

EFFICIENT PRECONDITIONERS FOR COUPLED STOKES–DARCY PROBLEMS*

PAULA STROHBECK[†] AND IRYNA RYBAK[†]

Abstract. Coupled systems of free flow and porous media arise in a variety of technical and environmental applications. For laminar flow regimes, such systems are described by the Stokes equations in the free-flow region and Darcy’s law in the porous medium. An appropriate set of coupling conditions is needed on the fluid–porous interface. Discretisations of the Stokes–Darcy problems yield large, sparse, ill-conditioned, and, depending on the interface conditions, non-symmetric linear systems. Therefore, robust and efficient preconditioners are needed to accelerate convergence of the applied Krylov method. In this work, we develop and investigate block diagonal, block triangular and constraint preconditioners for the coupled Stokes–Darcy problems. We apply two classical sets of coupling conditions considering the Beavers–Joseph and the Beavers–Joseph–Saffman condition for the tangential velocity. For the Beavers–Joseph interface condition, the resulting system is non-symmetric, therefore GMRES method is used. Spectral and field-of-values bounds independent of the grid width are derived for the exact versions of the preconditioners. Furthermore, we develop efficient inexact versions of the preconditioners. We demonstrate the effectiveness and robustness of the proposed preconditioners in numerical experiments.

Key words. Stokes–Darcy problem, MAC scheme, GMRES, preconditioner, spectral analysis, field-of-values analysis

MSC codes. 65F08, 65N08, 76D07, 76S05

1. Introduction. Coupled systems of free flow and porous media appear in a variety of technical applications and environmental settings such as industrial filtration, water/gas management in fuel cells, and surface–subsurface interactions. For low Reynolds numbers, the free flow is governed by the Stokes equations, and flow in the porous-medium region is described by Darcy’s law. Interface conditions are needed to couple these models at the fluid–porous interface. In this work, we consider the classical set of coupling conditions, which consists of the conservation of mass across the interface, the balance of normal forces and either the Beavers–Joseph or the Beavers–Joseph–Saffman interface condition on the tangential velocity, e.g. [3, 16, 26, 33].

Different discretisations for the coupled Stokes–Darcy problems have been investigated such as the finite element method [5, 13, 14, 16, 26], the finite volume method [31, 34], the discontinuous Galerkin method [15, 38, 39] or their combinations. These discretisations yield large, sparse, ill-conditioned, and, in case of the Beavers–Joseph coupling condition, non-symmetric linear systems. The Krylov methods are typically applied to efficiently solve large linear systems. Since the Beavers–Joseph coupling condition leads to non-symmetric matrices, we focus in this paper on the GMRES method. The convergence of the iterative method can be significantly enhanced by an appropriate choice of preconditioners [6, 30, 32, 37].

Preconditioners for GMRES method applied to solve the Stokes–Darcy problem, which is discretised by the finite element method, have been recently studied [5, 9, 13, 14]. In particular, a block diagonal and a block triangular preconditioner based on decoupling the Stokes–Darcy system were developed in [13] for the case of the Beavers–

*Submitted to the editors DATE.

Funding: The work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 327154368 – SFB 1313 and Project Number 490872182.

[†]Institute of Applied Analysis and Numerical Simulation, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany (paula.strohbeck@ians.uni-stuttgart.de, iryna.rybak@ians.uni-stuttgart.de).

Joseph–Saffman condition on the tangential velocity. A constraint preconditioner for this coupled problem was proposed in [14]. Spectral and field-of-values analysis for the block triangular preconditioner from [13] and the constraint preconditioners from [14] is conducted in [5].

In this paper, we focus on preconditioners for the Stokes–Darcy system discretised by the finite volume method on staggered grids (MAC scheme), since this discretisation scheme is mass conservative, stable and allows natural coupling across the fluid–porous interface, e.g. [20, 22, 31]. The discretisation yields a symmetric matrix \mathcal{A}_{BJS} for the Beavers–Joseph–Saffman condition and a nearly symmetric matrix \mathcal{A}_{BJ} for the Beavers–Joseph condition. The matrices are displayed in equation (2.10) and illustrated in Figure 1. Note that \mathcal{A}_{BJS} is a double saddle point matrix. There exist several efficient preconditioners for double saddle point problems in the literature, e.g. a block diagonal preconditioner was developed in [23] and several block triangular preconditioners were proposed in [4, 22, 23, 25]. The matrix \mathcal{A}_{BJS} can also be interpreted as a standard saddle point matrix. This implies that besides the results on the double saddle point problems, the results on the standard saddle point problems can be applied as well. Different types of preconditioners for such systems were established in [2, 10, 28, 30, 32].

In this work, we consider three main classes of preconditioners and construct one for each class, namely, one block diagonal, one block triangular, and one constraint preconditioner. In comparison to available literature, we developed preconditioners not only for the Beavers–Joseph–Saffman interface condition, but also for the more general Beavers–Joseph coupling condition on the tangential velocity. Besides establishing spectral bounds for the preconditioned systems, we prove the norm- and field-of-values (FOV) equivalence for the constructed preconditioners extending our previous work [36], where we provided only numerical results and considered other coupling conditions for the Stokes–Darcy problem. FOV theory [7, 27] states that the convergence of the GMRES method is then independent of the grid width. In addition, the robustness of preconditioners is desired such that the choice of the physical parameters does not influence the convergence [1, 9, 11, 12, 29].

The direct use of the exact preconditioners is computationally expensive. Thus, accurate and easily invertible approximations of preconditioners are required, e.g. [8, 30, 32]. In this work, we therefore present also inexact variants of the constructed preconditioners and illustrate their robustness and efficiency in numerical experiments.

The paper is structured as follows. In section 2, we present the coupled Stokes–Darcy problems with two sets of interface conditions and introduce the corresponding discrete systems. In section 3, we develop three preconditioners and propose their efficient inexact variants. We provide spectral and field-of-values analysis of the preconditioned systems in section 4. To demonstrate efficiency and robustness of the constructed preconditioners, numerical experiments are conducted in section 5. In section 6, we summarise the obtained results and present possible extensions of this work. Some useful definitions and results are provided in Appendix A.

2. Model formulation. In this paper, we consider the two-dimensional setting. The coupled flow domain $\Omega = \Omega_{\text{pm}} \cup \Omega_{\text{ff}} \subset \mathbb{R}^2$ consists of the free flow Ω_{ff} and the porous medium Ω_{pm} . The two flow regions are separated by the sharp fluid–porous interface Σ (Figure 1, left). We consider steady-state, incompressible, single-fluid-phase flows at low Reynolds numbers ($Re \ll 1$). The solid phase is non-deformable and rigid that leads to a constant porosity. We deal with homogeneous isotropic and orthotropic porous media. The whole flow system is assumed to be isothermal.

2.1. Mathematical model. Coupled flow formulation consists of two different flow models in the two domains and an appropriate set of coupling conditions on the

fluid–porous interface. Under the assumptions on the flow made above, the Stokes equations (2.1) are used in Ω_{ff} . Without loss of generality, we consider the Dirichlet boundary conditions on the external boundary

$$(2.1) \quad \nabla \cdot \mathbf{v}_{\text{ff}} = 0, \quad -\nabla \cdot \mathbf{T}(\mathbf{v}_{\text{ff}}, p_{\text{ff}}) = \mathbf{f}^{\text{ff}} \quad \text{in } \Omega_{\text{ff}},$$

$$(2.2) \quad \mathbf{v}_{\text{ff}} = \bar{\mathbf{v}} \quad \text{on } \partial\Omega_{\text{ff}} \setminus \Sigma,$$

where \mathbf{v}_{ff} is the fluid velocity, p_{ff} is the fluid pressure, \mathbf{f}^{ff} is a source term, e.g. body force, and $\bar{\mathbf{v}}$ is a given function. We consider the stress tensor $\mathbf{T}(\mathbf{v}_{\text{ff}}, p_{\text{ff}}) = \mu(\nabla \mathbf{v}_{\text{ff}} + (\nabla \mathbf{v}_{\text{ff}})^\top) - p_{\text{ff}} \mathbf{I}$, where μ is the dynamic viscosity and \mathbf{I} is the identity tensor.

Fluid flow in the porous-medium domain Ω_{pm} is based on Darcy’s law. Again, we consider the Dirichlet boundary conditions on the external boundary

$$(2.3) \quad \nabla \cdot \mathbf{v}_{\text{pm}} = f_{\text{pm}}, \quad \mathbf{v}_{\text{pm}} = -\mu^{-1} \mathbf{K} \nabla p_{\text{pm}} \quad \text{in } \Omega_{\text{pm}},$$

$$(2.4) \quad p_{\text{pm}} = \bar{p} \quad \text{on } \partial\Omega_{\text{pm}} \setminus \Sigma,$$

where \mathbf{v}_{pm} is the Darcy velocity, p_{pm} is the pressure, f_{pm} is a source term, \mathbf{K} is the intrinsic permeability tensor, and \bar{p} is a given function. The permeability tensor is symmetric, positive definite, and bounded. In this paper, we restrict ourselves to isotropic ($\mathbf{K} = k \mathbf{I}$, $k > 0$) and orthotropic ($\mathbf{K} = \text{diag}(k_{xx}, k_{yy})$, $k_{xx}, k_{yy} > 0$) porous media.

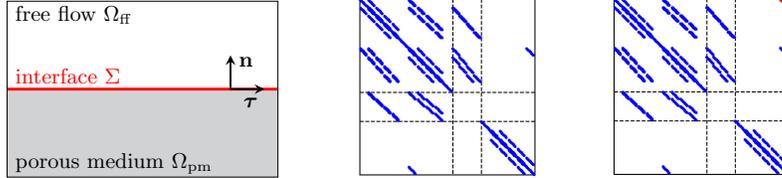


Fig. 1: Flow system description (left) and sparsity structures of \mathcal{A}_{BJS} (middle) and \mathcal{A}_{BJ} (right) for $h = 1/8$ with symmetric non-zero entries in blue (•) and non-symmetric non-zero entries in red (•)

In addition to the boundary conditions on the external boundary, appropriate coupling conditions have to be defined on the fluid–porous interface Σ . The classical set of coupling conditions consists of the conservation of mass across the interface (2.5), the balance of normal forces (2.6) and the Beavers–Joseph condition (2.7) on the tangential velocity [3]:

$$(2.5) \quad \mathbf{v}_{\text{ff}} \cdot \mathbf{n} = \mathbf{v}_{\text{pm}} \cdot \mathbf{n} \quad \text{on } \Sigma,$$

$$(2.6) \quad -\mathbf{n} \cdot \mathbf{T}(\mathbf{v}_{\text{ff}}, p_{\text{ff}}) \cdot \mathbf{n} = p_{\text{pm}} \quad \text{on } \Sigma,$$

$$(2.7) \quad (\mathbf{v}_{\text{ff}} - \mathbf{v}_{\text{pm}}) \cdot \boldsymbol{\tau} - \alpha^{-1} \sqrt{K} ((\nabla \mathbf{v}_{\text{ff}} + (\nabla \mathbf{v}_{\text{ff}})^\top) \cdot \mathbf{n}) \cdot \boldsymbol{\tau} = 0 \quad \text{on } \Sigma.$$

Here, $\mathbf{n} = -\mathbf{n}_{\text{ff}} = \mathbf{n}_{\text{pm}}$ is the unit vector normal to the fluid–porous interface Σ pointing outward from the porous-medium domain Ω_{pm} , $\boldsymbol{\tau}$ is the unit vector tangential to the interface (Figure 1), and $\alpha > 0$ is the Beavers–Joseph slip coefficient. Different approaches to compute \sqrt{K} exist in the literature. In this paper, we consider $\sqrt{K} = \sqrt{\boldsymbol{\tau} \cdot \mathbf{K} \cdot \boldsymbol{\tau}}$ as in [31].

Saffman [33] proposed a simplification of the Beavers–Joseph condition (2.7), where the tangential porous-medium velocity $\mathbf{v}_{\text{pm}} \cdot \boldsymbol{\tau}$ on the interface is neglected

$$(2.8) \quad \mathbf{v}_{\text{ff}} \cdot \boldsymbol{\tau} - \alpha^{-1} \sqrt{K} ((\nabla \mathbf{v}_{\text{ff}} + (\nabla \mathbf{v}_{\text{ff}})^\top) \cdot \mathbf{n}) \cdot \boldsymbol{\tau} = 0 \quad \text{on } \Sigma.$$

In the literature, the Stokes–Darcy problem with the Beavers–Joseph–Saffman condition (2.8) is usually studied, both from the analytical and numerical point of view. Only a few papers focus on the original Beavers–Joseph condition (2.7). In this work, we develop and analyse preconditioners for the Stokes–Darcy problem (2.1)–(2.4) with both sets of interface conditions, (2.5)–(2.7) and (2.5), (2.6), (2.8).

2.2. Discretisation. The coupled Stokes–Darcy problems (2.1)–(2.7) and (2.1)–(2.6), (2.8) are discretised with the second order finite volume method. The MAC scheme (finite volume method on staggered grids) is used for the Stokes equations, e.g. [20, 31]. The porous-medium model (2.3) is discretised in its primal form, where Darcy’s law is substituted to the mass balance equation. Here, the pressure p_{pm} is the primary variable which is defined in the control volume centres as well as on the fluid–porous interface and the external boundary of the domain. The porous-medium velocity components $\mathbf{v}_{\text{ff}} = (u_{\text{ff}}; v_{\text{ff}})$ are computed on the control volume faces in a post-processing step. This leads to the system of linear equations

$$(2.9) \quad \mathcal{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} = (\mathbf{v}_{\text{ff}}; p_{\text{ff}}; p_{\text{pm}})^\top, \quad \mathcal{A} \in \{\mathcal{A}_{\text{BJS}}, \mathcal{A}_{\text{BJ}}\},$$

where $\mathbf{v}_{\text{ff}} \in \mathbb{R}^n$, $p_{\text{ff}} \in \mathbb{R}^m$, $p_{\text{pm}} \in \mathbb{R}^l$ are the primary variables, and the matrices are given by

$$(2.10) \quad \mathcal{A}_{\text{BJS}} = \begin{pmatrix} A & B^\top & C^\top \\ B & 0 & 0 \\ C & 0 & -D \end{pmatrix}, \quad \mathcal{A}_{\text{BJ}} = \begin{pmatrix} A & B^\top & C_2^\top \\ B & 0 & 0 \\ C_1 & 0 & -D \end{pmatrix}.$$

Here, the blocks $A \in \mathbb{R}^{n \times n}$ and $D \in \mathbb{R}^{l \times l}$ are both symmetric and positive definite ($A = A^\top \succ 0$, $D = D^\top \succ 0$) and the matrix $B \in \mathbb{R}^{m \times n}$ has full row rank ($\text{rank}(B) = m$). For the case of the Beavers–Joseph–Saffman condition (2.8), the matrix \mathcal{A}_{BJS} is symmetric ($\mathcal{A}_{\text{BJS}} = \mathcal{A}_{\text{BJS}}^\top$), however for the Beavers–Joseph condition (2.7) it is not possible to get a completely symmetric matrix ($\mathcal{A}_{\text{BJ}} \neq \mathcal{A}_{\text{BJ}}^\top$). Note that the matrix \mathcal{A}_{BJS} is a double saddle point matrix.

The sparsity structure of \mathcal{A}_{BJS} and \mathcal{A}_{BJ} is presented in Figure 1 for the grid width $h_x = h_y = h = 1/8$. The first row in (2.10) corresponds to the discretised momentum balance equation in the Stokes system (2.1) and the second row is the incompressibility condition from (2.1). The third row is the discrete version of the porous-medium model (2.3) in its primal form, where Darcy’s law is substituted to the mass balance equation. The interface conditions (2.5), (2.6) and (2.8) are incorporated in the matrix $C \in \mathbb{R}^{l \times n}$, and conditions (2.5)–(2.7) are in the matrices $C_1 \in \mathbb{R}^{l \times n}$ and $C_2 \in \mathbb{R}^{l \times n}$.

The discretisation scheme (stencil in Figure 2, left) for the conservation of mass across the interface (2.5) reads

$$(2.11) \quad -h_x v_{\text{ff},P} - 2 \frac{k_{yy}}{\mu} \frac{h_x}{h_y} p_{\text{pm},s} + 2 \frac{k_{yy}}{\mu} \frac{h_x}{h_y} p_{\text{pm},P} = 0,$$

where the coefficient $-h_x$ in the first term in (2.11) goes to the matrices C and C_1 , respectively, and the two other coefficients $\pm 2k_{yy}h_x/(\mu h_y)$ enter the matrix D in (2.10).

We obtain the discrete form of the balance of normal forces (2.6) considering a

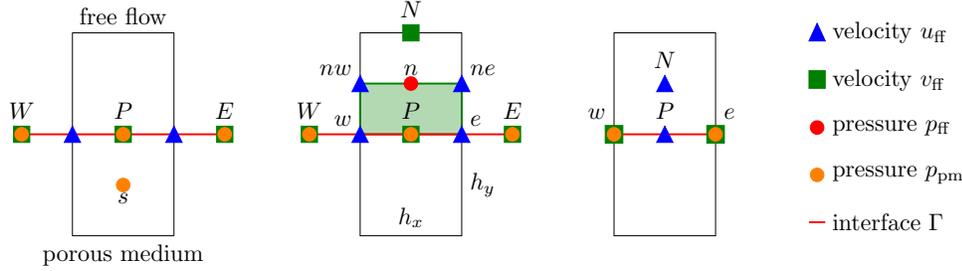


Fig. 2: Stencils and primary variables for the coupling conditions

control volume of size $h_x h_y / 2$ (Figure 2, middle in green):

$$(2.12) \quad \left(2\mu \frac{h_y}{h_x} + \mu \frac{h_x}{h_y} \right) v_{\text{ff},P} - 2\mu \frac{h_x}{h_y} v_{\text{ff},N} - \frac{1}{2}\mu \frac{h_y}{h_x} v_{\text{ff},W} - \frac{1}{2}\mu \frac{h_y}{h_x} v_{\text{ff},E} \\ + \mu u_{\text{ff},nw} - \mu u_{\text{ff},ne} - \mu u_{\text{ff},w} + \mu u_{\text{ff},e} + h_x p_{\text{ff},n} - h_x p_{\text{pm},P} = f_P^v \frac{h_x h_y}{2}.$$

The first eight coefficients in (2.12) contribute to the matrix A , the coefficient h_x to the matrix B , and the coefficient $-h_x$ to the matrices C^\top and C_2^\top , respectively. For discretisation of the Beavers–Joseph condition (2.7) we get

$$(2.13) \quad \left(\mu \frac{\alpha}{\sqrt{k_{xx}}} h_x + 2\mu \frac{h_x}{h_y} \right) u_{\text{ff},P} - 2\mu \frac{h_x}{h_y} u_{\text{ff},N} + \mu v_{\text{ff},w} - \mu v_{\text{ff},e} \\ - \alpha \frac{k_{yy}}{\sqrt{k_{xx}}} p_{\text{pm},w} + \alpha \frac{k_{yy}}{\sqrt{k_{xx}}} p_{\text{pm},e} = 0,$$

and for the simplification by Saffman (2.8) we end up with

$$(2.14) \quad \left(\mu \frac{\alpha}{\sqrt{k_{xx}}} h_x + 2\mu \frac{h_x}{h_y} \right) u_{\text{ff},P} - 2\mu \frac{h_x}{h_y} u_{\text{ff},N} + \mu v_{\text{ff},w} - \mu v_{\text{ff},e} = 0.$$

The first four coefficients in (2.13) and (2.14) enter the matrix A . The fifth and sixth coefficients $\pm \alpha k_{yy} / \sqrt{k_{xx}}$ in (2.13) contribute to the matrix C_2^\top and break the symmetry of the matrix \mathcal{A}_{BJ} .

3. Preconditioners. In this section, we propose three different types of preconditioners for coupled Stokes–Darcy problems, a block diagonal, a block triangular and a constraint one (subsection 3.1). For efficient numerical simulations of large systems we develop inexact variants of these preconditioners (subsection 3.2).

3.1. Exact preconditioners. We solve the coupled problems (2.9), (2.10) monolithically using flexible GMRES (FGMRES) method [32, chap. 9.4.1], which is applicable to non-symmetric matrices like \mathcal{A}_{BJ} . In FGMRES right preconditioning has to be used

$$(3.1) \quad \mathcal{A}\mathcal{P}^{-1}\bar{\mathbf{x}} = \mathbf{b}, \quad \bar{\mathbf{x}} = \mathcal{P}\mathbf{x}.$$

As a first step in the construction of preconditioners, we decouple the Stokes–Darcy system and consider the following matrix

$$(3.2) \quad \bar{\mathcal{A}} = \begin{pmatrix} A & B^\top & 0 \\ B & 0 & 0 \\ 0 & 0 & -D \end{pmatrix}.$$

We develop preconditioners for the matrix $\bar{\mathcal{A}}$ and show theoretically and numerically that they are also suitable to the matrix $\mathcal{A} \in \{\mathcal{A}_{\text{BJS}}, \mathcal{A}_{\text{BJ}}\}$ defined in (2.10). We propose block diagonal and block triangular preconditioners

$$(3.3) \quad \mathcal{P}_{\text{diag}} = \begin{pmatrix} A & 0 & 0 \\ 0 & -S_B & 0 \\ 0 & 0 & -D \end{pmatrix}, \quad \mathcal{P}_{\text{tri}} = \begin{pmatrix} A & B^\top & 0 \\ 0 & -S_B & 0 \\ 0 & 0 & -D \end{pmatrix},$$

where $S_B := BA^{-1}B^\top \in \mathbb{R}^{m \times m}$ is the Schur complement ($S_B = S_B^\top \succ 0$). A similar attempt to $\mathcal{P}_{\text{diag}}$ was proposed in [13], where the Schur complement is approximated as $S_B \approx \mu^{-1}I$, where I is the identity matrix. The block triangular preconditioner \mathcal{P}_{tri} given in (3.3) is a simplified version of the block triangular preconditioners developed in [4] for double saddle point problems.

Furthermore, we construct a constraint preconditioner \mathcal{P}_{con} . Here, the matrix $\bar{\mathcal{A}}$ can be interpreted as a standard saddle point matrix

$$(3.4) \quad \bar{\mathcal{A}} = \begin{pmatrix} A & \bar{B}^\top \\ \bar{B} & -\bar{D} \end{pmatrix} \quad \text{with} \quad \bar{B} = \begin{pmatrix} B \\ 0 \end{pmatrix}, \quad \bar{D} = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}.$$

Therefore, also preconditioners for standard saddle point problems are suitable. We modify and generalise the following constraint preconditioner proposed in [2]:

$$(3.5) \quad \mathcal{P}_{\text{con}} = \begin{pmatrix} G & \bar{B}^\top \\ \bar{B} & -\bar{D} \end{pmatrix}, \quad G = \text{diag}(A_{11}, A_{22}), \quad G = G^\top \succ 0,$$

where G is the preconditioner to the block $A = (A_{ij})_{i,j=1,2}$ in (2.10).

3.2. Inexact preconditioners. To obtain accurate numerical results for the coupled Stokes–Darcy system (2.1), (2.3) with suitable boundary and interface conditions, we need to consider small grid widths which yield large linear systems. Since exact versions of preconditioners are computationally expensive, they have to be replaced by efficient inexact variants. The approximations of the blocks A , G , S_B and D in (3.3) and (3.5) are marked by \hat{A} , \hat{G} , \hat{S}_B and \hat{D} , respectively, and the inexact versions of the corresponding preconditioners are $\hat{\mathcal{P}}_{\text{diag}}$, $\hat{\mathcal{P}}_{\text{tri}}$ and $\hat{\mathcal{P}}_{\text{con}}$. The approximations should be easily invertible to reduce the computational effort. We replace the inverses of A , G and D from (3.3) and (3.5) applied in (3.1) with the approximations $\hat{A}^{-1} = \hat{G}^{-1} = \text{diag}(\text{AMG}(A_{11}), \text{AMG}(A_{22}))$ and $\hat{D}^{-1} = \text{AMG}(D)$, respectively. Here, $\text{AMG}(A_{11})$, $\text{AMG}(A_{22})$ and $\text{AMG}(D)$ are algebraic multigrid methods.

When the Stokes–Darcy system (2.1), (2.3) is discretised with the MAC scheme, a common way to approximate the Schur complement S_B is to use the Stokes pressure mass matrix [19, Theorem 3.22]. To define the finite volume analogon of the pressure mass matrix, we consider the continuous version of the Schur complement

$$(3.6) \quad \mathcal{S}_B := -\nabla \cdot [-\nabla \cdot (\mu(\nabla + \nabla^\top))]^{-1} \nabla.$$

We note that $\mathcal{S}_B \approx (2\mu)^{-1}I$. Since we use the finite volume scheme, we approximate the Schur complement S_B by $\hat{S}_B = (2\mu)^{-1}h_x h_y I$.

4. Analysis. To analyse the preconditioned systems, we provide spectral and field-of-values (FOV) analysis. We show that the eigenvalues of the preconditioned matrices $\mathcal{A}\mathcal{P}^{-1}$ are clustered and bounded away from zero. These properties ensure fast convergence of iterative methods. To get an upper bound for the GMRES residuals, we need FOV bounds.

4.1. Spectral analysis. The goal of this section is to analyse the spectra of the preconditioned matrices $\sigma(\mathcal{AP}_{\text{diag}}^{-1})$, $\sigma(\mathcal{AP}_{\text{tri}}^{-1})$ and $\sigma(\mathcal{AP}_{\text{con}}^{-1})$ for $\mathcal{A} \in \{\mathcal{A}_{\text{BJS}}, \mathcal{A}_{\text{BJ}}\}$ defined in (2.10). Eigenvalues clustered around one and/or a clustered spectrum away from zero often provide fast convergence of the Krylov subspace methods [6]. We conduct the proofs for the matrix \mathcal{A}_{BJS} . The proofs can be extended to the more general case \mathcal{A}_{BJ} .

In this work, we use the following notation. For given vectors $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^l$, we define a column vector $(x; y; z)^\top \in \mathbb{R}^{n+m+l}$. For $H = H^\top \succ 0$, we define the vector norm and the corresponding induced matrix norm

$$\langle x, x \rangle_H = \langle Hx, x \rangle = x^\top Hx = \|x\|_H, \quad \|M\|_H = \max_{x \neq 0} \frac{\|Mx\|_H}{\|x\|_H}.$$

The extension of the induced matrix norm for $H_1 = H_1^\top \succ 0$, $H_2 = H_2^\top \succ 0$ is given by [24, Problem 5.6.P4]:

$$(4.1) \quad \|M\|_{H_1, H_2} = \max_{x \in \mathbb{R}^t \setminus \{0\}} \frac{\|Mx\|_{H_2}}{\|x\|_{H_1}}, \quad M \in \mathbb{R}^{k \times t}.$$

Moreover, we have the following equalities

$$(4.2) \quad \|H_2^{-1/2} M H_1^{-1/2}\|_2 = \|M\|_{H_1, H_2^{-1}} = \|M H_1^{-1}\|_{H_1^{-1}, H_2^{-1}} = \|H_2^{-1} M\|_{H_1, H_2}.$$

Note that for $H_1 = H_2 = H$ and $k = t$, we get the standard induced matrix norm $\|M\|_{H, H} = \|M\|_H$.

LEMMA 4.1. *Let $A = A^\top \succ 0$ and B has full rank, then the following equalities hold*

$$(i) \quad \max_{x \neq 0} \frac{x^\top B^\top (BA^{-1}B^\top)^{-1} Bx}{x^\top Ax} = 1,$$

$$(ii) \quad \frac{x^\top B^\top (BA^{-1}B^\top)^{-1} Bx}{x^\top Ax} = 1 \quad \text{for } x \notin \ker(BA^{-1}).$$

Proof. (i) Simple algebraic manipulations yield

$$\begin{aligned} \max_{x \neq 0} \frac{x^\top B^\top S_B^{-1} Bx}{x^\top Ax} &\stackrel{(4.1)}{=} \max_{x \neq 0} \frac{\|Bx\|_{S_B^{-1}}^2}{\|x\|_A^2} = \|B\|_{A, S_B^{-1}}^2 \stackrel{(4.2)}{=} \left\| S_B^{-1/2} B A^{-1/2} \right\|_2^2 \\ &= \left\| A^{-1/2} B^\top S_B^{-1/2} \right\|_2^2 \stackrel{(4.2)}{=} \|B^\top\|_{S_B, A^{-1}}^2 = \max_{y \neq 0} \frac{\|B^\top y\|_{A^{-1}}^2}{\|y\|_{S_B}^2} \stackrel{(4.1)}{=} \max_{y \neq 0} \frac{y^\top S_B y}{y^\top S_B y} = 1. \end{aligned}$$

(ii) We have

$$\frac{\langle B^\top S_B^{-1} Bx, x \rangle}{\langle Ax, x \rangle} = \frac{\langle B^\top S_B^{-1} B A^{-1} y, y \rangle}{\langle y, y \rangle}.$$

Since $B^\top S_B^{-1} B A^{-1} = I$ on $\text{range}(B^\top) \supset \text{range}(A^{-1}B^\top)$, see [17, Section 1], this equality also holds for $\mathbb{R}^m \setminus \ker(BA^{-1})$ due to $\mathbb{R}^m = \text{range}(A^{-1}B^\top) \oplus \ker(BA^{-1})$. This completes the proof. \square

THEOREM 4.2. *The preconditioned matrix $\mathcal{A}_{\text{BJS}} \mathcal{P}_{\text{diag}}^{-1}$ defined in (2.10) with the block diagonal preconditioner $\mathcal{P}_{\text{diag}}$ defined in (3.3) has either the eigenvalue $\lambda = 1$ or $|\lambda| \geq \tau$ and $\zeta \leq |\lambda - 1| \leq 1 + \omega$ for $\tau, \zeta, \omega > 0$. Furthermore, the eigenvalues cluster around $1/2$ and 1 for $h \rightarrow 0$.*

Proof. Let λ be an eigenvalue of $\mathcal{A}_{\text{BJS}}\mathcal{P}_{\text{diag}}^{-1}$ to the eigenvector $(x; y; z)^\top \neq 0$ such that

$$(4.3) \quad Ax + B^\top y + C^\top z = \lambda Ax,$$

$$(4.4) \quad Bx = -\lambda S_B y,$$

$$(4.5) \quad Cx - Dz = -\lambda Dz.$$

We obtain $\lambda = 1$ with the corresponding eigenvectors $(x; 0; 0)^\top$ for $0 \neq x \in \ker(C)$ and $(0; 0; z)^\top$ for $0 \neq z \in \ker(C^\top)$.

Now we assume $\lambda \neq 1$. From (4.4) and (4.5), we get $y = -\lambda^{-1}S_B^{-1}Bx$ and $z = 1/(1-\lambda)D^{-1}Cx$. Substitution of these vectors in (4.3) yields

$$(4.6) \quad \lambda(1-\lambda)^2q - (1-\lambda)r + \lambda p = 0,$$

where

$$(4.7) \quad q = x^\top Ax > 0, \quad r = x^\top B^\top S_B^{-1}Bx \geq 0, \quad p = x^\top C^\top D^{-1}Cx \geq 0.$$

For $p = 0$ ($x \in \ker(C)$), we obtain $\lambda = 1/2 \pm \sqrt{1/4 - r/q}$.

For $r = 0$ ($x \in \ker(B)$), we get $\lambda = 1 \pm \sqrt{-p/q}$.

For $r \neq 0, p \neq 0$, substitution of $\lambda = t + 1$ in (4.6) yields

$$(4.8) \quad qt^3 + qt^2 + (r+p)t + p = 0.$$

Applying [5, Theorem 2.5], we get the bounds

$$0 < \zeta := \min \left\{ \frac{p}{p+r}, \frac{r+p}{q}, 1 \right\} \leq |\lambda - 1| \leq \max \left\{ \frac{p}{p+r}, \frac{r+p}{q}, 1 \right\}.$$

For the upper bound it holds

$$\max \left\{ \frac{p}{p+r}, \frac{r+p}{q}, 1 \right\} \leq \max \left\{ \max_{x \neq 0} \frac{p}{p+r}, \max_{x \neq 0} \frac{r+p}{q}, 1 \right\} \leq 1 + \omega,$$

where we applied Lemma 4.1 with r and q defined in (4.7) to obtain $\max_{x \neq 0} (r/q) = 1$.

Using (4.6) we estimate

$$|\lambda| = |t+1| \geq \frac{r|t|}{q|t|^2 + p} \geq \frac{r\zeta}{q(1+\omega)^2 + p} =: \tau,$$

which gives us the lower bound on $|\lambda|$.

To show clustering around $1/2$ and 1 , we get $x = (A^{-1}B^\top y + A^{-1}C^\top z)/(\lambda - 1)$ from (4.3) and insert it into (4.4). Rearranging the terms and multiplying the expression with y^\top from the left, we obtain

$$(4.9) \quad a\lambda^2 - a\lambda + a + b = 0, \quad a = y^\top S_B y > 0, \quad b = y^\top B A^{-1} C^\top z = z^\top C A^{-1} B^\top y.$$

The solution of this equation is

$$(4.10) \quad \lambda_{1,2} = \left(1 \pm \sqrt{-3 - 4b/a} \right) / 2.$$

Analogously, inserting $x = (A^{-1}B^\top y + A^{-1}C^\top z)/(\lambda - 1)$ in (4.5), rearranging the terms and multiplying from the left with z^\top , we get

$$(4.11) \quad c\lambda^2 - 2c\lambda + c + b + d = 0, \quad c = z^\top Dz > 0, \quad d = z^\top C A^{-1} C^\top z.$$

Solving the equation above leads to

$$(4.12) \quad \lambda_{1,2} = 1 \pm \sqrt{-(b+d)/c}.$$

The terms b and d defined in (4.9) and (4.11) converge to zero for $h \rightarrow 0$, because the entries of C are of order $O(h)$ due to discretisation of the interface conditions (2.11) and (2.12). Therefore, the eigenvalues cluster around $1/2$ and 1 . \square

The eigenvalue distribution of the preconditioned matrix $\mathcal{A}_{\text{BJS}}\mathcal{P}_{\text{diag}}^{-1}$ (Figure 3) confirms the theoretical results from Theorem 4.2. Note that we get similar clustering for the eigenvalues of $\mathcal{A}_{\text{BJ}}\mathcal{P}_{\text{diag}}^{-1}$.

THEOREM 4.3. *The preconditioned matrix $\mathcal{A}_{\text{BJS}}\mathcal{P}_{\text{tri}}^{-1}$ defined in (2.10) with the block triangular preconditioner \mathcal{P}_{tri} defined in (3.3) has either the eigenvalue $\lambda = 1$ or $|\lambda| \geq \tau$ and $\zeta \leq |\lambda - 2| \leq 3$ for $\zeta, \tau, \omega > 0$. Furthermore, the eigenvalues cluster around 1 for $h \rightarrow 0$.*

Proof. Let λ be an eigenvalue of $\mathcal{A}_{\text{BJS}}\mathcal{P}_{\text{tri}}^{-1}$ to the eigenvector $(x; y; z)^\top \neq 0$ such that

$$(4.13) \quad Ax + B^\top y + C^\top z = \lambda(Ax + B^\top y),$$

$$(4.14) \quad Bx = -\lambda S_B y,$$

$$(4.15) \quad Cx - Dz = -\lambda Dz.$$

We get $\lambda = 1$ with the corresponding eigenvectors $(x; 0; 0)^\top$ for $0 \neq x \in \ker(C)$ and $(0; 0; z)^\top$ for $0 \neq z \in \ker(C^\top)$. Now, we assume $\lambda \neq 1$. From (4.14) and (4.15) we get $y = -S_B^{-1}Bx/\lambda$ and $z = D^{-1}Cx/(1-\lambda)$. Inserting y and z in (4.13) and rearranging the terms, we obtain

$$(4.16) \quad \lambda(1-\lambda)^2q - (1-\lambda)^2r + \lambda p = 0,$$

where q, r and p are defined in (4.7).

For $p = 0$ ($x \in \ker(C)$), we get $\lambda = r/q$.

For $r = 0$ ($x \in \ker(B)$), we get $\lambda = 1 \pm \sqrt{-p/q}$.

For $p \neq 0, r \neq 0$, substitution of $\lambda = t + 2$ in (4.16) yields

$$(4.17) \quad qt^3 + (4q - r)t^2 + (5q - 2r + p)t + 2q - r + 2p = 0.$$

Following a similar procedure as in Theorem 4.2 and taking Lemma 4.1(ii) into account, we obtain

$$(4.18) \quad \zeta := \min \left\{ \frac{2q - r + 2p}{5q - 2r + p}, \frac{5q - 2r + p}{4q - r} \right\} \leq |\lambda - 2| \leq \frac{4q - r}{q} = 3,$$

since the coefficients in (4.17) are positive and $q/r = 1$ for $r \neq 0$. We obtain the lower bound on $|\lambda|$ from (4.16) with $\lambda = s + 1$:

$$|\lambda| = |s + 1| \geq \frac{r|s|^2}{q|s|^2 + p} =: \tau.$$

To show clustering around 1, we get $x = -A^{-1}B^\top y - A^{-1}C^\top z/(1-\lambda)$ from (4.13), insert it into (4.14) and (4.15) and multiply the resulting expressions with y^\top and z^\top , respectively. This leads to

$$(4.19) \quad a\lambda^2 - 2a\lambda + a + b = 0, \quad a = y^\top S_B y, \quad b = y^\top B A^{-1} C^\top z = z^\top C A^{-1} B^\top y,$$

$$(4.20) \quad c\lambda^2 - (2c + b)\lambda + c + b + d = 0, \quad c = z^\top Dz, \quad d = z^\top CA^{-1}C^\top z.$$

Solutions of equations (4.19) and (4.20) are

$$\lambda_{1,2} = 1 \pm \sqrt{-b/a}, \quad \lambda_{1,2} = 1 + \left(b/c \pm \sqrt{(b/c)^2 - 4d/c} \right) / 2.$$

Since b and d converge to zero for $h \rightarrow 0$, the eigenvalues are clustered around 1. \square

In Figure 3, we observe clustering around 1 for both matrices $\mathcal{A}_{\text{BJ}}\mathcal{P}_{\text{tri}}^{-1}$ and $\mathcal{A}_{\text{BJS}}\mathcal{P}_{\text{tri}}^{-1}$.

THEOREM 4.4. *The preconditioned matrix $\mathcal{A}_{\text{BJS}}\mathcal{P}_{\text{con}}^{-1}$ defined in (2.10) with the constraint preconditioner \mathcal{P}_{con} defined in (3.5) has either the eigenvalue $\lambda = 1$ or $\lambda = \left(1 + \eta \pm \sqrt{(\eta - 1)^2 - \xi} \right) / 2$.*

Proof. Let λ be an eigenvalue of $\mathcal{A}_{\text{BJS}}\mathcal{P}_{\text{con}}^{-1}$ to the eigenvector $(x; y; z)^\top \neq 0$ such that

$$(4.21) \quad Ax + B^\top y + C^\top z = \lambda(Gx + B^\top y),$$

$$(4.22) \quad Bx = \lambda Bx,$$

$$(4.23) \quad Cx - Dz = -\lambda Dz.$$

We get $\lambda = 1$ with the corresponding eigenvector $(0; 0; z)^\top$ for $0 \neq z \in \ker(C^\top)$. If $\lambda \neq 1$, it holds $Bx = 0$. We obtain $z = D^{-1}Cx / (1 - \lambda)$ from (4.23), substitute it into (4.21) and multiply the resulting expression from the left with x^\top . Rearranging the terms yields

$$(4.24) \quad \lambda^2 - \lambda(\eta + 1) + (\eta + \xi) = 0, \quad \eta = \frac{x^\top Ax}{x^\top Gx} > 0, \quad \xi = \frac{x^\top C^\top D^{-1}Cx}{x^\top Gx} \geq 0,$$

with the roots

$$\lambda_{1,2} = \left(1 + \eta \pm \sqrt{(\eta - 1)^2 - \xi} \right) / 2. \quad \square$$

Note that ξ defined in (4.24) is of order $O(h^2)$, because the entries of C are of order $O(h)$. In Figure 3, we observe clustering around 1 and η , where η depends on the choice of G .

4.2. Field-of-values analysis. To guarantee convergence of the Krylov subspace methods (FGMRES) independent of the grid width h (dimension $n + m + l$) for non-symmetric matrices, a bounded spectrum of the preconditioned matrix may not be sufficient [21]. Therefore, a bound on the GMRES residuals depending on the H -field-of-values (H -FOV) is needed [27].

Here, we show $\mathcal{A}_{\text{BJS}} \sim_{H^{-1}} \mathcal{P}$ for $\mathcal{P} \in \{\mathcal{P}_{\text{diag}}, \mathcal{P}_{\text{tri}}, \mathcal{P}_{\text{con}}\}$ and $\mathcal{A}_{\text{BJS}} \approx_{H^{-1}} \mathcal{P}$ for $\mathcal{P} \in \{\mathcal{P}_{\text{tri}}, \mathcal{P}_{\text{con}}\}$ following the ideas from [7]. This yields convergence in H^{-1} -norm, which is the natural norm for the right preconditioning [27]. We consider

$$(4.25) \quad H = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix}, \quad H_1 = A, \quad H_2 = \begin{pmatrix} S_B & 0 \\ 0 & D \end{pmatrix}.$$

Since $A = A^\top \succ 0$, $S_B = S_B^\top \succ 0$ and $D = D^\top \succ 0$, the matrices H_1 and H_2 are also symmetric positive definite.

Remark 4.5. Note that if $\mathcal{A}_{\text{BJS}} \sim_{H^{-1}} H$ and $H \sim_{H^{-1}} \mathcal{P}$, then $\mathcal{A}_{\text{BJS}} \sim_{H^{-1}} \mathcal{P}$ due to the transitivity of the H -norm equivalence (Definition A.1).

LEMMA 4.6. *For the matrices \mathcal{A}_{BJS} and H defined in (2.10) and (4.25), respectively, it holds $\mathcal{A}_{\text{BJS}} \sim_{H^{-1}} H$.*

Proof. From (4.2) we get $\|H^{-1}\mathcal{A}_{\text{BJS}}\|_H = \|H^{-1/2}\mathcal{A}_{\text{BJS}}H^{-1/2}\|_2$ and the bound

$$(4.26) \quad \begin{aligned} \|H^{-1/2}\mathcal{A}_{\text{BJS}}H^{-1/2}\|_2 &\leq \|I\|_2 + \underbrace{\|A^{-1/2}B^\top S_B^{-1/2}\|_2}_{=1 \text{ (Lemma 4.1)}} + \|A^{-1/2}C^\top D^{-1/2}\|_2 \\ &\quad + \underbrace{\|S_B^{-1/2}BA^{-1/2}\|_2}_{=1 \text{ (Lemma 4.1)}} + \|D^{-1/2}CA^{-1/2}\|_2 + \|-I\|_2. \end{aligned}$$

Following similar algebraic manipulations as in the proof of Lemma 4.1(i), we obtain one for the second and fourth term in the right-hand side of equation (4.26). For the third and fifth term, we get

$$(4.27) \quad \|A^{-1/2}C^\top D^{-1/2}\|_2 = \|D^{-1/2}CA^{-1/2}\|_2 = \sqrt{\lambda_{\max}(A^{-1/2}C^\top D^{-1}CA^{-1/2})}.$$

Due to $\lambda_{\max}(A^{-1/2}C^\top D^{-1}CA^{-1/2}) = O(h^2)$, $\exists \epsilon > 0$ s.t. $\|A^{-1/2}C^\top D^{-1/2}\|_2 < \epsilon$.

To find a lower bound to $\|H^{-1/2}\mathcal{A}_{\text{BJS}}H^{-1/2}\|_2$, we split the matrix in the form $(H^{-1/2}\mathcal{A}_{\text{BJS}}H^{-1/2})^\top (H^{-1/2}\mathcal{A}_{\text{BJS}}H^{-1/2}) = \mathcal{M} + \mathcal{N}$ as

$$(4.28) \quad \mathcal{M} = \left(\begin{array}{cc|c} I + A^{-1/2}B^\top S_B^{-1}BA^{-1/2} & A^{-1/2}B^\top S_B^{-1/2} & 0 \\ S_B^{-1/2}BA^{-1/2} & I & 0 \\ \hline 0 & 0 & I \end{array} \right) = \begin{pmatrix} \overline{\mathcal{M}} & 0 \\ 0 & I \end{pmatrix},$$

$$\mathcal{N} = \begin{pmatrix} A^{-1/2}C^\top D^{-1}CA^{-1/2} & 0 & 0 \\ 0 & 0 & S_B^{-1/2}BA^{-1}C^\top D^{-1/2} \\ 0 & D^{-1/2}CA^{-1}B^\top S_B^{-1/2} & D^{-1/2}CA^{-1}C^\top D^{-1/2} \end{pmatrix}.$$

Using [24, Theorem 4.3.1], we obtain

$$(4.29) \quad \|H^{-1/2}\mathcal{A}_{\text{BJS}}H^{-1/2}\|_2^2 \geq \lambda_{\min}(\mathcal{M}) + \lambda_{\max}(\mathcal{N}) \geq \lambda_{\min}(\mathcal{M}).$$

The matrix \mathcal{M} given in (4.28) has the eigenvalue $\lambda = 1$ and the eigenvalues of the block $\overline{\mathcal{M}}$. Since the block $A^{-1/2}B^\top S_B^{-1}BA^{-1/2}$ in $\overline{\mathcal{M}}$ is a projector, we get

$$((\overline{\mathcal{M}} - I)^2 - (\overline{\mathcal{M}} - I))^2 = (\overline{\mathcal{M}} - I)^2 - (\overline{\mathcal{M}} - I).$$

Therefore, $\overline{\mathcal{M}}$ has at most the four distinct eigenvalues

$$\lambda_1 = (3 - \sqrt{5})/2, \quad \lambda_2 = 1, \quad \lambda_3 = 2, \quad \lambda_4 = (3 + \sqrt{5})/2,$$

leading to $\|H^{-1/2}\mathcal{A}_{\text{BJS}}H^{-1/2}\|_2 \geq \sqrt{\lambda_1}$. This completes the proof. \square

THEOREM 4.7. *For the matrices \mathcal{A}_{BJS} , $\mathcal{P}_{\text{diag}}$ and H defined in (2.10), (3.3) and (4.25), respectively, it holds $\mathcal{A}_{\text{BJS}} \sim_{H^{-1}} \mathcal{P}_{\text{diag}}$.*

Proof. Taking Remark 4.5 and Lemma 4.6 into account, it is sufficient to show $H \sim_{H^{-1}} \mathcal{P}_{\text{diag}}$. Using (4.2) we get $\|\mathcal{P}_{\text{diag}}H^{-1}\|_{H^{-1}} = \|H^{-1/2}\mathcal{P}_{\text{diag}}H^{-1/2}\|_2$ and $\|H\mathcal{P}_{\text{diag}}^{-1}\|_{H^{-1}} = \|H^{1/2}\mathcal{P}_{\text{diag}}^{-1}H^{1/2}\|_2$. A simple calculation yields

$$H^{-1/2}\mathcal{P}_{\text{diag}}H^{-1/2} = H^{1/2}\mathcal{P}_{\text{diag}}^{-1}H^{1/2},$$

which provides the upper bounds $\|H^{-1/2}\mathcal{P}_{\text{diag}}H^{-1/2}\|_2 = \|H^{1/2}\mathcal{P}_{\text{diag}}^{-1}H^{1/2}\|_2 \leq 1$. \square

THEOREM 4.8. *For the matrices \mathcal{A}_{BJS} , \mathcal{P}_{tri} and H defined in (2.10), (3.3) and (4.25), respectively, it holds $\mathcal{A}_{\text{BJS}} \sim_{H^{-1}} \mathcal{P}_{\text{tri}}$.*

Proof. As in Theorem 4.7, we show that $H \sim_{H^{-1}} \mathcal{P}_{\text{tri}}$. Again with (4.2), we obtain $\|\mathcal{P}_{\text{tri}}H^{-1}\|_{H^{-1}} = \|H^{-1/2}\mathcal{P}_{\text{tri}}H^{-1/2}\|_2$ and $\|H\mathcal{P}_{\text{tri}}^{-1}\|_{H^{-1}} = \|H^{1/2}\mathcal{P}_{\text{tri}}^{-1}H^{1/2}\|_2$. Algebraic manipulations yield

$$H^{-1/2}\mathcal{P}_{\text{tri}}H^{-1/2} = H^{1/2}\mathcal{P}_{\text{tri}}^{-1}H^{1/2}.$$

For the matrix $(H^{-1/2}\mathcal{P}_{\text{tri}}H^{-1/2})^\top (H^{-1/2}\mathcal{P}_{\text{tri}}H^{-1/2})$, we get at most four distinct eigenvalues using similar argumentation as in the proof of Lemma 4.6:

$$\lambda_1 = (3 - \sqrt{5})/2, \quad \lambda_2 = 1, \quad \lambda_3 = 2, \quad \lambda_4 = (3 + \sqrt{5})/2.$$

This leads to $\|H^{-1/2}\mathcal{P}_{\text{tri}}H^{-1/2}\|_2 = \|H^{1/2}\mathcal{P}_{\text{tri}}^{-1}H^{1/2}\|_2 \leq \sqrt{\lambda_4}$. \square

THEOREM 4.9. *For the matrices \mathcal{A}_{BJS} , \mathcal{P}_{con} , G and H defined in (2.10), (3.5) and (4.25), respectively, the following inference holds $G \sim_{H_1^{-1}} H_1 \Rightarrow \mathcal{A}_{\text{BJS}} \sim_{H^{-1}} \mathcal{P}_{\text{con}}$.*

Proof. Again, it is sufficient to show $H \sim_{H^{-1}} \mathcal{P}_{\text{con}}$. As above, we use (4.2) and get $\|\mathcal{P}_{\text{con}}H^{-1}\|_{H^{-1}} = \|H^{-1/2}\mathcal{P}_{\text{con}}H^{-1/2}\|_2$ and $\|H\mathcal{P}_{\text{con}}^{-1}\|_{H^{-1}} = \|H^{1/2}\mathcal{P}_{\text{con}}^{-1}H^{1/2}\|_2$. We consider

$$(4.30) \quad H^{-1/2}\mathcal{P}_{\text{con}}H^{-1/2} = \begin{pmatrix} H_1^{-1/2}GH_1^{-1/2} & H_1^{-1/2}\overline{B}^\top H_2^{-1/2} \\ H_2^{-1/2}\overline{B}H_1^{-1/2} & -H_2^{-1/2}\overline{D}H_2^{-1/2} \end{pmatrix},$$

where \overline{B} and \overline{D} are defined in (3.4). We obtain the following estimate

$$\begin{aligned} \|H^{-1/2}\mathcal{P}_{\text{con}}H^{-1/2}\|_2 &\leq \underbrace{\|H_1^{-1/2}GH_1^{-1/2}\|_2}_{\leq \beta_1} + \underbrace{\|H_1^{-1/2}\overline{B}^\top H_2^{-1/2}\|_2}_{=1 \text{ (Lemma 4.1)}} + \underbrace{\|H_2^{-1/2}\overline{B}H_1^{-1/2}\|_2}_{=1 \text{ (Lemma 4.1)}} \\ &\quad + \underbrace{\|H_2^{-1/2}\overline{D}H_2^{-1/2}\|_2}_{=1} \leq \beta_1 + 3, \end{aligned}$$

where the constant $\beta_1 > 0$ arises from the assumption $G \sim_{H_1^{-1}} H_1$:

$$(4.31) \quad \|GH_1^{-1}\|_{H_1^{-1}} \leq \beta_1, \quad \|H_1G^{-1}\|_{H_1^{-1}} \leq 1/\alpha_1.$$

To estimate an upper bound for $\|H\mathcal{P}_{\text{con}}^{-1}\|_{H^{-1}}$, we compute the inverse

$$\mathcal{P}_{\text{con}}^{-1} = \begin{pmatrix} G^{-1} - G^{-1}\overline{B}^\top (\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1}\overline{B}G^{-1} & G^{-1}\overline{B}^\top (\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1} \\ (\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1}\overline{B}G^{-1} & -(\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1} \end{pmatrix}$$

and obtain

$$\begin{aligned} \|H^{1/2}\mathcal{P}_{\text{con}}^{-1}H^{1/2}\|_2 &\leq \underbrace{\|H_1^{1/2}G^{-1}H_1^{1/2}\|_2}_{\leq 1/\alpha_1 \text{ (eq. (4.31))}} + \|H_1^{1/2}G^{-1}\overline{B}^\top (\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1}\overline{B}G^{-1}H_1^{1/2}\|_2 \\ &\quad + \|H_1^{1/2}G^{-1}\overline{B}^\top (\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1}H_2^{1/2}\|_2 + \|H_2^{1/2}(\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1}\overline{B}G^{-1}H_1^{1/2}\|_2 \\ &\quad + \|H_2^{1/2}(\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1}H_2^{1/2}\|_2. \end{aligned}$$

The last term is bounded as

$$(4.32) \quad \|H_2^{1/2}(\overline{D} + \overline{B}G^{-1}\overline{B}^\top)^{-1}H_2^{1/2}\|_2 \leq \|S_B^{1/2}(BG^{-1}B^\top)^{-1}S_B^{1/2}\|_2 + \|I\|_2 \leq \delta_1 + 1,$$

where the constant $\delta_1 > 0$ arises from the assumption $G \sim_{H_1^{-1}} H_1$. We bound the second term as follows

$$\begin{aligned} & \|H_1^{1/2} G^{-1} \bar{B}^\top (\bar{D} + \bar{B} G^{-1} \bar{B}^\top)^{-1} \bar{B} G^{-1} H_1^{1/2}\|_2 \\ & \leq \|H_1^{1/2} G^{-1} H_1^{1/2}\|_2 \|H_1^{-1/2} \bar{B}^\top H_2^{-1/2}\|_2 \|H_2^{1/2} (\bar{D} + \bar{B} G^{-1} \bar{B}^\top)^{-1} H_2^{1/2}\|_2 \times \\ & \quad \|H_2^{-1/2} \bar{B} H_1^{-1/2}\|_2 \|H_1^{1/2} G^{-1} H_1^{1/2}\|_2 \leq \alpha_1^{-1} \cdot 1 \cdot (\delta_1 + 1) \cdot 1 \cdot \alpha_1^{-1} = (\delta_1 + 1) \alpha_1^{-2}. \end{aligned}$$

The third and fourth term are bounded in a similar manner completing the proof. \square

In the next step, we show the H -FOV equivalence of the matrix \mathcal{A}_{BJS} and the proposed preconditioners \mathcal{P}_{tri} and \mathcal{P}_{con} . The case of the block diagonal preconditioner $\mathcal{P}_{\text{diag}}$ is beyond the scope of this work.

THEOREM 4.10. *For the matrices \mathcal{A}_{BJS} , \mathcal{P}_{tri} , H defined in (2.10), (3.3), (4.25), respectively, it holds $\mathcal{A}_{\text{BJS}} \approx_{H^{-1}} \mathcal{P}_{\text{tri}}$.*

Proof. Note that, the upper bound in Definition A.2 is already shown in Theorem 4.8. To obtain the lower bound $x^\top H^{-1} \mathcal{A}_{\text{BJS}} \mathcal{P}_{\text{tri}}^{-1} x \geq \alpha x^\top H^{-1} x$, we take $x = (x_1; x_2; x_3)^\top$ and compute

$$(4.33) \quad x^\top H^{-1} \mathcal{A}_{\text{BJS}} \mathcal{P}_{\text{tri}}^{-1} x = \underbrace{x_1^\top A^{-1} x_1}_{=\|x_1\|_{A^{-1}}^2} + x_2^\top S_B^{-1} B A^{-1} x_1 + \underbrace{x_2^\top S_B^{-1} x_2}_{=\|x_2\|_{S_B^{-1}}^2} + \underbrace{x_3^\top D^{-1} x_3}_{=\|x_3\|_{D^{-1}}^2}.$$

We bound the second term in (4.33) as

$$\begin{aligned} |x_2^\top S_B^{-1} B A^{-1} x_1| & \leq \|B A^{-1}\|_{A^{-1}, S_B^{-1}} \|x_1\|_{A^{-1}} \|x_2\|_{S_B^{-1}} \\ & = \|S_B^{-1/2} B A^{-1/2}\|_2 \|x_1\|_{A^{-1}} \|x_2\|_{S_B^{-1}} \stackrel{\text{Lemma 4.1}}{=} \|x_1\|_{A^{-1}} \|x_2\|_{S_B^{-1}}. \end{aligned}$$

Thus, with $\|x_1\|_{A^{-1}}^2 + \|x_2\|_{S_B^{-1}}^2 \geq 2\|x_1\|_{A^{-1}} \|x_2\|_{S_B^{-1}}$, we get

$$\begin{aligned} x^\top H^{-1} \mathcal{A}_{\text{BJS}} \mathcal{P}_{\text{tri}}^{-1} x & \geq \|x_1\|_{A^{-1}}^2 - \|x_1\|_{A^{-1}} \|x_2\|_{S_B^{-1}} + \|x_2\|_{S_B^{-1}}^2 + \|x_3\|_{D^{-1}}^2 \\ & \geq \frac{1}{2} \left(\|x_1\|_{A^{-1}}^2 + \|x_2\|_{S_B^{-1}}^2 + \|x_3\|_{D^{-1}}^2 \right). \quad \square \end{aligned}$$

THEOREM 4.11. *Let the matrices \mathcal{A}_{BJS} , \mathcal{P}_{con} , G be defined as in (2.10), (3.5) and let the block H_1 from (4.25) be scaled by $\rho > 0$ as $H = \text{diag}(\rho H_1, H_2)$. Assume $A \approx_{H_1^{-1}} G$ and there exists $\rho_1 > 0$ such that $\|I - A G^{-1}\|_{H_1^{-1}} \leq \rho_1$. Then, it holds $\mathcal{A}_{\text{BJS}} \approx_{H^{-1}} \mathcal{P}_{\text{con}}$ for all $\rho \geq \rho_1$.*

Proof. We obtain the upper bound following the proof of Theorem 4.9 and using the fact that the block H_1 is scaled by ρ (see equation (4.30)). To get the lower bound, we define $S_G := B G^{-1} B^\top$ and compute for $x = (x_1; x_2; x_3)^\top$:

$$\begin{aligned} x^\top H^{-1} \mathcal{A}_{\text{BJS}} \mathcal{P}_{\text{con}}^{-1} x & = \rho^{-1} x_1^\top A^{-1} A G^{-1} x_1 + \rho^{-1} x_1^\top A^{-1} (I - A G^{-1}) B^\top S_G^{-1} B G^{-1} x_1 \\ & \quad + x_1^\top A^{-1} (A G^{-1} - I) B^\top S_G^{-1} x_2 - \rho^{-1} x_1^\top A^{-1} C^\top D^{-1} x_3 + x_2^\top S_B^{-1} x_2 \\ & \quad + x_3^\top D^{-1} C G^{-1} (I - B^\top S_G^{-1} B G^{-1}) x_1 + x_3^\top D^{-1} C G^{-1} B^\top S_G^{-1} x_2 + x_3^\top D^{-1} x_3. \end{aligned}$$

For the first term in the right-hand side, due to the assumption $A \approx_{H_1^{-1}} G$, we have bounds (4.31) and there exists α_0 such that

$$(4.34) \quad x_1^\top H_1^{-1} A G^{-1} x_1 \geq \alpha_0 \|x_1\|_{H_1^{-1}}^2.$$

For the second term, we get the following bound

$$\begin{aligned}
& | \langle (I - AG^{-1})B^\top S_G^{-1}BG^{-1}x_1, x_1 \rangle_{H_1^{-1}} | \\
& \leq \underbrace{\|I - AG^{-1}\|_{H_1^{-1}}}_{\leq \rho_1} \underbrace{\|A^{-1/2}B^\top S_B^{-1/2}\|_2}_{=1 \text{ (Lemma 4.1)}} \underbrace{\|S_B^{1/2}S_G^{-1}S_B^{1/2}\|_2}_{\leq \delta_1 \text{ in (4.32)}} \underbrace{\|S_B^{-1/2}BA^{-1/2}\|_2}_{=1 \text{ (Lemma 4.1)}} \underbrace{\|AG^{-1}\|_{H_1^{-1}}}_{\leq \alpha_1^{-1} \text{ in (4.31)}} \|x_1\|_{H_1^{-1}}^2 \\
& \leq \rho_1 \delta_1 \alpha_1^{-1} \|x_1\|_{H_1^{-1}}^2 \leq \rho \delta_1 \alpha_1^{-1} \|x_1\|_{H_1^{-1}}^2.
\end{aligned}$$

We bound the rest of the terms in a similar manner. This results in

$$\begin{aligned}
x^\top H^{-1} \mathcal{A}_{\text{BJS}} \mathcal{P}_{\text{con}}^{-1} x & \geq (\rho^{-1} \alpha_0 - \delta_1 \alpha_1^{-1}) \|x_1\|_{A^{-1}}^2 - \delta_1 \|x_1\|_{A^{-1}} \|x_2\|_{S_B^{-1}} + \|x_2\|_{S_B^{-1}}^2 \\
& \quad - (\rho^{-1} \epsilon + \epsilon \alpha_1^{-1} + \epsilon \alpha_1^{-2} \delta_1) \|x_1\|_{A^{-1}} \|x_3\|_{D^{-1}} - \epsilon \alpha_1^{-1} \delta_1 \|x_2\|_{S_B^{-1}} \|x_3\|_{D^{-1}} + \|x_3\|_{D^{-1}}^2 \\
& = \frac{1}{2} \left(\epsilon \alpha_1^{-1} \delta_1 \|x_2\|_{S_B^{-1}} - \|x_3\|_{D^{-1}} \right)^2 \\
& \quad + \left(1 - \left(\frac{\epsilon \delta_1}{\sqrt{2} \alpha_1} \right)^2 \right) \|x_2\|_{S_B^{-1}}^2 + \frac{1}{2} \left(\frac{\alpha_0}{\rho} - \frac{\delta_1}{\alpha_1} \right) \|x_1\|_{A^{-1}}^2 - \delta_1 \|x_1\|_{A^{-1}} \|x_2\|_{S_B^{-1}} \\
& \quad + \frac{1}{2} \|x_3\|_{D^{-1}}^2 + \frac{1}{2} \left(\frac{\alpha_0}{\rho} - \frac{\delta_1}{\alpha_1} \right) \|x_1\|_{A^{-1}}^2 - \epsilon \left(\frac{1}{\rho} + \frac{1}{\alpha_1} + \frac{\delta_1}{\alpha_1^2} \right) \|x_1\|_{A^{-1}} \|x_3\|_{D^{-1}}.
\end{aligned}$$

We divide the second, the third and the fourth term in the right-hand side in the equation above by $\theta := 1 - (\epsilon \delta_1 / \sqrt{2} \alpha_1)^2$ and estimate these terms as follows

$$\frac{1}{2\theta} \left(\frac{\alpha_0}{\rho} - \frac{\delta_1}{\alpha_1} \right) \|x_1\|_{A^{-1}}^2 - \frac{\delta_1}{\theta} \|x_1\|_{A^{-1}} \|x_2\|_{S_B^{-1}} + \|x_2\|_{S_B^{-1}}^2 \geq \frac{1}{2} \left(\|x_1\|_{A^{-1}} + \|x_2\|_{S_B^{-1}}^2 \right),$$

where $\rho := \alpha_0 / (\theta + \delta_1 / \alpha_1 + \delta_1^2 / \theta)$. The remaining terms are estimated as follows

$$\begin{aligned}
& \frac{1}{2} \|x_3\|_{D^{-1}}^2 + \frac{1}{2} \left(\frac{\alpha_0}{\rho} - \frac{\delta_1}{\alpha_1} \right) \|x_1\|_{A^{-1}}^2 - \epsilon \left(\frac{1}{\rho} + \frac{1}{\alpha_1} + \frac{\delta_1}{\alpha_1^2} \right) \|x_1\|_{A^{-1}} \|x_3\|_{D^{-1}} \\
& \geq \frac{1}{2} \|x_3\|_{D^{-1}}^2 + \frac{1}{2} \left(\frac{\alpha_0}{\rho} - \frac{\delta_1}{\alpha_1} \right) \|x_1\|_{A^{-1}}^2 - \frac{\epsilon}{2} \left(\frac{1}{\rho} + \frac{1}{\alpha_1} + \frac{\delta_1}{\alpha_1^2} \right) (\|x_1\|_{A^{-1}}^2 + \|x_3\|_{D^{-1}}^2) \\
& = \frac{1}{2} \left(\frac{\alpha_0}{\rho} - \frac{\delta_1}{\alpha_1} - \frac{\epsilon}{\rho} - \frac{\epsilon}{\alpha_1} - \frac{\epsilon \delta_1}{\alpha_1^2} \right) \|x_1\|_{A^{-1}}^2 + \frac{1}{2} \left(1 - \frac{\epsilon}{\rho} - \frac{\epsilon}{\alpha_1} - \frac{\epsilon \delta_1}{\alpha_1^2} \right) \|x_3\|_{D^{-1}}^2 \\
& = \tau_1 \|x_1\|_{A^{-1}}^2 + \tau_2 \|x_3\|_{D^{-1}}^2.
\end{aligned}$$

Taking (4.27) into account, we get $\epsilon = O(h)$ and thus for $h \rightarrow 0$ we have $\tau_1, \tau_2 > 0$. In conclusion, we get $x^\top H^{-1} \mathcal{A}_{\text{BJS}} \mathcal{P}_{\text{con}}^{-1} x \geq \tau \left(\|x_1\|_{A^{-1}}^2 + \|x_2\|_{S_B^{-1}}^2 + \|x_3\|_{D^{-1}}^2 \right)$ with $\tau = \min\{\theta/2, \tau_2\}$. \square

For every spectrally equivalent preconditioner G to the block A , the established bounds in Theorem 4.9 and Theorem 4.11 hold true.

THEOREM 4.12. *Let G be spectrally equivalent to $H_1 = A$. Then $G \sim_{H_1^{-1}} H_1$ and $H_1 \approx_{H_1^{-1}} G$.*

Proof. The matrices G and H_1 are spectrally equivalent, i.e. there exist constants $\alpha_0, \beta_0 > 0$ such that for all $x \in \mathbb{R}^n \setminus \{0\}$:

$$\alpha_0 \leq \frac{\langle H_1 x, x \rangle}{\langle G x, x \rangle} \leq \beta_0 \quad \overset{x=G^{-1}y}{\iff} \quad \alpha_0 \leq \frac{\langle H_1 G^{-1}y, G^{-1}y \rangle}{\langle y, H_1^{-1}y \rangle} \frac{\langle y, H_1^{-1}y \rangle}{\langle y, G^{-1}y \rangle} \leq \beta_0.$$

The spectral equivalence of H_1 and G induces the spectral equivalence of the inverses H_1^{-1} and G^{-1} . We conclude that $\|H_1 G^{-1}\|_{H_1^{-1}}$ is bounded from below and above, i.e. $G \sim_{H_1^{-1}} H_1$. The H_1 -FOV of H_1 and G follows directly from the spectral equivalence of H_1 and G . \square

5. Numerical results. In this section, we present numerical simulation results for two coupled Stokes–Darcy problems (2.1)–(2.4): (i) *Problem \mathcal{A}_{BJ}* is completed with the conservation of mass (2.5), the balance of normal forces (2.6) and the Beavers–Joseph condition (2.7) on the tangential velocity, and (ii) *Problem \mathcal{A}_{BJS}* with the coupling conditions (2.5), (2.6) and the Beavers–Joseph–Saffman condition (2.8).

5.1. Benchmark problem. We consider the coupled flow domain $\bar{\Omega} = \bar{\Omega}_{\text{pm}} \cup \bar{\Omega}_{\text{ff}} \subset \mathbb{R}^2$ with $\bar{\Omega}_{\text{pm}} = [0, 1] \times [-0.5, 0]$ and $\bar{\Omega}_{\text{ff}} = [0, 1] \times [0, 0.5]$ separated by the flat fluid–porous interface $\Sigma = (0, 1) \times \{0\}$. We consider an isotropic porous medium, i.e. $\mathbf{K} = k\mathbf{I}$, $k > 0$. To investigate the robustness of the preconditioners, we consider different values for the dynamic viscosity μ , the intrinsic permeability k and the Beavers–Joseph slip coefficient α .

The exact solution of *Problem \mathcal{A}_{BJ}* and *Problem \mathcal{A}_{BJS}* is chosen as

$$(5.1) \quad \begin{aligned} u_{\text{ff}}(x_1, x_2) &= -\cos(\pi x_1) \sin(\pi x_2), & v_{\text{ff}}(x_1, x_2) &= \sin(\pi x_1) \cos(\pi x_2), \\ p_{\text{ff}}(x_1, x_2) &= x_2 \sin(\pi x_1)/2, & p_{\text{pm}}(x_1, x_2) &= x_2^2 \sin(\pi x_1)/2. \end{aligned}$$

The right-hand sides \mathbf{f}_{ff} , f_{pm} and the boundary conditions $\bar{\mathbf{v}}$, \bar{p} are defined by substitution of the chosen physical parameters and the exact solution (5.1) into the corresponding Stokes–Darcy problem.

5.2. Implementation. The Stokes–Darcy problems (*Problem \mathcal{A}_{BJ}* and *Problem \mathcal{A}_{BJS}*) are implemented using our in-house C++ code. To evaluate the eigenvalues, we use the `linalg.eig` method of NUMPY in Python 3.9. We solve the original and the preconditioned systems with FGMRES(20) considering exact and inexact versions of the preconditioners. The AMG method for the inexact versions is implemented using the MATLAB toolbox IFISS [18, 35].

The stopping criterion is the maximum number of iteration steps $n_{\text{max}} = 2000$ or $\|\mathcal{A}\mathbf{x}_n - \mathbf{b}\|_2 \leq \varepsilon_{\text{tol}} \|\mathbf{b}\|_2$ for the tolerance $\varepsilon_{\text{tol}} = 10^{-8}$. The initial solution is always $\mathbf{x}_0 = \mathbf{0}$. All computations are carried out on a laptop with an 12th Gen Intel(R) Core(TM) i7 1255U processor and 2× 16GB RAM using MATLAB.R2019b.

5.3. Numerical simulation results. In this section, we first present the eigenvalue distribution for the original and the preconditioned Stokes–Darcy systems. Then, we compare exact and inexact versions of the preconditioners. Finally, we provide the efficiency and robustness study of the proposed preconditioners.

5.3.1. Eigenvalue distribution. Here, we choose the parameters $\mu = 10^{-3}$, $k = 10^{-2}$ and $\alpha = 1$. In Figure 3, we plot the eigenvalues for the original matrices $\mathcal{A} \in \{\mathcal{A}_{BJ}, \mathcal{A}_{BJS}\}$ given in (2.10) and for the corresponding exact preconditioned matrices $\mathcal{A}\mathcal{P}^{-1}$ with $\mathcal{P} \in \{\mathcal{P}_{\text{diag}}, \mathcal{P}_{\text{tri}}, \mathcal{P}_{\text{con}}\}$ from (3.3), (3.5). As proven in subsection 4.1, the eigenvalues are clustered and bounded away from zero (Theorems 4.2 to 4.4). Therefore, all three developed preconditioners significantly improve the eigenvalue distributions of the original system.

5.3.2. Efficiency analysis. We consider here the same physical parameters as in the previous section: $\mu = 10^{-3}$, $k = 10^{-2}$ and $\alpha = 1$. To study the efficiency of the developed preconditioners, we plot the relative residuals $\|\mathcal{A}\mathbf{x}_n - \mathbf{b}\|_2 / \|\mathbf{b}\|_2$ for

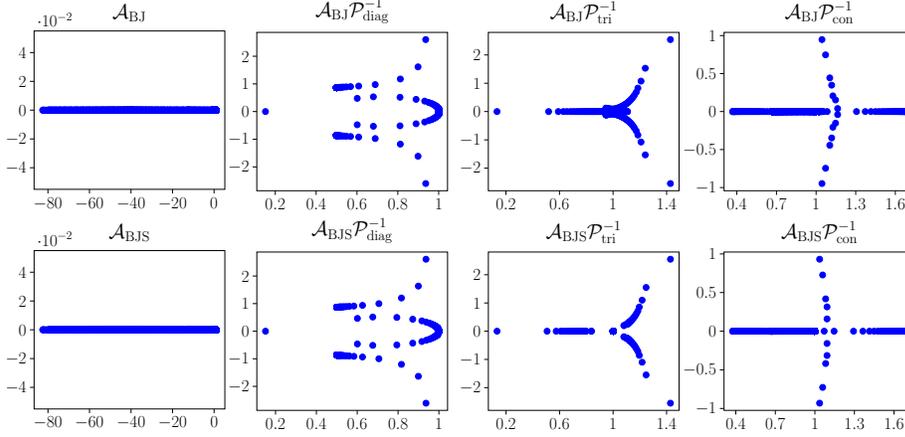


Fig. 3: Eigenvalue distributions for the matrices \mathcal{A}_{BJ} and \mathcal{A}_{BJS} and the corresponding preconditioned matrices for $h = 1/40$

$\mathcal{A} \in \{\mathcal{A}_{\text{BJ}}, \mathcal{A}_{\text{BJS}}\}$ against the number of iterations until the stopping criterion given in subsection 5.2 is reached (Figure 4). The number of iterations n and the CPU times are presented in Table 1. For the inexact versions of the preconditioners $\hat{\mathcal{P}}$, the CPU time is composed of two parts: (i) time to construct the algebraic grids for A_{11} , A_{22} and D given in (2.10) and (3.5), and (ii) time to solve the linear system (3.1) using the generated algebraic grids. Note that, even though the exact preconditioners $\mathcal{P}_{\text{diag}}$, \mathcal{P}_{tri} and \mathcal{P}_{con} require less iteration steps (Figure 4), the CPU times are significantly higher than for the inexact versions of the preconditioners $\hat{\mathcal{P}}_{\text{diag}}$, $\hat{\mathcal{P}}_{\text{tri}}$ and $\hat{\mathcal{P}}_{\text{con}}$ (Table 1).

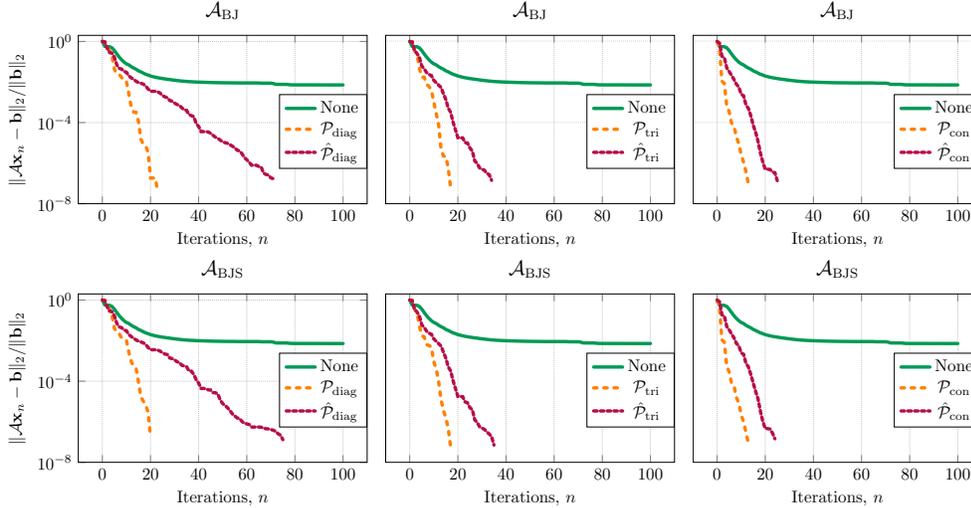


Fig. 4: Comparison of different preconditioners for *Problem \mathcal{A}_{BJ}* and *Problem \mathcal{A}_{BJS}* for $h = 1/80$: original system (None), exact preconditioner (\mathcal{P}) and inexact preconditioner ($\hat{\mathcal{P}}$)

Table 1: Computational costs to solve *Problem* \mathcal{A}_{BJ} and *Problem* \mathcal{A}_{BJS} for $h = 1/80$

| Preconditioner | \mathcal{A}_{BJ} | | \mathcal{A}_{BJS} | |
|----------------------------|--------------------|----------------------|---------------------|----------------------|
| | Iterations | CPU time [s] | Iterations | CPU time [s] |
| \mathcal{P}_{diag} | 22 | 19.00 | 21 | 18.81 |
| $\hat{\mathcal{P}}_{diag}$ | 68 | $0.38 + 0.32 = 0.7$ | 72 | $0.37 + 0.30 = 0.67$ |
| \mathcal{P}_{tri} | 17 | 18.67 | 17 | 18.98 |
| $\hat{\mathcal{P}}_{tri}$ | 33 | $0.38 + 0.19 = 0.57$ | 34 | $0.37 + 0.19 = 0.56$ |
| \mathcal{P}_{con} | 13 | 17.11 | 13 | 16.67 |
| $\hat{\mathcal{P}}_{con}$ | 24 | $0.38 + 0.13 = 0.51$ | 23 | $0.37 + 0.14 = 0.51$ |

5.3.3. Robustness analysis. An important property of preconditioners is the independence of the convergence rate of the iterative method from the grid width h . First, we fix the parameters $\mu = 10^{-3}$, $k = 10^{-2}$, $\alpha = 1$ and study the convergence for different grid widths h . Due to the large size of the linear systems, we provide the results only for the inexact versions of the preconditioners. As it can be seen in Table 2, the number of iterations stays nearly constant.

Table 2: Iterations to solve *Problem* \mathcal{A}_{BJ} and *Problem* \mathcal{A}_{BJS} for different grid widths h

| h | \mathcal{A}_{BJ} | | | \mathcal{A}_{BJS} | | |
|-------|----------------------------|---------------------------|---------------------------|----------------------------|---------------------------|---------------------------|
| | $\hat{\mathcal{P}}_{diag}$ | $\hat{\mathcal{P}}_{tri}$ | $\hat{\mathcal{P}}_{con}$ | $\hat{\mathcal{P}}_{diag}$ | $\hat{\mathcal{P}}_{tri}$ | $\hat{\mathcal{P}}_{con}$ |
| 1/10 | 74 | 35 | 27 | 75 | 35 | 26 |
| 1/20 | 75 | 35 | 26 | 72 | 35 | 27 |
| 1/40 | 71 | 34 | 26 | 71 | 34 | 26 |
| 1/80 | 68 | 33 | 24 | 72 | 34 | 23 |
| 1/160 | 68 | 28 | 20 | 71 | 28 | 20 |
| 1/320 | 53 | 27 | 19 | 54 | 27 | 19 |
| 1/640 | 51 | 24 | 18 | 51 | 24 | 18 |

To study robustness of the preconditioners with respect to physical parameters μ , k and α , we plot the number of iteration steps for *Problem* \mathcal{A}_{BJ} (Figure 5, left) and *Problem* \mathcal{A}_{BJS} (Figure 5, right). We consider the values of the intrinsic permeability $k \in \{10^{-3}, 10^{-2}, 10^{-1}\}$, the dynamic viscosity $\mu \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ and the Beavers–Joseph slip parameter $\alpha \in \{10^{-1}, 1, 10\}$. For highly permeable porous media with $k \in \{10^{-2}, 10^{-1}\}$, the number of iteration steps changes only slightly for different values of μ and α . However, this is not the case for low permeable porous media with $k = 10^{-3}$.

6. Conclusions. In this paper, we proposed and analysed three different preconditioners for coupled Stokes–Darcy systems: a block diagonal, a block triangular and a constraint preconditioner. We considered two classical sets of interface conditions with either the Beavers–Joseph (*Problem* \mathcal{A}_{BJ}) or the Beavers–Joseph–Saffman coupling condition (*Problem* \mathcal{A}_{BJS}). We applied the finite volume method on staggered grids (MAC scheme) to discretise the coupled Stokes–Darcy problems and used FGMRES(20) to solve the resulting linear systems.

We provided bounds on the spectrum and the field-of-values for the exact variants of the developed preconditioners that are independent of the grid width. To confirm the obtained theoretical results, we performed numerical experiments for two Stokes–Darcy problems. The numerical experiments show that both, the exact and

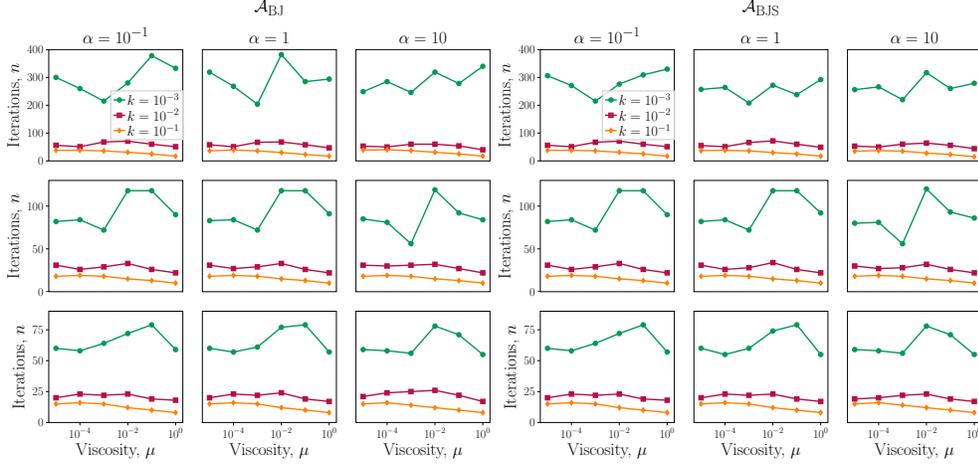


Fig. 5: Robustness analysis for the preconditioned system $\mathcal{A}_{\text{BJ}}\hat{\mathcal{P}}^{-1}$ (left) and $\mathcal{A}_{\text{BJS}}\hat{\mathcal{P}}^{-1}$ (right) with $\hat{\mathcal{P}}_{\text{diag}}$ (top), $\hat{\mathcal{P}}_{\text{tri}}$ (middle) and $\hat{\mathcal{P}}_{\text{con}}$ (bottom) for $h = 1/80$

inexact variants of the preconditioners, significantly improve the convergence rate of FGMRES. Even though the number of iteration steps is smaller for the exact variants of the preconditioners, the inexact versions yield remarkably smaller CPU times.

The preconditioners are robust with respect to varying grid width h . For highly permeable porous media, we obtain robustness for different values of the viscosity μ and the Beavers–Joseph slip coefficient α . However, with decreasing intrinsic permeability k the number of iteration steps increases. The development and analysis of robust preconditioners for low permeable porous media is the subject of future work.

Appendix A. Additional definitions and results.

DEFINITION A.1 (*H*-norm equivalence). Let $H = H^\top \succ 0$. Non-singular matrices $M, N \in \mathbb{R}^{k \times k}$ are *H*-norm equivalent ($M \sim_H N$) if there exist constants $\gamma, \Gamma > 0$, independent of dimension k such that [27]:

$$(A.1) \quad \gamma \leq \frac{\|Mx\|_H}{\|Nx\|_H} \leq \Gamma \quad \forall x \in \mathbb{R}^k \setminus \{0\}.$$

DEFINITION A.2 (*H*-FOV equivalence). Let $H = H^\top \succ 0$. Non-singular matrices $M, N \in \mathbb{R}^{k \times k}$ are *H*-FOV equivalent ($M \approx_H N$) if there exist constants $\gamma, \Gamma > 0$, independent of dimension k such that [27]:

$$(A.2) \quad \gamma \leq \frac{\langle x, MN^{-1}x \rangle_H}{\langle x, x \rangle_H}, \quad \frac{\|MN^{-1}x\|_H}{\|x\|_H} \leq \Gamma \quad \forall x \in \mathbb{R}^k \setminus \{0\}.$$

Note that if $M \approx_H N$ then $M \sim_H N$. The *H*-norm equivalence is reflexive, symmetric and transitive. For symmetric matrices $M = M^\top$ and $N = N^\top$, the *H*-FOV equivalence is symmetric, i.e. $M \approx_H N$ implies $N \approx_H M$.

THEOREM A.3. Let $M, N \in \mathbb{R}^{k \times k}$ be non-singular matrices and $H = H^\top \succ 0$. If $M \approx_H N$, the GMRES method converges with respect to $\langle \cdot, \cdot \rangle_H$ in a number of iter-

ation steps independent of dimension k . Moreover, the residuals satisfy [27, Alg. 2.2]:

$$\frac{\|r_s\|_H}{\|r_0\|_H} \leq \left(1 - \frac{\gamma^2}{\Gamma^2}\right)^{s/2},$$

where γ, Γ are the constants from Definition A.2 and s is the iteration step.

REFERENCES

- [1] P. ANTONIETTI, J. DE PONTI, L. FORMAGGIA, AND A. SCOTTI, *Preconditioning techniques for the numerical solution of flow in fractured porous media*, J. Sci. Comput., 86 (2021), <https://doi.org/10.1007/s10915-020-01372-0>.
- [2] O. AXELSSON AND M. NEYTCHIEVA, *Preconditioning methods for linear systems arising in constrained optimization problems*, Numer. Linear Algebra Appl., 10 (2003), pp. 3–31, <https://doi.org/10.1002/nla.310>.
- [3] G. BEAVERS AND D. JOSEPH, *Boundary conditions at a naturally permeable wall*, J. Fluid Mech., 30 (1967), pp. 197–207, <https://doi.org/10.1017/S0022112067001375>.
- [4] F. BEIK AND M. BENZI, *Iterative methods for double saddle point systems*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 902–921, <https://doi.org/10.1137/17M1121226>.
- [5] F. BEIK AND M. BENZI, *Preconditioning techniques for the coupled Stokes–Darcy problem: spectral and field-of-values analysis*, Numer. Math., 150 (2022), pp. 257–298, <https://doi.org/10.1007/s00211-021-01267-8>.
- [6] M. BENZI, *Preconditioning techniques for large linear systems: A survey*, J. Comput. Phys., 182 (2002), pp. 418–477, <https://doi.org/10.1006/jcph.2002.7176>.
- [7] M. BENZI, *Some uses of the field of values in numerical analysis*, Boll. Unione. Mat. Ital., 14 (2021), pp. 159–177, <https://doi.org/10.1007/s40574-020-00249-2>.
- [8] M. BENZI, G. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137, <https://doi.org/10.1017/S0962492904000212>.
- [9] W. BOON, T. KOCH, M. KUCHTA, AND K.-A. MARDAL, *Robust monolithic solvers for the Stokes–Darcy problem with the Darcy equation in primal form*, SIAM J. Sci. Comput., 44 (2022), pp. B1148–B1174, <https://doi.org/10.1137/21M1452974>.
- [10] J. BRAMBLE AND J. PASCIAK, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp., 50 (1988), pp. 1–17, <https://doi.org/10.2307/2007912>.
- [11] A. BUDIŠA AND X. HU, *Block preconditioners for mixed-dimensional discretization of flow in fractured porous media*, Comput. Geosci., 25 (2021), pp. 671–686, <https://doi.org/10.1007/s10596-020-09984-z>.
- [12] A. BUDIŠA, X. HU, M. KUCHTA, K.-A. MARDAL, AND L. ZIKATANOV, *Rational approximation preconditioners for multiphysics problems*, in Numerical Methods and Applications, I. Georgiev, M. Datcheva, K. Georgiev, and G. Nikolov, eds., Springer Nature Switzerland, 2023, pp. 100–113, https://doi.org/10.1007/978-3-031-32412-3_9.
- [13] M. CAI, M. MU, AND J. XU, *Preconditioning techniques for a mixed Stokes/Darcy model in porous media applications*, J. Comput. Appl. Math., 233 (2009), pp. 346–355, <https://doi.org/10.1016/j.cam.2009.07.029>.
- [14] P. CHIDYAGWAI, S. LADENHEIM, AND D. SZYLD, *Constraint preconditioning for the coupled Stokes–Darcy system*, SIAM J. Sci. Comput., 38 (2016), pp. A668–A690, <https://doi.org/10.1137/15M1032156>.
- [15] P. CHIDYAGWAI AND B. RIVIÈRE, *On the solution of the coupled Navier–Stokes and Darcy equations*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 3806–3820, <https://doi.org/10.1016/j.cma.2009.08.012>.
- [16] M. DISCACCIATI, E. MIGLIO, AND A. QUARTERONI, *Mathematical and numerical models for coupling surface and groundwater flows*, Appl. Num. Math., 43 (2002), pp. 57–74, [https://doi.org/10.1016/S0168-9274\(02\)00125-3](https://doi.org/10.1016/S0168-9274(02)00125-3).
- [17] H. ELMAN, *Preconditioning for the steady-state Navier–Stokes equations with low viscosity*, SIAM J. Sci. Comput., 20 (1999), pp. 1299–1316, <https://doi.org/10.1137/S1064827596312547>.
- [18] H. ELMAN AND D. SILVESTER, *Algorithm 866: IFISS, A Matlab toolbox for modelling incompressible flow*, ACM Trans. Math. Software, 33 (2007), pp. 14–es, <https://doi.org/10.1145/1236463.1236469>.
- [19] H. ELMAN, D. SILVESTER, AND A. WATHEN, *Finite Elements and Fast Iterative Solvers with Applications in Incompressible Fluid Dynamics*, Oxford Science Publications, 2014.

- [20] R. EYMARD, T. GALLOUËT, R. HERBIN, AND J.-C. LATCHÉ, *Convergence of the MAC scheme for the compressible Stokes equations*, SIAM J. Numer. Anal., 48 (2010), pp. 2218–2246, <https://doi.org/10.1137/090779863>.
- [21] A. GREENBAUM, V. PTÁK, AND Z. STRAKOŠ, *Any nonincreasing convergence curve is possible for GMRES*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 465–469, <https://doi.org/10.1137/S0895479894275030>.
- [22] C. GREIF AND Y. HE, *Block preconditioners for the marker-and-cell discretization of the Stokes–Darcy equations*, SIAM J. Matrix Anal. Appl., 44 (2023), pp. 1540–1565, <https://doi.org/10.1137/22M1518384>.
- [23] Y. HE, J. LI, AND L. MENG, *Three effective preconditioners for double saddle point problem*, AIMS Mathematics, 6 (2021), pp. 6933–6947, <https://doi.org/10.3934/math.2021406>.
- [24] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, 2013.
- [25] N. HUANG, Y.-H. DAI, D. ORBAN, AND M. SAUNDERS, *On GSOR, the generalized successive overrelaxation method for double saddle-point problems*, SIAM J. Sci. Comput., 45 (2023), pp. A2185–A220, <https://doi.org/10.1137/22M1515884>.
- [26] W. LAYTON, F. SCHIEWECK, AND I. YOTOV, *Coupling fluid flow with porous media flow*, SIAM J. Numer. Anal., 40 (2003), pp. 2195–2218, <https://doi.org/10.1137/S0036142901392766>.
- [27] D. LOGHIN AND A. WATHEN, *Analysis of preconditioners for saddle-point problems*, SIAM J. Sci. Comput., 25 (2004), pp. 2029–2049, <https://doi.org/10.1137/S1064827502418203>.
- [28] I. PERUGIA AND V. SIMONCINI, *Block-diagonal and indefinite symmetric preconditioners for mixed finite element formulations*, Numer. Linear Algebra Appl., 7 (2000), pp. 585–616, [https://doi.org/10.1002/1099-1506\(200010/12\)7:7/8<585::AID-NLA214>3.0.CO;2-F](https://doi.org/10.1002/1099-1506(200010/12)7:7/8<585::AID-NLA214>3.0.CO;2-F).
- [29] C. RODRIGO, F. GASPAR, J. ADLER, X. HU, P. OHM, AND L. ZIKATANOV, *Parameter-robust preconditioners for Biot’s model*, SeMA J., (2023), <https://doi.org/10.1007/s40324-023-00336-2>.
- [30] M. ROZLOŽNÍK, *Saddle-Point Problems and Their Iterative Solution*, Birkhäuser, 2018.
- [31] I. RYBAK, J. MAGIERA, R. HELMIG, AND C. ROHDE, *Multirate time integration for coupled saturated/unsaturated porous medium and free flow systems*, Comput. Geosci., 19 (2015), pp. 299–309, <https://doi.org/10.1007/s10596-015-9469-8>.
- [32] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, 2003.
- [33] P. SAFFMAN, *On the boundary condition at the surface of a porous medium*, Stud. Appl. Math., 50 (1971), pp. 93–101, <https://doi.org/10.1002/sapm197150293>.
- [34] J. SCHMALFUSS, C. RIETHMÜLLER, M. ALTENBERND, K. WEISHAUP, AND D. GÖDDEKE, *Partitioned coupling vs. monolithic block-preconditioning approaches for solving Stokes–Darcy systems*, in Proc. International Conference on Computational Methods for Coupled Problems in Science and Engineering (COUPLED PROBLEMS), E. Oñate, M. Papadrakakis, and B. Schrefler, eds., 2021, <https://doi.org/10.23967/coupled.2021.043>.
- [35] D. SILVESTER, H. ELMAN, AND A. RAMAGE, *Incompressible flow and iterative solver software (IFISS) version 3.5*, (2016), http://www.cs.umd.edu/~elman/ifiss3.5/ifiss_guide.3.5.pdf.
- [36] P. STROHBECK, C. RIETHMÜLLER, D. GÖDDEKE, AND I. RYBAK, *Robust and efficient preconditioners for Stokes–Darcy problems*, in Finite Volumes for Complex Applications X—Volume 1, Elliptic and Parabolic Problems, E. Franck, J. Fuhrmann, V. Michel-Dansac, and L. Navoret, eds., Springer Nature Switzerland, 2023, pp. 375–383, https://doi.org/10.1007/978-3-031-40864-9_32.
- [37] H. VAN DER VORST, *Iterative Krylov Methods for Large Linear Systems*, Cambridge University Press, 2003.
- [38] L. ZHAO AND E. PARK, *A lowest-order staggered DG method for the coupled Stokes–Darcy problem*, IMA J. Numer. Anal., 40 (2020), pp. 2871–2897, <https://doi.org/10.1093/imanum/drz048>.
- [39] G. ZHOU, T. KASHIWABARA, I. OIKAWA, E. CHUNG, AND M.-C. SHIUE, *Some DG schemes for the Stokes–Darcy problem using P1/P1 element*, Jpn. J. Ind. Appl. Math., 36 (2019), pp. 1101–1128, <https://doi.org/10.1007/s13160-019-00377-z>.