BIT Numerical Mathematics **43:** 961–974, 2003.

PRECONDITIONING TECHNIQUES FOR NEWTON'S METHOD FOR THE INCOMPRESSIBLE NAVIER–STOKES EQUATIONS*

H. C. ELMAN[1], D. LOGHIN[2,†] and A. J. WATHEN[2]

[1] *Department of Computer Science, University of Maryland, College Park, MD 20742, USA. email: elman@cs.umd.edu*

[2] *Oxford University Computing Laboratory, Parks Road, Oxford, OX1 3QD. e-mail: {daniel.loghin,wathen}@comlab.ox.ac.uk*

Abstract.

Newton's method for the incompressible Navier–Stokes equations gives rise to large sparse non-symmetric indefinite matrices with a so-called saddle-point structure for which Schur complement preconditioners have proven to be effective when coupled with iterative methods of Krylov type. In this work we investigate the performance of two preconditioning techniques introduced originally for the Picard method for which both proved significantly superior to other approaches such as the Uzawa method. The first is a block preconditioner which is based on the algebraic structure of the system matrix. The other approach uses also a block preconditioner which is derived by considering the underlying partial differential operator matrix. Analysis and numerical comparison of the methods are presented.

AMS subject classification (2000): 65F10, 65H10, 65F50, 65N30.

Key words: block preconditioners, Navier–Stokes, Newton's method.

1 Problem description.

Mixed finite element discretizations of Navier–Stokes equations give rise to nonlinear systems with the following saddle-point structure (see, e.g., [7, 14])

$$K(\mathbf{u})\begin{pmatrix}\mathbf{u}\\ \mathbf{p}\end{pmatrix}=\begin{pmatrix}F(\mathbf{u}) & B^t\\ B & 0\end{pmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{p}\end{pmatrix}=\mathbf{f}, \tag{1.1}$$

where $K(\mathbf{u})\in\mathbb{R}^{n\times n}$, $F(\mathbf{u})\in\mathbb{R}^{n_1\times n_1}$ is nonsymmetric and possibly indefinite and $B^t\in\mathbb{R}^{n_1\times n_2}$ has nontrivial kernel spanned by the constant vector.

Since K is large and sparse, a competitive class of solution methods is that of iterative solvers of Krylov type combined with suitable preconditioning. In this

* Received October 2002. Revised April 2003. Communicated by D. B. Szyld.

† This work was supported by a grant from EPSRC.

work we are interested in the case where a Newton linearization is employed to solve (1.1); this leads to an iteration of the form [14, §10.3]

$$J(\mathbf{u}^k)\begin{pmatrix}\delta\mathbf{u}^{k+1}\\ \delta\mathbf{p}^{k+1}\end{pmatrix}=\begin{pmatrix}F(\mathbf{u}^k)+M(\mathbf{u}^k) & B^t\\ B & 0\end{pmatrix}\begin{pmatrix}\delta\mathbf{u}^{k+1}\\ \delta\mathbf{p}^{k+1}\end{pmatrix}=\mathbf{r}_k, \tag{1.2}$$

where $M(\mathbf{u})=A_{\mathbf{u}}(\mathbf{u})\cdot\mathbf{u}$, $\mathbf{u}^{k+1}=\mathbf{u}^k+\delta\mathbf{u}^{k+1}$, $\mathbf{r}_k=\mathbf{f}-K(\mathbf{u}^k)\mathbf{u}^k$.

On the other hand, the Jacobian matrix $J(\mathbf{u}^k)$ is related to the system matrix arising from a Picard linearization of (1.1),

$$K(\mathbf{u}^k)\begin{pmatrix}\delta\mathbf{u}^{k+1}\\ \delta\mathbf{p}^{k+1}\end{pmatrix}=\begin{pmatrix}F(\mathbf{u}^k) & B^t\\ B & 0\end{pmatrix}\begin{pmatrix}\delta\mathbf{u}^{k+1}\\ \delta\mathbf{p}^{k+1}\end{pmatrix}=\mathbf{r}_k, \tag{1.3}$$

for which efficient preconditioners are available. Our aim is to adapt and analyze the preconditioned iterative solvers devised for the Picard iteration for problems where Newton's method is employed. Henceforth, we will use the notation

$$F(\rho,\mathbf{u}^k)=F(\mathbf{u}^k)+\rho M(\mathbf{u}^k); \tag{1.4}$$

where $\rho\in\{0,1\}$ to denote the matrices arising from the above two discretizations. To complete the picture, we note here that for flow problems in $\mathbb{R}^d$ $(d=2,3)$ there holds $F(0,\mathbf{u}^k)=I_d\otimes F(\mathbf{u}^k)$, where $F(\mathbf{u}^k)$ is nonsymmetric and positive-definite and comes from the discretization of an advection–diffusion operator

$$F(\mathbf{u}^k)=\nu L+N(\mathbf{u}^k),$$

where ν is the diffusion parameter and L is a discrete Laplacian. The matrix $M(\mathbf{u}^k)$ is also a $d\times d$ block-matrix and represents the discretization of a 'zero-order operator'. Hence it is well-conditioned; however, depending on the problem, it can be indefinite which leads to the possibility of $F(1,\mathbf{u}^k)$ being indefinite.

Finally, we note that the parameter ν is very important in practical applications. Its reciprocal, ν^{-1}, is the Reynolds number and given the size of the problem n, there are parameters ν for which the nonlinear problem (1.1) does not have a unique solution. These critical values of ν where non-uniqueness sets in can be detected in various ways, e.g., continuation methods. For this reason, optimality of our solvers will be measured in terms of dependence of the number of iterations on two parameters: n and ν.

Whether Picard, Newton or variations thereof, linearizations of (1.1) preserve the saddle-point structure of the system matrix for which Schur complement preconditioners perform optimally. More precisely, it is known that the key to efficient preconditioning is the Schur complement

$$S=-BF(\rho,\mathbf{u}^k)^{-1}B^t; \tag{1.5}$$

in particular, a block (right) preconditioner of the form

$$P=\begin{pmatrix}F(\rho,\mathbf{u}^k) & B^t\\ 0 & S\end{pmatrix} \tag{1.6}$$

leads to convergence of any Krylov subspace iterative solver in 2 iterations [13, 8]. The difficulty with this approach obviously relates to applying the action of the inverse of the Schur complement (1.5) as part of the preconditioning technique. It is usually the case that the Schur complement is not available, due to its dense structure and that suitable approximations (or factorizations) are required if we are to stand any chance to make this approach competitive. Two successful Schur complement approximations have been devised, analyzed and tested for the case of the Picard iteration ($\rho = 0$). We consider these approaches below.

The first attempt to approximate the Schur complement corresponding to the Picard iteration with a nonsymmetric matrix was presented by Elman in ([4]) who suggested the approximation

$$S^{-1} \approx (BB^t)^{-1}(BF(\mathbf{u}^k)B^t)(BB^t)^{-1}.$$

It is shown in [4] that the performance of the block preconditioner (1.6) with this choice of approximation has a mild dependence on both ν and n for the Picard iteration. We refer to [4] for the derivation and details; we note however that this choice of approximation is algebraic, i.e., it can be generalized to any problem of saddle-point type and is independent of the underlying differential operator corresponding to $F(\rho, \mathbf{u}^k)$. For this reason, we will take our first approximation of the Schur complement of the Newton method to be

$$S_1^{-1} = (BB^t)^{-1}(BF(1, \mathbf{u}^k)B^t)(BB^t)^{-1}.$$

The other approach we consider here is derived in [9]. In this approach, the Schur complement is seen as the discretization of the fundamental solution for the differential operator which gives rise to $K(\mathbf{u}^k)$, the Picard linearization matrix. The resulting approximation of the Schur complement is also defined as a factorization of sparse matrices

$$S_2^{-1} = M_p^{-1} F_p A_p^{-1},$$

where M_p, F_p, A_p are projections onto the pressure space of the identity, advection–diffusion and Laplace operators, respectively (see [9] for details). The cost of assembling the preconditioner in this fashion is justified by the improved performance which becomes linear in the size of the problem, although it still exhibits a mild dependence on ν.

Unlike the first choice above, this preconditioner depends strongly on what the underlying operator is: if $F(\rho, \mathbf{u}^k)$ changes, the new underlying operator will lead to a new fundamental solution which will have to be computed (if it exists) and discretized. Since $F(1, \mathbf{u}^k)$ corresponds to an operator with non-constant coefficients, the fundamental solution is not available to our knowledge. For this reason we have not been able to adapt this preconditioner to Newton's method. We therefore analyze and test the above Schur complement approximation S_2 (derived for the Picard iteration) on Newton's method.

2 Eigenvalue analysis.

In this section we present an analysis of the above preconditioning techniques for the case where the system matrix results from a mixed finite element discretization of the Navier–Stokes equations. These discretizations have the advantage of satisfying certain stability conditions which arise in the formulation of the problem. More precisely, we assume that the following conditions (known as the Babuška–Brezzi conditions) are satisfied by the Jacobian matrix $J(\mathbf{u}^k)$ for all $\mathbf{u}^k$ and for all ν

$$\max_{\mathbf{w}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\frac{\mathbf{w}^t J(\mathbf{u}^k)\mathbf{v}}{\|\mathbf{v}\|_H\|\mathbf{w}\|_H} \le \Gamma, \tag{2.1a}$$

$$\min_{\mathbf{w}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\max_{\mathbf{v}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\frac{\mathbf{w}^t J(\mathbf{u}^k)\mathbf{v}}{\|\mathbf{v}\|_H\|\mathbf{w}\|_H} \ge \gamma, \tag{2.1b}$$

where $\Gamma, \gamma > 0$. These conditions guarantee existence and uniqueness of a finite element approximation ([6, Chap. IV]). Here H is a symmetric positive-definite matrix which defines the discrete norm of the problem; this norm depends on the choice of function spaces where the solution is sought. For the case of the standard mixed finite-element approximations of the Navier–Stokes equations the norm-matrix H takes the form

$$H = \begin{pmatrix} A & 0 \\ 0 & M_p \end{pmatrix},$$

where $A = I_d \otimes L$ is a d-dimensional Laplacian matrix and M_p is a Grammian matrix (mass matrix) assembled on the pressure space.

Let us consider a general block-triangular preconditioner of the form (1.6) with $F(\rho, \mathbf{u}^k)$ and S replaced by the respective approximations $\widehat{F}, \widehat{S}$

$$P = \begin{pmatrix} \widehat{F} & B^t \\ 0 & \widehat{S} \end{pmatrix}. \tag{2.2}$$

The preconditioned matrix is then

$$JP^{-1} = \begin{pmatrix} F\hat{F}^{-1} & (I - F\hat{F}^{-1})B^t\hat{S}^{-1} \\ B\hat{F}^{-1} & S\hat{S}^{-1} \end{pmatrix}$$

and it appears that the derivation of eigenvalue bounds for the above system is not a simple task. For this reason we make recourse to the general analysis of block-preconditioners for saddle-point problems presented in [11]. We recall here the results of interest.

Definition 2.1 (H-norm equivalence). *Non-singular matrices $M, N \in \mathbb{R}^{n\times n}$ are said to be H-norm equivalent if there exist constants α, β independent of n such that for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$*

$$\alpha \le \frac{\|M\mathbf{x}\|_H}{\|N\mathbf{x}\|_H} \le \beta.$$

We write

$$M \sim_H N.$$

Thus, H-norm equivalence simply means that the H-singular values of MN^{-1} are bounded. An immediate consequence is that the eigenvalues of the preconditioned system are bounded independently of the size of the problem

$$\alpha < |\lambda(MN^{-1})| < \beta;$$

moreover, H-norm equivalence is an equivalence relation on $\mathbb{R}^{n\times n}$. We will need the following results proved in [11].

LEMMA 2.1. *Let* (2.1) *hold. Then*

$$H \sim_{H^{-1}} J.$$

Moreover, if for all $\mathbf{v} \in \mathbb{R}^{n_1\times n_1}$

$$\mathbf{v}^t F(1,\mathbf{u}^k)\mathbf{v} \geq \eta\|\mathbf{v}\|_A^2 \tag{2.3}$$

then the Schur complement $S = BF(1,\mathbf{u}^k)^{-1}B^t$ *satisfies*

$$S \sim_{M_p^{-1}} M_p. \tag{2.4}$$

LEMMA 2.2. *Let* (2.1) *hold with* H *defined as above and let* P *be defined in* (2.2). *Then* $P \sim_{H^{-1}} J$ *if*

$$\widehat{F} \sim_{A^{-1}} A, \qquad \widehat{S} \sim_{M_p^{-1}} M_p.$$

The above results simplify our task considerably; in particular, under standard assumptions on the matrix J, it is sufficient to perform our analysis on the equivalence of the diagonal blocks of H and P.

Let us consider first a norm-equivalent approximation $\widehat{F}$. Since $F(0,\mathbf{u}^k)$ is the discretization of a (positive-definite) advection–diffusion operator it satisfies (see, e.g., [15])

$$\mathbf{v}^t F(0,\mathbf{u}^k)\mathbf{v} \geq \eta\|\mathbf{v}\|_A^2 \quad \forall \mathbf{v} \in \mathbb{R}^{n_1\times n_1}. \tag{2.5}$$

This bound implies that $F(0,\mathbf{u}^k)$ satisfies a min-max bound of type (2.1b) with respect to the A-norm; since $F(0,\mathbf{u}^k)$ also satisfies a max-max bound with respect to the same norm, we can apply Lemma 2.1 to deduce that

$$F(0,\mathbf{u}^k) \sim_{A^{-1}} A.$$

It can be shown that there exists a value ν_0 such that a bound of type (2.3) holds also for $F(1,\mathbf{u}^k)$ [6, p. 300] and thus $F(1,\mathbf{u}^k) \sim_{A^{-1}} A$. We remark here that the value of ν_0 is not known *a priori* and that it depends on application. We will investigate this issue in the numerics section.

REMARK 2.1. In general, even the weaker bound

$$\min_{\mathbf{w}\in\mathbb{R}^{n_1}\setminus\{\mathbf{0}\}} \max_{\mathbf{v}\in\mathbb{R}^{n_1}\setminus\{\mathbf{0}\}} \frac{\mathbf{w}^t F(1,\mathbf{u}^k)\mathbf{v}}{\|\mathbf{v}\|_A\|\mathbf{w}\|_A} \geq \eta \tag{2.6}$$

is not implied by the stability conditions (2.1). In fact, the Babuška–Brezzi conditions imply that the above min-max bound only holds over $\ker B$ [1].

Since the inverse of the matrix $F(1,\mathbf{u}^k)$ can be quite expensive to compute in practice we consider below two block approximations. For this purpose, let us look at the structure of $F(1,\mathbf{u}^k)$ for two-dimensional problems (cf. (1.4))

$$F(1,\mathbf{u}^k) = \begin{pmatrix} F(\mathbf{u}^k) & 0 \\ 0 & F(\mathbf{u}^k) \end{pmatrix} + \begin{pmatrix} M_{11}(\mathbf{u}^k) & M_{12}(\mathbf{u}^k) \\ M_{21}(\mathbf{u}^k) & M_{22}(\mathbf{u}^k) \end{pmatrix}.$$

Simple approximations are given by

$$F^T(\mathbf{u}^k) = \begin{pmatrix} F(\mathbf{u}^k) + M_{11}(\mathbf{u}^k) & M_{12}(\mathbf{u}^k) \\ 0 & F(\mathbf{u}^k) + M_{22}(\mathbf{u}^k) \end{pmatrix} \tag{2.7}$$

and

$$F^D(\mathbf{u}^k) = \begin{pmatrix} F(\mathbf{u}^k) + M_{11}(\mathbf{u}^k) & 0 \\ 0 & F(\mathbf{u}^k) + M_{22}(\mathbf{u}^k) \end{pmatrix}. \tag{2.8}$$

Given their structure, $F^T(\mathbf{u}^k), F^D(\mathbf{u}^k)$ satisfy (2.1a) if $F(1,\mathbf{u}^k)$ does. If (2.3) holds for $F(1,\mathbf{u}^k)$ then $F^D(\mathbf{u}^k)$ satisfies (2.3) also and thus

$$F^D(\mathbf{u}^k) \sim_{A^{-1}} A. \tag{2.9}$$

If moreover $\|M_{12}(\mathbf{u}^k)\| < \eta$, then one can show that $\mathbf{v}^t F^T(\mathbf{u}^k)\mathbf{v} \geq \frac{1}{2}\eta\|\mathbf{v}\|_A^2$; thus

$$F^T(\mathbf{u}^k) \sim_{A^{-1}} A. \tag{2.10}$$

We note that for a quasi-uniform discretization (see below) $\|M_{12}(\mathbf{u}^k)\| = \mathrm{O}(h^2)$ so that the above restriction is satisfied if $\eta > ch^2$. However, $\eta = \mathrm{O}(\nu)$ and $\nu > \nu_0$ is a stronger restriction on $\|M_{12}(\mathbf{u}^k)\|$. Henceforth, we restrict our attention to the regime $\nu > \nu_0$, for which $F(1,\mathbf{u}^k), F^T(\mathbf{u}^k), F^D(\mathbf{u}^k)$ are all norm equivalent to A. This equivalence is useful in the analysis we present below.

In the following we assume a quasi-uniform discretization of our physical domain, i.e., we assume that we have a subdivision of our computational domain in which all the simplices do not exceed a diameter h. Moreover, the following bounds can be shown to hold for all $\mathbf{v} \in \mathbb{R}^{n_1}, \mathbf{q} \in \mathbb{R}^{n_2} \setminus \ker B^t$ and $\Omega \in \mathbb{R}^2$

$$c_1 h^2\|\mathbf{v}\|^2 \leq \mathbf{v}^t A\mathbf{v} \leq c_2\|\mathbf{v}\|^2, \tag{2.11a}$$

$$c_3 h^4\|\mathbf{q}\|^2 \leq \mathbf{q}^t BB^t\mathbf{q} \leq c_4 h^2\|\mathbf{q}\|^2, \tag{2.11b}$$

$$c_5 h^2\|\mathbf{q}\|^2 \leq \mathbf{q}^t M_p\mathbf{q} \leq c_6 h^2\|\mathbf{q}\|^2. \tag{2.11c}$$

The following results provide eigenvalue bounds for the system matrix preconditioned in the fashion described in the previous section.

THEOREM 2.1. *Let* (2.1) *hold and let* $\nu > \nu_0$ *be a viscosity parameter such that* $F(1, \mathbf{u}^k)$ *satisfies* (2.3). *Let*

$$P_i = \begin{pmatrix} F(1, \mathbf{u}^k) & B^t \\ 0 & S_i \end{pmatrix}, \tag{2.12}$$

where $S_i, i = 1, 2$ *are defined as above. Then there exist constants* $C_i, i = 1, \ldots, 4$ *independent of* n *such that*

$$\begin{aligned} C_1 &\leq |\lambda(JP_1^{-1})| \leq C_2 h^{-2}, \\ C_3 &\leq |\lambda(JP_2^{-1})| \leq C_4. \end{aligned}$$

PROOF. We first prove the bounds for P_2. Since by hypothesis (2.3) holds, by the above discussion $F(1, \mathbf{u}^k) \sim_{A^{-1}} A$; in particular, (2.1a) implies

$$\|F(1, \mathbf{u}^k)A^{-1}\|_{A^{-1}} = \max_{\mathbf{v} \in \mathbb{R}^{n_1} \setminus \mathbf{0}} \frac{\|F(1, \mathbf{u}^k)\mathbf{v}\|_{A^{-1}}}{\|A\mathbf{v}\|_{A^{-1}}} \leq \Gamma. \tag{2.13}$$

According to Lemma 2.2 we only have to check that $S_2 \sim_{M_p^{-1}} M_p$. This is equivalent to showing that there exist constants α, β such that

$$\alpha \leq \frac{\|A_p F_p^{-1} \mathbf{x}\|_{M_p^{-1}}}{\|\mathbf{x}\|_{M_p^{-1}}} \leq \beta.$$

On the other hand the above inequality holds with respect to the l_2-norm ([12], see also [10]) and the required result follows from the spectral equivalence between the mass matrix M_p and the identity [16].

The bound for P_1 follows in a similar way, the only difference being that we have to show that β is of order h^{-2} in this case. We have

$$\begin{aligned} \|M_p S_1^{-1}\|_{M_p^{-1}} &= \|M_p^{1/2} S_1^{-1} M_p^{1/2}\| \\ &\leq c_6 h^2 \|(BB^t)^{-1} B F(1, \mathbf{u}^k) B^t (BB^t)^{-1}\| \\ &\leq c_6 h^2 \|A^{-1/2} F(1, \mathbf{u}^k) A^{-1/2}\| \, \|(BB^t)^{-1} B A B^t (BB^t)^{-1}\| \\ &\leq c_6 \Gamma h^2 \|(BB^t)^{-1} B A B^t (BB^t)^{-1}\| \\ &\leq c_6 \Gamma h^2 \|A\| \, \|(BB^t)^{-1}\| \\ &\leq c_6 c_2 c_3^{-1} \Gamma h^{-2}, \end{aligned}$$

where we used (2.13) and the fact that $\|A^{-1/2} F(1, \mathbf{u}^k) A^{-1/2}\| = \|F(1, \mathbf{u}^k) A^{-1}\|_{A^{-1}}$.

For the lower bound we note that

$$\|S_1 M_p^{-1}\|_{M_p^{-1}} = \|M_p^{-1/2} S_1 M_p^{-1/2}\| \leq \frac{1}{\sigma_{\min}(M_p^{1/2} S_1^{-1} M_p^{1/2})}.$$

Let $S_0 = (BB^t)(BAB^t)^{-1}(BB^t)$ and let $\mathbb{R}^{n_2}_* = \mathbb{R}^{n_2} \setminus \ker B^t$. We have

$$\begin{aligned}
\sigma_{\min}(M_p^{1/2}S_1^{-1}M_p^{1/2}) &\geq \min_{\mathbf{q}\in\mathbb{R}^{n_2}_*} \frac{\mathbf{q}^t M_p^{1/2}S_1^{-1}M_p^{1/2}\mathbf{q}}{\mathbf{q}^t\mathbf{q}} \\
&= \min_{\mathbf{r}\in\mathbb{R}^{n_2}_*} \frac{\mathbf{r}^t S_1^{-1}\mathbf{r}}{\mathbf{r}^t S_0^{-1}\mathbf{r}} \frac{\mathbf{r}^t S_0^{-1}\mathbf{r}}{\mathbf{r}^t(BA^{-1}B^t)^{-1}\mathbf{r}} \frac{\mathbf{r}^t(BA^{-1}B^t)^{-1}\mathbf{r}}{\mathbf{r}^t M_p^{-1}\mathbf{r}} \\
&\geq \min_{\mathbf{z}\in\mathbb{R}^{n_2}_*} \frac{\mathbf{z}^t F(1,\mathbf{u}^k)\mathbf{z}}{\mathbf{z}^t A\mathbf{z}} \times \\
&\quad \times \min_{\mathbf{r}\in\mathbb{R}^{n_2}_*} \frac{\mathbf{r}^t S_0^{-1}\mathbf{r}}{\mathbf{r}^t(BA^{-1}B^t)^{-1}\mathbf{r}} \frac{\mathbf{r}^t(BA^{-1}B^t)^{-1}\mathbf{r}}{\mathbf{r}^t M_p^{-1}\mathbf{r}} \\
&\geq \eta \cdot 1 \cdot c,
\end{aligned}$$

where we used Lemma 2.1, the fact that $F(1,\mathbf{u}^k)$ satisfies (2.3) and the result proved in [4] that the smallest eigenvalue of

$$(BA^{-1}B^t)\mathbf{x} = \lambda S_0\mathbf{x}, \quad \mathbf{x}\in\mathbb{R}^{n_2}_*$$

is greater or equal to 1. □

Let P_i^T, P_i^D denote preconditioners of type (2.12) with $F(1,\mathbf{u}^k)$ replaced by $F^T(\mathbf{u}^k), F^D(\mathbf{u}^k)$ respectively. The following result shows that the bounds of Theorem 2.1 hold also for these approximations.

COROLLARY 2.1. *Let the conditions of Theorem* 2.1 *hold and let* P_i^* *denote either of* P_i^T, P_i^D. *Then there exist constants* $C_i^*, i = 1,\ldots,4$ *independent of* n *such that*

$$\begin{aligned}
C_1^* &\leq |\lambda(JP_1^{*-1})| \leq C_2^* h^{-2} \\
C_3^* &\leq |\lambda(JP_2^{*-1})| \leq C_4^*.
\end{aligned}$$

PROOF. The proof follows similarly by using the norm-equivalences (2.9), (2.10). □

REMARK 2.2. The above analysis relies on relation (2.3). Under this assumption, the analysis for the preconditioned matrix JP_2^{-1} is equivalent to that performed in [10] for preconditioning the Picard iteration. However, the general analysis in [11] allowed us to simplify that analysis and also to provide for the first time analytic bounds for the eigenvalues of JP_1^{-1}.

REMARK 2.3. The above result exhibits only the dependence of the eigenvalues on the mesh parameter. Since by hypothesis, the analysis is restricted to a sufficiently large ν, we will not pursue the dependence on ν here. For the Picard iteration, analyses and experiments on ν dependence are contained in [4, 5, 10].

Numerical experiments indicate that the bounds on the eigenvalues are tight. However, it is well-known that the eigenvalues alone do not always describe

convergence of nonsymmetric iterative solvers. This is also the case here: although the largest modulus eigenvalues of the preconditioned systems behave as described theoretically, they actually belong to a small group of outliers. We found experimentally that most of the eigenvalues are clustered in a region of the plane close to 1, with a small number of eigenvalues migrating from the cluster as the parameters ν, h are reduced. A general qualitative model for convergence of GMRES when clustering occurs can be found in [2]. We recall here their result for one cluster, since it appears to be pertinent to the distribution of eigenvalues resulting from our preconditioning techniques. Given a cluster of eigenvalues

$$\mathcal{C} = \{\lambda : |\lambda - c| < \rho|c|\}$$

to which there correspond $M = n - |\mathcal{C}|$ outliers, whose maximum distance δ away from $\mathcal{C}$ is

$$\delta = \max_{|z-c|=\rho|c|} \max_{1\le j\le M} \frac{|\lambda_j - z|}{|\lambda_j|}$$

the GMRES residual at step $k + d$ satisfies

$$\|\mathbf{r}_{k+d}\| \le C\rho^k \|\mathbf{r}_0\|,$$

where d is the degree of the minimum polynomial associated with the outliers and $C \sim \rho\delta^d$ is independent of k.

A specific numerical study of this phenomenon for preconditioning with P_2 is contained in [5]. In this case the location of the eigenvalues does not depend on the mesh-size and the migration of eigenvalues from the cluster occurs with reducing ν. This leads to a delay in convergence dependent on the number of eigenvalues outside the cluster.

A very similar behaviour is observed for fixed ν in the case of preconditioning with P_1. As described by Theorem 2.1, the maximum modulus eigenvalue of the preconditioned system is bounded from above by a factor of order $\mathrm{O}(h^{-2})$. This is in fact a descriptive bound. Table 2.1 exhibits this dependence for the case $\nu = 1/10$ for the driven cavity test problem (see next section). However, convergence is not entirely described by the behaviour of the largest modulus eigenvalues – indeed, the spectrum is mostly clustered in a small region of the complex plane with only a handful of outliers exhibiting the $\mathrm{O}(h^{-2})$ behaviour. Thus, there appears to occur a clustering similar to the case described above where we precondition with P_2 and where convergence is described through a

Table 2.1: Spectral information for JP_1^{-1}; $n^o(c, r)$ is the percentage of eigenvalues outside the circle of radius r centered at c

$h(n)$	$\max \mathrm{Re}(\lambda)$	$\max \mathrm{Im}(\lambda)$	$n^o(2,1)$	$n^o(3,2)$	$n^o(4,3)$	$n^o(5,4)$
1/16(289)	46.65	6.44	15.2	7.6	3.1	3.8
1/32(1089)	177.55	24.72	15.4	7.8	3.6	3.7
1/64(4096)	694.42	96.74	15.6	7.9	3.8	3.9

delay dependent on the (increasing) number of outliers. Table 2.1 presents also the percentage of (outlying) eigenvalues $n^o(c, r)$ outside the circle of radius r centered at c. We see indeed that the number of eigenvalues outside each disc close to the unity, though in small proportion, is increasing with n. This appears to match the behaviour exhibited in the tests below, where a slight dependence on the size of the problem is noticed, though never a linear one.

3 Numerical experiments.

In this section we present numerical experiments obtained for a standard test problem, the driven cavity flow. This involves solving the Navier–Stokes equations inside the unit square with Dirichlet boundary conditions equal to zero everywhere, except for the boundary $y = 1$ where the horizontal velocity has a positive prescribed profile.

Although our analysis considered the exact Newton's method, in practice it is of considerable advantage to use an inexact version for reasons of computational efficiency. The algorithm we use is a Newton–GMRES method which involves solving (1.2) using GMRES with preconditioning and the following stopping criterion suggested in [3]: writing (1.1) as $\mathcal{F}(w) = 0$, where $w = (\mathbf{u}, \mathbf{p})$, at each Newton step i we stop after k iterations of GMRES if the residual $\mathbf{r}_k$ satisfies

$$\|\mathbf{r}_k\|/\|\mathcal{F}(w^i)\| \leq c\|\mathcal{F}(w^i)\|^q, \tag{3.1}$$

where the choice $c = 10^{-2}, q = 1/4$ does not affect the number of nonlinear iterations in our tests. Moreover, the order of convergence is only marginally affected, as we show below.

We report below the performance of preconditioners $P_i, P_i^T, P_i^D, i = 1, 2$. The tests were performed on discretizations resulting from three successively refined meshes and a range of diffusion parameters ν. The value ν_0 in Thm 2.1 depends on the problem; for our example we found experimentally that $\nu_0 \sim 1/80$. As the results below demonstrate, the regime $\nu \geq \nu_0$ is also the regime where our preconditioners perform best. The finite element discretization is the so-called $Q2 - Q1$ discretization, using a regular mesh of rectangles with finite element bases of quadratic and linear piecewise polynomials for the velocity and pressure, respectively. The initial guess for the Newton iteration was the zero vector, except for the range $\nu \leq 1/640$ for which three steps of the Picard iteration were necessary to provide a initial iterate which ensured convergence of Newton's method. We note that this will lead to an apparent improvement of performance in this regime, although not to a qualitative change. We stopped the Newton (outer) iteration when $\|\mathcal{F}(w^i)\|/\|\mathcal{F}(w^0)\| \leq 10^{-6}$.

The computational cost of our preconditioners is dominated by the inversion of the (1,1)-block $F(1, \mathbf{u}^k)$. For the $Q2 - Q1$ discretization the cost of storing the Schur complement approximations S_i is approximately $25n_2(BB^t)$ $+45n_2(BFB^t)$ for S_1 and $3 \times 9n_2(A_p, F_p, M_p)$ for S_2. The cost of applying these preconditioners will of course depend on the choice of implementation. In the experiments below we implemented them exactly; however, iterative techniques

Table 3.1: Average number of GMRES iterations per Newton step for P_1

$\nu =$	1/10	1/20	1/40	1/80	1/160	1/320	1/640	1/1280
$n = 2{,}467$	13.7	14.2	15.5	17.7	21.4	29.3	31.8	43.2
9,539	20.2	21.2	21.5	23.5	29.2	36.8	40.0	53.2
37,507	29.7	31.7	32.7	33.7	41.2	52.0	62.6	72.3

Table 3.2: Average number of GMRES iterations per Newton step for P_2

$\nu =$	1/10	1/20	1/40	1/80	1/160	1/320	1/640	1/1280
$n = 2{,}467$	13.5	13	17.7	22.8	27.8	47.3	59.5	85.1
9,539	11.5	13.2	16.2	21.2	31.6	44.6	57.2	77.3
37,507	11.7	14.2	16.5	19.5	29.6	43.5	56.5	75.1

Table 3.3: Average number of GMRES iterations per Newton step for P_1^T

$\nu =$	1/10	1/20	1/40	1/80	1/160	1/320	1/640	1/1280
$n = 2{,}467$	13.7	14.2	16.2	19.5	26.0	32.6	37.6	49.1
9,539	20.2	22	22.2	25.7	43.2	44.3	44.5	59.7
37,507	30.2	32	34.5	37.7	49.2	64.0	69.8	80.5

are certainly possible [4, 9] in which case the choice S_2 becomes clearly more competitive.

The results for the 'exact' case, when preconditioners P_1, P_2 are employed are shown in Tables 3.1, 3.2. We see indeed that while the performance of P_2 is mesh-independent, the performance of P_1 exhibits a sub-linear dependence on the size of the problem. This we believe to be explained by the distribution of eigenvalues which was exhibited in Table 2.1, which showed an increasing number of outliers with decreasing mesh-size. As for the viscosity parameter ν, preconditioner P_1 appears to exhibit a milder dependence than P_2, which makes it comparable to P_1 for high-Reynolds number flows, despite its mesh-dependence.

The same behaviour is exhibited by the preconditioners $P_i^T, P_i^D, i = 1, 2$. In particular, the performance remains mesh-independent in the case of P_2^T, P_2^D as shown in Tables 3.4, 3.6; as a consequence, the number of extra iterations compared to the 'exact case' displayed in Table 3.2 is independent of the mesh. This is not the case for P_1^T, P_1^D as shown in Tables 3.3, 3.5, for which this difference (cf. Table 3.1) grows with the mesh parameter, although the number of iterations still grows sub-linearly with respect to the size of the problem.

From a practical point of view it is important to study the deterioration in performance when the exact preconditioners P_i are replaced with P_i^T, P_i^D, as solving systems with F can be quite costly. As the above numerics show, this deterioration is negligible for large ν and quite acceptable for smaller ν. To exemplify this we replaced the exact solution of systems involving $F(1, \mathbf{u}^k)$ with

Table 3.4: Average number of GMRES iterations per Newton step for P_2^T

$\nu =$	1/10	1/20	1/40	1/80	1/160	1/320	1/640	1/1280
$n = 2{,}467$	14.7	14	19.7	26.8	36.4	58.3	71.3	103.5
9,539	12.5	14.7	17.7	24.7	37.4	55.8	67.2	102.6
37,507	12.7	15.5	18.2	21.7	34.6	53.2	67.5	108.6

Table 3.5: Average number of GMRES iterations per Newton step for P_1^D

$\nu =$	1/10	1/20	1/40	1/80	1/160	1/320	1/640	1/1280
$n = 2{,}467$	14	15	17	21.5	28.4	36.2	39.2	51.4
9,539	20.7	23.2	24.5	28.5	37.6	46.8	45.3	59.2
37,507	30.7	33.7	36.5	41.2	54.0	69.2	69.5	80.83

Table 3.6: Average number of GMRES iterations per Newton step for P_2^D

$\nu =$	1/10	1/20	1/40	1/80	1/160	1/320	1/640	1/1280
$n = 2{,}467$	15.7	15.7	22.7	30.6	41.6	65.6	77.9	114.3
9,539	13	16	19.7	25.2	43	63	72.3	111.0
37,507	13.7	17	20.7	25	39.6	59.2	72.5	116.1

a GMRES iteration with preconditioners F^T, F^D. The stopping criterion was the reduction of the relative 2-norm residual by a factor of 10^6 which insured that this inner iteration did not affect the outer iteration count displayed in Tables 3.1, 3.2. The number of iterations was always mesh-independent. We present these results in Table 3.7. We see indeed that the modified preconditioners P_i^T, P_i^D (which require just one solve with F^T and F^D respectively) outperform by far P_i when the inner-GMRES iteration is used to approximate F. Thus, unless a fast solution algorithm is available for $F(1, \mathbf{u}^k)$, the 'ideal' performance presented in Tables 3.1, 3.2 may prove too expensive to achieve.

We end this section with a note on the convergence rate of Newton's method observed for our inexact Newton–GMRES algorithm. The choice of this algorithm and in particular of stopping criterion (3.1) was justified by both practical and theoretical considerations [3]. In Table 3.8 we display the average convergence rate observed in our tests. More precisely, we tabulated the mean of the ratio q_k of logs of consecutive residuals

$$q_k = \frac{\log \|\mathcal{F}(w^k)\|}{\log \|\mathcal{F}(w^{k-1})\|},$$

for the cases $q = 1/4$ and $q = 1$ in (3.1) and for the exact case ($q = \infty$). The choice of preconditioner does not affect the convergence rate displayed below.

Table 3.7: Average number of iterations per outer-GMRES step for inner-GMRES solver using preconditioners F^T, F^D

$\nu =$	1/10	1/20	1/40	1/80	1/160	1/320	1/640	1/1280
F^T	3.0	3.8	4.5	5.6	8.1	12.0	16.4	27.7
F^D	4.8	6.2	8.3	11.3	15.6	22.5	31.7	34.2

Table 3.8: Convergence rate of Newton's method: exact and inexact with $q = 1/4$ and $q = 1$; $n = 9,539$

$\nu =$	1/10	1/20	1/40	1/80	1/160	1/320
$q = 1/4$	1.78	1.77	1.70	1.58	1.43	1.28
$q = 1$	2.00	1.98	1.80	1.57	1.43	1.28
$q = \infty$	2.20	2.02	1.81	1.57	1.43	1.28

We see indeed that the choice of stopping criterion (i.e., q in (3.1)) does not affect significantly the rate of convergence of the method except possibly for the low Reynolds number (large ν) regime. In principle, one could recover this convergence rate by a stricter criterion ($q = 1$, say), at the cost of more iterations. Given our choice of tolerance for the outer nonlinear iteration (10^{-6}), $q = 1/4$ was sufficient to keep the same number of Newton steps as in the exact case. A stricter tolerance may require a larger value of q. However, for high Reynolds number flows a more economical stopping criterion is sufficient due to the sub-optimal convergence of Newton's method.

4 Summary.

The purpose of this work was to analyze and compare block preconditioners for the system matrix arising from a Newton iteration for the Navier–Stokes equations. The preconditioners are saddle-point preconditioners using an algebraic and operator based approximation of the Schur complement. They were introduced originally for the case of the Picard iteration. In each case we give analytic bounds for the spectrum of the preconditioned system under standard stability assumptions for the finite element discretization. The bounds require a further assumption of positive-definiteness of the $(1,1)$-block of the matrix, which holds for large viscosity parameters. The numerical experiments however show no qualitative change even for smaller values of viscosity. In each case the implementation is modular – we only require the action of inverses of sub-blocks which have to be constructed algebraically or assembled together with the system matrix. The performance is very similar to the case of the Picard iteration which has been recently reported. The operator-based preconditioner appears to be more robust and less costly than the algebraic preconditioner and especially the independence of the size of the problem is an attractive feature. However, optimality with respect to the Reynolds number is still to be achieved.

REFERENCES

1. F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York, 1991.
2. S. L. Campbell, I. C. F. Ipsen, C. T. Kelley, and C. D. Meyer, *GMRES and the minimal polynomial*, BIT, 36 (1996), pp. 664–675.
3. R. S. Dembo, S. C. Eisenstat, and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.
4. H. C. Elman, *Preconditioning for the steady-state Navier–Stokes equations with low viscosity*, SIAM J. Sci. Comp., 20 (1999), pp. 1299–1316.
5. H. C. Elman, D. J. Silvester, and A. J. Wathen, *Performance and analysis of saddle point preconditioners for the discrete steady-state Navier–Stokes equations*, Numer. Math., 90 (2001), pp. 665–688.
6. V. Girault and P. A. Raviart, *Finite Element Methods for Navier–Stokes Equations*, Springer-Verlag, 1986.
7. P. M. Gresho and R. L. Sani, *Incompressible Flow and the Finite Element Method: Isothermal Laminar Flow*, Wiley, Chichester, 1998.
8. I. C. F. Ipsen, *A note on preconditioning non-symmetric matrices*, SIAM J. Sci. Comput., 23 (2001), pp. 1050–1051.
9. D. Kay, D. Loghin, and A. J. Wathen, *A preconditioner for the steady-state Navier–Stokes equations*, SIAM J. Sci. Comput., 24 (2002), pp. 237–256.
10. D. Loghin, *Analysis of preconditioned Picard iterations for the Navier–Stokes equations*, submitted to Numer. Math., 2001.
11. D. Loghin and A. J. Wathen, *Analysis of block preconditioners for saddle-point problems*, Tech. Rep. 13, Oxford University Computing Laboratory, 2002. To appear in SISC.
12. T. A. Manteuffel and S. V. Parter, *Preconditioning and boundary conditions*, SIAM J. Numer. Anal., 27 (1989), pp. 656–694.
13. M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comp., 21 (2000), pp. 1969–1972.
14. A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer-Verlag, 1994.
15. H. G. Roos, M. Stynes, and L. Tobiska, *Numerical Methods for Singularly Perturbed Differential Equations*, Springer Series in Computational Mathematics, Springer, 1996.
16. A. J. Wathen, *Realistic eigenvalue bounds for the Galerkin mass matrix*, IMA J. Numer. Anal. (1987), pp. 449–457.