Inexact Uzawa 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 block saddle point problems. Specifically, we consider systems where the upper left block is invertable nonsymmetric 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 which utilize an "incomplete" or "approximate" evaluation of the inverse of the operator in the upper left block. Convergence results for the inexact algorithms are established in appropriate norms. The convergence of one of the algorithms is shown without the assumption of a sufficiently accurate approximation to the inverse operator. The other algorithm is shown to converge provided that the approximation to the inverse of the upper left hand block is of sufficient accuracy. Applications to the solution of steady-state nonlinear Navier–Stokes equations are discussed and finally, the results of numerical experiments involving the algorithms are presented.

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

AMS subject classifications. 65N22, 65N30, 65F10.

1 Introduction

This paper provides an analysis for the inexact Uzawa method applied to the solution of linear nonsymmetric saddle point systems. Such systems arise in certain discretizations

---

of Navier–Stokes equations, mixed discretizations of second order elliptic problems with convective terms (cf. [11], [14], [17], [20]). The theory in this paper is an extension of the theory 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 abstract saddle point problem:

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

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, not necessarily symmetric 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 $\mathbf{B}^T : H_2 \mapsto H_1$ is the adjoint of $\mathbf{B} : H_1 \mapsto H_2$.

In general, (1.1) may not even 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

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

is solvable. In the case of a symmetric and positive definite operator $\mathbf{A}$, the Ladyzhenskaya–Babuška–Brezzi (LBB) condition (cf. [6]) is 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 the LBB condition holds for the symmetric part of $\mathbf{A}$.

The papers [9], [18] propose solving $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ by preconditioned iteration. One common problem with this 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 [10]. Their preconditioner also requires the evaluation of $\mathbf{A}^{-1}$ during each step of the iteration.

The inexact Uzawa methods replace the exact inverse of $\mathbf{A}$ by an "incomplete" or "approximate" evaluation of $\mathbf{A}^{-1}$. Such algorithms are defined in Sections 3 and 4. In this paper we distinguish two types of inexact algorithms: (i) a linear one-step, where the action of the approximate inverse is provided by a linear preconditioner such as one sweep of a multigrid procedure; (ii) a multistep, where a sufficiently accurate approximation to $\mathbf{A}^{-1}$ is provided by some preconditioned iterative method, e.g., preconditioned GMRES [19] or preconditioned Lancos [15].

The inexact Uzawa 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 transformed trivially into an inexact algorithm. It is not surprising that the inexact Uzawa methods are widely used in the engineering community.

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 inexact Uzawa algorithm applied to (1.1). 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 from finite element approximations of the steady-state nonlinear Navier-Stokes equations. Finally, the results of numerical experiments involving the inexact Uzawa 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 applied to the solution of (1.2). Even though this algorithm is not very efficient for reasons already mentioned, the result of Theorem 2.2 below is important for the analysis of the inexact algorithms defined in the subsequent sections.

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

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

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

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

for some number α. Clearly, $\alpha \ge 1$. Moreover, since $\mathbf{A}_s$ is positive definite, such an α always exists. In many applications in 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 the pair of spaces H_1 and H_2, i.e.

$$(2.3) \qquad \sup_{\substack{U \in H_1 \\ U \ne 0}} \frac{(V, \mathbf{B}U)^2}{(\mathbf{A}_s U, U)} \ge c_0 \|V\|^2 \quad \text{for all} \quad V \in H_2,$$

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 replaced by $\mathbf{A}_s$. We will see that it also suffices in the case of nonsymmetric

A. To this end, we prove the following lemma which establishes that $(\mathbf{A}^{-1})_s$ is positive definite.

Lemma 2.1 *Let A be 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*

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

Proof: Clearly,

$$((\mathbf{A}_s)^{-1} W, W) = \sup_{\substack{U \in H_1 \\ U \ne 0}} \frac{(W, U)^2}{(\mathbf{A}_s U, U)} = \sup_{\substack{U \in H_1 \\ U \ne 0}} \frac{((A^{-1})^T W, \mathbf{A} U)^2}{(\mathbf{A}_s U, U)}$$

$$(2.5) \qquad\qquad \le \alpha^2 \sup_{\substack{U \in H_1 \\ U \ne 0}} \frac{\|(A^{-1})^T W\|_{\mathbf{A}_s}^2 \|U\|_{\mathbf{A}_s}^2}{\|U\|_{\mathbf{A}_s}^2} = \alpha^2 \|(A^{-1})^T W\|_{\mathbf{A}_s}^2$$

$$= \alpha^2 ((\mathbf{A}^{-1})_s W, W).$$

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

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

On the other hand,

$$((\mathbf{A}^{-1})_s U, U) = (\mathbf{A}^{-1} U, U) = (\mathbf{A}_s^{1/2} \mathbf{A}^{-1} U, (\mathbf{A}_s)^{-1/2} U)$$

$$\le \|\mathbf{A}^{-1} U\|_{\mathbf{A}_s} \|U\|_{(\mathbf{A}_s)^{-1}} = (\mathbf{A}^{-1} U, U) \|U\|_{(\mathbf{A}_s)^{-1}}.$$

Therefore,

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

This completes the proof of the lemma. $\square$

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

$$(\mathbf{B} \mathbf{A}^{-1} \mathbf{B}^T V, V) = ((\mathbf{A}^{-1})_s \mathbf{B}^T V, \mathbf{B}^T V)$$

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

Thus, we have proved the following theorem.

Theorem 2.1 *Let the linear operator $\mathbf{A}$ be invertible and let (2.3) hold. 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 $i = 1, 2, \ldots$, by*

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

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

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

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 particular applications effective preconditioners that satisfy (2.8) with γ bounded away from one are known.

Let

$$(2.9a) \qquad\qquad\qquad E_i^X = X - X_i$$

and

$$(2.9b) \qquad\qquad\qquad E_i^Y = Y - Y_i$$

be the iteration errors generated by the above method. It is an easy observation 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 2.2 *Let $\mathbf{A}$ be invertible with positive definite symmetric part $\mathbf{A}_s$ which satisfies (2.2). Let also (2.3) hold. In addition, let $\mathbf{Q}_B$ be a symmetric and positive definite operator satisfying (2.8). If τ is a positive parameter with $\tau \leq \dfrac{1 - \gamma}{\alpha^2}$, then*

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

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

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

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

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

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

In addition, using (2.4),

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

Taking $v = \mathbf{B}^T V$ and $w = \mathbf{B}^T W$ above gives

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

Next,

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

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

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

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

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

From this, the result of the theorem follows easily. $\square$

Remark 2.2 The inequalities (2.11) and (2.13) are the basis for developing the inexact algorithms in the subsequent sections.

3   Analysis of the linear one-step inexact method

In this section we define and analyze a linear one-step inexact Uzawa 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) result in an efficient and simple method for solving (1.1).

To this end, the exact inverse of $\mathbf{A}$ is replaced with an approximation of $\mathbf{A}^{-1}$ in order to improve the efficiency of Algorithm 2.1. Let $\mathbf{A}_0 : H_1 \mapsto H_1$ be a linear, symmetric and positive definite operator that satisfies

$$(3.1) \qquad (\mathbf{A}_0 V, V) \leq (\mathbf{A}_s V, V) \leq \beta(\mathbf{A}_0 V, V) \quad \text{for all} \quad V \in H_1,$$

for some positive $\beta \geq 1$.

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. Alternatively, preconditioners based on multigrid methods are often scaled appropriately by construction.

The linear inexact Uzawa algorithm is then defined as follows.

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

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

Here δ and τ are positive iteration 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 *Let $\mathbf{A}$ have a positive definite symmetric part $\mathbf{A}_s$ satisfying (2.2). Let also $\mathbf{Q}_B$ and $\mathbf{A}_0$ be symmetric and positive definite operators satisfying (2.8) and (3.1). Then the linear inexact Uzawa algorithm converges if $\delta \leq (3\alpha^2\beta^2)^{-1}$ and $\tau \leq (4\beta)^{-1}$. 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*

$$(3.2) \qquad \delta^{-1}\|E_i^X\|_*^2 + \tau^{-1}\|E_i^Y\|_{\mathbf{Q}_B}^2 \leq \bar{\rho}^i\left\{\delta^{-1}\|E_0^X\|_*^2 + \tau^{-1}\|E_0^Y\|_{\mathbf{Q}_B}^2\right\}$$

for any $i \geq 1$. Here $\|\cdot\|_*^2 = ((\mathbf{A}_0 - \delta\mathbf{A}_s)\cdot, \cdot)$ and

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

Remark 3.2 Convergence of the linear inexact Uzawa algorithm follows from (3.2). Indeed, a simple algebraic manipulation using the fact that $\tau(1-\gamma)$ is less than one gives

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

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

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.

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

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

$$(3.4) \qquad \mathcal{N}E_{i+1} = \mathcal{M}E_i,$$

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 now 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 $\mathcal{N}_s$ be the symmetric part of $\mathcal{N}$ and $\mathcal{M}_1$ 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

$$(3.5) \qquad \lambda\mathcal{N}_s\psi = \mathcal{M}_1\psi.$$

Since δ is less than $1/\beta$, $\mathcal{N}_s$ is symmetric and positive definite and the above problem is well defined. Obviously, (3.5) involves symmetric operators only so the eigenvalues λ are real.

Lemma 3.1 *The iteration error E_i satisfies*

$$(\mathcal{N}E_{i+1}, E_{i+1}) \leq \bar{\rho}(\mathcal{N}E_i, E_i),$$

where $\bar{\rho} = \max_i |\lambda_i|$, and λ_i are 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 $H_1 \times H_2$. Without loss of generality we may assume that the eigenvectors are normalized so that

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

where δ_{ij} denotes the Kronecker Delta Function. Then any arbitrary 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,

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

(3.6)

Obviously, $\mathcal{J}^2$ is the identity operator and hence $\mathcal{M} = \mathcal{J}\mathcal{M}_1$. Therefore, using (3.4) we get

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

The result of the lemma immediately follows. $\square$

Our proof of the theorem requires another lemma. We need to provide some control on the convergence of the related linear iterative process

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

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*

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

where

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

Remark 3.3 Clearly, $\bar{\delta}$ is less than one if

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

or

$$(3.9) \qquad \delta < \frac{1}{\alpha^2 \beta^2 + \beta} < \frac{1}{\beta}.$$

Note in addition, that

$$((\mathbf{A}_0 - \delta\mathbf{A}_s)V, V) \leq (1 - \delta)\|V\|_{\mathbf{A}_0}^2.$$

Thus, the lemma proves convergence of (3.7) provided that (3.9) holds.

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),

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

On the other hand,

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

In view of (3.1), we have

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

Also, (3.10) implies

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

Thus,

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

Using (3.12) and (3.13) in (3.11) yields (3.8). $\square$

Proof (of Theorem 3.1): To prove the theorem, we shall bound the positive and negative eigenvalues of (3.5) separately. 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 block equation by $\delta^{-1}\mathbf{A}_0(\mathbf{A}_0 - \delta\mathbf{A}^T)^{-1}$ gives

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

Applying $\delta \mathbf{B} \mathbf{A}_0^{-1}$ to the first equation and adding it to the second gives

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

Substituting this back into the first equation and taking an inner product with

$$\frac{\delta}{\lambda}(\mathbf{A}_0 - \delta\mathbf{A}^T)^{-1}(\mathbf{A}_0 - \delta\mathbf{A}_s)\chi$$

yields

(3.15)
$$-\frac{1}{\lambda}((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi) + \frac{\delta\tau}{1-\lambda}\|\mathbf{B}(\mathbf{A}_0 - \delta\mathbf{A}^T)^{-1}(\mathbf{A}_0 - \delta\mathbf{A}_s)\chi\|_{\mathbf{Q}_B^{-1}}^2$$
$$= \|(\mathbf{A}_0 - \delta\mathbf{A}^T)^{-1}(\mathbf{A}_0 - \delta\mathbf{A}_s)\chi\|_{\mathbf{A}_0}^2.$$

For convenience, the last equation can be abbreviated as

$$T_1 + T_2 = T_3.$$

For any $\phi \in H_1$,

(3.16)
$$(\mathbf{Q}_B^{-1}\mathbf{B}\phi, \mathbf{B}\phi) = \sup_{\substack{\zeta \in H_2 \\ \zeta \neq 0}} \frac{(\phi, \mathbf{B}^T\zeta)^2}{(\mathbf{Q}_B\zeta, \zeta)} = \sup_{\substack{\zeta \in H_2 \\ \zeta \neq 0}} \frac{((\mathbf{A}_s)^{1/2}\phi, (\mathbf{A}_s)^{-1/2}\mathbf{B}^T\zeta)^2}{(\mathbf{Q}_B\zeta, \zeta)}$$
$$\leq \sup_{\substack{\zeta \in H_2 \\ \zeta \neq 0}} \frac{(\mathbf{A}_s\phi, \phi)(\mathbf{B}(\mathbf{A}_s)^{-1}\mathbf{B}^T\zeta, \zeta)}{(\mathbf{Q}_B\zeta, \zeta)} \leq (\mathbf{A}_s\phi, \phi).$$

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

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

Using this in (3.15) gives

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

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

(3.17)
$$\left((\mathbf{A}_0 - \delta\mathbf{A}^T)\mathbf{A}_0^{-1}(\mathbf{A}_0 - \delta\mathbf{A})\phi, \phi\right) \leq \bar{\delta}((\mathbf{A}_0 - \delta\mathbf{A}_s)\phi, \phi).$$

This in turn implies that

(3.18)
$$((\mathbf{A}_0 - \delta\mathbf{A}_s)^{-1}\phi, \phi) \leq \bar{\delta}((\mathbf{A}_0 - \delta\mathbf{A})^{-1}\mathbf{A}_0(\mathbf{A}_0 - \delta\mathbf{A}^T)^{-1}\phi, \phi).$$

Hence,

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

Combining and using the fact that $\lambda < 0$ gives

$$-\frac{1}{\lambda}((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi) \geq \frac{1}{\bar{\delta}}\left(1 - \frac{\delta\tau\beta}{1 - \lambda}\right)((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi)$$

(3.19)

$$\geq \frac{1 - \delta\tau\beta}{\bar{\delta}}((\mathbf{A}_0 - \delta\mathbf{A}_s)\chi, \chi).$$

Now, if $\chi = 0$ then the first equation in (3.14) implies that $\mathbf{B}^T\xi = 0$. Then, from the second equation in (3.14), we get that $\xi = 0$. Hence, we can assume that $\chi \neq 0$ in (3.19). Thus,

(3.20)
$$-\lambda \leq \frac{\bar{\delta}}{1 - \delta\tau\beta}.$$

Let $\delta \leq \frac{1}{3\alpha^2\beta^2}$ and $\tau \leq \frac{1}{4\beta}$. Applying straightforward manipulations, we get

(3.21a)
$$\bar{\delta} = 1 - \delta + \frac{\alpha^2\beta^2\delta^2}{1 - \delta\beta} \leq 1 - \delta\left(1 - \frac{1/3}{1 - 1/3}\right) = 1 - \frac{\delta}{2}$$

and

(3.21b)
$$\frac{1}{1 - \delta\tau\beta} \leq \frac{1}{1 - \delta/4}.$$

Using (3.21) in (3.20) gives

(3.22)
$$-\lambda \leq \frac{1 - \delta/2}{1 - \delta/4} \leq 1 - \frac{\delta}{4},$$

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

Next, we bound the positive eigenvalues of (3.5). Let us 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 $\theta = 1 - \delta/2$. By definition, the largest eigenvalue of (3.5) is

$$\lambda = \sup_{\substack{\mathbf{w} \in H_1 \times H_2 \\ \mathbf{w} \neq 0}} \frac{(\mathcal{M}_1 \mathbf{w}, \mathbf{w})}{(\mathcal{N}_s \mathbf{w}, \mathbf{w})} = \sup_{\substack{\mathbf{w} \in H_1 \times H_2 \\ \mathbf{w} \neq 0}} \frac{(\mathcal{M}_2 \mathcal{D}\mathbf{w}, \mathcal{D}\mathbf{w})}{(\mathcal{N}_s \mathbf{w}, \mathbf{w})}.$$

We now show that for any vector $\begin{pmatrix} \phi \\ \zeta \end{pmatrix} \in H_1 \times H_2$,

$$(3.23) \qquad \left(\mathcal{M}_2 \begin{pmatrix} \phi \\ \zeta \end{pmatrix}, \begin{pmatrix} \phi \\ \zeta \end{pmatrix} \right) \leq \bar{\rho} \left[\delta^{-1} \|\phi\|^2 + \tau^{-1} \|\zeta\|_{\mathbf{Q}_B}^2 \right].$$

Let $\mathbf{L} = \mathbf{B}(\mathbf{A}_0)^{-1/2}$. Then

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

To prove (3.23), we need to estimate the largest eigenvalue of

$$(3.24a) \qquad -\theta\delta^{-1}\chi + \theta^{1/2}\mathbf{L}^T\xi = \lambda\delta^{-1}\chi$$

$$(3.24b) \qquad \theta^{1/2}\mathbf{L}\chi + (\tau^{-1}\mathbf{Q}_B - \delta\mathbf{L}\mathbf{L}^T)\xi = \lambda\tau^{-1}\mathbf{Q}_B\xi,$$

where $\begin{pmatrix} \chi \\ \xi \end{pmatrix}$ is an eigenvector. Solving for χ in (3.24a) we get

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

Substituting this in (3.24b) yields

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

Taking an inner product with ξ in the above equation gives

$$(3.25) \qquad (1 - \lambda)(\lambda + \theta)(\mathbf{Q}_B\xi, \xi) = \delta\tau\lambda(\mathbf{L}^T\xi, \mathbf{L}^T\xi).$$

If $\xi = 0$, then (3.24a) implies that either $\chi = 0$ or $\lambda = -\theta \leq 0$. Hence, we can assume that $\xi \neq 0$. In addition, by (3.1) and (2.8),

$$(\mathbf{L}^T\xi, \mathbf{L}^T\xi) = (\mathbf{A}_0^{-1}\mathbf{B}^T\xi, \mathbf{B}^T\xi) \geq ((\mathbf{A}_s)^{-1}\mathbf{B}^T\xi, \mathbf{B}^T\xi)$$
$$\geq (1 - \gamma)(\mathbf{Q}_B\xi, \xi).$$

Using this in (3.25) gives

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

or equivalently

$$\lambda^2 - \lambda(1 - \theta - \delta\tau(1 - \gamma)) - \theta \leq 0.$$

From here we obtain that

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

Next, we observe that for

$$
\begin{pmatrix} \phi \\ \zeta \end{pmatrix} = \mathcal{D} \begin{pmatrix} \chi \\ \xi \end{pmatrix}
$$

the following estimate holds:

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

Equivalently

$$
(3.28) \qquad \delta^{-1} \|\phi\|^2 + \tau^{-1} \|\zeta\|_{\mathbf{Q}_B}^2 \le \left(\mathcal{N}_s \begin{pmatrix} \chi \\ \xi \end{pmatrix}, \begin{pmatrix} \chi \\ \xi \end{pmatrix} \right) = \delta^{-1} \|\chi\|_*^2 + \tau^{-1} \|\xi\|_{\mathbf{Q}_B}^2.
$$

Indeed, (3.27) is a direct consequence of (3.17) and (3.21a). It is clear now that (3.23), (3.28), and (3.26) provide the bound for the positive part of the spectrum.

Finally, elementary inequalities imply that

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

which concludes the proof of the theorem. $\square$

4   Analysis of the multistep inexact algorithm

In this section we define and analyze an inexact Uzawa algorithm with $\mathbf{A}^{-1}$ replaced with 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 certain accuracy of the approximation to $\mathbf{A}^{-1}$ typically requires more computational work than the evaluation of the action of a linear one-step preconditioner. Second, depending on the way 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 is 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

$$
(4.1) \qquad \mathbf{A}\xi = \phi.
$$

More precisely, we assume that

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

for some positive δ with $\delta < 1$.

Notice that for any $\delta \in (0, 1)$, (4.2) can be satisfied by taking sufficiently many steps in some iterative method for solving (4.1) which reduces the error in a norm equivalent to $\|\cdot\|_{\mathbf{A}_s}$. For example, we already showed in the previous section that for appropriate choice of the corresponding iteration parameter, the linear iteration (3.7) converges (cf. Remark 3.3) to the solution of the linear system (4.1). Hence, an estimate of the type of (4.2) can be established easily for any $\delta < 1$, provided that sufficiently many iterations with (3.7) are performed.

In addition, in the case when $\mathbf{A}$ corresponds to a second order differential operator, there are preconditioners $\mathbf{B}$ based on multigrid (cf. [5], [21], [12]) or domain decomposition [7] which satisfy

$$(4.3) \qquad \|(\mathbf{I} - \mathbf{B}\mathbf{A})\chi\|_{\mathbf{A}_s} \leq \tilde{\delta} \|\chi\|_{\mathbf{A}_s} \quad \text{for all} \quad \chi \in H_1,$$

for some $\tilde{\delta} < 1$. Some of these preconditioners are even nonsymmetric. Typically, these methods require sufficiently fine coarse grid in order to work for a given small $\tilde{\delta}$. Taking $\chi = \mathbf{A}^{-1}\phi$ in (4.3) trivially implies (4.2), provided that $\tilde{\delta} \leq \delta$.

Another example for Ψ is a generalized Lanczos procedure [15] applied to (4.1) which converges to the solution ξ. In this case the resulting Uzawa algorithm will be nonlinear. Among the variety of conjugate gradient-like methods for solving (4.1) proposed in the literature, there are some for which convergence can be shown rigorously. In particular, a convergence of the following type is known to hold (cf. [19]) for the generalized minimal residual algorithm (GMRES):

$$\|\xi_n - \mathbf{A}^{-1}\phi\|_{\mathbf{A}_s} \leq \delta_n \|\mathbf{A}^{-1}\phi\|_{\mathbf{A}_s} \quad \text{for all} \quad \phi \in H_1,$$

where $\xi_n = \Psi(\phi)$ is the approximation to the solution computed at the n-th iteration and $\delta_n \to 0$ as n increases. Unlike the case when $\mathbf{A} = \mathbf{A}^T$, a rate of convergence for GMRES is generally not available even though this algorithm reaches the threshold $\delta_n < \delta$ eventually. In practice, GMRES may be a more efficient method for computing an approximation satisfying (4.2) than the linear iteration (3.7).

The variant of the inexact Uzawa algorithm 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 $i = 1, 2, \ldots,$ by*

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

Clearly, 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 δ introduced in (2.2), (2.8), and (4.2) respectively. The theorem below is a sufficient condition on δ for convergence of the algorithm.

Theorem 4.1 *Let* $\mathbf{A}$ *have a positive definite symmetric part* $\mathbf{A}_s$ *satisfying* (2.2) *and let* $\mathbf{Q}_B$ *be symmetric and positive definite operator satisfying* (2.8). *Assume that* (4.2) *holds and that the iteration parameter* τ *is chosen so that*

$$\tau \leq \frac{1-\gamma}{\alpha^2}.$$

Set

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

Then the multistep inexact Uzawa algorithm converges if

$$(4.4) \qquad \delta \leq \frac{1-\theta}{1+2\tau-\theta}.$$

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*

$$(4.5) \qquad \frac{\delta\tau}{1+\delta}\|E_{i+1}^X\|_{\mathbf{A}_s}^2 + \|E_{i+1}^Y\|_{\mathbf{Q}_B}^2 \leq \rho^{2(i+1)}\left(\frac{\delta\tau}{1+\delta}\|E_0^X\|_{\mathbf{A}_s}^2 + \|E_0^Y\|_{\mathbf{Q}_B}^2\right)$$

and

$$(4.6) \qquad \|E_{i+1}^X\|_{\mathbf{A}_s}^2 \leq \tau^{-1}(1+\delta)(1+2\delta)\rho^{2i}\left(\frac{\delta\tau}{1+\delta}\|E_0^X\|_{\mathbf{A}_s}^2 + \|E_0^Y\|_{\mathbf{Q}_B}^2\right),$$

where

$$(4.7) \qquad \rho = \frac{(1+\tau)\delta + \theta + \sqrt{((1+\tau)\delta+\theta)^2 + 4\delta(\tau-\theta)}}{2} < 1.$$

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

$$(4.8) \qquad \begin{aligned} E_{i+1}^X &= E_i^X - \Psi\left(\mathbf{A}E_i^X + \mathbf{B}^T E_i^Y\right), \\ E_{i+1}^Y &= E_i^Y + \tau\mathbf{Q}_B^{-1}\mathbf{B}E_{i+1}^X. \end{aligned}$$

The first equation above can be rewritten

$$(4.9) \qquad E_{i+1}^X = (\mathbf{A}^{-1} - \Psi)\left(\mathbf{A}E_i^X + \mathbf{B}^T E_i^Y\right) - \mathbf{A}^{-1}\mathbf{B}^T E_i^Y.$$

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

$$
\begin{aligned}
\|E_{i+1}^X\|_{\mathbf{A}_s} &\le \delta(\|E_i^X\|_{\mathbf{A}_s} + \|\mathbf{A}^{-1}\mathbf{B}^T E_i^Y\|_{\mathbf{A}_s}) + \|\mathbf{A}^{-1}\mathbf{B}^T E_i^Y\|_{\mathbf{A}_s} \\
&= \delta\|E_i^X\|_{\mathbf{A}_s} + (1+\delta)\|\mathbf{B}^T E_i^Y\|_{(\mathbf{A}^{-1})_s} \\
&\le \delta\|E_i^X\|_{\mathbf{A}_s} + (1+\delta)\|E_i^Y\|_{\mathbf{Q}_B}.
\end{aligned}
$$
(4.10)

Using (4.9) in the second equation of (4.8) gives

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

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

$$
\begin{aligned}
\|E_{i+1}^Y\|_{\mathbf{Q}_B} &\le \|(\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)E_i^Y\|_{\mathbf{Q}_B} \\
&\quad + \tau\|\mathbf{Q}_B^{-1}\mathbf{B}(\mathbf{A}^{-1} - \Psi)(\mathbf{A}E_i^X + \mathbf{B}^T E_i^Y)\|_{\mathbf{Q}_B}.
\end{aligned}
$$
(4.11)

Since $\tau \le \dfrac{1-\gamma}{\alpha^2}$, by (2.10) we have

$$
(4.12) \quad \|(\mathbf{I} - \tau\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)E_i^Y\|_{\mathbf{Q}_B} \le \left(1 - \tau\frac{1-\gamma}{\alpha^2}\right)^{1/2}\|E_i^Y\|_{\mathbf{Q}_B} = \theta\|E_i^Y\|_{\mathbf{Q}_B}.
$$

Because of (3.16), (4.2), the triangle inequality, and (2.8), the second term in the right-hand side of (4.11) is bounded as follows:

$$
(4.13) \qquad \|\mathbf{Q}_B^{-1}\mathbf{B}(\mathbf{A}^{-1} - \Psi)(\mathbf{A}E_i^X + \mathbf{B}^T E_i^Y)\|_{\mathbf{Q}_B} \le \delta(\|E_i^X\|_{\mathbf{A}_s} + \|E_i^Y\|_{\mathbf{Q}_B}).
$$

Using (4.12) and (4.13) in (4.11) yields

$$
(4.14) \qquad\qquad \|E_{i+1}^Y\|_{\mathbf{Q}_B} \le \theta\|E_i^Y\|_{\mathbf{Q}_B} + \delta\tau(\|E_i^X\|_{\mathbf{A}_s} + \|E_i^Y\|_{\mathbf{Q}_B}).
$$

Combining (4.10) and (4.14) gives

$$
\begin{aligned}
\|E_{i+1}^X\|_{\mathbf{A}_s} &\le \delta\|E_i^X\|_{\mathbf{A}_s} + (1+\delta)\|E_i^Y\|_{\mathbf{Q}_B} \\
\|E_{i+1}^Y\|_{\mathbf{A}_s} &\le \delta\tau\|E_i^X\|_{\mathbf{A}_s} + (\theta + \delta\tau)\|E_i^Y\|_{\mathbf{Q}_B}.
\end{aligned}
$$
(4.15)

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 (4.15) we obtain

$$
(4.16) \qquad \begin{pmatrix} \|E_{i+1}^X\|_{\mathbf{A}_s} \\ \|E_{i+1}^Y\|_{\mathbf{Q}_B} \end{pmatrix} \le \begin{pmatrix} \delta & 1+\delta \\ \delta\tau & \theta + \delta\tau \end{pmatrix} \begin{pmatrix} \|E_i^X\|_{\mathbf{A}_s} \\ \|E_i^Y\|_{\mathbf{Q}_B} \end{pmatrix}.
$$

Repeated application of (4.16) gives

$$(4.17) \qquad \begin{pmatrix} \|E_{i+1}^X\|_{\mathbf{A}_s} \\ \|E_{i+1}^Y\|_{\mathbf{Q}_B} \end{pmatrix} \le \mathcal{M}^{i+1} \begin{pmatrix} \|E_0^X\|_{\mathbf{A}_s} \\ \|E_0^Y\|_{\mathbf{Q}_B} \end{pmatrix}$$

where $\mathcal{M}$ is given by

$$\mathcal{M} = \begin{pmatrix} \delta & 1+\delta \\ \delta\tau & \theta + \delta\tau \end{pmatrix}.$$

We consider two dimensional Euclidean space with the inner product

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

A trivial computation shows that $\mathcal{M}$ is symmetric with respect to the inner product. It follows from (4.17) that

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

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

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

It is elementary to see that the spectral radius of $\mathcal{M}$ is equal to its positive eigenvalue which is given by (4.7).

Examining the expression for ρ given by (4.7) we see that the square root expression is nonnegative. Moreover, for any fixed positive τ and θ in the interval $[0,1]$, ρ is a function of δ only. It is straightforward to see that $\rho < 1$ if

$$\delta \le \frac{1-\theta}{1+2\tau-\theta}.$$

Finally, we prove (4.6). Multiplying both sides of the first inequality in (4.15) by $\tau^{1/2}$ and using the fact that $\tau < 1$ we obtain

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

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

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

Inequality (4.6) follows by taking $\eta = 1 + 1/\delta$ and applying (4.5). This completes the proof of the theorem. $\square$

We conclude this section with the following remarks.

Remark 4.1 The result of Theorem 4.1 is somewhat weaker than the result obtained in Section 3 for the linear case due to the threshold condition (4.4) on δ. In principle, it is possible to take sufficiently many iterations n so that (4.4) holds for any fixed γ, α, and τ. In applications involving partial differential equations, γ or α may depend on the discretization parameter h. If, however, γ and α can be bounded independently of h with γ also bounded away from one, then θ can be bounded away from one also. Hence, a fixed number (independent of h) of iterations of (3.7) are sufficient to guarantee convergence of Algorithm 4.1.

Remark 4.2 The result of Theorem 4.1 is similar to the result of Theorem 4.1 in [4], which considers the case of a multistep inexact Uzawa algorithm applied to a symmetric indefinite problem. The case of a nonsymmetric $\mathbf{A}$ however is inherently more difficult. Thus, in practice it is always more expensive computationally to satisfy (4.2) than its symmetric counterpart in [4] in order to guarantee convergence of the corresponding algorithm.

5   Application to Navier-Stokes problems

Here we consider an application of the algorithms developed in the previous sections to solving indefinite systems of linear equations arising from finite element approximations of the steady-state Navier-Stokes equations.

The Navier-Stokes equations provide the flow model of Newtonian fluids. This is the simplest and arguably the most useful model of viscous, incompressible fluid behavior. If the forces driving the flow are time independent, the flow is stationary. We consider the following model problem for the steady-state Navier-Stokes equations:

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

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

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

$$(5.1d) \qquad\qquad\qquad\qquad \int_\Omega p(x)\,dx = 0.$$

Here Ω is a the unit square in $\mathbb{R}^2$, $\mathbf{u}$ 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 of a Newtonian fluid 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 $H^1(\Omega)$ denote
the Sobolev space of order one on Ω ([8, 16]). The space $H_0^1(\Omega)$ consists of those
functions in Ω 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 vector 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

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

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

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

$$b(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \sum_{i,j=1}^{2} \int_\Omega \mathbf{u}_i(D_i\mathbf{v}_j)\mathbf{w}_j \, dx,$$

where $D_i \equiv \frac{\partial}{\partial x_i}$.

The existence of a solution to (5.2) has been shown (cf. [20], [11]). It is well
known that the Navier-Stokes equations have more that one solution unless the data
(the kinematic viscosity and the external forces) satisfy very stringent requirements (cf.
[11], [20]). 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. These solutions depend continuously on ν. Therefore, as ν varies
in a given interval, each solution describes an isolated branch. This means that the
bifurcation of solutions is rare and branches of solutions can be computed. We refer
the reader to [11] and [20] for additional discussion of the subject.

We next define our finite element approximation subspaces. The discussion here is
very closely related to the examples given in [2] and [3] where additional comments and
other applications can be found. We partition Ω into $2n \times 2n$ square shaped elements,
where n is a positive integer and define $h = 1/2n$. Let $x_i = ih$ and $y_j = jh$ for
$i, j = 1, \ldots, 2n$. Each of the square elements is further partitioned into two triangles
by connecting the lower left corner to the upper right corner. Let S_h be the space
of functions that vanish on $\partial\Omega$ and are continuous and piecewise linear with respect
to the triangulation just defined. We set $\mathbf{V}_h \equiv S_h \times S_h \subset \mathbf{V}$. The definition of the
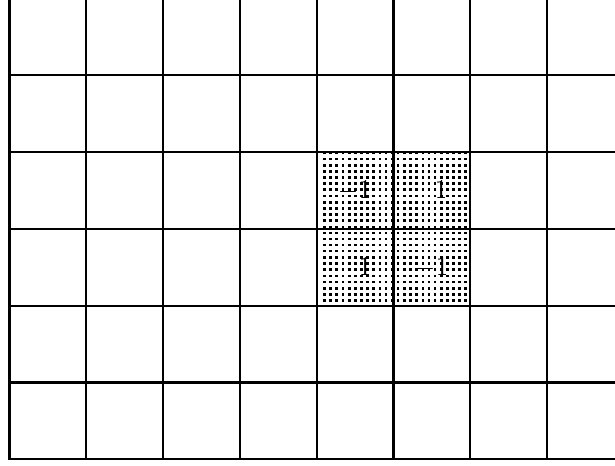
Figure 1: The square mesh used for $\tilde{H}_2$; the support
(shaded) and values for a typical ϕ_{ij}

approximation to Π is motivated by the observation [13] that 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

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

fails to hold with constant c_0 independent of the discretization parameter 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

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

The pair $\mathbf{V}_h \times \Pi_h$ now satisfies (5.3) with a constant c_0 independent of h [13]. Moreover,
the exclusion of the functions $\phi_{i,j}$ does not change the order of approximation for the
space since Π_h still contains the piecewise constant functions of size $2h$.

Following Temam [20], we introduce a modification $\tilde{b}(\cdot, \cdot, \cdot)$ of the trilinear form
$b(\cdot, \cdot, \cdot)$, given by

$$\tilde{b}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \sum_{i,j}^{2} \frac{1}{2} \int_{\Omega} \{ \mathbf{u}_i[(D_i \mathbf{v}_j)\mathbf{w}_j - v_j(D_i \mathbf{w}_j)] \} \, dx.$$

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

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

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

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. [20]). The uniqueness is again subject to imposing conditions on the data ν and $\mathbf{f}$.

To solve (5.4) we apply a Picard iteration of the following type (cf. [14]). Given an initial approximation X^0, we compute (X^i, Y^i), for $i = 1, 2, ...,$ as the solution of the linear system

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

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

The convergence analysis of this algorithm is beyond the scope of the present paper. It is shown in [14] 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

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

During 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 $\mathbf{f}$ into 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,

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

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. [20] and [11]). 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 one independently of h.

We still need to provide preconditioners for $\mathbf{A}_s$. However, $\mathbf{A}_s$ 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 It appears from the definition of the above operators that one has to invert Gram matrices in order to evaluate the action of $\mathbf{A}$, $\mathbf{B}^T$ and $\mathbf{B}$ on vectors from the corresponding spaces. In practice, the H_1 Gram matrix inversion is avoided by suitable definition of the preconditioner $\mathbf{Q}_A$. For the purpose of computation, the evaluation of $\mathbf{Q}_A^{-1}W$ for $W \in H_1$ is defined as a process which acts on the inner product data (W, ψ_i) where $\{\psi_i\}$ is the basis for H_1. Moreover, from the definition of the Uzawa-like algorithms in the previous sections, it is clear that every occurrence of $\mathbf{A}$ or $\mathbf{B}^T$ is followed by an evaluation of $\mathbf{Q}_A^{-1}$. Thus the inversion of the Gram matrix is avoided since the data for the computation of $\mathbf{Q}_A^{-1}$, $((\mathbf{B}^T Q, \psi_i)$ and $(\mathbf{A}V, \psi_i))$, for any $Q \in H_2$ and $V \in H_1$, can be computed by applying simple sparse matrices. In the case of this special choice of H_2, it is possible to compute the operator $\mathbf{B}$ in an economical way (see Remark 5 of [3]) and we can take $\mathbf{Q}_B$ to be $\nu^{-1}\mathbf{I}$. For more general spaces H_2, the inversion of Gram matrices can be avoided by introducing a preconditioner $\mathbf{Q}_B$ whose inverse is implemented acting on inner product data as in the H_1 case above.

Remark 5.2 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 5.3 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 branches of 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 experiments that illustrate the theory developed in the earlier sections. Our goals here are first to demonstrate the efficiency and the robustness of the new algorithms on the basis of a comparison between the implicit and the explicit Picard iteration applied to a Navier-Stokes problem with known analytic solution. Second, 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 5.2 are used in both cases.

Our first experiment compares the performance of the implicit and the explicit methods applied to the solution of (5.4) when the velocity X is given by

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

and the pressure Y is given by

$$(6.2) \qquad\qquad Y = x - \frac{1}{2}.$$

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

The implicit and explicit algorithms were tested for a set of different Reynolds numbers ($Re = 1, 10, 100, 1000$), and different mesh discretization parameters ($h = 1/8, 1/16, 1/32$). Clearly, the exact solution defined above is very smooth in Ω, without any singularities. The experiments described below show the asymptotic behavior of the error of the approximate solution computed by the two algorithms for the selected set of Reynolds numbers.

Four 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^{-15}). 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. Third, the Picard iteration was started with zero initial guess. Fourth, we defined $\mathbf{Q}_A^{-1}$ to be the operator which corresponds to one V–cycle sweep of variational multigrid with point Gauss-Seidel smoothing. The order of points in the Gauss-Seidel iteration was reversed in pre– and post–smoothing. The preconditioner $\mathbf{Q}_B$ was provided by an appropriate scale of the identity operator in the pressure space (cf. Remark 5.1).

In all experiments $\tau = 0.1$ was used when the nonsymmetric saddle point problem (5.4) was solved. This comes from the fact that β in (3.1) is independent of Re, because of the properties of the trilinear form $\tilde{b}(\cdot, \cdot, \cdot)$. The parameter δ in this case was set to $\delta = 1/Re$, where Re is the corresponding Reynolds number. Alternatively, δ and τ were set to one for the case of symmetric saddle point problem. These choices for δ provided the appropriate scaling of $\mathbf{Q}_A$ according to the requirements of the corresponding algorithm

(cf. (3.1) and [4]). In both cases this resulted in a preconditioner independent of the mesh parameter h.

The numerical results from these experiments are shown in Tables 1–4.

Table 1: 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	2.51e-3	1.93e-4	5.41e-3	4	5
1/32	6.17e-4	4.81e-5	2.99e-3	4	5

Table 2: 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	6	6
1/16	2.60e-3	1.93e-4	5.41e-3	6	6
1/32	6.43e-4	4.81e-5	2.99e-3	6	6

Table 3: 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	3.14e-2	7.86e-4	8.85e-3	10	30
1/16	7.78e-3	1.92e-4	5.41e-3	11	88*
1/32	1.94e-3	4.79e-5	2.99e-3	11	**

* − 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.

We note that the difference in the velocity error obtained for a given mesh parameter h in Tables 1–4 above is due to the nonsymmetric pressure (6.2), even though the velocity (6.1) is symmetric with respect to the spatial variables x and y in Ω. The computational results from the first experiment are in a good agreement with the theory developed in the paper. They show that the implicit method is a robust algorithm for solving Navier-Stokes equations in a wide range of Reynolds numbers. The number of inner iterations was independent of the mesh parameter h and exhibited a mild dependence on ν for $1 \leq \nu^{-1} \leq 100$. The actual number of iterations needed to solve the corresponding linear system exactly depends on the values of the iteration parameters τ and δ. In this regard, it appears that setting $\delta = 1/Re$ contradicts Theorem 3.1, in view of the definition of $\mathbf{A}_s$ in (5.6). In practice, for a given Re and h, one can select $\delta > 1/Re^2$ such that the algorithm still remains stable yet shows an improved performance of the linear solves. The key here is not to select a δ which is

Table 4: Errors and nonlinear iteration numbers for $Re = 1,000$
for the implicit and explicit methods.

h	Error (p)	Error $(\mathbf{u}_1)$	Error $(\mathbf{u}_2)$	Implicit	Explicit
1/8***	0.30	8.13e-4	8.82e-3	33	****
1/16***	7.74e-2	3.57e-4	5.40e-3	38	****
1/32***	1.85e-2	5.45e-5	2.99e-3	39	****

*** − 50,000 inner iterations were taken for each Picard iteration
in the implicit method because the inexact Uzawa algorithm could
not reduce the residual below 1.0e-15 after 50,000 iterations when
solving the nonsymmetric saddle point problem. The norm of the
residual was on the order of 1.0e-11 after the first few nonlinear
iterations and less than 5.0e-15 towards the last Picard iterations.

**** − the algorithm broke down.

"too far away" from the safe zone. Indeed, setting $\delta = \sqrt{1/Re}$ resulted in a divergent
linear solver during the Picard iteration which caused the whole solution process to
break down. Also, as $h \to 0$, the method becomes more sensitive with respect to
deviations from the hypothesis of Theorem 3.1. For example, the case of $h = 1/128$ and
$Re = 1,000$ in our second numerical test described below required $\delta = 0.0001$ to remain
stable and broke down if $\delta = 0.001$. It is possible to tune up the parameters δ and τ
for fixed ν and h so that the number of inner iterations is minimized. We, however, did
not pursue this issue in the numerical experiments presented here.

The implicit algorithm is well suited for calculations on finite precision computers
(double precision recommended). On the other hand, the explicit method is a rea-
sonable approach to solving Navier-Stokes problems only for low Reynolds numbers
($Re = 1, 10$). It is a quite efficient algorithm for such flow problems, outperforming
the implicit method by a factor of 10 to 1 or more. However, the stability of this al-
gorithm deteriorates very fast as Re increases and the method becomes unstable for
$Re = 100$ and $Re = 1,000$. The case of $Re = 1,000$ is a very difficult computational
problem which could only be solved by performing a large number of inner iterations
for each Picard iteration. Clearly, the properties of $\mathbf{A}$ in this case are dominated by
its skew-symmetric part. This in turn means that $\mathbf{Q}_A^{-1}$ is a poor approximation of
$\mathbf{A}^{-1}$. Nevertheless, the implicit method converged to the analytic solution branch for
all values of h, showing the proper asymptotic behavior of the error. An efficient precon-
ditioned iterative method for approximating the inverse of the strongly nonsymmetric
operator $\mathbf{A}$ combined with the multistep algorithm from Section 4 could result in a
better method for solving the steady-state Navier-Stokes equations with high Reynolds
numbers.

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
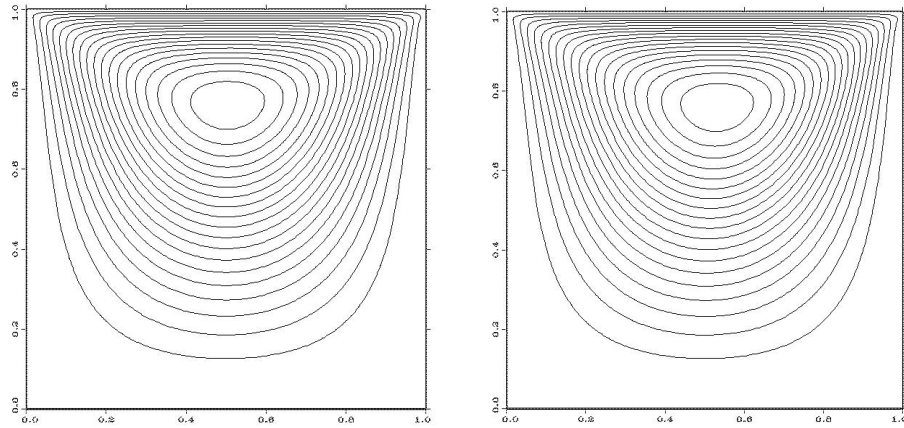are independent of time, the flow in this case limits to a steady-state which is modeled

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

by (5.1) with corresponding changes in the boundary conditions (5.1c) In particular, the solution **u** on the boundary is zero everywhere except on the boundary segment $y = 1$, where $\mathbf{u} = (v, 0)$ with v given. The rescaled pressure form of these equations is used here (cf. Remark 5.2).
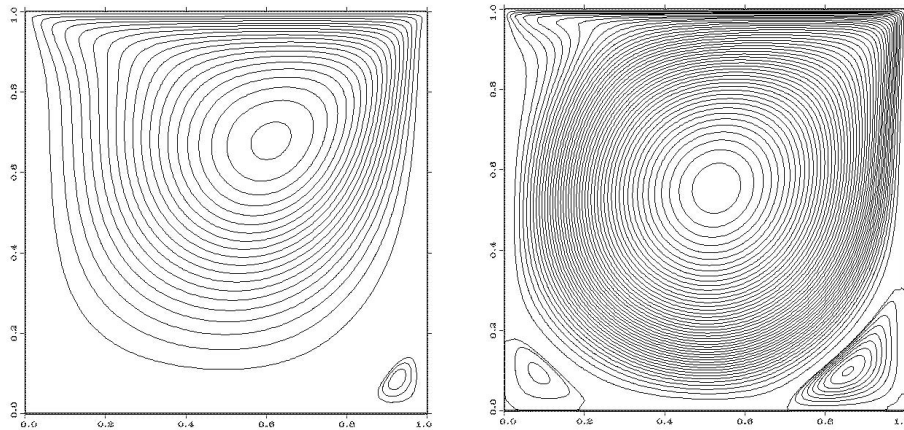

Figure 3: Streamlines for $h = 1/64$, $v = 1$, $Re = 100$ (left); $Re = 1,000$ (right).

To discretize this flow problem similarly to (5.4), we set $\mathbf{u} = \mathbf{u}_0 + \hat{\mathbf{u}}$, where $\mathbf{u}_0$ vanishes on the boundary of Ω and $\hat{\mathbf{u}}$ is a known function which satisfies the Dirichlet boundary conditions of **u**. 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 $\nabla \cdot \hat{X} = 0$.

Next, the implicit Picard iteration for this nonlinear problem is given by the following. Let $\hat{X}$ be as defined above. Then, given an initial guess X^0, we compute (X_0^i, Y^i), for $i = 1, 2, ...,$ as the solution of the linear system

$$D(X_0^i, V) + Re\,\tilde{b}(X^{i-1}, X_0^i, V) + (Y^i, \nabla \cdot V) = Re\,(\mathbf{f}, V) - D(\hat{X}, V) - Re\,\tilde{b}(X^{i-1}, \hat{X}, V),$$
$$(\nabla \cdot X_0^i, W) = 0,$$

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

The streamlines of the velocity field X computed using this algorithm for a wide range of Reynolds numbers are shown in Figures 2–3. The effect of the Reynolds number on the flow pattern is clearly seen there. The flow for low Reynolds numbers (see Fig. 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 Fig. 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 Fig. 3, the case of $Re = 1,000$).

Again, the case of $Re = 1,000$ 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 $Re = 1,000$ 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. An advantage of this method is that it solves the discrete system (5.4) without the need for additional stabilization terms in contrast to the class of penalty algorithms (cf. [11], [20]). The typical penalty methods add stabilization terms to (5.4). The bigger these terms are, the easier it is to solve the corresponding system. However, these stabilization terms change the discrete equations that one solves. In particular, their presence effectively reduces the Reynolds number for the corresponding flow causing different flow behavior to be computed. On the other hand, such a problem does not exists with the implicit method because it does not need any additional stabilization terms. The convergence of the

linear iteration at each Picard iteration is guaranteed only by the appropriate scaling of $\mathbf{Q}_A$ and the appropriate choice of the parameters δ and τ.

References

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

[2] J.H. Bramble and J.E. Pasciak. Iterative Techniques for Time Dependent Stokes Problems. *Inter. Jour. Computers and Math. with Applic.* (to appear).

[3] 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.

[4] 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.

[5] J.H. Bramble, Z. Leyk, and J.E. Pasciak. Iterative schemes for non-symmetric and indefinite elliptic boundary value problems. *Math. Comp.*, 60:1–22, 1993.

[6] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New York, 1991.

[7] X.-C. Cai. An additive Schwarz algorithm for nonselfadjoint elliptic equations. In T. Chan, R. Glowinski, J. Periaux, and O. Widlund, editors, *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*. SIAM, Phil. PA, 1990.

[8] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, New York, 1978.

[9] 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.

[10] H. Elman and D. Silvester. Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations. Technical Report CS-TR-3283, Univ. Maryland, College Park, June 1994.

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

[12] W. Hackbush. *Multi-Grid Methods and Applications*. Springer-Verlag, Berlin, 1985.

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

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

[15] 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.

[16] J.L. Lions and E. Magenes. *Problèmes aux Limites non Homogènes et Applications*, volume 1. Dunod, Paris, 1968.

[17] 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.

[18] 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.

[19] 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.

[20] R. Temam. *Navier-Stokes Equations*. North-Holland Publishing Co., New York, 1977.

[21] J. Xu. Two-grid dicretization techniques. *SIAM J. Numer. Anal.*, 33:1759–1777, 1996.