

RESEARCH ARTICLE

# On the optimization of the fixed-stress splitting for Biot's equations

Erlend Storvik<sup>1</sup> | Jakub W. Both<sup>1</sup> | Kundan Kumar<sup>1,2</sup> | Jan M. Nordbotten<sup>1,3</sup> | Florin A. Radu<sup>1</sup> 

<sup>1</sup>Department of Mathematics, University of Bergen, Bergen, Norway

<sup>2</sup>Department of Mathematics and Computer Science, Karlstad University, Karlstad, Sweden

<sup>3</sup>Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

## Correspondence

Florin A. Radu, Department of Mathematics, University of Bergen, Allégaten 41, 5007 Bergen, Norway.  
Email: florin.radu@uib.no

## Funding information

Norges Forskningsråd, Grant/Award Number: 250223

## Summary

In this work, we are interested in efficiently solving the quasi-static, linear Biot model for poroelasticity. We consider the fixed-stress splitting scheme, which is a popular method for iteratively solving Biot's equations. It is well known that the convergence properties of the method strongly depend on the applied stabilization/tuning parameter. We show theoretically that, in addition to depending on the mechanical properties of the porous medium and the coupling coefficient, they also depend on the fluid flow and spatial discretization properties. The type of analysis presented in this paper is not restricted to a particular spatial discretization, although it is required to be inf-sup stable with respect to the displacement-pressure formulation. Furthermore, we propose a way to optimize this parameter that relies on the mesh independence of the scheme's optimal stabilization parameter. Illustrative numerical examples show that using the optimized stabilization parameter can significantly reduce the number of iterations.

## KEYWORDS

Biot model, convergence analysis, fixed-stress splitting, geomechanics, poroelasticity

## 1 | INTRODUCTION

There is currently a strong interest in the numerical simulation of poroelasticity, ie, fully coupled porous media flow and mechanics. This is due to its high number of societal relevant applications, such as geothermal energy extraction, life sciences, or CO<sub>2</sub> storage, to name a few. The most commonly used mathematical model for poroelasticity is the quasi-static, linear Biot model. It is the coupled problem arising when considering the balance of linear momentum for the porous medium allowing for only small deformations (1) and mass conservation and Darcy's law for the fluid flow (2) (see, eg, the work of Coussy<sup>1</sup>): find  $(\mathbf{u}, p)$  such that

$$-\nabla \cdot (2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\nabla \cdot \mathbf{u}\mathbf{I}) + \alpha\nabla p = \mathbf{f}, \quad (1)$$

$$\frac{\partial}{\partial t} \left( \frac{p}{M} + \alpha\nabla \cdot \mathbf{u} \right) - \nabla \cdot (\kappa(\nabla p - \mathbf{g}\rho)) = S_f, \quad (2)$$

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *International Journal for Numerical Methods in Engineering* Published by John Wiley & Sons, Ltd.

where  $\mathbf{u}$  is the displacement;  $\varepsilon(\mathbf{u}) := \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T)$  is the (linear) strain tensor;  $\mu$  and  $\lambda$  are the Lamé parameters;  $\alpha$  is the Biot-Willis constant;  $p$  and  $\rho$  are the fluid's pressure and density, respectively;  $1/M$  is the compressibility constant;  $\mathbf{g}$  is the gravitational vector; and  $\kappa$  is the permeability. The source terms  $\mathbf{f}$  and  $S_f$  represent the density of applied body forces and a forced fluid extraction or injection process, respectively.

A lot of work has been done concerning the discretization of Biot's equations (1) and (2). Various spatial discretizations, combined with the backward Euler method as temporal discretization, have been proposed and analyzed. We mention cell-centered finite volumes,<sup>2</sup> continuous Galerkin for the mechanics and mixed finite elements for the flow,<sup>3-6</sup> mixed finite elements for flow and mechanics,<sup>4,7</sup> nonconforming finite elements,<sup>8</sup> the MINI element,<sup>9</sup> continuous or discontinuous Galerkin,<sup>10-12</sup> or multiscale methods.<sup>13-15</sup> Continuous and discontinuous higher-order Galerkin space-time finite elements were proposed in the work of Bause et al.<sup>16</sup> Adaptive computations were considered, for example, in the work of Ern and Meunier.<sup>17</sup> A Monte Carlo approach was proposed in the work of Rahrah and Vermolen.<sup>18</sup> For a discussion on the stability of different spatial discretizations, we refer to the recent papers.<sup>19,20</sup>

Independently of the chosen discretization, there are two popular alternatives for solving Biot's equations: monolithically or by using an iterative splitting algorithm. The former has the advantage of being unconditionally stable, whereas a splitting method is much easier to implement, typically building on already available, tailored, separate numerical codes for porous media flow and for mechanics. However, a naive splitting of Biot's equations will lead to an unstable scheme.<sup>21</sup> To overcome this, one adds a stabilization term in either the mechanics equation (the so-called *undrained splitting scheme*<sup>22</sup>) or the flow equation (the *fixed-stress splitting scheme*).<sup>23</sup> The splitting methods have very good convergence properties, making them a valuable alternative to monolithic solvers for simulation of the linear Biot model (see, eg, the works of Both et al,<sup>5</sup> Kim et al,<sup>21</sup> Settari and Mourits,<sup>23</sup> and Mikelić and Wheeler<sup>24</sup>). In the present work, we will discuss the fixed-stress splitting scheme. For other splitting schemes, see, for example, the works of Turska and Schrefler<sup>25</sup> and Turska et al.<sup>26</sup>

After applying the backward Euler method in time to (1) and (2) and discretizing in space (using finite elements or finite volumes), one has to solve a fully coupled, discrete system at each time step. The fixed-stress splitting scheme is an iterative splitting scheme to solve this system. Let  $i$  denote the iteration index, and look for a pair  $(\mathbf{u}^i, p^i)$  to converge to the solution  $(\mathbf{u}, p)$ , when  $i \rightarrow +\infty$ . Algorithmically, one first solves the flow equation (2) using the displacement from the previous iteration, and then, one solves the mechanics equation (1) with the updated pressure and iterates until convergence is achieved. To ensure convergence,<sup>5,21,24</sup> one needs to add a stabilizing term  $L(p^i - p^{i-1})$  to the flow equation (2). The free-to-be-chosen parameter  $L \geq 0$  is called the stabilization or tuning parameter. Choosing the value of this parameter is of major importance to the performance of the algorithm, because the number of iterations strongly depends on its value (see the works of Both et al,<sup>5</sup> Bause et al,<sup>16</sup> Both and Köcher,<sup>27</sup> Mikelić et al,<sup>28</sup> and Dana et al<sup>29</sup>). Moreover, a too small or too big  $L$  will lead to slow or no convergence.

The initial derivation of the fixed-stress splitting scheme had a physical motivation<sup>21,23</sup>: one “fixes the (volumetric) stress,” ie, imposes  $K_{\text{dr}} \nabla \cdot \mathbf{u}^i - \alpha p^i = K_{\text{dr}} \nabla \cdot \mathbf{u}^{i-1} - \alpha p^{i-1}$  and uses this to replace  $\alpha \nabla \cdot \mathbf{u}^i$  in the flow equation. Here,  $K_{\text{dr}}$  is the physical drained bulk modulus. The resulting stabilization parameter  $L$ , called from now on the *physical* stabilization parameter, is  $L_{\text{phys}} = \frac{\alpha^2}{K_{\text{dr}}}$  (depending on the mechanical properties and the Biot coefficient). In 2013, a rigorous mathematical analysis of the fixed-stress splitting scheme was performed for the first time in the work of Mikelić and Wheeler.<sup>24</sup> The authors show that the scheme is a contraction for any stabilization parameter  $L \geq \frac{L_{\text{phys}}}{2}$ . This analysis was confirmed in the work of Both et al<sup>5</sup> for heterogeneous media using a simpler technique, and the same result was obtained for both continuous and discontinuous Galerkin higher-order space-time finite elements in the works of Bause et al<sup>16</sup> and Bause,<sup>30</sup> implying that the value of the stabilization parameter does not depend on the order of the spatial discretization. The question of which stabilization parameter is the optimal one (in the sense that it requires the least number of iterations to converge) arises, and the aim of this paper is to answer this open question.

In a recent study,<sup>27</sup> the authors studied the convergence of the fixed-stress splitting scheme for different test cases with varying material parameters. They determined numerically the optimal stabilization parameter for each considered case. This study, together with the previous results presented in the works of Mikelić et al<sup>28</sup> and Both et al,<sup>5</sup> suggests that the optimal parameter actually is a value in the interval  $[\frac{L_{\text{phys}}}{2}, L_{\text{phys}}]$ , depending on the data. In particular, the optimal parameter depends on the problem's boundary conditions and flow parameters, and not only on its mechanical properties and coupling coefficient. Nevertheless, to the best of our knowledge, there exists no theoretical evidence for this in the literature so far.

In this paper, we propose for the first time that the optimal stabilization parameter for the fixed-stress splitting scheme lies in the interval  $[\frac{\alpha^2}{4\mu+2\lambda}, \frac{\alpha^2}{K_{\text{dr}}}] \supseteq [\frac{L_{\text{phys}}}{2}, L_{\text{phys}}]$  and depends also on the fluid flow properties and stability properties of

the spatial discretization. This is achieved through refining the proof techniques in the work of Both et al<sup>5</sup> to obtain an improved linear rate of convergence; minimizing this rate with respect to the stabilization parameter gives the “theoretical” optimal choice. Although the trends for the practical and the proposed theoretically optimal stabilization parameter are sound for varying material parameters, the theoretically calculated one does not show great practical promise in terms of being optimal (see the work of Storvik et al<sup>31</sup> for a supplementary numerical study). This is due to harsh bounds that have been used in the proof. Therefore, we propose a brute-force approach for optimizing the stabilization parameter, utilizing the newly found interval  $[\frac{\alpha^2}{4\mu+2\lambda}, \frac{\alpha^2}{K_{dr}})$ .

In contrast to previous works, the spatial discretization is required to be inf-sup stable, which essentially allows for the control of errors in the pressure by those in the stress. A novel consequence of our theoretical result is that under the use of an inf-sup-stable discretization, the fixed-stress splitting scheme also converges robustly in the limit case of incompressible fluids and impermeable porous media.

In Section 4, numerical experiments are performed, which show the soundness and efficiency of the proposed optimization technique. In particular, we show that the optimized stabilization parameter can be far superior to a naive choice among the classical stabilization parameters,  $L_{\text{phys}}$  or  $\frac{L_{\text{phys}}}{2}$ .

To summarize, the main contributions of this work are as follows:

- an improved, theoretical convergence result for the fixed-stress splitting scheme under the assumption of an inf-sup-stable discretization;
- the derivation of an explicit interval for the optimal stabilization parameter, depending solely on the material parameters;
- a brute-force approach for optimizing the stabilization parameter, relying on a nearly mesh-independent performance of the fixed-stress splitting.

We mention that the fixed-stress splitting scheme also can be applied to more involved extensions of Biot's equations, for example, including nonlinear water compressibility,<sup>32</sup> unsaturated poroelasticity,<sup>33,34</sup> the multiple-network poroelasticity theory,<sup>35,36</sup> finite-strain poroplasticity,<sup>37</sup> fractured porous media,<sup>38</sup> and fracture propagation.<sup>39,40</sup> For nonlinear problems, one combines a linearization technique, eg, the  $L$ -scheme,<sup>41,42</sup> with the splitting algorithm; the convergence of the resulting scheme can be proved rigorously.<sup>32,33</sup> Finally, we would like to mention some valuable variants of the fixed-stress splitting scheme: the multirate fixed-stress method,<sup>43</sup> the multiscale fixed-stress method,<sup>29</sup> and the parallel-in-time fixed-stress method.<sup>44</sup>

This paper is structured as follows. The notation, the discretization, and the fixed-stress splitting scheme are presented in Section 2. The theoretical analysis of the convergence and the optimization technique are the subject of Section 3. In Section 4, numerical experiments that test the optimization technique are presented. Finally, conclusions are given in Section 5.

## 2 | THE NUMERICAL SCHEME FOR SOLVING BIOT'S MODEL

In this paper, we use common notations in functional analysis. Let  $\Omega \subset \mathbb{R}^d$  be a Lipschitz domain where  $d$  is the spatial dimension. The space  $L^2(\Omega)$  is the Hilbert space of Lebesgue-measurable, square-integrable functions on  $\Omega$ , and  $H^1(\Omega)$  is the Hilbert space of functions in  $L^2(\Omega)$  with derivatives (in the weak sense) in  $L^2(\Omega)$ . The inner product and its associated norm in  $L^2(\Omega)$  are denoted by  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$ , respectively, and  $\|\cdot\|_{H^1(\Omega)}$  is the standard  $H^1(\Omega)$ -norm. Vectors and tensors are written bold, and, sometimes, the scalar product and the norm will be taken for vectors and tensors. Vectorial functions are written bold-italic.  $T$  will denote the final time.

Biot equations (1) and (2) are solved in the domain  $\Omega \times (0, T)$  together with (for simplicity) homogeneous Dirichlet boundary conditions and a given initial condition. In time, the backward Euler method is applied with a constant time-step size  $\tau := \frac{T}{N}$ ,  $N \in \mathbb{N}$ . Throughout this work, the index  $n$  will refer to the time level. For the spatial discretization, a two-field Galerkin finite element formulation is considered, and two generic discrete spaces  $\mathbf{V}_h$  and  $Q_h$ , associated with displacements and pressures, are introduced. Later, we require  $\mathbf{V}_h \times Q_h$  to be inf-sup stable with respect to the divergence operator; the most prominent inf-sup-stable example is the Taylor-Hood element, ie, P2-P1 for displacement and pressure.<sup>45</sup> Nevertheless, the analysis below can be extended without difficulties to a three-field formulation as, for example, in the works of Phillips and Wheeler,<sup>3</sup> Both et al,<sup>5</sup> and Berger et al.<sup>6</sup>

In this way, the fully discrete, weak problem reads: let  $n \geq 1$  and assume  $(\mathbf{u}_h^{n-1}, p_h^{n-1}) \in \mathbf{V}_h \times Q_h$  are given. Find  $(\mathbf{u}_h^n, p_h^n) \in \mathbf{V}_h \times Q_h$  such that

$$2\mu \langle \boldsymbol{\varepsilon}(\mathbf{u}_h^n), \boldsymbol{\varepsilon}(\mathbf{v}_h) \rangle + \lambda \langle \nabla \cdot \mathbf{u}_h^n, \nabla \cdot \mathbf{v}_h \rangle - \alpha \langle p_h^n, \nabla \cdot \mathbf{v}_h \rangle = \langle \mathbf{f}^n, \mathbf{v}_h \rangle, \quad (3)$$

$$\frac{1}{M} \langle p_h^n - p_h^{n-1}, q_h \rangle + \alpha \langle \nabla \cdot (\mathbf{u}_h^n - \mathbf{u}_h^{n-1}), q_h \rangle + \tau \langle \kappa \nabla p_h^n, \nabla q_h \rangle - \tau \langle \kappa \mathbf{g} \rho, \nabla q_h \rangle = \tau \langle S_f^n, q_h \rangle \quad (4)$$

for all  $\mathbf{v}_h \in \mathbf{V}_h, q_h \in Q_h$ . For  $n = 1$ , the functions  $(\mathbf{u}_h^{n-1}, p_h^{n-1})$  are obtained by using the initial condition.

The fixed-stress splitting scheme<sup>5,21,23,28</sup> is now introduced. Denote by  $i$  the iteration index. Iterate until convergence.

For  $i \geq 1$ , given a stabilization parameter  $L \geq 0$  and  $(\mathbf{u}_h^{n-1}, p_h^{n-1}), (\mathbf{u}_h^{n,i-1}, p_h^{n,i-1}) \in \mathbf{V}_h \times Q_h$ , find  $(\mathbf{u}_h^{n,i}, p_h^{n,i}) \in \mathbf{V}_h \times Q_h$  such that

$$2\mu \langle \boldsymbol{\varepsilon}(\mathbf{u}_h^{n,i}), \boldsymbol{\varepsilon}(\mathbf{v}_h) \rangle + \lambda \langle \nabla \cdot \mathbf{u}_h^{n,i}, \nabla \cdot \mathbf{v}_h \rangle - \alpha \langle p_h^{n,i}, \nabla \cdot \mathbf{v}_h \rangle = \langle \mathbf{f}^n, \mathbf{v}_h \rangle, \quad (5)$$

$$\begin{aligned} \frac{1}{M} \langle p_h^{n,i} - p_h^{n-1}, q_h \rangle + \alpha \langle \nabla \cdot (\mathbf{u}_h^{n,i-1} - \mathbf{u}_h^{n-1}), q_h \rangle + L \langle p_h^{n,i} - p_h^{n,i-1}, q_h \rangle \\ + \tau \langle \kappa \nabla p_h^{n,i}, \nabla q_h \rangle - \tau \langle \kappa \mathbf{g} \rho, \nabla q_h \rangle = \tau \langle S_f^n, q_h \rangle \end{aligned} \quad (6)$$

for all  $\mathbf{v}_h \in \mathbf{V}_h, q_h \in Q_h$ . The initial guess for the iterations is chosen to be the solution at the last time step, ie,  $(\mathbf{u}_h^{n,0}, p_h^{n,0}) := (\mathbf{u}_h^{n-1}, p_h^{n-1})$ . Notice that the mechanics and flow problems decouple, allowing for the use of separate simulators for both subproblems.

### 3 | CONVERGENCE ANALYSIS AND OPTIMIZATION

In this section, the convergence of the scheme (5)-(6) is analyzed. We are particularly interested in finding an *optimal* stabilization parameter  $L$ , in the sense that the scheme requires the least amount of iterations, ie, has the smallest possible convergence rate. Before we proceed with the main result, we need some preliminaries.

**Definition 1.** The mathematical bulk modulus,  $K_{\text{dr}}^* > 0$ , is defined as the largest constant such that

$$2\mu \|\boldsymbol{\varepsilon}(\mathbf{u}_h)\|^2 + \lambda \|\nabla \cdot \mathbf{u}_h\|^2 \geq K_{\text{dr}}^* \|\nabla \cdot \mathbf{u}_h\|^2 \quad \text{for all } \mathbf{u}_h \in \mathbf{V}_h. \quad (7)$$

By the Cauchy-Schwarz inequality, we get that the physical drained bulk modulus  $K_{\text{dr}} = \frac{2\mu}{d} + \lambda$  is a lower bound for  $K_{\text{dr}}^*$ . However, for effectively lower-dimensional situations, eg, a one-dimensional-like compression,  $d$  can be replaced by a value closer to 1. Lemma 1 below guarantees an upper bound for  $K_{\text{dr}}^*$ . Nevertheless, there is a strong indication (based on numerical experiments; see, eg, Section 4 and the work of Both and Köcher<sup>27</sup>) that  $K_{\text{dr}}^* \in [K_{\text{dr}} = \frac{2\mu}{d} + \lambda, 2\mu + \lambda]$ . We remark that the exact value, depending on the physical situation, can be computed as a generalized eigenvalue.

Throughout this paper, we make use of the following two assumptions.

**Assumption 1.** The constants  $\mu, \lambda, \alpha$ , and  $\rho$  are strictly positive, the constants  $1/M$  and  $\kappa$  are nonnegative, and the vector  $\mathbf{g}$  is constant.

**Assumption 2.** The discretization  $\mathbf{V}_h \times Q_h$  is inf-sup stable with respect to the bilinear form  $b(\mathbf{v}_h, q_h) = \langle \nabla \cdot \mathbf{v}_h, q_h \rangle$ .

From Assumption 2 follows Lemma 1 by applying corollary 4.1.1 in the work of Boffi et al,<sup>45</sup> which states as follows.

**Corollary 1.** Let  $V$  and  $Q$  be Hilbert spaces, and let  $B$  be a linear continuous operator from  $V$  to  $Q'$ ; here,  $Q'$  denotes the dual space of  $Q$ . Denote by  $B^t$  the transposed operator of  $B$ . Then, the following two statements are equivalent.

- $B^t$  is bounding:  $\exists \gamma > 0$  such that  $\|B^t q\|_{V'} \geq \gamma \|q\|_Q \quad \forall q \in Q$ .
- $\exists L_B \in \mathcal{L}(Q', V)$  such that  $B(L_B(\xi)) = \xi \quad \forall \xi \in Q'$  with  $\|L_B\| = \frac{1}{\gamma}$ .

**Lemma 1.** Assume Assumption 2. There exists  $\beta > 0$  such that, for any  $p_h \in Q_h$ , there exists  $\mathbf{u}_h \in \mathbf{V}_h$  satisfying  $\langle \nabla \cdot \mathbf{u}_h, q_h \rangle = \langle p_h, q_h \rangle$  for all  $q_h \in Q_h$  and

$$2\mu \|\boldsymbol{\varepsilon}(\mathbf{u}_h)\|^2 + \lambda \|\nabla \cdot \mathbf{u}_h\|^2 \leq \beta \|p_h\|^2. \quad (8)$$

*Proof.* Consider Corollary 1. Let the continuous linear function  $B : \mathbf{V}_h \rightarrow Q'_h$  be defined by  $B(\mathbf{u}_h)(q_h) = \langle \nabla \cdot \mathbf{u}_h, q_h \rangle$ . The first statement of Corollary 1 is a characterization of an inf-sup-stable discretization Assumption 2, with inf-sup constant  $\gamma$ . Hence, the second statement of Corollary 1 holds; there exists a linear function  $L_B \in \mathcal{L}(Q'_h, \mathbf{V}_h)$  such that  $B(L_B(\langle p_h, \cdot \rangle)) = \langle p_h, \cdot \rangle$  for all  $p_h \in Q_h$  with  $\|L_B\| = 1/\gamma$ . In particular,  $L_B$  is mapping  $p_h \in Q_h$  to the corresponding  $\mathbf{u}_h \in \mathbf{V}_h$  such that

$$\langle \nabla \cdot \mathbf{u}_h, q_h \rangle = B(L_B(\langle p_h, \cdot \rangle))(q_h) = \langle p_h, q_h \rangle$$

for all  $q_h \in Q_h$ . Additionally, the following chain of inequalities holds true:

$$2\mu \|\varepsilon(\mathbf{u}_h)\|^2 + \lambda \|\nabla \cdot \mathbf{u}_h\|^2 \leq C \|\mathbf{u}_h\|_{H^1(\Omega)}^2 \leq C \|L_B\|^2 \|p_h\|^2,$$

where the first inequality follows from Young's inequality with  $C$  depending only on the Lamé parameters, and the second inequality results from the operator norm, ie,

$$\|L_B\| = \sup_{0 \neq p_h \in Q_h} \frac{\|L_B(\langle p_h, \cdot \rangle)\|_{H^1(\Omega)}}{\|\langle p_h, \cdot \rangle\|_{L^2(\Omega)'}} = \sup_{\substack{0 \neq p_h \in Q_h \\ \mathbf{u}_h = L_B(\langle p_h, \cdot \rangle)}} \frac{\|\mathbf{u}_h\|_{H^1(\Omega)}}{\|p_h\|}.$$

We obtain our desired inequality, as follows:

$$2\mu \|\varepsilon(\mathbf{u}_h)\|^2 + \lambda \|\nabla \cdot \mathbf{u}_h\|^2 \leq \frac{C}{\gamma^2} \|p_h\|^2 = \beta \|p_h\|^2.$$

□

*Remark 1.* The constant  $\beta$  above depends on  $\mu$ ,  $\lambda$ , and the domain  $\Omega$  and on the choice of the finite-dimensional spaces  $\mathbf{V}_h$  and  $Q_h$ . Similar to  $K_{\text{dr}}^*$ ,  $\beta$  can be computed as a generalized eigenvalue.

We can now give our main convergence result.

**Theorem 1.** Assume that Assumptions 1 and 2 hold true, and let  $\delta \in (0, 2]$ . Define the iteration errors as  $\mathbf{e}_u^{n,i} := \mathbf{u}_h^{n,i} - \mathbf{u}_h^n$  and  $e_p^{n,i} := p_h^{n,i} - p_h^n$ , where  $(\mathbf{u}_h^{n,i}, p_h^{n,i})$  is a solution to (5) and (6), and  $(\mathbf{u}_h^n, p_h^n)$  is a solution to (3) and (4). The fixed-stress splitting scheme (5)-(6) converges linearly for any  $L \geq \frac{\alpha^2}{\delta K_{\text{dr}}^*}$ , with a convergence rate given by

$$\text{rate}(L, \delta) = \frac{L}{L + \frac{2}{M} + \frac{2\tau\kappa}{C_\Omega^2} + (2 - \delta) \frac{\alpha^2}{\beta}}, \quad (9)$$

through the error inequalities

$$\|e_p^{n,i}\|^2 \leq \text{rate}(L, \delta) \|e_p^{n,i-1}\|^2, \quad (10)$$

$$2\mu \left\| \varepsilon(\mathbf{e}_u^{n,i}) \right\|^2 + \lambda \|\nabla \cdot \mathbf{e}_u^{n,i}\|^2 \leq \frac{\alpha^2}{K_{\text{dr}}^*} \|e_p^{n,i}\|^2, \quad (11)$$

where  $C_\Omega$  is the Poincaré constant and  $\beta$  is the constant from (8).

*Proof.* Subtract (5) and (6) from (3) and (4), respectively, to obtain the error equations

$$\begin{cases} (i) & 2\mu \left\langle \varepsilon(\mathbf{e}_u^{n,i}), \varepsilon(\mathbf{v}_h) \right\rangle + \lambda \left\langle \nabla \cdot \mathbf{e}_u^{n,i}, \nabla \cdot \mathbf{v}_h \right\rangle - \alpha \left\langle e_p^{n,i}, \nabla \cdot \mathbf{v}_h \right\rangle = 0, \\ (ii) & \frac{1}{M} \left\langle e_p^{n,i}, q_h \right\rangle + \alpha \left\langle \nabla \cdot \mathbf{e}_u^{n,i-1}, q_h \right\rangle + L \left\langle e_p^{n,i} - e_p^{n,i-1}, q_h \right\rangle + \tau \left\langle \kappa \nabla e_p^{n,i}, \nabla q_h \right\rangle = 0, \end{cases} \quad (12)$$

holding for all  $(\mathbf{v}_h, q_h) \in \mathbf{V}_h \times Q_h$ . To prove (11), test (12)(i) with  $\mathbf{v}_h = \mathbf{e}_u^{n,i}$ , and apply the Cauchy-Schwarz inequality and Young's inequality to the pressure term to obtain

$$2\mu \left\| \varepsilon(\mathbf{e}_u^{n,i}) \right\|^2 + \lambda \|\nabla \cdot \mathbf{e}_u^{n,i}\|^2 \leq \frac{\alpha^2}{2K_{\text{dr}}^*} \|e_p^{n,i}\|^2 + \frac{K_{\text{dr}}^*}{2} \|\nabla \cdot \mathbf{e}_u^{n,i}\|^2. \quad (13)$$

We now get (11) by applying (7).

In order to prove (10), test (12) with  $q_h = e_p^{n,i}$  and  $\mathbf{v}_h = \mathbf{e}_u^{n,i}$ , add the resulting equations, and use the algebraic identity

$$\langle e_p^{n,i} - e_p^{n,i-1}, e_p^{n,i} \rangle = \frac{1}{2} \left( \|e_p^{n,i} - e_p^{n,i-1}\|^2 + \|e_p^{n,i}\|^2 - \|e_p^{n,i-1}\|^2 \right)$$

to get

$$\begin{aligned} & 2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} \right) \right\|^2 + \lambda \left\| \nabla \cdot \mathbf{e}_u^{n,i} \right\|^2 + \frac{1}{M} \|e_p^{n,i}\|^2 - \alpha \langle e_p^{n,i}, \nabla \cdot (\mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1}) \rangle + \tau \kappa \left\| \nabla e_p^{n,i} \right\|^2 + \frac{L}{2} \|e_p^{n,i} - e_p^{n,i-1}\|^2 + \frac{L}{2} \|e_p^{n,i}\|^2 \\ & = \frac{L}{2} \|e_p^{n,i-1}\|^2. \end{aligned}$$

Using now Equation (12)(i), tested with  $\mathbf{v}_h = \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1}$  in the above, yields

$$\begin{aligned} & 2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} \right) \right\|^2 + \lambda \left\| \nabla \cdot \mathbf{e}_u^{n,i} \right\|^2 + \frac{1}{M} \|e_p^{n,i}\|^2 + \tau \kappa \left\| \nabla e_p^{n,i} \right\|^2 + \frac{L}{2} \|e_p^{n,i}\|^2 + \frac{L}{2} \|e_p^{n,i} - e_p^{n,i-1}\|^2 \\ & = \frac{L}{2} \|e_p^{n,i-1}\|^2 + 2\mu \langle \varepsilon \left( \mathbf{e}_u^{n,i} \right), \varepsilon \left( \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1} \right) \rangle + \lambda \langle \nabla \cdot \mathbf{e}_u^{n,i}, \nabla \cdot \left( \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1} \right) \rangle. \end{aligned} \quad (14)$$

By applying Young's inequality in (14), we obtain that, for any  $\delta > 0$ , there holds

$$\begin{aligned} & 2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} \right) \right\|^2 + \lambda \left\| \nabla \cdot \mathbf{e}_u^{n,i} \right\|^2 + \frac{1}{M} \|e_p^{n,i}\|^2 + \tau \kappa \left\| \nabla e_p^{n,i} \right\|^2 + \frac{L}{2} \|e_p^{n,i}\|^2 + \frac{L}{2} \|e_p^{n,i} - e_p^{n,i-1}\|^2 \\ & = \frac{L}{2} \|e_p^{n,i-1}\|^2 + \frac{\delta}{2} \left( 2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} \right) \right\|^2 + \lambda \left\| \nabla \cdot \mathbf{e}_u^{n,i} \right\|^2 \right) + \frac{1}{2\delta} \left( 2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1} \right) \right\|^2 + \lambda \left\| \nabla \cdot \left( \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1} \right) \right\|^2 \right). \end{aligned} \quad (15)$$

To take care of the last term in (15), consider Equation (12)(i), subtract iteration  $i - 1$  from iteration  $i$ , let  $\mathbf{v}_h = \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1}$  in the result, and apply the Cauchy-Schwarz inequality to get

$$2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} \right) - \varepsilon \left( \mathbf{e}_u^{n,i-1} \right) \right\|^2 + \lambda \left\| \nabla \cdot \left( \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1} \right) \right\|^2 \leq \alpha \left\| e_p^{n,i} - e_p^{n,i-1} \right\| \left\| \nabla \cdot \left( \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1} \right) \right\|. \quad (16)$$

By using (7), (16) implies

$$K_{\text{dr}}^* \left\| \nabla \cdot \left( \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1} \right) \right\| \leq \alpha \left\| e_p^{n,i} - e_p^{n,i-1} \right\|. \quad (17)$$

Inserting (17) into (16) yields

$$2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} \right) - \varepsilon \left( \mathbf{e}_u^{n,i-1} \right) \right\|^2 + \lambda \left\| \nabla \cdot \left( \mathbf{e}_u^{n,i} - \mathbf{e}_u^{n,i-1} \right) \right\|^2 \leq \frac{\alpha^2}{K_{\text{dr}}^*} \left\| e_p^{n,i} - e_p^{n,i-1} \right\|^2. \quad (18)$$

By rearranging terms and inserting (18) into (15), we immediately get

$$\begin{aligned} & \left( 1 - \frac{\delta}{2} \right) \left( 2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} \right) \right\|^2 + \lambda \left\| \nabla \cdot \mathbf{e}_u^{n,i} \right\|^2 \right) + \frac{1}{M} \|e_p^{n,i}\|^2 + \tau \kappa \left\| \nabla e_p^{n,i} \right\|^2 + \frac{L}{2} \|e_p^{n,i}\|^2 + \frac{L}{2} \|e_p^{n,i} - e_p^{n,i-1}\|^2 \\ & \leq \frac{L}{2} \|e_p^{n,i-1}\|^2 + \frac{\alpha^2}{2\delta K_{\text{dr}}^*} \|e_p^{n,i} - e_p^{n,i-1}\|^2. \end{aligned}$$

Using that  $L \geq \frac{\alpha^2}{\delta K_{\text{dr}}^*}$  and the Poincaré inequality, we obtain from the above

$$\left( 1 - \frac{\delta}{2} \right) \left( 2\mu \left\| \varepsilon \left( \mathbf{e}_u^{n,i} \right) \right\|^2 + \lambda \left\| \nabla \cdot \mathbf{e}_u^{n,i} \right\|^2 \right) + \left( \frac{1}{M} + \frac{L}{2} + \frac{\tau \kappa}{C_\Omega^2} \right) \|e_p^{n,i}\|^2 \leq \frac{L}{2} \|e_p^{n,i-1}\|^2. \quad (19)$$

The result, (19), already implies that we have convergence of the scheme. In previous works, particularly that of Both et al<sup>5</sup> (where the proof so far is very similar), the conclusion at this point is that  $L = \frac{\alpha^2}{2K_{\text{dr}}^*}$  is the optimal parameter. However, this does not consider the influence of the first term in (19). By Lemma 1, we get that there exists  $\mathbf{v}_h \in V_h$  such that  $e_p^{n,i} = \nabla \cdot \mathbf{v}_h$  in a weak sense and

$$2\mu \|\varepsilon(\mathbf{v}_h)\|^2 + \lambda \|\nabla \cdot \mathbf{v}_h\|^2 \leq \beta \|e_p^{n,i}\|^2. \quad (20)$$

By testing now (12)(i) with this  $\mathbf{v}_h$ , we get

$$\alpha \|e_p^{n,i}\|^2 = 2\mu \left\langle \varepsilon(\mathbf{e}_u^{n,i}), \varepsilon(\mathbf{v}_h) \right\rangle + \lambda \left\langle \nabla \cdot \mathbf{e}_u^{n,i}, \nabla \cdot \mathbf{v}_h \right\rangle. \quad (21)$$

From (20) and (21) and the Cauchy-Schwarz inequality, we immediately obtain

$$\frac{\alpha^2}{\beta} \|e_p^{n,i}\|^2 \leq 2\mu \|\varepsilon(\mathbf{e}_u^{n,i})\|^2 + \lambda \|\nabla \cdot \mathbf{e}_u^{n,i}\|^2, \quad (22)$$

which, together with (19), implies

$$\left( \frac{1}{M} + \frac{L}{2} + \frac{\tau\kappa}{C_\Omega^2} + \left(1 - \frac{\delta}{2}\right) \frac{\alpha^2}{\beta} \right) \|e_p^{n,i}\|^2 \leq \frac{L}{2} \|e_p^{n,i-1}\|^2.$$

This gives the following rate of convergence, for  $\delta \in (0, 2]$  and  $L \geq \frac{\alpha^2}{\delta K_{\text{dr}}^*}$ :

$$\text{rate}(L, \delta) = \frac{L}{L + \frac{2}{M} + \frac{2\tau\kappa}{C_\Omega^2} + (2 - \delta) \frac{\alpha^2}{\beta}}.$$

□

*Remark 2.* Assumptions 1 and 2 are valid in various relevant physical situations. Therefore, our analysis has a wide range of applications. One can easily extend the result to heterogeneous media, ie,  $\kappa = \kappa(\mathbf{x})$  as long as  $\kappa$  is bounded from below by  $\kappa_m \geq 0$ . Moreover, any of the other parameters can be chosen spatially dependent as long as they are bounded from below by appropriate constants satisfying Assumption 1.

### 3.1 | Optimality

Consider the rate obtained in (9). As  $\text{rate}(L, \delta)$  is an increasing function of  $L$ , it follows that, for all  $\delta \in (0, 2]$ , its minimum is obtained at  $L = \frac{\alpha^2}{\delta K_{\text{dr}}^*}$ , giving the rate

$$\text{rate}(\delta) = \frac{\frac{\alpha^2}{K_{\text{dr}}^*}}{\frac{\alpha^2}{K_{\text{dr}}^*} + \delta \left( \frac{2}{M} + \frac{2\tau\kappa}{C_\Omega^2} + (2 - \delta) \frac{\alpha^2}{\beta} \right)}. \quad (23)$$

Minimizing (23) with respect to  $\delta$  corresponds to maximizing

$$\delta \left( \frac{2}{M} + \frac{2\tau\kappa}{C_\Omega^2} + (2 - \delta) \frac{\alpha^2}{\beta} \right).$$

Let  $A := \frac{2}{M} + \frac{2\tau\kappa}{C_\Omega^2} + 2\frac{\alpha^2}{\beta}$  and  $B := \frac{\alpha^2}{\beta}$ . It is easily seen that the maximum of  $\delta(A - \delta B)$  is attained at  $\delta = \frac{A}{2B}$ . Therefore, the minimizer of  $\text{rate}(\delta)$  is

$$\delta = \min \left\{ \frac{A}{2B}, 2 \right\} \in (1, 2], \quad (24)$$



since  $A \geq 2B$ . This suggests that the theoretical optimal choice of  $L$  is

$$L = \frac{\alpha^2}{K_{\text{dr}}^* \min \left\{ \frac{A}{2B}, 2 \right\}} \in \left[ \frac{\alpha^2}{2K_{\text{dr}}^*}, \frac{\alpha^2}{K_{\text{dr}}^*} \right) \subset \left[ \frac{\alpha^2}{4\mu + 2\lambda}, \frac{\alpha^2}{\frac{2\mu}{d} + \lambda} \right). \quad (25)$$

*Remark 3* (Consequence for low-compressible fluids and low-permeable porous media).

Previous convergence results in the literature for the fixed-stress splitting scheme have not predicted or guaranteed any robust convergence in the limit cases  $M \rightarrow \infty$  and  $\kappa \rightarrow 0$  (for a fixed time-step size  $\tau$ ). Now, by Theorem 1, for inf-sup-stable discretizations, robust convergence of the fixed-stress splitting scheme is guaranteed, even in the limit case. This was studied numerically in the work of Storvik et al.<sup>31</sup> Convergence was showed to be robust with respect to material parameters for P2-P1 elements and deteriorating for P1-P1.

### 3.2 | Brute-force optimization of the stabilization parameter

The rate obtained in Theorem 1 is not necessarily sharp, and it is rather viewed as theoretical evidence that the optimal stabilization parameter resides in the interval  $[\frac{\alpha^2}{4\mu+2\lambda}, \frac{\alpha^2}{\frac{2\mu}{d}+\lambda})$ . Additionally, convergence is predicted to be robust with respect to the mesh size. It can be, indeed, verified numerically that the performance of the fixed-stress splitting scheme is nearly mesh independent (see, for instance, the numerical examples in Section 4 or in the work of Adler et al<sup>46</sup>). Based on that, we propose the following brute-force search for optimizing the stabilization parameter for a fixed test case: test the fixed-stress splitting scheme using different stabilization parameters in the interval  $[\frac{\alpha^2}{4\mu+2\lambda}, \frac{\alpha^2}{\frac{2\mu}{d}+\lambda})$  for a coarse mesh and a single time step. Choose the parameter that gives the fewest number of iterations, and employ it for any arbitrary mesh. Section 4 shows the effectiveness of the proposed method.

## 4 | NUMERICAL EXAMPLES

In this section, we demonstrate the effectiveness of the proposed brute-force method for optimizing the stabilization parameter for the fixed-stress splitting scheme. In particular, we show for several numerical test cases that the optimal stabilization parameter is close to being mesh independent and that the method for choosing it optimally, as described in Section 3.2, indeed yields a preferable alternative to the classical choices of  $L = \frac{\alpha^2}{2K_{\text{dr}}}$  and  $L = \frac{\alpha^2}{K_{\text{dr}}}$ .

We consider four different test cases, as follows:

1. a unit square domain;
2. an L-shaped domain;
3. Mandel's problem;
4. three-dimensional (3D) footing problem on the unit cube.

For the implementation of the numerical examples, we use modules from the DUNE project,<sup>47</sup> particularly dune-functions.<sup>48,49</sup> If not mentioned otherwise, the inf-sup-stable Taylor-Hood pair P2-P1 is utilized as spatial discretization. As stopping criteria, we have applied relative  $L_2$ -norms for the pressure, ie, iterations stop when  $\|p_h^i - p_h^{i-1}\| \leq \epsilon_r \|p_h^{i-1}\|$ , consistent with Theorem 1. Constant material and fluid parameters are applied and given for each individual test case.

### 4.1 | Notations

During the numerical experiments, we apply some specific choices of stabilization parameters several times. Therefore, we give them names here. Recall the definition of the physical drained bulk modulus  $K_{\text{dr}} = \frac{2\mu}{d} + \lambda$ . The original stabilization parameter will be called the physical one due to the fixed-stress splitting scheme's physical origin, ie,  $L_{\text{phys}} = \frac{\alpha^2}{K_{\text{dr}}}$ . The other classical choice of stabilization parameter will be named after Mikelić and Wheeler due to their paper,<sup>24</sup> ie,  $L_{\text{MW}} = \frac{L_{\text{phys}}}{2} = \frac{\alpha^2}{2K_{\text{dr}}}$ . The stabilization parameter obtained by the brute-force method described in Section 3.2 will be called  $L_{\text{opt}}$ . The final parameter is the one that is proposed to be the smallest possible choice in Section 3.1, ie,  $L_{\text{min}} = \frac{\alpha^2}{4\mu+2\lambda}$  (see Table 1).



Name	$L_{\text{phys}}$	$L_{\text{MW}}$	$L_{\text{opt}}$	$L_{\text{min}}$
Value	$\frac{\alpha^2}{K_{\text{dr}}}$	$\frac{\alpha^2}{2K_{\text{dr}}}$	Section 3.2	$\frac{\alpha^2}{4\mu+2\lambda}$

TABLE 1 Names of specific stabilization parameters

Name	Symbol	Value	Unit
Shear modulus	$\mu$	$41.667 \cdot 10^9$	Pa
First Lamé parameter	$\lambda$	$27.778 \cdot 10^9$	Pa
Permeability	$\kappa$	$10^{-13}$	$\text{m}^2$
Compressibility	$\frac{1}{M}$	$10^{-11}$	$\text{Pa}^{-1}$
Initial time	$t_0$	0	s
Time-step size	$\tau$	0.1	s
Stop time	$T$	1	s
Biot-Willis coefficient	$\alpha$	1	–
Relative error tolerance	$\epsilon_r$	$10^{-6}$	–
Inverse of mesh size <sup>a</sup>	$1/h$	16, 32, 64, 128, 512	$\text{m}^{-1}$

TABLE 2 Parameters used in Sections 4.2 and 4.5

<sup>a</sup> Mesh sizes are only used in Section 4.2.

## 4.2 | Dependence on boundary conditions—the unit square

We consider two test cases differing solely in the applied boundary conditions. Common for both, the domain is the unit square discretized by structured triangles, and the constant material parameters from Table 2 are considered. Moreover, we employ source terms corresponding to the analytical solution

$$u_1(x, y, t) = u_2(x, y, t) = \frac{1}{p_{\text{ref}}} p(x, y, t) = txy(1-x)(1-y), \quad (x, y) \in (0, 1)^2, \quad t \in (0, 1),$$

of the continuous problem (1)–(2). The pressure,  $p$ , is scaled by  $p_{\text{ref}} = 10^{11}$  Pa in order to balance the magnitude of the mechanical and fluid stresses for the chosen physical parameters. Regarding the different sets of boundary conditions, we consider the following.

- BC1: homogeneous Dirichlet data on the entire boundary for displacement and pressure.
- BC2: homogeneous Dirichlet data for the pressure; homogeneous Neumann data on top in the mechanics equation and homogeneous Dirichlet data everywhere else for the displacement.

Solutions after 10 time steps using a mesh size of  $h = 1/128$  are displayed in Figures 1 and 2.

To motivate the brute-force approach from Section 3.2, the performance of the fixed-stress splitting scheme has been measured for a variety of stabilization parameters and mesh sizes (see Figure 3). We observe that the numbers of iterations vary significantly for different stabilization parameters but that the optimal choice is within our proposed interval  $[L_{\text{min}}, L_{\text{phys}}]$ . Additionally, for fixed stabilization parameters, we observe that the numbers of iterations are close to constant with respect to the mesh size.

Now, we test the brute-force approach of Section 3.2. In order to calculate  $L_{\text{opt}}$ , we start by applying the fixed-stress splitting scheme for 11 equidistant stabilization parameters in  $[L_{\text{min}}, L_{\text{phys}}]$  while only computing one time step for a mesh

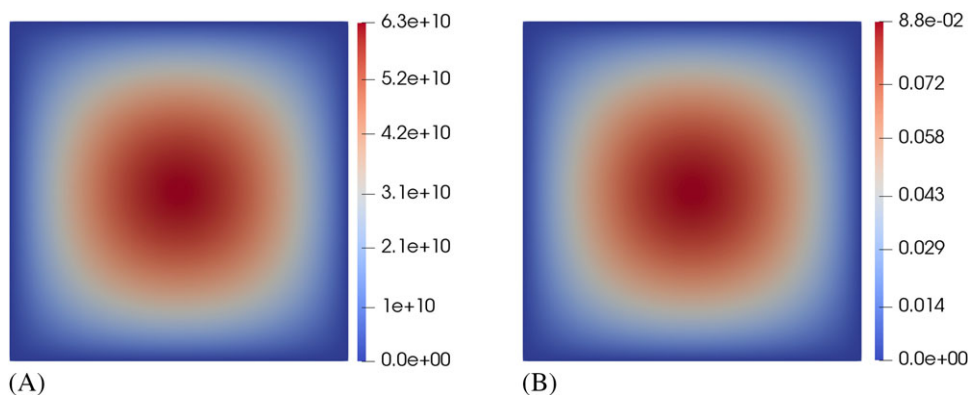
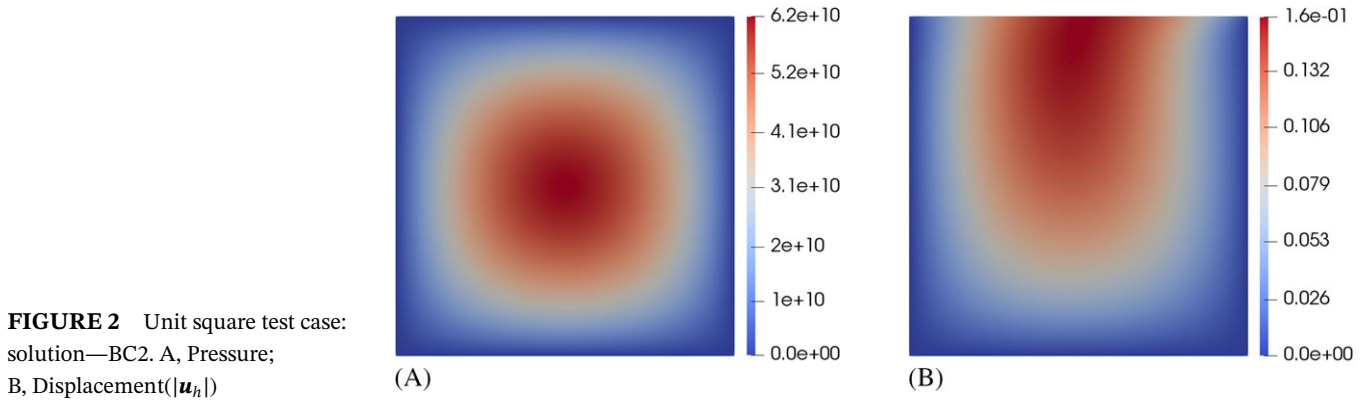
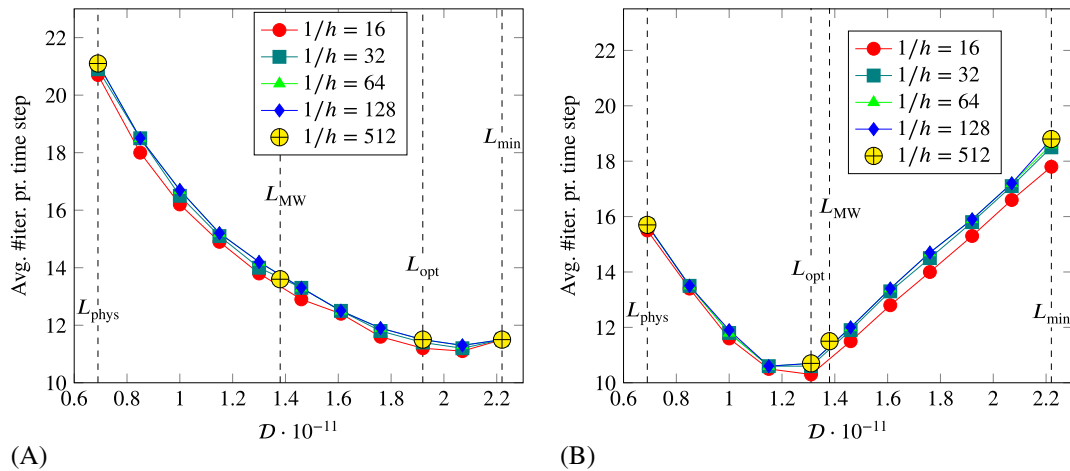


FIGURE 1 Unit square test case: solution—BC1. A, Pressure; B, Displacement( $|\mathbf{u}_h|$ )



**FIGURE 2** Unit square test case: solution—BC2. A, Pressure; B, Displacement( $|\mathbf{u}_h|$ )



**FIGURE 3** Unit square test case: average number of iterations per time step for different stabilization parameters,  $L = \frac{\alpha^2}{D}$ , using parameters from Table 2. The largest value of  $D$  corresponds to  $L_{min}$ , whereas the smallest value of  $D$  corresponds to  $L_{phys}$ . Recall that  $L_{opt}$  is calculated using only one time step, and therefore, there is a slight deviation between  $L_{opt}$  and the actual optimal choice. A, BC1; B, BC2

size of  $h = 1/16$ . Then, using the stabilization parameter that needed the least amount of iterations to converge, we apply the fixed-stress splitting scheme for the full problem using a mesh size of  $h = 1/512$ . In Figure 3, the average numbers of iterations over 10 time steps are displayed for this “optimal” stabilization parameter, for the two classical choices  $L_{phys}$  and  $L_{MW}$ , and for the stabilization parameter that we consider to be the smallest possible choice, ie,  $L_{min}$ . We see that the optimized stabilization parameter requires the least amount of iterations for both boundary conditions. It is also worth noticing that the optimal choice differs considerably for the two sets of boundary conditions.

### 4.3 | Dependence on Poisson's ratio—L-shaped domain

To further analyze the proposed brute-force optimization of the stabilization parameter for the fixed-stress splitting scheme, we test it on an L-shaped domain as well. The L-shaped domain is considered as a subdomain of the unit square domain where the top-right quarter square has been removed, ie,  $L = [0, 1]^2 \setminus (0.5, 1]^2$ . The material and implementation parameters from Table 3 are applied, whereas the right-hand side is the same as for the unit square test case. Zero Dirichlet boundary conditions are applied everywhere, but at the top boundary ( $[0, 0.5] \times \{1\}$ ) for the mechanics equation where zero Neumann conditions are considered. A solution to this problem after 10 time steps with  $\nu = 0$  and mesh size  $1/h = 128$  is given in Figure 4.

Given Young's modulus  $E$  and Poisson's ratio  $\nu$ , the corresponding Lamé parameters have been determined by

$$\mu = \frac{E}{2(1 + \nu)} \quad \text{and} \quad \lambda = \frac{E\nu}{(1 + \nu)(1 - 2\nu)}. \quad (26)$$

Name	Symbol	Value	Unit
Young's modulus	$E$	$10^{11}$	Pa
Poisson's ratio	$\nu$	0, 0.2, 0.4	–
Permeability	$\kappa$	$10^{-13}$	$\text{m}^2$
Compressibility	$\frac{1}{M}$	$10^{-11}$	$\text{m}^{-1}$
Initial time	$t_0$	0	s
Time-step size	$\tau$	0.1	s
Stop time	$T$	1	s
Biot-Willis coefficient	$\alpha$	1	–
Relative error tolerance	$\epsilon_r$	$10^{-6}$	–
Inverse of mesh size	$1/h$	16, 32, 64, 128, 512	$\text{m}^{-1}$

TABLE 3 Parameters used in Section 4.3

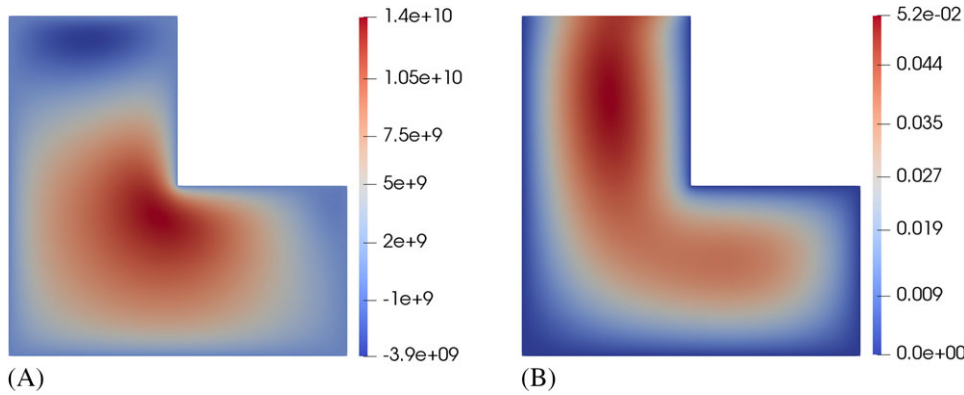


FIGURE 4 L-shaped domain test case: solution for  $\nu = 0$ . A, Pressure; B, Displacement ( $|\mathbf{u}_h|$ )

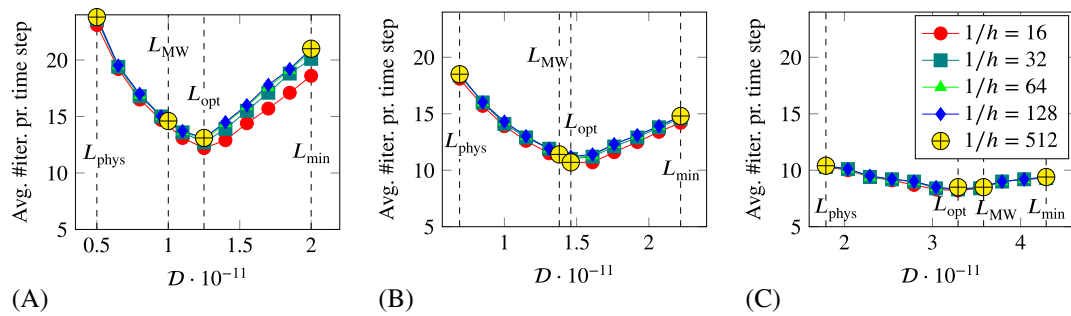


FIGURE 5 L-shaped domain test case: number of iterations for different stabilization parameters,  $L = \frac{l^2}{D}$ , using parameters from Table 3. The largest value of  $D$  corresponds to  $L_{\min}$ , whereas the smallest value of  $D$  corresponds to  $L_{\text{phys}}$ . Notice that the axes are different. A,  $\nu = 0$ ; B,  $\nu = 0.2$ ; C,  $\nu = 0.4$

Again, as for the unit square test case, we test the brute-force optimization technique that is described in Section 3.2, but now for three different Poisson's ratios. In Figure 5, the fixed-stress splitting scheme is applied to a variety of mesh sizes and with a variety of stabilization parameters to three problems with different Poisson's ratios. There are several key observations to make. First, the scheme is close to being mesh independent for all mesh sizes, stabilization parameters, and Poisson's ratios. Second, we see that the optimal stabilization parameter is in the proposed interval  $[L_{\min}, L_{\text{phys}}]$  for all Poisson's ratios and all mesh sizes. The final observation is that when the Poisson's ratio increases, the choice of stabilization parameter becomes less important. This is due to the fact that an increase in the Poisson's ratio can be seen as an effective decrease in the coupling strength.

To calculate the optimal stabilization parameter, we follow the recipe of Section 3.2. We apply 11 equidistant stabilization parameters in the interval  $[L_{\min}, L_{\text{phys}}]$  for the fixed-stress splitting scheme on a coarse mesh ( $1/h = 16$ ) for only one time step. Counting the numbers of iterations it takes to reach convergence, we choose the parameter that corresponds to the smallest number and use this for the finer mesh ( $1/h = 512$ ) and more time steps (10). We see that the parameter that is the optimal choice for the coarse mesh is also the optimal one for the finer mesh for all Poisson's ratios.

#### 4.4 | Mandel's problem

Here, we consider Mandel's problem, a relevant two-dimensional problem with a known analytical solution that is often used as a benchmark problem for discretizations. The analytical solution is derived in the works of Coussy<sup>1</sup> and Aboalsleiman et al.,<sup>50</sup> and its expressions for pressure and displacement are given by

$$p = \frac{2FB(1 + \nu_u)}{3a} \sum_{n=1}^{\infty} \frac{\sin(\alpha_n)}{\alpha_n - \sin(\alpha_n) \cos(\alpha_n)} \left( \cos\left(\frac{\alpha_n x}{a}\right) - \cos(\alpha_n) \right) e^{-\frac{\alpha_n^2 c_f t}{a^2}}, \quad (27)$$

$$u_x = \left[ \frac{F\nu}{2\mu a} - \frac{F\nu_u}{\mu a} \sum_{n=1}^{\infty} \frac{\sin(\alpha_n) \cos(\alpha_n)}{\alpha_n - \sin(\alpha_n) \cos(\alpha_n)} e^{-\frac{\alpha_n^2 c_f t}{a^2}} \right] x + \frac{F}{\mu} \sum_{n=1}^{\infty} \frac{\cos(\alpha_n)}{\alpha_n - \sin(\alpha_n) \cos(\alpha_n)} \sin\left(\frac{\alpha_n x}{a}\right) e^{-\frac{\alpha_n^2 c_f t}{a^2}}, \quad (28)$$

$$u_y = \left[ \frac{-F(1 - \nu)}{2\mu a} + \frac{F(1 - \nu_u)}{\mu a} \sum_{n=1}^{\infty} \frac{\sin(\alpha_n) \cos(\alpha_n)}{\alpha_n - \sin(\alpha_n) \cos(\alpha_n)} e^{-\frac{\alpha_n^2 c_f t}{a^2}} \right] y, \quad (29)$$

where  $\alpha_n, n \in \mathbb{N}$ , correspond to the positive solutions of the equation

$$\tan(\alpha_n) = \frac{1 - \nu}{\nu_u - \nu} \alpha_n,$$

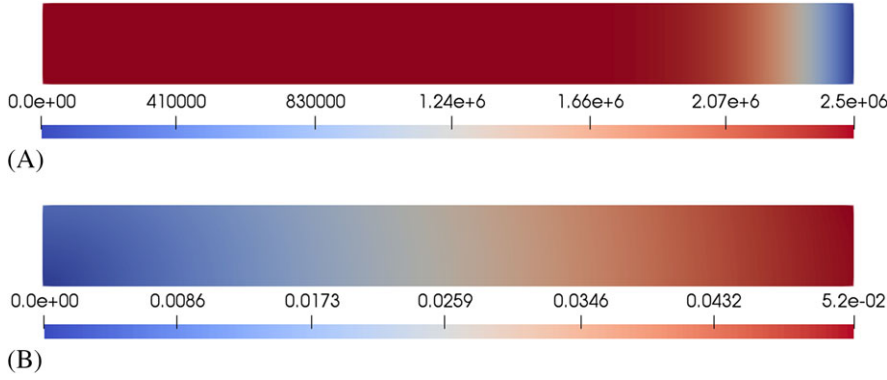
and  $\nu_u, F, B, c_f$ , and  $a$  are input parameters, partially depending on the physical problem parameters. Here, we apply the values listed in Table 4. For a thorough explanation of the problem and the coefficients in (27)-(29), we refer to the works of Coussy<sup>1</sup> and Phillips and Wheeler.<sup>3</sup>

We consider the domain,  $\Omega = (0, 100) \times (0, 10)$ , discretized by a regular triangular mesh. An equidistant partition of the time interval is applied with time-step size  $\tau = 10$  from  $t_0 = 0$  to  $T = 100$ . Initial conditions are inherited from the analytic solutions (27)-(29). As boundary conditions, we apply exact Dirichlet boundary conditions for the normal displacement on the top, left, and bottom boundaries. For pressure, we apply homogeneous boundary conditions on the right boundary. On the remaining boundaries, homogeneous natural boundary conditions are applied. The tolerance  $\epsilon_r$  is set to  $10^{-6}$ . The solution after 10 time steps with 80 vertical and horizontal nodes is displayed in Figure 6.

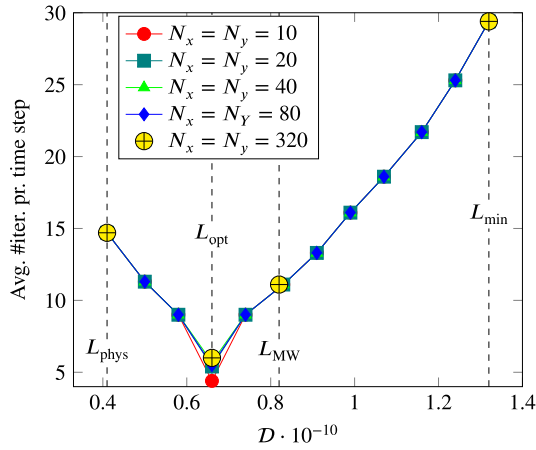
Similar to the unit square and L-shaped domain test cases, we test the mesh independence and the brute-force optimization technique for Mandel's problem. This time, the parameters from Table 4 are applied. In Figure 7, the mesh dependence of the fixed-stress splitting scheme is tested, and it is clear that the performance of the scheme is independent of this choice. At the same time, we confirm that the optimal stabilization parameters actually are in the proposed interval  $[L_{\min}, L_{\text{phys}})$ .

**TABLE 4** Parameters for Mandel's problem

Name	Symbol	Value	Unit
Young's modulus	$E$	$5.94 \cdot 10^9$	Pa
Poisson's ratio	$\nu$	0.2	–
Skempton coefficient	$B$	0.833	–
Undrained Poisson's ratio	$\nu_u$	0.44	–
Applied force	$F$	$6 \cdot 10^8$	N
Biot-Willis constant	$\alpha$	1	–
Compressibility coefficient	$M$	$1.650 \cdot 10^{10}$	Pa
Fluid diffusivity constant	$c_f$	0.47	m <sup>2</sup> /s
Permeability	$\kappa$	$10^{-10}$	m <sup>2</sup>
Width of domain	$a$	100	m
Height of domain	$b$	10	m
Horizontal number of nodes	$N_x$	10, 20, 40, 80, 320	–
Vertical number of nodes	$N_y$	10, 20, 40, 80, 320	–
Time-step size	$\tau$	10	s
Initial time	$t_0$	0	s
Final time	$T$	100	s
Relative error tolerance	$\epsilon_r$	$10^{-9}$	–



**FIGURE 6** Mandel's problem: solution after 10 time steps with  $N_x = N_y = 80$ . A, Pressure; B, Displacement( $|\mathbf{u}_h|$ )



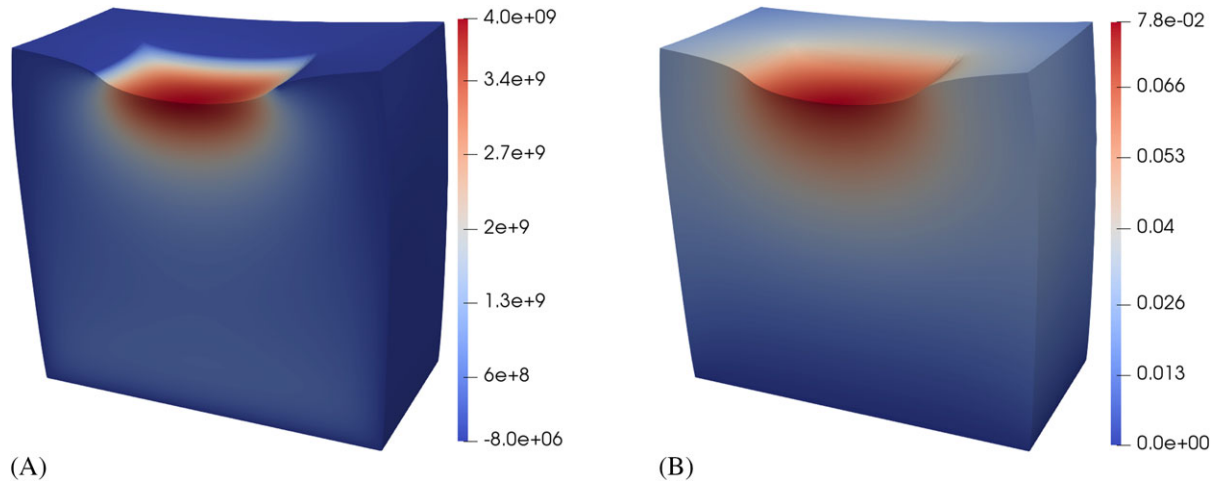
**FIGURE 7** Mandel's problem: number of iterations for different stabilization parameters,  $L = \frac{\alpha^2}{D}$ , using parameters from Table 4. The largest value of  $D$  corresponds to  $L_{\min}$ , whereas the smallest value of  $D$  corresponds to  $L_{\text{phys}}$

To calculate the optimal stabilization parameter, we have applied the optimization technique of Section 3.2. First, the fixed-stress splitting scheme is applied for one time step using a coarse mesh with 10 horizontal and 10 vertical nodes for 11 different stabilization parameters in the interval  $[L_{\min}, L_{\text{phys}}]$ . Choosing the parameter that yields the lowest number of iterations, we apply the scheme for finer meshes and count the number of iterations. As for the other test cases, we see that the optimal parameter indeed is optimal. Moreover, a poor choice of stabilization parameter can result in a huge number of iterations.

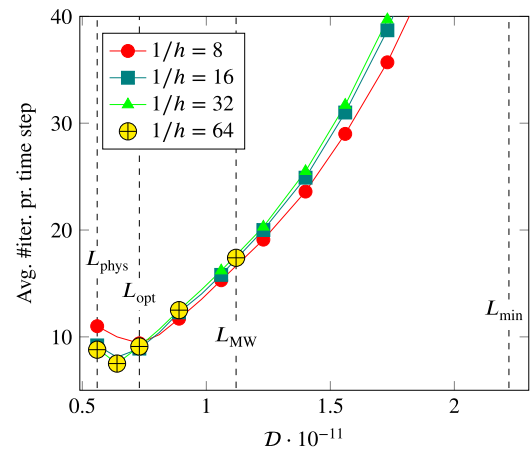
#### 4.5 | 3D footing problem

The numerical section is concluded with a three-dimensional example, ie, a footing problem similar to a test case studied in the work of Adler et al.<sup>46</sup> We consider a unit cube subject to normal compression, ramped in time  $\sigma_n(t) = t \cdot 10^{10} \text{ N} \cdot \text{m}^2/\text{s}$ , applied to a part of the top boundary  $\Gamma_N := [0.25, 0.75] \times [0.25, 0.75] \times \{1\}$ . The bottom is fixed in all directions, and the remaining boundary is considered to be stress free. A no-flow boundary condition is applied at the compression zone  $\Gamma_N$ , and zero pressure is enforced on the remaining boundary. Furthermore, zero body forces are applied. The medium is considered isotropic with the same material parameters as used in Section 4.2 (cf Table 2). For the numerical discretization, we consider a set of four meshes with mesh size  $h \in \{1/8, 1/16, 1/32, 1/64\}$  and employ the inf-sup-stable MINI element.<sup>51</sup> The simulation result for the final time step is visualized in Figure 8.

Due to high computational cost, optimizing the stabilization parameter of the fixed-stress splitting becomes tedious for fine meshes in 3D. Motivated by the previous results, the optimal stabilization parameter is assumed to be nearly mesh independent. This allows for a brute-force search for the optimal, practical stabilization parameter utilizing the coarsest grid (cf Section 3.2). For validation of the optimization strategy, the performance of the splitting scheme is measured in the range  $[L_{\min}, L_{\text{phys}}]$  suggested by Theorem 1; for the finest mesh, we restrict the validation only to a neighborhood of the optimized stabilization parameter. The performance measured in terms of the number of iterations is presented in Figure 9. A large contrast in the performance can be observed for different stabilization parameters, emphasizing the need for a suitable stabilization parameter. Finally, as before, we observe that, indeed, the optimal, practical stabilization parameter is only slightly mesh dependent; it is close to the physical bulk modulus  $K_{\text{dr}} = \frac{2\mu}{d} + \lambda$ . All in all, the brute-force



**FIGURE 8** Three-dimensional footing problem: solution with a deformed configuration magnified by a factor of 2 at the final time  $T = 1$ . Notice that the figure only displays half of the domain but that the other half is symmetric. A, Pressure; B, Displacement( $|\mathbf{u}_h|$ )



**FIGURE 9** Three-dimensional footing problem: average number of iterations per time step for different stabilization parameters,  $L = \frac{\alpha^2}{D}$ , using parameters from Table 2. The largest value of  $D$  corresponds to  $L_{\min}$ , whereas the smallest value of  $D$  corresponds to  $L_{\text{phys}}$

search strategy from Section 3.2 has, again, been confirmed to be a suitable method to obtain a satisfactory stabilization parameter for finer meshes.

## 5 | CONCLUSIONS

In this work, we have considered the quasi-static, linear Biot model for poroelasticity and studied theoretically and numerically the convergence of the fixed-stress splitting scheme. An improved convergence result has been proved, indicating the nontrivial dependence of the optimal stabilization parameters on not only mechanical properties but also fluid flow properties and discretization properties. We observe numerically that the fixed-stress splitting scheme is close to being mesh independent and determine a novel domain in which the optimal stabilization/tuning parameter is found, ie,  $[\frac{\alpha^2}{4\mu+2\lambda}, \frac{\alpha^2}{\frac{2\mu}{d}+\lambda}]$ . On the basis of these observations, we propose a brute-force method with low cost for choosing the optimal stabilization parameter, ie, the parameter that corresponds to the smallest amount of fixed-stress iterations. Through numerical experiments, we have showed that this optimization method results in a much faster fixed-stress splitting scheme than those obtained by choosing the classical stabilization parameters  $L = \frac{\alpha^2}{K_{\text{dr}}}$  and  $L = \frac{\alpha^2}{2K_{\text{dr}}}$ .

## ORCID



## REFERENCES

1. Coussy O. *Poromechanics*. Hoboken, NJ: John Wiley & Sons; 2004.
2. Nordbotten JM. Stable cell-centered finite volume discretization for Biot equations. *SIAM J Numer Anal*. 2016;54(2):942-968. <https://doi.org/10.1137/15M1014280>
3. Phillips PJ, Wheeler MF. A coupling of mixed and continuous Galerkin finite element methods for poroelasticity I: the continuous in time case. *Computational Geosciences*. 2007;11(2):131. <https://doi.org/10.1007/s10596-007-9045-y>
4. Yi S-Y, Bean ML. Iteratively coupled solution strategies for a four-field mixed finite element method for poroelasticity. *Int J Numer Anal Methods Geomech*. 2016;41(2):159-179. <https://onlinelibrary.wiley.com/doi/abs/10.1002/nag.2538>
5. Both JW, Borregales M, Nordbotten JM, Kumar K, Radu FA. Robust fixed stress splitting for Biot's equations in heterogeneous media. *Appl Math Lett*. 2017;68:101-108. <http://www.sciencedirect.com/science/article/pii/S0893965917300034>
6. Berger L, Bordas R, Kay D, Tavener S. A stabilized finite element method for finite-strain three-field poroelasticity. *Computational Mechanics*. 2017;60(1):51-68. <https://doi.org/10.1007/s00466-017-1381-8>
7. Lee JJ. Robust error analysis of coupled mixed methods for Biot's consolidation model. *J Sci Comput*. 2016;69(2):610-632. <https://doi.org/10.1007/s10915-016-0210-0>
8. Hu X, Rodrigo C, Gaspar FJ, Zikatanov LT. A nonconforming finite element method for the Biot's consolidation model in poroelasticity. *J Comput Appl Math*. 2017;310:143-154. <http://www.sciencedirect.com/science/article/pii/S0377042716302734>
9. Rodrigo C, Gaspar FJ, Hu X, Zikatanov LT. Stability and monotonicity for some discretizations of the Biot's consolidation model. *Comput Methods Appl Mech Eng*. 2016;298:183-204. <http://www.sciencedirect.com/science/article/pii/S0045782515003138>
10. Chaabane N, Rivière B. A splitting-based finite element method for the Biot poroelasticity system. *Comput Math Appl*. 2018;75(7):2328-2337. <http://www.sciencedirect.com/science/article/pii/S0898122117307721>
11. Chaabane N, Rivière B. A sequential discontinuous Galerkin method for the coupling of flow and geomechanics. *J Sci Comput*. 2018;74(1):375-395. <https://doi.org/10.1007/s10915-017-0443-6>
12. Simoni L, Secchi S, Schrefler BA. Numerical difficulties and computational procedures for thermo-hydro-mechanical coupled problems of saturated porous media. *Computational Mechanics*. 2008;43(1):179-189. <https://doi.org/10.1007/s00466-008-0302-2>
13. Dana S, Wheeler MF. Convergence analysis of fixed stress split iterative scheme for anisotropic poroelasticity with tensor Biot parameter. *Computational Geosciences*. 2018;22(5):1219-1230. <https://doi.org/10.1007/s10596-018-9748-2>
14. Castelletto N, Klevtsov S, Hajibeygi H, Tchelepi HA. Multiscale two-stage solver for Biot's poroelasticity equations in subsurface media. *Computational Geosciences*. 2018;23(2):207-224. <https://doi.org/10.1007/s10596-018-9791-z>
15. Castelletto N, Hajibeygi H, Tchelepi HA. Multiscale finite-element method for linear elastic geomechanics. *J Comput Phys*. 2017;331:337-356. <http://www.sciencedirect.com/science/article/pii/S0021999116306362>
16. Bause M, Radu FA, Köcher U. Space-time finite element approximation of the Biot poroelasticity system with iterative coupling. *Comput Methods Appl Mech Eng*. 2017;320:745-768. <http://www.sciencedirect.com/science/article/pii/S0045782516316164>
17. Ern A, Meunier S. A posteriori error analysis of Euler-Galerkin approximations to coupled elliptic-parabolic problems. *ESAIM Math Model Numer Anal*. 2009;43(2):353-375.
18. Rahrah M, Vermolen F. Monte Carlo assessment of the impact of oscillatory and pulsating boundary conditions on the flow through porous media. *Transp Porous Media*. 2018;123(1):125-146. <https://link.springer.com/article/10.1007/s11242-018-1028-z>
19. Haga JB, Osnes H, Langtangen HP. On the causes of pressure oscillations in low-permeable and low-compressible porous media. *Int J Numer Anal Methods Geomech*. 2012;36(12):1507-1522. <https://onlinelibrary.wiley.com/doi/abs/10.1002/nag.1062>
20. Rodrigo C, Hu X, Ohm P, Adler JH, Gaspar FJ, Zikatanov LT. New stabilized discretizations for poroelasticity and the Stokes' equations. *Comput Methods Appl Mech Eng*. 2018;341:467-484. <http://www.sciencedirect.com/science/article/pii/S0045782518303347>
21. Kim J, Tchelepi HA, Juanes R. Stability and convergence of sequential methods for coupled flow and geomechanics: fixed-stress and fixed-strain splits. *Comput Methods Appl Mech Eng*. 2011;200(13-16):1591-1606. <http://www.sciencedirect.com/science/article/pii/S0045782510003786>
22. Kim J, Tchelepi HA, Juanes R. Stability and convergence of sequential methods for coupled flow and geomechanics: drained and undrained splits. *Comput Methods Appl Mech Eng*. 2011;200(23-24):2094-2116. <http://www.sciencedirect.com/science/article/pii/S0045782511000466>
23. Settari A, Mourits FM. A coupled reservoir and geomechanical simulation system. *Soc Petroleum Eng*. 1998;3:219-226.
24. Mikelić A, Wheeler MF. Convergence of iterative coupling for coupled flow and geomechanics. *Computational Geosciences*. 2013;17(3):455-461. <https://doi.org/10.1007/s10596-012-9318-y>
25. Turska E, Schrefler BA. On convergence conditions of partitioned solution procedures for consolidation problems. *Comput Methods Appl Mech Eng*. 1993;106(1-2):51-63. <http://www.sciencedirect.com/science/article/pii/004578259390184Y>
26. Turska E, Wisniewski K, Schrefler BA. Error propagation of staggered solution procedures for transient problems. *Comput Methods Appl Mech Eng*. 1994;114(1-2):177-188. <http://www.sciencedirect.com/science/article/pii/0045782594901686>
27. Both JW, Köcher U. Numerical investigation on the fixed-stress splitting scheme for Biot's equations: Optimality of the tuning parameter. In: *Numerical Mathematics and Advanced Applications ENUMATH 2017*. Cham, Switzerland: Springer; 2019:789-797.
28. Mikelić A, Wang B, Wheeler MF. Numerical convergence study of iterative coupling for coupled flow and geomechanics. *Computational Geosciences*. 2014;18(3-4):325-341. <https://doi.org/10.1007/s10596-013-9393-8>
29. Dana S, Ganis B, Wheeler MF. A multiscale fixed stress split iterative scheme for coupled flow and poromechanics in deep subsurface reservoirs. *J Comput Phys*. 2018;352:1-22. <http://www.sciencedirect.com/science/article/pii/S002199911730709X>



30. Bause M. Iterative coupling of mixed and discontinuous Galerkin methods for poroelasticity. In: *Numerical Mathematics and Advanced Applications ENUMATH 2017*. Cham, Switzerland: Springer; 2019:551-560.
31. Storvik E, Both JW, Kumar K, Nordbotten JM, Radu FA. On the optimization of the fixed-stress splitting for Biot's equations. arXiv preprint arXiv:1811.06242. 2018.
32. Borregales M, Radu FA, Kumar K, Nordbotten JM. Robust iterative schemes for non-linear poromechanics. *Computational Geosciences*. 2018;22(4):1021-1038. <https://doi.org/10.1007/s10596-018-9736-6>
33. Both JW, Kumar K, Nordbotten JM, Radu FA. Anderson accelerated fixed-stress splitting schemes for consolidation of unsaturated porous media. *Comput Math Appl*. 2019;77(6):1479-1502. <http://www.sciencedirect.com/science/article/pii/S0898122118304048>
34. Both JW, Kumar K, Nordbotten JM, Radu FA. Iterative methods for coupled flow and geomechanics in unsaturated porous media. *Poromechanics VI*. 2017:411-418. <https://ascelibrary.org/doi/abs/10.1061/9780784480779.050>
35. Hong Q, Kraus J, Lymbery M, Philo F. Conservative discretizations and parameter-robust preconditioners for Biot and multiple-network flux-based poroelastic models. arXiv preprint arXiv: 1806.00353v2. 2018.
36. Lee JJ, Piersanti E, Mardal K-A, Rognes ME. A mixed finite element method for nearly incompressible multiple-network poroelasticity. *SIAM J Sci Comput*. 2019;41(2):A722-A747. <https://doi.org/10.1137/18M1182395>
37. Kim J. A new numerically stable sequential algorithm for coupled finite-strain elastoplastic geomechanics and flow. *Comput Methods Appl Mech Eng*. 2018;335:538-562.
38. Girault V, Kumar K, Wheeler MF. Convergence of iterative coupling of geomechanics with flow in a fractured poroelastic medium. *Computational Geosciences*. 2016;20(5):997-1011. <https://doi.org/10.1007/s10596-016-9573-4>
39. Giovanardi B, Formaggia L, Scotti A, Zunino P. Unfitted FEM for modelling the interaction of multiple fractures in a poroelastic medium. In: Bordas SPA, Burman E, Larson MG, Olshanskii MA, eds. *Geometrically Unfitted Finite Element Methods and Applications*. Cham, Switzerland: Springer International Publishing; 2017:331-352.
40. Lee S, Wheeler MF, Wick T. Iterative coupling of flow, geomechanics and adaptive phase-field fracture including level-set crack width approaches. *J Comput Appl Math*. 2017;314:40-60. <http://www.sciencedirect.com/science/article/pii/S0377042716305118>
41. List F, Radu FA. A study on iterative methods for solving richards' equation. *Computational Geosciences*. 2016;20(2):341-353. <https://doi.org/10.1007/s10596-016-9566-3>
42. Pop IS, Radu FA, Knabner P. Mixed finite elements for the richards' equation: linearization procedure. *J Comput Appl Math*. 2004;168(1):365-373. <http://www.sciencedirect.com/science/article/pii/S037704270301001X>
43. Almani T, Kumar K, Dogru A, Singh G, Wheeler MF. Convergence analysis of multirate fixed-stress split iterative schemes for coupling flow with geomechanics. *Comput Methods Appl Mech Eng*. 2016;311:180-207. <http://www.sciencedirect.com/science/article/pii/S0045782516308180>
44. Borregales M, Kumar K, Radu FA, Rodrigo C, Gaspar FJ. A partially parallel-in-time fixed-stress splitting method for Biot's consolidation model. *Comput Math Appl*. 2018;77(6):1466-1478. <http://www.sciencedirect.com/science/article/pii/S0898122118305091>
45. Boffi D, Brezzi F, Fortin M. *Mixed Finite Element Methods and Applications*. Berlin, Germany: Springer; 2013. *Springer Series in Computational Mathematics*; vol. 44.
46. Adler JH, Gaspar FJ, Hu X, Rodrigo C, Zikatanov LT. Robust block preconditioners for Biot's model. In: *Domain Decomposition Methods in Science and Engineering XXIV*. Cham, Switzerland: Springer; 2017:3-16.
47. Blatt M, Burchardt A, Dedner A, et al. The distributed and unified numerics environment, version 2.4. *Arch Numer Softw*. 2016;4(100):13-29.
48. Engwer C, Gräser C, Müthing S, Sander O. The interface for functions in the dune-functions module. arXiv preprint arXiv:1512.06136. 2015.
49. Engwer C, Gräser C, Müthing S, Sander O. Function space bases in the dune-functions module. arXiv preprint arXiv:1806.09545. 2018.
50. Abousleiman Y, Cheng AH-D, Cui L, Detournay E, Roegiers J-C. Mandel's problem revisited. *Géotechnique*. 1996;46(2):187-195. <https://doi.org/10.1680/geot.1996.46.2.187>
51. Arnold DN, Brezzi F, Fortin M. A stable finite element for the stokes equations. *CALCOLO*. 1984;21(4):337-344. <https://doi.org/10.1007/BF02576171>

**How to cite this article:** Storvik E, Both JW, Kumar K, Nordbotten JM, Radu FA. On the optimization of the fixed-stress splitting for Biot's equations. *Int J Numer Methods Eng*. 2019;120:179-194. <https://doi.org/10.1002/nme.6130>