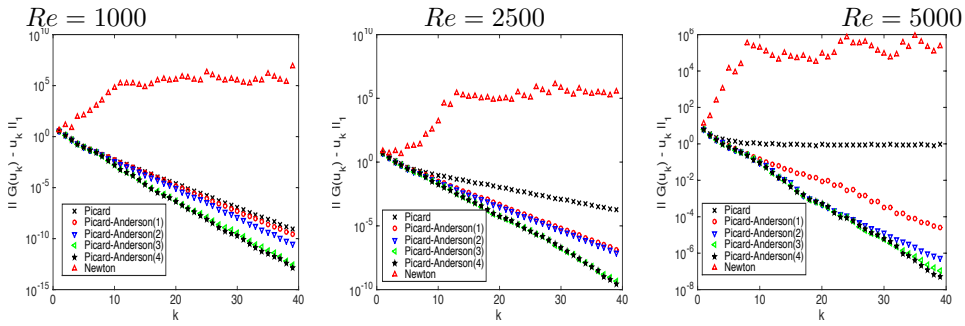

FIG. 2. *Convergence of the various nonlinear solvers for the 2D cavity test at varying Re.*

Convergence results for Re =1000, 2500, and 5000 are shown in Figure 2. In all cases, we observe an improvement from Anderson acceleration for the Picard method, with an increase in improvement for higher Reynolds numbers. That is, while Anderson acceleration offers just a modest gain for $Re = 1000$, for $Re = 2500$ the gain is much greater, and for $Re = 5000$ Picard appears to fail (or at least will take many, many iterations to converge to a reasonable tolerance). The Newton solver fails in all 3 cases, as we observe the Newton residuals are very large and show no signs of getting small by $n = 40$ (this is basic Newton with no relaxation; it is a subject of our future work to see how Anderson acceleration will help the convergence of Newton iterations). We observe the best Anderson performance in all cases with $m = 4$, however, the convergence behaviors with $m = 3$ and $m = 4$ are generally close for all 3 cases. Figure 3 shows the computed gain θ_k for each optimization problem, for each value of m and Re investigated. In general, we see smaller values of θ_k (greater gain) with increasing m. We compare our numerical results with the theoretical ones by comparing median values of the gain θ_k and convergence rate κ taken over all iterations k for each m and value of Re investigated. We estimate κ by taking the median of $\frac{\|G(u_k)-u_k\|_1}{\|G(u_{k-1})-u_{k-1}\|_1}$ over all iterations in the $m = 0$ (Picard) case. For $m > 0$ the convergence rate is computed by the same respective ratio over all iterations past the mth one. Table 1 shows the computed θ_{med}, and Table 2 compares the theoretical convergence rate approximated to first order by $\kappa_{med}\theta_{med}$ to the computed mean convergence rate taken over all iterations. We find the computed rates bounded below the theoretical ones, with a better prediction for lower values of Re.

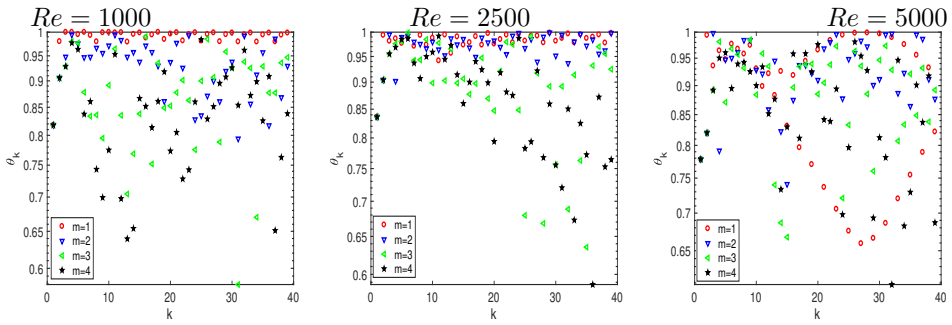

FIG. 3. *The θ_k versus k for varying m for the $Re = 2500$ driven cavity simulation.*

TABLE 1
Shown below are median values of θ_k for the 2D driven cavity simulations.

	$Re = 1000$	$Re = 2500$	$Re = 5000$
m	θ_{med}	θ_{med}	θ_{med}
1	0.9976	0.9872	0.9306
2	0.9684	0.9783	0.9368
3	0.9370	0.9137	0.9051
4	0.9229	0.8987	0.8919

TABLE 2
Shown below are median values of the convergence rates (median of successive residual ratios), and an estimate of the predicted rate of our theory, using the product of the median gain of the optimization θ^m_{med} with the median convergence rate of the Picard iteration, for varying Re and m.

	$Re = 1000$	$Re = 1000$	$Re = 2500$	$Re = 2500$	$Re = 5000$	$Re = 5000$
m	Conv rate	$\theta^m_{med} \cdot 0.5843$	Conv rate	$\theta^m_{med} \cdot 0.7852$	Conv rate	$\theta^m_{med} \cdot 0.9614$
0	0.5843	-	0.7852	-	0.9614	-
1	0.5533	0.5829	0.6417	0.7751	0.7041	0.7307
2	0.5162	0.5658	0.6372	0.7682	0.6866	0.7356
3	0.4330	0.5475	0.5687	0.7175	0.6239	0.7107
4	0.4301	0.5392	0.5365	0.7056	0.5920	0.7003

Theorem 2.5 and Corollary 2.6 are expected to give an overly pessimistic view of whether the algorithm should converge and at what rate, assuming the optimization coefficients are sufficiently small in comparison to the gain from the optimization. However we next see in these examples that the coefficients do not lie far outside the range required from the theory. For $m = 1$ where there is less cancellation error between the iterates to take into account, the theory predicts convergence in the far preasymptotic regime, i.e., regardless of the magnitude of the residual. Figure 4 shows for the simulation with $Re = 1000$ and $m = 1, 2, 3$, the computed θ_k and η_k, the latter of which for $m = 1$ is simply $|\alpha^{k+1}_{k-1}|$. For comparison, also shown are the theoretical values of θ from Corollary 2.6 which guarantee convergence in the sense $\|u_{k+1} - u_k\|_1 \leq q^k \|u_1 - u_0\|_1$ at rate $q = \kappa$ and, respectively, $q < 1$. The top line of the plot showing θ_k sufficient for $q < 1$ is in places greater than one, which shows the optimization coefficients from the $Re = 1000$ simulation using $\kappa = 0.58$ are small enough to guarantee global convergence regardless of the value of $\theta_k \leq 1$. As expected, the theory better describes the convergence behavior for smaller m, as the lack of optimality due to not accounting for cancellation error makes these estimates far from sharp.

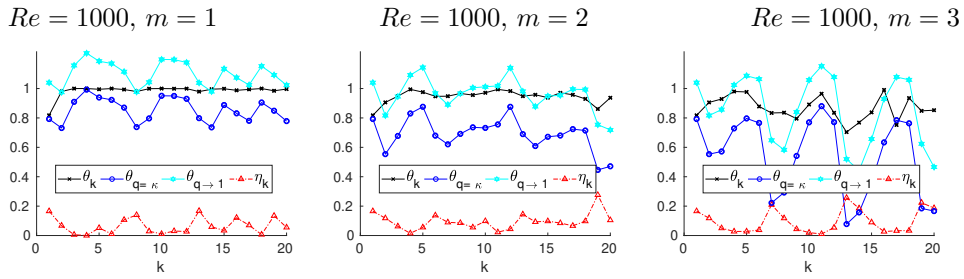

FIG. 4. *The gain θ_k at each step compared to the the theoretical gain $\theta_{q=\kappa}$ and $\theta_{q\to 1}$ sufficient to show convergence at rate $q = \kappa$ and $q < 1$ from Corollary 2.6. Also shown, the optimization parameters η_k for $Re = 1000$ with $m = 1, 2, 3$.*

TABLE 3

Shown below are timing (in seconds) for assembly, linear solve, and optimization steps for each test (averaged over all iterations).

	$Re = 1000$			$Re = 2500$			$Re = 5000$		
m	Assembly	Solve	opt.	Assembly	Solve	opt.	Assembly	Solve	Opt.
1	8.04e-0	2.45e-0	1.03e-1	8.02e-0	2.85e-0	1.47e-1	8.32e-0	2.87e-0	1.58e-1
2	7.94e-0	2.72e-0	2.80e-1	7.97e-0	2.95e-0	3.75e-1	7.77e-0	2.59e-0	4.85e-1
3	7.72e-0	2.20e-0	8.63e-1	8.11e-0	2.02e-0	7.78e-1	8.10e-0	2.57e-0	9.25e-1
4	8.19e-0	2.30e-0	1.72e-0	8.14e-0	2.57e-0	1.40e-0	7.73e-0	2.23e-0	2.21e-0

Last, for this test, we show the timings for the assembly, linear solve, and optimization steps in Table 3. Here, we display the mean values for each, taken over all iterations. We note that the assembly was not parallelized. While the assembly and solve times do not appear to scale with m, the optimization step does, and appears to roughly double when m is increased by 1. While the optimization step timing is an order of magnitude less than that of the linear solve when $m = 1$, by $m = 4$ the optimization step is no longer negligible, and for $Re = 5000$ has a cost the same as that of the linear solver.

5.2. 3D lid driven cavity. Next, we test AAPINSE on the 3D lid driven cavity problem. This problem is similar to the 2D case, and uses no-slip boundary conditions on all walls, $u = \langle 1, 0, 0\rangle^T$ on the moving lid, no forcing, and $\nu = \frac{1}{400}$. We compute with (P_3, P_2^{disc}) Scott–Vogelius elements on a barycenter refined uniform tetrahedral mesh that provides 206,874 total degrees of freedom. We tested the algorithm with different levels of optimization, all with initial guesses of zero in the interior but satisfying the boundary conditions. Figure 5 shows a visualization of the computed solution with $m = 4$, which is in good agreement with [18]. To solve the linear systems that arose at each iteration, we decomposed the saddle point matrix with a block LU factorization via

(5.1)
$$\begin{pmatrix} A_k & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} \hat{u}_k \\ \hat{p}_k \end{pmatrix} = \begin{pmatrix} A_k & 0 \\ B^T & -B^T A_k^{-1} B \end{pmatrix} \begin{pmatrix} I & A_k^{-1} B \\ 0 & I \end{pmatrix} \begin{pmatrix} \hat{u}_k \\ \hat{p}_k \end{pmatrix} = \begin{pmatrix} \hat{f} \\ \hat{g} \end{pmatrix},$$

which leads to two solves with coefficient matrix A_k, and one solve with the Schur complement $B^T A_k^{-1} B$ as the coefficient matrix. A sparse direct solver was used to solve systems with coefficient matrix A_k, and we did this by creating and reusing an LUPQ factorization in MATLAB at each nonlinear iteration. For the Schur complement system, we used BICGSTAB as an outer solver with tolerance 1e-10, and to

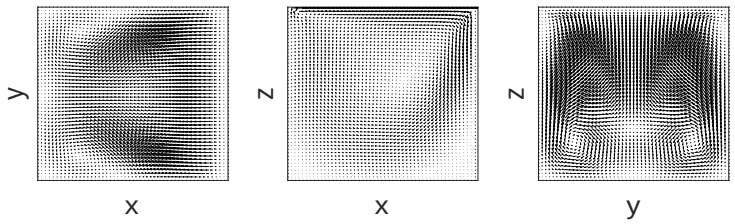

FIG. 5. *Shown above are mid-slice plane plots for the 3D driven cavity simulations at* $Re = 400$ *using the Picard–Anderson* (4) *method, these plots are in good agreement with* [18].

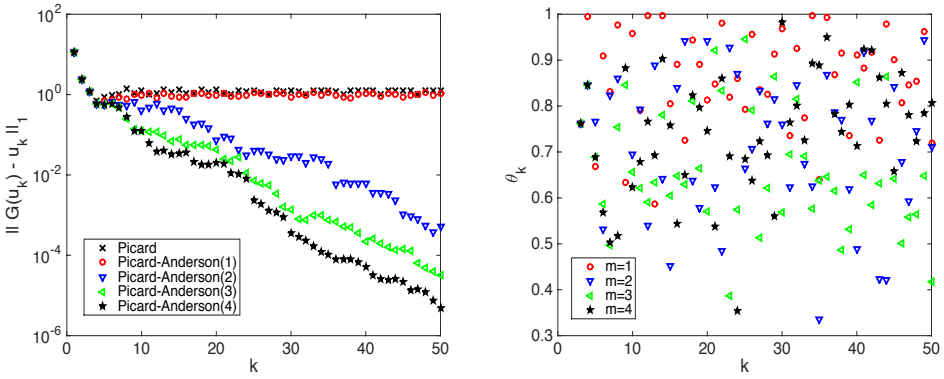

FIG. 6. *Convergence (left) and* θ_k *(right) for AAPINSE for the 3D Cavity test at* $Re = 400$.

precondition it we used the pressure mass matrix and adjusted the saddle point linear system by adding grad-div stabilization with parameter 100 (which has no effect on the solution since divergence-free elements are being used). We note this is in the same spirit as was done in [3] for the treatment of the Schur complement, where a purely algebraic augmentation was used. The inner Schur complement solves were accomplished using the factorization of A_k already created from the first solve using A_k at each iteration. The typical number of outer BICGSTAB iterations needed for each Schur complement solve was just 1 or 2, and this did not vary with m.

Figure 6 shows the convergence of residuals (left) and values of θ_k (right) for AAPINSE with varying m, for the first 50 iterations. From the convergence plot, we observe that the Picard iteration and AAPINSE with $m = 1$ essentially fail, but a dramatic improvement is obtained using $m \geq 2$, sufficient to provide convergence. Clear improvements in convergence are observed as m is increased from 2 to 3, and 3 to 4. The θ_k plot shows the computed gain for each optimization problem and each value of m. Here it is noted that the larger values of θ_k for each m tend to be further from unity as m increases. Overall, there is a considerable spread in the θ_k values for each m. Table 4 summarizes the computed median values of θ_k for the different m, and we observe lower values for $m \geq 2$. Interestingly, the median θ_k for $m = 3$ is much lower than for $m = 4$, even though $m = 4$ converges faster; this suggests the median may not be the best measure to use: since we essentially have exponential convergence with rate $\kappa\theta_k$ at each step, one very small θ_k can make up for several larger

TABLE 4
Shown below are median values of θ_k for the 3D driven cavity simulations.

	$Re = 400$
m	θ_{med}
1	0.8314
2	0.7532
3	0.6400
4	0.7598

TABLE 5
Shown below are median values of the convergence rates (median of ratios of successive residuals), and an estimate of the predicted rate of our theory, using the product of the median gain of the optimization θ^m_{med} with the median convergence rate of the Picard iteration $\kappa_{med} \approx 0.9936$.

m	Conv rate	$\theta^m_{med} \cdot \kappa_{med}$
0	0.9936	-
1	0.9752	0.8261
2	0.8748	0.7484
3	0.8120	0.6359
4	0.8136	0.7550

TABLE 6
Shown below are timing (in seconds) for assembly, linear solve, and optimization steps for each test for the 3D cavity problem (averaged over all iterations).

m	Assembly	Solve	Opt.
1	8.43e+1	1.75e+1	8.68e-1
2	8.39e+1	2.02e+1	3.12e+0
3	8.74e+1	1.82e+1	8.37e+0
4	8.28e+2	1.72e+1	1.45e+1

ones. Table 5 compares the computed median convergence rate over all iterations to the theoretical convergence approximated by $\kappa_{med}\theta_{med}$, where κ_{med} is calculated by taking the median of $\frac{\|G(u_k)-u_k\|_1}{\|G(u_{k-1})-u_{k-1}\|_1}$ over all iterations in the $m = 0$ (Picard) case. For $m > 0$ the convergence rate is computed by the same respective ratio over all iterations past the mth one. These results differ from the 2D case in that the computed rates are not bounded above by the approximated theoretical rates (although for $m = 4$ the values are close). However, in this case the computed median convergence rate is approximately 1 with residuals not monotonically decreasing, which suggests the this computation does not satisfy the small data condition (the operator G is not contractive). In particular, (4.16) no longer implies the key estimates (4.12)–(4.15) in the $m = 2$ case and, similarly, (4.6) does not imply (4.3) for the $m = 1$ analysis.

Last, for this test, we give the timings for the assembly, linear solve, and optimization steps in Table 6. We display the median values for each, taken over all iterations. We note that as in the 2D tests, the assembly was not parallelized. Results here are similar to those found in the 2D tests: While the assembly and solve times do not appear to scale with m, the optimization step does, and appears to roughly double with each increase in m. For small m, the optimization cost is negligible compare to the linear solve time, however, by $m = 4$ it is close the cost of a linear solve.

6. Conclusions. In this paper, we showed that Anderson acceleration applied to the Picard iteration can provide a significant, and sometimes dramatic, improvement in convergence behavior. We proved this analytically and, to our knowledge, this is the first proof of Anderson acceleration providing (essentially) guaranteed im-

proved convergence for a fixed point iteration and, in particular, for a nonlinear fluid system. We also give results of several numerical tests that show the gains provided by Anderson acceleration for this problem can even be an enabling technology in the sense that it allows for convergence when both the Picard and Newton iterations fail. The presented theory is based on characterizing the improvement in the fixed-point convergence rate by the gain from the optimization problem. Our numerical results appear to capture the highest order effects of our theory, and clearly reveal dramatic improvement created by Anderson acceleration.

Important future work includes extending these ideas to the recently proposed Incremental Picard Yosida variant of the Picard iteration for the steady NSE [14], which has similar convergence properties of Picard but has linear systems that are much easier to solve. We also plan to explore whether Anderson acceleration can be used to aid in the convergence of Newton iterations for steady NSE, since (basic) Newton tends to fail for higher Re. Applying Anderson acceleration to steady multiphysics problems such as MHD may also be a fruitful pursuit.

Acknowledgment. The authors would like to thank the anonymous referees whose suggestions improved the clarity of this paper.

REFERENCES

[1] D. G. Anderson, *Iterative procedures for nonlinear integral equations*, J. ACM, 12 (1965), pp. 547–560, https://doi.org/10.1145/321296.321305.

[2] D. Arnold and J. Qin, *Quadratic velocity/linear pressure Stokes elements*, in Advances in Computer Methods for Partial Differential Equations VII, R. Vichnevetsky, D. Knight, and G. Richter, eds., IMACS, New Brunswick, NJ, 1992, pp. 28–34.

[3] M. Benzi and M. Olshanskii, *An augmented Lagrangian-based approach to the Oseen problem*, SIAM J. Sci. Comput., 28 (2006), pp. 2095–2113, https://doi.org/10.1137/050646421.

[4] C.-H. Bruneau and M. Saad, *The 2d lid-driven cavity problem revisited*, Comput. & Fluids, 35 (2006), pp. 326–348, https://doi.org/10.1016/j.compfluid.2004.12.004.

[5] R. S. Falk and M. Neilan, *Stokes complexes and the construction of stable finite elements with pointwise mass conservation*, SIAM J. Numer. Anal., 51 (2013), pp. 1308–1326, https://doi.org/10.1137/120888132.

[6] H. Fang and Y. Saad, *Two classes of multisecant methods for nonlinear acceleration*, Numer. Linear Algebra Appl., 16 (2009), pp. 197–221, https://doi.org/10.1002/nla.617.

[7] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier–Stokes Equations: Theory and Algorithms*, Springer, Berlin, 1986.

[8] J. Guzmán and M. Neilan, *Conforming and divergence-free Stokes elements on general triangular meshes*, Math. Comp., 83 (2014), pp. 15–36, https://doi.org/10.1090/S0025-5718-2013-02753-6.

[9] E. Jenkins, V. John, A. Linke, and L. Rebholz, *On the parameter choice in grad-div stabilization for the Stokes equations*, Adv. Comput. Math., 40 (2014), pp. 491–516, https://doi.org/10.1007/s10444-013-9316-1.

[10] C. Kelley, *Numerical methods for nonlinear equations*, Acta Numer., 27 (2018), pp. 207–287, https://doi.org/10.1017/S0962492917000113.

[11] W. Layton, *An Introduction to the Numerical Analysis of Incompressible Viscous Flows*, SIAM, Philadelphia, 2008.

[12] P. A. Lott, H. F. Walker, C. S. Woodward, and U. M. Yang, *An accelerated Picard method for nonlinear systems related to variably saturated flow*, Adv. Water Resour., 38 (2012), pp. 92–101, https://doi.org/10.1016/j.advwatres.2011.12.013.

[13] M. A. Olshanskii and A. Reusken, *Grad-div stabilization for the Stokes equations*, Math. Comp., 73 (2004), pp. 1699–1718, https://doi.org/10.1090/S0025-5718-03-01629-6.

[14] L. Rebholz, A. Viguerie, and M. Xiao, *Efficient nonlinear iteration schemes based on algebraic splitting for the incompressible Navier-Stokes equations*, Math. Comput., to appear.

[15] R. Temam, *Navier-Stokes Equations*, Elsevier, North-Holland, 1991.

[16] A. Toth and C. T. Kelley, *Convergence analysis for Anderson acceleration*, SIAM J. Numer. Anal., 53 (2015), pp. 805–819, https://doi.org/10.1137/130919398.

[17] H. F. Walker and P. Ni, *Anderson acceleration for fixed-point iterations*, SIAM J. Numer. Anal., 49 (2011), pp. 1715–1735, https://doi.org/10.1137/10078356X.

[18] K. Wong and A. Baker, *A* 3*d incompressible Navier-Stokes velocity-vorticity weak form finite element algorithm*, Internat. J. Numer. Methods Fluids, 38 (2002), pp. 99–123, https://doi.org/10.1002/fld.204.

[19] S. Zhang, *A new family of stable mixed finite elements for the* 3*d Stokes equations*, Math. Comp., 74 (2005), pp. 543–554, https://doi.org/10.1090/S0025-5718-04-01711-9.