

# ROBUST PRECONDITIONERS FOR PERTURBED SADDLE-POINT PROBLEMS AND CONSERVATIVE DISCRETIZATIONS OF BIOT'S EQUATIONS UTILIZING TOTAL PRESSURE\*

WIETSE M. BOON<sup>†</sup>, MIROSLAV KUCHTA<sup>‡</sup>, KENT-ANDRE MARDAL<sup>§</sup>, AND  
RICARDO RUIZ-BAIER<sup>¶</sup>

**Abstract.** We develop robust solvers for a class of perturbed saddle-point problems arising in the study of a second-order elliptic equation in mixed form (in terms of flux and potential), and of the four-field formulation of Biot's consolidation problem for linear poroelasticity (using displacement, filtration flux, total pressure, and fluid pressure). The stability of the continuous variational mixed problems, which hinges upon using adequately weighted spaces, is addressed in detail; and the efficacy of the proposed preconditioners, as well as their robustness with respect to relevant material properties, is demonstrated through several numerical experiments.

**Key words.** operator preconditioning, mixed finite element methods, perturbed saddle-point problems, equations of linear poroelasticity

**AMS subject classifications.** 65F08, 65N30, 65N12, 35B35

**DOI.** 10.1137/20M1379708

**1. Introduction.** Disparity of model parameters is a phenomenon commonly encountered in a variety of applications, and it is of paramount importance that the formulation of multiphysics problems and the design of discretizations and efficient solvers are robust with respect to at least some of the parameters with wide value ranges. We will consider here the equations of linear poroelasticity, where fluid flows in isothermal deformable porous media, assuming that the solid skeleton undergoes small strains. The poroelastic Biot equations that form the subject of this study are

$$(1.1a) \quad -\nabla \cdot (2\mu\varepsilon(\mathbf{u}) + (\lambda\nabla \cdot \mathbf{u} - \alpha p)\mathbf{I}) = \mathbf{b} \quad \text{in } \Omega \times (0, t_{\text{end}}],$$

$$(1.1b) \quad K^{-1}\mathbf{w} + \nabla p = \mathbf{0} \quad \text{in } \Omega \times (0, t_{\text{end}}],$$

$$(1.1c) \quad \partial_t(cp + \alpha\nabla \cdot \mathbf{u}) + \nabla \cdot \mathbf{w} = f \quad \text{in } \Omega \times (0, t_{\text{end}}],$$

equipped with suitable boundary (and initial) conditions to be specified later. Here,  $\mathbf{u}(t) : \Omega \rightarrow \mathbb{R}^d$  is the solid displacement vector,  $\mathbf{w}(t) : \Omega \rightarrow \mathbb{R}^d$  is the Darcy flux or

---

\*Submitted to the journal's Computational Methods in Science and Engineering section November 11, 2020; accepted for publication (in revised form) April 5, 2021; published electronically August 4, 2021.

<https://doi.org/10.1137/20M1379708>

**Funding.** The first author's work was supported by the Dahlquist Research Fellowship, funded by Comsol AB. The second author's work was supported by the Research Council of Norway (NFR), grant 280709. The third author's work was supported by the Research Council of Norway, grants 300305 and 301013. The fourth author's work was supported by the Monash Mathematics Research Fund S05802-3951284 and by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital biodesign and personalized healthcare" 075-15-2020-926.

<sup>†</sup>Department of Mathematics, KTH Royal Institute of Technology, Lindstedtsvägen 25, 114 28 Stockholm, Sweden (wietse@kth.se).

<sup>‡</sup>Simula Research Laboratory, 1325 Lysaker, Norway (miroslav@simula.no).

<sup>§</sup>Department of Mathematics, Division of Mechanics, University of Oslo, 0316 Oslo, Norway, and Simula Research Laboratory, 1325 Lysaker, Norway (kent-and@math.uio.no).

<sup>¶</sup>School of Mathematics, Monash University, 9 Rainforest Walk, Clayton 3800 VIC, Australia, and Institute of Computer Science and Mathematical Modeling, Sechenov University, 119435, Moscow, Russian Federation, and Universidad Adventista de Chile, Casilla 7-D, 3780000, Chillán, Chile (ricardo.ruizbaier@monash.edu).

percolation velocity,  $p(t) : \Omega \rightarrow \mathbb{R}$  is the fluid pressure, the symbol  $\partial_t$  denotes the partial derivative with respect to time,  $\mathbf{b}(t) : \Omega \rightarrow \mathbb{R}^d$  is a prescribed body force per unit volume of the porous medium, the symmetric part of the displacement (row-wise) gradient defines the infinitesimal strain tensor  $\boldsymbol{\varepsilon}(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^\top)$ , the parameters  $\lambda, \mu$  are the Lamé constants of the solid,  $K$  is the hydraulic conductivity (ratio between the material permeability and fluid viscosity),  $f(t) : \Omega \rightarrow \mathbb{R}$  is a source or sink of fluid mass, and  $c, \alpha$  are the total storage capacity and Biot–Willis poroelastic coefficient, respectively.

Several types of discretizations for (1.1) are available from the literature, including mixed and continuous elements, least-squares mixed, stabilized  $\mathbf{H}(\text{div})$ -conforming and other nonconforming schemes, adaptive mixed methods, weak Galerkin, enriched Lagrangian, and hybrid finite-volume finite element methods (see, e.g., [30, 4, 44, 24, 34, 41, 20, 42, 36, 16] and the references therein).

A main challenge for these equations is the construction of solvers that scale properly for nearly incompressible solids where the Lamé dilation modulus tends to infinity, as well as in the case of nearly incompressible fluids, for which the constrained specific storage coefficient approaches zero, or the nearly impermeable regime where the hydraulic conductivity is very small. These scenarios entail not only a complication at the practical and implementation level, but also a difficulty inherent to the functional setting of the abstract formulation (see, e.g., [31, 18, 43]). In more detail, for almost incompressible solids ( $\lambda \gg \mu$ ), the primal form of the elasticity equation, used in (1.1a), here scaled by  $\lambda$ ,

$$-\nabla \cdot \left( 2\frac{\mu}{\lambda} \boldsymbol{\varepsilon}(\mathbf{u}) + (\nabla \cdot \mathbf{u}) \mathbf{I} \right) = \mathbf{b} \quad \text{in } \Omega,$$

is known to suffer from locking when using standard elements such as Lagrange elements. The reason is that the problem is a singular perturbation problem, where stability in  $\mathbf{H}^1(\Omega)$  decays as  $\mu/\lambda$  tends to zero and where stability can only be obtained in  $\mathbf{H}(\text{div}, \Omega)$ . A remedy is to use elements that are stable in both  $\mathbf{H}(\text{div}, \Omega)$  and  $\mathbf{H}^1(\Omega)$  such as in [18, 19] where stabilized Brezzi–Douglas–Marini (BDM) elements are employed. Another alternative is to employ a technique similar to Herrmann’s method [17] where an additional solid pressure,  $p = \lambda \nabla \cdot \mathbf{u}$ , is introduced. It has been shown that a straightforward application of Herrmann’s method is unstable, but that the technique can be adjusted such that the discretization becomes stable [23, 29] for displacement-pressure formulations, the so-called total-pressure formulation. The method was extended to conservative formulations, i.e., displacement-flux-pressure, in [22], but robustness with respect to all parameters was not established.

A second singular perturbation problem occurs when the hydraulic conductivity ( $K$ ) tends to zero. To prevent nonphysical pressure oscillations, mixed schemes involving both flux and pressure are often used, i.e., ignoring for the moment the elastic deformation, the equations read

$$\begin{aligned} K^{-1} \mathbf{w} + \nabla p &= \mathbf{0} && \text{in } \Omega \times (0, t_{\text{end}}], \\ \partial_t c p + \nabla \cdot \mathbf{w} &= f && \text{in } \Omega \times (0, t_{\text{end}}]. \end{aligned}$$

Hence, upon time-discretization, this system is a mixed Darcy problem with a lower-order perturbation term for the pressure and we will consider the cases where one or both of  $K \rightarrow 0$  and  $c \rightarrow 0$  is allowed. It is seen that if the perturbation is sufficiently small (such that the additional term is bounded by the pressure norm), then the perturbed problem is well posed if a weighted  $L^2$ -norm, i.e.,  $K^{\frac{1}{2}} L^2(\Omega)$ , is used for the

pressure. This observation has been frequently employed [18, 32, 39, 40] in various porous media flow applications. However, if the hydraulic conductivity ratio  $K$  is small, then (the fluid part of) the perturbation cannot be bounded by the  $K^{\frac{1}{2}}L^2(\Omega)$ -norm, preventing a robust stability result. With this in mind, a convenient rescaling of the employed norms seems to produce better results, as recently suggested in [3].

We also mention that for conservative Biot formulations such as (1.1a)–(1.1c), the stability, i.e., the inf-sup condition, of the porous media problem can be weakened, as observed in [20, 25]. We will show, for the total pressure formulation of the conservative form of Biot's equations (1.1a)–(1.1c), that the scaling of the fluid pressure cannot be chosen independently of the coupling to the solid displacement and that the stability of the fluid pressure in  $L^2(\Omega) + K^{\frac{1}{2}}H^1(\Omega)$  is crucial.

A key tool for our stability analysis is the seminal paper [8] which analyzed saddle-point problems with penalty terms corresponding to singular perturbation problems. Therein, it is shown that, depending on the penalty term, the perturbation may either stabilize or destabilize the saddle-point problem. The Biot equations in study here involve two saddle-point problems with penalties corresponding to two singular perturbation problems that may be strongly coupled. The analysis leads us to utilize nonstandard Sobolev spaces to untangle the precise stability problems required in both the continuous and discrete settings.

This paper is structured as follows. The motivating problem of Biot consolidation and its variational formulation are presented in the remainder of this section. Then in section 2 we give an overview of the analysis of perturbed saddle-point problems following [8]. In section 3, this theoretical framework is used to show that a generalized Poisson equation in mixed form with Dirichlet boundary conditions is stable in appropriately weighted norms, and there we also discuss the case of Neumann boundary conditions. Section 4 contains an application of the theory to the four-field formulation of Biot equations. In section 5 we make precise the norms and spaces required at the discrete level, and in section 6 we collect numerical results that test the performance of the proposed block preconditioners for the modified Poisson equation and the Biot consolidation system.

**1.1. Problem formulation.** Let us consider the time domain  $t \in (0, t_{\text{end}}]$  and an open, bounded connected Lipschitz spatial domain  $\Omega \subset \mathbb{R}^d$ ,  $d = 2, 3$ , on which the Biot equations in quasi-static form, (1.1), are posed.

We introduce the total pressure (the sum of the volumetric contributions to the poroelastic Cauchy stress; cf. [23, 29]) as

$$(1.2) \quad p_T = \lambda \nabla \cdot \mathbf{u} - \alpha p.$$

We substitute (1.2) in the momentum balance equation and use it to rewrite the volumetric term in the mass balance equation:

$$\alpha \nabla \cdot \mathbf{u} = \frac{\alpha}{\lambda} p_T + \frac{\alpha^2}{\lambda} p.$$

This leads to the four-field formulation of Biot's equation (see, e.g., [18, 22]) written in operator form

$$\begin{bmatrix} -\nabla \cdot (2\mu \boldsymbol{\varepsilon}) & -\nabla & \nabla \\ K^{-1} & \frac{1}{\lambda} & \frac{\alpha}{\lambda} \\ -\nabla \cdot & \nabla \cdot & (c + \frac{\alpha^2}{\lambda}) \partial_t \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \\ p_T \\ p \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \\ 0 \\ f \end{bmatrix}.$$

Regarding boundary conditions, we assume that the boundary  $\partial\Omega = \Gamma^{\boldsymbol{u}} \cup \Gamma^{\boldsymbol{\sigma}}$ , with  $\Gamma^{\boldsymbol{u}} \cap \Gamma^{\boldsymbol{\sigma}} = \emptyset$  and  $|\Gamma^{\boldsymbol{u}}| \neq 0 \neq |\Gamma^{\boldsymbol{\sigma}}|$ , splits into two subregions:  $\Gamma^{\boldsymbol{u}}$  where displacement and normal filtration flux are prescribed (the solid is clamped and the fluid slips), and  $\Gamma^{\boldsymbol{\sigma}}$  where we set zero total traction and zero fluid pressure:

$$(1.3a) \quad \boldsymbol{u} = \mathbf{0} \quad \text{and} \quad \boldsymbol{\nu} \cdot \boldsymbol{w} = 0 \quad \text{on } \Gamma^{\boldsymbol{u}} \times (0, t_{\text{end}}],$$

$$(1.3b) \quad [2\mu\varepsilon(\boldsymbol{u}) - p_T \mathbf{I}] \boldsymbol{\nu} = \mathbf{0} \quad \text{and} \quad p = 0 \quad \text{on } \Gamma^{\boldsymbol{\sigma}} \times (0, t_{\text{end}}],$$

where  $\boldsymbol{\nu}$  is the unit normal vector on the boundary  $\partial\Omega$ . We also suppose that the system is initially at rest:

$$(1.3c) \quad \boldsymbol{u} = \mathbf{0}, \quad p = 0 \quad \text{in } \Omega \times \{0\}.$$

Let us consider a backward Euler time discretization of the four-field formulation with time step  $\tau$ . We let  $\boldsymbol{w}_\tau := \tau \boldsymbol{w}$ , and group the displacement and flux unknowns into a vector  $\vec{\boldsymbol{u}}$ , and the total pressure and fluid pressure into  $\vec{p}$  so that the vector of unknowns (at the current time step) is  $(\vec{\boldsymbol{u}}, \vec{p})^t = (\boldsymbol{u}, \boldsymbol{w}_\tau, p_T, p)^t$ . After a rescaling of the equations similar to [18], we have the operator:

$$(1.4) \quad \mathcal{A} \begin{bmatrix} \vec{\boldsymbol{u}} \\ \vec{p} \end{bmatrix} := \begin{bmatrix} -\nabla \cdot (2\mu\varepsilon) & -\nabla & \nabla \\ -\nabla \cdot & (\tau K)^{-1} & \frac{1}{\lambda} \\ \nabla \cdot & \frac{1}{\lambda} & \frac{\alpha}{\lambda} \\ & \frac{\alpha}{\lambda} & c + \frac{\alpha^2}{\lambda} \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{w}_\tau \\ p_T \\ p \end{bmatrix}.$$

Note that from the time-discrete formulation (1.4) and from the setup of boundary conditions (1.3a)–(1.3b), the natural trial and test spaces (before scaling) for displacement, filtration flux, total pressure, and fluid pressure, are, respectively,

$$\mathbf{H}_{\Gamma^{\boldsymbol{u}}}^1(\Omega), \quad \mathbf{H}_{\Gamma^{\boldsymbol{u}}}(\text{div}, \Omega), \quad L^2(\Omega), \quad L^2(\Omega).$$

Note also that system (1.4) adopts the structure

$$(1.5) \quad \begin{bmatrix} A & -B^t \\ B & C \end{bmatrix} \begin{bmatrix} \vec{\boldsymbol{u}} \\ \vec{p} \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix},$$

with  $A$  and  $C$  symmetric, positive semidefinite operators, and where the right-hand side vectors  $f, g$  contain contributions from the body load and volumetric source, as well as from quantities in the previous time step that arise from the discretization in time. More precisely, we have the weak formulation: Find  $(\vec{\boldsymbol{u}}, \vec{p}) \in (\mathbf{H}_{\Gamma^{\boldsymbol{u}}}^1(\Omega) \times \mathbf{H}_{\Gamma^{\boldsymbol{u}}}(\text{div}, \Omega)) \times (L^2(\Omega) \times L^2(\Omega))$  such that

$$\begin{aligned} a(\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}}) - b(\vec{\boldsymbol{v}}, \vec{p}) &= F(\vec{\boldsymbol{v}}) & \forall \vec{\boldsymbol{v}} \in \mathbf{H}_{\Gamma^{\boldsymbol{u}}}^1(\Omega) \times \mathbf{H}_{\Gamma^{\boldsymbol{u}}}(\text{div}, \Omega), \\ b(\vec{\boldsymbol{u}}, \vec{q}) + c(\vec{p}, \vec{q}) &= G(\vec{q}) & \forall \vec{q} \in L^2(\Omega) \times L^2(\Omega), \end{aligned}$$

where  $\vec{\boldsymbol{v}} = (\boldsymbol{v}, \boldsymbol{\zeta}, q_T, q)$ , and the bilinear forms and functionals adopt the form

$$\begin{aligned} a(\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}}) &:= 2\mu \int_{\Omega} \varepsilon(\boldsymbol{u}) : \varepsilon(\boldsymbol{v}) + \frac{1}{\tau K} \int_{\Omega} \boldsymbol{w}_\tau \cdot \boldsymbol{\zeta}, \\ b(\vec{\boldsymbol{v}}, \vec{q}) &:= - \int_{\Omega} \nabla \cdot \boldsymbol{v} q_T + \int_{\Omega} \nabla \cdot \boldsymbol{\zeta} q, \\ c(\vec{p}, \vec{q}) &:= \frac{1}{\lambda} \int_{\Omega} p_T q_T + \frac{\alpha}{\lambda} \int_{\Omega} p q_T + \frac{\alpha}{\lambda} \int_{\Omega} p_T q + \left[ c + \frac{\alpha^2}{\lambda} \right] \int_{\Omega} p q, \\ G(\vec{q}) &:= \int_{\Omega} \left( \tau f + \left[ c + \frac{\alpha^2}{\lambda} \right] p^n + \frac{\alpha}{\lambda} p_T^n \right) q, \quad F(\vec{\boldsymbol{v}}) := \int_{\Omega} \boldsymbol{b} \cdot \boldsymbol{v}, \end{aligned}$$

where  $p^n, p_T^n$  denote the approximations of fluid and total pressure on the previous iteration of backward Euler's method.

Let us point out that using  $(\vec{u}, \vec{p})^t$  as a test function, we are led to the following *poroelastic energy norm*:

(1.6)

$$\langle \mathcal{A}(\vec{u}, \vec{p}), (\vec{u}, \vec{p}) \rangle = 2\mu \|\boldsymbol{\varepsilon}(\mathbf{u})\|_{0,\Omega}^2 + (\tau K)^{-1} \|\mathbf{w}_\tau\|_{0,\Omega}^2 + \frac{1}{\lambda} \|p_T + \alpha p\|_{0,\Omega}^2 + c \|p\|_{0,\Omega}^2.$$

However, an issue with writing a global multilinear form and trying to analyze its stability is that this naturally induced seminorm does not take into account the term  $\nabla \cdot \mathbf{w}$ , and therefore, one loses separate control over  $p_T$  and  $p$  if  $c = 0$ . In particular, the operators  $A$  and  $C$  do not possess sufficient coercivity to ensure this (see, e.g., [6, 15]). We thus require a more involved strategy in order to obtain a stability bound in a stronger norm than (1.6). This will be presented in section 4, for which we first need to discuss theoretical aspects of perturbed saddle-point systems, which we exemplify with a simpler problem.

We also point out that the parabolic-elliptic nature of the coupled system may suggest, as an alternative to the monolithic approach leading to (1.4), to use operator splitting techniques that allow one to solve smaller and better conditioned systems in an iterative manner, as studied in, e.g., [1, 7, 27]; however, we do not address those lines here.

**2. Abstract analysis of perturbed saddle-point problems.** Typically, the stability analysis of perturbed saddle-point problems of type (1.5), posed on  $V \times Q$ , assumes a given norm on the space  $Q$  and uses this norm in its assumptions on the bilinear forms. However, we obtain two different types of control of the solution  $p \in Q$ , through the operators  $B$  and  $C$ , respectively. It is essential in the context of robust preconditioning to understand these two effects so that the dependencies on model parameters can be properly captured. This section, therefore, presents an analysis of (1.5) with the use of two different (semi)norms, reflecting the roles that  $B$  and  $C$  play. For this, we rely on the analysis presented in [8].

We start by introducing notation. Let  $V$  and  $Q_b$  be two Hilbert spaces endowed with norms  $\|\cdot\|_V$  and  $\|\cdot\|_{Q_b}$  that are possibly parameter-dependent. Let  $Q$  be a dense linear subspace of  $Q_b$ . We have three bilinear forms  $a : V \times V \rightarrow \mathbb{R}$ ,  $b : V \times Q_b \rightarrow \mathbb{R}$ , and  $c : Q \times Q \rightarrow \mathbb{R}$ , of which we assume that  $a$  and  $b$  are continuous and that  $a$  and  $c$  are symmetric and positive (semi)definite, i.e.,

$$\begin{aligned} a(u, v) &\lesssim \|u\|_V \|v\|_V, & b(u, q) &\lesssim \|u\|_V \|q\|_{Q_b} & \forall u, v \in V, \forall q \in Q, \\ a(u, v) &= a(v, u), & a(v, v) &\geq 0 & \forall u, v \in V, \\ c(p, q) &= c(q, p), & c(q, q) &\geq 0 & \forall p, q \in Q. \end{aligned}$$

Here, we use the notation  $x \lesssim y$  to denote that a constant  $c_0 > 0$  exists, independent of model parameters such that  $x \leq c_0 y$ . The relation  $\gtrsim$  has analogous meaning, and we denote  $x \approx y$  if  $x \lesssim y \lesssim x$ .

It is important to note that  $c$  may potentially be unbounded with respect to the norm  $\|\cdot\|_{Q_b}$ , as is the case in the examples we consider below. Let  $c$  generate the (semi)norm

$$|p|_c^2 := c(p, p) \quad \forall p \in Q,$$

and we assume that  $Q$  is complete with respect to the norm  $\|p\|_Q^2 := \|p\|_{Q_b}^2 + |p|_c^2$ . We emphasize that completeness of  $Q$  is assumed with respect to the stronger norm  $\|\cdot\|_Q$  whereas the density of  $Q$  in  $Q_b$  concerns the weaker norm  $\|\cdot\|_{Q_b}$ .

The linear operators associated to  $a(\cdot, \cdot)$ ,  $b(\cdot, \cdot)$ , and  $c(\cdot, \cdot)$  are denoted by  $A : V \rightarrow V'$ ,  $B : V \rightarrow Q'_b$ , and  $C : Q \rightarrow Q'$ , respectively. Let  $B^t : Q_b \rightarrow V'$  be the adjoint of  $B$ . Letting  $0 \leq t \leq 1$  be a scaling parameter, we consider the following problem: Find  $(u, p) \in V \times Q_b$  such that

$$(2.1) \quad \begin{bmatrix} A & -B^t \\ B & t^2 C \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}.$$

We assume that the following bounds, known as the Brezzi conditions, hold:

$$(2.2a) \quad \forall v \in \text{Ker } B, \quad a(v, v) \gtrsim \|v\|_V^2,$$

$$(2.2b) \quad \forall p \in Q_b, \quad \sup_{v \in V} \frac{b(v, p)}{\|v\|_V} \gtrsim \|p\|_{Q_b}.$$

Note the use of the norm  $\|\cdot\|_{Q_b}$  on  $p$  in (2.2b). This distinguishes our analysis from the convention in which the norm  $\|\cdot\|_Q$  is used in (2.2b) instead; see, e.g., [6, 37].

Finally, we introduce the parameter-dependent energy norm

$$(2.3) \quad \|(v, q)\|^2 := \|v\|_V^2 + \|q\|_{Q_b}^2 + t^2|q|_c^2.$$

Note that  $t$  in this norm reflects the additional stability obtained from the  $C$ -block. For  $t = 0$ , we obtain stability directly from the Brezzi conditions. However, for the range  $t \in [0, 1]$ , we require an additional inf-sup condition, as presented in the following theorem.

**THEOREM 2.1** (Brezzi–Braess). *Let the bilinear forms  $a$  and  $b$  satisfy the Brezzi conditions (2.2). If, moreover,*

$$(2.4) \quad \forall u \in V, \quad \sup_{(v, q) \in V \times Q} \frac{a(u, v) + b(u, q)}{\|(v, q)\|} \gtrsim \|u\|_V,$$

*then problem (2.1) is stable in the energy norm (2.3).*

*Proof.* See [8, Lemma 3] for the proof.  $\square$

In the following, we refer to (2.4) as the *Braess* condition. We will now demonstrate the use of this result for an exemplary problem concerning the modified Poisson equation, followed by the four-field formulation of Biot's equations (1.4).

### 3. A generalized Poisson (or simplified Helmholtz) equation in mixed form.

**3.1. Dirichlet boundary conditions.** Let us consider the following elliptic problem (here referred to as generalized Poisson equation, or also *modified/simplified* Helmholtz equation because the squared wavenumber is taken with the opposite sign):

$$(3.1) \quad -\nabla \cdot (K \nabla p) + \alpha p = f \quad \text{in } \Omega, \quad p = 0 \quad \text{on } \partial\Omega,$$

with coefficient matrix  $K$ , prescribed right-hand side  $f$ , and scalar parameter  $0 \leq \alpha \leq 1$  (the squared wavenumber), and its mixed formulation in operator form, given as follows:

Find  $\mathbf{u} \in V$  and  $p \in Q$  such that

$$(3.2) \quad \begin{bmatrix} K^{-1}I & \nabla \\ \nabla \cdot & \alpha I \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ f \end{bmatrix}.$$

Note that this problem has the structure of (1.5) (see similar mixed and mixed-hybrid formulations using Raviart–Thomas elements in, e.g., [11, 28]). We now define the appropriate function spaces and energy norm (2.3) using the properties of the operators  $A$ ,  $B$ , and  $C$ .

Starting with the Brezzi conditions (2.2), we follow the theory presented in [3] and consider the spaces

$$(3.3) \quad V := K^{-\frac{1}{2}} \mathbf{L}^2(\Omega) \cap \mathbf{H}(\text{div}, \Omega), \quad Q_b := K^{\frac{1}{2}} H^1(\Omega) + L^2(\Omega).$$

These intersection and summation spaces are defined by the parameter-dependent norms

$$(3.4a) \quad \|u\|_V^2 := \|K^{-\frac{1}{2}} u\|_{0,\Omega}^2 + \|\nabla \cdot u\|_{0,\Omega}^2,$$

$$(3.4b) \quad \|p\|_{Q_b}^2 := \inf_{r \in K^{\frac{1}{2}} H^1(\Omega)} \left( \|p - r\|_{0,\Omega}^2 + \|K^{\frac{1}{2}} \nabla r\|_{0,\Omega}^2 \right).$$

As shown in [3], both the inf-sup (2.2b) and coercivity (2.2a) conditions hold in these norms. For more information on summation spaces, we refer the reader to [5].

Letting  $\alpha$  play the role of  $t^2$  from section 2, we have  $C = I$  and thus

$$\|p\|_c^2 := c(p, p) = \|p\|_{0,\Omega}^2,$$

and we remark that  $\|p\|_c \not\lesssim \|p\|_{Q_b}$ . However, we have  $Q := L^2(\Omega) \cap Q_b = L^2(\Omega)$ , and it remains to show that  $Q$  is dense in  $Q_b$ . But this is immediate from the fact that  $Q^{\perp_{Q_b}} = \{0\}$ .

**LEMMA 3.1.** *Given the spaces  $V$  and  $Q$ , their associated norms, and the bilinear forms in (3.2), the assumptions of Theorem 2.1 then hold. In turn, the problem is stable in the energy norm (2.3).*

*Proof.* The validity of the Brezzi conditions (2.2) is shown in [3]. Hence, it remains to show the Braess condition (2.4) and we proceed as follows. Given  $u \in V$ , let  $v = u \in V$  and  $q = \nabla \cdot u \in Q$ . It follows that

$$a(u, v) + b(u, q) = \|K^{-\frac{1}{2}} u\|_{0,\Omega}^2 + \|\nabla \cdot u\|_{0,\Omega}^2 = \|u\|_V^2.$$

Moreover, we have

$$\|v, q\|^2 := \|v\|_V^2 + \|q\|_{Q_b}^2 + \alpha \|q\|_c^2 \lesssim \|u\|_V^2 + \|\nabla \cdot u\|_{0,\Omega}^2 \lesssim \|u\|_V^2. \quad \square$$

**Remark 3.2.** For scalar  $K$ , it seems natural to consider the norms  $\|v\|_V^2 := \|K^{-\frac{1}{2}} v\|_{0,\Omega}^2 + \|\nabla \cdot K^{-\frac{1}{2}} v\|_{0,\Omega}^2$  and  $\|q\|_{Q_b} := \|K^{\frac{1}{2}} q\|_{0,\Omega}$  instead, similar to [39]. Following the proof of Lemma 3.1, we would then choose  $v = u$  and  $q = -\nabla \cdot K^{-1} u$ , such that

$$(3.5) \quad a(u, v) - b(u, q) = \|K^{-\frac{1}{2}} u\|_{0,\Omega}^2 + \|\nabla \cdot K^{-\frac{1}{2}} u\|_{0,\Omega}^2 = \|u\|_V^2.$$

However, for the second bound, we obtain

$$(3.6) \quad \alpha \|q\|_c^2 = \alpha \|\nabla \cdot K^{-1} u\|_0^2 \leq \alpha K^{-1} \|u\|_V^2.$$

For the case  $\alpha > 0$ , the bound (3.6) cannot be improved with a constant independent of  $K$ . This explains the suboptimal performance of the preconditioner  $\mathcal{B}_{VV}$  for  $K < \alpha$  observed in Table 6.1 of section 6.

*Remark 3.3.* Strictly speaking, the bilinear form  $c(\cdot, \cdot)$  is not continuous in  $Q_b \times Q_b$  since the continuity constant would necessarily depend on  $K$ . For this reason, Theorem 2.1 is more appropriate than the analysis presented in, e.g., [6, 37], where continuity of  $c$  is assumed.

We remark that the energy norm is given by

$$(3.7) \quad \begin{aligned} \|(\mathbf{u}, p)\|^2 &:= \|\mathbf{u}\|_V^2 + \|p\|_{Q_b}^2 + \alpha \|p\|_c^2 \\ &\asymp \|\mathbf{u}\|_V^2 + \inf_{r \in K^{\frac{1}{2}} H^1(\Omega)} \left( (1 + \alpha) \|p - r\|_{0,\Omega}^2 + \alpha \|r\|_{0,\Omega}^2 + \|K^{\frac{1}{2}} \nabla r\|_{0,\Omega}^2 \right). \end{aligned}$$

Based on this norm, we use the theory from [26] to propose the following preconditioner for problem (3.2):

$$(3.8) \quad \mathcal{B} = \begin{bmatrix} (K^{-1} I - \nabla \nabla \cdot)^{-1} & 0 \\ 0 & ((1 + \alpha) I)^{-1} + (\alpha I - K \Delta)^{-1} \end{bmatrix}.$$

**3.2. Neumann boundary conditions.** Consider the following generalized Poisson problem with homogeneous Neumann boundary conditions for  $0 \leq \alpha \leq 1$ :

$$(3.9) \quad \nabla \cdot (-\nabla p) + \alpha p = f, \quad -\boldsymbol{\nu} \cdot \nabla p|_{\partial\Omega} = 0.$$

In the limit case of  $\alpha = 0$ , the solution  $p$  is only defined for compatible  $f$ . We thus restrict this section to the case where  $f$  has zero mean, i.e.,

$$\bar{f} = \Pi_{\mathbb{R}} f = 0,$$

with  $\Pi_{\mathbb{R}}$  the  $L^2$ -projection onto constants. Applying this projection to the original equation, we immediately obtain that  $\alpha \bar{p} = 0$ . In the limit case, we have the freedom to choose  $p$  with zero mean so this implies that  $\bar{p} = 0$  for all  $\alpha \geq 0$ . This property is usually treated by searching the solution in the restricted function space  $L^2(\Omega)/\mathbb{R}$ . However, this can be cumbersome to discretize so we present an alternative approach, based on the observations from section 2.

Let us consider the following equivalent problem:

Find  $\mathbf{u} \in \mathbf{H}_0(\text{div}, \Omega)$  and  $p \in L^2(\Omega)$  such that

$$(3.10) \quad \begin{bmatrix} I & \nabla \\ \nabla \cdot & \alpha I + (1 - \alpha) \Pi_{\mathbb{R}} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ f \end{bmatrix}.$$

**LEMMA 3.4.** *The solution  $(\mathbf{u}, p)$  to (3.10) exists uniquely and satisfies*

$$\|\mathbf{u}\|_{\text{div}, \Omega} + (1 + \sqrt{\alpha}) \|p\|_{0,\Omega} \lesssim \|f\|_{Q'}.$$

*Proof.* We first consider existence. Letting  $p$  be the solution to (3.9) and  $\mathbf{u} = -\nabla p$ , it follows that  $(\mathbf{u}, p)$  solves (3.10). Uniqueness, on the other hand, follows by establishing the bound on the solution.

We decompose the solution  $p$  into its mean  $\bar{p} \in \mathbb{R}$  and the deviation  $\hat{p} \in L^2(\Omega)/\mathbb{R}$ . Let us consider these components separately. First, by applying  $\Pi_{\mathbb{R}}$  to the second equation, we note that  $\bar{p}$  is given by

$$\bar{p} = \bar{f} = 0.$$

Second,  $(\mathbf{u}, \mathring{p}) \in \mathbf{H}_0(\text{div}, \Omega) \times L^2(\Omega)/\mathbb{R}$  solves

$$(3.11) \quad \begin{bmatrix} I & \nabla \\ \nabla \cdot & \alpha I \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathring{p} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathring{f} \end{bmatrix}.$$

This problem can be analyzed in the context of Theorem 2.1. We define the spaces  $V := \mathbf{H}_0(\text{div}, \Omega)$  and  $Q_b = Q := L^2(\Omega)/\mathbb{R}$  and introduce the norms

$$(3.12) \quad \|\mathbf{u}\|_V := \|\mathbf{u}\|_{\text{div}, \Omega}, \quad \|p\|_{Q_b} = |p|_c := \|p\|_{0, \Omega}.$$

The Brezzi conditions (2.2) are well known to be satisfied for these spaces and norms. Moreover, the Braess condition (2.4) can be verified by assuming given  $\mathbf{u}$ , setting  $(\mathbf{v}, \mathring{q}) = (\mathbf{u}, \nabla \cdot \mathbf{u})$ , and noting that

$$(3.13a) \quad a(\mathbf{u}, \mathbf{v}) + b(\mathbf{u}, q) = \|\mathbf{u}\|_V^2,$$

$$(3.13b) \quad \|\langle \mathbf{v}, q \rangle\|^2 = \|\mathbf{u}\|_V^2 + (1 + \alpha)\|\nabla \cdot \mathbf{u}\|_{0, \Omega}^2 \lesssim \|\mathbf{u}\|_V^2.$$

Hence, the assumptions of Theorem 2.1 are satisfied, and we obtain the result by noting that  $p = \mathring{p}$  and  $f = \mathring{f}$ .  $\square$

*Remark 3.5.* Unlike (3.9), problem (3.10) does not require compatible data, and is uniquely solvable for all  $f \in L^2(\Omega)$ . The ‘‘incompatibility’’ is captured by the mean of  $p$  since  $\bar{p} = \bar{f}$ . The stability result of Lemma 3.4 remains valid since we trivially have  $\|\bar{p}\|_Q = \|\bar{f}\|_{Q'}$ .

As a direct consequence of Lemma 3.4, we propose the following preconditioner for (3.10):

$$(3.14) \quad \mathcal{B} = \begin{bmatrix} (I - \nabla \nabla \cdot)^{-1} & 0 \\ 0 & ((1 + \alpha)I)^{-1} \end{bmatrix}.$$

Before returning to the problem of linear poroelasticity, we stress that a large class of problems can be put in the framework developed in this section. As an example, in Appendix C we discuss the application into the discretization of Herrmann’s formulation of linear elasticity [17], where the additional unknown of solid pressure is added to avoid volumetric locking.

**4. Back to the four-field formulation of Biot equations.** At this point, we want to apply the same strategy as in section 3 to construct a preconditioner for the Biot system  $\mathcal{A}$  from (1.4) (endowed with the boundary and initial conditions (1.3)). Let us consider the function spaces

$$\begin{aligned} V &:= 2\mu \mathbf{H}_{\Gamma u}^1(\Omega) \times [(\tau K)^{-\frac{1}{2}} \mathbf{L}^2(\Omega) \cap \mathbf{H}_{\Gamma u}(\text{div}, \Omega)], \\ Q_b &:= \mu^{-1} L^2(\Omega) \times [(\tau K)^{\frac{1}{2}} H^1(\Omega) + L^2(\Omega)], \end{aligned}$$

and introduce the following (semi)norms:

$$(4.1a) \quad \|\bar{\mathbf{u}}\|_V^2 = 2\mu \|\boldsymbol{\varepsilon}(\mathbf{u})\|_{0, \Omega}^2 + (\tau K)^{-1} \|\mathbf{w}_\tau\|_{0, \Omega}^2 + \|\nabla \cdot \mathbf{w}_\tau\|_{0, \Omega}^2,$$

$$(4.1b) \quad \|\bar{p}\|_{Q_b}^2 = \mu^{-1} \|p_T\|_{0, \Omega}^2 + \|p\|_{L^2(\Omega) + (\tau K)^{\frac{1}{2}} H^1(\Omega)}^2,$$

$$(4.1c) \quad |\bar{p}|_c^2 = \frac{1}{\lambda} \|p_T + \alpha p\|_{0, \Omega}^2 + c \|p\|_{0, \Omega}^2.$$

We define  $Q$  as the subspace of  $Q_b$  consisting of elements  $\vec{q}$  with  $|\vec{q}|_c < \infty$ . Density of  $Q$  in  $Q_b$  follows once more from the fact that  $Q^{\perp Q_b} = \{0\}$ .

Recall that  $0 \leq t \leq 1$  is the perturbation parameter in (2.1). The energy norm is given by (2.3), and we repeat it here for convenience:

$$(4.2) \quad \|(\vec{v}, \vec{q})\|^2 := \|\vec{v}\|_V^2 + \|\vec{q}\|_{Q_b}^2 + t^2 |\vec{q}|_c^2.$$

Although the relevant physical limits are not exactly given by a single parameter  $t$ , we do note that  $t = 1$  corresponds to the original problem (1.4). Moreover, the limit case with  $(\lambda, c) \rightarrow (\infty, 0)$  is equivalent to setting  $t = 0$ . Since Theorem 2.1 covers both cases, it forms the fundamental ingredient in our main result, presented in the following theorem.

**THEOREM 4.1.** *Problem (1.4) is well posed and the solution  $(\vec{u}, \vec{p}) \in V \times Q$  satisfies*

$$(4.3) \quad \|\vec{u}\|_V^2 + \|\vec{p}\|_{Q_b}^2 + |\vec{p}|_c^2 \lesssim \|F\|_{V'} + \|G\|_{Q'}.$$

*Proof.* We show that the assumptions of Theorem 2.1 are satisfied. Thus, let us consider the two Brezzi conditions (2.2) and the additional Braess condition (2.4):

- *Coercivity of  $A$  on  $\text{Ker } B$ .* For  $\vec{v} = (\vec{u}, \vec{w}_\tau) \in \text{Ker } B$ , we have  $\nabla \cdot \vec{w}_\tau = 0$ . It then directly follows that

$$a(\vec{v}, \vec{v}) = 2\mu \|\varepsilon(\vec{u})\|_{0,\Omega}^2 + (\tau K)^{-1} \|\vec{w}_\tau\|_{0,\Omega}^2 = \|\vec{v}\|_V^2 \quad \forall \vec{v} \in \text{Ker } B.$$

- *Inf-sup of  $B^\dagger$ .* Let  $\vec{p} = (p_T, p) \in Q_b$  be given. The usual inf-sup condition of Stokes problems, after a scaling by  $\mu$ , gives us that

$$\sup_{\vec{v}} \frac{(\nabla \cdot \vec{v}, p_T)_\Omega}{\|\mu^{\frac{1}{2}} \vec{v}\|_{1,\Omega}} \gtrsim \|\mu^{-\frac{1}{2}} p_T\|_{0,\Omega}.$$

Moreover, it was shown in [3] that

$$\sup_{\vec{w}_\tau} \frac{(\nabla \cdot \vec{w}_\tau, p)_\Omega}{\|\vec{w}_\tau\|_{(\tau K)^{-\frac{1}{2}} \mathbf{L}^2(\Omega) \cap \mathbf{H}(\text{div}, \Omega)}} \gtrsim \|p\|_{L^2(\Omega) + (\tau K)^{\frac{1}{2}} H^1(\Omega)}.$$

Combining the above, we obtain

$$\sup_{\vec{v}} \frac{b(\vec{v}, \vec{p})}{\|\vec{v}\|_V} = \sup_{\vec{v}} \frac{-(\nabla \cdot \vec{v}, p_T)_\Omega + (\nabla \cdot \vec{w}_\tau, p)_\Omega}{\|\vec{v}\|_V} \gtrsim \|\vec{p}\|_{Q_b}.$$

- *Inf-sup of  $[A]_{B^\dagger}$ .* Let  $\vec{u} = (\vec{u}, \vec{w}_\tau) \in V$  be given. We then choose  $\vec{v} = \vec{u}$  and  $\vec{q} = (0, \nabla \cdot \vec{w}_\tau)$  to derive

$$\begin{aligned} a(\vec{u}, \vec{v}) + b(\vec{u}, \vec{q}) &= 2\mu \|\varepsilon(\vec{u})\|_{0,\Omega}^2 + (\tau K)^{-1} \|\vec{w}_\tau\|_{0,\Omega}^2 + \|\nabla \cdot \vec{w}_\tau\|_{0,\Omega}^2 = \|\vec{u}\|_V^2, \\ \|(\vec{v}, \vec{q})\|^2 &= \|\vec{u}\|_V^2 + \|\nabla \cdot \vec{w}_\tau\|_{L^2(\Omega) + (\tau K)^{\frac{1}{2}} H^1(\Omega)}^2 + \left( \frac{\alpha^2}{\lambda} + c \right) \|\nabla \cdot \vec{w}_\tau\|_{0,\Omega}^2 \\ &\lesssim \|\vec{u}\|_V^2 + \|\nabla \cdot \vec{w}_\tau\|_{0,\Omega}^2 \\ &\lesssim \|\vec{u}\|_V^2. \end{aligned}$$

The proof is finalized by invoking Theorem 2.1 and noting that  $t = 1$  forms a special case.  $\square$

Given that Theorem 4.1 establishes a parameter-robust stability, we can straightforwardly use the general approach from [26] to construct the following preconditioner involving the specific norms (4.1):

$$(4.4) \quad \mathcal{B} = \begin{bmatrix} (-\nabla \cdot (2\mu\varepsilon))^{-1} & 0 & 0 & 0 \\ 0 & ((\tau K)^{-1}I - \nabla \nabla \cdot)^{-1} & 0 & 0 \\ 0 & 0 & \mathcal{P} & 0 \\ 0 & 0 & 0 & \mathcal{P} \end{bmatrix},$$

where the fluid and total pressure preconditioner,  $\mathcal{P}$ , is a  $2 \times 2$  operator defined as

$$\mathcal{P} = \left( \begin{bmatrix} \frac{1}{\mu} I & 0 \\ 0 & I \end{bmatrix} + \mathcal{C} \right)^{-1} + \left( \begin{bmatrix} \frac{1}{\mu} I & 0 \\ 0 & -\tau K \Delta \end{bmatrix} + \mathcal{C} \right)^{-1} \quad \text{and} \quad \mathcal{C} = \begin{bmatrix} \frac{1}{\lambda} & \frac{\alpha}{\lambda} \\ \frac{\alpha}{\lambda} & c + \frac{\alpha^2}{\lambda} \end{bmatrix}.$$

We expect such a preconditioner to be robust in the sense that the condition number of the (left-)preconditioned matrix  $\mathcal{B}\mathcal{A}$  is bounded uniformly in the parameters  $\{\mu, K, \tau, \lambda, \alpha, c\}$ .

## 5. Discrete stability.

**5.1. Abstract setting.** In order to define a finite element method, let  $\mathcal{T}_h$  be a conforming simplicial partition of  $\bar{\Omega}$ , constituted by tetrahedra (or triangles in two dimensions)  $K$  of diameter  $h_K$ , with mesh size  $h := \max\{h_K : K \in \mathcal{T}_h\}$ . The mesh is considered shape-regular. Given an integer  $s \geq 0$  and a generic element  $K \in \mathcal{T}_h$ , the symbol  $\mathbb{P}_s(K)$  will denote the space of polynomial functions defined locally on the element  $K$  and being of degree no greater than  $s$ .

For generic and conforming finite-dimensional subspaces  $V_h \subset V$ ,  $Q_h \subset Q$ , let us consider the Galerkin scheme arising from the discretization of (2.1),

$$(5.1a) \quad a(u_h, v_h) - b(v_h, p_h) = F(v_h) \quad \forall v_h \in V_h,$$

$$(5.1b) \quad b(u_h, q_h) + t^2 c(p_h, q_h) = G(q_h) \quad \forall q_h \in Q_h,$$

for which the following direct consequence of Theorem 2.1 holds.

**LEMMA 5.1.** *Let  $V_h, Q_h$  be endowed with norms  $\|\cdot\|_{V_h}, \|\cdot\|_{Q_{b,h}}$ , respectively, and let  $\|\cdot, \cdot\|_h$  be defined analogously to (2.3). Assume the following discrete counterparts of the Brezzi conditions are fulfilled:*

$$(5.2a) \quad \forall v_h \in \text{Ker } B_h, \quad a(v_h, v_h) \gtrsim \|v_h\|_{V_h}^2,$$

$$(5.2b) \quad \forall p_h \in Q_{b,h}, \quad \sup_{v_h \in V_h} \frac{b(v_h, p_h)}{\|v_h\|_{V_h}} \gtrsim \|p_h\|_{Q_{b,h}},$$

together with the discrete analogue of the Braess condition (2.4),

$$(5.3) \quad \forall u_h \in V_h, \quad \sup_{(v_h, q_h) \in V_h \times Q_h} \frac{a(u_h, v_h) + b(u_h, q_h)}{\|(v_h, q_h)\|_h} \gtrsim \|u_h\|_{V_h},$$

with all constants in (5.2a), (5.2b), and (5.3) independent of the mesh size  $h$  and of the perturbation parameter  $t$ . Then there exists a unique solution  $(u_h, p_h) \in V_h \times Q_h$  to (5.1), which is stable in the energy norm  $\|\cdot, \cdot\|_h$ .

**5.2. Discrete generalized Poisson problem.** According to Lemma 5.1, the stability of discretizations to (3.2) holds as long as the discrete Brezzi conditions and the discrete Braess condition are met by the chosen finite-dimensional spaces  $V_h, Q_h$ .

Following [3], let us first denote by  $\widehat{V}_h$  the space of  $\mathbf{H}(\text{div}, \Omega)$ -conforming vector functions approximated using BDM elements of order  $s + 1$  [10], or Raviart–Thomas elements of order  $s$  [33], and let  $\widehat{Q}_h$  be the space of discontinuous piecewise polynomials of order  $s$ . Then we can use the discrete gradient operator  $\nabla_h : \widehat{Q}_h \rightarrow \widehat{V}_h$  defined by

$$(5.4) \quad (\nabla_h q_h, \mathbf{v}_h) = -(q_h, \nabla \cdot \mathbf{v}_h),$$

to define a discrete  $H^1$ -norm as  $\|\nabla_h q_h\|_{0,\Omega}$ . This suffices to construct the approximation space for the potential as

$$(5.5) \quad Q_{b,h} = L_h^2 + K^{\frac{1}{2}} H_h^1,$$

where  $L_h^2$  corresponds to  $\widehat{Q}_h$  equipped with the usual  $L^2$ -norm and  $H_h^1$  denotes the space conformed by the set  $\widehat{Q}_h$  in combination with the discrete  $H^1$ -norm. Therefore, the norm associated with (5.5) is

$$\|q_h\|_{Q_{b,h}}^2 = \inf_{r_h \in \widehat{Q}_h} \left( \|q_h - r_h\|_{0,\Omega}^2 + \|K^{\frac{1}{2}} \nabla_h r_h\|_{0,\Omega}^2 \right).$$

On the other hand, for the flux we consider

$$(5.6) \quad V_h = K^{-\frac{1}{2}} \mathbf{L}_h^2 \cap \mathbf{H}_h(\text{div}),$$

where  $\mathbf{L}_h^2$  corresponds to  $\widehat{V}_h$  equipped with the usual  $\mathbf{L}^2$ -norm and  $\mathbf{H}_h(\text{div})$  denotes the space conformed by the set  $\widehat{V}_h$  in combination with the usual  $\mathbf{H}(\text{div})$ -norm. Since  $V_h \subseteq V$ , we endow  $V_h$  with the  $V$ -norm (3.4a).

**LEMMA 5.2.** *There exists a unique solution  $(u_h, p_h) \in V_h \times Q_h$  to the discretization of problem (3.2), which is stable in the energy norm  $\|\cdot\|_h$ .*

*Proof.* We verify the conditions of Lemma 5.1. First, the Brezzi conditions are shown in [3]. Second, for the Braess condition, we follow the proof of Lemma 3.1 in the discrete setting and note that we need  $\nabla \cdot \mathbf{u}_h \in Q_h$ , which holds for the chosen pairs of mixed finite element spaces.  $\square$

**5.3. Discrete mixed Biot consolidation system.** For the case of Biot equations, considering again Lemma 5.1 but in the context of Theorem 4.1, we can identify conditions for the discrete solvability and robust stability. It turns out that the approximation spaces for displacement and total pressure need to be inf-sup stable in the sense of the Brezzi conditions, and also the pair of spaces for filtration velocity and fluid pressure need to satisfy the discrete Brezzi conditions plus the additional requirement that  $(0, \nabla \cdot \mathbf{w}_{\tau,h}) \in Q_{b,h}$ ; however, it is not required that the divergence of discrete displacements belongs to the space of discrete total pressures.

As feasible choices for the approximation spaces for two-dimensional problems, we therefore take overall continuous and vector-valued, piecewise polynomials of degree  $s + 2$  to approximate displacements (denoted  $\widehat{V}_h$ ), and discontinuous and piecewise polynomials of degree  $s$  for the total pressure (denoted  $\widehat{Q}_h$ , as before). Alternatively, we may choose the Taylor–Hood pair for both two- and three-dimensional problems. Likewise, discrete inf-sup stability is required for the fluid flux-pressure pair, for which

we consider  $\mathbf{H}(\text{div}, \Omega)$ -conforming discretizations of the percolation fluxes using BDM elements of order  $s+1$  or Raviart–Thomas elements of order  $s$  (denoted  $\widehat{V}_h$  as before), and piecewise polynomials (overall discontinuous) of degree  $s$  for the fluid pressure.

Similar to the modified Helmholtz system, in the spaces

$$(5.7) \quad V_h = 2\mu \mathbf{H}_{\Gamma^u,h}^1 \times \left[ (\tau K)^{-\frac{1}{2}} \mathbf{L}_h^2 \cap \mathbf{H}_{\Gamma^u,h}(\text{div}) \right], \quad Q_{b,h} = \mu^{-1} L_h^2 \times \left[ (\tau K)^{\frac{1}{2}} H_h^1 + L_h^2 \right],$$

the discrete space  $\mathbf{H}_{\Gamma^u,h}^1$  corresponds to  $\widetilde{V}_h$  (restricted to discrete functions vanishing on  $\Gamma^u$ ) equipped with the usual  $\mathbf{H}^1$ -norm and  $\mathbf{H}_{\Gamma^u,h}(\text{div})$  denotes the space conformed by the set  $\widehat{V}_h$  (restricted to discrete functions with normal traces vanishing on  $\Gamma^u$ ) in combination with the usual  $\mathbf{H}(\text{div})$ -seminorm. On  $V_h$  we can use the  $V$ -norm defined in (4.1), while for  $Q_h$  we employ the norm

$$\begin{aligned} \|\vec{q}_h\|_{Q_h}^2 := & \frac{1}{\mu} \|q_{T,h}\|_{0,\Omega}^2 + \inf_{r_h \in \widehat{Q}_h} \left( \|q_h - r_h\|_{0,\Omega}^2 + \|K^{\frac{1}{2}} \nabla_h r_h\|_{0,\Omega}^2 \right) \\ & + \frac{1}{\lambda} \|q_{T,h} + \alpha q_h\|_{0,\Omega}^2 + c \|q_h\|_{0,\Omega}^2. \end{aligned}$$

**LEMMA 5.3.** *There exists a unique solution  $(\vec{u}_h, \vec{p}_h) \in V_h \times Q_h$  to the discretization of problem (1.4), which is stable in the energy norm  $\|\cdot\|_h$ .*

*Proof.* Let us consider the conditions of Lemma 5.1. Following the same steps as in Theorem 4.1, the Brezzi conditions follow directly from the choice of mixed finite elements. Moreover, we note that  $(0, \nabla \cdot \mathbf{w}_{\tau,h}) \in Q_{b,h}$  by the assumptions in the beginning of this section. Taking  $\vec{v}_h = \vec{u}_h$  and  $\vec{q}_h = (0, \nabla \cdot \mathbf{w}_{\tau,h})$  then gives us the Braess condition.  $\square$

**6. Numerical results.** We demonstrate robustness of the proposed preconditioners by considering spectra of the preconditioned systems. More precisely, for given problem operator  $\mathcal{A}$  of the form (2.1), let  $\tilde{\mathcal{A}}$  be the symmetric operator obtained by negating the second row of  $\mathcal{A}$ . We are interested in stability of the condition numbers  $|\lambda_{h,\max}|/|\lambda_{h,\min}|$ , where  $\lambda_{h,\max}$ ,  $\lambda_{h,\min}$  are the largest and smallest (in magnitude) eigenvalues of the generalized eigenvalue problem  $\tilde{\mathcal{A}}_h x = \lambda_h \mathcal{B}_h x$  with  $\tilde{\mathcal{A}}_h$ ,  $\mathcal{B}_h$  being the respective finite element approximations of  $\tilde{\mathcal{A}}$  and the preconditioner  $\mathcal{B}$ .

We remark that rather than the discrete  $H^1$ -norm defined in terms of the discrete gradient operator (5.4), we use an equivalent (see [35]), more implementation-friendly, norm defined in terms of the bilinear form  $\Delta_h : \widehat{Q}_h \times \widehat{Q}_h \rightarrow \mathbb{R}$  given by

$$\Delta_h(p_h, q_h) = \sum_{K \in \mathcal{T}_h} \int_K \nabla p_h \cdot \nabla q_h dx + \sum_{E \in \mathcal{E}_I} \int_E \frac{1}{\{h_E\}} [p_h] [q_h] ds + \sum_{E \in \mathcal{E}_D} \int_E \frac{1}{h_E} p_h q_h ds.$$

Here,  $\mathcal{E}_I$  are the interior facets of  $\mathcal{T}_h$  while  $\mathcal{E}_D$  are the external facets associated with pressure (Dirichlet) boundary conditions. The jump and average values of  $p_h \in \widehat{Q}_h$  are defined as  $\{p_h\} = \frac{1}{2}(p_h|_{K^+} + p_h|_{K^-})$  and  $[p_h] = p_h|_{K^+} - p_h|_{K^-}$ , respectively, with  $K^\pm$  the two elements that share the internal facet.

**6.1. Robust preconditioners for the generalized Poisson equation.** In the following we let  $\Omega = (0, 1)^2$  and  $\mathcal{T}_h$  is a uniform structured triangulation of the domain.

**Dirichlet boundary conditions.** Using the (stable) discretization given by  $\mathbb{RT}_0 \times \mathbb{P}_0$  elements, the robustness of (3.8) for the Dirichlet problem (3.1) can be seen

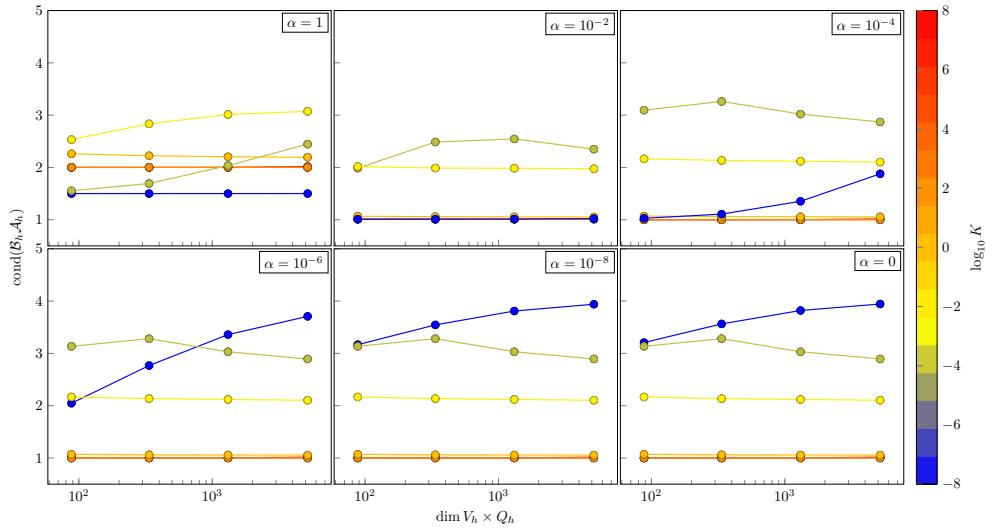


FIG. 6.1. Performance of preconditioner (3.8) for the generalized Poisson problem with pressure boundary conditions. Discretization by  $\mathbb{RT}_0\text{-}\mathbb{P}_0$  elements.

TABLE 6.1

Condition numbers of (6.1) for the generalized Poisson (or modified Helmholtz) problem and  $\alpha = 1$ . Pressure boundary conditions are prescribed and  $\mathbb{RT}_0\text{-}\mathbb{P}_0$  is used for discretization. Condition numbers exceeding  $10^4$  are indicated as  $-$ . In contrast, our proposed preconditioner (3.8) does perform robustly for  $K < \alpha$ , as shown in Figure 6.1.

$K$	$h$			
	$2^{-2}$	$2^{-3}$	$2^{-4}$	$2^{-5}$
$10^{-8}$	577	2306	9216	-
$10^{-6}$	577	2300	9133	-
$10^{-4}$	545	1874	4797	7867
$10^{-2}$	86	97	100	101
$10^0$	2.00	2.00	2.00	2.00
$10^2$	1.05	1.05	1.05	1.05
$10^4$	1.05	1.05	1.05	1.05
$10^6$	1.05	1.05	1.05	1.05
$10^8$	1.05	1.05	1.05	1.05

in Figure 6.1. To strengthen the numerical evidence, the experiments were carried out also with the lowest-order BDM and  $\mathbb{RT}_1\text{-}\mathbb{P}_1$  elements. The results are given in Figures A.1 and A.2 in Appendix A.

Before proceeding further, we address two aspects of the analysis in section 3. To compare (3.8) with alternatives, we recall a well-known  $K$ -robust preconditioner for the Darcy problem (i.e.,  $\alpha = 0$  in (3.2)) proposed in [39]. Extending it directly to the case of the generalized Poisson problem leads to

$$(6.1) \quad \mathcal{B}_{VV} = \begin{bmatrix} (K^{-1}I - K^{-1}\nabla\nabla\cdot)^{-1} & 0 \\ 0 & (KI + \alpha I)^{-1} \end{bmatrix}.$$

However, as noted in Remark 3.2, Table 6.1 shows that  $K < \alpha$  leads to a lack of robustness in  $K$ .

Finally, Table 6.2 illustrates the necessity of the assumption of a small perturbation, i.e.,  $\alpha \leq 1$ , for stability of the preconditioner (3.8). Indeed, by setting  $\alpha = 10^2$ ,

TABLE 6.2

Condition numbers of (3.8) for the generalized Poisson (modified Helmholtz) problem (3.2), setting  $\alpha = 10^2$ . Pressure boundary conditions are prescribed and  $\mathbb{RT}_0\text{-}\mathbb{P}_0$  is used for discretization. These results demonstrate the necessity of assuming small perturbations, i.e.,  $\alpha \leq 1$ .

$K$	$h$			
	$2^{-2}$	$2^{-3}$	$2^{-4}$	$2^{-5}$
$10^{-8}$	1.99	1.99	1.99	1.99
$10^{-6}$	1.99	1.99	2.01	2.06
$10^{-4}$	2.09	2.43	3.78	9.02
$10^{-2}$	11	38	96	158
$10^0$	166	190	197	198
$10^2$	151	151	151	152
$10^4$	151	152	152	152
$10^6$	151	152	152	152
$10^8$	151	152	152	152

TABLE 6.3

Condition numbers obtained for the Neumann problem (3.10) using the preconditioner (3.14).

$\alpha$	$h$			
	$2^{-2}$	$2^{-3}$	$2^{-4}$	$2^{-5}$
0	1.10	1.10	1.10	1.10
$10^{-8}$	1.10	1.10	1.10	1.10
$10^{-6}$	1.10	1.10	1.10	1.10
$10^{-4}$	1.10	1.10	1.10	1.10
$10^{-2}$	1.10	1.10	1.10	1.10
1	2.00	2.00	2.00	2.00

the sensitivity of the condition numbers for  $K < 1$  becomes evident.

**Neumann boundary conditions.** Parameter stability of preconditioner (3.14) for the Neumann problem (3.10) is illustrated in Table 6.3.

Regarding the Neumann problem (3.10) as a special case of

$$(6.2) \quad \begin{bmatrix} K^{-1}I & \nabla \\ \nabla \cdot & \alpha I + (1 - \alpha)\Pi_{\mathbb{R}} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ f \end{bmatrix},$$

let us finally address the preconditioning of problem (6.2). Combining the analysis developed in sections 3.1 and 3.2, we propose

$$(6.3) \quad \mathcal{B} = \begin{bmatrix} (K^{-1}I - \nabla\nabla \cdot)^{-1} & 0 \\ 0 & (\alpha I + (1 - \alpha)\Pi_{\mathbb{R}} + I)^{-1} + (\alpha I + (1 - \alpha)\Pi_{\mathbb{R}} - K\Delta)^{-1} \end{bmatrix}$$

as a preconditioner for (6.2). The robustness of such a preconditioner is demonstrated in Figure 6.2, and exemplified further in Figure A.3.

**6.2. Robust preconditioners for the mixed Biot system.** Due to its larger parameter space we restrict numerical experiments for the Biot system to a single type of finite element discretization, namely, we shall use continuous piecewise quadratic vector valued ( $\mathbb{P}_2$ ) functions for the displacement and lowest-order Raviart–Thomas elements for the percolation velocity. The total pressure will be discretized using continuous piecewise linear Lagrange elements ( $\mathbb{P}_1$ ) if  $\Omega \subset \mathbb{R}^3$  while  $\mathbb{P}_0$  is used in the two-dimensional case. Finally, the fluid pressure shall be approximated by piecewise constants.

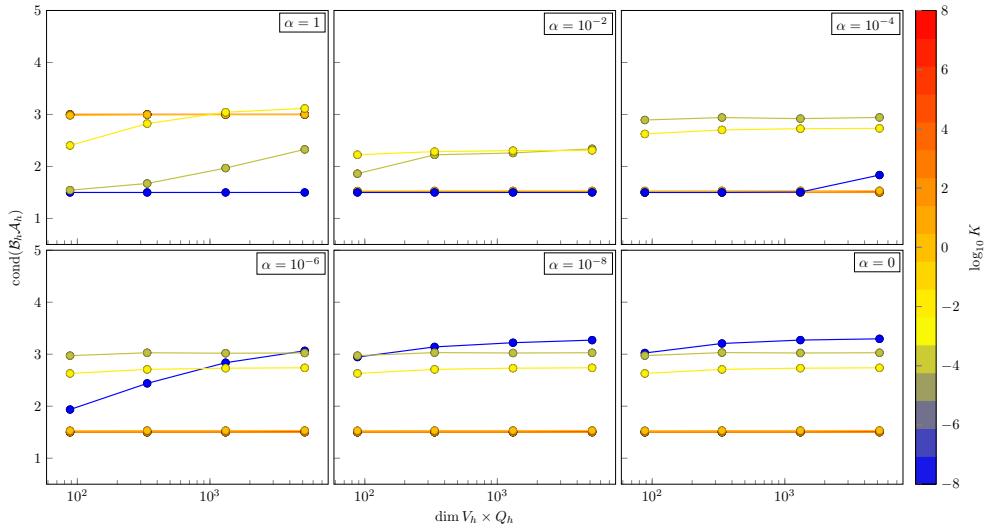


FIG. 6.2. Performance of preconditioner (6.3) for the generalized Poisson (simplified Helmholtz) equation (6.2) with flux boundary conditions. Discretization by  $\text{RT}_0\text{-}\mathbb{P}_0$  elements.

**Parameter robustness.** As in the case of the generalized Poisson problem we consider a two-dimensional problem (1.4) with domain  $\Omega = (0, 1)^2$ , and  $\Gamma^u = \{(x, y) \in \partial\Omega : x = 0 \text{ or } x = 1\}$ ,  $\Gamma^\sigma = \partial\Omega \setminus \Gamma^u$ . The domain is discretized by a uniform mesh.

Figure 6.3 shows variations of the condition numbers for the (4.4)-preconditioned Biot problem across the parameter ranges  $10^{-12} \leq K \leq 1$ ,  $1 \leq \lambda \leq 10^{16}$ ,  $0 \leq \alpha \leq 1$ ,  $c < 0 \leq 1$ . It can be seen that the preconditioner yields bounded condition numbers (not exceeding 8 in the experiments). We remark that the time step  $\tau$  was kept fixed at  $\tau = 1$  as its variations effectively translate to a modified hydraulic conductivity  $K\tau$ .

**Scalable realization of the preconditioner.** Numerical experiments presented thus far have utilized the exact Biot preconditioner, that is, each of the blocks was computed by LU factorization. As such a construction is of limited interest in practical/large scale applications, we next briefly discuss the realization of (4.4) in terms of off-the-shelf scalable components.

Indeed, the displacement block of the preconditioner is a standard operator which can be efficiently realized by, e.g., multigrid [9, 38]. Similarly, geometric (see [2]) and algebraic (see [21]) multigrid methods have been proposed for the Riesz map with respect to the weighted  $\mathbf{H}(\text{div})$ -inner product, which in our context corresponds to the flux preconditioner in (4.4). These methods have been shown to be robust in the respective parameters (cf. experiments in [21] for algebraic and [13, section 4.1] for the geometric multigrid case). Finally, to the best of the authors' knowledge, the approximation of the pressure block  $\mathcal{P}$  of the Biot preconditioner has not been studied in the literature. In this case, as the operator consists of two inverses of symmetric elliptic operators, we expect multigrid methods to perform well. We remark, however, that the approximation might not be robust with respect to the model parameters.

In order to illustrate the performance of the multigrid realization of the Biot preconditioner we consider the three-dimensional footing problem (see, e.g., [14, section

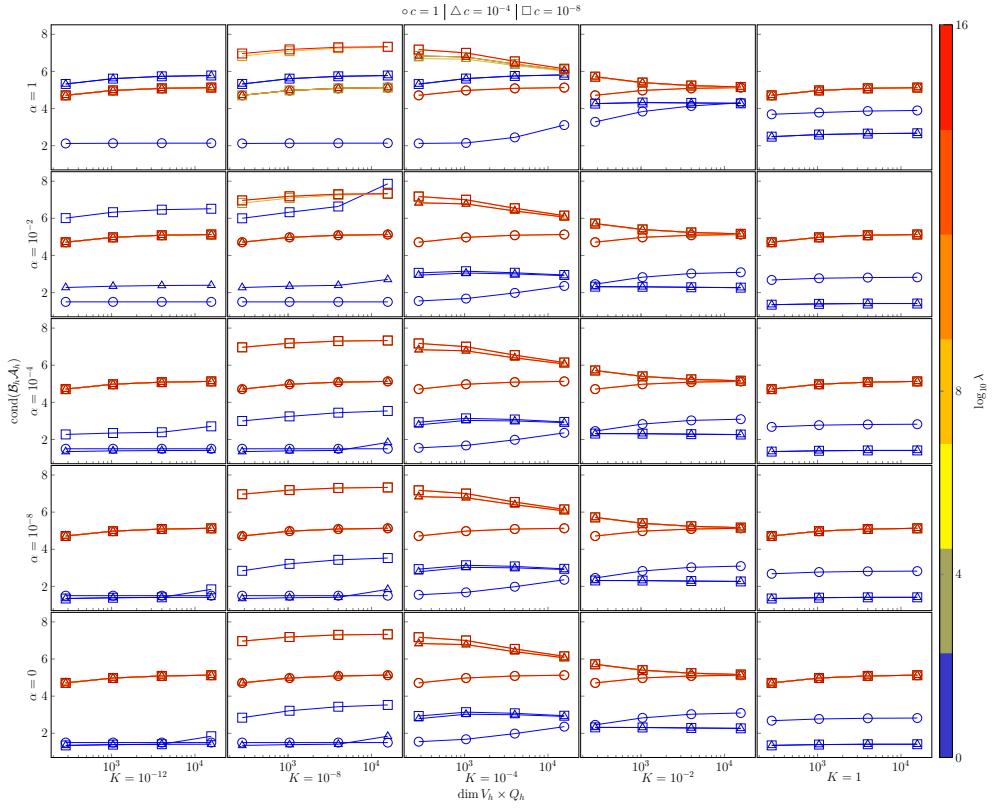


FIG. 6.3. Performance of the preconditioner (4.4) for the Biot system (1.4). Discretization by  $\mathbb{P}_2\text{-}\mathbb{RT}_0\text{-}\mathbb{P}_0\text{-}\mathbb{P}_0$  elements. Values of  $\alpha$  and  $K$  vary along the vertical, respectively horizontal, axis, while color encodes Lamé constant  $\lambda$ , and storage capacity is depicted with markers. In most cases, the condition numbers for  $\lambda > 1$  are very similar, leading to an overlap of the curves.

TABLE 6.4

Number of MinRes iterations obtained by using a multigrid realization of the Biot preconditioner (4.4) for the three-dimensional footing problem. For the coarsest mesh the number in brackets indicates the iteration count using an exact preconditioner with blocks computed by a direct solver. Discretization by  $\mathbb{P}_2\text{-}\mathbb{RT}_0\text{-}\mathbb{P}_1\text{-}\mathbb{P}_0$  elements.

$h^{-1}$	$\dim V_h \times Q_h$	$t$				
		0.1	0.2	0.3	0.4	0.5
16	$1.88 \times 10^5$	85(27)	92(31)	36(16)	36(16)	36(16)
32	$1.46 \times 10^6$	92	99	38	39	39
48	$4.86 \times 10^6$	101	110	41	41	41

5.2.2]) and set  $\alpha = c = 1/2$ . Here the displacement preconditioner and the components of  $\mathcal{P}$  employ a single V-cycle of algebraic multigrid (implemented in hypre [12]), while the flux preconditioner is realized with geometric multigrid using a hierarchy of three meshes in combination with the star smoother [2]. The implementation has been carried out using the PCPATCH framework [13].

Fixing the time step to  $\tau = 0.1$ , Table 6.4 displays the number of MinRes iterations needed to reduce the preconditioned residual norm by a factor of  $10^6$  at each step of the simulation. Taking the coarsest mesh for comparison, the use of multigrid approximately doubles the number of solver iterations as compared to the exact pre-

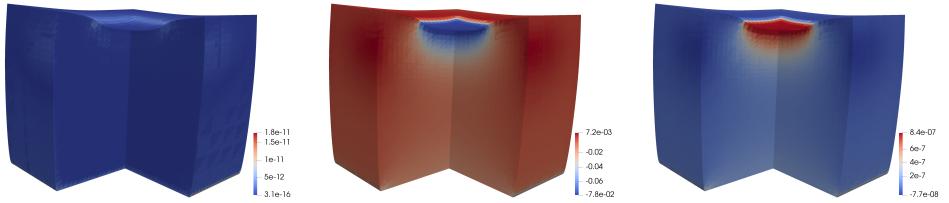


FIG. 6.4. Approximate solutions for the footing problem at  $t = 0.5$  s. The percolation velocity magnitude (left), total pressure (center), and fluid pressure (right) are shown on a domain deformed by the computed displacement (scaled by a factor of  $10^6$ ).

conditioner. However, the iterations appear to be bounded in the mesh size. Samples of the approximate solution at the final time can be seen in Figure 6.4.

**Appendix A. Stability of the preconditioner for the generalized Poisson problem.** This section presents results of numerical experiments showing robustness of the proposed preconditioners when different (from those in the main article text) finite elements are used for the discretization.

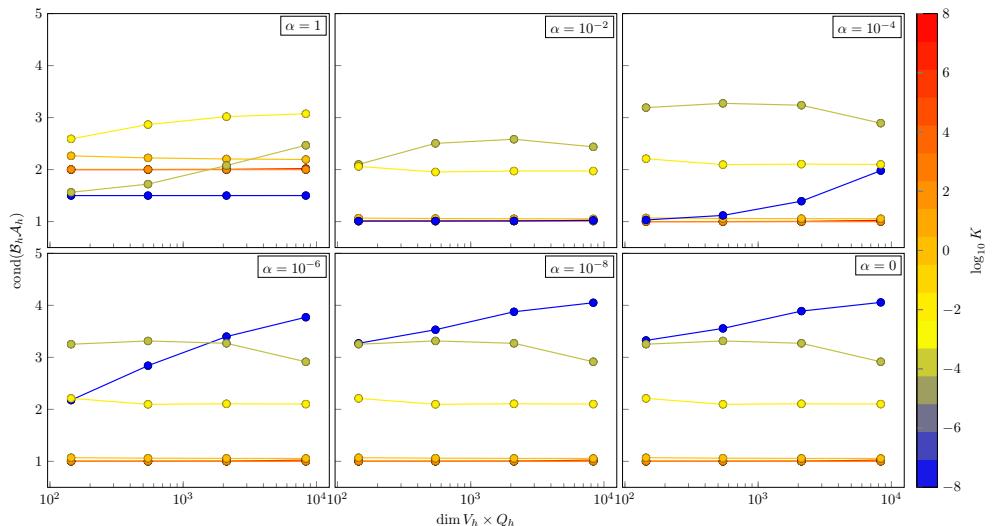


FIG. A.1. Performance of preconditioner (3.8) for the simplified Helmholtz equation with potential boundary conditions. Discretization by  $\mathbb{BDM}_1\text{-}\mathbb{P}_0$  elements.

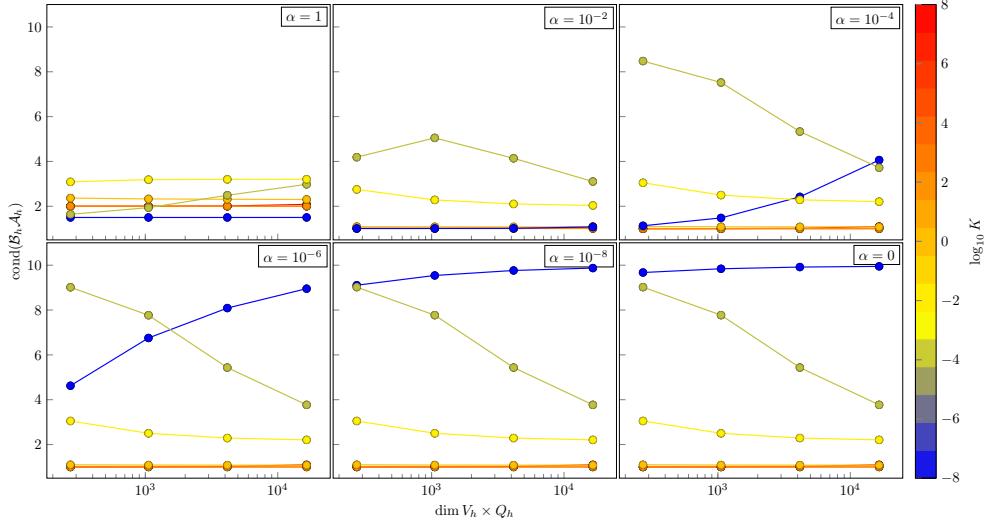


FIG. A.2. Performance of preconditioner (3.8) for the generalized Poisson equation with potential boundary conditions. Discretization by  $\mathbb{RT}_1\text{-}\mathbb{P}_1$  elements.

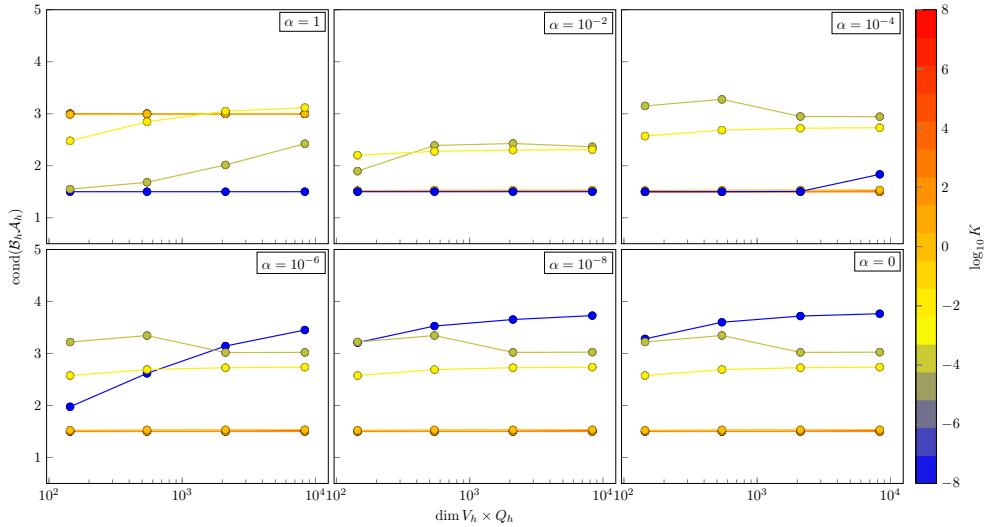


FIG. A.3. Performance of preconditioner (6.3) for the generalized Poisson equation (6.2) with flux boundary conditions. Discretization by  $\mathbb{BDM}_1\text{-}\mathbb{P}_0$  elements.

**Appendix B. Suboptimal preconditioners for Biot equations.** Here we present a few preconditioners that might be the natural and intuitive choices if one starts from the original formulation (1.4), but that do not retain robustness with respect to model parameters. For simplicity, we let  $\tau = 1$  and  $\mu = 1$  and only focus on stability with respect to the remaining model parameters. Numerical experiments then utilize the same (two-dimensional) setup as the robustness study of section 6.2.

First, one can suggest a preconditioner being the inverse of

$$\mathcal{B}_1 = \begin{bmatrix} -\nabla \cdot (2\mu\varepsilon) & 0 & 0 & 0 \\ 0 & K^{-1}I - \nabla\nabla \cdot & 0 & 0 \\ 0 & 0 & \mu^{-1}I & 0 \\ 0 & 0 & 0 & I \end{bmatrix},$$

and one expects that the performance might be compromised for some combinations of the model parameters  $\alpha$ ,  $c$ ,  $K$ ,  $\lambda$ ; cf. Table B.1. Therefore, we consider other options, for instance, exploring preconditioners based on the inverse of

$$\mathcal{B}_2 = \begin{bmatrix} -\nabla \cdot (2\mu\varepsilon) & 0 & 0 & 0 \\ 0 & K^{-1}(I - \nabla\nabla \cdot) & 0 & 0 \\ 0 & 0 & \mu^{-1}I & 0 \\ 0 & 0 & 0 & KI \end{bmatrix},$$

or, alternatively, using

$$\mathcal{B}_3 = \begin{bmatrix} -\nabla \cdot (2\mu\varepsilon) & 0 & 0 & 0 \\ 0 & K^{-1}I - \nabla\nabla \cdot & 0 & 0 \\ 0 & 0 & \mu^{-1}I & 0 \\ 0 & 0 & 0 & (I + (-\nabla \cdot K\nabla)^{-1})^{-1} \end{bmatrix}.$$

TABLE B.1

*Condition numbers of  $\mathcal{B}_1$ -preconditioned Biot problem (1.4) with  $\alpha = 1$ ,  $\mu = 1$ . The preconditioner seems robust in  $K$  and  $\lambda$  for  $c = 1$  but not for  $c < 1$ .*

$c$	$K$	$\lambda$	$h$			$c$	$K$	$\lambda$	$h$		
			$2^{-2}$	$2^{-3}$	$2^{-4}$				$2^{-2}$	$2^{-3}$	$2^{-4}$
1	$10^{-8}$	1	5.25	5.28	5.30	$10^{-8}$	$10^3$	$10^9$	17.62	18.02	18.20
		$10^3$	6.75	7.08	7.22				139	144	146
		$10^9$	6.77	7.11	7.25				153	158	160
	$10^{-4}$	1	5.25	5.30	5.42	$10^{-2}$	$10^{-4}$	$10^3$	17.53	17.91	18.17
		$10^3$	6.75	7.08	7.22				128	132	134
		$10^9$	6.77	7.11	7.25				139	144	146
1	1	1	6.19	6.27	6.31	1	$10^3$	$10^9$	3.91	3.94	3.95
		$10^3$	7.13	7.23	7.28				6.75	7.08	7.22
		$10^9$	7.15	7.26	7.30				6.77	7.11	7.25

Finally, we can consider the following modification of  $\mathcal{B}_2$ :

$$\mathcal{B}_4 = \begin{bmatrix} \nabla \cdot (2\mu\varepsilon) & 0 & 0 & 0 \\ 0 & K^{-1}(I - \nabla\nabla \cdot) & 0 & 0 \\ 0 & 0 & (\mu^{-1} + \frac{1}{\lambda})I & \frac{\alpha}{\lambda}I \\ 0 & 0 & \frac{\alpha}{\lambda}I & (K + c + \frac{\alpha^2}{\lambda})I \end{bmatrix}.$$

However, these preconditioners yield suboptimal performance, as evidenced in Table B.2.

TABLE B.2

Condition numbers of (1.4) with  $\mathcal{B}_2$  preconditioner (left),  $\mathcal{B}_3$  preconditioner (center), and  $\mathcal{B}_4$  preconditioner (right). In all cases  $\alpha = 1$ ,  $\mu = 1$ , and  $c = 1$ . The preconditioners are not  $K, \lambda$ -robust.

$K$	$\lambda$	$h$			$h$			$h$		
		$2^{-2}$	$2^{-3}$	$2^{-4}$	$2^{-2}$	$2^{-3}$	$2^{-4}$	$2^{-2}$	$2^{-3}$	$2^{-4}$
$10^{-8}$	1	$1 \times 10^{11}$	$5 \times 10^{11}$	$2 \times 10^{12}$	$9 \times 10^7$	$8 \times 10^7$	$7 \times 10^7$	721	$3 \times 10^3$	$1 \times 10^4$
	$10^3$	$5 \times 10^{10}$	$2 \times 10^{11}$	$9 \times 10^{11}$	$1 \times 10^8$	$1 \times 10^8$	$1 \times 10^8$	833	$4 \times 10^3$	$1 \times 10^4$
	$10^9$	$5 \times 10^{10}$	$2 \times 10^{11}$	$9 \times 10^{11}$	$1 \times 10^8$	$1 \times 10^8$	$1 \times 10^8$	833	$4 \times 10^3$	$1 \times 10^4$
$10^{-4}$	1	$1 \times 10^7$	$4 \times 10^7$	$1 \times 10^8$	$9 \times 10^3$	$8 \times 10^3$	$7 \times 10^3$	693	$3 \times 10^3$	$7 \times 10^3$
	$10^3$	$5 \times 10^6$	$2 \times 10^7$	$5 \times 10^7$	$1 \times 10^4$	$1 \times 10^4$	$1 \times 10^4$	790	$3 \times 10^3$	$8 \times 10^3$
	$10^9$	$5 \times 10^6$	$2 \times 10^7$	$5 \times 10^7$	$1 \times 10^4$	$1 \times 10^4$	$1 \times 10^4$	790	$3 \times 10^3$	$8 \times 10^3$
1	1	6.19	6.27	6.31	7.35	7.23	7.18	3.33	3.46	3.52
	$10^3$	7.13	7.23	7.28	8.37	8.25	8.19	6.75	7.08	7.23
	$10^9$	7.15	7.26	7.30	8.39	8.28	8.22	6.77	7.11	7.25

**Appendix C. Herrmann formulation of linear elasticity.** Let us consider a two-dimensional domain on which the equations of linear elasticity are written as

$$\begin{aligned} -\nabla \cdot (2\mu\varepsilon(\mathbf{u}) + pI) &= \mathbf{f} && \text{in } \Omega, \\ \nabla \cdot \mathbf{u} - \lambda^{-1}p &= 0 && \text{in } \Omega, \end{aligned}$$

with pure displacement boundary conditions. In view of constructing locking-free solvers, one is interested in discretizations that are robust with respect to  $\lambda$  (and also with respect to  $\mu$ ).

The following preconditioner is robust:

$$(C.1) \quad \mathcal{B}_H = \begin{bmatrix} -\nabla \cdot (2\mu\varepsilon(\mathbf{u})) & 0 \\ 0 & \mu^{-1}(I - \Pi_{\mathbb{R}}) + \lambda^{-1}I \end{bmatrix}^{-1},$$

as we can see in Table C.1.

TABLE C.1  
Condition numbers associated with the preconditioner (C.1) for the Herrmann problem.

$\lambda$	$\mu$	$h$			$\lambda$	$\mu$	$h$		
		$2^{-2}$	$2^{-3}$	$2^{-4}$			$2^{-2}$	$2^{-3}$	$2^{-4}$
1	$10^{-6}$	18.13	18.19	18.20	$10^4$	$10^{-6}$	18.13	18.19	18.20
	$10^{-4}$	18.11	18.17	18.18		$10^{-4}$	18.13	18.19	18.20
	$10^{-2}$	16.22	16.27	16.28		$10^{-2}$	18.13	18.19	18.20
	1	2.19	2.19	2.19		1	18.11	18.17	18.18
	$10^2$	1.01	1.01	1.01		$10^2$	16.22	16.27	16.28
	$10^4$	1.00	1.00	1.00		$10^4$	2.19	2.19	2.19
	$10^8$	1.00	1.00	1.00		$10^8$	1.00	1.00	1.00
	$10^{10}$	1.00	1.00	1.00		$10^{10}$	1.00	1.00	1.00
$10^2$	$10^{-6}$	18.13	18.19	18.20	$10^8$	$10^{-6}$	18.13	18.19	18.20
	$10^{-4}$	18.13	18.19	18.20		$10^{-4}$	18.13	18.19	18.20
	$10^{-2}$	18.11	18.17	18.18		$10^{-2}$	18.13	18.19	18.20
	1	16.22	16.27	16.28		1	18.13	18.19	18.20
	$10^2$	2.19	2.19	2.19		$10^2$	18.13	18.19	18.20
	$10^4$	1.01	1.01	1.01		$10^4$	18.11	18.17	18.18
	$10^8$	1.00	1.00	1.00		$10^8$	2.19	2.19	2.19
	$10^{10}$	1.00	1.00	1.00		$10^{10}$	1.01	1.01	1.01

In regard to Theorem 2.1, it is clear that for  $\lambda = \infty$ , the Brezzi conditions are satisfied as the problem is then reduced to the Stokes problem on  $\mu^{1/2}\mathbf{H}_0^1(\Omega) \times$

$\mu^{-1/2}L_0^2(\Omega)$ . Furthermore, as the bilinear form  $a(\cdot, \cdot)$  is coercive in the whole space  $\mu^{1/2}\mathbf{H}_0^1(\Omega)$  the Braess condition is automatically satisfied.

## REFERENCES

- [1] E. AHMED, J. M. NORDBOTSEN, AND F. A. RADU, *Adaptive asynchronous time-stepping, stopping criteria, and a posteriori error estimates for fixed-stress iterative schemes for coupled poromechanics problems*, J. Comput. Appl. Math., 364 (2020), 112312.
- [2] D. N. ARNOLD, R. S. FALK, AND R. WINTHER, *Multigrid in  $H(\text{div})$  and  $H(\text{curl})$* , Numer. Math., 85 (2000), pp. 197–217.
- [3] T. BÆRLAND, M. KUCHTA, K.-A. MARDAL, AND T. THOMPSON, *An observation on the uniform preconditioners for the mixed Darcy problem*, Numer. Methods Partial Differential Equations, 36 (2020), pp. 1718–1734.
- [4] L. BERGER, R. BORDAS, D. KAY, AND S. TAVENER, *Stabilized lowest-order finite element approximation for linear three-field poroelasticity*, SIAM J. Sci. Comput., 37 (2015), pp. A2222–A2245, <https://doi.org/10.1137/15M1009822>.
- [5] J. BERGH AND J. LÖFSTRÖM, *Interpolation Spaces: An Introduction*, Grundlehren der Mathematischen Wissenschaften 223, Springer-Verlag, Berlin, New York, 2012.
- [6] D. BOFFI, F. BREZZI, AND M. FORTIN, *Mixed Finite Element Methods and Applications*, Springer Ser. Comput. Math. 44, Springer-Verlag, Heidelberg, 2013.
- [7] J. W. BOTH, M. BORREGALES, J. M. NORDBOTSEN, K. KUMAR, AND F. A. RADU, *Robust fixed stress splitting for Biot's equations in heterogeneous media*, Appl. Math. Lett., 68 (2017), pp. 101–108.
- [8] D. BRAESS, *Stability of saddle point problems with penalty*, RAIRO Modél. Math. Anal. Numér., 30 (1996), pp. 731–742.
- [9] A. BRANDT, *Algebraic multigrid theory: The symmetric case*, Appl. Math. Comput., 19 (1986), pp. 23–56.
- [10] F. BREZZI, J. DOUGLAS, JR., AND L. D. MARINI, *Two families of mixed finite elements for second order elliptic problems*, Numer. Math., 47 (1985), pp. 217–235.
- [11] T. CHAUMONT-FRELET, *Mixed finite element discretizations of acoustic Helmholtz problems with high wavenumbers*, Calcolo, 56 (2019), 49.
- [12] R. D. FALGOUT AND U. M. YANG, *hypre: A library of high performance preconditioners*, in Proceedings of the International Conference on Computational Science, Lecture Notes in Comput. Sci. 2331, Springer, Berlin, Heidelberg, 2002, pp. 632–641.
- [13] P. E. FARRELL, M. G. KNEPLEY, F. WECHSUNG, AND L. MITCHELL, *PCPATCH: Software for the Topological Construction of Multigrid Relaxation Methods*, preprint, <https://arxiv.org/abs/1912.08516>, 2019.
- [14] F. J. GASPAR, J. L. GRACIA, F. J. LISBONA, AND C. W. OOSTERLEE, *Distributive smoothers in multigrid for problems with dominating grad-div operators*, Numer. Linear Algebra Appl., 15 (2008), pp. 661–683.
- [15] G. N. GATICA, *A Simple Introduction to the Mixed Finite Element Method*, Springer-Verlag, Berlin, 2014.
- [16] G. HARPER, R. WANG, J. LIU, S. TAVENER, AND R. ZHANG, *A locking-free solver for linear elasticity on quadrilateral and hexahedral meshes based on enrichment of Lagrangian elements*, Comput. Math. Appl., 80 (2020), pp. 1578–1595.
- [17] L. R. HERRMANN, *Elasticity equations for incompressible and nearly incompressible materials by a variational theorem*, AIAA J., 3 (1965), pp. 1896–1900.
- [18] Q. HONG AND J. KRAUS, *Parameter-robust stability of classical three-field formulation of Biot's consolidation model*, Electron. Trans. Numer. Anal., 48 (2018), pp. 202–226.
- [19] Q. HONG, J. KRAUS, M. LYMBERY, AND F. PHILO, *Conservative discretizations and parameter-robust preconditioners for Biot and multiple-network flux-based poroelasticity models*, Numer. Linear Algebra Appl., 26 (2019), e2242.
- [20] X. HU, C. RODRIGO, F. J. GASPAR, AND L. T. ZIKATANOV, *A nonconforming finite element method for the Biot's consolidation model in poroelasticity*, J. Comput. Appl. Math., 310 (2017), pp. 143–154.
- [21] T. V. KOLEV AND P. S. VASSILEVSKI, *Parallel auxiliary space AMG solver for  $H(\text{div})$  problems*, SIAM J. Sci. Comput., 34 (2012), pp. A3079–A3098, <https://doi.org/10.1137/110859361>.
- [22] S. KUMAR, R. OYARZÚA, R. RUIZ-BAIER, AND R. SANDILYA, *Conservative discontinuous finite volume and mixed schemes for a new four-field formulation in poroelasticity*, ESAIM Math. Model. Numer. Anal., 54 (2020), pp. 273–299.
- [23] J. LEE, K.-A. MARDAL, AND R. WINTHER, *Parameter-robust discretization and preconditioning*

- of Biot's consolidation model*, SIAM J. Sci. Comput., 39 (2017), pp. A1–A24, <https://doi.org/10.1137/15M1029473>.
- [24] Y. LI AND L. T. ZIKATANOV, *Residual-based a posteriori error estimates of mixed methods for a three-field Biot's consolidation model*, IMA J. Numer. Anal., (2020), <https://doi.org/10.1093/imanum/draa074>.
- [25] K.-A. MARDAL, M. E. ROGNES, AND T. B. THOMPSON, *Accurate Discretization of Poroelasticity Without Darcy Stability—Stokes-Biot Stability Revisited*, preprint, <https://arxiv.org/abs/2007.10012>, 2020.
- [26] K.-A. MARDAL AND R. WINTHER, *Preconditioning discretizations of systems of partial differential equations*, Numer. Linear Algebra Appl., 18 (2011), pp. 1–40.
- [27] A. MIKELIĆ, B. WANG, AND M. F. WHEELER, *Numerical convergence study of iterative coupling for coupled flow and geomechanics*, Comput. Geosci., 18 (2014), pp. 325–341.
- [28] P. MONK, J. SCHÖBERL, AND A. SINWEL, *Hybridizing Raviart-Thomas elements for the Helmholtz equation*, Electromagnetics, 30 (2010), pp. 149–176.
- [29] R. OYARZÚA AND R. RUIZ-BAIER, *Locking-free finite element methods for poroelasticity*, SIAM J. Numer. Anal., 54 (2016), pp. 2951–2973, <https://doi.org/10.1137/15M1050082>.
- [30] P. J. PHILLIPS AND M. F. WHEELER, *A coupling of mixed and continuous Galerkin finite element methods for poroelasticity. I. The continuous in time case*, Comput. Geosci., 11 (2007), pp. 131–144.
- [31] E. PIERSANTI, J. J. LEE, T. THOMPSON, K.-A. MARDAL, AND M. E. ROGNES, *Parameter Robust Preconditioning by Congruence for Multiple-network Poroelasticity*, preprint, <https://arxiv.org/abs/2003.09641>, 2020.
- [32] C. E. POWELL AND D. SILVESTER, *Optimal preconditioning for Raviart–Thomas mixed formulation of second-order elliptic problems*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 718–738, <https://doi.org/10.1137/S0895479802404428>.
- [33] P. A. RAVIART AND J. M. THOMAS, *A mixed finite element method for 2-nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, I. Galligani and E. Magenes, eds., Lecture Notes in Math. 606, Springer, Berlin, 1977, pp. 292–315.
- [34] R. RUIZ-BAIER AND I. LUNATI, *Mixed finite element—discontinuous finite volume element discretization of a general class of multicontinuum models*, J. Comput. Phys., 322 (2016), pp. 666–688.
- [35] T. RUSTEN, P. VASSILEVSKI, AND R. WINTHER, *Interior penalty preconditioners for mixed finite element approximations of elliptic problems*, Math. Comput., 65 (1996), pp. 447–466.
- [36] M. SUN AND H. RUI, *A coupling of weak Galerkin and mixed finite element methods for poroelasticity*, Comput. Math. Appl., 73 (2017), pp. 804–823.
- [37] A. TOSELLI AND O. WIDLUND, *Domain Decomposition Methods — Algorithms and Theory*, Springer Ser. Comput. Math. 34, Springer-Verlag, Berlin, Heidelberg, 2006.
- [38] P. VANĚK, J. MANDEL, AND M. BREZINA, *Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems*, Computing, 56 (1996), pp. 179–196.
- [39] P. S. VASSILEVSKI AND U. VILLA, *A block-diagonal algebraic multigrid preconditioner for the Brinkman problem*, SIAM J. Sci. Comput., 35 (2013), pp. S3–S17, <https://doi.org/10.1137/120882846>.
- [40] P. S. VASSILEVSKI AND U. VILLA, *A mixed formulation for the Brinkman problem*, SIAM J. Numer. Anal., 52 (2014), pp. 258–281, <https://doi.org/10.1137/120884109>.
- [41] M. F. WHEELER, G. XUE, AND I. YOTOV, *Coupling multipoint flux mixed finite element methods with continuous Galerkin methods for poroelasticity*, Comput. Geosci., 18 (2007), pp. 57–75.
- [42] S.-Y. YI, *A coupling of nonconforming and mixed finite element methods for Biot's consolidation model*, Numer. Methods Partial Differential Equations, 29 (2013), pp. 1749–1777.
- [43] S.-Y. YI, *A study of two modes of locking in poroelasticity*, SIAM J. Numer. Anal., 55 (2017), pp. 1915–1936, <https://doi.org/10.1137/16M1056109>.
- [44] Y. ZENG, M. CAI, AND F. WANG, *An  $H(\text{div})$ -conforming finite element method for the Biot consolidation model*, East Asian J. Appl. Math., 9 (2019), pp. 558–579.