ANALYSIS OF THE INEXACT UZAWA ALGORITHM FOR SADDLE POINT PROBLEMS*

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

Abstract. In this paper, we consider the so-called "inexact Uzawa" algorithm for iteratively solving block saddle point problems. Such saddle point problems arise, for example, in finite element and finite difference discretizations of Stokes equations, the equations of elasticity and mixed finite element discretization of second order problems. We consider both the linear and nonlinear variants of the inexact Uzawa algorithm. We show that the linear method always converges as long as the preconditioners defining the algorithm are properly scaled. Bounds for the rate of convergence are provided in terms of the rate of convergence for the preconditioned Uzawa algorithm and the reduction factor corresponding to the preconditioner for the upper left hand block. In the nonlinear case, the inexact Uzawa algorithm is shown to converge provided that the nonlinear process approximating the inverse of the upper left hand block is of sufficient accuracy. Bounds for the nonlinear iteration are given in terms of this accuracy parameter and the rate of convergence of the preconditioned Uzawa algorithm. Applications to the Stokes equations and mixed finite element discretization of second order elliptic problems are discussed and finally, the results of numerical experiments involving the algorithms are presented.

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

AMS(MOS) subject classifications. 65N30, 65F10

1. Introduction.

This paper provides a new analysis for the inexact Uzawa method applied to the solution of saddle point systems which arise in the discretization of various systems of partial differential equations. Such systems typically are obtained when "multiplier" or mixed discretization techniques are employed. Examples of these include the discrete equations which result from approximation of elasticity problems, Stokes equations and sometimes linearizations of Navier–Stokes equations [4], [14], [15], [16]. In addition, these systems result from Lagrange multiplier [2], [3], [24] and mixed formulations of second order elliptic problems [8], [21], [24].

We shall consider iterative solution of an abstract saddle point problem. 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

* This manuscript has been authored 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-9007185 and by the U.S. Army Research Office through the Mathematical Sciences Institute, Cornell University.

† Department of Mathematics, Cornell University, Ithaca, New York 14853 and Department of Mathematics, Texas A&M University, College Station, TX 77843.

‡ Brookhaven National Laboratory, Upton, NY 11973.

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

the type of functions appearing. We consider the abstract saddle point problem:

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

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, symmetric, and positive definite operator. In addition, the linear map $\mathbf{B}^T : H_2 \mapsto H_1$ is the adjoint of $\mathbf{B}$: $H_1 \mapsto H_2$. Applying block elimination to (1.1) yeilds

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

Clearly, $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ is symmetric and nonnegative and a straightforward computation shows that

$$(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T V, V) = \sup_{U \in H_1} \frac{(V, \mathbf{B}U)^2}{(\mathbf{A}U, U)}. \tag{1.3}$$

Consequently, a necessary and sufficient condition for the unique solvability of (1.1) is that the Ladyzhenskaya–Babuška–Brezzi condition hold, i.e.

$$\sup_{U \in H_1} \frac{(V, \mathbf{B}U)^2}{(\mathbf{A}U, U)} \geq c_0 \|V\|^2 \quad \text{for all } V \in H_2\,, \tag{1.4}$$

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

One could iteratively solve (1.2) for Y by conjugate gradient (or preconditioned conjugate gradient) iteration [12]. Then X is obtained by $X = \mathbf{A}^{-1}(F - \mathbf{B}^T Y)$. The Uzawa method [1] is a particular implementation of a linear iterative method for solving (1.2). One common problem with the methods just described is that they require the evaluation of the action of the operator $\mathbf{A}^{-1}$ in each step of the iteration. For many applications, this operation is expensive and is also implemented as an iteration. The inexact Uzawa methods replace the exact inverse in the Uzawa algorithm by an "incomplete" or "approximate" evaluation of $\mathbf{A}^{-1}$. These algorithms are defined in Section 2 and 4. They were also studied in [11].

There are other general iterative techniques for solving saddle point problems of the form of (1.1), e.g., [5], [26]. In [5], a preconditioner for $\mathbf{A}$ is introduced and system (1.1) is reformulated as a well conditioned symmetric and positive definite algebraic system which may be solved efficiently by applying the conjugate gradient algorithm. In [26], the authors consider the convergence properties when the minimal residual algorithm is applied to a more direct preconditioned reformulation of (1.1). Both of the above mentioned techniques incorporate preconditioning and avoid the inversion of $\mathbf{A}$.

There is also a variety of application specific techniques that depend strongly on the particular approximation spaces, geometry of the domain etc. In the case of the mixed approximation of second order problems, those include domain decomposition techniques [17], a reduction technique involving the use of additional Lagrange multipliers [9], as well as an indefinite preconditioner [13].

The inexact Uzawa algorithms are of interest because they are simple and have minimal computer memory requirements. This could be important in large scale scientific applications implemented for todays computing architectures. In addition, an Uzawa algorithm implemented as a double iteration can be transformed trivially into an inexact Uzawa algorithm. It is not surprising that the inexact Uzawa methods are widely used in the engineering community.

In this paper we present new estimates for the inexact Uzawa algorithm both in the linear and nonlinear case. In the former case, the evaluation of $\mathbf{A}^{-1}$ is replaced by the inverse of a linear preconditioner. Theorem 3.1 shows that the resulting algorithm always converges and gives bounds on the rate of convergence provided that the preconditioner is properly scaled. The inexact Uzawa algorithm in the nonlinear case replaces the evaluation of $\mathbf{A}^{-1}$ by some approximate nonlinear process. Theorem 4.1 shows that the resulting algorithm converges provided that the nonlinear approximation to $\mathbf{A}^{-1}$ is suitably accurate. More restrictive results for variants of the inexact Uzawa algorithms have already appeared in the literature [11], [23].

The outline of the remainder of the paper is as follows. In Section 2, we define and motivate the linear version of the inexact Uzawa algorithm. Section 3 provides an analysis of this algorithm. In Section 4, the nonlinear version of the inexact Uzawa algorithm is defined and analyzed. Section 5 discusses a model application to the Stokes problem while Section 6 considers a model application to a mixed finite element discretization of a second order problem boundary value problem. Finally, the results of numerical experiments involving the inexact Uzawa algorithms are given in Section 7. A comparison with some other methods is presented as well.

2. The abstract inexact Uzawa algorithm. In this section, we define the inexact Uzawa method when linear preconditioners are used. This algorithm is motivated by first considering the Uzawa iteration [1] which can be defined as follows.

ALGORITHM 2.1 (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*

$$\begin{aligned} X_{i+1} &= X_i + \mathbf{A}^{-1}\left(F - (\mathbf{A}X_i + \mathbf{B}^T Y_i)\right), \\ Y_{i+1} &= Y_i + \tau(\mathbf{B}X_{i+1} - G), \end{aligned} \tag{2.1}$$

with τ a given real number.

Let $E_i^Y = Y - Y_i$ be the iteration error generated by the above method. It is easy to show that

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

Let c_1 denote the largest eigenvalue of $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$. Then, Y_i converges to Y if τ is chosen such that

$$\rho = \max(1 - c_0\tau, c_1\tau - 1) < 1.$$

In this case, X_i and Y_i converge respectively to X and Y with a rate of convergence bounded by ρ^i.

One problem with the above method is that it may converge slowly if $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ is not well conditioned. Thus, it is natural to introduce a preconditioner $\mathbf{Q}_B : H_2 \mapsto H_2$. We assume that $\mathbf{Q}_B$ is linear, symmetric and positive definite and define the preconditioned Uzawa algorithm as follows.

ALGORITHM 2.2 (PRECONDITIONED 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*

$$\begin{aligned} X_{i+1} &= X_i + \mathbf{A}^{-1}\left(F - (\mathbf{A}X_i + \mathbf{B}^T Y_i)\right), \\ Y_{i+1} &= Y_i + \mathbf{Q}_B^{-1}(\mathbf{B}X_{i+1} - G). \end{aligned} \tag{2.2}$$

For convenience of notation, we have absorbed the parameter τ into the preconditioner $\mathbf{Q}_B$. Accordingly, we assume that $\mathbf{Q}_B$ is scaled so that

$$(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T W, W) \leq (\mathbf{Q}_B W, W) \quad \text{for all } W \in H_2. \tag{2.3}$$

Note that since $\mathbf{Q}_B$ is positive definite, it follows that

$$(1-\gamma)(\mathbf{Q}_B W, W) \leq (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T W, W) \quad \text{for all } W \in H_2, \tag{2.4}$$

holds for some γ in the interval $[0, 1)$. In practice, effective preconditioners satisfy (2.4) with γ bounded away from one.

If $E_i^Y = Y - Y_i$ where Y_i is generated by (2.2) then

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

Clearly, $\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ is symmetric with respect to the inner product

$$< V, W >= (\mathbf{Q}_B V, W) \quad \text{for all } V, W \in H_2.$$

Let $\|\cdot\|_{Q_B}$ denote the corresponding norm

$$\|W\|_{Q_B} =< W, W >^{1/2}.$$

Then by (2.3) and (2.4),

$$\left\|E_i^Y\right\|_{Q_B} \leq \gamma^i \left\|E_0^Y\right\|_{Q_B}.$$

Here and in the sequel, for a symmetric and positive definite linear operator $\mathbf{L}$ on H_j, $j = 1, 2$, $\|\cdot\|_L$ will denote the norm $(\mathbf{L}\cdot, \cdot)^{1/2}$.

One problem with the above algorithms is that they require the computation of the action of the operator $\mathbf{A}^{-1}$ at each step of the iteration. For many of the applications, this is an expensive operation which is also done iteratively. This leads to a two level iteration, an inner iteration for computing the action of $\mathbf{A}^{-1}$ coupled with the outer Uzawa iteration (2.1) or (2.2). The inexact Uzawa method replaces the action of $\mathbf{A}^{-1}$ by a preconditioner. A preconditioner $\mathbf{Q}_A$ is a linear operator $\mathbf{Q}_A : H_1 \mapsto H_1$ which is symmetric and positive definite. In practice, good preconditioners are relatively cheap

to invert. For example, the computational cost for one evaluation of $\mathbf{Q}_A^{-1}$ should be comparable with the cost of evaluating the action of $\mathbf{A}$ (not $\mathbf{A}^{-1}$). The inexact Uzawa algorithm is then given as follows (this algorithm was also studied in [11]).

ALGORITHM 2.3 (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*

$$\begin{aligned} X_{i+1} &= X_i + \mathbf{Q}_A^{-1}\left(F - \left(\mathbf{A}X_i + \mathbf{B}^T Y_i\right)\right), \\ Y_{i+1} &= Y_i + \mathbf{Q}_B^{-1}(\mathbf{B}X_{i+1} - G). \end{aligned} \tag{2.5}$$

One step of the inexact Uzawa algorithm involves an evaluation of each of the operators, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{B}^T$, $\mathbf{Q}_A^{-1}$ and $\mathbf{Q}_B^{-1}$. In contrast to Krylov space minimization algorithms such as conjugate residual, there are no discrete inner products involved in the iteration. This makes this algorithm very well suited for implementation on contemporary massively parallel computer architectures.

3. Analysis of the inexact Uzawa algorithm. In this section, we investigate the stability and convergence rate of the inexact Uzawa algorithm defined above. The main theorem will show that the inexact Uzawa algorithm will always converge provided that the preconditioners are properly scaled. By this we mean that (2.3) holds and that

$$(\mathbf{A}W, W) < (\mathbf{Q}_A W, W) \tag{3.1}$$

for all $W \in H_1$ with $W \neq 0$. The strict inequality above will be replaced by

$$(\mathbf{A}W, W) \leq (\mathbf{Q}_A W, W) \quad \text{for all } W \in H_1, \tag{3.2}$$

in a subsequent corollary.

Bounds for the rates of iterative convergence will be provided in terms two natural parameters. The first parameter has already been defined and is the convergence factor γ (see (2.4)) for the preconditioned Uzawa algorithm. The second parameter is the rate δ at which the preconditioned iteration

$$U_{i+1} = U_i + \mathbf{Q}_A^{-1}(W - \mathbf{A}U_i)$$

converges to the solution of

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

If $E_i^A = U - U_i$ then

$$E_{i+1}^A = (\mathbf{I} - \mathbf{Q}_A^{-1}\mathbf{A})E_i^A.$$

Clearly $\mathbf{Q}_A^{-1}\mathbf{A}$ is a symmetric operator with respect to the inner product $(\mathbf{Q}_A \cdot, \cdot)$ and hence the convergence rate δ is the largest eigenvalue of $\mathbf{I} - \mathbf{Q}_A^{-1}\mathbf{A}$. Alternatively, δ is the smallest number for which the inequality

$$(1 - \delta)(\mathbf{Q}_A W, W) \leq (\mathbf{A}W, W) \quad \text{for all } W \in H_1 \tag{3.3}$$

is satisfied. It will sometimes be convenient to rewrite (3.3) as

$$((\mathbf{Q}_A - \mathbf{A})W, W) \leq \delta(\mathbf{Q}_A W, W) \quad \text{for all } W \in H_1. \tag{3.4}$$

The first convergence estimate will be provided in terms of a norm on $H_1 \times H_2$ which we shall now define. Consider the bilinear form on $H_1 \times H_2$ given by

$$\left[\begin{pmatrix} U \\ V \end{pmatrix}, \begin{pmatrix} R \\ S \end{pmatrix}\right] = ((\mathbf{Q}_A - \mathbf{A})U, R) + (\mathbf{Q}_B V, S). \tag{3.5}$$

By (3.1), $[\cdot, \cdot]$ generates a norm on $H_1 \times H_2$ which we shall denote by

$$|\!|\!| T |\!|\!| = [T, T]^{1/2}, \quad \text{for all } T \in H_1 \times H_2.$$

We can now state the main result of this section.

THEOREM 1. *ssume that (2.3) and (3.1) hold and that γ and δ satisfy (2.4) and (3.3) respectively. Let $\{X, Y\}$ be the solution pair for (1.1), $\{X_i, Y_i\}$ be defined by the inexact Uzawa algorithm and set*

$$e_i = \begin{pmatrix} X - X_i \\ Y - Y_i \end{pmatrix}.$$

Then, for $i = 1, 2, \ldots,$

$$|\!|\!| e_i |\!|\!| \leq \rho^i |\!|\!| e_0 |\!|\!|, \tag{3.6}$$

where

$$\rho = \frac{\gamma(1-\delta) + \sqrt{\gamma^2(1-\delta)^2 + 4\delta}}{2}. \tag{3.7}$$

Remark 3.1. It is elementary to see that

$$\rho \leq 1 - \frac{1}{2}(1-\gamma)(1-\delta).$$

Thus the inexact Uzawa method converges if (2.3) and (3.1) hold. As expected, the convergence rate deteriorates as either γ or δ approach one. In addition, if δ tends to zero (and thus, $\mathbf{Q}_A$ tends to $\mathbf{A}$) then ρ (defined by (3.7)) tends to γ, the convergence rate of the preconditioned Uzawa algorithm.

Proof.[Theorem 1] We first derive a relationship between the errors e_{i+1} and e_i. The components of the corresponding errors are denoted by $E_i^X = X - X_i$ and $E_i^Y = Y - Y_i$. From (1.1) and (2.5) we see that the errors satisfy the recurrence

$$\begin{aligned} E_{i+1}^X &= \left(\mathbf{I} - \mathbf{Q}_A^{-1}\mathbf{A}\right) E_i^X - \mathbf{Q}_A^{-1}\mathbf{B}^T E_i^Y, \\ E_{i+1}^Y &= E_i^Y + \mathbf{Q}_B^{-1}\mathbf{B} E_{i+1}^X. \end{aligned} \tag{3.8}$$

Replacing E_{i+1}^X in the second equation with its expression from the first gives

$$\begin{aligned}\begin{pmatrix} E_{i+1}^X \\ E_{i+1}^Y \end{pmatrix} &= \begin{pmatrix} (\mathbf{I}-\mathbf{Q}_A^{-1}\mathbf{A}) & -\mathbf{Q}_A^{-1}\mathbf{B}^T \\ \mathbf{Q}_B^{-1}\mathbf{B}\left(\mathbf{I}-\mathbf{Q}_A^{-1}\mathbf{A}\right) & (\mathbf{I}-\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T) \end{pmatrix} \begin{pmatrix} E_i^X \\ E_i^Y \end{pmatrix} \\ &\equiv \mathcal{M} \begin{pmatrix} E_i^X \\ E_i^Y \end{pmatrix}. \end{aligned} \tag{3.9}$$

This can be rewritten as

$$e_{i+1} = \mathcal{M} e_i. \tag{3.10}$$

The proof of the theorem will be complete if we can show that the operator norm

$$[\![\mathcal{M}]\!] = \sup_{x \in H_1 \times H_2} \frac{[\![\mathcal{M}x]\!]}{[\![x]\!]}$$

is bounded by ρ given by (3.7).

The operator $\mathcal{M}$ can be written in the form

$$\begin{aligned}\mathcal{M} &= \begin{pmatrix} -\mathbf{I} & 0 \\ 0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} -(\mathbf{I}-\mathbf{Q}_A^{-1}\mathbf{A}) & \mathbf{Q}_A^{-1}\mathbf{B}^T \\ \mathbf{Q}_B^{-1}\mathbf{B}\left(\mathbf{I}-\mathbf{Q}_A^{-1}\mathbf{A}\right) & (\mathbf{I}-\mathbf{Q}_B^{-1}\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T) \end{pmatrix} \\ &\equiv \mathcal{E}\mathcal{M}_1. \end{aligned}$$

It is straightforward to check that both $\mathcal{E}$ and $\mathcal{M}_1$ are symmetric in the $[\cdot,\cdot]$-inner product. Let $\mathcal{M}^*$ denote the adjoint of $\mathcal{M}$ with respect to $[\cdot,\cdot]$. Then we have

$$\mathcal{M}^* = (\mathcal{E}\mathcal{M}_1)^* = \mathcal{M}_1\mathcal{E}$$

and

$$\mathcal{M}^*\mathcal{M} = \mathcal{M}_1\mathcal{E}^2\mathcal{M}_1 = \mathcal{M}_1^2.$$

Consequently,

$$\begin{aligned}[\![\mathcal{M}]\!]^2 &= \sup_{x \in H_1 \times H_2} \frac{[\mathcal{M}x, \mathcal{M}x]}{[x,x]} = \sup_{x \in H_1 \times H_2} \frac{[\mathcal{M}^*\mathcal{M}x, x]}{[x,x]} \\ &= \sup_{x \in H_1 \times H_2} \frac{[\mathcal{M}_1^2 x, x]}{[x,x]} = \sup_{\lambda_i \in \sigma(\mathcal{M}_1)} |\lambda_i|^2. \end{aligned}$$

Therefore, to estimate the norm of $\mathcal{M}$, it suffices to bound the spectrum $\sigma(\mathcal{M}_1)$ of $\mathcal{M}_1$. Since $\mathcal{M}_1$ is symmetric with respect to the $[\cdot,\cdot]$ inner product, its eigenvalues are real. We shall bound the positive and negative eigenvalues of $\mathcal{M}_1$ separately.

We first provide a bound for the positive eigenvalues of $\mathcal{M}_1$. The operator $\mathbf{I}-\mathbf{Q}_A^{-1}\mathbf{A}$ is symmetric with respect to the inner product $((\mathbf{Q}_A-\mathbf{A})\cdot,\cdot)$. Moreover, it follows from (3.1) that it is positive definite and its positive square root is well defined. Let

$$\mathcal{D}=\begin{pmatrix}\delta^{-1/2}(\mathbf{I}-\mathbf{Q}_A^{-1}\mathbf{A})^{1/2} & 0\\ 0 & \mathbf{I}\end{pmatrix}.$$

It follows from (3.1) that $\mathcal{D}$ is invertible and from (3.3) that

$$[\![\mathcal{D}x]\!]\le[\![x]\!] \quad \text{for all } x\in H_1\times H_2. \tag{3.11}$$

Let $\mathcal{N}=\mathcal{D}^{-1}\mathcal{M}_1\mathcal{D}^{-1}$. Then

$$\mathcal{N}=\begin{pmatrix}-\delta\mathbf{I} & \delta^{1/2}\mathbf{L}\\ \delta^{1/2}\mathbf{L}^* & (\mathbf{I}-\mathbf{L}^*\mathbf{L})\end{pmatrix} \tag{3.12}$$

where $\mathbf{L}=(\mathbf{I}-\mathbf{Q}_A^{-1}\mathbf{A})^{-1/2}\mathbf{Q}_A^{-1}\mathbf{B}^T$ and $\mathbf{L}^*=\mathbf{Q}_B^{-1}\mathbf{B}(\mathbf{I}-\mathbf{Q}_A^{-1}\mathbf{A})^{1/2}$.

The largest eigenvalue λ_m of $\mathcal{M}_1$ satisfies

$$\begin{aligned}\lambda_m &= \sup_{x\in H_1\times H_2}\frac{[\mathcal{M}_1x,x]}{[x,x]}=\sup_{x\in H_1\times H_2}\frac{[\mathcal{N}\mathcal{D}x,\mathcal{D}x]}{[x,x]}\\ &= \sup_{x\in H_1\times H_2}\frac{[\mathcal{N}\mathcal{D}x,\mathcal{D}x]\,[\mathcal{D}x,\mathcal{D}x]}{[\mathcal{D}x,\mathcal{D}x]\,[x,x]}\le\sup_{y\in H_1\times H_2}\frac{[\mathcal{N}y,y]}{[y,y]}.\end{aligned}$$

We used (3.11) for the last inequality above. Since both $\mathcal{D}$ and $\mathcal{M}_1$ are symmetric with respect to $[\cdot,\cdot]$, it follows that $\mathcal{N}$ is also. Consequently, λ_m is bounded by the largest eigenvalue of $\mathcal{N}$.

Let λ be a nonnegative eigenvalue of $\mathcal{N}$ with corresponding eigenvector $\{\psi_1,\psi_2\}$, i.e.,

$$\begin{aligned}-\delta\psi_1+\delta^{1/2}\mathbf{L}\psi_2 &= \lambda\psi_1,\\ \delta^{1/2}\mathbf{L}^*\psi_1+(\mathbf{I}-\mathbf{L}^*\mathbf{L})\psi_2 &= \lambda\psi_2.\end{aligned} \tag{3.13}$$

Eliminating ψ_1 in the above equations gives

$$-\lambda\mathbf{L}^*\mathbf{L}\psi_2=(\lambda+\delta)(\lambda-1)\psi_2$$

and hence

$$-\lambda<\mathbf{L}^*\mathbf{L}\psi_2,\psi_2>=(\lambda+\delta)(\lambda-1)<\psi_2,\psi_2>. \tag{3.14}$$

By (3.3) and (2.4), it follows that

$$\begin{aligned}<\mathbf{L}^*\mathbf{L}\psi_2,\psi_2> &= (\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T\psi_2,\psi_2)\ge(1-\delta)(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T\psi_2,\psi_2)\\ &\ge(1-\delta)(1-\gamma)<\psi_2,\psi_2>.\end{aligned} \tag{3.15}$$

Since $\delta > 0$ and λ is nonnegative, we see from the first equation in (3.13) that if $\psi_2 = 0$ then $\psi_1 = 0$. Consequently, ψ_2 is not equal to zero. Thus, from (3.14) and (3.15), we get

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

from which it follows that $\lambda \leq \rho$ where ρ is given by (3.7). This provides the desired bound for the positive eigenvalues of $\mathcal{M}_1$.

We next estimate the negative eigenvalues of $\mathcal{M}_1$. Let λ be a negative eigenvalue of $\mathcal{M}_1$ with corresponding eigenvector (ψ_1, ψ_2), i.e.,

$$\begin{aligned} -\left(\mathbf{I} - \mathbf{Q}_A^{-1}\mathbf{A}\right)\psi_1 + \mathbf{Q}_A^{-1}\mathbf{B}^T\psi_2 &= \lambda\psi_1, \\ \mathbf{Q}_B^{-1}\mathbf{B}\left(\mathbf{I} - \mathbf{Q}_A^{-1}\mathbf{A}\right)\psi_1 + \left(\mathbf{I} - \mathbf{Q}_B^{-1}\mathbf{B}\mathbf{Q}_A^{-1}\mathbf{B}^T\right)\psi_2 &= \lambda\psi_2. \end{aligned} \tag{3.16}$$

The first equation in (3.16) together with (2.4) imply that if $\psi_1 = 0$ then $\psi_2 = 0$. Consequently, any eigenvector must have a nonzero component ψ_1.

Multiplying the first equation of (3.16) by $\mathbf{Q}_B^{-1}\mathbf{B}$ from the left and adding it to the second one yields

$$(1-\lambda)\psi_2 = \lambda\mathbf{Q}_B^{-1}\mathbf{B}\psi_1. \tag{3.17}$$

Substituting (3.17) into the first equation of (3.16) and taking an inner product with $\mathbf{Q}_A\psi_1$ gives

$$-((1-\lambda)((1+\lambda)\mathbf{Q}_A - \mathbf{A})\psi_1, \psi_1) + \lambda(\mathbf{Q}_B^{-1}\mathbf{B}\psi_1, \mathbf{B}\psi_1) = 0,$$

which we rewrite as

$$\lambda(\mathbf{Q}_B^{-1}\mathbf{B}\psi_1, \mathbf{B}\psi_1) = ((1-\lambda^2)(\mathbf{Q}_A\psi_1, \psi_1) - (1-\lambda)(\mathbf{A}\psi_1, \psi_1). \tag{3.18}$$

For any $V \in H_1$,

$$\begin{aligned} (\mathbf{Q}_B^{-1}\mathbf{B}V, \mathbf{B}V) &= \sup_{W\in H_2} \frac{(V, \mathbf{B}^T W)^2}{(\mathbf{Q}_B W, W)} = \sup_{W\in H_2} \frac{(\mathbf{A}^{1/2}V, \mathbf{A}^{-1/2}\mathbf{B}^T W)^2}{(\mathbf{Q}_B W, W)} \\ &\leq \sup_{W\in H_2} \frac{(\mathbf{A}V, V)(\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T W, W)}{(\mathbf{Q}_B W, W)} \leq (\mathbf{A}V, V). \end{aligned} \tag{3.19}$$

For the last inequality above we used (2.3). Applying (3.19) to the left hand side of (3.18) and (3.4) on the right hand side of (3.18) gives

$$\lambda(\mathbf{A}\psi_1, \psi_1) \leq (\delta - \lambda^2)(\mathbf{Q}_A\psi_1, \psi_1) + \lambda(\mathbf{A}\psi_1, \psi_1)$$

or

$$0 \leq (\delta - \lambda^2)(\mathbf{Q}_A\psi_1, \psi_1).$$

This implies that $\lambda \geq -\sqrt{\delta}$ since ψ_1 is nonzero. It is elementary to check that $\sqrt{\delta} \leq \rho$ where ρ is defined by (3.7). This completes the proof of the theorem.

The proof of Theorem 3.1 depended on (3.1) so that the inner product $[\cdot,\cdot]$ induced a norm. The next result shows that the inexact Uzawa method converges even when only (3.2) is assumed. It also provides an estimate for the error $E_i^X = X - X_i$ in a more natural norm.

COROLLARY 3.1. *Assume that (2.3) and (3.2) hold and that γ and δ satisfy (2.4) and (3.3) respectively. Let $\{X, Y\}$ be the solution pair for (1.1), let $\{X_i, Y_i\}$ be defined by the inexact Uzawa algorithm and set $E_i^X = X - X_i$ and $E_i^Y = Y - Y_i$. Then*

$$(\mathbf{Q}_B E_i^Y, E_i^Y)^{1/2} \leq \rho^i [\![e_0]\!] \tag{3.20}$$

where ρ is given by (3.7). In addition,

$$(\mathbf{A} E_i^X, E_i^X)^{1/2} \leq \rho^{i-1} [\![e_0]\!]. \tag{3.21}$$

The above inequalities hold for $i = 1, 2, \ldots$.

Proof. Taking the $(\cdot,\cdot)$-inner product of the first equation of (3.8) with $\mathbf{Q}_A e_{i+1}^X$, applying the Schwarz inequality, and (2.3) gives

$$\begin{aligned}
(\mathbf{Q}_A E_i^X, E_i^X) &= ((\mathbf{Q}_A - \mathbf{A}) E_{i-1}^X, E_i^X) - (\mathbf{B}^T E_{i-1}^Y, E_i^X) \\
&\leq ((\mathbf{Q}_A - \mathbf{A}) E_{i-1}^X, E_{i-1}^X)^{1/2} ((\mathbf{Q}_A - \mathbf{A}) E_i^X, E_i^X)^{1/2} \\
&\quad + (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T E_{i-1}^Y, E_{i-1}^Y)^{1/2} (\mathbf{A} E_i^X, E_i^X)^{1/2} \\
&\leq (((\mathbf{Q}_A - \mathbf{A}) E_{i-1}^X, E_{i-1}^X) + \left\| E_{i-1}^Y \right\|_{Q_B}^2)^{1/2} (\mathbf{Q}_A E_i^X, E_i^X)^{1/2}.
\end{aligned}$$

Thus, applying (3.2) gives

$$(\mathbf{A} E_i^X, E_i^X) \leq (\mathbf{Q}_A E_i^X, E_i^X) \leq [\![e_{i-1}]\!]^2. \tag{3.22}$$

Let $\mathbf{Q}_{A,\epsilon} = \epsilon \mathbf{I} + \mathbf{Q}_A$ for $0 < \epsilon < 1 - \delta$. Then (3.1) holds for $\mathbf{Q}_{A,\epsilon}$ and by (3.3),

$$(1 - \delta_\epsilon)(\mathbf{Q}_{A,\epsilon} W, W) \leq (\mathbf{A} W, W) \quad \text{for all } W \in H_1 \tag{3.23}$$

for $\delta_\epsilon = \delta + \epsilon$. Fix $(X_0, Y_0) \in H_1 \times H_2$ and consider the sequence of iterates $\{X_{\epsilon,i}, X_{\epsilon,i}\}$ generated by the inexact Uzawa algorithm which replaces $\mathbf{Q}_A$ in the first equation of (2.5) by $\mathbf{Q}_{A,\epsilon}$. Applying Theorem 3.1 gives that the error

$$e_{\epsilon,i} = \begin{pmatrix} X - X_{\epsilon,i} \\ Y - X_{\epsilon,i} \end{pmatrix}$$

satisfies

$$[\![e_{\epsilon,i}]\!]_\epsilon \leq \rho_\epsilon^i [\![e_{\epsilon,0}]\!]_\epsilon \tag{3.24}$$

where $[\![\cdot]\!]_\epsilon = [\cdot,\cdot]_\epsilon^{1/2}$,

$$\left[\begin{pmatrix} U \\ V \end{pmatrix}, \begin{pmatrix} R \\ S \end{pmatrix}\right]_\epsilon = ((\mathbf{Q}_{A,\epsilon} - \mathbf{A})U, R) + (\mathbf{Q}_B V, S),$$

and

$$\rho_\epsilon = \frac{\gamma(1-\delta_\epsilon) + \sqrt{\gamma^2(1-\delta_\epsilon)^2 + 4\delta_\epsilon}}{2}.$$

Clearly,

$$\left\| E_{\epsilon,i}^X \right\|_{Q_B} \leq [\![e_{\epsilon,i}]\!]_\epsilon. \tag{3.25}$$

Inequality (3.20) results from combining (3.24) and (3.25) and taking the limit as ϵ tends to zero.

In a similar manner we prove (3.21). Taking the limit in (3.24) as ϵ tends to zero gives

$$[\![e_{i-1}]\!] \leq \rho^{i-1} [\![e_0]\!]. \tag{3.26}$$

Combining (3.22) and (3.26) gives (3.21) and completes the proof of the corollary.

Remark 3.2. More restrictive convergence results (in these norms) were obtained by Queck [23]. He proved a convergence result which required stronger conditions with respect to the scaling of $\mathbf{Q}_A$ and $\mathbf{Q}_B$. In particular, there are cases which fail to satisfy the hypothesis of the theory of [23] yet convergence is guaranteed by the corollary above. In addition, there are many cases when the convergence estimates given above are substantially better than those of [23].

4. Analysis of the nonlinear inexact Uzawa algorithm. As was pointed out in Section 2, the Uzawa algorithm is often implemented as a two level iterative process, an inner iteration for computing $\mathbf{A}^{-1}$ coupled with the outer Uzawa iteration (2.1) or (2.2). In this section we investigate the stability and convergence rate of an abstract inexact Uzawa algorithm where the computation of the action of $\mathbf{A}^{-1}$ is replaced with that of an approximation to $\mathbf{A}^{-1}$ which results from applying a nonlinear iterative process for inverting $\mathbf{A}$. Two examples of such approximations come from defining the approximate inverse by a preconditioned conjugate gradient iteration or the operator which results from the application of a multigrid cycling algorithm with a nonlinear smoother.

The nonlinear approximate inverse is described as a map $\Psi : H_1 \mapsto H_1$. For $\phi \in H_1$, $\Psi(\phi)$ is an "approximation" to the solution ξ of

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

We shall assume that our approximation satisfies

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

for some $\delta < 1$. As will be seen below, (4.2) is a reasonable assumption which is satisfied by the approximate inverse associated with the preconditioned conjugate gradient algorithm. It also can be shown that (4.2) holds under reasonable assumptions for approximate inverses defined by one sweep of a multigrid algorithm with conjugate gradient smoothing.

Perhaps the most natural example of a nonlinear approximate inverse is defined in terms of the preconditioned conjugate gradient procedure [22]. Let $\mathbf{Q}_A$ be a symmetric and positive definite operator on H_1 and consider applying n steps of the conjugate gradient algorithm preconditioned by $\mathbf{Q}_A$ to solve the problem (4.1) with a zero starting iterate. We define $\Psi(\phi) = \xi_n$ where ξ_n is the resulting approximation to ξ. The preconditioned conjugate gradient algorithm (PCG) provides the best approximation (with respect to the norm corresponding to the $(A\cdot,\cdot)$-inner product) to the solution ξ in the space

$$K_n \;=\; \text{span}\left\{\phi, \mathbf{Q}_A^{-1}\mathbf{A}\phi, \ldots, (\mathbf{Q}_A^{-1}\mathbf{A})^{n-1}\phi\right\}.$$

It is well known that this implies [6]

$$\|\xi_n - \mathbf{A}^{-1}\phi\|_A \leq \delta\|\phi\|_{A^{-1}} \quad \text{for all } \phi \in H_1\ , \tag{4.3}$$

where

$$\delta = \delta_n \leq \frac{1}{\cosh(n\cosh^{-1}\eta)}.$$

Here $\eta = (\kappa(\mathbf{Q}_A^{-1}\mathbf{A}) + 1)/(\kappa(\mathbf{Q}_A^{-1}\mathbf{A}) - 1)$ and $\kappa(\mathbf{Q}_A^{-1}\mathbf{A})$ is the condition number of $\mathbf{Q}_A^{-1}\mathbf{A}$. Note that δ_n is a decreasing function of n and δ_1 is less than one. Thus, (4.2) holds in the PCG example. In fact,

$$\delta_n \leq 2\left(\frac{\kappa(\mathbf{Q}_A^{-1}\mathbf{A})^{1/2} - 1}{\kappa(\mathbf{Q}_A^{-1}\mathbf{A})^{1/2} + 1}\right)^n.$$

Since δ_n tends to zero as n tends to infinity, it is possible to make δ_n as small as we want by taking a suitably large number PCG iterations.

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

ALGORITHM 4.1 (NONLINEAR 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*

$$\begin{aligned} X_{i+1} &= X_i + \Psi\left(F - \left(\mathbf{A}X_i + \mathbf{B}^T Y_i\right)\right), \\ Y_{i+1} &= Y_i + \mathbf{Q}_B^{-1}(\mathbf{B}X_{i+1} - G). \end{aligned} \tag{4.4}$$

Clearly, (4.4) reduces to the preconditioned Uzawa algorithm (2.2) if $\Psi(f) = A^{-1}f$ for all $f \in H_1$ and (4.4) reduces to the inexact Uzawa algorithm if Ψ is a linear operator $\mathbf{Q}_A^{-1}$.

We will provide bounds for the rate of convergence for the above algorithm in terms of two parameters, the convergence factor γ for the preconditioned Uzawa algorithm defined in (2.4) and the parameter δ of (4.2). The main result of this section provides a sufficient condition on δ for convergence of the nonlinear Uzawa algorithm and bounds for the resulting rate of convergence.

THEOREM 2. *ssume that (2.3) and (4.2) hold and that γ satisfies (2.4). Let $\{X, Y\}$ be the solution pair for (1.1) and $\{X_i, Y_i\}$ be defined by the nonlinear Uzawa algorithm (4.4). Then X_i and Y_i converge to X and Y respectively if*

$$\delta < \frac{1-\gamma}{3-\gamma}. \tag{4.5}$$

In this case the following inequalities hold:

$$\begin{aligned} \frac{\delta}{1+\delta}(\mathbf{A}E_i^X, E_i^X) &+ (\mathbf{Q}_B E_i^Y, E_i^Y) \\ &\leq \rho^{2i}\left(\frac{\delta}{1+\delta}(\mathbf{A}E_0^X, E_0^X) + (\mathbf{Q}_B E_0^Y, E_0^Y)\right) \end{aligned} \tag{4.6}$$

and

$$(\mathbf{A}E_i^X, E_i^X) \leq (1+\delta)(1+2\delta)\rho^{2i-2}\left(\frac{\delta}{1+\delta}(\mathbf{A}E_0^X, E_0^X) + (\mathbf{Q}_B E_0^Y, E_0^Y)\right) \tag{4.7}$$

where

$$\rho = \frac{2\delta + \gamma + \sqrt{(2\delta+\gamma)^2 + 4\delta(1-\gamma)}}{2}. \tag{4.8}$$

Remark 4.1. The result of Theorem 2 is somewhat weaker than the results obtained in Section 3 for the linear case due to the threshold condition (4.5) on δ. In the case of PCG, it is possible to take sufficiently many iterations n so that (4.5) holds for any fixed γ and $\kappa(\mathbf{Q}_A^{-1}\mathbf{A})$. In applications involving partial differential equations, γ and $\kappa(\mathbf{Q}_A^{-1}\mathbf{A})$ may depend on the discretization parameter h. If, however, $\kappa(\mathbf{Q}_A^{-1}\mathbf{A})$ can be bounded and γ can be bounded away from one independently of h then by Theorem 2, a fixed number (independent of h) of iterations of PCG are sufficient to guarantee convergence of the nonlinear Uzawa algorithm.

Remark 4.2. An analysis of (4.4) is given in [10] and [11] in the case of applications to Stokes problems. The sufficient condition for convergence derived there is that the iterate X_{i+1} satisfies

$$\|F - \mathbf{B}^T Y_i - \mathbf{A}X_{i+1}\| \leq \tau \|\mathbf{B}X_i - G\|_{Q_A^{-1}}, \tag{4.9}$$

where τ is independent of the mesh size. The above norms are not natural for procedures such as PCG and multigrid with nonlinear smoothing. PCG does not give rise to

monotone error behavior in the norm $\|\cdot\|$ even though convergence is guaranteed by the canonical bound (4.3),

$$\|F - \mathbf{B}^T Y_i - \mathbf{A} X_{i+1}\|_{A^{-1}} \leq \delta \|F - \mathbf{B}^T Y_i - \mathbf{A} X_i\|_{A^{-1}}$$

and equivalence of norms in finite dimensional spaces. Such norm equivalences depend on the mesh parameter h. A second problem with the requirement (4.9) is that the norm on the right hand side converges to zero as X_i converges to the solution X. This means that even thought τ is fixed independent of h, considerably more iterations of PCG may be required to satisfy (4.9) as the approximate solution converges.

Proof.[Theorem 2] We start by deriving norm inequalities involving the errors E_i^X and E_i^Y. As in (3.8),

$$\begin{aligned} E_{i+1}^X &= E_i^X - \Psi \left(\mathbf{A} E_i^X + \mathbf{B}^T E_i^Y \right), \\ E_{i+1}^Y &= E_i^Y + \mathbf{Q}_B^{-1} \mathbf{B} E_{i+1}^X. \end{aligned} \tag{4.10}$$

The first equation above can be rewritten

$$E_{i+1}^X = (\mathbf{A}^{-1} - \Psi) \left(\mathbf{A} E_i^X + \mathbf{B}^T E_i^Y \right) - \mathbf{A}^{-1} \mathbf{B}^T E_i^Y. \tag{4.11}$$

It follows from the triangle inequality, (4.2) and (2.3) that

$$\begin{aligned} \left\| E_{i+1}^X \right\|_A &\leq \delta \left(\left\| E_i^X \right\|_A + (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T E_i^Y, E_i^Y)^{1/2} \right) \\ &\quad + (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T E_i^Y, E_i^Y)^{1/2} \\ &\leq \delta \left\| E_i^X \right\|_A + (1+\delta) \left\| E_i^Y \right\|_{Q_B}. \end{aligned} \tag{4.12}$$

Using (4.11) in the second equation of (4.10), we obtain

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

Since $\mathbf{Q}_B^{-1} \mathbf{B} \mathbf{A}^{-1} \mathbf{B}^T$ is a symmetric operator in the $< \cdot, \cdot >$-inner product, it follows from (2.4) that

$$\left\| \left(\mathbf{I} - \mathbf{Q}_B^{-1} \mathbf{B} \mathbf{A}^{-1} \mathbf{B}^T \right) E_i^Y \right\|_{Q_B} \leq \gamma \left\| E_i^Y \right\|_{Q_B}.$$

Thus, by the triangle inequality, (2.3), (3.19) and (4.2),

$$\begin{aligned} \left\| E_{i+1}^Y \right\|_{Q_B} &\leq \gamma \left\| E_i^Y \right\|_{Q_B} + \left\| \mathbf{Q}_B^{-1} \mathbf{B} \left(\mathbf{A}^{-1} - \Psi \right) \left(\mathbf{A} E_i^X + \mathbf{B}^T E_i^Y \right) \right\|_{Q_B} \\ &\leq \gamma \left\| E_i^Y \right\|_{Q_B} + \left\| (\mathbf{A}^{-1} - \Psi) \left(\mathbf{A} E_i^X + \mathbf{B}^T E_i^Y \right) \right\|_A \\ &\leq \gamma \left\| E_i^Y \right\|_{Q_B} + \delta \left(\left\| E_i^X \right\|_A + (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T E_i^Y, E_i^Y)^{1/2} \right) \\ &\leq (\gamma + \delta) \left\| E_i^Y \right\|_{Q_B} + \delta \left\| E_i^X \right\|_A. \end{aligned} \tag{4.13}$$

Let us adopt the notation

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

for vectors of nonnegative numbers x_1, x_2, y_1, y_2 if $x_1 \leq x_2$ and $y_1 \leq y_2$. Repeated application of (4.12) and (4.13) gives

$$\begin{pmatrix} \left\| E_i^X \right\|_A \\ \left\| E_i^Y \right\|_{Q_B} \end{pmatrix} \leq M^i \begin{pmatrix} \left\| E_0^X \right\|_A \\ \left\| E_0^Y \right\|_{Q_B} \end{pmatrix} \tag{4.14}$$

where M is given by

$$M = \begin{pmatrix} \delta & 1+\delta \\ \delta & \gamma+\delta \end{pmatrix}.$$

We consider two dimensional Euclidean space with the inner product

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

A trivial computation shows that M is symmetric with respect to the $\lfloor \cdot, \cdot \rfloor$-inner product. It follows from (4.14) that

$$\begin{aligned} \frac{\delta}{1+\delta}(\mathbf{A}E_i^X, E_i^X) \quad + \quad & (\mathbf{Q}_B E_i^Y, E_i^Y) = \left\lfloor \begin{pmatrix} \left\| E_i^X \right\|_A \\ \left\| E_i^Y \right\|_{Q_B} \end{pmatrix}, \begin{pmatrix} \left\| E_i^X \right\|_A \\ \left\| E_i^Y \right\|_{Q_B} \end{pmatrix} \right\rfloor \\ \leq \quad & \left\lfloor M^i \begin{pmatrix} \left\| E_0^X \right\|_A \\ \left\| E_0^Y \right\|_{Q_B} \end{pmatrix}, M^i \begin{pmatrix} \left\| E_0^X \right\|_A \\ \left\| E_0^Y \right\|_{Q_B} \end{pmatrix} \right\rfloor \\ \leq \quad & \rho^{2i} \left(\frac{\delta}{1+\delta}(\mathbf{A}E_0^X, E_0^X) + (\mathbf{Q}_B E_0^Y, E_0^Y) \right) \end{aligned}$$

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

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

It is elementary to see that the spectral radius of M is equal to its positive eigenvalue which is given by (4.8).

Examining the expression for ρ given by (4.8) we see that ρ is an increasing function of δ for any fixed $\gamma \in [0, 1]$. Moreover, $\rho = 1$ for

$$\delta = \frac{1-\gamma}{3-\gamma}.$$

This completes the proof of the (4.6).

To prove (4.7) we apply the arithmetic-geometric mean inequality to (4.12) and get for any positive η,

$$\left\|E_i^X\right\|_A^2 \leq (1+\eta)\delta^2 \left\|E_{i-1}^X\right\|_A^2 + (1+\eta^{-1})(1+\delta)^2 \left\|E_{i-1}^Y\right\|_{Q_B}^2 .$$

Inequality (4.7) follows taking $\eta = 1+1/\delta$ and applying (4.6). This completes the proof of the theorem.

5. Application to a Stokes problem . In this section we consider an application of the theory developed in the previous sections to solving indefinite systems of linear equations arising from finite element approximations of the Stokes equations. For simplicity we restrict our discussion to the following model problem: Find $\mathbf{u}$ and p such that

$$\begin{aligned} -\Delta\mathbf{u} - \nabla p &= \mathbf{g} \quad \text{in } \Omega, \\ \nabla\cdot\mathbf{u} &= f \quad \text{in } \Omega, \\ \mathbf{u} &= 0 \quad \text{on } \partial\Omega, \\ \int_\Omega p(x)\,dx &= 0, \end{aligned} \tag{5.1}$$

where Ω is the unit cube in R^d, d=2, 3, Δ is the componentwise Laplace operator, $\mathbf{u}$ is a vector valued function representing the velocity, and the pressure p is a scalar function. Generalizations to domains with more complex geometry and variable coefficients equations are possible.

Let $L_0^2(\Omega)$ 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 (cf., [18], [20]). The space $H_0^1(\Omega)$ consists of those functions in Ω whose traces vanish on $\partial\Omega$, the boundary of Ω. Also, $(H_0^1(\Omega))^d$ 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 $(H_0^1(\Omega))^d$ and $L_0^2(\Omega)$ respectively and integrate over Ω to get

$$\begin{aligned} D(\mathbf{u},\mathbf{v}) + (p, \nabla\cdot\mathbf{v}) &= (\mathbf{g},\mathbf{v}) \quad \text{for all } \mathbf{v}\in(H_0^1(\Omega))^d, \\ (\nabla\cdot\mathbf{u}, q) &= (f,q) \quad \text{for all } q\in L_0^2(\Omega). \end{aligned} \tag{5.2}$$

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

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

We next identify approximation subspaces of $(H_0^1(\Omega))^d$ and $L_0^2(\Omega)$. In order to avoid unnecessary complexity of the presentation only a two dimensional example will be considered. The discussion here is very closely related to the examples given in [4] and [5] where additional comments and other applications can be found. We partition Ω

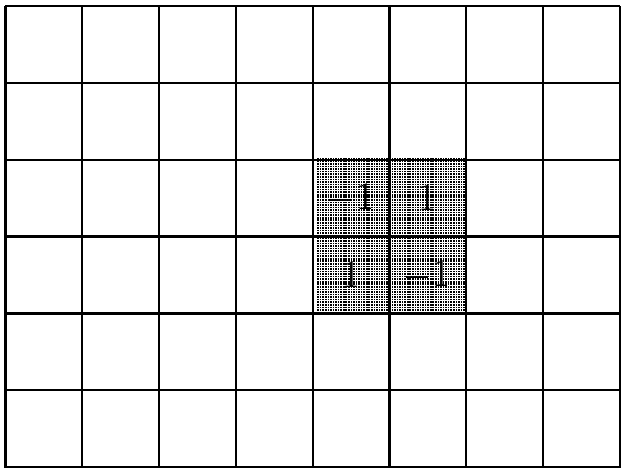

FIG. 5.1. *The square mesh used for $\tilde{H}_2$; the support (shaded) and values for a typical ϕ_{ij}.*

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 right corner to the upper left 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 thus defined. We set $H_1 \equiv S_h \times S_h \subset (H_0^1(\Omega))^2$. The choice of H_2 is motivated by the observation [19] that the space $\tilde{H}_2$ of functions that are piecewise constant with respect to the square elements together with H_1 as defined above form an unstable pair of approximation spaces. This means that the functions from $H_1 \times \tilde{H}_2$ do not satisfy (1.4) with a constant c_0 independent of the discretization parameter h. To overcome this problem, one may consider a smaller space defined as follows. Let η_{kl} for $k, l = 1, \ldots, 2n$ be the function that is one on the square element $[x_{k-1}, x_k] \times [y_{l-1}, y_l]$ and vanishes elsewhere. Define $\phi_{ij} \in \tilde{H}_2$ 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 5.1). The space H_2 is then defined by

$$H_2 \equiv \left\{ W \in \tilde{H}_2 \ : \ (W, \phi_{ij}) = 0 \text{ for } i, j = 1, \ldots, n \right\}.$$

The pair $H_1 \times H_2$ now satisfies (1.4) with a constant c_0 independent of h [19]. Moreover, the exclusion of the functions $\phi_{i,j}$ does not change the order of approximation for the space since the H_2 still contains the piecewise constant functions of size $2h$.

The approximation to the solution of (5.2) is defined as the unique pair $(X, Y) \in H_1 \times H_2$ satisfying

$$\begin{aligned} D(X, V) + (Y, \nabla \cdot V) &= (\mathbf{g}, V) \quad \text{for all } V \in H_1, \\ (\nabla \cdot X, W) &= (f, W) \quad \text{for all } W \in H_2. \end{aligned} \tag{5.3}$$

Obviously, (5.3) is a system of linear equations whose unique solvability is guaranteed by (1.4).

The system (5.3) can be reformulated in terms of operators as follows. Let

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

It follows that the solution (X, Y) of (5.3) satisfies (1.1) with F equal to the $L^2(\Omega)$ projection of f into H_2 and G equal to the $(L^2(\Omega))^2$ projection of $\mathbf{g}$ into H_1.

It is straightforward to check that (2.3) holds for $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{B}^T$ as above. Moreover, it follows from (1.4) that (2.4) holds with γ independent of the mesh size 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 $\mathbf{Q}_A^{-1} f$ for $f \in H_1$ is defined as a process which acts on the inner product data (f, 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 $(\mathbf{A}V, \psi_i))$, 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 [5]) and we can take $\mathbf{Q}_B$ to be the identity. 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.

We still need to provide preconditioners for $\mathbf{A}$. However, $\mathbf{A}$ 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. In our examples Section 7, we will use a preconditioning operator which results from a V-cycle variational multigrid algorithm. Such a preconditioner is known to be scaled so that both (3.2) holds and (2.4) holds with γ bounded away from one independently of the mesh parameter h.

6. Applications to mixed finite element discretizations of elliptic problems. In this section we discuss applications of the algorithms analyzed in Section 3 to solving indefinite systems arising from mixed finite element discretizations of second order partial differential equations. For this application, it will be relatively easy to construct preconditioners $\mathbf{Q}_A$ while the development of a suitable operator $\mathbf{Q}_B$ is more difficult.

The basic problem we consider here is

$$\begin{aligned}
-\nabla \cdot \mathbf{K} \nabla p &= f \quad \text{in } \Omega,\\
p &= 0 \quad \text{on } \partial\Omega,
\end{aligned} \tag{6.1}$$

where $\mathbf{K} = \{k_{i,j}\}_{i,j=1}^{d}$ is a symmetric positive definite matrix whose entries are bounded functions of the spatial variable, Ω is a bounded domain with polygonal or polyhedral boundary in d–dimensional Euclidean space for $d = 2$ or 3. This is a classical model problem in continuum mechanics or fluid flow in porous media.

Introducing a new variable $\mathbf{u}$, (6.1) can be written as a first order system as follows:

$$\begin{aligned} \mathbf{u} &= \mathbf{K}\nabla p \quad \text{in } \Omega, \\ \nabla \cdot \mathbf{u} &= -f \quad \text{in } \Omega, \\ p &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{6.2}$$

In the typical applications $\mathbf{K}$ is the elasticity/permeability tensor, $\mathbf{u}$ usually represents the stress/velocity, p is the displacement/pressure. The mixed method naturally takes into account constraints that appear in the variational formulation of a given differential problem, e.g., $\nabla \cdot \mathbf{u} = f$, and provides direct approximations to the two variables of interest: $\mathbf{u}$ and p. Often these features are more attractive then those corresponding to the standard finite element method.

Then the weak formulation of (6.2) is

$$\begin{aligned} (\mathbf{K}^{-1}\mathbf{u}, V) + (p, \nabla \cdot V) &= 0, && \text{for all } V \in H_{div}(\Omega), \\ (\nabla \cdot \mathbf{u}, W) &= -(f, W), && \text{for all } W \in L^2(\Omega). \end{aligned} \tag{6.3}$$

The space $H_{div}(\Omega)$ is the set of vector functions in $(L^2(\Omega))^d$ whose divergences are also in $L^2(\Omega)$. Here, as in the previous section, $(\cdot, \cdot)$ denotes the $L^2(\Omega)$ or $(L^2(\Omega))^2$ inner product. The mixed discretizations involve the introduction of two approximation subspaces, $H_1 \subset H_{div}(\Omega)$ for the velocities and $H_2 \subset L^2(\Omega)$ for the pressures. To illustrate this type of application, we will only discuss the simplest mixed finite element discretization of (6.1), namely the lowest order Raviart-Thomas spaces [24]. We assume some familiarity with the mixed approximation approach and only give limited detail. Detailed development can be found in [8], [21], and [24].

Let T_h be a partitioning of Ω into simplices of quasi-uniform size h. The space H_1 is defined to be the vector valued functions which are linear on the simplices and have a continuous constant normal component on each of the face of the mesh. The space H_2 consists of the set functions which are constant (discontinuous across the faces) on each of the simplices. The mixed finite element approximation is defined to be the pair (X, Y) satisfying

$$\begin{aligned} (\mathbf{K}^{-1}X, V) + (Y, \nabla \cdot V) &= 0, && \text{for all } V \in H_1, \\ (\nabla \cdot X, W) &= -(f, W), && \text{for all } W \in H_2. \end{aligned}$$

The operators $\mathbf{B}$ and $\mathbf{B}^T$ are defined as in the previous section. However, for this application, the operator $\mathbf{A} : H_1 \mapsto H_1$ is defined by

$$(\mathbf{A}U, V) = (\mathbf{K}^{-1}U, V) \quad \text{for all } U, V \in H_2.$$

In terms of these operators, we get a discrete system of linear equations of type (1.1) with $F = 0$ and $G = f_h$, where f_h is the $L^2(\Omega)$ orthogonal projection of f into H_2.

The operator A is well conditioned and hence a simple multiple of the identity provides an effective $\mathbf{Q}_A$. On the other hand, the operator $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ is not uniformly well conditioned. In fact, it exhibits a condition number growth like h^{-2} and should be preconditioned in order to get an efficient algorithm of type (2.2) or (2.5). It is well known that $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ behaves like a discretization of a second order operator. In some applications, it can be preconditioned by cell centered techniques [25], multigrid [7], or incomplete Choleski factorization of $\mathbf{B}\mathbf{B}^T$ [26].

7. Numerical examples. In this section we present the results from numerical experiments that illustrate the theory developed in the earlier sections. We also report similar results obtained from applying the conjugate gradient algorithm for saddle point problems introduced in [5].

Even though the most effective algorithms result from the use of good preconditioners, we shall initially present results using one of the worst possible preconditioners, the identity operator. This is important since in some engineering applications, good preconditioners may not be readily available. We also report results when effective preconditioners are employed.

The test problem was (5.1) with $\Omega \equiv (0,1)^2$, $\mathbf{g} = 0$ and $f = 0$. Clearly, its exact solution was zero for both pressure and velocity. We started the iterations with an arbitrary but fixed initial iterate. All of the iterative methods considered are functions of the error and thus, iterating for a problem with a zero solution and a nonzero starting guess is equivalent to solving a related problem with a nonzero solution and a zero initial guess. We used the discretization described in Section 5.

Our objectives in conducting the numerical experiments were to establish experimentally the conclusions from the theoretical analysis of the algorithms tested and to assess their effectiveness in terms of error reduction after fixed number of iterations. The same nonzero initial iterate was used for all algorithms. As discussed in Section 5, we used $\mathbf{Q}_B \equiv \mathbf{I}$. The experimental results are organized in four tables.

In Table 7.1 we give results for three algorithms using $\mathbf{Q}_A$ equal to an appropriate multiple of the identity. The algorithms are described as follows.

UID : The algorithm (2.5) with $\mathbf{Q}_A = \bar{\lambda}_{max}\mathbf{I}$ and $\mathbf{Q}_B = \mathbf{I}$. Here $\bar{\lambda}_{max}$ is an upper bound for the largest eigenvalue of $\mathbf{A}$.

USTD : The algorithm (4.4) with $\mathbf{Q}_B = \mathbf{I}$ and Ψ defined by one step of the steepest descent method (SDM) applied to approximate the action of $\mathbf{A}^{-1}$.

BPID : The preconditioned conjugate gradient algorithm for saddle point problems given in [5] with $\mathbf{Q}_A = \bar{\lambda}_{min}\mathbf{I}$, where $\bar{\lambda}_{min}$ is a lower bound for the smallest eigenvalue of $\mathbf{A}$ and $\mathbf{Q}_B = \mathbf{I}$. Notice that the scaling required by Theorem 1 of [5] is in the opposite direction of (3.1).

The reported error values in Table 7.1 represent the relative error norm after i iterations

computed by

$$Error_i = \left(\frac{D(E_i^X, E_i^X) + \left\| E_i^Y \right\|^2}{D(E_0^X, E_0^X) + \left\| E_0^Y \right\|^2} \right)^{1/2} . \tag{7.1}$$

Clearly, this is not the norm which appears in the theory and one cannot expect the errors to behave in a monotone way. This explains the increase in the reported error for **UID** when $h = 1/32$ and $h = 1/64$. That the **USTD** method appears convergent for $h \leq 32$ is surprising since (4.5) is not satisfied for these applications. The **BPID** method converges considerably faster in these examples since the saddle point method of [5] is known to give a rate of convergence which exhibits square root acceleration in cases when poor preconditioners are employed. As expected, all methods deteriorate due to lack of preconditioning as the mesh size is decreased.

TABLE 7.1
Errors in **UID**, **USTD** *and* **BPID** *by (7.1)*

h	200 iterations		
	UID	**USTD**†	**BPID**
1/8	4.2×10^{-3}	5.1×10^{-6}	‡6.5×10^{-12}
1/16	0.4	5.8×10^{-2}	2.9×10^{-10}
1/32	1.5	0.2	1.1×10^{-4}
1/64	2.7	4.5	2.0×10^{-2}

† *one SDM step per inexact Uzawa iteration.*
‡ *for 109* **BPID** *iterations.*

In order to establish experimentally the convergence of **UID** and **USTD**, we ran these two algorithms for 2000 iterations. The results are shown in Table 7.2. Even though improved convergence is observed in all cases when compared to Table 7.1, the **UID** algorithm still appears unstable for $h = 1/64$. We ran **UID** for 10000 iterations and observed an error of .0048. Although convergent, the inexact Uzawa method with such a poor preconditioner converges too slowly to be of practical use.

The above results may at first appear to contradict the validity of the theory of Section 4. The reason that the methods appear divergent at a relatively low numbers of iterations is that the theorems guarantee monotonicity of the errors in norms which are different from those used in (7.1). Our next experiment was designed to illustrate the monotone convergence of **UID** and **BPID** predicted by Theorem 1 and Theorem 1 in [5]. Accordingly, we measured the errors in the norms appearing in the theorems. In the case of **UID**, we use

$$Error_i = \left(\frac{\bar{\lambda}_{max} \left\| E_i^X \right\|^2 - D(E_i^X, E_i^X) + \left\| E_i^Y \right\|^2}{\bar{\lambda}_{max} \left\| E_0^X \right\|^2 - D(E_0^X, E_0^X) + \left\| E_0^Y \right\|^2} \right)^{1/2} . \tag{7.2}$$

TABLE 7.2
Errors in **UID** *and* **USTD** *by (7.1)*

h	2000 iterations	
	UID	**USTD**†
1/8	0	2.0×10^{-23}
1/16	3.7×10^{-6}	3.9×10^{-16}
1/32	2.5×10^{-2}	2.1×10^{-4}
1/64	1.5	8.7×10^{-2}

† *one SDM step per inexact Uzawa iteration.*

In the case of **BPID**, we used

$$Error_i = \left(\frac{D(E_i^X, E_i^X) - \bar{\lambda}_{min} \left\| E_i^X \right\|^2 + \left\| E_i^Y \right\|^2}{D(E_0^X, E_0^X) - \bar{\lambda}_{min} \left\| E_0^X \right\|^2 + \left\| E_0^Y \right\|^2} \right)^{1/2} . \tag{7.3}$$

The convergence results in these norms are reported in Table 7.3. Note that all of the reported errors are less than one. We made additional runs at lower number of iterations. All runs reflected the monotone error behavior in these norms as guaranteed by the theory.

TABLE 7.3
Errors in **UID** *and* **BPID** *by (7.1) and (7.3)*

h	200 iterations	
	UID	**BPID**
1/8	4.29×10^{-3}	‡2.1×10^{-12}
1/16	0.18	3.1×10^{-10}
1/32	0.52	1.1×10^{-4}
1/64	0.77	2.0×10^{-2}

‡ *for 109* **BPID** *iterations.*

The last experiment given in this section is intended to illustrate the performance of the algorithms when effective preconditioners are used. In this case, we define $\mathbf{Q}_A^{-1}$ 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. Note that $\mathbf{Q}_A$ automatically satisfies (3.2) and satisfies (3.3) with δ independent of h. We consider the following two algorithms:

UMG : The algorithm (2.5) with $\mathbf{Q}_B = \mathbf{I}$ and $\mathbf{Q}_A^{-1}$ being the action of multigrid.

BPMG : The algorithm from [5] with the $\mathbf{A}$ block preconditioned by $.5\mathbf{Q}_A^{-1}$ and $\mathbf{Q}_B = \mathbf{I}$.

Table 7.4 contains the error reductions for this example. The effect of applying a better preconditoner $\mathbf{Q}_A$ is clearly seen when we compare the results from **UID** (Tables 7.1 and 7.2) with those from **UMG**. Notice that the **UMG** data in Table 7.4 show little, if any, deterioration as the mesh size becomes small.

TABLE 7.4
Errors in **UMG** *and* **BPMG** *by (7.1)*

h	40 iterations	
	UMG†	**BPMG**†
1/8	1.6×10^{-5}	1.0×10^{-11}
1/16	9.4×10^{-7}	6.9×10^{-9}
1/32	1.6×10^{-6}	1.3×10^{-8}
1/64	2.2×10^{-6}	4.5×10^{-9}

† *one multigrid V–cycle per iteration.*

In all of the reported results, the reformulation method of [5] shows faster convergence. Nevertheless, the inexact Uzawa methods are still of interest since they are robust, simple to implement, have minimal memory requirements and avoid the necessity of computing inner products. These properties may make the inexact Uzawa methods attractive in certain applications.

REFERENCES

[1] K. ARROW, L. HURWICZ, AND H. UZAWA, *Studies in Nonlinear Programming*, Stanford University Press, Stanford, CA, 1958.

[2] I. BABUŠKA, *The finite element method with Lagrangian multipliers*, Numer. Math., 20 (1973), pp. 179–192.

[3] J. BRAMBLE, *The Lagrange multiplier method for Dirichlet's problem*, Math. Comp., 37 (1981), pp. 1–12.

[4] J. BRAMBLE AND J. PASCIAK, *Iterative techniques for time dependent Stokes problems.* Brookhaven National Laboratory Report BNL-49970 (1994).

[5] ——, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp., 50 (1988), pp. 1–18.

[6] J. H. BRAMBLE, *Multigrid Methods*, Pitman Research Notes in Mathematics Series, Longman Scientific & Technical, London. Copublished with John Wiley & Sons, Inc., New York, 1993.

[7] J. BRAMBLE, J. PASCIAK AND J. XU, *The analysis of multigrid algorithms with non-nested spaces or non-inherited quadratic forms*, Math. Comp., 56 (1991), pp. 1–34.

[8] F. BREZZI AND M. FORTIN, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York, 1991.

[9] L. COWSAR, J. MANDEL, AND M. WHEELER, *Balancing domain decomposition for mixed finite elements*, Math. Comp. (to appear).

[10] H. ELMAN, *Multigrid and Krylov subspace methods for the discrete Stokes equations.* Univ. of Maryland. Dept. of Comp. Sci. Rep. CS-TR-3302 (1994).

[11] H. ELMAN AND G. GOLUB, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal. (to appear).

[12] R. EWING AND M. WHEELER, *Computational aspects of mixed finite element methods*, in Numerical methods for scientific computing, R. Steelman, ed., Amsterdam, 1983, North-Holland, pp. 163–172.

[13] R. EWING, R. LAZAROV, P. LU, AND P. VASSILEVSKI, *Preconditioning indefinite systems arising from mixed finite element discretization of second-order elliptic problems*, in Preconditioned Conjugate Gradient Methods, Lecture Notes in Mathematics, O. Axelsson and L. Kolotina, eds., vol. 1457, Berlin, 1990, Springer-Verlag, pp. 280–343.

[14] R. FALK, *An analysis of the finite element method using Lagrange multipliers for the stationary Stokes equations*, Math. Comp., 30 (1976), pp. 241–269.

[15] R. FALK AND J. OSBORN, *Error estimates for mixed methods*, R.A.I.R.O. Numerical Analysis, 14 (1980), pp. 249–277.

[16] V. GIRAULT AND P. RAVIART, *Finite Element Approximation of the Navier-Stokes Equations*, Lecture Notes in Math. # 749, Springer-Verlag, New York, 1981.

[17] R. GLOWINSKI AND M. WHEELER, *Domain decomposition and mixed finite element methods for elliptic problems*, in First Inter. Symp. on Domain Decomposition Methods for Partial Differential Equations, R. Glowinski, G. Golub, G. Meurant, and J. Periaux, eds., Phil. PA, 1988, SIAM, pp. 144–172.

[18] P. GRISVARD, *Elliptic Problems in Nonsmooth Domains*, Pitman, Boston, 1985.

[19] C. JOHNSON AND J. PITKÄRANTA, *Analysis of some mixed finite element methods related to reduced integration*, Math. Comp., 38 (1982), pp. 375–400.

[20] J. LIONS AND E. MAGENES, *Problèmes aux Limites non Homogènes et Applications*, vol. 1, Dunod, Paris, 1968.

[21] J. NEDELEC, *Elements finis mixtes incompressibles pour l'equation de Stokes dans R^3.*, Numer. Math., 39 (1982), pp. 97–112.

[22] W. PATTERSON, 3RD, *Iterative Methods for the Solution of a Linear Operator Equation in Hilbert Space - A Survey*, Lecture Notes in Mathematics 394, Springer-Verlag, New York, 1974.

[23] W. QUECK, *The convergence factor of preconditioned algorithms of the arrow-hurwicz type*, SIAM J. Numer. Anal., 26 (1989), pp. 1016–1030.

[24] P. RAVIART AND J. THOMAS, *A mixed finite element method for 2-nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, New York, 1977, Springer-Verlag, pp. 292–315. Lecture Notes in Mathematics, #606 (Eds. I. Galligani and E. Magenes).

[25] T. RUSSELL AND M. WHEELER, *Finite element and finite difference methods for continuous flows in porous media*, in The Mathematics of Reservoir Simulation, R. Ewing, ed., Phil, PA, 1983, SIAM, pp. 35–106.

[26] T. RUSTEN AND R. WINTHER, *A preconditioned iterative method for saddle point problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 887–904.