SIAM J. MATRIX ANAL. APPL.
Vol. 23, No. 2, pp. 317–338

AN ITERATIVE METHOD WITH VARIABLE RELAXATION PARAMETERS FOR SADDLE-POINT PROBLEMS*

QIYA HU† AND JUN ZOU‡

Abstract. In this paper, we propose an inexact Uzawa method with variable relaxation parameters for iteratively solving linear saddle-point problems. The method involves two variable relaxation parameters, which can be updated easily in each iteration, similar to the evaluation of the two iteration parameters in the conjugate gradient method. This new algorithm has an advantage over most existing Uzawa-type algorithms: it is always convergent without any a priori estimates on the spectrum of the preconditioned Schur complement matrix, which may not be easy to achieve in applications. The rate of the convergence of the inexact Uzawa method is analyzed. Numerical results of the algorithm applied for the Stokes problem and a purely linear system of algebraic equations are presented.

Key words. saddle-point, inexact Uzawa method, indefinite systems, preconditioning

AMS subject classifications. 65F10, 65N20

PII. S0895479899364064

1. Introduction.

The major interest of this paper is to solve the indefinite system of equations

$$\begin{pmatrix} A & B \\ B^t & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{1.1}$$

where A is a symmetric and positive definite $n \times n$ matrix, and B is an $n \times m$ matrix with $m \leq n$. We assume that the global coefficient matrix

$$M = \begin{pmatrix} A & B \\ B^t & 0 \end{pmatrix}$$

is nonsingular, which is equivalent to the positive definiteness of the Schur complement matrix

$$C = B^t A^{-1} B. \tag{1.2}$$

Linear systems such as (1.1) are called saddle-point problems, which may arise from finite element discretizations of Stokes equations and Maxwell equations [6], [8], [12]; mixed finite element formulations for second order elliptic problems [2], [6]; or from Lagrange multiplier formulations for optimization problems [1], [13], for parameter identification, and domain decomposition problems [9], [14], [15].

*Received by the editors November 18, 1999; accepted for publication (in revised form) by L. Eldén March 5, 2001; published electronically August 8, 2001.

http://www.siam.org/journals/simax/23-2/36406.html

†Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing 100080, China (hqy@lsec.cc.ac.cn). The work of this author was partially supported by National Natural Science Foundation grant 19801030 and a grant from the Institute of Mathematical Sciences of the Chinese University of Hong Kong.

‡Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong (zou@math.cuhk.edu.hk). The work of this author was partially supported by Hong Kong RGC grants CUHK4004/98P and CUHK4292/00P and the Visiting Scholar Foundation of Key Laboratory in University (China).

In recent years, there has been a rapidly growing interest in preconditioned iterative methods for solving the indefinite system of equations like (1.1); see [3], [4], [5], [7], [11], [14], [16], [17], and [18]. In particular, the inexact Uzawa-type algorithms have attracted wide attention; see [3], [4], [7], [11], [17], and the references therein. The main merit of these Uzawa-type algorithms is that they preserve the minimal memory requirement and do not need actions of the inverse matrix A^{-1}.

Let $\hat{A}$ and $\hat{C}$ be two positive definite matrices, which are assumed to be the preconditioners of the matrices A and C, respectively. Also let R^l be the usual l-dimensional Euclidean space. For any $l \times l$ positive definite matrix G, we use $\|x\|_G$ to denote the G-induced norm, i.e., $\|x\|_G = (Gx,\ x)^{1/2}$ for all $x \in R^l$. However, we write $\|x\|$ (the Euclidean norm) when G is the identity. Then the standard inexact Uzawa algorithm can be described as follows (cf. [4] and [11]).

Algorithm 1.1 (inexact Uzawa). Given $x_0 \in R^n$ and $y_0 \in R^m$, the sequence $\{x_i, y_i\} \subset R^n \times R^m$ is defined for $i = 1, 2, \ldots$ by

$$x_{i+1} = x_i + \hat{A}^{-1}[f - (Ax_i + By_i)] \tag{1.3}$$

and

$$y_{i+1} = y_i + \hat{C}^{-1}(B^t x_{i+1} - g). \tag{1.4}$$

There are several earlier versions of the above algorithm; see, e.g., [3] and [17]. The existing convergence results indicate that these algorithms are convergent by assuming some good knowledge of the spectrum of the preconditioned matrices $\hat{A}^{-1}A$ and $\hat{C}^{-1}C$ or under some proper scalings of the preconditioners $\hat{A}$ and $\hat{C}$. This "preprocessing" may not be easy to achieve in some applications.

To avoid the proper estimate of the generalized eigenvalues of $\hat{C}$ with respect to $B^t\hat{A}^{-1}B$, the Uzawa-type algorithm proposed in [3] introduced a preconditioned conjugate gradient (PCG) algorithm as an inner iteration of (1.4) and proved that when the number of the PCG iteration is suitably large this Uzawa-type algorithm converges. However, it requires subtle skill in implementations to determine when to terminate this inner iteration.

The preconditioned minimal residual method is always convergent, but its convergence depends on the ratio of the smallest eigenvalue of $\hat{A}^{-1}A$ over the smallest eigenvalue of $\hat{C}^{-1}(B^t\hat{A}^{-1}B)$ (cf. [18]). Hence one should have some good knowledge of the smallest eigenvalues of these preconditioned matrices in order to achieve a practical convergence rate. Without a good scaling based on some a priori estimate of these smallest eigenvalues, the condition number of the (global) preconditioned system still may be very large even if the condition numbers of the matrices $\hat{A}^{-1}A$ and $\hat{C}^{-1}(B^t\hat{A}^{-1}B)$ are small (cf. [18]). In this case, the convergence of this iterative method may be slow (see section 4).

In this paper we propose a new variant of the inexact Uzawa algorithm to relax some aforementioned drawbacks by introducing two variable relaxation parameters in the algorithm (1.3)–(1.4). That is, we define the sequence $\{x_i, y_i\}$ for $i = 1, 2, \ldots$ by

$$x_{i+1} = x_i + \omega_i\hat{A}^{-1}[f - (Ax_i + By_i)] \tag{1.5}$$

and

$$y_{i+1} = y_i + \tau_i\hat{C}^{-1}(B^t x_{i+1} - g). \tag{1.6}$$

The parameters ω_i and τ_i above can be computed effectively, similar to the evaluation of the two iteration parameters in the conjugate gradient method. It will be shown that our algorithm always converges provided the preconditioner $\hat{A}$ for A is properly scaled so that the eigenvalues of $A^{-1}\hat{A}$ are bounded by one. It is very interesting to know whether this is a technical or necessary assumption, a question to which we still do not have a definite answer. But the numerical experiments of section 4 seem to imply that the proposed algorithm converges even when this assumption is violated. Furthermore, it is important to remark that the convergence of the new algorithm is independent of the constant scalings of the preconditioners $\hat{A}$ and $\hat{C}$ while the convergences of the preconditioned minimum residual (MINRES) method and Algorithm 1.1 are strongly affected by such constant scalings; see section 4 for some numerical verifications. Also the new algorithm is always convergent for general preconditioners $\hat{C}$, while the convergences of most existing Uzawa-type algorithms are guaranteed only under certain conditions on the extreme eigenvalues of the preconditioned matrix $\hat{C}^{-1}C$ or $\hat{C}^{-1}H$ (cf. [3] and [4]).

The rest of the paper is arranged as follows. In section 2, we describe the algorithm and its convergence results, which indicate that the algorithm converges with an optimal rate (independent of mesh sizes) if the preconditioned matrices $\hat{A}^{-1}A$ and $\hat{C}^{-1}C$ or $\hat{C}^{-1}(B^t\hat{A}^{-1}B)$ are well-conditioned. The analysis of convergence rates will be given in section 3. In section 4, we apply the proposed algorithm for solving the Stokes problem and a linear system of purely algebraic equations.

2. Algorithm and main results. We start with some illustrations about how to choose the relaxation parameters ω_i and τ_i in (1.5)–(1.6). We first claim that it is impractical to determine these two parameters by the standard steepest descent method. To see this, let $\{x, y\}$ be the true solution of the saddle-point problem (1.1) and set

$$e_i^x = x - x_i, \qquad e_i^y = y - y_i,$$

$$f_i = f - (Ax_i + By_i), \quad g_i = B^t x_{i+1} - g.$$

Consider two arbitrary symmetric and positive definite $n \times n$ and $m \times m$ matrices A_0 and C_0. Suppose we choose the parameters ω_i and τ_i such that the errors

$$\|e_{i+1}^x\|_{A_0}^2 = \|e_i^x\|_{A_0}^2 - 2\,\omega_i(e_i^x, \hat{A}^{-1}f_i)_{A_0} + \omega_i^2\|\hat{A}^{-1}f_i\|_{A_0}^2$$

and

$$\|e_{i+1}^y\|_{C_0}^2 = \|e_i^y\|_{C_0}^2 - 2\,\tau_i(e_i^y, \hat{C}^{-1}g_i)_{C_0} + \tau_i^2\|\hat{C}^{-1}g_i\|_{C_0}^2$$

are minimized; then we have

$$\omega_i = \frac{(A_0e_i^x, \hat{A}^{-1}f_i)}{\|\hat{A}^{-1}f_i\|_{A_0}^2}, \quad f_i \neq 0; \quad \tau_i = \frac{(C_0e_i^y, \hat{C}^{-1}g_i)}{\|\hat{C}^{-1}g_i\|_{C_0}^2}, \quad g_i \neq 0.$$

This requires the evaluations of $A_0e_i^x = A_0x - A_0x_i$ and $C_0e_i^y = C_0y - C_0y_i$. Clearly such evaluations are usually very expensive no matter how we choose A_0 and C_0, since the action of A^{-1} is always involved. This verifies our claim.

Now, we are going to find a more efficient way to compute the parameters ω_i and τ_i. Note that the exact version of the inner iteration (1.3) is

$$x_{i+1} = x_i + A^{-1}f_i.$$

Comparing this with the inexact iteration (1.5), we see that ω_i may be chosen such that the norm

$$\|A^{-1}f_i - \omega_i \hat{A}^{-1}f_i\|_A^2$$

is minimized. A direct computation yields that

$$\omega_i = \begin{cases} \frac{(f_i, \hat{A}^{-1}f_i)}{\|\hat{A}^{-1}f_i\|_A^2}, & f_i \neq 0, \\ 1, & f_i = 0. \end{cases} \tag{2.1}$$

With this parameter ω_i, the outer iteration (1.4) is changed to

$$y_{i+1} = y_i + \hat{C}^{-1}(b_i - \omega_i B^t \hat{A}^{-1} B y_i)$$

with

$$b_i = B^t x_i + \omega_i B^t \hat{A}^{-1}(f - Ax_i) - g,$$

which is independent of y_i. When replacing $\hat{C}$ by $\omega_i B^t \hat{A}^{-1} B$, we get the exact version of this outer iteration:

$$y_{i+1} = y_i + (\omega_i B^t \hat{A}^{-1} B)^{-1} g_i.$$

Comparing this with the inexact form (1.6), we see that the parameter τ_i may be chosen such that the norm

$$\|(\omega_i B^t \hat{A}^{-1} B)^{-1} g_i - \tau_i \hat{C}^{-1} g_i\|_{(\omega_i B^t \hat{A}^{-1} B)}^2$$

is minimized. A direct calculation gives

$$\tau_i = \begin{cases} \omega_i^{-1} \frac{(\hat{C}^{-1}g_i,\ g_i)}{\|\hat{C}^{-1}g_i\|_{B^t\hat{A}^{-1}B}^2}, & g_i \neq 0; \\ 1, & g_i = 0. \end{cases} \quad \text{or} \quad \tau_i = \begin{cases} \omega_i^{-1} \frac{(\hat{C}^{-1}g_i,\ g_i)}{\|B\hat{C}^{-1}g_i\|_{\hat{A}^{-1}}^2}, & g_i \neq 0; \\ 1, & g_i = 0. \end{cases} \tag{2.2}$$

Unfortunately, such a relaxation parameter τ_i chosen as in (2.2) may cause the corresponding algorithm (1.5)–(1.6) to diverge, especially when ω_i is very small. This has been confirmed by our numerical experiments. Also we will see from the subsequent analysis that the factor ω_i^{-1} in (2.2) needs to be corrected appropriately to guarantee the convergence.

With the above preparations, we are now ready to formulate a new inexact Uzawa algorithm.

Algorithm 2.1 (Uzawa algorithm with variable relaxation parameters). Given the initial guesses $x_0 \in R^n$ and $y_0 \in R^m$, compute the sequences $\{x_i, y_i\}$ for $i = 1, 2, \ldots$ as follows.

Step 1. Compute $f_i = f - (Ax_i + By_i)$, $r_i = \hat{A}^{-1}f_i$, and

$$\omega_i = \begin{cases} \frac{(f_i,\ r_i)}{(Ar_i, r_i)}, & f_i \neq 0, \\ 1, & f_i = 0. \end{cases}$$

Set

$$x_{i+1} = x_i + \omega_i\, r_i\,. \tag{2.3}$$

Step 2. Compute $g_i = B^t x_{i+1} - g,\ \ d_i = \hat{C}^{-1} g_i$, and

$$\tau_i = \begin{cases} \frac{(g_i,\ d_i)}{(\hat{A}^{-1}Bd_i,\ Bd_i)}, & g_i \neq 0, \\ 1, & g_i = 0\,. \end{cases}$$

Set

$$y_{i+1} = y_i + \theta_i \tau_i \, d_i \tag{2.4}$$

with

$$\theta_i = \frac{1 - \sqrt{1 - \omega_i}}{2}. \tag{2.5}$$

Remark 2.1. Intuitively, it is not easy to see why one needs to introduce the additional parameter θ_i in (2.4), but its presence is essential to guarantee the convergence of Algorithm 2.1. This will become transparent from our subsequent convergence proof. Also, the choices of θ_i in (2.4) are not unique. In fact, θ_i can be chosen to be any real numbers such that

$$0 < \theta_i \leq \frac{1 - \sqrt{1 - \omega_i}}{2}.$$

We refer to the remarks at the end of section 3 for more details.

Remark 2.2. It is clear that when both f_i and g_i vanish, the vectors x_i and y_i are the exact solution of the system (1.1). In this case Algorithm 2.1 terminates.

Now we are ready to state our main results. Let $H = B^t \hat{A}^{-1} B$ and

$$\kappa_1 = \mathrm{cond}(\hat{A}^{-1}A), \quad \alpha = \frac{\kappa_1 - 1}{\kappa_1 + 1},$$

$$\kappa_2 = \mathrm{cond}(\hat{C}^{-1}H), \quad \beta = \frac{\kappa_2 - 1}{\kappa_2 + 1}\,.$$

We shall frequently use a new norm $|||\cdot|||$ given by

$$|||v||| = \left(\|v_1\|^2 + \|v_2\|_C^2\right)^{\frac{1}{2}}, \quad v = \{v_1, v_2\} \in R^n \times R^m.$$

Without loss of generality, from now on we will always assume that $\alpha > 0$, and the preconditioner $\hat{A}$ for A is properly scaled so that

$$(\hat{A}v, v) \leq (Av, v) \quad \text{for all } v \in R^n. \tag{2.6}$$

The numerical experiments of section 4 indicate that Algorithm 2.1 still converges when the condition (2.6) is violated. But our convergence proof will make use of this assumption, and it is still an open question whether the convergence of Algorithm 2.1 is guaranteed without this assumption.

The following two theorems summarize the main results of the paper, and their proofs will be given in section 3.

THEOREM 2.1. *With the assumption* (2.6), *there is a positive number* $\rho < 1$ *such that*

$$|||E_{i+1}||| \leq \rho \, |||E_i|||$$

with $E_i = \{\sqrt{\alpha}A^{-\frac{1}{2}}f_i,\ e_i^y\}$. *Also the positive number* ρ *can be estimated by*

$$\rho \le \rho_0 = \frac{|c(\gamma,\alpha)| + \sqrt{c(\gamma,\alpha)^2 + 4\alpha}}{2} \tag{2.7}$$

with

$$\gamma \equiv \frac{(1-\beta)(\sqrt{\lambda_0} - \sqrt{\lambda_0 - 1})}{2\lambda_0\sqrt{\lambda_0}} < 1 - \alpha, \quad c(\gamma,\alpha) = 1 - \gamma - \alpha(1+\gamma).$$

Here λ_0 *is any positive number such that*

$$(Av, v) \le \lambda_0(\hat{A}v, v) \quad \textit{for all } v \in R^n. \tag{2.8}$$

Moreover, we have

$$\rho_0 < \begin{cases} 1 - \frac{1}{2}\gamma(1+\alpha), & 0 < \gamma \le \frac{1-\alpha}{1+\alpha}, \\ 1 - \frac{1}{2}(1-\alpha)^2, & \frac{1-\alpha}{1+\alpha} < \gamma < 1 - \alpha. \end{cases} \tag{2.9}$$

Theorem 2.2. *With the assumption* (2.6), *Algorithm* 2.1 *converges, and we have*

$$\|e_i^x\|_A \le (\sqrt{1+4\alpha} + \rho)\rho^{i-1}|||E_0|||, \quad i = 1, 2, \ldots,$$

and

$$\|e_i^y\|_C \le \rho^i|||E_0|||, \quad i = 1, 2, \ldots.$$

Remark 2.3. There always exists a λ_0 such that (2.8) holds. It follows from (2.6) that $\lambda_0 \ge 1$.

Remark 2.4. Theorem 2.2 indicates that Algorithm 2.1 is always convergent for general preconditioners $\hat{C}$. This seems to be a big advantage over most existing inexact Uzawa-type algorithms for saddle-point problems, whose convergences are guaranteed only under certain conditions on the extreme eigenvalues of the preconditioned matrix $\hat{C}^{-1}C$ or $\hat{C}^{-1}H$; see, for example, [3] and [4].

3. Analysis of the convergence rate. This section will focus on the proofs of our main results stated in Theorems 2.1 and 2.2. Unless otherwise specified, the notation below will be the same as that defined in section 2. In our subsequent proofs we will often use the following well-known inequality:

$$\frac{(v,v)\,(v,v)}{(Gv,v)\,(G^{-1}v,v)} \ge \frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2} \quad \text{for all } v \in R^l, \tag{3.1}$$

where λ_1 and λ_2 are the smallest and largest eigenvalues of the $l \times l$ symmetric positive definite matrix G. First we will show some auxiliary lemmas.

For $f_i \ne 0$, let α_i denote the following ratio:

$$\alpha_i = \frac{\|(I - \omega_i A^{\frac{1}{2}}\hat{A}^{-1}A^{\frac{1}{2}})A^{-\frac{1}{2}}f_i\|}{\|A^{-\frac{1}{2}}f_i\|}.$$

Lemma 3.1. *With the assumption* (2.6), *the above ratio* α_i *and the parameter* ω_i *given in Algorithm* 2.1 *can be bounded above and below as follows:*

$$\lambda_0^{-1} \le \omega_i \le 1 - \alpha_i^2, \quad 0 \le \alpha_i \le \alpha.$$

Proof. By the definition of the parameter ω_i, we have

$$\begin{aligned}\|(I-\omega_i A^{\frac12}\hat{A}^{-1}A^{\frac12})A^{-\frac12}f_i\|^2 &= \|A^{-1}f_i-\omega_i\hat{A}^{-1}f_i\|_A^2\\ &= \|A^{-1}f_i\|_A^2-\omega_i(f_i,\hat{A}^{-1}f_i)\\ &= \left(1-\omega_i\frac{(f_i,\hat{A}^{-1}f_i)}{(f_i,A^{-1}f_i)}\right)\|A^{-1}f_i\|_A^2. \end{aligned}\tag{3.2}$$

Using the Cauchy–Schwarz inequality and assumption (2.6), we obtain

$$\begin{aligned}(A^{-1}f_i,f_i)=(\hat{A}(A^{-1}f_i),\hat{A}^{-1}f_i) &\le \|A^{-1}f_i\|_{\hat{A}}\,\|\hat{A}^{-1}f_i\|_{\hat{A}}\\ &\le \|A^{-1}f_i\|_A\,\|\hat{A}^{-\frac12}f_i\|=(A^{-1}f_i,f_i)^{\frac12}\,(\hat{A}^{-1}f_i,f_i)^{\frac12}.\end{aligned}$$

Thus

$$(A^{-1}f_i,f_i)\le(\hat{A}^{-1}f_i,f_i),$$

and this with (3.2) leads to $\alpha_i^2\le 1-\omega_i$ or $\omega_i\le 1-\alpha_i^2$. The desired lower bound of ω_i is a direct consequence of (2.8) and the definition of ω_i.

We next show that $0\le\alpha_i\le\alpha$. It follows from (3.1) that

$$\begin{aligned}\omega_i\frac{(f_i,\hat{A}^{-1}f_i)}{(f_i,A^{-1}f_i)} &= \frac{(f_i,\hat{A}^{-1}f_i)^2}{(A\hat{A}^{-1}f_i,\hat{A}^{-1}f_i)\,(f_i,A^{-1}f_i)}\\ &= \frac{(\hat{A}^{-\frac12}f_i,\hat{A}^{-\frac12}f_i)^2}{(\hat{A}^{-\frac12}A\hat{A}^{-\frac12}(\hat{A}^{-\frac12}f_i),\hat{A}^{-\frac12}f_i)\,(\hat{A}^{\frac12}A^{-1}\hat{A}^{\frac12}(\hat{A}^{-\frac12}f_i),\hat{A}^{-\frac12}f_i)}\\ &\ge\frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2},\end{aligned}$$

where λ_1 and λ_2 are the minimal and maximal eigenvalues of the matrix $\hat{A}^{-\frac12}A\hat{A}^{-\frac12}$, respectively. This with (3.2) implies that

$$\alpha_i^2\le 1-\frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}=\alpha^2. \qquad \square$$

The following lemma introduces an auxiliary matrix Q_{Bi} which plays an important role in the subsequent spectral estimates of the propagation matrix associated with Algorithm 2.1.

LEMMA 3.2. *With the assumption* (2.6), *for any natural number i, there is a symmetric and positive definite $m\times m$ matrix Q_{Bi} such that*

(i) $Q_{Bi}^{-1}g_i=\theta_i\tau_i\hat{C}^{-1}g_i$ *with* $g_i=B^t x_{i+1}-g$ *as defined in Algorithm* 2.1;

(ii) *all eigenvalues of the matrix* $Q_{Bi}^{-1}C$ *lie in the interval* $[\frac{\theta_i(1-\beta)}{\lambda_0},\theta_i(1+\beta)]$.

Proof. If $g_i=0$, $Q_{Bi}=[\theta_i(1+\beta)]^{-1}C$ is the desired matrix. We next consider the case with $g_i\neq 0$. Using $H=B^t\hat{A}^{-1}B$, we can write

$$\|B\hat{C}^{-1}g_i\|_{\hat{A}^{-1}}^2=\|\hat{C}^{-1}g_i\|_H^2;$$

then by the definition of the parameter τ_i we have

$$\|\tau_i\hat{C}^{-1}g_i-H^{-1}g_i\|_H^2=\|H^{-1}g_i\|_H^2-\tau_i(g_i,\hat{C}^{-1}g_i)=\left(1-\tau_i\frac{(g_i,\hat{C}^{-1}g_i)}{(g_i,H^{-1}g_i)}\right)\|H^{-1}g_i\|_H^2\,.$$

It follows from (3.1) that

$$\tau_i \frac{(g_i, \hat{C}^{-1} g_i)}{(g_i, H^{-1} g_i)} = \frac{(\hat{C}^{-\frac{1}{2}} g_i, \hat{C}^{-\frac{1}{2}} g_i)^2}{(\hat{C}^{-\frac{1}{2}} H \hat{C}^{-\frac{1}{2}} (\hat{C}^{-\frac{1}{2}} g_i), \hat{C}^{-\frac{1}{2}} g_i)\,(\hat{C}^{\frac{1}{2}} H^{-1} \hat{C}^{\frac{1}{2}} (\hat{C}^{-\frac{1}{2}} g_i), \hat{C}^{-\frac{1}{2}} g_i)}$$
$$\geq \frac{4\lambda_1' \lambda_2'}{(\lambda_1' + \lambda_2')^2},$$

where λ_1' and λ_2' are the minimal and maximal eigenvalues of the matrix $\hat{C}^{-\frac{1}{2}} H \hat{C}^{-\frac{1}{2}}$, respectively. Hence we obtain

$$\|\tau_i \hat{C}^{-1} g_i - H^{-1} g_i\|_H \leq \left\{ 1 - \frac{4\lambda_1' \lambda_2'}{(\lambda_1' + \lambda_2')^2} \right\}^{\frac{1}{2}} \|H^{-1} g_i\|_H = \beta \|H^{-1} g_i\|_H.$$

This implies the existence of a symmetric positive definite $m \times m$ matrix G_{Bi} such that

$$G_{Bi}^{-1} g_i = \tau_i \hat{C}^{-1} g_i$$

and

$$\|I - H^{\frac{1}{2}} G_{Bi}^{-1} H^{\frac{1}{2}}\| \leq \beta. \tag{3.3}$$

See Lemma 9 in [3], for example, for the existence of such a matrix G_{Bi}.

Now set $Q_{Bi}^{-1} = \theta_i G_{Bi}$; then

$$Q_{Bi}^{-1} g_i = \theta_i \tau_i \hat{C}^{-1} g_i,$$

and we know from (3.3) that all eigenvalues of the matrix $H^{\frac{1}{2}} Q_{Bi}^{-1} H^{\frac{1}{2}}$ lie in the interval $[\theta_i(1-\beta), \theta_i(1+\beta)]$.

To prove result (ii), let ϕ be an eigenvector of the matrix $Q_{Bi}^{-1} C$ corresponding to the eigenvalue λ. Then we can write

$$(C\phi, \phi) = \lambda (Q_{Bi}\phi, \phi),$$

or equivalently,

$$(\hat{A}^{\frac{1}{2}} A^{-1} \hat{A}^{\frac{1}{2}} (\hat{A}^{-\frac{1}{2}} B\phi), (\hat{A}^{-\frac{1}{2}} B\phi)) = \lambda (Q_{Bi}\phi, \phi).$$

Using inequalities (2.6) and (2.8), we immediately derive

$$\lambda_0^{-1} (\hat{A}^{-\frac{1}{2}} B\phi, \hat{A}^{-\frac{1}{2}} B\phi) \leq \lambda (Q_{Bi}\phi, \phi) \leq (\hat{A}^{-\frac{1}{2}} B\phi, \hat{A}^{-\frac{1}{2}} B\phi).$$

This can be written as

$$\lambda_0^{-1} (H\phi, \phi) \leq \lambda (Q_{Bi}\phi, \phi) \leq (H\phi, \phi).$$

Note that $Q_{Bi}^{-1} H$ has the same eigenvalues as the matrix $H^{\frac{1}{2}} Q_{Bi}^{-1} H^{\frac{1}{2}}$; thus by (3.3) we have

$$\lambda_0^{-1} \theta_i (1-\beta) (Q_{Bi}\phi, \phi) \leq \lambda (Q_{Bi}\phi, \phi) \leq \theta_i (1+\beta) (Q_{Bi}\phi, \phi),$$

which yields the desired eigenvalue bound. □

The two functions $F(z)$ and $\varphi(z)$ to be introduced below and their properties are very helpful in achieving some sharper estimates in the subsequent convergence rate analysis. $F(z)$ is defined for two given positive numbers $\alpha, \gamma \in (0,1)$ as follows:

$$F(z) = \frac{1}{2}\Big(az + b + \sqrt{(az+b)^2 - 4z}\Big), \quad z \in [0,1),$$

where $a = (1+\gamma)^2 + \gamma^2/\alpha$ and $b = \alpha\gamma^2 + (1-\gamma)^2$, and it has the following properties.

LEMMA 3.3. *The function F(z) can be bounded below and above as follows:*

$$\alpha\gamma^2 + (1-\gamma)^2 \leq F(z) \leq F(\alpha^2) = \Big(|c(\gamma,\alpha)| + \sqrt{c(\gamma,\alpha)^2 + 4\alpha}\Big)^2/4 \tag{3.4}$$

for all $z \in [0,\ \alpha^2]$. Here $c(\gamma,\alpha)$ is as given in Theorem 2.1.

Proof. Set $f(z) = az + b$. Then

$$F(z) = \frac{1}{2}[f(z) + \sqrt{f^2(z) - 4z}].$$

Moreover, we have

$$f(\alpha^2) = \alpha^2(1+\gamma)^2 + 2\alpha\gamma^2 + (1-\gamma)^2 = c(\gamma,\alpha)^2 + 2\alpha;$$

therefore

$$\sqrt{f^2(\alpha^2) - 4\alpha^2} = \sqrt{[f(\alpha^2) - 2\alpha][f(\alpha^2) + 2\alpha]} = |c(\gamma,\alpha)|\sqrt{c(\gamma,\alpha)^2 + 4\alpha}.$$

Note that $f(\alpha^2)$ can be written as

$$f(\alpha^2) = \frac{1}{2}c(\gamma,\alpha)^2 + \frac{1}{2}\{c(\gamma,\alpha)^2 + 4\alpha\};$$

then

$$F(\alpha^2) = \frac{1}{2}[f(\alpha^2) + \sqrt{f^2(\alpha^2) - 4\alpha^2}] = \left(\frac{|c(\gamma,\alpha)| + \sqrt{c(\gamma,\alpha)^2 + 4\alpha}}{2}\right)^2.$$

It is easy to see that (3.4) is equivalent to

$$F(0) \leq F(z) \leq F(\alpha^2),$$

so it suffices to prove that $F(z)$ is a real and monotone increasing function in the interval $[0,\ 1)$. First we see that

$$\begin{aligned} ab &= [(1+\gamma)^2 + \gamma^2/\alpha]\,[\alpha\gamma^2 + (1-\gamma)^2] \\ &= \alpha\gamma^2(1+\gamma)^2 + (1-\gamma^2)^2 + \gamma^4 + \frac{\gamma^2(1-\gamma)^2}{\alpha} \\ &= 1 + \left[\sqrt{\alpha}\gamma(1+\gamma) - \frac{\gamma(1-\gamma)}{\sqrt{\alpha}}\right]^2; \end{aligned}$$

thus $ab \geq 1$, and

$$(az+b)^2 - 4z = (az + 2\sqrt{z} + b)\left[\left(\sqrt{az} - \frac{1}{\sqrt{a}}\right)^2 + \frac{ab-1}{a}\right] \geq 0,$$

which indicates that $F(z)$ is real in the interval $[0,\ 1)$.

On the other hand, taking the derivative of F, we have

$$F'(z) = \frac{f'(z)[f(z) + \sqrt{f^2(z) - 4z}] - 2}{2\sqrt{f^2(z) - 4z}}, \quad z \in [0,\ 1);$$

then the condition that $F'(z) \geq 0$ is equivalent to

$$f'(z)[\sqrt{f^2(z) - 4z}] \geq 2 - f'(z)f(z), \quad z \in [0,\ 1). \tag{3.5}$$

Using $ab \geq 1$, we obtain (note that $f'(z) = a$)

$$z[f'(z)]^2 - f(z)f'(z) + 1 = a^2 z - a(az + b) + 1 = 1 - ab \leq 0, \quad z \in [0,\ 1).$$

This implies

$$[f'(z)]^2[f^2(z) - 4z] \geq [2 - f'(z)f(z)]^2, \quad z \in [0,\ 1),$$

which guarantees the inequality (3.5). (Note that $f'(z)\sqrt{f^2(z) - 4z} \geq 0$.) □

LEMMA 3.4. *Let γ be defined as in Theorem 2.1 and $\varphi(z) = \alpha z^2 + (1 - z)^2$; then*

$$\varphi(z) \leq \varphi(\gamma) \quad \textit{for all} \quad z \in \left[\frac{1-\beta}{2\lambda_0},\ \frac{1+\beta}{2}\right].$$

Proof. We can directly verify that

$$\varphi'(z) \begin{cases} < 0, & z < (1+\alpha)^{-1}; \\ = 0, & z = (1+\alpha)^{-1}; \\ > 0, & z > (1+\alpha)^{-1}. \end{cases}$$

So the maximum value of $\varphi(z)$ is

$$\max\left\{\varphi\left(\frac{1-\beta}{2\lambda_0}\right),\ \varphi\left(\frac{1+\beta}{2}\right)\right\}.$$

By the direct calculations we have

$$\varphi\left(\frac{1-\beta}{2\lambda_0}\right) = 1 - \frac{1-\beta}{\lambda_0} + \frac{(1+\alpha)(1-\beta)^2}{4\lambda_0^2}$$

and

$$\varphi\left(\frac{1+\beta}{2}\right) = 1 - (1+\beta) + \frac{(1+\alpha)(1+\beta)^2}{4}.$$

Thus

$$\varphi\left(\frac{1-\beta}{2\lambda_0}\right) - \varphi\left(\frac{1+\beta}{2}\right) = \left[1 - \frac{1+\alpha}{4}\left(1 + \beta + \frac{1-\beta}{\lambda_0}\right)\right]\left[(1+\beta) - \frac{1-\beta}{\lambda_0}\right].$$

Note that $\lambda_0 \geq 1$ and $\alpha < 1$; hence

$$\frac{1-\beta}{\lambda_0} \leq 1 - \beta \leq 1 + \beta$$

and

$$\frac{1+\alpha}{4}\left(1+\beta+\frac{1-\beta}{\lambda_0}\right) \le \frac{1+\alpha}{4}\left(1+\beta+1-\beta\right) < 1,$$

and we have

$$\varphi\left(\frac{1-\beta}{2\lambda_0}\right) - \varphi\left(\frac{1+\beta}{2}\right) \ge 0. \tag{3.6}$$

So $\varphi(z)$ reaches its maximum at $z = (1-\beta)/(2\lambda_0)$. By the definition of γ it is easy to see that

$$\frac{1-\beta}{2\lambda_0} \ge \gamma;$$

this and the monotonicity of φ implies the desired estimate of Lemma 3.4. □

The following spectral bounds will be directly used in the spectral estimates of the propagation matrix associated with Algorithm 2.1.

LEMMA 3.5. *Let Q be a given symmetric positive definite matrix with its eigenvalues lying in the interval* $[\frac{\theta_i(1-\beta)}{\lambda_0}, \theta_i(1+\beta)]$ (*cf. Lemma* 3.2(ii)), *and F_i is a matrix given by*

$$F_i = \begin{pmatrix} \alpha_i(I+Q) & -\sqrt{\alpha}Q \\ \sqrt{\alpha}^{-1}\alpha_i Q & (I-Q) \end{pmatrix}.$$

Then the spectrum of F_i is bounded by ρ_0 (*defined in* (2.7)), *i.e.*, $\|F_i\| \le \rho_0$.

Proof. Let $\{\lambda_j\}_{j=1}^m$ be the positive eigenvalues of the matrix Q. It is easy to verify that

$$\|F_i\| = \max_{1\le j\le m}\left\|\begin{pmatrix} \alpha_i(1+\lambda_j) & -\sqrt{\alpha}\lambda_j \\ \sqrt{\alpha}^{-1}\alpha_i\lambda_j & 1-\lambda_j \end{pmatrix}\right\|. \tag{3.7}$$

To estimate $\|F_i\|$, it suffices to estimate the maximum eigenvalue of the matrix $\mathcal{F}_i^t\mathcal{F}_i$ with

$$\mathcal{F}_i = \begin{pmatrix} \alpha_i(1+\lambda_j) & -\sqrt{\alpha}\lambda_j \\ \sqrt{\alpha}^{-1}\alpha_i\lambda_j & 1-\lambda_j \end{pmatrix}.$$

The determinant of the matrix $\mathcal{F}_i^t\mathcal{F}_i$ can be simplified as follows:

$$\begin{aligned}
&[\alpha_i^2(1+\beta_j)^2 + \alpha^{-1}\alpha_i^2\beta_j^2]\,[(1-\beta_j)^2 + \alpha\beta_j^2] - \{\sqrt{\alpha}^{-1}\alpha_i\beta_j[1-\beta_j-\alpha(1+\beta_j)]\}^2 \\
&= \alpha_i^2(1-\beta_j^2)^2 + \alpha\alpha_i^2\beta_j^2(1+\beta_j)^2 + \alpha^{-1}\alpha_i^2\beta_j^2(1-\beta_j)^2 + \alpha_i^2\beta_j^4 \\
&\quad -\alpha^{-1}\alpha_i^2\beta_j^2[(1-\beta_j)^2 - 2\alpha(1-\beta_j^2) + \alpha^2(1+\beta_j)^2] \\
&= \alpha_i^2(1-\beta_j^2)^2 + \alpha\alpha_i^2\beta_j^2(1+\beta_j)^2 + \alpha^{-1}\alpha_i^2\beta_j^2(1-\beta_j)^2 + \alpha_i^2\beta_j^4 \\
&\quad -\alpha^{-1}\alpha_i^2\beta_j^2(1-\beta_j)^2 + 2\alpha_i^2\beta_j^2(1-\beta_j^2) - \alpha\alpha_i^2\beta_j^2(1+\beta_j)^2 \\
&= \alpha_i^2[(1-\beta_j^2)^2 + \beta_j^4 + 2\beta_j^2(1-\beta_j^2)] = \alpha_i^2[(1-\beta_j^2)+\beta_j^2]^2 = \alpha_i^2;
\end{aligned}$$

hence the characteristic equation of $\mathcal{F}_i^t\mathcal{F}_i$ is

$$\lambda^2 - [\alpha_i^2(1+\lambda_j)^2 + \alpha^{-1}\alpha_i^2\lambda_j^2 + (1-\lambda_j)^2 + \alpha\lambda_j^2]\lambda + \alpha_i^2 = 0.$$

Then the desired maximum eigenvalue is

$$\lambda^* = \left(f(\alpha_i, \lambda_j) + \sqrt{f^2(\alpha_i, \lambda_j) - 4\alpha_i^2} \right) /2 \tag{3.8}$$

with $f(\alpha_i, z)$ defined by

$$f(\alpha_i, z) = \alpha_i^2 (1+z)^2 + \alpha^{-1}\alpha_i^2\, z^2 + (1-z)^2 + \alpha\, z^2.$$

For a fixed α_i, the equation $f'(\alpha_i, z) = 0$ has a unique solution:

$$z = \beta_0 \equiv \frac{\alpha(1-\alpha_i^2)}{\alpha\alpha_i^2 + \alpha_i^2 + \alpha^2 + \alpha}.$$

Moreover, we have $f'(\alpha_i, z) < 0$ for $z < \beta_0$ and $f'(\alpha_i, z) > 0$ for $z > \beta_0$. Thus using the assumption on the range of the eigenvalues of Q, we have

$$\max_{1\le j\le m} \{f(\alpha_i, \lambda_j)\} \le \max \left\{ f\left(\alpha_i, \frac{\theta_i(1-\beta)}{\lambda_0}\right), f(\alpha_i, \theta_i(1+\beta)) \right\}. \tag{3.9}$$

Noting that

$$\alpha\alpha_i^2 + \alpha_i^2 + \alpha^2 + \alpha \le \alpha(1+\alpha)(1+\alpha_i) < 2\alpha(1+\alpha_i),$$

it follows from Lemma 3.1 that

$$\theta_i = \frac{1-\sqrt{1-\omega_i}}{2} \le \frac{1-\alpha_i}{2} \le \frac{\alpha(1-\alpha_i^2)}{\alpha\alpha_i^2 + \alpha_i^2 + \alpha^2 + \alpha}. \tag{3.10}$$

Using this, one can verify directly that

$$f(\alpha_i, \theta_i(1-\beta)) \ge f(\alpha_i, \theta_i(1+\beta)),$$

which, with the fact that $\lambda_0 \ge 1$, yields

$$f\left(\alpha_i, \frac{\theta_i(1-\beta)}{\lambda_0}\right) \ge f(\alpha_i, \theta_i(1+\beta)). \tag{3.11}$$

On the other hand, Lemma 3.1 implies that $\sqrt{1-\omega_i} \le \sqrt{1-\lambda_0^{-1}}$; hence

$$\theta_i = \frac{1-\sqrt{1-\omega_i}}{2} \ge \frac{1-\sqrt{1-\lambda_0^{-1}}}{2}$$

or

$$\frac{\theta_i(1-\beta)}{\lambda_0} \ge \frac{(1-\beta)}{2\lambda_0}\left(1-\sqrt{1-\lambda_0^{-1}}\right) = \gamma$$

with the γ given in Theorem 2.1. Therefore,

$$f\left(\alpha_i, \frac{\theta_i(1-\beta)}{\lambda_0}\right) \le f(\alpha_i, \gamma);$$

this together with (3.9) and (3.11) leads to

$$f(\alpha_i, \lambda_j) \le f(\alpha_i, \gamma), \quad j = 1, \ldots, m. \tag{3.12}$$

By (3.8), (3.12), and the definitions of $f(\alpha_i, \gamma)$ and $F(z)$, we have $\lambda^* \le F(\alpha_i^2)$. This result together with (3.7), Lemma 3.1, and the second inequality of Lemma 3.3 implies $\|F_i\| \le \rho_0$. □

With the help of Lemmas 3.1–3.5 above, we are now ready to show the convergence results in Theorems 2.1 and 2.2.

Proof of Theorem 2.1. As is true for classical iterative methods, the convergence proofs for most existing inexact Uzawa-type iterative methods are carried out with the natural error vectors $e_i^x = x - x_i$ and $e_i^y = y - y_i$ (cf. [3], [4], [17]). But this traditional analysis seems to be very difficult to follow in our current case with variable relaxation parameters, which is much more complicated technically. It is essential that we shall first estimate the residual f_i instead of the error vector e_i^x. Clearly, the residuals f_i and g_i can be represented in terms of e_i^x and e_i^y:

$$f_i = Ae_i^x + Be_i^y, \quad g_i = -B^t e_{i+1}^x. \tag{3.13}$$

By (2.3) and (3.13) we have

$$A^{\frac{1}{2}} e_{i+1}^x = A^{\frac{1}{2}}(e_i^x - \omega_i \hat{A}^{-1} f_i) = (I - \omega_i A^{\frac{1}{2}} \hat{A}^{-1} A^{\frac{1}{2}}) A^{-\frac{1}{2}} f_i - A^{-\frac{1}{2}} B e_i^y. \tag{3.14}$$

Using (2.4), Lemma 3.2(i), and (3.14) we obtain

$$\begin{aligned} A^{-\frac{1}{2}} B e_{i+1}^y &= A^{-\frac{1}{2}} B(e_i^y - \theta_i \tau_i \hat{C}^{-1} g_i) = A^{-\frac{1}{2}} B(e_i^y + Q_{Bi}^{-1} B^t e_{i+1}^x) \\ &= A^{-\frac{1}{2}} B[e_i^y + Q_{Bi}^{-1} B^t A^{-\frac{1}{2}} \left((I - \omega_i A^{\frac{1}{2}} \hat{A}^{-1} A^{\frac{1}{2}}) A^{-\frac{1}{2}} f_i - A^{-\frac{1}{2}} B e_i^y)\right] \\ &= A^{-\frac{1}{2}} B Q_{Bi}^{-1} B^t A^{-\frac{1}{2}} (I - \omega_i A^{\frac{1}{2}} \hat{A}^{-1} A^{\frac{1}{2}}) A^{-\frac{1}{2}} f_i \\ &\quad + (I - A^{-\frac{1}{2}} B Q_{Bi}^{-1} B^t A^{-\frac{1}{2}}) A^{-\frac{1}{2}} B e_i^y, \end{aligned} \tag{3.15}$$

while using (3.14) and (3.15) we have

$$\begin{aligned} A^{-\frac{1}{2}} f_{i+1} &= A^{\frac{1}{2}} e_{i+1}^x + A^{-\frac{1}{2}} B e_{i+1}^y \\ &= (I + A^{-\frac{1}{2}} B Q_{Bi}^{-1} B^t A^{-\frac{1}{2}})(I - \omega_i A^{\frac{1}{2}} \hat{A}^{-1} A^{\frac{1}{2}}) A^{-\frac{1}{2}} f_i \\ &\quad - (A^{-\frac{1}{2}} B Q_{Bi}^{-1} B^t A^{-\frac{1}{2}}) A^{-\frac{1}{2}} B e_i^y. \end{aligned} \tag{3.16}$$

Now let

$$B^t A^{-\frac{1}{2}} = U \Sigma V^t \tag{3.17}$$

with $\Sigma = (\Sigma_0 \; 0)$ being the singular value decomposition of the matrix $B^t A^{-\frac{1}{2}}$. As usual, U is an orthogonal $m \times m$ matrix and V is an orthogonal $n \times n$ matrix. The diagonal entries of the matrix Σ_0 are the singular values of $B^t A^{-\frac{1}{2}}$. Define

$$E_i^{xy} = \sqrt{\alpha} V^t A^{-\frac{1}{2}} f_i, \quad E_i^y = \Sigma^t U^t e_i^y.$$

By (3.15) and (3.16), we obtain

$$\begin{aligned} E_{i+1}^{xy} &= (I + V^t A^{-\frac{1}{2}} B Q_{Bi}^{-1} B^t A^{-\frac{1}{2}} V) V^t (I - \omega_i A^{\frac{1}{2}} \hat{A}^{-1} A^{\frac{1}{2}}) V E_i^{xy} \\ &\quad - \sqrt{\alpha} (V^t A^{-\frac{1}{2}} B Q_{Bi}^{-1} B^t A^{-\frac{1}{2}} V) E_i^y \end{aligned} \tag{3.18}$$

and

$$
\begin{aligned}
E^y_{i+1} =& \frac{1}{\sqrt{\alpha}}(V^tA^{-\frac12}BQ^{-1}_{Bi}B^tA^{-\frac12}V)V^t(I-\omega_iA^{\frac12}\hat{A}^{-1}A^{\frac12})VE^{xy}_i \\
&+(I-V^tA^{-\frac12}BQ^{-1}_{Bi}B^tA^{-\frac12}V)E^y_i. \qquad (3.19)
\end{aligned}
$$

Set

$$Q_{1i}\equiv V^t(I-\omega_iA^{\frac12}\hat{A}^{-1}A^{\frac12})V$$

and

$$Q_{2i}\equiv \Sigma^tU^tQ^{-1}_{Bi}U\Sigma = V^tA^{-\frac12}BQ^{-1}_{Bi}B^tA^{-\frac12}V\,;$$

then the propagation relations (3.18) and (3.19) may be written in the matrix form

$$\begin{pmatrix} E^{xy}_{i+1} \\ E^y_{i+1}\end{pmatrix} = \begin{pmatrix} (I+Q_{2i})Q_{1i} & -\sqrt{\alpha}Q_{2i} \\ \sqrt{\alpha}^{-1}Q_{2i}Q_{1i} & (I-Q_{2i})\end{pmatrix}\begin{pmatrix} E^{xy}_i \\ E^y_i\end{pmatrix}. \qquad (3.20)$$

Let E^{0y}_i and Q^0_{2i} denote the nonzero part of E^y_i and Q_{2i}, respectively, namely,

$$E^{0y}_i=\Sigma_0U^te^y_i,\quad Q^0_{2i}=\Sigma_0U^tQ^{-1}_{Bi}U\Sigma_0,$$

and set $\hat{Q}_{2i}=(Q^0_{2i},\ 0)^t$. Then we have from (3.20) that

$$\begin{pmatrix} E^{xy}_{i+1} \\ E^{0y}_{i+1}\end{pmatrix} = \begin{pmatrix} (I+Q_{2i})Q_{1i} & -\sqrt{\alpha}\hat{Q}_{2i} \\ \sqrt{\alpha}^{-1}\hat{Q}^t_{2i}Q_{1i} & (I-Q^0_{2i})\end{pmatrix}\begin{pmatrix} E^{xy}_i \\ E^{0y}_i\end{pmatrix}. \qquad (3.21)$$

Next we estimate the spectrum of the propagation matrix in (3.21). We first consider two cases: $f_i=0$; $f_i\neq 0$ but $\alpha_i=0$. Then we have by the definition of E^{xy}_i and α_i that

$$Q_{1i}E^{xy}_i=0 \quad\text{for}\quad f_i=0 \quad\text{or}\quad \alpha_i=0.$$

So we can write (3.21) as

$$\begin{pmatrix} E^{xy}_{i+1} \\ E^{0y}_{i+1}\end{pmatrix} = \begin{pmatrix} 0 & -\sqrt{\alpha}\hat{Q}_{2i} \\ 0 & (I-Q^0_{2i})\end{pmatrix}\begin{pmatrix} E^{xy}_i \\ E^{0y}_i\end{pmatrix} \equiv F_{0i}\begin{pmatrix} E^{xy}_i \\ E^{0y}_i\end{pmatrix}.$$

For the case that $f_i\neq 0$ but $\alpha_i=0$, an estimate of the norm $\|F_{0i}\|$ can be obtained directly later on, so we consider only the case that $f_i=0$ at the moment. Since

$$
\begin{aligned}
F^t_{0i}F_{0i} =& \begin{pmatrix} 0 & 0 \\ -\sqrt{\alpha}\hat{Q}^t_{2i} & (I-Q^0_{2i})\end{pmatrix}\begin{pmatrix} 0 & -\sqrt{\alpha}\hat{Q}_{2i} \\ 0 & (I-Q^0_{2i})\end{pmatrix} \\
=& \begin{pmatrix} 0 & 0 \\ 0 & \alpha(Q^0_{2i})^2+(I-Q^0_{2i})^2\end{pmatrix},
\end{aligned}
$$

it suffices to estimate the maximum eigenvalue of the matrix

$$Q_{0i}=\alpha(Q^0_{2i})^2+(I-Q^0_{2i})^2. \qquad (3.22)$$

Using (1.2) and (3.17), we have

$$Q^{-1}_{Bi}C=Q^{-1}_{Bi}U\Sigma V^tV\Sigma^tU^t=Q^{-1}_{Bi}U\Sigma^2_0U^t=(\Sigma_0U^t)^{-1}Q^0_{2i}(\Sigma_0U^t). \qquad (3.23)$$

Thus the matrix Q_{2i}^0 has the same eigenvalues as the matrix $Q_{Bi}^{-1}C$, and Lemma 3.2(ii) implies that the maximum eigenvalue of the matrix Q_{0i} defined in (3.22) is bounded above by the maximum of the function

$$\varphi(z) = \alpha z^2 + (1-z)^2, \quad z \in \left[\frac{(1-\beta)}{2\lambda_0}, \frac{(1+\beta)}{2}\right].$$

Here we have used the fact that $\theta_i = \frac{1}{2}$ for $f_i = 0$ by definition. Using (3.22), (3.4), and Lemmas 3.3 and 3.4 we have

$$\|F_{0i}\|^2 \le \alpha\gamma^2 + (1-\gamma)^2 \le F(\alpha^2) = \rho_0^2 \quad (\text{when} \quad f_i = 0). \tag{3.24}$$

Next, we consider the case that $f_i \neq 0$ and $\alpha_i > 0$. Write (3.21) in the form

$$\begin{pmatrix} E_{i+1}^{xy} \\ E_{i+1}^{0y} \end{pmatrix} = \begin{pmatrix} \alpha_i(I + Q_{2i}) & -\sqrt{\alpha}\hat{Q}_{2i} \\ \sqrt{\alpha}^{-1}\alpha_i\hat{Q}_{2i}^t & (I - Q_{2i}^0) \end{pmatrix} \begin{pmatrix} \alpha_i^{-1}Q_{1i} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} E_i^{xy} \\ E_i^{0y} \end{pmatrix}.$$

By the definitions of Q_{1i}, E_i^{xy}, and α_i, we have (note that V^t is an orthogonal matrix)

$$\begin{aligned} \|\alpha_i^{-1}Q_{1i}E_i^{xy}\|^2 &= \|\alpha_i^{-1}\sqrt{\alpha}V^t(I - \omega_i A^{\frac{1}{2}}\hat{A}^{-1}A^{\frac{1}{2}})A^{-\frac{1}{2}}f_i\|^2 \\ &= \alpha_i^{-2}\alpha\|(I - \omega_i A^{\frac{1}{2}}\hat{A}^{-1}A^{\frac{1}{2}})A^{-\frac{1}{2}}f_i\|^2 \\ &= \alpha_i^{-2}\alpha\alpha_i^2\|A^{-\frac{1}{2}}f_i\|^2 \\ &= \|\sqrt{\alpha}V^tA^{-\frac{1}{2}}f_i\|^2 = \|E_i^{xy}\|^2. \end{aligned}$$

Thus

$$\begin{aligned} \left\| \begin{pmatrix} \alpha_i^{-1}Q_{1i} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} E_i^{xy} \\ E_i^{0y} \end{pmatrix} \right\| &= \left\| \begin{pmatrix} \alpha_i^{-1}Q_{1i}E_i^{xy} \\ E_i^{0y} \end{pmatrix} \right\| \\ &= \left(\|\alpha_i^{-1}Q_{1i}E_i^{xy}\|^2 + \|E_i^{0y}\|^2 \right)^{\frac{1}{2}} \\ &= \left\| \begin{pmatrix} E_i^{xy} \\ E_i^{0y} \end{pmatrix} \right\|. \end{aligned}$$

Therefore

$$\left\| \begin{pmatrix} E_{i+1}^{xy} \\ E_{i+1}^{0y} \end{pmatrix} \right\| \le \left\| \begin{pmatrix} \alpha_i(I + Q_{2i}) & -\sqrt{\alpha}\hat{Q}_{2i} \\ \sqrt{\alpha}^{-1}\alpha_i\hat{Q}_{2i}^t & (I - Q_{2i}^0) \end{pmatrix} \right\| \left\| \begin{pmatrix} E_i^{xy} \\ E_i^{0y} \end{pmatrix} \right\|.$$

It is clear that

$$\begin{pmatrix} \alpha_i(I + Q_{2i}) & -\sqrt{\alpha}\hat{Q}_{2i} \\ \sqrt{\alpha}^{-1}\alpha_i\hat{Q}_{2i}^t & (I - Q_{2i}^0) \end{pmatrix} = \begin{pmatrix} \alpha_i(I + Q_{2i}^0) & 0 & -\sqrt{\alpha}Q_{2i}^0 \\ 0 & \alpha_i I & 0 \\ \sqrt{\alpha}^{-1}\alpha_i Q_{2i}^0 & 0 & (I - Q_{2i}^0) \end{pmatrix}.$$

Let F_i be the matrix defined in Lemma 3.5 but with Q replaced by Q_{2i}^0; then we have

$$\left\| \begin{pmatrix} \alpha_i(I + Q_{2i}) & -\sqrt{\alpha}\hat{Q}_{2i} \\ \sqrt{\alpha}^{-1}\alpha_i\hat{Q}_{2i}^t & (I - Q_{2i}^0) \end{pmatrix} \right\| = \left\| \begin{pmatrix} \alpha_i I & 0 \\ 0 & F_i \end{pmatrix} \right\| = \max\{\alpha_i, \|F_i\|\} \le \max\{\alpha, \|F_i\|\}.$$

Noting that $\alpha \le \rho_0$ by the definition of ρ_0 and $|c(\gamma, \alpha)| \ge 0$, the desired estimate now follows from Lemma 3.5.

For the case that $f_i \neq 0$ and $\alpha_i = 0$, F_{0i} has the same form as F_i. Thus $\|F_{0i}\| \leq \rho_0$ by Lemma 3.5. This proves (2.7) for all possible cases.

Finally we show (2.9). We first claim that

$$|1-\gamma-\alpha(1+\gamma)| < 1-\alpha. \tag{3.25}$$

In fact, since

$$\lambda_0 \geq \kappa_1 = \frac{1+\alpha}{1-\alpha},$$

we have

$$\sqrt{1-\frac{1}{\lambda_0}} \geq \sqrt{\frac{2\alpha}{1+\alpha}} \geq \alpha.$$

Thus

$$\gamma = \frac{1-\beta}{2\lambda_0}\left(1-\sqrt{1-\frac{1}{\lambda_0}}\right) < 1-\alpha,$$

which implies (3.25) using $\gamma > 0$ and $\alpha < 1$. Now by (3.25) and the definition of ρ_0 in (2.7)

$$\rho_0 < \frac{|1-\gamma-\alpha(1+\gamma)|+(1+\alpha)}{2} = \begin{cases} \frac{1-\alpha-\gamma(1+\alpha)+(1+\alpha)}{2}, & 0<\gamma\leq\frac{1-\alpha}{1+\alpha}, \\ \frac{\gamma(1+\alpha)-(1-\alpha)+(1+\alpha)}{2}, & \frac{1-\alpha}{1+\alpha}<\gamma<1-\alpha\,. \end{cases}$$

This completes the proof of Theorem 2.1. □

Proof of Theorem 2.2. For ease of notation, we let

$$\tilde{Q}_{1i} = I - \omega_i A^{\frac{1}{2}}\hat{A}^{-1}A^{\frac{1}{2}}, \quad \tilde{Q}_{2i} = A^{-\frac{1}{2}}BQ_{Bi}^{-1}B^tA^{-\frac{1}{2}}.$$

Then (3.16) can be written as (replacing i by $i-1$)

$$A^{-\frac{1}{2}}f_i = (I+\widetilde{Q}_{2i})\tilde{Q}_{1i}A^{-\frac{1}{2}}f_{i-1} - \widetilde{Q}_{2i}A^{-\frac{1}{2}}Be_{i-1}^y.$$

Applying Young's inequality, we obtain for any positive η that

$$\|A^{-\frac{1}{2}}f_i\|^2 \leq (1+\eta)\|(I+\widetilde{Q}_{2i})\tilde{Q}_{1i}A^{-\frac{1}{2}}f_{i-1}\|^2 + (1+\eta^{-1})\|\widetilde{Q}_{2i}A^{-\frac{1}{2}}Be_{i-1}^y\|^2. \tag{3.26}$$

By the proof of Theorem 2.1 we know that $\widetilde{Q}_{2i}$ has the same positive eigenvalues as the matrix $Q_{Bi}^{-1}C$. Hence, Lemma 3.2(ii) infers that the eigenvalues of $\widetilde{Q}_{2i}$ lie in the interval [0, 1], namely,

$$\|\widetilde{Q}_{2i}\| \leq 1, \quad \|I+\widetilde{Q}_{2i}\| \leq 2;$$

combining with (3.26) and Lemma 3.1, this leads to

$$\begin{aligned}\|A^{-\frac{1}{2}}f_i\|^2 &\leq (1+\eta)4\alpha^2\|A^{-\frac{1}{2}}f_{i-1}\|^2 + (1+\eta^{-1})\|A^{-\frac{1}{2}}Be_{i-1}^y\|^2 \\ &= 4\alpha(1+\eta)\|\sqrt{\alpha}A^{-\frac{1}{2}}f_{i-1}\|^2 + (1+\eta^{-1})\|e_{i-1}^y\|_C^2\,;\end{aligned}$$

taking $\eta = (4\alpha)^{-1}$ and using Theorem 2.1, we have

$$\|A^{-\frac{1}{2}}f_i\| \leq \sqrt{1+4\alpha}\rho^{i-1}|||E_0|||.$$

Now Theorem 2.2 follows immediately from the identity $A^{\frac{1}{2}}e_i^x = A^{-\frac{1}{2}}f_i - A^{-\frac{1}{2}}Be_i^y$, the triangle inequality, and Theorem 2.1. □

We end this section with some remarks on the selection of the parameter θ_i in Algorithm 2.1. As we see, the parameter θ_i has been used in the convergence rate analysis (cf. the inequality (3.10)). We next illustrate in a more direct manner why we have to introduce such a parameter and why we suggest choosing θ_i using (2.5). It is easy to find out from the proof of Theorem 2.1 that the sufficient and necessary condition for Algorithm 2.1 to converge is $\|F_i\| < 1$, where F_i is essentially the propagation matrix of Algorithm 2.1. This is equivalent to the condition that $\lambda^* < 1$ (cf. 3.8), that is,

$$\sqrt{f^2(\alpha_i,\lambda_j) - 4\alpha_i^2} < 2 - f(\alpha_i,\lambda_j)$$

or

$$f^2(\alpha_i,\lambda_j) - 4\alpha_i^2 < 4 - 4f(\alpha_i,\lambda_j) + f^2(\alpha_i,\lambda_j), \quad f(\alpha_i,\lambda_j) \le 2.$$

Namely,

$$f(\alpha_i,\lambda_j) < 1 + \alpha_i^2.$$

By the definition of $f(\alpha_i,\lambda_j)$, this condition is equivalent to

$$0 < \lambda_j < \frac{2\alpha(1-\alpha_i^2)}{\alpha\alpha_i^2 + \alpha_i^2 + \alpha^2 + \alpha}\,. \tag{3.27}$$

From Lemma 3.2(ii) and (3.23) we know that $\lambda_j \in [\theta_i(1-\beta)/\lambda_0, \theta_i(1+\beta)]$. Clearly (3.27) holds if θ_i is chosen such that

$$0 < \theta_i < \frac{2\alpha(1-\alpha_i^2)}{(\alpha\alpha_i^2 + \alpha_i^2 + \alpha^2 + \alpha)(1+\beta)}. \tag{3.28}$$

But since the paramaters α, β, and α_i are not easily computable, it is impractical to choose θ_i using the criterion (3.28). To find a more practical way of choosing θ_i, we further relax the condition (3.27). By Lemma 3.1, we know $\alpha_i \le \alpha$; hence

$$\alpha\alpha_i^2 + \alpha_i^2 + \alpha^2 + \alpha = (1+\alpha)\,\alpha\left(1 + \frac{\alpha_i}{\alpha}\alpha_i\right) < 2\alpha(1+\alpha_i), \tag{3.29}$$

so (3.27) is still satisfied if

$$0 < \lambda_j \le 1 - \alpha_i, \quad j = 1,\ldots,m. \tag{3.30}$$

For this we need to choose θ_i such that

$$0 < \theta_i(1+\beta) \le 1 - \alpha_i, \quad j = 1,\ldots,m; \tag{3.31}$$

this, with the relation $\alpha_i < \sqrt{1-\omega_i}$ from Lemma 3.1, yields the following selection criterion for θ_i:

$$\theta_i \le \frac{1-\sqrt{1-\omega_i}}{2}. \tag{3.32}$$

Namely, any positive θ_i satisfying (3.32) guarantees the convergence of Algorithm 2.1. However, using (3.8) and the monotone decreasing property of $f(\alpha_i, z)$ for $z < \beta_0$ we know that the larger the parameter θ_i is, the faster Algorithm 2.1 converges, namely, the choice

$$\theta_i < \frac{1-\sqrt{1-\omega_i}}{2} \quad \left(\le \frac{1-\alpha_i}{2} \le \beta_0\right)$$

will result in a convergence slower than the equality case. This is why we choose the equality case for θ_i in Theorem 2.1.

Note that the condition (3.32) is very conservative and it is obtained under the worst case: $\alpha \to 1^-$ (cf. (3.29)) and $\beta \to 1^-$ (cf. (3.31)). Therefore the choice

$$\theta_i > \frac{1-\sqrt{1-\omega_i}}{2}$$

is also possible. We omit the detailed discussion about this possibility here.

Finally, we add the additional observation that when α is small the condition (3.27) becomes $0 < \lambda_j < 2$ (the last term of (3.27) tends to 2^- as $\alpha \to 0$), which is satisfied if $\theta_i(1+\beta) < 2$ or $\theta_i \le 1$. Thus we can take $\theta_i = \omega_i \le 1$ to speed up the convergence of Algorithm 2.1 in this case.

Summarizing the above, and noting that

$$0.25\omega_i < \frac{1-\sqrt{1-\omega_i}}{2} < 0.5\omega_i\,,$$

we can conclude that the convergence of Algorithm 2.1 will speed up in the following order:

$$\theta_i = 0.25\omega_i\,,\ \frac{1-\sqrt{1-\omega_i}}{2},\ 0.5\omega_i\,,\ \omega_i$$

in the case that Algorithm 2.1 converges with $\theta_i = 0.5\omega_i$ and ω_i. This matches well with our numerical results; see Tables 4.1 and 4.2.

4. Numerical experiments. In this section, we apply our new Algorithm 2.1 of section 2, Algorithm 1.1 of [4], and the preconditioned MINRES method [18] to solve the two-dimensional generalized Stokes problem and a system of purely algebraic equations. Let Ω be the unit square in R^2, and $L_0^2(\Omega)$ be the set of all square integrable functions with zero mean values over Ω, and let $H^1(\Omega)$ be the usual Sobolev space of order one. The space $H_0^1(\Omega)$ consists of those functions in $H^1(\Omega)$ with vanishing traces on $\partial\Omega$.

Our first example is the generalized Stokes problem whose variational formulation reads as follows: Find $(u,p) \in (H_0^1(\Omega))^2 \times L_0^2(\Omega)$ such that

$$(\mu(x)\nabla u, \nabla v) - (p, \nabla\cdot v) = (f, v), \quad \text{for all } v \in (H_0^1(\Omega))^2, \tag{4.1}$$

$$(q, \nabla\cdot u) = (q, g), \quad \text{for all } q \in L_0^2(\Omega), \tag{4.2}$$

where $f \in (L^2(\Omega))^2$, $g \in L^2(\Omega)$, and $\mu \in L^\infty(\Omega)$ with $\mu(x) \ge c > 0$ almost everywhere in Ω.

We use one of the well-known conforming Taylor–Hood elements, which have been widely used in engineering, to solve the system (4.1)–(4.2). For any positive integer N, a triangulation $\mathcal{T}^h$ of Ω is obtained by dividing Ω into $N \times N$ subsquares with side lengths of $h = 1/N$. Let $X_h \subset H_0^1(\Omega)$ and $M_h \subset H^1(\Omega) \cap L_0^2(\Omega)$ be the usual continuous Q_2 and Q_1 finite element spaces defined on $\mathcal{T}^h$, respectively

Table 4.1
Number of iterations for Algorithm 2.1.

N	$\theta_i = \omega_i^{-1}$	$\theta_i = 1$	$\theta_i = \omega_i$	$\theta_i = 0.5\omega_i$	$\theta_i = \frac{1-\sqrt{1-\omega_i}}{2}$	$\theta_i = 0.25\omega_i$
8	638	203	35	39	41	46
16	154	44	36	41	42	46
32	153	45	36	40	42	46
48	154	45	37	40	41	47
64	154	44	36	41	42	46

Table 4.2
Number of iterations for Algorithm 1.1 *(left) and the MINRES method (right).*

N	8	16	32	48	64	N	8	16	32	48	64
Alg. 1.1	917	300	58	93	95	MINRES	63	55	51	50	50
N	8	16	32	48	64	N	8	16	32	48	64
Alg. 1.1	92	85	76	75	75	MINRES	56	65	65	66	66

(cf. [6, 10]). The total number of unknowns for this finite element is $n+m = [2(2N-1)^2] + [(N+1)^2 - 1]$; e.g., the total unknowns are 36482 for $N = 64$. The finite element approximation of the above Stokes system can be formulated as follows: Find $(u_h, p_h) \in X_h^2 \times M_h$ such that

$$(\mu(x)\nabla u_h, \nabla v) - (p_h, \nabla \cdot v) = (f, v), \qquad \text{for all } v \in X_h^2, \tag{4.3}$$

$$(q, \nabla \cdot u_h) = (q, g), \qquad \text{for all } q \in M_h. \tag{4.4}$$

It is known that the inf-sup condition is satisfied by the pair (X_h^2, M_h) (see [6]), thus the Schur complement matrix $C = B^tA^{-1}B$ associated with the system (4.3)–(4.4) has a condition number independent of h. As in [5], [18], we take the variable coefficient μ to be $\mu = 1 + x_1x_2 + x_1^2 - x_2^2/2$. We know that the corresponding matrix A is block diagonal with two copies of a discrete Laplace operator on the diagonal if $\mu = 1$, and so it can be solved by the fast Poisson solver. Therefore it is natural to choose this fast solver $\hat{A}$ as the preconditioner of A. In fact, the matrix $\hat{A}^{-1}A$ is well-conditioned since we have

$$0.5\,(\hat{A}z, z) \le (Az, z) \le 2.5\,(\hat{A}z, z). \tag{4.5}$$

Thus the matrix $B^t\hat{A}^{-1}B$ is also well-conditioned. In fact, it is spectrally equivalent to h^2I (cf. [19]); that is, we can choose $\hat{C} = h^2I$.

In most applications, the condition numbers κ_1 and κ_2 are not very large; otherwise all iterative methods for the saddle-point problems perform without any essential difference. It is clear that the parameter ω_i has a small range in this case, and we can roughly estimate the maximum and minimum eigenvalues of the matrix $\hat{A}^{-1}A$ based on several values of ω_i. In fact, when the system (4.3)–(4.4) is solved by Algorithm 2.1 with these preconditioners, the computational results (set $\theta_i = 1$) indicate that the parameter ω_i lies between 0.46 and 0.93 for $1 \le i \le 4$, which reflects roughly the range of the eigenvalues of the matrix $\hat{A}^{-1}A$.

In order to see whether assumption (2.6) is necessary for the convergence of Algorithm 2.1, we do not scale the preconditioner $\hat{A}$, so condition (2.6) is violated. The numerical results show that our Algorithm 2.1 converges well; the number of iterations is listed in Table 4.1. Note that all the initial guesses for the algorithms tested in this section are taken to be zero and the algorithms are terminated when the following relative error reaches 1.0×10^{-5}:

$$\varepsilon = \frac{\|Mu_i - b\|}{\|Mu_0 - b\|},$$

where M and $b = (b_1\ b_2)^t$ are the coefficient matrix and the right-hand side vector of the algebraic system corresponding to (4.3)–(4.4) and $u_i = (x_i\ y_i)^t$ is the ith iterate of the algorithms to be tested. Here we take the vector $b = Mu$ with the solution $u = (x\ y)^t$, and x and y are two vectors with all components being 1.0 and 0.5, respectively. From Table 4.1 we can see the importance of choosing a different θ_i other than $\theta_i = \omega_i^{-1}$. Also, one can find out that the convergence of Algorithm 2.1 is nearly independent of the mesh size h.

The inexact Uzawa Algorithm 1.1 is convergent if the two preconditioners for A and C satisfy the conditions (3.2) and (2.3) of [4]. Using (4.5), one can verify that these two conditions are indeed satisfied if we take the two preconditioners to be $2.5\hat{A}$ and $2I$ for A and C, respectively. Thus, we can also apply Algorithm 1.1 to solve the system (4.3)–(4.4). However, the convergence is a bit slow; see Table 4.2 (upper left). When the preconditioner $2I$ for C is replaced by $h^2 I$, which is spectrally equivalent to C (cf. [19]), Algorithm 1.1 converges slightly faster; see Table 4.2 (lower left). The main reason for the slow convergence in this case is that the parameter γ defined by (2.4) of [4] is close to one. Also it is difficult to achieve an accurate estimate on this parameter γ because of the difficulty of estimating the maximum eigenvalue of the matrix $\hat{C}^{-1}C$.

Then we applied the preconditioned MINRES method (cf. [16], [18]) with a block diagonal preconditioner with diagonal blocks being $\hat{A}$ and $\hat{C} = 0.01I$ or $\hat{C} = h^2 I$ (spectrally equivalent to C; cf. [19]) to solve the system (4.3)–(4.4). The number of iterations is listed in the upper right of Table 4.2 for $\hat{C} = 0.01I$ and in the lower right for $\hat{C} = h^2 I$. We remark that different constant scalings for $\hat{C}$ affect the convergence of the MINRES method greatly; see the comments at the end of this section.

Our second example is a system of purely algebraic equations. We define the matrices $A = (a_{ij})_{n\times n}$ and $B = (b_{ij})_{n\times m}$ $(n \geq m)$ in (1.1) as follows:

$$a_{ij} = \begin{cases} i+1, & i = j, \\ 1, & |i-j| = 1, \\ 0, & \text{otherwise}; \end{cases} \qquad b_{ij} = \begin{cases} j, & i = j + n - m, \\ 0, & \text{otherwise.} \end{cases}$$

The preconditioners $\hat{A} = (\hat{a}_{ij})_{n\times n}$ and $\hat{C} = (\hat{c}_{ij})_{m\times m}$ are defined by

$$\hat{a}_{ij} = \begin{cases} i+2, & i = j\,, \\ 0, & i \neq j\,; \end{cases} \qquad \hat{c}_{ij} = \begin{cases} k\,(i^2+3), & i = j\,, \\ 0, & i \neq j\,, \end{cases}$$

where k is a scaling constant. The right-hand side vectors f and g are taken such that the exact solutions x and y are both vectors with all components being 1.

Assumption (2.6) is violated again with this example. However, Algorithm 2.1 still converges well; see the number of iterations listed in Table 4.3. The convergence of Algorithm 1.1 and the preconditioned MINRES method with two different scaling constants, $k = 1, 1/200$, are reported in Tables 4.4 and 4.5.

TABLE 4.3
Number of iterations for Algorithm 2.1.

n	m	$\theta_i = \omega_i^{-1}$	$\theta_i = 1$	$\theta_i = \omega_i$	$\theta_i = 0.5\omega_i$	$\theta_i = \frac{1-\sqrt{1-\omega_i}}{2}$	$\theta_i = 0.25\omega_i$
200	150	15	15	15	17	19	38
400	300	16	16	16	17	18	38
800	600	17	17	17	18	18	38
1600	1200	17	17	17	17	18	39

TABLE 4.4
Iterations for Algorithm 1.1 *with different scalings: $k = 1,\ 1/200$.*

n	200	400	800	1600
m	150	300	600	1200
$k = 1$	1892	3759	> 5000	> 5000

n	200	400	800	1600
m	150	300	600	1200
$k = 1/200$	diverge	24	34	71

TABLE 4.5
Iterations for the preconditioned MINRES method with different scalings $k = 1,\ 1/200$.

n	200	400	800	1600
m	150	300	600	1200
$k = 1$	33	35	38	39

n	200	400	800	1600
m	150	300	600	1200
$k = 1/200$	22	22	22	23

From the above numerical results and many more tests we have not reported here, one can observe that different scalings for the preconditioner $\hat{C}$ greatly affect the convergence of Algorithm 1.1 and the preconditioned MINRES method. For example, Algorithm 1.1 converges (slowly) when the scaling constant $k = 1$, but it may diverge (the errors do not decrease) when $k = 1/200$; see Table 4.4. Such behaviors also happen for the preconditioned MINRES method (cf. [16], [18] and also see Table 4.5), whose convergence rate is known to depend on the ratio $\lambda_{\min}/\lambda'_{\min}$, where $\lambda_{\min}$ and $\lambda'_{\min}$ are, respectively, the minimal eigenvalues of $\hat{A}^{-1}A$ and $\hat{C}^{-1}H$ with $H = B^t\hat{A}^{-1}B$ (cf. [18]). So it is important for these algorithms to have good a priori estimates on the minimum or maximum eigenvalues of the matrix $\hat{C}^{-1}C$ or $\hat{C}^{-1}H$ in order to find an effective scaling for the preconditioner $\hat{C}$. But such a priori estimates are usually very difficult to achieve in practical applications, even when $\hat{C}^{-1}C$ is well-conditioned; e.g., this is the case with our first example; see the system (4.3)–(4.4). One of the advantages of our Algorithm 2.1 is to have relieved such a troublesome estimate, and different scalings for the preconditioner $\hat{C}$ do not affect the convergence of our Algorithm 2.1, which is easily seen from the algorithm itself.

Acknowledgments. The authors wish to thank the anonymous referees for many constructive comments that improved the paper greatly.

REFERENCES

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

[2] O. AXELSSON, *Numerical algorithms for indefinite problems*, in Elliptic Problem Solvers, Academic Press, New York, 1984, pp. 219–232.

[3] R. BANK, B. WELFERT, AND H. YSERENTANT, *A class of iterative methods for solving saddle point problems*, Numer. Math., 56 (1990), pp. 645–666.

[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 (1997), pp. 1072–1092.

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

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

[7] X. Chen, *On preconditioned Uzawa methods and SOR methods for saddle-point problems*, J. Comput. Appl. Math., 100 (1998), pp. 207–224.

[8] Z. Chen, Q. Du, and J. Zou, *Finite element methods with matching and nonmatching meshes for Maxwell equations with discontinuous coefficients*, SIAM J. Numer. Anal., 37 (2000), pp. 1542–1570.

[9] Z. Chen and J. Zou, *An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems*, SIAM J. Control Optim., 37 (1999), pp. 892–910.

[10] P. Ciarlet, *Basic error estimates for elliptic problems*, in Handbook of Numerical Analysis, Vol. II, P. Ciarlet and J.-L. Lions, eds., North-Holland, Amsterdam, 1991, pp. 17–351.

[11] H. C. Elman and G. H. Golub, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1645–1661.

[12] V. Girault and P. Raviart, *Finite Element Methods for Navier–Stokes Equations*, Springer-Verlag, Berlin, 1986.

[13] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, SIAM Stud. Appl. Math. 9, SIAM, Philadelphia, 1989.

[14] Q. Hu, G. Liang, and P. Sun, *Solving parabolic problems by domain decomposition methods with Lagrangian multipliers*, Math. Numer. Sin., 22 (2000), pp. 241–256.

[15] Y. Keung and J. Zou, *Numerical identifications of parameters in parabolic systems*, Inverse Problems, 14 (1998), pp. 83–100.

[16] A. Klawonn, *An optimal preconditioner for a class of saddle point problems with a penalty term*, SIAM J. Sci. Comput., 19 (1998), pp. 540–552.

[17] W. Queck, *The convergence factor of preconditioned algorithms of the Arrow–Hurwicz type*, SIAM J. Numer. Anal., 26 (1989), pp. 1016–1030.

[18] T. Rusten and R. Winther, *A preconditioned iterative method for saddlepoint problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 887–904.

[19] A. Wathen and D. Silvester, *Fast iterative solution of stabilised Stokes systems. Part* I: *Using simple diagonal preconditioners*, SIAM J. Numer. Anal., 30 (1993), pp. 630–649.