
This version is available at https://strathprints.strath.ac.uk/55440/

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (https://strathprints.strath.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk
ON THE EIGENVALUES AND EIGENVECTORS OF BLOCK TRIANGULAR PRECONDITIONED BLOCK MATRICES

JENNIFER PESTANA

Abstract. Block lower triangular matrices and block upper triangular matrices are popular preconditioners for 2 × 2 block matrices. In this note we show that a block lower triangular preconditioner gives the same spectrum as a block upper triangular preconditioner and that the eigenvectors of the two preconditioned matrices are related.

Key words. block triangular preconditioner, convergence, eigenvalues, eigenvectors, iterative method, saddle point system

AMS subject classifications. 65F08, 65F10, 65F50, 65N22

1. Introduction. Nonsingular block matrices of the form

\[ A = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix}, \]

where \( A \in \mathbb{C}^{n \times n}, B \in \mathbb{C}^{m \times n} \) with \( \text{rank}(B) = m \), \( C \in \mathbb{C}^{m \times m} \), \( m \leq n \) arise in a number of applications, many of which are discussed in the survey paper by Benzi, Golub and Liesen [5, Section 2]. Of particular interest are block matrices for which \( C = 0 \) and/or for which \( A \) is symmetric positive definite and \( C \) is symmetric positive semidefinite [5],[10, Chapters 5 and 7].

In many applications \( A \) in (1.1) is large and sparse, in which case linear systems with \( A \) as the coefficient matrix are typically solved by a preconditioned iterative method. Two popular preconditioners are the block lower triangular matrix \( P_L = \begin{bmatrix} P_A & 0 \\ B & P_S \end{bmatrix} \)

and block upper triangular matrix \( P_U = \begin{bmatrix} P_A & B^T \\ 0 & P_S \end{bmatrix} \),

where \( P_A \in \mathbb{C}^{n \times n} \) and \( P_S \in \mathbb{C}^{m \times m} \) (and, consequently, \( P_L \) and \( P_U \)) are nonsingular.

When \( A \) and \( C \) are Hermitian semidefinite it is known that \( P_U^{-1}A \) and \( P_L^{-1}A \) are similar [14, Remark 2]. (The case in which \( A \) is positive definite was also recently treated by Notay [16, Theorem 3.1].) For non-Hermitian matrices, Bai and Ng [3] analysed the minimal polynomials of \( P_L^{-1}A \) and \( P_U^{-1}A \) when \( P_A = A \) or \( P_S \) is the Schur complement, while Bai [1] obtained identical eigenvalue bounds for \( P_L^{-1}A \) and \( P_U^{-1}A \) in the more general case of inexact \( P_A \) and \( P_S \). Additionally, Bai and Ren [4] applied block triangular preconditioned GMRES [18] and BiCGStab [22] to nonsymmetric 2 × 2 block systems arising from discretizations of third-order ODEs. They

\*This publication is based on work supported by Award No. KUK-C1-013-04, made by King Abdullah University of Science and Technology (KAUST).

1Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK (pestana@maths.ox.ac.uk)

In fact the results of Bai and Ng [3] and Bai [1] are more general than ours since they do not assume that the (1,2) and (2,1) blocks of \( A \) are transposes of each other.
found that iteration counts for upper and lower triangular preconditioners were similar, while eigenvalue plots for \( P_L^{-1}A \) and \( P_U^{-1}A \) were indistinguishable. Additionally, for M-matrices arising from Markov chains Benzi and Uçar [7] noticed little difference between the performances of block lower triangular and block upper triangular preconditioners.

In this note we extend the theoretical results to the non-Hermitian case. We show that \( P_L^{-1}A \) and \( P_U^{-1}A \) have identical spectra and relate the corresponding eigenvectors. When \( C = 0 \) and \( P_L^{-1}A \) and \( P_U^{-1}A \) are diagonalizable, we bound the difference between the condition numbers of the eigenvector matrices; this gives some insight into when we might expect certain iterative methods to converge similarly for the block lower and block upper triangular preconditioned systems. Our results are illustrated on a numerical example.

Throughout \( I_p \in \mathbb{C}^{p \times p} \) denotes the identity matrix of dimension \( p \) and \( \| \cdot \|_2 \) represents the Euclidean vector norm or the corresponding induced matrix norm. The conjugate transpose of a matrix \( E \) is denoted by \( E^* \), its range by \( \text{range}(E) \), its nullspace by \( \text{null}(E) \), and its Moore-Penrose pseudoinverse by \( E^+ \).

2. Eigenvalue, eigenvector and condition number relationships. In this section we state our main results, starting with the equivalence of the spectra of \( P_U^{-1}A \) and \( P_L^{-1}A \).

**Proposition 1.** Let \( A, P_L \) and \( P_U \) be invertible. Then the spectra of \( P_L^{-1}A \) and \( P_U^{-1}A \) are identical, as are the spectra of \( AP_L^{-1} \) and \( AP_U^{-1} \).

**Proof.** Let

\[
Y(\lambda) = \begin{bmatrix} A - \lambda P & B^T \\ (1 - \lambda)B & -(C + \lambda P_S) \end{bmatrix} \quad \text{and} \quad Z(\lambda) = \begin{bmatrix} A - \lambda P & (1 - \lambda)B^T \\ B & -(C + \lambda P_S) \end{bmatrix}.
\]  

Then the eigenvalues \( \lambda_L \) of \( P_L^{-1}A \) must be roots of \( \det(Y(\lambda_L)) = 0 \) while the eigenvalues \( \lambda_U \) of \( P_U^{-1}A \) must satisfy \( \det(Z(\lambda_U)) = 0 \). Setting

\[
J(\lambda) = \begin{bmatrix} I_n & 0 \\ 0 & (1 - \lambda)I_m \end{bmatrix},
\]

we see that, for any \( \lambda \neq 1, 0 = \det(Y(\lambda_L)) = \det(J(\lambda_L)^{-1}Y(\lambda_L)) = \det(Z(\lambda_L)). \) Thus, the non-unit roots of \( \det(Y(\lambda)) = 0 \) and \( \det(Z(\lambda)) = 0 \) coincide. Additionally, \( \det(Y(1)) = \det(Z(1)) = (-1)^m \det(A - P_A) \det(C + P_S) \). The results for right preconditioning follow from the similarity of \( P_L^{-1}A \) and \( AP_L^{-1} \) and of \( P_U^{-1}A \) and \( AP_U^{-1} \).

The above result shows that if eigenvalues alone are important, there is nothing to distinguish \( P_L^{-1}A, AP_L^{-1}, P_U^{-1}A \) and \( AP_U^{-1} \). This may be the case when, for example, we precondition to achieve self-adjointness and positive definiteness with respect to a nonstandard inner product [8, 14, 17]. However, in many situations the eigenvectors will also have an effect on convergence. We relate the eigenvectors of \( P_L^{-1}A \) and \( P_U^{-1}A \) in the following proposition.

**Proposition 2.** Let \( A, P_L \) and \( P_U \) be nonsingular.

1. Suppose that \( \lambda \neq 1 \) is an eigenvalue of \( P_U^{-1}A \). If \( [u_L^T, v_U^T]^T \) is the corresponding eigenvector, then \( J(\lambda)[u_L^T, v_U^T]^T = [u_L^T, (1 - \lambda)v_U^T]^T \) is an eigenvector of \( P_L^{-1}A \) corresponding to this eigenvalue.

2. Suppose that \( \lambda = 1 \) is an eigenvalue of \( P_L^{-1}A \).

   - If \( P_S + C \) is nonsingular then the eigenvectors of \( P_L^{-1}A \) corresponding to \( \lambda \) are of the form \( [u^T, \Theta^T]^T \), where \( u \) is any eigenvector of the matrix...
$P_A^{-1}A$ that corresponds to its unit eigenvalue. Moreover, the eigenvectors of $P_U^{-1}A$ corresponding to $\lambda$ are of the form $[u^T, (P_S + C)^{-1}Bw]^T$.

- If $P_A - A$ is nonsingular then the eigenvectors of $P_U^{-1}A$ corresponding to $\lambda$ are of the form $[0^T, v^T]^T$ where $v$ is any eigenvector of $-P_S^{-1}C$ that corresponds to its unit eigenvalue. Moreover, the eigenvectors of $P_U^{-1}A$ corresponding to $\lambda$ are of the form $[(P_A - A)^{-1}B^Tv]^T, v^T]^T$.

Proof. The first part follows immediately from Proposition 1, since if $\lambda \neq 1$ then $Z(\lambda) = J(\lambda)^{-1}Y(\lambda)J(\lambda)$ where $J(\lambda)$ is as in (2.2).

If $\lambda = 1$, then $Y(1)u_L = 0$ is equivalent to

$$Au_L + B^Tv_L = P_Au_L,$$  \hspace{1cm} (2.3)

$$P_Sv_L + Cv_L = 0,$$  \hspace{1cm} (2.4)

while $Z(1)w_U = 0$ is equivalent to

$$Au_U - P_Au_U = 0,$$  \hspace{1cm} (2.5)

$$Bu_U - Cv_U = P_Sv_U.$$  \hspace{1cm} (2.6)

Note that $-P_S^{-1}C$ does not have an eigenvalue at 1 if and only if $P_S + C$ is nonsingular. In this case, (2.3) and (2.4) show that $v_L = 0$ and that $u_L$ must be an eigenvector of $P_U^{-1}A$ corresponding to the eigenvalue 1. Meanwhile, (2.5) and (2.6) imply $u_U \neq 0$, since otherwise nonsingularity of $P_S + C$ means that $\nu_U = 0$. Thus, $u_L$ is also an eigenvector of $P_U^{-1}A$ corresponding to the eigenvalue 1 and $Bu_U = (C + P_S)v_U$. The last case is proved similarly. \hfill \square

Remark 1. The case that $\lambda = 1$ is an eigenvalue of both $P_A^{-1}A$ and $-P_S^{-1}C$ can also be worked out but is of less interest so we omit it here.

Remark 2. It may be preferable to use right preconditioning rather than left preconditioning, since doing so preserves the residual norm for methods such as GMRES [18]. However, if $X$ is an eigenvector matrix of $P^{-1}A$ then $PX$ is an eigenvector matrix of $AP^{-1}$ for any invertible preconditioner $P$. This allows the eigenvectors of the right preconditioned block matrix to be determined. However, the eigenvector matrices of $AP_L^{-1}$ and $AP_U^{-1}$ do not have as straightforward a relationship as those of $P_L^{-1}A$ and $P_U^{-1}A$ and we do not consider them here.

Particularly important is the case $C \equiv 0 \in A$, for which the eigenvectors of $P_L^{-1}A$ and $P_U^{-1}A$ are more simply related.

Corollary 3. Let $A$, $P_L$ and $P_U$ be nonsingular and let $C \equiv 0$ in (1.1).

1. Suppose that $\lambda \neq 1$ is an eigenvalue of $P_U^{-1}A$. If $[u_L^T, v_U^T]^T$ is the corresponding eigenvector, then $[u_L^T, (1 - \lambda)v_U^T]^T$ is an eigenvector of $P_L^{-1}A$ corresponding to this eigenvalue.

2. Otherwise, $\lambda = 1$ is an eigenvalue of $P_U^{-1}A$ with corresponding eigenvector $[u_T^T, 0^T]^T$, where $u$ is any eigenvector of the matrix $P_A^{-1}A$ that corresponds to its unit eigenvalue. Moreover, the eigenvectors of $P_L^{-1}A$ corresponding to $\lambda$ are of the form $[u^T, (P_S^{-1}Bu)^T]^T$.

Proof. If $C = 0$, then $P_S^{-1}C = 0$ and the last case in Proposition 2 does not apply. \hfill \square

Of interest for Krylov methods such as GMRES is the condition number of the eigenvector matrix [18] when it is invertible.

Corollary 4. Let $A$, $P_L$ and $P_U$ be nonsingular and let $C \equiv 0$ in $A$. Assume that both $P_L^{-1}A = X_L\Lambda X_L^{-1}$ and $P_U^{-1}A = X_U\Lambda X_U^{-1}$ are diagonalizable with $p$
eigenvalues equal to 1, where

\[ X_U = \begin{bmatrix} U^{(1)} & U^{(2)} \\ V^{(1)} & V^{(2)} \end{bmatrix} \quad \text{and} \quad X_L = \begin{bmatrix} U^{(1)} & U^{(2)} \\ 0 & V^{(2)}(I - \Lambda^{(2)}) \end{bmatrix}, \]  

(2.7)

with \( U^{(1)} \in \mathbb{C}^{n \times p} \), \( U^{(2)} \in \mathbb{C}^{n \times (n+m-p)} \), \( V^{(1)} \in \mathbb{C}^{m \times p} \), \( V^{(2)} \in \mathbb{C}^{m \times (n+m-p)} \) and \( \Lambda = \text{diag}(I_p, \Lambda^{(2)}) \). \( \Lambda^{(2)} = \text{diag}(\lambda_{p+1}, \ldots, \lambda_{n+m}) \). For any matrix \( E \), let \( P_E = (I - EE^\dagger) \) and \( Q_E = (I - E^\dagger E) \) be orthogonal projectors onto \( \text{null}(E) \) and \( \text{null}(E^\dagger) \), respectively. Then the 2-norm condition numbers \( \kappa_2(X_L) \) of \( X_L \) and \( \kappa_2(X_U) \) of \( X_U \) are related by

\[ \frac{1}{\alpha} \leq \frac{\kappa_2(X_L)}{\kappa_2(X_U)} \leq \alpha \]  

(2.8)

where, when \( p = 0 \),

\[ \alpha = (1 + \beta(1 + \|V^{(2)}(I - \Lambda^{(2)})(U^{(2)})^\dagger\|_2))(1 + \beta(1 + \|V^{(2)}(U^{(2)})^\dagger\|_2)), \]

\[ \beta = \|V^{(2)}\Lambda^{(2)}\|_2(\|(U^{(2)})^\dagger\|_2 + \|Q_{U^{(2)}}(V^{(2)})^\dagger S^{-1}\|_2), \]

and when \( p \geq 1 \)

\[ \alpha = (1 + \|V^{(1)}(U^{(1)})^\dagger\|_2 + \|F\|_2\beta_1)(1 + \|V^{(1)}(U^{(1)})^\dagger\|_2 + \|F\|_2\beta_2), \]

\[ \beta_1 = \|G^\dagger\|_2 + \|H^\dagger\|_2(1 + \|V^{(2)}(I - \Lambda^{(2)})\|_2\|G^\dagger\|_2), \]

\[ \beta_2 = \|G^\dagger\|_2 + \|H^\dagger\|_2(1 + \|V^{(2)} - V^{(1)}(U^{(1)})^\dagger U^{(2)}\|_2\|G^\dagger\|_2 + \|V^{(1)}(U^{(1)})^\dagger\|_2). \]

Here, \( F = V^{(2)}\Lambda^{(2)} - V^{(1)}(U^{(1)})^\dagger U^{(2)} \), \( G = P_{U^{(2)}}U_2 \), \( H_1 = V^{(2)}(I - \Lambda^{(2)})Q_G \), \( H_2 = (V^{(2)} - V^{(1)}(U^{(1)})^\dagger U^{(2)})Q_G \) and \( S = V^{(2)}Q_{U^{(2)}}(V^{(2)})^\dagger \).

Proof. It is clear from (2.7) that \( X_U = X_L + K \), where

\[ K = \begin{bmatrix} 0 & 0 \\ V^{(1)} & V^{(2)}\Lambda^{(2)} \end{bmatrix}. \]

From Ipsen [12, Corollary 3.3], we have that

\[ \frac{\sigma_{\text{max}}(X_L)}{\sigma_{\text{max}}(X_U)} \leq 1 + \|KX_U^{-1}\|_2 \quad \text{and} \quad \frac{\sigma_{\text{min}}(X_U)}{\sigma_{\text{min}}(X_L)} \leq 1 + \|KX_L^{-1}\|_2, \]  

(2.9)

where \( \sigma_{\text{max}}(X) = \|X\|_2 \) and \( \sigma_{\text{min}}(X) = 1/\|X^{-1}\|_2 \) are the largest and smallest singular values of \( X \). Combining these results gives

\[ \frac{\kappa_2(X_L)}{\kappa_2(X_U)} \leq (1 + \|KX_U^{-1}\|_2)(1 + \|KX_L^{-1}\|_2). \]  

(2.10)

We now obtain expressions for \( KX_U^{-1} \) and \( KX_L^{-1} \), starting with the case \( p = 0 \). Let \( X_{L,U} \) denote \( X_L \) or \( X_U \) as appropriate and let \( \tilde{V}^{(2)} = V^{(2)} \) for \( X_U \) and \( \tilde{V}^{(2)} = V^{(2)}(I - \Lambda^{(2)}) \) for \( X_L \). Then the inverse of \( X_{L,U} \) is \( X_{L,U}^{-1} = X_{L,U}^\dagger(X_{L,U}X_{L,U}^\dagger)^{-1} \), where

\[ X_{L,U}X_{L,U}^\dagger = \begin{bmatrix} I_n & (U^{(2)})(U^{(2)})^\dagger \end{bmatrix} \begin{bmatrix} U^{(1)} & U^{(2)} \\ 0 & V^{(2)} \end{bmatrix} \begin{bmatrix} S \end{bmatrix} \begin{bmatrix} I_n \end{bmatrix} \begin{bmatrix} (U^{(2)})^\dagger \end{bmatrix}^\dagger \]
with $S = \tilde{V}^T Q U (\tilde{V}^T)^\dagger$. Since $C = 0$, $P^{-1}_U A X U = X U A$ implies that $V(2)^T \Lambda(2) = P^{-1}_S B U(2)$. Thus, $V(2)^T \Lambda(2) Q U(2) = 0$ and $S = V(2)^T Q U(2) (V(2))^\dagger$. Straightforward calculation then yields

$$K X^{-1}_L = \begin{bmatrix} 0 & 0 \\ V(2) \Lambda(2) \end{bmatrix} \left[ \begin{array}{cc} I_n & - (U(1))^\dagger U(2) \\ 0 & I_{n+m-p} \end{array} \right] \left[ \begin{array}{cc} (U(1))^\dagger & 0 \\ (I - H_1^2 V(2)(I - \Lambda(2))) G^\dagger & H_1^2 \end{array} \right] = \begin{bmatrix} V(1)(U(1))^\dagger + F(I - H_2^2 (V(2) - V(1)(U(1))^\dagger U(2))) G^\dagger - H_2^2 V(1)(U(1))^\dagger & 0 \\ 0 & I_m \end{bmatrix}.$$ 

and the upper bounds are obtained by bounding $\|K X^{-1}_L\|_2$ and $\|K X^{-1}_U\|_2$. When $p \geq 1$ we use Theorem 2.1 in Tian and Takane [21]. Since $U(1)^\dagger$ has linearly independent columns, $Q U(1) = 0$, and

$$K X^{-1}_U = \begin{bmatrix} 0 & 0 \\ V(2) \Lambda(2) \end{bmatrix} \left[ \begin{array}{cc} I_p & 0 \\ 0 & I_{n+m-p} \end{array} \right] \left[ \begin{array}{cc} (U(1))^\dagger & 0 \\ (I - H_1^2 V(2)(I - \Lambda(2))) G^\dagger & H_1^2 \end{array} \right] = \begin{bmatrix} V(1)(U(1))^\dagger + F(I - H_2^2 (V(2) - V(1)(U(1))^\dagger U(2))) G^\dagger - H_2^2 V(1)(U(1))^\dagger & 0 \\ 0 & I_m \end{bmatrix}.$$ 

The upper bound on the condition number again follows from bounding the norms. Then, the lower bound on the condition number is achieved by bounding $\kappa_2(X_U) / \kappa_2(X_L)$ from above. □

Although the expressions in Corollary 4 are quite complicated, they highlight that the difference between the condition numbers depends not only on the matrices that vary between $X_L$ and $X_U$, namely $V(1)$, $V(2)$ and $V(2)(I - \Lambda(2))$, but also on the conditioning of $U(1)$ and $U(2)$. This is not surprising since $U(1)$ and $U(2)$ affect the conditioning of $X_L$ and $X_U$.

More specifically, when $p = 0$, $\alpha$ is smaller when $\|(U(2))^\dagger\|_2$ is small, i.e., when $U(2)$ is well conditioned in the sense that its smallest (nonzero) singular value is not too small, and when the rows of $V(2)$ are almost orthogonal to the rows of $U(2)$, so that $\|V(2) U(2)^\dagger\|_2$ is small. Also, since $Q U(2)(V(2))^\dagger S^{-1}$ is a right inverse of $V(2)$, we expect its norm to be large when the rows of $V(2)$ are almost linearly dependent. This confirms that well conditioned $X_L$ and $X_U$ and a small perturbation $V(2) \Lambda(2)$ ensure that $\kappa_2(X_L)$ and $\kappa_2(X_U)$ are close.

If $p > 0$, $\alpha$ depends on $\|V(1)(U(1))^\dagger\|_2$, $\|F\|_2$, $\beta_1$ and $\beta_2$. The term $\|V(1)(U(1))^\dagger\|_2$ is small when $\|V(1)\|_2$ is small but the smallest singular value of $U(1)$ is not. Additionally, $\|F\|_2$ is small when $\|V(2) \Lambda(2)\|_2$ and $\|V(1)\|_2$ are small, and the columns of $U(1)$ are almost orthogonal to those of $U(2)$. If range$(U(2)) \subset$ range$(U(1))$, then $G = 0$, $G^\dagger = I$ and a sufficient condition for $\beta_1$ and $\beta_2$ to be small is that $\|V(2)(I - \Lambda(2))\|_2$, $\|V(2)\|_2$ and $\|V(1)\|_2$ are small. Otherwise, $\|G^\dagger\|_2$ is not too large if the columns of $U(1)$ are orthogonal to those of $U(2)$ with $U(2)$ well conditioned. The terms $H_1$ and $H_2$ are the most difficult to analyse. Although $H_1 = 0$ when range$(V(2)(I - \Lambda(2))^\dagger) \subset$ range$(G^\dagger)$ and $H_2 = 0$ when range$(V(2) - V(1)(U(1))^\dagger U(2)^\dagger) \subset$ range$(G^\dagger)$, these conditions may not hold in general. Alternatively, if $V(1)(U(1))^\dagger U(2)$ is small, a condition for $H_1$ and $H_2$ to be well conditioned is that $V(2)(I - \Lambda(2))$ and $V(2)$ are well conditioned on null$(U(2))$. Considering all these
conditions together, we see that again $\alpha$ is small when $V^{(1)}$ and $V^{(2)}\Lambda^{(2)}$ are small in norm and $X_L$ and $X_U$ are well conditioned. We note that additional special cases might also result in well conditioned matrices $X_L$ and $X_U$ and that Corollary 4 may be useful for checking these.

Of course, neither $X_U$ and $X_L$, nor their condition numbers, are uniquely defined (see, for example, the discussion in Bai, Benzi and Chen [2, Remark 3.1]). In our experiment we consider the eigenvector matrices for which each eigenvector has unit norm. To this end let us fix the columns $[u_i^T, v_i^T]^T$, $i = 1, \ldots, n + m$, of $X_U$ in (2.7) to have unit length, so that $\|u_i\|_2^2 + \|v_i\|_2^2 = 1$. Then, the column scaling that transforms $X_L$ to $\hat{X}_L$, an eigenvector matrix with unit-length columns, is given by the diagonal matrix $D = \text{diag}(d_{11}, \ldots, d_{n+m,n+m})$, where

$$d_{ii}^{-2} = \begin{cases} 1 - \|v_i\|^2, & i \leq p, \\ 1 + (|\lambda_i|^2 - 2\Re(\lambda_i))\|v_i\|^2, & i > p. \end{cases} \quad (2.11)$$

3. Numerical example. In our experience with incompressible Stokes and Navier-Stokes examples and preconditioners in IFISS [9, 19] and with time-harmonic Maxwell equations [11] we find, similarly to [4, 7], that there is often little difference between iteration counts achieved with $P_L^{-1}A$ and $P_U^{-1}A$. This seems to be true regardless of whether left or right preconditioning is used. However, there are certainly examples for which the condition numbers differ as we now show.

Our linear system comes from an incompressible Stokes problem that describes a flow over a backward facing step in two dimensions and is described in detail in Elman et al. [10, Example 5.1.2]. The equations are discretized by $Q_2 - Q_1$ finite elements in Matlab using IFISS with default parameters. We apply left-preconditioned GMRES with a zero initial guess and terminate when the preconditioned residual decreases by eight orders of magnitude—although it may be desirable to consider the unpreconditioned residual, the preconditioned residual is more closely connected with the theory of Section 2. In both of our preconditioners $P_S$ is the diagonal of the pressure mass matrix. The first choice for $P_A$ is a no-fill incomplete Cholesky (IC) factorization (produced by the Matlab command \texttt{ichol}) of $A_0$, the vector Laplace matrix obtained with natural boundary conditions. The second is the algebraic multigrid (AMG) preconditioner implemented in IFISS. The eigenvector matrix $X_U$ is computed by the Matlab function \texttt{eig} while $X_L$ is computed from $X_U$ using Corollary 3. Both $X_L$ and $X_U$ are scaled to have unit-length vectors.

We first consider the incomplete Cholesky preconditioner. The eigenvalues of $P_L^{-1}A$ (and $P_U^{-1}A$) lie in $[-2.3, -0.046] \cup [0.19, 1.2]$. Additionally, there are 20 eigenvalues within $10^{-14}$ of 1 that we assume are unit eigenvalues. The iterations for different mesh sizes are given in Table 3.1, from which we see that the iteration counts are consistently lower for $P_L$ than $P_U$. For the problem on the $8 \times 24$ grid this

<table>
<thead>
<tr>
<th>Grid</th>
<th>IC</th>
<th>AMG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4 \times 12$</td>
<td>132</td>
<td>135</td>
</tr>
<tr>
<td>$8 \times 24$</td>
<td>194</td>
<td>202</td>
</tr>
<tr>
<td>$16 \times 48$</td>
<td>313</td>
<td>323</td>
</tr>
<tr>
<td>$32 \times 96$</td>
<td>553</td>
<td>590</td>
</tr>
</tbody>
</table>
We note that there are other examples for which (2.8) is tight—these are typically estimating $\kappa$ condition numbers increases, the bound (2.8) is usually not as tight, with (2.10) overmain causes of the difference between $\|\cdot\|_2$ and $\kappa$. We have shown that the eigenvectors of $P_L$ and $P_U$ are uniquely defined, the conditioning could be related to the particular choice of matrix.

In summary, when the eigenvalues are important, or when the eigenvector matrices $X_L$ and $X_U$ are fairly well conditioned, there is no benefit in choosing $P_L$ over $P_U$ or vice versa. (Note that since the eigenvector matrix is not uniquely defined, the conditioning could be related to the particular choice of matrix.) We have shown that the eigenvectors of $P_L^{-1}A$ and $P_U^{-1}A$ can differ significantly, is reflected in the condition numbers of $X_L$ and $X_U$ (see Table 3.2). To further investigate the disparity in these condition numbers, we also list in Table 3.2 quantities related to Corollary 4. Since $\|X_L\|_2 = 2.9$ and $\|X_U\|_2 = 5.7$, the perturbations $V^{(1)}$ and $V^{(2)}A^{(2)}$ are relatively large in norm. Moreover, $\|\{(U^{(1)})\}^T\|_2 = 9.1$, so that $U_1$ is not so well conditioned, and as a consequence $\|V^{(1)}(U^{(1)})\|_2$ is large. Relative to $\|V^{(1)}(U^{(1)})\|_2$, the remaining terms in $\alpha$ are reasonably small and we conclude that the most significant contributions to the difference between $\kappa_2(X_L)$ and $\kappa_2(X_U)$ are $U^{(1)}$ and the perturbations $V^{(1)}$ and $V^{(2)}A^{(2)}$. We note that although the bounds in Corollary 4 are not quantitatively descriptive of this problem they give insight into why the condition numbers of $X_L$ and $X_U$ differ.

The eigenvalues of the AMG-preconditioned matrix lie in $[-2.5, -0.01] \cup [0.009, 58]$ and no eigenvalue is within $10^{-14}$ of 1. The iteration counts appear to be mesh-independent and are much lower than for the incomplete Cholesky preconditioner in spite of the wider distribution of eigenvalues (see Table 3.1). Additionally, $P_L^{-1}A$ and $P_U^{-1}A$ give similar iteration counts, with the latter performing slightly better for larger problems. Thus, the eigenvector condition numbers are not necessarily good predictors of convergence for this problem since, at least for the $8 \times 24$ grid, $\kappa_2(X_L) > \kappa_2(X_U)$. Nevertheless, we can investigate why the condition numbers differ by applying Corollary 4. Compared to $\|X_L\|_2 = 4.5$ and $\|X_U\|_2 = 2.9$, $\|V^{(2)}A^{(2)}\|_2$ and $\|\{(U^{(1)})\}^T\|_2 = 20.6$ are large. Consequently, $\|V^{(2)}(U^{(1)})\|_2$ and $\|V^{(2)}(I - A^{(2)})^T(U^{(2)})\|_2$ are large relative to the other terms in $\alpha$. From this we deduce that the norm of $(U^{(1)})^T$ and the size of the perturbation $V^{(2)}A^{(2)}$ are the main causes of the difference between $\kappa_2(X_L)$ and $\kappa_2(X_U)$.

For this problem the condition number bound (2.8) overestimates $\kappa_2(X_U)/\kappa_2(X_L)$. We note that there are other examples for which (2.8) is tight—are these typically problems for which $\kappa_2(X_L)$ and $\kappa_2(X_U)$ are close. When the difference between the condition numbers increases, the bound (2.8) is usually not as tight, with (2.10) overestimating $\kappa_2(X_U)/\kappa_2(X_L)$ and the subsequent bounds on $\|K^{-1}\|_2$ and $\|K^{-1}\|_2$ then causing (2.10) to be overestimated.

### Table 3.2 Condition numbers of eigenvector matrices and norms of quantities in Corollary 4 for the problem on the $8 \times 24$ grid and the incomplete Cholesky preconditioner.

<table>
<thead>
<tr>
<th>$\kappa_2(X_L)$</th>
<th>$\kappa_2(X_U)$</th>
<th>$|V^{(2)}A^{(2)}|_2$</th>
<th>$|V^{(2)}(U^{(1)})|_2$</th>
<th>$|V^{(2)}(I - A^{(2)})^T(U^{(2)})|_2$</th>
<th>$|\cdot|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>53</td>
<td>11.1</td>
<td>4.7</td>
<td>6.6</td>
<td>0.6</td>
</tr>
<tr>
<td>$|G|_2$</td>
<td>$|H|_2$</td>
<td>$|H|_2$</td>
<td>$|\cdot|_2$</td>
<td>$|\cdot|_2$</td>
<td>$|\cdot|_2$</td>
</tr>
<tr>
<td>6.9</td>
<td>3.1</td>
<td>3.1</td>
<td>6.6</td>
<td>$\alpha = 3.8 \times 10^5$</td>
<td>$\alpha = 3.8 \times 10^{17}$</td>
</tr>
</tbody>
</table>

### Table 4.1 Condition numbers of eigenvector matrices and norms of quantities in Corollary 4 for the problem on the $8 \times 24$ grid and the algebraic multigrid preconditioner.

<table>
<thead>
<tr>
<th>$\kappa_2(X_L)$</th>
<th>$\kappa_2(X_U)$</th>
<th>$|V^{(2)}A^{(2)}|_2$</th>
<th>$|V^{(2)}(U^{(1)})|_2$</th>
<th>$|V^{(2)}(I - A^{(2)})^T(U^{(2)})|_2$</th>
<th>$|\cdot|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>205</td>
<td>99</td>
<td>13.9</td>
<td>11.7</td>
<td>14.1</td>
<td>1.6</td>
</tr>
<tr>
<td>$|S|_2$</td>
<td>$\alpha = 1.7 \times 10^1$</td>
<td>$|\cdot|_2$</td>
<td>$|\cdot|_2$</td>
<td>$|\cdot|_2$</td>
<td>$|\cdot|_2$</td>
</tr>
</tbody>
</table>
and this can affect the condition numbers of the eigenvector matrices (when $P_L^{-1}A$ and $P_U^{-1}A$ are diagonalizable). However, the convergence rate of GMRES remains weakly sensitive to these factors for the problems examined, and for others in the literature. When $C = 0$ and $P_L^{-1}A$ and $P_U^{-1}A$ are diagonalizable, we can bound the ratio of the condition numbers of the eigenvector matrices $X_L$ and $X_U$, which depend not only on difference between $X_L$ and $X_U$, contained in $V^{(1)}$ and $V^{(2)}A^{(2)}$, but also on $U^{(1)}$ and $U^{(2)}$.

Acknowledgments. The author would like to thank the editor and referees, who offered many helpful comments and suggestions, and Yvan Notay, whose manuscript on symmetric saddle point matrices inspired this work.

REFERENCES

22. H. A. van der Vorst, Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for