Eigenvalue bounds for double saddle-point systems

Susanne Bradley* and Chen Greif

Department of Computer Science, The University of British Columbia, Vancouver, Canada V6T 1Z4
*Corresponding author: smbrad@cs.ubc.ca

We derive bounds on the eigenvalues of a generic form of double saddle-point matrices. The bounds are expressed in terms of extremal eigenvalues and singular values of the associated block matrices. Inertia and algebraic multiplicity of eigenvalues are considered as well. The analysis includes bounds for preconditioned matrices based on block diagonal preconditioners using Schur complements, and it is shown that in this case the eigenvalues are clustered within a few intervals bounded away from zero. Analysis for approximations of Schur complements is included. Some numerical experiments validate our analytical findings.

Keywords: double saddle-point systems; eigenvalue bounds; eigenvalues and singular values; iterative methods; preconditioners; Schur complement.

1. Introduction

Given positive integer dimensions $n \geq m \geq p$, consider the $(n+m+p) \times (n+m+p)$ double saddle-point system

$$\mathcal{K} u = b, \tag{1.1}$$

where

$$\mathcal{K} = \begin{bmatrix} A & B^T & 0 \\ B & -D & C^T \\ 0 & C & E \end{bmatrix}; \quad u = \begin{bmatrix} x \\ y \\ z \end{bmatrix}; \quad b = \begin{bmatrix} p \\ q \\ r \end{bmatrix}. \tag{1.2}$$

In (1.2), $A \in \mathbb{R}^{n \times n}$, $D \in \mathbb{R}^{m \times m}$ and $E \in \mathbb{R}^{p \times p}$ are positive semidefinite, and $B \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{p \times m}$. In many applications, A is positive definite and B and C have full row rank.

Under the assumptions stated above, $\mathcal{K}$ is symmetric and indefinite, and solving (1.1) presents several numerical challenges. When the linear system is too large for direct solvers to work effectively, iterative methods (Saad, 2003) are preferred; this is the scenario that is of interest to us in this work. A minimum residual Krylov subspace solver such as MINRES (Paige & Saunders, 1975) is a popular choice due to its optimality and short recurrence properties. Other solvers may be highly effective as well.

Linear systems of the form (1.1)–(1.2) appear frequently in multiphysics problems, and their numerical solution is of increasing importance and interest. There is a large number of relevant applications (Cai *et al.*, 2009; Keyes *et al.*, 2013; Sogn & Zulehner, 2018); liquid crystal problems (Ramage & Gartland, 2013; Ali Beik & Benzi, 2018), Darcy–Stokes equations (Cai *et al.*, 2021; Holter *et al.*, 2021), coupled poromechanical equations (Ferronato *et al.*, 2019), magma-mantle dynamics (Rhebergen *et al.*, 2015) and partial differential equation (PDE)-constrained optimization problems

(Rees *et al.*, 2010; Pearson & Wathen, 2012) are just a small subset of linear systems that fit into this framework.

The matrices D and E are often associated with regularization (or stabilization). We refer to the special case of $D = E = 0$ as an *unregularized form* of $\mathcal{K}$, and denote it by $\mathcal{K}_0$:

$$\mathcal{K}_0 = \begin{bmatrix} A & B^T & 0 \\ B & 0 & C^T \\ 0 & C & 0 \end{bmatrix}. \tag{1.3}$$

This simpler form has been analyzed in various ways in Sogn & Zulehner (2018); Huang & Ma (2019); Pearson & Potschka (2021a) and is of much potential interest, as it may be considered a direct generalization of the standard saddle-point form for 2×2 block matrices:

$$\tilde{\mathcal{K}}_0 = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}. \tag{1.4}$$

Matrices of the form (1.4) have been extensively studied, and their analytical and numerical properties are well understood; see Benzi *et al.* (2005); Rozložník (2018) for excellent surveys. Some properties of $\mathcal{K}$ follow from appropriately reordering and partitioning the block matrix, and then using known results for block-2×2 saddle-point matrices (as given in, for example, Rusten & Winther, 1992; Silvester & Wathen, 1994; Gould & Simoncini, 2009; Ruiz *et al.*, 2018). Specifically, $\mathcal{K}$ can be reordered and partitioned into a 2×2 block matrix

$$\tilde{\mathcal{K}} = \left[\begin{array}{cc|c} A & 0 & B^T \\ 0 & E & C \\ \hline B & C^T & -D \end{array} \right]. \tag{1.5}$$

While this approach is often effective, we may benefit from considering the block-3×3 formulation $\mathcal{K}$ directly, without resorting to (1.5). When E is rank-deficient or zero (as occurs, for example, in (Gatica & Heuer, 2000; Langer *et al.*, 2007)), the leading 2×2 block of $\tilde{\mathcal{K}}$ is singular, even if A is full rank. It is then more challenging to develop preconditioners for $\tilde{\mathcal{K}}$ or to obtain bounds on its eigenvalues, as we are restricted to methods that can handle singular leading blocks. Additionally, by considering the block-3×3 formulation in deriving eigenvalue bounds, we will see in this paper that we can derive effective bounds using the singular values of B and C (along with the eigenvalues of the diagonal blocks). Analyzing $\tilde{\mathcal{K}}$ using established results for block-2×2 matrices (such as those given by Rusten & Winther, 1992) instead requires singular values of the larger off-diagonal block $\begin{bmatrix} B & C^T \end{bmatrix}$, which may be more difficult to obtain than singular values of B and C. (We can estimate the singular values of $\begin{bmatrix} B & C^T \end{bmatrix}$ in terms of those of B and C, but if these estimates are loose it will result in loose eigenvalue bounds.)

There has been a recent surge of interest in the iterative solution of multiple saddle-point problems, and our work adds to the increasing body of literature that considers these problems. Recent papers that provide interesting analysis are, for example, Cai *et al.* (2009); Ali Beik & Benzi (2018); Sogn & Zulehner (2018); Beigl *et al.* (2020); Pearson & Potschka (2021a).

The distribution of eigenvalues of $\mathcal{K}$ plays a central role in determining the efficiency of iterative solvers. It is therefore useful to gain an understanding of the spectral structure as part of the selection and employment of solvers. Effective preconditioners are instrumental in accelerating convergence for sparse and large linear systems; see Benzi (2002); Saad (2003); Wathen (2015); Pearson & Pestana (2020) for general overviews, and Benzi *et al.* (2005); Pestana & Wathen (2015); Rozložník (2018) for a useful overview of solvers and preconditioners for saddle-point problems. For multiphysics problems,

it has been demonstrated that exploiting the properties of the underlying discrete differential operators and other characteristics of the problem at hand is beneficial in the development of robust and fast solvers; see, e.g., Ferronato *et al.* (2019).

There are several potential approaches to the development of preconditioners for the specific class of problem that we are considering. Monolithic preconditioners, which work on the entire matrix, have been recently shown to be extremely effective. Recent work such as Adler *et al.* (2021) has pushed the envelope towards scalable solvers based on this methodology. Another increasingly important approach is operator preconditioning based on continuous spaces; see Hiptmair (2006); Mardal & Winther (2011). This approach relies on the properties of the underlying continuous differential operators, and uses tools such as Riesz representation and natural norm considerations to derive block diagonal preconditioners. Block diagonal preconditioners can also be derived directly by linear algebra considerations accompanied by properties of discretized PDEs; see, for example, Elman *et al.* (2005); Pestana & Wathen (2015).

When A is positive definite, we consider the preconditioner:

$$\mathcal{M} := \begin{bmatrix} A & 0 & 0 \\ 0 & S_1 & 0 \\ 0 & 0 & S_2 \end{bmatrix}, \tag{1.6}$$

where

$$S_1 = D + BA^{-1}B^T; \quad S_2 = E + CS_1^{-1}C^T. \tag{1.7}$$

We assume that S_1 and S_2 are both positive definite.

The preconditioner $\mathcal{M}$ is based on Schur complements. It has been considered in, for example, Cai *et al.* (2009); Sogn & Zulehner (2018); Huang & Ma (2019), and is a natural extension of Murphy *et al.* (2000); Ipsen (2001) for block-2×2 matrices of the form (1.4). In practice, the Schur complements S_1 and S_2 defined in (1.7) are too expensive to form and invert exactly. It is therefore useful to consider approximations to those matrices when a practical preconditioner is to be developed, and we include an analysis of that scenario.

The properties of the preconditioning approach (1.7) have been the focus of several recent articles. The paper by Sogn & Zulehner (2018) analyzes the performance of a block-$n \times n$ block diagonal preconditioner analogous to $\mathcal{M}$ defined in (1.6), concentrating on the spectral properties of the continuous preconditioned operator. The paper by Pearson & Potschka (2021a) provides additional analytical results and bounds, and in Pearson & Potschka (2021b), the concentration of eigenvalues near zero is discussed and an alternative preconditioning approach is offered for multiple saddle-point systems with a larger number of blocks. The papers by Cai *et al.* (2009); Huang & Ma (2019) focus their analyses on the case where all diagonal blocks of $\mathcal{K}$ except A are zero. See also Remark 3.4.

Existing analyses of the eigenvalues of unpreconditioned block-3×3 matrices have often been restricted to specific problems, such as interior-point methods in constrained optimization; see Greif *et al.* (2014); Morini *et al.* (2016). Our goal in this paper is to provide a general framework for analysis of eigenvalue bounds for the double saddle-point case, with a rather minimal set of assumptions on the matrices involved.

This paper extends the work in the literature in a few useful ways. We use energy estimates and other analytical tools to prove our results in a variety of cases. Our proof techniques enable a refined

TABLE 1   *Summary of notation for eigenvalues/singular values of matrix blocks*

Matrix	Size	Type	Number	Notation	max	min
A	$n \times n$	eigenvalues	n	$\mu_i^A,\ i = 1, \ldots, n$	$\mu_{\max}^A$	$\mu_{\min}^A$
B	$m \times n$	singular values	m	$\sigma_i^B,\ i = 1, \ldots, m$	$\sigma_{\max}^B$	$\sigma_{\min}^B$
C	$p \times m$	singular values	p	$\sigma_i^C,\ i = 1, \ldots, p$	$\sigma_{\max}^C$	$\sigma_{\min}^C$
D	$m \times m$	eigenvalues	m	$\mu_i^D,\ i = 1, \ldots, m$	$\mu_{\max}^D$	$\mu_{\min}^D$
E	$p \times p$	eigenvalues	p	$\mu_i^D,\ i = 1, \ldots, p$	$\mu_{\max}^E$	$\mu_{\min}^E$

analysis of the eigenvalues of the unpreconditioned matrix when only $D = 0$, which is an interesting case because it arises commonly in applications such as PDE-constrained optimization (Rees *et al.*, 2010). Here, we obtain bounds that cannot be easily derived from the analysis of the case in which both D and E are (potentially) nonzero. Finally and most significantly, we provide eigenvalue bounds for the preconditioned system when approximations of the leading block and Schur complement inversions are used in the challenging case where $D, E \neq 0$. Our assumptions are minimal, and the analysis is broader than the analysis in Huang & Ma (2019) for the unregularized matrix $\mathcal{K}_0$.

In Section 2, we derive bounds on the eigenvalues of $\mathcal{K}$. In Section 3, we turn our attention to preconditioned matrices of the form $\mathcal{M}^{-1}\mathcal{K}$. In Section 4, we allow the preconditioners to contain approximations of Schur complements and an approximate inversion of the leading block, and show how the bounds are affected as a result. In Section 5, we validate some of our analytical observations with numerical experiments. Finally, in Section 6, we draw some conclusions.

Notation. The $n + m + p$ eigenvalues of the unpreconditioned and preconditioned double saddle-point matrices will be denoted by λ. We will use μ with appropriate matrix superscripts and subscripts to denote eigenvalues of the matrix blocks, and σ will accordingly signify singular values. The eigenvalues of A, for example, will be denoted by

$$\mu_i^A, \quad i = 1, \ldots, n,$$

and in terms of ordering we will assume that

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n > 0.$$

To increase clarity, we will use $\mu_{\max}^A$ to denote μ_1^A and $\mu_{\min}^A$ to denote μ_n^A, and so on.

Based on the above conventions, we use the following notation for eigenvalues and singular values of the matrices comprising $\mathcal{K}$: finally, in some parts of our analysis, we denote a symmetric positive semidefinite matrix X by $X \succeq 0$.

2. Eigenvalue bounds for $\mathcal{K}$

2.1 *Inertia and solvability conditions*

We first discuss the inertia and conditions for nonsingularity of $\mathcal{K}$ defined in (1.2). Recall that the inertia of a matrix is the triplet denoting the number of its positive, negative and zero eigenvalues (Horn & Johnson, 1985, Definition 4.5.6).

PROPOSITION 2.1   The following conditions are necessary for $\mathcal{K}$ to be invertible:

  (i) $\ker(A) \cap \ker(B) = \{0\}$;

  (ii) $\ker(B^T) \cap \ker(D) \cap \ker(C) = \{0\}$;

  (iii) $\ker(C^T) \cap \ker(E) = \{0\}$.

A sufficient condition for $\mathcal{K}$ to be invertible is that A, S_1 and S_2 are invertible.

Proof.   We begin with the proof of statement (i) by assuming to the contrary that the intersection of the kernels is not empty—namely, there exists a nonzero vector x such that $Ax = Bx = 0$. This would mean that the block vector $\begin{bmatrix} x^T & 0 & 0 \end{bmatrix}^T$ was a null vector of $\mathcal{K}$, which would imply that $\mathcal{K}$ was singular. Similar reasoning proves (ii) and (iii).

For the sufficient condition, we observe that when A, S_1 and S_2 are invertible we can write a block-LDL^T factorization of $\mathcal{K}$:

$$
\begin{bmatrix} A & B^T & 0 \\ B & -D & C^T \\ 0 & C & E \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ BA^{-1} & I & 0 \\ 0 & -CS_1^{-1} & I \end{bmatrix} \underbrace{\begin{bmatrix} A & 0 & 0 \\ 0 & -S_1 & 0 \\ 0 & 0 & S_2 \end{bmatrix}}_{=:\mathscr{D}} \begin{bmatrix} I & A^{-1}B^T & 0 \\ 0 & I & -S_1^{-1}C^T \\ 0 & 0 & I \end{bmatrix}. \tag{2.1}
$$

When A and S_1 are invertible, S_2 is well defined and $\mathcal{K}$ is invertible if and only if $\mathscr{D}$ is invertible. The stated result follows. $\square$

Throughout the rest of this paper, we assume that the sufficient condition holds: namely, that A, S_1 and S_2 are invertible. Given our assumptions that D and E are semidefinite, this is equivalent to A, S_1 and S_2 being positive definite. This allows us to obtain the following result on the inertia of $\mathcal{K}$, which will be useful in deriving our bounds.

LEMMA 2.1   (Inertia of a double saddle-point matrix). If A, S_1 and S_2 are positive definite, the matrix $\mathcal{K}$ has $n + p$ positive eigenvalues and m negative eigenvalues.

Proof.   Sylvester's Law of Inertia tells us the inertia of $\mathcal{K}$ is the same as that of $\mathscr{D}$ defined in (2.1); the stated result follows. $\square$

COROLLARY 2.1   Let a, b, c, d and e be scalars with $a > 0$, $d, e \geq 0$, $s_1 := d + \frac{b^2}{a} > 0$ and $s_2 := e + \frac{c^2}{s_1} > 0$. Any cubic polynomial of the form

$$
p(\lambda) = \lambda^3 + (d - a - e)\lambda^2 + (ae - ad - de - b^2 - c^2)\lambda + (ade + ac^2 + b^2 e)
$$

has two positive real roots and one negative real root.

Proof.   Consider the 3×3 matrix

$$
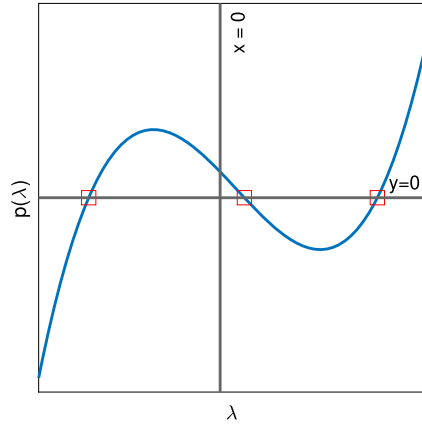P = \begin{bmatrix} a & b & 0 \\ b & -d & c \\ 0 & c & e \end{bmatrix}.
$$

Fig. 1. Plot of a cubic polynomial $p(\lambda)$ of the form described in Corollary 2.1, with two positive roots and one negative root.

Using the characteristic polynomial of P, it is straightforward to confirm that

$$\det(\lambda I - P) = \lambda^3 - Tr(P)\lambda^2 - \frac{1}{2}\left(Tr(P^2) - Tr^2(P)\right)\lambda - \det(P)$$

$$= p(\lambda).$$

Because P is symmetric its eigenvalues are real, and because P is a double saddle-point matrix with $n = m = p = 1$, the two positive and one negative root follow by Lemma 2.1. See Fig. 1 for a graphical illustration. □

2.2 *Derivation of bounds*

Let us define three cubic polynomials, as follows:

$$p(\lambda) = \lambda^3 + (\mu_{\max}^D - \mu_{\min}^A)\lambda^2 - \left(\mu_{\min}^A\mu_{\max}^D + (\sigma_{\max}^B)^2 + (\sigma_{\min}^C)^2\right)\lambda + \mu_{\min}^A(\sigma_{\min}^C)^2; \qquad (2.2a)$$

$$\begin{aligned} q(\lambda) = \lambda^3 &+ (\mu_{\min}^D - \mu_{\max}^A - \mu_{\max}^E)\lambda^2 \\ &+ \left(\mu_{\max}^A\mu_{\max}^E - \mu_{\max}^A\mu_{\min}^D - \mu_{\min}^D\mu_{\max}^E - (\sigma_{\max}^B)^2 - (\sigma_{\max}^C)^2\right)\lambda \\ &+ \left(\mu_{\max}^A\mu_{\min}^D\mu_{\max}^E + \mu_{\max}^A(\sigma_{\max}^C)^2 + (\sigma_{\max}^B)^2\mu_{\max}^E\right); \end{aligned} \qquad (2.2b)$$

$$\begin{aligned} r(\lambda) = \lambda^3 &+ (\mu_{\max}^D - \mu_{\min}^A - \mu_{\min}^E)\lambda^2 \\ &+ \left(\mu_{\min}^A\mu_{\min}^E - \mu_{\min}^A\mu_{\max}^D - \mu_{\max}^D\mu_{\min}^E - (\sigma_{\max}^B)^2 - (\sigma_{\max}^C)^2\right)\lambda \\ &+ (\mu_{\min}^A\mu_{\max}^D\mu_{\min}^E + \mu_{\min}^A(\sigma_{\max}^C)^2 + (\sigma_{\max}^B)^2\mu_{\min}^E). \end{aligned} \qquad (2.2c)$$

All three of these polynomials are of the form described in Corollary 2.1. Thus, all roots are real and each polynomial has two positive roots and one negative root. For notational convenience, we will denote the negative root of $p(\lambda)$, for example, by p^-, and use subscripts *max* and *min* to distinguish between the two positive roots. For example, $p_{\max}^+$ will denote the largest positive root and $p_{\min}^+$ will denote the smallest positive root. The same notational rules apply to $q(\lambda)$ and $r(\lambda)$.

THEOREM 2.2 (Eigenvalue bounds, matrix $\mathscr{K}$). Suppose A is symmetric positive definite. Using the notation established in (2.2), the eigenvalues of $\mathscr{K}$ are bounded within the intervals

$$\left[r^-, \frac{\mu_{\max}^A - \sqrt{(\mu_{\max}^A)^2 + 4(\sigma_{\min}^B)^2}}{2} \right] \bigcup [p_{\min}^+, q_{\max}^+]. \tag{2.3}$$

Proof. *Upper bound on positive eigenvalues.* We let $v = \begin{bmatrix} x^T & y^T & z^T \end{bmatrix}^T$ be a vector with $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^p$. Because $\mathscr{K}$ is symmetric, we can derive an upper bound on the eigenvalues of $\mathscr{K}$ by bounding the value of $\frac{v^T \mathscr{K} v}{v^T v}$. We can write

$$v^T \mathscr{K} v = x^T A x - y^T D y + z^T E z + 2 x^T B^T y + 2 y^T C^T z. \tag{2.4}$$

We use Cauchy–Schwarz to bound the mixed bilinear forms — such as $x^T B^T y$ – by, for example,

$$- \|x\| \cdot \|B^T y\| \le x^T B^T y \le \|x\| \cdot \|B^T y\|.$$

Using this and the eigenvalues/singular values of the block of $\mathscr{K}$, we can bound (2.4) from above by

$$v^T \mathscr{K} v \le \mu_{\max}^A \|x\|^2 - \mu_{\min}^D \|y\|^2 + \mu_{\max}^E \|z\|^2 + 2\sigma_{\max}^B \|x\| \cdot \|y\| + 2\sigma_{\max}^C \|y\| \cdot \|z\|$$

$$= \begin{bmatrix} \|x\| & \|y\| & \|z\| \end{bmatrix} \underbrace{\begin{bmatrix} \mu_{\max}^A & \sigma_{\max}^B & 0 \\ \sigma_{\max}^B & -\mu_{\min}^D & \sigma_{\max}^C \\ 0 & \sigma_{\max}^C & \mu_{\max}^E \end{bmatrix}}_{=:R} \begin{bmatrix} \|x\| \\ \|y\| \\ \|z\| \end{bmatrix}.$$

An upper bound on $\frac{v^T \mathscr{K} v}{v^T v}$ is therefore given by the maximal eigenvalue of R. The largest positive eigenvalue of $\mathscr{K}$ is therefore less than or equal to the largest root of the characteristic polynomial $\det(\lambda I - R)$, which yields the desired result.

Lower bound on negative eigenvalues. The proof is similar to that for the upper bound on the positive eigenvalues. Using Cauchy–Schwarz and the eigenvalues/singular values of the blocks, we bound (2.4) from below by writing

$$v^T \mathscr{K} v \ge \mu_{\min}^A \|x\|^2 - 2\sigma_{\max}^B \|x\| \cdot \|y\| - \mu_{\max}^D \|y\|^2 - 2\sigma_{\max}^C \|y\| \cdot \|z\| + \mu_{\min}^E \|z\|^2$$

$$= \begin{bmatrix} \|x\| & \|y\| & \|z\| \end{bmatrix} \underbrace{\begin{bmatrix} \mu_{\min}^A & -\sigma_{\max}^B & 0 \\ -\sigma_{\max}^B & -\mu_{\max}^D & -\sigma_{\max}^C \\ 0 & -\sigma_{\max}^C & \mu_{\min}^E \end{bmatrix}}_{=:R} \begin{bmatrix} \|x\| \\ \|y\| \\ \|z\| \end{bmatrix}.$$

A lower bound on $\frac{v^T \mathscr{K} v}{v^T v}$ is therefore given by the smallest eigenvalue of R. Taking the characteristic polynomial of R yields the stated result.

Upper bound on negative eigenvalues. We derive an upper bound on the negative eigenvalues of $\mathscr{K}$ by finding a lower bound on the negative eigenvalues of $\mathscr{K}^{-1}$. We begin by partitioning $\mathscr{K}$ as

$$\mathscr{K} = \begin{bmatrix} A & B^T & 0 \\ B & -D & C^T \\ 0 & C & E \end{bmatrix} = \begin{bmatrix} \mathscr{K}_2 & \bar{C}^T \\ \bar{C} & E \end{bmatrix}, \tag{2.5}$$

where $\mathscr{K}_2 = \begin{bmatrix} A & B^T \\ B & -D \end{bmatrix}$ and $\bar{C} = \begin{bmatrix} 0 & C \end{bmatrix}$. By Benzi *et al.* (2005, Equation (3.4)),

$$\mathscr{K}^{-1} = \begin{bmatrix} \mathscr{K}_2^{-1} + \mathscr{K}_2^{-1} \bar{C}^T S_2^{-1} \bar{C} \mathscr{K}_2^{-1} & -\mathscr{K}_2^{-1} \bar{C} S_2^{-1} \\ -S_2^{-1} \bar{C} \mathscr{K}_2^{-1} & S_2^{-1} \end{bmatrix},$$

where $S_2 = E - \bar{C} \mathscr{K}_2^{-1} \bar{C}^T = E + C S_1^{-1} C^T$. Notice that

$$\mathscr{K}^{-1} = \begin{bmatrix} \mathscr{K}_2^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \mathscr{K}_2^{-1} \bar{C} \\ -I \end{bmatrix} S_2^{-1} \begin{bmatrix} \bar{C}^T \mathscr{K}_2^{-1} & -I \end{bmatrix}.$$

Because the second term is positive semidefinite, we conclude that the eigenvalues of $\mathscr{K}^{-1}$ are greater than or equal to the eigenvalues of $\begin{bmatrix} \mathscr{K}_2^{-1} & 0 \\ 0 & 0 \end{bmatrix}$. Thus, a lower bound on the negative eigenvalues of $\mathscr{K}_2^{-1}$ is also a lower bound on the negative eigenvalues of $\mathscr{K}^{-1}$. This means that an upper bound on the negative eigenvalues of $\mathscr{K}_2$ is also an upper bound on the negative eigenvalues of $\mathscr{K}$. The desired result now follows from Silvester & Wathen (1994, Lemma 2.2).

Lower bound on positive eigenvalues. We begin by noting that, because E is positive semidefinite, the eigenvalues of $\mathscr{K}$ are greater than or equal to those of

$$\mathscr{K}_{E=0} = \begin{bmatrix} A & B^T & 0 \\ B & -D & C^T \\ 0 & C & 0 \end{bmatrix}.$$

Moreover, $\mathscr{K}$ and $\mathscr{K}_{E=0}$ have the same inertia, by Lemma 2.1. Thus, the smallest positive eigenvalue of $\mathscr{K}$ is greater than or equal to the smallest positive eigenvalue of $\mathscr{K}_{E=0}$. We therefore obtain a lower bound on the positive eigenvalues of $\mathscr{K}$ by using energy estimates with $\mathscr{K}_{E=0}$. The eigenvalue problem associated with $\mathscr{K}_{E=0}$ is

$$\begin{bmatrix} A & B^T & 0 \\ B & -D & C^T \\ 0 & C & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{2.6}$$

The third block row of (2.6) gives $z = \frac{1}{\lambda} Cy$. The first block row gives

$$B^T y = (\lambda I - A)x.$$

We now consider two cases based on the value of λ.

Case I: $\lambda < \mu_{\min}^A$: If $\lambda < \mu_{\min}^A$, then $(\lambda I - A)$ is negative definite, so we can write

$$x = (\lambda I - A)^{-1} B^T y.$$

Substituting this into the second block row of (2.6), pre-multiplying by y^T and rearranging gives

$$\lambda y^T y = y^T B (\lambda I - A)^{-1} B^T y - y^T D y + \frac{1}{\lambda} y^T C^T C y \tag{2.7a}$$

$$\geq \frac{(\sigma_{\max}^B)^2}{\lambda - \mu_{\min}^A} y^T y - \mu_{\max}^D y^T y + \frac{(\sigma_{\min}^C)^2}{\lambda} y^T y. \tag{2.7b}$$

Dividing by $y^T y$, using the fact that $\lambda > 0$ and $\lambda - \mu_{\min}^A < 0$ and rearranging yields

$$\underbrace{\lambda^3 + (\mu_{\max}^D - \mu_{\min}^A)\lambda^2 - \left(\mu_{\min}^A \mu_{\max}^D + (\sigma_{\max}^B)^2 + (\sigma_{\min}^C)^2\right)\lambda + \mu_{\min}^A (\sigma_{\min}^C)^2}_{=p(\lambda)} \leq 0.$$

By applying Corollary 2.1 with $a = \mu_{\min}^A$, $b = \sigma_{\max}^B$, $c = \sigma_{\min}^C$ and $d = \mu_{\max}^D$, we know that this polynomial has two positive roots. Moreover, $p(\lambda)$ is negative between these two roots: this follows from the fact that $p(0) > 0$ and $\lim_{\lambda \to \infty} p(\lambda) = \infty$. Therefore, we conclude that in this case λ is greater than or equal to the smaller positive root of $p(\lambda)$, namely $p_{\min}^+$.

Case II: $\lambda \geq \mu_{\min}^A$: If $\lambda \geq \mu_{\min}^A$, then $(\lambda I - A)$ may be indefinite (and possibly singular). However, we can still obtain a lower bound for λ by observing that $\mu_{\min}^A$ is greater than $p_{\min}^+$. As stated earlier, $p(\lambda)$ has two positive roots and the value of $p(\lambda)$ is negative between those roots. Moreover, these are the only positive values of λ for which $p(\lambda) < 0$. We observe, after simplification, that

$$p(\mu_{\min}^A) = -(\sigma_{\max}^B)^2 \mu_{\min}^A < 0.$$

Therefore, the bound $\lambda \geq p_{\min}^+$ also holds in this case, which completes the proof. $\qquad\square$

REMARK 2.3 From Theorem 2.2, we see that when B and C are rank deficient, the internal bounds are zero. Similarly, when A is rank deficient, the lower positive bound is zero. Under mild conditions on the ranks and kernels of the matrix blocks, the statement of the theorem and its proof may be revised to obtain nonzero internal bounds in these cases. However, doing so is rather technical, and since the case of full rank A, B and C is common, further details on these end cases are omitted.

The matrix $\mathcal{K}_0$ is a special case of $\mathcal{K}$, and bounds on its eigenvalues can be obtained as a direct consequence of Theorem 2.2.

COROLLARY 2.2 (Eigenvalue bounds, matrix $\mathcal{K}_0$). Define the following three cubic polynomials as special cases of p, q and r defined in (2.2) with $D = E = 0$:

$$\hat{p}(\lambda) = \lambda^3 - \mu_{\min}^A \lambda^2 - \left((\sigma_{\max}^B)^2 + (\sigma_{\min}^C)^2 \right) \lambda + \mu_{\min}^A (\sigma_{\min}^C)^2; \tag{2.8a}$$

$$\hat{q}(\lambda) = \lambda^3 - \mu_{\max}^A \lambda^2 - \left((\sigma_{\max}^B)^2 + (\sigma_{\max}^C)^2 \right) \lambda + \mu_A^{\max} (\sigma_{\max}^C)^2; \tag{2.8b}$$

$$\hat{r}(\lambda) = \lambda^3 - \mu_{\min}^A \lambda^2 - \left((\sigma_{\max}^B)^2 + (\sigma_{\max}^C)^2 \right) \lambda + \mu_{\min}^A (\sigma_{\max}^C)^2. \tag{2.8c}$$

Using (2.8), the eigenvalues of $\mathcal{K}_0$ are bounded within the intervals

$$\left[\hat{r}^-, \frac{\mu_{\max}^A - \sqrt{(\mu_{\max}^A)^2 + 4(\sigma_{\min}^B)^2}}{2} \right] \bigcup \left[\hat{p}_{\min}^+, \hat{q}_{\max}^+ \right]. \tag{2.9}$$

The proof of Corollary 2.2 is omitted; it is similar to the proof of Theorem 2.2, with simplifications arising from setting $D = E = 0$.

2.3 *Tightness of the bounds*

While the tightness of the bounds we have obtained depends on the problem at hand, all bounds presented in this section are attainable, as we will demonstrate with small examples. For the extremal bounds (upper positive and lower negative), we note that both bounds hold for all double saddle-point matrices with $n = m = p = 1$, as the characteristic polynomial of the 3×3 matrix is the same as the polynomial given in the upper positive and lower negative bounds of Theorem 2.2.

For the upper bound on negative eigenvalues, we consider the example (with $n = m = 2$, and $p = 1$):

$$\mathcal{K} = \left[\begin{array}{cc|cc|c} \mu_{\max}^A & 0 & \sigma_{\min}^B & 0 & 0 \\ 0 & \mu_{\max}^A & 0 & \sigma_{\min}^B & 0 \\ \hline \sigma_{\min}^B & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\min}^B & 0 & -\mu^D & \sigma^C \\ \hline 0 & 0 & 0 & \sigma^C & \mu^E \end{array} \right].$$

We can permute the rows and columns of $\mathcal{K}$ to obtain a block diagonal matrix:

$$\mathcal{K}_D = \left[\begin{array}{cc|ccc} \mu_{\max}^A & \sigma_{\min}^B & 0 & 0 & 0 \\ \sigma_{\min}^B & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \mu_{\max}^A & \sigma_{\min}^B & 0 \\ 0 & 0 & \sigma_{\min}^B & -\mu^D & \sigma^C \\ 0 & 0 & 0 & \sigma^C & \mu^E \end{array} \right].$$

The upper left block of $\mathcal{K}_D$ has as an eigenvalue $\frac{\mu_{\max}^A - \sqrt{(\mu_{\max}^A)^2 + 4(\sigma_{\min}^B)^2}}{2}$, which is the bound given by Theorem 2.2.

Finally, for the lower positive bound, we consider the matrix (with $n = m = p = 2$)

$$
\mathcal{K} = \left[
\begin{array}{cc|cc|cc}
\mu^A_{\min} & 0 & \sigma^B_{\max} & 0 & 0 & 0 \\
0 & \mu^A_{\min} & 0 & \sigma^B_{\max} & 0 & 0 \\
\hline
\sigma^B_{\max} & 0 & -\mu^D_{\max} & 0 & \sigma^C_{\min} & 0 \\
0 & \sigma^B_{\max} & 0 & -\mu^D_{\max} & 0 & \sigma^C_{\min} \\
\hline
0 & 0 & \sigma^C_{\min} & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma^C_{\min} & 0 & \mu^E
\end{array}
\right].
$$

As in the previous example, we can permute the rows and columns to obtain a block diagonal matrix:

$$
\mathcal{K}_D = \left[
\begin{array}{ccc|ccc}
\mu^A_{\min} & \sigma^B_{\max} & 0 & 0 & 0 & 0 \\
\sigma^B_{\max} & -\mu^D_{\max} & \sigma^C_{\min} & 0 & 0 & 0 \\
0 & \sigma^C_{\min} & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & \mu^A_{\min} & \sigma^B_{\max} & 0 \\
0 & 0 & 0 & \sigma^B_{\max} & -\mu^D_{\max} & \sigma^C_{\min} \\
0 & 0 & 0 & 0 & \sigma^C_{\min} & \mu^E
\end{array}
\right].
$$

The characteristic polynomial of the upper left block is precisely $p(\lambda)$ defined in (2.2a); thus, the bound of Theorem 2.2 is obtained.

3. Eigenvalue bounds for $\mathcal{M}^{-1}\mathcal{K}$

We now derive bounds for the preconditioned matrix $\mathcal{M}^{-1}\mathcal{K}$, with $\mathcal{M}$ defined in (1.6). To ensure that the preconditioner $\mathcal{M}$ is well defined and invertible, we assume that A, S_1 and S_2 are positive definite.

3.1 *Inertia and eigenvalue multiplicity*

We begin with some observations on inertia and eigenvalue multiplicity. To simplify the presentation and proof of this result, we restrict ourselves to the case that B and C have full row rank. In some later results, we will lift this restriction to allow for rank-deficient B and C.

THEOREM 3.1 (Inertia and algebraic multiplicity, matrix $\mathcal{M}^{-1}\mathcal{K}$). Let $\mathcal{K}$ be defined as in (1.2) and $\mathcal{M}$ as in (1.6), and suppose that B and C have full row rank. The preconditioned matrix $\mathcal{M}^{-1}\mathcal{K}$ has:

(i) m negative eigenvalues;

(ii) p eigenvalues in $(0, 1)$;

(iii) $n - m$ eigenvalues equal to 1; and

(iv) m eigenvalues greater than 1.

Proof. Because $\mathcal{M}$ is symmetric positive definite, $\mathcal{M}^{1/2}$ exists and is invertible, and the inertias of $\mathcal{M}^{-1/2}\mathcal{K}\mathcal{M}^{-1/2}$ and $\mathcal{M}^{-1}\mathcal{K}$ are equal to the inertia of $\mathcal{K}$. Thus, Lemma 2.1 establishes that $\mathcal{M}^{-1}\mathcal{K}$ has $n + p$ positive and m negative eigenvalues, which proves (i).

For (ii)–(iv), we split the $n + p$ positive eigenvalues into eigenvalues less than, equal to, or greater than 1. For this, we compute the inertia of the shifted, split-preconditioned matrix

$$\mathscr{M}^{-1/2}\mathscr{K}\mathscr{M}^{-1/2} - I = \begin{bmatrix} 0 & \tilde{B}^T & 0 \\ \tilde{B} & -\tilde{D} - I & \tilde{C}^T \\ 0 & \tilde{C} & \tilde{E} - I \end{bmatrix},$$

where $\tilde{B} = S_1^{-1/2}BA^{-1/2}$, $\tilde{C} = S_2^{-1/2}CS_1^{-1/2}$, $\tilde{D} = S_1^{-1/2}DS_1^{-1/2}$, $\tilde{E} = S_2^{-1/2}ES_2^{-1/2}$ and I is an identity matrix of appropriate dimension (with slight abuse of notation, we use I to denote an identity matrix of any size). The number of positive, negative and zero eigenvalues of $\mathscr{M}^{-1/2}\mathscr{K}\mathscr{M}^{-1/2} - I$ will be equal to the number of eigenvalues of $\mathscr{M}^{-1}\mathscr{K}$ greater than, less than or equal to 1, respectively.

Noting that $\tilde{E} + \tilde{C}\tilde{C}^T = I$, we write

$$\mathscr{M}^{-1/2}\mathscr{K}\mathscr{M}^{-1/2} - I = \begin{bmatrix} 0 & \mathscr{B} \\ \mathscr{B}^T & \mathscr{T} \end{bmatrix},$$

where

$$\mathscr{T} := \begin{bmatrix} -\tilde{D} - I & \tilde{C}^T \\ \tilde{C} & -\tilde{C}\tilde{C}^T \end{bmatrix} \quad \text{and} \quad \mathscr{B} = \begin{bmatrix} \tilde{B}^T & 0 \end{bmatrix}.$$

The matrix $\mathscr{B}$ is in $\mathbb{R}^{n \times (m+p)}$ and has rank m. We define a matrix

$$N := \begin{bmatrix} \mathbf{0}_{m \times p} \\ I_p \end{bmatrix},$$

where $\mathbf{0}_{m \times p}$ is the $m \times p$ zero matrix and I_p is the $p \times p$ identity matrix. The columns of N form a basis for ker($\mathscr{B}$). Denote the inertia of M by $In(M) = (n_+, n_-, n_0)$. It is well known (see, e.g., Gould, 1985, Lemma 3.4) that

$$In(\mathscr{M}^{-1/2}\mathscr{K}\mathscr{M}^{-1/2} - I) = In(N^T\mathscr{T}N) + (m, m, n - m).$$

Because $N^T\mathscr{T}N = -\tilde{C}\tilde{C}^T$ is negative definite, this gives $In(\mathscr{M}^{-1/2}\mathscr{K}\mathscr{M}^{-1/2} - I) = (m, m+p, n-m)$, which yields (ii)–(iv). □

3.2 *Derivation of bounds*

It is possible to obtain eigenvalue bounds on the preconditioned system $\mathscr{M}^{-1}\mathscr{K}$ by using the results of Theorem 2.2 on the (symmetric) preconditioned system $\mathscr{M}^{-1/2}\mathscr{K}\mathscr{M}^{-1/2}$; however, some of the resulting bounds will be loose. The reason for this is that Theorem 2.2 assumes no relationships between the blocks of $\mathscr{K}$ and considers each block individually. In this section, we will derive tight eigenvalue bounds using energy estimates, which allow us to fully exploit the relationships between the blocks of the preconditioned system.

We begin by recalling a result that follows (after minor notational adjustments) from Horn & Johnson (1985, Theorem 7.7.3), which will be useful throughout the analysis that follows.

LEMMA 3.1 Let M and N be symmetric positive semidefinite matrices such that $M + N$ is positive definite. Then all eigenvalues of $(M + N)^{-1}M$ are in $[0, 1]$.

The following result follows directly from Sogn & Zulehner (2018, Lemma A.1) and simplifies the presentation of some of the subsequent results in this section.

LEMMA 3.2 The roots of the cubic polynomial $\lambda^3 - \lambda^2 - 2\lambda + 1$ are given by

$$\lambda \in \left\{ 2\cos\left(\frac{\pi}{7}\right), \ 2\cos\left(\frac{3\pi}{7}\right), \ 2\cos\left(\frac{5\pi}{7}\right) \right\},$$

which are approximately equal to 1.8019, 0.4450 and -1.2470, respectively.

The case $D = E = 0$

When D and E are both zero, $\mathcal{M}^{-1}\mathcal{K}$ has six distinct eigenvalues given by: 1, $\frac{1\pm\sqrt{5}}{2}$, $2\cos\left(\frac{\pi}{7}\right)$, $2\cos\left(\frac{3\pi}{7}\right)$ and $2\cos\left(\frac{5\pi}{7}\right)$. The last three are the roots of the cubic polynomial $\lambda^3 - \lambda^2 - 2\lambda + 1$, per Lemma 3.2. The proof follows directly from Sogn & Zulehner (2018, Theorem 2.3). See also Cai *et al.* (2009); Huang & Ma (2019) for later proofs.

The case $D = 0$ and $E \succeq 0$

When $D = 0$, it is necessary that B have full row rank in order for S_1 to be invertible; however, C may be rank deficient. Suppose that C^T has nullity k. The following result holds.

THEOREM 3.2 (Eigenvalue bounds, matrix $\mathcal{M}^{-1}\mathcal{K}$, $D = 0$, $E \succeq 0$). When $D = 0$ and $E \succeq 0$, the eigenvalues of $\mathcal{M}^{-1}\mathcal{K}$ are given by: $\lambda = 1$ with multiplicity $n - m + k$; $\lambda = \frac{1\pm\sqrt{5}}{2}$, each with multiplicity $m - p + k$; and $p - k$ eigenvalues located in each of the three intervals:

- $I_1 = \left[2\cos\left(\frac{5\pi}{7}\right), \frac{1-\sqrt{5}}{2} \right) \approx [-1.618, -0.618)$

- $I_2 = \left[2\cos\left(\frac{3\pi}{7}\right), 1 \right) \approx [0.4450, 1)$

- $I_3 = \left(\frac{1+\sqrt{5}}{2}, 2\cos\left(\frac{\pi}{7}\right) \right] \approx (1.618, 1.8019]$.

Proof. We write out the (left-)preconditioned operator

$$\mathcal{M}^{-1}\mathcal{K} = \begin{bmatrix} I & A^{-1}B^T & 0 \\ S_1^{-1}B & 0 & S_1^{-1}C^T \\ 0 & S_2^{-1}C & S_2^{-1}E \end{bmatrix},$$

with corresponding eigenvalue equations

$$x + A^{-1}B^T y = \lambda x; \tag{3.1a}$$

$$S_1^{-1}Bx + S_1^{-1}C^T z = \lambda y; \tag{3.1b}$$

$$S_2^{-1}Cy + S_2^{-1}Ez = \lambda z. \tag{3.1c}$$

We obtain $n - m$ eigenvectors for $\lambda = 1$ by choosing $x \in \ker(B)$ and $y, z = 0$. By considering $y \in \ker(C)$ and $z = 0$, we obtain eigenvalues $\lambda = \frac{1 \pm \sqrt{5}}{2}$, each with geometric multiplicity $m - p + k$.

For the remaining eigenvalues, we assume that $z \neq 0$ and $\lambda \notin \{1, \frac{1 \pm \sqrt{5}}{2}\}$. From (3.1a), we obtain

$$x = \frac{1}{\lambda - 1} A^{-1} B^T y,$$

which we substitute into (3.1b) and rearrange to get

$$y = \frac{\lambda - 1}{\lambda^2 - \lambda - 1} S_1^{-1} C^T z.$$

Substituting this into (3.1c) gives

$$\frac{\lambda - 1}{\lambda^2 - \lambda - 1} S_2^{-1} C S_1^{-1} C^T z + S_2^{-1} E z = \lambda z. \tag{3.2}$$

Because $S_2 = E + C S_1^{-1} C^T$, we can write $S_2^{-1} E = I - S_2^{-1} C S_1^{-1} C^T$. We substitute this into (3.2) and rearrange to obtain

$$\left(-\lambda^2 + 2\lambda \right) S_2^{-1} C S_1^{-1} C^T z = \left(\lambda^3 - 2\lambda^2 + 1 \right) z.$$

Let $\{\delta_j, v_j\}$, for $1 \leq j \leq p$, denote an eigenpair of $S_2^{-1} C S_1^{-1} C^T$. For k of these eigenpairs corresponding to $v_j \in \ker(C^T)$, we note that

$$S_2^{-1} E v_j = (I - S_2^{-1} C S_1^{-1} C^T) v_j = v_j,$$

and therefore $\begin{bmatrix} 0 & 0 & v_j^T \end{bmatrix}^T$ is an eigenvector of $\mathcal{M}^{-1} \mathcal{K}$ with $\lambda = 1$. For the $p - k$ remaining eigenpairs, we have $0 < \delta_j \leq 1$ (by Lemma 3.1). We can then write

$$\left(-\lambda_j^2 + 2\lambda_j \right) \delta_j z_j = \left(\lambda_j^3 - 2\lambda_j^2 + 1 \right) z_j,$$

where λ_j corresponds to an eigenpair $\{\delta_j, z_j\}$ of $S_2^{-1} C S_1^{-1} C^T$. Because $z_j \neq 0$ this implies that

$$\lambda_j^3 - (2 - \delta_j)\lambda_j^2 - 2\delta_j \lambda_j + 1 = 0. \tag{3.3}$$

Thus, each of the $p - k$ positive eigenvalues δ_j of $S_2^{-1} C S_1^{-1} C^T$ yields three distinct corresponding eigenvalues $\lambda_j^{(1)}, \lambda_j^{(2)}$ and $\lambda_j^{(3)}$ of $\mathcal{M}^{-1} \mathcal{K}$, corresponding to the roots of the cubic polynomial (3.3).

These $3(p-k)$ eigenvalues, combined with the eigenvalues described earlier, account for all eigenvalues of $\mathscr{M}^{-1}\mathscr{K}$. Using continuity arguments, the extremum must be obtained at the endpoints of the range for δ_j. Substituting $\delta_j = 0$ and $\delta_j = 1$ into (3.3) and using Lemma 3.2 gives us the three intervals for these eigenvalues stated in the theorem. □

The case $D \succeq 0$ and $E = 0$

When $D \succeq 0$ and $E = 0$, the bounds are the same as when D and E are both nonzero, which are given next.

The case $D, E \succeq 0$

THEOREM 3.3 (Eigenvalue bounds, matrix $\mathscr{M}^{-1}\mathscr{K}$, $D, E \succeq 0$). The eigenvalues of $\mathscr{M}^{-1}\mathscr{K}$ are bounded within the intervals

$$\left[-\frac{1+\sqrt{5}}{2}, \frac{1-\sqrt{5}}{2} \right] \cup \left[2\cos\left(\frac{3\pi}{7}\right), 2\cos\left(\frac{\pi}{7}\right) \right],$$

which are approximately $[-1.618, -0.618] \cup [0.4450, 1.8019]$.

REMARK 3.4 A proof for the upper bound on the positive eigenvalues is presented in Sogn & Zulehner (2018, Theorem 2.1), and one for the lower bound on the positive eigenvalues is given in Sogn & Zulehner (2018, Theorem 2.2). A different proof for the four bounds appears in Pearson & Potschka (2021b, Theorem 5.3). We provide below an alternative technique of proof based on energy estimates.

Proof. Upper bound on positive eigenvalues. We know from Theorem 3.1 that the upper bound on positive eigenvalues is greater than 1, so we assume here that $\lambda > 1$. We begin by writing the eigenvalue equations as

$$Ax + B^T y = \lambda Ax; \tag{3.4a}$$

$$Bx - Dy + C^T z = \lambda S_1 y; \tag{3.4b}$$

$$Cy + Ez = \lambda S_2 z. \tag{3.4c}$$

From (3.4a), we get

$$x = \frac{1}{\lambda - 1} A^{-1} B^T y, \tag{3.5}$$

and from (3.4c), we get $z = (\lambda S_2 - E)^{-1} Cy$ (because $\lambda > 1$, we are guaranteed that $\lambda S_2 - E = (\lambda - 1)E + \lambda C S_1^{-1} C^T$ is positive definite). Substituting these values back into (3.4b) and pre-multiplying by y^T gives

$$\left(\frac{1}{\lambda - 1} \right) y^T B A^{-1} B^T y + y^T (D + \lambda S_1) y - y^T C^T (\lambda S_2 - E)^{-1} Cy = 0.$$

Recalling that $S_1 = D + BA^{-1}B^T$ we can rewrite this as

$$\left(\frac{1}{1-\lambda} + \lambda\right) y^T S_1 y + \left(1 - \frac{1}{1-\lambda}\right) y^T D y - y^T C^T (\lambda S_2 - E)^{-1} C y = 0. \tag{3.6}$$

Because $\lambda > 1$, we have

$$y^T C^T (\lambda S_2 - E)^{-1} C y \le y^T C^T \left(\lambda(S_2 - E)\right)^{-1} C y$$
$$= \frac{1}{\lambda} y^T C^T \left(C S_1^{-1} C^T\right)^{-1} C y.$$

Therefore, (3.6) gives

$$\left(\frac{1}{1-\lambda} + \lambda\right) y^T S_1 y + \left(1 - \frac{1}{1-\lambda}\right) y^T D y - \frac{1}{\lambda} y^T C^T (C S_1^{-1} C^T)^{-1} C y \le 0. \tag{3.7}$$

Next, we let $\tilde{y} = S_1^{1/2} y$ and rewrite the first and third terms in (3.7) in terms of $\tilde{y}$:

$$\left(\frac{1}{1-\lambda} + \lambda\right) \tilde{y}^T \tilde{y} + \left(1 - \frac{1}{1-\lambda}\right) y^T D y - \frac{1}{\lambda} \tilde{y}^T \underbrace{S_1^{-1/2} C^T (C S_1^{-1} C^T)^{-1} C S_1^{-1/2}}_{:=P} \tilde{y} \le 0. \tag{3.8}$$

Because P defined in (3.8) is an orthogonal projector, we have $\tilde{y}^T P \tilde{y} \le \tilde{y}^T \tilde{y}$. And because $\lambda > 1$, we have $1 - \frac{1}{1-\lambda} > 0$, which means that the second term $\left(1 - \frac{1}{1-\lambda}\right) y^T D y$ is non-negative and can be dropped from the inequality. The inequality (3.8) therefore becomes

$$\left(\frac{1}{1-\lambda} + \lambda\right) \tilde{y}^T \tilde{y} - \frac{1}{\lambda} \tilde{y}^T y \le 0,$$

which we can divide by $\tilde{y}^T \tilde{y}$ rearrange to give

$$\lambda^3 - \lambda^2 - 2\lambda + 1 \le 0.$$

This, combined with the assumption that $\lambda > 1$, gives us the stated result that λ is less than or equal to the largest root of $\lambda^3 - \lambda^2 - 2\lambda + 1$, which is defined in Lemma 3.2.

Lower bound on negative eigenvalues. We begin from (3.6) and note that when $\lambda < 0$, by similar reasoning as was shown for the upper bound on positive eigenvalues,

$$\left(\frac{1}{1-\lambda} + \lambda\right) y^T S_1 y + \left(1 - \frac{1}{1-\lambda}\right) y^T D y - \frac{1}{\lambda} y^T C^T (C S_1^{-1} C^T)^{-1} C y \ge 0.$$

Rewriting the inequality in terms of $\tilde{y} = S_1^{1/2}y$ gives

$$\left(\frac{1}{1-\lambda} + \lambda\right)\tilde{y}^T\tilde{y} + \left(1 - \frac{1}{1-\lambda}\right)\tilde{y}^T S_1^{-1/2}DS_1^{-1/2}\tilde{y} - \frac{1}{\lambda}\tilde{y}^T P\tilde{y} \geq 0,$$

where P is the orthogonal projector defined in (3.8). For the second term, note that $S_1^{-1/2}DS_1^{-1/2}$ is similar to $S_1^{-1}D$, which has all eigenvalues between 0 and 1 (by Lemma 3.1). Thus, both $\tilde{y}^T S_1^{-1/2}DS_1^{-1/2}\tilde{y}$ and $\tilde{y}^T P\tilde{y}$ are less than or equal to $\tilde{y}^T\tilde{y}$ and we can therefore write

$$\left(\frac{1}{1-\lambda} + \lambda\right)\tilde{y}^T\tilde{y} + \left(1 - \frac{1}{1-\lambda}\right)\tilde{y}^T\tilde{y} - \frac{1}{\lambda}\tilde{y}^T\tilde{y} \geq 0,$$

which, after dividing by $\tilde{y}^T\tilde{y}$ and simplifying, yields

$$\lambda^2 + \lambda - 1 \leq 0.$$

This along with the assumption that $\lambda < 0$ gives the desired bound of $\lambda \geq -\frac{1+\sqrt{5}}{2}$.

Lower bound on positive eigenvalues. Assume that $0 < \lambda < 1$ (we know from Theorem 3.1 that the lower bound is in this interval). Substituting (3.5) into (3.4b) and solving for y gives

$$y = \underbrace{\left(\frac{1}{1-\lambda}BA^{-1}B^T + D + \lambda S_1\right)^{-1}}_{=:Q} C^T z. \tag{3.9}$$

When $0 < \lambda < 1$, the value $\frac{1}{1-\lambda}$ is positive, so we are guaranteed that Q in (3.9) is positive definite. If $z \in \ker(C^T)$, (3.9) gives $y = 0$ and (3.5) gives $x = 0$, and we can see from the eigenvalue equations 3.4a3.4c that this eigenvector corresponds to $\lambda = 1$ (because $Ez = S_2z$ for $z \in \ker(C^T)$). This contradicts our assumption that $\lambda < 1$; thus, we assume $z \notin \ker(C^T)$.

We can then write (3.4c) as

$$C\left(\frac{1}{1-\lambda}BA^{-1}B^T + D + \lambda S_1\right)^{-1}C^T z + (1-\lambda)Ez - \lambda CS_1^{-1}C^T z = 0. \tag{3.10}$$

When $0 < \lambda < 1$, we have $\frac{1}{1-\lambda} > 1$. Therefore, if we take the inner product of z^T with (3.10), replace D by $\frac{1}{1-\lambda}D$ and drop the non-negative term $(1-\lambda)z^T Ez$, we obtain the inequality

$$z^T C\left(\left(\frac{1}{1-\lambda} + \lambda\right)S_1\right)^{-1}C^T z - \lambda z^T CS_1^{-1}C^T z \leq 0.$$

After simplifying and dividing by $z^T CS_1^{-1}C^T z$, we obtain

$$\lambda^3 - \lambda^2 - 2\lambda + 1 \leq 0. \tag{3.11}$$

This, combined with the assumption that $0 < \lambda < 1$, gives us that λ must be greater than or equal to the smaller positive root of (3.11), as required.

Upper bound on negative eigenvalues. Assume that $\lambda < 0$. We begin from Equation (3.6) and note that $\lambda S_2 - E$ is negative definite. Therefore,

$$\left(\frac{1}{1-\lambda} + \lambda\right) y^T S_1 y + \left(1 - \frac{1}{1-\lambda}\right) y^T D y \leq 0.$$

Since $1 - \frac{1}{1-\lambda} > 0$, the $y^T D y$ term is non-negative and can be dropped while keeping the inequality, giving us

$$\left(\frac{1}{1-\lambda} + \lambda\right) y^T S_1 y \leq 0.$$

We thus require

$$\frac{1}{1-\lambda} + \lambda \leq 0 \text{ with } \lambda < 0,$$

which leads to the desired bound $\lambda \leq \frac{1-\sqrt{5}}{2}$. □

REMARK 3.5 The bounds in Theorem 3.3 are different than the bounds of Theorem 3.2: the lower bound on the negative eigenvalues is looser, and the inclusion set for positive eigenvalues is a single interval that strictly contains and is larger than the union of the two positive intervals in Theorem 3.2, $I_2 \cup I_3$. While the lower bound on negative eigenvalues of Theorem 3.2 may be obtained in the proof of Theorem 3.3 by assuming $D = 0$, the two positive intervals in Theorem 3.2 are obtained thanks to the specific technique of proof we use for that theorem, and we believe that they cannot easily be obtained from the proof of Theorem 3.3 or by other means.

4. Bounds for preconditioners with approximations of Schur complements

In practice, it is too expensive to invert A, S_1 and S_2 exactly. In this section, we examine eigenvalue bounds on matrices of the form $\tilde{\mathcal{M}}^{-1}\mathcal{K}$, where $\tilde{\mathcal{M}}$ uses symmetric positive definite and ideally spectrally equivalent approximations for A, S_1 and S_2. Specifically, we consider the approximate block diagonal preconditioner

$$\tilde{\mathcal{M}} = \begin{bmatrix} \tilde{A} & 0 & 0 \\ 0 & \tilde{S}_1 & 0 \\ 0 & 0 & \tilde{S}_2 \end{bmatrix}, \tag{4.1}$$

with $\tilde{A}$, $\tilde{S}_1$ and $\tilde{S}_2$ satisfying the following:

ASSUMPTION 4.1 Let $\Lambda(\cdot)$ denote the spectrum of a matrix. The diagonal blocks of the approximate preconditioner $\tilde{\mathcal{M}}$, given by $\tilde{A}$, $\tilde{S}_1$ and $\tilde{S}_2$, satisfy:

(i) $\Lambda(\tilde{A}^{-1}A) \in [\alpha_0, \beta_0]$;

(ii) $\Lambda(\tilde{S}_1^{-1}S_1) \in [\alpha_1, \beta_1]$;

(iii) $\Lambda(\tilde{S}_2^{-1}S_2) \in [\alpha_2, \beta_2]$,

where $0 < \alpha_i \leq 1 \leq \beta_i$.

We note that to obtain spectral equivalence we seek approximations that yield values of α_i independent of the mesh size and bounded uniformly away from zero. It is also worth noting that the values of α_i and β_i are typically not explicitly available. We briefly address this at the end of this section, following our derivation of the bounds.

To simplify our analyses, we define:

$$\tilde{A}^{-1}A =: Q_0; \tag{4.2a}$$

$$\tilde{S}_1^{-1}S_1 =: Q_1; \tag{4.2b}$$

$$\tilde{S}_2^{-1}S_2 =: Q_2. \tag{4.2c}$$

To derive the eigenvalue bounds, we consider the split preconditioned matrix

$$\tilde{\mathcal{M}}^{-1/2}\mathcal{K}\tilde{\mathcal{M}}^{-1/2} = \begin{bmatrix} \tilde{Q}_0 & \tilde{B}^T & 0 \\ \tilde{B} & -\tilde{D} & \tilde{C}^T \\ 0 & \tilde{C} & \tilde{E} \end{bmatrix}, \tag{4.3}$$

where $\tilde{Q}_0 = \tilde{A}^{-1/2}A\tilde{A}^{-1/2}$, $\tilde{B} = \tilde{S}_1^{-1/2}B\tilde{A}^{-1/2}$, $\tilde{C} = \tilde{S}_2^{-1/2}C\tilde{S}_1^{-1/2}$, $\tilde{D} = \tilde{S}_1^{-1/2}D\tilde{S}_1^{-1/2}$ and $\tilde{E} = \tilde{S}_2^{-1/2}E\tilde{S}_2^{-1/2}$. We proceed by bounding the eigenvalues and singular values of the blocks of $\tilde{\mathcal{M}}^{-1/2}\mathcal{K}\tilde{\mathcal{M}}^{-1/2}$, and then applying the results for general double saddle-point matrices presented in Section 2.2. In order to avoid providing internal eigenvalue bounds equal to zero (see Remark 2.3), we assume here that B and C have full row rank.

LEMMA 4.1   When B and C have full row rank, bounds on the eigenvalues/singular values of the blocks of the matrix $\tilde{\mathcal{M}}^{-1/2}\mathcal{K}\tilde{\mathcal{M}}^{-1/2}$ are as follows:

- $\tilde{Q}_0$: eigenvalues are in $[\alpha_0, \beta_0]$.

- $\tilde{B}$: singular values are in $\left[\sqrt{\frac{\alpha_0\alpha_1}{1+\eta_D}}, \sqrt{\beta_0\beta_1}\right]$, where η_D is the maximal eigenvalue of $(BA^{-1}B^T)^{-1}D$.

- $\tilde{C}$: singular values are in $\left[\sqrt{\frac{\alpha_1\alpha_2}{1+\eta_E}}, \sqrt{\beta_1\beta_2}\right]$, where η_E is the maximal eigenvalue of $(CS_1^{-1}C^T)^{-1}E$.

- $\tilde{D}$: eigenvalues are in $[0, \beta_1]$.

- $\tilde{E}$: eigenvalues are in $[0, \beta_2]$.

Proof.   The eigenvalue bounds on $\tilde{Q}_0$ follow from the fact that $\tilde{Q}_0$ is similar to Q_0. For $\tilde{D}$ and $\tilde{E}$, the lower bounds follow from the fact that D and E are semidefinite (if D and/or E are definite, this bound will be loose, but we use the zero bound to simplify some results that we present in this section). For the upper bound on $\tilde{D}$, we note that $\tilde{D} = \tilde{S}_1^{-1/2}D\tilde{S}_1^{-1/2}$, which is similar to $\tilde{S}_1^{-1}D = Q_1S_1^{-1}D$. The

eigenvalues of $S_1^{-1}D$ are less than or equal to 1 by Lemma 3.1, meaning that those of $Q_1 S_1^{-1}D$ are less than or equal to β_1. Analogous reasoning gives the upper bound for $\tilde{E}$.

We now present the results for $\tilde{B}$. For the upper bound, note that the matrix $\tilde{B}\tilde{B}^T = \tilde{S}_1^{-1/2} B\tilde{A}^{-1}B^T \tilde{S}_1^{-1/2}$ is similar to $\tilde{S}_1^{-1} B\tilde{A}^{-1}B^T = Q_1 S_1^{-1} B\tilde{A}^{-1}B^T$. Because the eigenvalues of Q_1 are in $[\alpha_1, \beta_1]$, we need only bound the eigenvalues of $S_1^{-1} B\tilde{A}^{-1}B^T$. These are the same as the nonzero eigenvalues of

$$\tilde{A}^{-1}B^T S_1^{-1}B = Q_0 A^{-1}B^T S_1^{-1}B.$$

The nonzero eigenvalues of $A^{-1}B^T S_1^{-1}B$ are the same as those of $S_1^{-1}BA^{-1}B^T$, which are all less than or equal to 1 by Lemma 3.1. Thus, the eigenvalues of $Q_0 A^{-1}B^T S_1^{-1}B$ are less than or equal to β_0, from which we conclude that the eigenvalues of $\tilde{B}\tilde{B}^T$ are less than or equal to $\beta_0\beta_1$, giving an upper singular value bound of $\sqrt{\beta_0\beta_1}$. Similarly, a lower bound on the eigenvalues of $\tilde{B}\tilde{B}^T$ is given by $\alpha_0\alpha_1$ times a lower bound on the eigenvalues of $S_1^{-1}BA^{-1}B^T$. Because we have assumed that B is full rank, $BA^{-1}B^T$ is invertible, implying that

$$S_1^{-1}BA^{-1}B^T = \left((BA^{-1}B^T)^{-1}S_1\right)^{-1} = \left(I + (BA^{-1}B^T)^{-1}D\right)^{-1}.$$

The stated result then follows because the eigenvalues of $S_1^{-1}BA^{-1}B^T$ are greater than or equal to $\frac{1}{1+\eta_D}$, where η_D is the maximal eigenvalue of $(BA^{-1}B^T)^{-1}D$. Thus, a lower bound on the singular values of $\tilde{B}$ is given by the square root of this value.

The bounds for $\tilde{C}$ are obtained in the same way as those for $\tilde{B}$.   □

We can now present bounds on the eigenvalues of $\tilde{\mathscr{M}}^{-1/2}\mathscr{K}\tilde{\mathscr{M}}^{-1/2}$. We define three cubic polynomials:

$$u(\lambda) = \lambda^3 + (\beta_1 - \alpha_0)\lambda^2 - \left(\alpha_0\beta_1 + \frac{\alpha_1\alpha_2}{1+\eta_E} + \beta_0\beta_1\right)\lambda + \frac{\alpha_0\alpha_1\alpha_2}{1+\eta_E}; \qquad (4.4a)$$

$$v(\lambda) = \lambda^3 - (\beta_0 + \beta_2)\lambda^2 + (\beta_0\beta_2 - \beta_0\beta_1 - \beta_1\beta_2)\lambda + 2\beta_0\beta_1\beta_2; \qquad (4.4b)$$

$$w(\lambda) = \lambda^3 + (\beta_1 - \alpha_0)\lambda^2 - (\alpha_0\beta_1 + \beta_0\beta_1 + \beta_1\beta_2)\lambda + \alpha_0\beta_1\beta_2. \qquad (4.4c)$$

These polynomials all have two positive roots and one negative root, by Corollary 2.1. As in Section Section 2, we let u^- denote the (single) negative root of a polynomial u, and let $u_{\min}^+$ and $u_{\max}^+$ respectively denote the smallest and largest positive roots.

THEOREM 4.1 (Eigenvalue bounds, matrix $\tilde{\mathscr{M}}^{-1/2}\mathscr{K}\tilde{\mathscr{M}}^{-1/2}$). When B and C have full row rank, the eigenvalues of $\tilde{\mathscr{M}}^{-1/2}\mathscr{K}\tilde{\mathscr{M}}^{-1/2}$ are bounded within the intervals

$$\left[w^-, \frac{\beta_0 - \sqrt{\beta_0^2 + \frac{4\alpha_0\alpha_1}{1+\eta_D}}}{2}\right] \bigcup \left[u_{\min}^+, v_{\max}^+\right]. \qquad (4.5)$$

Proof. The stated bounds follow from Lemma 4.1 and Theorem 2.2.   □

REMARK 4.2 Obtaining η_D and η_E requires computation of eigenvalues of two matrices related to the Schur complements S_1 and S_2: $(S_1 - D)^{-1}D = (BA^{-1}B^T)^{-1}D$ and $(S_2 - E)^{-1}E = (CS_1^{-1}C^T)^{-1}E$, respectively. As we have previously mentioned, in practice when solving (1.1) the matrices S_1 and S_2 would typically be approximated by sparse and easier-to-invert matrices, rather than formed and computed explicitly. Therefore, we cannot expect to compute η_D and η_E exactly. Spectral equivalence relations may be helpful in providing reasonable approximations here. For example, in common formulations of the Stokes–Darcy problem (Cai *et al.*, 2021), the matrix A is a discrete negative Laplacian and B and C are discrete divergence operators (or scaled variations thereof). In such a case, for certain finite element discretizations $BA^{-1}B^T = S_1 - D$ is spectrally equivalent to the mass matrix in the pressure space (Elman *et al.*, 2005, Section 5.5). Denote it by Q. Then, estimating the maximal eigenvalue of η_D amounts to computing an approximation to the maximal eigenvalue of $Q^{-1}D$, which is computationally straightforward given the favorable spectral properties of Q. If the above-mentioned Schur complement is approximated by the mass matrix, then it can be shown that $CS_1^{-1}C^T = S_2 - E$ is strongly related to the scalar Laplacian, and therefore maximal eigenvalue of η_E would be relatively easy to approximate as well.

When $D = 0$, η_D and the maximal eigenvalue of $\tilde{D}$ are zero. Similarly, when $E = 0$, η_E and the maximal eigenvalue of $\tilde{E}$ are zero. In these cases, we can simplify some of the bounds of Theorem 4.1. Some of the cubic polynomials that define the bounds will change in these cases. We will use a bar (e.g., $\bar{u}$) to denote cubic polynomials where $D = 0$, a hat (e.g., $\hat{u}$) to denote the polynomials where $E = 0$ and both (e.g., $\hat{\bar{u}}$) to denote both D and E being zero. We define the following cubic polynomials:

$$\bar{u}(\lambda) = \lambda^3 - \alpha_0\lambda^2 - \left(\frac{\alpha_1\alpha_2}{1 + \eta_E} + \beta_0\beta_1\right)\lambda + \frac{\alpha_0\alpha_1\alpha_2}{1 + \eta_E}; \tag{4.6a}$$

$$\hat{u}(\lambda) = \lambda^3 + (\beta_1 - \alpha_0)\lambda^2 - (\alpha_0\beta_1 + \alpha_1\alpha_2 + \beta_0\beta_1)\lambda + \alpha_0\alpha_1\alpha_2; \tag{4.6b}$$

$$\hat{\bar{u}}(\lambda) = \lambda^3 - \alpha_0\lambda^2 - (\alpha_1\alpha_2 + \beta_0\beta_1)\lambda + \alpha_0\alpha_1\alpha_2; \tag{4.6c}$$

$$\hat{v}(\lambda) = \lambda^3 - \beta_0\lambda^2 - (\beta_0\beta_1 + \beta_1\beta_2)\lambda + \beta_0\beta_1\beta_2; \tag{4.6d}$$

$$\bar{w}(\lambda) = \lambda^3 - \alpha_0\lambda^2 - (\beta_0\beta_1 + \beta_1\beta_2)\lambda + \alpha_0\beta_1\beta_2. \tag{4.6e}$$

The following results then follow directly from Theorem 4.1; the proof is omitted.

COROLLARY 4.1 **In the case $D = 0$, $E \succeq 0$:** the eigenvalues of $\mathscr{M}^{-1/2}\mathscr{K}\mathscr{M}^{-1/2}$ are bounded within the intervals

$$\left[\bar{w}^-, \frac{\beta_0 - \sqrt{\beta_0^2 + 4\alpha_0\alpha_1}}{2}\right] \bigcup [\bar{u}_{\min}^+, v_{\max}^+]. \tag{4.7}$$

In the case $D \succeq 0$, $E = 0$: the eigenvalues are bounded in

$$\left[w^-, \frac{\beta_0 - \sqrt{\beta_0^2 + \frac{4\alpha_0\alpha_1}{1+\eta_D}}}{2}\right] \bigcup [\hat{u}_{\min}^+, \hat{v}_{\max}^+]. \tag{4.8}$$

In the case $D = 0$, $E = 0$: the eigenvalues are bounded in

$$\left[\bar{w}^-, \frac{\beta_0 - \sqrt{\beta_0^2 + 4\alpha_0\alpha_1}}{2} \right] \bigcup \left[\hat{\bar{u}}_{\min}^+, \hat{v}_{\max}^+ \right]. \tag{4.9}$$

REMARK 4.3 There is some looseness in the bounds of Theorem 4.1 when applied to the exact preconditioning case (i.e., $\alpha_i = \beta_i = 1$). This is a consequence of the fact that Theorem 4.1 is based on the bounds for unpreconditioned matrices, which consider each matrix block individually, as opposed to the energy estimates approach in Section 3, which fully exploits the relationships between the blocks of the preconditioned matrix. For the extremal (lower negative and upper positive) bounds, the looseness is minor: Theorem 4.1 gives a bound of 2 for the positive eigenvalues and approximately -1.9 on the negative eigenvalues, while we know from Theorem 3.3 that tight bounds are approximately 1.8 and -1.6, respectively. For the interior bounds, however, we note that the bounds may be quite loose if η_D or η_E is very large. Fortunately, this is often not a concern in practical settings as having a large η_D or η_E generally implies that the spectral norm of D or E is large relative to that of $BA^{-1}B^T$ or $CS_1^{-1}C^T$, respectively. Often, D and/or E are regularization terms, which tend to have a fairly small norm. Nonetheless, it should be acknowledged that the interior bounds may be pessimistic for some problems.

REMARK 4.4 The values of α_i and β_i are rarely available in practical applications; they are typically coercivity constants or related quantities, which are proven to be independent of the mesh size, but are not known explicitly. The values of β_i should not typically generate a difficulty, and approximating them by 1 or a value close to 1 should provide a reasonable approximation for the bounds.

For the α_i, let us provide some (partial) observations. For the solutions of the quadratic equations in Corollary 4.1, we can use the Taylor approximation $\sqrt{1+x} \lessgtr 1 + \frac{x}{2}$ for $0 < x \ll 1$ to conclude that

$$\frac{\beta_0 - \sqrt{\beta_0^2 + 4\alpha_0\alpha_1}}{2} \gtrless -\frac{\alpha_0\alpha_1}{\beta_0}.$$

This means that if α_0 and α_1 are small, then the above displayed expression would be a valid (albeit slightly less tight) upper negative bound, and if the α_i are uniformly bounded away from zero, then so is the bound.

5. Numerical experiments

In this section, we consider two slightly different variants of a Poisson control problem as in Rees *et al.* (2010). First, we consider a distributed control problem with Dirichlet boundary conditions:

$$\min_{u,f} \frac{1}{2}||u - \hat{u}||_{L_2(\Omega)}^2 + \frac{\beta}{2}||f||_{L_2(\Omega)}^2 \tag{5.1a}$$

$$\text{s.t.} \quad -\nabla^2 u = f \ \text{ in } \Omega, \tag{5.1b}$$

$$u = g \ \text{ on } \partial\Omega, \tag{5.1c}$$

where u is the state, $\hat{u}$ is the desired state, $0 < \beta \ll 1$ is a regularization parameter, f is the control and Ω is the domain with boundary $\partial\Omega$. Upon discretization, we obtain the system

$$\underbrace{\begin{bmatrix} M & K & 0 \\ K & 0 & -M \\ 0 & -M & \beta M \end{bmatrix}}_{=:\mathscr{K}} \begin{bmatrix} u_h \\ \lambda \\ f_h \end{bmatrix} = \begin{bmatrix} b \\ d \\ 0 \end{bmatrix}, \tag{5.2}$$

where M is a symmetric positive definite mass matrix and K is a symmetric positive definite discrete Laplacian. All blocks of $\mathscr{K}$ are square (i.e., $n = m = p$).

As a second experiment we consider a boundary control problem:

$$\min_{u,f} \frac{1}{2}||u - \hat{u}||^2_{L_2(\Omega)} + \frac{\beta}{2}||g||^2_{L_2(\Omega)} \tag{5.3a}$$

$$\text{s.t.} \ -\nabla^2 u = 0 \ \text{in} \ \Omega, \tag{5.3b}$$

$$\frac{\partial u}{\partial n} = g \ \text{on} \ \partial\Omega, \tag{5.3c}$$

which after discretization yields the linear system

$$\underbrace{\begin{bmatrix} M & K & 0 \\ K & 0 & -E \\ 0 & -E^T & \beta M_b \end{bmatrix}}_{=:\mathscr{K}_\partial} \begin{bmatrix} u_h \\ \lambda \\ g_h \end{bmatrix} = \begin{bmatrix} b_\partial \\ d_\partial \\ 0 \end{bmatrix}, \tag{5.4}$$

where $M_b \in \mathbb{R}^{n_b \times n_b}$ (with $n_b < n$) is a boundary mass matrix. Thus, in the distributed control problem, the mass matrix in the (2,3)/(3,2)-block is square and in the boundary control problem we consider a rectangular version of it.

In all experiments that follow, we set Ω to be the unit square. We use uniform **Q1** finite elements and set $\beta = 10^{-3}$. The MATLAB code of Rees (2010) was used to generate the linear systems.

5.1 *Eigenvalues of unpreconditioned matrices*

Here we compare the eigenvalues of $\mathscr{K}$ (5.2) and $\mathscr{K}_\partial$ (5.4) to the eigenvalue bounds predicted by Theorem 2.2. We use MATLAB's eigs/svds functions to compute the minimum and maximum eigenvalues/singular values of the matrix blocks M, K, M_b and E.

We note that for the distributed control matrix $\mathscr{K}$, all blocks are square and the (3,3)-block is positive definite; therefore, we can also use our results to obtain bounds on the re-ordered matrix

$$\mathscr{K}_{\text{flip}} = \begin{bmatrix} \beta M & -M & 0 \\ -M & 0 & K \\ 0 & K & M \end{bmatrix}. \tag{5.5}$$

Both orderings generate the same extremal bounds, but different interior bounds.

(a) Positive eigenvalues of $\mathcal{K}$

(b) Negative eigenvalues of $\mathcal{K}$

(c) Positive eigenvalues of $\mathcal{K}_\partial$

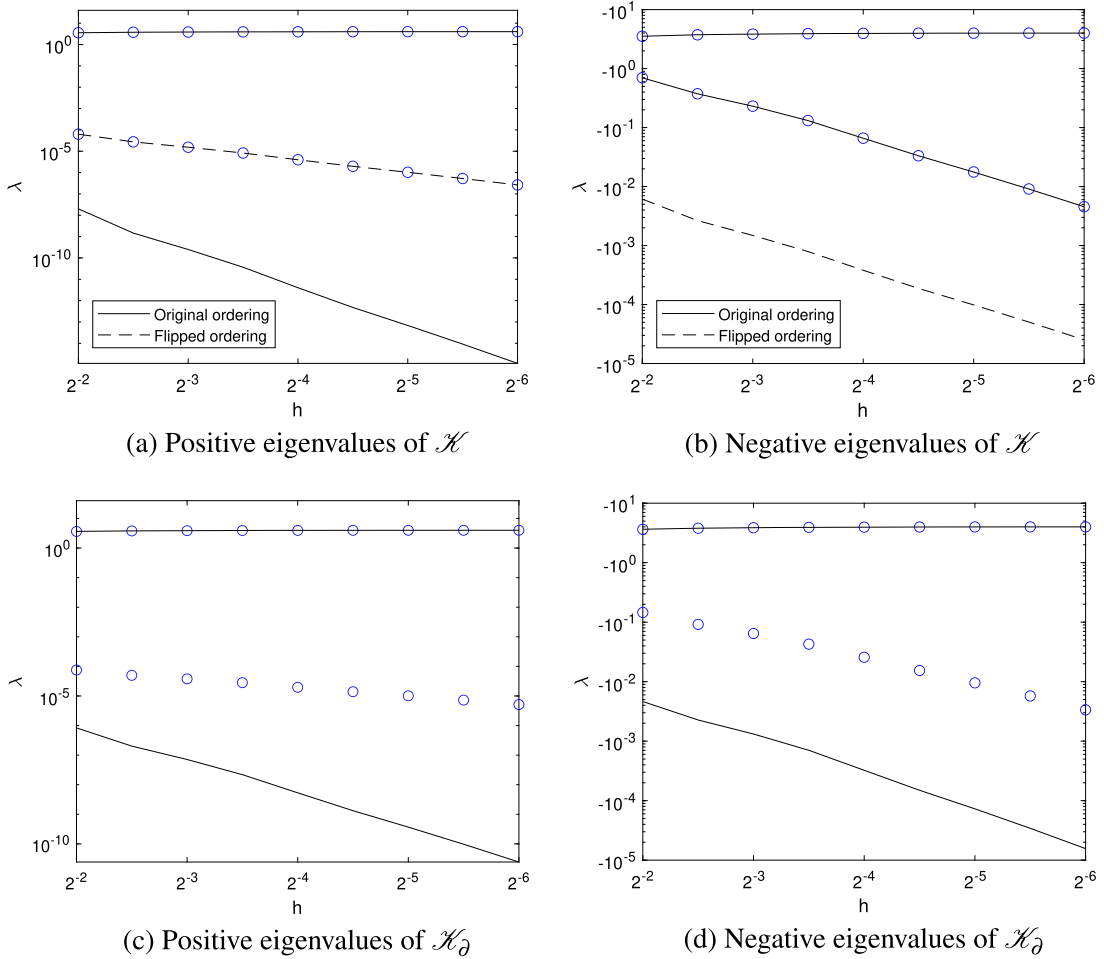(d) Negative eigenvalues of $\mathcal{K}_\partial$

FIG. 2. Largest and smallest positive (left) and negative (right) eigenvalues of the distributed control matrix $\mathcal{K}$ (top) and the boundary control matrix $\mathcal{K}_\partial$ (bottom). Circles indicate the eigenvalues, and lines the bounds given by Theorem 2.2. For $\mathcal{K}$, the dashed lines indicate the bounds obtained by applying Theorem 2.2 to the reordered matrix $\mathcal{K}_{\text{flip}}$.

Comparisons of the predicted eigenvalue bounds to the actual eigenvalues are shown in Figure 2. In the distributed control case, we show the bounds obtained for both the original matrix $\mathcal{K}$ and those for the reordered matrix $\mathcal{K}_{\text{flip}}$. In all cases, the bounds for the extremal eigenvalues are quite tight: this is because the K block has the largest eigenvalues ($O(1)$ compared to $O(h^2)$ for the others—see Elman *et al.*, 2005, Proposition 1.29 and Theorem 1.32). Therefore, the largest positive and smallest negative eigenvalue of both $\mathcal{K}$ and $\mathcal{K}_\partial$ tend towards $\mu_{\max}^K$ and $-\mu_{\max}^K$, respectively. Referring to the proofs of the extremal bounds in Theorem 2.2, the 3×3 matrices R will contain a $\mu_{\max}^K$ (or $-\mu_{\max}^K$) term, with other lower-order terms. This means that the extremal eigenvalues of R (and therefore the predicted eigenvalue bounds) will also be close to $\pm\mu_{\max}^K$.

It is more difficult to capture the interior bounds. In the distributed control case, each ordering (either the original ordering with M in the leading block or the flipped ordering with βM in the leading block) gives one bound that is quite tight and one that is loose. In the boundary control case, both interior bounds are loose.

5.2 *Eigenvalues of preconditioned matrices*

We now consider preconditioning strategies for PDE-constrained optimization. We examine eigenvalue bounds with both exact and approximate Schur complements.

For the distributed control problem, we work with the reordered matrix $\mathcal{K}_{\text{flip}}$ (5.5). The Schur complement preconditioner for $\mathcal{K}_{\text{flip}}$ is

$$\mathcal{M} = \begin{bmatrix} \beta M & 0 & 0 \\ 0 & \frac{1}{\beta} M & 0 \\ 0 & 0 & M + \beta K M^{-1} K \end{bmatrix}. \tag{5.6}$$

The first and second blocks are mass matrices, which are cheap to invert, so we leave these terms as they are. For the second Schur complement $S_2 = M + \beta K M^{-1} K$, we use the approximation $\tilde{S}_2 = \left(M + \sqrt{\beta} K \right) M^{-1} \left(M + \sqrt{\beta} K \right)$ proposed by Pearson and Wathen (Pearson & Wathen, 2012) to obtain the preconditioner:

$$\tilde{\mathcal{M}} = \begin{bmatrix} \beta M & 0 & 0 \\ 0 & \frac{1}{\beta} M & 0 \\ 0 & 0 & \left(M + \sqrt{\beta} K \right) M^{-1} \left(M + \sqrt{\beta} K \right) \end{bmatrix}. \tag{5.7}$$

Per Pearson & Wathen (2012, Theorem 4), the eigenvalues of $\tilde{S}_2^{-1} S_2$ satisfy $\Lambda(\tilde{S}_2^{-1} S_2) \in \left[\frac{1}{2}, 1 \right]$. Thus, $\tilde{\mathcal{M}}$ satisfies Theorem 4.1 with $\alpha_0 = \beta_0 = \alpha_1 = \beta_1 = 1$, $\alpha_2 = \frac{1}{2}$ and $\beta_2 = 1$.

Plots of the preconditioned eigenvalues for $h = 2^{-4}$ are shown in Section 3. The value η_E, defined as the maximal eigenvalue of

$$(C S_1^{-1} C^T)^{-1} E = \beta (K M^{-1} K)^{-1} M,$$

is approximately 2.6×10^{-7}. We note that for two-dimensional problems with uniform **Q1** finite element discretizations, the value η_E is $O(\beta h^4)$ and will thus be small in general.

Comparing Figs 3a and 3b, we notice that the bounds on the negative eigenvalue bounds do not change when we use the approximate Schur complement, but the lower positive bound becomes smaller (from 0.4450 with the exact Schur complement to 0.2929 for the approximate Schur complement), and the upper positive bound becomes larger (1.8019 in the exact case and 2 in the approximate case). We note that the eigenvalues appear to be very close to the predicted bounds, except for the upper positive eigenvalues of $\tilde{\mathcal{M}}^{-1} \mathcal{K}$. As discussed in Remark 4.3, this kind of minor looseness in the upper bound can happen when we have highly accurate Schur complement approximations (as in this case, where we are exactly inverting the two M blocks).
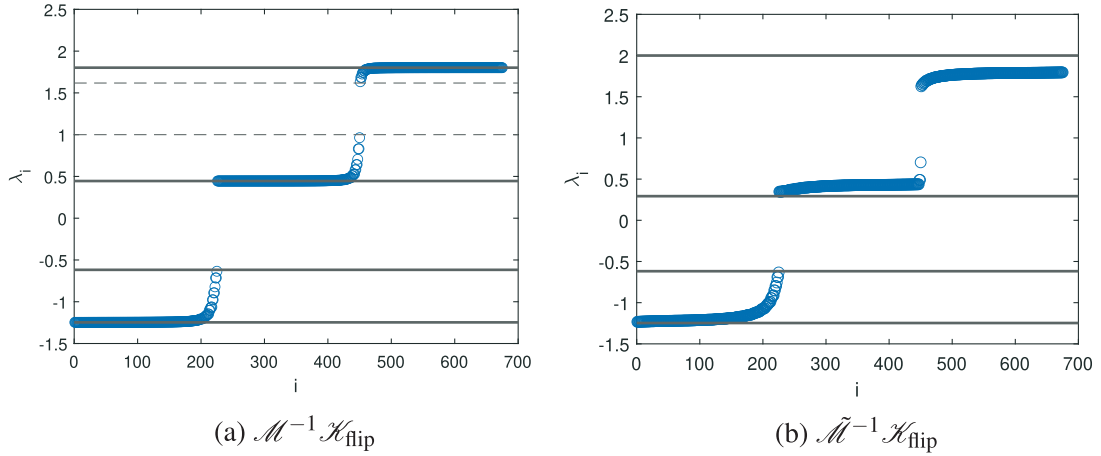
FIG. 3. Eigenvalue plots for reordered distributed Poisson control matrix $\mathcal{K}_{\text{flip}}$, with exact preconditioner $\mathcal{M}$ (left) and approximate preconditioner $\tilde{\mathcal{M}}$ (right). Eigenvalues are shown by the circles; eigenvalue bounds (from Theorem 3.2 on left and Corollary 4.1 on right) are shown by lines.

For the matrix $\mathcal{K}_\partial$ arising from the boundary control problem, the Schur complement preconditioner is

$$
\mathcal{M}_\partial = \begin{bmatrix} M & 0 & 0 \\ 0 & KM^{-1}K & 0 \\ 0 & 0 & \beta M_b + E^T \left(KM^{-1}K \right)^{-1} E \end{bmatrix}. \tag{5.8}
$$

In practice, the first Schur complement $KM^{-1}K$ can be inverted approximately with, for example, a multigrid method. For the second Schur complement, we note that for the two-dimensional Poisson control problem on a uniform **Q1** grid, the eigenvalues of K are between $O(h^2)$ and $O(1)$, while those of the mass matrices are all $O(h^2)$. Therefore, the term βM_b will dominate $E^T \left(KM^{-1}K \right)^{-1} E$ for all but very small values of β (which are not commonly used in practice). Therefore, an approximate preconditioner for $\mathcal{K}_\partial$ is given by

$$
\tilde{\mathcal{M}}_\partial = \begin{bmatrix} M & 0 & 0 \\ 0 & KM^{-1}K & 0 \\ 0 & 0 & \beta M_b \end{bmatrix}. \tag{5.9}
$$

We note that this preconditioner is presented in Rees *et al.* (2010), though it is derived there from the block-2 × 2 formulation of $\mathcal{K}$.

Plots of the preconditioned eigenvalues for $h = 2^{-4}$ are shown in Fig. 4. For $\tilde{\mathcal{M}}_\partial^{-1}\mathcal{K}_\partial$, we note that there is a single large-magnitude positive eigenvalue and a large-magnitude negative eigenvalue, whose absolute values are nearly 200, and they make it difficult to see how the other eigenvalues compare to to those of $\mathcal{M}_\partial^{-1}\mathcal{K}_\partial$. Thus in Fig. 4b, we omit the largest and smallest eigenvalues of $\tilde{\mathcal{M}}_\partial^{-1}\mathcal{K}_\partial$ and overlay the others on the corresponding eigenvalues of $\mathcal{M}_\partial^{-1}\mathcal{K}_\partial$. We notice that most other eigenvalues of $\tilde{\mathcal{M}}_\partial^{-1}\mathcal{K}_\partial$ remain close to those of $\mathcal{M}_\partial^{-1}\mathcal{K}_\partial$. We also note that the performance of this preconditioner
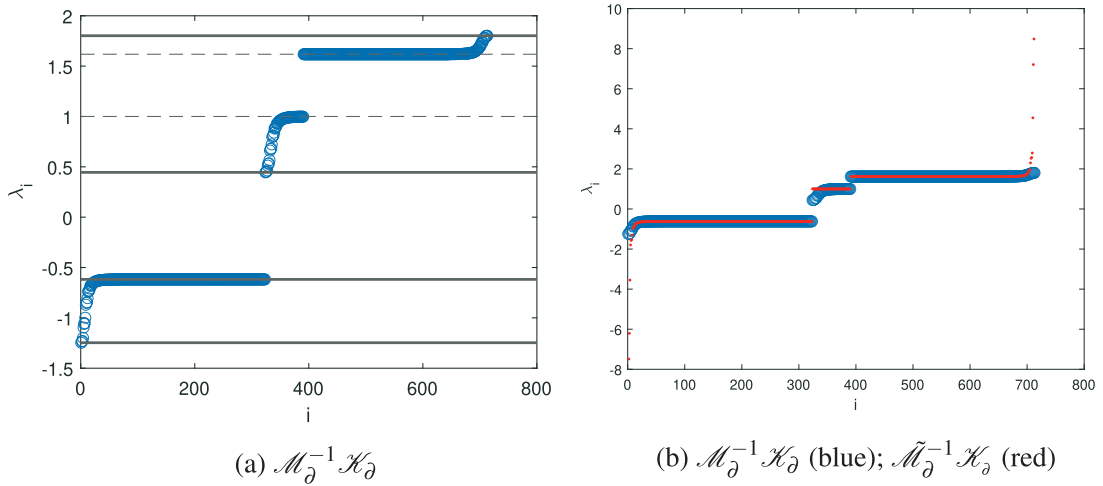
FIG. 4. Eigenvalue plots for the boundary control problem $\mathcal{K}_\partial$, with exact preconditioner $\mathcal{M}$ and approximate preconditioner $\tilde{\mathcal{M}}$. On left, eigenvalue bounds (from Theorem 3.2) are shown by horizontal lines. Right: all but the single smallest and largest eigenvalues for both preconditioners; values for $\mathcal{M}$ are shown by large blue circles, and those for $\tilde{\mathcal{M}}$ are shown by small red circles.

depends on β: in particular, the performance of the Schur complement approximation deteriorates for small β. We refer to Pearson & Wathen (2012); Rees *et al.* (2010) for further discussion of this.

It is evident from Fig. 4a that our bounds are tight and effective. Unlike in the boundary control example, the Schur complement approximation $\tilde{S}_2$ used here does not have β- and h-independent constants α_2, β_2 such that $\Lambda(\tilde{S}_2^{-1} S_2) \in [\alpha_2, \beta_2]$, so our analyses on Schur complement approximations in Section 4 are difficult to apply. Nonetheless, we observe from Fig. 4b that most of the eigenvalues of $\tilde{\mathcal{M}}_\partial^{-1} \mathcal{K}_\partial$ remain very close to those of $\mathcal{M}_\partial^{-1} \mathcal{K}_\partial$. Thus, we see that the eigenvalue bounds for the 'ideal' Schur complement preconditioner may still be of use, provided that we have an effective approximation of the Schur complement.

6. Conclusions

The increasing importance of double saddle-point systems requires attention to spectral properties of the matrices involved. We have shown that energy estimates are an effective tool for obtaining eigenvalue bounds in this case. The assumptions we make are rather general, and the analysis covers a large class of problems.

There are several directions for potential future work. Following Remark 4.3, specific assumptions on the magnitudes of the norms of the matrices D and E may yield additional results and insights on the bounds. The rank structure of the blocks may also have a significant effect, and it may be useful to eliminate the positive definiteness requirement of A and/or consider rank-deficient B and C. It may also be useful to consider nonsymmetric double saddle-point systems, and block triangular rather than block diagonal preconditioners, although when symmetry is lost eigenvalue bounds may not be a sufficient tool for predicting convergence rates of Krylov subspace solvers.

The tightness of our bounds indicates that spectral analysis is beneficial in capturing the properties of the matrices involved. This, in turn, makes it possible to effectively predict the convergence rate of Krylov subspace solvers for this important class of problems.

Acknowledgements

References

ADLER, J. H., BENSON, T. R., CYR, E. C., FARRELL, P. E., MACLACHLAN, S. P. & TUMINARO, R. S. (2021) Monolithic multigrid methods for magnetohydrodynamics. *SIAM J. Sci. Comput.*, **43**, S70–S91.

ALI BEIK, F. & BENZI, M. (2018) Iterative methods for double saddle point systems. *SIAM J. Matrix Anal. Appl.*, **39**, 902–921.

BEIGL, A., SOGN, J. & ZULEHNER, W. (2020) Robust preconditioners for multiple saddle point problems and applications to optimal control problems. *SIAM J. Matrix Anal. Appl.*, **41**, 1590–1615.

BENZI, M. (2002) Preconditioning techniques for large linear systems: a survey. *J. Comput. Phys.*, **182**, 418–477.

BENZI, M., GOLUB, G. H. & LIESEN, J. (2005) Numerical solution of saddle point problems. *Acta Numer.*, **14**, 1–137.

CAI, M., JU, G. & LI, J. (2021) Schur complement based preconditioners for twofold and block tridiagonal saddle point problems. https://arxiv.org/abs/2108.08332.

CAI, M., MU, M. & XU, J. (2009) Preconditioning techniques for a mixed Stokes/Darcy model in porous media applications. *J. Comput. Appl. Math.*, **233**, 346–355.

ELMAN, H. C., SILVESTER, D. J. & WATHEN, A. J. (2005) Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics. *Numerical Mathematics and Scientific Computation*. New York: Oxford University Press.

FERRONATO, M., FRANCESCHINI, A., JANNA, C., CASTELLETTO, N. & TCHELEPI, H. A. (2019) A general preconditioning framework for coupled multiphysics problems with application to contact- and poro-mechanics. *J. Comput. Phys.*, **398**, 108887.

GATICA, G. & HEUER, N. (2000) A dual–dual formulation for the coupling of mixed-FEM and BEM in hyperelasticity. *SIAM J. Numer. Anal.*, **38**, 380–400.

GOULD, N. I. M. (1985) On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem. *Math. Programming*, **32**, 90–99.

GOULD, N. I. M. & SIMONCINI, V. (2009) Spectral analysis of saddle point matrices with indefinite leading blocks. *SIAM J. Matrix Anal. Appl.*, **31**, 1152–1171.

GREIF, C., MOULDING, E. & ORBAN, D. (2014) Bounds on eigenvalues of matrices arising from interior-point methods. *SIAM J. Optim.*, **24**, 49–83.

HIPTMAIR, R. (2006). Operator preconditioning. *Comput. Math. Appl.*, **52**, 699–706.

HOLTER, K. E., KUCHTA, M. & MARDAL, K.-A. (2021) Robust preconditioning for coupled Stokes–Darcy problems with the Darcy problem in primal form. *Comput. Math. Appl.*, **91**, 53–66.

HORN, R. A. & JOHNSON, C. R. (1985) *Matrix Analysis*, 1st edn. Cambridge, UK: Cambridge University Press .

HUANG, N. & MA, C.-F. (2019) Spectral analysis of the preconditioned system for the 3×3 block saddle point problem. *Numer. Algorithms*, **81**, 421–444.

IPSEN, I. C. F. (2001) A note on preconditioning nonsymmetric matrices. *SIAM J. Sci. Comput.*, **23**, 1050–1051.

KEYES, D. E., MCINNES, L. C., WOODWARD, C. S., GROPP, W., MYRA, E., PERNICE, M., BELL, J., BROWN, J., CLO, A., CONNORS, J. M., CONSTANTINESCU, E. M., ESTEP, D. J., EVANS, K., FARHAT, C., HAKIM, A., HAMMOND, G., HANSEN, G., HILL, J., ISAAC, T., JIAO, X., JORDAN, K. E., KAUSHIK, D. K., KAXIRAS, E., KONIGES, A. E.,

LEE, K., LOTT, A., LU, Q., MAGERLEIN, J., MAXWELL, R., McCOURT, M., MEHL, M., PAWLOWSKI, R. P., RANDLES, A. P., REYNOLDS, D. R., RIVIÈRE, B., RÜDE, U., SCHEIBE, T. D., SHADID, J. N., SHEEHAN, B., SHEPHARD, M. S., SIEGEL, A. R., SMITH, B., TANG, X., WILSON, C. & WOHLMUTH, B. I. (2013) Multiphysics simulations: challenges and opportunities. *Int. J. High Perform. Comput. Appl.*, **27**, 4–83.

LANGER, U., OF, G., STEINBACH, O. & ZULEHNER, W. (2007) Inexact data-sparse boundary element tearing and interconnecting methods. *SIAM J. Sci. Comput.*, **29**, 290–314.

MARDAL, K.-A. & WINTHER, R. (2011) Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebra Appl.*, **18**, 1–40.

MORINI, B., SIMONCINI, V. & TANI, M. (2016) Spectral estimates for unreduced symmetric KKT systems arising from interior point methods. *Numer. Linear Algebra Appl.*, **23**, 776–800.

MURPHY, M. F., GOLUB, G. H. & WATHEN, A. J. (2000) A note on preconditioning for indefinite linear systems. *SIAM J. Sci. Comput.*, **21**, 1969–1972.

PAIGE, C. C. & SAUNDERS, M. A. (1975) Solutions of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, **12**, 617–629.

PEARSON, J. W. & PESTANA, J. (2020) Preconditioners for Krylov subspace methods: an overview. *GAMM-Mitt.*, **43**, e202000015, 35.

PEARSON, J. W. & POTSCHKA, A. (2021a) On symmetric positive definite preconditioners for multiple saddle-point systems. https://arxiv.org/abs/2106.12433v3.

PEARSON, J. W. & POTSCHKA, A. (2021b) A preconditioned inexact active-set method for large-scale nonlinear optimal control problems. https://arxiv.org/abs/2112.05020.

PEARSON, J. W. & WATHEN, A. J. (2012) A new approximation of the Schur complement in preconditioners for PDE-constrained optimization. *Numer. Linear Algebra Appl.*, **19**, 816–829.

PESTANA, J. & WATHEN, A. J. (2015) Natural preconditioning and iterative methods for saddle point systems. *SIAM Rev.*, **57**, 71–91.

RAMAGE, A. & GARTLAND, Jr., E. C. (2013) A preconditioned nullspace method for liquid crystal director modeling. *SIAM J. Sci. Comput.*, **35**, B226–B247.

REES, T. (2010) Github–tyronerees/poisson-control. https://github.com/tyronerees/poisson-control. Accessed 22 October 2021.

REES, T., DOLLAR, H. S. & WATHEN, A. J. (2010) Optimal solvers for PDE-constrained optimization. *SIAM J. Sci. Comput.*, **32**, 271–298.

RHEBERGEN, S., WELLS, G. N., WATHEN, A. J. & KATZ, R. F. (2015) Three-field block preconditioners for models of coupled magma/mantle dynamics. *SIAM J. Sci. Comput.*, **37**, A2270–A2294.

ROZLOŽNÍK, M. (2018) *Saddle-Point Problems and Their Iterative Solution*. Nečas Center Series. Birkhäuser/Springer, Cham.

RUIZ, D., SARTENAER, A. & TANNIER, C. (2018) Refining the lower bound on the positive eigenvalues of saddle point matrices with insights on the interactions between the blocks. *SIAM J. Matrix Anal. Appl.*, **39**, 712–736.

RUSTEN, T. & WINTHER, R. (1992) A preconditioned iterative method for saddlepoint problems. *SIAM J. Matrix Anal. Appl.*, **13**, 887–904.

SAAD, Y. (2003) *Iterative Methods for Sparse Linear Systems*, 2nd edn. Philadelphia, PA: Society for Industrial and Applied Mathematics.

SILVESTER, D. & WATHEN, A. (1994) Fast iterative solution of stabilised Stokes systems part II: using general block preconditioners. *SIAM J. Numer. Anal.*, **31**, 1352–1367.

SOGN, J. & ZULEHNER, W. (2018) Schur complement preconditioners for multiple saddle point problems of block tridiagonal form with application to optimization problems. *IMA J. Numer. Anal.*, **39**, 1328–1359.

WATHEN, A. J. (2015) Preconditioning. *Acta Numer.*, **24**, 329–376.