MATHEMATICS OF COMPUTATION
Volume 69, Number 230, Pages 667–689
S 0025-5718(99)01152-7
Article electronically published on May 17, 1999

UZAWA TYPE ALGORITHMS FOR NONSYMMETRIC SADDLE POINT PROBLEMS

JAMES H. BRAMBLE, JOSEPH E. PASCIAK, AND APOSTOL T. VASSILEV

ABSTRACT. In this paper, we consider iterative algorithms of Uzawa type for solving linear nonsymmetric saddle point problems. Specifically, we consider systems, written as usual in block form, where the upper left block is an invertible linear operator with positive definite symmetric part. Such saddle point problems arise, for example, in certain finite element and finite difference discretizations of Navier–Stokes equations, Oseen equations, and mixed finite element discretization of second order convection-diffusion problems. We consider two algorithms, each of which utilizes a preconditioner for the operator in the upper left block. Convergence results for the algorithms are established in appropriate norms. The convergence of one of the algorithms is shown assuming only that the preconditioner is spectrally equivalent to the inverse of the symmetric part of the operator. The other algorithm is shown to converge provided that the preconditioner is a sufficiently accurate approximation of the inverse of the upper left block. Applications to the solution of steady-state Navier–Stokes equations are discussed, and, finally, the results of numerical experiments involving the algorithms are presented.

1. INTRODUCTION

This paper provides an analysis for Uzawa type methods applied to the solution of linear nonsymmetric saddle point systems. Such systems arise in certain discretizations of Navier–Stokes equations and mixed discretizations of second order elliptic problems with convective terms (cf. [9], [11], [14], [17]). The theory in this paper is an extension of that for symmetric saddle point problems developed in [4].

Let H_1 and H_2 be finite dimensional Hilbert spaces with inner products which we shall denote by $(\cdot,\cdot)$. There is no ambiguity even though we use the same notation for the inner products on both of these spaces, since the particular inner product will be identified by the type of functions appearing. We consider the system

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix}, \tag{1.1}$$

Received by the editor July 8, 1997 and, in revised form, June 23, 1998.

1991 *Mathematics Subject Classification.* Primary 65N22, 65N30, 65F10.

Key words and phrases. Indefinite systems, iterative methods, preconditioners, saddle point problems, nonsymmetric saddle point systems, Navier-Stokes equations, Oseen equations, Uzawa algorithm.

This manuscript was written under contract number DE-AC02-76CH00016 with the U.S. Department of Energy. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. This work was also supported in part under the National Science Foundation Grant No. DMS-9626567 and by the Schlumberger Technology Corporation.

where $F \in H_1$ and $G \in H_2$ are given and $X \in H_1$ and $Y \in H_2$ are the unknowns. Here $\mathbf{A} : H_1 \mapsto H_1$ is assumed to be a linear, nonsymmetric operator. $\mathbf{A}^T : H_1 \mapsto H_1$ is the adjoint of $\mathbf{A}$ with respect to the $(\cdot,\cdot)$–inner product. In addition, the linear map $\mathbf{B}$ takes H_1 into H_2 and its adjoint, $\mathbf{B}^T$, takes H_2 into H_1.

In general, (1.1) may not be solvable unless additional conditions on the operators $\mathbf{A}$ and $\mathbf{B}$ and the spaces H_1 and H_2 are imposed. Throughout this paper we assume that $\mathbf{A}$ has a positive definite symmetric part. Under this assumption, (1.1) is solvable if and only if the reduced problem

$$\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T Y = \mathbf{B}\mathbf{A}^{-1}F - G \tag{1.2}$$

is solvable. In the case of a symmetric and positive definite operator $\mathbf{A}$, the Ladyzhenskaya–Babuška–Brezzi (LBB) condition (cf. [5]) is a necessary and sufficient condition for solvability of this problem. As we shall see, the solvability of (1.1) in the nonsymmetric case is guaranteed provided that the LBB condition holds for the symmetric part of $\mathbf{A}$.

The papers [7], [15] propose solving $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ by a preconditioned iteration. One common problem with these approaches is that the evaluation of the action of the operator $\mathbf{A}^{-1}$ is required in each step of the iteration. For many applications, this operation is expensive and is also implemented as an iteration. The Uzawa method [1] is a particular implementation of a linear iterative method for solving (1.2). It is an exact algorithm in the sense that the action of $\mathbf{A}^{-1}$ is required for the implementation. An alternative method which solves (1.1) by preconditioned iteration was proposed in [8]. Their preconditioner also requires the evaluation of $\mathbf{A}^{-1}$ at each step of the iteration.

The Uzawa type methods studied here replace the exact inverse of $\mathbf{A}$ by an "approximate" evaluation of $\mathbf{A}^{-1}$ or a preconditioner for its symmetric part. Such algorithms are defined in Sections 3 and 4. In this paper we distinguish two types of algorithms: (i) a linear one-step method, where the action of the inverse is replaced by a linear preconditioner such as one sweep of a multigrid procedure; (ii) a multistep method, where a sufficiently accurate approximation to $\mathbf{A}^{-1}$ is provided by some iterative method, e.g., preconditioned GMRES [16] or preconditioned Lanczos [12].

The Uzawa type algorithms applied to nonsymmetric problems are of interest because they are simple, efficient, and have minimal computer memory requirements. They can be applied to the solution of difficult practical problems such as the Navier–Stokes equation. In addition, an exact Uzawa algorithm implemented as a double iteration can be easily modified to be an algorithm of the type studied here.

The paper is organized as follows. In Section 2 we establish sufficient conditions for solvability of the abstract saddle point problem and analyze an exact Uzawa algorithm for solving it. In Section 3 we define and analyze a linear one-step Uzawa type algorithm. Next, a multistep inexact method is defined and analyzed in Section 4. Section 5 provides applications of the algorithms from Section 3 and Section 4 to the solution of indefinite systems of linear equations arising in finite element approximations of the steady-state Navier-Stokes equations. Finally, the results of numerical computations involving the algorithms are given in Section 6.

2. Analysis of the exact method

In this section we establish sufficient conditions for solvability of (1.2) and analyze the exact Uzawa algorithm for the computation of its solution. The analysis of this method and, in particular, the result of Theorem 2.2 below is important for the analysis of the algorithms defined in the subsequent sections.

The symmetric part $\mathbf{A}_s$ of the operator $\mathbf{A}$ is defined by

$$\mathbf{A}_s = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T). \tag{2.1}$$

In the remainder of this paper a subscript s will be used to denote the symmetric part of various operators, defined as in (2.1). We assume that $\mathbf{A}_s$ is positive definite and satisfies

$$(\mathbf{A}X, Y) \leq \alpha(\mathbf{A}_s X, X)^{1/2}(\mathbf{A}_s Y, Y)^{1/2}, \quad \text{for all} \quad X, Y \in H_1, \tag{2.2}$$

for some number α. Clearly, $\alpha \geq 1$. Moreover, since $\mathbf{A}_s$ is positive definite, such an α always exists. In many applications involving the numerical solution of partial differential equations, the constant α can be chosen independently of the mesh parameter.

In addition, the Ladyzhenskaya–Babuška–Brezzi condition is assumed to hold for the pair of spaces H_1 and H_2, i.e.

$$\sup_{U \in H_1} \frac{(V, \mathbf{B}U)^2}{(\mathbf{A}_s U, U)} \geq c_0 \|V\|^2, \quad \text{for all} \quad V \in H_2\,, \tag{2.3}$$

for some positive number c_0. Here $\|\cdot\|$ denotes the norm in the space H_2 (or H_1) corresponding to the inner product $(\cdot,\cdot)$.

As is well known, the condition (2.3) is sufficient to guarantee solvability of (1.1) when $\mathbf{A}$ is symmetric. We will see that it also suffices in the case of nonsymmetric $\mathbf{A}$. To this end, we prove the following lemma.

Lemma 2.1. *Suppose that* $\mathbf{A}$ *is an invertible linear operator with positive definite symmetric part* $\mathbf{A}_s$ *that satisfies* (2.2). *Then* $(\mathbf{A}^{-1})_s$ *is positive definite and satisfies*

$$((\mathbf{A}^{-1})_s W, W) \leq ((\mathbf{A}_s)^{-1} W, W) \leq \alpha^2((\mathbf{A}^{-1})_s W, W), \quad \textit{for all} \quad W \in H_1. \tag{2.4}$$

Proof. Clearly,

$$\begin{aligned} ((\mathbf{A}_s)^{-1}W, W) &= \sup_{U \in H_1} \frac{(W, U)^2}{(\mathbf{A}_s U, U)} = \sup_{U \in H_1} \frac{((\mathbf{A}^{-1})^T W, \mathbf{A}U)^2}{(\mathbf{A}_s U, U)} \\ &\leq \alpha^2 \sup_{U \in H_1} \frac{\|(\mathbf{A}^{-1})^T W\|^2_{\mathbf{A}_s} \|U\|^2_{\mathbf{A}_s}}{\|U\|^2_{\mathbf{A}_s}} = \alpha^2 \|(\mathbf{A}^{-1})^T W\|^2_{\mathbf{A}_s} \\ &= \alpha^2((\mathbf{A}^{-1})_s W, W). \end{aligned} \tag{2.5}$$

Here $\|\cdot\|^2_{\mathbf{A}_s} = (\mathbf{A}_s \cdot, \cdot)$. In the above inequalities we have used the Schwarz inequality, (2.2), and the fact that

$$(\mathbf{A}_s U, U) = (\mathbf{A}U, U), \quad \text{for all} \quad U \in H_1. \tag{2.6}$$

On the other hand,

$$\begin{aligned} ((\mathbf{A}^{-1})_s U, U) = (\mathbf{A}^{-1}U, U) &= (\mathbf{A}_s^{1/2}\mathbf{A}^{-1}U, (\mathbf{A}_s)^{-1/2}U) \\ &\leq \|\mathbf{A}^{-1}U\|_{\mathbf{A}_s} \|U\|_{(\mathbf{A}_s)^{-1}} = (\mathbf{A}^{-1}U, U)^{1/2} \|U\|_{(\mathbf{A}_s)^{-1}}. \end{aligned}$$

Therefore,

$$((\mathbf{A}^{-1})_s U, U) \le ((\mathbf{A}_s)^{-1} U, U). \tag{2.7}$$

This completes the proof of the lemma. □

It is now clear that Lemma 2.1 and (2.3) guarantee solvability of (1.2). Indeed,

$$\begin{aligned}(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T V, V) &= ((\mathbf{A}^{-1})_s \mathbf{B}^T V, \mathbf{B}^T V)\\ &\ge \alpha^{-2}((\mathbf{A}_s)^{-1}\mathbf{B}^T V, \mathbf{B}^T V) \ge \alpha^{-2} c_0 \|V\|^2.\end{aligned}$$

Thus, we have proved the following theorem.

Theorem 2.1. *Suppose that the linear operator* $\mathbf{A}$ *is invertible and that* (2.3) *holds. Then the reduced problem* (1.2), *or equivalently* (1.1), *is solvable.*

Next, we turn to the analysis of the exact Uzawa algorithm applied to the solution of (1.2). The preconditioned variant of the exact Uzawa algorithm (cf. [1, 4]) is defined as follows.

Algorithm 2.1 (Preconditioned exact Uzawa). *For* $X_0 \in H_1$ *and* $Y_0 \in H_2$ *given, the sequence* $\{(X_i, Y_i)\}$ *is defined, for* $i = 0, 1, 2, \ldots,$ *by*

$$\begin{aligned}X_{i+1} &= X_i + \mathbf{A}^{-1}\left(F - (\mathbf{A}X_i + \mathbf{B}^T Y_i)\right),\\ Y_{i+1} &= Y_i + \tau \mathbf{Q}_B^{-1}(\mathbf{B}X_{i+1} - G).\end{aligned}$$

Here the preconditioner $\mathbf{Q}_B : H_2 \mapsto H_2$ is a symmetric positive definite linear operator satisfying

$$\gamma(\mathbf{Q}_B W, W) \le (\mathbf{B}(\mathbf{A}_s)^{-1}\mathbf{B}^T W, W) \le (\mathbf{Q}_B W, W), \quad \text{for all} \quad W \in H_2, \tag{2.8}$$

for some γ in the interval $(0, 1]$, and τ is a positive parameter. Notice that this condition implies appropriate scaling of $\mathbf{Q}_B$. In many applications effective preconditioners that satisfy (2.8) with γ bounded away from zero are known.

Let

$$E_i^X = X - X_i \tag{2.9a}$$

and

$$E_i^Y = Y - Y_i \tag{2.9b}$$

be the iteration errors generated by the above method. Note that

$$E_{i+1}^Y = (\mathbf{I} - \tau \mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T) E_i^Y.$$

Therefore, the convergence of Algorithm 2.1 is governed by the properties of the operator $\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ summarized in the following theorem.

Theorem 2.2. *Suppose that* $\mathbf{A}$ *is invertible with positive definite symmetric part* $\mathbf{A}_s$ *which satisfies* (2.2). *Suppose also that* (2.3) *holds. In addition, let* $\mathbf{Q}_B$ *be a symmetric and positive definite operator satisfying* (2.8). *If* τ *is a positive parameter with* $\tau \le \frac{\gamma}{\alpha^2}$, *then*

$$\|(\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)U\|_{\mathbf{Q}_B}^2 \le \left(1 - \frac{\gamma}{\alpha^2}\tau\right)\|U\|_{\mathbf{Q}_B}^2, \quad \textit{for all} \quad U \in H_2. \tag{2.10}$$

Remark 2.1. If $\mathbf{A} = \mathbf{A}^T$, then τ may be taken equal to one and (2.8) implies (cf. [4]) that

$$\|(\mathbf{I} - \mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)U\|_{\mathbf{Q}_B}^2 \leq (1-\gamma)^2\|U\|_{\mathbf{Q}_B}^2.$$

Hence, Theorem 2.2 is not optimal in the limit when $\alpha \to 1$.

Proof of Theorem 2.2. The proof is based on Lemma 2.1. Let $\mathcal{L} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$. Then, by (2.8) and Lemma 2.1,

$$\begin{aligned}\gamma\|V\|_{\mathbf{Q}_B}^2 &\leq ((\mathbf{A}_s)^{-1}\mathbf{B}^TV, \mathbf{B}^TV)\\ &\leq \alpha^2(\mathcal{L}V, V).\end{aligned} \tag{2.11}$$

In addition,using (2.4),

$$\begin{aligned}(\mathbf{A}^{-1}v, w) &= ((\mathbf{A}_s)^{1/2}\mathbf{A}^{-1}v, (\mathbf{A}_s)^{-1/2}w)\\ &\leq (\mathbf{A}^{-1}v, v)^{1/2}((\mathbf{A}_s)^{-1}w, w)^{1/2}\\ &\leq ((\mathbf{A}_s)^{-1}v, v)^{1/2}((\mathbf{A}_s)^{-1}w, w)^{1/2}.\end{aligned} \tag{2.12}$$

Taking $v = \mathbf{B}^TV$ and $w = \mathbf{B}^TW$ above gives

$$(\mathcal{L}V, W) \leq \|V\|_{\mathbf{Q}_B}\|W\|_{\mathbf{Q}_B}. \tag{2.13}$$

Next,

$$\|(\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathcal{L})V\|_{\mathbf{Q}_B}^2 = \|V\|_{\mathbf{Q}_B}^2 - 2\tau(\mathcal{L}V, V) + \tau^2(\mathcal{L}V, \mathbf{Q}_B^{-1}\mathcal{L}V). \tag{2.14}$$

By (2.12) and (2.11), the last term in the right hand side of (2.14) is estimated by

$$\begin{aligned}(\mathcal{L}V, \mathbf{Q}_B^{-1}\mathcal{L}V) &\leq \|V\|_{\mathbf{Q}_B}\|\mathbf{Q}_B^{-1}\mathcal{L}V\|_{\mathbf{Q}_B}\\ &= \|V\|_{\mathbf{Q}_B}(\mathcal{L}V, \mathbf{Q}_B^{-1}\mathcal{L}V)^{1/2}.\end{aligned} \tag{2.15}$$

Using (2.11) and (2.15) in (2.14) yields

$$\|(\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathcal{L})V\|_{\mathbf{Q}_B}^2 \leq \left(1 - \frac{2\tau\gamma}{\alpha^2} + \tau^2\right)\|V\|_{\mathbf{Q}_B}^2.$$

This concludes the proof of the theorem. □

3. Analysis of the linear one-step method

In this section we define and analyze a linear one-step Uzawa type algorithm applied to (1.1). This section contains the main result of the paper. We show that, under the minimal assumptions needed to guarantee solvability (cf. Section 2), appropriately scaled linear preconditioners (cf. (2.8) and (3.1) below) lead to an efficient and simple method for solving (1.1).

The exact inverse of $\mathbf{A}$ is replaced by a preconditioner for the symmetric part of $\mathbf{A}$. Let $\mathbf{A}_0 : H_1 \mapsto H_1$ be a linear, symmetric, positive definite operator that satisfies

$$(\mathbf{A}_0V, V) \leq (\mathbf{A}_sV, V) \leq \beta(\mathbf{A}_0V, V), \quad \text{for all} \quad V \in H_1, \tag{3.1}$$

for some $\beta \geq 1$.

Remark 3.1. The inequalities (2.8) and (3.1) respectively imply scaling of $\mathbf{Q}_B$ and $\mathbf{A}_0$. In practice, the proper scaling these operators can be achieved using even crude estimates for the largest eigenvalues of $\tilde{\mathbf{A}}_0^{-1}\mathbf{A}_s$ and $\tilde{\mathbf{Q}}_B^{-1}\mathbf{B}(\mathbf{A}_s)^{-1}\mathbf{B}^T$, where $\tilde{\mathbf{A}}_0$ and $\tilde{\mathbf{Q}}_B$ are unscaled preconditioners. Usually, a few iterations of the power method are enough for obtaining such estimates.

The linear Uzawa type algorithm is defined as follows.

Algorithm 3.1 (Linear one-step method). *For $X_0 \in H_1$ and $Y_0 \in H_2$ given, the sequence $\{(X_i, Y_i)\}$ is defined, for $i = 0, 1, 2, \dots$, by*

$$\begin{aligned} X_{i+1} &= X_i + \delta \mathbf{A}_0^{-1} \left(F - (\mathbf{A} X_i + \mathbf{B}^T Y_i) \right), \\ Y_{i+1} &= Y_i + \tau \mathbf{Q}_B^{-1} (\mathbf{B} X_{i+1} - G). \end{aligned}$$

Here δ and τ are positive parameters.

We will assume that $\delta < 1/\beta$. It then follows from (3.1) that $\mathbf{A}_0 - \delta \mathbf{A}_s$ is positive definite. The following theorem is the main result of this paper.

Theorem 3.1. *Suppose that $\mathbf{A}$ has a positive definite symmetric part, $\mathbf{A}_s$, satisfying (2.2). Suppose also that $\mathbf{Q}_B$ and $\mathbf{A}_0$ are symmetric positive definite operators satisfying (2.8) and (3.1). Then Algorithm 3.1 converges if $0 < \delta \le (3\alpha^2\beta^2)^{-1}$ and $0 < \tau \le (4\beta)^{-1}$. Moreover, if (X, Y) is the solution of (1.1) and (X_i, Y_i) is the approximation defined by Algorithm 3.1, then the iteration errors E_i^X and E_i^Y defined in (2.9) satisfy*

(3.2)

$$\{\delta^{-1} \|E_i^X\|^2_{\mathbf{A}_0 - \delta \mathbf{A}_s} + \tau^{-1} \|E_i^Y\|^2_{\mathbf{Q}_B}\}^{1/2} \le \bar{\rho}^i \left\{\delta^{-1} \|E_0^X\|^2_{\mathbf{A}_0 - \delta \mathbf{A}_s} + \tau^{-1} \|E_0^Y\|^2_{\mathbf{Q}_B}\right\}^{1/2}$$

for any $i \ge 1$. Here

$$\bar{\rho} = \frac{\delta/2 - \delta\tau\gamma + \sqrt{(\delta/2 - \delta\tau\gamma)^2 + 4(1 - \delta/2)}}{2}.$$

Remark 3.2. Convergence of Algorithm 3.1 follows from (3.2). Indeed, a simple algebraic manipulation using the fact that $\tau\gamma \le 1/4$ gives

$$\bar{\rho} \equiv \frac{\delta/2 - \delta\tau\gamma + \sqrt{(\delta/2 - \delta\tau\gamma)^2 + 4(1 - \delta/2)}}{2} < 1 - \frac{\delta\tau\gamma}{2}.$$

The quantity on the right hand side is clearly less than one.

Remark 3.3. The use of a symmetric preconditioner in Algorithm 3.1 results in a fundamental change in the convergence properties. For example, if we take $\tau = \frac{\gamma}{\alpha^2}$ in Algorithm 2.1, then Theorem 2.2 gives the reduction of $1 - \gamma^2/\alpha^4$. In contrast, if we set $\delta = (3\alpha^2\beta^2)^{-1}$ and $\tau = (4\beta)^{-1}$, then Theorem 3.1 gives a convergence rate which is bounded by

$$1 - \frac{\gamma}{24\alpha^2\beta^3}.$$

This behaves significantly better in applications involving large α.

In order to analyze Algorithm 3.1 we reformulate it in terms of the iteration errors defined in (2.9). It is easy to see that E_i^X and E_i^Y satisfy the following equations:

$$\begin{aligned} E_{i+1}^X &= E_i^X - \delta \mathbf{A}_0^{-1} \left(\mathbf{A} E_i^X + \mathbf{B}^T E_i^Y) \right), \\ E_{i+1}^Y &= \left(\mathbf{I} - \delta\tau \mathbf{Q}_B^{-1} \mathbf{B} \mathbf{A}_0^{-1} \mathbf{B}^T \right) E_i^Y + \tau \mathbf{Q}_B^{-1} \mathbf{B} (\mathbf{I} - \delta \mathbf{A}_0^{-1} \mathbf{A}) E_i^X. \end{aligned}$$

For convenience, these equations can be written in matrix form as

$$(3.3) \qquad \begin{pmatrix} E_{i+1}^X \\ E_{i+1}^Y \end{pmatrix} = \begin{pmatrix} (\mathbf{I} - \delta \mathbf{A}_0^{-1} \mathbf{A}) & -\delta \mathbf{A}_0^{-1} \mathbf{B}^T \\ \tau \mathbf{Q}_B^{-1} \mathbf{B} (\mathbf{I} - \delta \mathbf{A}_0^{-1} \mathbf{A}) & (\mathbf{I} - \tau\delta \mathbf{Q}_B^{-1} \mathbf{B} \mathbf{A}_0^{-1} \mathbf{B}^T) \end{pmatrix} \begin{pmatrix} E_i^X \\ E_i^Y \end{pmatrix}.$$

Straightforward manipulations of (3.3) give

$$\mathcal{N} E_{i+1} = \mathcal{M} E_i, \tag{3.4}$$

where

$$E_i = \begin{pmatrix} E_i^X \\ E_i^Y \end{pmatrix},$$

$$\mathcal{N} = \begin{pmatrix} \delta^{-1}(\mathbf{A}_0 - \delta \mathbf{A}^T) & 0 \\ 0 & \tau^{-1}\mathbf{Q}_B \end{pmatrix},$$

and

$$\mathcal{M} = \begin{pmatrix} \delta^{-1}(\mathbf{A}_0 - \delta \mathbf{A}^T)\mathbf{A}_0^{-1}(\mathbf{A}_0 - \delta \mathbf{A}) & -(\mathbf{A}_0 - \delta \mathbf{A}^T)\mathbf{A}_0^{-1}\mathbf{B}^T \\ \mathbf{B}\mathbf{A}_0^{-1}(\mathbf{A}_0 - \delta \mathbf{A}) & (\tau^{-1}\mathbf{Q}_B - \delta \mathbf{B}\mathbf{A}_0^{-1}\mathbf{B}^T) \end{pmatrix}.$$

It is clear that we can study the convergence of Algorithm 3.1 by investigating the properties of the linear operators $\mathcal{M}$ and $\mathcal{N}$. We shall reduce this problem to estimation of the spectral radius of related symmetric operators.

Let $\mathcal{M}_1$ be the symmetric matrix defined by

$$\mathcal{M}_1 = \mathcal{J}\mathcal{M},$$

where

$$\mathcal{J} = \begin{pmatrix} -\mathbf{I} & 0 \\ 0 & \mathbf{I} \end{pmatrix}.$$

Our next lemma reduces the proof of the theorem to the estimation of the eigenvalues of the generalized eigenvalue problem

$$\lambda \mathcal{N}_s \psi = \mathcal{M}_1 \psi. \tag{3.5}$$

Since δ is less than $1/\beta$, $\mathcal{N}_s$ is positive definite and the above problem is well defined. Because $\mathcal{N}_s$ and $\mathcal{M}_1$ are symmetric, the eigenvalues λ are real.

Lemma 3.1. *The iteration error E_i satisfies*

$$(\mathcal{N}_s E_{i+1}, E_{i+1})^{1/2} \le \bar{\rho} (\mathcal{N}_s E_i, E_i)^{1/2},$$

where $\bar{\rho} = \max_i |\lambda_i|$, with $\{\lambda_i\}$ the eigenvalues of (3.5).

Proof. Let $\{(\lambda_i, \psi_i)\}$ be the eigenpairs for (3.5). Since $\mathcal{N}_s$ is positive definite, $\{\psi_i\}$ spans the space $H_1 \times H_2$. We may choose the eigenvectors so that

$$(\mathcal{N}_s \psi_i, \psi_j) = \delta_{ij},$$

where δ_{ij} denotes the Kronecker delta. Now any vectors $\mathbf{v}$ and $\mathbf{w}$ in $H_1 \times H_2$ can be represented as $\mathbf{v} = \sum_i v_i \psi_i$ and $\mathbf{w} = \sum_i w_i \psi_i$. Thus,

$$\begin{aligned} (\mathcal{M}_1 \mathbf{v}, \mathbf{w}) &= \sum_{ij} v_i w_j (\mathcal{M}_1 \psi_i, \psi_j) = \sum_i v_i w_i \lambda_i \\ &\le \bar{\rho} \left(\sum_i v_i^2 \right)^{1/2} \left(\sum_i w_i^2 \right)^{1/2} \\ &= \bar{\rho} \|\mathbf{v}\|_{\mathcal{N}_s} \|\mathbf{w}\|_{\mathcal{N}_s}. \end{aligned} \tag{3.6}$$

Since $\mathcal{J}^2$ is the identity operator we have that $\mathcal{M} = \mathcal{J}\mathcal{M}_1$. Therefore, using (3.4) we see that

$$\begin{aligned}(\mathcal{N}_s E_{i+1}, E_{i+1}) &= (\mathcal{M} E_i, E_{i+1}) = (\mathcal{M}_1 E_i, \mathcal{J} E_{i+1}) \\ &\leq \bar{\rho} \| E_i \|_{\mathcal{N}_s} \| \mathcal{J} E_{i+1} \|_{\mathcal{N}_s} = \bar{\rho} \| E_i \|_{\mathcal{N}_s} \| E_{i+1} \|_{\mathcal{N}_s}.\end{aligned}$$

The lemma immediately follows. □

Our proof of Theorem 3.1 will require another lemma. We need to provide some control on the convergence of the related linear iteration

$$U_{i+1} = U_i + \delta \mathbf{A}_0^{-1}(W - \mathbf{A}U_i) \tag{3.7}$$

to the solution U of

$$\mathbf{A}U = W.$$

Lemma 3.2. *Let* $\mathbf{A}_0$ *satisfy* (3.1) *and* δ *be a positive number with* $\delta < 1/\beta$. *Then*

$$\|(\mathbf{I} - \delta \mathbf{A}_0^{-1}\mathbf{A})V\|_{\mathbf{A}_0}^2 \leq \bar{\delta}\left((\mathbf{A}_0 - \delta \mathbf{A}_s)V, V\right), \quad \textit{for all} \quad V \in H_1, \tag{3.8}$$

where

$$\bar{\delta} = 1 - \delta + \frac{\alpha^2\beta^2\delta^2}{1 - \delta\beta}.$$

Remark 3.4. If, in addition

$$\delta < \frac{1}{\alpha^2\beta^2 + \beta}, \tag{3.9}$$

so that

$$\frac{\alpha^2\beta^2\delta}{1 - \delta\beta} < 1,$$

then $\bar{\delta}$ is less than one.

Proof of Lemma 3.2. By (3.1),

$$(1 - \delta\beta)(\mathbf{A}_0 V, V) \leq ((\mathbf{A}_0 - \delta \mathbf{A}_s)V, V), \quad \text{for all} \quad V \in H_1.$$

Hence, by (2.2) and (3.1),

$$\begin{aligned}(\mathbf{A}V, W) &\leq \alpha(\mathbf{A}_s V, V)^{1/2}(\mathbf{A}_s W, W)^{1/2} \\ &\leq \frac{\alpha\beta}{(1 - \delta\beta)^{1/2}}(\mathbf{A}_0 V, V)^{1/2}((\mathbf{A}_0 - \delta \mathbf{A}_s)W, W)^{1/2}.\end{aligned} \tag{3.10}$$

On the other hand,

$$\begin{aligned}\|(\mathbf{I} - \delta \mathbf{A}_0^{-1}\mathbf{A})V\|_{\mathbf{A}_0}^2 &= \|V\|_{\mathbf{A}_0}^2 - 2\delta(\mathbf{A}V, V) + \delta^2(\mathbf{A}_0^{-1}\mathbf{A}V, \mathbf{A}V) \\ &= ((\mathbf{A}_0 - \delta \mathbf{A}_s)V, V) - \delta(\mathbf{A}V, V) + \delta^2(\mathbf{A}_0^{-1}\mathbf{A}V, \mathbf{A}V).\end{aligned} \tag{3.11}$$

In view of (3.1), we have

$$(\mathbf{A}V, V) \geq (\mathbf{A}_0 V, V) \geq ((\mathbf{A}_0 - \delta \mathbf{A}_s)V, V). \tag{3.12}$$

Also, (3.10) implies that

$$(\mathbf{A}_0^{-1}\mathbf{A}V, \mathbf{A}V) \leq \frac{\alpha\beta}{(1 - \delta\beta)^{1/2}}(\mathbf{A}_0^{-1}\mathbf{A}V, \mathbf{A}V)^{1/2}((\mathbf{A}_0 - \delta \mathbf{A}_s)V, V)^{1/2}.$$

Thus,

$$(\mathbf{A}_0^{-1}\mathbf{A}V,\mathbf{A}V) \le \frac{\alpha^2\beta^2}{1-\delta\beta}((\mathbf{A}_0-\delta\mathbf{A}_s)V,V). \tag{3.13}$$

Using (3.12) and (3.13) in (3.11) yields (3.8). □

Proof of Theorem 3.1. Let δ and τ satisfy the hypotheses of the theorem. Because of Lemma 3.1 it suffices to bound the eigenvalues of (3.5). We begin with the negative eigenvalues. Let (χ,ξ) be an eigenvector in $H_1 \times H_2$ with eigenvalue $\lambda < 0$. Then multiplying the first equation of (3.5) by $\mathbf{A}_0(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}$ gives

$$\lambda\delta^{-1}\mathbf{A}_0(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}(\mathbf{A}_0-\delta\mathbf{A}_s)\chi = -\delta^{-1}(\mathbf{A}_0-\delta\mathbf{A})\chi+\mathbf{B}^T\xi. \tag{3.14}$$

The second equation of (3.5) is

$$\lambda\tau^{-1}\mathbf{Q}_B\xi = \mathbf{B}\mathbf{A}_0^{-1}(\mathbf{A}_0-\delta\mathbf{A})\chi+(\tau^{-1}\mathbf{Q}_B-\delta\mathbf{B}\mathbf{A}_0^{-1}\mathbf{B}^T)\xi. \tag{3.15}$$

Applying $\delta\mathbf{B}\mathbf{A}_0^{-1}$ to (3.14) and adding it to (3.15), we obtain

$$(1-\lambda)\tau^{-1}\mathbf{Q}_B\xi = \lambda\mathbf{B}(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}(\mathbf{A}_0-\delta\mathbf{A}_s)\chi. \tag{3.16}$$

Eliminating ξ between (3.14) and (3.16), we find that

$$\begin{aligned}-\frac{1}{\lambda}(\mathbf{A}_0-\delta\mathbf{A})\chi+\frac{\delta\tau}{1-\lambda}\mathbf{B}^T\mathbf{Q}_B{}^{-1}\mathbf{B}(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}(\mathbf{A}_0-\delta\mathbf{A}_s)\chi\\ = \mathbf{A}_0(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}(\mathbf{A}_0-\delta\mathbf{A}_s)\chi.\end{aligned}$$

Taking the inner product with $(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}(\mathbf{A}_0-\delta\mathbf{A}_s)\chi$ yields

$$\begin{aligned}-\frac{1}{\lambda}((\mathbf{A}_0-\delta\mathbf{A}_s)\chi,\chi)+\frac{\delta\tau}{1-\lambda}\|\mathbf{B}(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}(\mathbf{A}_0-\delta\mathbf{A}_s)\chi\|^2_{\mathbf{Q}_B^{-1}}\\ = \|(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}(\mathbf{A}_0-\delta\mathbf{A}_s)\chi\|^2_{\mathbf{A}_0}.\end{aligned} \tag{3.17}$$

For convenience, the last equation can be abbreviated as

$$T_1+T_2=T_3.$$

In order to bound T_2 we note that for any $\phi \in H_1$,

$$\begin{aligned}(\mathbf{Q}_B^{-1}\mathbf{B}\phi,\mathbf{B}\phi) &= \sup_{\zeta\in H_2}\frac{(\phi,\mathbf{B}^T\zeta)^2}{(\mathbf{Q}_B\zeta,\zeta)} = \sup_{\zeta\in H_2}\frac{((\mathbf{A}_s)^{1/2}\phi,(\mathbf{A}_s)^{-1/2}\mathbf{B}^T\zeta)^2}{(\mathbf{Q}_B\zeta,\zeta)}\\ &\le \sup_{\zeta\in H_2}\frac{(\mathbf{A}_s\phi,\phi)(\mathbf{B}(\mathbf{A}_s)^{-1}\mathbf{B}^T\zeta,\zeta)}{(\mathbf{Q}_B\zeta,\zeta)} \le (\mathbf{A}_s\phi,\phi).\end{aligned} \tag{3.18}$$

For the last inequality above we used (2.8). Therefore,

$$T_2 \le \frac{\delta\tau}{1-\lambda}\|(\mathbf{A}_0-\delta\mathbf{A}^T)^{-1}(\mathbf{A}_0-\delta\mathbf{A}_s)\chi\|^2_{\mathbf{A}_s} \le \frac{\beta\delta\tau}{1-\lambda}T_3.$$

Using this in (3.17) gives

$$\left(1-\frac{\delta\tau\beta}{1-\lambda}\right)T_3 \le -\frac{1}{\lambda}((\mathbf{A}_0-\delta\mathbf{A}_s)\chi,\chi). \tag{3.19}$$

By Lemma 3.2, for any $\phi \in H_1$, we have

$$((\mathbf{A}_0-\delta\mathbf{A}^T)\mathbf{A}_0^{-1}(\mathbf{A}_0-\delta\mathbf{A})\phi,\phi) \le \bar{\delta}((\mathbf{A}_0-\delta\mathbf{A}_s)\phi,\phi), \tag{3.20}$$

which is equivalent to

$$((\mathbf{A}_0 - \delta\mathbf{A}_s)^{-1}\phi, \phi) \le \bar{\delta}((\mathbf{A}_0 - \delta\mathbf{A})^{-1}\mathbf{A}_0(\mathbf{A}_0 - \delta\mathbf{A}^T)^{-1}\phi, \phi) \tag{3.21}$$

for any $\phi \in H_1$. Taking $\phi = (\mathbf{A}_0 - \delta\mathbf{A}_s)\chi$ in (3.21), we obtain

$$T_3 \ge \frac{1}{\bar{\delta}}((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi).$$

Combining this with (3.19) and using the fact that $\lambda < 0$ gives

$$\begin{aligned} -\frac{1}{\lambda}((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi) &\ge \frac{1}{\bar{\delta}}\left(1 - \frac{\delta\tau\beta}{1-\lambda}\right)((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi) \\ &\ge \frac{1-\delta\tau\beta}{\bar{\delta}}((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi). \end{aligned} \tag{3.22}$$

Now if χ were equal to zero then (3.14) would imply that $\mathbf{B}^T\xi = 0$, but (3.15), in turn, would imply that $\xi = 0$. Hence, since (χ, ξ) is an eigenvector, $\chi \neq 0$ and therefore $((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi) \neq 0$. Thus, from (3.22),

$$-\lambda \le \frac{\bar{\delta}}{1 - \delta\tau\beta}. \tag{3.23}$$

Applying straightforward manipulations, we get

$$\bar{\delta} = 1 - \delta + \frac{\alpha^2\beta^2\delta^2}{1-\delta\beta} \le 1 - \delta\left(1 - \frac{1/3}{1-1/3}\right) = 1 - \frac{\delta}{2} \tag{3.24}$$

and

$$\frac{1}{1-\delta\tau\beta} \le \frac{1}{1-\delta/4}. \tag{3.25}$$

Using (3.24) and (3.25) in (3.23) gives

$$-\lambda \le \frac{1-\delta/2}{1-\delta/4} \le 1 - \frac{\delta}{4}, \tag{3.26}$$

which provides a bound for the negative part of the spectrum.

Next we obtain a bound for the positive eigenvalues of (3.5). To this end we factor $\mathcal{M}_1$ as

$$\mathcal{M}_1 = \mathcal{D}^T\mathcal{M}_2\mathcal{D},$$

where

$$\mathcal{D} = \begin{pmatrix} \theta^{-1/2}(\mathbf{A}_0)^{-1/2}(\mathbf{A}_0 - \delta\mathbf{A}) & 0 \\ 0 & \mathbf{I} \end{pmatrix},$$

$$\mathcal{M}_2 = \begin{pmatrix} -\delta^{-1}\theta\mathbf{I} & \theta^{1/2}(\mathbf{A}_0)^{-1/2}\mathbf{B}^T \\ \theta^{1/2}\mathbf{B}(\mathbf{A}_0)^{-1/2} & \tau^{-1}\mathbf{Q}_B - \delta\mathbf{B}\mathbf{A}_0^{-1}\mathbf{B}^T \end{pmatrix},$$

and $\theta = 1 - \delta/2$. The largest eigenvalue of (3.5) is given by

$$\Lambda = \sup_{\mathbf{w}\in H_1\times H_2} \frac{(\mathcal{M}_1\mathbf{w}, \mathbf{w})}{(\mathcal{N}_s\mathbf{w}, \mathbf{w})} = \sup_{\mathbf{w}\in H_1\times H_2} \frac{(\mathcal{M}_2\mathcal{D}\mathbf{w}, \mathcal{D}\mathbf{w})}{(\mathcal{N}_s\mathbf{w}, \mathbf{w})}.$$

In order to obtain an upper bound for Λ, we first note that by (3.20) and (3.24) it follows that

$$\theta^{-1}\|\mathbf{A}_0^{-1/2}(\mathbf{A}_0 - \delta\mathbf{A})\chi\|^2 \le ((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi), \tag{3.27}$$

and thus

$$\theta^{-1}\delta^{-1}\|\mathbf{A}_0^{-1/2}(\mathbf{A}_0-\delta\mathbf{A})\chi\|^2+\tau^{-1}\|\xi\|^2_{\mathbf{Q}_B}\le\left(\mathcal{N}_s\begin{pmatrix}\chi\\ \xi\end{pmatrix},\begin{pmatrix}\chi\\ \xi\end{pmatrix}\right).\tag{3.28}$$

Thus, it suffices to show that for any vector $(\phi,\zeta)\in H_1\times H_2$,

$$\begin{aligned}\left(\mathcal{M}_2\mathcal{D}\begin{pmatrix}\phi\\ \zeta\end{pmatrix},\mathcal{D}\begin{pmatrix}\phi\\ \zeta\end{pmatrix}\right)&\le\bar{\rho}\left[(\delta\theta)^{-1}\|\mathbf{A}_0{}^{-1/2}(\mathbf{A}_0-\delta\mathbf{A})\phi\|^2+\tau^{-1}\|\zeta\|^2_{\mathbf{Q}_B}\right]\\ &=\bar{\rho}\left(\begin{pmatrix}\delta^{-1}\mathbf{I} & 0\\ 0 & \tau^{-1}\mathbf{Q}_B\end{pmatrix}\mathcal{D}\begin{pmatrix}\phi\\ \zeta\end{pmatrix},\mathcal{D}\begin{pmatrix}\phi\\ \zeta\end{pmatrix}\right).\end{aligned}\tag{3.29}$$

Then $\bar{\rho}$ will be an upper bound for Λ. To this end let $\mathbf{L}=\mathbf{B}(\mathbf{A}_0)^{-1/2}$. Now $\mathcal{M}_2$ may be written as

$$\mathcal{M}_2=\begin{pmatrix}-\delta^{-1}\theta\mathbf{I} & \theta^{1/2}\mathbf{L}^T\\ \theta^{1/2}\mathbf{L} & \tau^{-1}\mathbf{Q}_B-\delta\mathbf{L}\mathbf{L}^T\end{pmatrix}.$$

The proof of (3.29) is now reduced to estimating the largest eigenvalue, λ, with eigenvector (χ,ξ) satisfying

$$-\theta\delta^{-1}\chi+\theta^{1/2}\mathbf{L}^T\xi=\lambda\delta^{-1}\chi\tag{3.30}$$

and

$$\theta^{1/2}\mathbf{L}\chi+(\tau^{-1}\mathbf{Q}_B-\delta\mathbf{L}\mathbf{L}^T)\xi=\lambda\tau^{-1}\mathbf{Q}_B\xi.\tag{3.31}$$

Solving for χ in (3.30), we get

$$\chi=\delta(\lambda+\theta)^{-1}\theta^{1/2}\mathbf{L}^T\xi.$$

Substituting this in (3.31) yields

$$(1-\lambda)(\lambda+\theta)\mathbf{Q}_B\xi=\delta\tau\lambda\mathbf{L}\mathbf{L}^T\xi.$$

Hence

$$(1-\lambda)(\lambda+\theta)(\mathbf{Q}_B\xi,\xi)=\delta\tau\lambda(\mathbf{L}^T\xi,\mathbf{L}^T\xi).\tag{3.32}$$

Now ξ cannot be zero, since if it were, then (3.30) would imply that either $\chi=0$ or $\lambda=-\theta\le0$. Hence, since λ is a positive eigenvalue, it follows that $\xi\ne0$. In addition, by (3.1) and (2.8),

$$\begin{aligned}(\mathbf{L}^T\xi,\mathbf{L}^T\xi)&=(\mathbf{A}_0^{-1}\mathbf{B}^T\xi,\mathbf{B}^T\xi)\ge((\mathbf{A}_s)^{-1}\mathbf{B}^T\xi,\mathbf{B}^T\xi)\\ &\ge\gamma(\mathbf{Q}_B\xi,\xi).\end{aligned}$$

Using this in (3.32) gives

$$(1-\lambda)(\lambda+\theta)\ge\delta\tau\lambda\gamma,$$

or equivalently

$$\lambda^2-\lambda(1-\theta-\delta\tau\gamma)-\theta\le0.$$

From here we obtain that

$$\begin{aligned}\lambda&\le\frac{1-\theta-\delta\tau\gamma+\sqrt{((1-\theta)-\delta\tau\gamma)^2+4\theta}}{2}\\ &=\frac{\delta/2-\delta\tau\gamma+\sqrt{(\delta/2-\delta\tau\gamma)^2+4(1-\delta/2)}}{2}.\end{aligned}\tag{3.33}$$

Finally, elementary inequalities imply that

$$1 - \frac{\delta}{4} \le \frac{\delta/2 - \delta\tau\gamma + \sqrt{(\delta/2 - \delta\tau\gamma)^2 + 4(1 - \delta/2)}}{2},$$

which concludes the proof of the theorem. □

4. Analysis of the multistep inexact algorithm

In this section we define and analyze an inexact Uzawa algorithm obtained by replacing $\mathbf{A}^{-1}$ with a sufficiently accurate approximation. Such an algorithm is essentially different from the linear one-step method developed in the previous section for two main reasons. First, achieving a certain accuracy of the approximation to $\mathbf{A}^{-1}$ typically requires more computational work than one evaluation of the action of a preconditioner. Second, depending on the manner in which the accurate approximate inverse is computed, the resulting inexact Uzawa algorithm may not be linear. In view of this, we shall approach the analysis of this method differently.

The approximate inverse may be described as a map $\Psi : H_1 \mapsto H_1$, not necessarily linear. In this section we shall assume that for any $\phi \in H_1$, $\Psi(\phi)$ is "close" to the solution ξ of

$$\mathbf{A}\xi = \phi. \tag{4.1}$$

More precisely, we assume that

$$\|\Psi(\phi) - \mathbf{A}^{-1}\phi\|_{\mathbf{A}_s} \le \epsilon \|\mathbf{A}^{-1}\phi\|_{\mathbf{A}_s}, \quad \text{for all} \quad \phi \in H_1, \tag{4.2}$$

for some positive ϵ with $\epsilon < 1$.

Notice that for any $\epsilon \in (0,1)$, (4.2) can be satisfied by taking sufficiently many steps of some iterative method for solving (4.1) which reduces the error in a norm equivalent to $\|\cdot\|_{\mathbf{A}_s}$. For example, for an appropriate choice of the iteration parameter, the linear iteration (3.7) converges to the solution of the linear system (4.1) (cf. Remark 3.4). Hence, an estimate of the type of (4.2) can be established for any $\epsilon < 1$, provided that sufficiently many iterations with (3.7) are performed.

Another example of Ψ results from a preconditioned generalized Lanczos procedure [12]. In this case the resulting Uzawa algorithm will be nonlinear. As an example, we consider the generalized minimal residual algorithm (GMRES). Specifically, let $\mathbf{A}_0$ satisfy (3.1) and consider the GMRES algorithm applied in the inner product $(\cdot,\cdot)_{\mathbf{A}_0} \equiv (\mathbf{A}_0\cdot,\cdot)$ to the preconditioned equation

$$\mathbf{A}_0^{-1}\mathbf{A}\xi = \mathbf{A}_0^{-1}\phi.$$

Using the initial iterate $\xi_0 = 0$, we set $\Psi(\phi) = \xi_n$, where ξ_n is the approximation obtained after n steps. The GMRES method computes the best approximation to ξ (in the norm $\|\mathbf{A}_0^{-1}\mathcal{A}\cdot\|_{\mathbf{A}_0}$) in the Krylov space K_n =span$\{(\mathbf{A}_0^{-1}\mathbf{A})^i\xi,\ i = 1,\dots,n\}$. Thus, it follows from Lemma 3.2 that

$$\|\mathbf{A}_0^{-1}\mathbf{A}(\xi_n - \xi)\|_{\mathbf{A}_0} \le \|(I - \delta\mathbf{A}_0^{-1}\mathbf{A})^n\mathbf{A}_0^{-1}\mathbf{A}\xi\|_{\mathbf{A}_0} \le \bar{\delta}^{n/2}\|\mathbf{A}_0^{-1}\mathbf{A}\xi\|_{\mathbf{A}_0}.$$

Applying Lemma 2.1 and (3.1) gives

$$\|\mathbf{A}_0^{-1}\mathbf{A}(\xi_n - \xi)\|_{\mathbf{A}_0} \ge \|\xi_n - \mathbf{A}^{-1}\phi\|_{\mathbf{A}_s}.$$

Similarly,

$$\|\mathbf{A}_0^{-1}\mathbf{A}\xi\|_{\mathbf{A}_0} \le \alpha\beta^{1/2}\|\mathbf{A}^{-1}\phi\|_{\mathbf{A}_s}.$$

Combining the above inequalities, we obtain

$$\|\xi_n - \mathbf{A}^{-1}\phi\|_{\mathbf{A}_s} \le \alpha\beta^{1/2}\bar{\delta}^{n/2}\|\mathbf{A}^{-1}\phi\|_{\mathbf{A}_s}.$$

This shows that, given ϵ, there exists n such that $\Psi(\phi) = \xi_n$ satisfies (4.2) provided that (3.9) holds.

The variant of the inexact Uzawa algorithm which we investigate in this section is defined as follows.

Algorithm 4.1 (Multistep inexact Uzawa). *For $X_0 \in H_1$ and $Y_0 \in H_2$ given, the sequence $\{(X_i, Y_i)\}$ is defined, for $i = 0, 1, 2, \ldots$, by*

$$\begin{aligned} X_{i+1} &= X_i + \Psi\left(F - (\mathbf{A}X_i + \mathbf{B}^T Y_i)\right), \\ Y_{i+1} &= Y_i + \tau\mathbf{Q}_B^{-1}(\mathbf{B}X_{i+1} - G). \end{aligned}$$

Algorithm 4.1 reduces to Algorithm 2.1 if $\Psi(\phi) = \mathbf{A}^{-1}\phi$ for all $\phi \in H_1$.

The main result of this section is a bound for the rate of convergence of the multistep algorithm in terms of the factors α, γ, and ϵ introduced in (2.2), (2.8), and (4.2) respectively. The theorem below gives a sufficient condition on ϵ for convergence of the algorithm.

Theorem 4.1. *Suppose that $\mathbf{A}$ has a positive definite symmetric part and satisfies (2.2), and $\mathbf{Q}_B$ is a symmetric positive definite operator satisfying (2.8). Assume that (4.2) holds and that the parameter τ is chosen so that*

$$0 < \tau \le \frac{\gamma}{\alpha^2}.$$

Set

$$\theta = \left(1 - \tau\frac{\gamma}{\alpha^2}\right)^{1/2}.$$

Then the multistep inexact Uzawa algorithm converges if

$$\epsilon < \frac{1-\theta}{1+2\tau-\theta}. \tag{4.3}$$

Moreover, if (X, Y) is the solution of (1.1) and (X_i, Y_i) is the approximation defined by Algorithm 4.1, then the iteration errors E_i^X and E_i^Y defined in (2.9) satisfy

$$\frac{\epsilon\tau}{1+\epsilon}\|E_{i+1}^X\|_{\mathbf{A}_s}^2 + \|E_{i+1}^Y\|_{\mathbf{Q}_B}^2 \le \rho^{2(i+1)}\left(\frac{\epsilon\tau}{1+\epsilon}\|E_0^X\|_{\mathbf{A}_s}^2 + \|E_0^Y\|_{\mathbf{Q}_B}^2\right) \tag{4.4}$$

and

$$\|E_{i+1}^X\|_{\mathbf{A}_s}^2 \le \tau^{-1}(1+\epsilon)(1+2\epsilon)\rho^{2i}\left(\frac{\epsilon\tau}{1+\epsilon}\|E_0^X\|_{\mathbf{A}_s}^2 + \|E_0^Y\|_{\mathbf{Q}_B}^2\right), \tag{4.5}$$

where

$$\rho = \frac{(1+\tau)\epsilon + \theta + \sqrt{((1+\tau)\epsilon+\theta)^2 + 4\epsilon(\tau-\theta)}}{2} < 1. \tag{4.6}$$

Proof. We start by deriving norm inequalities involving the errors E_i^X and E_i^Y. Similarly to the approach in the previous section, we can write

$$\begin{aligned} E_{i+1}^X &= E_i^X - \Psi\left(\mathbf{A}E_i^X + \mathbf{B}^T E_i^Y\right), \\ E_{i+1}^Y &= E_i^Y + \tau\mathbf{Q}_B^{-1}\mathbf{B}E_{i+1}^X. \end{aligned} \tag{4.7}$$

The first equation above can be rewritten

$$E^X_{i+1} = (\mathbf{A}^{-1} - \Psi)\left(\mathbf{A}E^X_i + \mathbf{B}^T E^Y_i\right) - \mathbf{A}^{-1}\mathbf{B}^T E^Y_i. \tag{4.8}$$

It follows from the triangle inequality, (4.2), (2.4), and (2.8) that

$$\begin{aligned} \|E^X_{i+1}\|_{\mathbf{A}_s} &\le \epsilon(\|E^X_i\|_{\mathbf{A}_s} + \|\mathbf{A}^{-1}\mathbf{B}^T E^Y_i\|_{\mathbf{A}_s}) + \|\mathbf{A}^{-1}\mathbf{B}^T E^Y_i\|_{\mathbf{A}_s} \\ &= \epsilon\|E^X_i\|_{\mathbf{A}_s} + (1+\epsilon)\|\mathbf{B}^T E^Y_i\|_{(\mathbf{A}^{-1})_s} \\ &\le \epsilon\|E^X_i\|_{\mathbf{A}_s} + (1+\epsilon)\|E^Y_i\|_{\mathbf{Q}_B}. \end{aligned} \tag{4.9}$$

Using (4.8) in the second equation of (4.7) gives

$$E^Y_{i+1} = (\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)E^Y_i + \tau\mathbf{Q}_B^{-1}\mathbf{B}(\mathbf{A}^{-1} - \Psi)(\mathbf{A}E^X_i + \mathbf{B}^T E^Y_i).$$

Applying the $\|\cdot\|_{\mathbf{Q}_B}$ norm to both sides of the above equation and using the triangle inequality yields

$$\begin{aligned} \|E^Y_{i+1}\|_{\mathbf{Q}_B} \le{}& \|(\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)E^Y_i\|_{\mathbf{Q}_B} \\ &+ \tau\|\mathbf{Q}_B^{-1}\mathbf{B}(\mathbf{A}^{-1} - \Psi)(\mathbf{A}E^X_i + \mathbf{B}^T E^Y_i)\|_{\mathbf{Q}_B}. \end{aligned} \tag{4.10}$$

Since $\tau \le \dfrac{\gamma}{\alpha^2}$, by (2.10) we have

$$\|(\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)E^Y_i\|_{\mathbf{Q}_B} \le \left(1 - \tau\frac{\gamma}{\alpha^2}\right)^{1/2} \|E^Y_i\|_{\mathbf{Q}_B} = \theta\|E^Y_i\|_{\mathbf{Q}_B}. \tag{4.11}$$

Because of (3.18), (4.2), the triangle inequality, and (2.8), the second term in the right-hand side of (4.10) is bounded as follows:

$$\|\mathbf{Q}_B^{-1}\mathbf{B}(\mathbf{A}^{-1} - \Psi)(\mathbf{A}E^X_i + \mathbf{B}^T E^Y_i)\|_{\mathbf{Q}_B} \le \epsilon(\|E^X_i\|_{\mathbf{A}_s} + \|E^Y_i\|_{\mathbf{Q}_B}). \tag{4.12}$$

Using (4.11) and (4.12) in (4.10) yields

$$\|E^Y_{i+1}\|_{\mathbf{Q}_B} \le \theta\|E^Y_i\|_{\mathbf{Q}_B} + \epsilon\tau(\|E^X_i\|_{\mathbf{A}_s} + \|E^Y_i\|_{\mathbf{Q}_B}). \tag{4.13}$$

Combining (4.9) and (4.13) gives

$$\begin{aligned} \|E^X_{i+1}\|_{\mathbf{A}_s} &\le \epsilon\|E^X_i\|_{\mathbf{A}_s} + (1+\epsilon)\|E^Y_i\|_{\mathbf{Q}_B}, \\ \|E^Y_{i+1}\|_{\mathbf{A}_s} &\le \epsilon\tau\|E^X_i\|_{\mathbf{A}_s} + (\theta + \epsilon\tau)\|E^Y_i\|_{\mathbf{Q}_B}. \end{aligned} \tag{4.14}$$

Let us adopt the notation

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \le \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

for vectors of nonnegative numbers x_1, x_2, y_1, y_2 if $x_1 \le x_2$ and $y_1 \le y_2$. Hence, from (4.14) we obtain

$$\begin{pmatrix} \|E^X_{i+1}\|_{\mathbf{A}_s} \\ \|E^Y_{i+1}\|_{\mathbf{Q}_B} \end{pmatrix} \le \begin{pmatrix} \epsilon & 1+\epsilon \\ \epsilon\tau & \theta+\epsilon\tau \end{pmatrix} \begin{pmatrix} \|E^X_i\|_{\mathbf{A}_s} \\ \|E^Y_i\|_{\mathbf{Q}_B} \end{pmatrix}. \tag{4.15}$$

Repeated application of (4.15) gives

$$\begin{pmatrix} \|E^X_{i+1}\|_{\mathbf{A}_s} \\ \|E^Y_{i+1}\|_{\mathbf{Q}_B} \end{pmatrix} \le \mathcal{M}^{i+1} \begin{pmatrix} \|E^X_0\|_{\mathbf{A}_s} \\ \|E^Y_0\|_{\mathbf{Q}_B} \end{pmatrix}, \tag{4.16}$$

where $\mathcal{M}$ is given by

$$\mathcal{R} = \begin{pmatrix} \epsilon & 1+\epsilon \\ \epsilon\tau & \theta+\epsilon\tau \end{pmatrix}.$$

We consider two dimensional Euclidean space with the inner product

$$\left\lfloor \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rfloor = \frac{\epsilon\tau}{1+\epsilon} x_1 x_2 + y_1 y_2.$$

A trivial computation shows that $\mathcal{R}$ is symmetric with respect to the inner product. It follows from (4.16) that

$$\begin{aligned} \frac{\epsilon\tau}{1+\epsilon}\|E_{i+1}^X\|_{\mathbf{A}_s}^2 + \|E_{i+1}^Y\|_{\mathbf{Q}_B}^2 &= \left\lfloor \begin{pmatrix} \|E_{i+1}^X\|_{\mathbf{A}_s} \\ \|E_{i+1}^Y\|_{\mathbf{Q}_B} \end{pmatrix}, \begin{pmatrix} \|E_{i+1}^X\|_{\mathbf{A}_s} \\ \|E_{i+1}^Y\|_{\mathbf{Q}_B} \end{pmatrix} \right\rfloor \\ &\le \left\lfloor \mathcal{R}^{i+1} \begin{pmatrix} \|E_0^X\|_{\mathbf{A}_s} \\ \|E_0^Y\|_{\mathbf{Q}_B} \end{pmatrix}, \mathcal{R}^{i+1} \begin{pmatrix} \|E_0^X\|_{\mathbf{A}_s} \\ \|E_0^Y\|_{\mathbf{Q}_B} \end{pmatrix} \right\rfloor \\ &\le \rho^{2(i+1)} \left(\frac{\epsilon\tau}{1+\epsilon}\|E_0^X\|_{\mathbf{A}_s}^2 + \|E_0^Y\|_{\mathbf{Q}_B}^2 \right), \end{aligned}$$

where ρ is the norm of the matrix $\mathcal{R}$ with respect to the $\lfloor\cdot,\cdot\rfloor$-inner product. Since $\mathcal{R}$ is symmetric in this inner product, its norm is equal to its spectral radius. The eigenvalues of $\mathcal{R}$ are the roots of

$$\lambda^2 - ((1+\tau)\epsilon + \theta)\lambda - \epsilon(\tau - \theta) = 0.$$

It is elementary to see that the root with largest absolute value is that given by (4.6). For any fixed positive τ and θ in the interval $[0,1]$, ρ is a function of ϵ only. It is straightforward to check that $\rho = 1$ only if

$$\epsilon = \frac{1-\theta}{1+2\tau-\theta}.$$

Moreover, $\rho = \theta$ if $\epsilon = 0$. Thus, $\rho < 1$ for $\epsilon \in [0, \frac{1-\theta}{1+2\tau-\theta})$.

Finally, we prove (4.5). Multiplying both sides of the first inequality in (4.14) by $\tau^{1/2}$ and using the fact that $0 < \tau < 1$, we obtain

$$\begin{aligned} \tau^{1/2}\|E_{i+1}^X\|_{\mathbf{A}_s} &\le \tau^{1/2}\epsilon\|E_i^X\|_{\mathbf{A}_s} + \tau^{1/2}(1+\epsilon)\|E_i^Y\|_{\mathbf{Q}_B} \\ &\le \tau^{1/2}\epsilon\|E_i^X\|_{\mathbf{A}_s} + (1+\epsilon)\|E_i^Y\|_{\mathbf{Q}_B}. \end{aligned}$$

We now apply the arithmetic-geometric mean inequality to the last inequality and get that, for any positive η,

$$\tau\|E_{i+1}^X\|_{\mathbf{A}_s}^2 \le (1+\eta)\tau\epsilon^2\|E_i^X\|_{\mathbf{A}_s}^2 + (1+\eta^{-1})(1+\epsilon)^2\|E_i^Y\|_{\mathbf{Q}_B}^2.$$

Inequality (4.5) follows by taking $\eta = 1 + 1/\epsilon$ and applying (4.4). This completes the proof of the theorem. □

We conclude this section with the following remarks.

Remark 4.1. If we fix all parameters except ϵ, then provided that the assumptions of the theorem are satisfied, $\rho \in [0,1)$. Moreover, ρ is a continuous function of ϵ which is equal to θ at $\epsilon = 0$, i.e., the theorem reproduces the result of Theorem 2.2 when $\epsilon = 0$. We clearly can achieve any convergence rate between one and θ with an appropriate choice of ϵ.

Remark 4.2. Theorem 4.1 is somewhat weaker than the result obtained in Section 3 for the linear case due to the threshold condition (4.3) on ϵ. As discussed above, it is possible to take sufficiently many iterations n, so that (4.3) holds for any fixed γ, α, and τ. In applications involving discretizations of partial differential equations,

β or α may depend on the discretization parameter h. If, however, these parameters can be bounded independently of h, then a fixed number of preconditioned GMRES iterations (independent of h) are sufficient to guarantee convergence of Algorithm 4.1.

5. Application to Navier-Stokes problems

In this section we consider an application of the algorithms developed in the previous sections to the problem of solving indefinite systems of linear equations arising from finite element approximations of the steady-state Navier-Stokes equations. We consider the following model problem for the steady-state Navier-Stokes equations:

$$\text{(5.1a)} \qquad -\nu\Delta\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} - \nabla p = \mathbf{f} \quad \text{in } \Omega,$$

$$\text{(5.1b)} \qquad \nabla\cdot\mathbf{u} = 0 \quad \text{in } \Omega,$$

$$\text{(5.1c)} \qquad \mathbf{u} = 0 \quad \text{on } \partial\Omega,$$

$$\text{(5.1d)} \qquad \int_\Omega p\, dx = 0.$$

Here Ω is a the unit square in $\mathbf{R}^2$, $\mathbf{u}$ is a vector valued function representing the fluid velocity, and ν is the kinematic viscosity of the flow. The fluid pressure p is a scalar function. The pressure is determined only up to an additive constant, so for uniqueness, we require (5.1d). Generalizations to more complex domains and nonhomogenious boundary conditions are possible. For example, we shall consider a problem with nonzero Dirichet boundary conditions in the next section.

Let Π be the set of functions in $L^2(\Omega)$ with zero mean value on Ω and let $H^1(\Omega)$ denote the Sobolev space of order one on Ω ([6, 13]). The space $H_0^1(\Omega)$ consists of those functions in $H^1(\Omega)$ whose traces vanish on $\partial\Omega$. Also, $\mathbf{V} = (H_0^1(\Omega))^2$ will denote the product space consisting of vector valued functions with each component in $H_0^1(\Omega)$.

In order to derive the weak formulation of (5.1) we multiply the first two equations of (5.1) by functions in $\mathbf{V}$ and Π respectively and integrate over Ω to get

$$\text{(5.2a)} \qquad \nu D(\mathbf{u},\mathbf{v}) + b(\mathbf{u},\mathbf{u},\mathbf{v}) + (p,\nabla\cdot\mathbf{v}) = (\mathbf{f},\mathbf{v}), \quad \text{for all} \quad \mathbf{v}\in\mathbf{V},$$

$$\text{(5.2b)} \qquad (\nabla\cdot\mathbf{u},q) = 0, \quad \text{for all} \quad q\in\Pi.$$

Here $(\cdot,\cdot)$ is the $L^2(\Omega)$ inner product and $D(\cdot,\cdot)$ denotes the Dirichlet form for vector functions on Ω defined by

$$D(\mathbf{v},\mathbf{w}) = \sum_{i=1}^{2}\int_\Omega \nabla v_i\cdot\nabla w_i\, dx.$$

The trilinear form $b(\cdot,\cdot,\cdot)$ for vector functions on Ω is given by

$$b(\mathbf{u},\mathbf{v},\mathbf{w}) = ((\mathbf{u}\cdot\nabla)\mathbf{v},\mathbf{w}).$$

The existence of a solution to (5.2) has been shown (cf. [17], [9]). It is well known that the Navier-Stokes equations may have more that one solution unless the data (the kinematic viscosity and the external forces) satisfy very stringent requirements (cf. [9], [17]). On the other hand, it has been shown that in many practical cases these solutions are mostly isolated, i.e. there exists a neighborhood of ν and $\mathbf{f}$ in which each solution is unique. We refer the reader to [9] and [17] for additional discussion.

We next define our finite element approximation subspaces. The discussion here is very closely related to the examples given in [3] and [2], where additional comments and other applications can be found. We partition Ω into $2n \times 2n$ squares, where n is a positive integer, and we define $h = 1/2n$. Let $x_i = ih$ and $y_j = jh$ for $i, j = 1, \ldots, 2n$. Each of the squares is further partitioned into two triangles by its diagonal with positive slope. Let S_h be the space of functions that are continuous and piecewise linear with respect to the triangulation just defined and vanish on $\partial\Omega$. We set $\mathbf{V}_h \equiv S_h \times S_h \subset \mathbf{V}$. The space $\tilde{\Pi}_h$ of functions that are piecewise constant with respect to the square elements and have zero mean value on Ω together with $\mathbf{V}_h$ as defined above form an unstable pair of approximation spaces. This means that the inequality

$$\|p\| \le c_0 \sup_{V \in \mathbf{V}_h} \frac{(\nabla \cdot V, p)}{D(V,V)^{1/2}}, \qquad \text{for all } p \in \tilde{\Pi}_h, \tag{5.3}$$

fails to hold. In fact, there is a function $p \in \tilde{\Pi}_h$ such that $(\nabla \cdot V, p) = 0$ for all $V \in \mathbf{V}_h$. Here $(\cdot,\cdot)$ denotes the inner product in $L^2(\Omega)$ and $\|\cdot\|$ is the corresponding norm. To get a divergence stable pair, we consider a smaller space defined as follows. Let η_{kl} for $k, l = 1, \ldots, 2n$ be the function that is 1 on the square element $[x_{k-1}, x_k] \times [y_{l-1}, y_l]$ and vanishes elsewhere. Define $\phi_{ij} \in \tilde{\Pi}_h$ for $i, j = 1, \ldots, n$ by

$$\phi_{ij} = \eta_{2i-1,2j-1} - \eta_{2i,2j-1} - \eta_{2i-1,2j} + \eta_{2i,2j}$$

(see Figure 1). The space Π_h is then defined by

$$\Pi_h \equiv \left\{ W \in \tilde{\Pi}_h \ : \ (W, \phi_{ij}) = 0 \text{ for } i, j = 1, \ldots, n \right\}.$$

Now (5.3), with $\tilde{\Pi}_h$ replaced by Π_h, is satisfied with a constant c_0 independent of h [10]. Moreover, the exclusion of the functions $\phi_{i,j}$ does not change the order of approximation, since Π_h still contains the piecewise constant functions on squares of size $2h$.

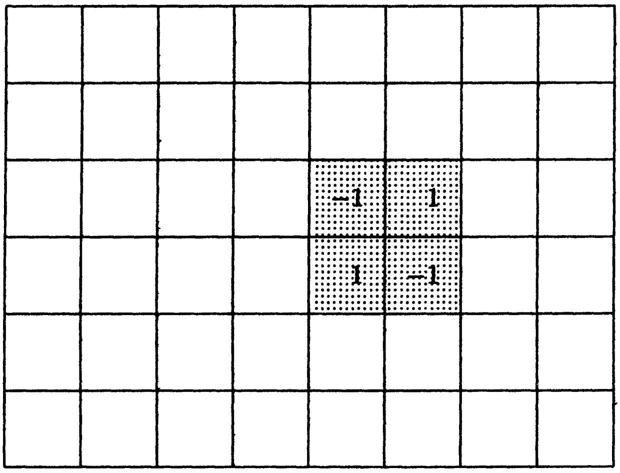

FIGURE 1. The square mesh used for $\tilde{H}_2$; the support (shaded) and values for a typical ϕ_{ij}

Following Temam [17], we introduce a modification $\tilde{b}(\cdot,\cdot,\cdot)$ of the trilinear form $b(\cdot,\cdot,\cdot)$, given by

$$\tilde{b}(\mathbf{u},\mathbf{v},\mathbf{w}) = ((\mathbf{u}\cdot\nabla)\mathbf{v},\mathbf{w}) - ((\mathbf{u}\cdot\nabla)\mathbf{w},\mathbf{v}).$$

The approximation to the solution of (5.2) is defined by the pair $(X,Y)\in \mathbf{V}_h\times\Pi_h$ satisfying

$$\nu D(X,V) + \tilde{b}(X,X,V) + (Y,\nabla\cdot V) = (\mathbf{f},V), \quad \text{for all} \quad V\in\mathbf{V}_h, \tag{5.4a}$$

$$(\nabla\cdot X,W) = 0, \qquad \text{for all} \quad W\in\Pi_h. \tag{5.4b}$$

Note that the use of $\tilde{b}(\cdot,\cdot,\cdot)$ above is justified by the observation that $\tilde{b}(\mathbf{u},\cdot,\cdot) = b(\mathbf{u},\cdot,\cdot)$ for functions $\mathbf{u}$ which are divergence free. The form of $\tilde{b}(\cdot,\cdot,\cdot)$ guarantees the existence of a solution to (5.4) (cf. [17]). The uniqueness requires conditions on the data ν and $\mathbf{f}$.

To solve (5.4) we apply a Picard iteration of the following type (cf. [11]). Given an initial approximation X^0, we compute (X^i,Y^i), for $i = 1, 2, ...,$ as the solution of the linear system

$$\nu D(X^i,V) + \tilde{b}(X^{i-1},X^i,V) + (Y^i,\nabla\cdot V) = (\mathbf{f},V), \quad \text{for all} \quad V\in\mathbf{V}_h, \tag{5.5a}$$

$$(\nabla\cdot X^i,W) = 0, \qquad \text{for all} \quad W\in\Pi_h. \tag{5.5b}$$

It is shown in [11] that the algorithm converges under the assumption that

$$\nu^2 c_a^2 > c_b\|\mathbf{f}\|_{-1},$$

where c_a and c_b are the coercivity and boundedness constants of the trilinear form $b(\cdot,\cdot,\cdot)$. Such an assumption is enough to guarantee a unique solution of (5.2).

The system (5.5) can be reformulated in the notation of the earlier sections. Set $H_1 = \mathbf{V}_h$ and $H_2 = \Pi_h$. Let

$$\mathbf{B}: H_1 \mapsto H_2, \quad (\mathbf{B}U,W) = (\nabla\cdot U,W), \quad \text{for all} \quad U\in H_1,\ W\in H_2,$$

and

$$\mathbf{B}^T: H_2 \mapsto H_1, \quad (\mathbf{B}^T W,V) = (W,\nabla\cdot V), \quad \text{for all} \quad V\in H_1,\ W\in H_2.$$

At each iterative step, X^{i-1} is fixed so that we can define

$$\mathbf{A}: H_1\mapsto H_1, \quad (\mathbf{A}U,V) = \nu D(U,V) + \tilde{b}(X^{i-1},U,V), \quad \text{for all} \quad U,\ V\in H_1.$$

It follows that the solution (X^i,Y^i) of (5.5) satisfies (1.1) with F equal to the $L^2(\Omega)$ projection of $\mathbf{f}$ onto H_1 and $G=0$. Notice also that

$$\tilde{b}(\mathbf{u},\mathbf{v},\mathbf{w}) = -\tilde{b}(\mathbf{u},\mathbf{w},\mathbf{v}).$$

Therefore,

$$\mathbf{A}_s: H_1\mapsto H_1, \quad (\mathbf{A}_s U,V) = \nu D(U,V), \quad \text{for all} \quad U,\ V\in H_1. \tag{5.6}$$

It is possible to show that (2.2) holds for $\mathbf{A}$ and $\mathbf{A}_s$ with a constant α proportional to ν^{-1} (cf. [17] and [9]). Moreover, it follows from (5.6) that (2.3) holds for $\mathbf{A}_s$, $\mathbf{B}$, and $\mathbf{B}^T$ as above with constant c_0 independent of the mesh size h. This implies that (2.8) is satisfied with $\mathbf{Q}_B = \nu^{-1}\mathbf{I}$ and γ bounded away from zero independently of h.

We still need to provide a preconditioner for $\mathbf{A}_s$. Clearly, $\mathbf{A}_s$ consists of two copies of the operator which results from a standard finite element discretization of Dirichlet's problem. There has been an intensive effort focused on the development and analysis of preconditioners for such problems. For the examples in Section 6, we will use a preconditioning operator which results from a V-cycle variational multigrid algorithm. Such a preconditioner can be scaled so that (3.1) holds with β independent of the mesh parameter h.

Remark 5.1. By rescaling p, one can rewrite (5.1a) in the form

$$-\Delta \mathbf{u} + Re(\mathbf{u} \cdot \nabla)\mathbf{u} - \nabla p = Re\, \mathbf{f},$$

where $Re = \nu^{-1}$ is the Reynolds number of the flow. This results in a different scaling of the discrete problem (5.4) which is better suited for implementation on finite precision machines. We use this scaling in our examples in the next section.

Remark 5.2. An alternative linearization of (5.4) can be defined by replacing $\tilde{b}(X, X, V)$ with $\tilde{b}(X^{i-1}, X^{i-1}, V)$ which provides a different Picard iteration. We will call this an explicit Picard iteration, because the nonlinear term is handled in an explicit fashion. This leads to a symmetric saddle point problem at each iteration. The inexact Uzawa methods analyzed in [4] can be used here. Even though the symmetric linear systems are easier to solve, this linearization is a less robust method for computing solutions to (5.4) than the implicit linearization defined above, because the explicit Picard iteration breaks down for values of ν where the implicit method converges. We shall provide a comparison of these two methods in the next section.

6. Numerical examples

In this section we present the results from numerical computations that illustrate the theory developed in the earlier sections. Our goals here are first to demonstrate the efficiency and the robustness of the algorithm, and also to provide a comparison between the implicit and the explicit Picard iteration applied to a Navier-Stokes problem with known analytic solution. In addition, we show results from computations of a classical flow problem. The finite element discretization defined in the previous section as well as the pressure rescaling according to Remark 5.1 are used in both cases.

The first computations are for the solution of (5.4) when the velocity X is given by

$$X = \begin{pmatrix} x(1-x)y(1-y) \\ x(1-x)y(1-y) \end{pmatrix}, \tag{6.1}$$

and the pressure Y is given by

$$Y = x - \frac{1}{2}. \tag{6.2}$$

Obviously, $\nabla \cdot X \neq 0$, so that the right-hand side of (5.4b) has to be adjusted appropriately.

We first give some results which indicate the behavior of the linear iteration method. Starting the implicit nonlinear iteration with zero initial guess, we report the number of iterations needed to solve the nonsymmetric linear system on the second nonlinear Picard iteration.

In all of our examples, we use one V–cycle sweep of variational multigrid with point Gauss-Seidel smoothing to define $\mathbf{A}_0^{-1}$. This results in (3.1) being satisfied with β independent of h. Thus, we can choose τ independent of h. The preconditioner $\mathbf{Q}_B$ was taken to be the identity.

TABLE 1. Linear iteration numbers on the second nonlinear iteration for the implicit method.

h	Re=1	Re=10	Re=100	Re=1000
1/8	39	50	590	37343
1/16	34	45	566	35964
1/32	30	41	527	35797
1/64	25	37	473	33827

Table 1 gives the number of iterations of the linear one-step method for a 10^{-8} normalized reduction of the residual. The nonsymmetric term was relatively weak for the case of $Re = 1$ and 10, so the discrete problem was dominated by its symmetric part. Thus, we set $\delta = 1/\beta$ and $\tau = 1$ consistent with the theory for the symmetric problem (cf. [4]). Note that in Algorithm 3.1, $\mathbf{A}_0^{-1}$ always appears multiplied by δ, so that using $\delta = 1/\beta$ with a preconditioner $\mathbf{A}_0$ satisfying (3.1) is equivalent to using $\delta = 1$ with the preconditioner $\tilde{\mathbf{A}}_0 = \beta \mathbf{A}_0$ satisfying

$$\frac{1}{\beta}(\tilde{\mathbf{A}}_0 V, V) \le (\mathbf{A}_s V, V) \le (\tilde{\mathbf{A}}_0 V, V), \tag{6.3}$$

for all $V \in H_1$. The unscaled multigrid preconditioner $\tilde{\mathbf{A}}_0$ automatically satisfies (6.3).

As the value of Re increased further, the values of τ and δ had to be adjusted to obtain stability of the iteration. The value $\tau = 0.1$ was chosen. To obtain stability in the case of $Re = 100$, we also needed to reduce δ to $0.1/\beta$. Finally, for stability in the case of $Re = 1000$, we needed to reduce δ to $0.001/\beta$. This is in agreement with the α^{-2} behavior required by the theory. The number of iterations was bounded independently of the mesh parameter h, as suggested by the theory.

The value of δ clearly depends on the strength of the nonsymmetric term. This, in turn, depends on the solution in nonlinear applications. For example, for the driven cavity results given below, the case of $h = 1/128$ and $Re = 1000$ required $\delta = 0.0001/\beta$ to remain stable, and diverged for $\delta = 0.001/\beta$.

The next set of experiments illustrates the differences between the implicit and explicit methods described in the previous section. Three conditions were common in all experiments. First, at each Picard iteration, the corresponding linear problem was solved exactly (i.e. the L^2 norm of the normalized residual was reduced until less than 10^{-8}). Second, the nonlinear iteration was considered to have converged when the L^2 norm of the difference $U_i - U_{i-1}$ was less than 10^{-6}. Here U consists of both velocity and pressure components. Finally, the Picard iteration was started with zero initial iterate. The numerical results from these experiments are shown in Tables 2–4. In the case of $Re = 1$ and 10, the nonlinearity is small, and both the explicit and implicit methods work well. In contrast, the explicit method fails to converge for $Re = 100$ while the implicit method behaves quite well.

Our second numerical experiment is the calculation of the flow in a cavity. The cavity domain Ω is the unit square and the flow is caused by a tangential velocity field applied to one of the square sides in the absence of other body forces. Since all forces are independent of time, the flow in this case limits to a steady-state which is modeled by (5.1) with corresponding changes in the boundary conditions (5.1c). In particular, the solution $\mathbf{u}$ on the boundary is zero everywhere except on the boundary segment $y = 1$, where $\mathbf{u} = (1, 0)$.

TABLE 2. Errors and nonlinear iteration numbers for $Re = 1$ for the implicit and explicit methods.

h	Error (p)	Error ($\mathbf{u}_1$)	Error ($\mathbf{u}_2$)	Implicit	Explicit
1/8	1.02e-2	7.87e-4	8.86e-3	4	4
1/16	2.50e-3	1.93e-4	5.41e-3	4	5
1/32	6.18e-4	4.81e-5	2.99e-3	4	5
1/64	1.93e-4	1.26e-5	1.57e-3	4	5

TABLE 3. Errors and nonlinear iteration numbers for $Re = 10$ for the implicit and explicit methods.

h	Error (p)	Error ($\mathbf{u}_1$)	Error ($\mathbf{u}_2$)	Implicit	Explicit
1/8	1.06e-2	7.87e-4	8.86e-3	5	6
1/16	2.60e-3	1.93e-4	5.41e-3	5	6
1/32	6.43e-4	4.81e-5	2.99e-3	5	6
1/64	1.65e-4	1.25e-5	1.57e-3	5	6

TABLE 4. Errors and nonlinear iteration numbers for $Re = 100$ for the implicit and explicit methods.

h	Error (p)	Error ($\mathbf{u}_1$)	Error ($\mathbf{u}_2$)	Implicit	Explicit
1/8	3.15e-2	7.87e-4	8.85e-3	8	30
1/16	7.83e-3	1.93e-4	5.42e-3	8	88*
1/32	1.95e-3	4.80e-5	2.99e-3	8	**
1/64	4.89e-4	1.20e-5	1.57e-3	8	**

* – the algorithm converged to a different solution with corresponding errors (p, $\mathbf{u}_1$, $\mathbf{u}_2$) 7.81e-3, 2.05e-4, 5.37e-3.

** – the algorithm could not converge to the solution.

The corresponding discrete problem in the spaces H_1 and H_2 as defined in the previous section is similar to (5.4) and is given by

$$D(X_0, V) + Re\,\tilde{b}(X, X_0, V) + (Y, \nabla \cdot V) = Re\,(\mathbf{f}, V) - D(\hat{X}, V) - Re\,\tilde{b}(X, \hat{X}, V),$$
$$(\nabla \cdot X_0, W) = 0,$$

for all $V \in H_1$ and $W \in H_2$. Here $X = X_0 + \hat{X}$, with $X_0 \in H_1$ and $\hat{X}$ satisfying the Dirichlet boundary conditions of $\mathbf{u}$ and vanishing at all interior vertex points from the triangulation of Ω. Note that $\nabla \cdot \hat{X} = 0$.

The implicit Picard iteration for this nonlinear problem is given as follows. Let $\hat{X}$ be as defined above. Then, given an initial iterate X_0^0, we compute (X_0^i, Y^i), for $i = 1, 2, \ldots$, by

$$\begin{aligned} D(X_0^i, V) &+ Re\,\tilde{b}(X^{i-1}, X_0^i, V) + (Y^i, \nabla \cdot V) \\ &= Re\,(\mathbf{f}, V) - D(\hat{X}, V) - Re\,\tilde{b}(X^{i-1}, \hat{X}, V), \end{aligned}$$

$$(\nabla \cdot X_0^i, W) = 0,$$

and set $X^i = X_0^i + \hat{X}$.

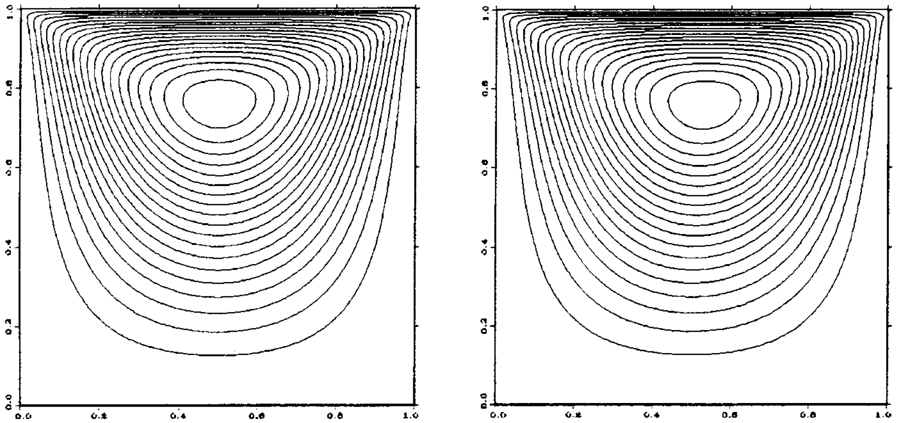

FIGURE 2. Streamlines for $h = 1/64$, and $Re = 1$ (left); $Re = 10$ (right).

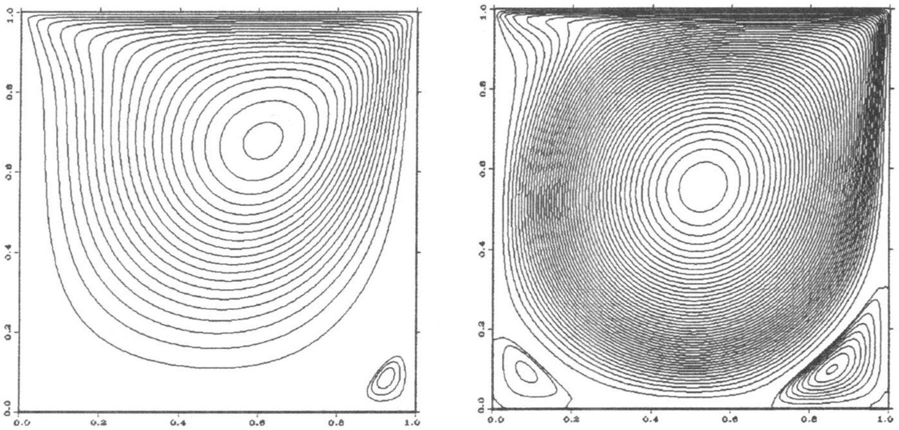

FIGURE 3. Streamlines for $h = 1/64$, $Re = 100$ (left); $Re = 1000$ (right).

The streamlines of the velocity field X computed using this algorithm for a range of Reynolds numbers are shown in Figures 2 and 3. The effect of the Reynolds number on the flow pattern is clearly seen. The flow for low Reynolds numbers (see Figure 2) has only one vortex center, located above the center of the domain (its location moves to the right as Re increases). As Re increases further, a second vortex center appears near the lower right corner (see Figure 3, the case of $Re = 100$), and, for even larger Reynolds numbers, a third vortex center develops near the lower left corner of the domain (see Figure 3, the case of $Re = 1000$).

Again, the case of $Re = 1000$ was the most difficult problem, requiring a large amount of work in the linear solver for each Picard iteration. The discretization with $h = 1/64$ was sufficiently fine for resolving the essential flow behavior for all Reynolds numbers tested. In contrast, the experimental results with $h = 1/16$ and $h = 1/32$ for $Re = 100$ did not show the vortex center near the lower right corner of the domain. The experiment with $h = 1/128$ and $Re = 1000$ resulted in a flow field whose streamlines were very similar to the ones from $h = 1/64$.

In conclusion, the implicit algorithm is a simple, robust and efficient method for solving Navier-Stokes equations for a wide range of Reynolds numbers. For each nonlinear iteration it requires the solution of a nonsymmetric saddle point problem which can be solved effectively with the inexact Uzawa algorithm 3.1.

References

[1] K. Arrow, L. Hurwicz, and H. Uzawa. *Studies in Linear and Nonlinear Programming.* Stanford University Press, Stanford, CA, 1958. MR **21:**7115

→ J.H. Bramble and J.E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Math. Comp.*, 50:1–18, 1988; 51:387–388, 1988. MR **89m:**65097

[3] J.H. Bramble and J.E. Pasciak. Iterative techniques for time dependent Stokes problems. *Computers Math. Applic.*, 33:13–30, 1997. MR **98e:**65091

→ J.H. Bramble, J.E. Pasciak, and A.T. Vassilev. Analysis of the inexact Uzawa algorithm for saddle point problems. *SIAM J. Numer. Anal.*, 34:1072–1092, 1997. MR **98c:**65182

[5] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag, New York, 1991. MR **92d:**65187

[6] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems.* North-Holland, New York, 1978. MR **58:**25001

[7] H. Elman. Preconditioning for the steady-state Navier-Stokes equations with low viscosity. Technical Report CS-TR-3712, Department of Computer Science, University of Maryland, College Park, MD 20742, 1996.

[8] H. Elman and D. Silvester. Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations. SIAM J. Sci. Comput., 17:33–46, 1996. MR **97e:**65119

[9] V. Girault and P.A. Raviart. *Finite Element Approximation of the Navier-Stokes Equations.* Lecture Notes in Math. # 749, Springer-Verlag, New York, 1981. MR **83b:**65122

[→ C. Johnson and J. Pitkäranta. Analysis of some mixed finite element methods related to reduced integration. *Math. Comp.*, 38:375–400, 1982. MR **83d:**65287

[→ O.A. Karakashian. On a Galerkin-Lagrange multiplier method for the stationary Navier-Stokes equations. *SIAM J. Numer. Anal.*, 19:909–923, 1982. MR **83j:**65107

[12] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. National Bureau of Standards*, 45:255–282, 1950. MR **13:**163d

[13] J.L. Lions and E. Magenes. *Problèmes aux Limites non Homogènes et Applications*, volume 1. Dunod, Paris, 1968. MR **40:**512

[14] M.M. Liu, J. Wang, and N.-N. Yan. New error estimates for approximate solutions of convection-diffusion problems by mixed and discontinuous Galerkin methods. *SIAM J. Numer. Anal.* Submitted.

[15] M.F. Murphy and A.J. Wathen. On preconditioning for the Oseen equations. Technical Report AM 95-07, Department of Mathematics, University of Bristol, 1995.

[16] Y. Saad and M.H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856 – 869, 1986. MR **87g:**65064

[17] R. Temam. *Navier-Stokes Equations.* North-Holland Publishing Co., New York, 1977. MR **58:**29439

Department of Mathematics, Texas A&M University, College Station, Texas 77843

Schlumberger, 8311 N. FM 620, Austin, Texas 78726