SIAM J. SCI. COMPUT.
Vol. 33, No. 5, pp. 2903–2926

PRECONDITIONING ITERATIVE METHODS FOR THE OPTIMAL CONTROL OF THE STOKES EQUATIONS*

TYRONE REES† AND ANDREW J. WATHEN‡

Abstract. Solving problems regarding the optimal control of partial differential equations (PDEs)—also known as PDE-constrained optimization—is a frontier area of numerical analysis. Of particular interest is the problem of flow control, where one would like to effect some desired flow by exerting, for example, an external force. The bottleneck in many current algorithms is the solution of the optimality system—a system of equations in saddle point form that is usually very large and ill conditioned. In this paper we describe two preconditioners—a block diagonal preconditioner for the minimal residual method and a block lower-triangular preconditioner for a nonstandard conjugate gradient method—which can be effective when applied to such problems where the PDEs are the Stokes equations. We consider only distributed control here, although we believe other problems could be treated in the same way. We give numerical results, and we compare these with those obtained by solving the equivalent forward problem using similar techniques.

Key words. saddle-point problems, PDE-constrained optimization, preconditioning, optimal control, linear systems, all-at-once methods, flow control, Stokes equations

AMS subject classifications. 49M25, 49K20, 65F10, 65N22, 65F50, 65N55

DOI. 10.1137/100798491

1. Introduction.

Suppose that we have a flow that satisfies the Stokes equations in some domain Ω with some given boundary condition, and that we have some mechanism—for example, the application of a magnetic field—to change the forcing term on the right-hand side of the PDE. Let $\widehat{\vec{v}}$ and $\hat{p}$ be given functions which are called the "desired states." Then the question is how do we choose the forcing term such that the velocity $\vec{v}$ and pressure p are as close as possible to $\widehat{\vec{v}}$ and $\hat{p}$, in some sense, while still satisfying the Stokes equations?

One way of formulating this problem is by minimizing a cost functional of tracking-type with the Stokes equations as a constraint, as follows:

$$\min_{\vec{v},p,\vec{u}} \frac{1}{2}\|\vec{v}-\widehat{\vec{v}}\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|p-\hat{p}\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|\vec{u}\|^2_{L^2(\Omega)} \tag{1.1}$$

$$\begin{aligned} \text{s.t.} -\nabla^2\vec{v} + \nabla p &= \vec{u} \quad \text{in } \Omega, \\ \nabla\cdot\vec{v} &= 0 \quad \text{in } \Omega, \\ \vec{v} &= \vec{w} \quad \text{on } \partial\Omega. \end{aligned}$$

Here $\vec{u}$ denotes the forcing term on the right-hand side, which is known as the control. In order for the problem to be well posed we also include the control in the cost functional, together with a Tikhonov regularization parameter β, which is usually chosen a priori. A constant $\alpha > 0$ is added in front of the desired pressure to enable us to penalize the pressure. We would normally take $\hat{p} = 0$. We specify a Dirichlet

*Received by the editors June 15, 2010; accepted for publication (in revised form) May 31, 2011; published electronically October 27, 2011.
http://www.siam.org/journals/sisc/33-5/79849.html
†Department of Computer Science, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada (tyronere@cs.ubc.ca).
‡Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, United Kingdom (wathen@maths.ox.ac.uk).

boundary condition with $\vec{v}$ taking some value $\vec{w}$—which may or may not be taken from the desired state—on the boundary.

There are two methods with which one can discretize this problem—we can either discretize the equations first and then optimize that system, or alternatively carry out the optimization first and then discretize the resulting optimality system. Since the Stokes equations are self-adjoint we will get the same discrete optimality system either way, provided the discretization methods are consistent between equations in the optimize-then-discretize technique. We will therefore only consider the discretize-then-optimize approach here.

Let $\{\vec{\phi}_j\}$, $j = 1, \ldots, n_v + n_\partial$ and $\{\psi_k\}$, $k = 1, \ldots, n_p$ be sets of finite element basis functions that form a stable mixed finite element discretization for the Stokes equations—see, for example, [13, Chapter 5] for further details—and let $\vec{v}_h = \sum_{i=1}^{n_v+n_\partial} V_i \vec{\phi}_i$ and $p_h = \sum_{i=1}^{n_p} P_i \psi_i$ be finite-dimensional approximations to $\vec{v}$ and p. Furthermore, let us also approximate the control from the velocity space, so $\vec{u}_h = \sum_{i=1}^{n_v} U_i \vec{\phi}_i$. The discrete Stokes equation is of the form

$$\begin{bmatrix} \underline{K} & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \underline{Q}_{\vec{v}} \\ 0 \end{bmatrix} \mathbf{u} + \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix},$$

where $\mathbf{v}$, $\mathbf{p}$, and $\mathbf{u}$ are the coefficient vectors in the expansions of $\vec{v}_h$, p_h, and $\vec{u}_h$ respectively,

$$\underline{K} = [\int_\Omega \nabla\vec{\phi}_i : \nabla\vec{\phi}_j], \qquad \mathbf{f} = \left[-\sum_{j=n_v+1}^{n_v+n_\partial} V_j \int_\Omega \nabla\vec{\phi}_i : \nabla\vec{\phi}_j \right],$$

$$B = [-\int_\Omega \psi_k \nabla\cdot\vec{\phi}_j],$$

$$\underline{Q}_{\vec{v}} = [\int_\Omega \vec{\phi}_i \cdot \vec{\phi}_j], \qquad \mathbf{g} = \left[\sum_{j=n_v+1}^{n_v+n_\partial} V_j \int_\Omega \psi_i \nabla\cdot\vec{\phi}_j \right].$$

Note that the coefficents V_j, $j = n_{v+1}, \ldots, n_v + n_\partial$ are fixed so that $\vec{v}_h$ interpolates the boundary data $\vec{w}$. In the above we have used the standard convention to denote Gram matrices obtained from the set of vector-valued basis functions by an underlined uppercase letter.

On discretizing, the cost functional becomes

$$\min \frac{1}{2}\mathbf{v}^T \underline{Q}_{\vec{v}} \mathbf{v} - \mathbf{v}^T\mathbf{b} + \frac{\alpha}{2}\mathbf{p}^T Q_p \mathbf{p} - \alpha\mathbf{p}^T\mathbf{d} + \frac{\beta}{2}\mathbf{u}^T \underline{Q}_{\vec{v}} \mathbf{u},$$

where $Q_p = [\int_\Omega \psi_i\psi_j]$, $\mathbf{b} = [\int_\Omega \widehat{\vec{v}}\vec{\phi}_i]$ and $\mathbf{d} = [\int_\Omega \widehat{p}\psi_i]$.

Let us introduce two vectors of Lagrange multipliers, $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. Then finding a critical point of the Lagrangian function gives the discrete optimality system of the form

$$\begin{bmatrix} \underline{Q}_{\vec{v}} & 0 & 0 & \underline{K} & B^T \\ 0 & \alpha Q_p & 0 & B & 0 \\ 0 & 0 & \beta\underline{Q}_{\vec{v}} & -\underline{Q}_{\vec{v}}^T & 0 \\ \underline{K} & B^T & -\underline{Q}_{\vec{v}} & 0 & 0 \\ B & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{p} \\ \mathbf{u} \\ \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \alpha\mathbf{d} \\ \mathbf{0} \\ \mathbf{f} \\ \mathbf{g} \end{bmatrix}. \tag{1.2}$$

It will be useful to relabel this system so that we group together blocks representing the PDE and the mass matrices which come from the states in the cost functional.

We label these blocks in calligraphic font. The system then becomes

$$\begin{bmatrix} \mathcal{Q} & 0 & \mathcal{K}^T \\ 0 & \beta \underline{Q}_{\vec{v}} & -\widehat{Q}^T \\ \mathcal{K} & -\widehat{Q} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{u} \\ \boldsymbol{\xi} \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \\ \mathbf{h} \end{bmatrix}, \tag{1.3}$$

where $\mathcal{Q} = \texttt{blkdiag}(\underline{Q}_{\vec{v}}, \alpha Q_p)$, $\mathcal{K} = \begin{bmatrix} K & B^T \\ B & 0 \end{bmatrix}$, $\widehat{Q} = [\underline{Q}_{\vec{v}}\ 0]^T$ and the vectors $\mathbf{w}$, $\boldsymbol{\xi}$, $\mathbf{c}$, and $\mathbf{h}$ take their obvious definitions. Note that we write $\mathcal{K}^T$ in (1.3) for clarity in the subsequent arguments, even though $\mathcal{K}$ is symmetric here. For more detail on the practicalities of discretizing control problems of this type, see, for example, Rees, Stoll, and Wathen [24]. Finding an efficient method to solve this system will be the topic of the remainder of the paper.

In section 2 we introduce two preconditioners that can be applied to this problem; one block diagonal, which we apply using the minimal residual method (MINRES) of Paige and Saunders [22], and one block lower triangular, which we use with the conjugate gradient (CG) method of Hestenes and Stiefel [17] applied with a nonstandard inner product. Both of these methods rely on good approximations to the $(1,1)$-block and the Schur complement, and we discuss suitable choices in sections 2.3 and 2.4, respectively. Finally, in section 3 we give numerical results.

2. Solution methods. The matrix in (1.3) is of saddle point form, i.e.,

$$\mathcal{A} := \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix}, \tag{2.1}$$

where $A := \texttt{blkdiag}(\mathcal{Q}, \beta \underline{Q}_{\vec{v}})$ and $C := [\mathcal{K}\ -\widehat{Q}]$. The matrix $\mathcal{A}$ is, in general, very large—the discrete Stokes equations are just one of its components—yet it is sparse. A good choice for solving such systems are iterative methods—in particular Krylov subspace methods. We will consider two such methods here—MINRES and CG in a nonstandard inner product—and extend the work of Rees, Dollar, and Wathen [23] and Rees and Stoll [25], respectively, to the case where the PDEs are the Stokes equations. Since the PDE is itself a saddle point problem, and hence the matrix representation is indefinite, significant complications arise here which are not present for simpler problems.

There is a large number of papers in the literature which deal with solving problems for the optimal control of PDEs. Below we comment on a few of these which share the philosophy of this paper. Most of these consider the model problem of the optimal control of Poisson's equation; it is not clear how easily they would be applied to the control of the Stokes equations and the additional difficulty this poses.

Schöberl and Zulehner [26] developed a preconditioner which is both optimal with respect to the problem size *and* with respect to the choice of regularization parameter β. This method was recently generalized slightly by Herzog and Sachs [16]. A multigrid-based preconditioner has also been developed by Biros and Dogan [3] which has both h- and β-independent convergence properties—where h is the mesh size—but it is not clear how their method would generalize to Stokes control. We note that the approximate reduced Hessian approximation used by Haber and Ascher [15] and Biros and Ghattas [4] also leads to a preconditioner with h-independence. Other solution methods employing multigrid for this and similar classes of problems were described by Borzì [5], Ascher and Haber [1], and Engel and Griebel [14].

2.1. Block diagonal preconditioners. It is well known that matrices of the form $\mathcal{A}$ are indefinite, and one choice of solution method for such systems is MINRES. For MINRES to be efficient for such a matrix we need to combine the method with a good preconditioner—i.e., a matrix $\mathcal{P}$ which is cheap to invert and which clusters the eigenvalues of $\mathcal{P}^{-1}\mathcal{A}$. One method that is often used—see [2, section 10.1.1] and the references therein—is to look for a block diagonal preconditioner of the form

$$\mathcal{P}_{\mathrm{bd}} = \begin{bmatrix} A_0 & 0 \\ 0 & S_0 \end{bmatrix}$$

for some matrices $A_0 \in \mathbb{R}^{2n_v+n_p}$, $S_0 \in \mathbb{R}^{n_v+n_p}$. Preconditioners of this form for the optimal control of Poisson's equation were discussed by Rees, Dollar, and Wathen [23].

It is well known (see, for example, [13, Theorem 6.6]) that if A, A_0, $CA^{-1}C^T$, and S_0 are positive definite matrices such that there exist constants δ, Δ, ϕ, and Φ such that the generalized Rayleigh quotients satisfy

$$\delta \le \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T A_0\mathbf{x}} \le \Delta, \quad \phi \le \frac{\mathbf{y}^T CA^{-1}C^T\mathbf{y}}{\mathbf{y}^T S_0\mathbf{y}} \le \Phi \tag{2.2}$$

for all vectors $\mathbf{x} \in \mathbb{R}^{2n_v+n_p}$ and $\mathbf{y} \in \mathbb{R}^{n_v+n_p}$, $\mathbf{x},\ \mathbf{y} \neq \mathbf{0}$, then the eigenvalues λ of ${\mathcal{P}_{\mathrm{bd}}}^{-1}\mathcal{A}$ are real and satisfy

$$\begin{gathered} \frac{\delta - \sqrt{\delta^2 + 4\Delta\Phi}}{2} \le \lambda \le \frac{\Delta - \sqrt{\Delta^2 + 4\phi\delta}}{2}, \\ \delta \le \lambda \le \Delta, \qquad \text{or} \\ \frac{\delta + \sqrt{\delta^2 + 4\delta\phi}}{2} \le \lambda \le \frac{\Delta + \sqrt{\Delta^2 + 4\Phi\Delta}}{2}. \end{gathered}$$

Therefore, if we can find matrices A_0 and S_0 that are cheap to invert and are good approximations to A and the Schur complement $CA^{-1}C^T$ in the sense that the constants in (2.2) are close to unity, then we will have a good preconditioner, since the eigenvalues of ${\mathcal{P}_{\mathrm{bd}}}^{-1}\mathcal{A}$ will be in three distinct clusters bounded away from 0. In the ideal case where $A_0 = A$ and $S_0 = CA^{-1}C^T$ we have $\delta = \Delta = \phi = \Phi = 1$. Then the preconditioned system will have precisely three eigenvalues, 1, $\frac{1+\sqrt{5}}{2}$, and $\frac{1-\sqrt{5}}{2}$, so MINRES would converge in three iterations [21].

2.2. Block lower-triangular preconditioners. Instead of MINRES we may want to use a CG method to solve a saddle point problem of the form (2.1). Since (2.1) is not positive definite, the standard CG algorithm cannot be used. However, the matrix

$$\begin{bmatrix} A_0 & 0 \\ C & -S_0 \end{bmatrix}^{-1} \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix}$$

is self-adjoint with respect to the inner product defined for $\mathbf{z}_1,\ \mathbf{z}_2 \in \mathbb{R}^{3n_v+2n_p}$ by $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle_{\mathcal{H}} := \mathbf{x}_1^T \mathcal{H}\mathbf{z}_2$, where

$$\mathcal{H} = \begin{bmatrix} A - A_0 & 0 \\ 0 & S_0 \end{bmatrix},$$

provided that this defines an inner product—i.e., when $A - A_0$ and S_0 are positive definite. Therefore, can we apply the CG algorithm with this inner product, along

with preconditioner

$$\mathcal{P}_{\text{lt}} = \begin{bmatrix} A_0 & 0 \\ C & -S_0 \end{bmatrix}.$$

This method was first described by Bramble and Pasciak in [9], and it has since generated a lot of interest—see, for example, [12, 18, 20, 26, 19, 28, 10]. This method was used in a control context by Rees and Stoll [25].

Convergence of this method again depends on the eigenvalue distribution of the preconditioned system—the clustering of the eigenvalues is given by, for example, Rees and Stoll [25, Theorem 3.1], and the relevant result is stated below in section 2.4. Note that in order to apply this preconditioner, only solves with A_0 and S_0 are needed; hence an implicit approximation (for example, multigrid) can be used. For more detail see, for example, Stoll [30].

One drawback of this method is that you need $A - A_0$ to be positive definite; this means that not just any approximation to A will do. This requirement usually results in having to find the eigenvalues of $A_0^{-1}A$ for a candidate A_0 and then adding an appropriate scaling ω so that $A > \omega A_0$—we will discuss this point further once we've described possible approximations A_0 in the following section.

2.3. Approximation of the (1,1) block. Suppose, for simplicity, that our domain $\Omega \subset \mathbb{R}^2$—the extension to three dimensions is obvious. If, as is usual, we use the same element space for all components in the velocity vector, and this has basis $\{\phi_i\}$, then $\underline{Q}_{\vec{v}} = \texttt{blkdiag}(Q_v, Q_v)$, where $Q_v = [\int_\Omega \phi_i \phi_j]$. Then the matrix A is just a block diagonal matrix composed of the mass matrices in the bases $\{\phi_i\}$ or $\{\psi_i\}$. Wathen [33] showed that for a general mass matrix, Q, if $D := \texttt{diag}(Q)$, then it is possible to calculate constants ξ and Ξ such that the eigenvalues of $D^{-1}Q$ are bounded below and above, respectively, by these constants. The values of ξ and Ξ depend on the elements used—for example, for $\mathbf{Q}_1$ elements $\xi = 1/4$, $\Xi = 9/4$ and for $\mathbf{Q}_2$ elements $\xi = 1/4$, $\Xi = 25/16$. The diagonal matrix itself would therefore be a reasonable candidate for A_0.

However, as A is in a sense "easy" to invert, it would help to have the best approximation here possible. Using the bounds described above we have all the information we need to use the relaxed Jacobi method accelerated by the Chebyshev semi-iteration, given as Algorithm 1. This is a method that is very cheap to use and, as demonstrated by Wathen and Rees in [32], is particularly effective in this case. In particular, since the eigenvalues of $D^{-1}Q$ are evenly distributed, there is very little difference between the convergence of this method and the CG method preconditioned with D. Note that since the CG algorithm is nonlinear, we cannot use it as a preconditioner for a stationary Krylov subspace method such as MINRES, unless run to convergence. The Chebyshev semi-iteration, on the other hand, is a linear method, and so a *fixed* number of iterations of this method can be used as a preconditioner.

Suppose we use this method to solve a system $Q\mathbf{x} = \hat{\mathbf{b}}$ for some right-hand side $\hat{\mathbf{b}}$. Then we can write every iteration as $\mathbf{x}^{(k)} = T_k^{-1}\hat{\mathbf{b}}$ for some matrix T_k implicitly defined by the method, which is independent of $\hat{\mathbf{b}}$. Let m denote the (fixed) number of Chebyshev semi-iterations. A larger m would make T_m a better approximation to Q, since $\mathbf{x}^{(m)}$ will be closer to the exact solution $\mathbf{x}$.

The upper and lower eigenvalue bounds can be obtained analytically—for example, Table I in Rees and Stoll [25] gives the upper and lower bounds for each m from 1 to 20 for a $\mathbf{Q}_1$ discretization. Let T_m^v and T_m^p denote the matrices defined implicitly by performing m steps of the Chebyshev semi-iteration on Q_v and Q_p. Then the reexist

ALGORITHM 1. m steps of the Chebyshev semi-iteration to approximate the solution of $Q\mathbf{x} = \hat{\mathbf{b}}$, where $\lambda(D^{-1}Q) \in [\xi, \Xi]$.

Choose $\mathbf{y}^{(0)}$, $w_0 = 0$, $(\mathbf{y}^{(-1)} = \mathbf{0})$
$\eta = (\xi + \Xi)/2$
$\rho = (\Xi - \xi)/(\xi + \Xi)$
for $k = 0, 1, \ldots, m-1$ **do**
  $w_{k+1} = \frac{1}{1 - \frac{\rho^2 w_k}{4}}$
  $\eta D\mathbf{z}^{(k)} = \hat{\mathbf{b}} - Q\mathbf{y}^{(k)}$
  $\mathbf{y}^{(k+1)} = w_{k+1}(\mathbf{z}^{(k)} + \mathbf{y}^{(k)} - \mathbf{y}^{(k-1)}) + \mathbf{y}^{(k-1)}$
end for

constants $\delta_m^v, \Delta_m^v, \delta_m^p$, and Δ_m^p independent of h such that $\delta_m^v \le \lambda((T_m^v)^{-1}Q_v) \le \Delta_m^v$ and $\delta_m^p \le \lambda((T_m^p)^{-1}Q_p) \le \Delta_m^p$. Hence, we can write

$$\delta_m \le \frac{\mathbf{x}^T A\mathbf{x}}{\mathbf{x}^T A_0 \mathbf{x}} \le \Delta_m \tag{2.3}$$

for all $\mathbf{x} \in \mathbb{R}^{2n_v+n_p}$, $\mathbf{x} \neq \mathbf{0}$, where $A_0 = \texttt{blkdiag}(T_m^v, T_m^v, \alpha T_m^p, \beta T_m^v, \beta T_m^v)$, $\delta_m = \min(\delta_m^v, \delta_m^p)$, and $\Delta_m = \max(\Delta_m^v, \Delta_m^p)$, both independent of the mesh size h. We therefore have an inexpensive way to make the bounds on $\lambda(A_0^{-1}A)$ as close to unity as required.

This choice of A_0 is all we need for the block diagonal preconditioner $\mathcal{P}_{\mathrm{bd}}$. However, for the block lower-triangular preconditioner $\mathcal{P}_{\mathrm{lt}}$ applied with CG in a nonstandard inner product we need $A - A_0 > 0$. This is not a problem here since we can work out these bounds accurately and inexpensively—even with a nonuniform mesh, it is just an $\mathcal{O}(n)$ calculation. Therefore, the scaling parameter ω, which ensures that $A > \omega A_0$, can be easily chosen; see Rees and Stoll [25] for more details.

2.4. Approximation of the Schur complement. Now consider the Schur complement $\frac{1}{\beta}\widehat{Q}\underline{Q}_{\vec{v}}^{-1}\widehat{Q}^T + \mathcal{K}\mathcal{Q}^{-1}\mathcal{K}^T =: S$. The dominant term in this sum, for all but the smallest values of β, is $\mathcal{K}\mathcal{Q}^{-1}\mathcal{K}^T$—the term that contains the PDE. Figure 2.1 shows the eigenvalue distribution for this approximation of S for a relatively coarse $\mathbf{Q}_2 - \mathbf{Q}_1$ discretization with $\beta = 0.01$. As we can see from the figure, the eigenvalues of $(\mathcal{K}\mathcal{Q}^{-1}\mathcal{K}^T)^{-1}S$ are nicely clustered, and so we could expect good convergence of MINRES if we took S_0 as $\mathcal{K}\mathcal{Q}^{-1}\mathcal{K}^T$. The effect of varying β is described in, for example, [31].

However, a preconditioner must be easy to invert, and solving a system with $\mathcal{K}\mathcal{Q}^{-1}\mathcal{K}^T$ requires two solves with the discrete Stokes matrix, which is not cheap. We therefore would like some matrix $\widetilde{\mathcal{K}}$—not necessarily symmetric—such that $\widetilde{\mathcal{K}}\mathcal{Q}^{-1}\widetilde{\mathcal{K}}^T$ approximates $\mathcal{K}\mathcal{Q}^{-1}\mathcal{K}^T$. In order to gain a theoretical understanding of this problem, we temporarily omit the mass matrices and look for a $\widetilde{\mathcal{K}}$ such that $\widetilde{\mathcal{K}}\widetilde{\mathcal{K}}^T$ approximates $\mathcal{K}\mathcal{K}^T$.

In order to achieve such an approximation, Braess and Peisker [7] show that it is *not* sufficient that $\widetilde{\mathcal{K}}$ approximates $\mathcal{K}$. Indeed, for the Stokes equations, Silvester and Wathen [27] showed that an ideal preconditioner is $\widehat{\mathcal{K}} = \texttt{blkdiag}(\underline{K}, M_p)$, but the eigenvalues of $(\widehat{\mathcal{K}}\widehat{\mathcal{K}}^T)^{-1}\mathcal{K}\mathcal{K}^T$ are not at all clustered, and the approximation of $\mathcal{K}\mathcal{K}^T$ is a poor one in this case. Suppose we wish to solve the equation $\mathcal{K}\tilde{\mathbf{w}} = \tilde{\mathbf{b}}$ for some right-hand side vector $\tilde{\mathbf{b}} \in \mathbb{R}^{n_v+n_p}$. Braess and Peisker, however, go on to

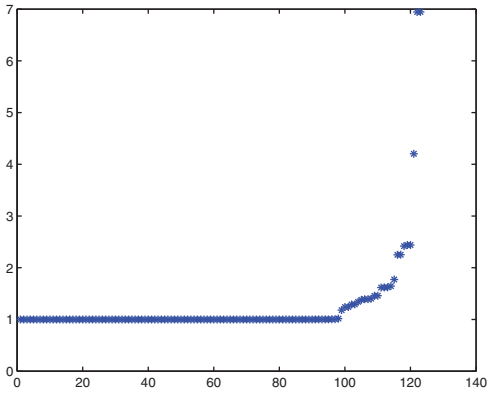


FIG. 2.1. *Eigenvalues of* $(\mathcal{K}\mathcal{Q}^{-1}\mathcal{K})^{-1}S$.

show that if we take an approximation $\mathcal{K}_n$ which is implicitly defined by an iteration $\mathbf{w}^{(n)} = \mathcal{K}_n^{-1}\tilde{\mathbf{b}}$, say, which converges to the solution $\tilde{w}$ in the sense that

$$\|\mathbf{w}^{(n)} - \tilde{\mathbf{w}}\| \leq \eta_n \|\tilde{\mathbf{w}}\|, \tag{2.4}$$

then $\eta_n = \|\mathcal{K}_n^{-1}\mathcal{K} - I\|$, where the matrix norm here is that induced from the vector norm in which we measure convergence—i.e., the spectral norm if we have convergence in the 2-norm. One can then show that for all $\mathbf{x} \in \mathbb{R}^{n_v+n_p},\ \mathbf{x} \neq \mathbf{0}$ [7, section 4],

$$(1-\eta_n)^2 \leq \frac{\mathbf{x}^T\mathcal{K}\mathcal{K}^T\mathbf{x}}{\mathbf{x}^T\mathcal{K}_n\mathcal{K}_n^T\mathbf{x}} \leq (1+\eta_n)^2. \tag{2.5}$$

Hence, approximation of $\mathcal{K}\mathcal{K}^T$ by $\mathcal{K}_n\mathcal{K}_n^T$ would be suitable in this case.

Of course, in practice we cannot simply ignore the mass matrices. Note that

$$\frac{\mathbf{x}^T\mathcal{K}\mathcal{Q}^{-1}\mathcal{K}^T\mathbf{x}}{\mathbf{x}^T\mathcal{K}_n\mathcal{Q}^{-1}\mathcal{K}_n^T\mathbf{x}} = \frac{\mathbf{y}^T\mathcal{Q}^{-1}\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \cdot \frac{\mathbf{x}^T\mathcal{K}\mathcal{K}^T\mathbf{x}}{\mathbf{x}^T\mathcal{K}_n\mathcal{K}_n^T\mathbf{x}} \cdot \frac{\mathbf{z}^T\mathbf{z}}{\mathbf{z}^T\mathcal{Q}^{-1}\mathbf{z}},$$

where $\mathbf{y} = \mathcal{K}^T\mathbf{x}$ and $\mathbf{z} = \mathcal{K}^T\mathbf{x}$. Hence, by applying a result which bounds the eigenvalues of a mass matrix, for example, [13, Theorem 1.29], we see that their addition will simply scale the lower and upper bounds by some constants $c_* < 1$ and $C^* > 1$. We therefore get

$$c_*(1-\eta_n)^2 \leq \frac{\mathbf{x}^T\mathcal{K}\mathcal{Q}^{-1}\mathcal{K}^T\mathbf{x}}{\mathbf{x}^T\mathcal{K}_n\mathcal{Q}^{-1}\mathcal{K}_n^T\mathbf{x}} \leq C^*(1+\eta_n)^2. \tag{2.6}$$

Note that MINRES cannot be used to approximate $\mathcal{K}$, unless run until convergence, since—like CG—MINRES is a Krylov subspace method, and hence nonlinear. We would therefore have to use a flexible outer method if we were to make use of an inner Krylov process as an approximation for the Stokes operator.

Consider a simple iteration of the form

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \mathcal{M}^{-1}\mathcal{K}\mathbf{r}^{(k)}, \tag{2.7}$$

where $\mathbf{r}^{(k)}$ is the residual at the kth step, and with a block lower-triangular splitting matrix

$$\mathcal{M} := \begin{bmatrix} K_0 & 0 \\ B & -Q_0 \end{bmatrix}, \tag{2.8}$$

where $\underline{K}_0$ approximates $\underline{K}$ and Q_0 approximates Q_p, which is itself spectrally equivalent to the Schur complement for the Stokes problem [13, section 6.2]. This iteration is the well-known inexact Uzawa method for solving saddle point problems [8, 11].

We know that if $\rho(I - \mathcal{M}^{-1}\mathcal{K}) < 1$, where ρ denotes the spectral radius, then the iteration (2.7) will converge in any norm, and hence $\eta_n < 1$ for sufficiently large n. In the remainder of this section we will prove some theoretical results about the iteration (2.7), and we will justify its use as a component in our preconditioners. We will take the following route:

- describe bounds for the eigenvalues of $\mathcal{M}^{-1}\mathcal{K}$;
- demonstrate that $\rho(I - \mathcal{M}^{-1}\mathcal{K}) < 1$ for the problem considered here;
- for the case where $\underline{K} - \underline{K}_0 > 0$, use our knowledge of $\rho(I - \mathcal{M}^{-1}\mathcal{K})$ to give an upper bound on the convergence rate when measured in the 2-norm—i.e., η_m in (2.6);
- show that, given a smallest possible h, we can pick the approximation to $\underline{K}$ such that η_n is independent of h.

2.4.1. Eigenvalues of $\mathcal{M}^{-1}\mathcal{K}$. First, we look at the generalized eigenvalues λ of

$$\begin{bmatrix} \underline{K} & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}} \\ \tilde{\mathbf{y}} \end{bmatrix} = \lambda \begin{bmatrix} \underline{K}_0 & 0 \\ B & -Q_0 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}} \\ \tilde{\mathbf{y}} \end{bmatrix}. \tag{2.9}$$

We ignore the one zero eigenvalue of $\underline{K}$ which is due to the hydrostatic pressure here, and in what follows, if we start an iteration orthogonal to this kernel, we will remain orthogonal to the kernel [13, section 2.3].

We consider two cases—$\underline{K} - \underline{K}_0$ positive definite and $\underline{K} - \underline{K}_0$ indefinite.

$\underline{K}_0 - \underline{K}$ positive definite. In the first case, it can be shown [25, Theorem 3.1], [34, Theorem 4.1] that if $\underline{K}_0$ and Q_0 are positive definite matrices such that

$$\upsilon \le \frac{\mathbf{x}^T \underline{K} \mathbf{x}}{\mathbf{x}^T \underline{K}_0 \mathbf{x}} \le \Upsilon, \quad \psi \le \frac{\mathbf{y}^T B \underline{K}^{-1} B^T \mathbf{y}}{\mathbf{y}^T Q_0 \mathbf{y}} \le \Psi, \tag{2.10}$$

where $\mathbf{x} \in \mathbb{R}^{n_v}$ and $\mathbf{y} \in \mathbb{R}^{n_p}$, $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$, then λ in (2.9) is real and positive, and moreover satisfies

$$\begin{aligned} \frac{(1+\psi)\Upsilon - \sqrt{(1+\psi)^2\Upsilon^2 - 4\psi\Upsilon}}{2} \le &\lambda \le \frac{(1+\Psi)\upsilon - \sqrt{(1+\Psi)^2\upsilon^2 - 4\Psi\upsilon}}{2}, \\ \upsilon \le &\lambda \le \Upsilon, \qquad \text{or} \\ \frac{(1+\psi)\upsilon + \sqrt{(1+\psi)^2\upsilon^2 - 4\psi\upsilon}}{2} \le &\lambda \le \frac{(1+\Psi)\Upsilon + \sqrt{(1+\Psi)^2\Upsilon^2 - 4\Psi\Upsilon}}{2}. \end{aligned}$$

$\underline{K}_0 - \underline{K}$ indefinite. The situation is more complicated in the case where $\underline{K} - \underline{K}_0$ is indefinite, as now the generalized eigenvalues of (2.9) will, in general, be complex. We still assume that $\underline{K}$, $\underline{K}_0$ and Q_0 are positive definite and satisfy (2.10).

Consider an eigenvector in (2.9) where $B\tilde{\mathbf{x}} = 0$. Then it is clear that $\tilde{\mathbf{y}} = \mathbf{0}$, and hence the associated eigenvalue is $\lambda = \frac{\tilde{\mathbf{x}}^T \underline{K} \tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^T \underline{K}_0 \tilde{\mathbf{x}}}$; hence the generalized eigenvalues associated with eigenvectors of this form must be real with

$$\upsilon \le \lambda \le \Upsilon.$$

Suppose now that $B\tilde{\mathbf{x}} \neq 0$. We can rearrange the second row in (2.9) to give

$$\tilde{\mathbf{y}} = \frac{\lambda - 1}{\lambda} Q_0^{-1} B \tilde{\mathbf{x}},$$

and substituting this into the first equation and rearranging gives

$$\lambda = \lambda^2 \frac{\tilde{\mathbf{x}}^T \underline{K}_0 \tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^T \underline{K} \tilde{\mathbf{x}}} + (1-\lambda) \frac{\tilde{\mathbf{x}}^T B^T Q_0^{-1} B \tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^T \underline{K} \tilde{\mathbf{x}}}.$$

If we define

$$\kappa := \kappa(\tilde{\mathbf{x}}) = \frac{\tilde{\mathbf{x}}^T \underline{K} \tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^T \underline{K}_0 \tilde{\mathbf{x}}}, \qquad \sigma := \sigma(\tilde{\mathbf{x}}) = \frac{\tilde{\mathbf{x}}^T B^T Q_0^{-1} B \tilde{\mathbf{x}}}{\tilde{\mathbf{x}}^T \underline{K} \tilde{\mathbf{x}}},$$

then we can write this as

$$\lambda^2/\kappa + (1-\lambda)\sigma - \lambda = 0,$$

or, alternatively,

$$\lambda^2 - (\sigma+1)\kappa\lambda + \sigma\kappa = 0.$$

Therefore, the eigenvalues satisfy

$$\lambda = \frac{(\sigma+1)\kappa \pm \sqrt{(\sigma+1)^2\kappa^2 - 4\sigma\kappa}}{2}.$$

We now consider the sign of the discriminant. As we have assumed that $\underline{K}$ and $\underline{K}_0$ are positive definite, $\kappa > 0$; hence the discriminant is only negative if

$$\kappa < \frac{4\sigma}{(1+\sigma)^2}.$$

We will complete our discussion by arguing for different values of κ.

$\underline{K} - \underline{K}_0$ indefinite, $\kappa > 1$. Here, the result given above for $\underline{K} - \underline{K}_0$ positive definite will still hold, and we have $\lambda \in \mathbb{R}$ which satisfies

$$\frac{(\psi+1)\Upsilon - \sqrt{(\psi+1)^2\Upsilon^2 - 4\psi\Upsilon}}{2} \le \lambda \le \frac{(\Psi+1)\Upsilon + \sqrt{(\Psi+1)^2\Upsilon^2 - 4\Psi\Upsilon}}{2}.$$

$\underline{K} - \underline{K}_0$ indefinite, $\kappa \in (0, \frac{4\sigma}{(1+\sigma)^2})$. Since the discriminant is negative in this case, there will be a pair of complex conjugate zeros. In this case,

$$\begin{aligned} \lambda &= \frac{(\sigma+1)\kappa \pm i\sqrt{4\sigma\kappa - (\sigma+1)^2\kappa^2}}{2} \\ \Rightarrow |\lambda|^2 &= \frac{(\sigma+1)^2\kappa^2 + 4\sigma\kappa - (\sigma+1)^2\kappa^2}{4} \\ &= \sigma\kappa. \end{aligned}$$

Therefore the complex eigenvalues satisfy

$$\sqrt{\upsilon\psi} \le |\lambda| \le \sqrt{\Psi}. \tag{2.11}$$

Moreover, $\mathrm{Re}(\lambda) = \frac{(\sigma+1)\kappa}{2} > 0$, so all the complex eigenvalues live in the right-hand plane. Also,

$$\begin{aligned} \frac{|\mathrm{Im}(\lambda)|}{\mathrm{Re}(\lambda)} &= \frac{\sqrt{4\sigma\kappa - (\sigma+1)^2\kappa^2}}{2} \cdot \frac{2}{(\sigma+1)\kappa} \\ &= \frac{\sqrt{4\sigma\kappa - (\sigma+1)^2\kappa^2}}{(\sigma+1)\kappa}. \end{aligned}$$

If we define

$$F(\sigma,\kappa) := \frac{\sqrt{4\sigma\kappa-(\sigma+1)^2\kappa^2}}{(\sigma+1)\kappa},$$

then

$$\frac{\partial F}{\partial \sigma} = \frac{2(\sigma-1)}{(\sigma+1)^2\sqrt{(4-2\kappa)\kappa\sigma-\kappa^2(\sigma^2+1)}},$$

so

$$\frac{\partial F}{\partial \sigma} = 0 \Rightarrow \sigma = 1.$$

This critical point is clearly a maximum. This means that, for any fixed κ, $F(\sigma,\kappa)$ has its maximum at $\sigma = 1$. Therefore,

$$\frac{|\mathrm{Im}(\lambda)|}{\mathrm{Re}(\lambda)} = F(\sigma,\kappa) \le \frac{\sqrt{\kappa-\kappa^2}}{\kappa} = \sqrt{\frac{1}{\kappa}-1} \le \sqrt{\frac{1}{\upsilon}-1}.$$

Therefore, putting this together with (2.11) above, the complex eigenvalues satisfy
(2.12)
$$\lambda \in \left\{ z = re^{i\theta} \in \mathbb{C} \ : \ \sqrt{\upsilon\psi} \le r \le \sqrt{\Psi},\ -\tan^{-1}(\sqrt{\upsilon^{-1}-1}) \le \theta \le \tan^{-1}(\sqrt{\upsilon^{-1}-1}) \right\}.$$

$\underline{K} - \underline{K}_0$ indefinite, $\kappa \in [\frac{4\sigma}{(1+\sigma)^2}, 1]$. What about the remaining case? Here too, the bounds given for when $\kappa > 1$ hold, since in the derivation we required no information about δ, all that is assumed is that $\lambda \in \mathbb{R}$—see [25, Theorem 3.1]. Verifying the inner bounds required that $\delta > 1$, so these do not carry over, but there is no such problem with the outer bounds.

We have proved the following theorem.

THEOREM 2.1. *Let λ be an eigenvalue associated with the generalized eigenvalue problem*

$$\begin{bmatrix} \underline{K} & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \lambda \begin{bmatrix} \underline{K}_0 & 0 \\ B & -Q_0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix},$$

where $\underline{K}$, $\underline{K}_0$ and Q_0 are positive definite and satisfy (2.10). *If $\lambda \in \mathbb{R}$, then it satisfies*

$$\frac{(1+\psi)\Upsilon - \sqrt{(1+\psi)^2\Upsilon^2 - 4\psi\Upsilon}}{2} \le \lambda \le \frac{(1+\Psi)\Upsilon + \sqrt{(1+\Psi)^2\Upsilon^2 - 4\Psi\Upsilon}}{2}$$
$$\text{or} \qquad \upsilon \le \lambda \le \Upsilon,$$

and if $\lambda \in \mathbb{C}$, then $\lambda = re^{i\theta}$, where r and θ satisfy

$$\sqrt{\upsilon\psi} \le r \le \sqrt{\Psi},\ -\tan^{-1}(\sqrt{\upsilon^{-1}-1}) \le \theta \le \tan^{-1}(\sqrt{\upsilon^{-1}-1}).$$

2.4.2. Spectral radius of $\mathcal{M}^{-1}\mathcal{K}$. The next step is to get bounds for $\rho(I - \mathcal{M}^{-1}\mathcal{A})$. Figure 2.2 is a diagram of the geometry which shows how this shift affects the complex eigenvalues. All the eigenvalues will be contained in the unit circle if the line d labeled on the diagram is less than unity. By the cosine rule:

$$d^2 = 1 + \Psi - 2\sqrt{\Psi}\cos\theta,$$

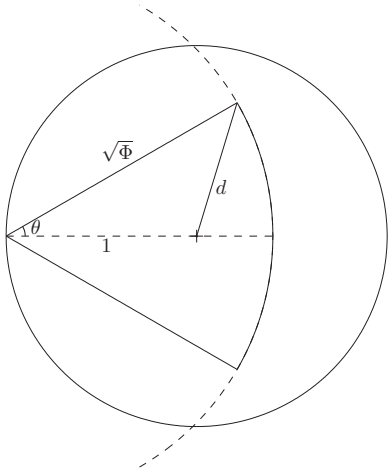


FIG. 2.2. *Diagram of the geometry containing the complex eigenvalues.* $\theta = \sqrt{\upsilon^{-1} - 1}$ *and* d *is the unknown length.*

where $\tan\theta = \sqrt{\upsilon^{-1} - 1}$. Therefore, all the complex eigenvalues are in the unit circle if

$$\frac{\sqrt{\Psi}}{2} < \cos\theta.$$

Note that, using the same argument, the distance from the origin to the point where the circle of radius $\sqrt{\psi\upsilon}$ and center -1 touches the ray that makes an angle θ with the x axis is

$$\sqrt{1 + \psi\upsilon - 2\sqrt{\psi\upsilon}\cos\theta}.$$

There follows Corollary 2.2.

COROLLARY 2.2. *Suppose that the eigenvalues of the generalized eigenvalue problem* (2.9) *are as described in Theorem* 2.1. *Define*

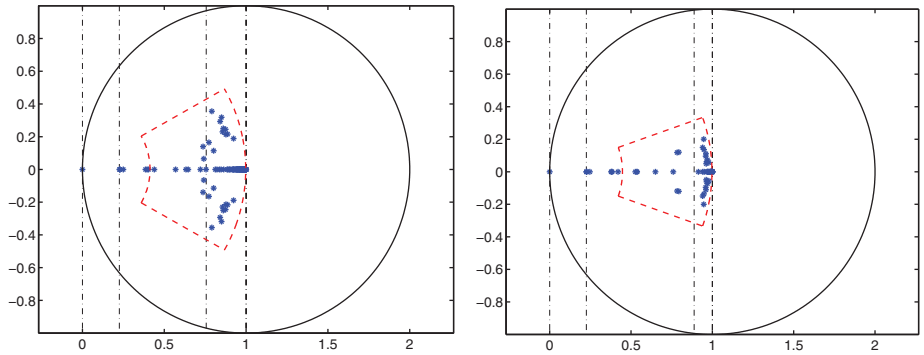
$$\xi := \max\left\{1 - \upsilon, \Upsilon - 1, 1 - \frac{(1+\psi)\Upsilon - \sqrt{(1+\psi)^2\Upsilon^2 - 4\psi\Upsilon}}{2}, \right.$$
$$\frac{(1+\Psi)\Upsilon + \sqrt{(1+\Psi)^2\Upsilon^2 - 4\Psi\Upsilon}}{2} - 1, \sqrt{1 + \Psi - 2\sqrt{\Psi}\cos\theta},$$
$$\left.\sqrt{1 + \psi\upsilon - 2\sqrt{\psi\upsilon}\cos\theta}\right\}.$$

Then a simple iteration with splitting matrix

$$\mathcal{M} = \begin{bmatrix} K_0 & 0 \\ B & -Q_0 \end{bmatrix}$$

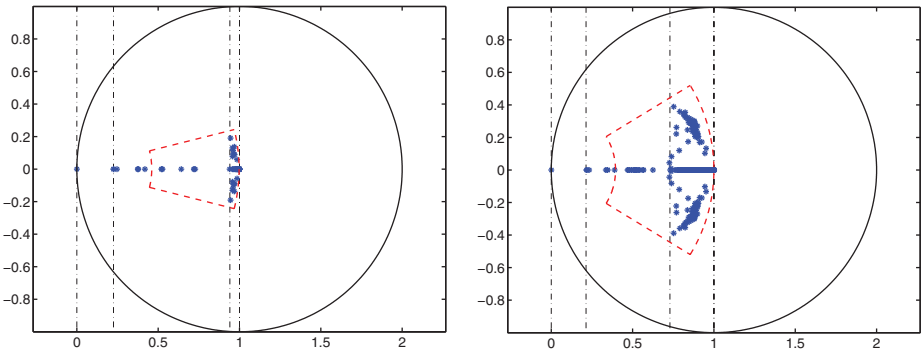
will converge if $\xi < 1$, *with the asymptotic convergence rate being* ξ.

Zulehner also derived an approximation to the convergence factor [34, Theorem 4.3]. Note that Corollary 2.2 differs slightly from the result in Zulehner—this is because neither the result given here nor in [34] are sharp with regards to the complex



(a) h = 0.25, $\underline{K}_0$ given by 1 AMG V-cycle with 1 pre- and 1 post-smoothing step

(b) h = 0.25, $\underline{K}_0$ given by 1 AMG V-cycle with 2 pre- and 2 post-smoothing steps



(c) h = 0.25, $\underline{K}_0$ given by 2 AMG V-cycles with 2 pre- and 2 post-smoothing steps

(d) h = 0.125, $\underline{K}_0$ given by 1 AMG V-cycle with 1 pre- and 1 post-smoothing step

FIG. 2.3. **'s denote computed eigenvalues of $\mathcal{M}^{-1}\mathcal{K}$ for different approximations to $\underline{K}$ with $Q = Q_0$. Lines, from left to right, are at $0, ((\psi+1)\Upsilon - \sqrt{(\psi+1)^2\Upsilon^2 - 4\psi\Upsilon})/2$, υ, Υ and $((\Psi+1)\Upsilon + \sqrt{(\Psi+1)^2\Upsilon^2 - 4\Psi\Upsilon})/2$, (the last two virtually coincide here). Dashed region is the bounds of Theorem 2.1 for the complex eigenvalues. Also shown is the unit circle centered at $z = 1$.*

eigenvalues. The two results are obtained in very different ways, and neither can be said to be a better approximation than the other one.

Figure 2.3 shows the bounds for the eigenvalues of $\mathcal{M}^{-1}\mathcal{K}$ predicted above—together with actual computed eigenvalues—for a number of approximations to the matrix $\underline{K}$. For clarity in producing this plot we have taken $Q_0 = Q$, with a direct solve used where needed.

2.4.3. Bounding η_n for $\underline{K} - \underline{K}_0 > 0$. The results so far have shown that we will get asymptotic convergence—i.e., there is some n such that (2.4) holds with $\eta_n < 1$. However, there may be some significant transient behavior in the convergence. For this iteration to be practical as a preconditioner, we need a good approximation from a small number of iterations.

Luckily, in practice we see good results from the first iteration. Also, the theory above is equally valid for the block *upper*-triangular approximation to the discrete Stokes matrix, whereas in practice we observe that it takes far more iterations with this upper-triangular splitting to converge. Below we explain why the lower-triangular splitting behaves so well.

Let us again return to the case where $\underline{K} - \underline{K}_0$ is positive definite. Then we know (c.f. section 2.2) that

$$\mathcal{M}^{-1}\mathcal{K} = \begin{bmatrix} \underline{K}_0 & 0 \\ B & -Q_0 \end{bmatrix}^{-1} \begin{bmatrix} \underline{K} & B^T \\ B & 0 \end{bmatrix}$$

is self-adjoint in the inner product defined by

$$\widehat{\mathcal{H}} = \begin{bmatrix} \underline{K} - \underline{K}_0 & 0 \\ 0 & Q_0 \end{bmatrix}.$$

If we define $\widehat{\mathcal{K}} := \mathcal{M}^{-1}\mathcal{K}$, then we have that $\widehat{\mathcal{K}}$ is $\widehat{\mathcal{H}}-$normal, i.e.,

$$\widehat{\mathcal{K}}^\dagger \widehat{\mathcal{K}} = \widehat{\mathcal{K}}\widehat{\mathcal{K}}^\dagger,$$

where $\widehat{\mathcal{K}}^\dagger = \widehat{\mathcal{H}}^{-1}\widehat{\mathcal{K}}^T\widehat{\mathcal{H}}$. The iteration matrix $I - \mathcal{M}^{-1}\mathcal{K}$ is therefore $\widehat{\mathcal{H}}-$normal, and so

$$\|I - \mathcal{M}^{-1}\mathcal{K}\|_{\widehat{\mathcal{H}}} = \rho(I - \mathcal{M}^{-1}\mathcal{K}),$$

which tells us that—if again $\mathcal{K}\mathbf{w} = \tilde{\mathbf{b}}$, say—the nth iterations satisfies

$$\|\mathbf{w}^{(n)} - \mathbf{w}\|_{\widehat{\mathcal{H}}} \leq \rho^n \|\mathbf{w}\|_{\widehat{\mathcal{H}}},$$

where $\rho = \rho(I - \mathcal{M}^{-1}\mathcal{K})$, the spectral radius of the iteration matrix. To apply the result of Braess and Peisker (2.5) we need a constant η_n such that the error converges the 2-norm, i.e.,

$$\|\mathbf{w}^{(n)} - \mathbf{w}\|_2 \leq \eta_n \|\mathbf{w}\|_2.$$

We know that over a finite-dimensional vector space all norms are equivalent, though the equivalence constants may be h-dependent for a discretized PDE problem. Thus, there exist positive constants γ and Γ such that

$$\sqrt{\gamma}\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_{\widehat{\mathcal{H}}} \leq \sqrt{\Gamma}\|\mathbf{x}\|_2$$

for all $\mathbf{x} \in \mathbb{R}^{n_v+n_p}$, and hence

$$\|\mathbf{x}_k - \mathbf{x}\|_2 \leq \frac{1}{\sqrt{\gamma}}\|\mathbf{x}_k - \mathbf{x}\|_{\widehat{\mathcal{H}}} \leq \frac{\rho^m}{\sqrt{\gamma}}\|\mathbf{x}\|_{\widehat{\mathcal{H}}} \leq \frac{\sqrt{\Gamma}\rho^m}{\sqrt{\gamma}}\|\mathbf{x}\|_2. \tag{2.13}$$

We can therefore bound the constant η_n above by ρ^m multiplied by the constant $\sqrt{\Gamma/\gamma}$. By the results in section 2.4.2 we know that ρ is independent of h—if Γ/γ can be shown to be independent of h, then the simple iteration (2.7) can be used as a component of an optimal (with respect to mesh size) preconditioner.

Recall standard bounds for two-dimensional finite element matrices—see, for example, Theorems 1.32 and 1.29 in [13]—we have that, under mild assumptions, there exist positive constants $c_1,\ C_1,\ c_2$, and C_2 such that

$$\begin{aligned} c_1 h^2 \mathbf{y}^T\mathbf{y} \leq \mathbf{y}^T\underline{K}\mathbf{y} \leq C_1\mathbf{y}^T\mathbf{y}, \\ c_2 h^2 \mathbf{z}^T\mathbf{z} \leq \mathbf{z}^T Q_p \mathbf{z} \leq C_2 h^2 \mathbf{z}^T\mathbf{z} \end{aligned}$$

for all $\mathbf{y} \in \mathbb{R}^{n_v}$ and $\mathbf{z} \in \mathbb{R}^{n_p}$. We will use this to give estimates for Γ and γ below.

The upper bound. First, condsider Γ such that $\mathbf{x}^T\widehat{\mathcal{H}}\mathbf{x} \leq \Gamma\mathbf{x}^T\mathbf{x}$. This means that

$$\mathbf{y}^T(\underline{K} - \underline{K}_0)\mathbf{y} + \mathbf{z}^T Q_0\mathbf{z} \leq \Gamma(\mathbf{y}^T\mathbf{y} + \mathbf{z}^T\mathbf{z}).$$

Therefore, if we can find constants Γ_1 and Γ_2 such that

$$\mathbf{y}^T(\underline{K} - \underline{K}_0)\mathbf{y} \leq \Gamma_1\mathbf{y}^T\mathbf{y} \quad \text{and} \quad \mathbf{z}^T Q_0\mathbf{z} \leq \Gamma_2\mathbf{z}^T\mathbf{z},$$

then we could take $\Gamma = \max(\Gamma_1, \Gamma_2)$.

First, note that from (2.10) we have that

$$\begin{aligned}
\mathbf{x}^T\underline{K}\mathbf{x} &\leq \Upsilon\mathbf{x}^T\underline{K}_0\mathbf{x},\\
\Upsilon\mathbf{x}^T\underline{K}\mathbf{x} - (\Upsilon-1)\mathbf{x}^T\underline{K}\mathbf{x} &\leq \Upsilon\mathbf{x}^T\underline{K}_0\mathbf{x},\\
\Upsilon(\mathbf{x}^T\underline{K}\mathbf{x} - \mathbf{x}^T\underline{K}_0\mathbf{x}) &\leq (\Upsilon-1)\mathbf{x}^T\underline{K}\mathbf{x},\\
\mathbf{x}^T(\underline{K} - \underline{K}_0)\mathbf{x} &\leq \frac{C_1(\Upsilon-1)}{\Upsilon}\mathbf{x}^T\mathbf{x},\\
\therefore\ \frac{\mathbf{x}^T(\underline{K} - \underline{K}_0)\mathbf{x}}{\mathbf{x}^T\mathbf{x}} &\leq \frac{C_1(\Upsilon-1)}{\Upsilon}.
\end{aligned}$$

Therefore,

$$\Gamma_1 = \frac{(\Upsilon-1)C_1}{\Upsilon}.$$

Let $Q_0 = T_m^p$ represent m steps of the Chebyshev semi-iteration, as defined in section 2.3, where

$$\delta_m^p \leq \frac{\mathbf{x}^T Q_p\mathbf{x}}{\mathbf{x}^T T_m^p\mathbf{x}} \leq \Delta_m^p.$$

Then

$$\frac{\mathbf{z}^T Q_0\mathbf{z}}{\mathbf{z}^T\mathbf{z}} = \frac{\mathbf{z}^T T_m^p\mathbf{z}}{\mathbf{z}^T\mathbf{z}} = \frac{\mathbf{z}^T T_m^p\mathbf{z}}{\mathbf{z}^T Q_p\mathbf{z}} \cdot \frac{\mathbf{z}^T Q_p\mathbf{z}}{\mathbf{z}^T\mathbf{z}} \leq \frac{C_2h^2}{\delta_m^p}.$$

Therefore, we can take $\Gamma_2 = C_2h^2$, and hence

$$\Gamma = \max\left(\frac{(\Upsilon-1)C_1}{\Upsilon}, \frac{C_2h^2}{\delta_m^p}\right) \tag{2.14}$$

satisfies $\mathbf{x}^T\widehat{\mathcal{H}}\mathbf{x} \leq \Gamma\mathbf{x}^T\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^{n_v+n_p}$.

The lower bound. Now we turn our attention to a lower bound. Similarly to above, we take $\gamma = \min(\gamma_1, \gamma_2)$, where γ_1 and γ_2 satisfy

$$\gamma_1\mathbf{y}^T\mathbf{y} \leq \mathbf{y}^T(\underline{K} - \underline{K}_0)\mathbf{y} \quad \text{and} \quad \gamma_2\mathbf{z}^T\mathbf{z} \leq \mathbf{z}^T Q_0\mathbf{z}$$

for all $\mathbf{y} \in \mathbb{R}^{n_v}$, $\mathbf{z} \in \mathbb{R}^{n_p}$. Again, we have from (2.10) that

$$\begin{aligned}
\upsilon\mathbf{y}^T\underline{K}_0\mathbf{y} &\leq \mathbf{y}^T\underline{K}\mathbf{y}\\
&= \upsilon\mathbf{y}^T\underline{K}\mathbf{y} + (1-\upsilon)\mathbf{y}^T\underline{K}\mathbf{y},\\
(\upsilon-1)\mathbf{y}^T\underline{K}\mathbf{y} &\leq \upsilon\mathbf{y}^T(\underline{K} - \underline{K}_0)\mathbf{y},\\
\frac{(\upsilon-1)c_1h^2}{\upsilon} &\leq \frac{\mathbf{y}^T(\underline{K} - \underline{K}_0)\mathbf{y}}{\mathbf{y}^T\mathbf{y}}.
\end{aligned}$$

Again arguing as above,

$$\frac{\mathbf{z}^T Q_0 \mathbf{z}}{\mathbf{z}^T \mathbf{z}} = \frac{\mathbf{z}^T T_m^p \mathbf{z}}{\mathbf{z}^T Q_p \mathbf{z}} \cdot \frac{\mathbf{x}^T Q_p \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \geq \frac{c_2 h^2}{\Delta_m^p}.$$

Therefore, we can take

$$\gamma = \min\left(\frac{(\upsilon - 1)c_1 h^2}{\upsilon}, \frac{c_2 h^2}{\Delta_m^p}\right), \tag{2.15}$$

which satisfies $\gamma \mathbf{x}^T \mathbf{x} \leq \mathbf{x}^T \widehat{\mathcal{H}} \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^{n_v + n_p}$.

2.4.4. An "optimal" precondtioner. By (2.13) the contraction constant for convergence in the 2-norm is given by $\rho^m \sqrt{\Gamma}/\sqrt{\gamma}$. It is clear that—irrespective of which term in (2.15) is smallest—$\sqrt{\gamma} = \nu h$ for some constant ν.

In order to have preconditioner that is robust with respect to mesh size we also need the numerator to be dependent on h—i.e., we need that

$$\frac{(\Upsilon - 1)C_1}{\Upsilon} < \frac{C_2 h^2}{\delta_m^p}. \tag{2.16}$$

At first glance, this seems unlikely, as the left-hand side above is a constant, whereas the right-hand side is dependent on the small parameter h. However, we have control over the value of Υ, as this measures the accuracy of the approximation to $\underline{K}$. Recall that $\underline{K}_0$ is a good approximation to $\underline{K}$ if Υ is close to unity. We will generally take $\underline{K}_0$ to be some iterative procedure—for example, a multigrid process—and hence we can make this parameter as close to 1 as required by simply taking more iterations. In the case of a multigrid approximation, this could mean using more V-cycles, better smoothing, etc.

Therefore, if we can ensure that $\underline{K}_0$ is a good enough approximation to $\underline{K}$ that (2.16) holds, then

$$\eta_n = \min\left(\frac{(\upsilon - 1)c_1}{\upsilon}, \frac{c_2}{\Delta_m^p}\right) \frac{\rho^n \delta_m^p}{C_2}, \tag{2.17}$$

which is a constant—at least up to some predetermined value of h. Note that we have knowledge of all the parameters involved, so given a smallest required value of h—which one will know a priori—one can pick an approximation $\underline{K}_0$ which gives a reasonable method. The quantity ρ^n also appears in the numerator, so convergence can be improved by taking more inexact Uzawa iterations.

The above argument only holds when $\underline{K} - \underline{K}_0$ is positive definite. Note that for any approximation $\underline{K}_0$ it is possible to scale it by a parameter $\hat{\omega}$—as in Bramble–Pasiak CG case—so that $\underline{K} > \hat{\omega}\underline{K}_0$. Since the eigenvalue problem to be solved to determine ω is much more expensive here than it was in section 2.2, such a scaling can add significantly to the cost of the method. In practice, if we take $\underline{K}_0$ to be defined implicitly by a multigrid method, we see the mesh-independent behavior described above without such a scaling, so we speculate that a similar result to (2.17) holds true for any $\underline{K}_0$ sufficiently close to $\underline{K}$. The fact that the complex eigenvalues in the general case don't stray too far into the complex plane for a high enough value of n—see Figure 2.3—gives some heuristic justification for this conjecture.

Since solving the approximation to $\mathcal{K}$ is particularly expensive here, it is worth getting the approximation to the mass matrix, Q_0, as close to Q as possible. Therefore, in the results that follow we take Q_0 to be defined implicity by 20 steps of the Chebyshev semi-iteration applied to the appropriate mass matrix.

The inexact Uzawa method can be improved with the introduction of a parameter τ in front of the approximation to the Schur complement [12]. In the inexact obtaining the optimal parameter is infeasible, but a good approximation is $(\phi + \Phi)/2$, where $\lambda(S_0^{-1}S) \in [\phi, \Phi]$. For $\mathbf{Q}_1$ elements and a Dirichlet problem, $\lambda(Q_p^{-1}S) \in [0.2, 1]$ [13, p. 271], so we take our scaling parameter as $\tau = 3/5$. Note that the preceding analysis in Theorem 2.1 remains vaild, simply by replacing Q_0 by τQ_0 and scaling the constants ϕ and Φ accordingly. On the basis of the theoretical results presented above, we advocate a practical splitting matrix for inexact Uzawa iteration (2.7) of

$$\mathcal{M} = \begin{bmatrix} \underline{K}_0 & 0 \\ B & -\tau Q_0 \end{bmatrix}. \tag{2.18}$$

2.5. Summary. Now that we have developed appropriate approximations for the blocks in the block diagonal preconditioner and block lower-triangular preconditioner—as described in sections 2.1 and 2.2, respectively—we can describe a practical preconditioner. Consider first the block diagonal case. Based on the results in the preceding sections, a matrix of the form

$$\mathcal{P}_{\text{bd}} := \begin{bmatrix} A_0 & 0 \\ 0 & \mathcal{K}_n \mathcal{Q}^{-1} \mathcal{K}_n^T \end{bmatrix},$$

where A_0 is composed of Chebyshev approximations and $\mathcal{K}_n$ denotes n steps of the simple iteration (2.7) based on the splitting matrix $\mathcal{M}$, as defined in (2.18), should therefore be an effective preconditioner for the matrix $\mathcal{A}$. This preconditioner is summarized as Algorithm 2 below.

Solving with the block lower-triangular preconditioner

$$\mathcal{P}_{\text{lt}} := \begin{bmatrix} \bar{A}_0 & 0 \\ C & \mathcal{K}_n \mathcal{Q}^{-1} \mathcal{K}_n^T \end{bmatrix}$$

is a slight modification of $\mathcal{P}_{bd}$ and is also described in Algorithm 2. There are two differences here. First, we need scale A_0 as described in Rees and Stoll [25] so that $A - \bar{A}_0 > 0$. Second, there is an extra matrix-vector multiply with C.

Algorithm 2 presupposes that we have two subroutines at our disposal: a Chebyshev semi-iteration routine `cheb_semi_it` (see 1) and some multigrid routine `mg`, both of which perform a fixed (m or t, respectively) number of iterations.

3. Numerical results. First, consider the following forward problem, which sets the boundary conditions that we will use for the control problem. This is a classic test problem in fluid dynamics, called leaky cavity flow, and a discussion is given by Elman, Silvester and Wathen [13, Example 5.1.3].

Example 3.1. Let $\Omega = [0, 1]^2$, and let $\vec{\mathbf{i}}$ and $\vec{\mathbf{j}}$ denote unit vectors in the direction of the x and y axis, respectively. Let $\vec{v}$ and p satisfy the Stokes equations

$$\begin{aligned} -\nabla^2 \vec{v} + \nabla p &= \vec{0} \quad \text{in } \Omega, \\ \nabla \cdot \vec{v} &= 0 \quad \text{in } \Omega, \end{aligned}$$

and let $\vec{v} = \vec{0}$ on the boundary except for on $x = 1$, $0 \le y \le 1$, where $\vec{v} = -\vec{\mathbf{j}}$.

ALGORITHM 2. An application of the preconditioner $\mathcal{P}_{\rm bd}$ or $\mathcal{P}_{\rm lt}$, where we solve $\mathcal{P}[\widehat{\mathbf{v}}^T\ \widehat{\mathbf{p}}^T\ \widehat{\mathbf{u}}^T\ \widehat{\boldsymbol{\lambda}}^T\ \widehat{\boldsymbol{\mu}}^T]^T = [\mathbf{v}^T\ \mathbf{p}^T\ \mathbf{u}^T\ \boldsymbol{\lambda}^T\ \boldsymbol{\mu}^T]^T$.

if preconditioner= $\mathcal{P}_{\rm lt}$ **then**
  Calculate scalings ω_p and $\omega_{\vec{v}}$
else
  $\omega_p = 1$, $\omega_{\vec{v}} = 1$
end if
$\widehat{\mathbf{v}} = \texttt{cheb\_semi\_it}(\omega_{\vec{v}}\underline{Q}_{\vec{v}}, \mathbf{v}, m)$
$\widehat{\mathbf{p}} = \frac{1}{\alpha}\texttt{cheb\_semi\_it}(\omega_p Q_p, \mathbf{p}, m)$
$\widehat{\mathbf{u}} = \frac{1}{\beta}\texttt{cheb\_semi\_it}(\omega_{\vec{v}}\underline{Q}_{\vec{v}}, \mathbf{u}, m)$
if preconditioner= $\mathcal{P}_{\rm lt}$ **then**

$$\begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \underline{K} & B^T & -\underline{Q}_{\vec{v}} \\ B & 0 & 0 \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{v}} \\ \widehat{\mathbf{p}} \\ \widehat{\mathbf{u}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix}$$

end if
$[\bar{\boldsymbol{\lambda}}^T\ \bar{\boldsymbol{\mu}}^T]^T = [\mathbf{0}^T\ \mathbf{0}^T]^T$
for i=1 ... n **do**

$$\begin{bmatrix} \mathbf{r}^\lambda \\ \mathbf{r}^\mu \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} - \begin{bmatrix} \underline{K} & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\lambda}} \\ \bar{\boldsymbol{\mu}} \end{bmatrix}$$

  $\bar{\boldsymbol{\lambda}} = \bar{\boldsymbol{\lambda}} + \texttt{mg}(\underline{K}, \mathbf{r}^\lambda, t)$
  $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}} - \frac{1}{\tau}\texttt{cheb\_semi\_it}(Q_p, \mathbf{r}^\mu - B\bar{\boldsymbol{\lambda}}, m)$
end for
$\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}} - (\sum_{i=1}^{n_p} \bar{\boldsymbol{\mu}}_i)/n_p) * \mathbf{1}$
$\bar{\boldsymbol{\lambda}} = \underline{Q}_{\vec{v}}\bar{\boldsymbol{\lambda}}$
$\bar{\boldsymbol{\mu}} = \alpha Q_p \bar{\boldsymbol{\mu}}$
$\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}} - (\sum_{i=1}^{n_p} \bar{\boldsymbol{\mu}}_i)/n_p) * \mathbf{1}$
$[\widehat{\boldsymbol{\lambda}}^T\ \widehat{\boldsymbol{\mu}}^T]^T = [\mathbf{0}^T\ \mathbf{0}^T]^T$
for i=1 ... n **do**

$$\begin{bmatrix} \mathbf{r}^\lambda \\ \mathbf{r}^\mu \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\lambda}} \\ \bar{\boldsymbol{\mu}} \end{bmatrix} - \begin{bmatrix} \underline{K} & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\lambda}} \\ \widehat{\boldsymbol{\mu}} \end{bmatrix}$$

  $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}} - \frac{1}{\tau}\texttt{cheb\_semi\_it}(Q_p, \mathbf{r}^\mu, m)$
  $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\lambda}} + \texttt{mg}(\underline{K}, \mathbf{r}^\lambda - B^T\widehat{\boldsymbol{\mu}}, t)$
end for

We discretize the Stokes problem using $\mathbf{Q}_2 - \mathbf{Q}_1$ elements and solve the resulting linear system using MINRES [22]. As a preconditioner we use the block diagonal matrix $\texttt{blkdiag}(\widehat{K}, T_{20})$, following Silvester and Wathen [27], where $\widehat{K}$ denotes one AMG V-cycle (using the HSL MI20 AMG routine [6] applied via a MATLAB interface), and T_{20}^{-1} is 20 steps of the Chebyshev semi-iteration applied with the pressure mass matrix. The problem was solved using MATLAB R2009b, and the number of iterations and the time taken for different mesh sizes is given in Table 3.1. The constant number of iterations independent of h and linear growth in CPU time (i.e., linear complexity of the solver) are well understood for this problem—see [13, Chapter 6].

TABLE 3.1
Number of MINRES iterations and time taken to solve the forward problem in Example 3.1.

h	size	CPU time (s)	Iterations
2^{-2}	187	0.015	25
2^{-3}	659	0.029	27
2^{-4}	2,467	0.076	28
2^{-5}	9,539	0.349	30
2^{-6}	37,507	1.504	30
2^{-7}	148,739	6.616	30

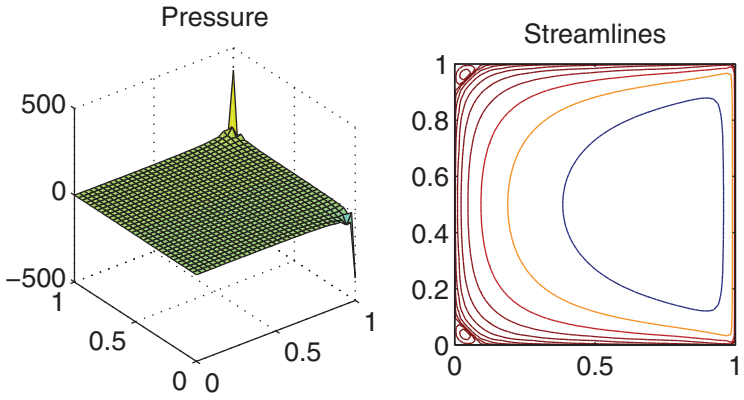


FIG. 3.1. *Solution of Example* 3.1.

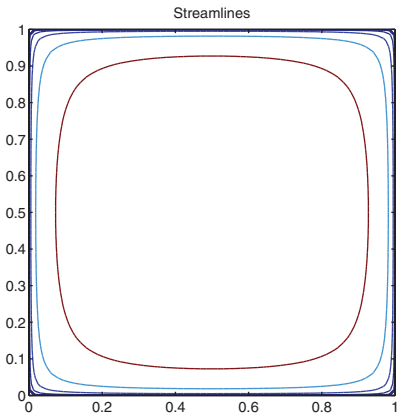


FIG. 3.2.

Figure 3.1 shows the streamlines and the pressure of the solution obtained. Note the small recirculations present in the lower corners—these are Moffatt eddies. Adding a forcing term that reduces these eddies will be the object of our control problem, Example 3.2.

Example 3.2. Let $\Omega = [0,1]^2$, and consider an optimal control problem of the form (1.1) with Dirichlet boundary conditions as given in Example 3.1 (leaky cavity flow). Take the desired pressure as $\widehat{p} = 0$, and let $\widehat{v} = y\vec{\mathbf{i}} - x\vec{\mathbf{j}}$. The exponentially distributed streamlines of the desired velocity are shown in Figure 3.2.

TABLE 3.2

Comparison of idealized *solution methods for solving Example* 3.2 *using MINRES preconditioned with* $\mathcal{P}_{\rm bd}$ *with* n *steps of inexact Uzawa with* exact Schur complement *approximating* $\mathcal{K}$ *and* t *AMG V-cycles approximating* $\underline{K}$.

h	size	Exact, $n=1$		$n=1$, $t=1$		$n=1$, $t=2$		$n=1$, $t=3$		$n=1$, $t=4$	
		time	its	time	its	time	its	time	its	time	its
2^{-2}	344	0.089	25	0.092	29	0.079	27	0.082	27	0.085	27
2^{-3}	1512	0.382	27	0.432	35	0.352	27	0.365	27	0.380	27
2^{-4}	6344	3.192	25	7.359	65	3.179	27	3.235	27	3.296	27
2^{-5}	25992	60.063	25	403.933	179	72.858	31	64.028	27	64.055	27
h	size	Exact, $n=2$		$n=2$, $t=1$		$n=2$, $t=2$		$n=2$, $t=3$		$n=2$, $t=4$	
		time	its	time	its	time	its	time	its	time	its
2^{-2}	344	0.073	21	0.100	27	0.099	25	0.096	23	0.101	23
2^{-3}	1512	0.408	23	0.429	29	0.400	25	0.423	25	0.450	25
2^{-4}	6344	3.466	23	3.954	31	3.347	25	3.193	23	3.319	23
2^{-5}	25992	57.284	21	98.885	39	65.489	25	60.051	23	61.398	23

We discretize (1.1) using $\mathbf{Q}_2-\mathbf{Q}_1$ elements, also using $\mathbf{Q}_2$ elements for the control. Table 3.2 shows the results for solving the problem using MINRES, with right-hand side as in Example 3.2 and with $\beta = 10^{-2}$ and $\alpha = 1$.

We take as our approximation to $\underline{K}$, $\underline{K}_0$, t HSL AMG MI20 V-cycles. First we would like to show some numerical experiments which point us towards the number of inexact Uzawa iterations, n, and the number of V-cycles, which also give numerical evidence for our conclusions in section 2.4. To this end, as a preconditioner we use the $\mathcal{P}_{\rm bd}$ with $\mathcal{K}$ approximated by n steps of the simple iteration with splitting matrix

$$\mathcal{M} = \begin{bmatrix} \underline{K}_0 & 0 \\ B & -S \end{bmatrix},$$

where $S = B\underline{K}^{-1}B^T$ is the *exact* Schur complement of the Stokes equation. This is *not* a practical preconditioner since it includes the exact Schur complement of the Stokes matrix—solved using a direct method. We can see clearly, however, that if the approximation $\underline{K}_0$ is not good enough we do not—*even in this idealized case*—get an optimal preconditioner. This phenomenon is explained by the theory in section 2.4. It is therefore vital that the approximation $\underline{K}_0$ is close enough to $\underline{K}$—i.e., Υ is close enough to unity—in order to get an effective practical preconditioner.

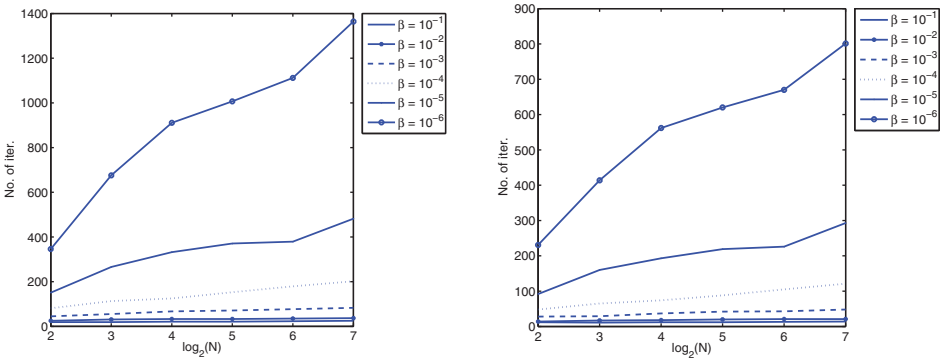
As we saw in section 2.4, a practical preconditioner can be obtained by replacing the exact Stokes Schur complement by the pressure mass matrix—or more generally, by something that approximates the pressure mass matrix. We take this to be 20 steps of the Chebyshev semi-iteration applied to the relevant matrix, as described in section 2.3. Experimentation suggests that taking two steps of the inexact Uzawa method, in which $\underline{K}_0^{-1}$ is given by three HSL MI20 AMG V-cycles, will give a good preconditioner. In the results that follow we take $\beta = 10^{-2}$, $\alpha = 1$ and solve to a tolerance of 10^{-6} in the appropriate norm.

As we see from Table 3.3, the overall technique which we have described seems to be a good method for solving the Stokes control problem. Comparing the results here with those to solve the forward problem in Table 3.1, the iteration numbers are not that much more, and they do not increase significantly with the mesh size; the solution times also scale roughly linearly. Solving the control problem using the block triangular preconditioner is just over a factor of 10 more expensive that solving a single forward problem for every grid size—an overhead that seems reasonable, given the increased complexity of the control problem in comparison to the forward problem.

TABLE 3.3

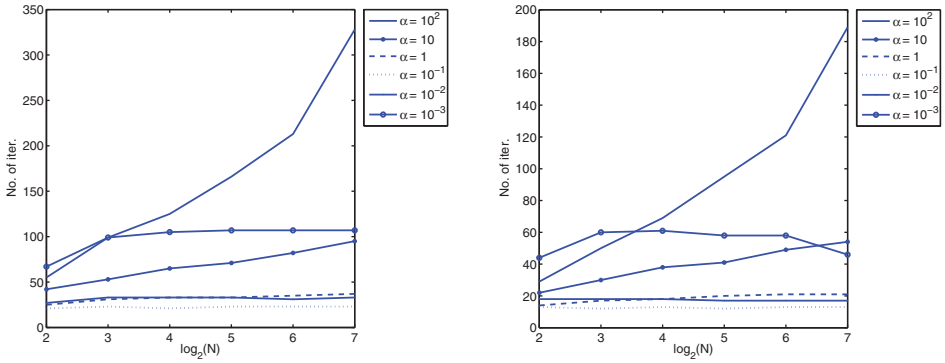
Comparison of solution methods for solving Example 3.2 using MINRES and BPCG preconditioned with the block diagonal and block lower-triangular preconditioners, respectively, with two steps of inexact Uzawa approximating $\mathcal{K}$ and three AMG V-cycles approximating $\underline{K}$.

h	size	MINRES		BPCG		`backslash`
		time	its	time	its	time
2^{-2}	344	1.366	21	1.157	14	0.043
2^{-3}	1,512	1.646	27	1.453	19	0.12
2^{-4}	6,344	3.161	29	2.527	20	1.2
2^{-5}	25,992	11.685	29	8.359	19	12.5
2^{-6}	105,224	46.813	31	34.875	22	—
2^{-7}	423,432	178.808	31	130.054	22	—
2^{-8}	1,698,824	837.691	35	587.139	24	—



(a) MINRES (b) BPCG

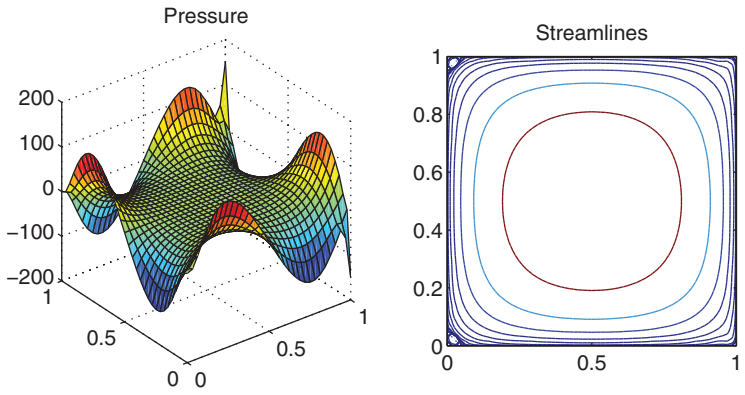
FIG. 3.3. *Plot of problem size vs iterations needed for different β, where $\alpha = 1$.*



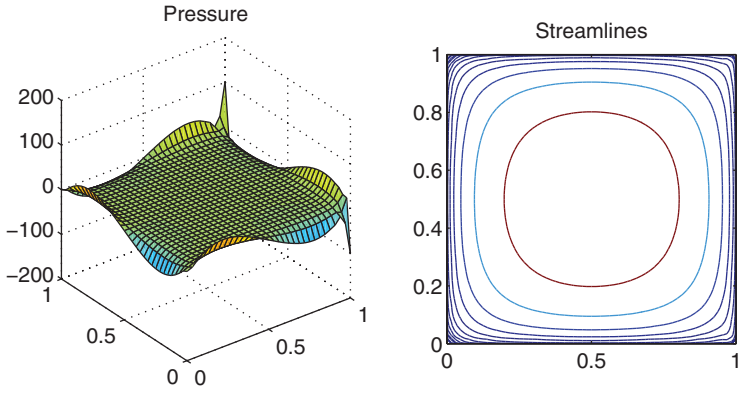
(a) MINRES (b) BPCG

FIG. 3.4. *Plot of problem size vs. iterations needed for different α, where $\beta = 10^{-2}$.*

Figures 3.3 and 3.4 show the number of iterations taken to solve this problem for different values of β and α in (1.1), respectively. These show that—as we might expect from the theory—decreasing β and increasing α increases the number of iterations required to solve the system using our methods. From the plots in Figures 3.5 and 3.6 it seems that the value $\alpha = 1$ gives a pressure of the same order as the uncontrolled



(a) Computed state for $\beta = 10^{-2}$, $\alpha = 1$



(b) Computed state for $\beta = 10^{-2}$, $\alpha = 10$

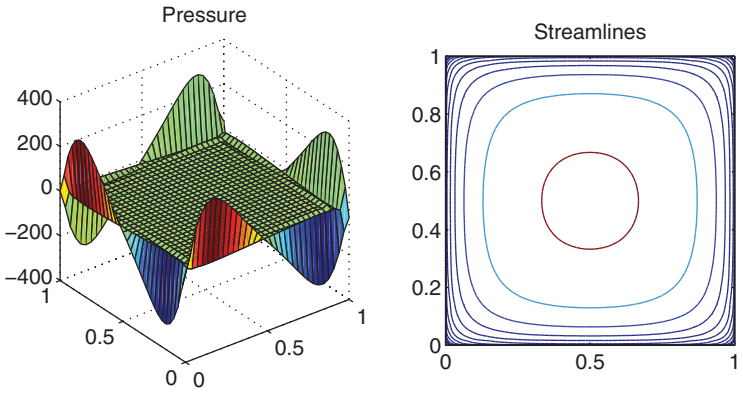
FIG. 3.5. *Computed states for Example* 3.2 *in two dimensions,* $\beta = 10^{-2}$.

problem, the solution of which is shown in Figure 3.1. However, one can conceive of situations where we require a tighter bound on the pressure, and hence a higher value of α.

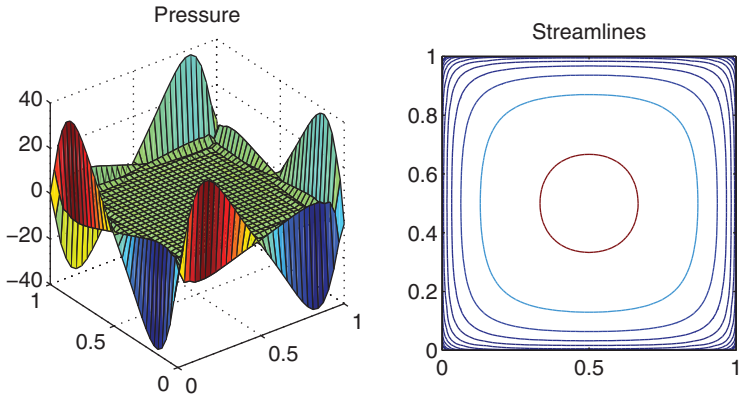
Example 3.3. Let $\Omega = [0,1]^3$, the unit cube, and consider an optimal control problem of the form (1.1) with the three-dimensional equivalent of the Dirichlet boundary conditions as given in Example 3.1; i.e., $\vec{v} = \vec{0}$ on the boundary except on the face where $y = 1$ and $z = 1$ when $\vec{v} = -\vec{\mathbf{j}}$. We take the desired pressure as $\widehat{p} = 0$ and $\widehat{\vec{v}} = y\vec{\mathbf{i}} - x\vec{\mathbf{j}} + z\vec{\mathbf{k}}$.

We solve this three-dimensional problem using the equivalent of the preconditioners employed for Example 3.2. Here $\underline{K}_0^{-1}$ is given by three geometric multigrid V-cycles, which use two steps of relaxed Jacobi as a pre- and post-smoother. The results are presented in Table 3.4. As we can see, the preconditioners perform as well—if not better—in three dimensions.

We have only presented a simple distributed control problem here. It is possible to solve other types of control problems using the same method—see [23] for a discussion in the simpler case of Poisson control. It is also possible to use this method together with bound constraints on the control—Stoll and Wathen [29] discuss this approach in consideration of the Poisson control problem.



(a) Computed state for $\alpha = 1$



(b) Computed state for $\alpha = 10$

FIG. 3.6. *Computed states for Example* 3.2 *in two dimensions,* $\beta = 10^{-5}$.

TABLE 3.4

Comparison of solution methods for solving Example 3.3 *using MINRES and BPCG preconditioned with the block diagonal and block lower-triangular preconditioners, respectively, with two steps of inexact Uzawa approximating* $\mathcal{K}$ *and three GMG V-cycles approximating* $\underline{K}$.

h	size	MINRES		BPCG	
		time	its	time	its
2^{-2}	3,337	3.286	33	1.940	18
2^{-3}	31,833	38.128	36	19.985	18
2^{-4}	277,945	536.566	40	189.921	18
2^{-5}	2,322,297	4660.230	42	2565.626	20

4. Conclusions. In this paper we have presented two preconditioners—one for MINRES, and one for CG in a nonstandard inner product—that can be used to solve problems in Stokes control. These both rely on effective approximations to the (1,1) block, which is composed of mass matrices, and to the Schur complement. We advocate using the Chebyshev semi-iteration used to accelerate a relaxed Jacobi iteration as an approximation to the (1,1) block, and an inexact Uzawa-based approximation for the Schur complement. We have given some theoretical justification for the effectiveness of such preconditioners and have given some numerical results in both two and three dimensions.

We compared these results with those for solving the equivalent forward problem, and the iteration count is only marginally higher in the control case, and it behaves in broadly the same way as the iterations taken to solve the forward problem as the mesh size decreases. These approximations therefore seem reasonable for problems of this type.

In practice one may only be able to control part of the domain, which would give a singular (1,1) block. At present our technique is unable to handle this situation, but we hope that—with further work—the same paradigm will be effective in this situation.

While the problems we have discussed are artificial, the ideas presented here have the potential to be extended to develop preconditioners for a variety of problems, with the additional constraints and features that real-world applications require.

REFERENCES

[1] U. M. Ascher and E. Haber, *A multigrid method for distributed parameter estimation problems*, Electron. Trans. Numer. Anal., 15 (2003), pp. 1–17.

[2] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.

[3] G. Biros and G. Dogan, *A multilevel algorithm for inverse problems with elliptic PDE constraints*, Inverse Problems, 24 (2008), article 034010.

[4] G. Biros and O. Ghattas, *Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part* I: *The Krylov–Schur solver*, SIAM J. Sci. Comput., 27 (2005), pp. 687–713.

[5] A. Borzì and V. Schulz, *Multigrid methods for PDE optimization*, SIAM Rev., 51 (2009), pp. 361–395.

[6] J. Boyle, M. D. Mihajlovic, and J. A. Scott, *HSL_MI20: An Efficient AMG Preconditioner*, Technical report RAL-TR-2007-021, Department of Computational and Applied Mathematics, Rutherford Appleton Laboratory, 2007.

[7] D. Braess and D. Peisker, *On the numerical solution of the biharmonic equation and the role of squaring matrices for preconditioning*, IMA J. Numer. Anal., 6 (1986), pp. 393–404.

[8] 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.

[9] J. H. Bramble and J. E. Pasciak, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp., 50 (1988), pp. 1–17.

[10] H. S. Dollar, N. I. M. Gould, M. Stoll, and A. J. Wathen, *Preconditioning saddle-point systems with applications in optimization*, SIAM J. Sci. Comput., 32 (2010), pp. 249–270.

[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] H. C. Elman, *Multigrid and Krylov subspace methods for the discrete Stokes equations*, Internat. J. Numer. Methods Fluids, 22 (1995), pp. 755–770.

[13] H. Elman, D. Silvester, and A. Wathen, *Finite elements and fast iterative solvers: With applications in incompressible fluid dynamics*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2005.

[14] M. Engel and M. Griebel, *A multigrid method for constrained optimal control problems*, J. Comput. Appl. Math., 235 (2011), pp. 4368–4388.

[15] E. Haber and U. Ascher, *Preconditioned all-at-once methods for large sparse parameter estimation problems*, Inverse Problems, 17 (2000), pp. 1847–1864.

[16] R. Herzog and E. Sachs, *Preconditioned conjugate gradient method for optimal control problems with control and state constraints*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2291–2317.

[17] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand., 49 (1952), pp. 409–436.

[18] A. Klawonn, *Block-triangular preconditioners for saddle point problems with a penalty term*, SIAM J. Sci. Comput., 19 (1998), pp. 172–184.

[19] J. Liesen and B. N. Parlett, *On nonsymmetric saddle point matrices that allow conjugate gradient iterations*, Numer. Math., 108 (2008), pp. 605–624.

[20] A. Meyer and T. Steidten, *Improvement and Experiments on the Bramble–Pasciak Type CG for Mixed Problems in Elasticity*, Technical report, TU Chemnitz, Germany, 2001.

[21] M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (2000), pp. 1969–1972.
[22] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
[23] T. Rees, H. S. Dollar, and A. J. Wathen, *Optimal solvers for PDE-constrained optimization*, SIAM J. Sci. Comput., 32 (2010), pp. 271–298.
[24] T. Rees, M. Stoll, and A. J. Wathen, *All-at-once preconditioning in PDE-constrained optimization*, Kybernetika, 46 (2010), pp. 341–360.
[25] T. Rees and M. Stoll, *Block triangular preconditioners for PDE-constrained optimization*, Numer. Linear Algebra Appl., 17 (2010), pp. 977–996.
[26] J. Schöberl and W. Zulehner, *Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization problems*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 752–773.
[27] D. Silvester and A. Wathen, *Fast iterative solution of stabilised Stokes systems. Part* II: *Using general block preconditioners*, SIAM J. Numer. Anal., 31 (1994), pp. 1352–1367.
[28] M. Stoll and A. Wathen, *Combination preconditioning and the Bramble–Pasciak+ preconditioner*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 582–608.
[29] M. Stoll and A. Wathen, *Preconditioning for Active Set and Projected Gradient Methods as Semi-smooth Newton Methods for PDE-constrained Optimization with Control Constraints*, Technical report 09/25, Oxford Centre for Collaborative Applied Mathematics, 2009.
[30] M. Stoll, *Solving Linear Systems Using the Adjoint*, Ph.D. thesis, University of Oxford, 2009.
[31] H. S. Thorne, *Properties of linear systems in PDE-constrained optimization. Part* I: *Distributed control*, Technical report RAL-TR-2009-017, Rutherford Appleton Laboratory, 2009.
[32] A. J. Wathen and T. Rees, *Chebyshev semi-iteration in preconditioning for problems including the mass matrix*, Electron. Trans. Numer. Anal., 34 (2009), pp. 125–135.
[33] A. J. Wathen, *Realistic eigenvalue bounds for the Galerkin mass matrix*, IMA J. Numer. Anal., 7 (1987), pp. 449–457.
[34] W. Zulehner, *Analysis of iterative methods for saddle point problems: A unified approach*, Math. Comp., 71 (2001), pp. 479–505.