Advances in Computational Mathematics (2024) 50:111
https://doi.org/10.1007/s10444-024-10200-0

# Higher-order iterative decoupling for poroelasticity

**Robert Altmann**[1] · **Abdullah Mujahid**[2] · **Benjamin Unger**[2]

Received: 28 November 2023 / Accepted: 30 September 2024 / Published online: 15 November 2024


**Abstract**
For the iterative decoupling of elliptic–parabolic problems such as poroelasticity, we introduce time discretization schemes up to order five based on the backward differentiation formulae. Its analysis combines techniques known from fixed-point iterations with the convergence analysis of the temporal discretization. As the main result, we show that the convergence depends on the interplay between the time step size and the parameters for the contraction of the iterative scheme. Moreover, this connection is quantified explicitly, which allows for balancing the single error components. Several numerical experiments illustrate and validate the theoretical results, including a three-dimensional example from biomechanics.

**Keywords** Poroelasticity · Iterative scheme · Decoupling · Higher-order discretization

**Mathematics Subject Classification (2010)** 65M12, 76S05, 65J10

## 1 Introduction

The poroelasticity equations are a set of coupled partial differential equations (PDEs) describing the flow of fluids in porous materials. Initially, this model was introduced in [1] for the consolidation of soil. The set of coupled PDEs in the quasi-static case

Communicated by: Gianluigi Rozza

✉ Benjamin Unger
benjamin.unger@simtech.uni-stuttgart.de

Robert Altmann
robert.altmann@ovgu.de

Abdullah Mujahid
abdullah.mujahid@simtech.uni-stuttgart.de

[1] Institute of Analysis and Numerics, Otto von Guericke University Magdeburg, Universitätsplatz 2, Magdeburg 39106, Germany

[2] Stuttgart Center for Simulation Science (SC SimTech), University of Stuttgart, Universitätsstr. 32, Stuttgart 70569, Germany

consists of an elliptic equation for the mechanics and a parabolic equation for the fluid flow. The overall system is thus called an *elliptic–parabolic PDE*. Among others, poroelasticity finds application in geomechanics [2–4] and the macroscopic modeling of organs such as the human brain [5, 6] or the human heart [7].

The well-posedness of such an elliptic–parabolic PDE is studied in [8]. With an implicit Euler discretization in time, the spatial adaptivity is considered in [9]. Spatial discretizations of this coupled problem result in block matrices, which have symmetric positive definite blocks on the diagonal. With fully coupled methods, the computational burden is tremendous in three-dimensional applications, especially when employing higher-order discretization methods. Hence, decoupling methods attempt to separate the flow and the mechanics equations—either iteratively or non-iteratively—such that efficient preconditioners for the subsystems can be used [10].

For *non-iterative methods*, first proposed and analyzed in [11], the convergence of the decoupled time iteration is guaranteed whenever a certain *weak coupling condition* is satisfied. Such conditions are described by restrictions on the parameters of the PDE that quantify the strength of the coupling. This framework also allows for a generalization to higher-order methods (in time) [12]. To extend the stability regions, i.e., to relax the weak coupling condition, a procedure with an a priori known number of additional inner iterations for each time step is introduced in [13, 14]. Extensions for corresponding nonlinear problems are considered in [15, 16].

*Iterative methods* decouple the system by either first solving the mechanics equation and then the flow equation or vice versa with additional stabilization terms in a fixed-point iteration framework. The operator splitting is motivated by physics and is known as *drained* and *undrained splits* [17] as well as *fixed stress* and *fixed strain splits* [18, 19]. The corresponding analysis shows the existence of a fixed point either in the fully continuous case or the fully discrete case. In the fully continuous case, the iterative solution is proven to converge to the solution of the fully coupled problem, hence accounting for the errors due to operator splitting. As an example, the fixed-point analysis for the fully continuous case, resulting in the popular fixed-stress and undrained split methods, is shown in [20]. In the fully discrete case, which considers either a first-order in time discretization [21] or the generalized midpoint rule [17, 18], it is shown that the iterative solution reaches the fixed point which is the solution when applying the same discretization to the fully coupled problem. Hence, one accounts again for the errors due to splitting alone of the discrete operators [21].

In this article, we focus on the interaction between time integration and operator splitting. For this, we consider the fixed-stress operator splitting to decouple the given elliptic–parabolic problem and employ higher-order time discretizations, using the *backward differentiation formulae* of $k$-th order (denoted as BDF-$k$) for $k \leq 5$. The choice of BDF methods over Runge–Kutta methods is more natural. This is due to the fact that higher-order Runge–Kutta schemes include the tensor product of the Butcher tableau with the decoupled block matrices resulting in larger linear systems. Moreover, symmetry and positive definiteness of the decoupled block matrices is destroyed, and transform-based methods would be needed to prove full convergence [22].

The results of this article are twofold. First, we present a rigorous convergence analysis, which consists of convergence in time for the higher-order time discretization combined with the fixed-point analysis of the operator splitting. In particular, this

accounts for the interaction between the two error components. As a direct consequence of the presented analysis, we propose strategies for optimizing the required number of inner iterations to reach a given tolerance based on the information on the order of the time discretization and the step size. Quantitatively, the relative tolerance TOL for the stopping criteria is proposed as $\mathrm{TOL} \approx \tau^{k+3/2}$ for the BDF-$k$ method where $\tau$ is the time step size. Second, the operator splitting is revisited with an academic toy example to study the optimal choice of the stabilization term.

The organization of this article is as follows. In Section 2, the problem setting is introduced and we recall stability results of BDF methods applied to PDEs. The BDF-$k$ time discretization of the iteratively decoupled problem is given in Section 3, and the main result, which is the convergence of this higher-order iterative solution to the true solution, is presented in Theorem 3.2. Moreover, we analyze and quantify the interplay between the time step size and the iterative parameters that control the iterative contraction. Finally, the expected convergence rates are demonstrated numerically in Section 4.1. Furthermore, we study the balancing of the iterative parameters and the time step size in Section 4.2 and present a real-world example from biomechanics in Section 4.4.

### Notation

Throughout the paper, we write $a \lesssim b$ to indicate that there exists a generic constant $C > 0$, independent of spatial and temporal discretization parameters, such that $a \leq Cb$. Further, the space of symmetric positive definite matrices of size $k$ is denoted by $\mathbb{S}^k_>$. The space of $k$-times continuously differentiable functions on the time interval $[0, T]$ with values in a Banach space $\mathcal{X}$ is denoted as $\mathcal{C}^k([0, T], \mathcal{X})$.

## 2 Preliminaries and problem formulation

Let $\Omega \subseteq \mathbb{R}^m$, $m \in \{2, 3\}$ be a bounded Lipschitz domain and $[0, T]$ the time interval where $0 < T < \infty$. We consider the quasi-static Biot poroelasticity model introduced in [1]. Therein, the task is to find the deformation $u\colon [0, T] \times \Omega \to \mathbb{R}^m$ and the pressure variable $p\colon [0, T] \times \Omega \to \mathbb{R}$, satisfying the system of coupled PDEs,

$$-\nabla \cdot \sigma(u) + \alpha \nabla p = \hat{f} \qquad \text{in } (0, T] \times \Omega, \tag{1a}$$

$$\partial_t \left(\alpha \nabla \cdot u + \tfrac{1}{M} p\right) - \nabla \cdot (\kappa \nabla p) = \hat{g} \qquad \text{in } (0, T] \times \Omega. \tag{1b}$$

Here, $\sigma$ denotes the linear stress tensor

$$\sigma(u) = \mu \left(\nabla u + (\nabla u)^{\mathsf{T}}\right) + \lambda (\nabla \cdot u)\, \mathrm{id}$$

with Lamé coefficients $\lambda$ and $\mu$. The quantity $\kappa$ denotes the permeability, $\alpha$ denotes the Biot–Willis fluid–solid coupling coefficient, and $M$ the Biot moduli. All these quantities are positive constants. Moreover, system (1) is equipped with appropriate initial data and boundary conditions. This model is based on a quasi-static assumption where

the inertial term is neglected in (1a) and, hence, results in a set of coupled elliptic–parabolic PDEs. The model can also be characterized as *partial differential–algebraic equation* (PDAE), since a spatial discretization yields a differential–algebraic equation (cf. [23]).

***Example 2.1*** (Multiple network case) A generalization of (1) is the so-called *multiple network poroelasticity theory* [24], which finds application in the modelling of the human brain [5, 25]. Instead of a single pressure variable, a set of $J \in \mathbb{N}$ pressure variables $p_i : [0, T] \times \Omega \to \mathbb{R}$, $i = 1, \dots, J$ are sought, satisfying

$$-\nabla \cdot \sigma(u) + \sum_{i=1}^{J} \alpha_i \nabla p_i = \hat{f} \qquad \text{in } (0, T] \times \Omega,$$

$$\partial_t \left( \alpha_i \nabla \cdot u + \tfrac{1}{M_i} p_i \right) - \nabla \cdot \left( \kappa_i \nabla p_i \right) + \sum_{j \neq i}^{J} \beta_{ij} (p_i - p_j) = \hat{g}_i \qquad \text{in } (0, T] \times \Omega.$$

Here, the parameters $\alpha_i$, $M_i$, and $\kappa_i$ denote the quantities for the corresponding fluid network and $\beta_{ij} > 0$ represent so-called exchange rates between the fluid networks. In practice, the system is closed with appropriate initial data and boundary conditions.

The single fluid network (1) is a special case of the multiple network case from Example 2.1 for $J = 1$. As illustrated in [23], however, the multiple network model has the same structure as the single network case. Hence, we restrict the following analysis to the single fluid network case and only mention relevant remarks about the multiple network case in the results. As a first step, we introduce an abstract formulation of the model as an elliptic–parabolic problem.

### 2.1 Abstract formulation as elliptic–parabolic problem

Consider the Hilbert spaces

$$\mathcal{V} := [H_0^1(\Omega)]^m, \qquad \mathcal{Q} := H_0^1(\Omega), \qquad \mathcal{H}_\mathcal{V} := [L^2(\Omega)]^m, \qquad \mathcal{H}_\mathcal{Q} := L^2(\Omega),$$

resulting in the Gelfand triples $\mathcal{V} \hookrightarrow \mathcal{H}_\mathcal{V} \simeq \mathcal{H}_\mathcal{V}^* \hookrightarrow \mathcal{V}^*$ and $\mathcal{Q} \hookrightarrow \mathcal{H}_\mathcal{Q} \simeq \mathcal{H}_\mathcal{Q}^* \hookrightarrow \mathcal{Q}^*$ (cf. [26, Sect. 23.4]). Moreover, we introduce the bilinear forms $a : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$, $b : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$, and $c : \mathcal{H}_\mathcal{Q} \times \mathcal{H}_\mathcal{Q} \to \mathbb{R}$, which we assume to be symmetric, continuous, and elliptic in their respective spaces. The corresponding ellipticity and continuity constants of the bilinear forms $\mathfrak{a} \in \{a, b, c\}$ are denoted by $c_\mathfrak{a}$, $C_\mathfrak{a} > 0$, respectively. Furthermore, let $d : \mathcal{V} \times \mathcal{H}_\mathcal{Q} \to \mathbb{R}$ be a bounded bilinear form, i.e., we assume that there exists a positive constant $C_d > 0$ such that $d(u, p) \leq C_d \|u\|_\mathcal{V} \|p\|_{\mathcal{H}_\mathcal{Q}}$ for all $u \in \mathcal{V}$, $p \in \mathcal{H}_\mathcal{Q}$. Based on these quantities, we define the parameter

$$\omega := \frac{C_d^2}{c_a c_c},$$

which quantifies the strength of the coupling between the elliptic and the parabolic equations.

Given sufficiently smooth source terms $f\colon [0,T] \to \mathcal{V}^*$ and $g\colon [0,T] \to \mathcal{Q}^*$, we now seek abstract functions $u\colon (0,T] \to \mathcal{V}$ and $p\colon (0,T] \to \mathcal{Q}$ such that for a.e. $t \in [0,T]$, it holds that

$$a(u,v) - d(v,p) = \langle f, v\rangle, \tag{2a}$$
$$d(\dot{u},q) + c(\dot{p},q) + b(p,q) = \langle g, q\rangle \tag{2b}$$

for all test functions $v \in \mathcal{V}$, $q \in \mathcal{Q}$. Furthermore, we assume *consistent* initial data

$$u(0) = u^0 \in \mathcal{V}, \qquad p(0) = p^0 \in \mathcal{H}_{\mathcal{Q}}, \tag{2c}$$

in the sense that we assume $a(u^0,v) - d(v,p^0) = \langle f(0), v\rangle$ for all $v \in \mathcal{V}$. Without the coupling term represented by the bilinear form $d$, (2a) is elliptic in $u$ and (2b) is parabolic in $p$. This explains why system (2) is called an elliptic–parabolic problem.

The bilinear forms for (1) are defined as

$$a(u,v) := \int_\Omega \sigma(u) : \varepsilon(v)\,\mathrm{d}x, \qquad b(p,q) := \int_\Omega \frac{\kappa}{\nu}\,\nabla p \cdot \nabla q\,\mathrm{d}x,$$
$$c(p,q) := \int_\Omega \frac{1}{M}\,p\,q\,\mathrm{d}x, \qquad d(u,q) := \int_\Omega \alpha\,(\nabla\cdot u)\,q\,\mathrm{d}x.$$

These definitions satisfy the assumed properties for the bilinear forms (cf. [9, 11]). The weak formulation (2) can also be written in terms of operators in the respective dual spaces. For this, let $\mathcal{A}\colon \mathcal{V} \to \mathcal{V}^*$, $\mathcal{B}\colon \mathcal{Q} \to \mathcal{Q}^*$, $\mathcal{C}\colon \mathcal{H}_{\mathcal{Q}} \to \mathcal{H}_{\mathcal{Q}}^*$, and $\mathcal{D}\colon \mathcal{V} \to \mathcal{H}_{\mathcal{Q}}^*$ be the operators associated with the bilinear forms $a$, $b$, $c$, and $d$, respectively. This then leads to the equivalent formulation

$$\mathcal{A}u - \mathcal{D}^* p = f \qquad \text{in } \mathcal{V}^*, \tag{3a}$$
$$\mathcal{D}\dot{u} + \mathcal{C}\dot{p} + \mathcal{B}p = g \qquad \text{in } \mathcal{Q}^*. \tag{3b}$$

Using operator matrices, this can also be written as

$$\begin{bmatrix} 0 & 0 \\ \mathcal{D} & \mathcal{C} \end{bmatrix} \begin{bmatrix} \dot{u} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} -\mathcal{A} & \mathcal{D}^* \\ 0 & -\mathcal{B} \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} + \begin{bmatrix} f \\ g \end{bmatrix},$$

which directly reveals the PDAE structure. As a consequence, explicit methods for time integration cannot be used [27, Ch. 5.2].

### 2.2 BDF operators and weighted inner product spaces

Since we consider higher-order time discretizations of (2) using BDF, we introduce the following notation: for a sequence $(y^n)_{n\in\mathbb{Z}}$ with $y^n \in \mathcal{X}$ in a given Hilbert space $\mathcal{X}$ and a step size $\tau$, we define the BDF-$k$ operator

$$\Xi_k(y^n) := \frac{1}{\tau}\left(\sum_{\ell=0}^{k} \xi_\ell\, y^{n-\ell}\right), \tag{4}$$

where the coefficients $\xi_\ell$ are determined from the relation

$$\xi(s) := \sum_{\ell=0}^{k} \xi_\ell s^\ell = \sum_{\ell=1}^{k} \tfrac{1}{\ell} (1-s)^\ell; \tag{5}$$

see for instance [28, Ch. III.3] and Table 1 for some example coefficients. By abuse of notation, we ignore the dependency on $k$ and simply write $\xi_\ell = \xi_{k,\ell}$. BDF methods up to order 6 are known to be stable [28, Ch. III.3], but for the analysis, we only consider BDF methods up to order 5, for which simple multipliers can be constructed to make it $A$–stable; see [29] and the following proposition.

To simplify the forthcoming analysis of the iterative schemes involving discretization via the BDF operator (4), we will use special inner products by means of the following preliminary result, taken as a special case from [29, 30] (see also [31, Ch. V. 6, Ch. V. 8]). For this, we introduce for a matrix $G = [g_{ij}] \in \mathbb{R}^{k\times k}$, $k \geq 1$, an operator $\mathcal{M}\colon \mathcal{X} \to \mathcal{X}$, and a sequence $(y^n)_{n\in\mathbb{Z}}$ with $y^n \in \mathcal{X}$ for $n \in \mathbb{Z}$ the Kronecker notation

$$(G \otimes \mathcal{M}) \begin{bmatrix} y^n \\ \vdots \\ y^{n+1-k} \end{bmatrix} := \begin{bmatrix} \sum_{j=1}^k g_{1j} \mathcal{M} y^{n+1-j} \\ \vdots \\ \sum_{j=1}^k g_{kj} \mathcal{M} y^{n+1-j} \end{bmatrix} \in \mathcal{X}^k.$$

With this, we obtain the following auxiliary result.

**Proposition 2.2** *Let $\mathcal{X}$ be a Hilbert space and $k \in \{1, \dots, 5\}$. Then, there exist a parameter $\eta \in [0, 1)$, a matrix $G = [g_{ij}] \in \mathbb{S}^k_>$, and a vector $\gamma \in \mathbb{R}^{k+1}$ such that for any self-adjoint and monotone operator $\mathcal{M}\colon \mathcal{X} \to \mathcal{X}$, any sequence $(y^n)_{n\in\mathbb{Z}}$ with $y^n \in \mathcal{X}$ for $n \in \mathbb{Z}$, and constant $\tau > 0$, we have*

$$\begin{aligned} &\left\langle \tau \mathcal{M}\, \Xi_k(y^n),\, y^n - \eta y^{n-1} \right\rangle_{\mathcal{X}} - \left\| \mathcal{M}^{1/2} \sum_{\ell=0}^{k} \gamma_{\ell+1} y^{n-\ell} \right\|^2_{\mathcal{X}} \\ &\quad = \left\langle (G \otimes \mathcal{M}) \begin{bmatrix} y^n \\ \vdots \\ y^{n+1-k} \end{bmatrix}, \begin{bmatrix} y^n \\ \vdots \\ y^{n+1-k} \end{bmatrix} \right\rangle_{\mathcal{X}^k} - \left\langle (G \otimes \mathcal{M}) \begin{bmatrix} y^{n-1} \\ \vdots \\ y^{n-k} \end{bmatrix}, \begin{bmatrix} y^{n-1} \\ \vdots \\ y^{n-k} \end{bmatrix} \right\rangle_{\mathcal{X}^k}. \end{aligned} \tag{6}$$

***Remark 2.3*** The first term on the left-hand side of (6) is independent of $\tau$, since $\Xi_k(\cdot)$ scales with $\frac{1}{\tau}$. Hence, the above statement holds true independently of $\tau$.

**Table 1** BDF-$k$ coefficients $\xi_\ell$, $\ell = 0, \dots, k$, for different values of $k$

| $k$ | $\xi_0$ | $\xi_1$ | $\xi_2$ | $\xi_3$ | $\xi_4$ | $\xi_5$ |
|---|---|---|---|---|---|---|
| 1 | 1 | $-1$ | | | | |
| 2 | $\frac{3}{2}$ | $-2$ | $\frac{1}{2}$ | | | |
| 3 | $\frac{11}{6}$ | $-3$ | $\frac{3}{2}$ | $-\frac{1}{3}$ | | |
| 4 | $\frac{25}{12}$ | $-4$ | 3 | $-\frac{4}{3}$ | $\frac{1}{4}$ | |
| 5 | $\frac{137}{60}$ | $-5$ | 5 | $-\frac{10}{3}$ | $\frac{5}{4}$ | $-\frac{1}{5}$ |

***Proof of Proposition 2.2*** For a self-adjoint and monotone operator, the square root $\mathcal{M}^{1/2}$ is well-defined and is itself a self-adjoint operator [32, Ch. 5, §3, Th. 3.35]. With this, we consider the sequence $(z^n)_{n\in\mathbb{Z}}$, given by $z^n := \mathcal{M}^{1/2} y^n \in \mathcal{X}$ for $n \in \mathbb{Z}$. Then, (6) is equivalent to

$$\begin{aligned}\left\langle \tau\, \Xi_k(z^n), z^n - \eta z^{n-1}\right\rangle_{\mathcal{X}} - \Big\| \sum_{\ell=0}^{k} \gamma_{\ell+1} z^{n-\ell} \Big\|_{\mathcal{X}}^2 \\ = \left\langle G \begin{bmatrix} z^n \\ \vdots \\ z^{n+1-k} \end{bmatrix}, \begin{bmatrix} z^n \\ \vdots \\ z^{n+1-k} \end{bmatrix} \right\rangle_{\mathcal{X}^k} - \left\langle G \begin{bmatrix} z^{n-1} \\ \vdots \\ z^{n-k} \end{bmatrix}, \begin{bmatrix} z^{n-1} \\ \vdots \\ z^{n-k} \end{bmatrix} \right\rangle_{\mathcal{X}^k}.\end{aligned}$$

Note that this is an algebraic criterion, which appears in proving the equivalence between $G$- and $A$-stability of the multistep methods. We refer to [31, Ch. V.6, Def. 6.3] for the definition of $G$-stability and to [31, Ch. V.1, Def. 1.1 and Th. 1.3] for the definition of $A$-stability and the corresponding criterion. Together with the assumption that the first inner product on the left-hand side is at most zero, the above criterion leads to $G$-stability for multistep methods. The $A$-stability criterion for the BDF-$k$ methods reads

$$\operatorname{Re} \frac{\xi(\zeta)}{1-\eta\zeta} \geq 0 \quad \text{for} \quad |\zeta| \geq 1$$

for appropriate multipliers $\eta$ and the polynomial $\xi$ from (5). The existence of $\eta \in [0, 1)$ for $k \leq 5$ is shown in [29]. The remaining part of the proof is the same as that in [31, Ch. V.6, Ch. V.8] albeit with an arbitrary inner product space instead of the finite-dimensional Euclidean space. □

***Example 2.4*** For BDF-1 ($\xi_0 = 1, \xi_1 = -1$), which equals the implicit Euler scheme, we immediately see that equality (6) is satisfied for the choice $\eta = 0$, $G = \frac{1}{2}$, and $\gamma = \left[\pm\frac{1}{\sqrt{2}}, \mp\frac{1}{\sqrt{2}}\right]$.

For BDF-2 with coefficients $\xi_0 = \frac{3}{2}, \xi_1 = -2, \xi_2 = \frac{1}{2}$, the choice

$$\eta = 0, \qquad G = \begin{bmatrix} \frac{5}{4} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{4} \end{bmatrix}, \qquad \gamma = \pm \begin{bmatrix} \frac{1}{2} & 1 & \frac{1}{2} \end{bmatrix}$$

leads to equality (6) (cf. [31, Ch. V. 6, Ex. 6.5]).

Under the assumptions of Proposition 2.2, the inner product used in (6) defines a norm, which we denote as

$$\|E\|_{G,\mathcal{M}}^2 := \big\langle (G \otimes \mathcal{M}) E, E \big\rangle_{\mathcal{X}^k}$$

for all $E \in \mathcal{X}^k$.

### 2.3 Discrete Grönwall-type inequality

We use the following perturbed Grönwall-type inequality in the proof of Theorem 3.2 below.

**Lemma 2.5** *Consider sequences of positive numbers* $\{a^n\}$, $\{b^n\}$, *and* $\{c^n\}$ *satisfying*

$$a^n - (1+\tau C)\, a^{n-1} + b^n - b^{n-1} \le c^n$$

*for some positive constants* $\tau, C > 0$. *Then, it holds that*

$$a^n + b^n \le e^{Cn\tau}\Big(a^0 + b^0 + \frac{1}{C\tau} \max_{1\le \ell \le n} c^\ell\Big).$$

***Proof*** The idea of the proof is to transform the perturbed sum into a telescoping sum; see also [33]. For this, we set $\tilde a^n := (1+\tau C)^{-n} a^n$ and $\tilde b^n := (1+\tau C)^{-n} b^n$. Then, the assumed inequality yields the estimate

$$\begin{aligned}
\tilde a^n - \tilde a^{n-1} &= (1+\tau C)^{-n}\big(a^n - (1+\tau C)\, a^{n-1}\big) \\
&\le (1+\tau C)^{-n} c^n - (1+\tau C)^{-n} b^n + (1+\tau C)^{-n} b^{n-1} \\
&\le (1+\tau C)^{-n} c^n - (1+\tau C)^{-n} b^n + (1+\tau C)^{-(n-1)} b^{n-1} \\
&\le (1+\tau C)^{-n} c^n - \tilde b^n + \tilde b^{n-1},
\end{aligned}$$

where we have used $(1+\tau C)^{-n} \le (1+\tau C)^{-n-1}$. Summing up from 1 to $n$, we obtain

$$\tilde a^n + \tilde b^n \le \tilde a^0 + \tilde b^0 + \sum_{\ell=1}^{n} (1+\tau C)^{-\ell} c^\ell \le \tilde a^0 + \tilde b^0 + \sum_{\ell=1}^{n} (1+\tau C)^{-\ell} \max_{1\le \ell \le n} c^\ell,$$

which simplifies to

$$a^n + b^n \le (1+\tau C)^n \Big(a^0 + b^0 + (1+\tau C)^{-n} \sum_{\ell=0}^{n-1} (1+\tau C)^\ell \max_{1\le \ell \le n} c^\ell\Big).$$

With the observations

$$(1+\tau C)^n \le e^{Cn\tau} \quad \text{and} \quad (1+\tau C)^{-n} \sum_{\ell=0}^{n-1} (1+\tau C)^\ell \le (1+\tau C)^{-n} \frac{(1+\tau C)^n - 1}{C\tau} \le \frac{1}{C\tau}$$

the assertion follows. □

## 3 Iterative decoupling and BDF time integration

To iteratively decouple (3), we consider sequences $(u^i)_{i\in\mathbb{N}}$, $(p^i)_{i\in\mathbb{N}}$ together with their derivatives and enforce

$$\mathcal{D}\dot u^i - \mathcal{D}\dot u^{i-1} = L\mathcal{I}(\dot p^i - \dot p^{i-1}) \qquad \text{in } \mathcal{Q}^*, \tag{7}$$

where $\mathcal{I}\colon \mathcal{Q}^* \to \mathcal{Q}^*$ is the identity operator, $L > 0$ a stabilization parameter, and the superscript denotes the (inner) iteration number. Including this equation in (3b), we obtain the decoupled system

$$\mathcal{A}u^i - \mathcal{D}^* p^i = f \qquad \text{in } \mathcal{V}^*, \tag{8a}$$

$$\mathcal{D}\dot{u}^{i-1} + \mathcal{C}\dot{p}^{i} + \mathcal{B}p^{i} + L\mathcal{I}(\dot{p}^{i} - \dot{p}^{i-1}) = g \qquad \text{in } \mathcal{Q}^*. \tag{8b}$$

This system is indeed decoupled, since (8b) can be used to first solve for $p^i$. Afterwards, $u^i$ is solved from (8a). System (8) is well-posed and the iterates converge globally to the true solution (cf. [20, 34]). We now consider a discretization in time and apply the BDF-$k$ scheme to (8), leading to

$$\mathcal{A}u^{n,i} - \mathcal{D}^* p^{n,i} = f^n \qquad \text{in } \mathcal{V}^*, \tag{9a}$$

$$\mathcal{D}\Xi_k(u^{n,i-1}) + \mathcal{C}\Xi_k(p^{n,i}) + \mathcal{B}p^{n,i} + L\mathcal{I}(\Xi_k(p^{n,i}) - \Xi_k(p^{n,i-1})) = g^n \qquad \text{in } \mathcal{Q}^*. \tag{9b}$$

Here, $y^{n,i}$ denotes the approximation of $y(t^n)$ after $i$ inner iterations (for $y \in \{u, p\}$), and the BDF-$k$ operator $\Xi_k(y^{n,i})$ is defined as

$$\Xi_k(y^{n,i}) := \frac{1}{\tau}\left(\xi_0 y^{n,i} + \sum_{\ell=1}^{k} \xi_\ell \, y^{n-\ell, J_{n-\ell}}\right).$$

Therein, $\xi_0, \ldots, \xi_k$ are the BDF coefficients (see Section 2.2) and $J_n$ denotes the (minimal) number of iterations such that the error between two iterates for the solution at time point $t^n$ goes below a prescribed tolerance TOL. More precisely, we set $J_n = i$, when

$$\tfrac{c_a}{2}\,\|u^{n,i} - u^{n,i-1}\|_{\mathcal{V}}^2 + \left(c_c + \tfrac{L}{2}\right)\|p^{n,i} - p^{n,i-1}\|_{\mathcal{H}_\mathcal{Q}}^2 + \tfrac{\tau}{\xi_0} c_b \,\|p^{n,i} - p^{n,i-1}\|_{\mathcal{Q}}^2 \le \mathrm{TOL}^2 \tag{10}$$

and stop the inner iterations. The initial guess for the iterates can be chosen to be the solution at the previous time step.

The choice of the norm on the left-hand side of (10) is motivated from the following lemma.

**Lemma 3.1** *Let the stabilization parameter $L$ satisfy $L \ge \frac{C_a^2 C_d^2}{2c_a^3}$. Then, the discrete iterative problem (9) is contractive and, hence, well-defined.*

***Proof*** Similar to the first-order in time case [21], it can be easily shown that (9) is contractive for all $n \ge k$. To see this, define

$$\Delta_u^{n,i} := u^{n,i} - u^{n,i-1} \quad \text{and} \quad \Delta_p^{n,i} := p^{n,i} - p^{n,i-1}.$$

Considering the difference between two iterates of (9) and multiplying the last of the resulting second subsystem by $\frac{\tau}{\xi_0}$, we get

$$\mathcal{A}\Delta_u^{n,i} - \mathcal{D}^*\Delta_p^{n,i} = 0 \qquad \text{in } \mathcal{V}^*, \tag{11a}$$

$$\mathcal{D}\Delta_u^{n,i-1} + \mathcal{C}\Delta_p^{n,i} + \tfrac{\tau}{\xi_0}\mathcal{B}\Delta_p^{n,i} + L\mathcal{I}(\Delta_p^{n,i} - \Delta_p^{n,i-1}) = 0 \qquad \text{in } \mathcal{Q}^*. \tag{11b}$$

With the choice of test functions $\Delta_u^{n,i}$ in (11a) and $\Delta_p^{n,i}$ in (11b) and adding the resulting two equations, we obtain

$$\begin{aligned}\langle \mathcal{A}\Delta_u^{n,i}, \Delta_u^{n,i}\rangle + \langle \mathcal{C}\Delta_p^{n,i}, \Delta_p^{n,i}\rangle + \tfrac{\tau}{\xi_0}\langle \mathcal{B}\Delta_p^{n,i}, \Delta_p^{n,i}\rangle + L\langle \Delta_p^{n,i} - \Delta_p^{n,i-1}, \Delta_p^{n,i}\rangle \\ = \langle \mathcal{D}^*\Delta_p^{n,i}, \Delta_u^{n,i} - \Delta_u^{n,i-1}\rangle \\ = \langle \mathcal{A}\Delta_u^{n,i}, \Delta_u^{n,i} - \Delta_u^{n,i-1}\rangle.\end{aligned}$$

With the assumed properties of the bilinear forms, it follows together with the weighted Young inequality with an arbitrary positive constant $\delta > 0$ that

$$\begin{aligned}c_a \|\Delta_u^{n,i}\|_{\mathcal{V}}^2 + c_c \|\Delta_p^{n,i}\|_{\mathcal{H}_\wp}^2 + \tfrac{\tau}{\xi_0} c_b \|\Delta_p^{n,i}\|_{\mathcal{Q}}^2 \\ + \tfrac{L}{2}\left(\|\Delta_p^{n,i}\|_{\mathcal{H}_\wp}^2 - \|\Delta_p^{n,i-1}\|_{\mathcal{H}_\wp}^2 + \|\Delta_p^{n,i} - \Delta_p^{n,i-1}\|_{\mathcal{H}_\wp}^2\right) \\ \le C_a \|\Delta_u^{n,i}\|_{\mathcal{V}} \|\Delta_u^{n,i} - \Delta_u^{n,i-1}\|_{\mathcal{V}} \\ \le \frac{\delta}{2}\|\Delta_u^{n,i}\|_{\mathcal{V}}^2 + C_a^2 \frac{1}{2\delta}\|\Delta_u^{n,i} - \Delta_u^{n,i-1}\|_{\mathcal{V}}^2.\end{aligned}$$

From (11a), we get

$$\|\Delta_u^{n,i}\|_{\mathcal{V}}^2 \le \tfrac{C_d^2}{c_a^2}\|\Delta_p^{n,i}\|_{\mathcal{H}_\wp}^2 \qquad \text{and} \qquad \|\Delta_u^{n,i} - \Delta_u^{n,i-1}\|_{\mathcal{V}}^2 \le \tfrac{C_d^2}{c_a^2}\|\Delta_p^{n,i} - \Delta_p^{n,i-1}\|_{\mathcal{H}_\wp}^2.$$

Combining the above estimates yields

$$\begin{aligned}\left(c_a - \tfrac{\delta}{2}\right)\|\Delta_u^{n,i}\|_{\mathcal{V}}^2 + \left(c_c + \tfrac{L}{2}\right)\|\Delta_p^{n,i}\|_{\mathcal{H}_\wp}^2 + \tfrac{\tau}{\xi_0} c_b \|\Delta_p^{n,i}\|_{\mathcal{Q}}^2 \\ + \left(\tfrac{L}{2} - \tfrac{1}{2\delta}\tfrac{C_a^2 C_d^2}{c_a^2}\right)\|\Delta_p^{n,i} - \Delta_p^{n,i-1}\|_{\mathcal{H}_\wp}^2 \le \tfrac{L}{2}\|\Delta_p^{n,i-1}\|_{\mathcal{H}_\wp}^2. \quad (12)\end{aligned}$$

We need $\delta < 2c_a$ and $L > \frac{C_a^2 C_d^2}{2c_a^3}$ to ensure positivity of the terms on the left-hand side. Choosing $\delta = c_a$ and adding positive terms on the right-hand side gives

$$\begin{aligned}\left(\tfrac{c_a}{2}\|\Delta_u^{n,i}\|_{\mathcal{V}}^2 + \left(c_c + \tfrac{L}{2}\right)\|\Delta_p^{n,i}\|_{\mathcal{H}_\wp}^2 + \tfrac{\tau}{\xi_0} c_b \|\Delta_p^{n,i}\|_{\mathcal{Q}}^2\right) + \tfrac{1}{2}\left(L - \tfrac{C_a^2 C_d^2}{c_a^3}\right)\|\Delta_p^{n,i} - \Delta_p^{n,i-1}\|_{\mathcal{H}_\wp}^2 \\ \le \frac{\frac{L}{2}}{c_c + \frac{L}{2}}\left(\tfrac{c_a}{2}\|\Delta_u^{n,i-1}\|_{\mathcal{V}}^2 + \left(c_c + \tfrac{L}{2}\right)\|\Delta_p^{n,i-1}\|_{\mathcal{H}_\wp}^2 + \tfrac{\tau}{\xi_0} c_b \|\Delta_p^{n,i-1}\|_{\mathcal{Q}}^2\right). \quad (13)\end{aligned}$$

Choosing the stabilization parameter $L = \frac{C_a^2 C_d^2}{c_a^3}$, we obtain

$$\mathfrak{d}\big((u^{n,i}, p^{n,i}), (u^{n,i-1}, p^{n,i-1})\big) \le \gamma\, \mathfrak{d}\big((u^{n,i-1}, p^{n,i-1}), (u^{n,i-2}, p^{n,i-2})\big)$$

where

$$\gamma^2 = \frac{\frac{L}{2}}{c_c + \frac{L}{2}} = \frac{C_a^2 C_d^2}{C_a^2 C_d^2 + 2c_a^3 c_c} < 1$$

and

$$\mathfrak{d}\big((u^{n,i}, p^{n,i}), (u^{n,i-1}, p^{n,i-1})\big) := \Big(\tfrac{c_a}{2}\|\Delta_u^{n,i}\|_{\mathcal{V}}^2 + \big(c_c + \tfrac{L}{2}\big)\|\Delta_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}}^2 + \tfrac{\tau}{\xi_0} c_b \|\Delta_p^{n,i}\|_{\mathcal{Q}}^2\Big)^{1/2}.$$

□

Let $\epsilon^{n,1} = \mathfrak{d}((u^{n,1}, p^{n,1}), (u^{n,0}, p^{n,0}))$ denote the norm of the error at time point $t^n$ due to the initial guess for solving (9). From the termination criterion (10) and the contraction, we have

$$\gamma^{J_n - 1}\epsilon^{n,1} \leq \mathrm{TOL}$$

such that the number of needed inner iterations is given by

$$J_n = \left\lceil \frac{\ln \mathrm{TOL} - \ln \epsilon^{n,1}}{\ln \gamma} \right\rceil + 1. \tag{14}$$

Minimizing the computational work goes along with minimizing $J_n$. An optimal choice balances the errors corresponding to the operator splitting and the time discretization. To realize this, we need to know the relevant rates. This is the subject of the following theorem, which is the main result of this article. Therein, we show that the iterates of (9) converge to the fixed point, which is the exact solution of (3).

**Theorem 3.2** *(Convergence result) Let $u \in \mathcal{C}^k([0,T],\mathcal{V})$, $p \in \mathcal{C}^k([0,T],\mathcal{H}_\mathcal{Q})$ be the solution of (2) for sufficiently smooth right-hand sides $f$ and $g$. Consider $k \in \{1,\dots,5\}$ and given initial data $(u^\ell, p^\ell) \in \mathcal{V}\times\mathcal{H}_\mathcal{Q}$ for $\ell = 0,\dots,k-1$. Let $(u^{n,J_n}, p^{n,J_n})$ for $n \geq k$ be the solution of (9) where $J_n$ denotes the number of inner iterations at the time point $t^n$ as defined in (10) for a prescribed tolerance* $\mathrm{TOL} > 0$. *Then, it holds that*

$$\big\|u^{n,J_n} - u(t^n)\big\|_{\mathcal{V}}^2 + \big\|p^{n,J_n} - p(t^n)\big\|_{\mathcal{H}_\mathcal{Q}}^2 \lesssim \frac{\mathrm{TOL}^2}{\tau^3} + \tau^{2k} + \sum_{\ell=0}^{k-1}\|p^\ell - p(t^\ell)\|_{\mathcal{H}_\mathcal{Q}}^2 \tag{15}$$

*with a hidden constant including a factor of the form $e^{Ct_n}$ with some constant $C > 0$.*

***Proof*** We define the errors of the iterates with respect to the true solution and the defects on approximating the true derivative with BDF-$k$, namely

$$\begin{aligned} e_u^{n,i} &:= u^{n,i} - u(t^n), \qquad & \tau d_u^n &:= \xi_0 u(t^n) + \cdots + \xi_k u(t^{n-k}) - \tau\dot{u}(t^n),\\ e_p^{n,i} &:= p^{n,i} - p(t^n), \qquad & \tau d_p^n &:= \xi_0 p(t^n) + \cdots + \xi_k p(t^{n-k}) - \tau\dot{p}(t^n). \end{aligned}$$

An application of the Taylor series shows that the defects satisfy the estimates

$$\|d_u^n\|_{\mathcal{V}} \lesssim \tau^k, \qquad \|d_p^n\|_{\mathcal{H}_\mathcal{Q}} \lesssim \tau^k.$$

Now consider the difference between (9) and (3). Making use of

$$\Xi_k(y^{n,i}) - \Xi_k(y^{n,i-1}) = \tfrac{\xi_0}{\tau}\left(y^{n,i} - y^{n,i-1}\right)$$

for $y \in \{u, p\}$, one can show that the errors and the defects satisfy the system

$$\mathcal{A}e_u^{n,i} - \mathcal{D}^* e_p^{n,i} = 0 \qquad \text{in } \mathcal{V}^*, \tag{16a}$$

$$\mathcal{D}\big(\tau\,\Xi_k(e_u^{n,i})\big) + \xi_0 \mathcal{D}\big(e_u^{n,i-1} - e_u^{n,i}\big) + \mathcal{C}\big(\tau\,\Xi_k(e_p^{n,i})\big) + \tau\mathcal{B}e_p^{n,i} + \xi_0 L\mathcal{I}\big(e_p^{n,i} - e_p^{n,i-1}\big) + \tau\mathcal{D}d_u^n + \tau\mathcal{C}d_p^n = 0 \qquad \text{in } \mathcal{Q}^*. \tag{16b}$$

Testing (16a) with $e_u^{n,i} \in \mathcal{V}$ and using the properties of the operators, it is easy to see that

$$\|e_u^{n,i}\|_{\mathcal{V}} \leq \tfrac{C_d}{c_a} \|e_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}}. \tag{17}$$

Note that (16a) holds for all $i > 0$, hence solving for $e_u^{n,i} \in \mathcal{V}$ and $e_u^{n,i-1} \in \mathcal{V}$ and inserting this into (16b) gives

$$(\mathcal{C} + \mathcal{D}\mathcal{A}^{-1}\mathcal{D}^*)\tau\,\Xi_k(e_p^{n,i}) + (\xi_0 L\mathcal{I} - \xi_0 \mathcal{D}\mathcal{A}^{-1}\mathcal{D}^*)\big(e_p^{n,i} - e_p^{n,i-1}\big) + \tau\mathcal{B}e_p^{n,i} + \tau\mathcal{D}d_u^n + \tau\mathcal{C}d_p^n = 0 \qquad \text{in } \mathcal{Q}^*.$$

Let $\mathcal{M} := \mathcal{C} + \mathcal{D}\mathcal{A}^{-1}\mathcal{D}^*$ and note that $\mathcal{D}\mathcal{A}^{-1}\mathcal{D}^*$ is self-adjoint and monotone. Therefore, together with the Riesz isomorphism in $\mathcal{H}_\mathcal{Q}$, the operator $\mathcal{M}\colon \mathcal{H}_\mathcal{Q} \to \mathcal{H}_\mathcal{Q}$ is self-adjoint and elliptic with ellipticity constant $c_c$. Testing with $e_p^{n,i} - \eta e_p^{n-1,J_{n-1}} \in \mathcal{Q}$, where $\eta$ is the appropriate multiplier from Proposition 2.2 such that the algebraic criterion (6) is satisfied for some $G \in \mathbb{S}^k_>$, we get

$$T_1 + T_2 = T_3 + T_4, \tag{18}$$

where

$$\begin{aligned}
T_1 &:= \tau\,\big\langle \mathcal{M}\,\Xi_k(e_p^{n,i}),\, e_p^{n,i} - \eta e_p^{n-1,J_{n-1}}\big\rangle,\\
T_2 &:= \tau\,\big\langle \mathcal{B}e_p^{n,i},\, e_p^{n,i} - \eta e_p^{n-1,J_{n-1}}\big\rangle,\\
T_3 &:= \big\langle \xi_0(\mathcal{D}\mathcal{A}^{-1}\mathcal{D}^* - L\mathcal{I})\big(e_p^{n,i} - e_p^{n,i-1}\big),\, e_p^{n,i} - \eta e_p^{n-1,J_{n-1}}\big\rangle,\\
T_4 &:= -\tau\,\big\langle \mathcal{D}d_u^n + \mathcal{C}d_p^n,\, e_p^{n,i} - \eta e_p^{n-1,J_{n-1}}\big\rangle.
\end{aligned}$$

Define the vector of error terms

$$E_p^{n,i} := \begin{bmatrix} e_p^{n,i} \\ e_p^{n-1,J_{n-1}} \\ \vdots \\ e_p^{n+1-k,J_{n+1-k}} \end{bmatrix} \in \mathcal{H}_\mathcal{Q}^k,$$

where the vector takes the same superscript as that of the first error. Using Proposition 2.2, we have

$$T_1 \geq \|E_p^{n,i}\|_{G,\mathcal{M}}^2 - \|E_p^{n-1,J_{n-1}}\|_{G,\mathcal{M}}^2. \tag{19a}$$

By the ellipticity of $\mathcal{B}$, we have

$$\begin{aligned} 2\,T_2 &= 2\tau \left\langle \mathcal{B} e_p^{n,i},\, e_p^{n,i} - \eta\, e_p^{n-1,J_{n-1}} \right\rangle \\ &= \tau\, \|e_p^{n,i}\|_b^2 - \tau\, \|\eta\, e_p^{n-1,J_{n-1}}\|_b^2 + \tau\, \|e_p^{n,i} - \eta\, e_p^{n-1,J_{n-1}}\|_b^2 \\ &\geq \tau\, \|e_p^{n,i}\|_b^2 - \tau\eta^2\, \|e_p^{n-1,J_{n-1}}\|_b^2. \end{aligned} \tag{19b}$$

For the third term, we apply the continuity of the appearing operator and the weighted Young inequality with a positive constant $\delta > 0$, leading to

$$\begin{aligned} T_3 &\leq \xi_0\big(\tfrac{C_d^2}{c_a} + L\big)\|e_p^{n,i} - e_p^{n,i-1}\|_{\mathcal{H}_\mathcal{Q}} \|e_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}} \\ &\quad + \eta\, \xi_0\big(\tfrac{C_d^2}{c_a} + L\big)\|e_p^{n,i} - e_p^{n,i-1}\|_{\mathcal{H}_\mathcal{Q}} \|e_p^{n-1,J_{n-1}}\|_{\mathcal{H}_\mathcal{Q}} \\ &\leq \tau\, \frac{\delta}{2}\|e_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}}^2 + \frac{1}{2\tau\delta}\xi_0^2\big(\tfrac{C_d^2}{c_a} + L\big)^2 \|e_p^{n,i} - e_p^{n,i-1}\|_{\mathcal{H}_\mathcal{Q}}^2 \\ &\quad + \tau\, \frac{1}{2}\|e_p^{n-1,J_{n-1}}\|_{\mathcal{H}_\mathcal{Q}}^2 + \frac{1}{2\tau}\eta^2\xi_0^2\big(\tfrac{C_d^2}{c_a} + L\big)^2 \|e_p^{n,i} - e_p^{n,i-1}\|_{\mathcal{H}_\mathcal{Q}}^2. \end{aligned} \tag{19c}$$

Finally, a similar estimate for the fourth term yields

$$\begin{aligned} T_4 &\leq \tau\, C_d \|d_u^n\|_{\mathcal{V}} \|e_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}} + \tau\, C_d\, \eta\, \|d_u^n\|_{\mathcal{V}} \|e_p^{n-1,J_{n-1}}\|_{\mathcal{H}_\mathcal{Q}} \\ &\quad + \tau\, C_c \|d_p^n\|_{\mathcal{H}_\mathcal{Q}} \|e_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}} + \tau\, C_c\, \eta\, \|d_p^n\|_{\mathcal{H}_\mathcal{Q}} \|e_p^{n-1,J_{n-1}}\|_{\mathcal{H}_\mathcal{Q}} \\ &\leq \tau\, \frac{\delta}{2}\|e_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}}^2 + \tau\, \frac{1}{2}\|e_p^{n-1,J_{n-1}}\|_{\mathcal{H}_\mathcal{Q}}^2 \\ &\quad + \tau\, \Big(\frac{1}{\delta} + \eta^2\Big)\big(C_d^2\, \|d_u^n\|_{\mathcal{V}}^2 + C_c^2\, \|d_p^n\|_{\mathcal{H}_\mathcal{Q}}^2\big). \end{aligned} \tag{19d}$$

The combination of (19) and (18) gives

$$\begin{aligned} &\|E_p^{n,i}\|_{G,\mathcal{M}}^2 - \|E_p^{n-1,J_{n-1}}\|_{G,\mathcal{M}}^2 + \tau\, \tfrac{1}{2}\, \|e_p^{n,i}\|_b^2 - \tau\, \tfrac{1}{2}\eta^2\, \|e_p^{n-1,J_{n-1}}\|_b^2 \\ &\quad \leq \tau\delta \|e_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}}^2 + \tau \|e_p^{n-1,J_{n-1}}\|_{\mathcal{H}_\mathcal{Q}}^2 + \tau\, \Big(\frac{1}{\delta} + \eta^2\Big)\big(C_d^2\, \|d_u^n\|_{\mathcal{V}}^2 + C_c^2\, \|d_p^n\|_{\mathcal{H}_\mathcal{Q}}^2\big) \\ &\qquad + \frac{1}{2\tau}\Big(\frac{1}{\delta} + \eta^2\Big)\xi_0^2\big(\tfrac{C_d^2}{c_a} + L\big)^2 \|e_p^{n,i} - e_p^{n,i-1}\|_{\mathcal{H}_\mathcal{Q}}^2. \end{aligned}$$

To absorb the first two terms on the right-hand side, observe that

$$\|e_p^{n,i}\|_{\mathcal{H}_\mathcal{Q}}^2 \leq \frac{C_{\mathcal{Q}\hookrightarrow\mathcal{H}_\mathcal{Q}}^2}{c_b}\|e_p^{n,i}\|_b^2 \quad \text{and} \quad c_G \|e_p^{n-1,J_{n-1}}\|_{\mathcal{H}_\mathcal{Q}} \leq \|E_p^{n-1,J_{n-1}}\|_{G,\mathcal{M}}$$

where $c_G$ is the smallest eigenvalue of $G$. For the last term, we use

$$\|e_p^{n,i} - e_p^{n,i-1}\|_{\mathcal{H}_\mathcal{Q}}^2 \leq C_{\mathcal{Q}\hookrightarrow\mathcal{H}_\mathcal{Q}} \|e_p^{n,i} - e_p^{n,i-1}\|_{\mathcal{Q}}^2$$

and (10), i.e, with the consideration of $i = J_n$ inner iteration steps, we obtain

$$\begin{aligned}\|E_p^{n,J_n}\|_{G,\mathcal{M}}^2 - \big(1 + \tfrac{1}{c_G}\tau\big)\|E_p^{n-1,J_{n-1}}\|_{G,\mathcal{M}}^2 + \tau\left(\tfrac{1}{2} - \delta\tfrac{C_{\mathcal{Q}\hookrightarrow\mathcal{H}_\mathcal{Q}}^2}{c_b}\right)\|e_p^{n,J_n}\|_b^2 - \tfrac{1}{2}\tau\eta^2\,\|e_p^{n-1,J_{n-1}}\|_b^2 \\ \leq \frac{1}{2}\Big(\frac{1}{\delta} + \eta^2\Big)\xi_0^2\big(\tfrac{C_d^2}{c_a} + L\big)^2\frac{\mathrm{TOL}^2}{\tau^2} + \Big(\frac{1}{\delta} + \eta^2\Big)\big(C_d^2 + C_c^2\big)\tau^{2k+1}.\end{aligned}$$

Now, we set

$$\delta = \frac{1}{2}\frac{c_b}{C_{\mathcal{Q}\hookrightarrow\mathcal{H}_\mathcal{Q}}^2}\big(1 - \eta^2\big)$$

in order to make the last two terms on the left-hand side amenable for telescoping sum. Using the discrete Grönwall inequality from Lemma 2.5 yields the estimate

$$\|E_p^{n,J_n}\|_{G,\mathcal{M}}^2 + \tau\,\|e_p^{n,J_n}\|_b^2 \lesssim e^{Ct_n}\Big(\frac{\mathrm{TOL}^2}{\tau^3} + \tau^{2k} + \|p^{k-1} - p(t^{k-1})\|_b^2 + \sum_{\ell=0}^{k-1}\|p^\ell - p(t^\ell)\|_{\mathcal{H}_\mathcal{Q}}^2\Big).$$

With this and (17), assertion (15) follows. □

The theorem states that the overall error is bounded by the initial error, a term of order $\tau^k$, and a term of order $\mathrm{TOL}/\tau^{3/2}$. Assuming sufficiently good initial data, the optimal choice for the tolerance is hence $\mathrm{TOL} = C\,\tau^{k+3/2}$ for some positive constant $C$. We will investigate this result in more detail numerically in Section 4.2.

***Remark 3.3*** In the derivation of the above error estimate, if we drop the index for the iteration and hence not use the contraction condition, we recover the derivation of error estimates for the fully implicit time discretization. In more detail, let $u^n$, $p^n$ denote the solution of fully implicit time discretization. Then,

$$\big\|u^n - u(t^n)\big\|_{\mathcal{V}}^2 + \big\|p^n - p(t^n)\big\|_{\mathcal{H}_\mathcal{Q}}^2 \lesssim \tau^{2k} + \sum_{\ell=0}^{k-1}\|p^\ell - p(t^\ell)\|_{\mathcal{H}_\mathcal{Q}}^2$$

with the same hidden constants as in Theorem 3.2 only depending on the BDF method and the operators of the original problem (3).

***Remark 3.4*** The above proof does not require the operator $\mathcal{B}$ to be self-adjoint. In particular, the above result holds true also for the multiple network case considered in Example 2.1 where the operator $\mathcal{B}$ is modified to include the network exchange terms with additional assumptions to make it elliptic (cf. [23, Lem. 2.3]).

## 4 Numerical experiments

In this section, we gather a number of numerical experiments illustrating the convergence results of the paper derived in Sections 4.1 and 4.2. Moreover, we do a

parametric study of the number of inner iteration steps in Section 4.3 and apply the proposed decoupling strategy to a three-dimensional brain model in Section 4.4. We use the FEniCSx [35] finite element library for the experiments and use direct solvers (standard LU decomposition from the PETSc library) for all the experiments.

### 4.1 Convergence studies

In this first experiment, we consider a manufactured two-dimensional problem similar to the one in [9] but with homogeneous Dirichlet boundary conditions. On the unit square $\Omega = (0, 1)^2$, we set

$$u(t, x, y) = -10\mathrm{e}^{-\frac{1}{5}t} \sin(\pi x) \sin(\pi y) \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad p(t, x, y) = 10\mathrm{e}^{-\frac{1}{5}t} \sin(\pi x) \sin(\pi y).$$

The poroelasticity parameters are chosen as in Table 2.

To demonstrate the convergence of the (decoupled) BDF-$k$ methods for $1 \leq k \leq 5$, we need to ensure that the errors due to the spatial discretization are very small. Therefore, we choose continuous $(P_6, P_5)$ finite elements for $(u, p)$ with mesh size $h = 2^{-6}$. The manufactured solution is also used to set the needed initial data for all the multistep methods.

The tolerance TOL is chosen sufficiently small, i.e., $10^{-3-3k}$ for each $1 \leq k \leq 5$, respectively. The maximum of the errors over a finite number of time points in the interval [0, 10] is plotted in Fig. 1 and clearly indicates the expected rates.

As a benchmark, the errors are compared with respect to fully coupled methods, i.e., implicit BDF-$k$ schemes applied to the fully coupled problem (see Remark 3.3).

### 4.2 Error balancing

As already discussed in the end of Section 3, the optimal choice for TOL in order to obtain a balanced overall error is $\mathrm{TOL} = C\,\tau^{k+3/2}$. To verify this choice with $C = 1$, we consider the same manufactured problem as in Section 4.1 but with $T = 1$. Here, however, for the spatial discretization, we pick $(P_{k+2}, P_{k+1})$ finite elements for the pair $(u, p)$ with $h = 2^{-7}$ which is sufficient to keep the spatial error components small. Let

$$\mathrm{error} = \max_n \left( \|u(t^n) - u^n\|_{\mathcal{V}} + \|p(t^n) - p^n\|_{\mathcal{H}_{\mathcal{Q}}} \right)$$

denote the maximum of the errors between a discrete solution: either fully implicit or iterative (with the index of iteration dropped) and the true solution computed on a finite number of time points. In Fig. 2, the value of this *error* is plotted for the choices $\mathrm{TOL} = \tau^s$, where $s \in \{k, k + \frac{1}{2}, k + 1, k + \frac{3}{2}, k + 2\}$. Exemplary for BDF-1, BDF-2, we see that this prediction indeed holds true. When the tolerances are bigger, there is

**Table 2** Physical parameters for the two-dimensional poroelasticity convergence study

| $\lambda$ | $\mu$ | $\frac{\kappa}{\nu}$ | $\frac{1}{M}$ | $\alpha$ |
|---|---|---|---|---|
| 0.5 | 0.125 | 0.05 | 4 | 0.75 |

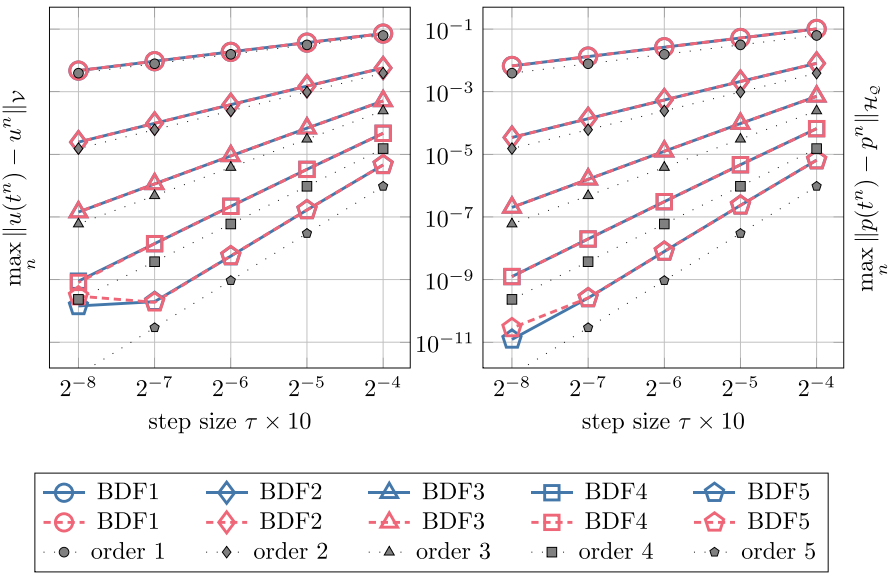



**Fig. 1** Errors for BDF-$k$ for $1 \leq k \leq 5$ for a fixed spatial mesh with mesh size $h = 2^{-6}$. The solid blue curves correspond to the newly introduced decoupled methods whereas the dashed red curves correspond to the fully coupled methods (original BDF schemes). Left: $\mathcal{V}$-error in the displacement $u$. Right: $\mathcal{H}_\mathcal{Q}$-error in the pressure $p$

no convergence. Whenever the TOL is at most $\tau^{k+3/2}$, convergence is observed and the errors are close to the errors which arise for a fully implicit method. For BDF-3, we do not see the sharpness.

In Table 3, the average number of inner iterations $J_n$ per time step is reported. When either the time step size decreases or the order of the method increases, the error that can be reached becomes smaller. This is reflected in a growing number of inner iterations required to reach these potentially smaller errors in the time discretiza-

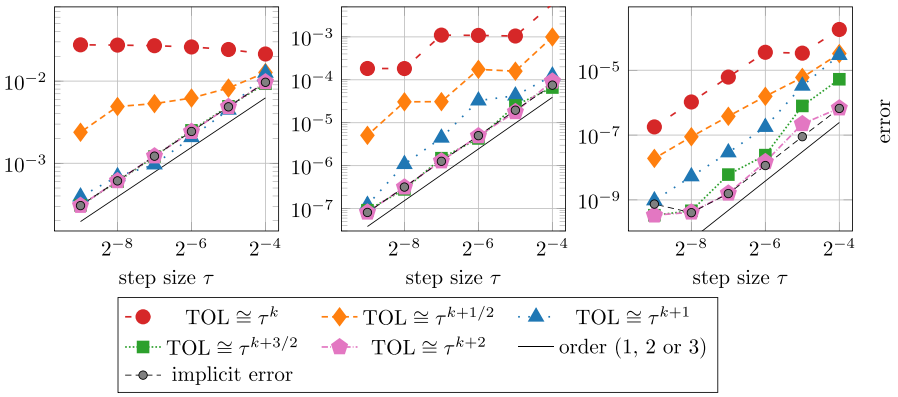



**Fig. 2** Errors for BDF-1 (left), BDF-2 (middle), and BDF-3 (right) for a fixed spatial mesh size $h = 2^{-7}$. The gray line indicates orders 1 (left), 2 (middle), and 3 (right)

**Table 3** Average number of inner iteration steps $J_n$ per time step

| | $\tau$ | $2^{-4}$ | $2^{-5}$ | $2^{-6}$ | $2^{-7}$ | $2^{-8}$ | $2^{-9}$ |
|---|---|---|---|---|---|---|---|
| BDF-1 | TOL $= \tau^{k+1}$ | 5 | 5 | 5 | 6 | 6 | 7 |
| | TOL $= \tau^{k+3/2}$ | 5 | 6 | 7 | 7 | 8 | 8 |
| | TOL $= \tau^{k+2}$ | 6 | 7 | 8 | 8.7 | 9 | 10 |
| BDF-2 | TOL $= \tau^{k+1}$ | 6 | 7 | 8 | 8.83 | 9 | 10 |
| | TOL $= \tau^{k+3/2}$ | 7 | 8 | 9 | 10 | 11 | 12 |
| | TOL $= \tau^{k+2}$ | 8 | 9 | 10 | 11 | 12.1 | 14 |

tion. Furthermore, picking a too small tolerance results in unnecessary iterations. One more parameter that helps in controlling the number of iterations is the stabilization parameter $L$, whose optimality is discussed in [21]. The true optimal choice, however, depends on the elasticity operator, the fluid operators, and the time step size. We investigate this in the next subsection with a toy problem.

### 4.3 Needed iteration steps to reach TOL

To study the sharpness of the estimates for the number of iterations (14) as a function of the stabilization parameter, we consider the scalar toy problem

$$a(u,v) = v^\top A u, \qquad d(v,p) = \sqrt{\omega}\, p D v, \qquad c(p,q) = qCp, \qquad b(p,q) = qBp,$$

where

$$A = \frac{1}{2-\sqrt{2}} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \qquad B = 1, \qquad C = 1, \qquad D = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

As right-hand sides, we set

$$f(t) \equiv \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad g(t) = 100\sin(t).$$

In this case, system (11) simplifies to

$$\Delta_p^{n,i} = \gamma \Delta_p^{n,i-1} \qquad \text{with} \qquad \gamma = \frac{L - \omega D\, A^{-1} D^\top}{L + C + \frac{\tau}{\xi_0} B}. \tag{20}$$

For a stronger contraction, one can prescribe a smaller contraction constant $\gamma$, but then the stabilization parameter has to be chosen accordingly. Rewriting (20), we see that the stabilization parameter should be set as

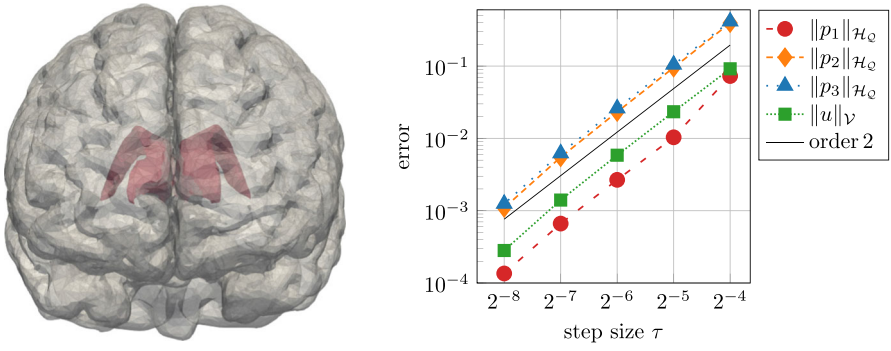
$$L = \frac{\omega}{1-\gamma} D\, A^{-1} D^\top + \frac{\gamma}{1-\gamma}\left(C + \tfrac{\tau}{\xi_0} B\right). \tag{21}$$

**Table 4** Average number of inner iterations per time step needed as a function of the time step size $\tau$ for different methods, coupling strengths $\omega$, and contraction constants $\gamma$. The resulting stabilization parameter $L$ is given by (21)

| | | $\tau$ | $2^{-3}$ | $2^{-4}$ | $2^{-5}$ | $2^{-6}$ | $2^{-7}$ | $2^{-8}$ |
|---|---|---|---|---|---|---|---|---|
| BDF-1 | $\omega = 2$ | $\gamma = 1/2$ | 9 | 13 | 13 | 16 | 19 | 22 |
| | | $\gamma = 1/10$ | 5 | 6 | 7 | 7 | 8 | 9 |
| | $\omega = 4$ | $\gamma = 1/2$ | 9 | 12 | 15 | 17 | 20 | 23 |
| | | $\gamma = 1/10$ | 5 | 6 | 7 | 7 | 8 | 9 |
| BDF-2 | $\omega = 2$ | $\gamma = 1/2$ | 13 | 15 | 18 | 22 | 26 | 30 |
| | | $\gamma = 1/10$ | 6 | 7 | 8 | 9 | 11 | 12 |
| | $\omega = 4$ | $\gamma = 1/2$ | 12 | 16 | 20 | 24 | 28 | 32 |
| | | $\gamma = 1/10$ | 6 | 7 | 8 | 9 | 10 | 12 |

For BDF-1 and BDF-2 and a given coupling strength $\omega$, we study the average number of inner iterations (per time step) needed as a function of the time step size $\tau$ and the contraction constant $\gamma$. This is reported in Table 4. For this toy example, the absolute tolerance TOL for the methods is set based on the average error achieved by a fully implicit BDF method of the same order, which happens to be different for different $\omega$. The initial data for BDF-2 is computed from the implicit BDF-1 method, which is locally accurate up to order two.

We make the following observations. First, we observe less inner iterations per time step if the contraction constant $\gamma$ is decreased, which yields a corresponding stabilization parameter $L$ as in (21). Second, if we fix $\gamma$ but increase the order of the method, also the number of inner iterations increases. Third, if the time step size is decreased, then the number of inner iterations increases. This is due to the choice of TOL. Finally, as the stabilization parameter depends on the coupling strength $\omega$, the number of inner iterations is almost identical for different $\omega$. All observations are as expected from the previous discussion.

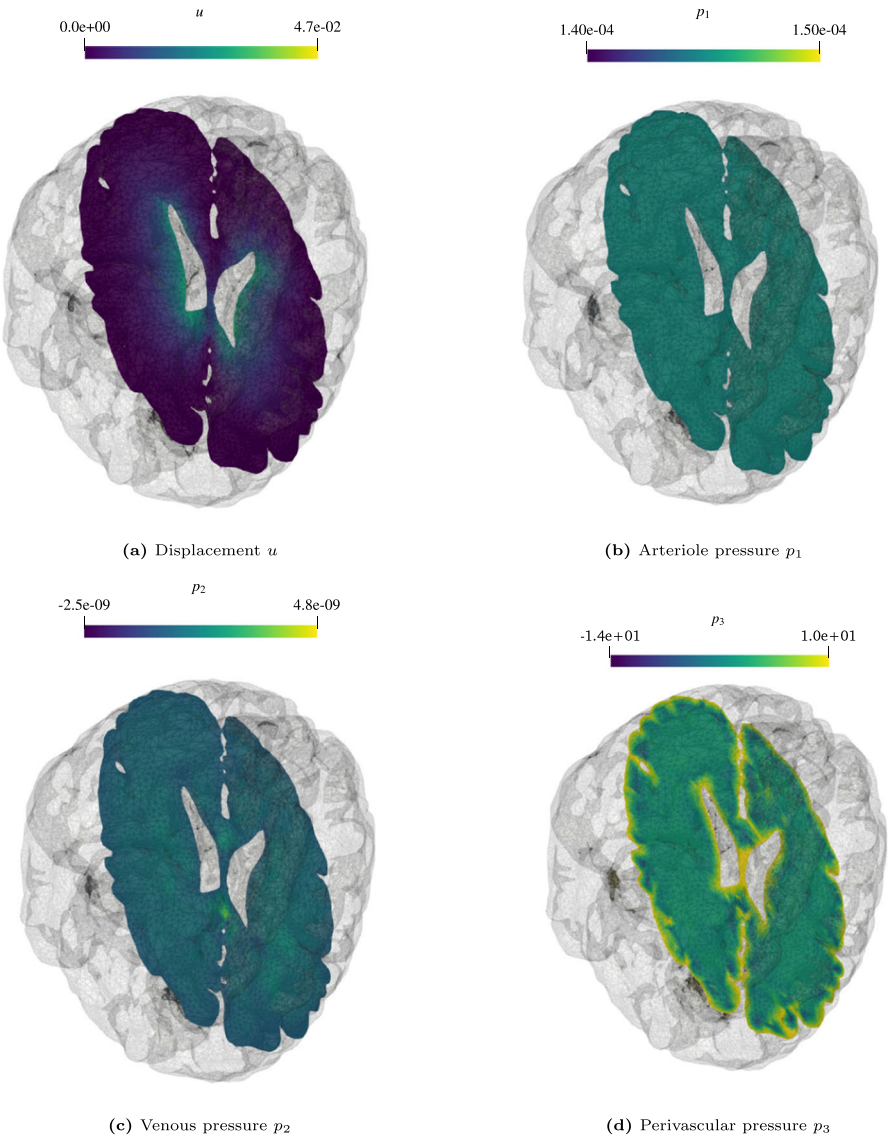



**(a)** Computational mesh **(b)** Convergence history

**Fig. 3** **a** The red part corresponds to ventricular boundary $\partial\Omega_v$, whereas the outer boundary is $\partial\Omega_s$. **b** Convergence history for all four fields

**Table 5** Parameters for the three-dimensional multiple network problem

| | Arteriole network | Venous network | Perivascular network |
|---|---|---|---|
| $\lambda = 9.08\text{e+}3$ | $\kappa_1/\nu_1 = 3.74\text{e}{-8}$ | $\kappa_2/\nu_2 = 3.74\text{e}{-8}$ | $\kappa_3/\nu_3 = 1.43\text{e}{-7}$ |
| $\mu = 5.48\text{e+}2$ | $1/M_1 = 2.90\text{e}{-4}$ | $1/M_2 = 1.50\text{e}{-5}$ | $1/M_3 = 2.90\text{e}{-4}$ |
| $\beta_{12} = 1.0\text{e}{-3}$ | $\alpha_1 = 0.4$ | $\alpha_2 = 0.2$ | $\alpha_3 = 0.4$ |
| $\beta_{13} = 1.0\text{e}{-4}$ | | | |




(a) Displacement $u$

(b) Arteriole pressure $p_1$

(c) Venous pressure $p_2$

(d) Perivascular pressure $p_3$

**Fig. 4** Illustration of the deformation and the three pressure variables at the final time $T = 1$

### 4.4 Biomechanics multiple network problem

To illustrate the tuning of decoupling for practical applications, we pick a network Problem—see Example 2.1—with three fluid pressures from biomechanics (similar to the one in [5]) and apply a BDF-2 method. The three fluid networks are the so-called *arteriole*, *venous*, and *perivascular* networks, and the fluid is allowed to flow from the arteriole to the venous and perivascular networks. In Lemma 3.1, we have $\gamma^2 = \frac{L}{2c_c+L}$ for a single pressure network. For the network case, we set $c_c$ to be the minimum over $c_c$ of each fluid network. Since BDF-2 is a multistep method, we pick BDF-1 (which is second-order accurate locally) to initialize the needed data. For the computational domain, a processed coarse brain mesh is taken from [36, 37], which has 99 605 cells and 29 037 vertices; see Fig. 3a. The boundary is such that $\partial\Omega = \partial\Omega_s \cup \partial\Omega_v$, where $\partial\Omega_s$ is the boundary near the skull and $\partial\Omega_v$ is the inner boundary that encloses the ventricle.

For the source terms, we set $g_1(t, x) = \frac{1}{2}(1 - \cos(2\pi t))$, $g_2 \equiv g_3 \equiv 0$, and $f \equiv 0$. The initial data for all fields are assumed to be zero. Finally, the boundary conditions read

$$\begin{aligned} u &= 0 \quad \text{on } \partial\Omega_s, & (\sigma(u) - \textstyle\sum_{j=1}^{3} \alpha_j p_j) \cdot \mathbf{n} &= 10\mathbf{n} \quad \text{on } \partial\Omega_v, \\ \nabla p_1 \cdot \mathbf{n} &= 0 \quad \text{on } \partial\Omega, & p_2 &= 0 \quad \text{on } \partial\Omega, \\ p_3 &= 10 \quad \text{on } \partial\Omega. \end{aligned}$$

with $\mathbf{n}$ denoting the outer unit normal vector. The material parameters are given in Table 5 in SI units (see [5] for units). We use P1 finite elements for all the variables. Solving using a direct solver over the interval [0, 1] with 64 time steps, the fields at the final time are plotted in Fig. 4. With a reference solution computed using the fully implicit BDF-2 method with $\tau = 2^{-9}$, the convergence history is shown in Fig. 3b.

## 5 Conclusions

In this article, a combined analysis of operator splitting together with higher-order BDF time discretization is studied. For an optimal balancing of the two error components, one needs to specify the tolerance for the termination criterion of the operator splitting in the specified form.

The results are thoroughly demonstrated in a number of numerical experiments and applied to a real-world example from biomechanics.

A similar combined analysis of operator splitting together with Runge–Kutta methods is deferred for a future work.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 467107679. Parts of this work were carried out while RA and AM were affiliated with the Institute of Mathematics, University of Augsburg. RA was also affiliated with the Centre for Advanced Analytics and Predictive Sciences (CAAPS). Moreover, AM and BU acknowledge support by the Stuttgart Center for Simulation Science (SimTech).

## Declarations

**Conflict of interest** The authors declare no Conflict of interest.



## References

1. Biot, M.A.: General theory of three-dimensional consolidation. J. Appl. Phys. **12**(2), 155–164 (1941). https://doi.org/10.1063/1.1712886
2. Zoback, M.D.: Reservoir geomechanics. Cambridge University Press, Cambridge (2010). https://doi.org/10.1017/CBO9780511586477
3. Kosloff, D., Scott, R.F., Scranton, J.: Finite element simulation of Wilmington oil field subsidence: I. Linear modelling. Tectonophysics. **65**(3–4), 339–368 (1980). https://doi.org/10.1016/0040-1951(80)90082-7
4. White, J.A., Borja, R.I.: Stabilized low-order finite elements for coupled solid-deformation/fluid-diffusion and their application to fault zone transients. Comput. Method. Appl. M. **197**(49–50), 4353–4366 (2008). https://doi.org/10.1016/j.cma.2008.05.015
5. Eliseussen, E., Rognes, M.E., Thompson, T.B.: A posteriori error estimation and adaptivity for multiple-network poroelasticity. ESAIM Math. Model. Numer. Anal. **57**(4), 1921–1952 (2023). https://doi.org/10.1051/m2an/2023033
6. Corti, M., Antonietti, P.F., Dede', L., Quarteroni, A.M.: Numerical modelling of the brain poromechanics by high-order discontinuous Galerkin methods. (2022). https://doi.org/10.48550/arXiv.2210.02272
7. Barnafi Wittwer, N.A., Gregorio, S.D., Dede', L., Zunino, P., Vergara, C., Quarteroni, A.: A multiscale poromechanics model integrating myocardial perfusion and the epicardial coronary vessels. SIAM J. Appl. Math. **82**(4), 1167–1193 (2022). https://doi.org/10.1137/21M1424482
8. Showalter, R.E.: Diffusion in poro-elastic media. J. Math. Anal. Appl. **251**(1), 310–340 (2000). https://doi.org/10.1006/jmaa.2000.7048
9. Ern, A., Meunier, S.: A posteriori error analysis of Euler-Galerkin approximations to coupled elliptic-parabolic problems. ESAIM Math. Model. Numer. Anal. **43**(2), 353–375 (2009). https://doi.org/10.1051/m2an:2008048
10. Lee, J.J., Mardal, K.-A., Winther, R.: Parameter-robust discretization and preconditioning of Biot's consolidation model. SIAM J. Sci. Comput. **39**(1), 1–24 (2017). https://doi.org/10.1137/15M1029473
11. Altmann, R., Maier, R., Unger, B.: Semi-explicit discretization schemes for weakly-coupled elliptic-parabolic problems. Math. Comp. **90**, 1089–1118 (2021). https://doi.org/10.1090/mcom/3608
12. Altmann, R., Maier, R., Unger, B.: Semi-explicit integration of second order for weakly coupled poroelasticity. BIT Numer. Math. **64**(20) (2024). https://doi.org/10.1007/s10543-024-01021-0
13. Altmann, R., Deiml, M.: A novel iterative time integration scheme for linear poroelasticity. Electron. Trans. Numer. Anal. **60**, 256–275 (2024). https://doi.org/10.1553/etna_vol60s256
14. Altmann, R., Deiml, M.: A second-order iterative time integration scheme for linear poroelasticity. (2024). arXiv:2403.12699
15. Altmann, R., Maier, R., Unger, B.: A semi-explicit integration scheme for weakly-coupled poroelasticity with nonlinear permeability. Proc. Appl. Math. Mech. **20**, 202000061 (2021). https://doi.org/10.1002/pamm.202000061
16. Altmann, R., Maier, R.: A decoupling and linearizing discretization for poroelasticity with nonlinear permeability. SIAM J. Sci. Comput. **44**(3), 457–478 (2022). https://doi.org/10.1137/21M1413985

17. Kim, J., Tchelepi, H.A., Juanes, R.: Stability and convergence of sequential methods for coupled flow and geomechanics: drained and undrained splits. Comput. Meth. Appl. Mech. Eng. **200**(23), 2094–2116 (2011). https://doi.org/10.1016/j.cma.2011.02.011
18. Kim, J., Tchelepi, H.A., Juanes, R.: Stability and convergence of sequential methods for coupled flow and geomechanics: fixed-stress and fixed-strain splits. Comput. Meth. Appl. Mech. Eng. **200**(13), 1591–1606 (2011). https://doi.org/10.1016/j.cma.2010.12.022
19. Wheeler, M.F., Gai, X.: Iteratively coupled mixed and Galerkin finite element methods for poro-elasticity. Numer. Meth. Part. D. E. **23**(4), 785–797 (2007). https://doi.org/10.1002/num.20258
20. Mikelić, A., Wheeler, M.F.: Convergence of iterative coupling for coupled flow and geomechanics. Comput. Geosci. **17**(3), 455–461 (2013). https://doi.org/10.1007/s10596-012-9318-y
21. Storvik, E., Both, J.W., Kumar, K., Nordbotten, J.M., Radu, F.A.: On the optimization of the fixed-stress splitting for Biot's equations. Internat. J. Numer. Methods Engrg. **120**(2), 179–194 (2019). https://doi.org/10.1002/nme.6130
22. Lubich, C., Ostermann, A.: Runge-Kutta approximation of quasi-linear parabolic equations. Math. Comp. **64**(210), 601–627 (1995). https://doi.org/10.2307/2153442
23. Altmann, R., Mehrmann, V., Unger, B.: Port-Hamiltonian formulations of poroelastic network models. Math. Comput. Model. Dyn. Sys. **27**(1), 429–452 (2021). https://doi.org/10.1080/13873954.2021.1975137
24. Barenblatt, G.I., Zheltov, I.P., Kochina, I.N.: Basic concepts in the theory of seepage of homogeneous liquids in fissured rocks [strata]. J. Appl. Math. Mech. **24**(5), 1286–1303 (1960). https://doi.org/10.1016/0021-8928(60)90107-6
25. Vardakis, J.C., Chou, D., Tully, B.J., Hung, C.C., Lee, T.H., Tsui, P.H., Ventikos, Y.: Investigating cerebral oedema using poroelasticity. Med. Eng. Phys. **38**(1), 48–57 (2016). https://doi.org/10.1016/j.medengphy.2015.09.006
26. Zeidler, E.: Nonlinear functional analysis and its applications. 2A: Linear Monotone Operators. Springer, New York, Berlin, Heidelberg (1990). https://doi.org/10.1007/978-1-4612-0985-0
27. Kunkel, P., Mehrmann, V.: Differential-algebraic equations. Analysis and numerical solution. European Mathematical Society, Zürich (2006). https://doi.org/10.4171/017
28. Hairer, E., Nørsett, S.P., Wanner, G.: Solving Ordinary Differential Equations I: Nonstiff Problems, 2nd rev. edn. Springer, Heidelberg, London (2009). https://doi.org/10.1007/978-3-540-78862-1
29. Nevanlinna, O., Odeh, F.: Multiplier techniques for linear multistep methods. Numer. Funct. Anal. Optim. **3**(4), 377–423 (1981). https://doi.org/10.1080/01630568108816097
30. Dahlquist, G.: G-stability is equivalent to A-stability. BIT Numer. Math. **18**(4), 384–401 (1978). https://doi.org/10.1007/BF01932018
31. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems. Springer, Berlin, Heidelberg (1996). https://doi.org/10.1007/978-3-642-05221-7
32. Kato, T.: Perturbation theory for linear operators. Springer, Berlin, Heidelberg (1995). https://doi.org/10.1007/978-3-642-66282-9
33. Emmrich, E.: Discrete versions of Gronwall's lemma and their application to the numerical analysis of parabolic problems. Preprint 637, Technische Universität Berlin, Germany. (1999). https://doi.org/10.14279/depositonce-14714
34. Mujahid, A.: Monolithic, non-iterative and iterative time discretization methods for linear coupled elliptic-parabolic systems. GAMM Arch. Stud. **4**(1) (2022). https://doi.org/10.14464/gammas.v4i1.500
35. Baratta, I.A., Dean, J.P., Dokken, J.S., Habera, M., Hale, J.S., Richardson, C.N., Rognes, M.E., Scroggs, M.W., Sime, N., Wells, G.N.: DOLFINx: the next generation FEniCS problem solving environment. Zenodo (2023). https://doi.org/10.5281/zenodo.10447666
36. Lee, J.J., Piersanti, E., Mardal, K.-A., Rognes, M.E.: A mixed finite element method for nearly incompressible multiple-network poroelasticity. SIAM J. Sci. Comput. **41**(2), 722–747 (2019). https://doi.org/10.1137/18M1182395
37. Piersanti, E., Rognes, M.E.: Supplementary Material (code) for 'A Mixed Finite Element Method for Nearly Incompressible Multiple-network Poroelasticity' by J. J. Lee, E. Piersanti, K.-A. Mardal and M. E. Rognes. https://doi.org/10.5281/zenodo.1215636