NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. (2011)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/nla.806

On eigenvalue distribution of constraint-preconditioned symmetric saddle point matrices

Luca Bergamaschi*,†

Department of Mathematical Methods and Models for Scientific Applications, University of Padova, Padova, Italy

SUMMARY

This paper is devoted to the analysis of the eigenvalue distribution of two classes of block preconditioners for the generalized saddle point problem. Most of the bounds developed improve those appeared in previously published works. Numerical results onto a realistic test problem give evidence of the effectiveness of the estimates on the spectrum of preconditioned matrices. Copyright © 2011 John Wiley & Sons, Ltd.

Received 7 August 2010; Revised 14 July 2011; Accepted 19 July 2011

KEY WORDS: saddle Point problems; iterative methods; inexact constraint preconditioners; eigenvalues

1. INTRODUCTION

The solution of the (generalized) saddle point linear system of the form

$$\mathcal{A}\boldsymbol{x} = \boldsymbol{b}, \quad \text{where} \quad \mathcal{A} = \begin{bmatrix} A & B^\top \\ B & -C \end{bmatrix} \tag{1}$$

where A is symmetric positive definite (SPD), C is symmetric semi-positive definite, and B, a full-rank rectangular matrix, is encountered in many fields such as constrained optimization, least squares, coupled consolidation problems and Navier–Stokes equations to mention a few (see [3] for a review of such applications). Iterative solution is recommended against direct factorization methods due to the extremely large size of these systems. However, well established iterative methods such as Krylov subspace methods are very slow or even fail to converge if not conveniently preconditioned.

The exact constraint preconditioner (ECP) for Krylov solvers in the solution of Equation (1) is defined as $\mathcal{K}^{-1}$, where

$$\mathcal{K} = \begin{bmatrix} P_A & B^\top \\ B & -C \end{bmatrix} = \begin{bmatrix} I & 0 \\ BP_A^{-1} & I \end{bmatrix} \begin{bmatrix} P_A & 0 \\ 0 & -S \end{bmatrix} \begin{bmatrix} I & P_A^{-1}B^\top \\ 0 & I \end{bmatrix}, \tag{2}$$

where P_A is a suitable approximation to the $(1,1)$ block A. These preconditioners have been studied by a number of authors [3–7]. In most of the aforementioned references, the preconditioner is obtained from $\mathcal{A}$ with the $(1,1)$ block A well approximated and replaced by its diagonal. In the most general case, where A is ill-conditioned, a better approximation is required to ensure convergence. Application of the ECP requires the solution at each iteration of an 'inner' linear system with the negative Schur complement $S = BP_A^{-1}B^\top + C$ as the coefficient matrix that makes a single iteration very costly.

*Correspondence to: Luca Bergamaschi, Dept. Mathematical Methods and Models for Scientific Applications, University of Padova, via Trieste 63, 35121 Padova, Italy.

†E-mail: berga@dmsa.unipd.it

To overcome this problem, a class of Inexact Constraint Preconditioners (ICP) [1, 7, 8] have been developed that compute another approximation P_S to the Schur complement matrix with the aim of reducing the cost of the Schur complement inversion. P_A^{-1} and P_S^{-1} can be regarded as multiplicative preconditioners for matrices A and S, respectively. Because, in our approach, the Schur complement matrix has to be explicitly computed, the inverse of P_A has to be known at least in factorized form. This is not the only option, as in many cases the Schur complement is approximated rather than computed explicitly. We mention the Stokes problem, where the pressure mass matrix represents a very good quality approximation of the product $BA^{-1}B^\top$.

Once P_A and P_S are available, we can define, among the others, two ICPs $\mathcal{M}_1^{-1}$ and $\mathcal{M}_2^{-1}$, where now

$$\mathcal{M}_1 = \begin{bmatrix} I & 0 \\ BP_A^{-1} & I \end{bmatrix} \begin{bmatrix} P_A & 0 \\ 0 & -P_S \end{bmatrix} \begin{bmatrix} I & P_A^{-1}B^\top \\ 0 & I \end{bmatrix} = \begin{bmatrix} P_A & B^\top \\ B & BP_A^{-1}B^\top - P_S \end{bmatrix} \tag{3}$$

defining the **Full Inexact Constraint Preconditioner (FICP) and**

$$\mathcal{M}_2 = \begin{bmatrix} P_A & 0 \\ 0 & -P_S \end{bmatrix} \begin{bmatrix} I & P_A^{-1}B^\top \\ 0 & I \end{bmatrix} = \begin{bmatrix} P_A & B^\top \\ 0 & -P_S \end{bmatrix} \tag{4}$$

defining the **Triangular Inexact Constraint Preconditioner (TICP)**.

The FICP has been employed, for example, in [9], where the approximation to A^{-1} is provided by the approximate inverse preconditioner (AINV) [10], and S is preconditioned by its incomplete Cholesky factorization with no fill-in. Block triangular preconditioners have been used in the solution of the Stokes problem [11–13]; their theoretical properties have been studied in [2].

Spectral properties of block preconditioners for saddle point problems have been investigated by many authors, (see for instance [14] and [15]). In these two papers, and in many others, preconditioners are considered that maintain the problem symmetric, and hence, making it solvable by, for example, the MINRES method. On the contrary, no block diagonal preconditioner has been able to successfully accelerate Krylov subspace methods in the solution of our realistic consolidation problem [16]. Recently, nonsymmetric saddle point matrices have attracted some interest—in [17], some conditions are developed for these matrices to be diagonalizable and to have a real and positive spectrum.

The aim of this paper is to give a detailed spectral analysis of FICP and TICP, showing that the convergence properties of these block preconditioners for iterative methods are strictly connected with the approximation introduced by P_A and P_S. Most of the bounds developed improves those of [1] and [2]. It is found that the eigenvalues, most of which complex ones, are well clustered around unity, provided that eigenvalues of $P_A^{-1}A$ and $P_S^{-1}S$ are so. On the other hand, eigenvalue information alone may fail to predict the convergence behavior of Krylov subspace methods. A well-known upper bound for the residual norm of a minimum residual iteration such as GMRES involves the condition number of the eigenvector matrix V. Although we do not have theoretical estimates for the condition number of V we experimentally noticed that its value was indeed modest.

The bounds on real and complex eigenvalues are verified on a test case obtained by a Finite Element discretization of a Coupled Consolidation Problem [16]. It is found that convergence of iterative solvers is driven mostly by the ratio between the largest and the smallest *real* eigenvalue of the preconditioned matrix.

The paper is organized as follows. In Section 2, we characterize the spectral distribution of a class of matrices strictly related to the preconditioned saddle point matrices, which are scaled in Section 3 to make the spectral analysis easier. In Section 4, we perform a complete analysis of eigenvalues of $\mathcal{M}_1^{-1}\mathcal{A}$ (FICP), whereas in Section 5, bounds on eigenvalues of $\mathcal{M}_2^{-1}\mathcal{A}$ (TICP) are provided. In Section 6 we report a brief description of the test case together with numerical results that accounts for the effectiveness of the bounds obtained in the previous sections. The conclusions are drawn in Section 7.

DOI: 10.1002/nla

2. EIGENVALUES OF M_-

Given an SPD $n \times n$ matrix A, a rectangular $m \times n$ matrix B, with $m < n$, and an SPD $m \times m$ matrix C, we are interested in the eigenvalues of

$$\mathcal{M}_- \boldsymbol{u} = \lambda \boldsymbol{u} \quad \text{where} \quad M_- = \begin{bmatrix} A & B^\top \\ -B & C \end{bmatrix} \tag{5}$$

or, exploiting the blocks matrices,

$$\begin{aligned} A\boldsymbol{u}_1 + B^\top \boldsymbol{u}_2 &= \lambda \boldsymbol{u}_1 \\ -B\boldsymbol{u}_1 + C\boldsymbol{u}_2 &= \lambda \boldsymbol{u}_2. \end{aligned} \tag{6}$$

We will use the following notation regarding the eigenvalues of SPD matrices A, $BB^\top$, and C:

$$\begin{aligned} \alpha_A = \lambda_{\min}(A), &\qquad \beta_A = \lambda_{\max}(A), \\ 0 \leqslant \alpha_S = \lambda_{\min}(BB^\top), &\quad \beta_S = \lambda_{\max}(BB^\top), \\ \alpha_C = \lambda_{\min}(C), &\qquad \beta_C = \lambda_{\max}(C). \end{aligned}$$

It is well known (see for instance Proposition 2.2 in [1]) that matrix M_- has at most $n - m$ eigenvalues satisfying

$$\alpha_A \leqslant \lambda \leqslant \beta_A.$$

with eigenvectors $\boldsymbol{u} = \left(\boldsymbol{u}_1^\top, 0\right)^\top$, and $B\boldsymbol{u}_1 = 0$.

The following Theorem will provide a further bound on the additional real eigenvalues of Equation (5). Note that the lower bound improves that of [1]. We define, for some $\boldsymbol{s}$, $\boldsymbol{u}_2 \neq 0$,

$$\eta_A = \frac{\boldsymbol{s}^\top A \boldsymbol{s}}{\boldsymbol{s}^\top \boldsymbol{s}} \in [\alpha_A, \beta_A], \quad \eta_C = \frac{\boldsymbol{u}_2^\top C \boldsymbol{u}_2}{\boldsymbol{u}_2^\top \boldsymbol{u}_2} \in [\alpha_C, \beta_C], \quad \eta_S = \frac{\boldsymbol{u}_2^\top BB^\top \boldsymbol{u}_2}{\boldsymbol{u}_2^\top \boldsymbol{u}_2} \in [\alpha_S, \beta_S]. \tag{7}$$

The proof of Theorem 1 is based on the following result.

Lemma 1
Let $\lambda \notin [\alpha_A, \beta_A]$. Then, for every $z \neq 0$, there exists a vector $\boldsymbol{s} \neq 0$ such that

$$\frac{z^\top (A - \lambda I)^{-1} z}{z^\top z} = \left(\frac{\boldsymbol{s}^\top A \boldsymbol{s}}{\boldsymbol{s}^\top \boldsymbol{s}} - \lambda \right)^{-1} = (\eta_A - \lambda)^{-1}.$$

Proof
$A - \lambda I$ is SPD or minus SPD. If $\lambda < \alpha_A$ it can be diagonalized as $A - \lambda I = Q(\Lambda - \lambda I)Q^\top$, where Λ is the diagonal matrix with the positive eigenvalues of A. Therefore, setting $\boldsymbol{p} = (\Lambda - \lambda I)^{-1/2} Q^\top z)$, we have

$$\frac{z^\top (A - \lambda I)^{-1} z}{z^\top z} = \frac{\boldsymbol{p}^\top \boldsymbol{p}}{\boldsymbol{p}^\top (\Lambda - \lambda I) \boldsymbol{p}} = \left(\frac{\boldsymbol{p}^\top \Lambda \boldsymbol{p}}{\boldsymbol{p}^\top \boldsymbol{p}} - \lambda \right)^{-1} = (\eta_A - \lambda)^{-1},$$

where $\boldsymbol{s} = Q\,\boldsymbol{p}$. The case $\lambda > \beta_A$ is analogous and leads to the same result. □

Theorem 1
The real eigenvalues of Equation (5) not lying in $[\alpha_A, \beta_A]$ satisfy

$$\alpha_C + \frac{\alpha_S}{\beta_A} \leqslant \eta_C + \frac{\eta_S}{\eta_A} \leqslant \lambda \leqslant \eta_C \leqslant \beta_C.$$

Proof
Let $\lambda \in \mathbb{R}^\top$ with $\lambda \notin [\alpha_A, \beta_A]$ and let $\boldsymbol{u}$ such that $B\boldsymbol{u}_1 \neq 0$ and $B^\top \boldsymbol{u}_2 \neq 0$. Because $A - \lambda I$ is invertible, we can compute $\boldsymbol{u}_1$ from the first of Equation (6)

$$\boldsymbol{u}_1 = -(A - \lambda I)^{-1} B^\top \boldsymbol{u}_2.$$

DOI: 10.1002/nla

Substituting in the second one, we obtain

$$B\,(A-\lambda I)^{-1}\,B^{\top}\boldsymbol{u}_2 + C\,\boldsymbol{u}_2 - \lambda\boldsymbol{u}_2 = 0.$$

Now, multiplying by $\frac{\boldsymbol{u}_2^{\top}}{\boldsymbol{u}_2^{\top}\boldsymbol{u}_2}$ yields

$$\frac{\boldsymbol{u}_2^{\top}B\,(A-\lambda I)^{-1}\,B^{\top}\boldsymbol{u}_2}{\boldsymbol{u}_2^{\top}\boldsymbol{u}_2} + \eta_C - \lambda = 0. \tag{8}$$

Observe that, defining $z = B^{\top}\boldsymbol{u}_2$, and applying Lemma 1,

$$\frac{\boldsymbol{u}_2^{\top}B\,(A-\lambda I)^{-1}\,B^{\top}\boldsymbol{u}_2}{\boldsymbol{u}_2^{\top}\boldsymbol{u}_2} = \frac{z^{\top}\,(A-\lambda I)^{-1}\,z}{z^{\top}z}\,\frac{\boldsymbol{u}_2^{\top}BB^{\top}\boldsymbol{u}_2}{\boldsymbol{u}_2^{\top}\boldsymbol{u}_2} = (\eta_A-\lambda)^{-1}\,\eta_S.$$

We are now able to rewrite Equation (8) as $(\eta_A-\lambda)^{-1}\eta_S + \eta_C - \lambda = 0$, or equivalently

$$\lambda^2 - (\eta_C+\eta_A)\lambda + \eta_S + \eta_A\eta_C = 0. \tag{9}$$

The larger solution of Equation (9) is

$$\begin{aligned}\lambda_2 &= \frac{\eta_A+\eta_C+\sqrt{(\eta_A+\eta_C)^2-4(\eta_A\eta_C+\eta_S)}}{2}\\ &= \frac{\eta_A+\eta_C+\sqrt{(\eta_A-\eta_C)^2-4\eta_S}}{2} \leqslant \max\{\eta_A,\eta_C\} = \eta_C \leqslant \beta_C.\end{aligned} \tag{10}$$

To bound the smallest eigenvalue, we consider the smaller solution of Equation (9).

$$\begin{aligned}\lambda_1 &= \frac{\eta_A+\eta_C-\sqrt{(\eta_A+\eta_C)^2-4(\eta_A\eta_C+\eta_S)}}{2}\\ &= \frac{2(\eta_A\eta_C+\eta_S)}{\eta_A+\eta_C+\sqrt{(\eta_A-\eta_C)^2-4\eta_S)}}\\ &\geqslant \frac{2(\eta_A\eta_C+\eta_S)}{2\max\{\eta_A,\eta_C\}} = \eta_C + \frac{\eta_S}{\eta_A}.\end{aligned} \tag{11}$$

The last equation follows from the inequality $\eta_C < \eta_A$ (otherwise we would have $\lambda_1 > \eta_A > \alpha_A$ against the assumption). Hence,

$$\lambda_1 \geqslant \eta_C + \frac{\eta_S}{\eta_A} \geqslant \alpha_C + \frac{\alpha_S}{\beta_A}. \tag{12}$$

□

We have then proved the following result

Corollary 1
The real eigenvalues of Equation (5) satisfy

$$\min\left\{\alpha_A, \alpha_C + \frac{\alpha_S}{\beta_A}\right\} \leqslant \lambda \leqslant \max\{\beta_A,\beta_C\}.$$

Remark 1
This result improves that of Proposition 2.12 in [1], which provides a lower bound for $\lambda : \lambda \geqslant \min\{\alpha_A,\alpha_C\}$.

In the sequel, we will denote any complex eigenvalue as

$$\lambda = \lambda_R + i\,\lambda_I.$$

DOI: 10.1002/nla

Proposition 1
The complex eigenvalues of Equation (5) satisfy

$$\frac{\alpha_A + \alpha_C}{2} \leqslant \lambda_R \leqslant \frac{\beta_A + \beta_C}{2}, \quad |\lambda_I| \leqslant \sqrt{\beta_S}.$$

Proof
It is part of the proof of Proposition 2.12 in [1]. □

The bound on imaginary part is not sharp and can be improved. In fact, if $\gamma \notin [\alpha_A, \beta_A]$ and $C = \gamma I$, a bound for λ_I can be devised in terms of γ.

Theorem 2
The imaginary part of the complex eigenvalues of

$$M_- = \begin{bmatrix} A & B^\top \\ -B & \gamma I \end{bmatrix} \tag{13}$$

satisfy

$$\begin{cases} |\lambda_I| \leqslant \sqrt{\beta_S - \dfrac{(\gamma - \alpha_A)^2}{4}} & \text{if} \quad 0 \leqslant \gamma < \alpha_A \\ |\lambda_I| \leqslant \sqrt{\beta_S - \dfrac{(\gamma - \beta_A)^2}{4}} & \text{if} \quad \gamma > \beta_A \end{cases}.$$

Proof
From

$$\begin{aligned} A\boldsymbol{u}_1 + B^\top \boldsymbol{u}_2 &= \lambda \boldsymbol{u}_1 \\ -B\boldsymbol{u}_1 + \gamma I \boldsymbol{u}_2 &= \lambda \boldsymbol{u}_2 \end{aligned} \tag{14}$$

we have

$$\boldsymbol{u}_2 = \frac{1}{\gamma - \lambda} B\boldsymbol{u}_1. \tag{15}$$

Substituting Equation (15) in the first of Equation (14) we obtain.

$$A\boldsymbol{u}_1 - \lambda \boldsymbol{u}_1 + \frac{1}{\gamma - \lambda} B^\top B\boldsymbol{u}_1 = 0.$$

Then premultiplying by $\frac{\boldsymbol{u}_1^*}{\boldsymbol{u}_1^* \boldsymbol{u}_1}$ and setting $\eta_A = \frac{\boldsymbol{u}_1^* A \boldsymbol{u}_1}{\boldsymbol{u}_1^* \boldsymbol{u}_1}$ and $\eta'_S = \frac{\boldsymbol{u}_1^* B^\top B \boldsymbol{u}_1}{\boldsymbol{u}_1^* \boldsymbol{u}_1}$ (now $\eta'_S \in (0, \beta_S]$), we obtain

$$\lambda^2 - (\gamma + \eta_A)\lambda + \eta'_S + \eta_A \gamma = 0,$$

which gives for $\gamma < \eta_A + 2\sqrt{\eta'_S}$

$$\lambda_I = \sqrt{\frac{\eta'_S - (\gamma - \eta_A)^2}{4}}.$$

□

Remark 2
The outcome of the previous theorem points out that if $\gamma \geqslant \gamma_0 = \beta_A + 2\sqrt{\beta_S}$ no complex eigenvalue occur. The bound of Proposition (1) is poor for $\gamma \approx 0$ or $\gamma \approx \gamma_0$, whereas it is sharp if $\gamma = \eta_A$.

DOI: 10.1002/nla

Corollary 2
The real eigenvalues of the matrix

$$M_-^0 = \begin{bmatrix} A & B^\top \\ -B & 0 \end{bmatrix} \tag{16}$$

satisfy

$$\min\left\{\alpha_A, \frac{\alpha_S}{\beta_A}\right\} \leqslant \lambda \leqslant \beta_A$$

Proof
Immediately following from Theorem 1 by setting $\alpha_C = 0$. □

Remark 3
The lower bound for M_-^0 can be further refined by using Equation (11) in the proof of Theorem 1.

$$\lambda_1 = \frac{2\eta_S}{\eta_A + \sqrt{\eta_A^2 - 4\eta_S}} \geqslant \frac{2\alpha_S}{\beta_A + \sqrt{\beta_A^2 - 4\alpha_S}}.$$

Proposition 2
The complex eigenvalues of Equation (16) satisfy

$$\frac{\alpha_A}{2} \leqslant \lambda_R \leqslant \frac{\beta_A}{2} \qquad |\lambda_I| \leqslant \sqrt{\beta_S - \frac{\alpha_A^2}{4}}.$$

Proof
Following from Theorem 2 by setting $\gamma = 0$. □

To show the behavior of these bounds, let us consider the following:

Example 1

$$M_- = \begin{pmatrix} \beta_A & 0 & 1 \\ 0 & \alpha_A & 1 \\ -1 & -1 & 0 \end{pmatrix}, \qquad \alpha_S = \beta_S = 2.$$

If $\beta_A = 3$, $\alpha_A = 2.9$ the eigenvalues of M_- are $\sigma(M_-) = \{1.0576, 1.8887, 2.9537\}$. The lower bound given by Remark 3 is $\lambda \geqslant \frac{4}{3+1} = 1$. If, instead, we choose $\beta_A = 3$, and $\alpha_A = 2.6$, the spectrum of M_- is $\sigma(M_-) = \{2.8530, 1.3735 + 0.2764i, 1.3735 - 0.2764i\}$. In this case, we may compare the bounds given by Proposition 1 and 2, respectively. In the first case we have $|\lambda_I| \leqslant \sqrt{2} \approx 1.4142$ while in the second $|\lambda_I| \leqslant \sqrt{2 - 6.76/4} \approx 0.5568$.

3. PRECONDITIONING

The results of the previous section will be used to investigate the spectral properties of two classes of preconditioners for the solution of the linear system (1). Let P_A and P_S be SPD approximations of A and $S = C + BP_A^{-1}B^\top$, respectively. P_A^{-1} and P_S^{-1} can also be viewed as preconditioners for the corresponding matrices, so that we can define the following SPD preconditioned matrices:

$$A_P = P_A^{-1/2} A P_A^{-1/2} \quad \text{and} \quad S_P = P_S^{-1/2} S P_S^{-1/2}.$$

DOI: 10.1002/nla

Let us assume that

$$\begin{aligned} 0 < \alpha_A = \lambda_{\min}(A_P) < 1 < \lambda_{\max}(A_P) = \beta_A, \\ 0 < \alpha_S = \lambda_{\min}(S_P) < 1 < \lambda_{\max}(S_P) = \beta_S, \\ 0 \leqslant \alpha_C = \lambda_{\min}\left(\widehat{C}\right) < \lambda_{\max}\left(\widehat{C}\right) = \beta_C, \end{aligned} \tag{17}$$

where $\widehat{C} = P_S^{-1/2} C P_S^{-1/2}$. The conditions $1 \in [\alpha_A, \beta_A]$ and $1 \in [\alpha_S, \beta_S]$ are very often fulfilled in practice because preconditioners P_A and P_S are expected to cluster eigenvalues around unit.

To characterize the eigenvalues of the preconditioned matrices $\mathcal{M}_1^{-1}\mathcal{A}$ and $\mathcal{M}_2^{-1}\mathcal{A}$ where $\mathcal{M}_1$ and $\mathcal{M}_2$ are defined in Equations (3) and (4), respectively, it is useful to define a matrix $\mathcal{P}$ as

$$\mathcal{P} = \begin{bmatrix} P_A^{-1/2} & 0 \\ 0 & P_S^{-1/2} \end{bmatrix}. \tag{18}$$

The problem of finding the eigenvalues of $\mathcal{M}_1^{-1}\mathcal{A}$ and $\mathcal{M}_2^{-1}\mathcal{A}$ is, therefore, equivalent to solve $\mathcal{P}\mathcal{A}\mathcal{P}\boldsymbol{v} = \lambda\mathcal{P}\mathcal{M}_1\mathcal{P}\boldsymbol{v}$, and $\mathcal{P}\mathcal{A}\mathcal{P}\boldsymbol{v} = \lambda\mathcal{P}\mathcal{M}_2\mathcal{P}\boldsymbol{v}$. Exploiting the blocks

$$\text{FICP}: \mathcal{P}\mathcal{A}\mathcal{P}\boldsymbol{v} = \lambda\mathcal{P}\mathcal{M}_1\mathcal{P}\boldsymbol{v} \longrightarrow \begin{bmatrix} A_P & R^\top \\ R & -\widehat{C} \end{bmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} = \lambda \begin{bmatrix} I & R^\top \\ R & RR^\top - I \end{bmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix}, \tag{19}$$

where $R = P_S^{-1/2} B P_A^{-1/2}$. Note that $RR^\top = S_P - \widehat{C}$.

$$\text{TICP}: \mathcal{P}\mathcal{A}\mathcal{P}\boldsymbol{v} = \lambda\mathcal{P}\mathcal{M}_2\mathcal{P}\boldsymbol{v} \longrightarrow \begin{bmatrix} A_P & R^\top \\ R & -\widehat{C} \end{bmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} = \lambda \begin{bmatrix} I & R^\top \\ 0 & -I \end{bmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix}. \tag{20}$$

The eigenvalues of the matrix $A_R = (RR^T)^{-1} R A_P R^T$ will also be important in the spectral analysis of the preconditioned matrices. First, they are all real positive as A_R is a product of two SPD matrices. Second, it is easy to show that $\left[\alpha_A^R, \beta_A^R\right] \subset [\alpha_A, \beta_A]$, where $\alpha_A^R = \lambda_{\min}(A_R)$, and $\beta_A^R = \lambda_{\max}(A_R)$ and that the eigenvalues of A_R do not depend on P_S. Finally, extremal eigenvalues of A_R can be computed by observing that they are the same as those of the generalized eigenproblem $BP_A^{-1}AB^\top\boldsymbol{u} = \lambda BB^\top\boldsymbol{u}$.

The following two sections are devoted to the spectral analysis of the preconditioned matrix in the two cases in terms of eigenvalues of $A_P(A_R)$ and S_P. Not all the obtained results are completely new. Some of them can be also found in [1].

4. SPECTRAL ANALYSIS OF FICP

This analysis will follow that of [1] to apply the new results developed in Section 2.

The inverse of the left hand side in Equation (19) can be written as

$$\mathcal{M}_1^{-1} = \begin{bmatrix} I & -R^\top \\ 0 & I \end{bmatrix}\begin{bmatrix} I & 0 \\ R & -I \end{bmatrix} = \mathcal{U}\mathcal{L}$$

so that the eigenvalues of Equation (19) are the same as those of $\mathcal{L}\mathcal{A}\mathcal{U}\boldsymbol{u} = \lambda\boldsymbol{u}$ which reads:

$$\begin{pmatrix} A_P & (I - A_P)R^\top \\ -R(I - A_P) & R(2I - A_P)R^\top + \widehat{C} \end{pmatrix}\begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{pmatrix} = \lambda \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{pmatrix}. \tag{21}$$

If $\beta_A^R < 2$, the aforementioned matrix satisfies the hypotheses of Theorem 1.

Throughout this section we will use the following **notation**:

$$\theta_S = \frac{\boldsymbol{u}_2^\top S_P \boldsymbol{u}_2}{\boldsymbol{u}_2^\top \boldsymbol{u}_2},\ \theta_A^R = \frac{\boldsymbol{u}_2^\top R A_P R^\top \boldsymbol{u}_2}{\boldsymbol{u}_2^\top RR^\top \boldsymbol{u}_2},\ \theta_A = \frac{\boldsymbol{s}^\top A_P \boldsymbol{s}}{\boldsymbol{s}^\top \boldsymbol{s}},\ \theta_C = \frac{\boldsymbol{u}_2^\top \widehat{C}\boldsymbol{u}_2}{\boldsymbol{u}_2^\top \boldsymbol{u}_2},$$

for some $\boldsymbol{s}, \boldsymbol{u}_2 \neq 0$. It follows that $\theta_A^R \in \left[\alpha_A^R, \beta_A^R\right]$ and $\frac{\boldsymbol{u}_2^\top RR^\top \boldsymbol{u}_2}{\boldsymbol{u}_2^\top \boldsymbol{u}_2} = \theta_S - \theta_C\,(> 0)$.

DOI: 10.1002/nla

Theorem 3 provides bounds on the eigenvalues of Equation (21) using the results of the following

Lemma 2
Let $H = R(2I - A_P)R^\top + \widehat{C}$, $P = R(I - A_P)^2 R^\top$. If $\beta_A^R < 2$, then

$$\lambda(H) \in \left[\alpha_S\left(2-\beta_A^R\right) + \alpha_C\left(\beta_A^R - 1\right), \beta_S\left(2-\alpha_A^R\right) - \alpha_C\left(1-\alpha_A^R\right)\right]$$
$$\lambda(P) \leqslant (\beta_S - \alpha_C)\max\left\{\left(1-\alpha_A^R\right)^2, \left(\beta_A^R - 1\right)^2\right\}.$$

Proof
$\lambda(H) \in [\min q(\boldsymbol{u}_2, H), \max q(\boldsymbol{u}_2, H)]$, where

$$q(\boldsymbol{u}_2, H) = \frac{\boldsymbol{u}_2^\top\left(R(2I - A_P)R^\top + \widehat{C}\right)\boldsymbol{u}_2}{\boldsymbol{u}_2^\top\boldsymbol{u}_2} = (\theta_S - \theta_C)\left(2 - \theta_A^R\right) + \theta_C.$$

Because the function on the right hand side is decreasing in θ_A, then

$$\min q(\boldsymbol{u}_2, H) \geqslant (\theta_S - \theta_C)\left(2-\beta_A^R\right) + \theta_C \geqslant \alpha_S\left(2-\beta_A^R\right) + \alpha_C\left(\beta_A^R - 1\right)$$
$$\max q(\boldsymbol{u}_2, H) \leqslant (\theta_S - \theta_C)\left(2-\alpha_A^R\right) + \theta_C \leqslant \beta_S\left(2-\alpha_A^R\right) - \alpha_C\left(1-\alpha_A^R\right).$$

The bound for $\lambda(P)$ follows from

$$\lambda(P) = q(\boldsymbol{u}_2, P) = \frac{\boldsymbol{u}_2^\top\left(R(I - A_P)^2 R^\top\right)\boldsymbol{u}_2}{\boldsymbol{u}_2^\top\boldsymbol{u}_2} = (\theta_S - \theta_C)\left(1 - \theta_A^R\right)^2.$$

□

Theorem 3
Let $\beta_A^R < 2$. The real eigenvalues of Equation (21) satisfy

$$\min\left\{\alpha_A, \frac{\alpha_S}{\beta_A} + \frac{\alpha_C(\beta_A - 1)}{\beta_A}\right\} \leqslant \lambda \leqslant \max\left\{\beta_A, \left(2-\alpha_A^R\right)\beta_S - \alpha_C\left(1-\alpha_A^R\right)\right\}. \tag{22}$$

If $\lambda_I \neq 0$, then

$$\frac{\alpha_A + \alpha_S\left(2-\beta_A^R\right) + \alpha_C\left(\beta_A^R - 1\right)}{2} \leqslant \lambda_R \leqslant \frac{\beta_A + \beta_S\left(2-\alpha_A^R\right) + \alpha_C\left(1-\alpha_A^R\right)}{2} \tag{23}$$
$$|\lambda_I| \leqslant \sqrt{\beta_S - \alpha_C}\max\left\{1-\alpha_A^R, \beta_A^R - 1\right\}. \tag{24}$$

Proof
We will first give bound for the real eigenvalues. The largest one satisfies, according to Corollary 1 and the Lemma 2

$$\lambda < \max\{\beta_A, \lambda_{max}(H)\} \leqslant \max\left\{\beta_A, \beta_S\left(2-\alpha_A^R\right) - \alpha_C\left(1-\alpha_A^R\right)\right\}.$$

To derive a lower bound for real eigenvalues we first observe that eigenvectors like $\left(\boldsymbol{u}_1^\top, 0\right)^\top$ produce eigenvalues satisfying $\lambda \geqslant \alpha_A$; moreover, off-diagonal blocks of Equation (21) do not have maximum rank because A_P may possess the unit eigenvalue, therefore the matrix in Equation (21) allows eigenvectors of the form $\left(0, \boldsymbol{u}_2^\top\right)^\top$. These eigenvectors satisfy $(I - A_P)R^\top\boldsymbol{u}_2 = 0$ and hence, from the second of Equation (21) $\left(RR^\top + \widehat{C}\right)\boldsymbol{u}_2 = \lambda\boldsymbol{u}_2$, which implies $\lambda \in [\alpha_S, \beta_S]$.

To give bound for the remaining eigenvalues, we write η_A, η_S, η_C of Equation (7) as

$$\eta_A = \theta_A,\ \eta_S = (\theta_S - \theta_C)\left(1 - \theta_A^R\right)^2,\ \eta_C = (\theta_S - \theta_C)\left(2 - \theta_A^R\right) + \theta_C.$$

DOI: 10.1002/nla

Then using Theorem 1 and observing that the function $f(t) = 2 - t + \frac{(1-t)^2}{\theta_A}$, with $t \in [0, \theta_A]$ is decreasing and hence satisfies $f(t) \geqslant f(\theta_A) = \frac{1}{\theta_A}$,

$$\lambda \geqslant \eta_S + \frac{\eta_C}{\eta_A} = (\theta_S - \theta_C)\left(2 - \theta_A^R + \frac{\left(1-\theta_A^R\right)^2}{\theta_A}\right) + \theta_C$$
$$\geqslant \frac{\theta_S - \theta_C}{\theta_A} + \theta_C \geqslant \frac{\theta_S - \theta_C}{\beta_A} + \theta_C = \frac{\theta_S}{\beta_A} + \theta_C \frac{\beta_A - 1}{\beta_A} \geqslant \frac{\alpha_S}{\beta_A} + \alpha_C \frac{\beta_A - 1}{\beta_A}.$$

The thesis holds by finally observing that $\frac{\alpha_S}{\beta_A} + \alpha_C \frac{\beta_A - 1}{\beta_A} \leqslant \alpha_S$. To obtain the bounds for the complex eigenvalues, we refer to Proposition 2.12 in [1] and to Lemma 2. □

If $C \equiv 0$ the bounds of Theorem 3 simplify as stated in the following

Corollary 3
Let $\beta_A^R < 2$. The real eigenvalues of Equation (21) with $\widehat{C} = 0$ satisfy

$$\min\left\{\alpha_A, \frac{\alpha_S}{\beta_A}\right\} \leqslant \lambda \leqslant \max\left\{\beta_A, \left(2 - \alpha_A^R\right)\beta_S\right\}.$$

If $\lambda_I \neq 0$ then

$$\frac{\alpha_A + \alpha_S\left(2 - \beta_A^R\right)}{2} \leqslant \lambda_R \leqslant \frac{\beta_A + \beta_S\left(2 - \alpha_A^R\right)}{2}.$$

Proof
Immediately follows from Theorem 3 by setting $\alpha_C = 0$. □

4.1. Other bounds for complex eigenvalues

The results regarding complex eigenvalues of inexact constraint preconditioners can be further refined. This is the outcome of the following theorem. We first define

$$g(a) = \frac{a-1}{2-a} \text{ and } \delta = \max_{a \in [\alpha_A, \beta_A]} |g(a)| = \max\left\{\frac{1-\alpha_A}{2-\alpha_A}, \frac{\beta_A - 1}{2-\beta_A}\right\}$$

Theorem 4
If $\beta_A < 2$ then the complex eigenvalues of FICP satisfy

$$|\lambda - 1| < 2\delta, \quad \lambda_I \leqslant \delta, \quad \frac{\alpha_A}{2-\alpha_A} < \lambda_R < \frac{\beta_A}{2-\beta_A}.$$

Proof
Multiplying the first equation of Equation (19) by $\frac{\boldsymbol{v}_1^*}{\|\boldsymbol{v}_1\|^2}$ and the second one by $\frac{\boldsymbol{v}_2^*}{\|\boldsymbol{v}_1\|^2}$, we obtain

$$\begin{cases} \theta_A + \phi = \lambda + \lambda\phi \\ \bar{\phi} - \theta_C \rho = \lambda\bar{\phi} + \lambda(\theta_S - 1)\rho, \end{cases} \tag{25}$$

where we set

$$\rho = \frac{\|\boldsymbol{v}_2\|}{\|\boldsymbol{v}_1\|}, \ \theta_A = \frac{\boldsymbol{v}_1^* A_P \boldsymbol{v}_1}{\boldsymbol{v}_1^* \boldsymbol{v}_1}, \theta_S = \frac{\boldsymbol{v}_2^* S_P \boldsymbol{v}_2}{\boldsymbol{v}_2^* \boldsymbol{v}_2}, \theta_C = \frac{\boldsymbol{v}_2^* \widehat{C} \boldsymbol{v}_2}{\boldsymbol{v}_2^* \boldsymbol{v}_2} \text{ and } \phi = \frac{\boldsymbol{v}_1^* R^\top \boldsymbol{v}_2}{\|\boldsymbol{v}_1\|^2}. \tag{26}$$

Now, from the second equation $\phi = \frac{(\bar{\lambda}(\theta_S - 1) + \theta_C)\rho}{1 - \bar{\lambda}}$, and substituting in the first one yields

$$\theta_A - \lambda = (\lambda - 1)\frac{\left(\bar{\lambda}\left(\theta_S - 1\right) + \theta_C\right)\rho}{1 - \bar{\lambda}},$$

DOI: 10.1002/nla

and

$$\theta_A\left(1-\bar{\lambda}\right)-\lambda+|\lambda|^2=\left(|\lambda|^2-\bar{\lambda}\right)(\theta_S-1)\,\rho+(\lambda-1)\theta_C\rho. \tag{27}$$

The imaginary part gives

$$\lambda_I(\theta_A-1)=\lambda_I(\theta_S+\theta_C-1)\rho,$$

which gives for every complex eigenvalue

$$\theta_A-1=(\theta_S+\theta_C-1)\rho. \tag{28}$$

After substituting Equation (28) in Equation (27) and taking the real part, we obtain

$$\theta_A(1-\lambda_R)-\lambda_R+|\lambda|^2=(|\lambda|^2-\lambda_R)(\theta_A-1-\theta_C\rho)+(\lambda_R-1)\theta_C\rho,$$

which can be written as

$$(|\lambda|^2-\lambda_R)(\theta_A-2-\theta_C\rho)=(1-\lambda_R)(\theta_C\rho+\theta_A).$$

After some algebra the last equation can be written as

$$|\lambda-1|^2=(\lambda_R-1)\frac{2(\theta_A-1)}{2-\theta_A+\theta_C\rho}$$

or

$$\lambda_I^2=-(\lambda_R-1)^2+2(\lambda_R-1)g_C \tag{29}$$

with $g_C=\frac{(\theta_A-1)}{2-\theta_A+\theta_C\rho}$.

The largest value of the right-hand-side of Equation (29) is taken for $\lambda_R-1=g_C$ in which $\lambda_I^2=g_c^2<g(\theta_A)^2\leqslant\delta^2$. To bound the real part, let us consider first the case $\theta_A>1$. The right-hand-side of Equation (29) is positive if and only if

$$0<\lambda_R-1<2g_C<2g(\theta_A)\leqslant 2g(\beta_A)\implies\lambda_R<\frac{\beta_A}{2-\beta_A}.$$

In the case $\theta_A<1$, the right-hand-side of Equation (29) is positive if and only if

$$0>\lambda_R-1>2g_C>2g(\theta_A)\geqslant 2g(\alpha_A)\implies\lambda_R>\frac{\alpha_A}{2-\alpha_A}.$$

□

5. SPECTRAL ANALYSIS OF TICP

We now analyze the eigenvalues of Equation (20)

$$\begin{bmatrix} A_P & R^\top \\ R & -\widehat{C} \end{bmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix}=\lambda\begin{bmatrix} I & R^\top \\ 0 & -I \end{bmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix}. \tag{30}$$

If an eigenvector of Equation (30) takes the form $\left(\boldsymbol{v}_1^\top,0\right)^\top$, where $\boldsymbol{v}_1$ is such that $R\boldsymbol{v}_1=0$, then the corresponding eigenvalue λ satisfies

$$\alpha_A\leqslant\lambda\leqslant\beta_A.$$

The following Lemma and subsequent theorem will provide a further bound on the extremal eigenvalues of Equation (30).

DOI: 10.1002/nla

Lemma 3
The real eigenvalues of Equation (30), which do not lie in $[\alpha_A, \beta_A]$ (if any) satisfy the following equation:

$$\lambda^2 - (\theta_S + \theta_A)\,\lambda + \theta_S + (\theta_A - 1)\theta_C = 0 \tag{31}$$

where

$$\theta_S = \frac{\boldsymbol{v}_2^\top S_P \boldsymbol{v}_2}{\boldsymbol{v}_2^\top \boldsymbol{v}_2},\ \ \theta_A = \frac{\boldsymbol{s}^\top A_P \boldsymbol{s}}{\boldsymbol{s}^\top \boldsymbol{s}},\ \ \theta_C = \frac{\boldsymbol{v}_2^\top \widehat{C} \boldsymbol{v}_2}{\boldsymbol{v}_2^\top \boldsymbol{v}_2},\quad \text{for some } \boldsymbol{s}, \boldsymbol{v}_2 \neq 0.$$

Proof
Let $\lambda \in \mathbb{R}$ with $\lambda < \alpha_A$ or $\lambda > \beta_A$. From the first equation in Equation (30),

$$\boldsymbol{v}_1 = (A_P - \lambda I)^{-1}\,(\lambda - 1) R^\top \boldsymbol{v}_2.$$

Substituting in the second one, we obtain

$$R\,(A_P - \lambda I)^{-1}\,(\lambda - 1) R^\top \boldsymbol{v}_2 - \widehat{C}\,\boldsymbol{v}_2 + \lambda \boldsymbol{v}_2 = 0.$$

Now, multiplying by $\frac{\boldsymbol{v}_2^\top}{\boldsymbol{v}_2^\top \boldsymbol{v}_2}$ yields

$$(\lambda - 1)\frac{\boldsymbol{v}_2^\top R\,(A_P - \lambda I)^{-1}\,R^\top \boldsymbol{v}_2}{\boldsymbol{v}_2^\top \boldsymbol{v}_2} + \lambda - \theta_C = 0. \tag{32}$$

Observe that defining $z = R^\top \boldsymbol{v}_2$ and applying Lemma 1 to matrix A_P, we obtain

$$\frac{\boldsymbol{v}_2^\top R\,(A_P - \lambda I)^{-1}\,R^\top \boldsymbol{v}_2}{\boldsymbol{v}_2^\top \boldsymbol{v}_2} = \frac{z^\top\,(A_P - \lambda I)^{-1}\,z}{z^\top z}\,\frac{\boldsymbol{v}_2^\top R R^\top \boldsymbol{v}_2}{\boldsymbol{v}_2^\top \boldsymbol{v}_2} = (\theta_A - \lambda)^{-1}\,(\theta_S - \theta_C),$$

so that Equation (32) can be rewritten as

$$(\lambda - 1)\,(\theta_A - \lambda)^{-1}\,(\theta_S - \theta_C) + \lambda - \theta_C = 0. \tag{33}$$

By simple algebra, from Equation (33), we get the thesis. □

The proof of the next theorem, being very technical, is postponed in Appendix A.

Theorem 5
The eigenvalues of Equation (20) satisfy the following bounds. If $\lambda_I \neq 0$ then

$$|\lambda - 1| \leqslant \sqrt{(1 - \alpha_A)(1 - \alpha_C)} \quad \text{and} \quad \frac{\alpha_A + \alpha_C}{2} \leqslant \lambda_R \leqslant 2 - \frac{\alpha_A + \alpha_C}{2}. \tag{34}$$

The real eigenvalues satisfy

$$\min\left\{\alpha_A, \frac{\alpha_C\,(\beta_A - 1) + \alpha_S}{\beta_A + \alpha_S}\right\} \leqslant \lambda \leqslant \beta_S + \beta_A. \tag{35}$$

5.1. *Case* $C \equiv 0$

Contrary to the Full ICP, with TICP, the case $C \equiv 0$ can be handled separately to provide more refined bounds for complex eigenvalues. In this case, the eigenvalue problem (30) reads

$$\begin{bmatrix} A_P & R^\top \\ R & 0 \end{bmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} = \lambda \begin{bmatrix} I & R^\top \\ 0 & -I \end{bmatrix}\begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix}. \tag{36}$$

We first need to prove the following Lemma.

DOI: 10.1002/nla

Lemma 4
All the eigenvalues of Equation (36) with $\boldsymbol{v}_1 \neq 0$ satisfy the following equation

$$\lambda^2 - \left(\theta'_S + \theta_A\right)\lambda + \theta'_S = 0 \tag{37}$$

where $\alpha_A \leqslant \theta_A \leqslant \beta_A$ whereas $0 < \theta'_S \leqslant \beta_S$.

Proof
To prove Equation (37), let us write $\boldsymbol{v}_2$ from the second set of equations in Equation (30) as

$$\boldsymbol{v}_2 = -\frac{1}{\lambda} R\boldsymbol{v}_1 \tag{38}$$

and substitute in the first set.

$$A_P\boldsymbol{v}_1 - \lambda\boldsymbol{v}_1 + \frac{\lambda - 1}{\lambda} R^\top R\boldsymbol{v}_1 = 0.$$

Then premultiplying by $\frac{\boldsymbol{v}_1^*}{\boldsymbol{v}_1^*\boldsymbol{v}_1}$ and setting $\theta'_A = \frac{\boldsymbol{v}_1^* A_P \boldsymbol{v}_1}{\boldsymbol{v}_1^*\boldsymbol{v}_1}$ and $\theta'_S = \frac{\boldsymbol{v}_1^* R^\top R\boldsymbol{v}_1}{\boldsymbol{v}_1^*\boldsymbol{v}_1}$, we obtain

$$(\theta_A - \lambda) + \theta'_S \frac{\lambda - 1}{\lambda} = 0,$$

which is equivalent to in Equation (37). □

We are now ready to derive eigenvalue bounds for TICP in the case $C \equiv 0$.

Theorem 6
The eigenvalues of Equation (36) satisfy the following bounds. If $\lambda_I \neq 0$ then

$$|\lambda - 1| \leqslant \sqrt{1 - \alpha_A}, \quad \text{and} \quad \frac{\alpha_A}{2} \leqslant \lambda_R \leqslant \min\left\{\frac{1 + \beta_S}{2}, 2 - \frac{\alpha_A}{2}\right\}. \tag{39}$$

The real eigenvalues satisfy

$$\min\left\{\alpha_A, \frac{\alpha_S}{\beta_A + \alpha_S}\right\} \leqslant \lambda_R \leqslant \beta_S + \beta_A.$$

Proof
The bounds regarding real eigenvalues follow from Theorem 5, by setting $\alpha_C = 0$.

If now $\lambda_I \neq 0$ then from Lemma 4, we know that $|\lambda|^2 = \theta'_S$ and $\lambda_R = \frac{\theta_A + \theta'_S}{2}$ so that

$$|\lambda - 1|^2 = |\lambda|^2 - 2\lambda_R + 1 = \theta'_S - \theta_A - \theta'_S + 1 = 1 - \theta_A \leqslant 1 - \alpha_A.$$

It follows that $\theta_A < 1$ must hold to yield complex eigenvalues. The real part, hence, must satisfy $\lambda_R \leqslant \frac{1+\beta_S}{2}$. Combining this with the bounds of Theorem 5 (with $\alpha_C = 0$), we finally obtain

$$\frac{\alpha_A}{2} \leqslant \lambda_R \leqslant \min\left\{2 - \frac{\alpha_A}{2}, \frac{1 + \beta_S}{2}\right\}.$$

□

6. NUMERICAL RESULTS

6.1. Finite Element coupled consolidation equations

The system of partial differential equations governing the 3D coupled consolidation process in fully saturated porous media is derived from the classical Biot's formulation [18] and successive modifications as

$$\begin{cases} (\lambda + \mu)\dfrac{\partial \epsilon}{\partial \vec{x}} + \mu\nabla^2\vec{u} = \alpha\dfrac{\partial p}{\partial \vec{x}} \\ \dfrac{1}{s_w}\nabla\left(\vec{\vec{K}}\nabla p\right) = \left[\phi\beta + c_{br}(\alpha - \phi)\right]\dfrac{\partial p}{\partial t} + \alpha\dfrac{\partial \epsilon}{\partial t} \end{cases} \tag{40}$$

DOI: 10.1002/nla

where c_{br} and β are the volumetric compressibility of solid grains and water, respectively, ϕ is the porosity, $\vec{\vec{K}}$ the medium hydraulic conductivity, ϵ the medium volumetric dilatation, α the Biot coefficient, λ and μ are the Lamé constant and the shear modulus of the porous medium, respectively, s_w is the specific weight of water, ∇ the gradient operator, t is time, and p and $\vec{u}$ are the incremental pore pressure and the incremental displacement, respectively.

Use of FE in space yields a system of first order differential equations, which can be integrated by the Crank–Nicolson scheme. The resulting linear system has to be repeatedly solved to obtain the transient displacements and pore pressures. The nonsymmetric matrix controlling the solution scheme reads

$$\mathcal{W} = \begin{bmatrix} K/2 & -Q/2 \\ \dfrac{Q^\top}{\Delta t} & H/2 + \dfrac{P}{\Delta t} \end{bmatrix}, \tag{41}$$

where K, H, P and Q are the elastic stiffness, flow stiffness, flow capacity and flow-stress coupling matrices, respectively. The matrix $\mathcal{W}$ can be readily symmetrized by multiplying the upper set of equations by 2 and the lower set by $-\Delta t$, thus obtaining the sparse 2×2 block symmetric indefinite matrix (1), where $A = K$, $B = -Q^\top$ and $C = \Delta t H/2 + P$.

A major difficulty in the repeated solution to system (1) is the likely ill-conditioning of $\mathcal{A}$ caused by the large difference in magnitude between the coefficients of blocks A, B and C. The ill-conditioning of $\mathcal{A}$ is then basically dependent on the size of Δt [16]. In long-term simulations, a small Δt is typically needed in the early stage of the consolidation process, whereas larger values may be used as the system approaches the steady state. Hence, the initial steps are the most critical ones, with the convergence expected to improve as the simulation proceeds.

6.2. *Test problem*

A vertical cross-section of the cylindrical porous volume used as a test problem is shown in Figure 1. The medium consists of a sequence of alternating sandy and clayey layers, with the hydraulic conductivity $k_{\text{sand}} = 10^{-5}$ m/s and $k_{\text{clay}} = 10^{-8}$ m/s, the porosity $\phi = 0.20$, the Poisson ratio $\nu = 0.25$, and the Young modulus $E = 833.33$ MPa, corresponding to a uniaxial vertical compressibility $c_M = 10^{-3}$ MPa^{-1}. Standard Dirichlet conditions are prescribed, with fixed outer and bottom boundaries, and zero pore pressure variation on the top and outer surfaces (see Figure 1). The upper boundary is a traction-free plane. This sample problem is solved using a fully three-dimensional grid. The medium is discretized into linear tetrahedral elements by projecting a plane triangulation made of 209 nodes and 400 triangles onto 17 layers located at different depths [9]. The

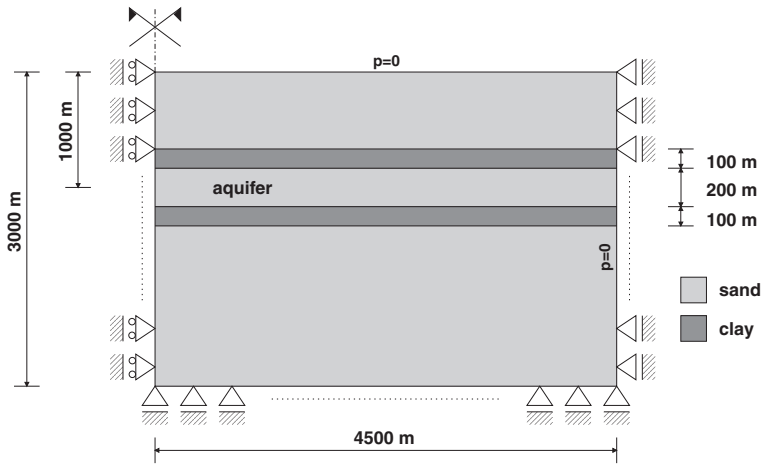

Figure 1. Schematic representation of a vertical cross-section of the stratified porous medium used as a test problem.

DOI: 10.1002/nla

grid totals $m = 3088$ nodes with a global matrix size N equal to 12352 and 707504 nonzeros. We use $\Delta t = 1$, which produces the most ill-conditioned situation.

6.3. Mixed Constraint Preconditioner

The Mixed Constraint Preconditioner (MCP) proposed in [16] is based on an approximation of the (1,1) block A, $P_A = L_A L_A^\top$ using the ILLT approach ([19]), and on an approximation of its inverse $\left(\hat{P}_A^{-1} = Z_A Z_A^\top\right)$, computed following the AINV approach, [10, 20]), which is needed in the explicit construction of the Schur complement matrix. S is then preconditioned by a simple IC(0) preconditioner. In detail,

$$\widehat{S} = B Z_A Z_A^\top B^\top + C, \qquad P_S = L_S L_S^\top.$$

Note that now the preconditioned Schur complement $S_P = P_S^{-1/2}\left(B\left(L_A L_A^\top\right)^{-1} B^\top + C\right)P_S^{-1/2}$ is the result of two approximation because L_S is the Cholesky factor of an already approximated Schur complement matrix $\widehat{S}$.

The MCP application requires first the explicit calculation of the $\widehat{S} = B Z_A Z_A^\top B^\top + C$ and then its incomplete triangular factor. Forming $\widehat{S}$ may be time and memory consuming being the result of two sparse matrix–matrix products and one sparse sum of matrices. However, it may be noted that the evaluation of $S_0 = B Z_A Z_A^\top B^\top$, which involves the main computational burden of $\widehat{S}$, is independent of the time step Δt, and, therefore, can be carried out just once at the beginning of the simulation.

The MCP approach is applicable both to FICP and TICP preconditioner whose inverses can be written as

$$\mathcal{M}_1^{-1} = \begin{bmatrix} \left(L_A L_A^T\right)^{-1} & \left(L_A L_A^T\right)^{-1} B^\top \left(L_S L_S^T\right)^{-1} \\ 0 & -\left(L_S L_S^T\right)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -B\left(L_A L_A^T\right)^{-1} & I \end{bmatrix} \tag{42}$$

$$\mathcal{M}_2^{-1} = \begin{bmatrix} \left(L_A L_A^T\right)^{-1} & \left(L_A L_A^T\right)^{-1} B^\top \left(L_S L_S^T\right)^{-1} \\ 0 & -\left(L_S L_S^T\right)^{-1} \end{bmatrix}. \tag{43}$$

6.4. Comparisons between spectra of preconditioned matrices and bounds

We report the eigenvalue distribution of the preconditioned matrices and the performance of BiCGSTAB in the solution of system (1) using the two preconditioners just defined. The set of numerical tests are obtained by using different parameters of the two approximations of the (1,1) block. They are summarized in Table I, where we also report the extremal eigenvalues of $A_P(A_R)$ and S_P together with the smallest eigenvalue of $\widehat{C}$. Note that in all the tests $\beta_A^R < 2$. The parameters rely on ILLT decomposition of block A (τ_A the dropping threshold and `lfill` the maximum allowed fill-in per row) and τ_Z, the dropping parameter related to the AINV of A again.

In all the combinations of parameters, the BiCGSTAB has been stopped whenever the relative residual was below $tol = 10^{-12}$. This very low tolerance is required to have a relative error of order 10^{-6}. The initial solution has been set to $\boldsymbol{x}_0 = \mathcal{M}^{-1}\boldsymbol{b}$. The CPU times (in seconds) refer to running a Fortran 90 code on an IBM Power6 with 4.7 GHz RAM.

In Table II, we summarized the results of our preconditioners in terms of storage, number of iterations and CPU time. In particular, we report a measure ρ of the density of the preconditioning factors provided as $\rho = \frac{\mathrm{nnz}(L_A)+\mathrm{nnz}(L_S)}{\mathrm{nnz}(\mathcal{A})}$, which gives an indication of the additional core memory needed for computing and storing the preconditioner. In the total CPU time, we neglect the CPU time needed to construct the incomplete Cholesky factor L_A ($T(L_A)$ in the table), the AINV factor Z of A and the constant part (S_0) of the Schur complement matrix, all of them regarded as *preprocessing* as explained before. Instead, we give the CPU time to compute the P_S preconditioner (T_p), the CPU time for the iterative solution (T_{sol}) and the total time (CPU $= T_p + T_{sol}$).

An analysis of this table reveals that FICP is slightly better than TICP; moreover, the optimal combination of parameter is $(\tau_A, \mathtt{lfill}) = (10^{-4}, 50)$, that is, very accurate 'implicit' ILLT

DOI: 10.1002/nla

Table I. Parameters used in the 12 sample tests together with the extremal eigenvalues of $A_P\,(A_R)$, S_P and $\widehat{C}$.

# run	τ_A	lfill	τ_Z	α_A, β_A $\left(\alpha_A^R, \beta_A^R\right)$	α_S	β_S	α_C
1	10^{-4}	50	0.30	0.255 1.255	0.110	23.731	0.008
2	10^{-4}	50	0.10		0.107	7.743	0.002
3	10^{-4}	50	0.05	(0.761 1.217)	0.313	5.111	0.007
4	10^{-4}	50	0.01		0.708	4.112	0.012
5	10^{-2}	30	0.30	0.059 1.662	0.110	23.323	0.008
6	10^{-2}	30	0.10		0.107	7.577	0.012
7	10^{-2}	30	0.05	(0.411 1.423)	0.312	3.100	0.007
8	10^{-2}	30	0.01		0.680	1.583	0.012
9	10^{-1}	10	0.30	0.027 2.046	0.110	23.019	0.008
10	10^{-1}	10	0.10		0.107	7.500	0.002
11	10^{-1}	10	0.05	(0.384 1.590)	0.312	3.059	0.007
12	10^{-1}	10	0.01		0.419	1.388	0.012

Table II. Iteration number and CPU times for BiCGSTAB preconditioned by FICP and TICP.

			FICP				TICP			
# run	$T(L_A)$	ρ	iter	T_p	T_{sol}	CPU	iter	T_p	T_{sol}	CPU
1	0.61	0.93	104	0.08	1.14	1.25	100	0.08	0.96	1.06
2		1.58	65	0.23	0.77	1.04	59	0.24	0.60	0.88
3		2.22	30	0.39	0.38	0.81	29	0.39	0.31	0.75
4		4.91	22	0.83	0.31	1.21	23	0.84	0.28	1.19
5	0.07	0.60	133	0.08	1.14	1.24	119	0.08	0.91	1.00
6		1.25	80	0.22	0.74	1.00	76	0.23	0.64	0.89
7		1.88	51	0.38	0.50	0.91	55	0.38	0.48	0.90
8		4.57	35	0.81	0.38	1.27	48	0.82	0.48	1.37
9	0.01	0.45	181	0.07	1.30	1.39	187	0.08	1.19	1.29
10		1.10	112	0.22	0.89	1.14	129	0.23	0.92	1.17
11		1.73	78	0.37	0.65	1.07	93	0.37	0.71	1.13
12		4.42	67	0.81	0.64	1.52	94	0.82	0.68	1.50

approximation to the structural block together with $\tau_Z = 0.1$ that is sparser AINV approximation of A which produces a rather sparse Schur complement matrix.

The eigenvalue distribution of the preconditioned matrices are very important to analyze the preconditioners properties. In Figures 2 and 3, we depict the eigenvalue distribution in the complex plane of $\mathcal{M}^{-1}\mathcal{A}$ for test cases # 4 and # 6. In the first case, the parameters used for ILLT produce a very accurate preconditioner, and the iterative solver converges in a few iterations with the two preconditioners. The eigenvalues are well clustered around one, the smallest eigenvalue being far away from zero and with the imaginary part of the complex ones more pronounced in the TICP preconditioner.

From Figure 3, we see that the smallest eigenvalue gets close to zero and that the imaginary part of the complex eigenvalues grows for both the preconditioners. However, the TICP preconditioner does not seem to be affected by the larger imaginary part of its complex eigenvalues.

In Table III, the extremal real eigenvalues (l, L), the smallest and the largest real part (l_R, L_R) and the largest imaginary part (L_I) of the eigenvalues of $\mathcal{M}^{-1}\mathcal{A}$ are provided together with the bounds computed following the previous theorems.

Comparisons with the estimates provided by [1] **and** [2]**.** As anticipated in the previous sections, particularly the estimates on real eigenvalues and real part of complex eigenvalues improve those of [1] and [2]. As an example, in the first test case, bounds for TICP taken from Theorem 6 in [2] give $\Re(\lambda) \in [0.0018, 47.46]$ (instead of $[0.089, 29.40]$ given by Table III) without distinguishing between real and complex eigenvalues. Regarding eigenvalues of FICP our paper slightly improves

DOI: 10.1002/nla

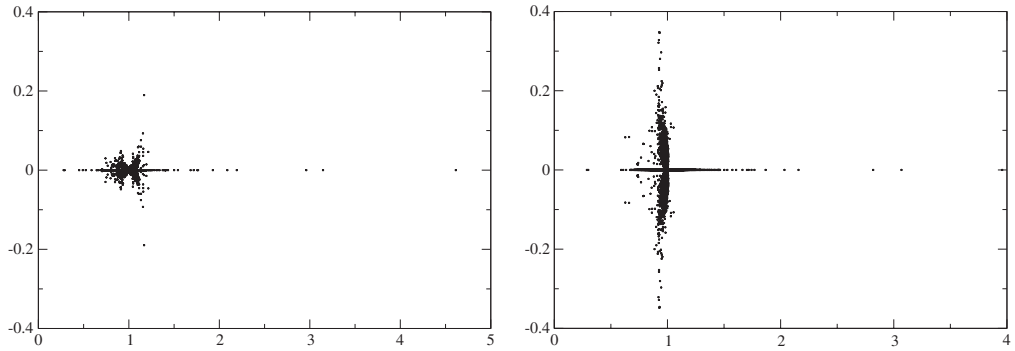

Figure 2. Eigenvalue distribution in the complex plane of $\mathcal{M}^{-1}\mathcal{A}$ for FICP (left figure) and TICP (right figure). Test case # 4.

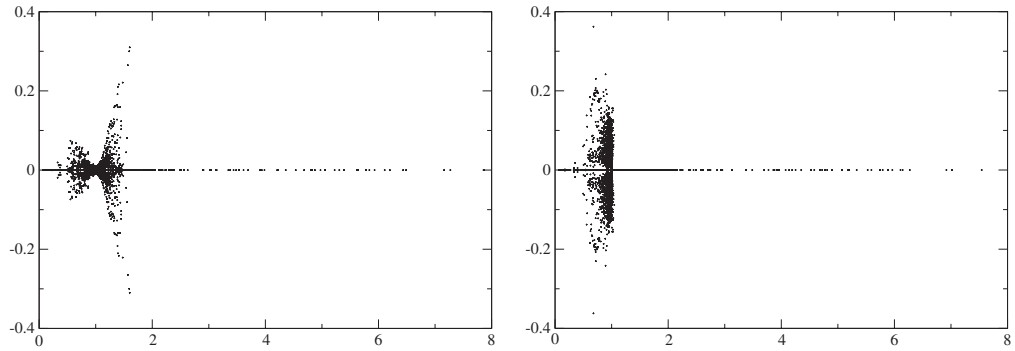

Figure 3. Eigenvalue distribution in the complex plane of $\mathcal{M}^{-1}\mathcal{A}$ for FICP (left figure) and TICP (right figure). Test case # 6.

the bound on the smallest real eigenvalue given in Corollary 5.1 in [1]. When $C \equiv 0$ the bound given in the cited paper can be written as (using our notation) $\lambda \geqslant \min\{\alpha_A, (2-\beta_A)\alpha_S)\} \equiv \chi_0$ while from Corollary 3 we have here $\lambda \geqslant \min\left\{\alpha_A, \frac{\alpha_S}{\beta_A}\right\} \equiv \chi_1$, with obviously $\chi_0 \leqslant \chi_1$. In particular, for test case # 1, $\chi_0 = 0.082 < \chi_1 = 0.089$.

Table III shows that the bounds are very tight, especially for the smallest real eigenvalue for both preconditioners and for the largest one for TICP. Only the bounds for the imaginary part (and in some instances also for the real part) of the complex eigenvalues is for FICP somewhat far from the true value especially for problems # 5–# 12, where the bounds for L_I incorrectly suggest that TICP may perform better than FICP.

However, by comparing Table III with Table II, we note that the complex eigenvalues hardly influence the convergence. The TICP eigenvalues have larger imaginary part than the FICP ones, however the number of iterations of the two preconditioner are roughly the same. The only tests which display a significant difference are the runs # 8 and # 12 where TICP requires 40 – 50 % more iterations than FICP. These cases are characterized by a low quality approximation for P_A provided by the ILLT preconditioner together with a high-quality AINV preconditioner.

The better performance of FICP can be related to the real eigenvalues, namely with a sort of "real" condition number, the ratio $\kappa = \frac{L}{I}$. On problem # 12, for instance, $\kappa = 52$ for FICP while $\kappa = 80$ for TICP. If an SPD problem be solved by PCG, we would expect a number of iteration proportional to the square root of $\lambda_{\max}/\lambda_{\min}$. Here our preconditioned BiCGSTAB behaves in a similar way depending on κ as shown in the next Figure 4 where for each test case and both preconditioners we plot the number of iterations and the value $7\sqrt{\kappa}$. The almost perfect correspondence between these

DOI: 10.1002/nla

Table III. Computed eigenvalues and corresponding bounds of the preconditioned matrices.

#	precondition	computed eigenvalues					bounds				
		l, L		l_R, L_R		L_I	l, L		l_R, L_R		L_I
1	FICP	0.110	24.84	0.44	1.33	0.02	0.089	29.40	0.17	1.68	0.43
1	TICP	0.110	23.69	0.54	1.01	0.12	0.082	24.99	0.13	1.87	0.86
2	FICP	0.107	8.20	0.59	1.34	0.04	0.086	9.59	0.17	1.68	0.43
2	TICP	0.107	7.69	0.57	1.01	0.11	0.079	9.00	0.13	1.87	0.86
3	FICP	0.278	5.68	0.64	1.36	0.07	0.251	6.33	0.25	1.68	0.43
3	TICP	0.287	4.98	0.57	1.02	0.17	0.201	6.37	0.13	1.87	0.86
4	FICP	0.279	4.61	0.71	1.22	0.19	0.255	5.09	0.41	1.68	0.43
4	TICP	0.288	3.96	0.63	1.05	0.35	0.255	5.37	0.13	1.87	0.86
5	FICP	0.073	24.27	0.14	1.86	0.50	0.059	37.06	0.06	4.92	1.96
5	TICP	0.074	23.29	0.17	1.03	0.20	0.059	24.99	0.03	1.97	0.97
6	FICP	0.073	7.86	0.14	1.60	0.31	0.059	12.04	0.06	4.92	1.62
6	TICP	0.074	7.54	0.14	1.05	0.36	0.059	9.24	0.03	1.97	0.97
7	FICP	0.073	3.19	0.31	1.70	0.37	0.059	4.92	0.12	3.30	1.04
7	TICP	0.074	3.09	0.38	1.06	0.43	0.059	4.76	0.03	1.97	0.97
8	FICP	0.073	1.74	0.65	1.49	0.19	0.059	2.51	0.23	2.09	0.74
8	TICP	0.074	2.03	0.59	1.00	0.51	0.059	3.24	0.04	1.96	0.96
9	FICP	0.031	23.87	0.11	1.77	0.71	0.027	37.19	0.04	19.62	2.95
9	TICP	0.031	22.99	0.11	1.12	0.49	0.027	25.06	0.02	1.98	0.98
10	FICP	0.031	7.71	0.11	1.78	0.43	0.027	12.12	0.04	7.08	1.69
10	TICP	0.031	7.48	0.21	1.03	0.55	0.027	9.55	0.01	1.99	0.99
11	FICP	0.031	3.12	0.39	1.58	0.35	0.027	4.94	0.08	3.50	1.08
11	TICP	0.031	3.05	0.38	1.00	0.57	0.027	5.11	0.02	1.98	0.98
12	FICP	0.031	1.61	0.51	1.31	0.36	0.027	2.24	0.11	2.15	0.72
12	TICP	0.031	2.49	0.41	1.00	0.50	0.027	3.43	0.02	1.98	0.98

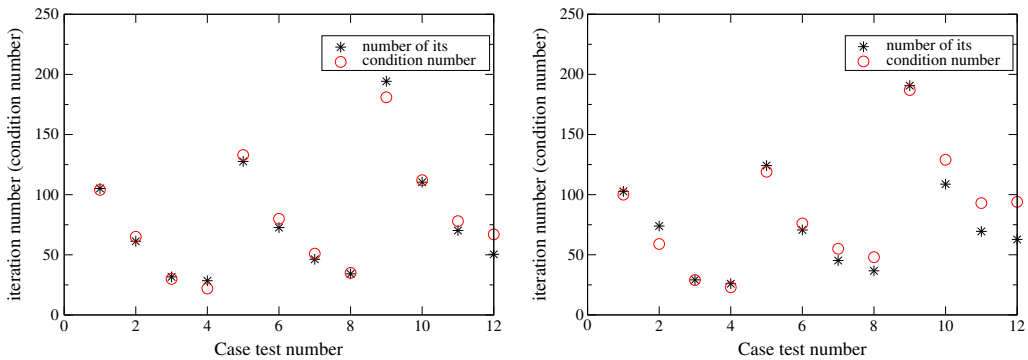

Figure 4. Plot of scaled square root of condition number $7\sqrt{\kappa}$ and number of iteration for each test case. Full Inexact Constraint Preconditioner (FICP) on the left, Triangular Inexact Constraint Preconditioner (TICP) on the right.

two numbers (the scale factor 7 has been chosen heuristically) gives experimental evidence that the performance of the analyzed preconditioners is very much dependent on extremal real eigenvalues.

7. CONCLUSION

In this paper, the spectral properties of some ICP have been investigated in the solution of generalized saddle point linear systems. The extremal eigenvalues of the preconditioned matrices have been put in connection with those of two matrices having positive eigenvalues: $P_A^{-1}A$ and $P_S^{-1}S$; P_A and P_S being two SPD preconditioners for the (1,1) block A and for a suitable Schur complement matrix S, respectively. Results obtained onto a realistic problem reveal that the bounds are very tight especially for the real eigenvalues.

DOI: 10.1002/nla

APPENDIX A: PROOF OF THEOREM 5

Proof
Recalling Lemma 3 the real eigenvalues of Equation (30), which do not lie in $[\alpha_A, \beta_A]$ (if any), satisfy

$$\lambda^2 - (\theta_S + \theta_A)\,\lambda + \theta_S + (\theta_A - 1)\theta_C = 0.$$

The larger root is bounded by

$$\lambda_2 = \frac{\theta_S + \theta_A + \sqrt{(\theta_S + \theta_A)^2 - 4(\theta_A - 1)\theta_C - 4\theta_S}}{2} \leqslant \theta_S + \theta_A \leqslant \beta_S + \beta_A.$$

Regarding the lower bound for λ_1, we consider separately the cases $\theta_A \leqslant 1$ and $\theta_A > 1$. If $\theta_A < 1$

$$\begin{aligned}\lambda_1 &= \frac{\theta_S + \theta_A - \sqrt{(\theta_S + \theta_A)^2 - 4(\theta_A - 1)\theta_C - 4\theta_S}}{2} \\ &= \frac{\theta_S + \theta_A - \sqrt{(\theta_S - \theta_A)^2 - 4(1 - \theta_A)(\theta_S - \theta_C)}}{2} \geqslant \min\{\theta_S, \theta_A\} \geqslant \min\{\alpha_S, \alpha_A\} \geqslant \alpha_S.\end{aligned}$$

If $\theta_A > 1$ then if $\theta_C \geqslant 1$ direct calculation show that $\lambda_1 \geqslant 1 > \alpha_S$. If, instead, $\theta_C \leqslant 1$ we find that λ_1 is an increasing function in θ_S and it is decreasing in θ_A:

$$\begin{aligned}\frac{\partial \lambda_1}{\partial \theta_S} \geqslant 0 &\Longleftrightarrow \sqrt{(\theta_S + \theta_A)^2 - 4(\theta_A - 1)\theta_C - 4\theta_S} \geqslant \theta_A + \theta_S - 2 \\ &\Longleftarrow (\theta_S + \theta_A)^2 + 4\theta_C - 4\theta_A\theta_C - 4\theta_S \geqslant (\theta_A + \theta_S)^2 - 4\theta_A - 4\theta_S + 4 \\ &\Longleftrightarrow \theta_C(1 - \theta_A) \geqslant 1 - \theta_A \Longleftrightarrow \theta_C \leqslant 1 \\ \frac{\partial \lambda_1}{\partial \theta_A} \leqslant 0 &\Longleftrightarrow \sqrt{(\theta_S + \theta_A)^2 - 4(\theta_A - 1)\theta_C - 4\theta_S} \leqslant \theta_A + \theta_S - 2\theta_C \\ &\Longleftarrow (\theta_S + \theta_A)^2 + 4\theta_C - 4\theta_A\theta_C - 4\theta_S \leqslant (\theta_A + \theta_S)^2 - 4\theta_C\theta_A - 4\theta_S\theta_C + 4\theta_C^2 \\ &\Longleftrightarrow \theta_C - \theta_S \leqslant \theta_C\,(\theta_C - \theta_S) \Longleftrightarrow \theta_C \leqslant 1.\end{aligned}$$

Hence,

$$\begin{aligned}\lambda_1 &= \frac{2\,(\theta_S + (\theta_A - 1)\theta_C)}{\theta_S + \theta_A + \sqrt{(\theta_S + \theta_A)^2 - 4\theta_A\theta_C - 4(\theta_S - \theta_C)}} \\ &\geqslant \frac{2\,(\alpha_S + (\beta_A - 1)\alpha_C)}{\alpha_S + \beta_A + \sqrt{(\alpha_S + \beta_A)^2 - 4\beta_A\alpha_C - 4(\alpha_S - \alpha_C)}} \geqslant \frac{\alpha_S}{\alpha_S + \beta_A} + \frac{(\beta_A - 1)\alpha_C}{\alpha_S + \beta_A}.\end{aligned}$$

The thesis holds by finally observing that

$$\frac{\alpha_S}{\alpha_S + \beta_A} + \frac{(\beta_A - 1)\alpha_C}{\alpha_S + \beta_A} = \frac{(\alpha_S - \alpha_C)(1 - \beta_A) + \alpha_S\beta_A}{\alpha_S + \beta_A} < \frac{\alpha_S\beta_A}{\alpha_S + \beta_A} < \alpha_S.$$

The bounds regarding complex eigenvalues follow from the proof of Theorem 2 in [2]. Because these results are not explicitly given in that paper, we briefly derive them. Multiplying the first equation of Equation (20) by $\frac{\boldsymbol{v}_1^*}{\|\boldsymbol{v}_1\|^2}$ and the second by $\frac{\boldsymbol{v}_2^*}{\|\boldsymbol{v}_2\|^2}$, we obtain

$$\begin{cases} \theta_A + \phi = \lambda + \lambda\phi \\ \bar{\phi} - \theta_C\,\rho = -\lambda\rho, \end{cases} \tag{44}$$

where θ_A, θ_C, ϕ and ρ are as in Equation (26). Now, from the second equation $\phi = \left(\theta_C - \bar{\lambda}\right)\rho$, and substituting in the first one yields

$$\theta_A - \lambda = (\lambda - 1)\left(\theta_C - \bar{\lambda}\right)\rho \quad \text{or} \quad \theta_A - \lambda = (-|\lambda|^2 + \bar{\lambda} + \theta_C(\lambda - 1))\rho. \tag{45}$$

DOI: 10.1002/nla

The imaginary part gives $\rho = \frac{1}{1-\theta_C}$, which implies $\theta_C < 1$; substituting in Equation (45) and taking the real part yields

$$(1-\theta_C)(\theta_A - \lambda) = -|\lambda|^2 + \bar{\lambda} + \theta_C(\lambda - 1) \tag{46}$$

and after some algebra, we finally obtain

$$|\lambda - 1|^2 = (1-\theta_A)(1-\theta_C) \leqslant \text{ (since also } \theta_A \text{ must be less than 1) } \leqslant (1-\alpha_A)(1-\alpha_C).$$

The bound for the real part comes from $(\lambda_R - 1)^2 < (1-\alpha_A)(1-\alpha_C)$ and using the elementary result $\left(\frac{a+c}{2} - 1\right)^2 \geqslant (1-a)(1-c)$. □

ACKNOWLEDGEMENTS

We wish to thank the anonymous reviewers whose constructive criticism helped improve the presentation of the paper. This work has been funded by the Italian MIUR project (PRIN) 'Advanced numerical methods and models for environmental fluid-dynamics and geomechanics' and CINECA project 'Parallel preconditioners for large scale engineering problems'

REFERENCES

1. Benzi M, Simoncini V. On the eigenvalues of a class of saddle point matrices. *Numerische Mathematik* 2006; **103**:173–196.
2. Simoncini V. Block triangular preconditioners for symmetric saddle-point problems. *Applied Numerical Mathematics* 2004; **49**:63–80.
3. Benzi M, Golub GH, Liesen J. Numerical solution of saddle point problems. *Acta Numerica* 2005; **14**:1–137.
4. Bergamaschi L, Gondzio J, Zilli G. Preconditioning indefinite systems in interior point methods for optimization. *Computational Optimization and Applications* 2004; **28**:149–171.
5. Keller C, Gould NIM, Wathen AJ. Constraint preconditioning for indefinite linear systems. *SIAM Journal on Matrix Analysis and Applications* 2000; **21**:1300–1317.
6. Lukšan L, Vlček J. Indefinitely preconditioned inexact Newton method for large sparse equality constrained nonlinear programming problems. *Numerical Linear Algebra with Applications* 1998; **5**:219–247.
7. Perugia I, Simoncini V. Block-diagonal and indefinite symmetric preconditioners for mixed finite elements formulations. *Numerical Linear Algebra with Applications* 2000; **7**:585–616.
8. Bergamaschi L, Gondzio J, Venturin M, Zilli G. Inexact constraint preconditioners for linear systems arising in interior point methods. *Computational Optimization and Applications* 2007; **36**:136–147.
9. Bergamaschi L, Ferronato M, Gambolati G. Novel preconditioners for the iterative solution to FE-discretized coupled consolidation equations. *Computer Methods in Applied Mechanics and Engineering* 2007; **196**:2647–2656.
10. Benzi M, Tůma M. A comparative study of sparse approximate inverse preconditioners. *Applied Numerical Mathematics* 1999; **30**:305–340.
11. Elman HC, Silvester DJ, Wathen AJ. Performance and analysis of saddle point preconditioners for the discrete steady-state Navier-Stokes equations. *Numerische Mathematik* 2002; **90**:665–688.
12. Silvester D, Elman H, Kay D, Wathen A. Efficient preconditioning of the linearized Navier-Stokes equations for incompressible flow. *Journal of Computational and Applied Mathematics* 2001; **128**:261–279. Numerical analysis 2000, Vol. VII, Partial differential equations.
13. Silvester D, Wathen A. Fast iterative solution of stabilised Stokes systems. II. Using general block preconditioners. *SIAM Journal on Numerical Analysis* 1994; **31**:1352–1367.
14. Axelsson O, Neytcheva M. Eigenvalue estimates for preconditioned saddle point matrices. *Numerical Linear Algebra with Applications* 2006; **13**:339–360.
15. Rusten T, Winther R. A Preconditioned Iterative Method for Saddlepoint Problems. *SIAM Journal on Matrix Analysis and Applications* 1992; **13**:887–904.
16. Bergamaschi L, Ferronato M, Gambolati G. Mixed constraint preconditioners for the solution to FE coupled consolidation equations. *Journal of Computational Physics* 2008; **227**:9885–9897.
17. Liesen J, Parlett BN. On nonsymmetric saddle point matrices that allow conjugate gradient iterations. *Numerische Mathematik* 2008; **108**:605–624.
18. Biot MA. General theory of three-dimensional consolidation. *Journal of Applied Physics* 1941; **12**:155–164.
19. Saad Y. ILUT: A dual threshold incomplete LU factorization. *Numerical Linear Algebra with Applications* 1994; **1**:387–402.
20. Benzi M, Cullum JK, Tůma M. Robust approximate inverse preconditioning for the conjugate gradient method. *SIAM Journal on Scientific Computing* 2000; **22**:1318–1332.

DOI: 10.1002/nla