Kumulative Habilitationsschrift

Modelling and Simulation of Fluid-Structure Interactions with Contact and Multiple Scales in Time

Dr. Stefan Frei

Universität Konstanz

Preface

Fluid-structure interaction (FSI) describes the interplay of a flow and a solid mechanical system. Due to the many applications in industry, technology and science, FSI has been a very active field of research in recent decades, both from a mathematical and from an engineering perspective. Mathematically, the combination of fluid and solid mechanical equations poses several challenges, for example due to the different coordinate systems involved, time-dependent subdomains and interfaces and the different characters of the underlying partial differential equations.

More recently, the additional difficulties that arise when the solid comes into contact with another solid or an exterior wall have been the subject of different publications. Among others, additional challenges are given by the resulting topology change in the fluid domain, the Navier-Stokes *no-collision paradoxon* and the nonlinearly changing interface conditions in the transition to contact.

This habilitation thesis consists of nine journal articles [F1]-[F9] that deal with the modelling and simulation of fluid-structure interaction problems and contact [F1]-[F3], the development and analysis of related discretisation techniques in space and time [F4]-[F7] and the efficient numerical simulation for FSI problems with multiple scales in time [F8, F9]. To distinguish them from other references they are referred to as [F1]-[F9] compared to pure numbers [1]-[96]. In the following introduction the reader is guided through these publications and the most relevant results are summarised. For details, proofs and most of the numerical results, we refer to the respective articles that follow in the second part of this thesis.

The results in these papers have been obtained in a joint effort of the respective authors with comparable scientific contributions. The numerical results in [F1, F5] and [F9] have been obtained by the author, those in [F2] partially by the author and partially by F. Gerosa, in [F3] by the author, H. von Wahl and T. Richter and in [F4], [F6], [F7] and [F8] by the collaborators G. Judakova, M.K. Singh, M.W. Scroggs and T. Richter, respectively. In all papers essential contributions in the analysis and the development of the methods, as well as in the redaction of the articles, have been made by the author.

Acknowledgements *I thank my mentors Thomas Richter and Erik Burman for the excellent research environments they provided during my PhD resp. my Postdoc time, including many stimulating and motivating discussions, and for their continuous support, even years after leaving their groups. Moreover, I would like to thank the Department of Mathematics and Statistics in Konstanz, in particular Stefan Volkwein, for an excellent working environment with many scientific possibilities and support in all kind of matters. Last but not least, my biggest gratitude goes to my parents for their unconditional and lifelong support and to my wife Lisseth for all her sacrifice, backing and continuous encouragement over the last years.*

Contents

Publications of this Thesis

[F1] E Burman, MA Fernández, S Frei: *A Nitsche-based formulation for fluid-structure interactions with contact*, ESAIM Mathematical Modelling and Numerical Analysis (M2AN) 54(2), 531-564 (2020)

[F2] S Frei, FM Gerosa, E Burman, MA Fernández: *A mechanically consistent model for fluid-structure interactions with contact including seepage*, Computer Methods in Applied Mechanics and Engineering (CMAME) 392, 114637 (2022)

[F3] H von Wahl, T Richter, S Frei, T Hagemeier: *Falling balls in a viscous fluid with contact: Comparing numerical simulations with experimental data*, Physics of Fluids 33, 033304 (2021)

[F4] S Frei, G Judakova, T Richter: *A locally modified second-order finite element method for interface problems and its implementation in 2 dimensions*, ESAIM Mathematical Modelling and Numerical Analysis (M2AN) 57 (3), 1355-1380 (2023)

[F5] E Burman, S Frei, A Massing: *Eulerian time-stepping schemes for the non-stationary Stokes equations on time-dependent domains*, Numerische Mathematik 150, 423-478 (2022)

[F6] S Frei, MK Singh: *An implicitly extended Crank-Nicolson scheme for the heat equation on time-dependent domains*, submitted journal article (2023), preprint: `https://arxiv.org/abs/2203.06581`

[F7] E Burman, S Frei, MW Scroggs: *Weak imposition of Signorini boundary conditions on the boundary element method*, SIAM Journal on Numerical Analysis 58(4), 2334-2350 (2020)

[F8] S Frei, T Richter: *Efficient approximation of flow problems with multiple scales in time*, SIAM Multiscale Modeling and Simulation 18(2), 942–969 (2020)

[F9] S Frei, A Heinlein: *Towards parallel time-stepping for the numerical simulation of atherosclerotic plaque growth*, submitted journal article (2022), preprint: `http://arxiv.org/abs/2203.06526`

Remark. *Out of the nine journal articles, the articles [F1]-[F5], [F7] and [F8] have already been published. The articles [F6] and [F9] are currently under review. Permissions to print original versions of the articles have been obtained by EDP Sciences ([F1], [F4]), Elsevier [F2], AIP Publishing [F3], Springer Nature [F5] and SIAM ([F7] and [F8]).*

Part I

Introduction

The interaction of a fluid and a solid mechanical system plays a crucial role in many applications in technology, industry and science. Prominent examples range from aerodynamical applications involving for example airplanes, parachutes or vehicles to biomedical problems such as blood flow through the cardiovascular system or airflow within the respiratory system.

A key problem in the modelling of fluid-structure interactions (FSI) are the different coordinate systems that are involved. While solid equations are typically formulated in the *Lagrangian* (material) coordinate framework on a fixed domain $\hat{\Omega}_s$, fluid equations are usually formulated and solved in *Eulerian* coordinates; in the context of transient FSI this means on a time-dependent domain $\Omega_f(t)$.

The most popular approach to combine the coordinate frameworks is the *Arbitrary Lagrangian Eulerian* (ALE) framework [27, 37, 62, 75]. The idea is to map the fluid equations back to an *arbitrary* fixed reference domain $\hat{\Omega}_f$ that matches the solid reference domain $\hat{\Omega}_s$ at its boundary. Therefore, a smooth map $\xi_f : \hat{\Omega}_f \to \Omega_f(t)$ is required (precisely ξ_f needs to be a $C^{1,1}$-diffeomorphism), see e.g. [77]. A sufficiently regular map does, however, not exist in the case of topology changes within the fluid domain, that occur necessarily when the solid comes into contact with another solid body or an exterior wall. The modelling and simulation of such FSI-contact problems is the subject of Section 1 and the articles [F1, F2] and [F3] described therein. Our approach to combine the FSI coupling and the contact conditions is based on a weak imposition of both by means of Nitsche's method [17, 24, 56, 74]. Moreover, we introduce a benchmark configuration for FSI with contact in [F3], including a comparison with experimental data.

Instead of the ALE approach, a so-called *Fully Eulerian* approach [28, 29, 38, 76] is applied in these works. There, all equations, including the solid equations, are formulated in the current *Eulerian* coordinate framework, i.e. on $\Omega_f(t)$ and $\Omega_s(t)$. While this enables an elegant variational formulation with the possibility of topology changes in the fluid domain, discretisation becomes more cumbersome due to the time dependence of the subdomains $\Omega_f(t), \Omega_s(t)$. In fact, an accurate discretisation needs to resolve the position and the movement of the interface both in space and time. This is the subject of the first three articles in Section 2. First, we introduce a second-order *fitted* finite element method for the spatial discretisation of interface problems in Section 2.1 resp. [F4]. Then, we analyse *Eulerian* time-stepping schemes based on an *unfitted* finite element discretisation in Section 2.2 resp. [F5] and [F6]. While in [F4]-[F6] finite elements are considered for spatial discretisation, we show in Section 2.3 resp. [F7], how contact conditions can be incorporated in a Boundary Element Formulation by means of Nitsche's method.

In Section 3 resp. [F8] and [F9], we study temporal multi-scale problems for the numerical simulation of plaque growth in arteries. In this application, plaque growth takes place over a long time scale of months to several years. On the other hand, short-time effects related to the wall shear stress, which depends on the periodic heart beat with a scale of milliseconds to seconds, have an essential effect on the plaque growth, i.e. the long-term dynamics. First, in [F8], a temporal two-scale algorithm is developed. A complete numerical analysis is given for a strongly simplified flow problem with a moving boundary. Then, in Section 3.2 resp. [F9] a temporal parallelisation based on the parareal algorithm is developed for the two-scale algorithm of [F8] and the method is applied to a full FSI-plaque growth problem.

1 Fluid-Structure Interactions with Contact

In the first section, we introduce recently developed numerical models for the simulation of FSI and contact [F1, F2] and describe an experimental benchmark configuration to validate the numerical results [F3].

1.1 Development of a Suitable Numerical Model [F1]

1.1.1 Fluid-Structure Interactions

We assume that the entire domain of interest $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ is fixed and consider a partitioning into time-dependent sub-domains

$$\Omega = \Omega_f(t) \cup \Gamma_{\text{fsi}}(t) \cup \Omega_s(t),$$

where $\Omega_f(t)$ and $\Omega_s(t)$ denote the fluid and solid subdomain, respectively, and $\Gamma_{\text{fsi}}(t)$ is the $(d-1)$-dimensional FSI interface, see Figure 1.1, left sketch. As we will use an Eulerian model for the complete FSI system all equations are formulated in the current domains $\Omega_f(t), \Omega_s(t)$ and $\Gamma_{\text{fsi}}(t)$.

In [F1] a novel numerical framework for the modelling and simulation of FSI and contact is introduced and -as a first step- applied to linear fluid and solid mechanical sub-systems. Precisely, we consider the linear incompressible Stokes equations in the fluid sub-domain $\Omega_f(t)$

$$\partial_t \mathbf{u} - \operatorname{div} \boldsymbol{\sigma}_f(\mathbf{u}, p) = \mathbf{f}_f, \quad \operatorname{div} \mathbf{u} = 0,$$

where the Cauchy stress tensor $\boldsymbol{\sigma}_f$ is defined by

$$\boldsymbol{\sigma}_f(\mathbf{u}, p) = \nu_f(\nabla \mathbf{u}^T + \nabla \mathbf{u}) - pI, \tag{1.1}$$

$\mathbf{u}$ denotes the fluid velocity, p the fluid pressure and $\nu_f > 0$ is a constant viscosity. In the solid-subdomain $\Omega_s(t)$, we assume a linear elastic material

$$\partial_t \dot{\mathbf{d}} - \operatorname{div} \boldsymbol{\sigma}_s(\mathbf{d}) = \mathbf{f}_s, \quad \partial_t \mathbf{d} = \dot{\mathbf{d}}, \tag{1.2}$$

where the Cauchy stress tensor $\boldsymbol{\sigma}_s$ is given by

$$\boldsymbol{\sigma}_s(\mathbf{d}) = 2\mu_s E(\mathbf{d}) + \lambda_s \operatorname{tr}(E(\mathbf{d}))I, \quad E(\mathbf{d}) = \frac{1}{2}\left(\nabla \mathbf{d} + \nabla \mathbf{d}^T\right), \tag{1.3}$$

$\dot{\mathbf{d}} = \partial_t \mathbf{d}$ denotes the solid velocity and $\lambda_s, \mu_s > 0$ are the so-called Lamé constants.

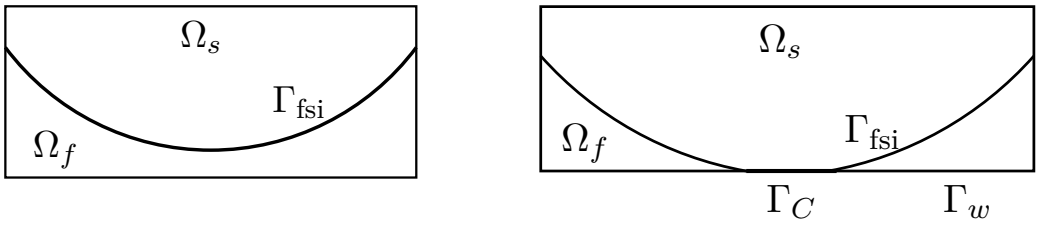


Figure 1.1. Left: Fluid-structure interaction problem without contact. **Right**: Body in contact with a plane wall Γ_w.

For the coupling across the fluid-solid interface $\Gamma_{\text{fsi}}(t)$, continuity of velocities and normal stresses

$$\dot{\mathbf{d}} = \mathbf{u}, \quad \boldsymbol{\sigma}_f(\mathbf{u}, p)\mathbf{n} = \boldsymbol{\sigma}_s(\mathbf{d})\mathbf{n} \quad \text{on} \quad \Gamma_{\text{fsi}}(t) \tag{1.4}$$

is typically assumed for viscous fluids, where $\mathbf{n} = \mathbf{n}_s$ denotes the outer normal vector of the solid domain $\Omega_s(t)$. When it comes to contact, however, these *no-slip* conditions do in general not lead to a reasonable approximation of the underlying physics. Theoretical works show, that the Navier-Stokes equations in combination with no-slip conditions do not allow for contact [59], contrary to what is observed in experiments [53]. This changes when Navier-slip [73] or pure slip-conditions are used [51]. Thus, we also consider slip conditions in [F1], where the continuity across $\Gamma_{\text{fsi}}(t)$ is only imposed for the normal velocity

$$\mathbf{u} \cdot \mathbf{n} = \dot{\mathbf{d}} \cdot \mathbf{n}, \quad (I - \mathbf{n} \otimes \mathbf{n})\boldsymbol{\sigma}_f \mathbf{n} = 0, \quad \boldsymbol{\sigma}_f \mathbf{n} = \boldsymbol{\sigma}_s \mathbf{n} \quad \text{on} \quad \Gamma_{\text{fsi}}(t). \tag{1.5}$$

For suitable initial conditions at $t = 0$ and exterior boundary conditions on $\partial\Omega$, we refer to [F1].

We close this section by formulating the corresponding semi-discrete variational formulation on suitable finite element spaces $\mathcal{V}_h, \mathcal{Q}_h, \mathcal{W}_h$ (see Section 2) using Nitsche's method. We present the formulation exemplarily for the case of slip-conditions on Γ_{fsi}:

Variational Formulation 1. *Find* $\mathbf{u} \in \mathcal{V}_h, p \in \mathcal{Q}_h, \mathbf{d} \in \mathcal{W}_h$*, such that* $\dot{\mathbf{d}} = \partial_t \mathbf{d}$ *and*

$$\mathcal{A}_{\text{fsi, slip}}(\mathbf{u}, p, \mathbf{d}, \dot{\mathbf{d}})(\mathbf{v}, q, \mathbf{w}) = (\mathbf{f}_f, \mathbf{v})_{\Omega_f(t)} + (\mathbf{f}_s, \mathbf{w})_{\Omega_s(t)} \qquad \forall\, \mathbf{v}, q, \mathbf{w} \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h,$$

where

$$\begin{aligned} &\mathcal{A}_{\text{fsi, slip}}(\mathbf{u}, p, \mathbf{d}, \dot{\mathbf{d}})(\mathbf{v}, q, \mathbf{w}) \\ &\quad := (\partial_t \mathbf{u}, \mathbf{v})_{\Omega_f(t)} + (\boldsymbol{\sigma}_f(\mathbf{u}, p), \nabla \mathbf{v})_{\Omega_f(t)} + (\operatorname{div} \mathbf{u}, q)_{\Omega_f(t)} + S(p, q) \\ &\qquad + (\partial_t \mathbf{d}, \mathbf{w})_{\Omega_s(t)} + (\boldsymbol{\sigma}_s(\mathbf{d}), \nabla \mathbf{w})_{\Omega_s(t)} - \Big(T_f(\mathbf{u}, p, \dot{\mathbf{d}}) \cdot \mathbf{n}, (\mathbf{w} - \mathbf{v}) \cdot \mathbf{n}\Big)_{\Gamma(t)} \\ &\qquad\quad - \Big((\dot{\mathbf{d}} - \mathbf{u}) \cdot \mathbf{n}, \mathbf{n}^T \boldsymbol{\sigma}_f(\mathbf{v}, -q)\mathbf{n}\Big)_{\Gamma(t)} \end{aligned} \tag{1.6}$$

and

$$T_f(\mathbf{u}, p, \dot{\mathbf{d}}) := \boldsymbol{\sigma}_f(\mathbf{u}, p) - \gamma_{\text{fsi}}(\dot{\mathbf{d}} - \mathbf{u}) \cdot \mathbf{n} \tag{1.7}$$

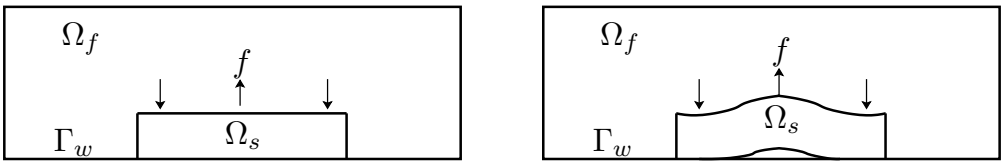


Figure 1.2. Illustrative example to motivate the role of seepage in fluid-structure interaction with contact: Contact of a solid body with the lower wall (*left sketch*) is released in the central part of the contact surface, when a specific force f is applied. Without seepage on Σ_p a vacuum would be created in the cavity (*right sketch*) (Figure taken from [F2] with permission from *Elsevier*).

denotes the discrete fluid fluxes with $\gamma_{\text{fsi}} = \gamma_{\text{fsi}}^0 h^{-1}$ and a parameter $\gamma_{\text{fsi}}^0 > 0$. Moreover, we assume that $S : \mathcal{Q}_h \times \mathcal{Q}_h \to \mathbb{R}$ is a pressure stabilisation term that leads to a well-posed fluid subsystem. We note that Variational Formulation 1 goes back to [17], where an analogous Nitsche formulation was developed for no-slip interface conditions and fixed domains Ω_s and Ω_f.

1.1.2 FSI and Contact

Now we consider the case that the solid can come into contact with an exterior wall Γ_w. We assume that no contact with Γ_w takes place at $t = 0$ and denote by $\Gamma(t) \subset \partial\Omega_s$ the surface which was initially the FSI interface $\Gamma(0) = \Gamma_{\text{fsi}}(0)$ separating fluid and solid. When $\Gamma(t)$ comes into contact with Γ_w in a part $\Gamma_c(t) \subset \Gamma(t)$, there is no fluid layer attached to $\Gamma(t)$ anymore. Thus, the balance of forces $\boldsymbol{\sigma}_f \mathbf{n} = \boldsymbol{\sigma}_s \mathbf{n}$ in (1.4) resp. (1.5) can not hold anymore, as the left-hand side is not even defined.

On the other hand, discarding the fluid forces on the contact surface can also lead to undesired and non-physical effects, as discussed in [F1] and [2]. One issue is that contact could be released in the interior of the contact surface without any resistive forces, which would create non-physical islands of void (if no seepage through the contact surface is considered). An example is given in Figure 1.2.

The solution chosen in [F1, F2] and also by Ager et al. in [2, 3] is to extend the fluid stresses $\boldsymbol{\sigma}_f \mathbf{n}$ onto the contact surface $\Gamma_c(t)$. This is motivated by the fact that from a microscopic perspective a solid body will never be exactly in contact with the wall throughout the contact surface (for example due to surface roughness) and some seepage of fluid through the surface can occur. Different possibilities to extend the fluid forces are described in the following Subsection 1.1.3 and also in Section 1.2 resp. [F2]. By a slight abuse of notation we will denote the extension of $\boldsymbol{\sigma}_f \mathbf{n}$ onto $\Gamma_c(t)$ still by $\boldsymbol{\sigma}_f \mathbf{n}$.

Nevertheless, we can not expect that the balance of forces $\boldsymbol{\sigma}_f \mathbf{n} = \boldsymbol{\sigma}_s \mathbf{n}$ will hold during contact for these extensions in general. Thus, we introduce a Lagrange multiplier that corresponds to a contact force acting normal to Γ_w

$$\lambda := [\![\boldsymbol{\sigma}_n]\!] \leq 0 \quad \text{ on } \Gamma(t), \tag{1.8}$$

where $\boldsymbol{\sigma}_{s,n} = \mathbf{n}_w^T \boldsymbol{\sigma}_s \mathbf{n}$, $\boldsymbol{\sigma}_{f,n} = \mathbf{n}_w^T \boldsymbol{\sigma}_f \mathbf{n}$, $[\![\boldsymbol{\sigma}_n]\!] := \boldsymbol{\sigma}_{s,n} - \boldsymbol{\sigma}_{f,n}$ and $\mathbf{n}_w$ denotes the exterior normal vector of Ω on Γ_w. The resulting contact conditions read

$$\mathbf{d} \cdot \mathbf{n}_w - g_0 \leq 0, \quad \lambda \leq 0, \quad \lambda(\mathbf{d} \cdot \mathbf{n}_w - g_0) = 0 \quad \text{ on } \Gamma(t), \tag{1.9}$$

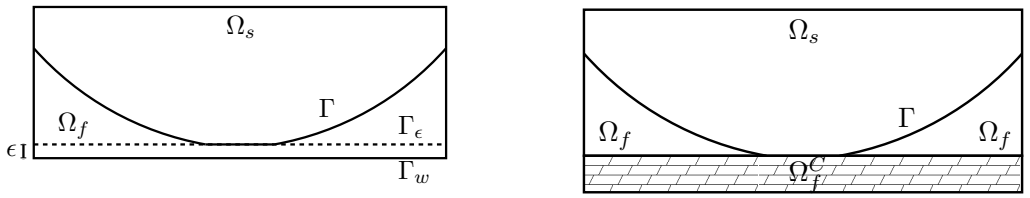


Figure 1.3. Illustration of the two contact approaches. **Left**: Relaxed contact formulation: Body in contact with an obstacle line Γ_ϵ close to Γ_w. **Right**: Artificial fluid domain Ω_f^C below Γ_w (Figure taken from [F1] with permission from *ESAIM*).

where g_0 denotes the initial distance between interface $\Gamma(0)$ and wall Γ_w.

It can be shown that the three conditions in (1.9) are for an arbitrary $\gamma_C > 0$ equivalent to

$$\lambda = -\gamma_C \big[\mathbf{d}\cdot\mathbf{n}_w - g_0 - \frac{1}{\gamma_C}\lambda\big]_+ \quad \text{on } \Gamma(t), \tag{1.10}$$

where $[f]_+ := \max\{f, 0\}$, see e.g. [4, 24]. Using (1.8) the Lagrange multiplier can be further eliminated to get

$$[\![\boldsymbol{\sigma}_n]\!] = -\gamma_C \big[P_\gamma([\![\boldsymbol{\sigma}_n]\!], \mathbf{d})\big]_+, \tag{1.11}$$

where $P_\gamma([\![\boldsymbol{\sigma}_n]\!], \mathbf{d}) := \mathbf{d}\cdot\mathbf{n}_w - g_0 - \gamma_C^{-1}\boldsymbol{\sigma}_n$. Equation (1.11) can be used to incorporate the contact condition into the variational formulation by variational principles, see below.

In fact, a detailed analysis of the corresponding contact conditions in the discrete formulation reveals that it is more convenient to use the discrete Nitsche fluxes

$$[\![\tilde{\boldsymbol{\sigma}}_n]\!] := \mathbf{n}^T(\boldsymbol{\sigma}_s - T_f(\mathbf{u}, p, \dot{\mathbf{d}}))\mathbf{n}\,(\mathbf{n}\cdot\mathbf{n}_w).$$

(see (1.7)) in the discrete setting. For the details we refer to [F1]. As before we can derive the equivalent FSI-contact condition

$$[\![\widetilde{\boldsymbol{\sigma}}_n]\!] = -\gamma_C \big[P_\gamma([\![\widetilde{\boldsymbol{\sigma}}_n]\!], \mathbf{d})\big]_+. \tag{1.12}$$

Note that (1.12) combines both the FSI coupling and the contact conditions on a joint FSI-contact surface $\Gamma(t)$. The switch between both conditions is included implicitly by means of the sign of $P_\gamma([\![\boldsymbol{\sigma}_n]\!], \mathbf{d})$. Using Nitsche's method, (1.12) can be incorporated easily into a monolithic variational formulation:

Variational Formulation 2. *Find* $\mathbf{u} \in \mathcal{V}_h, p \in \mathcal{Q}_h, \mathbf{d} \in \mathcal{W}_h$ *such that* $\dot{\mathbf{d}} = \partial_t \mathbf{d}$ *and*

$$\begin{aligned}\mathcal{A}_{\text{fsi,slip}}(\mathbf{u}, p, \mathbf{d}, \dot{\mathbf{d}})(\mathbf{v}, q, \mathbf{w}) + \gamma_C \left([P_\gamma([\![\widetilde{\boldsymbol{\sigma}}_n]\!], \mathbf{d})]_+, \mathbf{w}\cdot\mathbf{n}_w\right)_{\Gamma(t)} \\ = (\mathbf{f}_f, \mathbf{v})_{\Omega_f(t)} + (\mathbf{f}_s, \mathbf{w})_{\Omega_s(t)} \quad \forall \mathbf{v}, q, \mathbf{w} \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h.\end{aligned} \tag{1.13}$$

The contact parameter is chosen as $\gamma_C = \gamma_C^0 h^{-1}$ depending on the mesh size h, see [24].

1.1.3 Extension of Fluid Forces

In [F1] we consider two possibilities to extend the fluid forces onto the contact surface. The first idea is to keep a thin fluid layer of thickness $\epsilon > 0$ between fluid and solid at any time. This can be achieved by a *relaxation* of the no-penetration condition in (1.9) to

$$\mathbf{d} \cdot \mathbf{n}_w - g_\epsilon \leq 0, \qquad g_\epsilon = g_0 + \epsilon. \tag{1.14}$$

We will call this a *relaxation approach* in the following. This is equivalent to imposing the no-penetration condition already at a virtual barrier Γ_ϵ at an ϵ-distance to Γ_w, see Figure 1.3, left sketch. In the numerical results we choose $\epsilon(h) = ch$ depending on the mesh size h, such that $\epsilon(h) \to 0$ for $h \to 0$. The variational formulation is given by Variational Formulation 2 with g_0 replaced by g_ϵ in the definition of P_γ.

As a second possibility we consider an extension of the fluid domain $\Omega_f(t)$ by an artificial fluid domain Ω_f^C across the boundary Γ_w, see Figure 1.3, right sketch. In [F1], an extension of the fluid forces $\boldsymbol{\sigma}_f \mathbf{n}$ is constructed by solving a Stokes problem including an additional L^2-penalty term that avoids penetration of fluid into Ω_f^C for $h \to 0$

$$\partial_t \mathbf{u} - \operatorname{div} \boldsymbol{\sigma}_f(\mathbf{u}, p) + \gamma_a \mathbf{u} = 0, \quad \operatorname{div} \mathbf{u} = 0 \quad \text{in } \Omega_f^C, \tag{1.15}$$

where $\gamma_a = \gamma_a^0 h^{-2}$. The artificial fluid is coupled by continuity of velocities $\mathbf{u}$ and normal stresses $\boldsymbol{\sigma}_f \mathbf{n}$ across Γ_w to $\Omega_f(t)$. We will call this the *artificial fluid approach* in the following.

In fact the system (1.15) can be seen as a Brinkman model describing a porous medium with asymptotically vanishing permeability $K = \gamma_a^{-1} h^2$ for $h \to 0$. A mathematical justification of the approach has been given by Angot [6], who showed convergence of the Stokes-Brinkman coupled model to a Stokes-rigid body system for $h \to 0$.

In fact the *relaxation approach* can be cast in a framework similar to the *artificial fluid approach* by considering the fluid layer below Γ_ϵ as an artificial fluid domain Ω_f^C. The main difference is the type of equations used in the articifial fluid layer (Stokes versus Brinkman). Obviously, the usage of a pure Stokes equation (without penalising penetration) can only yield reasonable approximations in a very thin layer Ω_f^C.

Considering slip interface conditions the variational formulation for the *artificial fluid approach* reads:

Variational Formulation 3. *Find* $\mathbf{u} \in \mathcal{V}_h, p \in \mathcal{Q}_h, \mathbf{d} \in \mathcal{W}_h$ *such that* $\dot{\mathbf{d}} = \partial_t \mathbf{d}$ *and*

$$\begin{aligned} \mathcal{A}^C_{\text{fsi,slip}}(\mathbf{u}, p, \mathbf{d}, \dot{\mathbf{d}})(\mathbf{v}, q, \mathbf{w}) + \gamma_C \left([P_\gamma([\![\widetilde{\boldsymbol{\sigma}}_n]\!], \mathbf{d})]_+, \mathbf{w} \cdot \mathbf{n}_w \right)_{\Gamma(t)} \\ = (\mathbf{f}_f, \mathbf{v})_{\Omega_f(t)} + (\mathbf{f}_s, w)_{\Omega_s(t)} \quad \forall \mathbf{v}, q, \mathbf{w} \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h, \end{aligned} \tag{1.16}$$

where

$$\begin{aligned} &\mathcal{A}^C_{\text{fsi,slip}}(\mathbf{u}, p, \mathbf{d}, \dot{\mathbf{d}})(\mathbf{v}, q, \mathbf{w}) \\ &\quad := (\partial_t \mathbf{u}, \mathbf{v})_{\Omega_f(t) \cup \Omega_f^C} + (\boldsymbol{\sigma}_f(\mathbf{u}, p), \nabla \mathbf{v})_{\Omega_f(t) \cup \Omega_f^C} + (\operatorname{div} \mathbf{u}, q)_{\Omega_f(t) \cup \Omega_f^C} \\ &\qquad + S(p, q) + \gamma_a (\mathbf{u}, \mathbf{v})_{\Omega_f^C} + \left(\partial_t \dot{\mathbf{d}}, \mathbf{w} \right)_{\Omega_s(t)} + (\boldsymbol{\sigma}_s(\mathbf{d}), \nabla \mathbf{w})_{\Omega_s(t)} \\ &\qquad\quad - (T_f(\mathbf{u}, p, \mathbf{d}) \cdot \mathbf{n}, (\mathbf{w} - \mathbf{v}) \cdot \mathbf{n})_{\Gamma(t)} - \left((\dot{\mathbf{d}} - \mathbf{u}) \cdot \mathbf{n}, \mathbf{n}^T \boldsymbol{\sigma}_f(\mathbf{v}, -q) \mathbf{n} \right)_{\Gamma(t)}. \end{aligned} \tag{1.17}$$

We refer to [F1] for a detailed derivation, as well as for the corresponding form $\mathcal{A}_{\text{fsi,no-slip}}$ for no-slip conditions.

We will see below that -due to the weak imposition of the contact conditions by means of Nitsche's method- the discrete interface $\Gamma(t)$ can slightly overlap with Γ_ϵ in the *relaxation approach* resp. with Γ_w in the *artificial fluid approach*, which increases the stability of the numerical method essentially.

Similar methods have also been applied by different authors recently. Ager et al. [2] used a relatively complex porous model based on Biot's equations in a thin layer between the contacting structures, which yields a well-defined and physically motivated extension of the fluid forces. While they used a Lagrange multiplier approach in the aforementioned publication, the same authors followed a Nitsche-based approach later in the context of FSI with two-body contact [3], but with a very simple constant extension of the fluid forces. The recent work of Formaggia, Gatti and Zonca uses a *relaxation approach* similar to the one mentioned above, but with a Navier-slip FSI interface condition near the contact region [36]. In [F2] resp. Section 1.2, we will present another physically motivated model for the fluid extension, based on the consideration of seepage through the contact surface.

1.1.4 Stability Result

One remarkable feature of the variational contact formulations derived above is that they allow for an elegant numerical stability analysis. Before formulating the result, let us introduce a generalised contact formulation motivated by the work of Chouly *et al.* [25] for a purely solid mechanical contact problem including a parameter $\theta \in [0,1]$.

Variational Formulation 4. *Find* $\mathbf{u}(t) \in \mathcal{V}_h, p(t) \in \mathcal{Q}_h, \mathbf{d}(t) \in \mathcal{W}_h$, *such that* $\dot{\mathbf{d}} = \partial_t \mathbf{d}$ *and*

$$\begin{aligned}
&\mathcal{A}^*_\theta(\mathbf{u},p,\mathbf{d})(\mathbf{v},q,\mathbf{w}) \\
&\quad := \mathcal{A}^*_{\text{fsi,slip}}(\mathbf{u},p,\mathbf{d})(\mathbf{v},q,\mathbf{w}) + \gamma_C \left([P_\gamma(\lambda,d)]_+, \mathbf{w}\cdot\mathbf{n}_w\right)_{\Gamma(t)} \\
&\qquad -\theta \left(\gamma_C[P_\gamma(\lambda,\mathbf{d})]_+\lambda, \lambda_s(\mathbf{w})\right)_{\Gamma(t)} - \theta \left(\partial_t\left(\gamma_C[P_\gamma(\lambda,\mathbf{d})]_+ + \lambda\right), \lambda_f(\mathbf{v},q,\mathbf{w})\right)_{\Gamma(t)} \\
&\qquad = (\mathbf{f}_f, v)_{\Omega_f(t)} + (\mathbf{f}_s, w)_{\Omega_s(t)} \quad \forall \mathbf{v},q,\mathbf{w} \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h,
\end{aligned} \tag{1.18}$$

where

$$\lambda = [\![\widetilde{\boldsymbol{\sigma}}_n]\!], \quad \lambda_s(\mathbf{w}) = \boldsymbol{\sigma}_{s,n}(\mathbf{w})(\mathbf{n}\cdot\mathbf{n}_w), \quad \lambda_f(\mathbf{v}) = (T_f(v,q,w)\cdot\mathbf{n})(\mathbf{n}\cdot\mathbf{n}_w)$$

and $\mathcal{A}^*_{\text{fsi,slip}}$ *is one of the bilinear forms* $\mathcal{A}_{\text{fsi,slip}}$ *or* $\mathcal{A}^C_{\text{fsi,slip}}$ *introduced above.*

For $\theta = 0$ we recover Variational Formulation 2 resp. 3. As in [25] we were able to show a stability result for the symmetric formulation ($\theta = 1$) and local stability (around the unknown solution) in the general case (including the case $\theta = 0$ considered in the previous sections). For the stability analysis, we will assume infinitesimal displacements, i.e. the sub-domains Ω_f and Ω_s as well as the interface Γ are fixed. We introduce the notation $\widetilde{\Omega}_f$ for the combined fluid and artificial fluid domain in the case of $\mathcal{A}^C_{\text{fsi,slip}}$, while $\widetilde{\Omega}_f = \Omega_f$ in the relaxation approach, as well as the energy norm

$$E(T) := \|\mathbf{u}(T)\|^2_{\widetilde{\Omega}_f} + \|\dot{\mathbf{d}}(T)\|^2_{\Omega_s} + \|\mathbf{d}(T)\|^2_{H^1(\Omega_s)} + \theta \left\|\gamma_C^{-1/2}\lambda(T) + \gamma_C^{1/2}[P_\gamma(\lambda,\mathbf{d})]_+(T)\right\|^2_\Gamma.$$

Theorem 1.1. *Let $\mathbf{u}, p, \mathbf{d} \in V_h$ and $\dot{\mathbf{d}} = \partial_t \mathbf{d}$. Under the assumptions made above, the following stability result holds for the form $\mathcal{A}_\theta^*$ defined in (1.18), where $\theta \in [0,1]$ and γ_C^0 sufficiently large*

$$E(T) + \int_0^T \left(\nu_f \|\nabla \mathbf{u}\|^2_{\widetilde{\Omega}_f} + S_p(p,p) + \gamma_a \|\mathbf{u}\|^2_{\Omega_f^C} \right) + \gamma_{fsi} \|(\dot{\mathbf{d}} - \mathbf{u}) \cdot \mathbf{n}\|^2_\Gamma \, dt$$
$$\leq c \Bigg(E(0) + \int_0^T \mathcal{A}_\theta^*(\mathbf{u}, p, \mathbf{d}; \mathbf{u}, p, \dot{\mathbf{d}}) \; - (1-\theta)\gamma_C \underbrace{\left([P_\gamma(\lambda, \mathbf{d})]_+, \dot{\mathbf{d}} \cdot \mathbf{n}_w \right)_\Gamma}_{(I)} \, dt \Bigg).$$

Proof. See [F1] □

For $\theta = 1$ we get control over the energy norm $E(T)$ at the end time and some fluid dissipation over the whole time interval. For $\theta = 0$, i.e. the variational formulation introduced above, the additional term (I) remains on the right-hand side. This term vanishes for the continuous solution $(\mathbf{u}, p, \mathbf{d})$, as on $\Gamma(t)$ we have either $[P_\gamma(\lambda, \mathbf{d})]_+ = 0$ in points with a positive distance to Γ_w or $\dot{\mathbf{d}} \cdot \mathbf{n}_w = 0$ in points in contact. Thus, Theorem 1.1 can be seen as a local stability result for $\theta = 0$, as we have control over $E(T)$ in a certain neighborhood around the continuous solution.

1.1.5 A Summary of the Numerical Results

In the numerical results in [F1] we compare the two approaches on different mesh levels, observing that for $h \to 0$ the results are in reasonable agreement. Moreover, detailed convergence studies in time and space, as well as sensitivity studies with respect to the contact parameters are provided.

Finally, the numerical approaches are compared to a simple ad-hoc treatment of the dynamically changing conditions on $\Gamma(t)$, where the interface is split explicitly in each time-step $t_{m-1} \to t_m$ in a part $\Gamma_C(t_{m-1})$ which is at $t = t_{m-1}$ in contact with Γ_w, such that contact conditions are imposed, and a part $\Gamma_{\text{fsi}}(t_{m-1})$ with FSI coupling conditions. We observe that the latter approach leads to severe oscillations after the impact, while the approaches presented above are (almost perfectly) stable.

1.2 A Model including Seepage [F2]

The two strategies to extend the fluid forces in [F1] resp. Section 1.1.3 are easy to implement, but lack a rigorous physical foundation. It is unclear a priori, whether they reproduce the correct physics, in particular concerning the questions if, when and where contact is released.

On the other hand, the simplicity and versatility of the *relaxed* contact formulation is particularly appealing, since for example extensions to two- or multi-body contact are straight-forward. As topology changes are avoided, it is in principle even possible to impose no-slip conditions on the interface $\Gamma(t)$ and the boundary Γ_w, as typically used for viscous fluids in fluid mechanics (in absence of contact). Numerical results show, however, that in this case the fluid forces in the layer get very large (due to a singularity in the pressure for $\epsilon_g \to 0$) and might influence the contact dynamics significantly. The physically correct interface and boundary conditions, when it comes to contact, are still under investigation.

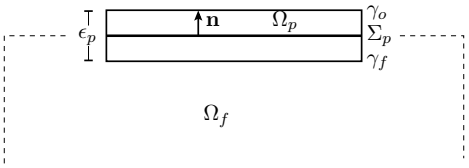


Figure 1.4. Porous medium domain Ω_p with interface γ_f to Ω_f and exterior boundary γ_o (Figure taken from [F2] with permission from *Elsevier*).

For these reasons, we introduce a physically motivated model for the fluid extension in [F2]. The idea is that due to surface roughness two bodies are never exactly in contact with each other throughout their surfaces. Thus, we allow for seepage of fluid through the contact surface (cf. Figure 1.2). To this purpose we introduce a $(d-1)$ dimensional porous layer Σ_p at those parts of the wall Γ_w, where contact might take place and derive an appropriate model based on a Darcy equation on Σ_p. The paradoxon explained in Figure 1.2 is resolved, as fluid con flow through the porous layer into the cavity and moreover, the porous pressure provides a physically motivated resistance against the lift-off of the body.

1.2.1 Model for the Porous Layer

The derivation of a model for seepage through a porous layer given in [F2] goes back to Martin, Jaffre & Robert [70]. Let us consider the configuration sketched in Figure 1.4, where a thin porous layer $\Omega_\mathrm{p} = \Sigma_\mathrm{p} \times (-\frac{\epsilon_\mathrm{p}}{2}, \frac{\epsilon_\mathrm{p}}{2}) \in \mathbb{R}^d$ $(d = 2, 3)$ with midsurface Σ_p is coupled to a surrounding fluid in a fixed domain $\Omega_f \subset \mathbb{R}^d$. The fluid in Ω_f is governed by the Navier-Stokes equations

$$\rho_\mathrm{f}\big(\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}\big) - \operatorname{div} \boldsymbol{\sigma}_f(\mathbf{u}, p) = \mathbf{0}, \qquad \operatorname{div} \mathbf{u} = 0 \quad \text{in} \quad \Omega_f. \tag{1.19}$$

In the porous domain Ω_p, we assume a Darcy law

$$\mathbf{u}_\mathrm{l} + \boldsymbol{K} \nabla p_\mathrm{l} = 0, \qquad \nabla \cdot \mathbf{u}_\mathrm{l} = 0 \quad \text{in} \quad \Omega_\mathrm{p}, \tag{1.20}$$

where $\mathbf{u}_\mathrm{l}$ denotes the Darcy velocity, p_l the Darcy pressure and $\boldsymbol{K}$ is a $d \times d$ matrix which can be decomposed into $\boldsymbol{K}\nabla p_\mathrm{l} = K_\tau \nabla_\tau p_\mathrm{l} + K_n \boldsymbol{\partial}_n p_\mathrm{l}$, where $K_\tau, K_n \in \mathbb{R}^+$. Here, $\mathbf{n}$ is the unit normal vector of the mid-surface Σ_p that points towards the exterior boundary γ_o, $\boldsymbol{\partial}_n = \mathbf{n}\partial_n$ and $\nabla_\tau := P_\tau \nabla$ stands for the corresponding tangential part of the gradient, where $P_\tau := (\boldsymbol{I} - \mathbf{n} \otimes \mathbf{n})$.

Furthermore, we assume that the porous layer is very thin compared to the size of the solid and consider the limit case $\epsilon_\mathrm{p} \to 0$. Let the outer boundary of Ω_p be denoted by γ_o and the interior boundary connecting to the fluid domain Ω_f by γ_f, see Figure 1.4. We assume zero normal velocity $(\mathbf{u}_\mathrm{l} \cdot \mathbf{n} = 0)$ on the outer boundary γ_o and continuity of normal velocities and normal stresses on γ_f. For the tangential fluid stresses, we consider the Beavers-Joseph-Saffman coupling conditions [81], such that the coupling conditions between porous medium Ω_p and fluid Ω_f are given by

$$\begin{cases} \boldsymbol{\sigma}_{\mathrm{f},n} = -p_\mathrm{l} & \text{on} \quad \gamma_\mathrm{f}, \\ \mathbf{u} \cdot \mathbf{n} = \mathbf{u}_\mathrm{l} \cdot \mathbf{n} & \text{on} \quad \gamma_\mathrm{f}, \\ \boldsymbol{\sigma}_{\mathrm{f},\tau} = -\dfrac{\alpha}{\sqrt{K_\tau \epsilon_\mathrm{p}}} \mathbf{u}_\tau & \text{on} \quad \gamma_\mathrm{f}, \end{cases} \tag{1.21}$$

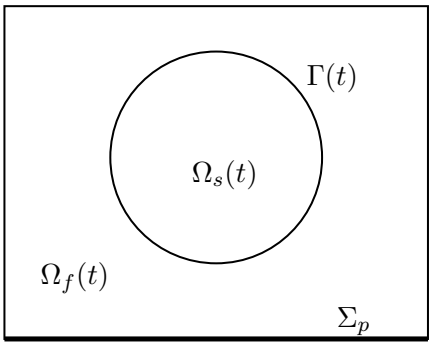


Figure 1.5. Geometric configurations of fluid ($\Omega_f(t)$) and solid domain ($\Omega_s(t)$) and porous layer Σ_p.

where $\mathbf{u}_\tau := P_\tau \mathbf{u}$ denotes the tangential part of the velocity vector and $\boldsymbol{\sigma}_{\mathrm{f},n} = \mathbf{n}^T \boldsymbol{\sigma}_f \mathbf{n}$ and $\boldsymbol{\sigma}_{\mathrm{f},\tau} = P_\tau \boldsymbol{\sigma}_f \mathbf{n}$ are the normal and tangential part of the fluid Cauchy stress tensor $\boldsymbol{\sigma}_f$. We note that the condition for the tangential stresses in the last line of (1.21) corresponds to a Navier-slip boundary condition for the fluid. However, in contrast to the typical Navier-slip boundary condition for the fluid, the normal velocity $\mathbf{u} \cdot \mathbf{n}$ is not zero here, as the fluid can enter the porous layer. For a discussion on the appropriate choice of α, we refer to [F2].

Introducing the averaged porous pressure P_l as

$$P_\mathrm{l} = \frac{1}{2}\left(p_\mathrm{l}|\gamma_\mathrm{f} + p_\mathrm{l}|\gamma_\mathrm{o}\right) \quad \text{in} \quad \Sigma_\mathrm{p}, \tag{1.22}$$

the following equations can be derived in the limit case $\epsilon_\mathrm{p} \to 0$ (see [22, 70])

$$\begin{cases} -\nabla_\tau \cdot (\epsilon_\mathrm{p} K_\tau \nabla_\tau P_\mathrm{l}) = \mathbf{u} \cdot \mathbf{n} & \text{on} \quad \Sigma_\mathrm{p}, \\ \boldsymbol{\sigma}_{\mathrm{f},n} = -P_\mathrm{l} - \dfrac{\epsilon_\mathrm{p} K_n^{-1}}{4} \mathbf{u} \cdot \mathbf{n} & \text{on} \quad \Sigma_\mathrm{p}, \\ \boldsymbol{\sigma}_{\mathrm{f},\tau} = -\dfrac{\alpha}{\sqrt{K_\tau \epsilon_\mathrm{p}}} \mathbf{u}_\tau & \text{on} \quad \Sigma_\mathrm{p}. \end{cases} \tag{1.23}$$

Note that the only remaining porous medium variable is the averaged pressure P_l. In the limit $K_n, K_\tau \to 0$, the coupling conditions turn into a Navier-slip boundary condition for the fluid on Σ_p.

Similar ideas based on the consideration of surface roughness have been introduced by Ager et al. in [2]. The model proposed in their work based on Biot's equations in a d-dimensional porous layer is, however, much more complicated and includes several (possibly unknown) parameters.

1.2.2 Fluid-Structure-Porous-Contact Interaction Problem

In [F2] we consider two cases, namely a coupling with a hyperelastic d-dimensional solid as in Section 1.1 on one hand, and a coupling with a thin $(d-1)$-dimensional solid, e.g. a shell or beam model, on the other hand. For simplicity, we will only present the model for the d-dimensional solid here and refer to [F2] for the $(d-1)$-dimensional case. We will use a relaxed treatment of the contact conditions.

As an example consider the sketch given in Figure 1.5. In the fluid domain $\Omega_f(t)$, we assume the incompressible Navier-Stokes equations (1.19), in the solid domain $\Omega_s(t)$ a linear elastic material law, as in (1.2). In absence of contact the equations are coupled by means of no-slip interface conditions

$$\dot{\mathbf{d}} = \mathbf{u}, \quad \boldsymbol{\sigma}_f(\mathbf{u}, p)\mathbf{n} = \boldsymbol{\sigma}_s(\mathbf{d})\mathbf{n} \quad \text{on} \quad \Gamma_{\text{fsi}}(t).$$

To model contact, we introduce again a Lagrange multiplier $\lambda = \boldsymbol{\sigma}_{s,n} - \boldsymbol{\sigma}_{f,n}$ which can get negative during contact, where $\boldsymbol{\sigma}_{s,n} = \mathbf{n}^T\boldsymbol{\sigma}_s\mathbf{n}$. As in (1.10)-(1.11) the *relaxed* form of the contact conditions can be written as

$$[\![\tilde{\boldsymbol{\sigma}}_n]\!] = -\gamma_{\text{C}}\big[P_{\gamma_{\text{C}}}(\mathbf{d}\cdot\mathbf{n}, [\![\tilde{\boldsymbol{\sigma}}_n]\!])\big]_+ \quad \text{on} \quad \Gamma(t), \tag{1.24}$$

for $\gamma_{\text{C}} > 0$, where $P_{\gamma_{\text{C}}}(\mathbf{d}\cdot\mathbf{n}, [\![\tilde{\boldsymbol{\sigma}}_n]\!]) = \mathbf{d}\cdot\mathbf{n} - g_\epsilon - \gamma_{\text{C}}^{-1}[\![\tilde{\boldsymbol{\sigma}}_n]\!]$ and g_ϵ as defined in (1.14).

In addition, the fluid is coupled to the porous layer on $\Sigma_p \subset \Gamma_w$ by the coupling conditions

$$\boldsymbol{\sigma}_{\text{f},n} = \underbrace{-P_{\text{l}} - \frac{\epsilon_{\text{p}}K_n^{-1}}{4}\mathbf{u}\cdot\mathbf{n}}_{=:\boldsymbol{\sigma}_{\text{p}}}, \qquad \boldsymbol{\sigma}_{\text{f},\tau} = -\frac{\alpha}{\sqrt{K_\tau\epsilon_{\text{p}}}}u_\tau \qquad \text{on} \quad \Sigma_{\text{p}}. \tag{1.25}$$

The porous pressure P_l is determined by the equation $(1.23)_1$, i.e.

$$-\nabla_\tau\cdot(\epsilon_{\text{p}}K_\tau\nabla_\tau P_{\text{l}}) = \mathbf{u}\cdot\mathbf{n} \quad \text{on} \quad \Sigma_{\text{p}}. \tag{1.26}$$

Note that due to the *relaxed* treatment of contact, the solid is never fully in contact with Σ_p such that the fluid forces $\boldsymbol{\sigma}_{f,n}$ remain well-defined for all times on $\Sigma(t)$. Moreover, the porous layer gives also a physical meaning to $\boldsymbol{\sigma}_{f,n}$ in the fluid layer during contact, as for small ϵ_g

$$\boldsymbol{\sigma}_{f,n} \approx \boldsymbol{\sigma}_p \circ \pi \quad \text{on} \quad \Gamma_C(t), \tag{1.27}$$

where π denotes a closest-point projection from $\Gamma_C(t)$ to Σ_p and σ_p is the physically motivated porous stress on Σ_p.

The coupled model (1.2), (1.19) and (1.24)-(1.26) can again be incorporated in a variational formulation using Nitsche's method and suitable finite element spaces ($\mathcal{V}_h, \mathcal{Q}_h, \mathcal{W}_h$ and $\mathcal{S}_h$ (see Section 2):

Variational Formulation 5. *Find* $(\mathbf{u}, p, \dot{\mathbf{d}}, \mathbf{d}, P_l) \in (\mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h \times \mathcal{W}_h \times \mathcal{S}_h)$ *with* $\dot{\mathbf{d}} = \partial_t\mathbf{d}$ *and*

$$\begin{aligned}\mathcal{A}_{\text{fsi,no-slip}}(\mathbf{u}, p, \mathbf{d}, \dot{\mathbf{d}})(\mathbf{v}, q, \mathbf{w}) + \frac{\alpha}{\sqrt{K_\tau\epsilon_p}}(\mathbf{u}_\tau, \mathbf{v}_\tau)_{\Sigma_p} - (\boldsymbol{\sigma}_p, \mathbf{v}\cdot\mathbf{n})_{\Sigma_p} \\ + (\epsilon_p K_\tau\nabla_\tau P_l, \nabla_\tau q_l)_{\Sigma_p} - (\mathbf{u}\cdot\mathbf{n}, q_l)_{\Sigma_p} = 0\end{aligned} \tag{1.28}$$

for all $(\mathbf{v}, q, \mathbf{w}, q_l) \in (\mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h \times \mathcal{S}_h)$*, with the porous stress* $\boldsymbol{\sigma}_p$ *defined in* (1.25) *and*

$$\begin{aligned}&\mathcal{A}_{\text{fsi,no-slip}}(\mathbf{u}, p, \mathbf{d}, \dot{\mathbf{d}})(\mathbf{v}, q, \mathbf{w}) \\ &\quad := \rho_f(\partial_t\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u}, \mathbf{v})_{\Omega_f} + (\boldsymbol{\sigma}_f(\mathbf{u}, p), \nabla\mathbf{v})_{\Omega_f} + (\operatorname{div}\mathbf{u}, q)_{\Omega_f} \\ &\qquad + \rho_s(\partial_t\dot{\mathbf{d}}, \mathbf{w})_{\Omega_s} + (\boldsymbol{\sigma}_s(\mathbf{d}), \nabla\mathbf{d}, \mathbf{w})_{\Omega_s} + \frac{\gamma_C E}{h}([P_{\gamma_c}(\mathbf{d})]_+, \mathbf{w}\cdot\mathbf{n})_\Gamma \\ &\qquad\quad - (\boldsymbol{\sigma}_f(\mathbf{u}, p)\mathbf{n}, \mathbf{v} - \mathbf{w})_\Gamma - (\mathbf{u} - \dot{\mathbf{d}}, \boldsymbol{\sigma}_f(\mathbf{v}, -q))\mathbf{n})_\Gamma + \frac{\gamma_{fsi}\nu_f}{h}(\mathbf{u} - \dot{\mathbf{d}}, \mathbf{v} - \mathbf{w})_\Gamma.\end{aligned}$$

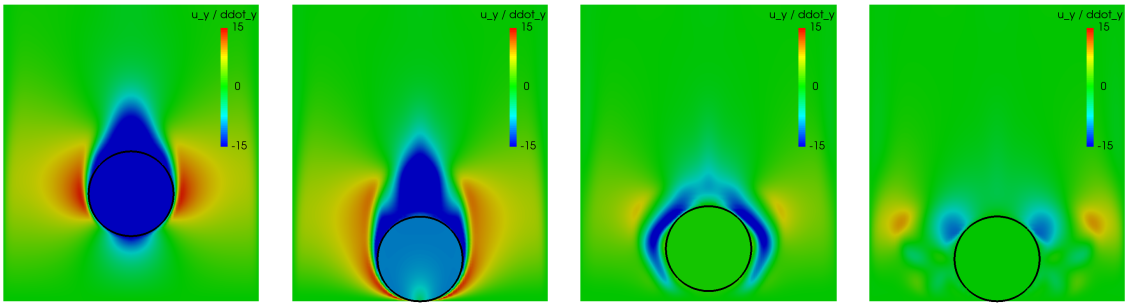


Figure 1.6. Illustration of the vertical velocities in the falling ball example at $t = 0.28$, $t = 0.382$, $t = 0.47$ and $t = 0.63$s (taken from [F2] with permission from *Elsevier*).

We note that the porous medium and the structure are always coupled with the fluid only and never directly to each other. This avoids switches in the variational formulation, which would be necessary in the transition between a fluid-solid and solid-porous interaction (see [22]), when the fluid layer vanishes. On the other hand, the solid perceives indirectly the presence of the porous layer through the fluid stresses and velocity during contact by means of (1.27) and (1.24). The resulting numerical approach is highly competitive in terms of computational costs compared to approaches using Lagrange multipliers and/or active-sets.

1.2.3 A Summary of the Numerical Results

In [F2] the model and the numerical framework are tested in two very different model problems in two space dimensions. One one hand, we investigate the problem of a falling elastic ball within a viscous fluid, including contact with the ground and subsequent rebounces, see Figure 1.6. On the other hand, we consider the closure and opening of two symmetric $(d-1)$-dimensional valves modelled by a non-linear Reissner-Mindlin curved beam model within a fluid flow, which is driven by an applied pressure on the fluid boundary, see Figure 1.7.

Moreover, we show the versatility of the approach by using different numerical techniques. In the case of a bouncing ball, a *Fully Eulerian* approach is used in combination with a fitted finite element discretisation and a dG(0) time-stepping scheme, as in Section 1.1. In the valve example, we use an *immersed* mixed-coordinate approach, where the fluid equations are formulated in Eulerian and the $(d-1)$ dimensional solid equations in Lagrangian coordinates. For discretisation we use an unfitted finite element discretisation [17, 21, 56] with the time-stepping scheme that will be introduced in Section 2.2.

A sensitivity analysis with respect to the conductivity $K = K_n = K_\tau$ of the porous layer reveals convergence to pure slip (valve example) resp. Navier-slip (ball example) boundary conditions for the fluid for $K \to 0$, as in the limit no fluid can escape through the porous layer anymore. For large K fluid can flow abundantly through the porous layer and the resisitive forces that act "against contact" are much smaller. This means that contact happens earlier in both examples and is released very late in the valve example, see Figure 1.7. In the ball example the bouncing height is higher for large K, as the impact velocity is higher.

Moreover, we provide similar sensitivity studies with respect to the parameters as well as convergence studies for space and time discretisation in [F2]. Finally, we compare the solution in the ball example obtained with the porous model against the solution of the

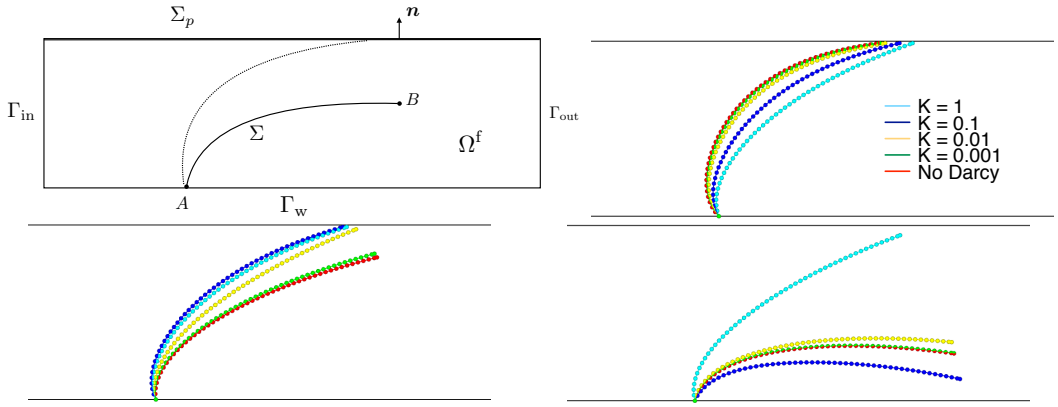


Figure 1.7. Top left: Geometric configuration of the valve in the initial state (Σ) and a state in contact. Due to symmetry, the domain is reduced to the lower half of the valve. **Top right** to **bottom right**: Position of the valves for different K at $t = 0.25$ (in contact), $t = 0.45$ (contact is released for $K \leq 10^{-2}$) and $t = 1$ (Figure taken from [F2] with permission from *Elsevier*).

same problem using a relaxation approach without a porous layer (as in Section 1.1.2) and no-slip resp. Navier-slip boundary conditions $\boldsymbol{\sigma}_{\text{f},\tau} = -\frac{\alpha}{\sqrt{K\epsilon_{\text{p}}}} u_\tau$ for the fluid on Σ_p for $\alpha = 1$, $K = 10^{-2}, \epsilon_p = 10^{-4}$. As the factor $\frac{\alpha}{\sqrt{K_\tau \epsilon_{\text{p}}}}$ is large, the solutions for no-slip and Navier-slip conditions are almost identical. For $h \to 0$, the bounces get smaller and the solution without the porous layer converges to a ball that does not bounce. This is in agreement with analytical findings for no-slip conditions (the so-called *no-collision paradoxon*, see e.g. [59]). On the other hand, using a porous layer with the same parameters yields convergence of solutions towards a bouncing ball with a bouncing height of approximately 3mm, when the ball is dropped from a height of 5cm. The experimental results described in the next section resp. [F3] indicate that this behaviour is indeed the physical one.

1.3 A Benchmark Configuration for FSI and Contact [F3]

To validate the different approaches to model FSI and contact, we have proposed a benchmark configuration with experimental data in [F3]. The experiments have been conducted by our collaborator Thomas Hagemeier [53]. Such a benchmark configuration is particularly important, since the correct physical model for FSI and contact is still unclear and different approaches have been proposed recently [2, 3, 36, F1, F2]. To our knowledge this is the first benchmark for FSI with contact including an experimental comparison.

1.3.1 Configuration of Benchmark and Experiment

We consider an elastic sphere which is dropped within a glass cylinder filled with a viscous water-glycerine mixture at equal volume fraction, see Figure 1.8. Due to gravity the sphere is accelerated and falls down. It comes to contact with the bottom of the glass cylinder and several bounces are observed afterwards. The experiments in [53] contain results for three different solid materials, two of which have been considered in [F3], namely a Teflon ball of

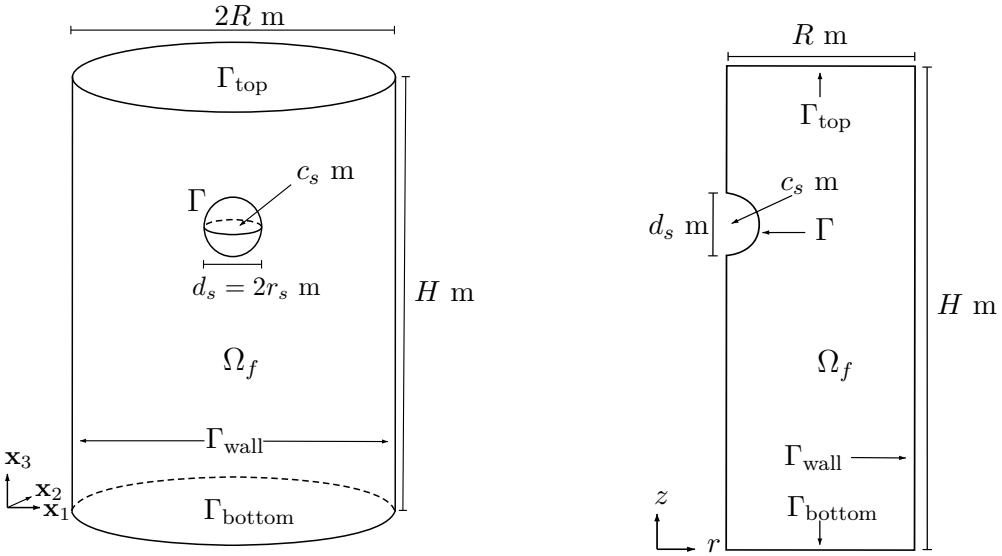


Figure 1.8. The initial spatial configuration. **Left**: Three-dimensional domain, **right**: rotationally reduced domain (taken from [F3] with permission from *AIP Publishing*).

diameter 6mm (in the following PTFE6) and a rubber ball with diameter 22mm (Rubber22). The resulting flow around the balls is in the transitional regime between creeping and turbulent flow. The position of the ball has been recorded by a high-speed CMOS-camera. For the details, we refer to [53].

If we assume that there is no horizontal deflection of the sphere, the setting is rotationally symmetric and the equations can be reduced to a two-dimensional geometry, see the right sketch of Figure 1.8.

Quantities of Interest

We consider six quantities of interest for comparison. The first three quantities are defined to quantify the settling process, namely

t_* The time after release when the ball is a distance of one diameter d_s to the bottom: $\operatorname{dist}(\Gamma, \Gamma_{\text{bottom}}) = d_s$

v_* The velocity of the ball in the vertical direction at time t_*,

f_* The vertical component of the force $\mathbf{F}$ acting on the ball at time t_*,

and three quantities for the contact dynamics:

$t_{\textsf{cont}}$ The time of the first solid contact,

$t_{\textsf{jump}}$ The time between first and second contact,

$d_{\textsf{jump}}$ The maximum of $\operatorname{dist}(\Gamma, \Gamma_{\text{bottom}})$ after contact, i.e., the height of the bounce.

1.3.2 Equations

In the numerical simulations in [F3], we consider models of different complexity for the solid. As deformations are relatively small, in particular before contact, a rigid-body model is

considered on one hand. The solid model is reduced to one degree of freedom, namely the vertical velocity $\dot{\mathbf{d}}_z$, which -according to Newton's law of motion- fulfills the equation

$$\partial_t \dot{\mathbf{d}}_z(t) = \frac{\rho_s - \rho_f}{\rho_s} g + \frac{\mathbf{F}_3}{\rho_s |\Omega_s|}, \tag{1.29}$$

where $g = 9.807 \frac{g}{ms^2}$ is the gravity force and $\mathbf{F}_3$ is the third component of the fluid forces

$$\mathbf{F} = \int_\Gamma \boldsymbol{\sigma}_f \mathbf{n} \, \mathrm{d}s.$$

The solid's motion couples back to the fluid equations through the boundary condition at the interface Γ, by requiring continuity of the velocities, i.e., $\mathbf{u}|_\Gamma = (0, 0, \dot{\mathbf{d}}_z)$.

On the other hand, a linear elastic behaviour is considered as in Section 1.1, with the respective elasticity parameters of PTFE and rubber. While the elasticity modulus $E_s \approx 670$ MPa is known in a good approximation for the PTFE material, it may lie in the range $E_s \in [1.7, 20.7]$ MPa for the rubber material. In the fluid domain, we assume the incompressible Navier-Stokes equations with no-slip interface conditions on $\Gamma_{\text{fsi}}(t)$. For appropriate boundary conditions and their numerical treatment, we refer to [F3].

In the case of an elastic material, we we use the simple *relaxation approach* developed in Section 1.1. In the case of a rigid solid such techniques are not available. Here, we use a basic contact avoidance scheme based on an artificial (lubrication) force, which is widely used in literature [38, 52, 82, 93].

Reduced 2-Dimensional Model

The experimental setup is symmetric with respect to cylindrical coordinates. If we assume that the ball falls down vertically and that there is no rotational component ($\mathbf{u}_\theta = 0$), the equations can be transformed to a two-dimensional setting. In the experiments the horizontal deflection of the ball is not exactly zero, but relatively small (max. 2mm for the PTFE case and 0.75mm in the rubber case) compared to the initial distances of 16.2 resp. 14.6cm from the bottom.

By means of an integral transformation, we obtain the fluid bilinear form

$$\begin{aligned} A_f(\mathbf{u}, p; \mathbf{v}, q) \coloneqq & \rho_f \int_{\Omega_f} r(\partial_t \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u}) \cdot \mathbf{v} \, \mathrm{d}x + \mu_f \int_{\Omega_f} r \nabla \mathbf{u} : \nabla \mathbf{v} + \frac{1}{r} \mathbf{u}_r \mathbf{v}_r \mathrm{d}x \\ & - \int_{\Omega_f} p(\mathbf{v}_r + r \nabla \cdot \mathbf{v}) \, \mathrm{d}x - \int_{\Omega_f} q(\mathbf{u}_r + r \nabla \cdot \mathbf{u}) \, \mathrm{d}x = 0, \end{aligned} \tag{1.30}$$

where r denotes the radial coordinate in the two-dimensional coordinate system and $\mathbf{u}_r, \mathbf{v}_r$ the radial component of the variables $\mathbf{u}, \mathbf{v}$.

Rigid-Body Approximation

For the rigid-body approximation this is coupled to the solid motion by means of the ODE (1.29), where

$$\mathbf{F} = 2\pi \int_\Gamma r \boldsymbol{\sigma}_f \mathbf{n} \, \mathrm{d}s + f_c$$

and by the continuity of velocities at the interface $\mathbf{u}_\Gamma = (0, 0, \dot{\mathbf{d}}_z)$. The latter is imposed weakly by means of the Nitsche term

$$b_f(\mathbf{u}, p; \mathbf{v}, q) := \mu_f \gamma_{\text{fsi}} \int_\Gamma r(\mathbf{u} - \dot{\mathbf{d}}) \cdot \mathbf{v} \, \mathrm{d}s + \int_\Gamma r\boldsymbol{\sigma}_f(\mathbf{u}, p)\mathbf{n} \cdot \mathbf{v} \, \mathrm{d}s + \int_\Gamma r\boldsymbol{\sigma}_f(\mathbf{v}, q)\mathbf{n} \cdot (\mathbf{u} - \dot{\mathbf{d}}) \mathrm{d}s$$

for a given solid motion $\dot{\mathbf{d}} = (0, 0, \dot{\mathbf{d}}_z)$. The problem reads in semi-discretised variational formulation for $t \in I$:

Variational Formulation 6. *Find* $(\mathbf{u}, p, \dot{\mathbf{d}}_z) \in (\mathcal{V}_h \times \mathcal{Q}_h \times \mathbb{R})$ *such that for all* $(\mathbf{v}, q) \in (\mathcal{V}_h \times \mathcal{Q}_h)$ *it holds*

$$A_f(\mathbf{u}, p; \mathbf{v}, q) + b_f(\mathbf{u}, p; \mathbf{v}, q) = 0, \qquad \partial_t \dot{\mathbf{d}}_z(t) = \frac{\rho_s - \rho_f}{\rho_s} g + \frac{\mathbf{F}_3}{\rho_s |\Omega_s|}.$$

Eulerian FSI Model

Similarly the equations characterising FSI and contact in (1.13) can be transformed to cylindrical coordinates. We define the solid bilinear form

$$A_s(\mathbf{d}, \dot{\mathbf{d}}; \mathbf{w}, \mathbf{z}) := \rho_s \int_{\Omega_s} r(\partial_t \dot{\mathbf{d}} - (\dot{\mathbf{d}} \cdot \nabla)\dot{\mathbf{d}}) \cdot \mathbf{w} \, \mathrm{d}x + \int_{\Omega_s} r\nabla\boldsymbol{\sigma}_s(\mathbf{d}) : \nabla\mathbf{w} + \boldsymbol{\sigma}_{s,r}\mathbf{w}_r \, \mathrm{d}x + \int_{\Omega_s} (\partial_t \mathbf{d} - (\dot{\mathbf{d}} \cdot \nabla)\mathbf{d} - \dot{\mathbf{d}}) \cdot \mathbf{z} \, \mathrm{d}x,$$

where

$$\boldsymbol{\sigma}_s = 2\mu_s E(\mathbf{d}) + \lambda_s \left(\mathrm{tr}(E(\mathbf{d})) + \frac{1}{r}\mathbf{d}_r \right) I, \qquad \boldsymbol{\sigma}_{s,r} = \frac{2\mu_s + \lambda_s}{r}\mathbf{d}_r + \lambda_s \mathrm{tr}(E(\mathbf{d})).$$

and $E(\mathbf{d})$ defined in (1.3). The FSI coupling terms read

$$b_{\text{fsi}}(\mathbf{u}, p, \dot{\mathbf{d}}; \mathbf{v}, q, \mathbf{w}) := \mu_f \gamma_{\text{fsi}} \int_\Gamma r(\mathbf{u} - \dot{\mathbf{d}}) \cdot (\mathbf{v} - \mathbf{w}) \mathrm{d}s + \int_\Gamma r\boldsymbol{\sigma}_f(\mathbf{u}, p)\mathbf{n} \cdot (\mathbf{v} - \mathbf{w}) \mathrm{d}s + \int_\Gamma r\boldsymbol{\sigma}_f(\mathbf{v}, -q)\mathbf{n} \cdot (\mathbf{u} - \dot{\mathbf{d}}) \mathrm{d}s.$$

We obtain the variational formulation:

Variational Formulation 7. *Find* $(\mathbf{u}, p, \mathbf{d}, \dot{\mathbf{d}}) \in (\mathcal{V}_h(t) \times \mathcal{Q}_h(t) \times \mathcal{W}_h(t) \times \mathcal{Z}_h(t))$ *such that for all* $(\mathbf{v}, q, \mathbf{w}, \mathbf{z}) \in (\mathcal{V}_h(t) \times \mathcal{Q}_h(t) \times \mathcal{W}_h(t) \times \mathcal{Z}_h(t))$ *it holds*

$$A_f(\mathbf{u}, p; \mathbf{v}, q) + A_s(\mathbf{d}, \dot{\mathbf{d}}; \mathbf{w}, \mathbf{z}) + b_{\text{fsi}}(\mathbf{u}, p, \dot{\mathbf{d}}; \mathbf{v}, q, \mathbf{w}) + (r[P_\gamma(\mathbf{d}, [\![\boldsymbol{\sigma}_n]\!])]_+, \mathbf{w}_n)_\Gamma = (r\mathbf{f}_s, \mathbf{w})_{\Omega_s}.$$

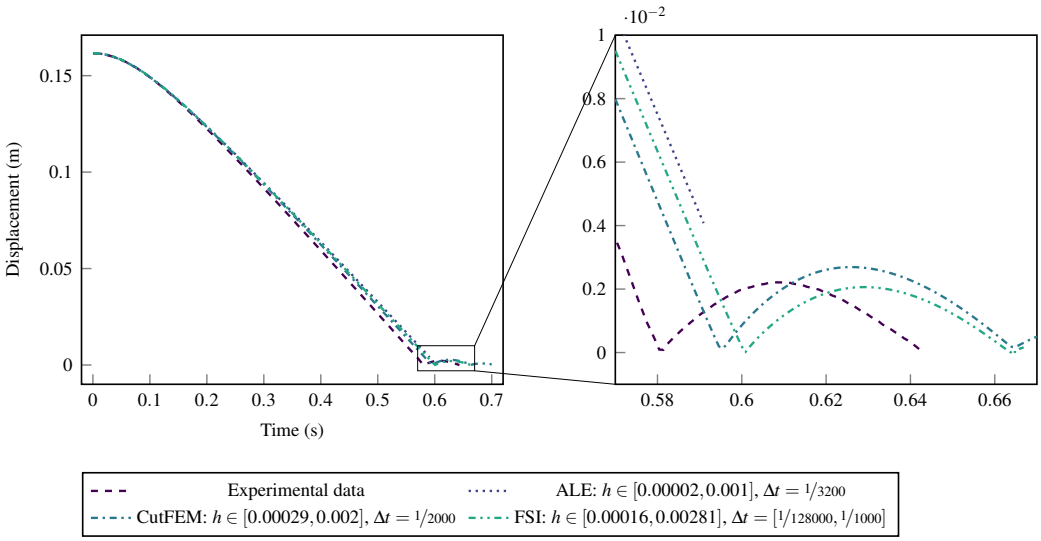


Figure 1.9. The distance between the bottom of the ball and the bottom of the tank: Experimental and numerical results for the PTFE6 set-up (Figure taken from [F3] with permission from *AIP Publishing*).

1.3.3 Summary of the Numerical Results

For the (two-dimensional) rigid-body approximation, we use an unfitted finite element approach in [F3], similar to the valve example in Section 1.2.3, but using *inf-sup-stable* Taylor-Hood elements and a BDF2 time discretisation [92]. This approach will be abbreviated as *CutFEM-RB* in the following. For the details, we refer to Section 2.2 and [F3]. For comparison of the settling process, we also consider a full three-dimensional and a reduced two-dimensional ALE approximation, see [77, F3] (called ALE3d-RB resp. ALE2d-RB). In each time-step, we iterate between equations (1.29) and (1.30) until the update in $\dot{\mathbf{d}}_z$ lies below a specified tolerance. The ALE method is expected to yield the most accurate results for the settling, but it is unable to simulate contact, as the deformed meshes degenerate before it comes to contact. For the (two-dimensional) full FSI model, we use again the equal-order locally modified finite element discretisation applied in Section 1.1, see also the following Section 2.1.

Settling

Concerning the trajectories of the particle and the three pre-contact quantities of interest, the four numerical approaches *ALE3d-RB*, *ALE2d-RB*, *CutFEM-RB* and *Eulerian-FSI* show reasonable agreement on the finest mesh levels, see Figure 1.9 and 1.10 as well as Table III and IV in [F3]. This shows that the reduced two-dimensional model yields (up to discretisation errors) the same results as the full three-dimensional model, and moreover, that a fluid-rigid body system can capture the pre-contact dynamics in this example as well as a full FSI model. The ALE approaches breaks down, when the ball reaches a distance $\text{dist}(\Gamma, \Gamma_{\text{bottom}}) \approx 1.1cm$ to the bottom.

Compared to the experimental values we observe a maximum relative deviation of 5.1% in the velocity functional v_* for the PTFE6 examples, and a smaller deviation of 0.6 resp. 2.1%

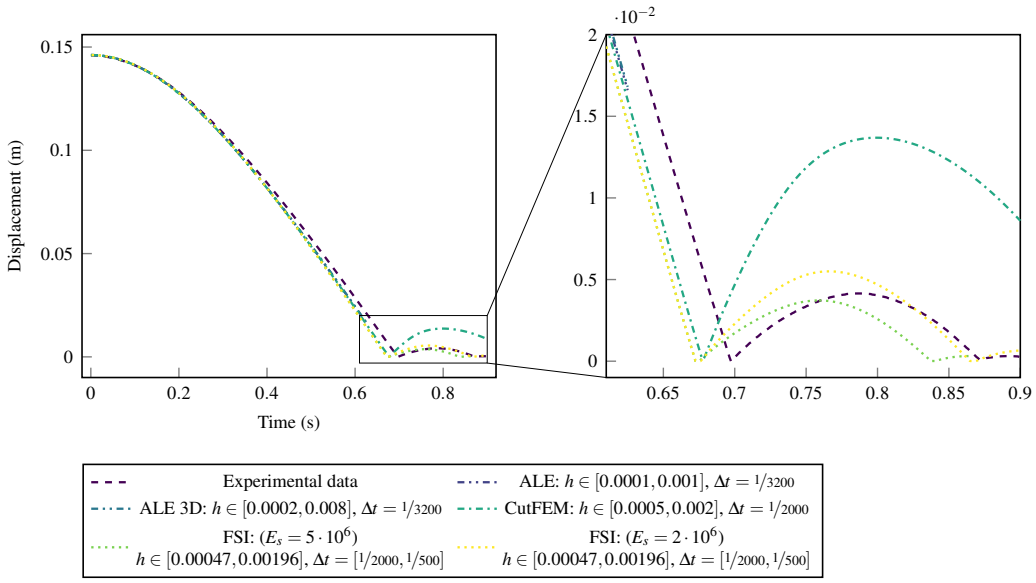


Figure 1.10. The distance between the bottom of the ball and the bottom of the tank: Experimental and numerical results for the Rubber22 set-up (Figure taken from [F3] with permission from *AIP Publishing*).

for the rubber example. Taking into account that these results ignore the 2mm- (PTFE6) resp. 0.75mm-deflection (Rubber22) from the z-axis observed in the experiment, we consider this to be acceptable, in particular since the results for rubber -where the horizontal effects are smaller- are much closer.

Contact Dynamics

The contact dynamics can only be reproduced by the *CutFEM-RB* and the *Eulerian-FSI* approach. While the contact dynamics are captured quite well with the full FSI model, it turns out that the *CutFEM-RB*-scheme is quite sensitive with respect to the artificial contact parameters. In [F3], these were chosen in such a way that the contact dynamics in the PTFE6 example are accurately reproduced on the finest mesh level. When applying the same parameters in the Rubber22 example the resulting rebound is, however, three times larger compared to the physical one, see Figure 1.10.

While the settling process is almost independent of the elasticity modulus E_s, the rebound height of the ball depends significantly on it. In the rubber case, where E_s is not known exactly, the experimental rebound height lies between the simulation results for $E_s = 2 \cdot 10^6$ and $E_s = 5 \cdot 10^6$ obtained with the *Eulerian FSI*-model, see Figure 1.10.

Three-Dimensional Computations Including Rotational Effects

Finally, a number of three-dimensional ALE simulations with random perturbations of the initial position, the initial velocity and rotational velocity of the ball are carried out in [F3], in order to investigate possible reasons for the horizontal deflections in the experiment. The rigid solids are further deflected from the center compared to their initial position,

which shows that small fluctuations during the release process can indeed explain the small horizontal displacement observed during the experiments.

2 Discretisation of Interface and Contact Problems

This section gives a summary of three of my contributions for the finite element discretisation of interface problems as well as a boundary element formulation for a Signorini contact problem using Nitsche's method.

For an accurate discretisation of interface problems the position of the interface $\Gamma(t)$ needs to be resolved within the spatial discretisation. If a domain $\Omega(t)$ is time-dependent such as $\Omega_f(t)$ and $\Omega_s(t)$ in the previous section, this means that new meshes would need to be created in each time step. *Meshing* is, however, not straight-forward, as (i) the construction of a new mesh at each time-step $t_i, i = 1, .., N$ might be costly by itself and (ii) a projection of the solution from one mesh to another needs to be computed, which is both costly and may introduce additional error contributions. Several techniques have been developed to circumvent these issues, as described in the following.

Let us first consider the finite element discretisation of a simple elliptic model problem with piecewise constant coefficient ν on two subdomains $\nu|_{\Omega_i} := \nu_i \, (i = 1, 2)$, $\nu_1 \neq \nu_2$:

$$-\nabla \cdot (\nu_i \nabla u) = f \text{ on } \Omega_i, \quad i = 1, 2, \tag{2.1}$$

$$[u] = 0, \ [\nu \partial_n u] = 0 \text{ on } \Gamma, \tag{2.2}$$

$$u = 0, \text{ on } \partial\Omega. \tag{2.3}$$

This model problem is considered in the contribution [F4]. The interface conditions (2.2) resemble the continuity of velocities and the balance of forces in the FSI system (1.4). The solution to (2.1)-(2.3) is continuous, its normal derivative is, however, discontinuous across the interface Γ.

If a discretisation does not resolve the interface, for example using a uniform Cartesian mesh, the convergence order of a standard finite element discretisation is limited to $\mathcal{O}(h^{1/2})$ in the energy norm and $\mathcal{O}(h)$ in the L^2-norm, regardless of the polynomial degree of the finite element space, see [8]:

$$||u - u_h||_\Omega = \mathcal{O}(h), \quad ||\nabla(u - u_h)||_\Omega = \mathcal{O}(h^{1/2}).$$

Strategies to improve the rate of convergence can be categorised in two classes, namely *fitted* and *unfitted* finite element techniques. In an *unfitted* finite element approach the mesh does not consider the position of the interface. To improve the accuracy, the finite element space is enriched in the interface cells by discrete functions functions that are non-smooth across Γ. Two of the most prominent examples are the *Extended Finite Element Method* (XFEM) [72] and the *Generalised Finite Element Method* (GFEM) [9].

A concentionally different *unfitted* finite element approach has been introduced by Hansbo & Hansbo in [56] called *Cut Finite Elements*. Here two overlapping finite element meshes corresponding to Ω_1 and Ω_2 are introduced and the corresponding finite element solutions

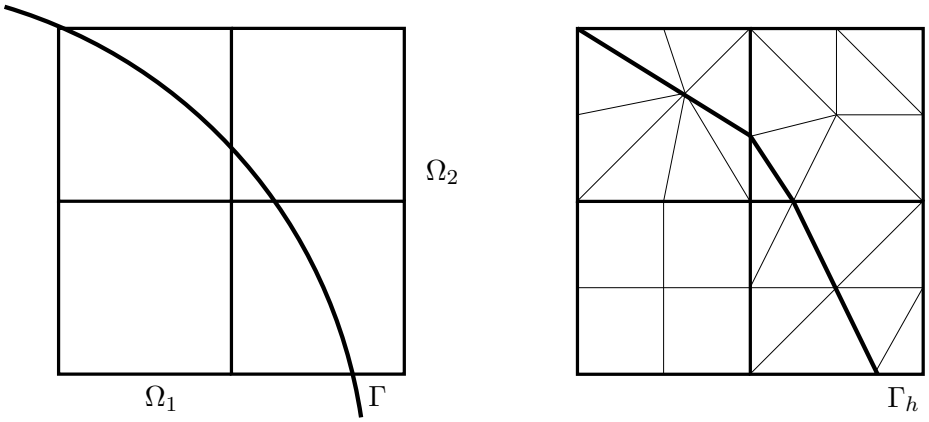


Figure 2.1. *Left: Unfitted* coarse triangulation of the domain Ω. *Right:* Subdivision of the coarse elements such that the interface $\Gamma(t)$ is resolved in a linear approximation by the discrete interface Γ_h (Figure taken from [F1] with permission from *Elsevier*).

u_1 and u_2 are coupled weakly by means of Nitsche's method. An elegant mathematical theory can be built for this approach allowing for arbitrary-order convergence estimates for different kind of interface problems [19, 21, 66]. In fact, Areias & Belytschko observed that this approach is equivalent to a discontinuous variant of XFEM [7]. In Section 2.2 resp. [F5] we will analyse time-stepping schemes for the Stokes equations on time-dependent domain based on *Cut Finite Elements*.

In a *fitted* finite element approach, on the other hand, the interface is resolved by means of the finite element mesh. In the case of moving interfaces, it is attractive to adjust the mesh only locally near the interface to avoid global changes in the finite element spaces. A simple approach in this category has been proposed by the author and Richter in [40] called *locally modified finite element method* (LMFEM). The idea is to use an *unfitted* coarse mesh consisting of quadrilaterals that does not consider the position of the interface $\Gamma(t)$. This coarse mesh (called *patch mesh* in the following) is further refined to resolve the interface in at least a linear approximation on the finer level, by dividing each coarse element either into four quadrilaterals or eight triangles. This is illustrated in Figure 2.1.

Using a combination of linear and bilinear finite elements on the so constructed sub-triangulation, a first-order finite element method results with optimal convergence orders

$$||u - u_h||_\Omega = \mathcal{O}(h^2), \quad ||\nabla(u - u_h)||_\Omega = \mathcal{O}(h).$$

However, even if higher-order polynomials are used, the convergence is still limited to first order (in the energy norm) due to the linear approximation of the interface, see Figure 2.1. In the following subsection resp. [F4], we develop a second-order approximation based on an isoparametric mapping and (where possible) a second-order approximation of the interface.

The (first-order) *locally modified finite element method* has been used by the author and co-workers [38, 43–45], and by Langer & Yang [63] for fluid-structure interaction (FSI) problems, including the transition between FSI and solid-solid contact [42, F1]. Holm et al. [60] and Gangl & Langer [50] used a corresponding approach based on triangular patches, the latter work being motivated by a topology optimisation problem. A pressure stabilisation technique for flow problems has been developed in [39] and a suitable (second-order) time

discretisation scheme in [41]. Details on the implementation in deal.ii and the corresponding source code have been published in [46, 48]. Extensions to three space dimensions have been developed by Langer & Yang [63] and by Höllbacher & Wittum [61].

For further *fitted* finite element methods, we refer to [8, 10, 15, 33, 34]. Some works are similar to the *locally modified finite element method* in the sense that only mesh elements close to the interface are altered [14, 94].

2.1 A Second-Order Locally Modified Finite Element Method [F4]

In the article [F4] with the collaborators G. Judakova and T. Richter, a second-order LMFEM is developed. The method is based on the same *unfitted* quadrilateral patch mesh used in the first-order variant [40] and a similar sub-division into triangular and rectangular sub-cells (cf. Figure 2.1). A combination of second-order (P_2 and Q_2) finite elements is used on these to define the discrete finite element space. A second-order interface approximation is constructed by moving all degrees of freedom that correspond to interface lines to the position of the continuous interface $\Gamma(t)$, in combination with an isoparametric mapping of order two, see Figure 2.2, left sketch. Unfortunately, this construction fails in certain pathological cases, as curved edges that correspond to a quadratic interface approximation might intersect other edges, which leads to degenerated sub-elements, see Figure 2.3 on the right. This difficulty is not present in the case of a linear interface approximation as in [40], where a maximum angle condition could be ensured in each cell, independently of the way the interface cuts a patch.

If such a case appears, we refrain from moving the corresponding degree of freedom to the interface and stick to a linear interface approximation in the affected patch. The maximum number of patches with a linear interface approximation (in the following denoted by n_l) remains bounded under refinement independently of the cell size $h \leq h_0$, if we assume that the curvature of the interface is bounded. We will see in the error analysis below that the additional error corresponding to a linear interface approximation in n_l patches does not disturb the second-order accuracy, if n_l is bounded independently of h.

Finite element space and discrete formulation The isoparametric finite element space $V_h \subset H_0^1(\Omega)$ is defined by

$$V_h := \{\varphi \in C(\Omega) \,|\, (\varphi \circ \xi_K^{-1}) \in \mathcal{P}_K^2(\hat{K}) \text{ for } K \in \mathcal{T}_h\}, \tag{2.4}$$

where

$$\mathcal{P}_K^2(\hat{K}) := \begin{cases} Q_2(\hat{K}), & K \text{ is a quadrilateral,} \\ P_2(\hat{K}), & K \text{ is a triangle,} \end{cases}$$

and ξ_K is a transformation from the reference element $\hat{K}$ to K. The map ξ_K resolves the interface with second order in all but n_l elements, where the approximation is only linear. Note that P_2 or Q_2 polynomials are used for the trial functions $\varphi \circ \xi_K^{-1}$, independently of the order of the interface approximation. The space V_h is continuous, as the restriction of a function in $Q_2(\hat{K})$ to a line $e \subset \partial K$ is in $P_2(\hat{K})$.

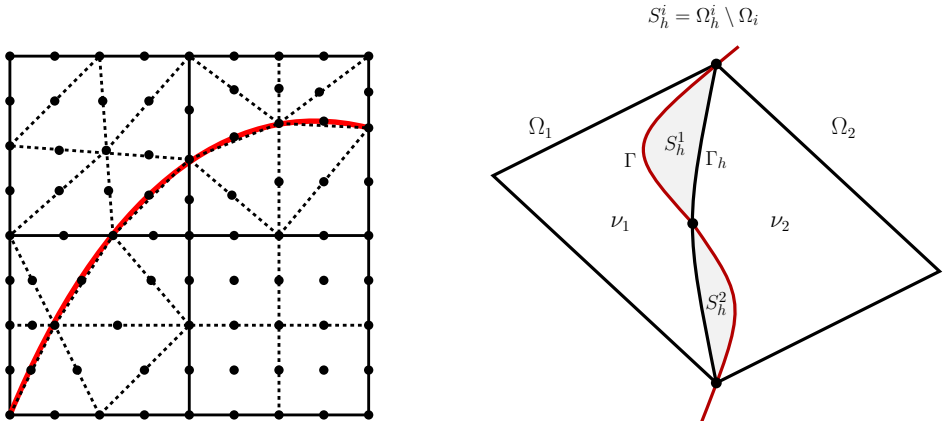


Figure 2.2. *Left*: Second-order approximation on a mesh consisting of four patches, three of which are cut by the interface. *Right*: Mismatch between Ω^i and Ω_h^i, $i = 1, 2$ at two elements along the curved interface.

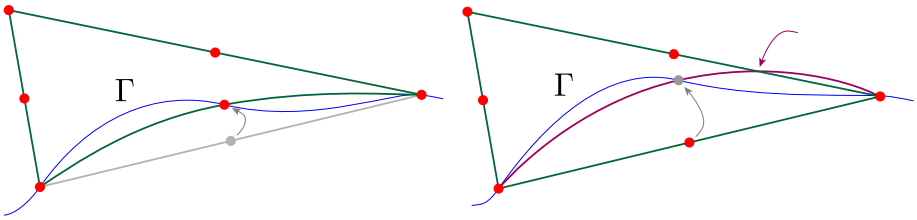


Figure 2.3. The interface Γ is approximated quadratically on all triangles. To achieve this, the midpoint of the edge that corresponds to the interface is pulled onto the curve. *Left*: This is a valid configuration where a quadratic approximation is possible. *Right*: In some configurations a fully quadratic interface approximation would result in a degenerate element with an interface that is leaving the triangle (see the mark on the upper edge). Such triangles are approximated linearly, which results in $n_l > 0$.

We consider the following discrete variational formulation: *Find* $u_h \in V_h$ *such that*

$$a_h(u_h, \phi_h) = (f_h, \phi_h)_\Omega \quad \forall \phi_h \in V_h, \tag{2.5}$$

where we set $f_h|_{\Omega_h^i} := f_i, i = 1, 2$ and f_i is a smooth extension of $f|_{\Omega_i}$ to Ω_h^i. The bilinear form is given by

$$a_h(u_h, \phi_h) := (\nu_h \nabla u_h, \nabla \phi_h)_\Omega, \qquad \text{where } \nu_h|_{\Omega_h^i} = \nu_i \, (i = 1, 2).$$

2.1.1 Error Analysis

In the error analysis the mismatch between discrete and continuous bilinear form is the predominant issue. The continuous solution u is regular in Ω_1 and Ω_2, while its normal derivative has a jump across Γ. Discrete functions can only have irregularities at the boundaries of cells ∂K, which means that -if Γ_h is sufficiently close to Γ- a discrete function can at best resemble a similar discontinuity across the discrete interface Γ_h instead of Γ.

As the complete analysis is relatively technical, we give here only a few hints for estimates that are needed repeatedly throughout the analysis, while the reader is referred to the article [F4] for further details. Therefore, let us denote the mismatch between Ω_h^i and Ω^i by S_h^i, $i = 1, 2$ (see Figure 2.2, right sketch)

$$S_h^1 := \Omega_h^1 \setminus \Omega_1 = \Omega_2 \setminus \Omega_h^2, \qquad S_h^2 := \Omega_h^2 \setminus \Omega_2 = \Omega_1 \setminus \Omega_h^1, \qquad S_h = S_h^1 \cup S_h^2.$$

Further, we split S_h into parts $S_{h,\text{lin}}$ with a linear approximation of the interface and parts $S_{h,\text{qu}}$ with a quadratic approximation.

Assuming that the interface Γ is of class C^{s+1}, the distance of the discrete and continuous interface is bounded by

$$\text{dist } (\Gamma_h \cap T; \Gamma \cap T) \leq c h_P^{s+1}, \tag{2.6}$$

where $s \in \{1, 2\}$ is the local approximation order of the interface and h_P is the size of the (regular) patch mesh. If the number of elements with a linear interface approximation is bounded by n_l, this implies that the areas of the regions $S_{h,\text{lin}}$ and $S_{h,\text{qu}}$ are bounded by

$$|S_{h,\text{lin}}| \leq c n_l h_P^3, \quad |S_{h,\text{qu}}| \leq c h_P^3, \tag{2.7}$$

see [77, F4]. This is used frequently in the error analysis in combination with a Hölder inequality to estimate terms of the form

$$\|\nabla u\|_{S_{h,\text{lin}}} \leq |S_{h,\text{lin}}|^{\frac{1}{2}-\frac{1}{p}} \|\nabla u\|_{L^p(S_{h,\text{lin}})} \leq c n_l h_P^{\frac{3}{2}-\frac{3}{p}} \|\nabla u\|_{L^p(S_{h,\text{lin}})}, \quad p > 2,$$

that appear for example when subtracting the continuous from the discrete bilinear form. Next, application of the Sobolev inequality $\|\nabla u\|_{L^p(\Omega_i)} \leq c p^{\frac{1}{2}} \|\nabla u\|_{H^1(\Omega_i)}$ (see [90]) yields for arbitrary $p \in [2, \infty)$

$$\|\nabla u\|_{S_{h,\text{lin}}} \leq c n_l p^{1/2} h_P^{\frac{3}{2}-\frac{3}{p}} \|u\|_{H^2(\Omega_1 \cup \Omega_2)}.$$

As the minimum of $J(p) := p^{1/2} h_P^{-\frac{3}{p}}$ is bounded above by $c|\ln(h_P)|^{1/2}$ (see [F4]), we obtain

$$\|\nabla u\|_{S_{h,\text{lin}}} \leq c n_l h_P^{\frac{3}{2}} |\ln(h_P)|^{1/2} \|u\|_{H^2(\Omega)}.$$

This gives rise to a logarithmic factor in h that will appear in some of the error estimates. To state the main result, we introduce discrete energy norm

$$|||u - u_h||| := \left(\|\nu_1 \nabla(\tilde{u}_1 - u_h^1)\|_{\Omega_h^1}^2 + \|\nu_2 \nabla(\tilde{u}_2 - u_h^2)\|_{\Omega_h^2}^2 \right)^{1/2},$$

where $\tilde{u}_i$ are smooth extensions of $u_i = u|_{\Omega_i}$ to Ω_h^i and $u_h^i := u_h|_{\Omega_h^i}$.

Theorem 2.1 (A priori estimate - Theorem 6 in [F4]). *Let $\Omega \subset \mathbb{R}^2$ be a convex domain with polygonal boundary, which is resolved (exactly) by the family of triangulations $\mathcal{T}_h$. We assume a splitting $\Omega = \Omega_1 \cup \Gamma \cup \Omega_2$, where Γ is a smooth interface with C^3-parametrisation and that the solution $u \in H_0^1(\Omega)$ to (2.1)-(2.3) belongs to $H^3(\Omega_1 \cup \Omega_2)$. Moreover, we denote*

by n_l the maximum number of elements $K \in \mathcal{T}_h$, where the interface is approximated linearly. For the locally modified finite element solution $u_h \in V_h$ to (2.5) *it holds*

$$|||u - u_h||| \leq \Big(c_l n_l^{\frac{1}{2}} |\ln(h_P)|^{1/2} + c_q\Big) h_P^2 \|u\|_{H^3(\Omega_1 \cup \Omega_2)}, \tag{2.8}$$

$$\|u - u_h\|_\Omega \leq \Big(c_l n_l |\ln(h_P)|^{1/2} + c_q\Big) h_P^3 \|u\|_{H^3(\Omega_1 \cup \Omega_2)}, \tag{2.9}$$

where c_l and c_q are generic constants that correspond to patches with a linear and a quadratic interface approximation, respectively. For $u \in W^{2,\infty}(\Omega_1 \cup \Omega_2)$ we have further

$$|||u - u_h||| \leq \Big(c_l n_l^{\frac{1}{2}} + c_q\Big) h_P^2 \left(\|u\|_{H^3(\Omega_1 \cup \Omega_2)} + \|u\|_{W^{2,\infty}(\Omega_1 \cup \Omega_2)}\right). \tag{2.10}$$

This means that we obtain the optimal error bounds ($\mathcal{O}(h_P^2)$ in the discrete energy norm and $\mathcal{O}(h_P^3)$ in the L^2-norm), when a quadratic interface approximation is possible in all elements. If a bounded number of elements requires a linear interface approximation, the convergence is reduced slightly by a logarithmic factor in both norms. In the discrete energy norm, the optimal convergence is recovered when additional regularity of the solution is assumed.

2.1.2 Numerical Results

The article [F4] presents two series of numerical examples on different mesh levels that confirm the error analysis. In each example the interface is moved gradually by $\delta \in [0, h_0]$ on each mesh level, such that all possible types of anistropies are present. While in the first series of examples a fully quadratic interface approximation is possible in all cells, a maximum of $n_l = 16$ patches with a linear interface approximation is required in the second example over all mesh levels.

With a full second-order interface approximation, we observe fully quadratic resp. cubic convergence in the discrete energy norm and the L^2-norm error, as shown in Theorem 2.1. In contrast, if a linear approximation is required in some patches, the contribution from the linearly approximated cells becomes visible and may dominate the error. However, in any case fully quadratic convergence was observed in the discrete energy norm (due to (2.10)), while the estimated order of convergence in the L^2-norm was slightly below cubic convergence (around 2.92), due to the logarithmic term in (2.9).

2.2 Eulerian Time-Stepping Schemes for Unfitted Finite Element Discretisation on Moving Domains ([F5], [F6])

In the articles [F5] with the collaborators E. Burman and A. Massing and [F6] with M.K. Singh we investigate Eulerian time-stepping schemes for the non-stationary Stokes equations resp. the heat equation on moving domains $\Omega(t)$. These considerations are motivated by certain interface problems and by flow problems on moving domains, where it is advantageous to stick to the Eulerian coordinate framework, such as fluid-structure interactions with large displacements, see Chapter 1, or multi-phase flows. For spatial discretisation, we use a *Cut Finite Element* method.

While the space discretisation of interface problems is extensively discussed in literature, much less works can be found for a proper time discretisation of problems with moving interfaces. In the case of moving domains, standard time discretisation based on the method of lines is not applicable in a straight-forward way. The reason is that the variables $u_h(t_k)$ are defined on different domains $\Omega(t_k)$ in each time-step.

As an example let us consider a first-order *finite difference* discretisation of the time derivative within a finite element formulation, which is used in many popular time-stepping schemes such as the *backward Euler* and the *Crank-Nicolson* method:

$$(\partial_t u_h(t_n), \phi_h^n)_{\Omega(t_n)} \overset{?}{\approx} \frac{1}{\Delta t}(u_h(t_n) - u_h(t_{n-1}), \phi_h^n)_{\Omega(t_n)}.$$

Note that $u_h(t_{n-1})$ is defined on $\Omega(t_{n-1})$, but is needed on $\Omega(t_n)$.

One solution to this dilemma are so-called characteristic-based approaches [58]. Similar time-stepping schemes result when applying the ALE method only locally within one time step and projecting the system back to the original reference frame after each step [26], or based on Galerkin time discretisations with modified Galerkin spaces [41]. The disadvantage of these approaches is the need for a projection between the domains $\Omega(t_{n-1})$ and $\Omega(t_n)$ that needs to be computed in each step.

Another possibility consists of space-time approaches [57, 64], where a $d+1$-dimensional domain is discretised if $\Omega(t) \subset \mathbb{R}^d$. The computational requirements of these approaches can, however, easily exceed the available computational resources, in particular within complex three-dimensional applications. Moreover, the implementation of higher-dimensional discretisations and accurate quadrature formulas pose additional challenges.

A simpler approach has been proposed in the dissertation of Schott [83] and by Lehrenfeld & Olshanskii [65]. Here, the idea is to define extensions of the solution $\mathbf{u}_h(t_{n-1})$ from the previous time-step to a domain that spans at least $\Omega(t_n)$. On the finite element level these extensions can be incorporated implicitly in the time-stepping scheme to a sufficiently large domain $\Omega_\delta(t_{n-1}) \supset \Omega(t_n)$ by using so-called *ghost penalty* stabilisations [16]. These techniques have originally been proposed to extend the coercivity of elliptic bilinear forms from the physical to the computational domain in the context of CutFEM or fictitious domain approaches [16].

While Schott used such an extension explicitly *after* each time step to define values for $\mathbf{u}_h(t_{n-1})$ in mesh nodes lying in $\Omega(t_n) \setminus \Omega(t_{n-1})$, Lehrenfeld & Olshanskii included the extension operator implicitly *within* each time step by solving a combined discrete system including the extension operator on the larger computational domain $\Omega_\delta(t_n)$. For the latter approach a complete analysis could be given for the corresponding backward Euler time discretisation of the heat equation, showing first-order convergence in time in the spatial energy norm [65]. Moreover, the authors gave hints on how to transfer the argumentation to a backward difference scheme (BDF(2)), which results in second-order convergence.

In Section 2.2.1 resp. [F5], we extend the method of Lehrenfeld & Olshanskii to the non-stationary Stokes equations, using both the BDF(1) (backward Euler) and the BDF(2) variant of the approach. In Section 2.2.2 resp. [F6], we investigate a Crank-Nicolson variant of the approach applied to the heat equation.

Equations

In [F5] we consider the non-stationary Stokes equations with homogeneous Dirichlet boundary conditions on a moving domain $\Omega(t) \subset \mathbb{R}^d, d = 2, 3$ for $t \in I = [0, t_{\text{fin}}]$

$$\begin{aligned} \partial_t \mathbf{u} - \Delta \mathbf{u} + \nabla p = \mathbf{f}, \qquad & \text{div } \mathbf{u} = 0 \qquad \text{in } \Omega(t), \\ \mathbf{u} = 0 \quad \text{on } \partial\Omega(t), \qquad & \mathbf{u}(x,0) = \mathbf{u}^0(x) \text{ in } \Omega(0). \end{aligned} \tag{2.11}$$

The model problem in [F6] is the following problem governed by the heat equation:

$$\partial_t u - \Delta u = f \quad \text{in } \Omega(t), \qquad u = 0 \quad \text{on } \partial\Omega(t), \qquad u(x,0) = u_0(x) \quad \text{in } \Omega(0). \tag{2.12}$$

In both cases the domain movement $\mathbf{T} : \Omega(0) \to \Omega(t)$ is assumed to be a sufficiently smooth diffeomorphism. For details on the assumptions, the corresponding variational formulations and the well-posedness of the continuous systems on the moving domains, we refer to [F5] and [F6], respectively.

Below we will use the following notations for the space-time domain Q and the space-time Bochner norms

$$Q := \bigcup_{t \in I} \Omega(t) \times \{t\}, \qquad \|u\|_{\infty,m} := \|u\|_{L^\infty(I, H^m(\Omega(t)))},$$

where $m \in \mathbb{N} \cup \{0\}$ and $H^0(\Omega(t)) := L^2(\Omega(t))$.

Discretisation

For discretisation in time, we split the time interval of interest $I = [0, t_{\text{fin}}]$ into time intervals $I_n = (t_{n-1}, t_n]$ of uniform step size $\Delta t = t_n - t_{n-1}$

$$0 = t_0 < t_1 < ... < t_N = t_{\text{fin}}.$$

Following Lehrenfeld & Olshanskii [65] we extend the domain $\Omega^n := \Omega(t_n)$ in each time point t_n by a strip of size δ to a domain Ω^n_δ, which is chosen large enough such that

$$\bigcup_{i=0}^{s} \Omega^{n+i} \subset \Omega^n_\delta, \tag{2.13}$$

see the left part of Figure 2.4. For the BDF schemes in [F5] the number s is the order of the BDF formula, i.e. $s = 1$ for backward Euler and $s = 2$ for the BDF(2) variant. For the Crank-Nicolson scheme in [F6], we set $s = 1$. We choose δ such that

$$s w_{\max} \Delta t \leq \delta \leq c s w_{\max} \Delta t, \quad \text{where } w_{\max} := \max_{t \in I, x \in \partial\Omega(0)} |\partial_t \mathbf{T}(x,t) \cdot \mathbf{n}| \tag{2.14}$$

is a bound for the maximum velocity of the boundary movement in normal direction.

In the spirit of the *Cut Finite Element* approach the discretisation is based on a fixed background mesh that covers $\Omega_\delta(t)$ for all times $t \in [0, t_{\text{fin}}]$. In each time-step the set of cells $K \in \mathcal{T}^n_{h,\delta}$ that lie at least partially in Ω^n_δ is used to define the finite element spaces, see 2.4, right sketch:

$$\mathcal{V}^n_h := \{\mathbf{v} \in C(\Omega^n_{h,\delta}),\, \mathbf{v}|_K \in P_m(K) \, \forall K \in \mathcal{T}^n_{h,\delta}\} \tag{2.15}$$

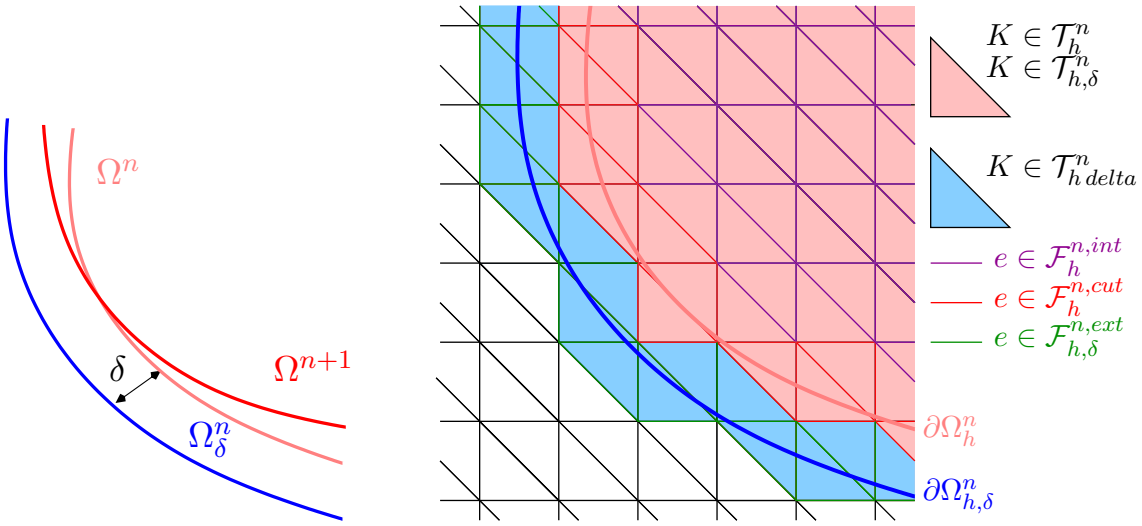


Figure 2.4. *Left:* Illustration of Ω^n_δ for $s = 1$. *Right:* Illustration of the discretisation and faces.

For the Stokes problem we use continuous equal-order finite elements of degree $m \geq 1$ for all variables and define additionally for the discretisation of pressure

$$\mathcal{L}^n_h := \{q \in C(\Omega^n_h),\, q|_K \in P_m(K)\,\forall K \in \mathcal{T}^n_h\}, \quad \mathcal{L}^n_{h,0} := \mathcal{L}^n_h \cap \mathcal{L}_0(t_n). \tag{2.16}$$

The finite element space $\mathcal{V}^n_h$ will be used for the discretisation of the heat equation (2.12) and for the velocities $\mathbf{u}$ in the Stokes problem (2.11) ; the space $\mathcal{L}^n_h$ for the corresponding pressure p, which is only needed in Ω^n_h.

Finally, we introduce the backward difference approximations $D^{(s)}_t$ of order $s = 1, 2$ for the time derivative ∂_t

$$D^{(1)}_t \mathbf{u}^n_h := \frac{1}{\Delta t}(\mathbf{u}^n_h - \mathbf{u}^{n-1}_h), \qquad D^{(2)}_t \mathbf{u}^n_h := \frac{1}{2\Delta t}(3\mathbf{u}^n_h - 4\mathbf{u}^{n-1}_h + \mathbf{u}^{n-2}_h).$$

2.2.1 Implicitly Extended Backward Difference Schemes for the Stokes Equations [F5]

For the BDF(s) approximation of the Stokes equations (2.11), we define the following time-stepping scheme in [F5] ($s = 1, 2$): *For $k = s, \ldots, n$ find* $\mathbf{u}^n_h \in \mathcal{V}^n_h, p^n_h \in \mathcal{L}^n_{h,0}$ such that

$$(D^{(s)}_t \mathbf{u}^n_h, \mathbf{v}_h)_{\Omega^n} + \mathcal{A}^n_h(\mathbf{u}^n_h, p^n_h; \mathbf{v}_h, q_h) = (\mathbf{f}, \mathbf{v}_h)_{\Omega^n} \quad \forall \mathbf{v}_h \in \mathcal{V}^n_h, q_h \in \mathcal{L}^n_h. \tag{2.17}$$

The bilinear form $\mathcal{A}^n_h$ is defined by

$$\mathcal{A}^n_h(\mathbf{u}_h, p_h; \mathbf{v}_h, q_h) := \mathcal{A}^n_S(\mathbf{u}_h, p_h; \mathbf{v}_h, q_h) + a^n_D(\mathbf{u}_h, p_h; \mathbf{v}_h, q_h) + \gamma_g g^n_h(\mathbf{u}_h, \mathbf{v}_h) + \gamma_p s^n_h(p_h, q_h).$$

It includes the Stokes part

$$\mathcal{A}^n_S(\mathbf{u}, p; \mathbf{v}, q) := (\nabla \mathbf{u}, \nabla \mathbf{v})_{\Omega^n} - (p, \operatorname{div} \mathbf{v})_{\Omega^n} + (\operatorname{div} \mathbf{u}, q)_{\Omega^n} \tag{2.18}$$

and Nitsche terms to weakly impose the Dirichlet boundary conditions

$$a_D^n(\mathbf{u}_h, p_h; \mathbf{v}_h, q_h) := -(\partial_n \mathbf{u}_h - p_h \mathbf{n}, \mathbf{v}_h)_{\partial\Omega^n} - (\mathbf{u}_h, \partial_n \mathbf{v}_h + q_h \mathbf{n})_{\partial\Omega^n} + \frac{\gamma_D}{h} (\mathbf{u}_h, \mathbf{v}_h)_{\partial\Omega^n} .$$

For the pressure stabilisation $s_h^n(\cdot,\cdot)$ we use the *Continuous Interior Penalty* method (CIP, [18]) in this work, although different pressure stabilisations are possible, see [F5].

The *ghost-penalty* term $g_h^n(\cdot,\cdot)$ is used to extend $\mathbf{u}_h^n$ to the region Ω_δ^n, where it might be needed in the following time steps. In [F5] we consider three different variants for g_h^n, see also [65]. Here, we only mention the "classical" fully consistent variant [16, 20, 71], where jumps of derivatives over element edges are penalised

$$g_h^{n,\text{jump}}(\mathbf{u}, \mathbf{v}) := \sum_{e \in \mathcal{F}_{h,\delta}^{n,g}} \sum_{k=1}^{m} h^{2k-1} ([\![\partial_n^k \mathbf{u}]\!], [\![\partial_n^k \mathbf{v}]\!])_e. \tag{2.19}$$

The required set of edges is defined as $\mathcal{F}_{h,\delta}^{n,g} = \mathcal{F}_h^{n,cut} \cup \mathcal{F}_h^{n,ext}$, see the illustration in Figure 2.4 on the right.

As initial value, we set (formally) $u_h^0 := E^1 u^0$ in (2.17) (which corresponds to a L^2-projection onto the discrete spaces), where E^1 denotes a smooth extension from Ω^0 to $\Omega_\delta^0 \supset \Omega^1$. For BDF(2) the iterate u_h^1 can be initialised with one BDF(1) step without compromising the (second-order) accuracy. For the well-posedness of the discrete system (2.17) we refer to [F5].

Stablity Results

To simplify the presentation of the analysis, we neglect geometry approximation errors related to the approximation of curved boundaries in [F5], which have been analysed in detail for parabolic problems in [65]. We introduce the triple norm

$$|||\mathbf{u}_h|||_{h,n} := \left(\|\nabla \mathbf{u}_h^n\|_{\Omega^n}^2 + \gamma_g g_h^n(\mathbf{u}_h^n, \mathbf{u}_h^n) + \frac{\gamma_D}{h} \|\mathbf{u}_h^n\|_{\partial\Omega^n}^2 \right)^{1/2} .$$

The error analysis in the next paragraph is based on the following stability result.

Theorem 2.2 (Corollary 4.5 and Lemma 4.4 in [F5])**.** *Let $\mathbf{u}_h = (\mathbf{u}_h^k)_{k=1}^N, p_h = (p_h^k)_{k=1}^N$ be the solutions of* (2.17) *and let γ_g and γ_D be sufficiently large. Under regularity assumptions stated in [F5], it holds for $s = 1$ and $n \geq 1$ that*

$$\begin{aligned} \|\mathbf{u}_h^n\|_{\Omega^n}^2 + \Delta t \sum_{k=1}^{n} \Big(|||\mathbf{u}_h^k|||_{h,k}^2 + \frac{1}{\Delta t} \|\mathbf{u}_h^k - \mathbf{u}_h^{k-1}\|_{\Omega^k}^2 \\ + \gamma_p s_h^k(p_h^k, p_h^k) + \min\{h^2, \Delta t\} \|\nabla p_h^k\|_{\Omega^k}^2 + \Delta t \|p_h^n\|_{\Omega^n}^2 \Big) \\ \leq \exp(c_1(w_{\max}) t_n) \left(c \|\mathbf{u}^0\|_{\Omega^0} + 2 t_n \|\mathbf{f}\|_{\infty,0}^2 \right). \end{aligned} \tag{2.20}$$

For $s = 2$, we have

$$\begin{aligned} \|\mathbf{u}_h^n\|_{\Omega^n}^2 + \Delta t \sum_{k=1}^{n} \left(|||\mathbf{u}_h^k|||_{h,k}^2 + \gamma_p s_h^k(p_h^k, p_h^k) + \min\{h^2, \Delta t^2\} \|\nabla p_h^k\|_{\Omega^k}^2 \right) \\ \leq \exp(c_1(w_{\max}) t_n) \left(c \|\mathbf{u}^0\|_{\Omega^0} + 2 t_n \|\mathbf{f}\|_{\infty,0}^2 \right). \end{aligned} \tag{2.21}$$

Compared to the case of fixed domains $\Omega(t) = \Omega$, Theorem 2.2 gives the same stability result for the velocities. For the H^1-seminorm of pressure an optimal stability estimate results under the inverse CFL condition $\Delta t \geq ch^2$ for $s = 1$ and the stronger condition $\Delta t \geq ch$ for $s = 2$. The stability bound in the H^1-seminorm is sufficient in the error analysis in the next paragraph to show optimal-order convergence results for the velocities.

In the L^2-norm of pressure, we obtain stability for the term $\Delta t^2 \sum_{k=1}^n \|p_h^k\|_{\Omega^k}^2$ on the left-hand side of (2.20) for $s = 1$, which is by a factor Δt weaker than in the case of fixed domains. The corresponding argumentation in the case of fixed domains and fixed discretisations uses that the term $(\operatorname{div} \mathbf{u}_h^{n-1}, \xi_h^n)_{\Omega^n}$ vanishes for $\xi_h^n \in \mathcal{L}_h^n$, see for example Besier & Wollner [11]. This is not true in the case of time-dependent domains, as $\mathbf{u}_h^{n-1}$ is not discrete divergence-free with respect to Ω^n:

$$(\operatorname{div} \mathbf{u}_h^{n-1}, \xi_h^n)_{\Omega^n} \neq 0 \quad \text{for certain } \xi_h^n \in \mathcal{L}_h^n.$$

This will lead to sub-optimal error estimates in the L^2-norm of pressure in the next paragraph.

Error Estimates

The error analysis is based on Galerkin orthogonality and the stability result of Theorem 2.2. For the BDF(1) variant, we obtain:

Theorem 2.3 (Theorems 5.3 and 5.12 in [F5])**.** *Let $\mathbf{u}_h = (\mathbf{u}_h^k)_{k=1}^n, p_h = (p_h^k)_{k=1}^n$ be the discrete solution of (2.17) for $s = 1$ and $(\mathbf{u}, p)$ the continuous solution of (2.11). Further, let γ_g, γ_D and γ_p sufficiently large and $\Delta t \geq ch^2$ for some $c > 0$. Under regularity assumptions stated in [F5], it holds for the error $\mathbf{e}_u^k = \mathbf{u}^k - \mathbf{u}_h^k, e_p^k = p^k - p_h^k$ for $n \geq 1$*

$$\begin{aligned}\|\mathbf{e}_u^n\|_{\Omega^n}^2 + \sum_{k=1}^n \Big\{ \|\mathbf{e}_u^k - \mathbf{e}_u^{k-1}\|_{\Omega^k}^2 + \Delta t \left(|||\mathbf{e}_u^k|||_{h,k}^2 + h^2 \|\nabla e_p^k\|_{\Omega^k}^2 \right) \Big\} \\ \leq c \exp(ct_n) \Big(\Delta t^2 \|\partial_t^2 \mathbf{u}\|_Q^2 + h^{2m} \left(\|\mathbf{u}\|_{\infty,m+1}^2 + \|\partial_t \mathbf{u}\|_{\infty,m}^2 + \|p\|_{\infty,m}^2 \right) \Big),\end{aligned}$$

where $\mathbf{e}_u^0 := 0$. In the $L^2(L^2)$-norm of velocities, we have further

$$\begin{aligned}\Delta t \sum_{k=1}^n \|\mathbf{e}_u^k\|_{\Omega^k}^2 \\ \leq c w_{\max} \exp(ct_n) \Big(\Delta t^2 \|\partial_t^2 \mathbf{u}\|_Q^2 + h^{2m+2} \left(\|\mathbf{u}\|_{\infty,m+1}^2 + \|\partial_t \mathbf{u}\|_{\infty,m}^2 + \|p\|_{\infty,m}^2 \right) \Big).\end{aligned}$$

Both estimates are optimal in the respective norms. For BDF(2), we obtain optimal-order estimates (i.e. second-order in time) under the stronger condition $\Delta t \geq ch$, see Remarks 5.5 and 5.13 in [F5].

For the L^2-norm error of pressure we can show for $s = 1, 2$ that

$$\Delta t \sum_{k=1}^n \|e_p^k\|_{\Omega^k}^2 \leq c \exp(ct_n) \Big(\Delta t^s \|\partial_t^2 \mathbf{u}\|_Q^2 + \frac{h^{2m}}{\Delta t^s} \left(\|\mathbf{u}\|_{\infty,m+1}^2 + \|\partial_t \mathbf{u}\|_{\infty,m}^2 + \|p\|_{\infty,m}^2 \right) \Big), \tag{2.22}$$

which is suboptimal by $\mathcal{O}(\Delta t^{s/2})$ (after taking the square roots in (2.22)) compared to the optimal estimates on fixed domains.

Summary of the Numerical Results

In [F5] three-dimensional numerical results are provided for the combinations P_1-BDF(1), P_2-BDF(1) and P_2-BDF(2) using the CutFEM library [21] which is based on FEniCS [5]. Detailed convergence studies confirm the proven convergence orders in the $L^2(H^1)$- and $L^2(L^2)$-norms of velocity in all cases. In the $L^2(H^1)$- and $L^2(L^2)$-norms of pressure, first-order convergence in time is clearly visible in the BDF(1) case. This confirms the error analysis in the $L^2(H^1)$-norm, while it indicates that the $L^2(L^2)$-norm estimate (2.22) might not yet be optimal. The estimated temporal convergence orders for BDF(2) are larger than 2 in the $L^2(H^1)$-norm and larger than 1.5 in the $L^2(L^2)$-norm of pressure in all cases, which leads to similar conclusions.

2.2.2 An Implicitly Extended Crank-Nicolson Scheme for the Heat Equation [F6]

In [F6] we consider a Crank-Nicolson time-stepping scheme with implicit extensions for the heat equation (2.12) on a moving domain. A complete a priori error analysis in space and time is given with second-order convergence in time under a parabolic CFL condition.

In contrast to the BDF2 method, the Crank-Nicolson scheme induces less numerical dissipation [68] and is thus more suitable for elasticity problems or fluid flows with low Reynolds number [13, 87]. Up to now, it has been largely open, if and under what conditions a Crank-Nicolson-type scheme can be used within an Eulerian time discretisation on moving domains. We give a detailed stability and convergence analysis, assuming a parabolic CFL condition of type $\Delta t \leq ch^2$.

We consider the time and space discretisation described above on page 28. The Crank-Nicolson method applied to (2.12) writes formally

$$\frac{u^n - u^{n-1}}{\Delta t} - \frac{1}{2}(\Delta u^n + \Delta u^{n-1}) = \frac{1}{2}(f^n + f^{n-1}), \quad x \in \Omega^n. \tag{2.23}$$

As in the previous subsection, the main issue is that u^{n-1} is needed on Ω^n, while it is defined on Ω^{n-1}. We will again add implicit extension operators to define u^n on $\Omega^n_\delta \supset \Omega^{n+1}$, where it is needed in the following time step. Similarly, f^{n-1} might be undefined on $\Omega^n \setminus \Omega^{n-1}$ and needs to be extended. To cover different possibilities, we do not want to restrict the analysis in this work to a particular extension, but assume only that f^{n-1} is smoothly extended to Ω^n.

For the stability and error analysis, we assume the following CFL condition depending on the polynomial degree m of the finite element approach:

Assumption 2.4 (CFL condition)**.** *We assume the parabolic CFL condition $\Delta t \leq c_{CFL} h^2$, where c_{CFL} is an arbitrary constant for $m = 1$, while we assume c_{CFL} sufficiently small for $m > 1$.*

Discrete Variational Formulation

We consider the following discrete variational formulation for $n = 1, 2, \ldots, N$ with $\mathcal{V}_h^n$ defined in (2.15): Find $u_h^n \in \mathcal{V}_h^n$ such that

$$\left(D_t^{(1)} u_h^n, v_h\right)_{\Omega^n} + \mathcal{A}(u_h^n, u_h^{n-1}; v_h) = (f_h^{n-\frac{1}{2}}, v_h), \quad \forall v_h \in \mathcal{V}_h^n, \tag{2.24}$$

where

$$\mathcal{A}(u_h^n, u_h^{n-1}; v_h) := \frac{1}{2}a_h^n(u_h^n, v_h) + \frac{1}{2}a_h^n(u_h^{n-1}, v_h) + \frac{\gamma_D}{h}(u_h^n, v_h)_{\partial\Omega^n} + \gamma_g g_h^n(u_h^n, v_h) \tag{2.25}$$

and

$$a_h^n(u_h^k, v_h) = (\nabla u_h^k, \nabla v_h)_{\Omega^n} - (\partial_n u_h^k, v_h)_{\partial\Omega^n}, \quad f^{n-\frac{1}{2}} = \frac{f^n + f^{n-1}}{2}.$$

As in the previous subsection, we assume that all integrals in (2.25) are evaluated exactly and refer to [65] for the consideration of additional quadrature errors. To incorporate the initial condition, we set $u_h^0 := E^1 u^0$ in (2.25) for $n = 1$, which corresponds to a Ritz projection of the initial value. For the ghost penalty $g_h^n(\cdot,\cdot)$, we consider again the classical variant given in (2.19).

Stability and Error Analysis

Compared to the BDF schemes in [65] and [F5], the analysis is complicated by the fact that not only the discrete function u_h^{n-1}, but also its gradient ∇u_h^{n-1} appear on Ω^n in the n-th time step, whereas bounds are only available for $\|u_h^{n-1}\|_{\Omega^{n-1}}$ and $\|\nabla u_h^{n-1}\|_{\Omega^{n-1}}$ from the previous time step.

Key to the stability proof is the following lemma, which is proven by means of several auxiliary estimates:

Lemma 2.5 (Lemma 3.2 in [F6])**.** *Under the CFL condition (Assumption 2.4) and a technical assumption on the finite element mesh stated in [F6] it holds for $v_h^n \in \mathcal{V}_h^n$, $v_h^{n-1} \in \mathcal{V}_h^{n-1}$ that*

$$\begin{aligned}\Delta t\|\nabla v_h^{n-1}\|^2_{\Omega^n\setminus\Omega^{n-1}} \quad &\leq \frac{\Delta t}{2}\|\nabla(v_h^n + v_h^{n-1})\|^2_{\Omega^n\setminus\Omega^{n-1}} + \|v_h^n - v_h^{n-1}\|^2_{\Omega^n} \\ &\quad + c\Delta t g_h^n(v_h^n, v_h^n) + c\Delta t g_h^{n-1}(v_h^{n-1}, v_h^{n-1}).\end{aligned} \tag{2.26}$$

Using this, the following stability result is shown in [F6]. We define the discrete energy as

$$\mathfrak{E}^n(u_h^n, u_h^{n-1}) = \left(\frac{1}{\Delta t}\|u_h^n - u_h^{n-1}\|^2_{\Omega^n} + \|\nabla(u_h^n + u_h^{n-1})\|^2_{\Omega^n} + \frac{\gamma_D}{h}\|u_h^n\|^2_{\partial\Omega^n} + \gamma_g g_h^n(u_h^n, u_h^n)\right)^{1/2}. \tag{2.27}$$

Theorem 2.6 (Stability result, Theorem 3.5 in [F6])**.** *Let the Assumptions of Lemma 2.5 be valid, let $f \in L^\infty(I, L^2(\Omega(t)))$, $u^0 \in H^1(\Omega^0)$ and let the mapping $\mathbf{T} : \Omega(0) \to \Omega(t)$ be a $W^{1,\infty}$-diffeomorphism for all $t \in [0, t_{fin}]$. For sufficiently large γ_g, γ_D the solution $\{u_h^k\}_{k=1}^n$ of the discrete problem (2.24) fulfills*

$$\begin{aligned}\|u_h^n\|^2_{\Omega^n} + \Delta t\|u_h^n\|^2_{\Omega^n} + \Delta t\sum_{k=1}^n \mathfrak{E}^k(u_h^k, u_h^{k-1})^2 \\ \leq c\exp(c\,t_n)\left(\|u^0\|^2_{\Omega^0} + t_n\|\nabla u^0\|^2_{\Omega^0} + t_n\|f\|^2_{\infty,0}\right).\end{aligned} \tag{2.28}$$

Based on the stability result and estimates for the consistency and interpolation errors, the following error estimate is proven in [F6]:

Theorem 2.7 (Theorem 4.5 in [F6]). *Let $u \in L^\infty(I_n, H^{m+1}(\Omega(t))) \cap W^{1,\infty}(I_n, H^m(\Omega(t)) \cap W^{3,\infty}(Q)$ be the solution of (2.12) and $\{u_h^k\}_{k=1}^n$ the discrete solution of (2.24), respectively. Under Assumptions 2.4 and further technical and regularity assumptions stated in [F6], the global error $e^k = u^k - u_h^k$ $(k = 1, ..., n)$ satisfies for γ_g and γ_D sufficiently large*

$$\|e^n\|_{\Omega^n}^2 + \Delta t \|\nabla e^n\|_{\Omega^n}^2 + \Delta t \sum_{k=1}^{n} \mathfrak{E}^k(e^k, e^{k-1})^2 \leq \exp(ct_n) \left(\Delta t^4 + h^{2m}\right) \mathcal{R}(u)^2, \tag{2.29}$$

where $\mathcal{R}(u) = \|u\|_{\infty,3} + \|u\|_{\infty,m+1} + \|\partial_t u\|_{\infty,m} + \|u(t_n)\|_{W^{3,\infty}(Q)}$.

We note that -independently of our work- a similar approach has been investigated by von Wahl, Richter & Lehrenfeld in [92], using an *inf-sup-stable* Taylor-Hood discretisation instead of the *equal-order* approach with pressure stabilisation used in our work. Interestingly, the absence of a pressure stabilisation leads to a sub-optimal convergence estimate in [92].

Summary of the Numerical Results

In [F6] numerical results in two and three space dimensions are provided to verify the theoretical findings and the practical behaviour of the numerical method. As in the previous subsection all numerical experiments have been obtained using the CutFEM library [21]. As a two-dimensional example we consider a circle travelling with constant velocity $\mathbf{w} = (1, 0)$ towards the right. The three-dimensional example is a three-dimensional rectangular channel with a moving upper and lower wall, inspired by a pump. In both cases exact solutions are available and computations are executed on different mesh levels h, for different time step-sizes Δt and for P_1 and P_2 finite elements.

In the two-dimensional example the observed spatial convergence orders are either in agreement with or even larger than proven in Theorem 2.7 and the estimated temporal convergence orders are close to two for sufficiently small h, in agreement with Theorem 2.7. In the three-dimensional example, the effect of the CFL condition becomes visible when using P_2 elements by means of slightly reduced convergence orders for small h.

2.3 A Boundary Element Formulation for a Signorini Contact Problem using Nitsche's Method [F7]

In this section resp. the article [F7] with the collaborators E. Burman and M.W. Scroggs, we apply a Nitsche technique within a Boundary Element method to impose Signorini (contact) conditions on a Laplace equation. The approach to handle the Signorini conditions is very similar to the one derived in Section 1.1.2.

The problem under consideration in [F7] writes: *Find u such that*

$$-\Delta u = 0 \quad \text{in } \Omega, \qquad u = g_D \quad \text{on } \Gamma_D, \tag{2.30}$$

$$u \leq g_C, \quad \partial_n u \leq \psi_C \qquad \text{and} \qquad (\partial_n u - \psi_C)(u - g_C) = 0 \quad \text{on } \Gamma_C. \tag{2.31}$$

Here, we focus on $\Omega \subset \mathbb{R}^3$ and assume that $\partial\Omega = \Gamma_D \cup \Gamma_C$.

The Boundary Element Method (BEM) is a very efficient discretisation technique for certain PDEs. The idea is to transform the equations from the domain Ω to its boundary $\partial\Omega$

by means of an analytically available Green's function. The advantage compared to a finite element approach is that a triangulation of the hypersurface $\partial\Omega$ requires typically much less elements compared to a triangulation of Ω. Suitable boundary element formulations have been derived for the Laplace equation, linear elasticity and for Stokes and Helmholtz equations under certains assumptions on the data, see for example the textbook of Steinbach [88] for an overview.

The article [F7] is based on a previous work of my collaborators [12], where a unified boundary element framework based on the Calderón projector was derived for a Laplace problem with Dirichlet, Neumann or Robin boundary conditions. The contributon of [F7] is to extend the framework to Signorini conditions. Boundary element methods for Signorini problems were first studied by Han [54]. A variational formulation based on the Calderón projector was first presented in [55] by the same author. Alternative boundary element methods are based on Steklov-Poincaré operators [89, 96].

2.3.1 Calderón Projector

To state a variational formulation on $\partial\Omega$, we need to introduce a number of boundary operators. First, we define the Green's function for the Laplace operator in $\mathbb{R}^3$ by

$$G(x,y) = \frac{1}{4|x-y|}. \tag{2.32}$$

In the standard fashion (see e.g. [88, Chapter 6]), we define the single layer potential operator, $V : H^{-1/2}(\partial\Omega) \to H^1(\Omega)$, and the double layer potential, $K : H^{1/2}(\partial\Omega) \to H^1(\Omega)$, for $v \in H^{1/2}(\partial\Omega)$, $\mu \in H^{-1/2}(\partial\Omega)$, and $x \in \Omega$ by

$$(\mathcal{V}\mu)(\boldsymbol{x}) := \int_{\partial\Omega} G(\boldsymbol{x},\boldsymbol{y})\mu(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}, \qquad (\mathcal{K}v)(\boldsymbol{x}) \qquad := \int_{\partial\Omega} \frac{\partial G(\boldsymbol{x},\boldsymbol{y})}{\partial \boldsymbol{\nu}_{\boldsymbol{y}}} v(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{2.33}$$

We denote by $\gamma_{\mathrm{D}} : H^1(\Omega) \to H^{1/2}(\partial\Omega)$ and $\gamma_{\mathrm{N}} : H^{3/2}(\Omega) \to H^{-1/2}(\partial\Omega)$ the Dirichlet and Neumann trace operators, see [F7]. If both traces of a harmonic function u are known on $\partial\Omega$, for example as solutions of a boundary element formulation, the function can be reconstructed in Ω as follows by the single and double layer potential operators

$$u = -\mathcal{K}(\gamma_{\mathrm{D}}u) + \mathcal{V}(\gamma_{\mathrm{N}}u). \tag{2.34}$$

The Calderón projector is based on single layer (K), double layer (V), adjoint double layer (W), and hypersingular (K') boundary integral operators given by

$$(\mathsf{K}v)(\boldsymbol{x}) := \{\gamma_{\mathrm{D}}\mathcal{K}v\}_{\partial\Omega}(\boldsymbol{x}), \qquad (\mathsf{V}\mu)(\boldsymbol{x}) := \{\gamma_{\mathrm{D}}\mathcal{V}\mu\}_{\partial\Omega}(\boldsymbol{x}), \tag{2.35a}$$

$$(\mathsf{W}v)(\boldsymbol{x}) := -\{\gamma_{\mathrm{N}}\mathcal{K}v\}_{\partial\Omega}(\boldsymbol{x}), \qquad (\mathsf{K}'\mu)(\boldsymbol{x}) := \{\gamma_{\mathrm{N}}\mathcal{V}\mu\}_{\partial\Omega}(\boldsymbol{x}), \tag{2.35b}$$

where $\boldsymbol{x} \in \partial\Omega$ and $\{\gamma_{\mathrm{D}}f\}_{\partial\Omega}$ and $\{\gamma_{\mathrm{N}}f\}_{\partial\Omega}$ denote the averages of the interior and exterior Dirichlet and Neumann traces of f, respectively. Now, the Calderón projector is defined by

$$\mathsf{C} := \begin{pmatrix} (1-\sigma)\mathsf{Id} - \mathsf{K} & \mathsf{V} \\ \mathsf{W} & \sigma\mathsf{Id} + \mathsf{K}' \end{pmatrix}, \tag{2.36}$$

where σ is defined as in [88, Equation 6.11] (in particular $\sigma = \frac{1}{2}$ almost everywhere). The Calderón projector has the important property that, if u is a solution of (2.30), then

$$\mathsf{C}\begin{pmatrix}\gamma_{\mathrm{D}} u\\ \gamma_{\mathrm{N}} u\end{pmatrix} = \begin{pmatrix}\gamma_{\mathrm{D}} u\\ \gamma_{\mathrm{N}} u\end{pmatrix}. \tag{2.37}$$

We write $\lambda = \gamma_{\mathrm{N}} u$ and $u = \gamma_{\mathrm{D}} u$ for a more compact notation and introduce the corresponding Calderón form

$$\mathcal{C}[(u,\lambda),(v,\mu)] := \left\langle (\tfrac{1}{2}\mathsf{Id} - \mathsf{K})u, \mu \right\rangle_{\partial\Omega} + \langle \mathsf{V}\lambda, \mu\rangle_{\partial\Omega} + \left\langle (\tfrac{1}{2}\mathsf{Id} + \mathsf{K}')\lambda, v \right\rangle_{\partial\Omega} + \langle \mathsf{W}u, v\rangle_{\partial\Omega}.$$

The equation (2.37) can be written in variational form as

$$\mathcal{C}[(u,\lambda),(v,\mu)] = \langle u, \mu\rangle_{\partial\Omega} + \langle \lambda, v\rangle_{\partial\Omega}. \tag{2.38}$$

While (2.38) could be used as basis for discretisation with a boundary element method, we use in [F7] the multitrace form given by

$$\mathcal{A}[(u,\lambda),(v,\mu)] := -\langle \mathsf{K}u, \mu\rangle_{\partial\Omega} + \langle \mathsf{V}\lambda, \mu\rangle_{\partial\Omega} + \langle \mathsf{K}'\lambda, v\rangle_{\partial\Omega} + \langle \mathsf{W}u, v\rangle_{\partial\Omega}. \tag{2.39}$$

Now, (2.38) can be equivalently written as

$$\mathcal{A}[(u,\lambda),(v,\mu)] = \tfrac{1}{2}\langle u, \mu\rangle_{\partial\Omega} + \tfrac{1}{2}\langle \lambda, v\rangle_{\partial\Omega}. \tag{2.40}$$

2.3.2 Discretisation and A Priori Error Estimate

We assume that the boundary $\partial\Omega$ consists of a finite set of faces denoted by $\{\Gamma_i\}_{i=1}^M$, that are discretised by a family of conforming, shape regular triangulations $\{\mathcal{T}_h\}_{h>0}$. We consider continuous finite element spaces to approximate the Dirichlet trace u on $\partial\Omega$

$$\mathrm{P}_h^k(\partial\Omega) := \{v_h \in C^0(\partial\Omega) : v_h|_{T_i} \in \mathbb{P}_k(T_i), \text{ for every } T_i \in \mathcal{T}_h\},$$

and spaces, that are either fully discontinuous or at least partially discontinuous between the faces Γ_i that constitute the boundary $\partial\Omega$, to approximate the Neumann trace λ

$$\begin{aligned}\mathrm{DP}_h^l(\partial\Omega) &:= \{v_h \in L^2(\partial\Omega) : v_h|_{T_i} \in \mathbb{P}_l(T_i), \text{ for every } T_i \in \mathcal{T}_h\},\\ \widetilde{\mathrm{DP}}_h^l(\partial\Omega) &:= \{v_h \in \mathrm{DP}_h^l(\partial\Omega) : v_h|_{\Gamma_i} \in C^0(\Gamma_i), \text{ for } i = 1,\dots,M\}.\end{aligned}$$

Alternatively, we consider the space $\mathrm{DUAL}_h^0(\partial\Omega)$ of piecewise constant functions on the barycentric dual grid for λ, which is illustrated in Figure 2.5.

Note that the formulation (2.40) does not contain any boundary conditions yet. In [12] Dirichlet boundary conditions were imposed weakly by adding a weighted boundary residual

$$R_{\Gamma_D}(u_h, \lambda_h) := \beta_D^{1/2}(g_D - u_h)$$

with a suitable parameter $\beta_D > 0$. In the case of pure Dirichlet conditions ($\Gamma_{\mathrm{D}} = \partial\Omega$) the discrete variational formulation reads: *Find* $u_h \in P_h^k(\partial\Omega)$, $\lambda_h \in \Lambda_h^l$, *such that*

$$\mathcal{A}[(u_h,\lambda_h),(v_h,\mu_h)] = \tfrac{1}{2}\langle u_h, \mu_h\rangle_{\partial\Omega} + \tfrac{1}{2}\langle \lambda_h, v_h\rangle_{\partial\Omega} + \left\langle R_{\Gamma_{\mathrm{D}}}(u_h,\lambda_h), \beta_D^{1/2} v_h + \beta_D^{-1/2}\mu_h \right\rangle_{\Gamma_{\mathrm{D}}}, \tag{2.41}$$

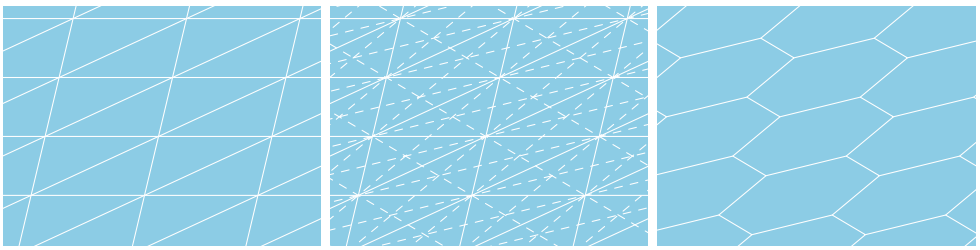


Figure 2.5. Illustration of a triangulation $\mathcal{T}_h$ (left) of a flat surface, the corresponding barycentric refinement of the grid (centre), and the dual grid (right). Figure taken from [F7] with permission from *SIAM*.

where Λ_h^l can be any of the discrete spaces $\mathrm{DP}_h^l(\partial\Omega)$, $\widetilde{\mathrm{DP}}_h^l(\partial\Omega)$ or $\mathrm{DUAL}_h^0(\partial\Omega)$.

For the incorporation of Signorini boundary conditions, we use the ideas discussed in Section 1.1.2. The three conditions in (2.31) are equivalent to

$$(\psi_C - \lambda_h) = [P^\tau(u_h, \lambda_h)]_+ , \qquad \text{where} \quad P^\tau(u_h, \lambda_h) := \tau(u_h - g_C) - (\lambda_h - \psi_C)$$

for arbitrary $\tau > 0$, see [4]. Thus, we define the boundary residual as

$$R_{\Gamma_C}(u_h, \lambda_h) = \tau^{-1} \left((\psi_C - \lambda_h) - [P^\tau(u_h, \lambda_h)]_+ \right). \tag{2.42}$$

The variational formulation for combined Signorini and Dirichlet conditions becomes: *Find* $u_h \in P_h^k(\partial\Omega)$, $\lambda_h \in \Lambda_h^l$, *such that*

$$\begin{aligned} \mathcal{A}[(u_h,\lambda_h),(v_h,\mu_h)] &= \tfrac{1}{2}\langle u_h, \mu_h\rangle_{\partial\Omega} + \tfrac{1}{2}\langle \lambda_h, v_h\rangle_{\partial\Omega} \\ &\quad + \left\langle R_{\Gamma_\mathrm{D}}(u_h,\lambda_h), \beta_D^{1/2} v_h + \beta_D^{-1/2}\mu_h \right\rangle_{\Gamma_\mathrm{D}} + \left\langle R_{\Gamma_\mathrm{C}}(u_h,\lambda_h), v_h + \tau^{-1}\mu_h \right\rangle_{\Gamma_\mathrm{C}} . \end{aligned} \tag{2.43}$$

For a proof of well-posedness of the discrete formulation (2.41) (based on Brouwer's fixed-point theorem), we refer to [F7].

A Priori Error Estimate

One of the main results of [F7] is the following a priori error estimate. We introduce the notation $\widetilde{\partial\Omega} = \cup_{i=1}^M \Gamma_i \setminus \partial\Gamma_i$ for the set of boundary points that lie in the interior of a face $\Gamma_i \subset \partial\Omega$.

Theorem 2.8 (Theorem 5.6 in [F7])**.** *Let* $(u,\lambda) \in H^s(\partial\Omega) \times H^r(\widetilde{\partial\Omega})$ *for some* $s \geq 1, r \geq 0$ *and* $(u_h, \lambda_h) \in \mathrm{P}_h^k(\partial\Omega) \times \Lambda_h^l$ *be the solutions of* (2.30) *and the discrete problem* (2.43)*, respectively. If there are positive constants* $c_i (i = 1, ..., 4)$ *such that* $c_1 < \beta_D < c_2 h^{-1}$ *and* $c_3 h^{-1} < \tau < c_4 h^{-1}$*, it holds that*

$$\|u - u_h\|_{H^{1/2}(\partial\Omega)} + \|\lambda - \lambda_h\|_{H^{-1/2}(\partial\Omega)} \leq c \left(h^{\zeta-1/2} |u|_{H^\zeta(\partial\Omega)} + h^{\xi+1/2} |\lambda|_{H^\xi(\widetilde{\partial\Omega})} \right),$$

where $\zeta = \min(k+1, s)$ *and* $\xi = \min(l+1, r)$ *for* $\Lambda_h^l \in \{\mathrm{DP}_h^l(\partial\Omega), \widetilde{\mathrm{DP}}_h^l(\partial\Omega)\}$ *and* $\zeta = \min(2, s)$ *and* $\xi = \min(\frac{1}{2}, r)$ *for* $\Lambda_h^l = \mathrm{DUAL}_h^0(\partial\Omega)$*. For the corresponding reconstructed functions on* Ω *defined by* (2.34)*, we have*

$$\|\tilde{u} - \tilde{u}_h\|_{H^1(\Omega)} \leq c \left(h^{\zeta-1/2} |u|_{H^\zeta(\partial\Omega)} + h^{\xi+1/2} |\lambda|_{H^\xi(\widetilde{\partial\Omega})} \right).$$

For the lowest-order variant $k = 1$, $l = 0$, Theorem 2.8 guarantees convergence order 1.5 when using the primal spaces $\mathrm{DP}_h^l(\partial\Omega), \widetilde{\mathrm{DP}}_h^l(\partial\Omega)$ if the solution is sufficiently regular. The order is reduced to one for the dual space $\mathrm{DUAL}_h^0(\partial\Omega)$, as the approximation property of the dual space is reduced on non-smooth domains $\partial\Omega$, see [84, Appendix 2].

2.3.3 Summary of the Numerical Results

A disadvantage of BEM is that the matrices in the resulting linear systems are typically both dense and ill-conditioned. Thus, good solvers and preconditioners are required for their solution. In [F7] a GMRES solver with a relatively simple mass matrix preconditioner is used within the library BEM++ [86].

In [F7], we show detailed numerical results for the pairs $P_h^1-\mathrm{DP}_h^0(\partial\Omega)$ and $P_h^1-\mathrm{DUAL}_h^0(\partial\Omega)$. A parameter study revealed that the choices $\beta_\mathrm{D} = 0.01$ and $\tau = 0.5/h$ led to the best results concerning the solvability of the system. When using the dual space $\mathrm{DUAL}_h^0(\partial\Omega)$ for the Neumann trace, we observe a convergence order slightly larger than one, which is slightly better than predicted by Theorem 2.8. For the primal space $\mathrm{DP}_h^0(\partial\Omega)$ the preconditioner was less effective and the accuracy of the GMRES solver decreased severely on finer meshes. Thus, we could not confirm the convergence order 1.5 shown in Theorem 2.8, but observed a slightly slower convergence behaviour with estimated convergence rates between 1 and 1.5. The construction of an efficient preconditioner, for example using the structure of the Calderón projector, is subject to future work.

3 Fluid-Structure Interactions with Multiple Scales in Time

This final section contains two articles that consider FSI with multiple scales in time, motivated by the numerical simulation of atherosclerotic plaque growth. In this application a complex cardiovascular FSI problem is coupled to a model for plaque growth, which is typically a system of convection-diffusion-reaction equations, see e.g. [35, 85, 91]. While plaque growth usually takes place over several months to years, this process depends on the wall shear stress in the blood vessel, which varies on a scale of milliseconds to seconds due to the pulsating heartbeat. A resolution of this short scale over a whole year would result in more than 10^9 time steps in the order of a few milliseconds. As each step corresponds to the solution of a complex FSI problem, this is computationally by far unfeasible.

In Section 3.1 resp. the publication [F8] with T. Richter, a temporal two-scale approach is developed for a slightly simplied flow problem. Then, in Section 3.2 resp. [F9] with the collaborator A. Heinlein, parallel time-stepping approaches are developed for this two-scale approach. Moreover, in [F9], the algorithms are applied to a full FSI problem coupled to a diffusion-reaction equation modelling plaque growth.

3.1 Efficient Approximation of Flow Problems with Multiple Scales in Time [F8]

Inspired by the temporal dynamics of atherosclerotic plaque growth, we consider flow through a channel which is deformed over a long (*macro*) time scale. The deformation is caused by an accumulation of foam cells within the vessel wall. In [F8] we use a strongly simplified ODE model for the evolution of the concentration $c_s(t)$ of these cells. For more complex plaque growth models, we refer to [35, 85, 91] and also Section 3.2 resp. [F9]. The coupled system considered in [F8] consists of the non-stationary Navier-Stokes equations for the flow and an ODE for the foam cell concentration

$$\mathbf{u}(0) = \mathbf{u}_0, \qquad \operatorname{div} \mathbf{u} = 0, \quad \rho(\partial_t \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u}) - \operatorname{div} \boldsymbol{\sigma}_f(\mathbf{u}, p) = \mathbf{f} \qquad \text{in } \Omega(c_s(t)) \tag{3.1a}$$

$$c_s(0) = 0, \qquad c_s' = \epsilon R(c_s, \mathbf{u}). \tag{3.1b}$$

Here, $\mathbf{u}$ and p are the fluid velocities and pressure and $\boldsymbol{\sigma}_f$ is the Cauchy stress tensor defined in (1.1). The concentration $c_s(t)$ influences the flow problem by means of the domain dependence $\Omega = \Omega(c_s(t))$. A prototypical configuration, which is also used in the numerical experiments in [F8], is sketched in Figure 3.1, with a parametrisation $\gamma(c_s)$ of the boundary. The reaction term $R(c_s, \mathbf{u}) \geq 0$ describes the influence of the fluid forces (namely the wall shear stress on the boundary Γ) on the foam cell concentration:

$$R(c_s, \mathbf{u}) \coloneqq (1 + c_s)^{-1}\left(1 + |\sigma_{WSS}(\mathbf{u})|^2\right)^{-1}, \quad \sigma_{WSS}(\mathbf{u}) \coloneqq \sigma_0^{-1} \int_\Gamma (I_d - \mathbf{n}\mathbf{n}^T)\boldsymbol{\sigma}_f \mathbf{n} \, do. \tag{3.2}$$

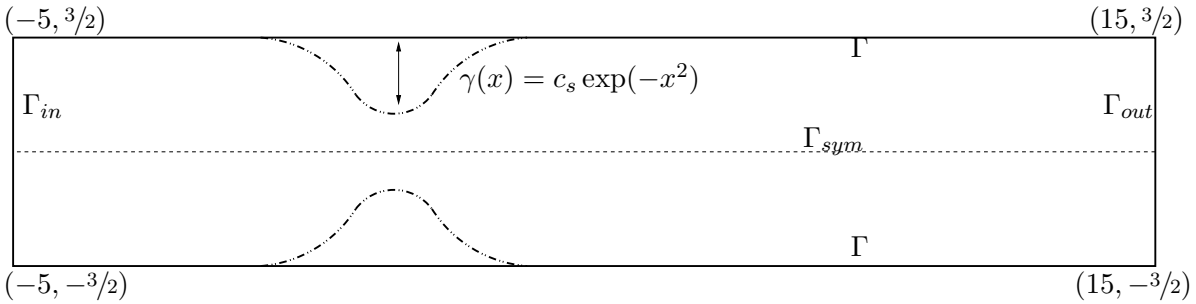


Figure 3.1. Configuration of the test case. We study flow in a channel with a boundary Γ that depends on a concentration variable c_s. c_s follows a simple reaction law with a right-hand side depending on the wall shear stress on Γ (Figure taken from [F8] with permission from *SIAM*.)

The parameter $\sigma_0 > 0$ is chosen such that $|\sigma_{WSS}(\mathbf{u})| = \mathcal{O}(1)$.

The flow problem is driven by a periodic oscillating inflow profile of period 1s:

$$\mathbf{u} = \mathbf{u}^D \text{ on } \Gamma_{in} \quad \text{ with } \mathbf{u}^D(t) = \mathbf{u}^D(t + 1s).$$

This period describes the short (*micro*) scale of the problem. The parameter $\epsilon \ll 1$ in (3.1b) defines the ratio between the short (*micro*) and the long (*macro*) time scales. We have $|c_s'| = \mathcal{O}(\epsilon)$ and $T = \mathcal{O}(\epsilon^{-1})$ is the expected long term horizon.

In a typical plaque growth problem, ϵ is in the range $\mathcal{O}(10^{-7}) - \mathcal{O}(10^{-6})$. A common numerical approach is to replace the short-scale problem by an averaged stationary problem using a fixed-in-time inflow profile [23, 95]. It is however widely accepted and also confirmed in numerical studies [45, 47] that such a simple averaging does not necessarily reproduce the correct dynamics. In [F8], we derive an improved algorithm in a mathematically rigorous way, including a detailed error analysis for both modelling and discretisation errors.

3.1.1 Derivation of the Two-Scale Scheme

We define the slow (*macro*) variable $\overline{c}_s(t)$ as average of the concentration c_s

$$\overline{c}_s(t) \coloneqq \int_t^{t+1s} c_s(s)\,\mathrm{d}s. \tag{3.3}$$

From (3.1b) we obtain the averaged equation

$$\overline{c}_s'(t) = \int_t^{t+1} \epsilon R(c_s(s), \mathbf{u}(s))\,\mathrm{d}s \tag{3.4}$$

Two approximation steps are performed to obtain an effective equation. First, the reaction term in (3.4) is evaluated in $\overline{c}_s(t)$ instead of $c_s(s)$. This is possible with a modelling error of $\mathcal{O}(\epsilon)$, as c_s varies slowly with respect to the micro scale. Second, the fast component $\mathbf{u}(s)$ is replaced by the localised solution $\mathbf{u}_{\overline{c}_s}(s)$ of a time-periodic problem in the micro interval $[0, 1\mathrm{s}]$ for a fixed value of $\overline{c}_s = \overline{c}_s(t)$:

$$\operatorname{div} \mathbf{u}_{\overline{c}_s} = 0, \quad \rho(\partial_t \mathbf{u}_{\overline{c}_s} + (\mathbf{u}_{\overline{c}_s} \cdot \nabla)\mathbf{u}_{\overline{c}_s}) - \operatorname{div} \boldsymbol{\sigma}_f(\mathbf{u}_{\overline{c}_s}, p_{\overline{c}_s}) = \mathbf{f}, \quad \mathbf{u}_{\overline{c}_s}(1) = \mathbf{u}_{\overline{c}_s}(0). \tag{3.5}$$

Algorithm 1: Abstract Two-Scale Algorithm

Let $0 = t_0 < t_1 < \cdots < t_N = T$ be a partition of the macroscopic time interval. Given $\overline{c}_{s,0}$, iterate for $n = 1, 2, \ldots$

1. Solve the time-periodic problem (3.5) for $\mathbf{u}_{\overline{c}_{s,n-1}}$.
2. Evaluate the reaction term $R_{n-1} \coloneqq \int_0^1 R(\overline{c}_{s,n-1}, \mathbf{u}_{\overline{c}_{s,n-1}}(s))\,\mathrm{d}s$
3. Forward the slow variable with a suitable (explicit) one-step scheme $\mathcal{G}$

$$\overline{c}_{s,n} = \mathcal{G}(\overline{c}_{s,n-1}; R_{n-1})$$

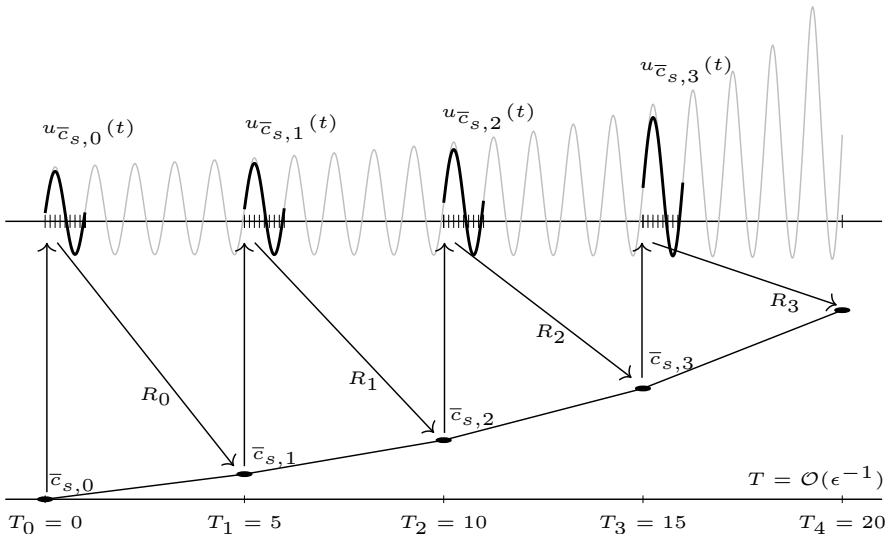


Figure 3.2. Construction of the two-scale scheme: 0. The slow variable $c_s(t)$ is discretised with a time-stepping scheme with macro time-step size $K \gg 1$. 1. In each step the current macro state $\overline{c}_{s,n-1}$ is transferred to the fast problem (top) and a periodic solution $\mathbf{u}_{\overline{c}_{s,n-1}}$ is computed on $[T_{n-1}, T_{n-1} + 1s]$ as approximation for $u(t)$. 2. The averaged reaction term $R_{n-1} = R(\overline{c}_{s,n-1}, \mathbf{u}_{\overline{c}_{s,n-1}})$ is computed from $\mathbf{u}_{\overline{c}_{s,n-1}}$ and transferred to the slow problem (bottom). 3. The slow variable $\overline{c}_{s,n}$ is updated by the macro step $T_{n-1} \to T_n$ (Figure taken from [F8] with permission from *SIAM*).

Given a suitable time discretisation of the macro time interval $[0, T]$, we obtain the following two-scale algorithm:

An illustration of the algorithm is given in Figure 3.2. In each macro time-step $t_{n-1} \to t_n$ (which is typically in the range of days for a plaque growth problem) a periodic flow problem is solved on the micro scale for one second (step 1 of Algorithm 1). The application of an explicit time integrator in step 3 of Algorithm 1 has the advantage that no additional loops

are necessary within one time step.

Algorithm 1 can be seen as a variant of the *Heterogeneous Multiscale Method* (HMM) [1, 30, 31]. The reason to use a time-periodic approximation in Step 1 is the lack of initial values on the micro scale, as the micro problem are not solved continuously in the whole interval $[T_n, T_{n+1}]$, but only once in the much smaller interval $[T_n, T_n + 1s]$. Within the HMM method, this issue is usually solved by introducing a relaxation time η to solve the micro problem in $[T_n, T_n + \eta]$ [1, 32]). The benefit of using a periodic micro problem is that efficient solution methods exist that exploit the periodicity. For details we refer to [F8, Section 4.3] and [78].

3.1.2 Analysis of the Two-Scale Scheme

The analysis of the two-scale scheme applied to the system (3.1a)-(3.1b) is complicated by several open questions regarding the existence and regularity theory of the Navier-Stokes equations in the periodic setting. For this reason and in order to improve the presentation, the analysis in [F8] is carried out for a strongly simplified model problem consisting of two ODEs instead of the full Navier-Stokes equations

$$u(0) = u_0, \qquad u'(t) + \lambda(c_s(t))u(t) = f(t), \tag{3.6a}$$

$$c_s(0) = 0, \qquad c_s'(t) = \epsilon R(c_s(t), u(t)), \tag{3.6b}$$

where $f(t) = f(t+1)$ is periodic, $\lambda(c_s) \geq \lambda_0 > 0$ is a function that depends smoothly on the concentration c_s and $R(\cdot, \cdot)$ is given by

$$R(c_s, u) \coloneqq (1 + c_s)^{-1}(1 + u^2)^{-1}. \tag{3.7}$$

For a discussion of the relation of (3.6a) to the Navier-Stokes equations (3.1b), we refer to [F8, Section 5.2]. By a slight abuse of notation, we use the notation $\overline{c}_s(t)$ also for the solution of the following averaged problem

$$\overline{c}_s'(t) = \int_t^{t+1} \epsilon R(\overline{c}_s(t), u_{\overline{c}_s(t)}(s))\,\mathrm{d}s, \quad \overline{c}_s(0) = 0, \tag{3.8}$$

where $u_{\overline{c}_s}$ is the solution of the corresponding time-periodic micro problem

$$u_{\overline{c}_s}'(t) + \lambda(\overline{c}_s)u_{\overline{c}_s}(t) = f(t), \quad u_{\overline{c}_s}(1) = u_{\overline{c}_s}(0), \tag{3.9}$$

for a fixed $\overline{c}_s \in [0, c_{max}]$. We obtain the following bound for the homogenisation error in terms of ϵ [F8, Lemma 10]:

Lemma 3.1. *Let $(c_s(t), u(t))$ and $(\overline{c}_s(t), u_{\overline{c}_s(t)}(t))$ be defined by (3.6a)-(3.6b), and (3.8)-(3.9), respectively, with the initial values $c_s(0) = \overline{c}_s(0) = 0$ and $u(0) = u_{\overline{c}_s(0)}(0)$. For $0 \leq t \leq T = \mathcal{O}(\epsilon^{-1})$ it holds*

$$|\overline{c}_s(t) - c_s(t)| \leq C\epsilon$$

with a constant $C > 0$.

Moreover, the interplay between the homogenisation error and the time discretisation error on both micro and macro scale is analysed in [F8]. The analysis is given exemplarily

for a discretisation with the second-order Adams-Bashforth scheme with uniform time-step K on the macro scale and a Crank-Nicolson scheme with uniform time step k on the micro scale. Moreover, we consider that in practice periodicity in step 1 of Algorithm 1 will only be fulfilled up to a certain (prescribed) tolerance tol_P:

$$\max_n |u_{\overline{c}_{s,n}}(1) - u_{\overline{c}_{s,n}}(0)| < \mathrm{tol}_P.$$

The resulting macro-scale solution is denoted by the vector $\{\overline{c}_{s,k}\}_{k=1}^n$. We obtain the following a priori error estimate:

Theorem 3.2 (A priori estimate for the two-scale algorithm)**.** *Let $I = [0,T]$ with $T = \mathcal{O}(\epsilon^{-1})$ and let $c_s \in C(I)$ and $\{\overline{c}_{s,k}\}_{k=1}^n$ be the solutions to problem (3.6) and the discrete equivalent of the effective equations (3.8)-(3.9), respectively. It holds that*

$$|c_s(t_n) - \overline{c}_{s,n}| = C\Big(k^2 + \epsilon^2 K^2 + \mathrm{tol}_P + \epsilon\Big),$$

where $C > 0$ does not depend on ϵ, K, k and tol_P.

The analysis reveals that an optimal choice for the time-step sizes on micro and macro level is given by $k \sim \epsilon K$.

3.1.3 Summary of the Numerical Results

In [F8] two numerical examples are considered, both using the coupled model (3.1a)-(3.1b) with the full Navier-Stokes equations. The two-dimensional geometry under consideration is illustrated in Figure 3.1. The fluid parameters are chosen as in a realistic blood flow configuration and the domain movement is incorporated by means of an ALE approach. For details on parameters and discretisation, we refer to [F8].

In the first example, we increase the scale separation parameter ϵ artificially to $\epsilon = 5 \cdot 10^{-5}$. For this value, the channel width decreases by approx. 66% after about $T = 51\,200s \approx 1$ day, such that a resolved simulation over the whole time frame $[0,T]$ is still feasible. We use this to validate the numerical approach and the convergence orders in Theorem 3.2. Therefore, we conduct simulations for three different micro time-step and five different macro time-step sizes and use a least-squares fit to approximate the parameters $\overline{c}, \alpha_k, \alpha_K, q_k, q_K$ in the postulated relation $c_s(k,K) = \overline{c} + \alpha_k k^{q_k} + \alpha_K K^{q_K}$. We find the relation

$$c_s(k,K) = \overline{c} - 1.12 \cdot k^{1.85} - 6.61 \cdot 10^{-10} \cdot K^{1.80}.$$

Even though the convergence estimates in Theorem 3.2 have been obtained for a simpler model, the numerical convergence orders are close to two and the constant $\alpha_K = 6.61 \cdot 10^{-10}$ is of order $\mathcal{O}(\epsilon^2)$, which indicates that the estimates in Theorem 3.2 are valid for the more complex configuration considered here as well.

In the second example, we decrease ϵ to the more realistic value $\epsilon = 10^{-6}$, such that the width of the artery is reduced by approx. 50% after $T \approx 30$ days. Besides the two time-step sizes k and K, we vary now also the mesh size $h \in \{0.04\mathrm{cm}, 0.08\mathrm{cm}, 0.16\mathrm{cm}\}$. A least-squares fit yields the relation

$$c_s(h,k,K) \approx 0.59076 + 7.6h^{2.4} - 1.7k^{2.2} - 0.04\epsilon^2 K^{1.9},$$

which indicates again that Theorem 3.2 is also valid in the flow configuration considered here. Moreover, we observe also a very good convergence behaviour in space.

Finally, a comparison of the runtimes of a two-scale simulation for mesh size $h = 0.08cm$ (approx. 1 hour) with the estimated time for a fully resolved simulation (almost 2 years) reveals a speed-up of up to 1:12 000 compared to the fully resolved simulation.

3.2 Parallel Time-Stepping for the Numerical Simulation of Atherosclerotic Plaque Growth [F9]

In the joint work [F9] with the collaborator A. Heinlein, we combine the temporal two-scale approach introduced in the previous section with parallel time-stepping algorithms. In the example studied above application of the temporal two-scale scheme led to a reduction of the total number of micro time steps to be solved by a factor of approx. 10^4. Considering that, in a realistic configuration, plaque growth typically takes place over more than a year (i.e. 10^8 to 10^9s), this means, however, that still 10^4 to 10^5 micro time steps remain to be solved.

Considering a complex three-dimensional FSI problem the simulation of up to 10^5 micro time steps is still computationally unfeasible. This motivates the development of parallelisation algorithms for the resulting two-scale algorithm. In [F9], we derive temporal parallelisation algorithms based on the parareal algorithm [49, 67] and apply them to prototypical plaque growth models including fluid-structure interactions.

Due to phase shifts in a coarse solution for hyperbolic partial differential equations (PDEs), such as the structural problem in FSI, it is generally challenging to apply parallel time-stepping methods to FSI problems; see, e.g., [69, 80]. Thus, it seems more promising to apply temporal parallelisation for the homogenised plaque growth problem on the macro scale, instead of the micro-scale FSI problem. This problem consists typically of a system of reaction-diffusion equations with parabolic character, see, e.g., [85, 95] and can thus be solved more efficiently by parallel time-stepping methods.

3.2.1 Equations

To compute the wall shear stress, we consider the following FSI system for the blood flow in the fluid domain $\Omega_f(t)$ and its interaction with the surrounding vessel wall in the solid part $\hat{\Omega}_s$ with interface $\Gamma(t)$:

$$\begin{aligned} \rho_f(\partial_t \mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u}) - \operatorname{div}\boldsymbol{\sigma}_f = 0, \qquad \operatorname{div}\mathbf{u} = 0 \qquad & \text{in } \Omega_f(t), \\ \rho_s \partial_t \dot{\hat{\mathbf{d}}} - \operatorname{div}(\hat{\mathbf{F}}_e\hat{\boldsymbol{\Sigma}}_e) = 0, \qquad & \text{in } \hat{\Omega}_s, \\ \boldsymbol{\sigma}_f\mathbf{n}_f + \boldsymbol{\sigma}_s\mathbf{n}_s = 0, \qquad \mathbf{u} = \dot{\mathbf{d}} \qquad & \text{on } \Gamma(t). \end{aligned} \tag{3.10}$$

We use here Lagrangian coordinates for the solid equations, indicated by the "hat" notation. For the meaning of the variables, we refer to Section 1. The solid Cauchy stress tensor $\boldsymbol{\sigma}_s$ is related to the second Piola–Kirchhoff stress $\hat{\Sigma}_e$ as follows:

$$\boldsymbol{\sigma}_s(x) = \hat{\boldsymbol{\sigma}}_s(\hat{x}) = (\hat{J}_e^{-1}\hat{\mathbf{F}}_e\hat{\boldsymbol{\Sigma}}_e\hat{\mathbf{F}}_e^T)(\hat{x}).$$

The influence of plaque growth is included in the FSI system by means of a multiplicative splitting of the deformation gradient $\hat{\mathbf{F}}_s$ into an elastic part $\hat{\mathbf{F}}_e$ and a growth function $\hat{\mathbf{F}}_g$ [45, 79, 95].

$$\hat{\mathbf{F}}_s = \hat{\mathbf{F}}_e \hat{\mathbf{F}}_g \quad \Leftrightarrow \quad \hat{\mathbf{F}}_e = \hat{\mathbf{F}}_s \hat{\mathbf{F}}_g^{-1} = [I + \hat{\nabla}\hat{\mathbf{u}}_s]\hat{\mathbf{F}}_g^{-1}. \tag{3.11}$$

For the foam cell concentration we consider two models of increasing complexity. First, we use the ODE model (3.1b)-(3.2) introduced in the previous section, in combination with the following dependence of the growth tensor $\mathbf{F}_g$ on the foam cell concentration

$$\hat{g}(\hat{x}, \hat{y}, t) = 1 + c_s \exp\left(-\hat{x}^2\right)(2 - |\hat{y}|), \quad \hat{\mathbf{F}}_g(\hat{x}, \hat{y}, t) := \hat{g}(\hat{x}, \hat{y}, t)\, I. \tag{3.12}$$

Here, shape and position of the plaque growth is prescribed in the center of the channel (cf. Figure 3.1), while the velocity of the growth depends on c_s. Second, we consider a reaction-diffusion PDE model with a space-dependent foam cell concentration $\hat{c}_s = \hat{c}_s(\hat{x}, t)$:

$$\begin{aligned} \partial_t \hat{c}_s - D_s \hat{\Delta} \hat{c}_s + R_s \hat{c}_s (1 - \hat{c}_s) = 0 \quad & \text{on } \hat{\Omega}_s, \\ D_s \hat{\partial}_n \hat{c}_s = \gamma(\boldsymbol{\sigma}_f^{WS}) := \epsilon \delta(x_1) \left(1 + \frac{\left\|\boldsymbol{\sigma}_f^{WS}\right\|^2}{\sigma_0^2}\right)^{-1} \quad & \text{on } \hat{\Gamma}, \\ \hat{c}_s = 0 \quad \text{on } \hat{\Gamma}_s, \qquad \hat{c}_s(\cdot, 0) = 0 \quad & \text{on } \hat{\Omega}_s. \end{aligned} \tag{3.13}$$

Here, σ_0 and ϵ are growth parameters as in the ODE model (3.1b)-(3.2) and D_s and R_s are positive diffusion and reaction coefficients, respectively. The function δ is given by $\delta(x_1) = \min\{0, (x_1 - 1)(x_1 + 1)\}^2$ and $\|\cdot\|$ is the Euclidean norm. The underlying idea is that the vessel wall is initially damaged in the central part of the interface around $\hat{x}_1 = 0$ (see Fig. 3.1) and monocytes can penetrate into $\hat{\Omega}_s$ in the part of the interface corresponding to $\hat{x}_1 \in (-1, 1)$. Here, the shape of the plaque growth is variable; the dependence of the growth tensor $\mathbf{F}_g$ on the foam cell concentration is given by

$$\hat{g}(\hat{x}, t) = 1 + \hat{c}_s(\hat{x}, t), \quad \hat{\mathbf{F}}_g(\hat{x}, t) := \hat{g}(\hat{x}, t)\, I. \tag{3.14}$$

3.2.2 Parallel Time-Stepping Algorithms

The parareal algorithm is based on a coarse discretisation of the (macro) time interval of interest $[0, T_{\text{end}}]$:

$$0 = T_0 < T_1 < ... < T_P = T_{\text{end}} \tag{3.15}$$

and a finer discretisation of each interval $[T_p, T_{p+1}]$ $(p = 0, ..., P-1)$:

$$T_p = t_{p,0} < t_{p,1} < ... < t_{p,n_p} = T_{p+1}.$$

The total number of fine time steps is denoted by $N_f = \sum_{p=1}^{P} n_p$. Note that the finer discretisation is still defined on the macro scale, i.e. the time step δt in this application is still in the scale of days, while the micro time step $\delta\tau$ is in the order of milliseconds.

The parareal algorithm consists of a coarse-scale propagator $\mathcal{C}(\overline{c}_s(T_p))$ and a a fine-scale propagator $\mathcal{F}(\overline{c}_s(T_p))$ [67], that both advance the foam cell concentration from T_p to T_{p+1}

Algorithm 2: Parareal algorithm

(I) **Initialisation**: Initialise $\{(\overline{c}_s^{(0)}(T_p)\}_{p=1}^{P}$ by the two-scale algorithm with a coarse macro time-step size $\delta T := (T_{p+1} - T_p)/N_p$ on the master process. Set $k \leftarrow 0$

(II) **while** $|c_s^{(k+1),fine}(T_{end}) - c_s^{(k),fine}(T_{end})| > \epsilon_{par}$ **do**

(II.a) **Fine problems** (in parallel for $p = 1, ..., P$)
Compute $\{(c_s^{(k+1),fine}(t_{p,q})\}_{q=1}^{N_p}$ by the two-scale algorithm with fine time-step size δt and set $\mathcal{F}(\overline{c}_s^{(k)}(T_p)) = c_s^{(k+1),fine}(t_{p,N_p})$

(II.b) **Coarse problem** (in serial)
for $p = 1, ..., P$ **do**

(i) Compute $\mathcal{C}(\overline{c}_s^{(k+1)}(T_p))$ by applying N_p time steps of the two-scale algorithm with coarse time-step size $\delta T = (T_{p+1} - T_p)/N_p$

(ii) Parareal update

$$\overline{c}_s^{(k+1)}(T_{p+1}) = \mathcal{C}(\overline{c}_s^{(k+1)}(T_p)) + \mathcal{F}(\overline{c}_s^{(k)}(T_p)) - \mathcal{C}(\overline{c}_s^{(k)}(T_p)).$$

$k \leftarrow k + 1$

with different accuracy. Precisely, given an iterate $\overline{c}_s^{(k)}$ for some $k \leq 0$, the parareal algorithm computes $\overline{c}_s^{k+1}$ by setting

$$\overline{c}_s^{(k+1)}(T_{p+1}) = \mathcal{C}(\overline{c}_s^{(k+1)}(T_p)) + \mathcal{F}(\overline{c}_s^{(k)}(T_p)) - \mathcal{C}(\overline{c}_s^{(k)}(T_p)) \quad \text{for } p = 0, ..., P-1. \tag{3.16}$$

This can be seen as a predictor-corrector scheme, where the coarse predictor $\mathcal{C}(\overline{c}_s^{(k+1)}(T_p))$ is corrected by the contribution $\mathcal{F}(\overline{c}_s^{(k)}(T_p)) - \mathcal{C}(\overline{c}_s^{(k)}(T_p))$ that depends only on the previous iterate $\overline{c}_s^{(k)}$ and can thus be computed in parallel.

A classical choice is to use a time-stepping scheme with a coarse time step $\delta T = (T_p - T_{p-1})/N_p$ for a small N_p in the coarse-scale propagator and a time-stepping scheme with a finer time step $\delta t = t_{p,q} - t_{p,q-1}$ in the fine-scale propagator. To keep the coarse propagator as cheap as possible, we will mainly focus on $N_p = 1$, such that the total number of coarse time steps

$$N_c := P \cdot N_p$$

is equal to P. A schematic algorithm is given as Algorithm 2. We note that in each time step of both coarse and fine propagator, a costly micro-scale computation is called. Before the first parareal iteration, we use one coarse propagation from T_0 to T_P to initialise the fine-scale problems (step (I)). More details on the algorithm are given in [F9].

Computational costs The main computational cost in Algorithm 2 lies in the solution of the non-stationary micro problems, which need to be solved in each time step of both coarse and fine-scale propagator. Considering a relatively coarse micro-scale discretisation of

$\delta\tau = 0.02\,\mathrm{s}$, as used in the numerical results in [F9], 50 time steps are necessary to compute a single period of the heart beat. The simulation of two or more cycles are usually necessary to obtain a near-periodic state, such that one micro problem consists of at least 100 FSI time-steps. Compared to these micro problems, we expect that the cost to advance the foam cell concentration by (3.1b) or (3.13) as well as the communication between the processes is negligible, see the detailed discussion in [F9]. Hence, it suffices to count the number of micro problems to be solved to estimate the computational cost of the algorithm.

In Algorithm 2, we need to solve N_c micro problems in step I, $k_{\mathrm{par}} \cdot \lceil N_l/P \rceil$ micro problems on each of the P processes in step II.a and $k_{\mathrm{par}} \cdot N_c$ micro problems within the coarse propagator in step II.b (i). As step II.a can be executed in parallel, this corresponds to the solution of

$$\underbrace{k_{\mathrm{par}} \cdot \lceil N_l/P \rceil}_{\text{fine level } (P \text{ parallel processes})} \quad + \quad \underbrace{(k_{\mathrm{par}}+1) \cdot N_c}_{\text{coarse level (1 serial process)}} \tag{3.17}$$

serial micro problems. This means that the computational cost tends to saturate for larger P (at least if we assume that the number of required parareal iterations k_{par} was independent of P). If we set $N_c = P$, the cost of the coarse propagator even increases with P and the choice $P \approx \sqrt{N_l}$ would be optimal in terms of the smallest computational time. Thus, the maximum possible speed-up is severely limited.

Re-usage of growth values In order to decrease the computational cost for larger P, we propose a modification of the coarse propagator. The idea is to re-use the growth values $\overline{\gamma}(\boldsymbol{\sigma}_f^{WS}(c_s^{(k+1),\mathit{fine}}(t_{p,i})))$, computed in the fine-scale propagator, on the coarse scale. To this end, all values $\overline{\gamma}_{p\cdot N_p+i} := \overline{\gamma}(\boldsymbol{\sigma}_f^{WS}(c_s^{(k+1),\mathit{fine}}(t_{p,i})))$ computed on the fine scale are stored and re-used as an approximation for $\overline{\gamma}(\boldsymbol{\sigma}_f^{WS}(c_s^{(k+1),\mathit{coarse}}(t_{p,i})))$ to advance the foam cell concentration in the coarse propagator. Hence, no micro problems need to be solved within the coarse propagator, at the cost of a slightly slower convergence behavior of the parareal algorithm. For the details of the re-usage algorithm, including an analytical convergence analysis, we refer to [F9].

The only coarse propagator that can not be modified is the initialisiation step, as in the beginning no growth values $\overline{\gamma}_{p\cdot N_p+i}$ are available. The computational cost in terms of micro problems to be solved in k_{par} iterations is thus

$$k_{\mathrm{par}} \cdot \lceil N_l/P \rceil + N_c.$$

The overall cost of the coarse propagators is reduced by a factor $(k_{\mathrm{par}}+1)$.

3.2.3 Numerical Results

In [F9] we present two-dimensional numerical results for both the ODE and the PDE growth model. An illustration of the configuration and the numerical example in the PDE case is given in Figure 3.3. As the focus in this work is not on large deformations, we use an ALE approach for the solution of the FSI problem (3.10). For details on discretisation and implementation, we refer to [F9].

In both examples, the standard parareal algorithm (Algorithm 2) converges quickly towards the reference solution. The respective stopping criterion (e.g. $|c_s^{(k+1),\mathit{fine}}(T_{\mathrm{end}}) -$

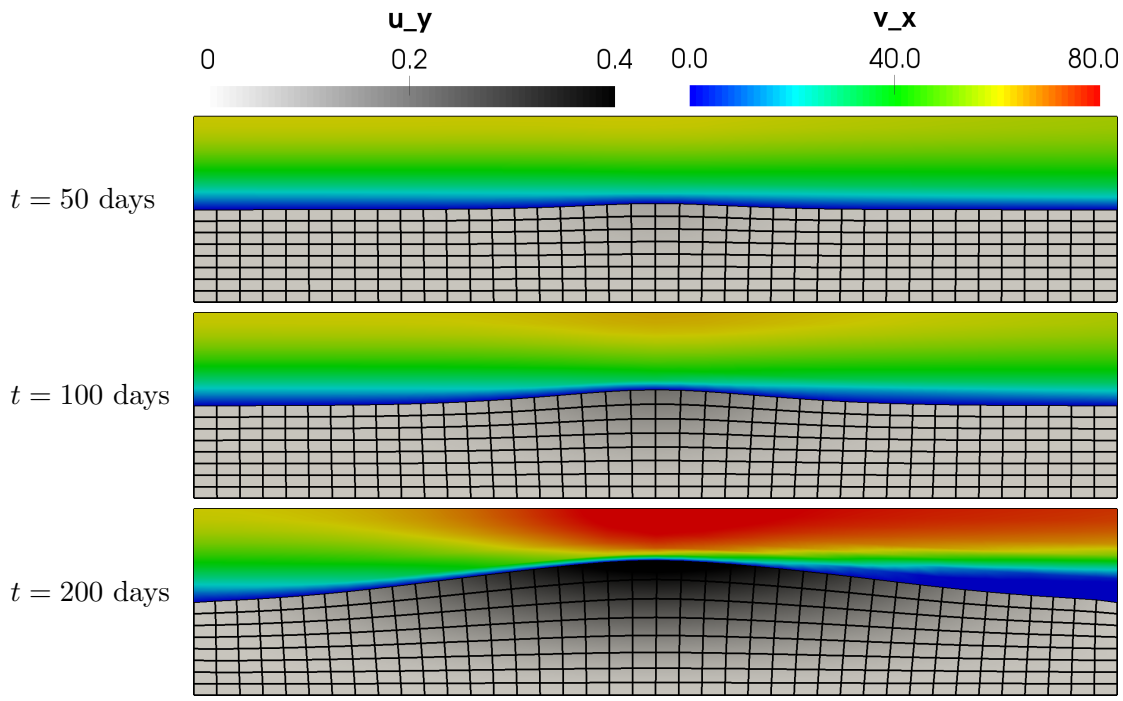


Figure 3.3. Visualisation of the flow and plaque growth on the lower half of the blood vessel at times $t = 50$ days, $t = 100$ days and $t = 200$ days. The horizontal velocity (in cm/s) and the vertical displacement (in cm) are shown on the deformed domain at micro time $\tau = 0.5\,\mathrm{s}$, i.e., the time of maximum inflow velocity. As the plaque growth evolves, significantly higher velocities arise in the central part.

$c_s^{(k),fine}(T_{\mathrm{end}})| < 10^{-3}$ for the ODE model) is reached in 3 or 4 parareal iterates, depending on the number of processes P. Due to the approximation used in the coarse-scale propagators, the modified parareal algorithm (re-usage of growth values) requires 1-2 additional iterations.

In the ODE model the lowest computational cost in terms of micro problems to be solved is 222 micro problems for $P = 30$ for the standard parareal algorithm and 128 micro problems for $P = 60$ in the re-usage variant This yields a speed-up of 4.5 resp. 7.8 compared to serial time-stepping (1 000 micro problems).

For the PDE model, the findings are similar. In Figure 3.4 we compare estimated parallel runtimes of both algorithms for different P, and illustrate the respective contributions of coarse- and fine-scale problems. For details on the computations and the estimation of parallel runtimes, we refer to [F9]. We observe that the cost of the coarse-scale propagators dominates for $P \geq 40$ in the standard parareal algorithm, while this contribution is much smaller in the re-usage variant, which is thus more efficient for $P \geq 40$. Due to a load imbalancing in the parallel fine-scale problems, the speed-ups with respect to actual runtimes are slightly smaller compared to the theoretical estimates based on the number of micro problems to be solved. Here, we obtain a maximum speed-up of 3.9 for the standard parareal algorithm for $P = 30$ and of 5.6 for the re-usage variant for $P = 70$.

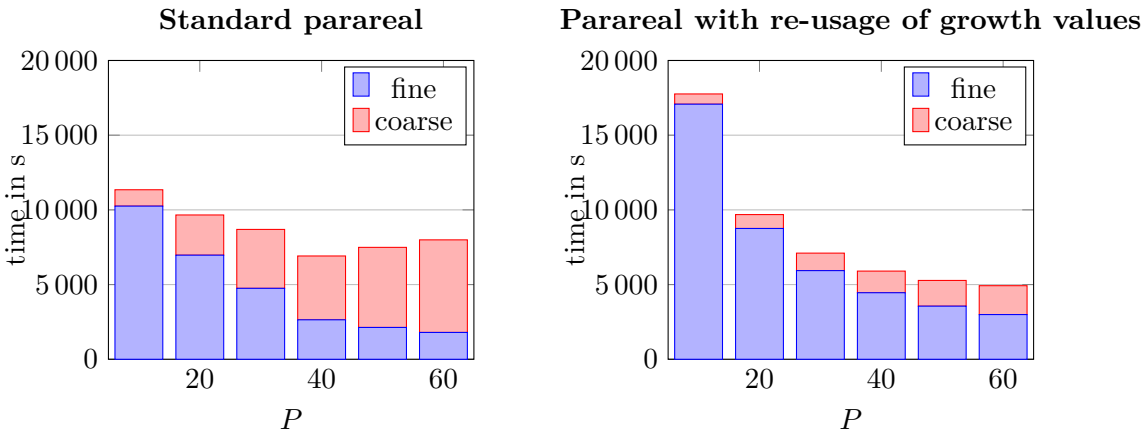


Figure 3.4. Illustration of the computational times spent within the coarse- and fine-scale problems for the PDE growth example.

Bibliography

[1] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden. The heterogeneous multiscale method. *Acta Numerica*, pages 1–87, 2012.

[2] C Ager, B Schott, AT Vuong, A Popp, and WA Wall. A consistent approach for fluid-structure-contact interaction based on a porous flow model for rough surface contact. *Int J Numer Methods Eng*, 119(13):1345–1378, 2019.

[3] C Ager, A Seitz, and WA Wall. A consistent and versatile computational approach for general fluid-structure-contact interaction problems. *Int J Numer Methods Eng*, 2020.

[4] P Alart and A Curnier. A mixed formulation for frictional contact problems prone to Newton like solution methods. *Comput Methods Appl Mech Eng*, 92(3):353–375, 1991. ISSN 0045-7825.

[5] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, and G.N. Wells. The fenics project version 1.5. *Arch. Numer. Soft.*, 3(100), 2015.

[6] P Angot. Analysis of singular perturbations on the brinkman problem for fictitious domain models of viscous flows. *Math Methods Appl Sci*, 22(16):1395–1412.

[7] P. Areias and T. Belytschko. A comment on the article "A finite element method for simulation of strong and weak discontinuities in solid mechanics" by A. Hansbo and P. Hansbo [Comput. Methods Appl. Mech. Engrg. 193 (2004) 3523-3540]. *Comput Methods Appl Mech Eng*, 195(9-12):1275–1276, February 2006.

[8] I Babuška. The finite element method for elliptic equations with discontinuous coefficients. *Computing*, 5:207–213, 1970.

[9] I Babuška, U Banarjee, and JE Osborn. Generalized finite element methods: Main ideas, results, and perspective. *Int J Comput Methods*, 1:67–103, 2004.

[10] S Basting and R Prignitz. An interface-fitted subspace projection method for finite element simulations of particulate flows. *Comput Methods Appl Mech Eng*, 267:133–149, 2013.

[11] M Besier and W Wollner. On the pressure approximation in nonstationary incompressible flow simulations on dynamically varying spatial meshes. *Int J Numer Methods Fluids*, 69(6):1045–1064, 2012.

[12] T Betcke, E Burman, and MW Scroggs. Boundary element methods with weakly imposed boundary conditions. *SIAM J Sci Comp*, 41(3):A1357–A1384, 2019.

[13] D Boffi, L Gastaldi, and S Wolf. Higher-order time-stepping schemes for fluid-structure interaction problems. *Discrete Contin. Dyn. Syst. - B*, 25(10):3807–3830, 2020.

[14] C Börgers. A triangulation algorithm for fast elliptic solvers based on domain imbedding. *SIAM J Numer Anal*, 27:1187–1196, 1990.

[15] JH Bramble and JT King. A finite element method for interface problems in domains with smooth boundaries and interfaces. *Adv Comput Math*, 6:109–138, 1996.

[16] E Burman. Ghost penalty. *CR Math*, 348(21-22):1217–1220, 2010.

[17] E Burman and MA Fernández. An unfitted Nitsche method for incompressible fluid–structure interaction using overlapping meshes. *Comput Methods Appl Mech Eng*, 279: 497–514, 2014.

[18] E Burman and P Hansbo. Edge stabilization for the generalized Stokes problem: a continuous interior penalty method. *Comput Methods Appl Mech Eng*, 195(19): 2393–2410, 2006.

[19] E Burman and P Hansbo. Fictitious domain finite element methods using cut elements: II. a stabilized Nitsche method. *Appl Numer Math*, 62(4):328–341, 2012.

[20] E Burman and P Hansbo. Fictitious domain methods using cut elements: III. A stabilized Nitsche method for Stokes' problem. *ESAIM: M2AN*, 48(3):859–874, 2014.

[21] E Burman, S Claus, P Hansbo, MG Larson, and A Massing. CutFEM: Discretizing geometry and partial differential equations. *Int J Numer Methods Eng*, 104(7):472–501, 2015.

[22] E Burman, MA Fernández, S Frei, and FM Gerosa. 3D-2D Stokes-Darcy coupling for the modelling of seepage with an application to fluid-structure interaction with contact. In *Numerical Mathematics and Advanced Applications ENUMATH 2019*, pages 215–223. Springer, 2021.

[23] CX Chen, Y Ding, and JA Gear. Numerical simulation of atherosclerotic plaque growth using two-way fluid-structural interaction. *ANZIAM Journal*, 53:278–291, 2012.

[24] F Chouly and P Hild. A Nitsche-based method for unilateral contact problems: numerical analysis. *SIAM J Numer Anal*, 51(2):1295–1307, 2013.

[25] F Chouly, P Hild, and Y Renard. Symmetric and non-symmetric variants of Nitsche's method for contact problems in elasticity: Theory and numerical experiments. *Math Comput*, 84(293):1089–1112, 2015.

[26] R Codina, G Houzeaux, H Coppola-Owen, and J Baiges. The fixed-mesh ALE approach for the numerical approximation of flows in moving domains. *J Comput Phys*, 228(5): 1591–1611, 2009.

[27] J Donea, P Fasoli-Stella, and S Giuliani. Lagrangian and Eulerian finite element techniques for transient fluid-structure interaction problems. In *Structural mechanics in reactor technology*. 1977.

[28] T Dunne. An Eulerian approach to fluid-structure interaction and goal-oriented mesh refinement. *Int J Numer Methods Eng*, 51:1017–1039, 2006.

[29] T Dunne. *Adaptive finite element approximation of fluid-structure interaction based on Eulerian and Arbitrary Lagrangian-Eulerian variational formulations.* PhD thesis, Univ of Heidelberg, 2007.

[30] W. E. *Principles of Multiscale Modeling.* Cambridge University Press, 2011.

[31] W. E and B. Engquist. The heterogenous multiscale method. *Comm. Math. Sci.*, 1(1): 87–132, 2003.

[32] B. Engquist and Y.-H. Tsai. Heterogeneous multiscale methods for stiff ordinary differential equations. *Math. Comp.*, 74(252):1707–1742, 2005.

[33] X Fang. An isoparametric finite element method for elliptic interface problems with nonhomogeneous jump conditions. *WSEAS Trans Math*, 12, 2013.

[34] M Feistauer and V Sobotíková. Finite element approximation of nonlinear problems with discontinuous coefficients. *ESAIM: M2AN*, 24:457–500, 1990.

[35] CA Figueroa, S Baek, CA Taylor, and JD Humphrey. A computational framework for fluid–solid-growth modeling in cardiovascular simulations. *Comput Methods Appl Mech Eng*, 198(45-46):3583–3602, 2009.

[36] L Formaggia, F Gatti, and S Zonca. An xfem/dg approach for fluid-structure interaction problems with contact. *Applicat Math*, 66(2):183–211, 2021.

[37] RM Franck and RB Lazarus. Mixed Eulerian-Lagrangian method. In *Methods in Computational Physics, vol.3: Fundamental methods in Hydrodynamics.* Academic Press: New York, 1964.

[38] S Frei. Eulerian finite element methods for interface problems and fluid-structure interactions. *PhD thesis, Heidelberg University, http://www.ub.uni-heidelberg.de/archiv/21590*, 2016.

[39] S Frei. An edge-based pressure stabilization technique for finite elements on arbitrarily anisotropic meshes. *Int J Numer Methods Fluids*, 89(10):407–429, 2019.

[40] S Frei and T Richter. A locally modified parametric finite element method for interface problems. *SIAM J Numer Anal*, 52(5):2315–2334, 2014.

[41] S Frei and T Richter. A second order time-stepping scheme for parabolic interface problems with moving interfaces. *ESAIM: M2AN*, 51(4):1539–1560, 2017.

[42] S Frei and T Richter. An accurate Eulerian approach for fluid-structure interactions. In S. Frei, B. Holm, T. Richter, T. Wick, and H. Yang, editors, *Fluid-Structure Interaction: Modeling, Adaptive Discretization and Solvers*, Radon Ser Comput Appl Math, pages 69–126. Walter de Gruyter, Berlin, 2017.

[43] S Frei, T Richter, and T Wick. Eulerian techniques for fluid-structure interactions: Part I–Modeling and simulation. In *Numerical Mathematics and Advanced Applications-ENUMATH 2013*, pages 745–753. Springer, 2015.

[44] S Frei, T Richter, and T Wick. Eulerian techniques for fluid-structure interactions: Part II–Applications. In *Numerical Mathematics and Advanced Applications-ENUMATH 2013*, pages 755–762. Springer, 2015.

[45] S Frei, T Richter, and T Wick. Long-term simulation of large deformation, mechano-chemical fluid-structure interactions in ALE and fully Eulerian coordinates. *J Comput Phys*, 321:874 – 891, 2016.

[46] S Frei, T Richter, and T Wick. An implementation of a locally modified finite element method for interface problems in deal.II. *Zenodo*, 2018. doi: 10.5281/zenodo.1457758. https://doi.org/10.5281/zenodo.1457758.

[47] S Frei, A Heinlein, and T Richter. On temporal homogenization in the numerical simulation of atherosclerotic plaque growth. *PAMM*, 21(1):e202100055, 2021.

[48] S Frei, T Richter, and T Wick. LocModFE: Locally modified finite elements for approximating interface problems in deal. ii. *Softw Impacts*, 8:100070, 2021.

[49] M. J. Gander. 50 years of Time Parallel Time Integration. In *Multiple Shooting and Time Domain Decomposition*. Springer, 2015. doi: 10.1007/978-3-319-23321-5_3.

[50] P Gangl. A local mesh modification strategy for interface problems with application to shape and topology optimization. In *Scientific Computing in Electrical Engineering*, pages 147–155. Springer, Cham, 2018.

[51] D Gérard-Varet, M Hillairet, and C Wang. The influence of boundary conditions on the contact problem in a 3d Navier–Stokes flow. *J Math Pures et Appl*, 103(1):1–38, 2015.

[52] R Glowinski, TW Pan, TI Hesla, and DD Joseph. A distributed Lagrange multiplier/fictitious domain method for particulate flows. *Int J Multiph Flow*, 25(5):755–794, 1999. doi: 10.1016/s0301-9322(98)00048-2.

[53] T Hagemeier, D Thévenin, and T Richter. Settling of spherical particles in the transitional regime. *Int J Multiphase Flow*, 138:103589, 2021. ISSN 0301-9322.

[54] H Han. The boundary finite element methods for Signorini problems. In YI Zhu and BY Guo, editors, *Numer Methods Part Diff Eq*, pages 38–49. Springer, 1987.

[55] H Han. A direct boundary element method for Signorini problems. *Math Comput*, 55 (191):115–128, 1990.

[56] A Hansbo and P Hansbo. An unfitted finite element method, based on Nitsche's method, for elliptic interface problems. *Comput Methods Appl Mech Eng*, 191(47-48):5537–5552, 2002.

[57] P Hansbo, MG Larson, and S Zahedi. A cut finite element method for coupled bulk-surface problems on time-dependent domains. *Comput Methods Appl Mech Engrg*, 307: 96–116, 2016.

[58] F Hecht and O Pironneau. An energy stable monolithic Eulerian fluid-structure finite element method. *Int J Numer Methods Fluids*, 85(7):430–446, 2017.

[59] M. Hillairet. Lack of collision between solid bodies in a 2d incompressible viscous flow. *Commun Part Diff Equ*, 32:1345–1371, 2007.

[60] J. Hoffman, B. Holm, and T. Richter. The locally adapted parametric finite element method for interface problems on triangular meshes. In S. Frei, B. Holm, T. Richter, T. Wick, and H. Yang, editors, *Fluid-Structure Interaction: Modeling, Adaptive Discretization and Solvers*, Radon Ser Comput Appl Math, pages 41–68. de Gruyter, 2017.

[61] S Höllbacher and G Wittum. A sharp interface method using enriched finite elements for elliptic interface problems. *Numer Math*, 147(4):759–781, 2021.

[62] TJR Hughes, WK Liu, and TK Zimmermann. Lagrangian-Eulerian finite element formulation for incompressible viscous flows. *Comput Methods Appl Mech Eng*, 29(3): 329–349, 1981.

[63] U Langer and H Yang. Numerical simulation of parabolic moving and growing interface problems using small mesh deformation. *RICAM Bericht-Nr. 2015-16*, 2015.

[64] C Lehrenfeld. The Nitsche XFEM-DG space-time method and its implementation in three space dimensions. *SIAM J Sci Comp*, 37(1):A245–A270, 2015.

[65] C Lehrenfeld and M Olshanskii. An Eulerian finite element method for pdes in time-dependent domains. *ESAIM: M2AN*, 53(2):585–614, 2019.

[66] C Lehrenfeld and A Reusken. Analysis of a high order unfitted finite element method for an elliptic interface problem. *IMA J Numer Anal*, 38:1351–1387, 2018.

[67] J.-L. Lions, Y. Maday, and G. Turinici. A "parareal" in time discretization of PDE's. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 332:661–668, 2001.

[68] M Luskin and R Rannacher. On the smoothing property of the Crank-Nicolson scheme. *Applicable Analysis*, 14:117–135, 1982.

[69] Margenberg, N and Richter, T. Parallel time-stepping for fluid-structure interactions. *Math Model Nat Phenom*, 16:20, 2021.

[70] V Martin, J Jaffré, and JE Roberts. Modeling fractures and barriers as interfaces for flow in porous media. *SIAM J Sci Comput*, 26(5):1667–1691, 2005.

[71] A Massing, MG Larson, A Logg, and ME Rognes. A stabilized Nitsche fictitious domain method for the Stokes problem. *J Sci Comput*, 61(3):604–628, 2014.

[72] N Moës, J Dolbow, and T Belytschko. A finite element method for crack growth without remeshing. *SIAM J Numer Anal*, 46:131–150, 1999.

[73] CLMH Navier. Mémoire sur les lois du mouvement des fluides. *Mémoires de l'Académie Royale des Sciences de l'Institut de France*, 6(1823):389–440, 1823.

[74] JA Nitsche. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh Math Sem Univ Hamburg*, 36:9–15, 1970.

[75] WF Noh. CEL: A time-dependent, two-space-dimensional, coupled Eulerian-lagrange code. Technical report, Lawrence Radiation Lab., Univ. of California, Livermore, 1963.

[76] T Richter. A fully Eulerian formulation for fluid-structure interactions. *J Comp Phys*, 233:227–240, 2013.

[77] T Richter. *Fluid-structure interactions: models, analysis and finite elements*, volume 118. Springer, 2017.

[78] T Richter. An averaging scheme for the efficient approximation of time-periodic flow problems. *Computers & Fluids*, 214:104769, 2021.

[79] EK Rodriguez, A Hoger, and AD McCulloch. Stress-dependent finite growth in soft elastic tissues. *J Biomechanics*, 4:455–467, 1994.

[80] D Ruprecht. Wave propagation characteristics of parareal. *Comput Vis Sci*, 19(1):1–17, 2018.

[81] PG Saffman. On the boundary condition at the surface of a porous medium. *Studies Appl Math*, 50(2):93–101, 1971.

[82] S Sathe and TE Tezduyar. Modeling of fluid–structure interactions with the space–time finite elements: contact problems. *Comput Mech*, 43(1):51–60, 2008.

[83] B Schott. *Stabilized cut finite element methods for complex interface coupled flow problems*. PhD thesis, Technische Universität München, 2017.

[84] MW Scroggs. *Efficient computation and applications of the Calderón projector*. PhD thesis, University College London, 2019.

[85] T. Silva, W. Jäger, M. Neuss-Radu, and A. Sequeira. Modeling of the early stage of atherosclerosis with emphasis on the regulation of the endothelial permeability. *J Theor Biol*, 496:110229, 2020.

[86] W Smigaj, T Betcke, S Arridge, J Phillips, and M Schweiger. Solving boundary integral problems with bem++. *ACM Trans Math Softw*, 41(2), 2015.

[87] A. Sohankar, C. Norberg, and L. Davidson. Numerical simulation of unsteady low-reynolds number flow around rectangular cylinders at incidence. *J. Wind Eng. Ind. Aerodyn.*, 69:189–201, 1997.

[88] O Steinbach. *Numerical approximation methods for elliptic boundary value problems*. Springer, 2008.

[89] Olaf Steinbach. Boundary element methods for variational inequalities. *Numer Math*, 126(1):173–197, 2014.

[90] K Tanaka, K Sekine, M Mizuguchi, and S Oishi. Estimation of Sobolev-type embedding constant on domains with minimally smooth boundary using extension operator. *J Inequ Appl*, 2015(1):1–23, 2015.

[91] MP Thon, A Hemmler, A Glinzer, M Mayr, M Wildgruber, A Zernecke-Madsen, and MW Gee. A multiphysics approach for modeling early atherosclerosis. *Biomech Model Mechanobiol*, 17(3):617–644, 2018.

[92] H von Wahl, T Richter, and C Lehrenfeld. An unfitted Eulerian finite element method for the time-dependent Stokes problem on moving domains. *IMA J Numer Anal*, 42(3): 2505–2544, 2021.

[93] D Wan and S Turek. Direct numerical simulation of particulate flow via multigrid FEM techniques and the fictitious boundary method. *Int J Numer Methods Fluids*, 51(5): 531–566, 2006.

[94] H Xie, K Ito, Z Li, and J Toivanen. A finite element method for interface problems with locally modified triangulation. *Contemp Math*, 466:179–190, 2008.

[95] Y Yang, W Jäger, M Neuss-Radu, and T Richter. Mathematical modeling and simulation of the evolution of plaques in blood vessels. *J Math Biol*, pages 1–24, 2014.

[96] S Zhang and X Li. An augmented Lagrangian method for the Signorini boundary value problem with BEM. *Bound Value Probl*, 2016(1):62, 2016.

Part II

Original Publications

ESAIM: M2AN 54 (2020) 531–564
https://doi.org/10.1051/m2an/2019072

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

A NITSCHE-BASED FORMULATION FOR FLUID-STRUCTURE INTERACTIONS WITH CONTACT

Erik Burman[1], Miguel A. Fernández[2,3] and Stefan Frei[1,*]

Abstract. We derive a Nitsche-based formulation for fluid-structure interaction (FSI) problems with contact. The approach is based on the work of Chouly and Hild (*SIAM J. Numer. Anal.* **51** (2013) 1295–1307) for contact problems in solid mechanics. We present two numerical approaches, both of them formulating the FSI interface and the contact conditions simultaneously in equation form on a joint interface-contact surface $\Gamma(t)$. The first approach uses a relaxation of the contact conditions to allow for a small mesh-dependent gap between solid and wall. The second alternative introduces an artificial fluid below the contact surface. The resulting systems of equations can be included in a consistent fashion within a monolithic variational formulation, which prevents the so-called "chattering" phenomenon. To deal with the topology changes in the fluid domain at the time of impact, we use a fully Eulerian approach for the FSI problem. We compare the effect of slip and no-slip interface conditions and study the performance of the method by means of numerical examples.

Mathematics Subject Classification. 65M60, 76D05, 73T05.

Received February 21, 2019. Accepted October 14, 2019.

1. Introduction

Contact problems have to be considered in many physical processes in engineering, medicine and nature. To name only a few consider for example the contact of balls and races in roller bearings, While an extensive amount of literature exists for the numerical simulation of contact in a purely solid mechanics context (see *e.g.* Wohlmuth [65] for an overview), *i.e.* disregarding the gas or liquid that mostly lies between contacting structures, much less works can be found considering full fluid-structure interaction with contact. The flow between contacting surfaces might however be of great importance for the contact dynamics. In the example of heart valves, the pulsating blood flow is even the driving force that enables opening and closure [4, 49]. In the case of ball bearings, fluid forces in the lubricant between ball and bearing may have a significant influence on the performance and wear of the bearing [11, 50].

Contact between different structures is typically formulated by means of variational inequalities [65]. In the context of full fluid-structure interaction (FSI), first results and algorithms can be found using either an

Keywords and phrases. Fluid-structure interaction, contact mechanics, Eulerian formalism, Nitsche's method, slip conditions.

[1] Department of Mathematics, University College London, Gower Street, WC1E 6BT London, UK.

[2] Inria Paris, 75012 Paris, France.

[3] Sorbonne Université & CNRS, UMR 7598 LJLL, 75005 Paris, France.

*Corresponding author: s.frei@ucl.ac.uk

artificial penalty force [63] or Lagrange multipliers [4,27,54] to obtain a well-posed and computationally feasible variational formulation.

However, these approaches have certain drawbacks: The use of a penalty force prevents real contact. The force is typically an artificial force and involves the choice of penalty parameters. If they are chosen too small, the structures might overlap in a numerical simulation. If they are chosen too large, the contact dynamics might be significantly perturbed [33]. In the case of Lagrange multipliers, additional variables are introduced on the contacting surfaces and an *inf-sup* condition is needed to ensure the well-posedness of the system. To tackle the variational inequality numerically, classical methods are based on *active-set strategies* including additional loops, which means in particular that the system of equations has to be solved several times in each time step. There are several possibilities to overcome these issues. Instead of an *active-set strategy*, semi-smooth Newton methods can be used to solve the system of equations [45]. In order to avoid additional variables, dual mortar methods, for example based on biorthogonal Lagrange multipliers, have been introduced [46].

Another possibility is to use Nitsche's method [57] to incorporate the contact conditions. This approach is based on an equivalent re-formulation of the contact conditions in equality form, which goes back to Alart and Curnier [1] and has been used in many works for the iterative solution of variational inequalities since then (see *e.g.* [45,46]). In the case of the contact of an elastic body with a wall, the mechanical contact conditions on the contact surface $\Gamma_C(t)$ read

$$d_n \leq 0, \quad \sigma_{s,n}(d) \leq 0, \quad \sigma_{s,n}(d) d_n = 0,$$

where d_n denotes the solid displacement in normal direction and $\sigma_{s,n}(d)$ is the normal stress component. It can be shown that these conditions are equivalent to the equality

$$\sigma_{s,n}(d) = -\frac{1}{\gamma} \max\{0, d_n - \gamma \sigma_{s,n}(d)\} \quad \text{on } \Gamma_C(t), \tag{1.1}$$

for arbitrary $\gamma > 0$ (see *e.g.* [1]). This equality can be incorporated weakly in the variational formulation using Nitsche's method. While this technique has originally been used as a solution technique for the iterative solution of variational inequalities, it has only recently been considered and analysed as a discretisation technique by Chouly and co-workers [20–22], including a complete analysis of the related discretisation errors.

The Nitsche approach has the advantage that it is fully consistent. Furthermore, no additional variables have to be introduced and no additional loop within each time-step is needed. Numerical convergence has been proven in a series of papers for friction-free and frictional contact [19,22,24]. Following these results, Burman and co-workers used the re-formulation (1.1) to derive a Galerkin Least Squares formulation in equality form for the obstacle problem [17] and a Galerkin Least Squares and a Lagrange multiplier formulation for membrane contact [18].

All these studies disregard however the fluid that usually lies between the contacting structures. The aim of the present paper is to derive a Nitsche-based formulation for FSI with contact. To this purpose, we will introduce Lagrange multipliers on the FSI interface which can then be eliminated following the arguments reported by Burman and Hansbo [16].

The mathematical modeling of contact in an FSI context brings along a further issue: It is unclear, whether the incompressible Navier–Stokes equations with standard boundary conditions are an appropriate model in the fluid part, when it comes to contact. Theoretical studies show, that for a smooth, rigid solid body, no contact with an exterior wall can happen, when no-slip conditions are used on the interface and the outer boundary of the fluid domain, see Hillairet [43] and Hesla [42] in 2 space dimensions and Hillairet *et al.* [36,44] in 3 space dimensions. This changes, when slip- or Navier-slip conditions are used on both the interface and the wall [36] or when the boundary of the solid is non-smooth [35,64]. Gerard-Varet and Hillairet [35] found in a model example that it comes to contact for a solid with a $C^{1,\alpha}$-parametrised boundary for $\alpha < 1/2$, while no contact happens for $\alpha \geq 1/2$. In the context of fluid-structure interactions, the regularity of the solid boundary depends on the solid displacement d, for which such regularity can usually not be guaranteed.

For a full FSI problem with a thin-walled structure, a no-collision result has been shown by Grandmont and Hillairet [38] in the no-slip case. For an overview on further results regarding existence of fluid-structure

interaction problems, we refer to Grandmont *et al.* [39]. Recently, Muha and Čanić [56] showed the well-posedness of a fluid-structure interaction system with slip-conditions.

Motivated by these theoretical results, we will study both no-slip and slip conditions on the FSI interface $\Gamma(t)$ in this work. It will turn out that the latter transits naturally into a "no-friction" condition when it comes to contact, while the prior leads to frictional contact. In this work, we will therefore consider friction-free contact, when a slip-condition is used on $\Gamma(t)$ and the specific frictional contact condition that follows from the transition of the interface conditions, when a no-slip condition is used on $\Gamma(t)$. For recent works on the incorporation of different friction laws (in particular Coulomb and Tresca friction), we refer to Chouly and co-workers [19, 24]. Moreover, we will study only contact of a deformable elastic structure with a fixed and straight wall for simplicity. Efficient algorithms to treat contact between more complex structures can be found, for example, in Puso [60], Yang *et al.* [66] and Chouly *et al.* [23, 55].

Concerning the governing equations, we focus on linear model equations for the fluid and solid part, *i.e.* the incompressible Stokes equations in the fluid and linear elasticity in the solid sub-domain. These simplifications must be seen as a first step towards the derivation of Nitsche-based contact formulations for complex FSI-contact problems. We consider, however, the case of a moving interface, which is a major challenge from the numerical point of view and leads to a non-linear FSI system, already in absence of contact.

The change of topology in the fluid domain causes additional numerical difficulties. Standard numerical approaches as the Arbitrary Lagrangian Eulerian method are not able to deal with topology changes, as the map from the reference domain to the Eulerian domain degenerates necessarily in this situation. The dynamics shortly before the impact can only be handled robustly when a Eulerian description of the fluid equations is used.

In the last years, several numerical approaches have been developed that are able to deal with topology changes. The methods can be split into two categories, according to the coordinate systems that are used for the solid system: Fully Eulerian approaches, where also the structure equations are formulated in Eulerian coordinates [26, 29, 33, 41, 61]; and Euler–Lagrangian techniques, where Lagrangian coordinates are used for the solid equations [2, 8, 13, 27, 37, 51, 53, 58, 67], such as in the *Immersed Boundary* or *Immersed Finite Element* methods.

Regarding the *Euler–Lagrangian* techniques, one can further distinguish between methods using Lagrange multipliers for the coupling of fluid and structure [37, 51] and methods based on Nitsche-techniques [2, 13, 40, 49, 53]. For the latter, a theoretical stability and convergence analysis has been derived [13]. The reader is referred to Boilevin-Kayl *et al.* [9] for a comparative study on the accuracy of some of these approaches.

The FSI approach we use here is based on the monolithic *Fully Eulerian* approach [26, 29, 33, 41, 61]. As the complete system of equations is formulated in Eulerian coordinates, the incorporation of contact conditions is straight-forward by means of variational principles. While some of the early works in this context suffered from stability and accuracy issues, see *e.g.* Dunne [28], accurate and robust discretisation and stabilisation techniques have been developed recently [32–34, 41]. We remark, however, that the algorithms we derive to incorporate contact can be combined in a straight-forward way with different FSI coupling techniques, *e.g.* Fictitious Domain or Immersed Boundary methods.

Concerning discretisation, we allow both for unfitted and fitted finite element approaches. For the unfitted case, so-called "ghost penalty" stabilisations can be used to guarantee the coercivity of the system [12, 13]. In order to simplify the presentation, we concentrate on fitted discretisations in this work and will comment on the unfitted case in a remark. In the numerical examples at the end of this paper, we will use the fitted *locally modified finite element method* [32].

The remainder of this paper is organised as follows: In Section 2, we first introduce the equations and the contact model. Before tackling the full contact problem, we derive a variational formulation for a model problem of a virtual obstacle within the fluid domain in Section 3, where we already have to deal with inequalities, but without any topology changes in the fluid domain. Then, we investigate contact with an exterior wall in Section 4. We show a stability result in Section 5 and present detailed numerical studies in Section 6, investigating the

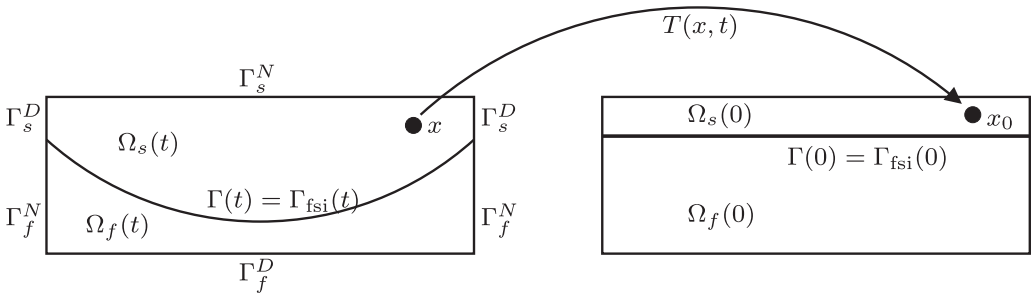


FIGURE 1. Illustration of an FSI problem without contact (left sketch). The domain affiliation in the current state can be determined by mapping back to the initial configuration, which is shown on on the right.

influence of contact parameters, interface conditions and different contact formulations as well as convergence under mesh refinement. We conclude in Section 7.

2. Model

We begin by presenting the models for the fluid part, the solid part and the fluid-structure interaction on one hand in Section 2.1 and on the other hand the contact model in Section 2.2. For both models, Nitsche-based variational formulations are introduced. Different possibilities to combine the two models will then be presented in Sections 3 and 4.

2.1. Fluid-structure interaction without contact

We consider a fluid-structure interaction problem that is given on an overall domain $\Omega \subset \mathbb{R}^2$ which is split into a (variable) sub-domain $\Omega_f(t)$ occupied by a viscous fluid, a sub-domain $\Omega_s(t)$ occupied by an elastic solid and a lower-dimensional interface $\Gamma(t)$ separating them, such that

$$\Omega = \Omega_f(t) \cup \Gamma(t) \cup \Omega_s(t).$$

The boundary of the fluid domain is partitioned as follows $\partial\Omega_f(t) = \Gamma_{\text{fsi}}(t) \cup \Gamma_f^D \cup \Gamma_f^N$, where $\Gamma_{\text{fsi}}(t)$ stands for the fluid-solid interface. As regards the solid boundary, we assume that $\partial\Omega_s(t) = \Gamma(t) \cup \Gamma_s^D \cup \Gamma_s^N$, where the boundary part $\Gamma(t) = \Gamma_{\text{fsi}}(t) \cup \Gamma_C(t)$ is decomposed into the terms of $\Gamma_{\text{fsi}}(t)$ and the contact zone $\Gamma_C(t)$ (see Fig. 1 for a configuration without contact and the left sketch of Fig. 2 for a configuration with contact). The restriction to two dimensions is made only to simplify the presentation. The models and the methods derived in this paper can be generalised conceptually in a straight-forward way to three space dimensions.

In this work we will use a Eulerian description for the complete FSI problem. As already mentioned in the introduction this is not necessary for the contact algorithms derived below, but one convenient way to deal with (possible) topology changes in the fluid domain $\Omega_f(t)$. In an Eulerian description, the solid sub-domain and the interface are implicitly defined by the (unknown) solid displacement d

$$\Omega_s(t) = \big\{x \in \Omega \,\big|\, T(x,t) \in \Omega_s(0)\big\}, \quad \Gamma(t) = \big\{x \in \Omega \,\big|\, T(x,t) \in \Gamma(0)\big\}, \tag{2.1}$$

where $T : \Omega(t) \to \Omega$ is a bijective map, that is given by $T(x,t) = x - d(x,t)$ in the solid domain $\Omega_s(t)$ and by an arbitrary (smooth) extension in $\Omega_f(t) = \Omega \setminus (\Omega_s(t) \cup \Gamma(t))$. For the details, we refer to the textbook of Richter [62] or to Frei [30].

In the variable fluid sub-domain $\Omega_f(t)$, we consider the linear incompressible Stokes equations

$$\partial_t u - \operatorname{div} \sigma_f(u,p) = f_f, \quad \operatorname{div} u = 0,$$

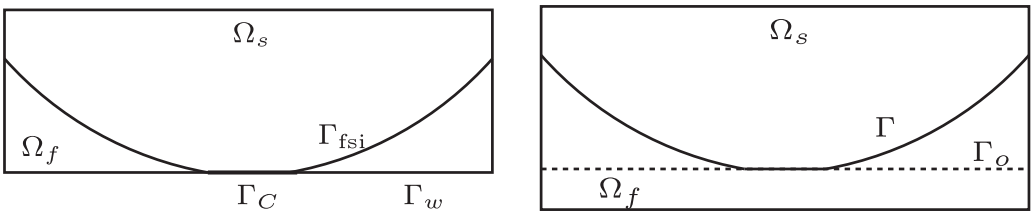


FIGURE 2. *Left*: body in contact with the wall Γ_w. *Right*: virtual obstacle line Γ_o within the fluid domain Ω_f.

where the Cauchy stress tensor σ_f is defined by

$$\sigma_f(u,p) = \nu_f\big(\nabla u^T + \nabla u\big) - pI,$$

u denotes the fluid velocity, p stands for pressure and $\nu_f > 0$ is a constant viscosity. In the solid-subdomain $\Omega_s(t)$, we assume a linear elastic material

$$\partial_t \dot{d} - \operatorname{div} \sigma_s(d) = f_s, \quad \partial_t d = \dot{d},$$

where the Cauchy stress tensor σ_s is given by

$$\sigma_s(d) = 2\mu_s E(d) + \lambda_s \operatorname{tr}(E(d))I, \quad E(d) = \frac{1}{2}\left(\nabla d + \nabla d^T\right),$$

$\dot{d} = \partial_t d$ denotes the solid velocity and $\lambda_s, \mu_s > 0$ are positive constants.

For the coupling across the fluid-solid interface $\Gamma_{\text{fsi}}(t)$, the continuity of velocities and normal stresses

$$\dot{d} = u, \quad \sigma_f(u,p)n = \sigma_s(d)n \quad \text{on} \quad \Gamma_{\text{fsi}}(t) \tag{2.2}$$

is typically considered for viscous fluids, where $n = n_s$ denotes the outer normal vector of the solid domain $\Omega_s(t)$. We recall that, since in this section there is no contact in the solid (*i.e.* $\Gamma_C(t) = \emptyset$), we have $\Gamma(t) = \Gamma_{\text{fsi}}(t)$. When it comes to contact, it is however questionable, whether the no-slip condition in (2.2) is still a reasonable approximation of the underlying physics, see the discussion on the no-collision paradox in the introduction. Therefore, we will study slip-conditions in this work as well (see Sect. 4.3), where the continuity across $\Gamma_{\text{fsi}}(t)$ is only imposed for the normal velocity

$$u \cdot n = \dot{d} \cdot n, \quad \tau^T \sigma_f n = 0, \quad \sigma_f n = \sigma_s n \quad \text{on} \quad \Gamma_{\text{fsi}}(t). \tag{2.3}$$

In order to close the system of equations, we define exterior boundary conditions for the fluid and solid

$$\begin{aligned} u = 0 \quad &\text{on} \quad \Gamma_f^D, \quad \sigma_f(u,p)n = 0 \quad \text{on} \quad \Gamma_f^N \\ d = 0 \quad &\text{on} \quad \Gamma_s^D, \quad \sigma_s(d)n = 0 \quad \text{on} \quad \Gamma_s^N \end{aligned}$$

and the initial conditions

$$u(x,0) = u^0(x) \text{ in } \Omega_f(0), \quad d(x,0) = d^0(x), \ \dot{d}(x,0) = \dot{d}^0(x) \text{ in } \Omega_s(0).$$

We introduce the finite element spaces $\mathcal{V}_h, \mathcal{Q}_h$ and $\mathcal{W}_h$ on a quasi-uniform family of triangulations $(\mathcal{T}_h)_{h>0}$ and use Nitsche's method to combine the equations and interface conditions into a monolithic variational formulation (see *e.g.* [13,40]). The monolithic system of equations reads in the no-slip case: *Find* $u(t) \in \mathcal{V}_h, p(t) \in \mathcal{Q}_h, d(t) \in \mathcal{W}_h$, *such that* $\dot{d} = \partial_t d$ *and*

$$\mathcal{A}_{\text{fsi, no-slip}}(u,p,d,\dot{d})(v,q,w) = (f_f, v)_{\Omega_f(t)} + (f_s, w)_{\Omega_s(t)} \quad \forall v,q,w \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h, \tag{2.4}$$

where

$$\begin{aligned}\mathcal{A}_{\text{fsi, no-slip}}(u,p,d,\dot{d})(v,q,w) := &\left(\partial_t u, v\right)_{\Omega_f(t)} + a_f(u,p;v,q) + \left(\partial_t \dot{d}, w\right)_{\Omega_s(t)} \\ &+ \left(\sigma_s(d), \nabla w\right)_{\Omega_s(t)} - \left(T_f(u,p,\dot{d}), w - v\right)_{\Gamma(t)} - \left(\dot{d} - u, \sigma_f(v,-q)n\right)_{\Gamma(t)}\end{aligned} \tag{2.5}$$

and

$$a_f(u,p;v,q) := \left(\sigma_f(u,p), \nabla v\right)_{\Omega_f(t)} + \left(\operatorname{div} u, q\right)_{\Omega_f(t)} + S(p,q).$$

The numerical fluid traction on the interface is defined by $T_f(u,p,\dot{d}) := \sigma_f(u,p)n - \gamma_{\text{fsi}}(\dot{d} - u)$. The Nitsche parameter is chosen as $\gamma_{\text{fsi}} := \gamma_{\text{fsi}}^0 \nu_f h^{-1}$. The term $S(p,q)$ stands for a pressure stabilisation term that is non-zero in case that the discrete fluid spaces do not fulfil a discrete *inf-sup condition*. Note that at the FSI interface, we have used the following relation for the interface terms arising from integration by parts

$$\left(\sigma_s n, w\right)_{\Gamma(t)} - \left(\sigma_f n, v\right)_{\Gamma(t)} = \left(\sigma_f n, w - v\right)_{\Gamma(t)} + \left(\llbracket \sigma n \rrbracket, w\right)_{\Gamma(t)}, \tag{2.6}$$

where we have dropped the dependencies of σ_f and σ_s for better readability. In the absence of contact, the jump of the stresses defined by

$$\llbracket \sigma n \rrbracket := \sigma_s n - \sigma_f n, \tag{2.7}$$

vanishes everywhere on $\Gamma(t)$. Furthermore, we have added the term

$$-\left(\dot{d} - u, \sigma_f(v,-q)n\right)_{\Gamma(t)}$$

as in Burman and Fernandez [13] for stability reasons.

Using a slip-condition at the FSI interface, the variational formulation reads: *Find* $u \in \mathcal{V}_h, p \in \mathcal{Q}_h, d \in \mathcal{W}_h$, *such that* $\dot{d} = \partial_t d$ *and*

$$\mathcal{A}_{\text{fsi, slip}}(u,p,d,\dot{d})(v,q,w) = \left(f_f, v\right)_{\Omega_f(t)} + \left(f_s, w\right)_{\Omega_s(t)} \qquad \forall v,q,w \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h,$$

where

$$\begin{aligned}\mathcal{A}_{\text{fsi, slip}}(u,p,d,\dot{d})(v,q,w) := &\left(\partial_t u, v\right)_{\Omega_f(t)} + a_f(u,p;v,q) + \left(\partial_t \dot{d}, w\right)_{\Omega_s(t)} + \left(\sigma_s(d), \nabla w\right)_{\Omega_s(t)} \\ &- \left(T_f(u,p,\dot{d})\cdot n, (w - v)\cdot n\right)_{\Gamma(t)} - \left((\dot{d} - u)\cdot n, n^T \sigma_f(v,-q)n\right)_{\Gamma(t)}.\end{aligned} \tag{2.8}$$

2.2. Contact model without fluid

We assume that the solid is at a positive distance to the boundary at initial time and that contact can only happen with the lower wall (see Fig. 2, left sketch, where the situation at contact is shown)

$$\Gamma_w = \left\{(x_1, x_2) \in \partial\Omega \,\middle|\, x_2 = 0\right\}.$$

We denote the outer normal vector of the fluid domain $\Omega_f(t)$ at Γ_w by $n_w = -e_2$. Moreover, let $g_0(x_2) > 0$ be the function describing the initial distance of a point $(x_1, x_2) \in \Gamma(t)$ to the wall Γ_w.

When contact with Γ_{w} occurs on a part $\Gamma_C(t) \subset \Gamma(t)$, suitable contact conditions are (see *e.g.* [1, 45])

$$d \cdot n_w \le g_0, \quad \sigma_{s,n_w} := n_w^T \sigma_s n_w \le 0, \quad (d \cdot n_w - g_0)\sigma_{s,n_w} = 0 \quad \text{on} \quad \Gamma(t). \tag{2.9}$$

The first inequality in (2.9) ensures that the solid can not pass though Γ_w, the second inequality describes that the normal stress is zero (in the absence of contact) or negative (during contact) and the third condition is a complementarity condition that guarantees that at least one of the inequalities is "active".

For arbitrary $\gamma_C > 0$ the first line in (2.9) is equivalent to (see *e.g.* [1])

$$\sigma_{s,n_w}(d) = -\gamma_C [P_\gamma(\sigma_{s,n_w}(d), d)]_+, \tag{2.10}$$

where $[f]_+ := \max\{f, 0\}$ and

$$P_\gamma(\lambda, d) := d \cdot n_w - g_0 - \frac{1}{\gamma_C}\lambda. \tag{2.11}$$

It remains to specify a contact condition for the tangential stresses. The simplest possibility is to consider frictionless contact

$$\tau_w^T \sigma_s n_w = 0 \quad \text{on} \quad \Gamma_C(t). \tag{2.12}$$

Choosing $\gamma_C = \gamma_C^0 \mu_s h^{-1}$, the variational formulation reads: *Find* $d(t) \in \mathcal{W}_h$ *such that* $\dot{d} = \partial_t d$ *and*

$$\left(\partial_t \dot{d}, w\right)_{\Omega_s(t)} + (\sigma_s(d), \nabla w)_{\Omega_s(t)} + \gamma_C \left([P_\gamma(\sigma_{s,n_w}, d)]_+, w \cdot n_w\right)_{\Gamma(t)} = (f_s, w)_{\Omega_s(t)} \quad \forall w \in \mathcal{W}_h. \tag{2.13}$$

We will discuss the tangential contact conditions in the context of fluid-structure interaction with contact below.

We close this section by mentioning that Chouly *et al.* [22] proposed a more general contact formulation that makes use of the consistency of the term

$$\left(\gamma_C [P_\gamma(\sigma_{s,n_w}(d), d)]_+ + \sigma_{s,n_w}(d), \sigma_{s,n_w}(w)\right)_{\Gamma(t)}.$$

For $\theta \in [-1, 1]$, the contact term can be generalised to

$$\begin{aligned} &\gamma_C \left([P_\gamma(\sigma_{s,n_w}(d), d)]_+, w \cdot n_w\right)_{\Gamma(t)} - \theta \left(\gamma_C [P_\gamma(\sigma_{s,n_w}(d), d)]_+ + \sigma_{s,n_w}(d), \sigma_{s,n_w}(w)\right)_{\Gamma(t)} \\ &\quad = (1-\theta)\gamma_C \left([P_\gamma(\sigma_{s,n_w}(d), d)]_+, w \cdot n_w\right)_{\Gamma(t)} + \theta \left(\gamma_C [P_\gamma(\sigma_{s,n_w}(d), d)]_+, P_\gamma(\sigma_{s,n_w}(w), w)\right)_{\Gamma(t)} \\ &\qquad - \theta \left(\sigma_{s,n_w}(d), \sigma_{s,n_w}(w)\right)_{\Gamma(t)}. \end{aligned} \tag{2.14}$$

For $\theta = 0$, we recover the formulation (2.13). Besides that, the case $\theta = 1$ is of particular interest, as it yields a symmetric formulation, for which a stability result has been shown [21].

3. Virtual obstacle within the fluid domain

Before considering full fluid-structure interaction with contact, we present our approach for a simplified model problem, where the movement of the solid is constraint by a virtual obstacle within the fluid domain, which is invisible to the fluid. This setting allows us to consider the numerical treatment of the interface conditions without accounting for the issues related to topology changes within the fluid. In Section 4, we will then show how this numerical setting can be extended to model contact of the solid with an exterior wall.

The setting of the model problem is shown on the right sketch of Figure 2. The fluid domain contains a horizontal obstacle line Γ_o which is invisible to the fluid, but an obstacle to the solid. One may consider for example a perfectly rigid thin membrane, that is laterally fixed and perfectly permeable for the fluid, but not for the solid, or a magnetic field below the obstacle that prevents the solid from crossing the line. We assume for simplicity that the obstacle line is parallel to the fluid boundary and denote its distance by $\alpha > 0$. The initial distance of the FSI interface $\Gamma(0)$ to the obstacle line is then given by

$$g_\alpha(x_2) := g_0(x_2) - \alpha.$$

We assume that $g_\alpha \geq 0$.

When considering fluid-structure interactions, the body is pre-loaded before it reaches the obstacle by means of the balance of normal forces

$$\sigma_s n = \sigma_f n \quad \text{on } \Gamma(t). \tag{3.1}$$

If the interface $\Gamma(t)$ reaches the obstacle Γ_o, the additional constraint $d \cdot n_w \leq g_\alpha$ has to be fulfilled. This gives rise to an additional surface force, that acts in the direction $-n_w = e_2$ (normal to Γ_o)

$$[\![\sigma n]\!] - \lambda n_w = 0 \quad \text{on} \quad \Gamma(t). \tag{3.2}$$

Note that we distinguish between the normal and tangential vectors of the wall n_w, τ_w and the normal and tangential vector of the FSI interface n, τ. The vectors are equal on $\Gamma_C(t)$, where the body is in contact with the wall, but different in the absence of contact.

The variable λ defined through (3.2) can be seen as a Lagrange multiplier, which is zero in absence of contact with the obstacle line (due to (3.1)) and can become negative during contact. This is described by the complementarity conditions

$$d \cdot n_w \leq g_\alpha, \quad \lambda \leq 0, \quad (d \cdot n_w - g_\alpha)\lambda = 0. \tag{3.3}$$

Splitting into normal and tangential contributions, (3.2) reads

$$n_w^T [\![\sigma n]\!] - \lambda = 0, \quad \tau_w^T [\![\sigma n]\!] = 0 \quad \text{on} \quad \Gamma(t). \tag{3.4}$$

Using the same trick as in Section 2.2, (3.3) is equivalent to

$$\lambda = -\gamma_C [P_\gamma(\lambda, d)]_+ \quad \text{on} \quad \Gamma(t), \tag{3.5}$$

with P_γ defined in (2.11).

Eliminating the Lagrange multiplier by using (3.4), this reads

$$\lambda = -\gamma_C [P_\gamma([\![\sigma_n(u,p,d)]\!], d)]_+ \quad \text{on} \quad \Gamma(t), \tag{3.6}$$

where we have used the abbreviations

$$\sigma_{s,n} := n_w^T \sigma_s n, \quad \sigma_{f,n} := n_w^T \sigma_f n, \quad [\![\sigma_n]\!] := n_w^T [\![\sigma n]\!].$$

The natural formulation in the discrete setting is to consider the numerical stresses T_f in (2.7) and in (3.2). Let us derive these conditions first for the no-slip case. Adding the additional surface force to the FSI-Nitsche formulation (2.4), the discrete variational formulation reads

$$\mathcal{A}_{\text{FSI, no-slip}}(u,p,d,\dot{d})(v,q,w) - (\lambda n_w, w)_{\Gamma(t)} = (f_f, v)_{\Omega_f(t)} + (f_s, w)_{\Omega_s(t)} \qquad \forall v,q,w \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h. \tag{3.7}$$

Due to the additional Nitsche interface terms in (3.2), this formulation includes the interface condition

$$[\![\widetilde{\sigma}_n(u,p,d)]\!] - \lambda = 0 \quad \text{on} \quad \Gamma(t), \tag{3.8}$$

where the numerical stress jump in the FSI-Nitsche formulation across the interface is given by

$$[\![\widetilde{\sigma}_n(u,p,d)]\!] := \sigma_{s,n}(d) - T_f(u,p,\dot{d}) \cdot n_{\mathrm{w}}. \tag{3.9}$$

Eliminating the Lagrange multiplier by means of (3.8), the identity (3.5) reads

$$\lambda = -\gamma_C [P_\gamma([\![\widetilde{\sigma}_n(u,p,d)]\!], d)]_+ \quad \text{on} \quad \Gamma(t). \tag{3.10}$$

For the definition of the numerical stresses in the slightly more complicated case of slip-interface conditions, we refer to Section 4.3. We will in the following analyse both contact formulations (3.6) and (3.10) and in particular their effect on the weakly imposed interface conditions. In order to avoid too much repetition, we use a general formulation that includes the Lagrange multiplier $\lambda = \lambda(u,p,d)$, keeping in mind that $\lambda(u,p,d)$ will be either chosen as the jump of normal stresses $[\![\sigma_n(u,p,d)]\!]$ or the jump of normal discrete stresses $[\![\widetilde{\sigma}_n(u,p,d)]\!]$.

Using (3.5), the variational formulation reads:

Variational Formulation 1. Find $u(t) \in \mathcal{V}_h, p(t) \in \mathcal{Q}_h, d(t) \in \mathcal{W}_h$ such that $\dot{d} = \partial_t d$ and

$$\begin{aligned} \mathcal{A}_{\text{fsi},*}(u,p,d,\dot{d})(v,q,w) + \gamma_C \left([P_\gamma(\lambda,d)]_+, w\cdot n_w\right)_{\Gamma(t)} \\ = (f_f, v)_{\Omega_f(t)} + (f_s, w)_{\Omega_s(t)} \quad \forall v,q,w \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h, \end{aligned} \tag{3.11}$$

where the bilinear form is one of the forms $\mathcal{A}_{\text{fsi, no-slip}}$ or $\mathcal{A}_{\text{fsi, slip}}$ defined in (2.5) and (2.8), respectively, and the contact parameter is chosen $\gamma_C = \gamma_C^0 \mu_s h^{-1}$, as in [20].

3.1. Weakly imposed interface conditions

Let us now analyse which interface conditions on $\Gamma(t)$ are implicitly included in Variational Formulation 1. For simplicity, we analyse the formulation with the bilinear form $\mathcal{A}_{\text{fsi, no-slip}}$ corresponding to no-slip conditions. Therefore, we integrate by parts in (3.11) and consider only the interface terms by formally neglecting the bulk and inter-element terms. For better readability, we drop all the dependencies of σ_f and σ_s. Collecting all terms with the fluid test function v, we obtain

$$-\sigma_f n + \sigma_f n - \gamma_{\text{fsi}}(\dot{d} - u) = -\gamma_{\text{fsi}}(\dot{d} - u) = 0 \quad \text{on} \quad \Gamma(t), \tag{3.12}$$

i.e. the kinematic condition $\dot{d} = u$. Next, we collect the interface terms for the solid part w and split into a normal part $(w\cdot n_w)$ and a tangential part $(w\cdot\tau_w)$. We recall that since the boundary Γ_w is flat, the extension to $\Gamma(t)$ of its tangential and normal vectors are trivial. For the tangential part, we obtain as usual for Nitsche-based FSI

$$\tau_w^T \sigma_s n - \tau_w^T \sigma_f n + \gamma_{\text{fsi}}(\dot{d} - u)\cdot\tau_w = 0 \quad \text{on} \quad \Gamma(t). \tag{3.13}$$

For the normal part, we have

$$\sigma_{s,n} - \sigma_{f,n} + \gamma_{\text{fsi}}(\dot{d} - u)\cdot n_w + \gamma_C [P_\gamma(d,\lambda)]_+ = 0 \quad \text{on} \quad \Gamma(t).$$

Let us first consider the case that the contact force is not active. We obtain, as in the standard FSI-Nitsche formulation

$$[\![\sigma_n]\!] + \gamma_{\text{fsi}}(\dot{d} - u)\cdot n_w = 0.$$

If the contact force is active, we get

$$0 = [\![\sigma_n]\!] + \gamma_{\text{fsi}}(\dot{d} - u)\cdot n_w + \gamma_C(d\cdot n_w - g_\alpha) - \lambda = \begin{cases} \gamma_C(d\cdot n_w - g_\alpha) + \gamma_{\text{fsi}}(\dot{d} - u)\cdot n_w, & \lambda = [\![\sigma_n]\!] \\ \gamma_C(d\cdot n_w - g_\alpha), & \lambda = [\![\widetilde{\sigma}_n]\!] \end{cases}. \tag{3.14}$$

In the first case, this is a combination of the "active" contact condition $d\cdot n_w = g$ and the continuity of velocities. As the continuity of velocities is imposed from the fluid side (3.12), this is not an issue for the model problem considered here. It will however lead to problems, when we consider contact of the solid with the lower wall Γ_w in Section 4. There, the second formulation, based on the discrete stresses (3.8) will be needed. As can be seen in (3.14), the pure contact condition $d\cdot n_w = g$ is valid from the solid side during contact.

4. Fluid-structure interaction with contact

In this section, we consider contact of the solid body with the lower fluid boundary Γ_w. The formulation of a macroscopically relevant computational formulation without introducing the full physical model of all the interacting scales and elements, which is too costly in many applications, is an open problem to our knowledge and rarely discussed in literature.

When the solid enters into contact with a rigid wall, the action of the wall on the solid is expressed as a constraint, giving rise to a variational inequality. In principle there is no longer a fluid between the contacting

solids ("dry contact"). However, the presence of the fluid can not be discarded in this zone, since otherwise contact could be released in the interior of the contact zone through deformation of the elastic solid, thus creating a void, without neither fluid nor solid. A solution to this problem is to design a model where a fluid pressure remains in the contact zone, either on the boundary or in the bulk domain and prevents unphysical lift off creating vacuum. Our objective in the present work is to consider finite element formulations for contact between an elastic solid and a rigid wall assuming the existence of such a pressure model.

For the purpose of discussion, and without claiming that these models are optimal, we consider two simple models of seepage that introduce a fluid pressure in the contact zone. The first is a rather widespread approach which consists in relaxing the contact condition by introducing a small gap (see *e.g.* [4,49]), so that the contact in the computational model takes place at an ϵ-distance from the wall (Sect. 4.1). Thus allowing for a thin fluid layer to remain. Here ϵ can be made to go to zero with reduced mesh-size. The second model consists in allowing the bulk fluid to penetrate into the solid wall in the form of a Navier–Stokes porous medium model where the permeability is driven to zero with reduced mesh-size (Sect. 4.2).

Our main objective is then to design a computational method that allows for contact in a consistent fashion, exploiting the presence of the modelled fluid contact pressure to get a smooth transition from contact to no contact without the appearance of non-physical inclusions or oscillations ("chattering"). To this end we study a nonlinear Nitsche-type coupling combining in a unfied way the FSI coupling and the contact conditions.

4.1. Relaxed contact formulation

The idea of the relaxed formulation is to impose contact using a virtual barrier Γ_ϵ for the solid at a very small distance ϵ to Γ_w. This may be used for instance when the numerical discretisation of all the scales of the macroscopic model is unfeasible. We consider two examples.

If the solid has some fine scale structure on an unresolved scale ϵ, for example due to surface roughness, it is reasonable to consider that contact already takes place when the distance between the two structures is $O(\epsilon)$. Since by the modelling assumption the elastic solid can not penetrate into the rigid solid, the total forces from the fluid captured in the ϵ-layer and of the rigid solid must be sufficient to prevent penetration. This gives rise to equation (3.8), where λ represents the additional force coming from the rigid solid below Γ_ϵ that is necessary to prevent penetration.

On the other hand, as noted in the introduction, the continuous formulation does not allow for contact, when no-slip conditions are used on at least one of the interface and/or the boundary, due to the Navier–Stokes contact paradox. To reproduce this phenomenon in numerical simulations, an extremely high mesh resolution in the remaining small fluid layer between solid and rigid wall is typically required, in order to approximate the fluid tensor σ_f, which might become singular for $h \to 0$. In many situations, this might exceed the computational limits and moreover, it is *a priori* not clear how fine the mesh needs to be. In principle the condition $\sigma_{s,n} = \sigma_{f,n}$ is valid in the continuous formulation in this situation. As the singular behaviour of $\sigma_{f,n}$ might however not be accurately captured on coarser grids, it makes sense to re-inforce the fluid force by the contact force $[P_\gamma]_+$, if the line Γ_ϵ is passed, which indicates a failure of the force $\sigma_{f,n}$ to prevent contact.

In both these situations our idea to impose a virtual barrier Γ_ϵ for the solid at a very small distance ϵ to Γ_w permits numerical simulations already on reasonably coarse grids. In the numerical examples we will place the obstacle line at a distance $\alpha = \epsilon(h)$ to Γ_w, see Figure 3 on the left. We assume that $\epsilon(h) \to 0$, as the mesh size h tends to zero. The variational formulation is given by (3.11), where $\alpha = \epsilon(h)$, *i.e.*

$$g_\alpha(x_2) = g_{\epsilon(h)}(x_2) = g_0(x_2) - \epsilon(h).$$

Besides its simplicity, the main advantages of this contact formulation are:

- The numerical difficulties related to a topology change of the fluid domain are simplified.
- No-slip conditions can be used on both $\Gamma(t)$ and the lower wall Γ_w.

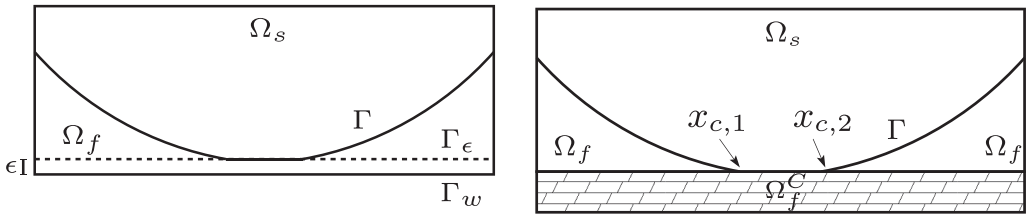


FIGURE 3. Illustration of the two approaches to include contact. *Left*: relaxed contact formulation: Body in contact with an obstacle line Γ_ϵ close to Γ_w. *Right*: introduction of an artificial fluid domain Ω_f^C below Γ_w.

4.2. Contact formulation using an artificial fluid

As a second possibility, we add an artificial fluid domain Ω_f^C below Γ_w, see the right sketch of Figure 3, and allow for a small, asymptotically for $h \to 0$ vanishing penetration into Ω_f^C. For the extension, we use a Stokes equation in combination with a penalty for the velocity u that drives the penetration to zero with $h \to 0$. This can be seen as a porous medium model with asymptotically vanishing permeability.
The variational formulation reads:

Variational Formulation 2. Find $u \in \mathcal{V}_h, p \in \mathcal{Q}_h, d \in \mathcal{W}_h$ such that $\dot{d} = \partial_t d$ and

$$\begin{aligned}\mathcal{A}^C_{FSI,*}(u,p,d,\dot{d})(v,q,w) + \gamma_C\left([P_\gamma(\lambda,d)]_+, w\cdot n_w\right)_{\Gamma(t)} \\ = (f_f, v)_{\Omega_f(t)} + (f_s, w)_{\Omega_s(t)} \quad \forall v,q,w \in \mathcal{V}_h \times \mathcal{Q}_h \times \mathcal{W}_h,\end{aligned} \tag{4.1}$$

where in the no-slip case

$$\begin{aligned}\mathcal{A}^C_{\text{fsi,no-slip}}(u,p,d,\dot{d})(v,q,w) := \left(\partial_t u, v\right)_{\Omega_f(t)\cup\Omega_f^C} + a_f^C(u,p;v,q) + \gamma_a(u,v)_{\Omega_f^C} + \left(\partial_t \dot{d}, w\right)_{\Omega_s(t)} \\ + (\sigma_s(d), \nabla w)_{\Omega_s(t)} - \left(T_f(u,p,\dot{d}), w - v\right)_{\Gamma(t)} - \left(\dot{d} - u, \sigma_f(v,-q)n\right)_{\Gamma(t)}\end{aligned} \tag{4.2}$$

and for slip interface conditions

$$\begin{aligned}\mathcal{A}^C_{\text{fsi,slip}}(u,p,d,\dot{d})(v,q,w) := \left(\partial_t u, v\right)_{\Omega_f(t)\cup\Omega_f^C} + a_f^C(u,p;v,q) + \gamma_a(u,v)_{\Omega_f^C} + \left(\partial_t \dot{d}, w\right)_{\Omega_s(t)} + (\sigma_s(d), \nabla w)_{\Omega_s(t)} \\ - \left(T_f(u,p,\dot{d})\cdot n, (w-v)\cdot n\right)_{\Gamma(t)} - \left((\dot{d}-u)\cdot n, n^T\sigma_f(v,-q)n\right)_{\Gamma(t)}.\end{aligned} \tag{4.3}$$

Note the presence of the penalty term $\gamma_a(u,v)_{\Omega_f^C}$ within the artificial fluid, where $\gamma_a := \gamma_a^0 h^\beta$ for $\beta > 0$. The fluid bilinear form is defined in $\Omega_f(t) \cup \Omega_f^C$

$$a_f^C(u,p;v,q) := (\sigma_f(u,p), \nabla v)_{\Omega_f(t)\cup\Omega_f^C} + (\text{div } u, q)_{\Omega_f(t)\cup\Omega_f^C} + S(p,q).$$

Remark 4.1 (Porous medium analogy)**.** The penalisation used in Variational Formulation 2 corresponds to the so-called penalty approach that is sometimes used for the coupling of free flow and flow through porous medium [25,48]. There, the Stokes equations in the fluid part Ω_f and the Darcy equations in the porous medium Ω_p (which corresponds to Ω_f^C) are formulated simultaneously in the whole domain $\Omega = \Omega_f \cup \Omega_p$ in the spirit of the volume penalty approach

$$\partial_t u - \mu\Delta u + \nabla p + \frac{\mu}{K} u\chi_{\Omega_p} = 0, \qquad \nabla\cdot u = 0 \quad \text{in } \Omega,$$

where K denotes the permeability of the porous medium and χ_{Ω_p} is the characteristic function of the domain Ω_p. In this sense the artificial fluid in our approach can be seen as a porous medium with asymptotically vanishing permeability $K = \mathcal{O}(h^\beta)$. A mathematical justification of this penalisation has been given by Angot [3]. Below we will choose the parameter $\beta = 2$, which is optimal in terms of the conditioning of the system. Following the argumentation of Angot a larger choice of β is also reasonable, with an optimal convergence rate reached for $\beta = 4$.

4.2.1. Discussion of the weakly imposed interface conditions: The no-slip case

We have already derived the weakly imposed interface conditions for the no-slip case in Section 3.1. Here, however, we have to consider that the fluid below the contact line is artificial, and we should in particular make sure that there is no feedback from the artificial fluid to the solid. In other words, we want that the artificial fluid acts as a *slave* to the solid during contact. Owing to (3.14), this naturally motivates the choice of λ in terms of the numerical stresses.

First, we note that the continuity of velocities (3.12) is imposed from the fluid side, such that no feedback to the solid is included from this equation. Considering the contact condition for the normal contact (3.14), we obtain

$$\gamma_C(d \cdot n_w - g_0) + \gamma_{\text{fsi}}(\dot{d} - u) \cdot n_w = 0 \tag{4.4}$$

when choosing $\lambda = [\![\sigma_n]\!]$. Instead of the condition $d \cdot n_w = g_0$, this induces an influence from the artificial velocity u from Ω_f^C onto the solid displacement. Moreover, if u is driven to zero in Ω_f^C, $\dot{d} \cdot n_w$ goes to zero as well, which might prevent the body from releasing from contact. On the other hand, using the jump of fluxes $\lambda = [\![\widetilde{\sigma}_n]\!]$, we obtain the "pure" contact condition

$$d \cdot n_w = g_0 \tag{4.5}$$

as desired.

The weakly imposed tangential contact condition (3.13) reads

$$\tau_w^T \sigma_s n + \gamma_{\text{fsi}} \dot{d} \cdot \tau_w = 0, \tag{4.6}$$

when considering that the fluid velocity u is driven to zero asymptotically. As $\gamma_{\text{fsi}} \to \infty$ for $h \to 0$, this means (asymptotically) that the solid is not allowed to slide along the line Γ_w. While this might seem restrictive at first sight, this condition is in fact in some sense inherited from the no-slip condition at $\Gamma(t)$ before contact. This is due to the continuity of velocities on $\Gamma_{\text{fsi}}(t)$ and the fact that the velocity is driven to zero in Ω_f^C (and hence on Γ_w). Moreover, the no-slip condition on the fluid part of Γ_w implies that the solid can not slide on the end points $x_{C,1}$ and $x_{C,2}$ of the contact interval (see Fig. 3, right sketch). Altogether, this shows in agreement with a number of theoretical works (*e.g.* [36]) that the no-slip interface conditions are not an appropriate model for the case that it comes to contact with an exterior wall.

Remark 4.2 (Relation of the two contact formulations)**.** The relaxed contact formulation derived in Section 4.1 can also be seen as an extension of the fluid forces σ_f to a region below the contact line (here Γ_ϵ), namely by using the Stokes equations in the extended domain. We will see in the numerical examples below that the two approaches yield similar results. In this way the use of the relaxed contact formulation might be justified also in cases, where real contact with the wall is expected. Moreover, it would be enough to use the extension in the artificial fluid approach only in a small layer of size $\mathcal{O}(\epsilon)$.

Remark 4.3 (Lagrange multiplier formulation)**.** A further possibility would be to keep the Lagrange multiplier λ in the variational formulation (3.7) and (3.8) and to discretise additionally the Lagrange multiplier space, as in [37,51] for the pure FSI case. Due to the difficulties concerning the discrete *inf-sup stability* and the additional computational effort, we will, however, not consider this alternative in the remainder of this work.

4.3. Slip conditions

Motivated by the above considerations, we study slip conditions in this section. The arguments that follow can be applied to the formulations proposed in Sections 3 and 4.

Observe that in the continuous case the discussion of (3.1) and (3.2) remains valid in the case of contact with slip conditions (noting that the tangential stresses vanish on both sides of the fluid-solid interfaces). However, at the discrete level, care has to be taken to use only the normal component of the numerical stress. Indeed, in this case the relation (3.2) becomes

$$(n^T[\![\sigma n]\!])n - \lambda n_w = 0 \quad \text{on} \quad \Gamma(t), \tag{4.7}$$

which yields $\lambda = (n^T[\![\sigma n]\!])(n \cdot n_{\mathrm{w}}) =: [\![\sigma_{n,\mathrm{slip}}]\!]$.

At the discrete level, using the numerical stress this expression translates to

$$\lambda = (n^T[\![\widetilde{\sigma} n]\!])(n \cdot n_{\mathrm{w}}) =: [\![\widetilde{\sigma}_{n,\mathrm{slip}}]\!], \tag{4.8}$$

where the jump operator is given by (3.9). The resulting discrete formulations are given by Variational Formulation 1 or Variational Formulation 2, respectively, with the respective choice of λ.

4.3.1. Weakly imposed interface conditions: The slip case

Let us consider again which interface conditions are implicitly included in the variational formulation,when using the numerical stresses (4.8) for λ. Considering the interface terms with fluid test function v yields as usual the interface conditions $\tau^T \sigma_f n = 0$ and $(\dot{d} - u) \cdot n = 0$. Let us therefore concentrate on the terms with the solid test function w

$$(\sigma_s n, w)_{\Gamma(t)} - (T_f \cdot n, w \cdot n)_{\Gamma(t)} + \gamma_C([\widetilde{P}_\gamma]_+, w \cdot n_w)_{\Gamma(t)} = 0.$$

In the case without contact, the last term vanishes and hence we retrieve the standard consistency of Nitsche's method for fluid-structure interaction with slip conditions. On the other hand, by developing the solid test function w in the local basis of $\Gamma(t)$ we have

$$(\tau^T \sigma_s n, w \cdot \tau)_{\Gamma(t)} + ((n^T[\![\widetilde{\sigma} n]\!]), w \cdot n)_{\Gamma(t)} + \gamma_C([P_\gamma(d, \lambda)]_+, w \cdot n_w)_{\Gamma(t)} = 0.$$

Now, we use the identity $n = (n \cdot n_w)n_w + (n \cdot \tau_w)\tau_w$ to get

$$\begin{aligned}\big(\tau^T \sigma_s n, w \cdot \tau\big)_{\Gamma(t)} + \big((n^T[\![\widetilde{\sigma} n]\!]), (w \cdot \tau_w)(n \cdot \tau_w)\big)_{\Gamma(t)} + \big((n^T[\![\widetilde{\sigma} n]\!]), (w \cdot n_w)(n \cdot n_w)\big)_{\Gamma(t)} \\ + \gamma_C([P_\gamma(d, \lambda)]_+, w \cdot n_w)_{\Gamma(t)} = 0.\end{aligned}$$

Hence, in the case of contact by using the definition of P_γ

$$(\tau^T \sigma_s n, w \cdot \tau)_{\Gamma(t)} + ((n^T[\![\widetilde{\sigma} n]\!]), (w \cdot \tau_w)(n \cdot \tau_w))_{\Gamma(t)} + \gamma_C(d \cdot n_w - g_0, w \cdot n_w)_{\Gamma(t)} = 0. \tag{4.9}$$

When it comes to contact, we have in the asymptotic limit $\tau_w = \tau$ and $n \cdot \tau_w = 0$, so that (4.9) reduces to the "no-friction" condition $\tau^T \sigma_s n = 0$ and the non-penetration condition $d \cdot n_w = g_0$. This is the desired contact condition, as in the no-slip case, see (4.5).

We observe that in (4.9) both the tangential and the normal components (with respect to τ_w and n_w) have asymptotically vanishing perturbations. It is possible to eliminate the perturbation in the normal component, from the solid stress term $(\tau^T \sigma_s n, w \cdot \tau)_{\Gamma(t)}$, by adding the corresponding term to the definition of λ, *i.e.*

$$\lambda = [\![\widetilde{\sigma}_{n,\mathrm{slip}}]\!] + (\tau^T \sigma_s n)(\tau \cdot n_w). \tag{4.10}$$

In this case the relation (4.9) takes the form

$$(\tau^T \sigma_s n, w \cdot \tau_w(\tau \cdot \tau_w))_{\Gamma(t)} + ((n^T[\![\widetilde{\sigma} n]\!]), (w \cdot \tau_w)(n \cdot \tau_w))_{\Gamma(t)} + \gamma_C(d \cdot n_w - g_0, w \cdot n_w)_{\Gamma(t)} = 0.$$

Testing with $w = n_w$ we see that here the non-penetration condition $d \cdot n_w = g_0$ is imposed unperturbed for all $h > 0$. By moving the perturbation to λ as in (4.10) it is instead the form P_γ that is perturbed, which implies a (weakly consistent) perturbation of the contact zone.

5. Stability

In this work, we will use fitted finite elements, *i.e.* we assume that both the interface $\Gamma(t)$ and the wall Γ_w are resolved by mesh lines. The approaches presented, however, can be combined with unfitted finite elements as well, with the only difference that some more stabilisation terms have to be added to the variational formulation. In order to simplify the presentation, we will concentrate on the fitted case first and discuss the extension to unfitted finite elements in a remark afterwards.

We will use equal-order finite elements in combination with a pressure stabilisation term $S_p(p,q)$ for the fluid equations. For the stability analysis, the only requirement on S_p is that it leads to a well-posed discrete fluid problem. Possibilities include the Brezzi–Pitkäranta stabilisation [10], local projections (LPS) [5], the pressure-stabilised Petrov–Galerkin approach (PSPG) [47] or the Continuous Interior Penalty method (CIP) [14]).

In order to present the stability analysis in a general setting, we introduce some further notation splitting the contact force variable λ into a fluid part λ_f and a solid part λ_s, such that $\lambda = \lambda_s - \lambda_f$. For the no-slip case, we define $\lambda_s(w) = \sigma_{s,n}(w)$ and

$$\lambda_f(v,q,w) = \begin{cases} \sigma_{f,n}(v,q), & \text{if } \lambda = [\![\sigma_n]\!], \\ T_f(v,q,w)\cdot n, & \text{if } \lambda = [\![\tilde{\sigma}_n]\!], \end{cases} \tag{5.1}$$

see (3.4) and (3.8). For the slip case, we define $\lambda_s(w) = \sigma_{s,n}(w)(n\cdot n_w)$ and

$$\lambda_f(v,q,w) = \begin{cases} \sigma_{f,n}(v,q)(n\cdot n_w), & \text{if } \lambda = [\![\sigma_{n,\text{slip}}]\!], \\ (T_f(v,q,w)\cdot n)(n\cdot n_w), & \text{if } \lambda = [\![\tilde{\sigma}_{n,\text{slip}}]\!], \end{cases} \tag{5.2}$$

see (4.7) and (4.8).

5.1. Generalised contact formulation

Before we conduct the stability analysis, let us introduce a generalised contact formulation, following the ideas of Chouly *et al.* [22]. We have already briefly discussed their ideas for the case of a pure solid with contact in (2.14). The generalisation of (2.14) to the FSI-contact system (Variational Formulation 2) would be to add the terms

$$-\left(\gamma_C[P_\gamma(\lambda,d)]_+ + \lambda, \lambda(\partial_t v, \partial_t q, w)\right)_{\Gamma(t)}. \tag{5.3}$$

The time derivatives on the test functions v and q are motivated by the stability analysis below, where we have to test the variational form with $v=u$, $q=p$ and $w=\partial_t d$, in order to show stability (see also Burman and Fernández [13]).

On the other hand, the term (5.3) is not usable within a time-stepping scheme due to the time derivatives on the test functions. A remedy is to shift the time derivatives to the first integrand (ignoring the boundary terms), *i.e.* adding the consistent terms

$$-\left(\gamma_C[P_\gamma(\lambda,d)]_+ + \lambda, \lambda_s(w)\right)_{\Gamma(t)} - \left(\partial_t\left(\gamma_C[P_\gamma(\lambda,d)]_+ + \lambda\right), \lambda_f(v,q,w)\right)_{\Gamma(t)}.$$

This yields the variational formulation:

Variational Formulation 3. Find $u(t)\in\mathcal{V}_h, p(t)\in\mathcal{Q}_h, d(t)\in\mathcal{W}_h$, such that $\dot{d}=\partial_t d$ and

$$\begin{aligned} \mathcal{A}(u,p,d)(v,q,w) :=\ & \mathcal{A}^*_{*,FSI}(u,p,d)(v,q,w) + \gamma_C\left([P_\gamma(\lambda,d)]_+, w\cdot n_w\right)_{\Gamma(t)} - \theta\left(\gamma_C[P_\gamma(\lambda,d)]_+ + \lambda, \lambda_s(w)\right)_{\Gamma(t)} \\ & - \theta\left(\partial_t\left(\gamma_C[P_\gamma(\lambda,d)]_+ + \lambda\right), \lambda_f(v,q,w)\right)_{\Gamma(t)} \\ =\ & (f_f, v)_{\Omega_f(t)} + (f_s, w)_{\Omega_s(t)} \quad \forall v,q,w\in\mathcal{V}_h\times\mathcal{Q}_h\times\mathcal{W}_h, \end{aligned} \tag{5.4}$$

where $\mathcal{A}^*_{*,\text{FSI}}$ is one of the bilinear forms $\mathcal{A}_{\text{no-slip,FSI}}, \mathcal{A}_{\text{slip,FSI}}, \mathcal{A}^C_{\text{no-slip,FSI}}$ or $\mathcal{A}^C_{\text{slip,FSI}}$.

5.2. A stability result

In this section, we will investigate stability of the discrete formulation for different values of θ. In particular, we will show a stability result for the symmetric formulation ($\theta = 1$) and stability up to a specific term in the general case (including $\theta = 0$). These results correspond to the results that have been obtained by Chouly *et al.* [22] for the pure solid case. For the stability analysis, we will assume infinitesimal displacements, *i.e.* the sub-domains Ω_f and Ω_s as well as the interface Γ are fixed. We introduce the notation $\widetilde{\Omega}_f = \Omega_f \cup \Omega_f^C$ for the combined fluid and artificial fluid domain and set $\Omega_f^C = \emptyset$ for the relaxed approach.

Theorem 5.1. *Let $u, p, d \in V_h$ and $\dot{d} = \partial_t d$. We have the following stability result for the form $\mathcal{A}$ defined in (5.4), where $\theta \in [-1, 1]$ and γ_C^0 sufficiently large*

$$\begin{aligned}
&\|u(T)\|^2_{\widetilde{\Omega}_f} + \|\dot{d}(T)\|^2_{\Omega_s} + \|d(T)\|^2_{H^1(\Omega_s)} + \int_0^T \left(\nu_f \|\nabla u\|^2_{\widetilde{\Omega}_f} + S_p(p,p) + \gamma_a \|u\|^2_{\Omega_f^C}\right) \mathrm{d}t \\
&\quad + \int_0^T \gamma_{fsi} \|(\dot{d} - u)\cdot n\|^2_\Gamma \,\mathrm{d}t + \theta \left\| \gamma_C^{-1/2} \lambda(T) + \gamma_C^{1/2} [P_\gamma(\lambda, d)]_+(T) \right\|^2_\Gamma \\
&\quad \leq C \Bigg(\int_0^T \mathcal{A}(u,p,d;u,p,\dot{d}) \; - (1-\theta)\gamma_C \left([P_\gamma(\lambda,d)]_+, \dot{d}\cdot n_w \right)_\Gamma \mathrm{d}t + \|u_0\|^2_{\widetilde{\Omega}_f} \\
&\qquad + \|\dot{d}_0\|^2_{\Omega_s} + \|d_0\|^2_{H^1(\Omega_s)} + \theta \left\| \gamma_C^{-1/2} \lambda_0 + \gamma_C^{1/2} [P_\gamma(\lambda_0, d_0)]_+ \right\|^2_\Gamma \Bigg),
\end{aligned}$$

where we have used the abbreviation $\lambda_0 := \lambda(u_0, p_0, d_0)$. For the no-slip case, the term $\|(\dot{d} - u)\cdot n\|_\Gamma$ can be replaced by $\|(\dot{d} - u)\|_\Gamma$.

Remark 5.2 (Contact terms)**.** The second line gives us control over the satisfaction of the FSI-contact condition for $\theta > 0$. In contrast to the work by Chouly *et al.* [22] for a pure solid problem, we obtain here discrete stability for $\theta = 1$, for the following positive discrete energy

$$E(T) := \|u(T)\|^2_{\widetilde{\Omega}_f} + \|\dot{d}(T)\|^2_{\Omega_s} + \|d(T)\|^2_{H^1(\Omega_s)} + \left\| \gamma_C^{-1/2} \lambda(T) + \gamma_C^{1/2} [P_\gamma(\lambda, d)]_+(T) \right\|^2_\Gamma .$$

For $\theta \neq 1$ on the other hand, the contact term $(1-\theta)\gamma_C \left([P_\gamma(\lambda,d)]_+, \dot{d}\cdot n_w \right)_\Gamma$ appears on the right-hand side. For $\theta < 0$ we have additionally a negative contribution on the left-hand side. Both issues are directly inherited from the pure solid mechanical case, see [21]. The last term on the right-hand side vanishes, if we assume that the contact conditions are fulfilled at initial time, for example if the solid is not in contact with at $t = 0$.

Proof. We test (5.4) with $w = \dot{d} = \partial_t d, v = u$ and $q = p$ and integrate in time. We start by deriving a lower bound for $\mathcal{A}^*_{*,\mathrm{FSI}}$. For the fluid part, we use the techniques from Burman and Fernandez [13], to show coercivity of the Stokes part including the coupling terms. For the no-slip case, the authors have shown that

$$\begin{aligned}
&a_f^C(u,p;u,p) + \gamma_a (u,u)_{\Omega_f^C} - (\sigma_f(u,p)n, \dot{d} - u)_\Gamma - (\dot{d} - u, \sigma_f(u,-p)n)_\Gamma + \gamma_{\mathrm{fsi}} \|\dot{d} - u\|^2_\Gamma \\
&\qquad \geq c \left(\nu_f \|\nabla u\|^2_{\widetilde{\Omega}_f} + \gamma_{\mathrm{fsi}} \|\dot{d} - u\|^2_\Gamma + S_p(p,p) + \gamma_a \|u\|^2_{\Omega_f^C} \right).
\end{aligned}$$

Analogously, one can show in the slip-case that

$$\begin{aligned}
&a_f^C(u,p;u,p) + \gamma_a (u,u)_{\Omega_f^C} - (n^T \sigma_f(u,p)n, (\dot{d} - u)\cdot n)_\Gamma - ((\dot{d} - u)\cdot n, n^T \sigma_f(u,-p)n)_\Gamma + \gamma_{\mathrm{fsi}} \|(\dot{d} - u)\cdot n\|^2_\Gamma \\
&\qquad \geq c \left(\nu_f \|\nabla u\|^2_{\widetilde{\Omega}_f} + \gamma_{\mathrm{fsi}} \|(\dot{d} - u)\cdot n\|^2_\Gamma + S_p(p,p) + \gamma_a \|u\|^2_{\Omega_f^C} \right).
\end{aligned}$$

Using the symmetry of σ_s, integration in time and a Korn's inequality, we obtain for the solid part

$$\int_0^T (\sigma_s(d), \nabla \dot{d})_{\Omega_s} \,\mathrm{d}t = \frac{1}{2}\int_0^T \partial_t (\sigma_s(d), \nabla d)_{\Omega_s} \,\mathrm{d}t = \frac{1}{2}\left((\sigma_s(d(T)), \nabla d(T))_{\Omega_s} - (\sigma_s(d(0)), \nabla d(0))_{\Omega_s}\right)$$
$$\geq c_1 \|\nabla d(T)\|^2_{\Omega_s} - c_2 \|\nabla d_0\|^2_{\Omega_s}.$$

Moreover, we have

$$\int_0^T (\partial_t u, u)_{\widetilde{\Omega}_f} + (\partial_t \dot{d}, \dot{d})_{\Omega_s} \,\mathrm{d}t = \frac{1}{2}\left(\|u(T)\|^2_{\widetilde{\Omega}_f} + \|\dot{d}(T)\|^2_{\Omega_s} - \|u_0\|^2_{\widetilde{\Omega}_f} - \|\dot{d}_0\|^2_{\Omega_s}\right).$$

Together, we have shown that

$$\|u(T)\|^2_{\widetilde{\Omega}_f} + \|\dot{d}(T)\|^2_{\Omega_s} + \|d(T)\|^2_{H^1(\Omega_s)} + \int_0^T \nu_f \|\nabla u\|^2_{\widetilde{\Omega}_f} + S_p(p,p) + \gamma_a \|u\|^2_{\Omega_f^C} + \gamma_{\mathrm{fsi}} \|\dot{d} - u\|^2_{\Gamma} \,\mathrm{d}t$$
$$\leq C\left(\int_0^T \mathcal{A}^*_{*,FSI}(u,p,d;u,p,\dot{d}) \,\mathrm{d}t + \|u_0\|^2_{\widetilde{\Omega}_f} + \|\dot{d}_0\|^2_{\Omega_s} + \|d_0\|^2_{H^1(\Omega_s)}\right).$$

Let us now estimate the contact terms. We split the principal contact term into

$$\int_0^T \gamma_C \left([P_\gamma(\lambda, d)]_+, \dot{d} \cdot n_w\right)_\Gamma \,\mathrm{d}t = \int_0^T \theta\gamma_C \left([P_\gamma(\lambda, d)]_+, \dot{d} \cdot n_w\right)_\Gamma + (1-\theta)\gamma_C \left([P_\gamma(\lambda, d)]_+, \dot{d} \cdot n_w\right)_\Gamma \,\mathrm{d}t.$$

We have to estimate the terms

$$\theta \int_0^T \underbrace{\gamma_C \left([P_\gamma(\lambda, d)]_+, \dot{d} \cdot n_w\right)_\Gamma}_{I_1} - \underbrace{\left(\gamma_C [P_\gamma(\lambda, d)]_+ + \lambda, \lambda_s(\dot{d})\right)_\Gamma}_{I_2}$$
$$- \underbrace{\left(\partial_t \left(\gamma_C [P_\gamma(\lambda, d)]_+ + \lambda\right), \lambda_f(u, p, \dot{d})\right)_\Gamma}_{I_3} \,\mathrm{d}t. \tag{5.5}$$

From the definition of P_γ we can write $d \cdot n_w = P_\gamma(\lambda, d) + \gamma_C^{-1}\lambda$. Hence, since the lower wall is assumed to be time independent, we have

$$\int_0^T I_1 \,\mathrm{d}t = \gamma_C \int_0^T \left([P_\gamma(\lambda, d)]_+, \partial_t (P_\gamma(\lambda, d) + \gamma_C^{-1}\lambda)\right)_\Gamma \,\mathrm{d}t$$
$$= \gamma_C \int_0^T \left([P_\gamma(\lambda, d)]_+, \partial_t ([P_\gamma(\lambda, d)]_+ + \gamma_C^{-1}\lambda)\right)_\Gamma \,\mathrm{d}t.$$

In the second line, we have used that [22]

$$\frac{1}{2}\partial_t [\phi]^2_+ = [\phi]_+ \partial_t [\phi]_+ = [\phi]_+ H(\phi) \partial_t [\phi]_+ = [\phi]_+ \partial_t \phi$$

where H is the Heaviside function. We insert $\pm\gamma_C^{-1}\lambda$ and integrate by parts

$$\begin{aligned}\int_0^T I_1\,\mathrm{d}t &= \gamma_C\int_0^T \big([P_\gamma(\lambda,d)]_+ + \gamma_C^{-1}\lambda, \partial_t([P_\gamma(\lambda,d)]_+ + \gamma_C^{-1}\lambda)\big)_\Gamma\,\mathrm{d}t \\ &\quad - \int_0^T \underbrace{\big(\lambda, \partial_t([P_\gamma(\lambda,d)]_+ + \gamma_C^{-1}\lambda)\big)_\Gamma}_{I_4}\,\mathrm{d}t \\ &= \underbrace{\frac{\gamma_C}{2}\|[P_\gamma(\lambda,d)(T)]_+ + \gamma_C^{-1}\lambda(T)\|_{0,\Gamma}^2}_{D_T} - \underbrace{\frac{\gamma_C}{2}\|[P_\gamma(\lambda_0,d_0)]_+ + \gamma_C^{-1}\lambda_0\|_{0,\Gamma}^2}_{D_0} - \int_0^T I_4\,\mathrm{d}t.\end{aligned}$$

Now the idea is to split the contribution from the term I_4 into fluid and solid stresses and to apply integrating by parts in time (only) in the solid stress contribution. By definition, we have

$$-\int_0^T I_4\,\mathrm{d}t = -\int_0^T \big(\lambda_s(d), \partial_t([P_\gamma(\lambda,d)]_+ + \gamma_C^{-1}\lambda)\big)_\Gamma + \big(\lambda_f(u,p,\dot d), \partial_t([P_\gamma(\lambda,d)]_+ + \gamma_C^{-1}\lambda)\big)_\Gamma\,\mathrm{d}t$$

and, by integrating by parts in the first term of the right-hand side, we have

$$\begin{aligned}-\int_0^T I_4\,\mathrm{d}t &= \underbrace{\int_0^T \big(\lambda_s(\dot d), [P_\gamma(\lambda,d)]_+ + \gamma_C^{-1}\lambda\big)_\Gamma\,\mathrm{d}t}_{I_5} \underbrace{-\big(\lambda_s(d(T)), [P_\gamma(\lambda,d)(T)]_+ + \gamma_C^{-1}\lambda(T)\big)_\Gamma}_{I_6} \\ &\quad + \underbrace{\big(\lambda_s(d_0), [P_\gamma(\lambda_0,d_0)]_+ + \gamma_C^{-1}\lambda_0\big)_\Gamma}_{I_7} + \underbrace{\int_0^T \big(\lambda_f(u,p,\dot d), \partial_t([P_\gamma(\lambda,d)]_+ + \gamma_C^{-1}\lambda)\big)_\Gamma\,\mathrm{d}t}_{I_8}.\end{aligned}$$

The terms I_5 and I_8 cancel with the terms I_2 and I_3 in (5.5). The term I_6 is treated in a standard fashion using Young's inequality, an inverse inequality and the dissipation provided by D_T and the elastic energy $\frac{1}{2}a^s(d(T),d(T))$ for γ_C^0 sufficiently large. For $\theta > 0$, we estimate

$$\begin{aligned}I_6 &\geq -\gamma_C^{-1}\|\lambda_s(d(T))\|_{0,\Gamma}^2 - \frac{\gamma_C}{4}\|[P_\gamma(\lambda(T),d(T))]_+ + \gamma_C^{-1}\lambda(T)\|_{0,\Gamma}^2 \\ &\geq -\frac{1}{4}\|d(T)\|_{H^1(\Omega_s)}^2 - \frac{1}{2}D_T.\end{aligned}$$

In the same way, we get the upper bound

$$I_6 \leq \frac{1}{4}\|d(T)\|_{H^1(\Omega_s)}^2 + \frac{1}{2}D_T,$$

which is needed in the case $\theta < 0$. For I_7 we obtain analogously

$$-\frac{1}{4}\|d_0\|_{H^1(\Omega_s)}^2 - \frac{1}{2}D_0 \;\leq\; I_7 \;\leq\; \frac{1}{4}\|d_0\|_{H^1(\Omega_s)}^2 + \frac{1}{2}D_0,$$

which completes the proof. □

Remark 5.3 (Unfitted finite elements)**.** When using unfitted finite elements [13, 15], additional stabilisation terms S_u and S_d are needed, if the interface $\Gamma(t)$ is not resolved by mesh lines. Their purpose is to extend the coercivity of the fluid system from Ω_f (resp. Ω_s) to the extended domains Ω_h^f (resp. Ω_h^s) that consists of all element $T \in \mathcal{T}_h$, with a non-empty intersection with the respective sub-domain ($T \cap \Omega_i \neq \emptyset$). Suitable "ghost penalty" operators have been defined in Burman [12]. The same stability result as in Theorem 5.1 can then be shown with an analogous argumentation.

Remark 5.4 (Newton convergence)**.** While the symmetric formulation ($\theta = 1$) seems beneficial from the theoretical point of view, the additional terms in (5.4) can cause severe difficulties for the non-linear solver. The reason is that the additional contact terms are not only highly non-linear, but also non-smooth, especially due to the time derivative acting on the maximum operator $\partial_t[\cdot]_+$. In our numerical tests, we were not able to obtain numerical convergence for different versions of generalised Newton methods [52]. The investigation of the case $\theta \neq 1$ by means of numerical tests and in particular the construction of a robust non-linear solver are subject to future research.

6. Numerical results

In this section, we show some numerical results to analyse and to compare the different contact formulations. As mentioned in Remark 5.4, we were not able to obtain results for the generalised contact formulation with $\theta \neq 0$, due to divergence of the generalised Newton-type methods we have tried. Therefore, we only show results for $\theta = 0$, where Newton convergence was not an issue, at least when the time step δt was chosen reasonably small. Following the standard approach for contact in solid mechanics, we could in this case simply ignore the non-differentiability of the maximum operator when computing the Newton derivatives, as the term P_γ inside the bracket $[\cdot]_+$ is in practice typically never exactly zero. For all other values of P_γ the derivatives are well-defined. In the computations made for this paper, the Newton algorithm needed 1–2 iterations per time step to reduce the initial residual by a factor of 10^{-7}, if the contact force was not getting active during the iteration, and 1–5 iterations per time step in and around the interval of contact. This makes the method highly competitive in terms of computational costs compared to approaches using Lagrange multipliers and/or active-sets.

We first give some details in Section 6.1 on the fitted, equal-order finite element discretisation and the stabilisations we use. Then, in Section 6.2, we study the problem of a virtual obstacle within the fluid domain introduced in Section 3. The purpose of this example is to isolate the effect of the contact terms from issues related to discretisation during contact and the topology change in the fluid domain $\Omega_f(t)$. Then, we study in Section 6.3 a model problem with contact with the boundary of the fluid domain, where we compare among other aspects the two contact formulations introduced in Section 4, the different possibilities to choose the fluxes λ and the effect of slip and no-slip boundary and interface conditions.

6.1. Details on discretisation and stabilisation

For the numerical results in this paper, we will use a monolithic *Fully Eulerian* approach on a global mesh $\mathcal{T}_h$ covering $\Omega(t)$. In order to resolve the interface $\Gamma(t)$ within the discretisation, we use the *locally modified finite element method* introduced by Frei and Richter [32]. The idea of this approach is to use a fixed coarse triangulation $\mathcal{T}_{2h}$ of the overall domain $\Omega = \widetilde{\Omega}_f(t) \cup \Gamma(t) \cup \Omega_s(t)$ that is independent of the position of the interface $\Gamma(t)$. Then, in each time step, this coarse grid is refined once by splitting each so-called "patch" element in either eight triangles or four quadrilaterals to resolve the interface in at least a linear approximation, see Figure 4 for an illustration.

The finite element space V_h is then defined as a combination of piece-wise linear and piece-wise bi-linear finite elements on the patches. It can be guaranteed that a maximum angle condition is fulfilled in each of the sub-cells, leading to optimal-order interpolation and error estimates [32].

For temporal discretisation, we split the time interval I into m equidistant-distant time intervals $I_j = (t_{j-1}, t_{j-1} + k]$ and use a time-stepping scheme that is based on a modified discontinuous Galerkin time discretisation of lowest order (dG(0)), see Frei and Richter [34]. The displacement-velocity relation $\partial_t d = \dot{d}$ is included by means of the L^2-projection

$$(\partial_t d, z)_{\Omega_s(t)} - (\dot{d}, z)_{\Omega_s(t)} = 0 \quad \forall z \in V_h^s,$$

where V_h^s denotes the (modified) finite element space that is spanned by the degrees of freedom of the elements in the solid part Ω_h^s.

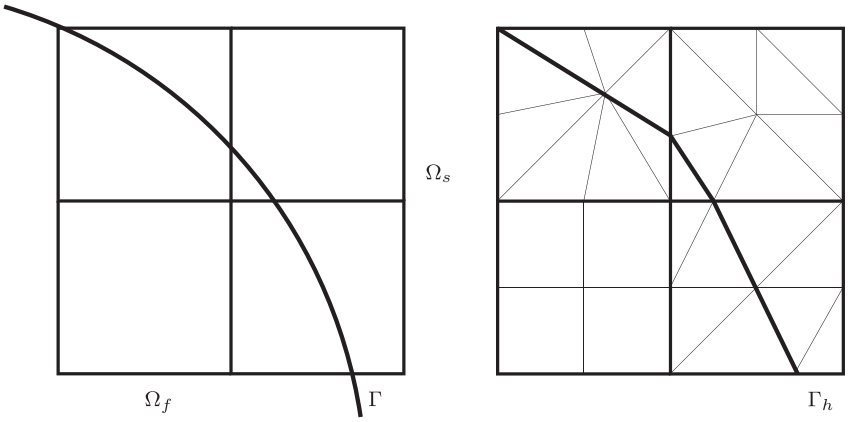


FIGURE 4. *Left:* fixed triangulation $\mathcal{T}_{2h}$ of the domain Ω. *Right:* subdivision of the patches $P \in \mathcal{T}_{2h}$ such that the interface $\Gamma(t)$ is resolved in a linear approximation by the discrete interface Γ_h.

The domain affiliation of a point $x \in \Omega(t)$ is determined by means of the *Initial Point Set/Backward Characteristics method* [26,29], that uses the displacement $d(t)$ in the solid domain and an extension to $\Omega_f(t)$ in order to trace back points to their initial position in $\Omega(0)$, following the definition (2.1).

For pressure stabilisation, we use an anisotropic variant of the Continuous Interior Penalty method, see Frei [30,31]. In addition, we add the temporal pressure stabilisation term

$$S_{pt}(p,q) = \gamma_{pt} h (p^m - p^{m-1}, q)_{\Gamma(t)}$$

in each time interval I_m. This additional stabilisation is needed, as the mesh $\mathcal{T}_h(t_m)$, and hence the finite element spaces, change from time-step to time-step. The solution u^{m-1} from the previous time-step t_{m-1} is therefore not discrete divergence-free with respect to the new mesh $\mathcal{T}_h(t_m)$, which gives rise to pressure oscillations, see for example Besier and Wollner [7].

All the following results have been obtained using the finite element library *Gascoigne 3d* [6].

6.2. Virtual obstacle within the fluid domain

We begin by investigating the simplified problem introduced in Section 3. The initial fluid and solid domains are defined as

$$\Omega_f(0) = (0,1) \times (0,0.5), \qquad \Omega_s(0) = (0,1) \times (0.5,0.6)$$

and a lower-dimensional obstacle $\Gamma_w = (0,1) \times 0.25$ within the fluid domain. We consider a moving interface $\Gamma(t)$, which is resolved using the locally modified finite element method. The sub-domains $\Omega_s(t)$ and $\Omega_f(t)$ and the interface $\Gamma(t)$ depend on the solid displacement $d(t)$, see (2.1).

The constraint for the solid displacement is given by

$$d \cdot n_w \leq 0.25 \quad (=: g_\alpha). \tag{6.1}$$

We use the elasticity parameters $\lambda_s = \mu_s = 2 \times 10^6$ and the fluid viscosity $\nu_f = 1$. The structure is pulled towards the bottom by fluid forces due to a prescribed pressure mean value at the left and right boundary of the fluid domain

$$\int_{\Gamma_{f,\text{left}}} p \, \mathrm{d}s = \int_{\Gamma_{f,\text{left}}} \overline{P} \, \mathrm{d}s, \qquad \int_{\Gamma_{f,\text{right}}} p \, \mathrm{d}s = \int_{\Gamma_{f,\text{right}}} \overline{P} \, \mathrm{d}s$$

where $\overline{P} := 1.3 \times 10^5$. We consider the Variational Formulation 1 with no-slip conditions and $\lambda = [\![\tilde{\sigma}_n]\!]$, see (3.8), on a Cartesian mesh that consists of 5120 elements and with a small time step $\delta t = 10^{-5}$. The Nitsche constant at the FSI interface is chosen $\gamma^0_{\text{fsi}} = 10^3$ and temporal pressure stabilisation with $\gamma_{pt} = 10^{-2}$ is used.

To analyse the results, we define the "minimal distance"

$$d_{\min} := \min_{x \in \Gamma(t)} x_2 - 0.25$$

of the interface $\Gamma(t)$ to Γ_w. To be precise the term "minimal distance" is only correct before contact, as $d_{\min}$ gets negative in case of an overlap. Moreover, we define the following functionals in order to analyse the pressure p and the contact force at the interface $\Gamma(t)$

$$J_p := \left| \int_{\Gamma(t)} p \, \mathrm{d}s \right|, \qquad J_{P_\gamma} := \gamma_C \int_{\Gamma(t)} [P_\gamma(\lambda, d)]_+ \, \mathrm{d}s.$$

In Figure 5, we plot these three functionals over time for two contact periods and the contact parameters $\gamma^0_C = 10, 10^2, 10^3$. In the top left plot, we observe that the solid is pulled down until it reaches Γ_w at $t \approx 1.5 \times 10^{-3}$. After a short *contact* period, it is released again due to its elastic properties before it reaches the obstacle for a second time at $t \approx 5 \times 10^{-3}$.

The contact condition $d \cdot n_w \leq 0.25$ is only significantly violated for the smallest contact parameter, where $d_{\min}$ reaches a minimum value of around -8×10^{-4}, see the zoom-in around the contact interval on the top left. This value is more than an order of magnitude smaller than the mesh size in vertical direction $h \approx 1.4 \times 10^{-2}$. For the larger values of γ^0_C, the minimal value of $d_{\min}$ is even much closer to zero. On the other hand, we observe that even for the largest value $\gamma^0_C = 10^3$, the contact condition is slightly relaxed, allowing for very small overlaps of solid and contact line.

In the second row of Figure 5, we observe that the pressure shows a peak at the beginning ($t \approx 1.5 \times 10^{-3}$ and $t \approx 5 \times 10^{-3}$) of the contact periods, followed by some small oscillations. The peak is caused by the fluid dynamics and will be discussed below. The oscillations get smoother for larger values of γ^0_C and are barely visible for $\gamma^0_C = 10^3$.

Similarly, the contact force J_{P_γ} shows oscillations for $\gamma^0_C = 10$ and a much smoother behaviour for $\gamma^0_C \geq 10^2$. Note that this does not contradict the stability result in Theorem 5.1, where we have assumed that γ^0_C is large enough. The relatively large value for γ^0_C that is needed here is due to the anisotropic cells that appear in some of the time-steps, when using the *locally modified finite element method*. In the absence of extreme anisotropies a value of $\gamma^0_C \approx 1$ seems to be enough to obtain stable numerical results. The optimal choice of the contact parameter γ_C in the context of anisotropic cells is subject to future research.

If γ^0_C is chosen large enough, the contact force is roughly of the same size for different γ^0_C. This is in agreement with the observations of Chouly *et al.* for the case of a pure solid problem [23], who showed that the consistency of the method makes the choice of the contact parameter much less sensitive compared to a pure penalty method.

6.2.1. Investigation of the pressure peak

The pressure peak at the beginning of the contact interval can be explained as follows. As the fluid does not "see" the obstacle before reaching it, the solid is pulled down towards it without reducing its velocity. At the moment when the obstacle is reached, its vertical velocity $\dot{d} \cdot n_w$ has to decrease to zero in an instant. Due to the continuity of velocities, the same happens for the fluid velocity $u \cdot n_w$ at the interface, and due to the incompressibility constraint the velocity has to change globally in the fluid domain $\Omega_f(t)$. The pressure can be seen as a Lagrange multiplier and more specifically as sensitivity of a (jumping) energy functional with respect to the incompressibility constraint, which explains the peak.

To substantiate this explanation numerically, we add an artificial penalty for the velocity on the sub-domain Ω^0_f below the contact line

$$S_a(u, v) := \gamma^0_a h^{-2} (u, v)_{\Omega^0_f}.$$

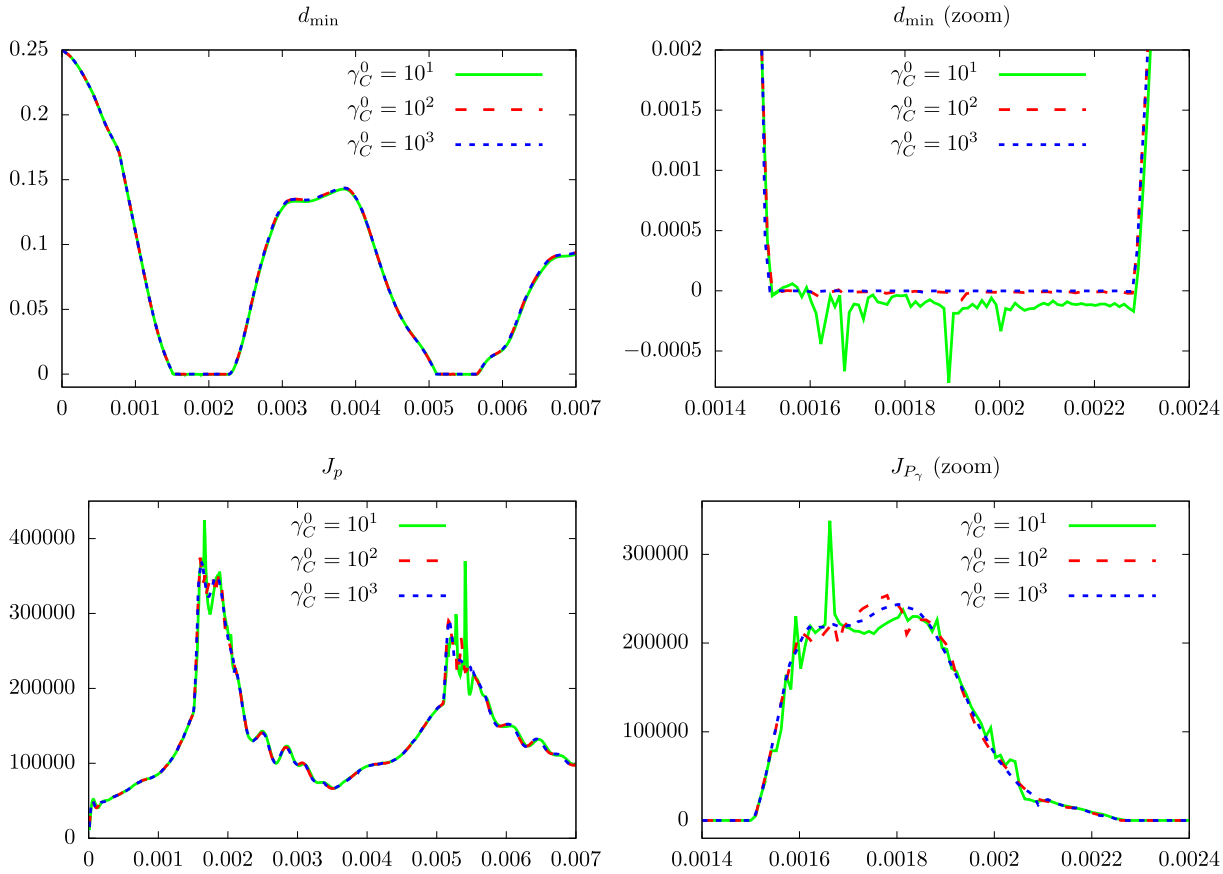


FIGURE 5. *Top row*: "Minimal distance" $d_{\min}$ to Γ_w over time with two contact periods. *Left*: total time interval. *Right*: zoom-in at the first contact interval. *Bottom left*: pressure functional J_p over time. *Bottom right*: functional J_{P_γ} measuring the contact force around the contact interval over time.

For $\gamma_a := \gamma_a^0 h^{-2} \to \infty$, the fluid velocity is driven to zero below the obstacle. As this is already the case before contact, no abrupt changes in the fluid velocity are expected at the moment of the impact. Note that the problem with $\gamma_a^0 > 0$ is purely artificial, as the pressure mean values are still applied on the whole fluid boundary, including the boundary of Ω_f^0.

In Figure 6 we compare the minimal distance and pressure functionals for computations without penalty ($\gamma_a^0 = 0$) to results for $\gamma_a^0 = 10$. First, we note that the vertical displacement is significantly influenced by the penalty, which has to be expected as the fluid dynamics are altered. Contact happens later at $t \approx 0.0022$ with the artificial penalty. Moreover, we observe indeed that the initial pressure peak at the time of impact is significantly reduced for $\gamma_a^0 = 10$.

6.3. Contact problem

Next, we study a problem, where it comes to real contact with the wall $\Gamma_w = \{(x, y) \in \Omega, y = 0.25\}$. At time $t = 0$, we define

$$\Omega_f(0) := (0, 1) \times (0.25, 0.5), \quad \Gamma(0) := (0, 1) \times 0.5, \quad \Omega_s(0) := (0, 1) \times (0.5, 0.6).$$

Below Γ_w, we define a fixed artificial fluid domain $\Omega_f^C := (0, 1) \times (0, 0.25)$.

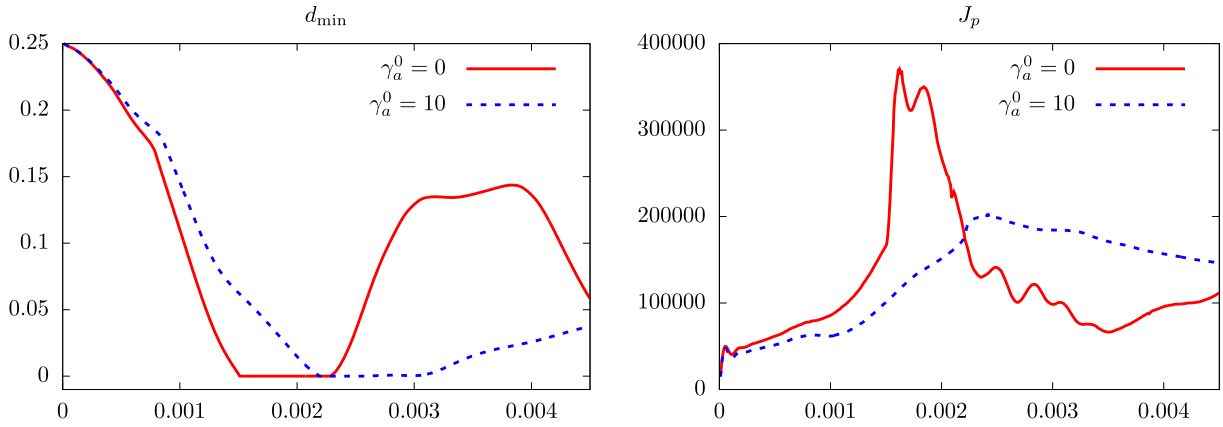


FIGURE 6. Minimal distance $d_{\min}$ to the contact line Γ_w (*left*) and mean pressure over the interface J_p (*right*), plotted over time with an artificial penalty for the velocity below the virtual obstacle.

We apply again a pressure mean value $\overline{P}$ on the lateral boundaries $\Gamma_{f,\text{left}}$ and $\Gamma_{f,\text{right}}$ of the fluid domain $\Omega_f(t)$. As the size of Γ_{left} and Γ_{right} is smaller and the viscous fluid forces acting against the closure of the channel are stronger than in the previous example, we have to set a larger pressure force $\overline{P} = 3 \times 10^5$ in order to obtain contact. On the other hand, the contact was never released again in our numerical experiments, when we used this constant boundary force for all times. Therefore, we decrease $\overline{P}$ linearly from $t = 10^{-3}$ on until it reaches zero at $t = 1.2 \times 10^{-3}$. In order to avoid the issues related to no-slip conditions and contact, we use slip-interface conditions first, *i.e.* the Variational Formulation 2 with $\mathcal{A}^C_{\text{slip,FSI}}$. Unless stated differently, λ is chosen as the jump of numerical stresses $[\![\widetilde{\sigma}_{n,\text{slip}}]\!]$. Moreover, we use again a Cartesian mesh that consists of 5120 elements, a time step $\delta t = 10^{-5}$ and temporal pressure stabilisation with $\gamma_{pt} = 10^{-2}$. Unless explicitly stated, the Nitsche parameters are chosen as $\gamma^0_{\text{fsi}} = \gamma^0_C = 10^3$ and the penalty in the artificial fluid as $\gamma^0_a = 10^2$.

The results on a coarser mesh are illustrated in Figure 7 at four time instants. Contact happens after the pressure on the lateral boundaries is released, as the solid continues moving downwards for some time. During contact, there is a very small overlap of the solid with the artificial fluid Ω^C_f. As the overlap is of order 10^{-5}, it can barely be seen in the bottom left picture. Notice however the triangular cells in Ω^C_f that are used only, when a patch is cut by $\Gamma(t)$.

6.3.1. Comparison of the two contact formulations

First, we compare the two contact strategies derived in Section 4, *i.e.* the relaxed contact formulation introduced in Section 4.1 with a small gap of size $\epsilon(h) = h/10$ between the solid and Γ_w and the strategy using an artificial fluid derived in Section 4.2. A comparison of the results for $d_{\min}$, the pressure norm $\|p\|_{L^2(\Omega_{f,\text{mid}}(t))}$, where

$$\Omega_{f,\text{mid}}(t) := \{x \in \Omega_f(t), 0.4 \leq x_1 \leq 0.6\}$$

denotes the central part of the fluid domain, and the contact force J_{P_γ} are shown in Figure 8 on two different meshes with 5120 and 20 480 elements, respectively.

First, we observe from the plots in the top row that the interface stays at a distance to Γ_w of about $\epsilon \approx h/10 \approx 1.4 \times 10^{-3}$ on the coarser and $\epsilon \approx h/10 \approx 7 \times 10^{-4}$ on the finer mesh for the relaxed formulation. The much smaller overlap with Ω^C_f in the artificial fluid formulation is not visible, not even in the zoom-in on the right.

While the curves for $d_{\min}$ look similar in the global picture (left), the zoom-in shows significant differences already before the impact. The contact happens earlier for the artificial fluid formulation: on the coarser mesh

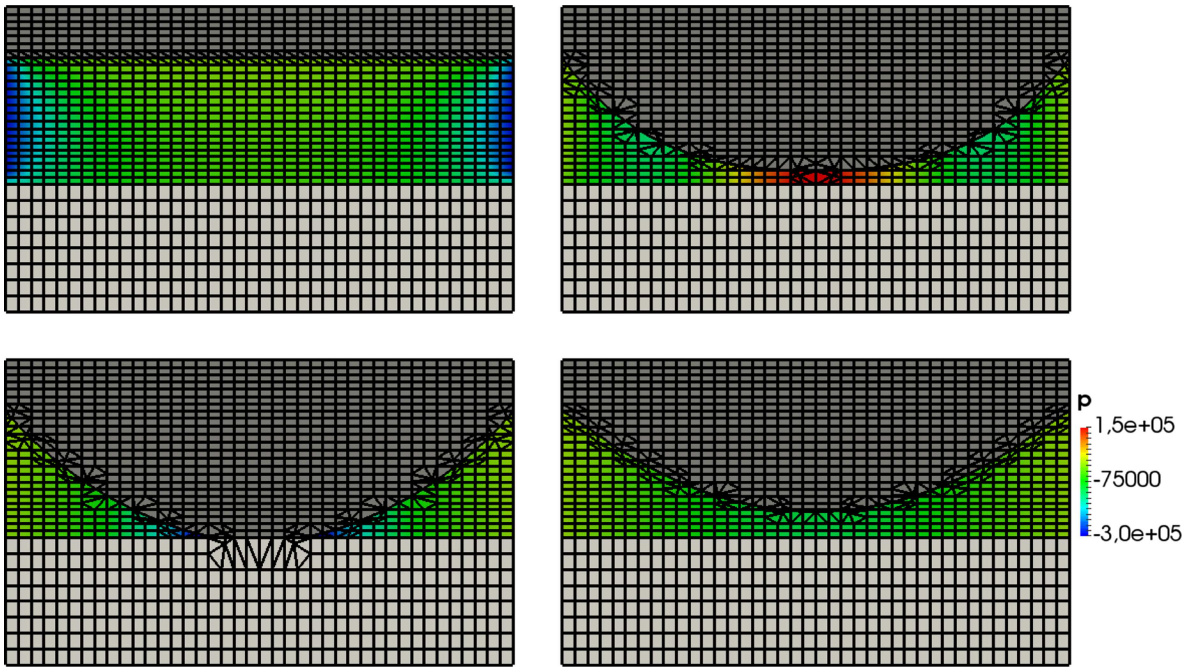


FIGURE 7. Illustration of the contact problem at four time instances on a coarse mesh: $t = 0$ (*top left*), $t = 1.2 \times 10^{-3}$ (*top right*), $t = 2 \times 10^{-3}$ (*bottom left*) and $t = 2.5 \times 10^{-3}$ (*bottom right*). The grey part corresponds to the structure $\Omega_s(t)$, the white part is the artificial fluid Ω_f^C. In the fluid domain $\Omega_f(t)$, values of the pressure p are visualised.

at time $t_{C,a} = 1.87 \times 10^{-3}$ compared to $t_{C,r} = 2.02 \times 10^{-3}$ for the relaxed formulation. This deviation is already much smaller on the finer mesh, where $t_{C,r} - t_{C,a} = 2 \times 10^{-5}$. The reason for this deviation is that in the artificial fluid formulation the wall Γ_w is only asymptotically for $\gamma_a \to \infty$ impermeable for the fluid. Therefore, in practice, the fluid forces acting against the contact, in particular the pressure p, are smaller for this formulation. As $\gamma_a = \gamma_a^0 h^{-2} \to \infty$ for $h \to 0$ the difference is significantly reduced on the finer mesh.

To substantiate this explanation, we plot the pressure norm $\|p\|_{L^2(\Omega_{f,\mathrm{mid}}(t))}$ over the central part of the fluid domain $\Omega_{f,\mathrm{mid}}(t)$ and the time period before the impact in the bottom left figure. The maximum value of the norm on the coarser mesh at time $t = 1.1 \times 10^{-3}$ is approximately 20.563 for the relaxed formulation and about 19.065 for the artificial fluid version. On the finer mesh, the pressure values are much closer. After that time the functional values decrease because the domain $\Omega_{f,\mathrm{mid}}(t)$ gets smaller.

In the next paragraph, we will study the performance of both contact formulations under mesh refinement.

6.3.2. Convergence under mesh refinement

We solve the same problem on three different meshes with 1.280, 5.120 and 20.480 mesh elements, where the finer meshes are obtained from the coarsest one by global mesh refinement. The plots of the functionals $d_{\min}$ and J_{P_γ} as well as the functionals

$$J_{\mathrm{contact}} := \frac{\left\| \gamma_C^{1/2} [P_\gamma(\lambda, d)]_+ + \gamma_C^{-1/2} \lambda \right\|_{\Gamma(t)}}{\overline{\|\lambda\|}_\Gamma}, \quad J_{\mathrm{vel,fsi}} := \frac{\|(\dot{d} - u) \cdot n\|_{\Gamma_{\mathrm{fsi}}(t)}}{\overline{\|u \cdot n\|}_{\Gamma_{\mathrm{fsi}}} + \overline{\|\dot{d} \cdot n\|}_{\Gamma_{\mathrm{fsi}}}}$$

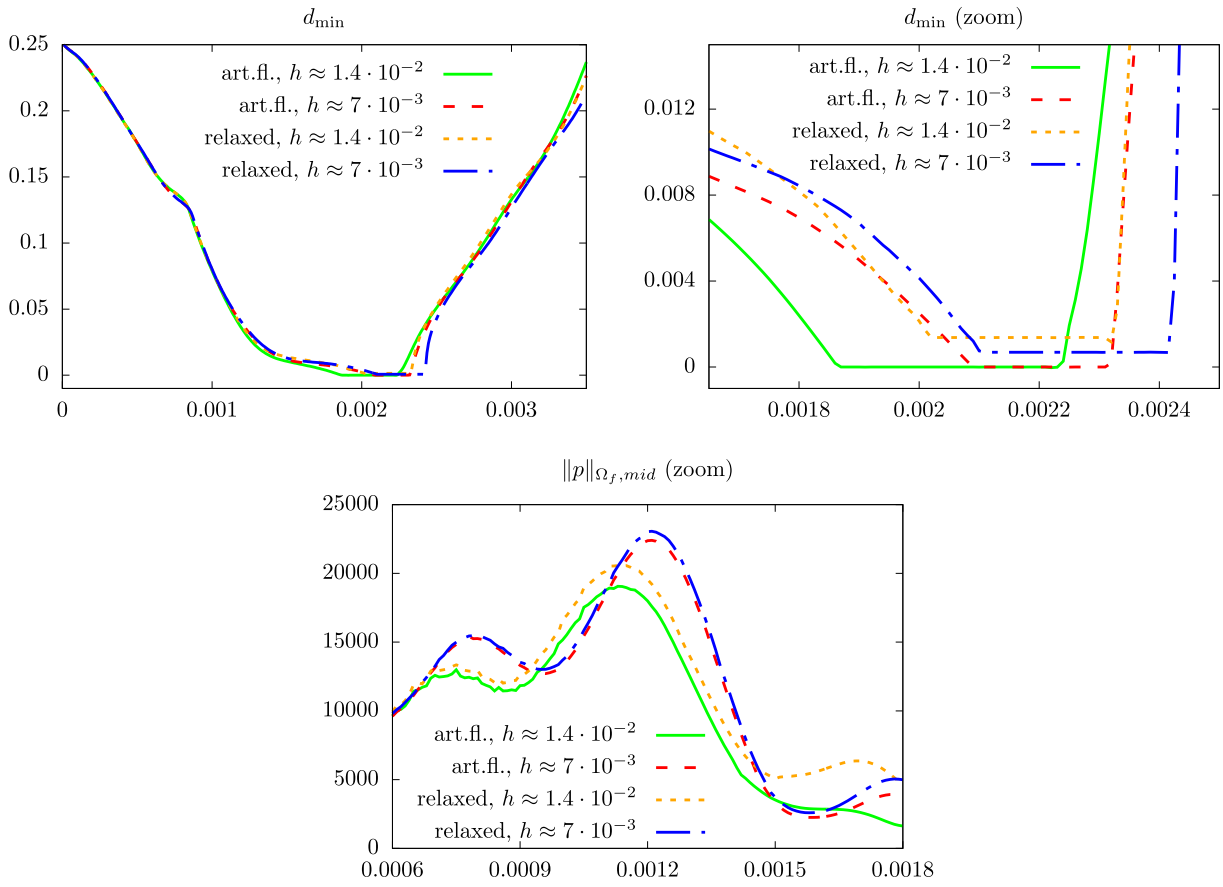


FIGURE 8. Comparison of the relaxed and the artificial fluid contact formulation. *Top*: minimal distance $d_{\min}$ to Γ_w. *Right:* zoom-in around the contact interval. *Bottom:* pressure norm $\|p\|_{L^2(\Omega_f,\text{mid})}$ over the central part of the fluid domain before contact over time.

measuring the fulfilment of the contact condition and the continuity of velocities on the part $\Gamma_{\text{fsi}}(t)$ of $\Gamma(t)$ that is not in contact with Γ_w

$$\Gamma_{\text{fsi}}(t) := \left\{ x \in \Gamma(t) \,\middle|\, P_\gamma(\lambda, d)(x) \leq 0 \right\}$$

are shown in Figure 9 for the artificial fluid formulation and in Figure 10 for the relaxed contact formulation over time. The quantities $\overline{\|\cdot\|}_*$ that are used to scale the functionals are temporal averages of the respective norms over the interval $I = [0, 0.004]$, computed on the finest grid.

First, we observe for both formulations in the plots on the top left that the contact happens later, the finer the discretisation is, as the fluid forces which act against the closure of the fluid channel are better resolved on the finer meshes (see also Fig. 8 and the related discussion above).

The curves for the contact force P_γ on the top right of both figures show significant differences between the two formulations. While the functional values seem to converge for the artificial fluid formulation (if we neglect the time shift), the contact force gets larger under mesh refinement for the relaxed formulation. The larger values for the relaxed formulation are due to the presence of the fluid forces $\sigma_{f,n}$ during the whole contact interval, that are not penalised in this formulation. As a Lagrange multiplier for the incompressibility constraint, the continuous pressure p gets singular when it comes to contact. The discrete pressure p_h gets larger and larger under mesh refinement in our computations.

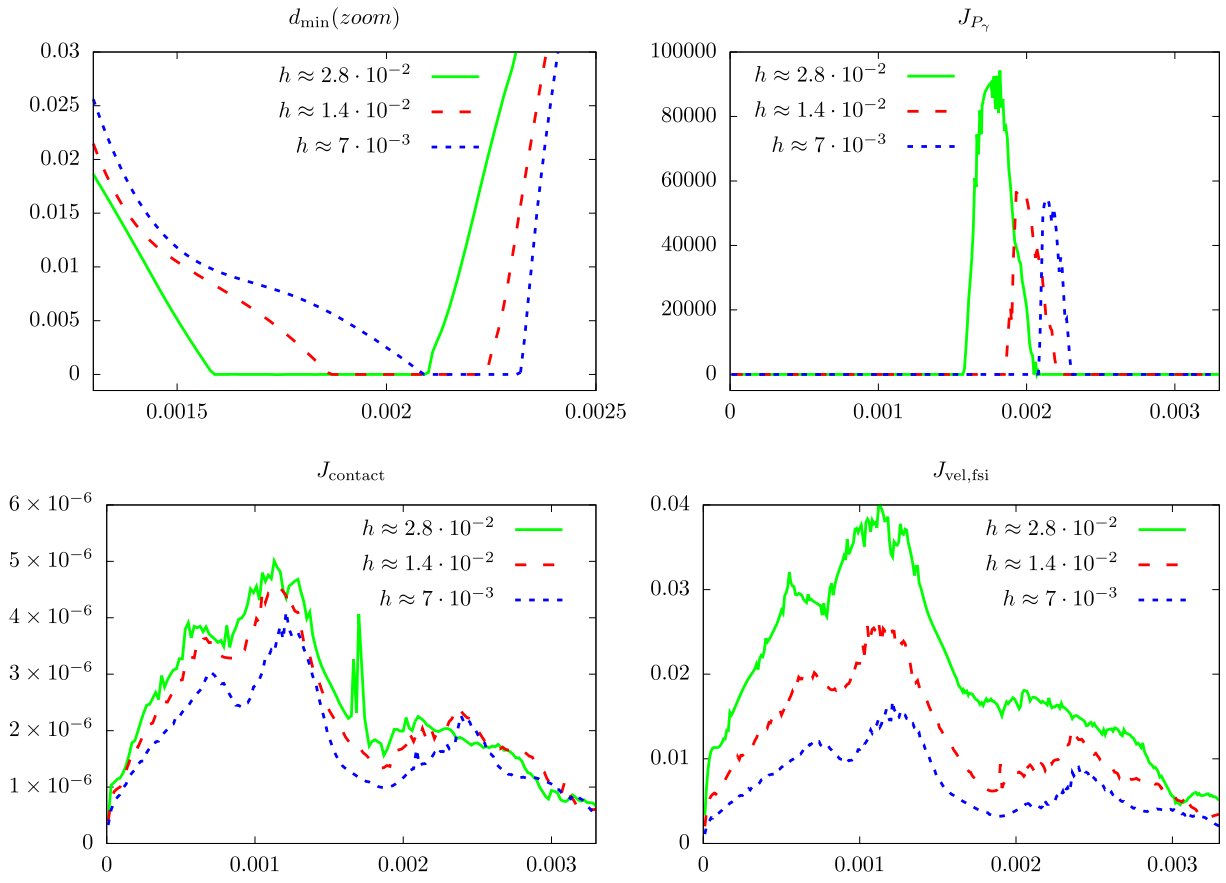


FIGURE 9. Convergence studies under mesh refinement for the *artificial fluid formulation* by means of the following functionals over time: *Top left:* minimal distance $d_{\min}$ of $\Gamma(t)$ to Γ_w, *top right*: contact force J_{P_γ}, *Bottom left*: fulfilment of the contact condition J_{contact}. *Bottom right*: continuity of velocities $J_{\text{vel,fsi}}$.

On the other hand, the functional J_{contact}, that measures the difference between $-\gamma_C^{1/2}[P_\gamma]_+$ and $\gamma_C^{-1/2}\lambda$ decreases under mesh refinement for both formulations. Besides the differences in the contact force J_{P_γ}, the functional values on each of the mesh levels are actually very similar. The reason must be that the fluid forces $\sigma_{f,n}$ enter in both λ and $[P_\gamma]_+$. We conclude that the increase in the functional J_{P_γ} seems not to be an issue for the contact dynamics.

Both J_{contact} and the functional $J_{\text{vel,fsi}}$ are controlled by the stability estimate in Theorem 5.1 for $\theta = 1$. Although the parameter $\theta = 0$ is used here, we observe that both functionals decrease with mesh refinement before and during contact. While the convergence for the contact functional is quite slow, the values of the velocity functional indicate a convergence order $\mathcal{O}(h^\alpha)$ with $0.5 \leq \alpha \leq 1$ for both formulations. Note that in contrast to the term J_{contact}, $J_{\text{vel,fsi}}$ is controlled in the stability estimate in Theorem 5.1 even with the pre-factor $(\gamma_{\text{fsi}}^0 \mu_f)^{1/2} h^{-1/2}$.

6.3.3. Flux formulations

Next, we compare the different choices for λ. We show results exemplarily for the artificial fluid formulation with slip interface conditions. We will compare results using the jump of stresses $\lambda = [\![\sigma_{n,\text{slip}}]\!]$ (4.7), the jump

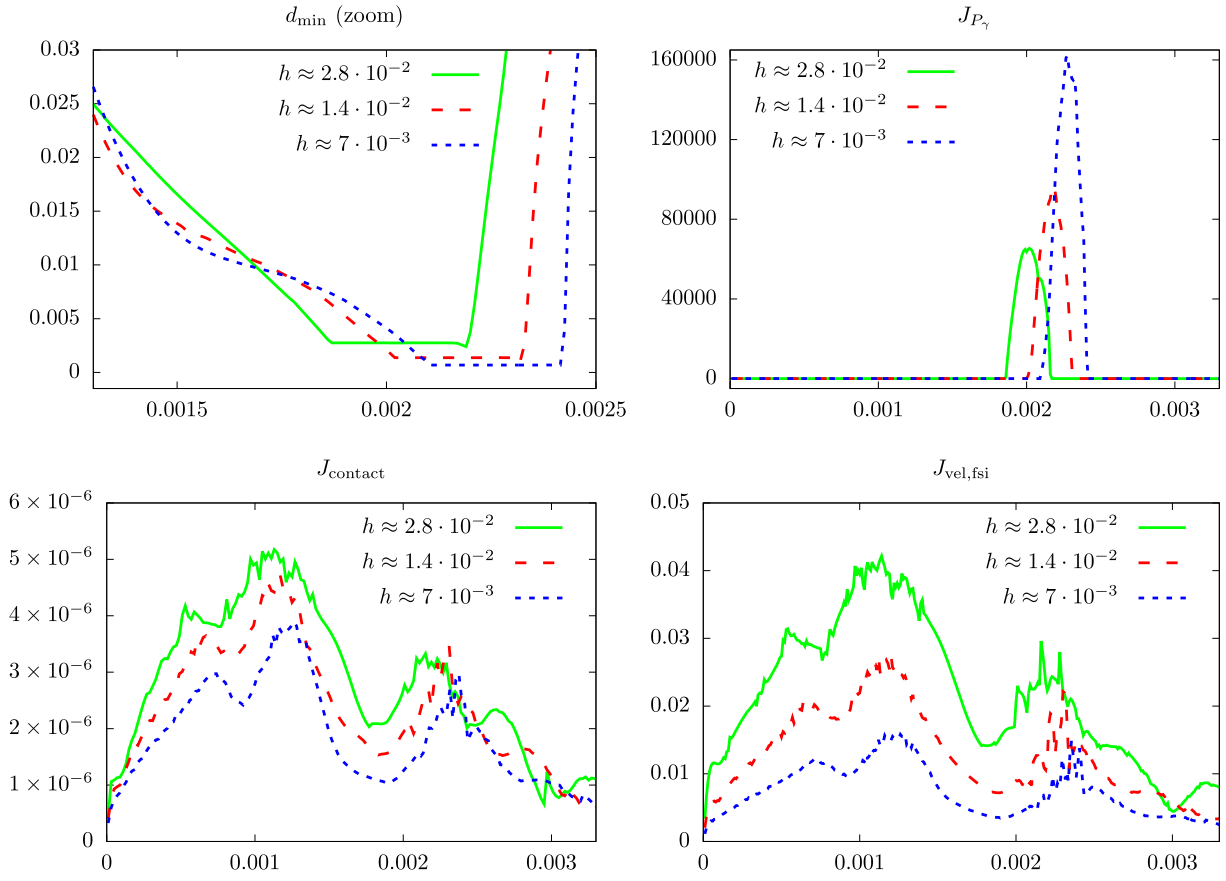


FIGURE 10. Convergence studies under mesh refinement for the *relaxed contact formulation* by means of the following functionals over time: *Top left:* minimal distance $d_{\min}$ of $\Gamma(t)$ to Γ_w, *top right*: contact force J_{P_γ}, *Bottom left*: fulfilment of the contact condition J_{contact}. *Bottom right*: continuity of velocities $J_{\text{vel,fsi}}$.

of the numerical fluxes $\lambda = [\![\widetilde{\sigma}_{n,\text{slip}}]\!]$ (4.8) and the extended fluxes $\lambda = [\![\widetilde{\sigma}_{n,\text{slip}}]\!] + \tau^T \sigma_s n(\tau \cdot n_w)$ (see (4.10)). As the results for the latter two choices are nearly identical in this example, we show here only plots for the jump of stresses and the (non-extended) numerical fluxes. We use the artificial fluid formulation (Variational Formulation 2) and the previously used mesh with 5120 elements.

In Figure 11, we show the minimal distance $d_{\min}$ to Γ_w, the contact force J_{P_γ} and the integral over the velocity difference across the contact part $\Gamma_C(t)$ of the interface over time

$$J_{\text{vel},C} := \int_{\Gamma_C(t)} (\dot{d} - u) \cdot n \, \mathrm{d}s, \qquad \Gamma_C(t) := \left\{ x \in \Gamma(t) \,\middle|\, P_\gamma(\lambda, d)(x) \leq 0 \right\}.$$

The fluid velocity u is here artificial as it comes from Ω_f^C. When choosing $\lambda = [\![\widetilde{\sigma}_{n,\text{slip}}]\!]$, we ensure that there is no feedback from this artificial velocity to the solid, see (4.5). For the jump of stresses $\lambda = [\![\sigma_{n,\text{slip}}]\!]$, we obtain a mixture of the solid contact condition and the continuity of normal velocities and a feedback might result. This follows analogously to the no-slip case, see (3.14).

In the left sketch of Figure 11, we see that the minimal distance in the stress-based formulation shows oscillations during the whole contact interval, especially in the second half. The interface jumps back and forth

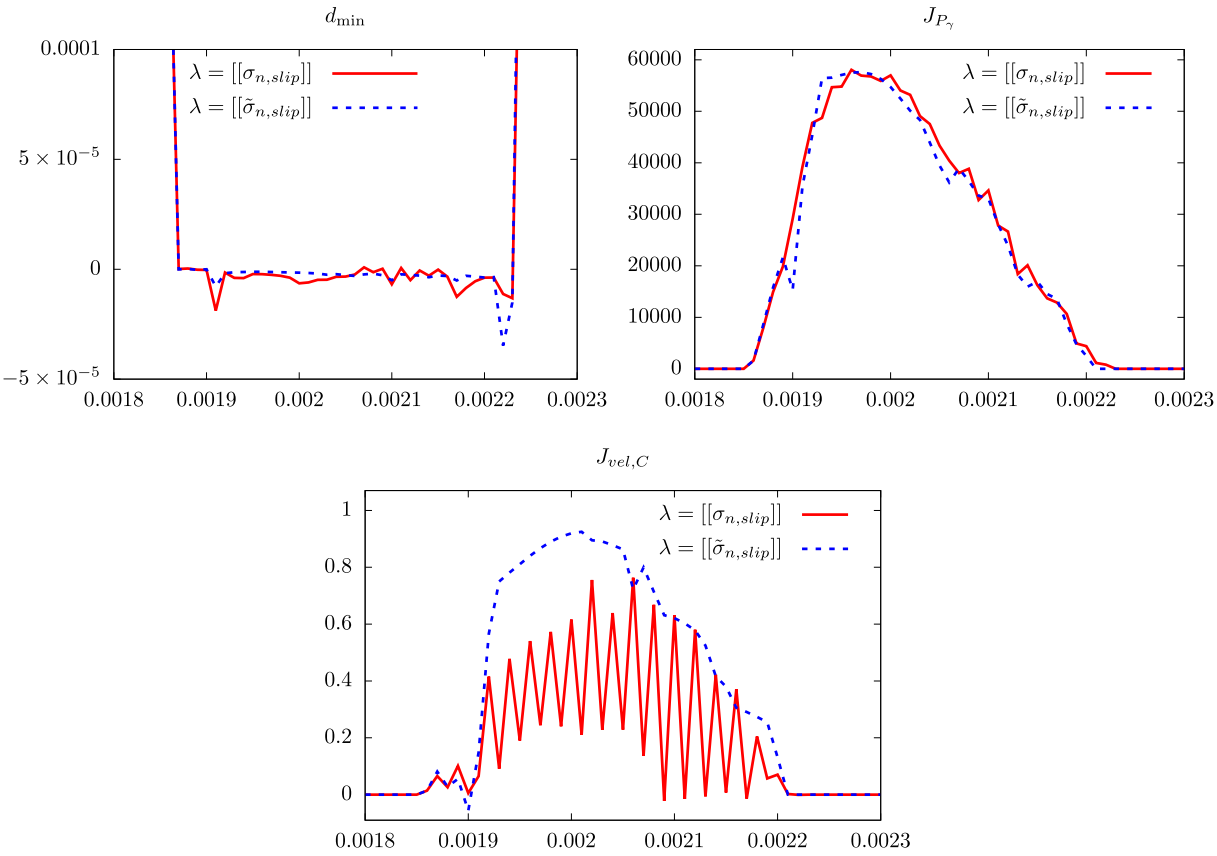


FIGURE 11. Comparison of the different possibilities to choose the fluxes λ. Minimal distance $d_{\min}$ to Γ_w (*top left*), contact force J_{P_γ} (*top right*) and velocity difference $J_{\text{vel},C}$ (*bottom*) integrated over the contact part $\Gamma_C(t)$ of $\Gamma(t)$ over time for computations with $\lambda = [\![\sigma_{n,\text{slip}}]\!]$ and $\lambda = [\![\widetilde{\sigma}_{n,\text{slip}}]\!]$ for the artificial fluid formulation.

over the contact line many times. The curve corresponding to the formulation using discrete fluxes is much smoother. Similarly, the contact force J_{P_γ} looks smoother, when the flux formulation is used. The reason for this behaviour is the mixture of the interface conditions during contact. On the bottom of Figure 11, we see that the velocity difference $J_{\text{vel},C}$ shows wild oscillations for the stress formulation, while it looks much smoother when using $[\![\widetilde{\sigma}_{n,\text{slip}}]\!]$. As the artificial velocity in Ω_f^C has no physical meaning, it is not a drawback that the absolute values of $J_{\text{vel},C}$ are larger. Due to the feedback of this velocity to the contact conditions, the oscillations appear in the displacement as well.

On the other hand, we should mention that the oscillations are relatively small. Especially those in $d_{\min}$ are almost by a factor 10^3 smaller than the mesh size $h \approx 1.4 \times 10^{-2}$ in vertical direction in this example and are therefore still acceptable.

6.3.4. Influence of the contact parameter γ_C^0

Next, we study the effect of different contact parameters γ_C^0 for the artificial fluid formulation and $\lambda = [\![\widetilde{\sigma}_{n,\text{slip}}]\!]$ on the mesh with 5120 elements. In Figure 12, we show the "minimal distance" $d_{\min}$ (top) and the contact force J_{P_γ} over time for different contact parameters γ_C^0. The results are similar to the corresponding results for the virtual obstacle problem in Figure 5. For the smallest contact parameter $\gamma_C^0 = 10$, the contact condition is

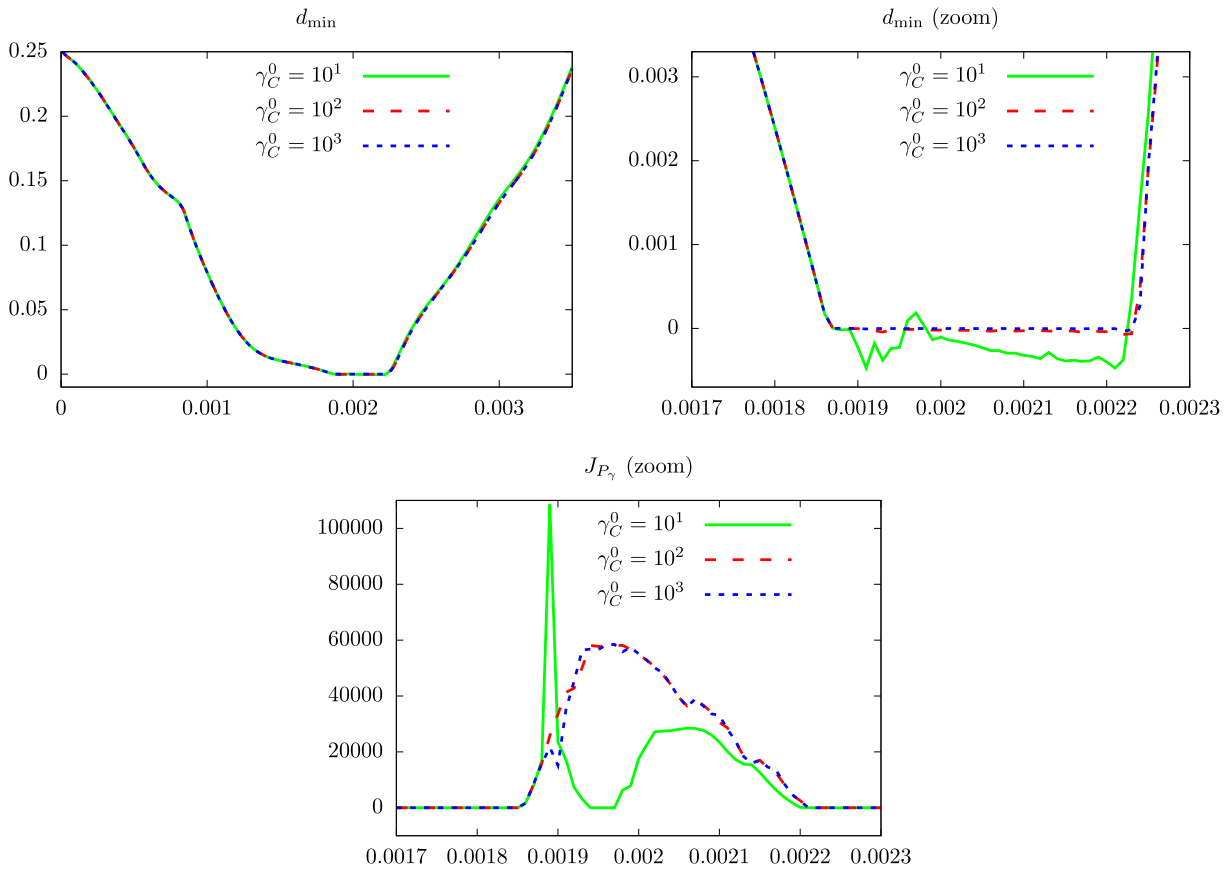


FIGURE 12. Parameter studies for the contact parameter γ_C^0: Minimal distance of $\Gamma(t)$ to the wall Γ_w (*Top left*: total time interval, *top right*: zoom around the contact interval) and contact force J_{P_γ} (*bottom*) over time.

violated throughout the contact interval ($d_{\min} < 0$). The maximum overlap into the artificial fluid domain is again approximately by a factor 30 smaller than the mesh size $h \approx 1.4 \times 10^{-2}$. This violation gets smaller, the larger the contact parameter is chosen. The instabilities for the smallest parameter are still much better visible in the contact force J_{P_γ}. At time $t = 1.89 \times 10^{-3}$ the functional shows a huge peak, as the contact condition $d \cdot n_w \leq g_0$ is severely violated and it vanishes from $t = 1.94 \times 10^{-3}$ to $t = 1.97 \times 10^{-3}$, when the contact is in fact shortly released.

For the larger values $\gamma_C^0 \geq 10^2$, the curves are relatively smooth and very similar. Altogether, this shows again that the assumption "γ_C^0 sufficiently large" in Theorem 5.1 is necessary in order to ensure stability.

6.3.5. Slip vs no-slip conditions

Next, we compare the effect of slip- and no-slip boundary and interface conditions in Figure 13. Due to the difficulties associated with the artificial fluid formulation and no-slip interface and boundary conditions (see the discussion at the end of Sect. 4.2), we use the relaxed contact formulation on the mesh with 5120 elements in this paragraph.

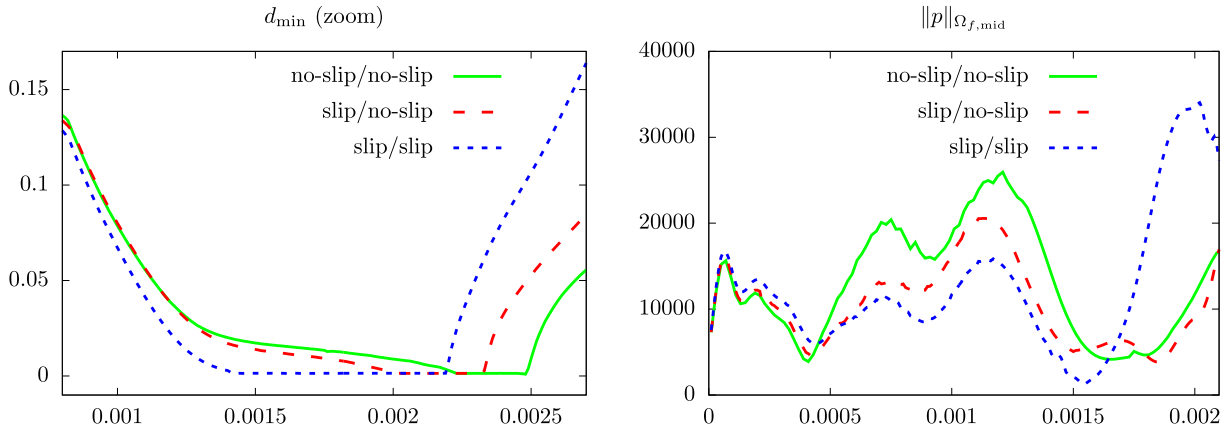


FIGURE 13. Comparison of slip- and no-slip interface/boundary conditions by means of the minimal distance $d_{\min}$ of the interface $\Gamma(t)$ to Γ_w around the contact interval (*left*) and the L^2-norm of the pressure over a region $\Omega_{f,\mathrm{mid}}(t)$ around the contact surface before contact (*right*) over time. Due to the larger pressure before contact, the impact happens later when using no-slip conditions.

We show results for

- Slip conditions on the interface $\Gamma(t)$ and the lower wall Γ_w
- A slip condition on $\Gamma(t)$ and a no-slip condition on Γ_w
- No-slip conditions on $\Gamma(t)$ and Γ_w.

Note that the second option is possible, as for the relaxed contact formulation $\Gamma_w \cap \Gamma(t) = \emptyset$.

We observe that the contact condition (or more precisely the relaxed condition $d \cdot n_w \leq g_\epsilon$) is earlier active, when using slip-conditions: at $t_C = 1.42 \times 10^{-3}$ for slip/slip conditions compared to $t_C = 2.02 \times 10^{-3}$ for slip interface and no-slip boundary conditions and at $t_C = 2.23 \times 10^{-3}$ for no-slip conditions on interface and boundary. The reason is that the fluid forces, and in particular the pressure, that act against the contact are larger for no-slip conditions, as the fluid can not "slip" out of the contact zone easily. This can be seen in the pressure plot on the right. The pressure is considerably larger from $t \approx 5 \times 10^{-4}$ for the no-slip conditions until contact is reached for the slip/slip case at $t_C = 1.42 \times 10^{-3}$.

As we are allowing for a small gap between the solid and the ground, these results do not contradict the theoretical results by Gerard-Varet *et al.* [36] discussed in Section 4.3, who showed that (in their configuration with a rigid body) contact can not happen, when no-slip conditions are used on the interface and/or the boundary. As discussed in Section 4.1, the basic assumption of the relaxed formulation is that a small or infinitesimal fluid layer remains during contact. On the contrary, the results confirm that contact is more likely to happen for slip-conditions, which is in agreement with the theoretical results.

6.3.6. Comparison with an explicit ad hoc approach

The probably simplest possibility to combine the FSI model introduced in Section 2.1 and the contact approach described in Section 2.2 is to split Γ explicitly in each time-step into a fluid-structure interface $\Gamma_{\mathrm{fsi}}(t_{m-1})$ and a contact surface $\Gamma_C(t_{m-1})$ based on the displacement $d(t_{m-1})$ of the previous time-step and to use the interface condition (3.1) on $\Gamma_{\mathrm{fsi}}(t_{m-1})$ and the contact condition (2.10) on $\Gamma_C(t_{m-1})$. A strategy of this type has been used by Hecht and Pironneau [41]. The system of equations reads in the slip case: *Find*

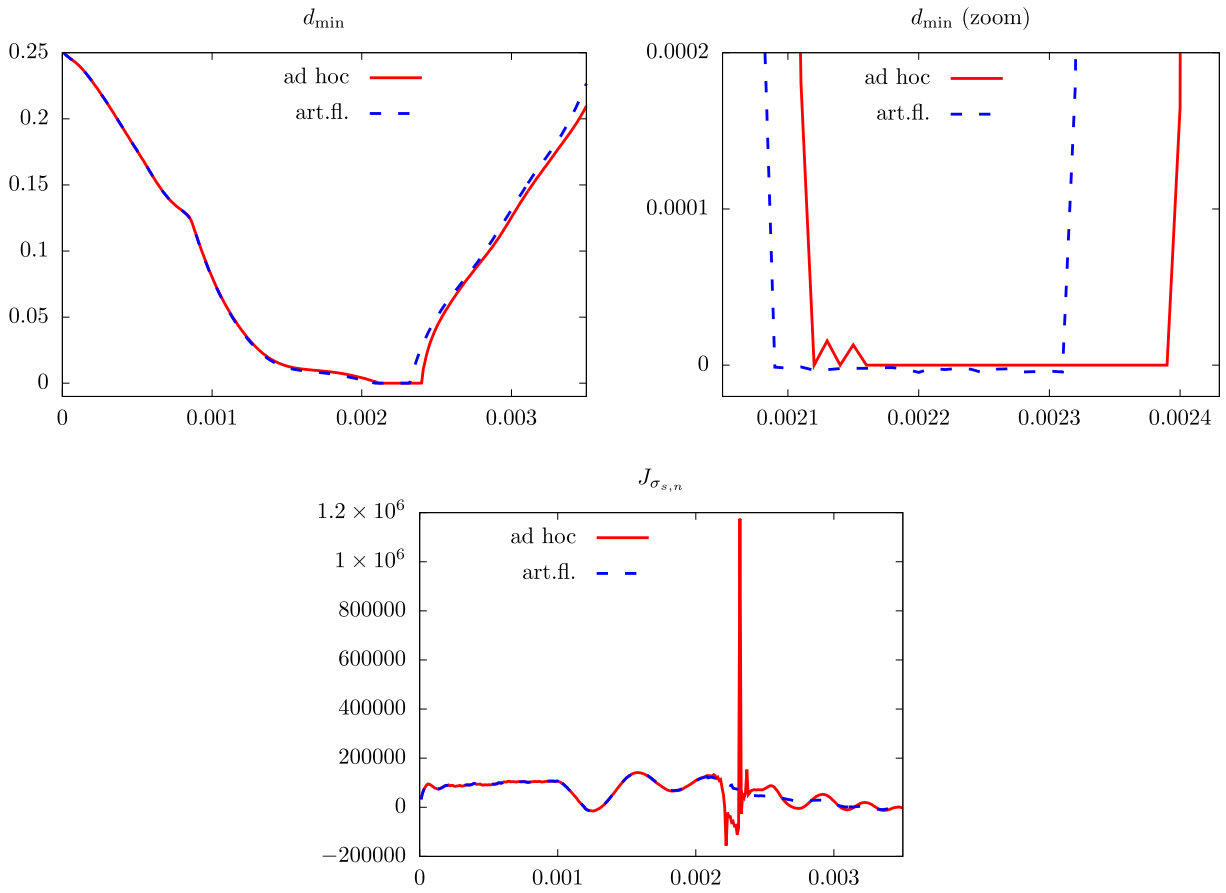


FIGURE 14. Comparison of an explicit *ad hoc* approach to include the contact and interface conditions with the approach using an artificial fluid described in Section 4.2. We show the minimal distance $d_{\min}$ of the interface $\Gamma(t)$ to Γ_w on the *top left* and a zoom-in on the *top right*. Chattering, *i.e.* an alteration between contact and no contact is visible for the *ad hoc* approach at the beginning of the contact interval. The integral over the normal solid stresses $J_{\sigma_{s,n}}$ shown in the graph on the bottom shows large instabilities for the *ad hoc* approach.

$u \in \mathcal{V}, p \in \mathcal{Q}, d \in \mathcal{W}$ *such that* $\dot{d} = \partial_t d$ *and*

$$
\begin{aligned}
&\left(\partial_t u, v\right)_{\Omega_f(t)} + \left(\sigma_f(u,p), \nabla v\right)_{\Omega_f(t)} + \left(\operatorname{div} u, q\right)_{\Omega_f(t)} + \left(\partial_t \dot{d}, w\right)_{\Omega_s(t)} + \left(\sigma_s(d), \nabla w\right)_{\Omega_s(t)} \\
&\quad - \left(n^T \sigma_f(u,p) n, (w-v)\cdot n\right)_{\Gamma_{\text{fsi}}(t_{m-1})} - \left((\dot{d}-u)\cdot n, n^T \sigma_f(v,-q) n\right)_{\Gamma_{\text{fsi}}(t_{m-1})} \\
&\quad + \gamma_{\text{fsi}} \left((\dot{d}-u)\cdot n, (w-v)\cdot n\right)_{\Gamma_{\text{fsi}}(t_{m-1})} + \gamma_C \left(P_\gamma(\sigma_{s,n}(d), d), w\cdot n_w\right)_{\Gamma_C(t_{m-1})} \\
&\quad = \left(f_f, v\right)_{\Omega_f(t)} + \left(f_s, w\right)_{\Omega_s(t)} \quad \forall v, q, w \in \mathcal{V} \times \mathcal{Q} \times \mathcal{W}.
\end{aligned}
\tag{6.2}
$$

We use the same numerical parameters as for the contact formulations presented in this work.

To compare this approach with the artificial fluid formulation we show the minimal distance to the ground $d_{\min}$ and the integral over the normal solid stresses over $\Gamma(t) = \Gamma_{\text{fsi}}(t) \cup \Gamma_C(t)$

$$J_{\sigma_{s,n}} = \int_{\Gamma(t)} \sigma_{s,n} \, \mathrm{d}s$$

on the finer mesh with 20 480 elements in Figure 14. While the curves for $d_{\min}$ over the total time interval shown on the top left look similar, a zoom-in on the right shows again that the presence of the artificial fluid leads to an earlier time of impact. Moreover, we observe chattering for the ad-hoc approach at the beginning of the contact interval, *i.e.* contact is released twice again before the solid stays in contact with Γ_w. The interface jumps back to the fluid domain, with a (relatively small) minimal distance of approximately 1.6×10^{-5}.

In fact, the functional $d_{\min}$ is not a good indicator to investigate stability for the *ad hoc* approach, as it is zero, as soon as one point of the interface lies on Γ_w. Note that this is different for the approaches presented in this work, where the interface can go beyond Γ_w (or Γ_ϵ for the relaxed approach). In the actual computation, the interface oscillates considerably in each time-step and contact is released and renewed frequently in different points. The functional $J_{\sigma_{s,n}}$ on the bottom left of Figure 14 serves to get a better impression of the instabilities during contact. It oscillates throughout the contact interval including a huge peak at $t = 2.32 \times 10^{-3}$. Moreover, we see that the elastic dynamics after the contact are also significantly influenced by these instabilities. Compared to the artificial fluid approach the oscillations in the displacement are significantly larger after contact.

We have also tried to iterate for the splitting into $\Gamma_C(t)$ and $\Gamma_{\text{fsi}}(t)$ within each time-step of the ad-hoc approach, which can be seen as an active-set strategy. This did however not cure the problem, as cycling between different active sets is not prevented.

7. Conclusions

We have presented two consistent formulations for fluid-structure interactions with contact, both including a continuous switch between the FSI interface and the contact condition depending on the contact force P_γ. In contrast to certain penalty approaches, the contact force is physically motivated and included in a consistent way in the variational formulations. Our numerical results indicate that the two proposed formulations have better stability properties than the usual *ad hoc* approaches and no chattering was observed in our computations.

Moreover, we have derived analytically a stability result for a generalised formulation including a parameter $\theta \in [-1, 1]$. As in the pure solid case [20], this result implies stability for $\theta = 1$ and stability up to a term including the contact force for $\theta \neq 1$. In our computations, we have however not observed any stability issues for the choice $\theta = 0$ either.

The contact formulations were derived here for the simplified configuration of contact with a fixed and straight wall and using linear models for the fluid and solid sub-problems. The algorithms can be applied to more complex contact configurations by using approaches from the literature to compute the projection and the distances between different surfaces [59,65,66]. In particular, the extension to the incompressible Navier–Stokes equations in the fluid and to non-linear elasticity in the solid can be addressed by combining the proposed approach with the arguments recently reported in Mlika *et al.* [55]. Moreover, Coulumb or Tresca friction can also be incorporated by following Chouly *et al.* [19,24].

In our numerical examples, we have studied a two-dimensional model problem on a relatively simple and smooth geometry. In future, we plan to apply the methods on more complex geometries, including a direct comparison with experiments, in order to further validate the numerical approach.

Acknowledgements. The first author acknowledges support by the EPSRC grant EP/P01576X/1. The third author was supported by the DFG Research Scholarship FR3935/1-1.

References

[1] P. Alart and A. Curnier, A mixed formulation for frictional contact problems prone to Newton like solution methods. *Comput. Methods Appl. Mech. Eng.* **92** (1991) 353–375.

[2] F. Alauzet, B. Fabréges, M.A. Fernández and M. Landajuela, Nitsche-XFEM for the coupling of an incompressible fluid with immersed thin-walled structures. *Comput. Methods Appl. Mech. Eng.* **301** (2016) 300–335.

[3] P. Angot, Analysis of singular perturbations on the brinkman problem for fictitious domain models of viscous flows. *Math. Methods Appl. Sci.* **22** (1999) 1395–1412.

[4] M. Astorino, J.F. Gerbeau, O. Pantz and K.F. Traoré, Fluid-structure interaction and multi-body contact: application to aortic valves. *Comput. Methods Appl. Mech. Eng.* **198** (2009) 3603–3612.

[5] R. Becker and M. Braack, A finite element pressure gradient stabilization for the Stokes equations based on local projections. *Calcolo* **38** (2001) 173–199.

[6] R. Becker, M. Braack, D. Meidner, T. Richter and B. Vexler, The finite element toolkit Gascoigne3d. http://www.gascoigne.uni-hd.de.

[7] M. Besier and W. Wollner, On the pressure approximation in nonstationary incompressible flow simulations on dynamically varying spatial meshes. *Int. J. Numer. Methods Fluids* **69** (2012) 1045–1064.

[8] D. Boffi and L. Gastaldi, A finite element approach for the immersed boundary method. *Comput. Struct.* **81** (2003) 491–501.

[9] L. Boilevin-Kayl, M.A. Fernández and J.F. Gerbeau, Numerical methods for immersed fsi with thin-walled structures. *Comput. Fluids* **179** (2019) 744–763.

[10] F. Brezzi and J. Pitkäranta, On the stabilization of finite element approximations of the stokes equations, edited by W. Hackbusch. In: *Efficient Solutions of Elliptic Systems*. Springer (1984) 11–19.

[11] V. Bruyere, N. Fillot, G.E. Morales-Espejel and P. Vergne, Computational fluid dynamics and full elasticity model for sliding line thermal elastohydrodynamic contacts. *Tribol. Int.* **46** (2012) 3–13.

[12] E. Burman, Ghost penalty. *C.R. Math.* **348** (2010) 1217–1220.

[13] E. Burman and M.A. Fernández, An unfitted Nitsche method for incompressible fluid–structure interaction using overlapping meshes. *Comput. Methods Appl. Mech. Eng.* **279** (2014) 497–514.

[14] E. Burman and P. Hansbo, Edge stabilization for the generalized Stokes problem: a continuous interior penalty method. *Comput. Methods Appl. Mech. Eng.* **195** (2006) 2393–2410.

[15] E. Burman and P. Hansbo, Fictitious domain finite element methods using cut elements: II. A stabilized Nitsche method. *Appl. Numer. Math.* **62** (2012) 328–341.

[16] E. Burman and P. Hansbo, Deriving robust unfitted finite element methods from augmented lagrangian formulations, edited by S.P.A. Bordas, E. Burman, M.G. Larson and M.A. Olshanskii. In: *Geometrically Unfitted Finite Element Methods and Applications – Proceedings of the UCL-workshop 2016*. Springer (2017) 1–24.

[17] E. Burman, P. Hansbo, M.G. Larson and R. Stenberg, Galerkin least squares finite element method for the obstacle problem. *Comput. Methods Appl. Mech. Eng.* **313** (2017) 362–374.

[18] E. Burman, P. Hansbo and M.G. Larson, Augmented lagrangian and galerkin least-squares methods for membrane contact. *Int. J. Numer. Methods Eng.* **114** (2018) 1179–1191.

[19] F. Chouly, An adaptation of Nitsche's method to the tresca friction problem. *J. Math. Anal. Appl.* **411** (2014) 329–339.

[20] F. Chouly and P. Hild, A Nitsche-based method for unilateral contact problems: numerical analysis. *SIAM J. Numer. Anal.* **51** (2013) 1295–1307.

[21] F. Chouly, P. Hild and Y. Renard, A Nitsche finite element method for dynamic contact: 1. space semi-discretization and time-marching schemes. *ESAIM: M2AN* **49** (2015) 481–502.

[22] F. Chouly, P. Hild and Y. Renard, Symmetric and non-symmetric variants of Nitsche's method for contact problems in elasticity: Theory and numerical experiments. *Math. Comput.* **84** (2015) 1089–1112.

[23] F. Chouly, R. Mlika and Y. Renard, An unbiased Nitsche's approximation of the frictional contact between two elastic structures. *Numer. Math.* **139** (2018) 593–631.

[24] F. Chouly, P. Hild, V. Lleras and Y. Renard, Nitsche-based finite element method for contact with coulomb friction, edited by F.A. Radu, K. Kumar, I. Berre, J.M. Nordbotten and I.S. Pop. In: *Numerical Mathematics and Advanced Applications ENUMATH 2017*. Springer International Publishing (2019) 839–847.

[25] F. Cimolin and M. Discacciati, Navier–Stokes/Forchheimer models for filtration through porous media. *Appl. Numer. Math.* **72** (2013) 205–224.

[26] G.H. Cottet, E. Maitre and T. Milcent, Eulerian formulation and level set models for incompressible fluid-structure interaction. *ESAIM: M2AN* **42** (2008) 471–492.

[27] N.D. dos Santos, J.F. Gerbeau and J.F. Bourgat, A partitioned fluid–structure algorithm for elastic thin valves with contact. *Comput. Methods Appl. Mech. Eng.* **197** (2008) 1750–1761.

[28] T. Dunne, *Adaptive finite element approximation of fluid-structure interaction based on Eulerian and Arbitrary Lagrangian-Eulerian variational formulations*. Ph.D. thesis, Heidelberg University (2007).

[29] T. Dunne and R. Rannacher, Adaptive finite element approximation of fluid-structure interaction based on an Eulerian variational formulation, edited by H.J. Bungartz and M. Schäfer. In: Fluid-Structure Interaction: Modeling, Simulation, Optimization. *Lect. Notes Comput. Sci. Eng.* Springer (2006) 110–145.

[30] S. Frei, *Eulerian finite element methods for interface problems and fluid-structure interactions*. Ph.D. thesis, Heidelberg University (2016). http://www.ub.uni-heidelberg.de/archiv/21590.

[31] S. Frei, An edge-based pressure stabilization technique for finite elements on arbitrarily anisotropic meshes. *Int. J. Numer. Methods Fluids* **89** (2019) 407–429.

[32] S. Frei and T. Richter, A locally modified parametric finite element method for interface problems. *SIAM J. Numer. Anal.* **52** (2014) 2315–2334.

[33] S. Frei and T. Richter, An accurate Eulerian approach for fluid-structure interactions, edited by S. Frei, B. Holm, T. Richter, T. Wick and H. Yang. In: Fluid-Structure Interaction: Modeling, Adaptive Discretization and Solvers. *Rad. Ser. Comput. Appl. Math.* Walter de Gruyter, Berlin (2017).

[34] S. Frei and T. Richter, A second order time-stepping scheme for parabolic interface problems with moving interfaces. *ESAIM: M2AN* **51** (2017) 1539–1560.

[35] D. Gérard-Varet and M. Hillairet, Regularity issues in the problem of fluid structure interaction. *Arch. Ration Mech. Anal.* **195** (2010) 375–407.

[36] D. Gerard-Varet, M. Hillairet and C. Wang, The influence of boundary conditions on the contact problem in a 3D Navier–Stokes flow. *J. Math. Pure Appl.* **103** (2015) 1–38.

[37] A. Gerstenberger and W.A. Wall, An extended finite element method/Lagrange multiplier based approach for fluid–structure interaction. *Comput. Methods Appl. Mech. Eng.* **197** (2008) 1699–1714.

[38] C. Grandmont and M. Hillairet, Existence of global strong solutions to a beam–fluid interaction system. *Arch. Ration Mech. Anal.* **220** (2016) 1283–1333.

[39] C. Grandmont, M. Lukáčová-Medvidóvá and Š. Nečasová, Mathematical and numerical analysis of some FSI problems, edited by T. Bodnár, G.P. Galdi and Š. Nečasová. In: Fluid-Structure Interaction and Biomedical Applications. Springer (2014) 1–77.

[40] P. Hansbo, J. Hermansson and T. Svedberg, Nitsche's method combined with space–time finite elements for ALE fluid–structure interaction problems. *Comput. Methods Appl. Mech. Eng.* **193** (2004) 4195–4206.

[41] F. Hecht and O. Pironneau, An energy stable monolithic eulerian fluid-structure finite element method. *Int. J. Numer. Methods Fluids* **85** (2017) 430–446.

[42] T.I. Hesla, *Collisions of smooth bodies in viscous fluids: A mathematical investigation.* Ph.D. thesis, Univ. of Minnesota (2004).

[43] M. Hillairet, Lack of collision between solid bodies in a 2D incompressible viscous flow. *Commun. Part Diff. Equ.* **32** (2007) 1345–1371.

[44] M. Hillairet and T. Takahashi, Collisions in three-dimensional fluid structure interaction problems. *SIAM J. Math. Anal.* **40** (2009) 2451–2477.

[45] M. Hintermüller, K. Ito and K. Kunisch, The primal-dual active set strategy as a semismooth newton method. *SIAM J. Opt.* **13** (2002) 865–888.

[46] S. Hüeber and B.I. Wohlmuth, A primal–dual active set strategy for non-linear multibody contact problems. *Comput. Methods Appl. Mech. Eng.* **194** (2005) 3147–3166.

[47] T.J.R. Hughes, L.P. Franca and M. Balestra, A new finite element formulation for computational fluid dynamics: V. circumventing the Babuška-Brezzi condition: a stable Petrov–Galerkin formulation of the Stokes problem accommodating equal-order interpolations. *Comput. Methods Appl. Mech. Eng.* **59** (1986) 85–99.

[48] O. Iliev and V. Laptev, On numerical simulation of flow through oil filters. *Comput. Visu. Sci.* **6** (2004) 139–146.

[49] D. Kamensky, M.C. Hsu, D. Schillinger, J.A. Evans, A. Aggarwal, Y. Bazilevs, M.S. Sacks and T.J.R. Hughes, An immersogeometric variational framework for fluid–structure interaction: Application to bioprosthetic heart valves. *Comput. Methods Appl. Mech. Eng.* **284** (2015) 1005–1053.

[50] S. Knauf, S. Frei, T. Richter and R. Rannacher, Towards a complete numerical description of lubricant film dynamics in ball bearings. *Comput. Mech.* **53** (2014) 239–255.

[51] A. Legay, J. Chessa and T. Belytschko, An Eulerian–Lagrangian method for fluid-structure interaction based on level sets. *Comput. Methods Appl. Mech. Eng.* **195** (2006) 2070–2087.

[52] S. Mandal, A. Ouazzi and S. Turek, Modified newton solver for yield stress fluids, edited by B. Karasözen, M. Manguoğlu, M. Tezer-Sezgin, S. Göktepe and Ö Uğur. In: *Numerical Mathematics and Advanced Applications ENUMATH 2015.* Springer International Publishing (2016) 481–490.

[53] A. Massing, M. Larson, A. Logg and M. Rognes, A nitsche-based cut finite element method for a fluid-structure interaction problem. *Comm. Appl. Math. Comput. Sci.* **10** (2015) 97–120.

[54] U.M. Mayer, A. Popp, A. Gerstenberger and W.A. Wall, 3D fluid–structure-contact interaction based on a combined XFEM FSI and dual mortar contact approach. *Comput. Mech.* **46** (2010) 53–67.

[55] R. Mlika, Y. Renard and F. Chouly, An unbiased Nitsche's formulation of large deformation frictional contact and self-contact. *Comput. Methods Appl. Mech. Eng.* **325** (2017) 265–288.

[56] B. Muha and S. Čanić, Existence of a weak solution to a fluid-elastic structure interaction problem with the Navier slip boundary condition. *J. Diff. Equ.* **260** (2016) 8550–8589.

[57] J.A. Nitsche, Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh. Math. Univ. Hamburg* **36** (1970) 9–15.

[58] C.S. Peskin, Flow patterns around heart valves: a numerical method. *J. Comput. Phys.* **10** (1972) 252–271.

[59] K. Poulios and Y. Renard, An unconstrained integral approximation of large sliding frictional contact between deformable solids. *Comput. Struct.* **153** (2015) 75–90.

[60] M.A. Puso, A 3D mortar method for solid mechanics. *Int. J. Numer. Methods Eng.* **59** (2004) 315–336.

[61] T. Richter, A fully Eulerian formulation for fluid-structure interactions. *J. Comput. Phys.* **233** (2013) 227–240.

[62] T. Richter, Finite elements for fluid-structure interactions. models, analysis and finite elements. In: Vol. 118 of *Lect Notes Comput. Sci. Eng.* Springer (2017).

[63] T.E. Tezduyar and S. Sathe, Modeling of fluid-structure interactions with the space-time finite elements: solution techniques. *Int. J. Numer. Methods Fluids* **54** (2007) 855–900.

[64] C. Wang, Strong solutions for the fluid–solid systems in a 2-D domain. *Asymptotic Anal.* **89** (2014) 263–306.

[65] B. Wohlmuth, Variationally consistent discretization schemes and numerical algorithms for contact problems. *Acta Numer.* **20** (2011) 569–734.

[66] B. Yang, T.A. Laursen and X. Meng, Two dimensional mortar contact methods for large deformation frictional sliding. *Int. J. Numer. Methods Eng.* **62** (2005) 1183–1225.

[67] L. Zhang, A. Gerstenberger, X. Wang and W.K. Liu, Immersed finite element method. *Comput. Methods Appl. Mech. Eng.* **193** (2004) 2051–2067.

Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 392 (2022) 114637
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

A mechanically consistent model for fluid–structure interactions with contact including seepage

Erik Burman[a], Miguel A. Fernández[b,c], Stefan Frei[d,*], Fannie M. Gerosa[b,c]

[a] *Department of Mathematics, University College London, London, UK–WC1E 6BT, United Kingdom*
[b] *Inria, 75012 Paris, France*
[c] *Sorbonne Université & CNRS, UMR 7598 LJLL, 75005 Paris, France*
[d] *Department of Mathematics & Statistics, University of Konstanz, Germany*

Received 25 July 2021; received in revised form 14 January 2022; accepted 14 January 2022
Available online 17 February 2022

Abstract

We present a new approach for the mechanically consistent modelling and simulation of fluid–structure interactions (FSI) with contact. The fundamental idea consists of combining a relaxation of the contact conditions with the modelling of seepage through a porous layer of co-dimension one during contact. For the latter, a Darcy model is considered in a thin porous layer attached to a solid boundary in the limit of infinitesimal thickness. The resulting computational model is both mechanically consistent and simple to implement. The FSI coupling and the contact conditions are imposed weakly by means of a unified Nitsche approach. We analyse the approach in detailed numerical studies with both thick- and thin-walled solids, within a fully Eulerian and an immersed approach and using fitted and unfitted finite element discretisations.

Keywords: Fluid–structure interaction; Contact mechanics; Nitsche's method; Seepage; Darcy equation; Mechanical consistency

1. Introduction

The design and analysis of computational methods for systems where solids are immersed in a fluid and that can come into contact is an outstanding problem. Already fluid–structure interaction (FSI) without contact is challenging due to the moving geometries and the stiff coupling between the solid and the fluid systems. If contact between solids is to be modelled as well, the complexity increases drastically. In particular the following additional issues need to be addressed:

- In combination with no-slip boundary conditions, FSI models are unable to predict contact (see, e.g., [1–3]), contrary to what is observed in experiments [4];
- The simple addition of a contact constraint (variational inequality) to an FSI model which allows for contact yields a mechanically inconsistent fluid–structure–contact interaction model (as discussed for example in [5]);
- Topological changes in the fluid domain need to be dealt with;

* Corresponding author.
E-mail address: stefan.frei@uni-konstanz.de (S. Frei).

https://doi.org/10.1016/j.cma.2022.114637
0045-7825/

- Non-linearly changing interface conditions: The interface condition changes from a fluid–solid interaction to a solid–solid contact problem which is described by variational inequalities;
- Important differences in the characteristic scales of the different physical phenomena: Contact represents a singular phenomenon in time and space, which needs to be resolved in space and time discretisation.

The first difficulty requires a modification of the boundary and interface conditions on the contacting walls (see, e.g., [6–8]) that reproduces the physically observed contact phenomena. However, as pointed out in the second issue, this does not necessarily lead to a mechanically consistent model, due to the possibility of unphysical void creation (when an elastic solid releases from contact) or to unbalanced stresses (in the case of relaxed contact modelling).

While nowadays several works on FSI–contact problems are available in literature, surprisingly few of them actually take care of the mechanical consistency of the contact treatment involving a fluid. In the vast majority of the cases, either an *ad-hoc* contact force is added to the model to prevent real contact (see, e.g., [9–13]), or contact is simply added as an additional constraint to the FSI problem (see, e.g., [14–18]).

A recent approach to circumvent the two difficulties discussed above is to consider a poroelastic modelling of the fluid seepage induced by the roughness of the contacting solids, which guarantees mechanical consistency at contact (see [5]). The price to pay is a model with a high computational complexity (notably, with respect to traditional FSI without contact). Alternatively, an artificial extension of the fluid stresses to the contact surface can be used [19,20]. The choice of the extension can, however, have an unphysical impact on the release of contact and moreover, in case of an unrelaxed contact treatment, very small fluid regions can get trapped between the solid and the contacting wall, which are difficult to treat numerically [20].

In the present work, we derive a simple, but mechanically consistent model for FSI with contact which circumvents these issues. Motivated by the fact that in reality two solid bodies will never be exactly in contact throughout their surfaces (for example due to microscopic roughness), the main idea is to consider seepage of the fluid through the contact surface. To fix ideas, let $\Omega \subset \mathbb{R}^d$ for $d = 2, 3$ be the overall domain consisting of fluid and solid subdomains. Seepage is modelled by the introduction of a $(d-1)$-dimensional porous layer that adheres to a solid boundary, where contact might take place. This can be considered as a generalised boundary condition, or a bulk surface coupling in the spirit of [21]. In our case, however, the free-flow Navier–Stokes' system is coupled to a surface Darcy equation. This model goes back to [22] and is of interest in its own right, as discussed in the note [23]. By combining this porous layer approach for seepage with a relaxed treatment of contact (see, e.g., [19]), we obtain a simple and mechanically consistent model for fluid–structure interactions with contact.

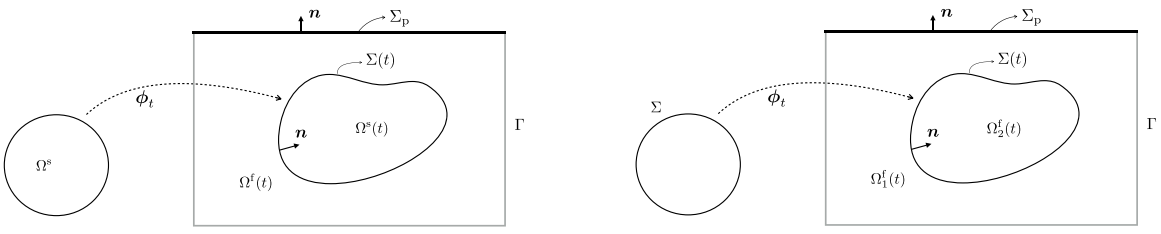
We implement the approach using different solid models, coordinate systems and discretisations, which are able to deal with topology changes in the fluid domain. Concerning coordinate systems, we consider both an Immersed approach going back to Peskin [24] as well as a Fully Eulerian approach [11,25–27]. For discretisation, we use the unfitted finite element method of [28–30] and the two-scale interface fitting approach of [31]. We illustrate the modelling capacity in a series of computational examples in two dimensions, including a beam solid model and a thick-walled solid model. To deal with the non-linearly changing interface conditions a unified Nitsche approach is applied which takes care of the FSI coupling and contact conditions simultaneously [19].

Indeed, depending on if no-slip conditions are imposed or if the porous medium approach proposed here is used, the approximations will converge to different solutions for $h \to 0$. If we consider the case of a bouncing ball the use of no-slip conditions will lead to a sequence of solutions that converge to a ball that does not bounce, whereas the solutions obtained with the porous medium approach converge to a certain bouncing height that depends on the parameters of the Darcy model. Recent comparisons of computational methods with experimental studies [4,13] confirm that the second behaviour is the physical one. Of course the parameters of the model need to be fixed through experimental studies, or otherwise.

An outline of the paper is as follows. In Section 2, we introduce the Navier–Stokes–Darcy coupling as well as the FSI–contact model. The variational formulation and the discretisation based on Nitsche's method is described in Section 3. In Section 4 we give detailed numerical studies both in the case of a thin- and thick-walled solid model. Finally, some conclusions are drawn in Section 5.

2. Equations

In this section, we derive the Navier–Stokes–Darcy coupling, and subsequently the equations for fluid–structure–porous–contact interaction. For simplicity, we will consider that contact takes place at a given fixed plane surface.



(a) Coupling with a thick-walled solid

(b) Coupling with a closed thin-walled solid

Fig. 1. Geometric configurations of the fluid and solid domains.

This can be either an exterior rigid wall or a symmetry boundary within the fluid domain, which is relevant for example in the case of contact between two symmetric valves. The case of two-body contact is not considered here.

The fluid equations in $\Omega^{\mathrm{f}}(t) \subset \mathbb{R}^d$ will be coupled to a fixed $(d-1)$-dimensional porous layer Σ_{p} on the exterior boundary, where contact might take place. The fluid is described by the Navier–Stokes equations in Eulerian formalism and the structure by a possibly non-linear solid model. We consider both $(d-1)$-dimensional thin-walled solids and d-dimensional thick-walled solids.

2.1. Problem setting

Let $\Omega = \overline{\Omega^{\mathrm{s}}(t)} \cup \Omega^{\mathrm{f}}(t) \subset \mathbb{R}^d$ be a current configuration of the complete domain of interest, with boundary $\partial\Omega := \Gamma \cup \Sigma_{\mathrm{p}}$, where Σ_{p} denotes the part of the boundary where contact might take place (see Fig. 1). There, a thin porous fluid layer is considered. The solid domain $\Omega^{\mathrm{s}}(t)$ can be either a surface (actually the solid mid-surface) or a domain with positive volume in $\mathbb{R}^d$ in the case of the coupling with a thick-walled solid. The current fluid–structure interface is denoted by $\Sigma(t)$ and coincides with $\Omega^{\mathrm{s}}(t)$ in the case of a thin-walled solid. The corresponding reference configurations are denoted by Σ and Ω^{s}.

The structure is allowed to move freely within the domain Ω. The current position of the interface $\Sigma(t)$ and the solid domain $\Omega^{\mathrm{s}}(t)$ are described in terms of a deformation map $\boldsymbol{\phi} : \Omega^{\mathrm{s}} \times \mathbb{R}^+ \longrightarrow \mathbb{R}^d$ such that $\Omega^{\mathrm{s}}(t) = \boldsymbol{\phi}(\Omega^{\mathrm{s}}, t)$ and $\Sigma(t) = \boldsymbol{\phi}(\Sigma, t)$, with $\boldsymbol{\phi} := \boldsymbol{I}_{\Omega^{\mathrm{s}}} + \boldsymbol{d}$ and where $\boldsymbol{d}$ denotes the solid displacement. To simplify the notation we will refer to $\boldsymbol{\phi}_t := \boldsymbol{\phi}(\cdot, t)$, so that we can also write $\Omega^{\mathrm{s}}(t) = \boldsymbol{\phi}_t(\Omega^{\mathrm{s}})$, $\Sigma(t) = \boldsymbol{\phi}_t(\Sigma)$. The fluid domain is time-dependent, namely $\Omega^{\mathrm{f}}(t) := \Omega \setminus (\Omega^{\mathrm{s}}(t) \cup \Sigma(t)) \subset \mathbb{R}^d$ with boundary $\partial\Omega^{\mathrm{f}}(t) = \Sigma(t) \cup \Gamma \cup \Sigma_p$. In the case of a closed thin-walled structure, the solid domain $\Omega^{\mathrm{s}}(t)$ divides $\Omega^{\mathrm{f}}(t)$ into two subdomains $\Omega^{\mathrm{f}}(t) = \Omega_1^{\mathrm{f}}(t) \cup \Omega_2^{\mathrm{f}}(t)$, with respective unit normals $\boldsymbol{n}_1 := \boldsymbol{n}$ and $\boldsymbol{n}_2 := -\boldsymbol{n}$, as shown in Fig. 1(b). Similarly, in the case of a thick-walled solid, the interface $\Sigma(t)$ divides Ω into a solid part $\Omega^{\mathrm{s}}(t)$ and a fluid part $\Omega^{f}(t)$. We write $H^1_\Gamma(\Omega)$ for the first-order Sobolev space with vanishing trace on $\Gamma \subset \partial\Omega$.

For a given field f defined in Ω (possibly discontinuous across the interface), we can define its one-sided restrictions, denoted by f_1 and f_2, as

$$f_1(\boldsymbol{x}) := \lim_{\xi \to 0^-} f(\boldsymbol{x} + \xi \boldsymbol{n}_1), \qquad f_2(\boldsymbol{x}) := \lim_{\xi \to 0^-} f(\boldsymbol{x} + \xi \boldsymbol{n}_2),$$

for all $\boldsymbol{x} \in \Sigma(t)$, and the following jump and average operators across $\Sigma(t)$:

$$[\![f]\!] := f_1 - f_2 \quad [\![f\boldsymbol{n}]\!] := f_1\boldsymbol{n}_1 + f_2\boldsymbol{n}_2, \quad \{\!\{f\}\!\} := \frac{1}{2}(f_1 + f_2). \tag{1}$$

In the case of a thin structure that has a boundary inside the fluid domain (for example with a tip), these quantities can be defined similarly. For the details, we refer to [30] and Remark 3.1.

While the fluid and solid equations are standard and will be introduced in Section 2.3, we give some details on the porous medium model in the following section.

2.2. Porous medium model and Navier–Stokes-Darcy coupling

We consider the configuration sketched in Fig. 2, where a thin porous layer $\Omega_{\mathrm{p}} = \Sigma_{\mathrm{p}} \times (-\frac{\epsilon_{\mathrm{p}}}{2}, \frac{\epsilon_{\mathrm{p}}}{2}) \in \mathbb{R}^d$ $(d = 2, 3)$ with midsurface Σ_{p} is coupled to a surrounding fluid in a fixed domain $\Omega^{\mathrm{f}} \subset \mathbb{R}^d$. The surrounding fluid is governed

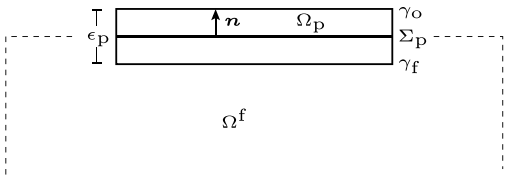


Fig. 2. Porous medium domain Ω_p with interface γ_f to Ω^f and exterior boundary γ_o.

by the Navier–Stokes equations

$$\begin{cases} \rho_f\big(\partial_t \boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u}\big) - \operatorname{div}\boldsymbol{\sigma}_f(\boldsymbol{u},p) = \rho_f \boldsymbol{f}_f & \text{in} \quad \Omega^f, \\ \operatorname{div}\boldsymbol{u} = 0 & \text{in} \quad \Omega^f. \end{cases} \tag{2}$$

Here, $\boldsymbol{u}$ denotes the fluid velocity, p the fluid pressure and $\boldsymbol{f}_f$ is a volume force defined in Ω^f. The Cauchy stress tensor is given by

$$\boldsymbol{\sigma}_f := 2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\boldsymbol{I}, \quad \boldsymbol{\varepsilon}(\boldsymbol{u}) = \frac{1}{2}\left(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^T\right), \tag{3}$$

where $\boldsymbol{I}$ denotes the identity matrix, ρ_f the fluid density, ν_f the kinematic and $\mu = \rho_f\nu_f$ the dynamic viscosity.

In the porous domain Ω_p, we assume a Darcy law

$$\begin{cases} \boldsymbol{u}_l + \boldsymbol{K}\nabla p_l = 0 & \text{in} \quad \Omega_p, \\ \nabla\cdot\boldsymbol{u}_l = 0 & \text{in} \quad \Omega_p, \end{cases} \tag{4}$$

where $\boldsymbol{u}_l$ denotes the Darcy velocity, p_l the Darcy pressure and $\boldsymbol{K}$ is a $d\times d$ matrix such that the following decomposition holds

$$\boldsymbol{K}\nabla p_l = K_\tau \nabla_\tau p_l + K_n \boldsymbol{\partial}_n p_l,$$

with $K_\tau, K_n \in \mathbb{R}^+$. According to [32] the conductivities K_n and K_τ are proportional to the inverse of the kinematic viscosity ν_f^{-1}. Here, $\boldsymbol{n}$ is the unit normal vector of the mid-surface Σ_p that points towards the exterior boundary γ_o, $\boldsymbol{\partial}_n = \boldsymbol{n}\partial_n$ and $\nabla_\tau := P_\tau\nabla$ stands for the corresponding tangential part of the gradient

$$P_\tau := (\boldsymbol{I} - \boldsymbol{n}\otimes\boldsymbol{n}).$$

We assume that the porous layer is very thin and consider the limit case $\epsilon_p \to 0$. Let the outer boundary of Ω_p be denoted by γ_o and the interior boundary connecting to the fluid domain Ω^f by γ_f, see Fig. 2. We assume zero normal velocity ($\boldsymbol{u}_l\cdot\boldsymbol{n} = 0$) on the outer boundary γ_o and continuity of normal velocities and normal stresses on γ_f. For the tangential fluid stresses, we consider the Beavers–Joseph–Saffman coupling conditions [33]. The coupling conditions between porous medium Ω_p and fluid Ω^f read

$$\begin{cases} \sigma_{f,n} = -p_l & \text{on} \quad \gamma_f, \\ \boldsymbol{u}\cdot\boldsymbol{n} = \boldsymbol{u}_l\cdot\boldsymbol{n} & \text{on} \quad \gamma_f, \\ \sigma_{f,\tau} = -\dfrac{\alpha}{\sqrt{K_\tau\epsilon_p}}\boldsymbol{u}_\tau & \text{on} \quad \gamma_f \end{cases} \tag{5}$$

where $\boldsymbol{u}_\tau := P_\tau\boldsymbol{u}$ denotes the tangential part of the velocity vector and $\sigma_{f,n} = \boldsymbol{n}^T\boldsymbol{\sigma}_f\boldsymbol{n}$ and $\sigma_{f,\tau} = P_\tau\boldsymbol{\sigma}_f\boldsymbol{n}$ are the normal and tangential part of the Cauchy stress tensor $\boldsymbol{\sigma}_f$ introduced above. We note that the condition for the tangential stresses in the last line of (5) corresponds to a Navier-slip boundary condition for the fluid. In contrast to this boundary condition for the fluid, the normal velocity $\boldsymbol{u}\cdot\boldsymbol{n}$ is not zero here, as the fluid can enter the porous layer.

The appropriate choice of the parameter α in the last line of (5) depends on the application. In the case that γ_f corresponds to a symmetry boundary within a larger fluid domain, where contact can take place, for example between two contacting valves, it is appropriate to set $\alpha = 0$ (pure slip). If the porous layer is, however, placed at a rigid wall, the Beavers–Joseph–Saffman condition with $\alpha > 0$ is more appropriate [33,34]. The parameter α depends on the structure of the porous layer. Values $0.01 < \alpha < 5$ have been suggested in [35], depending on the structure of the porous medium. We will consider both kinds of conditions in the numerical examples of Section 4.

Introducing the averaged porous pressure P_l as

$$P_\mathrm{l} = \frac{1}{2}\left(p_\mathrm{l}|\gamma_\mathrm{f} + p_\mathrm{l}|\gamma_\mathrm{o}\right) \quad \text{in} \quad \Sigma_\mathrm{p}, \tag{6}$$

the following equations can be derived in the limit case $\epsilon_\mathrm{p} \to 0$ (see [22,23])

$$\begin{cases} -\nabla_\tau \cdot \left(\epsilon_\mathrm{p} K_\tau \nabla_\tau P_\mathrm{l}\right) = \boldsymbol{u}\cdot\boldsymbol{n} & \text{on} \quad \Sigma_\mathrm{p}, \\ \sigma_{\mathrm{f},n} = -P_\mathrm{l} - \dfrac{\epsilon_\mathrm{p} K_n^{-1}}{4}\boldsymbol{u}\cdot\boldsymbol{n} & \text{on} \quad \Sigma_\mathrm{p}, \\ \sigma_{\mathrm{f},\tau} = -\dfrac{\alpha}{\sqrt{K_\tau \epsilon_\mathrm{p}}}\boldsymbol{u}_\tau & \text{on} \quad \Sigma_\mathrm{p}. \end{cases} \tag{7}$$

Note that the only remaining porous medium variable is the averaged pressure P_l. In the limit $K_n, K_\tau \to 0$, the coupling conditions turn into a no-slip boundary condition for the fluid on Σ_p.

2.3. Fluid–structure–porous–contact interaction model

We assume that $\Omega^\mathrm{f}(t)$ is filled with an incompressible fluid governed by the Navier–Stokes equations. The domain $\Omega^\mathrm{s}(t)$ is occupied by a solid medium described by a beam or shell solid model (given in terms of an abstract surface differential operator $\boldsymbol{L}$) on a $(d-1)$-dimensional domain Σ or by the elastodynamics equations in the case of a d-dimensional domain Ω^s. The fluid and solid equations are coupled with no-slip interface conditions on the fluid–structure interface $\Sigma(t)$. The solid is constrained to not penetrate into the porous medium Σ_p via the (relaxed) unilateral frictionless contact conditions

$$\boldsymbol{d}\cdot\boldsymbol{n} - g_\epsilon \le 0, \quad \lambda \le 0, \quad \lambda(\boldsymbol{d}\cdot\boldsymbol{n} - g_\epsilon) = 0 \quad \text{on } \Sigma. \tag{8}$$

Here, $g_\epsilon := g - \epsilon_g$, where g denotes the gap function to Σ_p and $\epsilon_g > 0$ is a small parameter. The symbol λ stands for the normal component of the contact traction, which corresponds to the Lagrange multiplier associated to the no-penetration condition.

The proposed fluid–structure–porous–contact interaction model is hence formulated as follows: Find the fluid velocity and pressure $\boldsymbol{u} : \Omega^\mathrm{f}\times\mathbb{R}^+ \to \mathbb{R}^d$, $p : \Omega^\mathrm{f}\times\mathbb{R}^+ \to \mathbb{R}$, the solid displacement and velocity $\boldsymbol{d} : \Omega^s\times\mathbb{R}^+ \to \mathbb{R}^d$, $\dot{\boldsymbol{d}} : \Omega^s \times \mathbb{R}^+ \to \mathbb{R}^d$, the Darcy porous pressure $P_\mathrm{l} : \Sigma_p \times \mathbb{R}^+ \to \mathbb{R}$ and the Lagrange multiplier $\lambda : \Sigma \times \mathbb{R}^+ \to \mathbb{R}$ such that, for all $t \in \mathbb{R}^+$, the following relations are satisfied

- Fluid problem:

$$\begin{cases} \rho_\mathrm{f}\left(\partial_t \boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u}\right) - \operatorname{div}\boldsymbol{\sigma}_\mathrm{f}(\boldsymbol{u}, p) = \rho_\mathrm{f}\boldsymbol{f}_f & \text{in} \quad \Omega^\mathrm{f}(t), \\ \operatorname{div}\boldsymbol{u} = 0 & \text{in} \quad \Omega^\mathrm{f}(t), \\ \boldsymbol{u} = \boldsymbol{0} & \text{on} \quad \Gamma, \end{cases} \tag{9}$$

- Porous layer:

$$\begin{cases} -\nabla_\tau \cdot \left(\epsilon_\mathrm{p} K_\tau \nabla_\tau P_\mathrm{l}\right) = \boldsymbol{u}\cdot\boldsymbol{n} & \text{on} \quad \Sigma_\mathrm{p}, \\ \epsilon_\mathrm{p} K_\tau \tau \cdot \nabla_\tau P_\mathrm{l} = 0 & \text{on} \quad \partial\Sigma_\mathrm{p}, \end{cases} \tag{10}$$

- Solid problem:

$$\begin{cases} \rho_\mathrm{s}\epsilon_\mathrm{s}\partial_t\dot{\boldsymbol{d}} + \boldsymbol{L}(\boldsymbol{d}) = \boldsymbol{T} & \text{on} \quad \Omega^\mathrm{s} = \Sigma, \\ \dot{\boldsymbol{d}} = \partial_t\boldsymbol{d} & \text{on} \quad \Omega^\mathrm{s} = \Sigma, \\ \boldsymbol{d} = \boldsymbol{0} & \text{on} \quad \partial\Omega^\mathrm{s}\cap\Gamma, \end{cases} \tag{11}$$

  in case of a thin-walled solid, or

$$\begin{cases} \rho_\mathrm{s}\partial_t\dot{\boldsymbol{d}} - \operatorname{div}\boldsymbol{\sigma}_\mathrm{s}(\boldsymbol{d}) = \rho_\mathrm{s}\boldsymbol{f}_s & \text{on} \quad \Omega^\mathrm{s}, \\ \dot{\boldsymbol{d}} = \partial_t\boldsymbol{d} & \text{on} \quad \Omega^\mathrm{s}, \\ \boldsymbol{d} = \boldsymbol{0} & \text{on} \quad \partial\Omega^\mathrm{s}\cap\Gamma, \end{cases} \tag{12}$$

  in case of a thick-walled solid.

- Contact conditions:

$$\boldsymbol{d}\cdot\boldsymbol{n}-g_\epsilon\le 0,\quad \lambda\le 0,\quad \lambda(\boldsymbol{d}\cdot\boldsymbol{n}-g_\epsilon)=0\quad \text{on } \Sigma. \tag{13}$$

- Fluid–structure coupling conditions:

$$\begin{cases}\boldsymbol{\phi}=\boldsymbol{I}_{\Omega^{\mathrm{s}}}+\boldsymbol{d},\quad \Omega_s(t)=\boldsymbol{\phi}_t(\Omega^{\mathrm{s}}),\quad \Omega^{\mathrm{f}}(t)=\Omega\backslash\Omega^{\mathrm{s}}(t),\\ \boldsymbol{u}=\dot{\boldsymbol{d}}\circ\boldsymbol{\phi}_t^{-1}\quad \text{on}\quad \Sigma(t),\end{cases} \tag{14}$$

and

$$\int_\Sigma(\boldsymbol{T}-\lambda\boldsymbol{n})\cdot\boldsymbol{w}=-\int_{\Sigma(t)}[\![\boldsymbol{\sigma}_{\mathrm{f}}(\boldsymbol{u},p)\boldsymbol{n}]\!]\cdot\boldsymbol{w}\circ\boldsymbol{\phi}_t^{-1}, \tag{15}$$

or

$$\int_\Sigma(\boldsymbol{\sigma}_{\mathrm{s}}-\lambda\boldsymbol{I})\boldsymbol{n}\cdot\boldsymbol{w}=\int_{\Sigma(t)}\boldsymbol{\sigma}_{\mathrm{f}}(\boldsymbol{u},p)\boldsymbol{n}\cdot\boldsymbol{w}\circ\boldsymbol{\phi}_t^{-1} \tag{16}$$

for all test functions $\boldsymbol{w}$, respectively in the case of the coupling with a thin- or thick-walled solid.

- Fluid–porous coupling conditions:

$$\begin{cases}\sigma_{\mathrm{f},n}=\underbrace{-P_{\mathrm{l}}-\dfrac{\epsilon_{\mathrm{p}}K_n^{-1}}{4}\boldsymbol{u}\cdot\boldsymbol{n}}_{=:\,\sigma_{\mathrm{p}}}\quad \text{on}\quad \Sigma_{\mathrm{p}},\\ \sigma_{\mathrm{f},\tau}=-\dfrac{\alpha}{\sqrt{K_\tau\epsilon_{\mathrm{p}}}}u_\tau\qquad \text{on}\quad \Sigma_{\mathrm{p}}.\end{cases} \tag{17}$$

Here, the symbol ρ_{s} stands for the solid density and $\boldsymbol{f}_f$ and $\boldsymbol{f}_s$ are volume forces in the fluid and solid, respectively. In the case of a thin-walled solid, the symbol ϵ_{s} denotes the solid thickness and $\boldsymbol{T}$ the surface force acting on the reference configuration of the thin-walled solid mid-surface. The relations in (14)–(16) enforce the geometrical compatibility and the kinematic and the dynamic coupling at the interface between the fluid and the solid, respectively. It should be noted that the no-penetration condition in (13) is already imposed at an ϵ_g-distance to the porous layer Σ_p. This modelling simplification circumvents most of the numerical difficulties associated with the topological change in the fluid domain induced by the exact contact condition (i.e., with $\epsilon=0$), such as switching between the contact and fluid–solid interfaces and presence of isolated small fluid regions (see [20]). Moreover, it also facilitates the explicit treatment of the geometric condition in the fluid–structure coupling (see Section 3).

2.3.1. Mechanical consistency

In the fluid–structure–porous–contact interaction model (9)–(17), a very thin fluid layer remains between the solid and porous medium during contact. Usually, this relaxation of the contact conditions comes with a drawback, which is that the fluid forces arising in the layer have no physical meaning, but might distort the contact dynamics. This is different when a porous layer is considered on Σ_p. Owing to relations (7), the behaviour of the fluid confined in the contact layer is expected to be very close to the one of the porous fluid. Indeed, this is a consequence of the kinematic-dynamics relations $(7)_{1,2}$, which are enforced both during and in absence of contact. If a part of $\Sigma(t)$ is in contact with Σ_{p} according to (8), the value of $\sigma_{\mathrm{f},n}$ (resp. $\boldsymbol{u}\cdot\boldsymbol{n}$) on this part $\Sigma(t)$ will be close to σ_{p} (resp. $\boldsymbol{u}_{\mathrm{l}}\cdot\boldsymbol{n}$) on the corresponding part of Σ_{p}. As a result, all the kinematic and dynamic relations acting on the solid during contact have a physical meaning, which guarantees the mechanical consistency of the proposed fluid–structure–contact interaction model.

More precisely, owing to (16), in the case of a thick-walled solid the Lagrange multiplier for the no-penetration condition will formally assume the form

$$\lambda=\underbrace{\sigma_{\mathrm{s},n}-\sigma_{\mathrm{f},n}}_{=:\,[\![\sigma_n]\!]}\approx\sigma_{\mathrm{s},n}-\sigma_{\mathrm{p}}\circ\pi\quad \text{on}\quad \Sigma,$$

where we write $\sigma_{\mathrm{s},n}:=\boldsymbol{n}^{\mathrm{T}}\boldsymbol{\sigma}_s\boldsymbol{n}$ for the solid normal traction and π denotes a (closest-point) projection from $\Sigma(t)$ to Σ_{p}. Both, the solid stresses $\sigma_{\mathrm{s},n}$ and the "porous stresses" σ_{p} have a physical meaning during solid-porous contact.

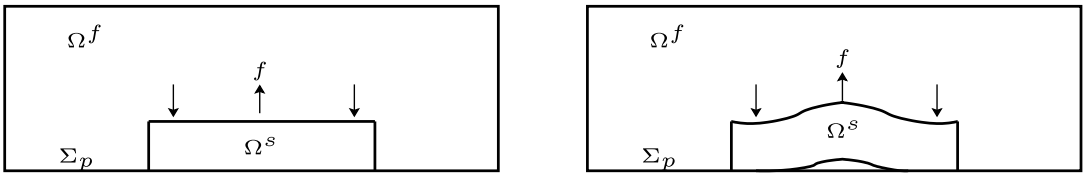


Fig. 3. Illustrative example to motivate the role of seepage in fluid–structure interaction with contact: Contact of a solid body with the lower wall is released in the central part of the contact surface, when a specific force f is applied. Without seepage on Σ_p a vacuum would be created.

Hence, this porous–contact approach gives a physical meaning to the stresses generated in the infinitesimal fluid layer, in contrast to the relaxed contact formulation in [19], where the fluid stresses $\sigma_{\mathrm{f},n}$ did not allow for a direct physical interpretation. A similar argumentation holds true for the thin-walled solid case, where

$$\lambda = \left(\rho_{\mathrm{s}}\epsilon^{\mathrm{s}}\partial_t\dot{\boldsymbol{d}} + \boldsymbol{L}(\boldsymbol{d})\right)\cdot\boldsymbol{n} + [\![\sigma_{\mathrm{f},n}]\!] \approx \left(\rho_{\mathrm{s}}\epsilon^{\mathrm{s}}\partial_t\dot{\boldsymbol{d}} + \boldsymbol{L}(\boldsymbol{d})\right)\cdot\boldsymbol{n} + \sigma_{\mathrm{p}}\circ\pi - \sigma_{\mathrm{f},n}|_2 \quad \text{on} \quad \Sigma.$$

In the spirit of [36], the Lagrange multiplier can further be eliminated, which in the case of a thick-walled solid results in the non-linear contact condition

$$[\![\sigma_n]\!] = -\gamma_{\mathrm{C}}\underbrace{\left[\boldsymbol{d}\cdot\boldsymbol{n} - g_\epsilon - \gamma_{\mathrm{C}}^{-1}[\![\sigma_n]\!]\right]_+}_{=:\, P_{\gamma_{\mathrm{C}}}(\boldsymbol{d}\cdot\boldsymbol{n},\, [\![\sigma_n]\!])} \quad \text{on} \quad \Sigma, \tag{18}$$

for $\gamma_{\mathrm{C}} > 0$. This can be embedded in an elegant way in the variational formulation using a Nitsche-based approach, see [19] and Section 3. For a thin solid, a similar approach is possible, with the additional difficulty that the normal solid traction on the mid-surface is given in terms of the normal PDE residual $\left(\rho_{\mathrm{s}}\epsilon^{\mathrm{s}}\partial_t\dot{\boldsymbol{d}} + \boldsymbol{L}(\boldsymbol{d})\right)\cdot\boldsymbol{n}$, which is rarely available at the discrete level. On the other hand, it has been shown that, for the case of a thin-walled solid, and P_1 finite elements a pure penalty approach (i.e. neglecting the normal traction λ in the term P_{γ_c}) leads to a first-order approximation, see [37,38]. The detailed variational formulations for both the case of a thick- and a thin-walled solid will be given in the next section.

Besides its mechanical consistency, the main advantage of this method is its simplicity. The porous medium and the structure are always coupled with the fluid only and never directly to each other. This avoids switches in the variational formulation, which would be necessary in the transition between fluid–solid and solid-porous interaction [23]. On the other hand, the solid perceives indirectly the presence of the porous layer through the fluid stresses and velocity during contact. The resulting numerical approach is highly competitive in terms of computational costs compared to approaches using Lagrange multipliers and/or active-sets.

2.3.2. Seepage

The proposed fluid–structure–porous–contact interaction model (9)–(17) allows for seepage in the sense that fluid can flow through the porous layer Σ_{p}, for example to connect a cavity in the central part of the contact surface with the exterior fluid. These could emerge when the impact of the structure happens in the lateral parts of the structure first or when contact of the solid is released in a central part of the contact surface only. This is an important aspect in the modelling of fluid–structure–contact interaction, as otherwise unphysical configurations might result. As an example consider the situation sketched in Fig. 3, where a solid body is in contact with the lower wall Σ_{p} at initial time (left sketch). When a (sufficiently strong) force $\boldsymbol{f}$ is applied in the central part of Ω^{s}, while the body is fixed at the lateral parts, contact will be released in the central part only. If no seepage along Σ_{p} is allowed, a vacuum would emerge between Ω^{s} and Σ_{p}. While one could argue that this paradox is already circumvented by using the relaxed contact conditions (8), we note that only the porous layer gives a physical meaning to the fluid filling the contact layer.

3. Numerical methods

This section is devoted to the numerical discretisation of the fluid–structure–porous–contact interaction model (9)–(17). Two numerical approaches will be considered which basically depend on thin- or thick-walled nature of

the solid model and on the formalism used for the fluid–structure coupling (mixed Lagrangian–Eulerian or fully Eulerian formalisms). For an accurate discretisation of the fluid–structure–porous–contact interaction model (9)–(17), two different strategies will be considered in the numerical examples reported in Section 4. The first strategy is based on an unfitted Nitsche-XFEM method, drawing on [29,30]. The second strategy is a fitted finite element method, following [31]. In both cases, we will use equal-order finite element methods.

To fix ideas, we will present the numerical approaches for two specific combinations of these components (mixed Lagrangian–Eulerian vs. Fully Eulerian, fitted vs. unfitted finite elements, thick-walled vs. thin-walled structures), namely a mixed Lagrangian–Eulerian approach with a thin-walled solid using unfitted finite elements in Section 3.1 and a fully Eulerian approach with a thick-walled solid using fitted finite element discretisation in Section 3.2. Different combinations are possible as well, but will not be considered in the remainder of this article.

3.1. Lagrange–Eulerian formalism with immersed thin-walled solids

In what follows, the parameter $\delta t > 0$ stands for the time-step length and $t_n := n\delta t$ denotes the time instant at time level $n \in \mathbb{N}$. The symbol x^n generally denotes an approximation of $x(t_n)$, for a given time valued function $x(t)$. We also introduced the notation

$$\partial_{\delta t} x^n := \frac{1}{\delta t}\left(x^n - x^{n-1}\right),$$

for the first-order backward difference.

We consider the fluid–structure–porous–contact interaction problem (9)–(17) in the case of the coupling with immersed thin-walled solids. The time discretisation is performed with a backward-Euler scheme, including a semi-implicit treatment of the convective term in (9) and an explicit treatment of the geometric coupling $(14)_1$. As regards the spatial discretisation, an unfitted finite element approximation with overlapping meshes is considered for the fluid–solid coupling, by drawing on the Nitsche-XFEM method reported in [29,30]. The Fluid–porous system is discretised by a standard fitted finite element approximation.

For the solid, we start by assuming that there exists a positive form $a^{\mathrm{s}} : \boldsymbol{W} \times \boldsymbol{W} \longrightarrow \mathbb{R}$, linear with respect to the second argument, such that

$$a^{\mathrm{s}}(\boldsymbol{d}, \boldsymbol{w}) = \left(\boldsymbol{L}(\boldsymbol{d}), \boldsymbol{w}\right)_{\Sigma}$$

for all $\boldsymbol{w} \in \boldsymbol{W} := [W^{1,\infty}{}_{\Gamma\cap\partial\Sigma}(\Sigma)]^d$, where $\boldsymbol{W}$ stands for the space of admissible displacements. Let $\{\mathcal{T}_h^{\mathrm{s}}\}_{0<h<1}$ be a family of triangulations of Σ. We consider the standard space of continuous piecewise affine functions

$$X_h^{\mathrm{s}} := \left\{v_h \in C^0(\overline{\Sigma}) \,\middle|\, v_{h|K} \in \mathbb{P}_1(K), \quad \forall K \in \mathcal{T}_h^{\mathrm{s}}\right\}, \qquad \boldsymbol{W}_h := [X_h^{\mathrm{s}}]^d \cap \boldsymbol{W}.$$

The contact condition (13) is approximated via a penalty method (see, e.g., [37]), by adding the following penalty term into the solid discrete problem:

$$\frac{\gamma_{\mathrm{c}} E \epsilon_{\mathrm{s}}}{h^2}\left(\left[\boldsymbol{d}_h^n \cdot \boldsymbol{n} - g_\varepsilon\right]_+, \boldsymbol{w}_h \cdot \boldsymbol{n}\right)_{\Sigma},$$

where E is the solid Young modulus and $\gamma_{\mathrm{c}} > 0$ is the (dimensionless) penalty parameter.

For a given discrete displacement approximation $\boldsymbol{d}_h^n \in \boldsymbol{W}_h$ at time t_n, we define its associated deformation map by $\boldsymbol{\phi}_h^n := \boldsymbol{I}_\Sigma + \boldsymbol{d}_h^n$. This map characterises the current solid configuration (i.e., at time level n), as $\Sigma^n := \boldsymbol{\phi}_h^n(\Sigma)$. As indicated above, we consider an explicit update for the physical fluid domain in $(14)_1$, namely,

$$\Omega^{\mathrm{f},n} := \Omega \setminus \Sigma^{n-1}, \tag{19}$$

which has the effect of removing the geometrical non-linearities in the fluid problem (9).

Let $\{\mathcal{T}_h\}_{0<h<1}$ a family of triangulations of Ω. Owing to (19), for each $\mathcal{T}_h$ we defined two overlapping meshes $\mathcal{T}_{h,i}^n$, $i = 1, 2$, such that $\mathcal{T}_{h,i}^n$ covers the ith fluid region $\Omega_i^{\mathrm{f},n}$ defined by Σ^{n-1} through (19). Note that each triangulation $\mathcal{T}_{h,i}^n$ is fitted to the exterior boundary $\Gamma \cup \Sigma_{\mathrm{p}}$, but in general not to Σ^{n-1} (nor $\mathcal{T}_h^{\mathrm{s}}$), see Fig. 4.

There will be duplicated elements, i.e., such that $K \in \mathcal{T}_{h,1}^n \cap \mathcal{T}_{h,2}^n$, but this is only allowed when $K \cap \Sigma^{n-1} \neq \emptyset$. We denote by $\Omega_{h,i}^{\mathrm{f},n}$ the domain covered by $\mathcal{T}_{h,i}^n$, viz.,

$$\Omega_{h,i}^{\mathrm{f},n} := \mathrm{int}\left(\bigcup_{K\in\mathcal{T}_{h,i}^n} K\right).$$

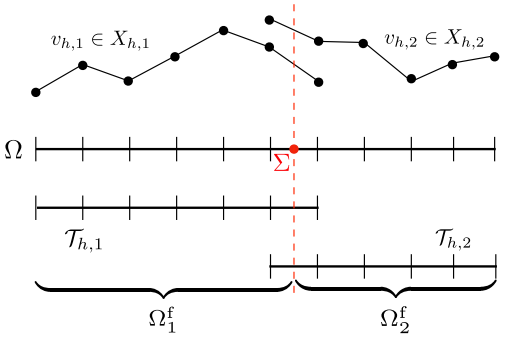


Fig. 4. One dimensional illustration of the overlapping meshes $\mathcal{T}_{h,i}^n$ and of the construction of the discrete spaces $X_{h,i}^n$.

We can hence introduce the following spaces of continuous piecewise affine functions

$$X_{h,i}^n := \left\{ v_h \in C^0(\overline{\Omega_{h,i}^{\mathrm{f},n}}) \,\middle|\, v_{h|K} \in \mathbb{P}_1(K), \quad \forall K \in \mathcal{T}_{h,i}^n \right\}.$$

and set

$$\boldsymbol{V}_{h,i}^n := [X_{h,i}^n]^d \cap [H_\Gamma^1(\Omega^{\mathrm{f},n})]^d, \quad Q_{h,i}^n := X_{h,i}^n \cap L_0^2(\Omega^{\mathrm{f},n}).$$

For the approximation of the fluid velocity and pressure we will consider the following time-dependent discrete product spaces

$$\boldsymbol{V}_h^n := \boldsymbol{V}_{h,1}^n \times \boldsymbol{V}_{h,2}^n, \quad Q_h^n := Q_{h,1}^n \times Q_{h,2}^n. \tag{20}$$

The functions in (20) are continuous in the physical fluid domain $\Omega_i^{\mathrm{f},n}$, but discontinuous across the interface location Σ^{n-1} (see Fig. 4).

Algorithm 1 Strongly coupled scheme for fluid–structure–porous–contact interaction (thin-walled solid).

For $n \geq 1$:

1. Interface update:

$$\boldsymbol{\phi}_h^{n-1} = \boldsymbol{I}_\Sigma + \boldsymbol{d}_h^{n-1}, \quad \Sigma^{n-1} = \boldsymbol{\phi}_h^{n-1}(\Sigma), \quad \Omega^{\mathrm{f},n} = \Omega \backslash \Sigma^{n-1}.$$

2. Find $\left(\boldsymbol{u}_h^n, p_h^n, \dot{\boldsymbol{d}}_h^n, \boldsymbol{d}_h^n, P_{\mathrm{l},h}^n\right) \in \boldsymbol{V}_h^n \times Q_h^n \times \boldsymbol{W}_h \times \boldsymbol{W}_h \times \boldsymbol{S}_h$ with $\dot{\boldsymbol{d}}_h^n = \partial_{\delta t} \boldsymbol{d}_h^n$ such that

$$\begin{aligned}
&\rho_{\mathrm{f}}\left(\partial_{\delta t} \boldsymbol{u}_h^n, \boldsymbol{v}_h\right)_{\Omega^n} + a_h^{\mathrm{f},n}\left(\boldsymbol{u}_h^{n-1}; (\boldsymbol{u}_h^n, p_h^n), (\boldsymbol{v}_h, q_h)\right) + \rho_{\mathrm{s}} \epsilon^{\mathrm{s}}\left(\partial_{\delta t} \dot{\boldsymbol{d}}_h^n, \boldsymbol{w}_h\right)_\Sigma + a^{\mathrm{s}}(\boldsymbol{d}_h^n, \boldsymbol{w}_h) \\
&\quad + \frac{\gamma_{\mathrm{c}} E \epsilon_{\mathrm{s}}}{h^2}\left(\left[\boldsymbol{d}_h^n \cdot \boldsymbol{n} - g_\epsilon\right]_+, \boldsymbol{w}_h \cdot \boldsymbol{n}\right)_\Sigma - \sum_{i=1}^2 \left(\boldsymbol{\sigma}_{\mathrm{f}}(\boldsymbol{u}_{h,i}^n, p_{h,i}^n)\boldsymbol{n}_i, \boldsymbol{v}_{h,i} - \boldsymbol{w}_h\right)_{\Sigma^{n-1}} \\
&\quad - \sum_{i=1}^2 \left(\boldsymbol{u}_{h,i}^n - \dot{\boldsymbol{d}}_h^n, \boldsymbol{\sigma}_{\mathrm{f}}(\boldsymbol{v}_{h,i}, -q_{h,i})\boldsymbol{n}_i\right)_{\Sigma^{n-1}} + \frac{\gamma \mu_{\mathrm{f}}}{h} \sum_{i=1}^2 \left(\boldsymbol{u}_{h,i}^n - \dot{\boldsymbol{d}}_h^n, \boldsymbol{v}_{h,i} - \boldsymbol{w}_h\right)_{\Sigma^{n-1}} \\
&\quad - (\sigma_{\mathrm{p}}, \boldsymbol{v}_h \cdot \boldsymbol{n})_{\Sigma_{\mathrm{p}}} + (\epsilon_{\mathrm{p}} K_\tau \nabla_\tau P_{\mathrm{l},h}^n, \nabla_\tau q_{\mathrm{l},h})_{\Sigma_{\mathrm{p}}} - \left(\boldsymbol{u}_h^n \cdot \boldsymbol{n}, q_{\mathrm{l},h}\right)_{\Sigma_{\mathrm{p}}} \\
&\quad + \frac{\alpha}{\sqrt{K_\tau \epsilon_{\mathrm{p}}}} (\boldsymbol{u}_{h,\tau}^n, \boldsymbol{v}_{h,\tau})_{\Sigma_{\mathrm{p}}} = (\rho_{\mathrm{f}} \boldsymbol{f}_f, \boldsymbol{v}_h)_{\Omega^n}
\end{aligned} \tag{21}$$

for all $(\boldsymbol{v}_h, q_h, \boldsymbol{w}_h, q_{\mathrm{l},h}) \in \boldsymbol{V}_h^n \times Q_h^n \times \boldsymbol{W}_h \times \boldsymbol{S}_h$, where the porous stress σ_{p} is given by

$$\sigma_{\mathrm{p}} := -P_{\mathrm{l},h}^n - \frac{\epsilon_{\mathrm{p}} K_n^{-1}}{4} \boldsymbol{u}_h^n \cdot \boldsymbol{n} \quad \text{on} \quad \Sigma_{\mathrm{p}}.$$

We can now introduce the corresponding fluid discrete tri-linear form (see [30]):

$$\begin{aligned} a_h^{\mathrm{f},n}\big(z_h;(\boldsymbol{u}_h,p_h),(\boldsymbol{v}_h,q_h)\big) :=& 2\mu_{\mathrm{f}}\big(\boldsymbol{\epsilon}(\boldsymbol{u}_h),\boldsymbol{\epsilon}(\boldsymbol{v}_h)\big)_{\Omega^{\mathrm{f},n}} + \rho_{\mathrm{f}}\big(z_h\cdot\nabla\boldsymbol{u}_h,\boldsymbol{v}_h\big)_{\Omega^{\mathrm{f},n}} + \frac{\rho_{\mathrm{f}}}{2}\big((\mathrm{div} z_h)\boldsymbol{u}_h,\boldsymbol{v}_h\big)_{\Omega^{\mathrm{f},n}} \\ &- \rho_{\mathrm{f}}\big(\{\!\{z_h\}\!\}\cdot\boldsymbol{n}[\![\boldsymbol{u}_h]\!],\{\!\{\boldsymbol{v}_h\}\!\}\big)_{\Sigma^{n-1}} - \frac{\rho_{\mathrm{f}}}{2}\big([\![z_h\cdot\boldsymbol{n}]\!],\{\!\{\boldsymbol{u}_h\cdot\boldsymbol{v}_h\}\!\}\big)_{\Sigma^{n-1}} \\ &- (p_h,\mathrm{div}\,\boldsymbol{v}_h)_{\Omega^{\mathrm{f},n}} + (\mathrm{div}\,\boldsymbol{u}_h,q_h)_{\Omega^{\mathrm{f},n}} \\ &+ s_{\mathrm{v},h}^n(z_h;\boldsymbol{u}_h,\boldsymbol{v}_h) + s_{\mathrm{p},h}^n(z_h;p_h,q_h) + g_h^n(\boldsymbol{u}_h,\boldsymbol{v}_h). \end{aligned}$$

For consistency the bulk terms are integrated in the physical domain $\Omega^{\mathrm{f},n}$, which requires a specific track of the interface intersections within the fluid domain (see e.g. [39–41]). The terms $s_{\mathrm{v},h}^n$ and $s_{\mathrm{p},h}^n$ respectively correspond to the continuous interior penalty velocity and pressure stabilisation operators, given by (see, e.g., [42]):

$$s_{\mathrm{v},h}^n(z_h;\boldsymbol{u}_h,\boldsymbol{v}_h) := \gamma_{\mathrm{v}}h^2\sum_{i=1}^{2}\sum_{F\in\mathcal{F}_{h,i}^n}\xi\big(\mathrm{Re}_F(z_h)\big)\|z_h\cdot\boldsymbol{n}\|_{L^\infty(F)}\big([\![\nabla\boldsymbol{u}_h]\!]_F,[\![\nabla\boldsymbol{v}_h]\!]_F\big)_F,$$

$$s_{\mathrm{p},h}^n(z_h;p_h,q_h) := \gamma_{\mathrm{p}}h^2\sum_{i=1}^{2}\sum_{F\in\mathcal{F}_{h,i}^n}\frac{\xi\big(\mathrm{Re}_F(z_h)\big)}{\|z_h\|_{L^\infty(F)}}\big([\![\nabla p_h]\!]_F,[\![\nabla q_h]\!]_F\big)_F,$$

where $\mathcal{F}_{h,i}^n$ denotes the set of interior edges or faces of $\mathcal{T}_{h,i}^n$, $\mathrm{Re}_F(z_h) := \rho_{\mathrm{f}}\|z_h\|_{L^\infty(F)}h\mu_{\mathrm{f}}^{-1}$ denotes the local Reynolds number, $\xi(x) := \min\{1,x\}$ is a cut-off function and γ_{p}, $\gamma_{\mathrm{v}} > 0$ are user-defined parameters. Finally, g_h^n is the so-called ghost-penalty operator, given by

$$g_h^n(\boldsymbol{u}_h,\boldsymbol{v}_h) := \gamma_{\mathrm{g}}\mu_{\mathrm{f}}h\sum_{i=1}^{2}\sum_{F\in\mathcal{F}_{i,h}^{\Sigma^{n-1}}}\big([\![\nabla\boldsymbol{u}_{i,h}]\!]_F,[\![\nabla\boldsymbol{v}_{i,h}]\!]_F\big)_F,$$

where $\mathcal{F}_{i,h}^{\Sigma^{n-1}}$ denotes the set of interior edges or faces of the elements intersected by Σ^{n-1}. This term is added to guarantee robustness independent of the way the interface cuts the fluid mesh. The underlying idea is to extend the coercivity of the bi-linear form to the whole computational domain, see [43] or [44] for different possibilities.

For the approximation of the porous system, we consider a family of triangulation $\{\mathcal{T}_h^{\mathrm{p}}\}_{h>0}$ of Σ_{p}, so that each $\mathcal{T}_h^{\mathrm{p}}$ is fitted $\mathcal{T}_h$. We then consider the standard space of continuous piecewise affine functions for the approximation of the porous pressure P_{l}

$$\boldsymbol{S}_h := \big\{v_h\in C^0(\overline{\Sigma_p})\,\big|\,v_{h|K}\in\mathbb{P}_1(K),\quad\forall K\in\mathcal{T}_h^{\mathrm{p}}\big\}.$$

In summary, the resulting fully discrete method is reported in Algorithm 1. We use the notation $\boldsymbol{u}_{h,i}^n$, $\boldsymbol{v}_{h,i}^n$, $p_{h,i}^n$, $q_{h,i}^n$ ($i=1,2$) introduced in (1) for the two parts of the discontinuous functions across Σ^{n-1}. Note that the kinematic-dynamic interface coupling $(14)_2$–(15) is enforced in a consistent and strongly coupled fashion through Nitsche's method (see [29,30]).

Remark 3.1. If the interface has a boundary inside the fluid domain (a so-called tip), we consider the construction of the fluid and solid discrete spaces proposed in [30] (see [45, Chapter 6] for an extension to the 3D case). A virtual interface $\widetilde{\Sigma}^{n-1}$ is introduced by connecting the interface tip with the fluid vertex opposite to the edge intersected by the interface and therefore the fluid domain is closed. Afterwards, we enforce the kinematic/dynamic continuity of the fluid on $\widetilde{\Sigma}^{n-1}$ in a discontinuous Galerkin fashion (see, e.g., [46]). More precisely, the following terms are added to (21)

$$-\big(\{\!\{\boldsymbol{\sigma}_{\mathrm{f}}(\boldsymbol{u}_h^n,p_h^n)\}\!\}\boldsymbol{n},[\![\boldsymbol{v}_h]\!]\big)_{\widetilde{\Sigma}^{n-1}} - \big(\{\!\{\boldsymbol{\sigma}_{\mathrm{f}}(\boldsymbol{v}_h,-q_h)\}\!\}\boldsymbol{n},[\![\boldsymbol{u}_h^n]\!]\big)_{\widetilde{\Sigma}^{n-1}} + \frac{\gamma\mu_{\mathrm{f}}}{h}\big([\![\boldsymbol{u}_h^n]\!],[\![\boldsymbol{v}_h]\!]\big)_{\widetilde{\Sigma}^{n-1}}.$$

3.2. Fully Eulerian formalism with immersed thick-walled solids

In a fully Eulerian approach the current displacement $\boldsymbol{d}(\boldsymbol{x},t)$ is defined by the relation

$$\boldsymbol{x}-\boldsymbol{d}(\boldsymbol{x},t) = \hat{\boldsymbol{x}}, \tag{22}$$

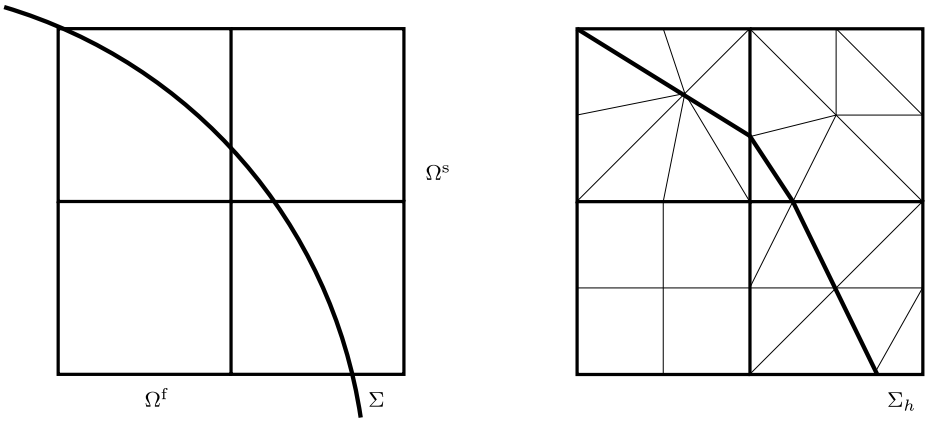


Fig. 5. Example of a locally fitted finite element discretisation with 4 coarse cells.

where $\hat{\boldsymbol{x}}$ is the corresponding point in the reference configuration Ω^{s}. This means that the displacement can be used to trace back points $\boldsymbol{x} \in \Omega^{\mathrm{s}}(t)$ to their reference position $\hat{\boldsymbol{x}}$ in Ω^{s} and hence to determine the domain affiliation of a point $\boldsymbol{x} \in \Omega$ at time t. As in the previous section, we use again an explicit approach to avoid the issues related to geometrical non-linearities

$$\Omega^{\mathrm{s},n} := \left\{\boldsymbol{x} \in \Omega,\ \left|\ \boldsymbol{x} - \boldsymbol{d}_h^{n-1} \in \Omega^{\mathrm{s}}\right.\right\}, \quad \Sigma^n := \left\{\boldsymbol{x} \in \Omega,\ \left|\ \boldsymbol{x} - \boldsymbol{d}_h^{n-1} \in \Sigma\right.\right\}, \quad \Omega^{\mathrm{f},n} := \Omega \setminus \left(\Omega^{\mathrm{s},n} \cup \Sigma^n\right).$$

Numerically, the domain affiliations can be determined by the Initial Point Set (backward characteristics) method (see, e.g., [11,25,26]). To evaluate the displacement $\boldsymbol{d}_h^{n-1}$ in points $\boldsymbol{x} \in \Omega^{\mathrm{f},n-1}$ near the interface Σ^{n-1}, that could belong to $\Omega^{\mathrm{s},n}$ in the next step, an extension of the solid displacement to a small layer around the interface is required (see, e.g., [27]). The domain affiliation can be computed in a separate step before setting up the variational formulation as shown in Algorithm 2, or "on-the-fly" while setting up the finite element formulation.

As an alternative to the unfitted finite element method presented in the previous section, we consider here a fitted finite element method for spatial discretisation. We briefly describe the locally modified finite element method as an example in two space dimensions (see [31]). The method is based on a fixed coarse triangulation $\mathcal{T}_{2h}$, which is independent of the interface position, and a further subtriangulation of the coarse elements $P \in \mathcal{T}_{2h}$, which resolves the interface, see Fig. 5. We restrict ourselves to linear finite elements and a linear interface approximation. A second-order approximation has been presented recently in [47].

In order to resolve the interface locally, we split each coarse cell cut by the interface into 8 subtriangles $K_1, \ldots, K_8$ and move some of the interior vertices to the interface, such that a linear interface approximation is obtained. The position of the 9 degrees of freedom $x_1, \ldots, x_9$ in each coarse cell are described by a piecewise linear reference map from the reference patch $\hat{P} = [0,1]^2$

$$\xi_P : \hat{P} \mapsto P, \qquad \xi_P \in Q_P := \left\{\phi_h \in C^0(\hat{P}) \,\middle|\, \xi|_{K_i} \in P_1(K_i),\ i = 1, \ldots, 8\right\}$$

that fulfils the 9 conditions $\xi_P(\hat{x}_i) = x_i$, $i = 1, \ldots, 9$, where $\hat{x}_i$ denotes the (fixed) Lagrangian points on the reference patch. In elements that are not affected by the interface piecewise bilinear shape functions can be used on four quadrilaterals alternatively, see the lower left coarse cell in Fig. 5. The locally modified finite element space is then given by

$$X_h^{\mathrm{lmfem},n}(\Omega) := \left\{\phi_h \in C^0(\overline{\Omega}),\ \left|\ (\phi_h \circ \xi^{-1})|_P \in Q_P\ \forall P \in \mathcal{T}_{2h}\right.\right\}.$$

By $X_h^{\mathrm{lmfem,n}}(\Omega^i)$ we denote the space that results by eliminating all degrees of freedom that do not lie in a subdomain $\Omega^i \subset \Omega$ or on its boundary. Then, we define the spaces

$$\boldsymbol{V}_h^n := \left[X_h^{\mathrm{lmfem},n}(\Omega^{\mathrm{f},n}) \cap H^1_\Gamma(\Omega^{\mathrm{f},n})\right]^2, \quad Q_h^n := X_h^{\mathrm{lmfem},n}(\Omega^{\mathrm{f},n}), \quad \boldsymbol{W}_h^n := \left[X_h^{\mathrm{lmfem},n}(\Omega^{\mathrm{s},n}) \cap H^1_\Gamma(\Omega^{\mathrm{s,n}})\right]^2.$$

The space S_h is defined as in Section 3.1. We use an implicit form of the fluid semi-linear form

$$\begin{aligned} a_h^{\mathrm{f},n}\left(\boldsymbol{u}_h, p_h; \boldsymbol{v}_h, q_h\right) := {} & 2\mu_{\mathrm{f}}\left(\boldsymbol{\epsilon}(\boldsymbol{u}_h), \boldsymbol{\epsilon}(\boldsymbol{v}_h)\right)_{\Omega^{\mathrm{f},n}} + \rho_{\mathrm{f}}\left(\boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h\right)_{\Omega^{\mathrm{f},n}} - (p_h, \operatorname{div} \boldsymbol{v}_h)_{\Omega^{\mathrm{f},n}} \\ & + (\operatorname{div} \boldsymbol{u}_h, q_h)_{\Omega^{\mathrm{f},n}} + s_{\mathrm{p},h}^n(z_h; p_h, q_h). \end{aligned}$$

Algorithm 2 Strongly coupled Eulerian approach for an FSI–contact problem with a thick-walled solid..

For $n \geq 1$:

1. Update the domain affiliations

$$\Omega^{\mathrm{s},n} := \left\{ \boldsymbol{x} \in \Omega,\ \left| \boldsymbol{x} - \boldsymbol{d}_h^{n-1} \in \Omega^{\mathrm{s}} \right. \right\}, \quad \Sigma^n := \left\{ x \in \Omega,\ \left| \boldsymbol{x} - \boldsymbol{d}_h^{n-1} \in \Sigma \right. \right\}, \quad \Omega^{\mathrm{f},n} := \Omega \setminus \left(\Omega^{\mathrm{s},n} \cup \Sigma^n \right).$$

2. Find $\left(\boldsymbol{u}_h^n, p_h^n, \dot{\boldsymbol{d}}_h^n, \boldsymbol{d}_h^n, P_{\mathrm{l},h}^n\right) \in \boldsymbol{V}_h^n \times Q_h^n \times \boldsymbol{W}_h^n \times \boldsymbol{W}_h^n \times \boldsymbol{S}_h$ with $\dot{\boldsymbol{d}}_h^n = \tilde{\partial}_{\delta t} \boldsymbol{d}_h^n$ and

$$\begin{aligned}
&\rho_{\mathrm{f}}\left(\tilde{\partial}_{\delta t} \boldsymbol{u}_h^n, \boldsymbol{v}_h\right)_{\Omega^{f,n}} + a_h^{\mathrm{f},n}\left(\boldsymbol{u}_h^n, p_h^n; \boldsymbol{v}_h, q_h\right) + \rho_{\mathrm{s}}\left(\tilde{\partial}_{\delta t} \dot{\boldsymbol{d}}_h^n, \boldsymbol{w}_h\right)_{\Omega^{\mathrm{s,n}}} + a^{\mathrm{s,n}}(\boldsymbol{d}_h^n, \boldsymbol{w}_h) \\
&\qquad + \frac{\gamma_{\mathrm{c}} E}{h}\left(\left[\tilde{P}_{\gamma_c}(\boldsymbol{d}_h^n)\right]_+, \boldsymbol{w}_h \cdot \boldsymbol{n}\right)_{\Sigma^{\mathrm{n}}} - \left(\boldsymbol{\sigma}_{\mathrm{f}}(\boldsymbol{u}_h^n, p_h^n)\boldsymbol{n}, \boldsymbol{v}_h - \boldsymbol{w}_h\right)_{\Sigma^n} \\
&\qquad - \left(\boldsymbol{u}_h^n - \dot{\boldsymbol{d}}_h^n, \boldsymbol{\sigma}_{\mathrm{f}}(\boldsymbol{v}_h, -q_h))\boldsymbol{n}\right)_{\Sigma^n} + \frac{\gamma \mu_{\mathrm{f}}}{h}\left(\boldsymbol{u}_h^n - \dot{\boldsymbol{d}}_h^n, \boldsymbol{v}_h - \boldsymbol{w}_h\right)_{\Sigma^n} \\
&\qquad - (\sigma_p, \boldsymbol{v}_h \cdot n)_{\Sigma_{\mathrm{p}}} + (\epsilon_{\mathrm{p}} K_\tau \nabla_\tau P_{\mathrm{l},h}^n, \nabla_\tau q_{\mathrm{l},h})_{\Sigma_p} - \left(\boldsymbol{u}_{h,n}^n, q_{\mathrm{l},h}\right)_{\Sigma_{\mathrm{p}}} \\
&\qquad + \frac{\alpha}{\sqrt{K_\tau \epsilon_{\mathrm{p}}}}(\boldsymbol{u}_{h,\tau}^n, \boldsymbol{v}_{h,\tau})_{\Sigma_{\mathrm{p}}} = (\rho_{\mathrm{f}} f_f, \boldsymbol{v}_h)_{\Omega^n} + (\rho_{\mathrm{s}} f_s, \boldsymbol{w}_h)_{\Omega^{s,n}} \qquad (23)
\end{aligned}$$

for all $(\boldsymbol{v}_h, q_h, \boldsymbol{w}_h, q_{\mathrm{l},h}) \in \boldsymbol{V}_h^n \times Q_h^n \times \boldsymbol{W}_h^n \times \boldsymbol{S}_h$, with the porous stress σ_{p} as defined in Algorithm 1 and the contact term $\tilde{P}_{\gamma_c}$ defined in (24).

To cope with the lack of inf–sup-stability of the discrete spaces, we use the (anisotropic) CIP pressure stabilisation developed in [48] for $s_{\mathrm{p},h}^n$. In contrast to 3.1, we have omitted the convection stabilisation, which will not be needed in the examples with a thick solid below. Moreover, a ghost-penalty term is not required, as we use a fitted finite element discretisation.

For the solid, we assume a hyperelastic material law with a corresponding variational formulation of the form

$$a^{\mathrm{s},n}(\boldsymbol{d}_h, \boldsymbol{w}_h) := (\boldsymbol{\sigma}_{\mathrm{s}}(\boldsymbol{d}_h), \nabla \boldsymbol{w}_h)_{\Omega^{\mathrm{s},n}},$$

where $\boldsymbol{\sigma}_{\mathrm{s}}$ denotes the Cauchy stress tensor.

Concerning time discretisation let us first note that the variables $\boldsymbol{u}_h^{n-1}$ and $\boldsymbol{d}_h^{n-1}$ are undefined on parts of $\Omega^{\mathrm{s,n}}$ and $\Omega^{\mathrm{f},n}$, respectively, as both Ω^{s} and Ω^{f} are time-dependent. Thus, the method of lines cannot be applied in a straight-forward way. To deal with this issue, we use the dG(0) variant of a family of Galerkin schemes that incorporates the characteristics of the domain movement in the Galerkin spaces [49]. For the dG(0) variant the difference to a standard backward Euler scheme lies solely in the discretisation of the time derivative. We introduce the notation

$$\tilde{\partial}_{\delta t} \boldsymbol{u}_h^n = \frac{\boldsymbol{u}_h^n - \left(\boldsymbol{u}_h^{n-1} \circ \Psi\right)}{\delta t} - \partial_t \Psi \cdot \nabla \boldsymbol{u}_h^n,$$

where Ψ is an (arbitrary) map defined in Ω that maps $\Omega^{\mathrm{f},n}$ to $\Omega^{\mathrm{f},n-1}$ and $\Omega^{\mathrm{s},n}$ to $\Omega^{\mathrm{s},n-1}$, respectively. Alternatively, one could use Eulerian time-stepping schemes with suitable extension operators. An implicit extension by means of ghost-penalties has been studied in [44,50].

Finally, let us note that due to the different meaning of the displacement $\boldsymbol{d}_h^n$ in the current frame $\Omega^{\mathrm{s},n}$ (see (22)), the no-penetration condition in (13) becomes

$$\boldsymbol{d}_h^n - \left(\boldsymbol{d}_h^{n-1} \circ \Psi\right) - \tilde{g}_\epsilon^n \leq 0$$

where $\tilde{g}_\epsilon^n$ denotes the current distance to the lower wall Σ_{p} minus ϵ and Ψ is a map from Σ^{n} to $\Sigma^{\mathrm{n}-1}$. The contact term takes the form

$$\tilde{P}_{\gamma_C}(\boldsymbol{d}_h^n) = \boldsymbol{d}_h^n - \left(\boldsymbol{d}_h^{n-1} \circ \Psi\right) - \tilde{g}_\epsilon^n - \gamma_C^{-1} [\![\sigma_n]\!], \qquad (24)$$

where $\gamma_C = \frac{\gamma_c E}{h}$, E denotes the elasticity modulus of the solid and γ_c is a (dimensionless) contact parameter. We note that in contrast to the $\mathcal{O}(h^{-2})$-weighting in the thin case, a weighting of $\mathcal{O}(h^{-1})$ is needed here. [51] The resulting numerical method is reported in Algorithm 2.

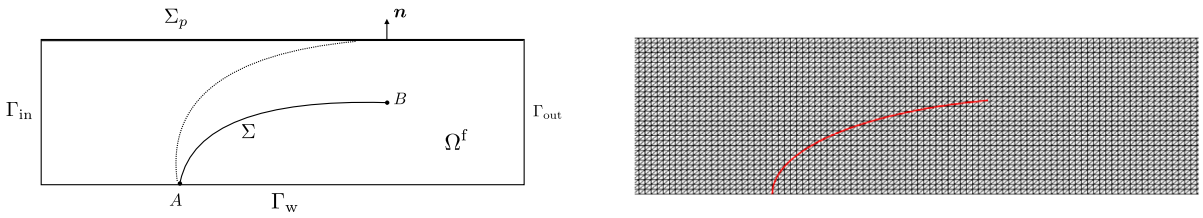


Fig. 6. (a) Geometric configuration of the idealised valve with contact, (b) Zoom of the leaflet mesh and fluid mesh.

Remark 3.2 (*Stability*)**.** In [19] a stability result has been derived for a very similar variational formulation with slip- or no-slip boundary conditions on Σ_p instead of the porous medium. An analogous result can easily be shown for both the variational formulations in (21) and (23) using the same technique.

Remark 3.3 (*Limit of Vanishing Fluid*)**.** In the limit of vanishing fluid density ρ_{f} and viscosity ν_{f} we recover the classical contact conditions in pure solid mechanics. This can be seen as follows: For $\rho_{\text{f}} \to 0$ the pressure p tends to zero, due to (2)–(3). For $\nu_{\text{f}} \to 0$ the viscous part of the fluid stresses $\boldsymbol{\sigma}_{\text{f}}$ vanishes as well. This means that in the limit we have $\lambda = \sigma_{s,n}$ in the case of a thick solid resp. $\lambda = T$ for a thin solid. The complementarity conditions (13) become the classical contact conditions with contact force (in the case of a thick solid)

$$\sigma_{s,n} = -\gamma_C \big[\boldsymbol{d} \cdot \boldsymbol{n} - g_\epsilon - \gamma_C^{-1} \sigma_{s,n}\big]_+.$$

4. Numerical experiments

In this section, we present two very different numerical examples to investigate the properties and the capabilities of the numerical approaches. First, we investigate the full fluid–structure–porous–contact problem for a thin-walled solid by means of a deflected thin elastic valve in Section 4.1. As introduced in Section 3.1, we use an unfitted discretisation and solve the solid equations on the reference domain $\Omega^{\text{s}} = \Sigma$. Then, we investigate the case of a thick solid in Section 4.2, namely an elastic ball that falls down and bounces within a viscous fluid. Here, a Fully Eulerian approach is used in combination with the locally fitted finite element method, as described in Section 3.2.

4.1. Idealised valve with contact

In this test, we consider a full FSI–contact problem with a thin-walled solid. This numerical example corresponds to the idealised valve test with possible contact on the porous layer Σ_p. The geometry is shown in Fig. 6(a). The computational domain is a rectangle $\Omega = [0, 8] \times [0, 0.805]$, where the upper boundary is a symmetry axis (we imagine a second symmetrical valve on top), which means that we impose the "slip" condition $\sigma_{f,\tau} = 0$ in (10), letting $\alpha = 0$. As reference configuration for the solid, Σ, we consider a curve segment with end points $A = (4, 0)$ and $B = (5.112, 0.483)$, parametrised by the analytical function

$$y(x) = \frac{1}{2}\sqrt{1 - \frac{(x - 11/2)^2}{(3/2)^2}}, \quad x \in [4, 5.112].$$

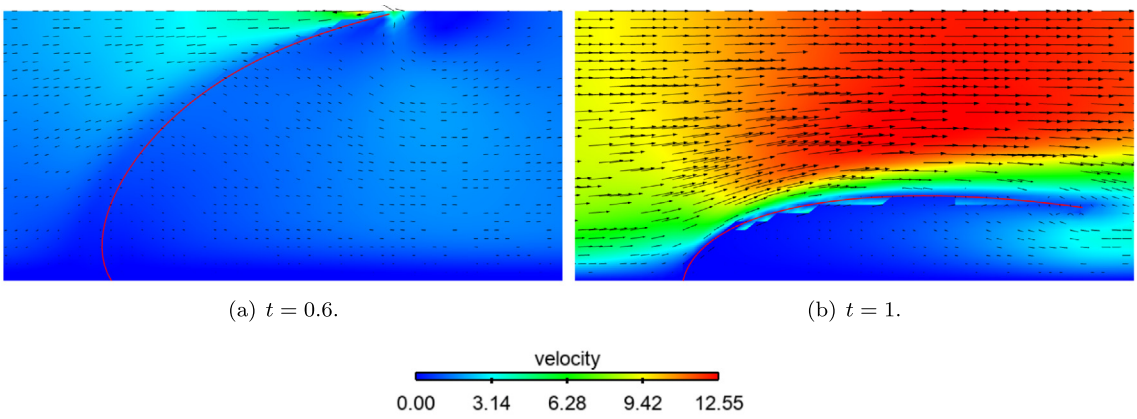
Regarding the boundary condition, a no-slip condition is enforced on the lower boundary Γ_w, zero traction on the outflow boundary Γ_{out} and a traction condition on Γ_{in}, in terms of the following time-dependent pressure:

$$p_{\text{in}}(t) = \begin{cases} -200 \tanh(100t) & \text{if} \quad 0 < t < 0.7, \\ 200 & \text{if} \quad t \geq 0.7. \end{cases}$$

The final time is $T = 1$, which corresponds to one full valve oscillation cycle. The fluid and the solid are initially at rest and the beam is pinched at the bottom tip A. In this test, the solid is described by a non-linear Reissner–Mindlin curved beam model with a MITC spatial discretisation. The physical parameters are given in Table 1 on the left and the numerical parameters on the right of the same table. Some of them will be varied in the following, in order

Table 1
Summary of the physical (left) and numerical (right) parameters used in Section 4.1.

Parameter	Symbol	Value (CGS)
Fluid density	ρ_{f}	1
Dynamic viscosity	μ_{f}	$3 \cdot 10^{-2}$
Solid density	ρ_{s}	1.2
Thickness (solid)	ϵ_{s}	$6.5 \cdot 10^{-2}$
Young's modulus	E	10^{7}
Poisson ration	ν_{f}	0.4
Conductivity (porous)	$K = K_n = K_\tau$	10^{-3}
Thickness (porous)	ϵ_{p}	10^{-2}
Parameter	Symbol	Value (CGS)
Ghost penalty	γ_p	1
CIP pressure stab.	γ_p	10^{-2}
CIP convection stab.	γ_v	10^{-2}
Nitsche par. (FSI)	γ	10^{2}
Contact parameter	γ_c	$5 \cdot 10^{-3}$
Relaxation param.	ϵ_g	$10^{-2} \approx \frac{h}{4}$
Time step	δt	10^{-3}
Mesh size	h	$\approx 10^{-4}$



(a) $t = 0.6$. (b) $t = 1$.

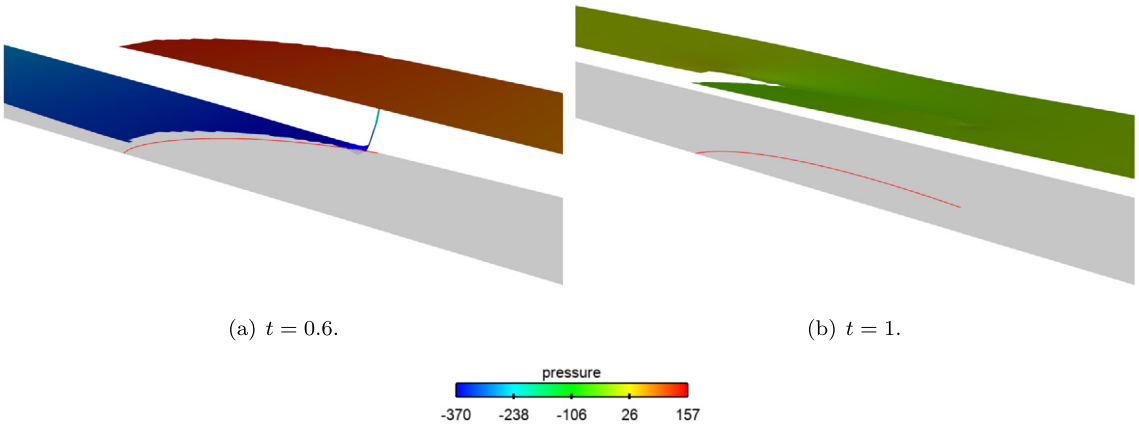
Fig. 7. Velocity magnitude snapshots.

to study the sensitivity of the approach with respect to certain parameters. The contact parameter $\gamma_{\mathrm{c}} = 5 \cdot 10^{-3}$ is chosen large enough to avoid penetration and small enough to maintain good convergence properties in the Newton solver (a very large parameter would lead to a stiffening of the solid system).

The fluid mesh has 16 384 triangles and the solid mesh 50 edges. We have $h \approx 0.04$. The zoom on both meshes is presented in Fig. 6(b). The time discretisation parameter is $\delta t = 10^{-3}$ and the Nitsche parameter is set to $\gamma = 100$.

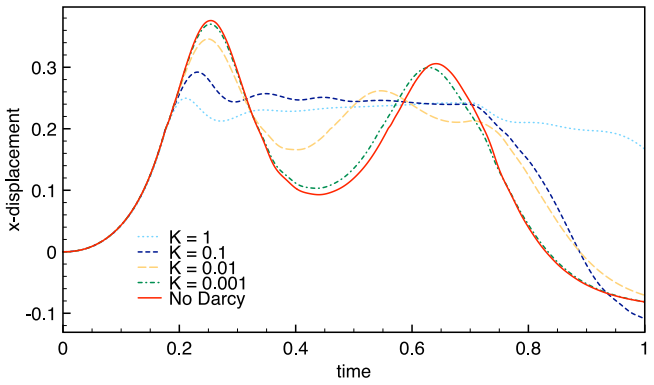
Let us first consider a test case with $K = K_\tau = K_n = 10^{-3}$. We report in Fig. 7 the velocity magnitude at two different instants. In Fig. 7(a) we report the approximation obtained at time $t = 0.6$. At this instant, the valve is in contact with the upper wall and the fluid velocity decreases globally as a consequence of the closing. The fluid enters into the porous layer and is transported from the right side to the left. At $t = 1$ the valve is open and far from Σ_p, therefore the fluid flow is reestablished and the velocity increases in the channel. The same comparison is performed in Fig. 8(a) and (b) for the pressure. We can see the high pressure jump when the valve is in contact with the wall (Fig. 8(a)), while at $t = 1$ the discontinuity between the two sides of the interface is weaker (Fig. 8(b)).

Next, we investigate the impact of the parameter $K = K_\tau = K_n$ on the contact dynamics. Fig. 9 presents the time history of the horizontal, Fig. 9(a), and vertical displacement, Fig. 9(b), at the upper solid point B for different

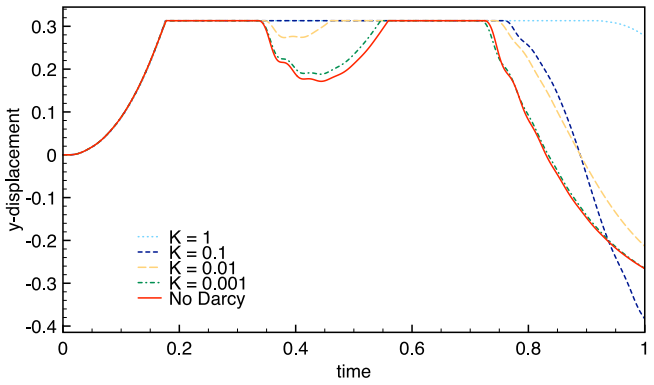


(a) $t = 0.6$. (b) $t = 1$.

Fig. 8. Pressure elevation snapshots.



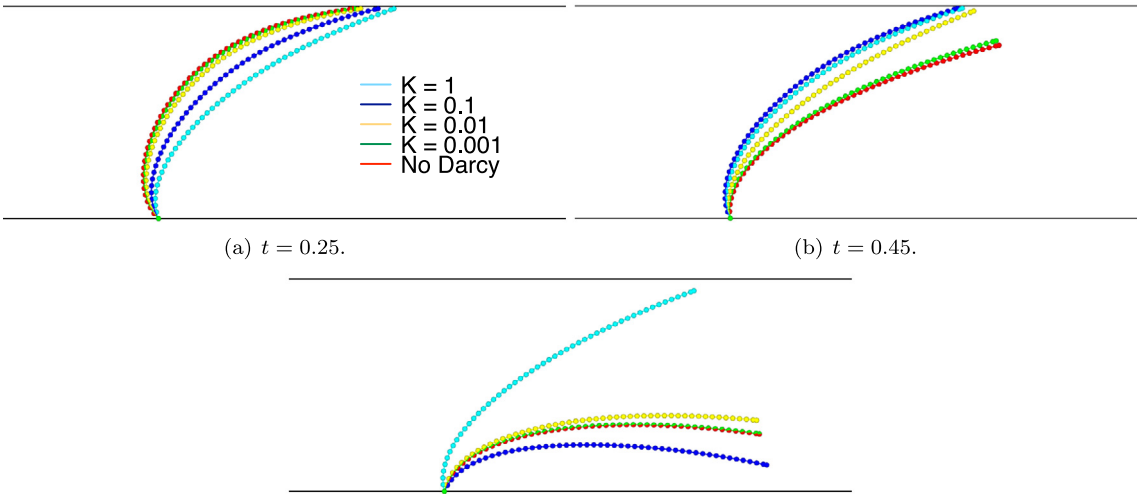
(a) x-displacement.



(b) y-displacement.

Fig. 9. Time evolution of the x and y-displacement for the structure endpoint B.

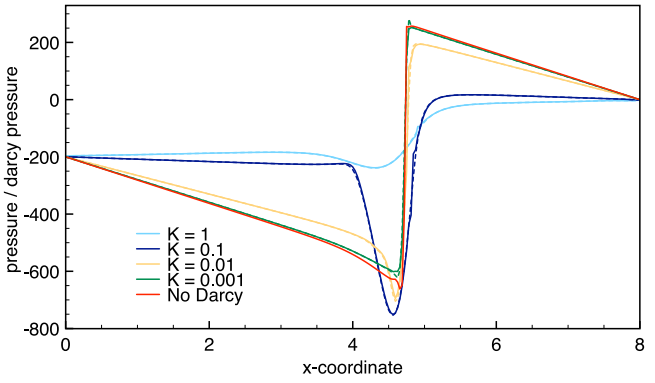
levels of conductivity. The non-penetration condition with the wall can be seen in Fig. 9(b), whereas Fig. 9(a) shows that the structure is sliding over the top wall. The interface is bouncing for all tests except the cases of $K_\tau = K_n = 1$ and 10^{-1}. In such cases, the structure reaches contact and the fluid flows abundantly into the porous layer, which prevents the release of contact. When the inlet pressure increases, the valve opens and the flow is restored. In all the other tests the interface is bouncing, but with a different reaction time, linked to the conductivity value. There



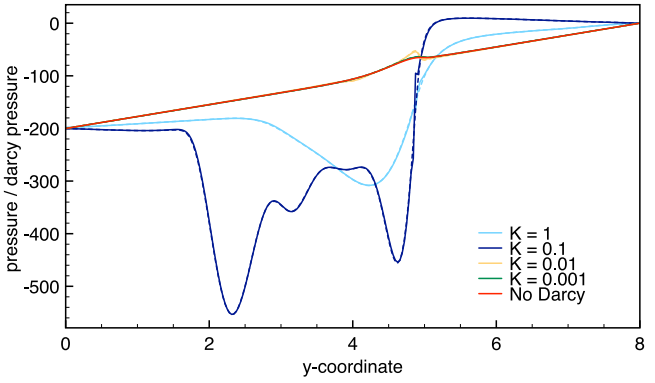
(a) $t = 0.25$. (b) $t = 0.45$.

(c) $t = 1$.

Fig. 10. Interfaces location at time $t = 0.25$ (a), $t = 0.45$ (b) and $t = 1$(c).

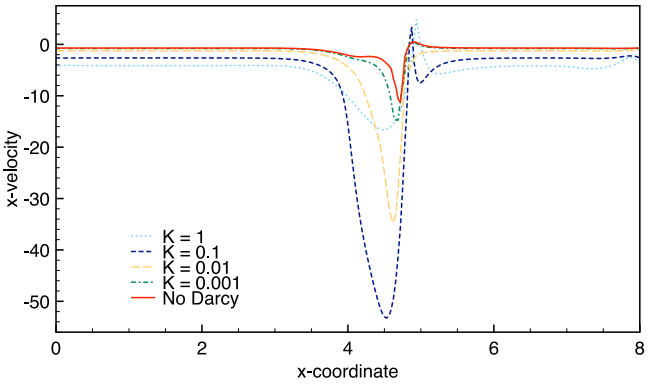


(a) $t = 0.25$.

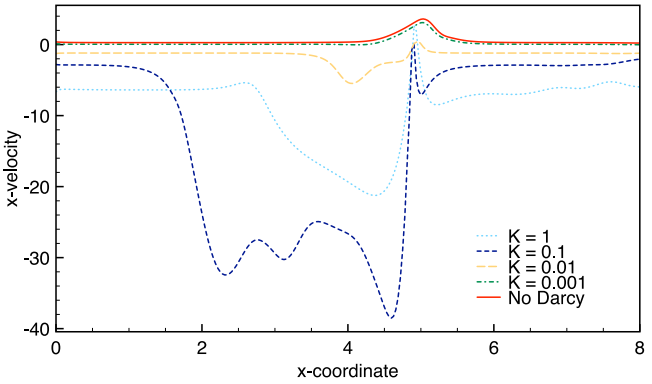


(b) $t = 0.45$.

Fig. 11. Fluid pressure (continuous line) and porous pressure (dash line) on Σ_p for different value of hydraulic conductivity at time $t = 0.25$ and $t = 0.45$.



(a) x-velocity at $t = 0.25$.



(b) x-velocity at $t = 0.45$.

Fig. 12. Fluid velocity on Σ_p.

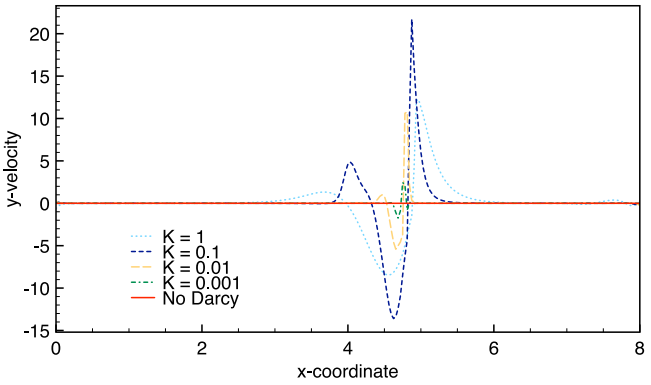
is a slight difference in the first release time, but the more visible differences are on the second bounce. Both, the second contact instant and the final release, are considerably sensitive to the changing in $K_\tau = K_n$. Finally, let notice that taking $K_\tau \to 0$ and $K_n \to 0$ we converge to the situation of no porous layer on Σ_p, as we can see in Fig. 9.

Similar observations can be inferred from Fig. 10, which shows the interface configuration at time $t = 0.25$ (during contact), $t = 0.45$ (after the first release) and $t = 1$ (when the flow is restored). We can see that for $K_\tau = K_n = 1$ and 10^{-1} the valve does not bounce, but it only releases once the inlet pressure increases (see Fig. 10(c)). Decreasing the conductivity of the porous medium increases the structure sliding at contact (see Fig. 10(a)) and the bouncing force applied on the structure (see Fig. 10(b)).

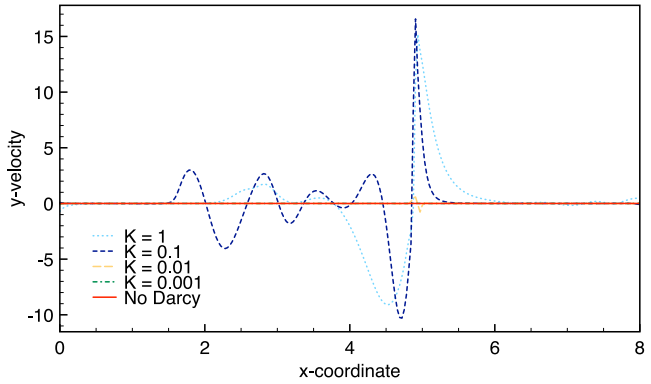
Fig. 11 displays the fluid pressure (continuous line) and the porous pressures (dashed line) at time $t = 0.25$ and $t = 0.45$. As expected, both pressures remain close. At time $t = 0.25$ the structure is in contact with the upper wall, therefore there is a high pressure gradient that decreases by increasing the conductivity.

Fig. 12 shows the fluid x-velocity along the porous layer Σ_p at two different instants. As we can see, the horizontal velocity is not zero also during contact as effect of the porous layer. As expected, the higher the conductivity the greater is the velocity magnitude and a larger area of the porous layer is leaking or pushing fluid inside the domain. In Fig. 13 we report the fluid y-velocity on Σ_p. The effect is more localised near the contact area except for cases of $K_\tau = K_n = 1, 0.1$, where the porous layer is still leaking and entering also far from the contact area.

We now explore the results when variations on the porous thickness ε_p are considered. The porous hydraulic conductivity is taken $K_\tau = K_n = 10^{-3}$. We explore results for $\varepsilon_p \in \{10^{-i}\}_{i=1}^{4}$. The outcome is shown in Fig. 14. For $\varepsilon_p \to 0$ the curves converge towards the results of no porous layer on the top wall.



(a) y-velocity at $t = 0.25$.



(b) y-velocity at $t = 0.45$.

Fig. 13. Fluid velocity on Σ_p.

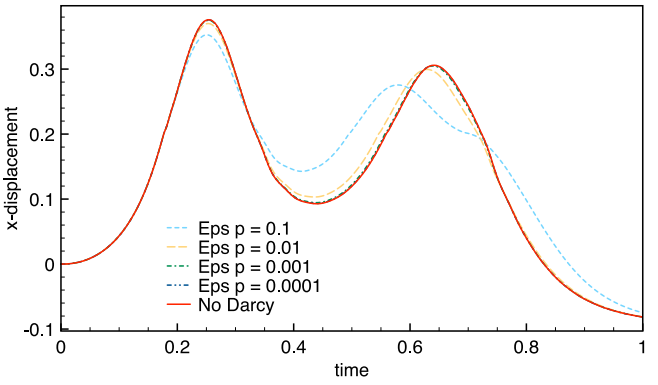
No particular differences are visible at first contact between the structure and the upper wall. During contact, the horizontal velocity is lower for higher values of ε_p, therefore, the bouncing force is also lower. In addition, the higher is ε_p, the later is the first release, the lower is the rebound force and, consequently, the earlier is the second contact and release. For illustration purposes, we report in Fig. 15 a zoom of the y-displacement between the first release and the second contact instants.

Refinement in space and time. We explore the convergence behaviour taking three levels of space and time refinement, namely $\left(\delta t, h\right) \in \left\{2 \cdot 10^{-3} \cdot 2^{-i}\,,\ 0.07 \cdot 2^{-i}\right\}_{i=0}^{2}$. The coarser fluid and solid meshes are made of 5 120 triangles and 26 segments, respectively. The second meshes consists of 20 480 triangles and 50 edges, while the finest one has 81 920 triangles and 102 segments. The porous conductivity is chosen $K_\tau = K_n = 10^{-3}$, and the contact relaxation parameter $\varepsilon_{\mathrm{g}} = \varepsilon_{\mathrm{g}}(h) \in \{0.02 \cdot 2^{-i}\}_{i=0}^{2}$.

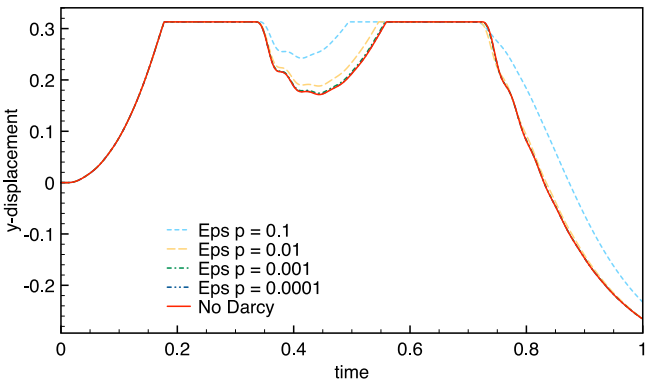
We show in Fig. 16 the results with these three refinement levels. We observe that the bouncing height is lower for the coarser mesh and that the intermediate level of refinement provides a reasonable approximation. We can also observe that, due to different contact relaxation parameters, contact and release occur at different instants and heights.

4.2. Falling and bouncing elastic ball

In this section, we consider the example of a falling and bouncing elastic ball in a cylinder, which is filled with a water–glycerin mixture. As we are interested in a detailed numerical study, we restrict ourselves to the two-dimensional case here and consider a box of size 6 cm × 7.5 cm. The ball has a radius of 1 cm and is kept initially at rest at a distance of 4 cm from the bottom. The ball falls down due to gravity $f_s = -9.81 \frac{\mathrm{m}}{\mathrm{s}^2}$ and bounces



(a) x-displacement.



(b) y-displacement.

Fig. 14. Time evolution of the x-displacement (a) and y-displacement (b) for the structure endpoint B for different values of ε_p.

back after the impact. Due to symmetry, we can reduce the computational domain to the right half by imposing symmetry boundary conditions on the midplane Γ^{sym}, see Fig. 17.

We use a Fully Eulerian approach to solve the coupled problem, as described in Section 3.2. For simplicity, we consider here a linear elastic material, where the Cauchy stress tensor $\boldsymbol{\sigma}_{\text{s}}$ is given by

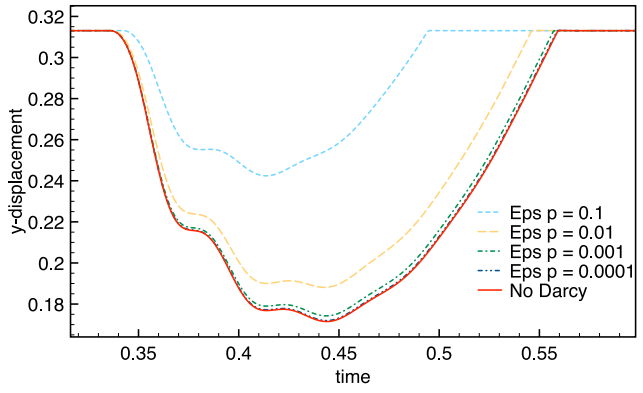
$$\sigma_s(d) = 2\mu_s E(d) + \lambda_s \text{tr}(E(d))I, \quad E(d) = \frac{1}{2}\left(\nabla d + \nabla d^T\right),$$

with Lamé parameters $\lambda_s = 7.64 \cdot 10^6 \frac{\text{kg}}{\text{ms}^2}$ and $\nu_s = 1.04 \cdot 10^6 \frac{\text{kg}}{\text{ms}^2}$.

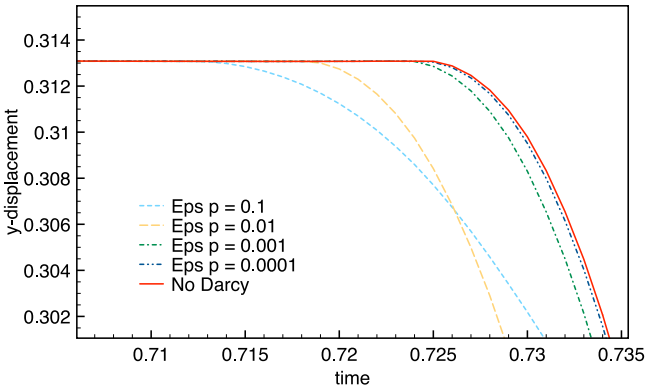
For time discretisation, we use a variant of the backward Euler method, namely a modified dG(0) time-stepping, see [12,49]. We start with a time-step size of $\delta t = 2 \cdot 10^{-3}$ s, which is reduced in a stepwise procedure up to the impact, where a small time-step size of $\delta t = 1.25 \cdot 10^{-4}$ s is reached.

The (kinematic) viscosity of the water–glycerol mixture is $\mu_{\text{f}} = 7 \cdot 10^{-6} \frac{\text{m}}{\text{s}^2}$, the density $\rho_f = 1141 \frac{\text{kg}}{\text{m}^3}$ and the solid density $\rho_s = 1351 \frac{\text{kg}}{\text{m}^3}$. Unless stated differently the parameters in the porous medium are chosen as $\epsilon_{\text{p}} = 10^{-4}$ and $K = K_n = K_\tau = 10^{-2}$ and the numerical contact parameters are $\gamma_c = 30\lambda_s$ and $\epsilon_g = \frac{h}{4}$. For the Navier–Stokes–Darcy coupling, we use the Beavers–Joseph–Saffman condition in (10) with $\alpha = 1$. All the results have been obtained with the finite element library Gascoigne3d [52]. We use a structured coarse grid $\mathcal{T}_{2h}$, which is highly refined close to Σ_p with 3201 vertices in total (unless specified differently). In Fig. 18, we illustrate the vertical velocities u_y and $\dot{d}_y$ of the falling ball at 6 instances of time.

Variation of the conductivity K. In Fig. 19 we compare the minimum distance to the ground during the fall and before and after the impact for different conductivities K with results obtained without a porous model ("No Darcy"), using either a full slip ($\sigma_{f,\tau} = 0$) or a Navier-slip condition $\sigma_{f,\tau} = \frac{\alpha}{\sqrt{K_\tau \epsilon_{\text{p}}}}$ with $\alpha = 1$ on Σ_p. Note that in the



(a) First release.



(b) Second release.

Fig. 15. Time evolution of the y-displacement for the structure endpoint B, between first release and second contact (a) and after second release (b).

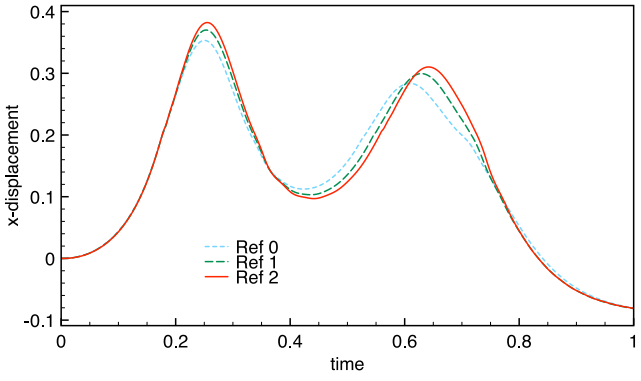
latter case this is exactly the same tangential condition which is imposed by the Beavers–Joseph–Saffman coupling. There, the normal velocity $\boldsymbol{u} \cdot \boldsymbol{n}$ is however not necessarily zero, as the flow can enter into the porous medium.

Depending on K, the ball bounces 4 or 5 times within the time interval $[0, 0.8\ \text{s}]$ with different bouncing heights (top right). The last bounces are barely visible in the graphs shown here due to a very small bouncing height. Moreover, we observe for $K \to 0$ that the results converge towards the results obtained with a Navier-slip-boundary conditions ("No Darcy"), the curve for $K = 10^{-4}$ showing no visible differences. For larger K the curves get slightly closer towards the results for a pure slip boundary condition.

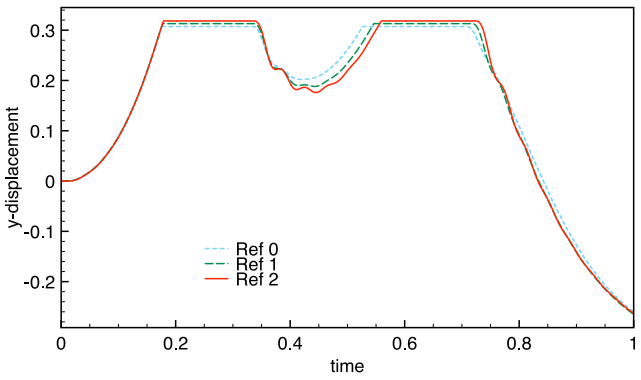
Concerning the time of impact (bottom left of Fig. 19), we observe that the impact happens slightly later, the smaller the conductivity K. The latest impact is observed for $K = 10^{-4}$ and the pure Navier-slip condition ("No Darcy"). This dependence on K is expected, as the resistive fluid forces that act against the contact are higher for smaller conductivities. The earliest impact is observed for the pure slip condition, followed by $K = 1$. Moreover, we observe that a small distance of about $4.3 \cdot 10^{-5}$ m, which lies slightly below the imposed gap distance of $\epsilon_g = 5 \cdot 10^{-5}$ m, remains in all cases.

From the upper right picture we can infer the bouncing height depending on K. It holds that the earlier the impact, the higher the impact velocity and hence, we observe a larger bouncing height. Consequently, we observe the largest bounce for pure slip-conditions, followed by $K = 1$, and the smallest one for pure no-slip conditions and $K = 10^{-4}$.

On the bottom right of Fig. 19 we show the corresponding vertical velocity $\dot{d}_y$ within the elastic ball, averaged in space. We see that the absolute value of the upwards velocity after the first impact is by more than 30% smaller than the absolute value of the impact velocity in all cases, which shows the dissipative impact of the fluid.



(a) x-displacement.



(b) y-displacement.

Fig. 16. Time evolution of the x-displacement (a) and y-displacement (b) for the structure endpoint B, with different levels of refinement.

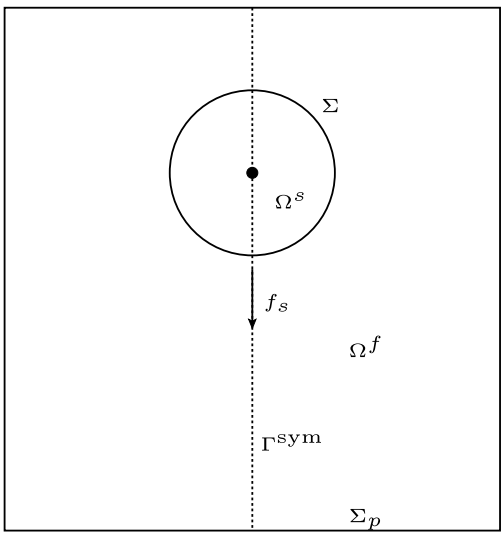


Fig. 17. Configuration of the numerical example with the falling ball. The computational domain is reduced to the right half due to symmetry on the axis Γ^{sym}.

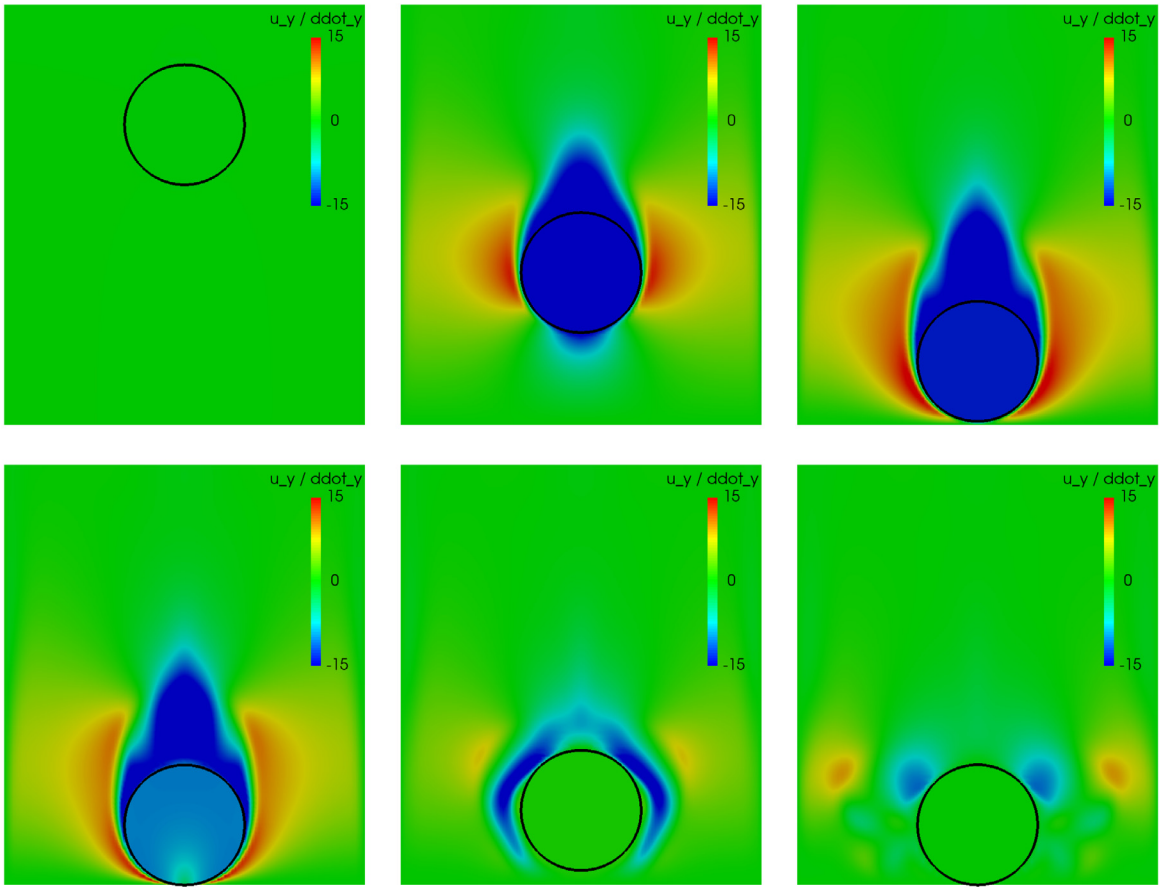


Fig. 18. Illustration of the vertical velocities in the falling ball example at $t = 0, 0.28, 0.378$ (top, left to right) and $t = 0.382, 0.47$ and 0.63 (bottom).

Variation of ϵ_p. In Fig. 20 we vary the thickness ϵ_p of the porous layer. We obtain 3 to 5 bounces with different heights depending on ϵ_p. For $\epsilon_p \to 0$, the curves converge towards the results for a pure Navier-slip condition ("No Darcy"), as the first equation in (10) implies $\boldsymbol{u} \cdot \boldsymbol{n} \to 0$. This is also what one expects from the physical model, as a smaller porous layer allows less fluid to diffuse through the layer. In the curve on the right of Fig. 20, we see that the contact happens earlier the larger ϵ_p is. This can again be explained by the smaller resistance of the fluid "against" the contact, when this is allowed to escape through the porous layer. The larger impact velocity for larger ϵ_p has again the effect that the bounce is higher for larger ϵ_p.

Variation of the contact parameter γ_c. In Fig. 21, we illustrate the influence of the contact parameter γ_c. As one would expect the violation of the relaxed contact condition is larger for a smaller γ_c, see the plot on the top right. For $\gamma_c \geq 10\lambda_s$ the curves are almost identical.

On the bottom, we show the contact force $\gamma_c[P_{\gamma_c}]_+$, which appears on the right-hand side of (18), for the first four bounces. We see that the values of the force are almost independent of the chosen contact parameter. While for the smallest contact parameter $\gamma_c = 0.1\lambda_s$ the contact times are slightly altered, there is (almost) no visible difference between the results for $\gamma_c = 10\lambda_s$ and $100\lambda_s$.

Time discretisation. In Fig. 22, we investigate the influence of the time-step size δt within and around the contact interval. In each simulation we start with a time-step of $\delta t = 2 \cdot 10^{-3}$ at $t = 0$, which is decreased successively by a factor of 2 depending on the distance to the ground. We see that a very small time-step is necessary to capture the contact dynamics. While for the largest time-step $\delta t_{\min} = 1.25 \cdot 10^{-4}$, the bounce is considerably reduced compared to the smaller time-step sizes, the curves seem to converge for $\delta t_{\min} \to 0$. The reason for the deviation can be

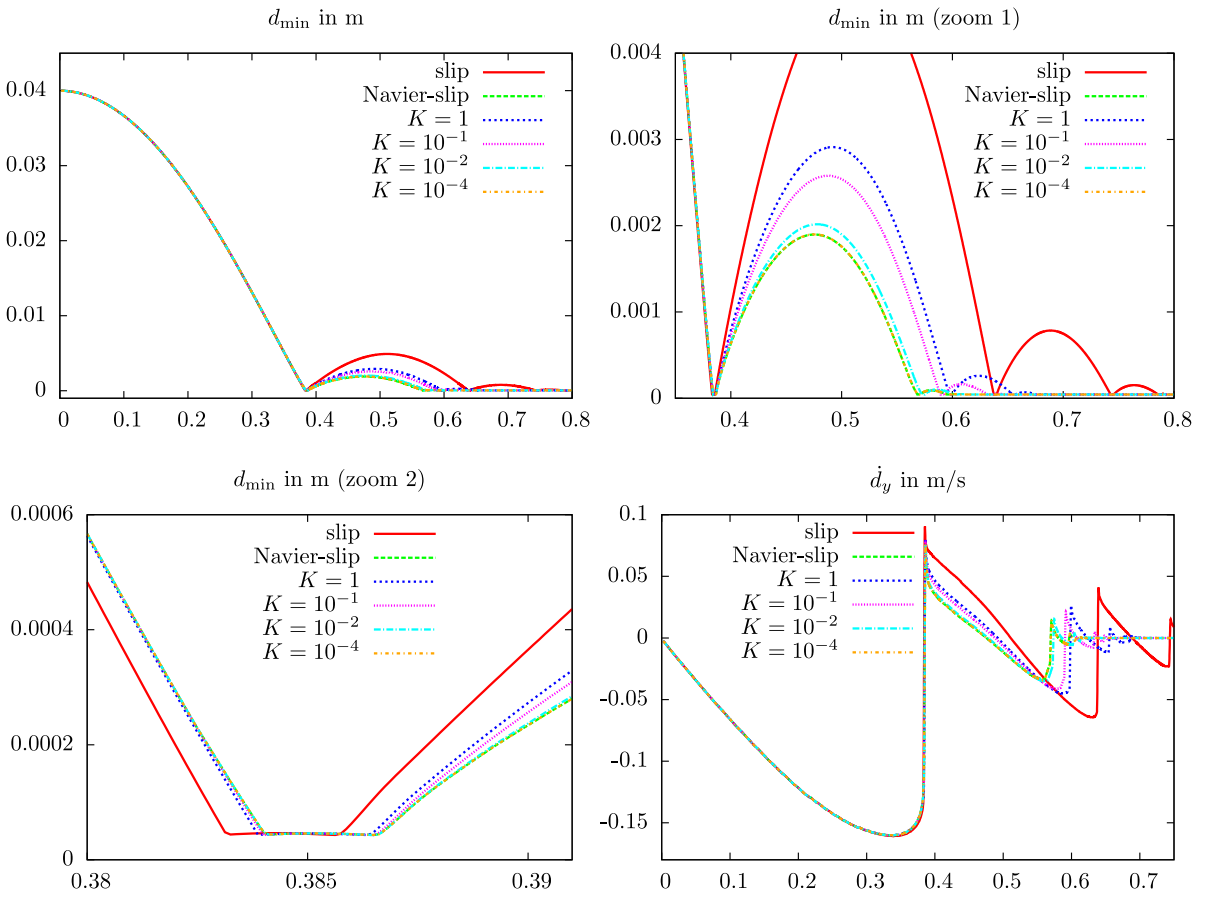


Fig. 19. Minimal distance $d_{\min}$ to the ground over time for different conductivities K and compared to pure slip and no-slip conditions on the lower boundary. After the curve over the full time interval $[0, 0.8\ \text{s}]$ on the top left, two different zooms are provided to illustrate the bounces (top right) and the time of impact (bottom left). *Bottom right*: Space-averaged vertical velocity of the elastic ball over time.

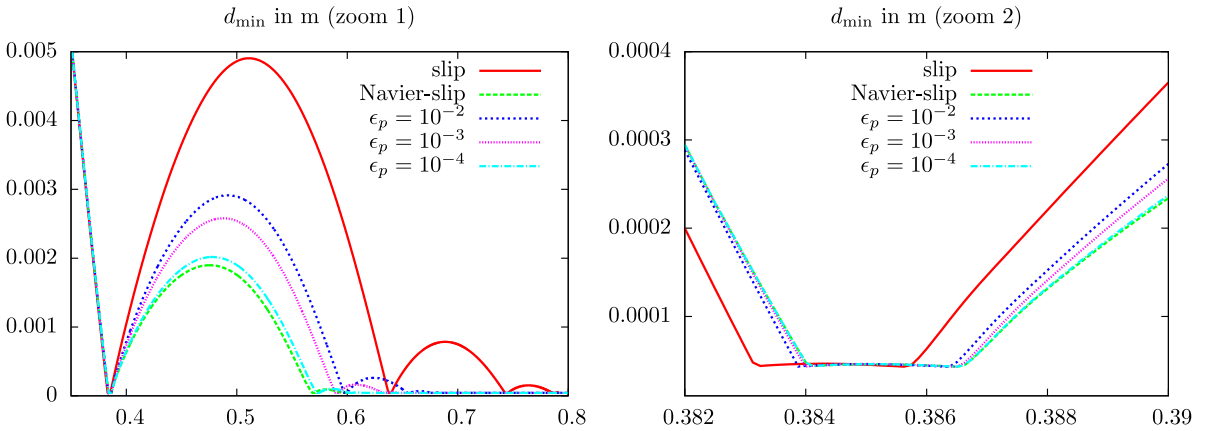


Fig. 20. Minimal distance $d_{\min}$ to the ground over time for different thicknesses ϵ_{p} of the porous layer and compared to pure slip and no-slip conditions on the lower boundary.

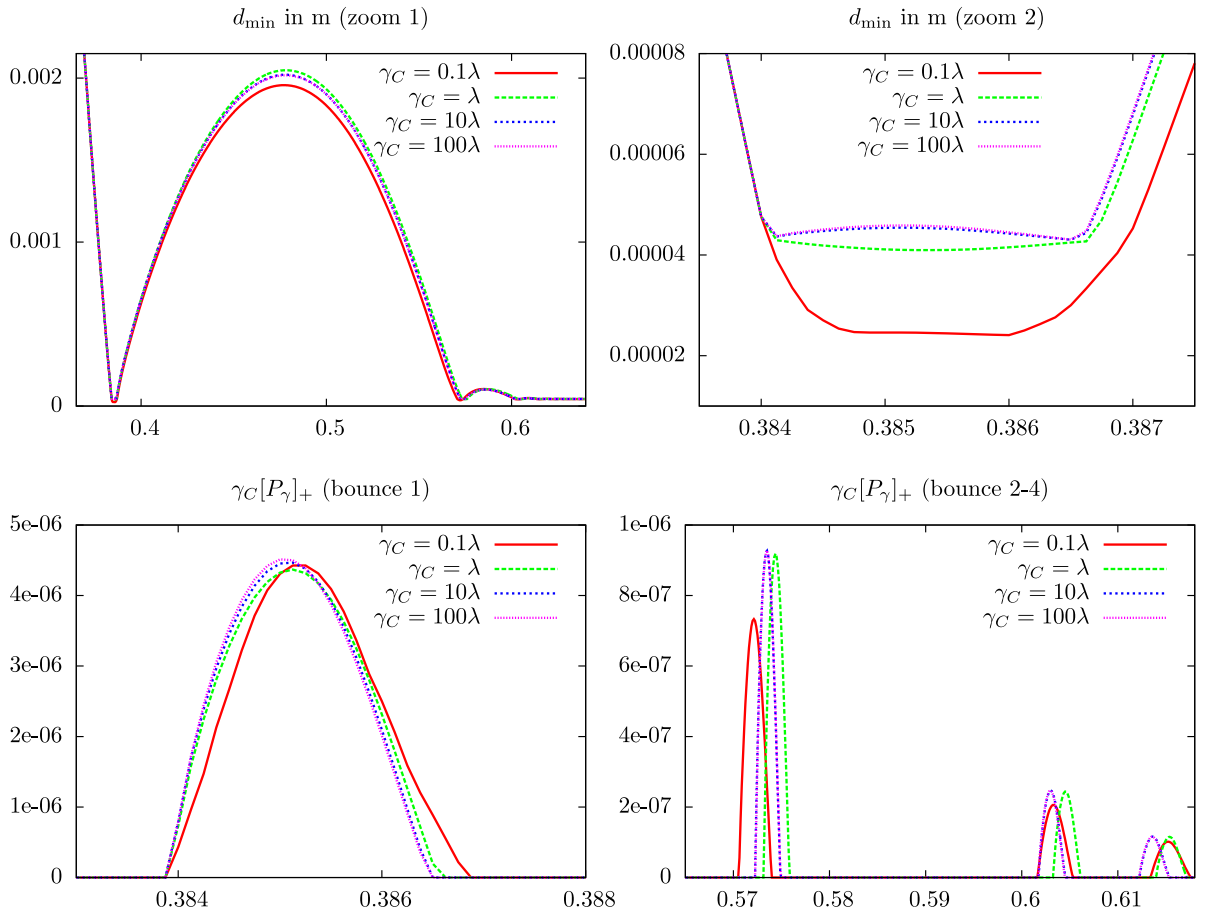


Fig. 21. *Top*: Minimal distance $d_{\min}$ to the ground over time for different contact parameters γ_c. *Bottom*: Contact force $\gamma_c[P_{\gamma_c}]_+$ over time for the first bounce (*left*) and the second to fourth bounce (*right*).

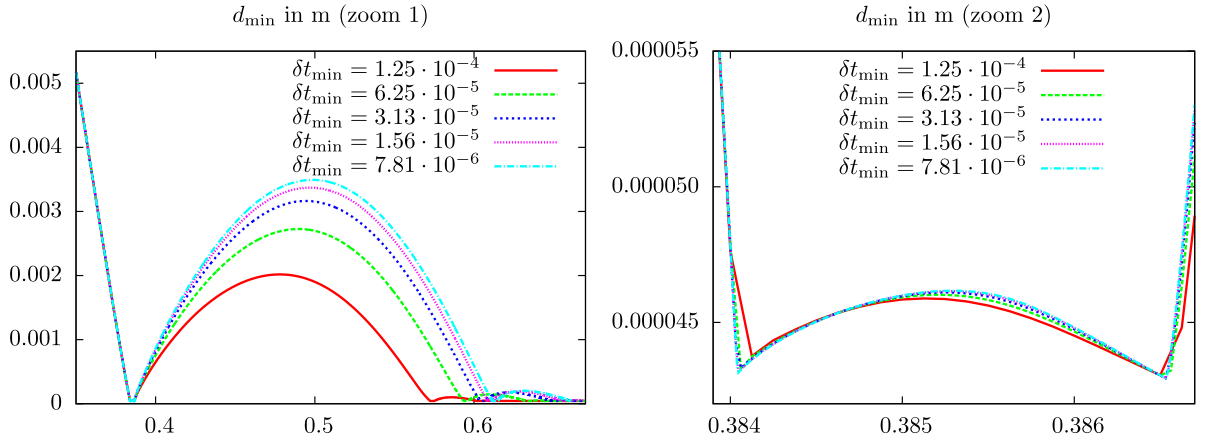


Fig. 22. Minimal distance $d_{\min}$ to the ground during the first two bounces (left) and at the first impact period (right) for different minimal time-step sizes $\delta t_{\min}$.

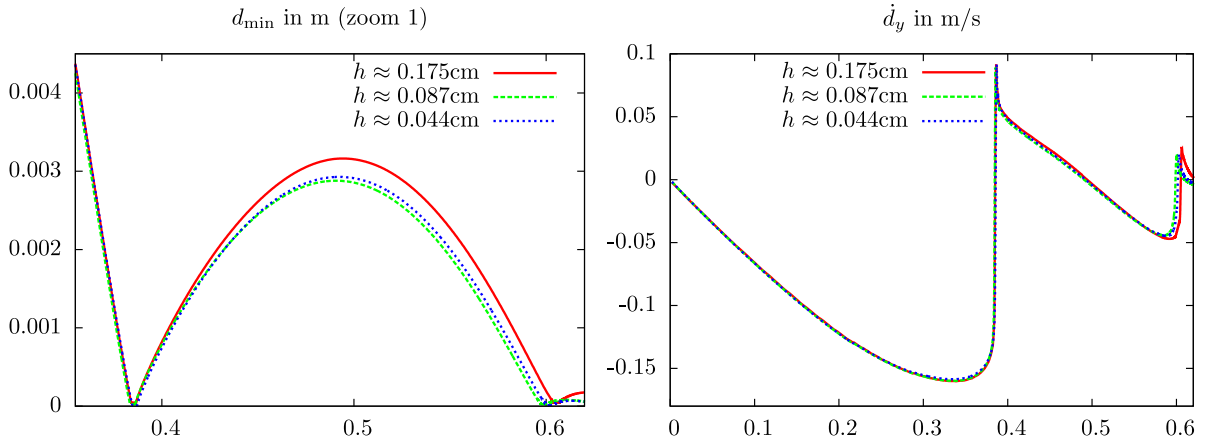


Fig. 23. Minimal distance $d_{\min}$ to the ground (left) and space-averaged solid velocity (right) over time on different mesh levels.

deduced from the right plot, which shows that for $\delta t_{\min} = 1.25 \cdot 10^{-4}$ the time of impact and release, where the curve shows a kink (i.e. the solution is non-smooth), is not captured accurately.

Space discretisation. In Fig. 23 we investigate the convergence behaviour under refinement of the finite element mesh. We fix the minimal time-step to $\delta t_{\min} = 3.125 \cdot 10^{-5}$ and consider a coarse mesh with a maximum cell size of $h \approx 0.175$cm and 3201 vertices and two finer meshes with 12 545 and 49 665 vertices that are constructed from the coarse mesh by global mesh refinement. We observe that the results both concerning minimal distance and vertical velocity are relatively close, even on the coarser mesh level, with an excellent agreement of the results on the finer meshes.

Comparison with a pure no-slip boundary condition. In Fig. 24, we compare the approach presented in this paper with a simple relaxed contact approach without porous medium ("No Darcy"), where a no-slip condition (resp. a Navier-slip condition) is imposed for the fluid on the bottom wall Σ_p. The no-slip condition is the boundary condition, which is typically used for viscous fluids in absence of contact. First, we note that the curves for the no-slip condition and the Navier-slip condition with (small) slip-length $\left(\frac{\alpha}{\sqrt{K_\tau \epsilon_p}}\right)^{-1}$ are almost identical. For this reason the latter curves are omitted in the following graphs.

As observed before, we see in the left picture that the curves obtained with the porous medium approach converge towards a certain bouncing height for $h \to 0$. Using a no-slip condition on Σ_p, the bounce get smaller and smaller and it is to be expected that for $h \to 0$ no bounce takes place at all (which is in agreement with the theoretical works on Navier–Stokes and contact [1,2,53]). The reason can be inferred from the zoom given on the right of Fig. 24, where we see that the fall is slowed down significantly right before the impact, while the curves for the two variants showed very good agreement until a distance of around 10^{-4} is reached. The reason are the strong fluid forces, in particular the pressure, that act against contact, when a pure no-slip condition is used. The finer the mesh, the better these forces are resolved. Interestingly, the results on the coarsest mesh ($h \approx 0.175$) show still a reasonable agreement, which might indicate that (only) on a very coarse mesh no-slip conditions could still yield physical results within a relaxed contact approach.

Finally, we show in Table 2 the spatially-averaged velocity of the solid at the time of impact t_i and the time of release t_r. Here we see quantitatively that the impact velocity is significantly reduced on the finer mesh levels when using a no-slip condition and thus, a much smaller rebound results.

5. Conclusion

We have introduced a physically consistent model to describe fluid–structure interactions with contact including seepage. For the latter a Darcy model is used on a thin porous layer of infinitesimal thickness. The approach can be used in a variety of different physical and numerical settings, including thick- and thin-walled solids, Eulerian or immersed (mixed-coordinate) descriptions, unfitted or fitted finite element discretisations, etc.

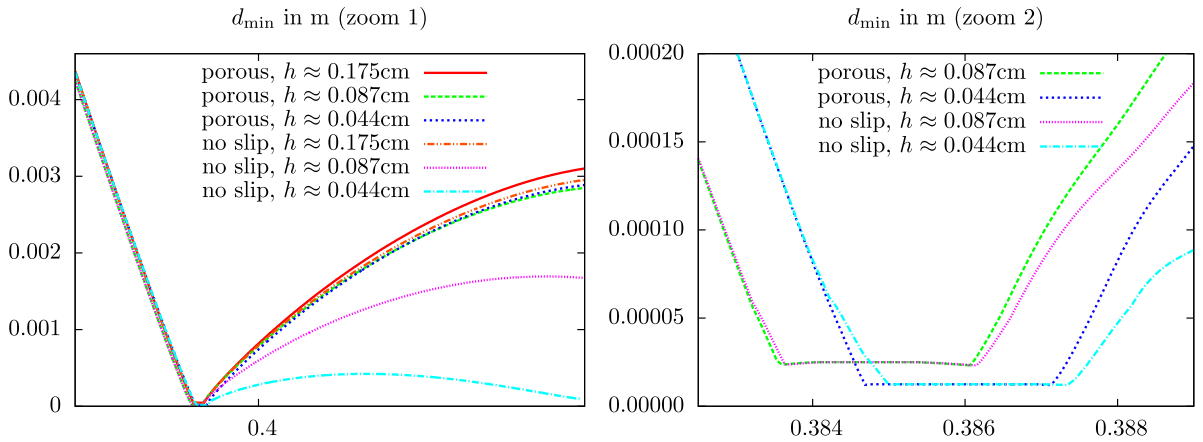


Fig. 24. Comparison of the contact approach with a porous layer with pure no-slip ("No Darcy") conditions. Minimal distance $d_{\min}$ to the ground over time on different mesh levels and two different zooms.

Table 2
Spatially averaged velocity $\overline{\dot{d}_y}(t) := |\Omega_s(t)|^{-1} \int_{\Omega_s(t)} \dot{d}_y(t)\,dx$ of the ball at the time of impact $t = t_i$ and at the time of release $t = t_r$ for relaxed contact algorithms with a porous medium model and pure no-slip or Navier-slip conditions on Σ_p on 3 different mesh levels.

h	Porous		No-slip		Navier-slip	
	$-\overline{\dot{d}_y}(t_i)$	$\overline{\dot{d}_y}(t_r)$	$-\overline{\dot{d}_y}(t_i)$	$\overline{\dot{d}_y}(t_r)$	$-\overline{\dot{d}_y}(t_i)$	$\overline{\dot{d}_y}(t_r)$
$1.75 \cdot 10^{-3}$	$1.12 \cdot 10^{-1}$	$8.87 \cdot 10^{-2}$	$1.11 \cdot 10^{-1}$	$8.81 \cdot 10^{-2}$	$1.11 \cdot 10^{-1}$	$8.81 \cdot 10^{-2}$
$8.77 \cdot 10^{-2}$	$1.05 \cdot 10^{-1}$	$8.48 \cdot 10^{-2}$	$9.74 \cdot 10^{-2}$	$7.59 \cdot 10^{-2}$	$9.74 \cdot 10^{-2}$	$7.60 \cdot 10^{-2}$
$4.39 \cdot 10^{-2}$	$1.03 \cdot 10^{-1}$	$8.91 \cdot 10^{-2}$	$7.25 \cdot 10^{-2}$	$5.98 \cdot 10^{-2}$	$7.26 \cdot 10^{-2}$	$5.98 \cdot 10^{-2}$

The numerical results show that the approach is numerically stable and (relatively) insensitive to variations of the numerical parameters, such as γ_c. The model parameters ϵ_p and K of the porous layer need to be chosen depending on the application, e.g., the surface properties of the contacting bodies. Moreover, the results indicate convergence in both space and time. The time-step $\delta t_{\min}$ needs to be chosen very small in and around the contact interval to resolve the contact dynamics accurately.

Due to the relaxation of the contact conditions the approach is relatively easy to implement, in particular in comparison to approaches where a full topology change in the discrete fluid domain takes place and small numerical errors can lead to technical issues like unphysical "islands of fluid" appearing within the contact area, see [20].

Future work might focus on the extension to contact between multiple elastic bodies, or to Tresca or Coulomb friction, see [54,55] Moreover, further developments within the time discretisation schemes on moving (sub-)domains are desirable including for example adaptive strategies for the time steps δt. or stabilised higher-order time-discretisation schemes, as discussed in [56].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

MF and FG were partially supported by the French National Research Agency (ANR), through the SIMR project (ANR-19-CE45-0020). EB was partially supported by the EPSRC, UK grants EP/P01576X/1 and EP/T033126/1. EB, MF and FG also acknowledge support by Inria, France through the Associated Team IMFIBIO.

References

[1] T.I. Hesla, Collisions of Smooth Bodies in Viscous Fluids: A Mathematical Investigation (Ph.D. thesis), University of Minnesota, 2004.

[2] M. Hillairet, Lack of collision between solid bodies in a 2D incompressible viscous flow, Comm. Partial Differential Equations 32 (9) (2007) 1345–1371.

[3] M. Hillairet, T. Takahashi, Collisions in three-dimensional fluid structure interaction problems, SIAM J. Math. Anal. 40 (6) (2009) 2451–2477.

[4] T. Hagemeier, D. Thévenin, T. Richter, Settling of spherical, non-wetting particles in a high viscous fluid, 2020, arXiv eprint 2009.02250v1.

[5] C. Ager, B. Schott, A.-T. Vuong, A. Popp, W.A. Wall, A consistent approach for fluid-structure-contact interaction based on a porous flow model for rough surface contact, Internat. J. Numer. Methods Engrg. 119 (13) (2019) 1345–1378.

[6] L. Bălilescu, J.S. Martín, T. Takahashi, Fluid-rigid structure interaction system with Coulomb's law, SIAM J. Math. Anal. 49 (9) (2017) 4625–4657.

[7] D. Gerard-Varet, M. Hillairet, The influence of boundary conditions on the contact problem in a 3d Navier-stokes flows, J. Math. Pures Appl. 103 (2015) 1–38.

[8] D. Gerard-Varet, M. Hillairet, C. Wang, The influence of boundary conditions on the contact problem in a 3d Navier-Stokes flows, Arch. Ration. Mech. Anal. 195 (2010) 375–407.

[9] T.E. Tezduyar, S. Sathe, Modeling of fluid-structure interactions with the space-time finite elements: solution techniques, Internat. J. Numer. Methods Fluids 54 (2007) 855–900.

[10] D. Kamensky, M. Hsu, D. Schillinger, J. Evans, A. Aggarwal, Y. Bazilevs, M. Sacks, T. Hughes, An immersogeometric variational framework for fluid–structure interaction: Application to bioprosthetic heart valves, Comput. Methods Appl. Mech. Engrg. 284 (2015) 1005–1053.

[11] S. Frei, Eulerian Finite Element Methods for Interface Problems and Fluid-Structure Interactions (Ph.D. thesis), Heidelberg University, 2016, http://www.ub.uni-heidelberg.de/archiv/21590.

[12] S. Frei, T. Richter, An accurate Eulerian approach for fluid-structure interactions, in: S. Frei, B. Holm, T. Richter, T. Wick, H. Yang (Eds.), Fluid-Structure Interaction: Modeling, Adaptive Discretization and Solvers, in: Rad Ser Computat Appl Math, Walter de Gruyter, Berlin, 2017.

[13] H. von Wahl, T. Richter, S. Frei, T. Hagemeier, Falling balls in a viscous fluid with contact: Comparing numerical simulations with experimental data, Phys. Fluids 33 (3) (2021) 033304.

[14] N.D. dos Santos, J.-F. Gerbeau, J.-F. Bourgat, A partitioned fluid–structure algorithm for elastic thin valves with contact, Comput. Methods Appl. Mech. Engrg. 197 (19) (2008) 1750–1761.

[15] M. Astorino, J.-F. Gerbeau, O. Pantz, K.-F. Traoré, Fluid-structure interaction and multi-body contact: Application to aortic valves, Comput. Methods Appl. Mech. Engrg. 198 (45–46) (2009) 3603–3612.

[16] U.M. Mayer, A. Popp, A. Gerstenberger, W.A. Wall, 3D fluid–structure-contact interaction based on a combined XFEM FSI and dual mortar contact approach, Comput. Mech. 46 (1) (2010) 53–67.

[17] F. Hecht, O. Pironneau, An energy stable monolithic Eulerian fluid-structure finite element method, Internat. J. Numer. Methods Fluids 85 (7) (2017) 430–446.

[18] S. Zonca, P.F. Antonietti, C. Vergara, A Polygonal Discontinuous Galerkin Formulation for Contact Mechanics in Fluid-Structure Interaction Problems, MOX-Report No. 26/2020, MOX, Dipartimento di Matematica Politecnico di Milano, 2020, https://www.mate.polimi.it/biblioteca/add/qmox/26-2020.pdf.

[19] E. Burman, M.A. Fernández, S. Frei, A Nitsche-based formulation for fluid-structure interactions with contact, ESAIM: M2AN 54 (2) (2020) 531–564.

[20] C. Ager, A. Seitz, W.A. Wall, A consistent and versatile computational approach for general fluid-structure-contact interaction problems, Int. J. Numer. Methods Eng. 122 (19) (2021) 5279–5312.

[21] C.M. Elliott, T. Ranner, Finite element analysis for a coupled bulk-surface partial differential equation, IMA J. Numer. Anal. 33 (2) (2013) 377–402.

[22] V. Martin, J. Jaffré, J. Roberts, Modeling fractures and barriers as interfaces for flow in porous media, SIAM J. Sci. Comput. 26 (5) (2005) 1667–1691.

[23] E. Burman, M.A. Fernández, S. Frei, F.M. Gerosa, 3D-2D Stokes-Darcy Coupling for the modelling of seepage with an application to fluid-structure interaction with contact, in: Numerical Mathematics and Advanced Applications ENUMATH 2019, Springer, 2021, pp. 215–223.

[24] C.S. Peskin, Flow patterns around heart valves: a numerical method, J. Comput. Phys. 10 (2) (1972) 252–271.

[25] T. Dunne, An Eulerian approach to fluid–structure interaction and goal-oriented mesh adaptation, Internat. J. Numer. Methods Fluids 51 (9–10) (2006) 1017–1039.

[26] G.-H. Cottet, E. Maitre, T. Milcent, Eulerian formulation and level set models for incompressible fluid-structure interaction, ESAIM: M2AN 42 (3) (2008) 471–492.

[27] T. Richter, A fully Eulerian formulation for fluid-structure interactions, J. Comput. Phys. 233 (2013) 227–240.

[28] P. Hansbo, J. Hermansson, T. Svedberg, Nitsche's method combined with space-time finite elements for ALE fluid-structure interaction problems, Comput. Methods Appl. Mech. Engrg. 193 (39–41) (2004) 4195–4206.

[29] E. Burman, M.A. Fernández, An unfitted Nitsche method for incompressible fluid–structure interaction using overlapping meshes, Comput. Methods Appl. Mech. Engrg. 279 (2014) 497–514.

[30] F. Alauzet, B. Fabréges, M.A. Fernández, M. Landajuela, Nitsche-XFEM for the coupling of an incompressible fluid with immersed thin-walled structures, Comput. Methods Appl. Mech. Engrg. 301 (2016) 300–335.

[31] S. Frei, T. Richter, A locally modified parametric finite element method for interface problems, SIAM J. Numer. Anal. 52 (5) (2014) 2315–2334.
[32] S.P. Neuman, Theoretical derivation of Darcy's law, Acta Mech. 25 (3) (1977) 153–170.
[33] P.G. Saffman, On the boundary condition at the surface of a porous medium, Stud. Appl. Math. 50 (2) (1971) 93–101.
[34] A. Mikelic, W. Jäger, On the interface boundary condition of Beavers, Joseph, and Saffman, SIAM J. Appl. Math. 60 (4) (2000) 1111–1127.
[35] D. Nield, A. Bejan, et al., Convection in Porous Media, Vol. 3, Springer, 2006.
[36] P. Alart, A. Curnier, A mixed formulation for frictional contact problems prone to Newton like solution methods, Comput. Methods Appl. Mech. Engrg. 92 (3) (1991) 353–375.
[37] R. Scholz, Numerical solution of the obstacle problem by the penalty method, Computing 32 (1984) 297–306.
[38] F. Chouly, P. Hild, On convergence of the penalty method for unilateral contact problems, Appl. Numer. Math. 65 (2012).
[39] A. Massing, M.G. Larson, A. Logg, Efficient implementation of finite element methods on nonmatching and overlapping meshes in three dimensions, SIAM J. Sci. Comput. 35 (1) (2013) C23–C47.
[40] F. Alauzet, B. Fabrèges, M. Fernández, M. Landajuela, Nitsche-XFEM for the coupling of an incompressible fluid with immersed thin-walled structures, Comput. Methods Appl. Mech. Engrg. 301 (2015).
[41] S. Zonca, C. Vergara, L. Formaggia, An unfitted formulation for the interaction of an incompressible fluid with a thick structure via an XFEM/DG approach, SIAM J. Sci. Comput. 40 (1) (2018) B59–B84.
[42] E. Burman, M.A. Fernández, P. Hansbo, Continuous interior penalty finite element method for Oseen's equations, SIAM J. Numer. Anal. 44 (3) (2006) 1248–1274.
[43] E. Burman, Ghost penalty, C. R. Math. Acad. Sci. Paris 348 (21–22) (2010) 1217–1220.
[44] C. Lehrenfeld, M. Olshanskii, An Eulerian finite element method for PDEs in time-dependent domains, ESAIM: M2AN 53 (2) (2019) 585–614.
[45] F.M. Gerosa, Immersed Boundary Methods for Fluid-Structure Interaction with Topological Changes (Ph.D. thesis), Sorbonne Université, 2021.
[46] D. Di Pietro, A. Ern, Mathematical Aspects of Discontinuous Galerkin Methods, in: Mathematics & Applications, vol. 69, Springer, Heidelberg, 2012.
[47] S. Frei, G. Judakova, T. Richter, A locally modified second-order finite element method for interface problems, 2020, arXiv e-print 2 007.13906.
[48] S. Frei, An edge-based pressure stabilization technique for finite elements on arbitrarily anisotropic meshes, Internat. J. Numer. Methods Fluids 89 (10) (2019) 407–429.
[49] S. Frei, T. Richter, A second order time-stepping scheme for parabolic interface problems with moving interfaces, ESAIM: M2AN 51 (4) (2017) 1539–1560.
[50] E. Burman, S. Frei, A. Massing, Eulerian time-stepping schemes for the non-stationary Stokes equations on time-dependent domains, Numerische Mathematik (2022) 1–56.
[51] F. Chouly, P. Hild, A nitsche-based method for unilateral contact problems: numerical analysis, SIAM J. Numer. Anal. 51 (2) (2013) 1295–1307.
[52] www.gascoigne.de R. Becker, M. Braack, D. Meidner, T. Richter, B. Vexler, The finite element toolkit Gascoigne3d.
[53] D. Gerard-Varet, M. Hillairet, C. Wang, The influence of boundary conditions on the contact problem in a 3d Navier-Stokes flow, J. Math. Pures Appl. 103 (2015) 1–38.
[54] F. Chouly, An adaptation of Nitsche's method to the tresca friction problem, J. Math. Anal. Appl. 411 (2014) 329–339.
[55] F. Chouly, P. Hild, V. Lleras, Y. Renard, Nitsche-based finite element method for contact with Coulomb friction, https://hal.archives-ouvertes.fr/hal-01654487, 2017,
[56] F. Chouly, P. Hild, Y. Renard, A nitsche finite element method for dynamic contact: 2. Stability of the schemes and numerical experiments, ESAIM: M2AN 49 (2) (2015) 503–528.

Falling balls in a viscous fluid with contact: Comparing numerical simulations with experimental data

Cite as: Phys. Fluids **33**, 033304 (2021); https://doi.org/10.1063/5.0037971
Submitted: 18 November 2020 • Accepted: 06 January 2021 • Published Online: 02 March 2021

Henry von Wahl, Thomas Richter, Stefan Frei, et al.

Physics of Fluids ARTICLE scitation.org/journal/phf

Falling balls in a viscous fluid with contact: Comparing numerical simulations with experimental data

Cite as: Phys. Fluids **33**, 033304 (2021); doi: 10.1063/5.0037971
Submitted: 18 November 2020 · Accepted: 6 January 2021 ·
Published Online: 2 March 2021

View Online Export Citation CrossMark

Henry von Wahl,[1,a)] Thomas Richter,[1] Stefan Frei,[2] and Thomas Hagemeier[3]

AFFILIATIONS

[1]Institute for Analysis and Numerics, Otto-von-Guericke Universität, Universitätsplatz 2, 39106 Magdeburg, Sachsen-Anhalt, Germany
[2]Department of Mathematics and Statistics, Universität Konstanz, Universitätsstraße 10, 78457 Konstanz, Baden-Württemberg, Germany
[3]Laboratory of Fluid Dynamics and Technical Flows, Otto-von-Guericke Universität, Universitätsplatz 2, 39106 Magdeburg, Sachsen-Anhalt, Germany

[a)]**Author to whom correspondence should be addressed:** henry.vonwahl@ovgu.de

ABSTRACT

We evaluate a number of different finite-element approaches for fluid–structure (contact) interaction problems against data from physical experiments. This consists of trajectories of single particles falling through a highly viscous fluid and rebounding off the bottom fluid tank wall. The resulting flow is in the transitional regime between creeping and turbulent flows. This type of configuration is particularly challenging for numerical methods due to the large change in the fluid domain and the contact between the wall and the particle. In the finite-element simulations, we consider both rigid body and linear elasticity models for the falling particles. In the first case, we compare the results obtained with the well-established Arbitrary Lagrangian–Eulerian (ALE) approach and an unfitted moving domain method together with a simple and common approach for contact avoidance. For the full fluid–structure interaction (FSI) problem with contact, we use a fully Eulerian approach in combination with a unified FSI-contact treatment using Nitsche's method. For higher computational efficiency, we use the geometrical symmetry of the experimental setup to reformulate the FSI system into two spatial dimensions. Finally, we show full three-dimensional ALE computations to study the effects of small perturbations in the initial state of the particle to investigate deviations from a perfectly vertical fall observed in the experiment. The methods are implemented in open-source finite element libraries, and the results are made freely available to aid reproducibility.

I. INTRODUCTION

Flows containing particles, i.e., particulate flows or particles settling in a fluid, have many industrial and biological applications. Examples range from the transport of blood cells in blood flows[1] to the simulation of fluidized bed reactors.[2]

We shall consider single elastic spherical particles falling freely in a viscous fluid and rebounding off the bottom wall of the fluid domain at Reynolds numbers in the transitional regime between creeping and turbulent flows. The multiphase and fluid–structure interaction (FSI) problem with solid contact posed by the settling in the fluid and rebounding off a wall is challenging from both an analytical and a numerical perspective.

From the theoretical point of view, the correct model for the transition to contact with the bottom wall is not yet fully understood. In the case where a rigid solid is assumed, most flow models lead to results contradicting real world observations. For example, if a creeping flow is assumed such that the linear Stokes equations are applicable, then contact can only occur under singular forces, cf. Ref. 3. When the non-linear incompressible Navier–Stokes together with no-slip boundary conditions are taken for the fluid model, then contact cannot occur and it is impossible to release contact.[4] This can however be overcome; if the boundary condition is modified to a free-slip condition,[5] the rough nature of the surface is taken into account[6] or the fluid is taken to be compressible.[7] If the solid model is changed to take the

elasticity of the body into account, then it is currently assumed that even with perfectly smooth boundaries and incompressibility, rebounding without contact can occur due to the storage of energy in the elastic solid.[8–10] This has been refined recently (in the Stokes setting),[11] where it has been shown that the internal storage of energy is not sufficient but that additionally a change in the "flatness" is necessary to achieve physically meaningful rebound without topological contact.

For numerical methods, the challenge lies in the discretization of the resultant FSI system.[12] It consists of a free boundary value problem with a moving interface. The most well-established method for this is the Arbitrary Lagrangian–Eulerian (ALE) approach.[13] This approach leads to very efficient and accurate computations in situations where the method is usable. However, its usage is limited as it breaks down when deformations with respect to the reference configuration become too large and when contact occurs.[14] The lattice-Boltzmann method is an alternative approach that is also able to describe the behavior of settling particles.[15] To deal with large deformations, overlapping mesh techniques have been developed.[16] Here, the background fluid domain and the region around the structure are meshed separately so that the fluid–solid interface is resolved. The two meshes are then coupled using unfitted approaches. This then allows a hybrid approach, where the solid and the near fluid are treated using the ALE framework, while the remaining fluid is treated in Eulerian coordinates.[17] To overcome both large deformations and contact, fully Eulerian approaches have lately become the focus of research. In the case of rigid bodies, a number of different approaches have been considered, for example, based on fictitious domain methods using Lagrange multipliers,[18] XFEM type approaches,[19,20] and most recently, CutFEM approaches[21] using Nitsche's method.[22,23] Here, a major issue remains of achieving a realistic rebound effect since an artificial contact/lubrication force is added to the equation governing the motion of the solid to prevent the overlap of the solid regions.[18] Nevertheless, topological changes appear to be unproblematic for the CutFEM type approaches.[24]

Considering full fluid–structure interactions, immersed approaches have become popular in recent years.[25–28] Here, the fluid and the solid are treated in their natural Eulerian and Lagrangian coordinate systems, respectively, and the sub-domains are meshed separately. The two meshes are then coupled by means of Nitsche's method[25,27] or using Lagrange multipliers.[26,28,29] Another possibility to handle large deformations and contact is by using fully Eulerian approaches, where both the solid and fluid equations are formulated in the Eulerian coordinate framework, which simplifies the coupling within monolithic algorithms.[30–34] All these approaches are however relatively new and require further development with respect to accuracy and robustness.

The aforementioned methods have been applied to different test cases for numerical validation, and *a priori* error estimates are also available in most cases. The established benchmarks for fluid-structure interaction problems such as those in Ref. 35 completely avoid contact since the methods that handle contact remain relatively new. For rigid-body motion, most numerical studies are interpreted qualitatively or compared to artificial, analytically derived solutions. Especially in the cases where artificial forces are introduced in order to avoid contact of rigid solids, real validation is near impossible as this introduces model parameters for which there is no *a priori* knowledge on a good choice. However, a number of FSI methods have recently become available that are able to resolve contact.[36–40] This then raises the question of how well the different modeling and discretization approaches depict the behavior of contact and rebound observed in physical experiments.

In this work, we take recently published data from experiments where different solid spherical particles were allowed to settle in a viscous fluid.[41,42] We then use a rigid-body ALE, a rigid-body Eulerian CutFEM, and a fully Eulerian FSI approach to simulate the scenarios presented by the physical experiments. This aims to show the validity and the applicability of these different approaches to the different aspects/problems posed by this process. Furthermore, we will illustrate how spatially reduced models are able to capture the behavior in comparison to full three-dimensional computations. To the best of our knowledge, there is currently no comparable benchmark that considers such a multiphase flow/fluid-structure interaction problem with contact, which is validated against experimental data.

The remainder of this paper is structured as follows: In Sec. II, we describe the considered problem, that is, a description of the physical experiment, the mathematical models used to describe the experiment, the specific setups we will simulate, and the quantities used to compare the numerical simulations with the experimental data. Section III then briefly covers the reduced formulation we apply to increase the computational efficiency in our numerical methods. The numerical computations are then presented in Sec. IV; we present the details of the different numerical approaches in Subsection IV A, and the results are then presented in Subsection IV B. We discuss the conclusions from these results in Sec. V and consider the aspects that remain open. Furthermore, we define and compute two simplified setups in Appendix B, designed to help others reproduce the presented computational results.

II. DESCRIPTION

We describe the experimental setup used to gather the data and the mathematical model that we will use to reproduce the behavior observed in the experiments, and we define relevant quantities used to compare the results quantitatively.

A. Physical experiment

The experiments in Ref. 42 capture the settling and impact process of spherical particles with different sizes and densities in a cylindrical tank. The latter contains a liquid mixture consisting of glycerine and water at equal volume fractions. At the bottom of the cylindrical tank, which is made of acrylic glass for optical access, a massive steel anvil serves as the impact object. Moreover, the cylinder is surrounded by a rectangular container, filled with refractive index matching liquid, to compensate for optical distortion coming from the curved cylinder walls. The filling level allows the observation of the particle settling along a vertical distance of 140 mm–160 mm, depending on the particle size. Initially, each particle is held in place and submerged in the liquid by using a vacuum tweezer. The particle is then released by switching off the vacuum pump. The particle is tracked during the settling process and the impact on the steel anvil, including the rebound, using a high-speed CMOS-camera. This acquires shadow images at a frame rate of 1000 fps and a scale factor of 8.89 pixel/mm. An image processing algorithm coded in MATLAB yields the in-plane particle

coordinates as a function of time and allows us to extract the instantaneous particle settling velocity.

The resulting data are available via Mendeley Data.[41] These data are the basis for our comparison and validation of the numerical code.

B. Mathematical model

We consider a bounded domain $\Omega \in \mathbb{R}^d$, with $d \in \{2, 3\}$, over a finite, non-empty, time interval $[0, T_{end}]$. This is divided into a d-dimensional fluid region $\mathcal{F}$, a d-dimensional solid $\mathcal{S}$, and a $d-1$-dimensional interface $\mathcal{I}$ dividing the solid and fluid regions. For these, we have $\Omega = \mathcal{F} \dot\cup \mathcal{I} \dot\cup \mathcal{S}$.

1. Fluid model

In the time dependent fluid region $\mathcal{F}(t)$, we consider the incompressible Navier–Stokes equations. Find a velocity $\boldsymbol{u}$ and a pressure p such that

$$\rho_f(\partial_t \boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}) - \operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}, p) = 0, \tag{1a}$$

$$\operatorname{div}(\boldsymbol{u}) = 0 \tag{1b}$$

with the non-symmetric stress tensor

$$\boldsymbol{\sigma}(\boldsymbol{u}, p) = \mu_f \nabla \boldsymbol{u} - p\mathrm{Id},$$

where $\mu_f = \rho_f \nu_f$ is the fluid's dynamic viscosity, ν_f is the kinematic viscosity, ρ_f is the fluid's density, and Id is the d-dimensional identity operator. Appropriate boundary conditions to complete this system will be discussed later. Note that we use bold face letters to denote vector/matrix valued quantities, while regular faced letters denote scalar objects.

2. Elastic solid and fluid–structure interaction

We consider a linear elastic solid model in $\mathcal{S}$ for the solid displacement $\boldsymbol{d}$ and the solid velocity $\dot{\boldsymbol{d}}$, given by

$$\rho_s \partial_t \dot{\boldsymbol{d}} - \operatorname{div}\boldsymbol{\sigma}_s(\boldsymbol{d}) = (\rho_s - \rho_f)g, \quad \dot{\boldsymbol{d}} = \partial_t \boldsymbol{d}, \tag{2}$$

with the acceleration due to gravity $g = -9.807\,\mathrm{m\,s^{-2}}$ and the Cauchy stress tensor $\boldsymbol{\sigma}_s$ defined by

$$\boldsymbol{\sigma}_s(\boldsymbol{d}) = 2\mu_s E(\boldsymbol{d}) + \lambda_s \operatorname{tr}(E(\boldsymbol{d}))I, \quad E(\boldsymbol{d}) = \frac{1}{2}(\nabla \boldsymbol{d} + \nabla \boldsymbol{d}^T),$$

where μ_s and λ_s denote the Lamé parameters. Note that we have subtracted the fluid gravitational force $\rho_f g$ on the right-hand side of the solid equation to be consistent with the equations for a rigid solid presented in the following paragraph. Alternatively, this could be added to the right-hand side of the fluid equations.

Solid and fluid are coupled by means of no-slip coupling conditions on $\mathcal{I}$,

$$\boldsymbol{u} = \dot{\boldsymbol{d}}, \qquad \boldsymbol{\sigma}(\boldsymbol{u}, p)\boldsymbol{n} = \boldsymbol{\sigma}_s(\boldsymbol{d})\boldsymbol{n}. \tag{3}$$

The current position of the interface $\mathcal{I}(t)$ is determined by the displacement variable $\boldsymbol{d}$.

3. Rigid solid

As the solid materials that we consider are relatively hard, the consideration of a rigid solid yields a good approximation of the FSI dynamics, at least up to the moment when the solid comes close to the lower wall. The movement of the solid is governed by Newton's second law of motion. Let $\boldsymbol{c}_{\mathcal{S}}(t)$ be the center of mass of the solid $\mathcal{S}$. Since we will consider spherical particles, this is then governed by

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\boldsymbol{c}_{\mathcal{S}}(t) \cdot m_{\mathcal{S}} = \boldsymbol{f}_s, \tag{4}$$

where $m_{\mathcal{S}}$ is the mass of the solid and $\boldsymbol{f}_s$ are the forces acting on the solid in the horizontal and vertical directions. For simplicity, we assume that the horizontal forces are negligible. The vertical forces are then the gravitational pull, buoyancy, and viscous drag,

$$\boldsymbol{f}_s = \begin{pmatrix} 0 \\ 0 \\ m_{\mathcal{S}} g - \operatorname{vol}(\mathcal{S})\rho_f g + \boldsymbol{F}_3 \end{pmatrix}, \tag{5}$$

where $\operatorname{vol}(\mathcal{S})$ is the volume of the solid and $\boldsymbol{F}_3$ is the viscous drag force in the vertical direction. This is the third component of

$$\boldsymbol{F} = \int_{\mathcal{I}} \boldsymbol{\sigma}(\boldsymbol{u}, p)\boldsymbol{n}\mathrm{d}s. \tag{6}$$

Note that we added the effects of buoyancy in (5). This would be naturally included in $\boldsymbol{F}$ if we added body force $\rho_f g \nabla \boldsymbol{x}_d$ to the right-hand side of the fluid equation (1a). However, since this would only affect the pressure, it is sufficient to consider the homogeneous equation (1a) and add the effect of buoyancy to (5). Furthermore, this approach is more accurate in our case since only pressure-robust methods are able to reflect this effect of gradient contributions in the forcing term on the pressure exactly on the numerical level.[43]

As a result of us neglecting the horizontal movement of the solid, (4) becomes a scalar ordinary differential equation (ODE). We can also simplify the terms in (5) so that, in total, we come to the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{v}_{\mathcal{S},3}(t) = \frac{\rho_s - \rho_f}{\rho_s} g + \frac{\boldsymbol{F}_3}{\operatorname{vol}(\mathcal{S})\rho_s}, \tag{7}$$

where $\boldsymbol{v}_{\mathcal{S}}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{c}_{\mathcal{S}}(t)$ is the solid's velocity.

The solid's motion couples back to the fluid equations through the boundary condition at the interface $\mathcal{I}$ by requiring the continuity of the velocity, i.e.,

$$\boldsymbol{u}|_{\mathcal{I}} = \boldsymbol{v}_{\mathcal{S}} = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{c}_{\mathcal{S}}(t). \tag{8}$$

We note that this model neglects rotational effects. As will be shown below, this does not have a major impact on the quality of the resulting approximations.

C. Domain description

Since we consider balls of different sizes, we shall keep the domain description general. The background domain is a cylinder $\Omega = \{\boldsymbol{x} = (x, y, z)^T \in \mathbb{R}^3 \mid x^2 + y^2 < R^2, 0 < z < H\}$ for a given radius R and a given height H. At $t = 0$, the solid domain is described

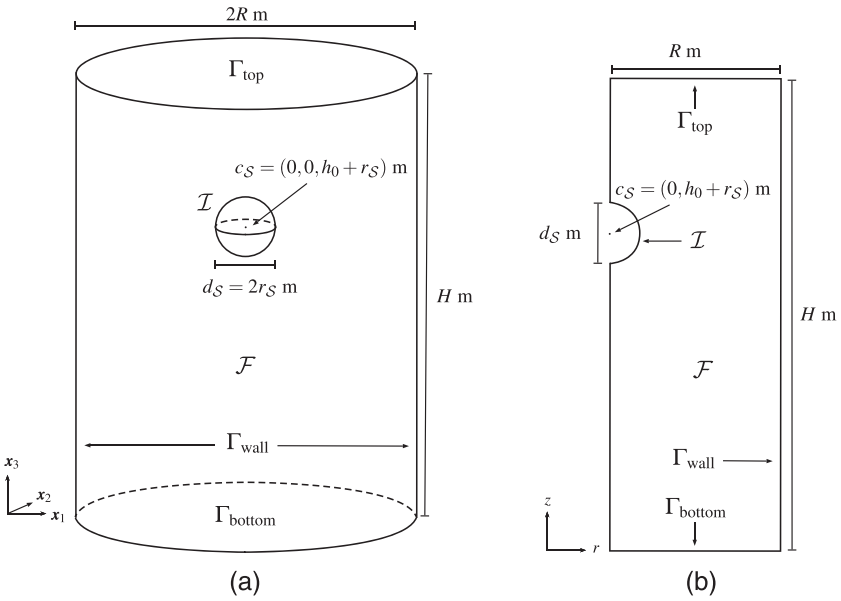


FIG. 1. The initial spatial configuration: (a) three-dimensional domain and (b) rotationally reduced domain.

by $\mathcal{S}(0) = \{\boldsymbol{x} = (x, y, z)^T \in \mathbb{R}^3 \mid x^2 + y^2 + (z - (h_0 + r_S))^2 < r_S^2\}$ for a given ball radius r_S and an initial height of the bottom of the ball h_0. Accordingly, the volume of the solid is given by $\mathrm{vol}(\mathcal{S}) = 4\pi r^3/3$. An illustration of this can be seen on the left of Fig. 1.

1. Boundary conditions

We denote the top boundary ($z = H$) of the cylinder as Γ_{top}, the bottom boundary ($z = 0$) as Γ_{bottom}, and the side of the cylinder ($x^2 + y^2 = R^2$) as Γ_{wall}.

On the interface $\mathcal{I}$ between the solid and the fluid, the Dirichlet boundary condition is given by the continuity of the velocity; see (3) and (8). On the wall and bottom boundaries $\Gamma_{\mathrm{bottom}} \cup \Gamma_{\mathrm{wall}}$, we shall impose homogeneous Dirichlet boundary conditions $\boldsymbol{u} = 0$. In order to approximate the free surface at the top of the water tank Ω, we impose a free-slip boundary condition $\boldsymbol{u}_3 = 0$ at Γ_{top}.

D. Setup

1. Spatial parameters

We consider two different particles: one with diameter $d_S = 6$ mm consisting of polytetrafluoroethylene/teflon (PFTE) and one with diameter $d_S = 22$ mm consisting of rubber. We shall refer to these two cases as PTFE6 and Rubber22, respectively.

Both particles are considered inside the same cylindrical fluid tank with radius $R = 0.055$ m and a height of $H = 0.2$ m. The spatial setup for both cases is summarized in Table I.

2. Material parameters

The numerical computation of the above-mentioned scenarios requires the following material parameters. For the fluid, these are the density and viscosity. For the solid balls, we require the density and

TABLE I. Geometrical parameters of the test cases.

	Geometry				Boundary conditions		
Experiment	R (m)	H (m)	r_S (m)	h_0 (m)	$\Gamma_{\mathrm{wall}} \cup \Gamma_{\mathrm{bottom}}$	Γ_{top}	$\mathcal{I}$
PTFE6	0.055	0.2	0.003	0.161 661 6	$\boldsymbol{u} = 0$	$\boldsymbol{u}_d = 0$	$\boldsymbol{u} = \boldsymbol{v}_S$
Rubber22			0.011	0.146 120 3			

TABLE II. Summary of the benchmark setup in standard units.

		Material parameters				
Experiment	g (m s^{-2})	μ_f (kg m^{-1} s^{-1})	ρ_f (kg m^{-3})	ρ_s (kg m^{-3})	λ_s (kg m^{-1} s^{-2})	μ_s (kg m^{-1} s^{-2})
PTFE6	−9.807	0.008	1141	2122	$2.638\,70 \times 10^9$	$2.294\,52 \times 10^8$
Rubber22				1361	$3.332\,89 \times 10^9$	$6.667\,11 \times 10^5$

the Lamé parameters. Both fluid and solid parameters are summarized in Table II.

Remark (source of material parameters). The fluid parameters and the solid densities are given in the original experimental paper.[42] We derived the Lamé parameters from the Young's modulus and Poisson ratio of the used materials. In the case of the PTFE6 ball, these are about 670 MPa $= 670\,000\,000\,\mathrm{kg\,m^{-1}\,s^{-2}}$ (https://www.kugelpompel.at/upload/2312783_Datenblatt%20Kunststoffkugel%20PTFE%20V1.01.pdf) and $\nu_s = 0.46$ (http://www.matweb.com/search/datasheet_print.aspx?matguid=4e0b2e88eeba4aaeb18e8820f1444cdb), respectively.

The chemical analysis of the Rubber22 material suggests this to be hydrogenated nitrile rubber (HNBR). The Young's modulus of this is ∼1.7 MPa to 20.7 MPa $= 1\,700\,000\,\mathrm{kg\,m^{-1}\,s^{2}}$–$20\,700\,000\,\mathrm{kg\,m^{-1}\,s^{2}}$ (https://eriks.de/content/dam/de/pdf/downloads/dichtungen/o-ringe/ERIKS_Technisches-Handbuch-O-Ringe_de.pdf, p. 20), and the Poisson ratio is $\nu_s = 0.4999$.[44]

E. Quantities of interest

We will use the following quantities to compare our numerical results with each other and with the experimental data:

- t_* let $t_0 = t|_{c_S = h_0}$ be the time at which the center of mass is at h_0, i.e., when the ball has traveled r_S vertically. We define t_* as the time after release when $\mathrm{dist}(\mathcal{I}, \Gamma_{\mathrm{bottom}}) = d_S$ relative to t_0, i.e., for PTFE6 $t_* = t|_{c_S=(0,0,0.009)} - t_0$ and for Rubber22 $t_* = t|_{c_S=(0,0,0.033)} - t_0$.
- v_* the velocity of the ball in the z-direction at $t = t_* + t_0$.
- f_* the vertical component of the force $\boldsymbol{F}$ acting on the ball at time $t_* + t_0$.
- t_{cont} the time of the first solid contact relative to t_0.
- t_{jump} the time between contact and the second change in direction is realized, i.e., the amount of time the balls travels upward after the first contact.
- d_{jump} the maximum of $\mathrm{dist}(\mathcal{I}, \Gamma_{\mathrm{bottom}})$ after contact, i.e., the size of the bounce.

An illustration of how these quantities are defined can be seen in Fig. 2.

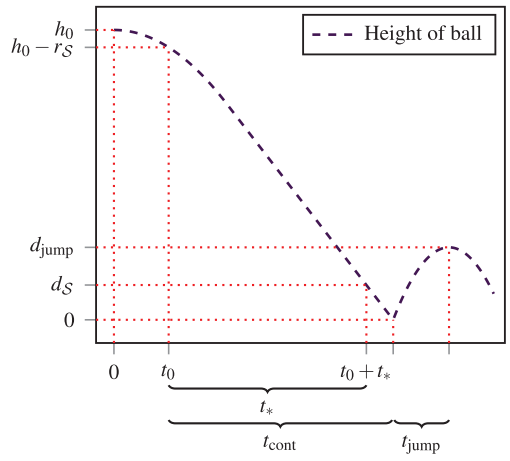


FIG. 2. Illustration for the definitions of the quantities of interest.

Remark (choice of reference values). In the experiments, we observed that the balls do not immediately start to fall after they are released. The settling process starts with some distance to the liquid surface but in the closest vicinity of the vacuum cup. In general, particles experience an increased drag force when they are moving toward or away from a solid wall and free fluid-surfaces. Accordingly, the early stage of the settling process here is dominated by an increased drag force coming from the vacuum cup and the liquid surface.

Due to the rather slow motion of the sphere at the beginning, the moment of release cannot be defined well. To be able to compare the numerical results with the experimental data, we therefore defined the above-mentioned quantities relative to the point in time at which the ball has traveled the distance of the ball's radius. Furthermore, since we have no measurement of the drag force in in the experiment, f_* will only be used to compare the computations directly.

To establish v_*, t_{cont}, d_{cont}, and d_{jump} from the experimental data, we interpolate the height data using a spline of order 3. This spline is then evaluated to establish the time at which $\boldsymbol{c}_S = (0, 0, 3r_S/2)$, the time of contact, as well as the time and height of the jump. The velocity is then taken as the first derivative of the spline representing the height. The resulting reference values for the quantities of interest are shown together with our numerical results in Tables III and IV, respectively.

The experimental study[42] was conducted so that the horizontal displacement of the particles was minimal. The dataset that we will compare against shows a maximal horizontal displacement of less than 2 mm and 0.75 mm in the PTFE6 and Rubber22 cases, respectively. This compares with the mean over time of the maximal deviation in the center location between experiment repetitions of 0.192 mm and 0.135 mm for the PTFE6 and Rubber22 cases, respectively. However, the experiment is only able to give the projection of the horizontal displacement onto the x–z plane. Thus, it is not possible to detect the true horizontal motion. Note that we have ignored the horizontal motion in the computation of the reference values. However, since overall the horizontal deflection is small, we consider this to be reasonable for the present purpose.

III. REDUCED MODEL

The setup described in Subsection II C is symmetric with respect to rotation if viewed in cylindrical coordinates. The experimental data presented in Refs. 41 and 42 show a rotational component in the motion of the solid, and they also show a small deflection of the center of mass c_S from the z axis. However, these effects are small, and since the material parameters are such that the resulting flow is in the low to intermediate Reynolds-number regime,[42] we assume that the solution is described sufficiently well by a rotationally symmetric flow in cylindrical coordinates. We use this in order to obtain a two-dimensional reduced formulation that is computationally cheaper compared to full three-dimensional computations. In Subsection IV B 3, we will also present a fully resolved three-dimensional simulation to have a closer look at these rotational effects. To distinguish between the full three-dimensional domains and the reduced two-dimensional domain, we shall denote objects stemming from the three-dimensional setup with a superscript "3d" if there is a potential for ambiguity. For the sake of readability, objects based on the reduced two-dimensional setup will not have a special notation.

TABLE III. Results for the PTFE6 setup. The experimental values have been reproduced with permission from T. Hagemeier, Particle settling-transitional regime, version 1, Mendeley Data, September 2020.

	Discretization				Results					
Method	$[h_{\min}, h_{\max}]$	Δt	dof	nze	t_*	v_*	f_*	t_{cont}	t_{jump}	d_{jump}
ALE	[0.000 08, 0.004]	1/200	7.54	0.34	0.542 134	−0.312 2077	$1.113\,93 \cdot 10^{-3}$	…	…	…
	[0.000 04, 0.002]	1/800	23.88	1.10	0.539 231	−0.313 9245	$1.119\,82 \cdot 10^{-3}$	…	…	…
	[0.000 02, 0.001]	1/3 200	83.07	3.90	0.539 026	−0.313 9931	$1.120\,08 \cdot 10^{-3}$	…	…	…
			Extrapolate order (in h)		0.539 010 3.8	−0.313 9960 4.6	$1.120\,21 \cdot 10^{-3}$ 4.0	… …	… …	… …
CutFEM	[0.001 14, 0.008]	1/2 000	14.50	0.43	0.579 181	−0.288 4462	$1.118\,97 \cdot 10^{-3}$	0.600 328	0.017 827	$1.339\,74 \cdot 10^{-3}$
	[0.000 57, 0.004]	1/2 000	45.80	1.35	0.536 698	−0.316 4086	$1.127\,23 \cdot 10^{-3}$	0.556 037	0.030 627	$2.706\,81 \cdot 10^{-3}$
	[0.000 29, 0.002]	1/2 000	152.06	4.50	0.530 537	−0.321 3762	$1.121\,22 \cdot 10^{-3}$	0.549 543	0.030 749	$2.692\,50 \cdot 10^{-3}$
FSI	[0.000 32, 0.00562]	[1/128 000, 1/1 000]	59.5	1.24	0.539 918	−0.314 7132	…	0.558 503	0.027 515	$1.817\,16 \cdot 10^{-3}$
	[0.000 16, 0.002 81]	[1/128 000, 1/1 000]	236.6	5.40	0.535 625	−0.315 9600	…	0.554 358	0.028 015	$2.061\,63 \cdot 10^{-3}$
Experiment					0.516 403	−0.330 987	…	0.534 503	0.027 92	$2.211\,70 \cdot 10^{-3}$

With the spaces $\boldsymbol{V}^{3d} = \{\boldsymbol{v} \in H^1(\mathcal{F}^{3d}) \,|\boldsymbol{v}|_{\mathcal{I}^{3d}} = \boldsymbol{u}_{\mathcal{S}^{3d}}, \boldsymbol{v}|_{\Gamma^{3d}_{\text{wall}} \cup \Gamma^{3d}_{\text{bottom}}} = 0 \text{ and } \boldsymbol{v}_z|_{\Gamma^{3d}_{\text{top}}} = 0\} \subset [\mathcal{H}^1(\mathcal{F}^{3d})]^3$ and $Q^{3d} = \mathcal{L}^2_0(\mathcal{F}^{3d})$, the weak formulation of the Navier–Stokes equation (1) is as follows:

Find $(\boldsymbol{u}, p) \in \boldsymbol{V}^{3d} \times Q^{3d}$ such that for all $(\boldsymbol{v}, q) \in \boldsymbol{V}^{3d} \times Q^{3d}$, it holds

$$A^{3d}_f(\boldsymbol{u}, p; \boldsymbol{v}, q) := m^{3d}(\partial_t \boldsymbol{u}, \boldsymbol{v}) + a^{3d}(\boldsymbol{u}, \boldsymbol{v}) + c^{3d}(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{v}) + b^{3d}(\boldsymbol{v}, p) + b^{3d}(\boldsymbol{u}, q) = 0 \tag{9}$$

with the multilinear forms

$$\begin{aligned} m^{3d}(\boldsymbol{u}, \boldsymbol{v}) &= \rho_f \int_{\mathcal{F}^{3d}} \boldsymbol{u} \cdot \boldsymbol{v} \mathrm{d}\boldsymbol{x}, \\ a^{3d}(\boldsymbol{u}, \boldsymbol{v}) &= \mu_f \int_{\mathcal{F}^{3d}} \nabla \boldsymbol{u} : \nabla \boldsymbol{v} \mathrm{d}\boldsymbol{x}, \\ c^{3d}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) &= \rho_f \int_{\mathcal{F}^{3d}} (\boldsymbol{u} \cdot \nabla) \boldsymbol{v} \cdot \boldsymbol{w} \mathrm{d}\boldsymbol{x}, \\ b^{3d}(q, \boldsymbol{v}) &= -\int_{\mathcal{F}^{3d}} q \nabla \cdot \boldsymbol{v} \mathrm{d}\boldsymbol{x}. \end{aligned} \tag{10}$$

In order to reduce this three-dimensional flow problem into a two-dimensional flow problem, we rewrite the problem into cylindrical coordinates r, ϕ, z. The rotational symmetry of a solution $\boldsymbol{u}$ means that $\partial_\phi \boldsymbol{u} = 0$. We use this to rotate the domain into the $r^+ - z$-plane and transform the weak formulation (9) accordingly. Now, let $\nabla = \begin{pmatrix} \partial_r \\ \partial_z \end{pmatrix}$ and $\Omega = \Omega^{3d} \cap (\mathbb{R}^+ \times \mathbb{R})$ be the reduced two-dimensional computational domain. A sketch of the three-dimensional domain transformed into two dimensions can be seen on the right of Fig. 1.

At the symmetry boundary ($r = 0$), the boundary conditions

$$\boldsymbol{u}_r = 0, \qquad \partial_r \boldsymbol{u}_z = 0 \tag{11}$$

are valid. The reduced spaces are then $\boldsymbol{V} = \{\boldsymbol{v} \in H^1(\mathcal{F}) \,|\boldsymbol{v}|_{\mathcal{I}} = \begin{pmatrix} 0 \\ \boldsymbol{v}_{\mathcal{S}_3} \end{pmatrix}, \boldsymbol{v}|_{\Gamma_{\text{wall}} \cup \Gamma_{\text{bottom}}} = 0, \boldsymbol{v}_z|_{\Gamma_{\text{top}}} = 0, \quad \text{and} \quad \boldsymbol{v}_r|_{r=0} = 0\} \subset [\mathcal{H}^1(\mathcal{F})]^2$ and $Q = \mathcal{L}^2_0(\mathcal{F})$. The reduced Navier–Stokes problem then reads as follows:

Find $(\boldsymbol{u}, p) \in \boldsymbol{V} \times Q$ such that for all $(\boldsymbol{v}, q) \in \boldsymbol{V} \times Q$, it holds

$$A_f(\boldsymbol{u}, p; \boldsymbol{v}, q) := m(\partial_t \boldsymbol{u}, \boldsymbol{v}) + a(\boldsymbol{u}, \boldsymbol{v}) + c(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) + b(\boldsymbol{u}, q) = 0 \tag{12}$$

with the transformed multilinear forms

$$\begin{aligned} m(\boldsymbol{u}, \boldsymbol{v}) &= 2\pi\rho_f \int_{\mathcal{F}} r \boldsymbol{u} \cdot \boldsymbol{v} \mathrm{d}\boldsymbol{x}, \\ a(\boldsymbol{u}, \boldsymbol{v}) &= 2\pi\mu_f \int_{\mathcal{F}} \nabla \boldsymbol{u} : \nabla \boldsymbol{v} + \frac{1}{r} \boldsymbol{u}_r \boldsymbol{v}_r \mathrm{d}\boldsymbol{x}, \\ c(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) &= 2\pi\rho_f \int_{\mathcal{F}} r (\boldsymbol{u} \cdot \nabla) \boldsymbol{v} \cdot \boldsymbol{w} \mathrm{d}\boldsymbol{x}, \\ b(q, \boldsymbol{v}) &= -\int_{\mathcal{F}} q (\boldsymbol{v}_r + r \nabla \cdot \boldsymbol{v}) \mathrm{d}\boldsymbol{x}, \end{aligned}$$

TABLE IV. Results for the Rubber22 setup. The experimental values have been reproduced with permission from T. Hagemeier, Particle settling-transitional regime, version 1, Mendeley Data, September 2020. Copyright 2020 Author(s), licensed under a Creative Commons Attribution 4.0 License.

Discretization					Results					
Method	$[h_{\min}, h_{\max}]$	Δt	dof	nze	t_*	v_*	f_*	t_{cont}	t_{jump}	d_{jump}
ALE	[0.000 4, 0.004]	1/200	6.9	0.31	0.455 904 5	−0.303 002	$1.132\,62 \cdot 10^{-2}$	…	…	…
	[0.000 2, 0.002]	1/800	21.5	0.99	0.455 355 1	−0.303 552	$1.131\,07 \cdot 10^{-2}$	…	…	…
	[0.000 1, 0.001]	1/3200	73.8	3.45	0.455 333 4	−0.303 616	$1.131\,17 \cdot 10^{-2}$	…	…	…
			Extrapolate		0.455 332 5	−0.303 625	…	…	…	…
			order (in h)		4.6	3.1	…	…	…	…
ALE 3D	[0.000 8, 0.032]	1/200	18.43	4.05	0.453 695	−0.304 252	$1.142\,52 \cdot 10^{-2}$	…	…	…
	[0.000 4, 0.016]	1/800	68.52	15.37	0.456 145	−0.302 170	$1.144\,93 \cdot 10^{-2}$	…	…	…
	[0.000 2, 0.008]	1/3200	304.4	69.33	0.455 929	−0.302 667	$1.135\,57 \cdot 10^{-2}$	…	…	…
CutFEM	[0.002 00, 0.008]	1/2000	11.46	0.34	0.453 455	−0.309 542 3	$1.090\,94 \cdot 10^{-2}$	0.524 979	0.098 086	$7.308\,15 \cdot 10^{-3}$
	[0.001 00, 0.004]	1/2000	38.98	1.16	0.454 081	−0.308 841 2	$1.110\,82 \cdot 10^{-2}$	0.525 502	0.110 672	$1.147\,72 \cdot 10^{-2}$
	[0.000 50, 0.002]	1/2000	140.96	4.20	0.453 789	−0.306 220 2	$1.129\,22 \cdot 10^{-2}$	0.526 010	0.121 606	$1.368\,74 \cdot 10^{-2}$
FSI ($E_s = 5 \cdot 10^6$)	[0.001 0, 0.004]	[1/2000, 1/500]	51.4	1.15	0.446 020	−0.320 688 9	…	0.515 197	0.079 14	$2.682\,98 \cdot 10^{-3}$
	[0.000 5, 0.002]	[1/2000, 1/500]	204.6	4.72	0.449 821	−0.311 385 1	…	0.521 487	0.083 323	$3.719\,86 \cdot 10^{-3}$
FSI ($E_s = 2 \cdot 10^6$)	[0.001 0, 0.004]	[1/2000, 1/500]	51.4	1.15	0.446 012	−0.320 715 1	…	0.516 197	0.092 5	$5.292\,47 \cdot 10^{-3}$
	[0.000 5, 0.002]	[1/2000, 1/500]	204.6	4.72	0.449 827	−0.311 375 9	…	0.522 087	0.090 5	$5.501\,46 \cdot 10^{-3}$
Experiment					0.469 137	−0.309 301	…	0.544 021	0.089 492	$4.414\,85 \cdot 10^{-3}$

cf. Ref. 45. Since the motion of the solid is partially driven by the fluid forces acting on it, we also need to transform these. Let $(\boldsymbol{u}^{3d}, p^{3d}) \in \boldsymbol{V}^{3d} \times Q^{3d}$ be a rotationally symmetric solution of (9) and $(\boldsymbol{u}, p) \in \boldsymbol{V} \times Q$ be the solution of (12). We can then transform the weak boundary integral formulations as

$$\int_{\mathcal{I}^{3d}} \boldsymbol{\sigma}^{3d}(\boldsymbol{u}^{3d}, p^{3d})\boldsymbol{n}^{3d} \cdot \boldsymbol{v}^{3d} \mathrm{d}S^{3d} = 2\pi \int_{\mathcal{I}} r(\mu_f \nabla \boldsymbol{u} - \mathrm{Id}p)\boldsymbol{n} \cdot \boldsymbol{v} \mathrm{d}S. \quad (13)$$

The forces $\boldsymbol{F}$ can then be computed by inserting the appropriate non-conforming test-functions $\boldsymbol{v}$ into this functional. Using this, we can compute the motion of the solid as before using (7).

IV. NUMERICAL COMPUTATIONS

We give the details of the different numerical approaches applied to attempt to reproduce the observed data and present the results attained with these methods.

The full result datasets of all methods and the source code to reproduce the rigid-body computations according to the descriptions in Subsections IV A 1 and IV A 2 can be found in the zenodo repository.[63] This also includes the preparatory examples presented in Appendix B.

A. Discretizations

We provide the details on the formulations of the different discretization approaches used.

1. Fluid–rigid body interaction in Arbitrary Lagrangian–Eulerian coordinates

To compute the spatially reduced coupled problem, we formulate the Navier–Stokes equations in Arbitrary Lagrangian–Eulerian (ALE) coordinates[46] by introducing a reference map,

$$T_{ALE}(\boldsymbol{c}_S) : \mathcal{F} \to \mathcal{F}(\boldsymbol{c}_S),$$

and by transforming the Navier–Stokes equation onto the reference domain $\mathcal{F}$, which is fixed for all times. This setting allows for a direct finite-element triangulation Ω_h of $\mathcal{F}$ that resolves the interface between fluid and solid well. ALE approaches are highly efficient and accurate and well established for fluid–solid interaction problems. ALE approaches will however fail if the deformation becomes too large[47] or if even contact of the solid with an outer boundary happens,[14] such as it is the case in our context. We refer to the literature on ALE[12,46] and to Appendix A 1 for details on our implementation.

2. Fluid–rigid body interaction in Eulerian coordinates

As in Subsection IV A 1, we consider the problem as a moving domain problem for the fluid, assume the solid to be a rigid body, and decouple the fluid and solid equations. For the resulting moving domain problem, we use an unfitted Eulerian finite-element method from Ref. 23 using BDF2 time stepping, together with Taylor–Hood elements in space, which are inf–sup stable in the CutFEM setting[48] with ghost-penalty stabilization.[49] In order to obtain the full convergence order of the Taylor–Hood finite-element pair, we use an isoparametric mapping introduced in Ref. 50 for stationary domains to realize the higher-order geometry approximation.

a. Transformed nitsche terms. In the CutFEM method used here, Dirichlet boundary conditions on unfitted boundaries are enforced using Nitsche's method.[51] For a consistent and stable method, we also transform these terms into the rotationally symmetric formulation. Using the standard derivation for Nitsche's method, we find that the consistency and penalty terms for the reduced formulation are

$$n_c(\boldsymbol{u}, \boldsymbol{v}) = -2\pi\mu_f \int_\Gamma r(\nabla \boldsymbol{u})\boldsymbol{n} \cdot \boldsymbol{v} \mathrm{d}S$$

and

$$n_s(\boldsymbol{u}, \boldsymbol{v}) = 2\pi\mu_f \sigma \frac{k^2}{h} \int_\Gamma r\boldsymbol{u} \cdot \boldsymbol{v} \mathrm{d}S,$$

respectively, with a penalty parameter $\sigma > 0$, the velocity space's polynomial order k, and the local mesh diameter h. Similarly, we find for the pressure-coupling operator that the Nitsche term is

$$n_p(\boldsymbol{v}, q) = \int_\Gamma rq\boldsymbol{v} \cdot \boldsymbol{n} \mathrm{d}S.$$

The necessary Nitsche term for a symmetric and consistent formulation to enforce the Dirichlet conditions $\boldsymbol{u} = \boldsymbol{g}$ is therefore

$$n(\boldsymbol{u}, p; \boldsymbol{v}, q) := n_c(\boldsymbol{u}, \boldsymbol{v}) + n_c(\boldsymbol{v} - \boldsymbol{g}, \boldsymbol{u}) + n_s(\boldsymbol{u} - \boldsymbol{g}, \boldsymbol{v}) + n_p(\boldsymbol{v}, p) + n_p(\boldsymbol{u} - \boldsymbol{g}, q).$$

b. Transformed ghost-penalty operators. The role of the ghost-penalty operator in the unfitted Eulerian time-stepping scheme used here is twofold. As in other CutFEM discretizations, it stabilizes arbitrary element cuts such that the method is stable and the resulting matrices are well conditioned.[49] Second, appropriately scaled ghosty-penalties provide the necessary implicit extension for the method-of-lines approach to the discretization of the time-derivative.[22–24]

We use the *direct-version* of the ghost-penalty operator.[52] To define this operator, let $\mathscr{F}_h$ be a set of facets between neighboring elements on which the ghost-penalty operator is to act on, cf. Refs. 23 and 24 for further details. For a facet $F \in \mathscr{F}_h$ such that $F = \overline{T}_1 \cap \overline{T}_2$, we define the facet patch $\omega_F = T_1 \cup T_2$. The velocity ghost-penalty operator for the rotationally symmetric formulation is then

$$i_h(\boldsymbol{u}, \boldsymbol{v}) = \gamma_{\boldsymbol{u}} \sum_{F \in \mathscr{F}_h} \int_{\omega_F} \frac{r}{h^2} (\boldsymbol{u}_1 - \boldsymbol{u}_2) \cdot (\boldsymbol{v}_1 - \boldsymbol{v}_2) + \frac{1}{r} (\boldsymbol{u}_{r,1} - \boldsymbol{u}_{r,2})(\boldsymbol{v}_{r,1} - \boldsymbol{v}_{r,2}) \mathrm{d}\boldsymbol{x},$$

where $\boldsymbol{v}_i = \mathcal{E}^{\mathbb{P}} \boldsymbol{v}|_{T_i}$ with the canonical extension of polynomials $\mathcal{E}^{\mathbb{P}} : \mathbb{P}T \to \mathbb{P}\mathbb{R}^d$. The pressure ghost-penalty operator is

$$j_h(p, q) = \gamma_p \sum_{F \in \mathscr{F}_h} \int_{\omega_F} r(p_1 - p_2)(q_1 - q_2) \mathrm{d}\boldsymbol{x}.$$

With these transformed ghost-penalty operators, it is easy to show the standard ghost-penalty results in the appropriately transformed norms.

The variational formulation of the CutFEM discretization then reads as follows:

Find $(\boldsymbol{u}, p) \in \boldsymbol{V}_h \times Q_h$ such that for all $(\boldsymbol{v}, q) \in \boldsymbol{V}_h \times Q_h$, it holds

$$A_f(\boldsymbol{u}, p; \boldsymbol{v}, q) + n(\boldsymbol{u}, p; \boldsymbol{v}, q) + \left(\mu_f + \frac{1}{\mu_f}\right) i_h(\boldsymbol{u}, \boldsymbol{v}) + \frac{1}{\mu_f} j_h(p, q) = 0.$$

c. Contact algorithm. We consider a very basic contact avoidance scheme, used widely in the literature.[18,32,53,54] The idea is to introduce an artificial (lubrication) force acting on the rigid body in the vicinity of the contact wall, which increases when the ball comes closer to the wall and acts in the direction away from the wall. This force is then added to the forces governing the motion of the rigid solid such that contact does not occur. We define this force as

$$f_c(\mathcal{S}) = \begin{cases} 0 & \text{if } \operatorname{dist}(\mathcal{I}, \Gamma_{\text{bottom}}) \geq dist_0 \\ \gamma_c \dfrac{dist_0 - \operatorname{dist}(\mathcal{I}, \Gamma_{\text{bottom}})}{\operatorname{dist}(\mathcal{I}, \Gamma_{\text{bottom}})} & \text{if } \operatorname{dist}(\mathcal{I}, \Gamma_{\text{bottom}}) < dist_0, \end{cases}$$

where $dist_0$ and γ_c are parameters to be chosen and $\operatorname{dist}(\mathcal{I}, \Gamma_{\text{bottom}})$ is the minimal distance between $\mathcal{I}$ and Γ_{bottom}. We then add this to the right-hand side of the ODE (4) and carry this thought so that the right-hand side of (7) becomes $\frac{\rho_S - \rho_F}{\rho_S} g + \frac{F_y + f_c}{\operatorname{vol}(\mathcal{S})\rho_S}$. The solution to this ODE then governs the position on the level set describing the solid and the Dirichlet condition enforced on the interface.

d. Implementation. This discretization is implemented using the finite-element library `Netgen/NGSolve` (see Refs. 55 and 56 and ngsolve.org) together with the add-on package `ngsxfem`[57] for unfitted finite-element functionality.

The background mesh is constructed by defining a local mesh parameter h_{inner} on the left of the reduced domain where $r < 2d_S/3$ and then creating a shape regular mesh with $h = h_{\max}$ in the remainder of the domain. This is to obtain more accurate boundary integrals, i.e., when computing the forces acting on the ball. In the Rubber22 setting, we choose $h_{\text{inner}} = h_{\max}/4$ and in the PTFE6 case as $h_{\text{inner}} = h_{\max}/7$. This is to obtain background meshes with a similar number of elements in each settings. On the active part of the mesh, we consider standard $\mathbb{P}^2/\mathbb{P}^1$ Taylor–Hood elements.

The Nitsche parameter is taken as $\sigma = 100$, the extension ghost-penalty parameter is $\gamma_{\boldsymbol{u},e} = 0.1$, the cut-stability ghost-penalty parameter $\gamma_{\boldsymbol{u},s} = \gamma_{p,s} = 0.01$, and the extension strip width parameter is $c_\delta = 4$. See Ref. 23 for details on these parameters.

The contact parameters are tuned with respect to the PTFE6 jump height since the model cannot be expected to resolve the elastic nature of rubber. For PTFE6, we take the contact model parameters $dist_0 = 2 \cdot 10^{-5}$ and $\gamma_c = 0.38$. Since the mass of the Rubber22 ball is ~31.6 times larger than the PTFE6 ball, we take contact model parameters that are appropriately larger such that the resulting acceleration acting on the balls is comparable. For the Rubber22 computations, we take $dist_0 = 2 \cdot 10^{-5}$ and $\gamma_c = 12$.

Each time step is iterated between the fluid system and the solid ODE until the system is solved. We consider the system as solved when the update of the ball velocity in an iteration is less than 10^{-8}.

3. Fluid–structure interaction in fully Eulerian coordinates

We consider the full fluid–structure interaction problem including contact with Γ_{bottom}. We adopt here a fully Eulerian approach for the FSI system in order to enable the transition to contact.[31,32,34] To allow for an implicit inclusion of the contact conditions into the system (see below), we use a Nitsche-based method for the FSI coupling as presented in Ref. 39.

a. Solid bilinear form and FSI coupling. The fluid bilinear form has already been detailed in (12). To introduce the solid form, we denote the reduced solid domain by $\mathcal{S}(t)$ and define the bilinear form corresponding to the linear elasticity equations in Eulerian coordinates (2) by

$$\begin{aligned} A_s(\boldsymbol{d}, \dot{\boldsymbol{d}}; \boldsymbol{w}, \boldsymbol{z}) := m_s(\partial_t \dot{\boldsymbol{d}} - \dot{\boldsymbol{d}} \cdot \nabla \dot{\boldsymbol{d}}, \boldsymbol{w}) + a_s(\boldsymbol{d}, \boldsymbol{w}) \\ + m_{\dot{\boldsymbol{d}}}(\partial_t \boldsymbol{d} - \dot{\boldsymbol{d}} \cdot \nabla \boldsymbol{d} + \dot{\boldsymbol{d}}, \boldsymbol{z}), \end{aligned}$$

where

$$\begin{aligned} m_s(\boldsymbol{d}, \boldsymbol{w}) &:= 2\pi\rho_s \int_{\mathcal{S}} r \boldsymbol{d} \cdot \boldsymbol{w} \ \mathrm{d}\boldsymbol{x}, \\ a_s(\boldsymbol{d}, \boldsymbol{w}) &:= 2\pi \int_{\mathcal{S}} r \nabla \sigma_s(d) : \nabla \boldsymbol{w} + \sigma_{s,r} \boldsymbol{w}_r \ \mathrm{d}\boldsymbol{x}, \\ m_{\dot{d}}(\boldsymbol{d}, \boldsymbol{w}) &:= 2\pi \int_{\mathcal{S}} \boldsymbol{d} \cdot \boldsymbol{w} \quad \mathrm{d}\boldsymbol{x}, \end{aligned}$$

and

$$\begin{aligned} \sigma_s &= 2\mu_s E(\boldsymbol{d}) + \lambda_s \left(\operatorname{tr}(E(\boldsymbol{d})) + \frac{1}{r} \boldsymbol{d}_r\right) I, \\ \sigma_{s,r} &= \frac{2\mu_s + \lambda_s}{r} \boldsymbol{d}_r + \lambda_s \operatorname{tr}(E(\boldsymbol{d})). \end{aligned}$$

Moreover, we make use of the Nitsche terms defined in Subsection IV A 2 to impose the FSI coupling conditions (3),

$$\begin{aligned} n(\boldsymbol{u}, p, \dot{\boldsymbol{d}}; \boldsymbol{v}, q, \boldsymbol{w}) := n_s(\boldsymbol{u} - \dot{\boldsymbol{d}}, \boldsymbol{v} - \boldsymbol{w}) + n_c(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{w}) \\ + n_c(\boldsymbol{v}, \boldsymbol{u} - \dot{\boldsymbol{d}}) + n_p(\boldsymbol{v} - \boldsymbol{w}, p) - n_p(\boldsymbol{u} - \dot{d}, q). \end{aligned}$$

Note the negative sign in front of the last term, which is required to ensure the stability of the FSI formulation; see Ref. 25.

b. Discretization and stabilization. In order to resolve the interface $\mathcal{I}$ within the discretization, we use the locally modified finite-element method introduced in Ref. 58. This *fitted* finite-element method is based on a coarse *unfitted* patch mesh, which is independent of the interface location. The coarse cells are then divided in such a way into sub-triangles and sub-quadrilaterals that the interface is resolved in a linear approximation. In this work, we use equal-order locally modified finite elements of first order, in combination with an anisotropic edge-oriented pressure stabilization term $s_p(p, q)$; see Ref. 59 for details. We denote the locally modified finite-element space of first order by X_h^1. The discrete spaces for fluid velocity and pressure $\boldsymbol{V}_h$ and Q_h and for solid displacement and velocity $\boldsymbol{W}_h$ and Z_h are given by applying the respective Dirichlet conditions to the locally modified finite-element space X_h^1 and by restricting the degrees of freedom to the fluid or solid sub-domain, respectively.

In addition, we add a stabilization term $s_{\boldsymbol{d}}(\dot{d}, z)$ of artificial diffusion type to the solid equations (see Ref. 32) and a consistent stabilization at the boundary Γ_{sym} corresponding to the rotational axis to

ensure that the second boundary condition (11) is accurately imposed in the discrete formulation,

$$s_d(\dot{d}, \boldsymbol{z}) = \alpha_d h^2 (r\nabla\dot{d}, \nabla\boldsymbol{z})_{\mathcal{S}},$$
$$s_r(\boldsymbol{u}, \boldsymbol{v}) := \alpha_{\text{sym}} \rho_f \mu_f \int_{\Gamma_{\text{sym}}} h_n^2 h_\tau \partial_r \boldsymbol{u}_z \partial_r \mathrm{d}S.$$

Here, h_n and h_τ refer to the cell-sizes in the normal and tangential directions, respectively. We summarize the stabilization terms in the bilinear form,

$$s(\boldsymbol{u}, p, \dot{d}; \boldsymbol{v}, q, \boldsymbol{z}) = s_p(p, q) + s_d(\dot{d}, \boldsymbol{z}) + s_r(\boldsymbol{u}, \boldsymbol{v}).$$

For time discretization, we use the modified dG(0) time discretization presented in Ref. 60, which can be seen as a variant of the dG(0)/backward Euler methods that considers the movement of the interface in each space–time slab. The interface positions and the domain affiliations are updated explicitly based on the displacement $d(t_{n-1})$ of the previous time step and using the initial point set/backward characteristics method.[31,32] This means that we set

$$\mathcal{I}^n := \mathcal{I}(d(t_{n-1})), \quad \mathcal{S}^n := \mathcal{S}(d(t_{n-1})), \quad \mathcal{F}^n := \mathcal{F}(d(t_{n-1})).$$

c. Contact treatment. When a part $\mathcal{I}_c := \mathcal{I} \cap \Gamma_{\text{bottom}}$ of $\mathcal{I}$ enters into contact with Γ_{bottom}, the FSI conditions need to be substituted with appropriate contact conditions. It has been noted in Refs. 39 and 40 that although the fluid layer between the ball and the lower wall vanishes (from a macroscopical perspective), an extension of the fluid forces to the contact surface $\mathcal{I}_c$ has to be considered to obtain a physically relevant contact formulation. Here, we use the simplest possible numerical approach, which is to relax the no-penetration condition by a small $\epsilon = \epsilon(h) > 0$ such that a very thin mesh-dependent fluid layer remains at all times.

The distance to Γ_{bottom} depends on the (Eulerian) displacement $\boldsymbol{d}(t_n)$ in the following way:

$$\begin{aligned}\text{dist}(\mathcal{I}(t_n), \Gamma_{\text{bottom}}) = \text{dist}(\mathcal{I}(t_{n-1}), \Gamma_{\text{bottom}}) \\ + \boldsymbol{d}_n(t_n) - \boldsymbol{d}_n(t_{n-1}).\end{aligned}$$

The contact conditions, relaxed by a small $\epsilon = \epsilon(h) > 0$, read as

$$\begin{aligned}\text{dist}(\mathcal{I}(t_n), \Gamma_{\text{bottom}}) \le \epsilon, \quad [\![\sigma_n]\!] \le 0, \\ [\![\sigma_n]\!](\text{dist}(\mathcal{I}(t_n), \Gamma_{\text{bottom}}) - \epsilon) = 0 \quad \text{on } \mathcal{I}^n = \mathcal{I}(t_{n-1}),\end{aligned} \tag{14}$$

where

$$[\![\sigma_n]\!] := n^T \sigma_s n - n^T \sigma_f n.$$

In other words, the relaxation means that the contact conditions are already applied at an ϵ-distance from Γ_{bottom}. The three conditions (14) can be equivalently formulated in equality form with an arbitrary $\gamma_C > 0$,[61]

$$[\![\sigma_n]\!] = -\gamma_C \underbrace{[\text{dist}(\mathcal{I}(t_n), \Gamma_{\text{bottom}}) - \epsilon - \gamma_C^{-1}[\![\sigma_n]\!]]_+}_{=:P_\gamma(\boldsymbol{d}, [\![\sigma_n]\!])} \quad \text{on } \mathcal{I}^n, \tag{15}$$

where $[\cdot]_+$ stands for the positive part of $(\cdot)$. The contact parameter γ_C will be chosen as $\gamma_C = \gamma_C^0 \lambda_s h^{-1}$ and the relaxation parameter will be chosen as $\epsilon = \epsilon_0 h_y$, where h_y denotes the cell size in the vertical direction at the bottom of the cylinder.

Note that (15) includes both the FSI coupling and the contact condition as in the absence of contact, it is exactly the FSI interface condition in the normal direction. For this reason, the transition between FSI coupling and contact conditions can be included easily in a fully implicit fashion in the variational formulation.

The final variational formulation reads as follows:

$$\begin{aligned}&\text{Find } (\boldsymbol{u}, p, \boldsymbol{d}, \dot{\boldsymbol{d}}) \in \boldsymbol{V}_h \times Q_h \times \boldsymbol{W}_h \times \boldsymbol{Z}_h \text{ such that for all} \\ &\quad (\boldsymbol{v}, q, \boldsymbol{w}, \boldsymbol{z}) \in \boldsymbol{V}_h \times Q_h \times \boldsymbol{W}_h \times \boldsymbol{Z}_h, \text{ it holds} \\ &A_f(\boldsymbol{u}, p; \boldsymbol{v}, q) + A_s(\boldsymbol{d}, \dot{\boldsymbol{d}}; \boldsymbol{w}, \boldsymbol{z}) + n(\boldsymbol{u}, p, \dot{\boldsymbol{d}}; \boldsymbol{v}, q, \boldsymbol{w}) \\ &\quad + s(\boldsymbol{u}, p, \dot{d}; \boldsymbol{v}, q, \boldsymbol{z}) + \left(r[P_\gamma(\boldsymbol{d}, [\![\sigma_n]\!])]_+, \boldsymbol{w}_n\right)_{\mathcal{I}} = (r\boldsymbol{f}_s, \boldsymbol{w})_\Omega.\end{aligned}$$

d. Implementation. The described algorithms and equations have been implemented in the finite-element library `Gascoigne3D`.[62] We use a Cartesian finite-element mesh, which is highly refined in the region where contact occurs.

Concerning time discretization, we start with a relatively coarse time step $\Delta t = 2 \cdot 10^{-3}$, which captures the essential dynamics of the case process. When the ball gets close to the lower wall, the time step is reduced in order to capture the contact dynamics and, in particular, to resolve the impact time accurately. We do this by reducing the time step by a factor of two each time the distance to the wall drops below certain thresholds d_i, $i = 0, \ldots, m$. In the Rubber22 case, we choose, for example, $d_0 = 10^{-2}$ m and $d_1 = 10^{-3}$ m. For the PTFE6 case, the contact interval is much shorter (around 10^{-4} s compared to $4 \cdot 10^{-3}$ s for rubber). For this reason, we specify seven thresholds to reduce the time step in seven steps until $\Delta t < 10^{-5}$ s.

In addition, we use the following numerical parameters:

$$\gamma_C^0 = 1, \quad \sigma = 10^5, \quad \alpha_d = 1, \quad \alpha_{\text{sym}} = 10^3.$$

The contact relaxation parameter is chosen as $\epsilon_0 = \frac{1}{8}$ for the Rubber22 test case and $\epsilon_0 = \frac{1}{4}$ for PTFE6. For a detailed sensitivity study of the influence of the contact parameters, we refer to Ref. 39.

B. Results

In the following, we shall abbreviate the methods described in Subsection IV A 1 as *ALE*, in Subsection IV A 2 as *CutFEM*, and in Subsection IV A 3 as *FSI*.

Remark. In order to indicate the computational effort needed to solve the (linearized) systems resulting from our computations, we give the size and sparsity of these with our quantitative results. The number fo free degrees of freedom is abbreviated as `dof`, and the number of non-zero entries in the system is denoted by `nze`.

In the CutFEM method, dof and nze can vary between time steps since different elements are active in different time steps. We therefore state the number of unconstrained degrees of freedom and non-zero entries of the linearized system in the first time step.

1. PTFE6

Our quantitative results for the PTFE6 setup are presented in Table III. Figure 3 shows the distance between the bottom of the ball to the bottom of the fluid domain over time from the experimental data and all three numerical methods.

If we look at the pre-contact quantities of interest in Table III, we see that all methods give very similar results for the given quantities. Looking at the velocity of the PTFE6 ball, we see that the numerical values are within a relative error of 5.1% of the experiment on the finest discretizations. Taking into account that these results ignore the 2 mm deflection from the z axis observed in the experimental data, we consider this to be acceptable. Since the elastic effects of the particle appear to be negligible in this phase of the problem and due to the known good approximation properties of the ALE method, we consider these to be the most accurate values for future comparison.

Looking at the quantities of interest in the later phase, we see that in both the CutFEM and FSI methods, contact occurs later than in the experiment. This is consistent with the smaller speed of the particle compared to the experiments as observed above. With respect to the jump, we see that both methods capture the rebound dynamics since both the point in time at which the peak of the rebound is realized and the size of the jump are consistent with the experiment. As the CutFEM contact parameters were tuned with respect to the size of this jump, this is unsurprising. However, since the contact force only acts for a very small number of time steps, the fact that the time at which the rebound is maximal is also captured well shows that even after contact, the system is still approximated well by the fluid–rigid body system. Nevertheless, it is clear that the FSI system captures the dynamics much more accurately and without the need of essentially unknowable and artificial parameters $dist_0$ and γ_c in the contact model.

2. Rubber22

Our results for the Rubber22 problem are presented in Table IV. The height of the ball over time can be seen in Fig. 4. To illustrate the applicability of our spatially reduced formulation, we solve the fluid–rigid body system in full three spatial dimensions using the ALE approach. The ALE approach has been chosen for this due to its significantly higher computational performance compared to the CutFEM and FSI methods. These results are also given in Table IV. To illustrate the flow solution, we show the result of the CutFEM simulation on the coarsest mesh, rotated into the x–z plane at $t = 0.575$ in Fig. 5.

We again start by inspecting the pre-contact results in Table IV. We observe that t_*, v_*, and f_* are again very similar for all methods. However, here, the discrepancy between the numeric and experimental velocity values is significantly smaller, with the relative difference being between 2.1% and 0.6% on the finest discretizations. We note at this point that the deviation from the z axis in this experiment was less than 0.75 mm. This shows that even for a more elastic material, a fluid–rigid body system can capture the pre-contact dynamics as well as a full FSI model. Furthermore, we see that with perfect initial data and spatial symmetry, our spatially reduced model, derived under the assumption of a rotationally symmetric solution, captures the dynamics as well as the significantly more computationally expensive full three-dimensional computation.

Looking at the numbers for the contact and rebound dynamics, we again see that the time of contact is similar for both the CutFEM and FSI methods, and this matches the experimental time with an error of less than 5%. For the CutFEM method, we clearly see that the PTFE tuned parameters are not able to capture the rebound dynamics well. In fact, the size of the rebound is approximately three times larger than the physical rebound. This shows that the use of an artificial lubrication force, as considered in a variety of other literature, can lead

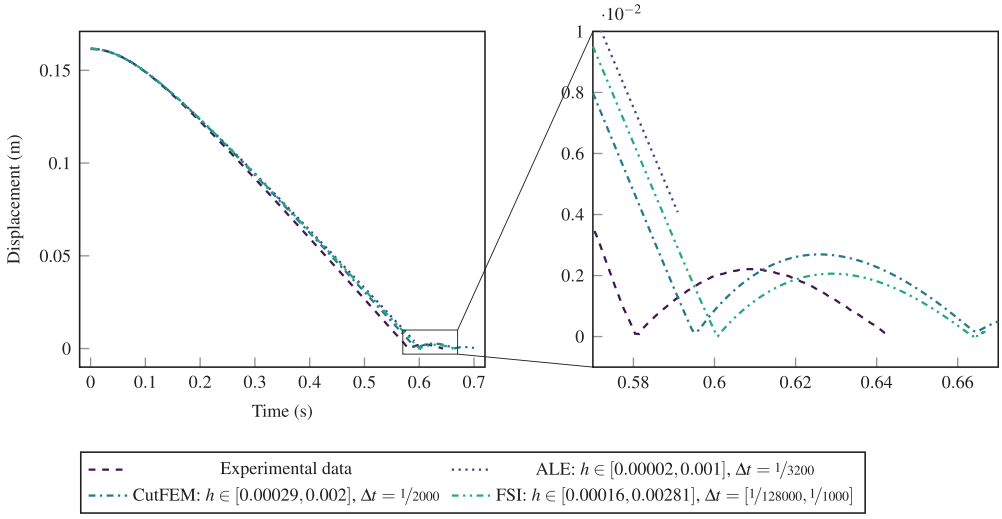


FIG. 3. The distance between the bottom of the ball and the bottom of the tank: Experimental and numerical results for the PFTE6 setup. The experimental data are reproduced with permission from T. Hagemeier, Particle settling-transitional regime, version 1, Mendeley Data, September 2020. Copyright 2020 Author(s), licensed under a Creative Commons Attribution 4.0 License.

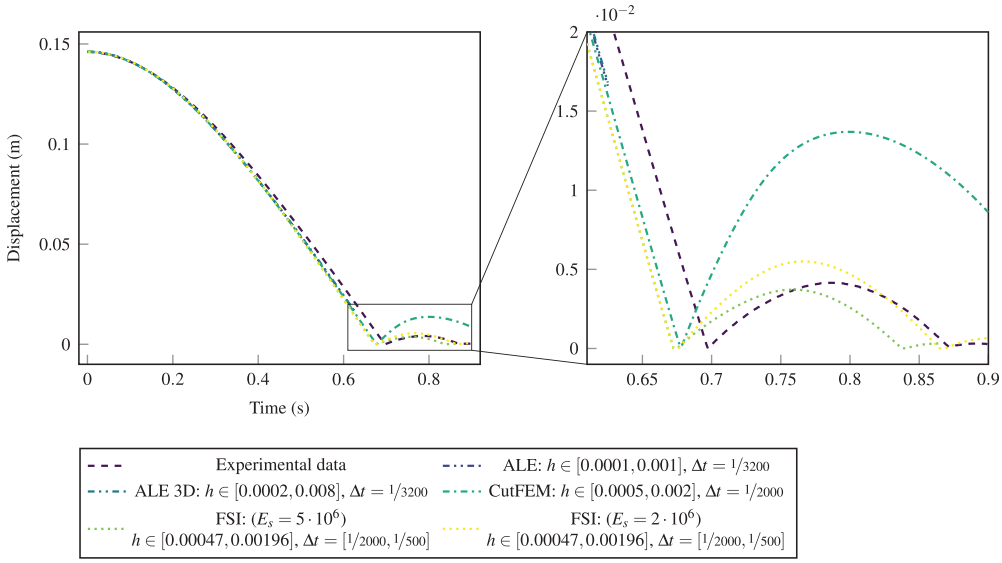


FIG. 4. The distance between the bottom of the ball and the bottom of the tank: experimental and numerical results for the Rubber22 setup. The experimental data are reproduced with permission from T. Hagemeier, Particle settling-transitional regime, version 1, Mendeley Data, September 2020. Copyright 2020 Author(s), licensed under a Creative Commons Attribution 4.0 License.

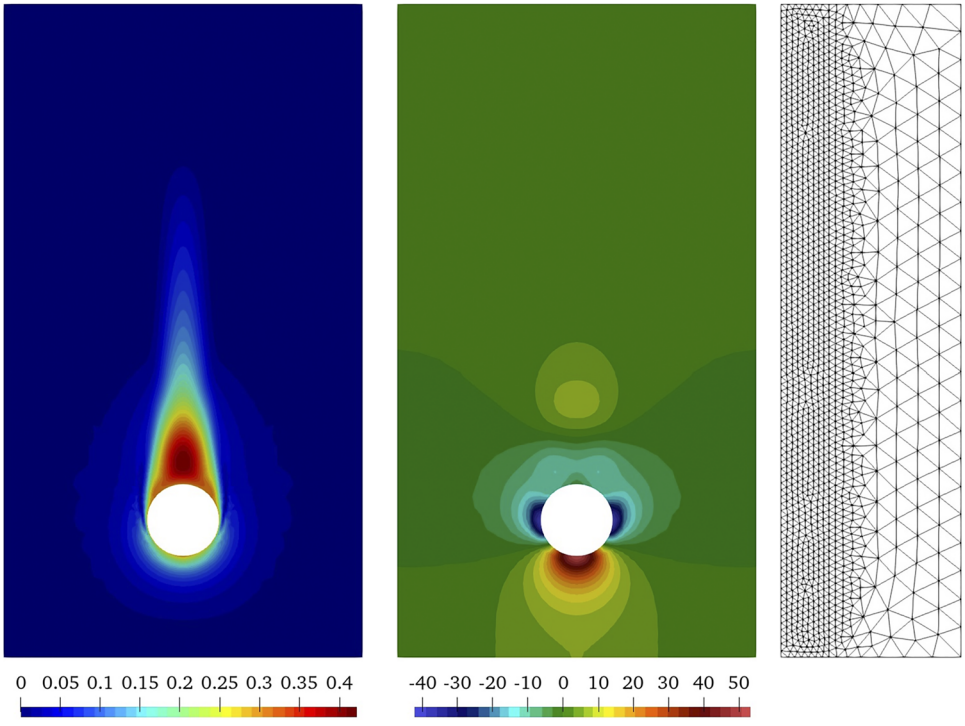


FIG. 5. Velocity solution (left) and pressure solution (center) at $t = 0.575$ s rotated into the x–z plane and computational mesh (right) for the Rubber22 case, resulting from the CutFEM computation with $h_{\max} = 0.008$ and $\Delta t = 0.005$ s.

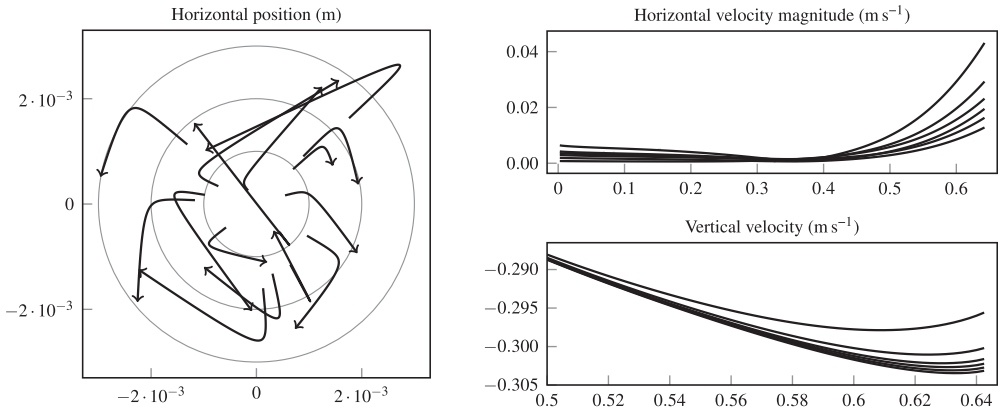


FIG. 6. Left: view from below. Deflection of the center of mass from the centerline (0, 0) for 15 experiments starting with random initial deviations each. Right: velocity of the particles for six experiments. The upper figure shows the horizontal velocity component $\sqrt{v_{S,1}^2 + v_{S,2}^2}$, and the lower plot gives the vertical velocity close to the bottom for $t \geq 0.5$ s.

to physically meaningful results but is heavily dependent on the "correct" choice of parameters for which there is no *a priori* knowledge. For the FSI model, we see that the overall dynamics are captured well. We also observe that—while the pre-contact dynamics are essentially independent of the elasticity parameters—a variation of the elasticity modulus E_s changes the rebound height d_{jump} significantly. Here, a softer material ($E_s = 2 \cdot 10^6$ Pa) leads to a larger rebound as more elastic energy is taken up through the deformation during the impact.

3. Three-dimensional computation including rotational effects

One of the challenges in performing the experimental study[42] was to limit the horizontal deflection of the falling objects, i.e., to keep them close to the centerline. Identifying the source of these three-dimensional effects is one of the intriguing questions for future research. The cause may be found in a complex solution pattern of the Navier–Stokes equations, in material inaccuracies such as non-uniform distribution of the mass or the surface roughness but also in experimental inaccuracies, e.g., during the release process.

Since the reduced formulation presented in Sec. III cannot depict 3D effects, we add further studies based on the full 3D formulation. Here, we choose the ALE representation again due to its much better computational performance, even if no contact can be modeled. However, this is not a significant limitation since effects of rotation and horizontal deflection happen during the phase of free fall. The three-dimensional simulations based on the ALE formulation presented above in Subsection IV B 2 did not show any three-dimensional effect if the configuration is fully symmetric and the ball is released at the centerline, i.e., $\boldsymbol{c}_S = (0, 0, d(0))$. To investigate the stability of the Navier–Stokes rigid-body system, we consider further numerical simulations based on the Rubber22 case with distorted initial values. We start the simulation with the initial data

$$\begin{aligned}\boldsymbol{c}_S &= \left(10^{-3}\chi_x, 10^{-3}\chi_y, 0.1461203\right)\text{m},\\ \boldsymbol{v}_S(0) &= \left(4 \cdot 10^{-3}\chi_x, 4 \cdot 10^{-3}\chi_y, 0\right)\text{ms}^{-1},\\ \boldsymbol{\omega}_S(0) &= (\chi_1, \chi_2, \chi_2)\ 2\pi \cdot 10^{-2}\ \text{s}^{-1},\end{aligned}$$

where $\chi_x, \chi_y, \chi_1, \chi_2, \chi_3 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ are normally distributed random numbers with mean zero and standard deviation 1.

Figure 6 shows the results for multiple experiments based on randomly chosen initial data. The left plot shows the projection of the center of mass onto the x–y plane. These results show that numerical simulations cannot predict a substantial deflection from the centerline if an initial deflection is prescribed. While the rigid solids are indeed further removed from the center, the effect is small and the objects remain within 3 mm off the center. On the right, we show the velocity of the particles. The upper figure shows the horizontal velocity component, while the lower plot gives the dominant vertical velocity. Here, we indeed see a substantial impact of the initial disturbances. When the solid comes close to the lower boundary, a deflection to the sides becomes visible. We note that these ALE simulations crash before contact is established. At final time $t \approx 0.64$ s, this distance between the lower boundaries is still slightly larger than radius $r_S = 0.011$ m. Since the horizontal velocity is beginning to increase significantly here, further simulations, in which the particle is able to get closer to the bottom boundary, are of interest. We deduce from these results that small fluctuations during the release process can indeed explain the small horizontal displacement observed during the experiments for the PTFE and rubber particles and the discrepancy between the experimental and numerical realizations of our quantities of interest. However, these computations also show that very large horizontal displacements, as observed in Refs. 41 and 42, for example, for the POM particle, cannot be solely explained through this.

V. SUMMARY AND CONCLUSION

We have presented two setups for a fluid–structure interaction problem with solid contact based on the physical experiments

described in Ref. 42 and the data available in Ref. 41. We computed these setups using a spatially reduced model under the assumption of rotational symmetry in cylindrical coordinates. For the discretization, we used a fitted ALE and unfitted CutFEM approach within a fluid–rigid body model and a fully Eulerian FSI approach in a fluid–elastic structure model capable of resolving the solid contact.

We showed how each of these discretizations is able to capture the pre-contact dynamics observed in the physical experiment within a margin of 5.1%–0.6%, even though this ignored any horizontal motion observed in the experimental data. Using a full three-dimensional ALE discretization, we saw that the spatially reduced approach did indeed result in meaningful results under the assumption of perfect initial conditions at a fraction of the computational cost. Furthermore, we presented computation with disturbed initial data. From these results, we deduced that the observed horizontal motion in the PTFE and rubber experiments is within the scope explainable by imperfect starting conditions. This shows that a fluid–rigid body model is suitable for this type of problem before solid contact occurs.

With respect to the contact dynamics, we were able to show that the Eulerian FSI discretization with contact treatment is able to reproduce the spatial and temporal dynamics observed in the experiments very well. This is even though the theoretical modeling of such contact dynamics is not yet fully understood. The moving domain CutFEM approach together with the artificial contact treatment showed that this type of contact treatment can result in physically meaningful results when the artificial parameters are chosen "correctly" and the extent to which the artificial parameters are material dependent.

The resulting datasets, the source code for the fluid–rigid body discretizations, as well as some simplified examples are available in the zenodo repository.[63]

We conclude that the two discussed setups are well suited for the validation of fluid–structure interaction models in the moderate Reynolds-number fluid regime both before and after contact. To the best of our knowledge, this is the first example of a computational FSI setup with rebound contact dynamics, which is validated by experimental data. We note that the PTFE6 setup is better suited for validation of contact and rebound models since the model parameters are known precisely. On the other hand, the Rubber22 scenario is well suited to validate models before or without contact since there is less deviation from the centerline in the data from the experiment.

For future research, it remains to be investigated that to what extent the free surface at the top of the fluid domain plays a role in system dynamics. Furthermore, an open question is that of the role of imperfections in the mass distribution within the solid, of the surface roughness of the solid, and whether these can explain larger horizontal displacements and rotation observed, for example, with the POM particle in Ref. 42.

ACKNOWLEDGMENTS

H.v.W. and T.R. acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant No. 314838170, GRK 2297 MathCoRe. T.R. further acknowledges support from the Federal Ministry of Education and Research of Germany (Project No. 05M16NMA).

APPENDIX A: DETAILS ON THE NUMERICAL REALIZATION

1. Fluid–rigid body interaction in arbitrary Lagrangian–Eulerian coordinates

The ALE formulation is based on the reference map

$$T_{ALE}(\boldsymbol{c}_S) : \mathcal{F} \to \mathcal{F}(\boldsymbol{c}_S),$$

where $\boldsymbol{c}_S$ is the solid's center of mass relative to the bottom boundary and where

$$\mathcal{F}(\boldsymbol{c}_S) = \big\{[0, 0.055\,\mathrm{m}] \times [0, 0.2\,\mathrm{m}]\big\} \backslash B_{r_S}(\boldsymbol{c}_S),$$

$B_{r_S}(\boldsymbol{c}_S)$ being the open ball of radius r with midpoint $\boldsymbol{c}_S$. As the reference domain, we set $\mathcal{F} := \mathcal{F}(0.05\,\mathrm{m})$ where the ball is centered at $\boldsymbol{c}_S = (0, 0, 0.5)$ such that the mesh distortion is not too extreme when the solid comes close to the lower boundary. The mapping $T_{ALE}(\boldsymbol{c}_S)$ is given analytically by

$$T_{ALE}(\boldsymbol{c}_S; r, z) = (r, z + (\boldsymbol{c}_S - 0.05\,\mathrm{m}) f_{ALE}(z)),$$

$$f_{ALE}(z) = \begin{cases} \dfrac{z}{0.05\,\mathrm{m} - 2r_S}, & z < 0.05\,\mathrm{m} - 2r_S \\ 1, & 0.05\,\mathrm{m} - 2r_S \le z \le 0.05\,\mathrm{m} + 2r_S \\ \dfrac{0.2\,\mathrm{m} - z}{0.15\,\mathrm{m} + 2r_S}, & z > 0.05\,\mathrm{m} + 2r_S \end{cases} \tag{A1}$$

such that $T_{ALE}(\boldsymbol{c}_S; r, z)$ is a pure translation for all $z \in [0.05\,\mathrm{m} - 2r_S, 0.05\,\mathrm{m} + 2r_S]$, where the reference ball is located. We make sure that the lines at $z = 0.05\,\mathrm{m} \pm 2r_S$ are resolved by the computational mesh such that T_{ALE} is differentiable within all elements. This construction of the domain map does not allow us to reduce the distance of the ball from the lower boundary to less than r_S.

The reference domain $\mathcal{F}$ is the basis for the finite-element discretization. In cylindrical coordinates, the ALE version of the variational formulation takes the form

$$m_{ALE}(\boldsymbol{c}_S; \boldsymbol{u}, \boldsymbol{v}) = 2\pi\rho_f \int_{\mathcal{F}} r J(\boldsymbol{c}_S) \boldsymbol{u} \cdot \boldsymbol{v} \mathrm{d}\boldsymbol{x},$$

$$a_{ALE}(\boldsymbol{c}_S; \boldsymbol{u}, \boldsymbol{v}) = 2\pi\mu_f \int_{\mathcal{F}} J(\boldsymbol{c}_S) \nabla \boldsymbol{u} \begin{pmatrix} 1 & 0 \\ 0 & J(\boldsymbol{c}_S)^{-1} \end{pmatrix} : \nabla \boldsymbol{v} \times \begin{pmatrix} 1 & 0 \\ 0 & J(\boldsymbol{c}_S)^{-1} \end{pmatrix} + \frac{J(\boldsymbol{c}_S)}{r} \boldsymbol{u}_r \boldsymbol{v}_r \mathrm{d}\boldsymbol{x},$$

$$c_{ALE}(\boldsymbol{c}_S; \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = 2\pi\rho_f \int_{\mathcal{F}} r J(\boldsymbol{c}_S) \nabla \begin{pmatrix} 1 & 0 \\ 0 & J(\boldsymbol{c}_S)^{-1} \end{pmatrix} \times (\boldsymbol{u} - \partial_t T_{ALE}(\boldsymbol{c}_S)) \cdot \boldsymbol{w} \mathrm{d}\boldsymbol{x},$$

$$b_{ALE}(\boldsymbol{c}_S; q, \boldsymbol{v}) = -2\pi \int_{\mathcal{F}} J(\boldsymbol{c}_S) q (\boldsymbol{v}_r + r\partial_r \boldsymbol{v}_1 + \partial_z \boldsymbol{v}_2 / J(\boldsymbol{c}_S)) \mathrm{d}\boldsymbol{x},$$

where $J(\boldsymbol{c}_S) = \det(\nabla T_{ALE})$ is the determinant of the deformation gradient.

The discretization is realized by means of quadratic equal-order finite elements for velocity and pressure on a quadrilateral mesh of the reference domain $\mathcal{F}$; we refer to Ref. 12, Sec. 4.4 for

details on the realization. To stabilize the inf–sup condition, we use the local projection scheme as introduced by Becker and Braack,[64] given, in the ALE formulation, as

$$s_{ALE}(\boldsymbol{c}_S; p, q) = \int_{\mathcal{F}} r \begin{pmatrix} \delta_r & 0 \\ 0 & \delta_z \end{pmatrix} \nabla \kappa_h(p) \cdot \begin{pmatrix} J(\boldsymbol{c}_S) & 0 \\ 0 & J(\boldsymbol{c}_S)^{-1} \end{pmatrix} \times \nabla \kappa_h(q) + \frac{J(\boldsymbol{c}_S)}{r} \delta_r \kappa_h(p) \kappa_h(q) \mathrm{d}\boldsymbol{x}.$$

Here, $\kappa_h := \mathrm{id} - i_{2h}$ is the coarse mesh fluctuation operator that subtracts the interpolation to the mesh with double spacing and where δ_r and δ_z are local stabilization parameters that depend on the element diameter h and the time step size k,

$$\delta_r = 0.1 \cdot \left(\frac{\mu_f}{\rho_f h^2} + \frac{1}{k} \right)^{-1}, \quad \delta_z = 0.1 \cdot \left(\frac{\mu_f}{\rho_f h^2 J(\boldsymbol{c}_S)^2} + \frac{1}{k} \right)^{-1}.$$

The different scaling in r- and z-directions reflects the anisotropy induced by the ALE transformation, cf. Ref. 65 or (Ref. 12, Sec. 5.3.3).

In time, we discretize the Navier–Stokes equations and the rigid-body problem with BDF2 time stepping in a decoupled iteration taking 10^{-8} as tolerance for the solid velocity and deformation update. The forces acting on the solid are evaluated by means of the Babuška–Miller trick[66,67] such that quadratic finite elements let us expect fourth order convergence. Hence, each step of spatial refinement will be accompanied by two refinements of the time step.

a. Fluid–rigid body interaction in arbitrary Lagrangian–Eulerian coordinates in three dimensions

To clarify the role of fixed body rotations of the solid and deflections of the center of mass from the z axis, both of which were observed in the experimental analysis,[41,42] we perform full three-dimensional simulations. The Navier–Stokes equations are formulated in Cartesian coordinates (1), and the motion of the rigid body is described by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{c}_S(t) &= \boldsymbol{v}_S(t), \\ \frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{v}_S(t) \cdot m_S &= \begin{pmatrix} 0 \\ 0 \\ m_S g - \mathrm{vol}(\mathcal{S}) \rho_f g \end{pmatrix} + \int_{\mathcal{I}} \boldsymbol{\sigma}(\boldsymbol{u}, p) \boldsymbol{n} \, \mathrm{d}s, \\ \frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{\omega}_S(t) \cdot I_S + \boldsymbol{\omega}_S(t) \times I_S \boldsymbol{\omega}_S(t) &= \int_{\mathcal{I}} (\boldsymbol{x} - \boldsymbol{c}_S(t)) \times \boldsymbol{\sigma}(\boldsymbol{u}, p) \boldsymbol{n} \, \mathrm{d}s, \end{aligned} \tag{A2}$$

where $\boldsymbol{v}_S(t) \in \mathbb{R}^3$ is the full velocity vector, $\boldsymbol{c}_S(t) \in \mathbb{R}^3$ is the solid's center of mass, and $\boldsymbol{\omega}_S(t) \in \mathbb{R}^3$ is the rotational velocity vector; cf. Eqs. (4)–(6). Assuming a homogeneous distribution of the density, the moment of inertia is given by

$$I_S = \frac{8}{15} \pi \rho_S r_S^5 \quad \mathrm{Id}.$$

It follows that the nonlinear term in the rotational ODE (A2) vanishes since $\boldsymbol{\omega}_S(t) \times I_S \boldsymbol{\omega}_S(t) = \frac{8}{15} \pi \rho_S r_S^5 \, \boldsymbol{\omega}_S(t) \times \boldsymbol{\omega}_S(t) = 0$. On the surface of the sphere, the Navier–Stokes Dirichlet conditions are used to describe the velocity

$$\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{v}_S(t) + \boldsymbol{\omega}_S(t) \times (\boldsymbol{x} - \boldsymbol{c}_S) \text{ on } \mathcal{I}.$$

To evaluate the torque correctly through the variational formulation, i.e., by using the Babuška–Miller trick,[66,67] the symmetric gradient is used instead in (10); hence,

$$a^{3d}(\boldsymbol{u}, \boldsymbol{v}) = \mu_f \int_{\Omega^{3d}} (\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T) : \nabla \boldsymbol{v} \, \mathrm{d}\boldsymbol{x}.$$

We cast the problem in standard ALE formulation and refer to Ref. 12, Chap. 5 for further details. The ALE map $T_{ALE}(t)$ is chosen similar to the reduced case (A1), but we must incorporate motion in the x–y plane and define

$$T_{ALE}(\boldsymbol{c}_S; x, y, z) = \begin{pmatrix} x + \boldsymbol{c}_{S,1} \cdot g_{ALE}(r(x, y)) \\ y + \boldsymbol{c}_{S,2} \cdot g_{ALE}(r(x, y)) \\ z + (\boldsymbol{c}_{S,3} - r_S - 0.05\,\mathrm{m}) f_{ALE}(z) \end{pmatrix},$$

where $f_{ALE}(\cdot)$ is defined in (A1), while $r(\cdot, \cdot)$ and $g_{ALE}(\cdot)$ are given by

$$r(x, y) = \max\left\{0, \min\left\{1, \frac{x - r_S}{R - r_S}\right\}\right\},$$
$$g_{ALE}(r) = 1 - \frac{1}{1 + \exp\left(\frac{1-2r}{r-r^2}\right)}.$$

The function $r(\cdot, \cdot)$ maps the x–y plane to $[0, 1]$ with $r(x, y) = 0$ for $\sqrt{x^2 + y^2} \le r_S$ and $r(x, y) = 1$ for $\sqrt{x^2 + y^2} \ge R$. Furthermore, g_{ALE} is a smooth transition function mapping $(0, 1)$ to $(0, 1)$, with all derivatives being zero at 0 and 1. On the other hand, the derivative of $f_{ALE}(z)$ is not defined at $z = 0.05\,\mathrm{m} \pm 2r_S$. These two surfaces are however resolved by the finite-element mesh to give good approximation characteristics of the ALE formulation.

b. Implementation

Both formulations, the reduced two-dimensional ALE formulation in cylindrical coordinates and the full three-dimensional ALE

TABLE V. Resulting reference quantities for the stationary test scenario.

	Discretization			Results
Method	$[h_{\min}, h_{\max}]$	dof	nze	F_z
ALE	[0.000 30, 0.004 9]	6.9	0.310	$-4.432\,34 \cdot 10^{-5}$
	[0.000 15, 0.002 4]	26.7	1.238	$-4.429\,92 \cdot 10^{-5}$
	[0.000 08, 0.001 2]	104.9	4.946	$-4.429\,75 \cdot 10^{-5}$
		Extrapolate		$-4.429\,74 \cdot 10^{-5}$
		order (in h)		3.8
CutFEM	[0.002 0, 0.008]	11.9	0.361	$-4.402\,54 \cdot 10^{-5}$
	[0.001 0, 0.004]	41.3	1.241	$-4.423\,55 \cdot 10^{-5}$
	[0.000 5, 0.002]	151.4	4.525	$-4.429\,19 \cdot 10^{-5}$

TABLE VI. Resulting reference quantities for the non-stationary moving domain test scenario.

	Discretization				Results	
Method	$[h_{\min}, h_{\max}]$	Δt	dof	nze	$F_{z,\max}$	$t_{z,\max}$
ALE	[0.000 30, 0.009 9]	1/5	11.1	0.49	$1.018\,38 \cdot 10^{-4}$	4.107 346 91
	[0.000 15, 0.005 0]	1/20	25.0	1.10	$1.017\,48 \cdot 10^{-4}$	4.113 563 25
	[0.000 08, 0.002 5]	1/80	58.7	2.61	$1.017\,20 \cdot 10^{-4}$	4.106 704 36
CutFEM	[0.002 0, 0.008]	1/25	11.9	0.361	$1.017\,00 \cdot 10^{-4}$	4.153
	[0.001 0, 0.004]	1/50	41.2	1.240	$1.013\,95 \cdot 10^{-4}$	4.138
	[0.000 5, 0.002]	1/100	151.3	4.526	$1.016\,71 \cdot 10^{-4}$	4.111

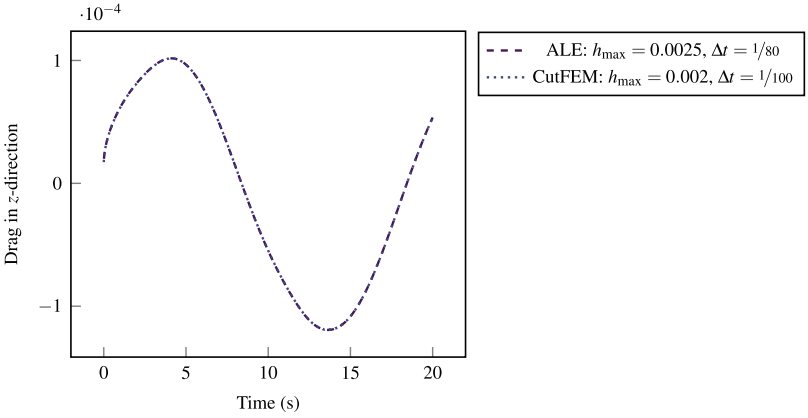


FIG. 7. Force functionals acting on the sphere with prescribed motion over time.

formulation, are implemented in the finite-element library Gascoigne3D.[62] The coupling between the Navier–Stokes equation and the rigid-body motion is resolved in a simple iteration until the discrepancy in velocity reached a threshold

$$\|\boldsymbol{v}|_{\mathcal{I}} - \boldsymbol{v}_{\mathcal{S}}\|_{\infty} < 10^{-8}.$$

The nonlinearity of the Navier–Stokes equation is solved by a Newton iteration, and the resulting linear systems of equations are approximated with a parallel generalized minimal residual method (GMRES) iteration, preconditioned by a geometric multigrid solver; see Ref. 68. The meshes are graded with a higher resolution close to the solid.

APPENDIX B: COMPUTATIONAL TEST CASES

We define two simplified test cases. This is intended to make it easier for others to reproduce the presented results using different methods and/or implementations.

1. Stationary flow test

For this stationary test, we modify the Rubber22 setup. The sphere is fixed at $c_{\mathcal{S}} = (0, 0, 0.1)$, i.e., the center of the cylinder. We impose an inflow boundary condition $\boldsymbol{u} = -0.01(1 - (x_1^2 + x_2^2)/R^2)$ on Γ_{top}, no-slip $\boldsymbol{u} = 0$ on $\Gamma_{\text{wall}} \cup \mathcal{I}$, and a homogeneous Neumann boundary condition $\boldsymbol{\sigma}(\boldsymbol{u}, p)\boldsymbol{n} = 0$ on Γ_{bottom}.

We consider the stationary Navier–Stokes problem on this domain. As reference quantities, we take the vertical stress acting on the sphere, i.e., testing the reduced formulation in (13) with the non-conforming, continuous test-functions $\widehat{\boldsymbol{w}} = (0, 1)^T$ on $\mathcal{I}$ and 0 on Γ, respectively.

We compute the problem based on the discretizations discussed in Subsections IV A 1 and IV A 2. The results can be seen in Table V.

2. Non-stationary flow test with prescribed motion

As a second test case, we keep the Rubber22 setup as the basis. The material parameters and the cylinder boundary conditions are as before, i.e., we consider no-slip on $\Gamma_{\text{wall}} \cup \Gamma_{\text{bottom}}$ and a free-slip condition on Γ_{top}. We prescribe the motion of the sphere as follows: Over the time interval $I = [0, 20]$, the sphere is located at

$$c_{\mathcal{S}} = (0, 0, d(t)) \quad \text{with} \quad d(t) = 0.1 + 0.05\cos(0.1\pi t).$$

The boundary condition on the interface is accordingly set to $\boldsymbol{u} = (0, 0, \partial_t d(t))$. Quantities of interest are the maximal value over

time of the z-component of the force functional in the reduced formulation.

We compute this again using the reduced formulation with the rigid-body discretizations. The quantitative results can be seen in Table VI, while the force functional is shown over time in Fig. 7.

DATA AVAILABILITY

The numerical data supporting the finding of this study are openly available in Zenodo, Ref. 63. The experimental data, on which this study is based, is openly available in Mendeley Data, Ref. 41.

REFERENCES

[1]G. Závodszky, B. van Rooij, B. Czaja, V. Azizi, D. de Kanter, and A. G. Hoekstra, "Red blood cell and platelet diffusivity and margination in the presence of cross-stream gradients in blood flows," Phys. Fluids **31**, 031903 (2019).

[2]J. O. Marston, J. P. K. Seville, Y.-V. Cheun, A. Ingram, S. P. Decent, and M. J. H. Simmons, "Effect of packing fraction on granular jetting from solid sphere entry into aerated and fluidized beds," Phys. Fluids **20**, 023301 (2008).

[3]H. Brenner, "The slow motion of a sphere through a viscous fluid towards a plane surface," Chem. Eng. Sci. **16**, 242–251 (1961).

[4]E. Feireisl, "On the motion of rigid bodies in a viscous incompressible fluid," in *Nonlinear Evolution Equations and Related Topics* (Birkhäuser, Basel, 2003), pp. 419–441.

[5]D. Gérard-Varet, M. Hillairet, and C. Wang, "The influence of boundary conditions on the contact problem in a 3D Navier–Stokes flow," J. Math. Pures Appl. **103**, 1–38 (2015).

[6]D. Gérard-Varet and M. Hillairet, "Existence of weak solutions up to collision for viscous fluid-solid systems with slip," Commun. Pure Appl. Math. **67**, 2022–2076 (2014).

[7]E. Feireisl, "On the motion of rigid bodies in a viscous compressible fluid," Arch. Ration. Mech. Anal. **167**, 281–308 (2003).

[8]R. H. Davis, J.-M. Serayssol, and E. J. Hinch, "The elastohydrodynamic collision of two spheres," J. Fluid Mech. **163**, 479–497 (1986).

[9]C. Grandmont and M. Hillairet, "Existence of global strong solutions to a beam-fluid interaction system," Arch. Ration. Mech. Anal. **220**, 1283–1333 (2016).

[10]M. Hillairet and T. Takahashi, "Collisions in three-dimensional fluid structure interaction problems," SIAM J. Math. Anal. **40**, 2451–2477 (2009).

[11]G. Gravina, S. Schwarzacher, O. Souček, and K. Tůma, "Contactless rebound of elastic bodies in a viscous incompressible fluid," arXiv:2011.01932v1 (2020).

[12]T. Richter, *Fluid-Structure Interactions: Models, Analysis and Finite Elements*, Lecture Notes in Computational Science and Engineering 118 (Springer, 2017).

[13]J. Donea, S. Giuliani, and J. P. Halleux, "An arbitrary Lagrangian-Eulerian finite element method for transient dynamic fluid-structure interactions," Comput. Methods Appl. Mech. Eng. **33**, 689–723 (1982).

[14]S. Frei, T. Richter, and T. Wick, "Long-term simulation of large deformation, mechano-chemical fluid-structure interactions in ALE and fully Eulerian coordinates," J. Comput. Phys. **321**, 874–891 (2016).

[15]T. Rezaee and K. Sadeghy, "Effect of porosity on the settling behavior of a 2D elliptic particle in a narrow vessel: A lattice-Boltzmann simulation," Phys. Fluids **31**, 123301 (2019).

[16]A. Johansson, M. G. Larson, and A. Logg, "High order cut finite element methods for the Stokes problem," Adv. Model. Simul. Eng. Sci. **2**, 24 (2015).

[17]B. Schott, C. Ager, and W. A. Wall, "A monolithic approach to fluid-structure interaction based on a hybrid Eulerian-ALE fluid domain decomposition involving cut elements," Int. J. Numer. Methods Eng. **119**, 208–237 (2019).

[18]R. Glowinski, T.-W. Pan, T. I. Hesla, and D. D. Joseph, "A distributed Lagrange multiplier/fictitious domain method for particulate flows," Int. J. Multiphase Flow **25**, 755–794 (1999).

[19]S. Court and M. Fournié, "A fictitious domain finite element method for simulations of fluid–structure interactions: The Navier–Stokes equations coupled with a moving solid," J. Fluids Struct. **55**, 398–408 (2015).

[20]S. Court, M. Fournié, and A. Lozinski, "A fictitious domain approach for the Stokes problem based on the extended finite element method," Int. J. Numer. Methods Fluids **74**, 73–99 (2013).

[21]E. Burman, S. Claus, P. Hansbo, M. G. Larson, and A. Massing, "CutFEM: Discretizing geometry and partial differential equations," Int. J. Numer. Methods Eng. **104**, 472–501 (2014).

[22]E. Burman, S. Frei, and A. Massing, "Eulerian time-stepping schemes for the non-stationary Stokes equations on time-dependent domains," arXiv:1910.03054v1 (2019).

[23]H. von Wahl, T. Richter, and C. Lehrenfeld, "An unfitted Eulerian finite element method for the time-dependent Stokes problem on moving domains," arXiv:2002.02352v1 (2020).

[24]C. Lehrenfeld and M. Olshanskii, "An Eulerian finite element method for PDEs in time-dependent domains," ESAIM: Math. Model. Numer. Anal. **53**, 585–614 (2019).

[25]E. Burman and M. A. Fernández, "An unfitted Nitsche method for incompressible fluid-structure interaction using overlapping meshes," Comput. Methods Appl. Mech. Eng. **279**, 497–514 (2014).

[26]A. Gerstenberger and W. A. Wall, "An eXtended finite element method/Lagrange multiplier based approach for fluid–structure interaction," Comput. Methods Appl. Mech. Eng. **197**, 1699–1714 (2008).

[27]P. Hansbo and J. Hermansson, "Nitsche's method for coupling non-matching meshes in fluid-structure vibration problems," Comput. Mech. **32**, 134–139 (2003).

[28]A. Legay, J. Chessa, and T. Belytschko, "An Eulerian–Lagrangian method for fluid–structure interaction based on level sets," Comput. Methods Appl. Mech. Eng. **195**, 2070–2087 (2006).

[29]F. P. T. Baaijens, "A fictitious domain/mortar element method for fluid–structure interaction," Int. J. Numer. Methods Fluids **35**, 743–761 (2001).

[30]G.-H. Cottet, E. Maitre, and T. Milcent, "Eulerian formulation and level set models for incompressible fluid-structure interaction," ESAIM: Math. Model. Numer. Anal. **42**, 471–492 (2008).

[31]T. Dunne, "An Eulerian approach to fluid–structure interaction and goal-oriented mesh adaptation," Int. J. Numer. Methods Fluids **51**, 1017–1039 (2006).

[32]S. Frei, "Eulerian finite element methods for interface problems and fluid-structure interactions," Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg, 2016.

[33]F. Hecht and O. Pironneau, "An energy stable monolithic Eulerian fluid–structure finite element method," Int. J. Numer. Methods Fluids **85**, 430–446 (2017).

[34]T. Richter, "A fully Eulerian formulation for fluid–structure-interaction problems," J. Comput. Phys. **233**, 227–240 (2013).

[35]S. Turek and J. Hron, "Proposal for numerical benchmarking of fluid–structure interaction between an elastic object and laminar incompressible flow," in *Fluid-Structure Interaction,* Lecture Notes in Computational Science and Engineering Vol. 53 (Springer Berlin Heidelberg, 2006), pp. 371–385.

[36]C. Ager, B. Schott, A. T. Vuong, A. Popp, and W. A. Wall, "A consistent approach for fluid-structure-contact interaction based on a porous flow model for rough surface contact," Int. J. Numer. Methods Eng. **119**, 1345–1378 (2019).

[37]C. Ager, A. Seitz, and W. A. Wall, "A consistent and versatile computational approach for general fluid-structure-contact interaction problems," Int. J. Numer. Methods Eng. (published online) (2020).

[38]E. Burman, M. A. Fernández, S. Frei, and F. M. Gerosa, "3D-2D Stokes-Darcy coupling for the modelling of seepage with an application to fluid-structure interaction with contact," arXiv:1912.08503 (2019).

[39]E. Burman, M. A. Fernández, and S. Frei, "A Nitsche-based formulation for fluid-structure interactions with contact," ESAIM: Math. Model. Numer. Anal. **54**, 531–564 (2020).

[40]S. Zonca, P. F. Antonietti, and C. Vergara, "A polygonal discontinuous Galerkin formulation for contact mechanics in fluid-structure interaction problems," MOX-Report No. 26/2020, 2020.

[41]T. Hagemeier, Particle settling-transitional regime, version 1, Mendeley Data, September 2020.

[42]T. Hagemeier, D. Thévenin, and T. Richter, "Settling of spherical particles in the transitional regime," Int. J. Multiph. Flow **138**, 103589 (2021).

[43]V. John, A. Linke, C. Merdon, M. Neilan, and L. G. Rebholz, "On the divergence constraint in mixed finite element methods for incompressible flows," SIAM Rev. **59**, 492–544 (2017).

[44]M. L. Anderson, P. H. Mott, and C. M. Roland, "The compression of bonded rubber disks," Rubber Chem. Technol. **77**, 293–302 (2004).
[45]C. Bernardi, M. Dauge, and Y. Maday, *Spectral Methods for Axisymetric Domain*, Series in Applied Mathematics (Gauthier-Villars, Paris, 1999).
[46]T. J. R. Hughes, W. K. Liu, and T. K. Zimmermann, "Lagrangian-Eulerian finite element formulation for incompressible viscous flows," Comput. Methods Appl. Mech. Eng. **29**, 329–349 (1981).
[47]T. Richter and T. Wick, "Finite elements for fluid–structure interaction in ALE and fully Eulerian coordinates," Comput. Methods Appl. Mech. Eng. **199**, 2633–2642 (2010).
[48]J. Guzmán and M. Olshanskii, "Inf-sup stability of geometrically unfitted Stokes finite elements," Math. Comput. **87**, 2091–2112 (2017).
[49]E. Burman, "Ghost penalty," C. R. Math. **348**, 1217–1220 (2010).
[50]C. Lehrenfeld, "High order unfitted finite element methods on level set domains using isoparametric mappings," Comput. Methods Appl. Mech. Eng. **300**, 716–733 (2016).
[51]J. Nitsche, "Über ein variationsprinzip zur lösung von Dirichlet-problemen bei verwendung von teilräumen, die keinen randbedingungen unterworfen sind," Abh. Math. Semin. Univ. Hambg. **36**, 9–15 (1971).
[52]J. Preuss, "Higher order unfitted isoparametric space-time FEM on moving domains," MA thesis, Georg-August Universität Göttingen, 2018.
[53]S. Sathe and T. E. Tezduyar, "Modeling of fluid-structure interactions with the space-time finite elements: Contact problems," Comput. Mech. **43**, 51–60 (2008).
[54]D. Wan and S. Turek, "Direct numerical simulation of particulate flow via multigrid FEM techniques and the fictitious boundary method," Int. J. Numer. Methods Fluids **51**, 531–566 (2006).
[55]J. Schöberl, "NETGEN an advancing front 2D/3D-mesh generator based on abstract rules," Comput. Vis. Sci. **1**, 41–52 (1997).
[56]J. Schöberl, "C++11 implementation of finite elements in NGSolve," ASC Technical Report Report No. 30/2014, Institute for Analysis and Scientific Computing, TU Wien, September 2014.
[57]Ngsxfem: An add-on to NGSolve for unfitted finite element discretizations, version 1.3.2008, 2020.
[58]S. Frei and T. Richter, "A locally modified parametric finite element method for interface problems," SIAM J. Numer. Anal. **52**, 2315–2334 (2014).
[59]S. Frei, "An edge-based pressure stabilization technique for finite elements on arbitrarily anisotropic meshes," Int. J. Numer. Methods Fluids **89**, 407–429 (2019).
[60]S. Frei and T. Richter, "A second order time-stepping scheme for parabolic interface problems with moving interfaces," ESAIM: Math. Model. Numer. Anal. **51**, 1539–1560 (2017).
[61]P. Alart and A. Curnier, "A mixed formulation for frictional contact problems prone to Newton like solution methods," Comput. Methods Appl. Mech. Eng. **92**, 353–375 (1991).
[62]R. Becker, M. Braack, D. Meidner, and T. Richter, The Finite Element Toolkit Gascoigne 3D, www.gascoigne.de (accessed February 24, 2021).
[63]H. vonWahl, T. Richter, S. Frei, and T. Hagemeier (2020), "Falling balls in a viscous fluid with contact: Comparing numerical simulations with experimental data," Zenodo. https://doi.org/10.5281/zenodo.3989604.
[64]R. Becker and M. Braack, "A finite element pressure gradient stabilization for the Stokes equations based on local projections," Calcolo **38**, 173–199 (2001).
[65]M. Braack and T. Richter, "Local projection stabilization for the Stokes system on anisotropic quadrilateral meshes," in Proceedings of ENUMATH 2005 the 6th European Conference on Numerical Mathematics and Advanced Applications, Santiago de Compostela, Spain, July 2005, edited by A. de Castro, D. Gómez, P. Quintela, and P. Salgado, 2006.
[66]I. Babuška and A. Miller, "The post-processing approach in the finite element method. I: Calculations of displacements, stresses and other higher derivatives of the displacements," Int. J. Numer. Methods Eng. **20**, 1085–1109 (1984).
[67]H. von Wahl, T. Richter, C. Lehrenfeld, J. Heiland, and P. Minakowski, "Numerical benchmarking of fluid-rigid body interactions," Comput. Fluids **193**, 104290 (2019).
[68]L. Failer and T. Richter, "A parallel Newton multigrid framework for monolithic fluid-structure interactions," J. Sci. Comput. **82**, 28 (2020).

ESAIM: M2AN 57 (2023) 1355–1380
https://doi.org/10.1051/m2an/2023022

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

A LOCALLY MODIFIED SECOND-ORDER FINITE ELEMENT METHOD FOR INTERFACE PROBLEMS AND ITS IMPLEMENTATION IN 2 DIMENSIONS

Stefan Frei[1,*], Gozel Judakova[2] and Thomas Richter[2]

Abstract. The locally modified finite element method, which is introduced in Frei and Richter [*SIAM J. Numer. Anal.* **52** (2014) 2315–2334], is a simple fitted finite element method that is able to resolve weak discontinuities in interface problems. The method is based on a fixed structured coarse mesh, which is then refined into sub-elements to resolve an interior interface. In this work, we extend the locally modified finite element method in two space dimensions to second order using an isoparametric approach in the interface elements. Thereby we need to take care that the resulting curved edges do not lead to degenerate sub-elements. We prove optimal *a priori* error estimates in the L^2-norm and in a discrete energy norm. Finally, we present numerical examples to substantiate the theoretical findings.

Mathematics Subject Classification. 65N30, 65N15.

Received July 16, 2022. Accepted March 3, 2023.

1. Introduction

In this paper, we extend the locally modified finite element method introduced in [1–3] to higher order. We investigate interface problems, where the normal derivative of the solution may have a jump over an interior interface. Examples of such interface problems are ubiquitous in technology, industry, science and medicine. Some of the most prominent examples include fluid-structure interactions [2,4] or multiphase flows [5]. Fluid-structure interactions arise in aerodynamical applications like flow around airplanes or parachutes [6], in biomedical problems such as blood flow through the cardiovascular system [7–9] or the airflow within the respiratory system [10] and even in tribological applications [11]. Multiphase problems include gas-liquid and particle-laden gas flows, rising bubbles [12], droplets in microfluidic devices [13] or the simulation of tumor growth [14]. Another field of application are shape or topology optimization problems including multi-component structures [15,16]. The simplest possible setting, which is the scope of the present paper, is a diffusion problem where the coefficient is discontinuous across an interior interface.

We assume that the domain $\Omega \subset \mathbb{R}^2$ is divided into $\Omega = \Omega_1 \cup \Gamma \cup \Omega_2$ with an interface $\Gamma \subset \partial\Omega_1 \cap \partial\Omega_2$, such that $\overline{\Gamma} = \partial\Omega_1 \cap \partial\Omega_2$, and a discontinuous diffusion coefficient $\nu > 0$ across Γ. In order to simplify the analysis we will assume that the outer domain Ω is a two-dimensional convex domain with polygonal boundary. Some

Keywords and phrases. Fitted finite elements, interface problem, *a priori* error estimates, weak discontinuities.

[1] Department of Mathematics & Statistics, University of Konstanz, Konstanz, Germany.
[2] Institut für Analysis und Numerik, Otto-von-Guericke-Universität Magdeburg, 39106 Magdeburg, Germany.
*Corresponding author: stefan.frei@uni-konstanz.de

remarks on corresponding three-dimensional methods will be given in Remark 2. We consider the equations

$$-\nabla \cdot (\nu_i \nabla u) = f \qquad \text{in } \Omega_i, \quad i = 1, 2, \tag{1}$$

$$[u] = 0, \ [\nu \partial_n u] = 0 \qquad \text{on } \Gamma, \tag{2}$$

$$u = 0 \qquad \text{on } \partial\Omega, \tag{3}$$

where $\nu|_{\Omega_i} := \nu_i$, $i = 1, 2$ and the jump $[w]$ at the interface Γ with normal vector $\boldsymbol{n}$ is defined by

$$[w](\boldsymbol{x}) := \lim_{s \searrow 0} w(\boldsymbol{x} + s\boldsymbol{n}) - \lim_{s \nearrow 0} w(\boldsymbol{x} + s\boldsymbol{n}), \quad \boldsymbol{x} \in \Gamma.$$

The corresponding variational formulation of the problem (1) is given by

$$u \in H_0^1(\Omega) : \ \sum_{i=1}^{2} (\nu_i \nabla u, \nabla \varphi)_{\Omega_i} = (f, \varphi)_\Omega \quad \forall \varphi \in H_0^1(\Omega). \tag{4}$$

This interface problem is intensively discussed in the literature. Babuška [17] shows that a standard finite element ansatz has low accuracy, regardless of the polynomial degree of the finite element space,

$$||u - u_h||_\Omega = \mathcal{O}(h), \quad ||\nabla(u - u_h)||_\Omega = \mathcal{O}(h^{1/2}),$$

where throughout the paper $\|\cdot\|_\Omega := \|\cdot\|_{L^2(\Omega)}$ denotes the L^2-norm. To improve the accuracy, the interface needs to be resolved within the discretization. Frei and Richter [1] presented a locally modified finite element method based on first-order polynomials with first-order accuracy in the energy norm and second order in the L^2-norm. The method is based on a fixed coarse *patch mesh* consisting of quadrilaterals, which is independent of the position of the interface. The patch elements are then divided into sub-elements, such that the interface is locally resolved. The discretization is based on piecewise linear finite elements which has a natural extension to higher order finite element spaces.

Due to the fixed background *patch mesh* this approach is particularly suitable for problems involving moving interfaces, where functions $u_h(t_{n-1})$ and $u_h(t_n)$ defined on different sub-meshes need to be integrated against each other within a time-stepping scheme [18]. Due to the implicit adaption of the finite element spaces within the locally modified finite element method, a costly re-meshing procedure is avoided. Similarly, the locally modified finite element method might be useful in shape or topology optimization problems, where problems need to be solved for different interface and boundary positions, while approaching the solution [15, 16].

The *locally modified finite element method* has been used by the authors and co-workers [3, 19–21], and by Langer and Yang [22] for fluid-structure interaction (FSI) problems, including the transition from FSI to solid-solid contact [23–25]. Holm *et al.* [26] and Gangl and Langer [15] used a corresponding approach based on triangular patches, the latter work being motivated by a topology optimization problem. A pressure stabilization technique for flow problems has been developed in [27] and a suitable (second-order) time discretization scheme in [18]. Details of the implementation in deal.ii and the corresponding source code have been published in [28,29]. Extensions to three space dimensions have been developed by Langer and Yang [30], where hexahedral coarse cells are divided into sub-elements consisting of hexahedra and tetrahedra, and by Höllbacher and Wittum, where a coarse mesh consisting of tetrahedra is sub-divided into hexahedrons, prisms and pyramids [31, 32].

Alternative approaches are unfitted methods, where the mesh is fixed and does not resolve the interface. Prominent examples are the extended finite element method (XFEM [33–36]) and the generalised finite element (GFEM [37]), where the finite element space is enriched by suitable functions that contain certain properties of the solution (for example discontinuities in the function or its derivative). Higher-order approximations within the XFEM approach have been developed by Cheng and Fries [38] and by Dréau *et al.* [39]. A conceptionally different unfitted approach are Cut Finite Elements (CutFEM) [40–44]. Here the main difficulty lies in the construction of quadrature formulas to represent the interface in the cut cells. Possibilities to obtain higher-order approximations include the definition of parametric mappings in the cut cells [45–47] or a boundary

value correction based on Taylor expansion [48]. Unfitted discontinuous Galerkin methods within the CutFEM paradigm have been developed in [49–51]. Areias and Belytschko noted that CutFEM and the discontinuous variant of the XFEM approach are in fact closely related [52]. A further unfitted approach that circumvents the problem to construct quadrature formulas is the shifted boundary method [53], where interface conditions are formulated on neighbouring edges instead of the interface Γ.

For further *fitted* finite element methods, we refer to [54–58]. Some works are similar to the *locally modified finite element methods* in the sense that only mesh elements close to the interface are altered [59, 60]. *Fitted* methods with higher order approximations have been developed by Fang [61] and by Omerović and Fries [62]. Recently, a method called *Universal Meshes* gained certain popularity [63]. Here the idea is to construct a suitable mapping for the elements in the interface region to resolve the interface.

After this introduction, we describe the locally modified high order finite element approach and show a maximum angle condition in Section 2. In Section 3, we derive the main results of this work, namely *a priori* error estimates in the L^2- and in a discrete energy norm. Section 4 gives some details on the implementation and in Section 5, we show different numerical examples. The conclusion of our work follows in Section 6.

2. Locally modified high order finite element method

In this section we review the first order approach proposed by Frei and Richter [1] and extend it to a second order discretization. The splitting into subelements that we propose is slightly different from the one presented in [1] and leads in general to a better bound for the maximal angles within the triangles.

Let $\mathcal{T}_P$ be a form and shape-regular quadrilateral mesh of the convex domain Ω with polygonal boundary. The elements $P \in \mathcal{T}_P$ are called patches and these do not necessarily resolve the partitioning. By a slight abuse of notation, we will in the following write P both for the elements of the triangulation and for the domain spanned by a patch P. The interface Γ may cut the patches. In this case we make the assumption:

Assumption 1 (Interface configuration)**.**

(1) *Each patch $P \in \mathcal{T}_P$ is either cut $P \cap \Gamma \neq \emptyset$ or not cut $P \cap \Gamma = \emptyset$. If it is cut, the cut goes through exactly two points on the boundary ∂P, see Figure 1 (left and top right).*

(2) *The interface does not cut the same edge multiple times and may not enter and leave the patch at the same edge, see Figure 1 (bottom right).*

Given a smooth interface Γ, this assumption is fulfilled after sufficient refinement. The patch mesh $\mathcal{T}_P$ is the fixed background mesh used in the parametric finite element method described below. We will introduce a further local refinement of the mesh, denoted $\mathcal{T}_h$, which resolves the interface. However, this refined mesh is only for illustrative purpose. The numerical realization is based on the fixed mesh $\mathcal{T}_P$ and the "refinement" is in fact only included in a parametric way within the reference map for each patch $P \in \mathcal{T}_P$.

If the interface is matched by one of the edges of the patch, then the patch is considered as not cut. We will split such patches into four quadrilaterals. If the interface cuts the patch, then the patch splits either in eight triangles or in four quadrilaterals. In both cases, the patch P is first split into four quadrilaterals, denoted by $K_1, \dots, K_4$, which are then possibly refined into two triangles each. The resulting sub-cells are denoted by $T_1, \dots, T_8$ in the case of triangular sub-cells and by $T_1, \dots, T_4$ in the case of quadrilaterals (In the latter case these are identical to $K_1, \dots K_4$). This two-step procedure will simplify the following proofs. We define the isoparametric finite element space $\tilde{V}_h \subset H_0^1(\Omega)$

$$\tilde{V}_h := \left\{ \varphi \in C(\Omega) \,|\, (\varphi \circ \xi_T) \in \mathcal{P}_T^r(\hat{T}) \text{ for } T \in \mathcal{T}_h \right\}, \tag{5}$$

where

$$\mathcal{P}_T^r(\hat{T}) := \begin{cases} Q_r(\hat{T}), & T \text{ is a quadrilateral,} \\ P_r(\hat{T}), & T \text{ is a triangle,} \end{cases}$$

and $\xi_T \in \mathcal{P}_T^1(\hat{T})$ is a transformation from the reference element $\hat{T}$ to T. The space $\tilde{V}_h$ is continuous, as the restriction of a function in $Q_r(\hat{T})$ to a line $e \subset \partial\hat{T}$ is in $P_r(\hat{T})$.

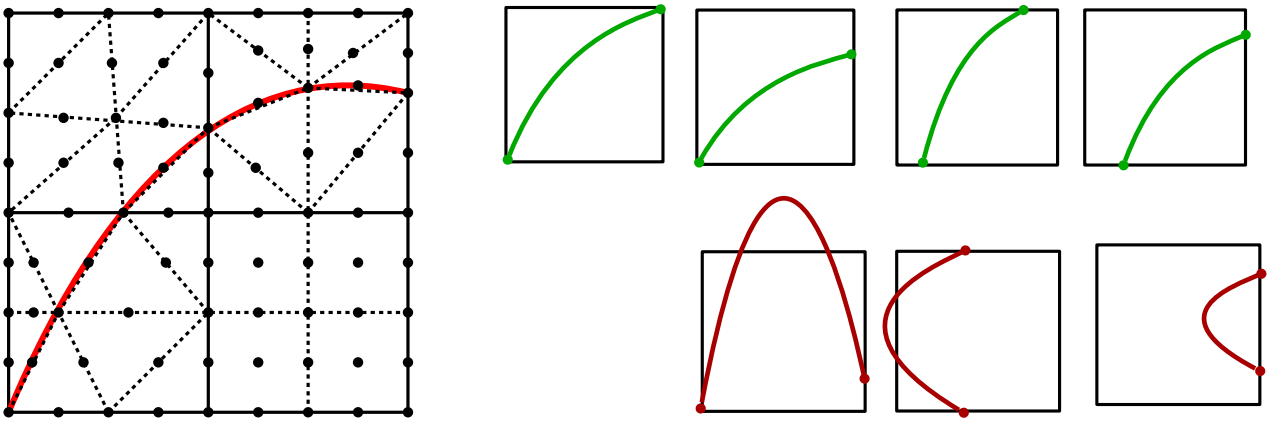


FIGURE 1. *Left*: mesh consisting of four patches, three of which are cut by the interface. *Right*: possible configurations (*top*), and configurations that are not allowed (*bottom*).

2.1. Maximum angle condition

In order to show optimal-order error estimates, the finite element mesh needs to fulfill a maximum angle condition in a fitted finite element method. We first analyse the maximum angles of the subtriangles in a Cartesian patch grid $\mathcal{T}_P$. A bound for a general regular patch mesh can be obtained by using the regularity of the patch mesh.

2.1.1. Linear interface approximation

We distinguish between five different types of interface cuts by the fact that the interface intersects a patch either in 1 or 2 exterior vertices (Config. B and A) or two opposite (C) or adjacent (D and E) edges, see Figure 2. Let $r, s \in (0,1)$ denote the relative cut positions on an edge e (see Fig. 2). In the case of adjacent edges, we distinguish further between the case that $r \leq \frac{1}{2}$ and $s \geq \frac{1}{2}$ (D) and the case that one these inequalities is violated (E). In all cases the patch element can be split in four large quadrilaterals $K_1, \ldots, K_4$ first, which are then divided into two sub-triangles, if the interface cuts through the patch. Details are given in the appendix.

Considering arbitrary interface positions, anisotropic elements can arise, when the relative cut position $r, s \in (0,1)$ on an edge e tends to 0 or 1 (see Fig. 2). We can not guarantee a minimum angle condition for the sub-triangles, but we can ensure that the maximum angles remain bounded away from 180°.

Lemma 1 (Linear approximation of the interface)**.** *All interior angles of the Cartesian patch elements shown in Figure 2 are bounded by* 135° *independently of the parameters* $r, s \in (0,1)$.

Proof. The proof follows by basic geometrical considerations, see Appendix B. □

Theorem 1. *We assume that the patch grid* $\mathcal{T}_P$ *is Cartesian. For all types of interface cuts (see Fig. 2), the interior angles of all subelements are bounded by* 135° *independently of the parameters* $r, s \in (0,1)$.

Proof. By means of Lemma 1 all interior angles on the reference patch are bounded by 135°. As all cells are Cartesian, the same bound holds for the elements $T \subset P$. □

Remark 1. We have assumed for simplicity that the underlying patch mesh is fully Cartesian. This assumption can, however, easily be weakened. Allowing more general form- and shape-regular patch meshes a geometric transformation of each patch to the unit patch will give a bound $\alpha < \alpha_{\max} < 180°$ for the interior angles α (with $\alpha_{\max}$ larger than 135°).

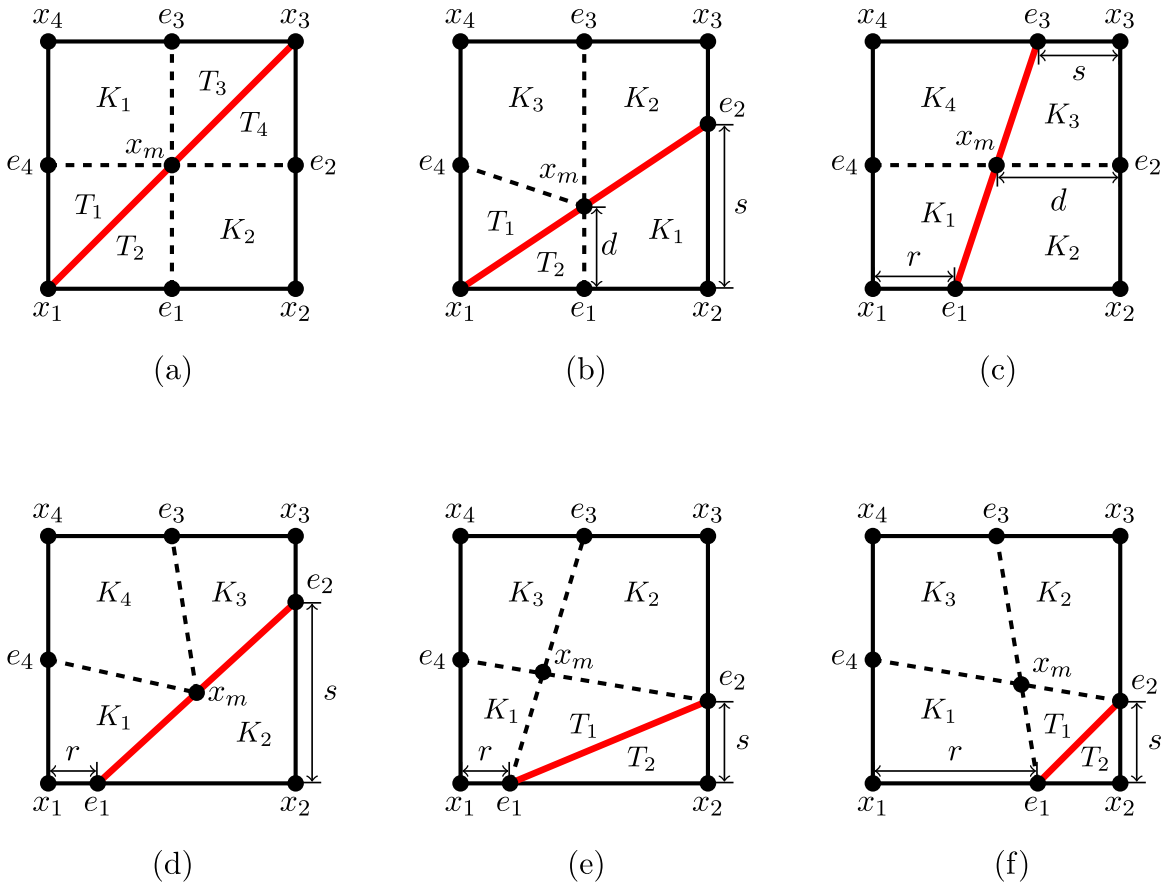


FIGURE 2. Different configurations and splitting into four large quadrilaterals $K_1, \ldots, K_4$. The red line shows a linear approximation of the interface. In quadrilaterals that are not split by the interface, we divide in such a way into subtriangles, that the largest angle is split. Note in particular that the degenerate quadrilateral K_2 in Configuration D will be split into two regular subtriangles that are used in the definition of finite element spaces. (a) Configuration A. (b) Configuration B. (c) Configuration C. (d) Configuration D. (e) Configuration E $r \in (0,1), s \in (0,1/2)$. (f) Configuration E $r \in (1/2,1), s \in (0,1)$.

2.1.2. Quadratic interface approximation

Next, we define a quadratic approximation of the interface. In each of the subtriangles obtained in the previous paragraph, we consider 6 degrees of freedom that lie on the vertices and edge midpints of the triangles (see the dots in Fig. 1, left). In order to guarantee a higher-order interface approximation those that lie on the discrete interface Γ_h need to be moved. The detailed algorithm is given in Section 4.

In certain "pathological" situations we can not guarantee that the angle conditions imposed above are fulfilled. This is due to the fact that the curved edges that correspond to a quadratic interface approximation might intersect other edges, see Figure 3 for an example.

In this case, we use a linear approximation of the interface in the affected patch (*i.e.* the reference maps ξ_T in the finite element space (5) are linear). We denote the set of patches, where a linear interface approximation is used by $\mathcal{T}_{P,\text{lin}}$ and the corresponding set of sub-cells with a linear reference map ξ_T by $\mathcal{T}_{h,\text{lin}}$. We will see in

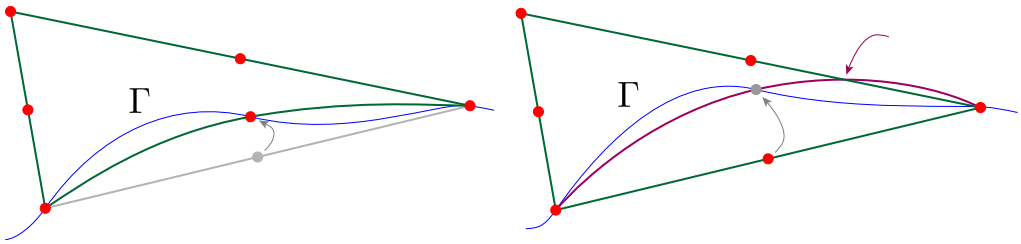


FIGURE 3. Each of the eight triangles approximates the interface Γ quadratically. Herefore, the midpoint of the edge that corresponds to the interface is pulled onto the curve. *Left*: this is a valid configuration where a quadratic approximation is possible. *Right*: in some configurations a fully quadratic interface approximation would result in a degenerate element with an interface that is leaving triangle (see the mark on the upper edge). Such triangles are approximated linearly and cause $n_l > 0$.

the numerical examples below that this happens rarely. Moreover, it is reasonable to assume that the maximum number of such patches remains bounded under refinement independently of $h \leq h_0$.

We give a heuristic argument for this assumption. Let us consider the situation sketched in Figure 3. The linear approximation of the interface will never leave the patch by definition. The maximum distance between a linear and quadratic interface approximation is bounded by $\mathcal{O}(h_P^2)$. In relation with the patch size $\mathcal{O}(h_P)$ this means that – considering arbitrary interface positions – the probability that the quadratic interface approximation leaves the patch is bounded by $\mathcal{O}(h_P)$. The number of interface patches, on the other hand, grows like $\mathcal{O}(h_P^{-1})$. Hence it is reasonable to assume that the number of affected patches behaves like $\mathcal{O}(1)$ for $h_P \to 0$. We will denote the maximum number of patches with a linear interface approximation by n_l.

2.2. Modified spaces and discrete variational formulation

We define the finite element space

$$V_h := \left\{ \varphi \in C(\Omega) \,|\, (\varphi \circ \xi_T) \in \mathcal{P}_T^r(\hat{T}) \text{ for } T \in \mathcal{T}_h \right\}, \tag{6}$$

where the map ξ_T resolves the interface with order r in all but n_l elements, where the approximation is only linear:

$$\xi_T \in \begin{cases} \mathcal{P}_T^1(\hat{T}), & T \in \mathcal{T}_{h,\text{lin}}, \\ \mathcal{P}_T^r(\hat{T}), & \text{else.} \end{cases}$$

The polynomial order of the trial functions $(\varphi \circ \xi_T)$ is r independent of the interface approximation.

We consider a C^3-parameterized interface Γ, which is not matched by the triangulation $\mathcal{T}_h$. The triangulation induces a discrete interface Γ_h, which is a quadratic (and in $\max \cdot n_l$ elements a linear) approximation to Γ. The discrete interface splits the triangulation in subdomains Ω_h^1 and Ω_h^2, such that each subcell $T \in \mathcal{T}_h$ is either completely included in Ω_h^1 or in Ω_h^2.

We consider the following discrete variational formulation: *Find* $u_h \in V_h$ *such that*

$$a_h(u_h, \phi_h) = (f_h, \phi_h)_\Omega \qquad \forall \phi_h \in V_h, \tag{7}$$

where we set $f_h|_{\Omega_h^i} := f_i, i = 1, 2$ and f_i is a smooth extension of $f|_{\Omega_i}$ to Ω_h^i. The bilinear form is given by

$$a_h(u_h, \phi_h) := (\nu_h \nabla u_h, \nabla \phi_h)_\Omega,$$

where ν_h is defined by

$$\nu_h = \begin{cases} \nu_1, & \boldsymbol{x} \in \Omega_h^1 \\ \nu_2, & \boldsymbol{x} \in \Omega_h^2. \end{cases}$$

Remark 2. The locally modified finite element method has straight-forward extensions to 3 space dimensions. One possibility is to use a hexahedral patch mesh, where each hexahedron is subdivided into 8 sub-hexahedra. The hexahedra affected by the interface are then further subdivided into 6 tetrahedra to resolve the interface. 4 different types of cuts have to be considered, based on the number of patch vertices that remain on each side of the interface (1 *vs.* 7, 2 *vs.* 6, 3 *vs.* 5 or 4 *vs.* 4). In order to guarantee a maximum angle condition in pathological situations some of the patch vertices can be moved, if necessary. Such an approach has been implemented by Langer and Yang, see [22]. Alternatively, the patches can also be subdivided into polyhedral sub-elements as in Höllbacher and Wittum [31,32]. For all variants the ideas as well as the analysis presented in this paper have a straight-forward extension to the three-dimensional method.

3. *A PRIORI* ERROR ANALYSIS

Let h_P be the maximum size of a patch element $P \in \mathcal{T}_P$ of the regular patch grid. We will denote the mismatch between Ω_h^i and Ω^i by S_h^i, $i = 1, 2$ (see Fig. 4)

$$S_h^1 := \Omega_h^1 \setminus \Omega_1 = \Omega_2 \setminus \Omega_h^2,$$
$$S_h^2 := \Omega_h^2 \setminus \Omega_2 = \Omega_1 \setminus \Omega_h^1.$$

Moreover, we denote the set of elements $T \in \mathcal{T}_h$ that contain parts of S_h^i by

$$S_T^i := \{T \in \mathcal{T}_h \,|\, T \cap S_h^i \neq 0\}, \qquad S_T := S_T^1 \cup S_T^2.$$

Further, we split S_h^i into parts $S_{h,\mathrm{lin}}^i$ with a linear approximation of the interface and parts $S_{h,\mathrm{qu}}^i$ with a quadratic approximation. Finally, by a slight abuse of notation, we will use the same notation, *e.g.* S_T, S_T^i, for the region that is spanned by the union of all elements in these sets.

By constants c we will denote in the following generic constants that are independent of the mesh size h_P, the position of the interface, the solution u and the number of linearly approximated elements n_l (but may depend on the parameters ν_i, $i = 1, 2$). We note that within a sequence of inequalities, c might even represent different values on different sides of the inequalities.

3.1. Auxiliary estimates

We begin with some technical estimates that will be needed in order to control the mismatch between continuous and discrete bilinear forms. To this purpose we will need the following Sobolev imbedding for $2 \leq p < \infty$

$$\|u\|_{L^p(\Omega)} \leq c p^{\frac{1}{2}} \|u\|_{H^1(\Omega)}, \tag{8}$$

which is valid with a constant c independent of p, see [64]. We will need the following technical result.

Lemma 2. *Let $\alpha \in \mathbb{N}$, $h \in \mathbb{R}_+$ and $J(p) := h^{-\frac{\alpha}{p}} p^{\frac{1}{2}}$. It holds that*

$$\min_{p \in [2,\infty]} J(p) \leq c |\ln(h)|^{\frac{1}{2}}.$$

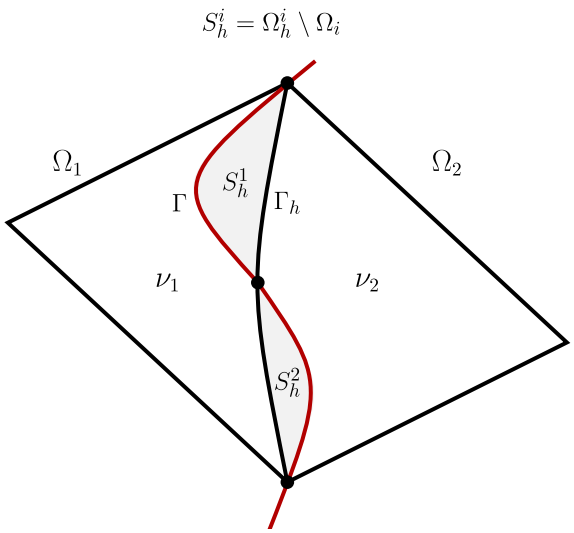


FIGURE 4. Mismatch between Ω^i and Ω_h^i, $i = 1, 2$ at two elements along the curved interface.

Proof. The necessary condition for a local minimum is

$$J'(p) = h^{-\frac{\alpha}{p}} \ln(h) \frac{\alpha}{p^2} p^{\frac{1}{2}} + \frac{1}{2} h^{-\frac{\alpha}{p}} p^{-\frac{1}{2}} = h^{-\frac{\alpha}{p}} p^{-\frac{1}{2}} \left(\ln(h) \frac{\alpha}{p} + \frac{1}{2} \right) \stackrel{!}{=} 0,$$

which yields $p = -2\alpha \ln(h)$. The minimum value is

$$J(-2\alpha \ln(h)) = -e^{-\frac{1}{2}} \sqrt{2\alpha} \ln(h)^{1/2}.$$

The fact that $\lim_{p\to\infty} J(p) = \infty$ and $J(2) > J(-2\alpha \ln(h))$ show that the local minimum is in fact a global one. □

The following lemma will be needed to estimate the mismatch between continuous and discrete bilinear forms.

Lemma 3 (Geometry approximation)**.** *Let $T \in S_T$ and let s be the local approximation order of the interface, i.e.*

$$\operatorname{dist}(\Gamma_h \cap T; \Gamma \cap T) \leq c h_P^{s+1}. \tag{9}$$

If the number of elements with a linear interface approximation is bounded by n_l, it holds for the areas of the regions $S_{h,\mathrm{lin}}$ and $S_{h,\mathrm{qu}}$ that

$$|S_{h,\mathrm{lin}}| \leq n_l h_P^3, \quad |S_{h,\mathrm{qu}}| \leq h_P^3. \tag{10}$$

For $u \in H^1(\Omega_1 \cup \Omega_2)$ and $\phi_h \in \mathcal{V}_h$ we have the bounds

$$\|\nabla \phi_h\|_{S_h \cap T} \leq c h_P^{\frac{s}{2}} \|\nabla \phi_h\|_T \tag{11}$$

$$\|u\|_{S_h \cap T} \leq c h_P^{\frac{s+1}{2}} \|u\|_{\Gamma \cap T} + c h_P^{s+1} \|\nabla u\|_{S_h \cap T}. \tag{12}$$

Moreover, we have for $u \in H^1(\Omega_1 \cup \Omega_2)$ and $v \in H^2(\Omega_1 \cup \Omega_2)$

$$\|u\|_{S_{h,\mathrm{lin}}} \leq c h_P \|u\|_{H^1(\Omega_1 \cup \Omega_2)}, \qquad \|u\|_{S_{h,\mathrm{qu}}} \leq c h_P^{\frac{3}{2}} \|u\|_{H^1(\Omega_1 \cup \Omega_2)} \tag{13}$$

and

$$\|u\|_{S_{h,\mathrm{lin}}} \leq cn_l^{\frac{1}{2}} h_P^{\frac{3}{2}} |\ln(h)|^{1/2} \|u\|_{H^1(\Omega_1\cup\Omega_2)}, \qquad \|v\|_{S_{h,\mathrm{lin}}} \leq cn_l^{\frac{1}{2}} h_P^{\frac{3}{2}} \|v\|_{H^2(\Omega_1\cup\Omega_2)}. \tag{14}$$

For functions $u \in H_0^1(\Omega)$ the H^1-norm on the right-hand side of (13) and (14) can be replaced by the H^1-seminorm.

Proof. Estimates (9), (11) and (12) have been shown in Lemmas 4.32 and 4.34 from [2][1] (10) follows from (9) and simple geometric arguments. For (11) and (12) a Poincaré-type estimate is used, see Lemma 4.34 of [2]

$$\|u\|^2_{S_h\cap T} \leq ch_P^{s+1} \|u\|^2_{\Gamma_h\cap T} + ch_P^{2s+2} \|\nabla u\|^2_{S_h\cap T}. \tag{15}$$

Summation over all elements in $S_{T,\mathrm{lin}}$ and $S_{T,\mathrm{qu}}$, respectively, and a global trace inequality for the interface terms yields (13). To show (14), we use a Hölder inequality for $p \in [2,\infty]$

$$\|u\|_{S_{h,\mathrm{lin}}} \leq |S_{h,\mathrm{lin}}|^{\frac{1}{2}-\frac{1}{p}} \|u\|_{L^p(S_{h,\mathrm{lin}})}. \tag{16}$$

Due to $|S_{h,\mathrm{lin}}| \leq cn_l h_P^3$ and the Sobolev imbedding (8) for $\Omega = \Omega_i$ we have for arbitrary $p \in [2,\infty)$

$$\|u\|_{S_{h,\mathrm{lin}}} \leq cp^{\frac{1}{2}} n_l^{\frac{1}{2}} h_P^{\frac{3}{2}-\frac{3}{p}} \|u\|_{H^1(\Omega_1\cup\Omega_2)}. \tag{17}$$

Using Lemma 2, we obtain the first estimate in (14). If $v \in H^2(\Omega_1 \cup \Omega_2)$ we can use (16) for $p = \infty$ due to the Sobolev imbedding $H^2(\Omega_i) \subset L^\infty(\Omega_i)$ and we obtain

$$\|u\|_{S_{h,\mathrm{lin}}} \leq cn_l^{\frac{1}{2}} h_P^{\frac{3}{2}} \|u\|_{H^2(\Omega_1\cup\Omega_2)}. \tag{18}$$

Finally, the norms on the right-hand side can be substituted by the H^1-seminorm for $u \in H_0^1(\Omega)$ by means of the Poincaré inequality. □

In the following estimates the mismatch between discrete and continuous bilinear form will be the predominant issue and will lead to some technicalities in the estimates. The continuous solution u is regular in Ω_1 and Ω_2, while its normal derivative has a jump across Γ. Discrete functions can only have irregularities at the boundaries of cells ∂T, which means that a discrete function can only resemble a similar discontinuity across the discrete interface Γ_h.

To cope with this difference, we will need a map $\pi : H^3(\Omega_1 \cup \Omega_2) \to H^3(\Omega_h^1 \cup \Omega_h^2)$. To define the map, let $u \in H^3(\Omega_1 \cup \Omega_2)$ and $u_i := u|_{\Omega_i} \in H^3(\Omega_i)$ the restriction to the subdomain $\Omega_i, i = 1, 2$. We use smooth extensions $\tilde{u}_i \in H^3(\Omega)(i = 1, 2)$ to the full domain Ω. Such an extension exists with the properties

$$\tilde{u}_i = u \quad \text{in } \Omega_i, \qquad \|\tilde{u}_i\|_{H^m(\Omega)} \leq C\|u\|_{H^m(\Omega_i)}, \quad i = 1, 2, \quad m = 2, 3, \tag{19}$$

as the interface Γ is smooth, see *e.g.* the textbook of Stein [65][Section VI.3.1]. We use these extensions to define a function $\pi u \in H^3(\Omega_h^1 \cup \Omega_h^2)$:

$$\pi u(\boldsymbol{x}) = \begin{cases} \tilde{u}_1(\boldsymbol{x}), & \boldsymbol{x} \in \Omega_h^1, \\ \tilde{u}_2(\boldsymbol{x}), & \boldsymbol{x} \in \Omega_h^2. \end{cases} \tag{20}$$

It should be noted that πu can be discontinuous across Γ_h.

The following estimate analyzes the difference between u and πu in the H^1-seminorm.

[1]The proof of Lemma 4.34 in [2] contains a typo. The Poincaré-type inequality is there given as $\|u_h\|^2_{S_h\cap T} \leq ch_P^{s+1}\|u\|^2_{\Gamma_h\cap T} + ch_P^{2s+1}\|\nabla u\|^2_{S_h\cap T}$ with the non-optimal order $2s+1$ instead of $2s+2$. The difference comes from a transmission error in the line above (4.17) in [2]. Integration of the inequality leading to (4.17) brings the factor h_P^{s+1} in addition to the factor h_P^{s+1}, which is already present by estimation of the distance between Γ and Γ_h. The statement of Lemma 4.34 in [2] is indeed correct and curing the proof is trivial by just correcting the typo.

Lemma 4. *Let $u \in H^3(\Omega_1 \cup \Omega_2)$, $\pi u \in H^3(\Omega_h^1 \cup \Omega_h^2)$ the function defined by (20) and n_l the maximum number of elements with a linear interface approximation. It holds that*

$$\|\nabla(u - \pi u)\|_\Omega \leq c h_P \Big(n_l^{1/2} + 1 \Big) \|u\|_{H^2(\Omega_1 \cup \Omega_2)} \tag{21}$$

$$\|\nabla(u - \pi u)\|_\Omega \leq c h_P^{3/2} \Big(n_l^{1/2} + 1 \Big) \|u\|_{H^3(\Omega_1 \cup \Omega_2)}. \tag{22}$$

Proof. u and πu differ only in the small strip S_h around the interface. For $u \in H^3(\Omega_i)$ we have, using the Sobolev embedding $H^3(\Omega_i) \subset W^{1,\infty}(\Omega_i)$ and the continuity of the extensions (19)

$$\begin{aligned} \|\nabla(u - \pi u)\|_\Omega = \|\nabla(u - \pi u)\|_{S_h} &\leq |S_h|^{\frac{1}{2}} \big(\|\nabla u\|_{L^\infty(\Omega)} + \|\nabla \pi u\|_{L^\infty(\Omega)} \big) \\ &\leq c |S_h|^{\frac{1}{2}} \|u\|_{H^3(\Omega_1 \cup \Omega_2)}. \end{aligned}$$

Equation (22) follows by means of (10).

To show (21), we note that $u - \pi u$ vanishes in cells $T \in \mathcal{T}_h \setminus S_T$. Thus, let $T \in S_T$ and let $s \in \{1, 2\}$ be the local approximation order of the interface in T. We use (12) and the fact that $s \geq 1$ to get

$$\begin{aligned} \|\nabla(u - \pi u)\|_T = \|\nabla(u - \pi u)\|_{S_h \cap T} &\leq c h_P^{\frac{1+s}{2}} \|\nabla(u - \pi u)\|_{\Gamma \cap T} + c h_P^{1+s} \big\|\nabla^2 (u - \pi u)\big\|_{S_h \cap T} \\ &\leq c h_P (\|\nabla u\|_{\Gamma \cap T} + \|\nabla \pi u\|_{\Gamma \cap T}) + c h_P^2 \Big(\big\|\nabla^2 u\big\|_{S_h \cap T} + \big\|\nabla^2 \pi u\big\|_{S_h \cap T} \Big), \end{aligned}$$

where the derivatives on Γ need to be seen from S_h.

After summation over all cells $T \in \mathcal{T}_h$ a global trace inequality and the continuity of the extension (19) yield

$$\|\nabla(u - \pi u)\|_\Omega \leq c h_P \big(\|u\|_{H^2(\Omega_1 \cup \Omega_2)} + \|\pi u\|_{H^2(\Omega_1 \cup \Omega_2)} \big) \leq c h_P \|u\|_{H^2(\Omega_1 \cup \Omega_2)}.$$

□

3.2. Interpolation

In this subsection, we will derive interpolation estimates for a Lagrangian interpolant I_h. Let $\mathcal{L}_T$ be the set of Lagrange points that belong to a cell $T \in \mathcal{T}_h$. In the case of a linear interface approximation, it can happen that some of these lie on Γ_h, but not on Γ. This means that there are elements with Lagrange points $\boldsymbol{x}_i \in \mathcal{L}_T$, that lie in different subdomains Ω_1 and Ω_2, see Figure 5. Defining the interpolant as $I_h u = \sum_{i \in \mathcal{L}_T} u(\boldsymbol{x}_i)$ would lead to a poor approximation order ($\mathcal{O}(h_P)$ in the H^1-norm), due to the discontinuity of ∇u across Γ. Each such point $\boldsymbol{x}_i$ lies, however, on a line between two points $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ on Γ. We use a linear interpolation of the values $u(\boldsymbol{x}_1^*)$ and $u(\boldsymbol{x}_2^*)$ in order to define $I_h u(\boldsymbol{x}_i) := \frac{1}{2}(u(\boldsymbol{x}_1^*) + u(\boldsymbol{x}_2^*))$, see also equation (6.13) from [3] and Figure 5.

We have the following approximation properties for this modified Lagrangian interpolant.

Lemma 5 (Interpolation)**.** *Let $u \in \mathcal{U} := \big[H_0^1(\Omega) \cap H^3(\Omega_1 \cup \Omega_2)\big]$ and $\tilde{u} = \pi u \in H^3(\Omega_h^1 \cup \Omega_h^2)$ the function resulting from the map π defined in (20). Moreover, we assume that Γ is a smooth interface with C^3-parametrization and that the interface is approximated with second order in all elements $T \in \mathcal{T}_h$, except for maximum n_l elements, where the interface approximation is linear. It holds for the Lagrangian interpolation operator $I_h : \mathcal{U} \to V_h$ that*

$$\|\nabla^m (u - I_h u)\|_\Omega \leq c h_P^{2-m} \|u\|_{H^2(\Omega_1 \cup \Omega_2)}, \quad m = 1, 2 \tag{23}$$

$$\|\nabla(\tilde{u} - I_h u)\|_\Omega \leq \Big(c_l n_l^{1/2} |\ln(h)|^{1/2} + c_q \Big) h_P^2 \|u\|_{H^3(\Omega_1 \cup \Omega_2)}. \tag{24}$$

where c_l and c_q are generic constants that correspond to patches with a linear and a quadratic interface approximation, respectively. For $u \in W^{2,\infty}(\Omega_1 \cup \Omega_2)$ we have further

$$\|\nabla(\tilde{u} - I_h u)\|_\Omega \leq \Big(c_l n_l^{1/2} + c_q \Big) h_P^2 \|u\|_{W^{2,\infty}(\Omega_1 \cup \Omega_2)}. \tag{25}$$

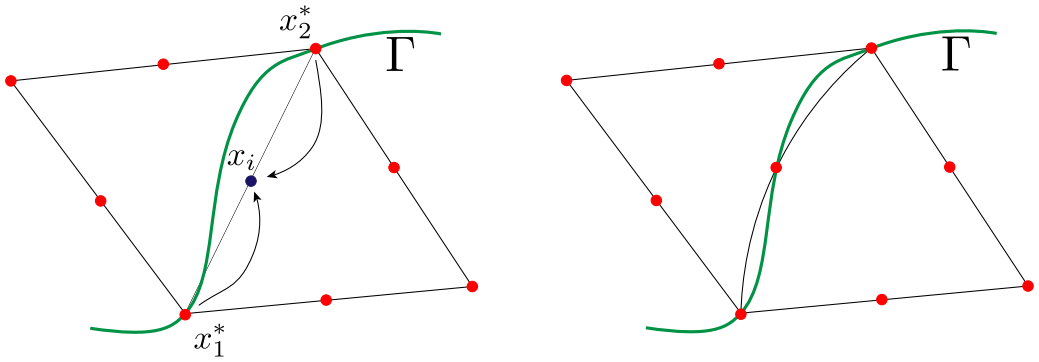


FIGURE 5. Interpolation operator at the interface. If the interface is approximated with second order (*right*) we use the standard nodal interpolation. For linear interface approximations (*left*) we replace the node in the middle of the interface edge by the mean of the two adjacend corner nodes.

Proof. First, we note that it holds $I_h u = I_h \tilde{u}$ by construction of the interpolant, as $u(\boldsymbol{x}_i) = \tilde{u}(\boldsymbol{x}_i)$ in all Lagrange points, that are used in the definition of the interpolant I_h.

Next, we note that the proof of all estimates is standard in all elements $T \in \mathcal{T}_h \setminus S_T$ that are not affected by the interface, since $\tilde{u}|_T = u|_T, u \in H^3(T)$ and

$$\|\nabla^m(u - I_h u)\|_T \leq c h_P^{2-m} \|u\|_{H^2(T)} \, (m = 1, 2), \qquad \|\nabla(u - I_h u)\|_T \leq c h_P^2 \|u\|_{H^3(T)}. \tag{26}$$

In elements $T \in S_{T,\mathrm{qu}}$ there are no Lagrange points on $\Gamma_h \setminus \Gamma$ and I_h is the standard Lagrangian interpolant. As $\tilde{u}$ is smooth in T, estimate (24) is also standard

$$\|\nabla(\tilde{u} - I_h u)\|_T = \|\nabla(\tilde{u} - I_h \tilde{u})\|_T \leq c h_P^2 \|\tilde{u}\|_{H^3(T)}.$$

In elements $T \in S_{T,\mathrm{lin}}$ the interpolation is only linear due to the modification described above. Let $T \in S^i_{T,\mathrm{lin}}$ with $i \in \{1, 2\}$ and let P be the patch that contains T. The following estimate has been shown in Lemma 6.14 of [3]

$$\|\nabla(\tilde{u} - I_h u)\|_T = \|\nabla(\tilde{u} - I_h \tilde{u})\|_T \leq c h_P \|\nabla^2 \tilde{u}_i\|_P, \tag{27}$$

where $\tilde{u}_i$ denotes the extension of u_i to Ω. We sum over all elements $T \in S^i_{T,\mathrm{lin}}$ and denote by $S^i_{P,\mathrm{lin}}$ the region, which is spanned by the patches P containing elements $T \in S^i_{T,\mathrm{lin}}$. It holds with $|S^i_{P,\mathrm{lin}}| = n_l \mathcal{O}(h_P^2)$

$$\begin{aligned}\|\nabla(\tilde{u} - I_h u)\|_{S^i_{T,\mathrm{lin}}} &\leq c h_P \left\|\nabla^2 \tilde{u}_i\right\|_{S^i_{P,\mathrm{lin}}} \\ &\leq c h_P \left|S^i_{P,\mathrm{lin}}\right|^{\frac{1}{2}-\frac{1}{p}} \|\tilde{u}_i\|_{W^{2,p}(\Omega_i)} \leq c n_l h_P^{2-\frac{2}{p}} \|\tilde{u}_i\|_{W^{2,p}(\Omega_i)}.\end{aligned}$$

The estimate (25) follows for $p = \infty$. To show (24), we can estimate further by using the Sobolev imbedding (8) for $p < \infty$

$$\|\nabla(\tilde{u} - I_h u)\|_{S^i_{T,\mathrm{lin}}} \leq c p^{1/2} n_l^{1/2} h_P^{2-2/p} \|u\|_{H^3(\Omega_i)}.$$

The estimate (24) follows by means of Lemma 2. To show (23) we split into

$$\|\nabla^m(u - I_h u)\|_T \leq \|\nabla^m(u - \tilde{u})\|_T + \|\nabla^m(\tilde{u} - I_h \tilde{u})\|_T. \tag{28}$$

For $m = 1$, the first term has been estimated in Lemma 4, for $m = 2$ we can use the stability of the extension (19). Equation (23) follows from (27) in $S_{T,\mathrm{lin}}$ and standard interpolation estimates elsewhere. □

3.3. *A priori* error estimate

We are now ready to prove the main result of this section. To this end, we introduce the discrete energy norm

$$|||u - u_h||| := \left(\left\| \nu_1^{1/2} \nabla (\tilde{u}_1 - u_h^1) \right\|_{\Omega_h^1}^2 + \left\| \nu_2^{1/2} \nabla (\tilde{u}_2 - u_h^2) \right\|_{\Omega_h^2}^2 \right)^{1/2},$$

where $\tilde{u}_i$ are smooth extensions of $u_i = u|_{\Omega_i}$ to Ω_h^i and $u_h^i := u_h|_{\Omega_h^i}$.

Theorem 2 (*A priori* estimate)**.** *Let $\Omega \subset \mathbb{R}^2$ be a convex domain with polygonal boundary, which is resolved (exactly) by the family of triangulations $\mathcal{T}_h$. We assume a splitting $\Omega = \Omega_1 \cup \Gamma \cup \Omega_2$, where Γ is a smooth interface with C^3-parametrization and that the solution u to (4) belongs to $H^3(\Omega_1 \cup \Omega_2)$. Moreover, we denote by n_l the maximum number of elements $T \in \mathcal{T}_h$, where the interface is approximated linearly. For the locally modified finite element solution $u_h \in V_h$ to (7) it holds*

$$|||u - u_h||| \leq \left(c_l n_l^{\frac{1}{2}} |\ln(h)|^{1/2} + c_q \right) h_P^2 \|u\|_{H^3(\Omega_1 \cup \Omega_2)}, \tag{29}$$

$$\|u - u_h\|_\Omega \leq \left(c_l n_l |\ln(h)|^{1/2} + c_q \right) h_P^3 \|u\|_{H^3(\Omega_1 \cup \Omega_2)}. \tag{30}$$

where c_l and c_q are generic constants that correspond to patches with a linear and a quadratic interface approximation, respectively. For $u \in W^{2,\infty}(\Omega_1 \cup \Omega_2)$ we have further

$$|||u - u_h||| \leq \left(c_l n_l^{\frac{1}{2}} + c_q \right) h_P^2 \left(\|u\|_{H^3(\Omega_1 \cup \Omega_2)} + \|u\|_{W^{2,\infty}(\Omega_1 \cup \Omega_2)} \right). \tag{31}$$

Proof. (i) First, we have the following perturbed Galerkin orthogonality by subtracting (7) from (4)

$$a(u, \phi_h) - a_h(u_h, \phi_h) = (f - f_h, \phi_h)_\Omega \qquad \forall \phi_h \in V_h. \tag{32}$$

We start by estimating the right-hand side in (32). The difference $f - f_h$ vanishes everywhere besides on S_h. We have

$$(f - f_h, \phi_h)_\Omega = (f - f_h, \phi_h)_{S_h} \leq (\|f_1\|_{S_h} + \|f_2\|_{S_h}) \|\phi_h\|_{S_h},$$

where f_i denotes a smooth extension of $f|_{\Omega_i}$ to $\Omega, i = 1, 2$.
We split the region S_h into parts with a quadratic interface approximation $S_{h,\mathrm{qu}}$ and parts with a linear approximation. Equations (13) and (14) yield

$$\begin{aligned} \|f_i\|_{S_{h,\mathrm{qu}}} + \|f_i\|_{S_{h,\mathrm{lin}}} &\leq c h_P \|f\|_{H^1(\Omega_1 \cup \Omega_2)} \leq c h_P \|u\|_{H^3(\Omega_1 \cup \Omega_2)} \quad \text{and} \\ \|f_i\|_{S_{h,\mathrm{qu}}} + \|f_i\|_{S_{h,\mathrm{lin}}} &\leq \left(c h_P^{\frac{3}{2}} + c n_l^{\frac{1}{2}} h_P^{\frac{3}{2}} |\ln(h)|^{1/2} \right) \|f\|_{H^1(\Omega_1 \cup \Omega_2)} \\ &\leq \left(c h_P^{\frac{3}{2}} + c n_l^{\frac{1}{2}} h_P^{\frac{3}{2}} |\ln(h)|^{1/2} \right) \|u\|_{H^3(\Omega_1 \cup \Omega_2)}. \end{aligned} \tag{33}$$

The second estimate yields

$$(f - f_h, \phi_h)_\Omega \leq c h_P^{\frac{3}{2}} \left(1 + n_l^{\frac{1}{2}} |\ln(h)|^{1/2} \right) \|u\|_{H^3(\Omega_1 \cup \Omega_2)} \|\phi_h\|_{S_h}. \tag{34}$$

(ii) For the energy norm estimate we start by splitting into interpolatory and discrete part

$$|||u - u_h||| \leq \left\| \nu_h^{1/2} \nabla (\tilde{u} - I_h u) \right\|_\Omega + \left\| \nu_h^{1/2} \nabla (I_h u - u_h) \right\|_\Omega. \tag{35}$$

The interpolatory part has already been estimated in Lemma 5. For the second term in (35), we use the perturbed Galerkin orthogonality (32) with $\varphi_h := I_h u - u_h$

$$\begin{aligned}\left\|\nu_h^{1/2}\nabla(I_h u - u_h)\right\|_\Omega^2 &= (\nu_h\nabla(I_h u - u_h), \nabla(I_h u - u_h))_\Omega \\ &= (\nu_h\nabla I_h u - \nu\nabla u, \nabla(I_h u - u_h))_\Omega + (f - f_h, I_h u - u_h)_\Omega. \end{aligned} \tag{36}$$

We split the first part in (36) further

$$\begin{aligned}(\nu_h\nabla I_h u - \nu\nabla u, \nabla(I_h u - u_h))_\Omega &= (\nu_h\nabla(I_h u - \tilde{u}), \nabla(I_h u - u_h))_\Omega \\ &\quad + (\nu_h\nabla\tilde{u} - \nu\nabla u, \nabla(I_h u - u_h))_\Omega. \end{aligned} \tag{37}$$

For the first part, we use (24) to get

$$(\nu_h\nabla(I_h u - \tilde{u}), \nabla(I_h u - u_h))_\Omega \le c h_P^2 \left(n_l^{\frac{1}{2}} |\ln(h)|^{1/2} + 1\right) \|u\|_{H^3(\Omega_1\cup\Omega_2)} \left\|\nu_h^{1/2}\nabla(I_h u - u_h)\right\|_\Omega. \tag{38}$$

The integrand in the second term on the right-hand side of (37) vanishes everywhere besides on S_h. We obtain by the Sobolev imbedding $H^3(\Omega_i) \subset W^{1,\infty}(\Omega_i)$, the continuity of the extension (19), (11) and (10) from Lemma 3

$$\begin{aligned}(\nu_h\nabla\tilde{u} - \nu\nabla u, \nabla(I_h u - u_h))_\Omega &= (\nu_h\nabla\tilde{u} - \nu\nabla u, \nabla(I_h u - u_h))_{S_h} \\ &\le c\left(\|\nabla\tilde{u}\|_{S_h} + \|\nabla u\|_{S_h}\right) \left\|\nu_h^{1/2}\nabla(I_h u - u_h)\right\|_{S_h} \\ &\le c|S_h|^{1/2}\|u\|_{W^{1,\infty}(\Omega_1\cup\Omega_2)} h_P^{1/2} \left\|\nu_h^{1/2}\nabla(I_h u - u_h)\right\|_{S_T} \\ &\le c h_P^2 \left(n_l^{1/2} + 1\right) \|u\|_{H^3(\Omega_1\cup\Omega_2)} \left\|\nu_h^{1/2}\nabla(I_h u - u_h)\right\|_\Omega. \end{aligned} \tag{39}$$

For the second term in (36), we use (33) and (13)

$$\begin{aligned}(f - f_h, I_h u - u_h)_\Omega &\le c h_P \|u\|_{H^3(\Omega_1\cup\Omega_2)} \|I_h u - u_h\|_{S_h} \\ &\le c h_P^2 \|u\|_{H^3(\Omega_1\cup\Omega_2)} \left\|\nu_h^{1/2}\nabla(I_h u - u_h)\right\|_{\Omega_1\cup\Omega_2}. \end{aligned}$$

Combining the estimates, we obtain

$$\left\|\nu_h^{\frac{1}{2}}\nabla(I_h u - u_h)\right\|_{\Omega_1\cup\Omega_2} \le c h_P^2 \|u\|_{H^3(\Omega_1\cup\Omega_2)}.$$

This completes the proof of (29). The proof of (31) follows exactly the same lines, with the only difference that we use (25) instead of (24) in (38) to get

$$(\nu_h\nabla(I_h u - \tilde{u}), \nabla(I_h u - u_h))_\Omega \le c h_P^2 \left(n_l^{\frac{1}{2}} + 1\right) \|u\|_{W^{2,\infty}(\Omega_1\cup\Omega_2)} \left\|\nu_h^{1/2}\nabla(I_h u - u_h)\right\|_\Omega. \tag{40}$$

(iii) To estimate the L^2-norm error, we define the following adjoint problem. Let $z \in H_0^1(\Omega)$ be the solution of

$$(\nu\nabla\varphi, \nabla z) = \|e_h\|^{-1}(e_h, \varphi)_\Omega \qquad \forall\varphi \in H_0^1(\Omega).$$

The solution z lies in in $H_0^1(\Omega) \cap H^2(\Omega_1 \cup \Omega_2)$ and satisfies

$$\|z\|_{H^2(\Omega_1\cup\Omega_2)} \le c_s.$$

By choosing $\varphi = u - u_h = e_h$ and adding and subtracting $\nu_h\nabla u_h$, we have

$$\|e_h\| = (\nu\nabla e_h, \nabla z)_\Omega = (\nu\nabla u - \nu_h\nabla u_h, \nabla z)_\Omega + ((\nu_h - \nu)\nabla u_h, \nabla z)_\Omega. \tag{41}$$

For the second term in (41), we have

$$((\nu_h - \nu)\nabla u_h, \nabla z)_\Omega = ((\nu_h - \nu)\nabla u_h, \nabla z)_{S_h} \leq C\big(\|\nu_h \nabla u_h\|_{S_h} \|\nabla z\|_{S_h}\big). \tag{42}$$

We split the first term on the right-hand side further and use the bound for the energy norm error as well as (14) (Lem. 3)

$$\begin{aligned}\|\nu_h \nabla u_h\|_{S_h} &\leq \|\nu_h \nabla (u_h - u)\|_{S_h} + \|\nu \nabla u\|_{S_h} \\ &\leq c\Big(n_l^{\frac{1}{2}} + 1\Big) h_P^{3/2} \|u\|_{H^3(\Omega_1 \cup \Omega_2)}.\end{aligned}$$

For the last term in (42), we obtain from (13) and (14)

$$\begin{aligned}\|\nabla z\|_{S_{h,\mathrm{lin}}} &\leq c h_P^{\frac{3}{2}} \Big(n_l^{\frac{1}{2}} |\ln(h)|^{1/2} + 1\Big) \|z\|_{H^2(\Omega_1 \cup \Omega_2)} \leq c h_P^{\frac{3}{2}} \Big(n_l^{\frac{1}{2}} |\ln(h)|^{1/2} + 1\Big), \\ \|\nabla z\|_{S_{h,\mathrm{qu}}} &\leq c h_P^{\frac{3}{2}} \|z\|_{H^2(\Omega_1 \cup \Omega_2)} \leq c h_P^{\frac{3}{2}}.\end{aligned}$$

Altogether, we obtain for the second term in (41)

$$((\nu_h - \nu)\nabla u_h, \nabla z)_\Omega \leq c h_P^3 \Big(n_l |\ln(h)|^{1/2} + 1\Big) \|u\|_{H^3(\Omega_1 \cup \Omega_2)}. \tag{43}$$

Concerning the first term in (41), we add and substract the interpolant $\nabla I_h z$, as well as $\pm \nu_h \tilde{u}$

$$\begin{aligned}(\nu \nabla u - \nu_h \nabla u_h, \nabla z)_\Omega = (\nu \nabla u - \nu_h \nabla \tilde{u}, \nabla (z - I_h z))_\Omega + (\nu_h \nabla (\tilde{u} - u_h), \nabla (z - I_h z))_\Omega \\ + (\nu \nabla u - \nu_h \nabla u_h, \nabla I_h z)_\Omega.\end{aligned} \tag{44}$$

For the first term on the right-hand side, we obtain as in (39)

$$(\nu \nabla u - \nu_h \nabla \tilde{u}, \nabla (z - I_h z))_\Omega \leq c h_P^{3/2} \Big(n_l^{1/2} + 1\Big) \|u\|_{H^3(\Omega_1 \cup \Omega_2)} \Big\|\nu_h^{1/2} \nabla (z - I_h z)\Big\|_{S_h}$$

We estimate the latter norm using (13), (14) and (23)

$$\begin{aligned}\Big\|\nu_h^{1/2} \nabla (z - I_h z)\Big\|_{S_h} &\leq c h_P^{3/2} \Big(n_l^{1/2} |\ln(h)|^{1/2} + 1\Big) \|z - I_h z\|_{H^2(\Omega_1 \cup \Omega_2)} \\ &\leq c h_P^{3/2} \Big(n_l^{1/2} |\ln(h)|^{1/2} + 1\Big).\end{aligned}$$

The second term in (42) is easily estimated with the bound for the energy norm and the interpolation error (23)

$$(\nu_h \nabla (\tilde{u} - u_h), \nabla (z - I_h z))_\Omega \leq c\Big(n_l^{\frac{1}{2}} + 1\Big) h_P^3 \|u\|_{H^3(\Omega_1 \cup \Omega_2)}.$$

For the third term in (44), we use the perturbed Galerkin orthogonality (32)

$$\begin{aligned}(\nu \nabla u - \nu_h \nabla u_h, \nabla I_h z)_\Omega &= (f - f_h, I_h z)_{S_h} \\ &\leq \|f_1 - f_2\|_{S_{h,\mathrm{lin}}} \|I_h z\|_{S_{h,\mathrm{lin}}} + \|f_1 - f_2\|_{S_{h,\mathrm{qu}}} \|I_h z\|_{S_{h,\mathrm{qu}}}.\end{aligned} \tag{45}$$

For the first part in both terms, we use (13) and (14), respectively

$$\begin{aligned}\|f_1 - f_2\|_{S_{h,\mathrm{lin}}} + \|f_1 - f_2\|_{S_{h,\mathrm{qu}}} &\leq c h_P^{\frac{3}{2}} \Big(n_l^{\frac{1}{2}} |\ln(h)|^{1/2} + 1\Big) \|f\|_{H^1(\Omega_1 \cup \Omega_2)} \\ &\leq c h_P^{\frac{3}{2}} \Big(n_l^{\frac{1}{2}} |\ln(h)|^{1/2} + 1\Big) \|u\|_{H^3(\Omega_1 \cup \Omega_2)}.\end{aligned}$$

For the remaining terms in (45), it is sufficient to consider the smallness of $|S_h|$, a Sobolev imbedding and the continuity of the extension (19)

$$\|I_h z\|_{S_{h,\mathrm{lin}}} \leq |S_{h,\mathrm{lin}}|^{\frac{1}{2}} \|I_h z\|_{L^\infty(\Omega)} \leq c n_l^{\frac{1}{2}} h_P^{\frac{3}{2}} \|z\|_{L^\infty(\Omega)} \leq c n_l^{\frac{1}{2}} h_P^{\frac{3}{2}} \|z\|_{H^2(\Omega_1 \cup \Omega_2)} \leq c n_l^{\frac{1}{2}} h_P^{\frac{3}{2}}$$
$$\|I_h z\|_{S_{h,\mathrm{qu}}} \leq |S_{h,\mathrm{qu}}|^{\frac{1}{2}} \|I_h z\|_{L^\infty(\Omega)} \leq c h_P^{\frac{3}{2}}.$$

Altogether this yields the following estimate for the term in (44), which completes the proof of the L^2-norm estimate

$$(\nu \nabla u - \nu_h \nabla u_h, \nabla z)_\Omega \leq c h_P^3 \Big(n_l |\ln(h)|^{1/2} + 1 \Big) \|u\|_{H^3(\Omega_1 \cup \Omega_2)}.$$

□

Remark 3 (Energy norm)**.** There are different possibilities to choose the energy norm in Theorem 2. The result (29) could also be shown in the corresponding norm defined on the continuous subdomains Ω_1 and Ω_2

$$|||u - u_h|||_2 := \left(\left\| \nu_1^{1/2} \nabla \big(u_1 - \tilde{u}_h^1\big) \right\|_{\Omega_1}^2 + \left\| \nu_2^{1/2} \nabla \big(u_2 - \tilde{u}_h^2\big) \right\|_{\Omega_2}^2 \right)^{1/2}, \tag{46}$$

where $u_i = u|_{\Omega_i}$ and $\tilde{u}_h^i$ denote the canonical extensions of $u_h^i := u_h|_{\Omega_h^i}$ to Ω_i. If one would consider the norm

$$|||u - u_h|||_3 := \left\| \nu_h^{1/2} \nabla (u - u_h) \right\|_\Omega \tag{47}$$

a reduced order of convergence, namely $\mathcal{O}(h_P^{\frac{3}{2}})$ would result, even for a fully quadratic interface approximation ($n_l = 0$). The reason is that ∇u shows a discontinuity across Γ, while ∇u_h is discontinuous across the discrete interface Γ_h. Hence, the error in the gradient is $\mathcal{O}(1)$ in the strip S_h between the interfaces, which is of size $|S_h|^{1/2} = \mathcal{O}(h_P^{3/2})$. This bound is already optimal in the estimate for $\|\nabla(u - \pi u)\|_\Omega$ in (22).

We have chosen the discrete energy norm $|||u - u_h|||$ in Theorem 2, as this is the only norm, which can be easily evaluated by numerical quadrature. A quadrature formula that evaluates the norms (46) or (47) accurately would need to resolve the strip S_h, which is non-trivial. Any standard approximation, such as a summed midpoint rule would lead to an additional quadrature error of $\mathcal{O}(h_P^{3/2})$, which would dominate the overall error.

Remark 4 (Regularity)**.** We have assumed the regularity $u \in H^3(\Omega_1 \cup \Omega_2)$ (resp. $u \in W^{2,\infty}(\Omega_1 \cup \Omega_2)$) in Theorem 2. This is guaranteed if both subdomains Ω_1 and Ω_2 are smooth (precisely $W^{3,\infty}$) and the right-hand side has regularity $f \in H^1(\Omega_1 \cup \Omega_2)$ (resp. $f \in L^\infty(\Omega_1 \cup \Omega_2)$). In this work the overall domain Ω is assumed polygonal in order to avoid additional technicalities associated with the approximation of exterior curved boundaries. For the latter we refer to the literature, for example [66].

4. Implementation

The locally modified finite element method is based on a patch-wise parametric approach. Let $\mathcal{T}_P$ be the triangulation in patches. We denote by $P \in \mathcal{T}_P$ the patches, which are quadrilaterals with 25 degrees of freedom (see Fig. 6). Depending on the location of the interface, we have two kinds of patches:

– If a patch is not cut by the interface, we divide it into four quadrilaterals $T_1, \ldots, T_4$. In this case we take the standard space of piecewise biquadratic functions as follows:

$$\hat{Q} = \Big\{ \phi \in C(\hat{P}), \phi|_{\hat{T}_i} \in \mathrm{span}\big\{1, x, y, x^2, xy, y^2, xy^2, x^2 y, x^2 y^2\big\},\ i = 1, \ldots, 4 \Big\},$$

where $\hat{P}$ is the reference patch on the unit square $(0,1)^2$ consisting of the four quadrilaterals $\hat{T}_1, \ldots, \hat{T}_4$.

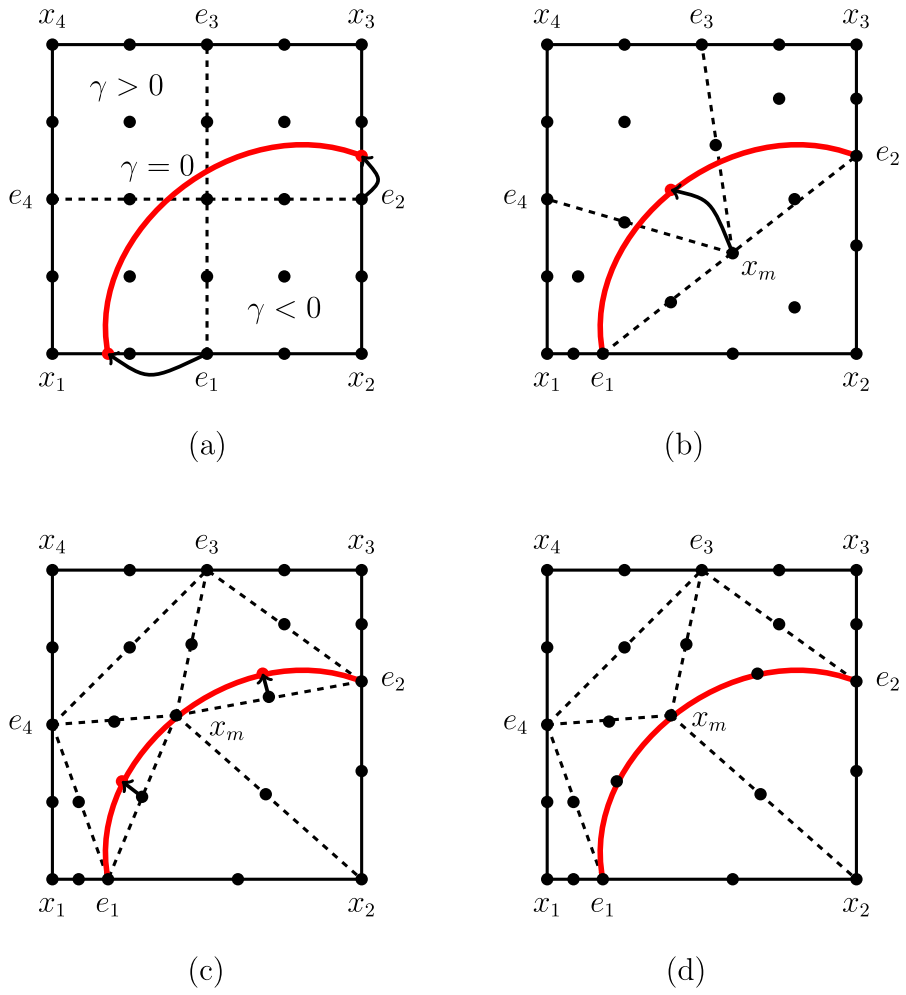


FIGURE 6. Rearrangement of the Lagrangian points on the interface. (a) Shifting cut points. (b) Shifting a middle point. (c) Shifting other points. (d) Points on the interface.

– If the patch is cut by the interface, we divide into eight triangles $T_1, \ldots, T_8$. Here we define the space of piecewise quadratic functions as follows:

$$\hat{Q} = \Big\{\phi \in C(\hat{P}),\ \phi|_{\hat{T}_i} \in \operatorname{span}\{1, x, y, x^2, xy, y^2\},\ i = 1, \ldots, 8\Big\},$$

where the reference patch $\hat{P}$ consists of eight triangles $\hat{T}_1, \ldots, \hat{T}_8$.

In both cases, we have locally 25 basis functions in each patch (see Fig. 6)

$$Q(P) := \operatorname{span}\{\phi_i\},\ \phi_i := \hat{\phi}_i \circ \hat{\xi}_P^{-1},\ \ i = 1, \ldots, 25.$$

$\hat{\xi}_P \in \hat{Q}$ is the reference patch map, which is defined in an isoparametric way by

$$\hat{\xi}_P(\hat{\boldsymbol{x}}) := \sum_{j=1}^{25} y_j\, \hat{\phi}_j \tag{48}$$

for the 25 vertices $y_i, i = 1, \ldots, 25$ of P.

4.1. Definining the patch type and movement of mesh nodes

We assume that the interface is given as zero level-set of an implicit level-set function $\gamma(\boldsymbol{x})$

$$\gamma(\boldsymbol{x}) = 0 \quad \Leftrightarrow \quad \boldsymbol{x} \in \Gamma.$$

The patch type and the edges that are cut are determined by the sign of $\gamma(\boldsymbol{x}_i)$ in the exterior vertices $\boldsymbol{x}_1, \dots, \boldsymbol{x}_4$, see Figure 6. An edge e is cut, if $\gamma(\boldsymbol{x}_1) \cdot \gamma(\boldsymbol{x}_2) < 0$ for its two end points $\boldsymbol{x}_1, \boldsymbol{x}_2$. The intersection of the interface with the edge can the be found by applying Newton's method locally to find the zero r of

$$\gamma\big(\boldsymbol{x}_1 + r(\boldsymbol{x}_2 - \boldsymbol{x}_1)\big) = 0, \tag{49}$$

see Figure 6a. The edge midpoints $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ will be moved to the respective position $\boldsymbol{x}_1 + r(\boldsymbol{x}_2 - \boldsymbol{x}_1)$. Next, we define a preliminary coordinate for the midpoint of the patch $\boldsymbol{x}_m$ as the midpoint of a segment $\boldsymbol{e}_1\boldsymbol{e}_2$, see Figure 6b. For a second-order interface approximation, it is necessary to move $\boldsymbol{x}_m$ to the interface Γ in the configurations A to D. We use again Newton's method to move $\boldsymbol{x}_m$ to the interface along a normal line, see Figure 6c. Second, we also move the midpoints of the segments $\boldsymbol{e}_1\boldsymbol{x}_m$ and $\boldsymbol{x}_m\boldsymbol{e}_2$ analogously, see Figure 6d. Finally, we need to specify a criteria to ensure that the resulting sub-triangles with curved boundaries fulfill a maximum angle condition. Details are given in Appendix C.

Remark 5. A disadvantage of the modified second-order finite element method described above is that the stiffness matrix can be ill-conditioned for certain anisotropies. In particular, the condition number depends not only on the mesh size, but also on how the interface intersects the triangulation (*e.g.* $s, r \to 0$). In Section 5 we consider two examples, where the condition number of the stiffness matrix is not bounded. For this reason a hierarchical finite element basis was introduced in [1] for linear finite elements and it was shown that the condition number of the stiffness matrix satisfies the usual bound $\mathcal{O}(h_P^{-2})$ with a constant that does not depend on the position of the interface. We extend this approach to the second-order finite element method below. We will see that the condition number for a scaled hierarchical basis is reduced significantly, although we can not guarantee the optimal bound for the method presented here.

Remark 6 (Comparison with unfitted finite element methods)**.** In contrast to unfitted finite element methods (*e.g.* [40]), continuity can be imposed strongly within the finite element spaces in the locally finite element method, while in unfitted methods a weak imposition based on Nitsche's method is typically used. Thus, an advantage of the locally modified finite element method is that it is parameter-free. The most tidious task in the implementation of unfitted finite element methods is the construction of suitable quadrature formulas. Usually, the cut cells are sub-divided into sub-cells [43], similarly to the subdivision used within the locally modified finite element method. For the purpose of quadrature no maximum angle condition is needed, which is required for the fitted method. This might be considered as an advantage of the unfitted approach, in particular concerning three dimensional problems. As a remeady in the fitted method, one could allow to move exterior patch vertices in certain "pathological" situations, as discussed in Remark 2.

5. Numerical examples

The higher order parametric finite element method is based on the finite element framework *Gascoigne 3d* [67]. The source code is freely available at `https://www.gascoigne.de` and published as Zenodo repository [68]. For reproducibility of the numerical results, the following two configurations are implemented and described in a separate Zenodo repository [69].

5.1. Example 1

We consider a square domain $\Omega = (-2, 2)^2$. The domain is split into two domains Ω_1 and Ω_2 by the interface $\Gamma = \{(x, y) \in \Omega \,|\, l(x, y) = 0\}$ with level-set function $l(x, y) = y - 2(x + \delta h)^2 + 0.5$, where $\delta \in [0, 1]$ and h is the

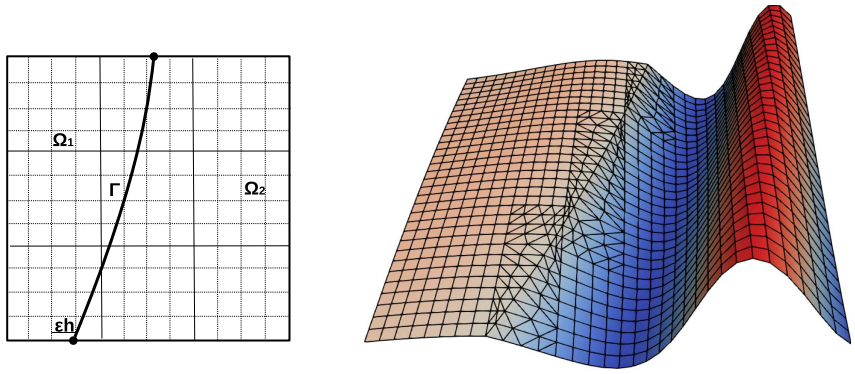


FIGURE 7. Example 1. *Left*: Configuration of the test problem. *Right*: Sketch of the exact solution.

TABLE 1. Example 1. Errors in the L^2-norm and the discrete energy norm, including an estimated order of convergence which is computed from two consecutive values in each row for Example 1 and $\delta = 0$.

h	L^2-error	EOC	Energy error	EOC
1/32	1.74×10^{-4}	–	2.08×10^{-2}	–
1/64	2.13×10^{-5}	3.023	5.22×10^{-3}	1.998
1/128	2.65×10^{-6}	3.006	1.31×10^{-3}	1.999
1/256	3.31×10^{-7}	3.004	3.26×10^{-4}	2.000

mesh size. We take $\nu_1 = 4$ and $\nu_2 = 1$ and choose the exact solution as

$$u(x,y) = \begin{cases} \frac{1}{\nu_1}\sin(l), & \text{in } \Omega_1, \\ \frac{1}{\nu_2}\sin(l), & \text{in } \Omega_2, \end{cases}$$

by setting the right-hand side $f_i = -\nu_i \Delta u$ and Dirichlet boundary data accordingly. We vary $\delta \in [0,1]$, such that this example includes different configurations with arbitrary anisotropies. The subdomains and the exact solution for this example are shown in Figure 7.

In this example the interface could be resolved with second order on all refinement levels ($n_l = 0$). Table 1 shows the discrete energy norm error $|||u - u_h|||$ and the L^2-norm error as well as estimated convergence orders on several levels of global mesh refinement for the fixed parameter $\delta = 0$. According to the *a priori* error estimate in Theorem 2, we observe fully quadratic convergence in the discrete energy norm and fully cubic convergence in the L^2-norm. In Figure 8, we plot the discrete energy norm error and the L^2-norm error for $\delta \in [0,1]$ on several levels of global mesh refinement and observe that the error is bounded independently of δ.

In Figure 9, we show how the condition number depends on the parameter $\delta \in [0,1]$ by moving the interface. We get the largest condition numbers at $\delta = 0.84$. Furthermore, we show a zoom-in of the numbers for $\delta \in [0.83, 0.85]$ in Figure 9, right. We see that the condition number is reduced by a factor of 100 using a scaled hierarchical basis, but that is not necessarily bounded for arbitrary anisotropies.

5.2. Example 2

We consider a square domain $\Omega = (-2,2)^2$ that is split into a ball $\Omega_1 = B_r(x_0, y_0)$ with $r = 0.3$ and $(x_0, y_0) = (1 + \delta h, 1.2)$, where $\delta \in [0,1]$, and $\Omega_2 = \Omega \setminus \bar{\Omega}_1$. We take the exact solution as in example 1, with the

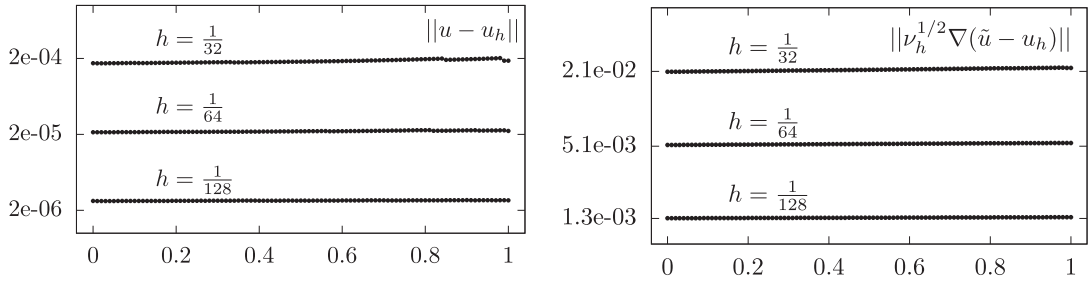


FIGURE 8. Example 1. L^2-norm and discrete energy norm errors for Example 1 with $x = 1.0+\delta h$ and $\delta \in [0, 1]$.

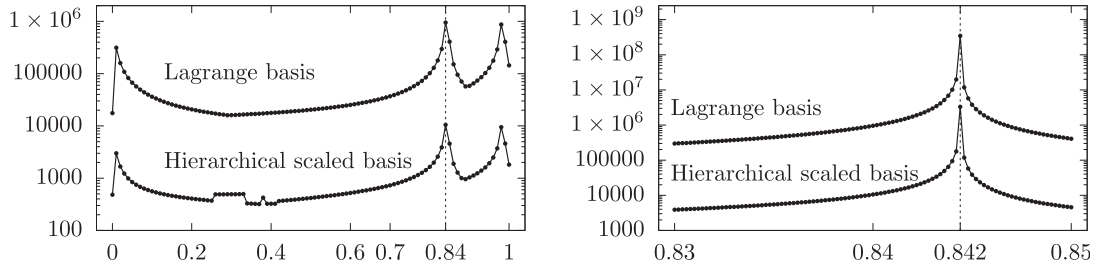


FIGURE 9. Example 1. Condition number of the stiffness matrix depending on the position of the interface δ. Comparison of the standard Lagrangian basis and a scaled hierarchical basis for $h = 1/32$. *Left*: $\delta \in [0, 1]$. *Right*: zoom-in for $\delta \in [0.83, 0.85]$.

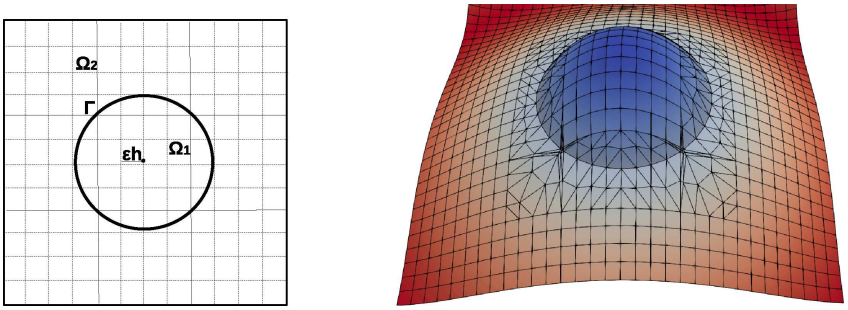


FIGURE 10. Example 2. *Left*: configuration of the test problem. *Right*: sketch of the exact solution

level set function replaced by $l(x, y) = (x-x_0)^2+(y-y_0)^2-r^2$. In Figure 10 we show the configuration and the exact solution of this example. For different $\delta \in [0, 1]$, this example includes all configurations A–E introduced above with different anisotropies.

The L^2-norm and the discrete energy norm errors are shown in Figure 11 for $\delta \in [0, 1]$ on several levels of global mesh refinement. We observe convergence in both norms for $\delta > 0$. The errors vary slightly depending on δ. Its magnitude depends mainly on the number of linearly approximated elements (n_l): We have $n_l = 0$ for $\delta = 0$ on all mesh levels, while $n_l > 0$ for all other values of δ. We observe that the errors increase from $\delta = 0$ to $\delta = 0.01$, as n_l increases from 0 to 8. Moreover, the number of linearly approximated elements increases for $h = 1/64$ once more, from $n_l = 8$ to $n_l = 16$ in the range $\delta \in [0.74, 0.81]$. Again, we observe a slight increase

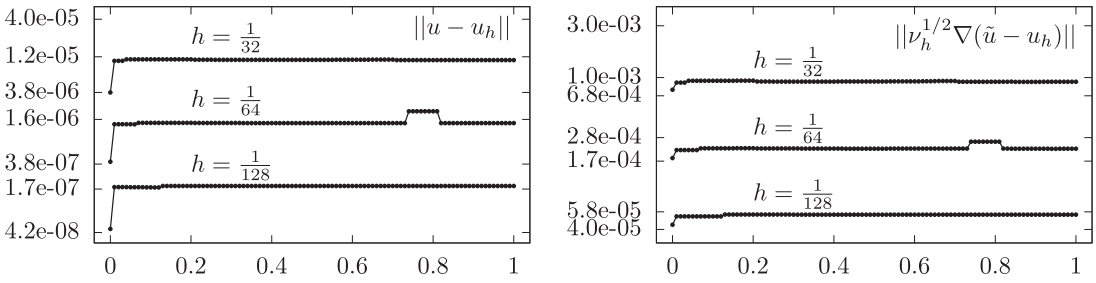


FIGURE 11. Example 2. L^2-norm and discrete energy-norm errors depending on $x = 1.0 + \delta h$ with $\delta \in [0, 1]$.

TABLE 2. Example 2. L^2-norm and modified energy norm errors, and convergence order for $\delta_0 = 0$ ($n_l = 0$).

h	L^2-error	EOC	Energy error	EOC
1/32	3.44×10^{-6}	–	4.36×10^{-4}	–
1/64	3.85×10^{-7}	3.159	9.69×10^{-5}	2.170
1/128	4.56×10^{-8}	3.078	2.28×10^{-5}	2.085
1/256	5.54×10^{-9}	3.040	5.52×10^{-6}	2.048

in the magnitude of the error within this range. This indicates that the constant $c_l n_l^{1/2}$ corresponding to the linearly approximated part in (31) is larger than the constant c_q arising from the quadratically approximated elements.

Tables 2–4 show the L^2-norm and the discrete energy norm errors obtained on several levels of global mesh refinement for the fixed positions $x_0 = 1.0 + \frac{\delta_0}{64}$ of the midpoint, with $\delta_0 \in \{0, 0.01, 0.8\}$, which results in three different cases ($n_l = 0, n_l = 8$ and $n_l = 16$) for $h = 1/64$.

In Table 2 ($\delta_0 = 0$) we observe fully quadratic (resp. cubic convergence) in the discrete energy norm (resp. the L^2-norm) as shown in Theorem 2, as no linearly approximated elements are present. This changes slightly for the other values of δ_0, see Tables 3 and 4.

In Table 3 ($\delta_0 = 0.01$), we see that 8 linearly approximated elements were required on all mesh levels. The convergence order in the discrete energy norm seems to be fully quadratic (according to (31)), while in the L^2-norm error the logarithmic factor $|\ln(h)|^{1/2}$ leads to a slightly reduced convergence, as predicted in Theorem 2.

For $\delta_0 = 0.8$, the number n_l increases from 8 to 16 between the coarsest and the second-coarsest refinement level and stays constant from then, see Table 4. This is again reflected in the magnitude of the error: The reduction factor between the coarsest mesh levels lies below 4 in the energy norm, and below 8 in the L^2-norm error, which shows again that the term $c_l n_l^{1/2} |\ln(h)|^{1/2}$ in front of the linearly approximated part is larger than the constant c_q in front of the quadratic counterpart. On the remaining mesh levels, the estimated convergence is again fully quadratic in the discrete energy norm and due to the logarithmic term slightly below 3 in the L^2-norm, in agreement with Theorem 2.

For $\delta_0 = 0.8$ and $h_P = \frac{1}{32}$, we show the resulting finite element mesh in Figure 12, where in 8 of the 18 patches, which are cut by the interface, a linear approximation was required, including a zoom around one linearly approximated patch on the right.

TABLE 3. Example 2. L^2- and discrete energy – norm errors for $\delta_0 = 0.01$, including estimated convergence orders obtained from two consecutive values. PN denotes the number of patches which are cut by the interface and n_l the number of the linear approximated elements.

h	L^2-error	EOC	Energy error	EOC	PN	n_l
1/32	1.05×10^{-5}	–	9.00×10^{-4}	–	18	8
1/64	1.37×10^{-6}	2.943	2.15×10^{-4}	2.066	36	8
1/128	1.81×10^{-7}	2.920	5.27×10^{-5}	2.028	76	8
1/256	2.39×10^{-8}	2.921	1.30×10^{-5}	2.015	154	8

TABLE 4. Example 2. L^2 and discrete energy – norm errors, including an estimated convergence order for $\delta = 0.8$.

h	L^2-error	EOC	Energy error	EOC	PN	n_l
1/32	1.09×10^{-5}	–	9.19×10^{-4}	–	18	8
1/64	2.07×10^{-6}	2.390	2.57×10^{-4}	1.839	36	16
1/128	2.73×10^{-7}	2.924	6.35×10^{-5}	2.016	76	16
1/256	3.59×10^{-8}	2.927	1.58×10^{-5}	2.007	152	16

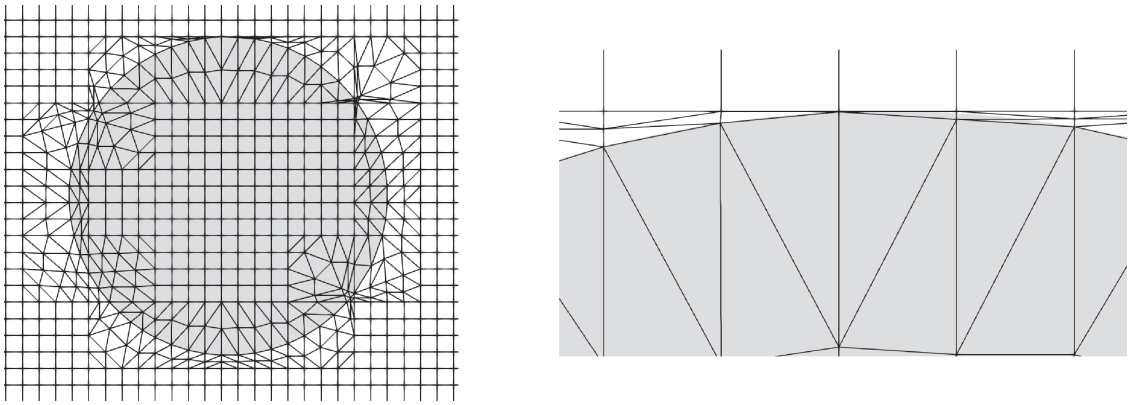


FIGURE 12. Example 2. *Left*: illustration of the sub-elements for $h = 1/32$ and $\delta = 0.8$. *Right*: zoom of the upper part with linearly approximated elements (*top right*).

In Figure 13, we show how the condition numbers depend on the parameter $\delta \in [0, 1]$ when moving the interface. We get the largest condition numbers at $\delta \approx 0.04$ for $h = 1/32$ and at $\delta \approx 0.07$ for $h = 1/64$, respectively. The condition numbers are again reduced by a factor of approx. 100 for the scaled hierarchical basis compared to the standard Lagrangian basis.

6. CONCLUSION

We have presented an extension of the locally modified finite element method for interface problems introduced in [1], to second order. We were able to show optimal-order error estimates of order two in a discrete energy

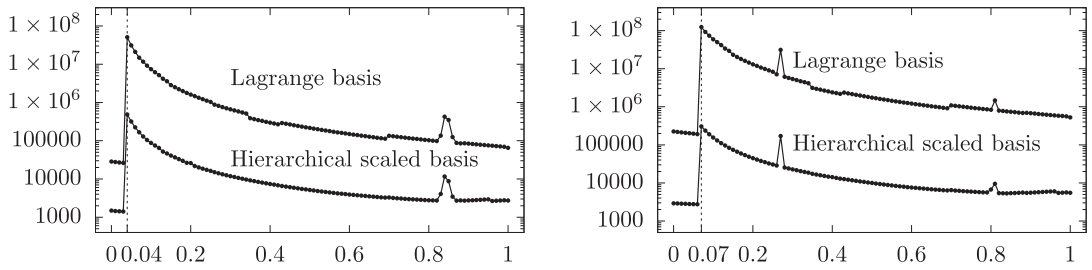


FIGURE 13. Example 2. Condition number of the stiffness matrix depending on the displacement of the circle. Comparison of the Lagrange and hierarchical scaled basis for $h = 1/32$ (*left*) and $h = 1/64$ (*right*).

norm and almost of order three (up to a logarithmic term) in the L^2-norm. Finally, we have presented different numerical examples that illustrate the convergence behaviour and the performance of the method. In future, we plan to extend the method to *inf-sup* stable finite elements for the discretization of interface problems including the Stokes- and Navier–Stokes equations.

APPENDIX A. LINEAR INTERFACE APPROXIMATION

We distinguish between the following five cases, see Figure 2:

- Configuration A: The patch is cut in two opposite nodes.
- Configuration B: The patch is cut at the interior of one edge and in one node.
- Configuration C: The patch is cut at the interior of two opposite edges.
- Configuration D: The patch is cut at the interior of two adjacent edges with $r \in \left(0, \frac{1}{2}\right)$, $s \in \left(\frac{1}{2}, 1\right)$.
- Configuration E: The patch is cut at the interior of two adjacent edges with
  - $r \in (0,1)$ and $s \in \left(0, \frac{1}{2}\right)$
  - $r \in \left(\frac{1}{2}, 1\right)$ and $s \in (0,1)$.

The subdivisions can be anisotropic with the parameters $r, s \in (0,1)$ in the configurations B, C, D and E. These parameters describe the relative position of the intersection points with the interface on the edges. We denote by $\boldsymbol{e}_i \in \mathbb{R}$, $i = 1,2,3,4$, the vertices on the edge. When the interface intersects an edge, we move the corresponding point $\boldsymbol{e}_i$, $i = 1,\dots,4$ on the intersected edge to the point of the intersection (see Fig. 2). If an edge is not intersected by the interface, we take $\boldsymbol{e}_i$ as midpoint of this edge. By $\boldsymbol{x}_m \in \mathbb{R}^2$ we denote the midpoint of the patch, which has different positions depending on the configurations. Precisely, it is chosen as intersection of the line connecting $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$ with the line connecting $\boldsymbol{e}_2$ and $\boldsymbol{e}_4$ for configurations A, C and E. For configuration B we choose the midpoint as intersection of the line connecting $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$ with the line connecting $\boldsymbol{x}_1$ and $\boldsymbol{e}_2$. The midpoint for the configuration D can be chosen as midpoint of the line segment $\boldsymbol{e}_1\boldsymbol{e}_2$.

In all configurations the patch is first divided into four quadrilaterals. We note that each of these has at least one right angle. The sub-quadrilaterals are then further divided into two triangles by either resolving the interface with an interior mesh line or – if this is not necessary – by splitting the largest interior angle of the quadrilateral.

APPENDIX B. PROOF OF LEMMA 1

Proof. First, the patch is split into four sub-quadrilaterals $K_1,\dots,K_4$, each of which is then split into two triangles. If we can show that all angles of the quadrilaterals are bounded by 135°, this applies for the sub-triangles as well. Moreover, if the splitting into triangles in K_i for some $i = 1,\dots,4$ is not determined by the

interface position, we split in such a way that the largest angle of K_i is divided. In this case the bound for the maximum angles of the sub-triangles can be further improved.

We consider the configurations A–E shown in Figure 2 separately. In all cases the angles at the vertices $\boldsymbol{x}_i, i = 1, \dots, 4$ are exactly 90° and the angles at the edge midpoints $\boldsymbol{e}_i, i = 1, \dots, 4$ lie between 45° and 135° (Note that this bound is not sharp, if we divide in an optimal way into sub-triangles).

In configuration A we have two squares and four right-angled triangles. This case is obvious and the maximum angle of the sub-triangles is 90°.

In configuration B and C each quadrilateral $K_i (i = 1, \dots, 4)$ has two right angles, as the positions of $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$, or $\boldsymbol{e}_2$ and $\boldsymbol{e}_4$, respectively, are fixed. Let us consider examplarily the configuration shown in Figure 2b. As mentioned above, the angles in $\boldsymbol{e}_2$ and $\boldsymbol{e}_4$ lie between 45° and 135°. By symmetry the angles around $\boldsymbol{x}_m$ are exactly the same. Therefore, in both configurations all angles of the sub-triangles lie below 135°.

In configuration D, we get one degenerate quadrilateral K_2 with a maximum angle of 180°. As this angle is divided by connecting $\boldsymbol{x}_m$ and $\boldsymbol{x}_2$ we have the following bounds for the triangles of K_2

$$\begin{aligned}\cos(\angle \boldsymbol{e}_1 \boldsymbol{x}_m \boldsymbol{x}_2) &= \frac{(\boldsymbol{e}_1 - \boldsymbol{x}_m) \cdot (\boldsymbol{x}_2 - \boldsymbol{x}_m)}{|\boldsymbol{e}_1 - \boldsymbol{x}_m| \cdot |\boldsymbol{x}_2 - \boldsymbol{x}_m|} = \frac{((r-1), -s) \cdot ((1-r), -s)}{(1-r)^2 + s^2} \\ &= \frac{-(1-r)^2 + s^2}{(1-r)^2 + s^2} \in \Big(-\frac{3}{5}, \frac{3}{5} \Big).\end{aligned}$$

such that $\angle \boldsymbol{e}_1 \boldsymbol{x}_m \boldsymbol{x}_2 \in (53°, 127°)$. The other angles in $\boldsymbol{x}_m$ are bounded above by 90°. The angles in $\boldsymbol{e}_1, \dots, \boldsymbol{e}_4$ are again bounded by 135°.

In configuration E, the angles in $\boldsymbol{e}_1, \dots, \boldsymbol{e}_4$ are all between 63° and 117°. A bound on the angles of the quadrilaterals at $\boldsymbol{x}_m$ is therefore given by $360° - 2 \cdot 63° - 90° = 144°$. This maximum is attained for $r \to 1, s \to 0$ (*cf.* Fig. 2f). The bound is further improved, as in K_1, K_2 and K_3 the largest angles are divided when splitting into sub-triangles, resulting in angles below 90°. For the angle of the subtriangle T_1 at $\boldsymbol{x}_m$ we have

$$\begin{aligned}\cos(\angle \boldsymbol{e}_1 \boldsymbol{x}_m \boldsymbol{e}_2) &= \frac{(\boldsymbol{e}_1 - \boldsymbol{e}_3) \cdot (\boldsymbol{e}_2 - \boldsymbol{e}_4)}{|\boldsymbol{e}_1 - \boldsymbol{e}_3| \cdot |\boldsymbol{e}_2 - \boldsymbol{e}_4|} = \frac{(r - 1/2, -1) \cdot (1, s - 1/2)}{\sqrt{1 + (r-1/2)^2} \cdot \sqrt{1 + (s-1/2)^2}} \\ &= \frac{r - s}{\sqrt{1 + (r-1/2)^2} \cdot \sqrt{1 + (s-1/2)^2}} \in \left(-\frac{1}{\sqrt{5}}, \frac{4}{5} \right)\end{aligned}$$

such that $\angle \boldsymbol{e}_1 \boldsymbol{x}_m \boldsymbol{e}_2 \in (36°, 117°)$. □

Appendix C. Quadratic interface approximation

For the elements with curved boundaries we need to ensure that all elements are allowed in the sense of Assumption 1 (see also Fig. 1) and that the maximum angle condition shown above remains valid. As described in Section 4, we move certain points to the interface in order to obtain a second-order interface approximation. This is possible if the following criteria are satisified. Otherwise, we leave them in their original positions and obtain a first-order interface approximation in the respective element. By $\alpha_\triangle$ we denote the largest angle in a triangle.

In the first step, we move the midpoint of the patch. If this is possible, we shift the other corresponding points in a second step (if possible). We use the following criteria for each configuration.

First step: Move the midpoints

- **Configuration A**: the midpoint of the patch can be moved along the normal line $\boldsymbol{n}$ (see Fig. 2a) if $\alpha_\triangle \leq \alpha_{\max} < 180°$.
- **Configuration B**: the midpoint of the patch can be moved along the line segment $\boldsymbol{e}_1 \boldsymbol{e}_3$, if the relative length $d = \frac{|\boldsymbol{e}_1 - \boldsymbol{x}_m|}{|\boldsymbol{e}_2 - \boldsymbol{e}_1|}$ of the line $\boldsymbol{e}_1 \boldsymbol{x}_m$ (see Fig. 2b) satisfies $\epsilon < d < 1 - \epsilon$ and $\alpha_\triangle \leq \alpha_{\max} < 180°$.

– **Configuration C**: the midpoint of the patch can be moved along the line segment $\boldsymbol{e}_2\boldsymbol{e}_4$, if the parameter d (see Fig. 2c) satisfies $\epsilon < d < 1 - \epsilon$.
– **Configuration D**: the midpoint of the patch can be moved along the normal line $\boldsymbol{n}$ (see Fig. 2d) if $\alpha_{\triangle} \leq \alpha_{\max} < 180°$.
– **Configuration E**: in this configuration we do not need to move the midpoint of the patch (see Figs. 2e and 2f).

Second step: Move other points

– In the second step, we investigate the other two points that need to be moved in order to obtain a second-order interface approximation. These are the points between the midpoint of the patch and the points where exterior edges are intersected. In all configurations, we obtain triangles with one curved edge (see Fig. 3). It can happen that this curved edge intersects other edges of the element T. Thus, we shift the corresponding points along the normal line to the interface, if and only if the curved edge of the triangle does not cut any other edges and $\alpha_{\triangle} \leq \alpha_{\max} < 180°$.

References

[1] S. Frei and T. Richter, A locally modified parametric finite element method for interface problems. *SIAM J. Numer. Anal.* **52** (2014) 2315–2334.
[2] T. Richter, Fluid Structure Interactions: Models, Analysis and Finite Elements. Springer, New York (2017).
[3] S. Frei, *Eulerian finite element methods for interface problems and fluid-structure interactions*. Ph.D. thesis, Heidelberg University, http://www.ub.uni-heidelberg.de/archiv/21590 (2016).
[4] Y. Bazilevs, K. Takizawa and T.E. Tezduyar, Computational Fluid-Structure Interaction: Methods and Applications. John Wiley & Sons (2013).
[5] S. Gross and A. Reusken, Numerical Methods for Two-phase Incompressible Flows. Vol. 40. Springer Science & Business Media (2011).
[6] K. Stein, R. Benney, V. Kalro, T.E. Tezduyar, J. Leonard and M. Accorsi, Parachute fluid–structure interactions: 3-d computation. *Comput. Methods Appl. Mech. Eng.* **190** (2000) 373–386.
[7] C.S. Peskin, Flow patterns around heart valves: a numerical method. *J. Comput. Phys.* **10** (1972) 252–271.
[8] F. Van de Vosse, J. De Hart, C. Van Oijen, D. Bessems, T. Gunther, A. Segal, B. Wolters, J. Stijnen and F. Baaijens, Finite-element-based computational methods for cardiovascular fluid-structure interaction. *J. Eng. Math.* **47** (2003) 335–368.
[9] L. Formaggia, A. Quarteroni and A. Veneziani, Cardiovascular Mathematics: Modeling and Simulation of the Circulatory System. Vol. 1. Springer Science & Business Media (2010).
[10] W.A. Wall and T. Rabczuk, Fluid-structure interaction in lower airways of CT-based lung geometries. *Int. J. Numer. Methods Fluids* **57** (2008) 653–675.
[11] S. Knauf, S. Frei, T. Richter and R. Rannacher, Towards a complete numerical description of lubricant film dynamics in ball bearings. *Comput. Mech.* **53** (2014) 239–255.
[12] S.-R. Hysing, S. Turek, D. Kuzmin, N. Parolini, E. Burman, S. Ganesan and L. Tobiska, Quantitative benchmark computations of two-dimensional bubble dynamics. *Int. J. Numer. Methods Fluids* **60** (2009) 1259–1288.
[13] S. Claus and P. Kerfriden, A CutFEM method for two-phase flow problems. *Comput. Methods Appl. Mech. Eng.* **348** (2019) 185–206.
[14] H. Garcke, K.F. Lam, R. Nürnberg and E. Sitka, A multiphase Cahn–Hilliard–Darcy model for tumour growth with necrosis. *Math. Models Methods Appl. Sci.* **28** (2018) 525–577.
[15] P. Gangl, A local mesh modification strategy for interface problems with application to shape and topology optimization, in Scientific Computing in Electrical Engineering. Springer, Cham (2018) 147–155.
[16] E. Burman, C. He and M.G. Larson, Comparison of shape derivatives using CutFEM for ill-posed Bernoulli free boundary problem. *J. Sci. Comput.* **88** (2021) 1–28.
[17] I. Babuška, The finite element method for elliptic equations with discontinuous coefficients. *Computing* **5** (1970) 207–213.
[18] S. Frei and T. Richter, A second order time-stepping scheme for parabolic interface problems with moving interfaces. *ESAIM: Math. Modell. Numer. Anal.* **51** (2017) 1539–1560.
[19] S. Frei, T. Richter and T. Wick, Long-term simulation of large deformation, mechano-chemical fluid-structure interactions in ALE and fully Eulerian coordinates. *J. Comput. Phys.* **321** (2016) 874–891.
[20] S. Frei, T. Richter and T. Wick, Eulerian techniques for fluid-structure interactions: Part I–Modeling and simulation, in Numerical Mathematics and Advanced Applications-ENUMATH 2013. Springer (2015) 745–753.
[21] S. Frei, T. Richter and T. Wick, Eulerian techniques for fluid-structure interactions: Part II–Applications, in *Numerical Mathematics and Advanced Applications-ENUMATH 2013*. Springer (2015) 755–762.
[22] U. Langer and H. Yang, Numerical simulation of parabolic moving and growing interface problems using small mesh deformation, in Johann Radon Institute for Computational and Applied Mathematics. Bericht-Nr. (2015–2016).

[23] E. Burman, M.A. Fernández and S. Frei, A Nitsche-based formulation for fluid-structure interactions with contact. *ESAIM: Math. Modell. Numer. Anal.* **54** (2020) 531–564.

[24] E. Burman, M.A. Fernández, S. Frei and F.M. Gerosa, A mechanically consistent model for fluid–structure interactions with contact including seepage. *Comput. Methods Appl. Mech. Eng.* **392** (2022) 114637.

[25] S. Frei and T. Richter, An accurate Eulerian approach for fluid-structure interactions, in Fluid-Structure Interaction: Modeling, Adaptive Discretization and Solvers. *Radon Series on Computational and Applied Mathematics*, edited by S. Frei, B. Holm, T. Richter, T. Wick and H. Yang. Walter de Gruyter, Berlin (2017) 69–126.

[26] J. Hoffman, B. Holm and T. Richter, The locally adapted parametric finite element method for interface problems on triangular meshes, in Fluid-Structure Interaction: Modeling, Adaptive Discretization and Solvers. *Radon Series on Computational and Applied Mathematics*, edited by S. Frei, B. Holm, T. Richter, T. Wick and H. Yang. de Gruyter (2017) 41–68.

[27] S. Frei, An edge-based pressure stabilization technique for finite elements on arbitrarily anisotropic meshes. *Int. J. Numer. Methods Fluids* **89** (2019) 407–429.

[28] S. Frei, T. Richter and T. Wick, LocModFE: locally modified finite elements for approximating interface problems in deal. II. *Softw. Impacts* **8** (2021) 100070.

[29] S. Frei, T. Richter and T. Wick, An implementation of a locally modified finite element method for interface problems in deal. II. *Zenodo* (2018). DOI: 10.5281/zenodo.1457758.

[30] U. Langer and H. Yang, Numerical simulation of parabolic moving and growing interface problems using small mesh deformation (2015). Preprint: arXiv:1507.08784 [math.NA].

[31] S. Höllbacher and G. Wittum, A sharp interface method using enriched finite elements for elliptic interface problems. *Numer. Math.* **147** (2021) 759–781.

[32] A. Vogel, S. Reiter, M. Rupp, A. Nägel and G. Wittum, UG 4: A novel flexible software system for simulating PDE based models on high performance computers. *Comput. Visual. Sci.* **16** (2013) 165–179.

[33] N. Moës, J. Dolbow and T. Belytschko, A finite element method for crack growth without remeshing. *Int. J. Numer. Methods Eng.* **46** (1999) 131–150.

[34] C. Daux, N. Moës, J. Dolbow, N. Sukumar and T. Belytschko, Arbitrary branched and intersecting cracks with the extended finite element method. *Int. J. Numer. Methods Eng.* **48** (2000) 1741–1760.

[35] J. Chessa and T. Belytschko, An extended finite element method for two-phase fluids. *J. Appl. Mech.* **70** (2003) 10–17.

[36] T.-P. Fries and T. Belytschko, The extended/generalized finite element method: an overview of the method and its applications. *Int. J. Numer. Methods Eng.* **84** (2010) 253–304.

[37] I. Babuška, U. Banarjee and J.E. Osborn, Generalized finite element methods: main ideas, results, and perspective. *Int. J. Comput. Methods* **1** (2004) 67–103.

[38] K. Cheng and T. Fries, Higher-order XFEM for curved strong and weak discontinuities. *Int. J. Numer. Methods Eng.* **82** (2010) 564–590.

[39] K. Dréau, N. Chevaugeon and N. Moës, Studied X-FEM enrichment to handle material interfaces with higher order finite element. *Comput. Methods Appl. Mech. Eng.* **199** (2010) 1922–1936.

[40] A. Hansbo and P. Hansbo, An unfitted finite element method, based on Nitsche's method, for elliptic interface problems. *Comput. Methods Appl. Mech. Eng.* **191** (2002) 5537–5552.

[41] E. Burman and P. Hansbo, Fictitious domain finite element methods using cut elements: II. A stabilized Nitsche method. *Appl. Numer. Math.* **62** (2012) 328–341.

[42] P. Hansbo, M. Larson and S. Zahedi, A cut finite element method for a Stokes interface problem. *Appl. Numer. Math.* **85** (2014) 90–114.

[43] E. Burman, S. Claus, P. Hansbo, M.G. Larson and A. Massing, CutFEM: discretizing geometry and partial differential equations. *Int. J. Numer. Methods Eng.* **104** (2015) 472–501.

[44] S. Zahedi, A space-time cut finite element method with quadrature in time, in Geometrically Unfitted Finite Element Methods and Applications. Springer (2017) 281–306.

[45] C. Lehrenfeld, High order unfitted finite element methods on level set domains using isoparametric mappings. *Comput. Methods Appl. Mech. Eng.* **300** (2016) 716–733.

[46] C. Lehrenfeld and A. Reusken, Analysis of a high order unfitted finite element method for an elliptic interface problem. *IMA J. Numer. Anal.* **38** (2018) 1351–1387.

[47] C. Lehrenfeld and A. Reusken, L^2-estimates for a high order unfitted finite element method for elliptic interface problems. *J. Numer. Math.* **27** (2018) 85–99.

[48] E. Burman, P. Hansbo and M. Larson, A cut finite element method with boundary value correction. *Math. Comput.* **87** (2018) 633–657.

[49] K. Fidkowski and D. Darmofal, An adaptive simplex cut-cell method for discontinuous Galerkin discretizations of the Navier–Stokes equations, in *AIAA Conference Paper* (2007).

[50] P. Bastian and C. Engwer, An unfitted finite element method using discontinuous Galerkin. *Int. J. Numer. Methods Eng.* **79** (2009) 1557–1576.

[51] R. Massjung, An unfitted discontinuous Galerkin method applied to elliptic interface problems. *SIAM J. Numer. Anal.* **50** (2012) 3134–3162.

[52] P. Areias and T. Belytschko, A comment on the article "A finite element method for simulation of strong and weak discontinuities in solid mechanics" by A. Hansbo and P. Hansbo [Comput. Methods Appl. Mech. Eng. 193 (2004) 3523–3540]. *Comput. Methods Appl. Mech. Eng.* **9** (2006) 1275–1276.

[53] A. Main and G. Scovazzi, The shifted boundary method for embedded domain computations. Part I: Poisson and Stokes problems. *J. Comput. Phys.* **372** (2018) 972–995.
[54] I. Babuška, The finite element method for elliptic equations with discontinuous coefficients. *Computing* **5** (1970) 207–213.
[55] S. Basting and R. Prignitz, An interface-fitted subspace projection method for finite element simulations of particulate flows. *Comput. Methods Appl. Mech. Eng.* **267** (2013) 133–149.
[56] J. Bramble and J. King, A finite element method for interface problems in domains with smooth boundaries and interfaces. *Adv. Comput. Math.* **6** (1996) 109–138.
[57] M. Feistauer and V. Sobotíková, Finite element approximation of nonlinear problems with discontinuous coefficients. *ESAIM: Math. Modell. Numer. Anal.* **24** (1990) 457–500.
[58] A. Ženíšek, The finite element method for nonlinear elliptic equations with discontinuous coefficients. *Numer. Math.* **58** (1990) 51–77.
[59] C. Börgers, A triangulation algorithm for fast elliptic solvers based on domain imbedding. *SIAM J. Numer. Anal.* **27** (1990) 1187–1196.
[60] H. Xie, K. Ito, Z.-L. Li and J. Toivanen, A finite element method for interface problems with locally modified triangulation. *Contemp. Math.* **466** (2008) 179–190.
[61] X. Fang, An isoparametric finite element method for elliptic interface problems with nonhomogeneous jump conditions. *WSEAS Trans. Math.* **12** (2013) 66–75.
[62] S. Omerović and T. Fries, Conformal higher-order remeshing schemes for implicitly defined interface problems. *Int. J. Numer. Methods Eng.* **109** (2017) 763–789.
[63] R. Rangarajan and A. Lew, Universal meshes: a method for triangulating planar curved domains immersed in nonconforming meshes. *Int. J. Numer. Methods Eng.* **98** (2014) 236–264.
[64] K. Tanaka, K. Sekine, M. Mizuguchi and S. Oishi, Estimation of Sobolev-type embedding constant on domains with minimally smooth boundary using extension operator. *J. Inequalities App.* **2015** (2015) 1–23.
[65] E. Stein, Singular Integrals and Differentiability Properties of Functions (PMS-30). Vol. 30. Princeton University Press, Princeton, NJ (2016).
[66] C. Bernardi, Optimal finite-element interpolation on curved domains. *SIAM J. Numer. Anal.* **26** (1989) 1212–1240.
[67] R. Becker, M. Braack, D. Meidner, T. Richter and B. Vexler, The finite element toolkit Gascoigne 3d. https://www.gascoigne.de (2021).
[68] R. Becker, M. Braack, D. Meidner, T. Richter and B. Vexler, The finite element toolkit gascoigne (v1.01). (2021). DOI: 10.5281/zenodo.5574969.
[69] T. Richter and G. Judakova, Locally modified second order finite elements. (2021). DOI: 10.5281/ZENODO.5575064.

Numerische Mathematik (2022) 150:423–478
https://doi.org/10.1007/s00211-021-01264-x

Numerische Mathematik

Eulerian time-stepping schemes for the non-stationary Stokes equations on time-dependent domains

Erik Burman[1] · Stefan Frei[2] · Andre Massing[3,4]

Received: 7 October 2019 / Revised: 30 November 2021 / Accepted: 13 December 2021 /
Published online: 9 January 2022

Abstract
This article is concerned with the discretisation of the Stokes equations on time-dependent domains in an Eulerian coordinate framework. Our work can be seen as an extension of a recent paper by Lehrenfeld and Olshanskii (ESAIM: M2AN 53(2):585–614, 2019), where BDF-type time-stepping schemes are studied for a parabolic equation on moving domains. For space discretisation, a geometrically unfitted finite element discretisation is applied in combination with Nitsche's method to impose boundary conditions. Physically undefined values of the solution at previous time-steps are extended implicitly by means of so-called *ghost penalty* stabilisations. We derive a complete a priori error analysis of the discretisation error in space and time, including optimal $L^2(L^2)$-norm error bounds for the velocities. Finally, the theoretical results are substantiated with numerical examples.

1 Introduction

Flows on moving domains $\Omega(t) \subset \mathbb{R}^d$ $(d = 2, 3)$ need to be considered in many different applications. Examples include particulate flows or flows around moving

✉ Stefan Frei
stefan.frei@uni-konstanz.de

Erik Burman
e.burman@ucl.ac.uk

Andre Massing
andre.massing@ntnu.no ; andre.masssing@umu.se

[1] Department of Mathematics, University College London, Gower Street, London WC1E 6BT, UK

[2] Department of Mathematics and Statistics, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany

[3] Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway

[4] Department of Mathematics and Mathematical Statistics, Umeå University, 90187 Umeå, Sweden

objects like biological or mechanical valves, wind turbines or parachutes. Strongly related problems are fluid-structure interactions or multi-phase flows.

There exists a vast literature on time discretisation of the non-stationary Stokes or Navier–Stokes equations on fixed domains, see for example the classical works of Girault and Raviart [30], Baker et al. [2] and Rannacher and Heywood [39], or more recently Bochev et al. [5] and Burman and Fernández [11,12] in the context of stabilised finite element methods. If the computational domain remains unchanged in each time-step, the same spatial discretisation can be used (unless adaptive mesh refinement is considered) and finite difference schemes based on the method of lines can be applied for time discretisation.

In the case of moderate domain movements, these techniques can be transferred to the moving framework by using the *Arbitrary Lagrangian Eulerian* (ALE) approach [22,24,50]. Here, the idea is to formulate an equivalent system of equations on a fixed reference configuration $\hat{\Omega}$, for example the initial configuration $\Omega(0)$, by means of a time-dependent map $\boldsymbol{T}(t) : \hat{\Omega} \to \Omega(t)$. This technique has been used widely for flows on moving domains, see e.g,. [19,22] and fluid-structure interactions [3,29,53]. The analysis of the time discretisation error is then very similar to the fixed framework, as all quantities and equations are formulated on the same reference domain $\hat{\Omega}$, see e.g. [54]. For a detailed stability analysis of ALE formulations, we refer to Nobile and Formaggia [49] and Boffi and Gastaldi [7].

On the other hand, it is well-known that the ALE method is less practical in the case of large domain deformations [25,53]. This is due to the degeneration of mesh elements both in a finite element and a finite difference context. A re-meshing of the domain $\Omega(t)$ becomes necessary. Moreover, topology changes, for example due to contact of particles within the flow or of a particle with an outer wall [18], are not allowed, as the map between $\hat{\Omega}$ and $\Omega(t)$ can not have the required regularity in this situation.

In such cases an Eulerian formulation of the problem formulated on the moving domains $\Omega(t)$ is preferable. This is also the standard coordinate framework for the simulation of multi-phase flows. In the last years a variety of space discretisation techniques have been designed to resolve curved or moving boundaries accurately. Examples include the cut finite element method [16,17,33,35,44–46] within a fictitious domain approach, extended finite elements [20,31,36,47] or locally fitted finite element techniques [27], to name such a few of the approaches.

Much less analysed is a proper time discretisation of the problem. In the case of moving domains, standard time discretisation based on the method of lines is not applicable in a straight-forward way. The reason is that the domain of definition of the variables changes from time step to time step.

As an example consider the *backward Euler* discretisation of the time derivative within a variational formulation

$$(\partial_t u_h(t_n), \phi_h^n)_{\Omega(t_n)} \approx \frac{1}{\Delta t}(u_h(t_n) - u_h(t_{n-1}), \phi_h^n)_{\Omega(t_n)}.$$

Note that $u_h(t_{n-1})$ is only well-defined on $\Omega(t_{n-1})$, but is needed on $\Omega(t_n)$.

One solution to this dilemma are so-called characteristic-based approaches [38]. Similar time-stepping schemes result when applying the ALE method only locally within one time-step and projecting the system back to the original reference frame after each step [21], or based on Galerkin time discretisations with modified Galerkin spaces [28]. The disadvantage of these approaches is a projection between the domains $\Omega(t_{n-1})$ and $\Omega(t_n)$ that needs to be computed within each or after a certain number of steps.

Another possibility consists of space-time approaches [37,41], where a $d+1$-dimensional domain is discretised if $\Omega(t) \subset \mathbb{R}^d$. The computational requirements of these approaches might, however, exceed the available computational resources, in particular within complex three-dimensional applications. Moreover, the implementation of higher-dimensional discretisations and accurate quadrature formulas pose additional challenges.

A simpler approach has been proposed recently in the dissertation of Schott [56] and by Lehrenfeld and Olshanskii [42]. Here, the idea is to define extensions of the solution $\boldsymbol{u}_h(t_{n-1})$ from the previous time-step to a domain that spans at least $\Omega(t_n)$. On the finite element level these extensions can be incorporated implicitly in the time-stepping scheme by using so-called *ghost penalty* stabilisations [10] to a sufficiently large domain $\Omega_\delta(t_{n-1}) \supset \Omega(t_n)$. These techniques have originally been proposed to extend the coercivity of elliptic bilinear forms from the physical to the computational domain in the context of CutFEM or fictitious domain approaches [10].

While Schott used such an extension explicitly after each time step to define values for $\boldsymbol{u}_h(t_{n-1})$ in mesh nodes lying in $\Omega(t_n)\backslash\Omega(t_{n-1})$, Lehrenfeld and Olshanskii included the extension operator implicitly within each time step by solving a combined discrete system including the extension operator on the larger computational domain $\Omega_\delta(t_n)$. For the latter approach a complete analysis could be given for the corresponding backward Euler time discretisation, showing first-order convergence in time in the spatial energy norm [42]. Moreover, the authors gave hints on how to transfer the argumentation to a backward difference scheme (BDF(2)), which results in second-order convergence. We should also mention that similar time discretisation techniques have been used previously in the context of surface PDEs [43,51], and mixed-dimensional surface-bulk problems [37] on moving domains.

In this work, we apply such an approach to the discretisation of the non-stationary Stokes equations on a moving domain, including a complete analysis of the space and time discretisation errors. Particular problems are related to the approximation of the pressure variable. It is well-known that stability of the pressure is lost in the case of fixed domains, when the discretisation changes from one time-step to another. This can already be observed, when the finite element mesh is refined or coarsened globally at some instant of time, see Besier and Wollner [4] and is due to the fact that the old solution $\boldsymbol{u}_h^{n-1} := \boldsymbol{u}_h(t_{n-1})$ is not discrete divergence free with respect to the new mesh. Possible remedies include the use of Stokes or Darcy projections [4,8] to pass $\boldsymbol{u}_h^{n-1}$ to the new mesh. Our analysis will reveal that similar issues hold true for the case of moving domains, even if the same discretisation is used on $\Omega(t_n) \cap \Omega(t_{n-1})$. The reason is that $\boldsymbol{u}_h^{n-1}$ is discrete divergence-free with respect to $\Omega(t_{n-1})$, but not with respect to $\Omega(t_n)$

$$(\operatorname{div} \boldsymbol{u}_h^{n-1}, \phi_h)_{\Omega(t_{n-1})} = 0, \quad \text{but } (\operatorname{div} \boldsymbol{u}_h^{n-1}, \phi_h)_{\Omega(t_n)} \neq 0$$

for certain $\phi_h \in \mathcal{V}_h$.

For space discretisation, we will use the Cut Finite Element framework [35]. The idea is to discretise a larger domain of simple structure in the spirit of the Fictitious Domain approach. The active degrees of freedom consist of all degrees of freedom in mesh elements with non-empty intersection with $\Omega_\delta(t_n)$. Dirichlet boundary conditions are incorporated by means of Nitsche's method [48].

We will consider both the BDF(1)/backward Euler and the BDF(2) variant of the approach. To simplify the presentation of the analysis, we will neglect geometry approximation errors related to the approximation of curved boundaries and, moreover, focus on the BDF(1) variant. The necessary modifications for the BDF(2) variant will be sketched within remarks. Finally, we will use a duality technique to prove an optimal $L^2(L^2)$-norm estimate for the velocities.

The structure of this article is as follows: In Sect. 2 we introduce the equations and sketch how to prove the well-posedness of the system. Then we introduce time and space discretisation in Sect. 3, including the extension operators and assumptions, that will be needed in the stability analysis of Sect. 4 and the error analysis in Sect. 5. Then, we give some three-dimensional numerical results in Sect. 6. We conclude in Sect. 7.

2 Equations

We consider the non-stationary Stokes equations with homogeneous Dirichlet boundary conditions on a moving domain $\Omega(t) \subset \mathbb{R}^d$, $d = 2, 3$ for $t \in I = [0, t_{\text{fin}}]$

$$\begin{aligned} \partial_t \boldsymbol{u} - \Delta \boldsymbol{u} + \nabla p &= \boldsymbol{f}, & \operatorname{div} \boldsymbol{u} &= 0 \quad \text{in } \Omega(t), \\ \boldsymbol{u} &= 0 \quad \text{on } \partial\Omega(t), & \boldsymbol{u}(x,0) &= \boldsymbol{u}^0(x) \text{ in } \Omega(0). \end{aligned} \tag{1}$$

We assume that the domain motion can be described by a $W^{1,\infty}$-diffeomorphism

$$\boldsymbol{T}(t) : \Omega(0) \to \Omega(t). \tag{2}$$

with the additional regularity

$$\boldsymbol{T} \in W^{1,\infty}(I, W^{2,\infty}(\Omega(0))). \tag{3}$$

and that the initial domain $\Omega(0)$ is piecewise smooth and Lipschitz. In order to formulate the variational formulation we define the spaces

$$\begin{aligned} \mathcal{V}(t) &:= H_0^1(\Omega(t))^d, \qquad \mathcal{L}(t) := L^2(\Omega(t)), \qquad \mathcal{L}_0(t) := L_0^2(\Omega(t)), \\ \mathcal{V}_I &:= \{\boldsymbol{u} \in L^2(I, \mathcal{V}(t))^d,\ \partial_t \boldsymbol{u} \in L^2(I, \mathcal{L}(t)^d)\}, \quad \mathcal{L}_{0,I} := L^2(I, \mathcal{L}_0(t)). \end{aligned} \tag{4}$$

and consider the variational formulation: *Find* $\boldsymbol{u} \in \mathcal{V}_I$, $p \in \mathcal{L}_{0,I}$ such that

$$\begin{aligned}(\partial_t \boldsymbol{u}, \boldsymbol{v})_{\Omega(t)} + \mathcal{A}_S(\boldsymbol{u}, p; \boldsymbol{v}, q) = (\boldsymbol{f}, \boldsymbol{v})_{\Omega(t)} \quad \forall \boldsymbol{v} \in \mathcal{V}(t),\ q \in \mathcal{L}(t) \quad \text{a.e. in } t \in I,\\ \boldsymbol{u}(x, 0) = \boldsymbol{u}^0(x) \qquad \text{a.e. in } \Omega(0),\end{aligned} \tag{5}$$

where

$$\mathcal{A}_S(\boldsymbol{u}, p; \boldsymbol{v}, q) := (\nabla \boldsymbol{u}, \nabla \boldsymbol{v})_{\Omega(t)} - (p, \operatorname{div} \boldsymbol{v})_{\Omega(t)} + (\operatorname{div} \boldsymbol{u}, q)_{\Omega(t)}. \tag{6}$$

We assume that $\boldsymbol{f} \in L^\infty(I, \mathcal{L}(t)^d)$ a.e. in $t \in I$ and $\boldsymbol{u}^0 \in H^1(\Omega(0))^d$.

Remark 2.1 (Boundary conditions) It might seem unnatural at first sight to use homogeneous Dirichlet boundary conditions for a Stokes problem on a moving domain $\Omega(t)$. In fact the assumption that the flow follows the domain motion on $\partial\Omega(t)$ would be a more realistic boundary condition, i.e.

$$\begin{aligned}\partial_t \tilde{\boldsymbol{u}} - \Delta \tilde{\boldsymbol{u}} + \nabla p = \tilde{\boldsymbol{f}}, \qquad \operatorname{div} \tilde{\boldsymbol{u}} = 0 \qquad \text{in } \Omega(t)\\ \tilde{\boldsymbol{u}} = \partial_t \boldsymbol{T}(\boldsymbol{T}^{-1}) \text{ on } \partial\Omega(t).\end{aligned} \tag{7}$$

Note, however, that for a sufficiently smooth map $\boldsymbol{T}(t)$ that fulfils $\operatorname{div}(\partial_t \boldsymbol{T}(\boldsymbol{T}^{-1})) = 0$, one obtains (1) from (7) for $\boldsymbol{u} = \tilde{\boldsymbol{u}} - \partial_t \boldsymbol{T}(\boldsymbol{T}^{-1})$ and $\boldsymbol{f} := \tilde{\boldsymbol{f}} + \partial_t^2 \boldsymbol{T}(\boldsymbol{T}^{-1}) - \Delta(\partial_t \boldsymbol{T}(\boldsymbol{T}^{-1}))$. For this reason and in order to simplify the presentation of the error analysis, we will consider homogeneous Dirichlet conditions in the remainder of this article.

2.1 Well-posedness

As the spaces in (4) are lacking a tensor product structure, the proof of well-posedness of (5) is more complicated than on a fixed domain. In the case of a fixed domain existence and uniqueness of solutions can be shown under weaker regularity assumptions on the data f and u^0 and the domain $\Omega(t)$ in the velocity space

$$\tilde{\mathcal{V}}_I := \{\boldsymbol{u} \in L^2(I, \mathcal{V})^d,\ \partial_t \boldsymbol{u} \in L^2(I, (\mathcal{V}^*)^d)\},$$

where $\mathcal{V}^*$ is the dual space to $\mathcal{V}$. Low regularity is, however, not of interest for the present paper, as we will require additional regularity of the solution in the error estimates. On the other hand, working with the space $\mathcal{V}_I$ under the additional regularity assumptions made above simplifies the proof of well-posedness of (5) significantly.

The well-posedness of the Navier–Stokes problem on time-dependent domains, including an additional nonlinear convective term has in fact been the subject of a number of papers in literature [6,55]. In order to deal with the additional nonlinearity, additional assumptions on the regularity of the domains are typically made. For completeness, we give a proof of the following Lemma (5) in the "Appendix" under the regularity assumptions made above.

Lemma 2.2 *Let $\Omega(0)$ be piecewise smooth and Lipschitz and $\boldsymbol{T}(t)$ a $W^{1,\infty}(\Omega(0))$ diffeomorphism with regularity $\boldsymbol{T} \in W^{1,\infty}(I, W^{2,\infty}(\Omega(0))$. For $\boldsymbol{f} \in L^\infty(I, \mathcal{L}(t)^d)$ and $\boldsymbol{u}^0 \in H^1(\Omega(0))^d$, Problem (5) has a unique solution $\boldsymbol{u} \in \mathcal{V}_I$, $p \in \mathcal{L}_{0,I}$.*

Proof A proof is given in the "Appendix". □

3 Discretisation

For discretisation in time, we split the time interval of interest $I = [0, t_{\text{fin}}]$ into time intervals $I_n = (t_{n-1}, t_n]$ of uniform step size $\Delta t = t_n - t_{n-1}$

$$0 = t_0 < t_1 < \cdots < t_N = t_{\text{fin}}.$$

We follow the work of Lehrenfeld and Olshanskii [42] for parabolic problems on moving domains and use BDF(s) discretisation for $s = 1, 2$, where $s = 1$ corresponds to a backward Euler time discretisation. Higher-order BDF formulae are not considered here, due to their lack of A-stability [34]. Following Lehrenfeld and Olshanskii [42] we extend the domain $\Omega^n := \Omega(t_n)$ in each time point t_n by a strip of size δ to a domain Ω_δ^n, which is chosen large enough such that

$$\bigcup_{i=0}^{s} \Omega^{n+i} \subset \Omega_\delta^n, \tag{8}$$

see also the left part of Fig. 1. In particular, we will allow

$$s w_{\max} \Delta t \ \leq \ \delta \leq c_\delta s w_{\max} \Delta t, \tag{9}$$

where

$$w_{\max} := \max_{t \in I, x \in \partial\Omega(0)} \|\partial_t T(x, t) \cdot \boldsymbol{n}\|$$

is the maximum velocity of the boundary movement in normal direction in the Euclidean norm $\|\cdot\|$ and $c_\delta > 1$ is a constant. If we assume that the domain map $\boldsymbol{T}$ lies in $W^{1,\infty}(I, L^\infty(\Omega(0)))$ (see Assumption 3.2 below), the lower bound on δ in (9) guarantees (8).

The space-time slabs defined by the time discretisation and the space-time domain are denoted by

$$Q^n := \bigcup_{t \in I_n} \{t\} \times \Omega(t), \qquad Q_\delta^n := \bigcup_{t \in I_n} \{t\} \times \Omega_\delta(t), \qquad Q := \bigcup_{t \in I} \{t\} \times \Omega(t).$$

In what follows we denote by c generic positive constants. These are in particular independent of space and time discretisation (Δt, N and h) and of domain velocity $w_{\max}$ and δ, unless such a dependence is explicitly mentioned.

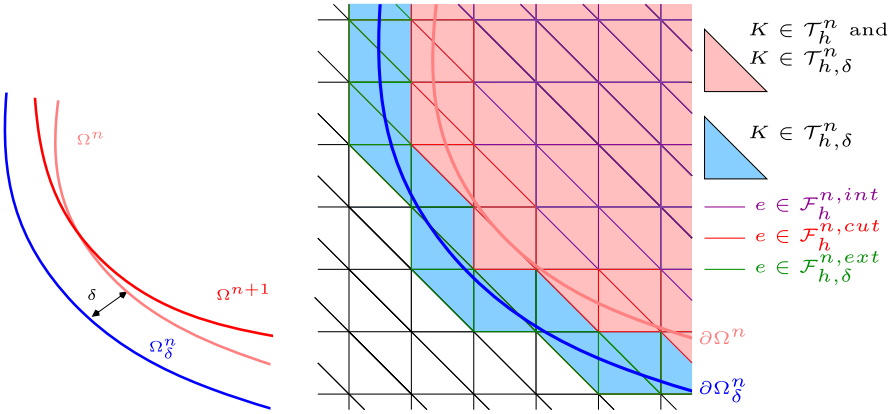


Fig. 1 Left: Illustration of Ω^n_δ for $s = 1$. Right: Illustration of the discretisation and faces

3.1 Space discretisation

Let $\mathcal{T}^n_{h,\delta}$ be a family of (possibly unfitted) quasi-uniform spatial discretisations of Ω^n_δ into simplices with maximum cell size h. We assume that $\mathcal{T}^n_{h,\delta}$ is based on a common background triangulation $\mathcal{T}_h$ for all n and may differ only in the elements outside $\Omega(t_n)$ that are not present in $\mathcal{T}^k_{h,\delta}$ for $k \neq n$. Further, we assume that $\mathcal{T}^n_{h,\delta}$ consists only of elements K with non-empty intersection with Ω^n_δ, i.e. $K \cap \Omega^n_\delta \neq \emptyset$. The subset of cells with non-empty intersection with Ω is denoted by $\mathcal{T}^n_h$. An illustration is given in Fig. 1. By $\Omega^n_{h,\delta}$ we denote the domain spanned by all cells $K \in \mathcal{T}^n_{h,\delta}$ and by Ω^n_h the domain spanned by all cells $K \in \mathcal{T}^n_h$.

Further, let $\mathcal{F}^n_{h,\delta}$ denote the set of interior faces e of $\mathcal{T}^n_{h,\delta}$. We split the faces into three parts: By $\mathcal{F}^{n,int}_h$, we denote the faces that belong exclusively to elements $K \in \mathcal{T}^n_{h,\delta}$ that lie in the interior of Ω^n. By $\mathcal{F}^{n,cut}_h$ we denote the set of faces that belong to some element $K \in \mathcal{T}^n_{h,\delta}$ with $K \cap \partial\Omega^n \neq \emptyset$ and by $\mathcal{F}^{n,ext}_{h,\delta}$ the set of the remaining faces in $\mathcal{F}^n_{h,\delta}$, see Fig. 1. Finally, we write $\mathcal{F}^{n,g}_{h,\delta}$ for the union of $\mathcal{F}^{n,cut}_h$ and $\mathcal{F}^{n,ext}_{h,\delta}$, which will be used to define the ghost penalty extensions.

For spatial discretisation, we use continuous equal-order finite elements of degree $m \geq 1$ for all variables

$$\mathcal{V}^n_h := \{\boldsymbol{v} \in C(\Omega^n_{h,\delta}),\ \boldsymbol{v}|_K \in P_m(K)\,\forall K \in \mathcal{T}^n_{h,\delta}\}$$
$$\mathcal{L}^n_h := \{q \in C(\Omega^n_h),\ q|_K \in P_m(K)\,\forall K \in \mathcal{T}^n_h\}, \quad \mathcal{L}^n_{h,0} := \mathcal{L}^n_h \cap \mathcal{L}_0(t_n).$$

Note that for the pressure space $\mathcal{L}^n_h$ an extension beyond Ω^n_h is not required.

To deal with the *inf-sup* stability, we will add a pressure stabilisation term s^n_h to the variational formulation. In order to simplify the presentation, we will restrict ourselves to the Continuous Interior Penalty method (CIP [14]) in this work, although different pressure stabilisations are possible. We define the CIP pressure stabilisation as

$$s_h^n(p;q) := \sum_{e\in\mathcal{F}_h^{n,int}} h^3([\![\partial_n p]\!],[\![\partial_n q]\!])_e + \sum_{e\in\mathcal{F}_h^{n,cut}}\sum_{k=1}^{m} h^{2k+1}([\![\partial_n^k p]\!],[\![\partial_n^k q]\!])_e.$$

The higher derivatives in the boundary elements are necessary to control the derivatives ∇p_h on the extended computational domain $\Omega_h^n\backslash\Omega^n$ in the spirit of the ghost penalty stabilisation [10].

We summarise the properties of the pressure stabilisation that will be needed in the following: There exists an operator $C_h^n : \mathcal{V}_h^n \cup L^2(\Omega^n) \to \mathcal{V}_h^n$, such that the following properties are fulfilled for $n = 1,\dots,N$

$$s_h^n(q,r) \le s_h^n(q,q)^{1/2} s_h^n(r,r)^{1/2} \qquad \forall q,r \in \mathcal{V}_h^n \cup H^2(\Omega^n), \tag{10}$$

$$s_h^n(q,q) \le ch^2 \|q\|^2_{H^1(\Omega^n)} \qquad \forall q \in \mathcal{V}_h^n \cup H^2(\Omega^n), \tag{11}$$

$$h^2\|\nabla p_h - C_h^n \nabla p_h\|^2_{\Omega_h^n} \le c s_h^n(p_h^n, p_h^n) \qquad \forall p_h \in \mathcal{V}_h^n, \tag{12}$$

$$\|C_h^n \nabla p_h\|^2_{\Omega_h^n} \le c\left(\|\nabla p_h\|^2_{\Omega^n} + h^{-2} s_h^n(p_h^n, p_h^n)\right) \qquad \forall p_h \in \mathcal{V}_h^n. \tag{13}$$

A suitable projector C_h^n for the CIP stabilisation is given by the the Oswald or Clément interpolation [14]. For $m \ge 2$, we have additionally the consistency property

$$s_h^n(p,p) = 0 \qquad \forall p \in H^m(\Omega^n). \tag{14}$$

Remark 3.1 (Pressure stabilisation) In general any pressure stabilisation operator that leads to a well-posed discrete problem and that fulfils the assumptions (10)–(14) can be used. The consistency condition $s_h^n(p,p) = 0$ can be relaxed to a weak consistency of order $m_s > 0$

$$s_h^n(p,p) \le ch^{2m_s} \|p\|^2_{H^{m_s}(\Omega^n)}.$$

which will limit the spatial convergence order in the error estimates. One possibility is the Brezzi–Pitkäranta stabilisation [9] with order $m_s = 1$. We refer to Burman and Fernández for a review of further possibilities for pressure stabilisation [12].

3.2 Variational formulation

To cope with the evolving geometry from one time-step to another, we extend the velocity variable $\boldsymbol{u}_h^n$ to Ω_δ^n, which will be needed in the following time-step, by using so-called *ghost penalty terms* g_h^n. We will describe different possibilities to define g_h^n in the next subsection. For $k = s,\dots,n$ we define the following time-stepping scheme: *Find* $\boldsymbol{u}_h^n \in \mathcal{V}_h^n$, $p_h^n \in \mathcal{L}_{h,0}^n$ such that

$$(D_t^{(s)}\boldsymbol{u}_h^n, \boldsymbol{v}_h)_{\Omega^n} + \mathcal{A}_h^n(\boldsymbol{u}_h^n, p_h^n; \boldsymbol{v}_h, q_h) = (\boldsymbol{f}, \boldsymbol{v}_h)_{\Omega^n} \quad \forall \boldsymbol{v}_h \in \mathcal{V}_h^n, q_h \in \mathcal{L}_h^n, \tag{15}$$

where $D_t^{(s)}$ is an approximation of the time derivative ∂_t by the BDF(s) backward difference formula, i.e.

$$D_t^{(1)}\boldsymbol{u}_h^n := \frac{1}{\Delta t}(\boldsymbol{u}_h^n - \boldsymbol{u}_h^{n-1}),$$
$$D_t^{(2)}\boldsymbol{u}_h^n := D_t^{(1)}\boldsymbol{u}_h^n + \frac{1}{2}(D_t^{(1)}\boldsymbol{u}_h^n - D_t^{(1)}\boldsymbol{u}_h^{n-1}) = \frac{1}{2\Delta t}(3\boldsymbol{u}_h^n - 4\boldsymbol{u}_h^{n-1} + \boldsymbol{u}_h^{n-2}).$$

The bilinear form $\mathcal{A}_h$ is defined by

$$\begin{aligned}\mathcal{A}_h^n(\boldsymbol{u}_h^n, p_h^n; \boldsymbol{v}_h, q_h) := \mathcal{A}_S^n(\boldsymbol{u}_h^n, p_h^n; \boldsymbol{v}_h, q_h) + a_D^n(\boldsymbol{u}_h^n, p_h^n; \boldsymbol{v}_h, q_h)\\ + \gamma_g g_h^n(\boldsymbol{u}_h^n, \boldsymbol{v}_h) + \gamma_p s_h^n(p_h^n, q_h).\end{aligned} \tag{16}$$

It includes the Stokes part

$$\mathcal{A}_S^n(\boldsymbol{u}, p; \boldsymbol{v}, q) := (\nabla \boldsymbol{u}, \nabla \boldsymbol{v})_{\Omega^n} - (p, \operatorname{div} \boldsymbol{v})_{\Omega^n} + (\operatorname{div} \boldsymbol{u}, q)_{\Omega^n} \tag{17}$$

and Nitsche terms to weakly impose the Dirichlet boundary conditions

$$\begin{aligned}a_D^n(\boldsymbol{u}_h^n, p_h^n; \boldsymbol{v}_h, q_h) := -(\partial_n \boldsymbol{u}_h^n - p_h^n \boldsymbol{n}, \boldsymbol{v}_h)_{\partial\Omega^n} - (\boldsymbol{u}_h^n, \partial_n \boldsymbol{v}_h + q_h \boldsymbol{n})_{\partial\Omega^n}\\ + \frac{\gamma_D}{h}\left(\boldsymbol{u}_h^n, \boldsymbol{v}_h\right)_{\partial\Omega^n}.\end{aligned} \tag{18}$$

In (18) the last term can be seen as a penalty term to weakly impose the homogeneous Dirichlet condition for the velocities. The first term on the right-hand side makes the variational formulation consistent (in space). Finally, the second term, which vanishes for $u_h^n = 0$, yields a formulation, which is symmetric for the velocities, but skew-symmetric for the pressure. The skew-symmetry in the pressure variable leads to a stable variational formulation, as the pressure terms cancel out by diagonal testing ($\boldsymbol{v}_h^n = \boldsymbol{u}_h^n, q_h^n = p_h^n$), see for example [13]. The parameters γ_D, γ_p and γ_g are positive constants.

To include the initial condition, we set $\boldsymbol{u}_h^0 := \pi_h^1 E\boldsymbol{u}^0$, where π_h^n denotes the $L^2(\Omega^n)$-projection onto $\mathcal{T}_h^n$ and E denotes an L^2-stable extension operator, which is introduced in the next section. Summing over $k = 1, \ldots, n$ in time, the complete system reads for $s = 1$

$$\begin{aligned}\sum_{k=1}^{n}\left\{\frac{1}{\Delta t}\left(\boldsymbol{u}_h^k - \boldsymbol{u}_h^{k-1}, \boldsymbol{v}_h^k\right)_{\Omega^k} + \mathcal{A}_h^k(\boldsymbol{u}_h^k, p_h^k; \boldsymbol{v}_h^k, q_h^k)\right\} + \frac{1}{\Delta t}\left(\boldsymbol{u}_h^0, \boldsymbol{v}_h^1\right)_{\Omega^1}\\ = \frac{1}{\Delta t}\left(E\boldsymbol{u}^0, \boldsymbol{v}_h^1\right)_{\Omega^1} + \sum_{k=1}^{n}(\boldsymbol{f}, \boldsymbol{v}_h^k)_{\Omega^k} \quad \forall \boldsymbol{v}_h^k \in \mathcal{V}_h^k, q_h^k \in \mathcal{L}_h^k, \quad k = 1, \ldots n.\end{aligned} \tag{19}$$

In order to simplify the presentation, we will assume that the integrals in (19) are evaluated exactly. If the integrals are only roughly approximated, for example due to a discrete level-set function $\phi_h^n \in \mathcal{V}_h^n$ which is only an approximation of a continuous function ϕ^n, an additional geometry approximation error needs to be considered. We

refer to the work of Lehrenfeld and Olshanskii [42], where these additional error contributions have been analysed in detail for parabolic problems on moving domains. An advantage of the CutFEM methodology compared to standard finite elements is that besides the geometry approximation no additional discretisation errors related to the approximation of curved boundaries within the finite element spaces need to be considered.

To initialise the BDF(2) scheme the value $\boldsymbol{u}_h^1$ needs to be computed with sufficient accuracy before the first full BDF(2) step can be made. We will comment on the specific requirements and on different possibilities below in Remark 5.6.

3.3 Extension operators

Due to the evolution of the domain, we will frequently need to extend variables defined on smaller domains to larger ones. Therefore, we will use $W^{k,p}$-stable extension operators $E^n : \Omega^n \to \Omega_\delta^n$ to extend functions $\boldsymbol{u}(t_n) \in W^{k,p}(\Omega^n)$. We make the following assumption for the regularity of the domains $\Omega(t)$ and the domain movement, depending on the polynomial degree m of the finite element spaces.

Assumption 3.2 We assume that the boundary of the initial domain $\Omega(0)$ is piecewise smooth and Lipschitz, and that the domain motion $\boldsymbol{T}(t)$ is a $W^{1,\infty}$-diffeomorphism for each t and smooth in the sense that $\boldsymbol{T} \in L^\infty(I, W^{m+1,\infty}(\Omega(0))) \cap W^{1,\infty}(I, W^{m,\infty}(\Omega(0)))$.

If Assumption 3.2 is fulfilled for $m \in \mathbb{N}$, suitable extension operators $E^n : \Omega^n \to \Omega_\delta^n$ exist with the properties

$$\|E^n\boldsymbol{u} - \boldsymbol{u}\|_{W^{m+1,p}(\Omega)} = 0, \qquad \|E^n\boldsymbol{u}\|_{W^{m+1,p}(\Omega_\delta^n)} \leq c\|\boldsymbol{u}\|_{W^{m+1,p}(\Omega^n)}, \tag{20}$$

$$\|\partial_t(E^n\boldsymbol{u})\|_{H^m(\Omega_\delta^n)} \leq c\left(\|\boldsymbol{u}\|_{H^{m+1}(\Omega^n)} + \|\partial_t\boldsymbol{u}\|_{H^m(\Omega^n)}\right), \tag{21}$$

$$\|\partial_t^2(E^n\boldsymbol{u})\|_{Q_\delta^n} \leq c\|\boldsymbol{u}\|_{H^2(Q^n)}. \tag{22}$$

For a proof of (20) we refer to Stein [57], Theorem 6 in Chapter VI. The estimate (21) has been shown in [42], Lemma 3.3. The estimate (22) follows by the same argumentation. In order to alleviate the notation we will in the following skip the operator E^n frequently and denote the extension also by $\boldsymbol{u}(t_n)$.

3.3.1 Ghost penalty extension

The discrete quantities are extended implicitly by adding so-called ghost penalty terms to the variational formulation. We will consider three variants for the ghost penalty stabilisation, and refer to [15,32] for a more abstract approach on how to design suitable ghost penalties for a PDE problem at hand. The first "classical" variant [10,15,45] is to penalise jumps of derivatives over element edges

$$g_h^{n,\text{jump}}(\boldsymbol{u},\boldsymbol{v}) := \sum_{e\in\mathcal{F}_{h,\delta}^{n,g}} \sum_{k=1}^{m} h^{2k-1}(\llbracket\partial_n^k\boldsymbol{u}\rrbracket, \llbracket\partial_n^k\boldsymbol{v}\rrbracket)_e.$$

This variant has the advantage that it is fully consistent, i.e. it vanishes for $\boldsymbol{u} \in H^{m+1}(\Omega_\delta^n)^d$, which implies $[\![\partial_n^k \boldsymbol{u}]\!]|_e = 0$ for $k \leq m$. A disadvantage is that higher derivatives need to be computed for polynomial degrees $m > 1$.

To define two further variants, let us introduce the notation $K_{e,1}$ and $K_{e,2}$ for the two cells surrounding a face $e \in \mathcal{F}_{h,\delta}^{n,g}$, such that

$$e = \overline{K}_{e,1} \cap \overline{K}_{e,2}.$$

We denote the union of both cells by $w_e := K_{e,1} \cup K_{e,2}$ and use the L^2-projection $\pi_{w_e} : L^2(\Omega_\delta^n) \to P_m(w_e)$, which is defined by

$$(\boldsymbol{u} - \pi_{w,e}\boldsymbol{u}, \boldsymbol{v})_{w_e} = 0 \quad \forall \boldsymbol{v} \in P_m(w_e).$$

We define the "projection variant" of the ghost penalty stabilisation [10]

$$\begin{aligned} g_h^{n,\mathrm{proj}}(\boldsymbol{u}_h, \boldsymbol{v}_h) &:= \frac{1}{h^2} \sum_{e \in \mathcal{F}_{h,\delta}^{n,g}} \left(\boldsymbol{u}_h - \pi_{w_e}\boldsymbol{u}_h, \boldsymbol{v}_h - \pi_{w_e}\boldsymbol{v}_h\right)_{w_e} \\ &= \frac{1}{h^2} \sum_{e \in \mathcal{F}_{h,\delta}^{n,g}} \left(\boldsymbol{u}_h - \pi_{w_e}\boldsymbol{u}_h, \boldsymbol{v}_h\right)_{w_e}. \end{aligned}$$

The last equality is a direct consequence of the definition of the L^2-projection.

The third variant, which has first been used in [52], uses canonical extensions of polynomials to the neighbouring cell instead of the projection $\pi_{w_e}\boldsymbol{u}$. Let us therefore denote the polynomials that define a function $\boldsymbol{u} \in \mathcal{V}_h^n$ in a cell $K_{e,i}$ by $\boldsymbol{u}_{e,i} = \boldsymbol{u}|_{K_{e,i}}$. We use the same notation for the canonical extension to the neighbouring cell, such that $\boldsymbol{u}_{e,i} \in P_m(w_e)$. Using this notation, we define the so-called "direct method" of the ghost penalty stabilisation

$$g_h^{n,\mathrm{dir}}(\boldsymbol{u}, \boldsymbol{v}) := \frac{1}{h^2} \sum_{e \in \mathcal{F}_{h,\delta}^{n,g}} \left(\boldsymbol{u}_{e,1} - \boldsymbol{u}_{e,2}, \boldsymbol{v}_{e,1} - \boldsymbol{v}_{e,2}\right)_{w_e}.$$

For the analysis, we extend the definition of the stabilisation to functions $\boldsymbol{u}, \boldsymbol{v} \in L^2(\Omega_\delta^n)$. Here, we set $\boldsymbol{u}_{e,i} := \pi_{K_{e,i}}\boldsymbol{u}|_{K_{e,i}}$ for $i = 1, 2$, where $\pi_{K_{e,i}}$ denotes the L^2-projection to $P_m(K_{e,i})$ and extend this polynomial canonically to the neighbouring cell. In contrast to the classical variant, $g_h^{n,\mathrm{proj}}$ and $g_h^{n,\mathrm{dir}}$ are only weakly consistent, i.e. they fulfil the estimate

$$g_h^n(\boldsymbol{u}, \boldsymbol{u}) \leq c h^{2m} \|\boldsymbol{u}\|_{H^{m+1}(\Omega_\delta^n)}, \qquad \text{for } \boldsymbol{u} \in H^{m+1}(\Omega_\delta^n)^d.$$

We will summarise the properties of these stabilisation terms, that we will need below, in the following lemma. Therefore, we assume that from each cell $K \in \mathcal{T}_{h,\delta}^n$ with $K \cap \Omega^n = \emptyset$, there exists a path of cells K_i, $i = 1, \ldots, m$, such that two subsequent cells share one common face $e = \overline{K}_i \cap \overline{K}_{i+1}$, and the final element lies in the interior

of Ω^n, i.e. $K_m \subset \Omega^n$. In addition the path shall fulfil the following properties. Let $\mathcal{K}$ be the maximum number of cells needed in the path among all cells $K \in \mathcal{T}_{h,\delta}^n$. We assume that

$$\mathcal{K} \leq (1 + \delta/h) \leq 1 + \frac{c_\delta s w_{\max} \Delta t}{h}, \tag{23}$$

where the second inequality follows from (9). Moreover, we assume that the number of cases in which a specific interior element $K_m \subset \Omega^n$ is used as a final element among all the paths is bounded independently of Δt and h. These assumptions are reasonable, as one can choose for example the final elements by a projection of distance δ towards the interior. For a detailed justification, we refer to Lehrenfeld and Olshanskii [42], Remark 5.2.

Lemma 3.3 *For $\boldsymbol{v}_h \in \mathcal{V}_h^n$ and the three variants $g_h^n \in \{g_h^{n,\text{jump}}, g_h^{n,\text{proj}}, g_h^{n,\text{dir}}\}$ it holds that*

$$\|\boldsymbol{v}_h\|_{\Omega_\delta^n}^2 \leq c\|\boldsymbol{v}_h\|_{\Omega^n}^2 + \mathcal{K}h^2 g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h), \qquad \|\nabla \boldsymbol{v}_h\|_{\Omega_\delta^n}^2 \leq c\|\nabla \boldsymbol{v}_h\|_{\Omega^n}^2 + \mathcal{K} g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h)$$

Further, it holds for $\boldsymbol{u}, \boldsymbol{v} \in H^{m+1}(\Omega_\delta^n)$ for $m \geq 1$ and $\boldsymbol{v}_h \in \mathcal{V}_h^n$ that

$$\begin{aligned} g_h^n(\boldsymbol{u}, \boldsymbol{v}) \leq g_h^n(\boldsymbol{u}, \boldsymbol{u})^{1/2} g_h^n(\boldsymbol{v}, \boldsymbol{v})^{1/2}, \;\; g_h^n(\boldsymbol{u}, \boldsymbol{u}) \leq ch^{2m} \|\boldsymbol{u}\|_{H^{m+1}(\Omega_\delta^n)}^2, \;\; g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h) \\ \leq c\|\nabla \boldsymbol{v}_h\|_{\Omega_\delta^n}^2. \end{aligned} \tag{24}$$

Proof The first four properties have been proven for the three possibilities introduced above by Lehrenfeld and Olshanskii [42]. The last inequality in (24) follows similarly. □

3.4 Properties of the bilinear form

We start with a continuity result for the combined bilinear form including the Nitsche terms in the functional spaces.

Lemma 3.4 (Continuity in the functional spaces) *For functions $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{V}(t_n) \cap H^2(\Omega^n)^d$ and $p, q \in \mathcal{L}(t_n) \cap H^1(\Omega^n)$, we have*

$$\begin{aligned} &\left|(\mathcal{A}_S^n + a_D^n)(\boldsymbol{u}, p; \boldsymbol{v}, q)\right| + \left|(\mathcal{A}_S^n + a_D^n)(\boldsymbol{v}, q; \boldsymbol{u}, p)\right| \\ &\quad \leq c \left(\|\nabla \boldsymbol{u}\|_{\Omega^n} + h^{-1/2}\|\boldsymbol{u}\|_{\partial\Omega^n} + h\|\nabla p\|_{\Omega^n} + h^{1/2}\|\partial_n \boldsymbol{u}\|_{\partial\Omega^n} \right) \\ &\qquad \cdot \left(\|\nabla \boldsymbol{v}\|_{\Omega^n} + h^{-1}\|\boldsymbol{v}\|_{\Omega^n} + \|q\|_{\Omega^n} + h^{-1/2}\|\boldsymbol{v}\|_{\partial\Omega^n} + h^{1/2}\left(\|\partial_n \boldsymbol{v}\|_{\partial\Omega^n} + \|q\|_{\partial\Omega^n}\right) \right). \end{aligned}$$

Proof We apply integration by parts in (17)

$$\begin{aligned}\mathcal{A}_S^n(\boldsymbol{u}, p; \boldsymbol{v}, q) &= (\nabla \boldsymbol{u}, \nabla \boldsymbol{v})_{\Omega^n} + (\nabla p, \boldsymbol{v})_{\Omega^n} - (p, \boldsymbol{v}\cdot\boldsymbol{n})_{\partial\Omega^n} + (\operatorname{div}\boldsymbol{u}, q)_{\Omega^n} \\ &\leq c\left(\|\nabla \boldsymbol{u}\|_{\Omega^n} + h\|\nabla p\|_{\Omega^n}\right)\Big(\|\nabla \boldsymbol{v}\|_{\Omega^n} + h^{-1}\|\boldsymbol{v}\|_{\Omega^n} \\ &\quad + \|q\|_{\Omega^n}\Big) - (p, \boldsymbol{v}\cdot\boldsymbol{n})_{\partial\Omega^n}.\end{aligned}$$

For the Nitsche terms standard estimates result in

$$\begin{aligned}a_D^n(\boldsymbol{u}, p; \boldsymbol{v}, q) \leq\ & c\left(h^{-1/2}\|\boldsymbol{u}\|_{\partial\Omega^n} + h^{1/2}\|\partial_n \boldsymbol{u}\|_{\partial\Omega^n}\right)\left(h^{-1/2}\|\boldsymbol{v}\|_{\partial\Omega^n} + h^{1/2}\left(\|\partial_n \boldsymbol{v}\|_{\partial\Omega^n} + \|q\|_{\partial\Omega^n}\right)\right) \\ & + (p, \boldsymbol{v}\cdot\boldsymbol{n})_{\partial\Omega^n}. \end{aligned} \tag{25}$$

The estimate for $(\mathcal{A}_S^n + a_D^n)(\boldsymbol{v}, q; \boldsymbol{u}, p)$ can be shown in exactly the same way by inverting the role of test and trial functions. □

Next, we show continuity and coercivity of the discrete bilinear form. To this end, we introduce the triple norm

$$|||\boldsymbol{u}_h|||_{h,n} := \left(\|\nabla \boldsymbol{u}_h^n\|_{\Omega^n}^2 + \gamma_g g_h^n(\boldsymbol{u}_h^n, \boldsymbol{u}_h^n) + \frac{\gamma_D}{h}\|\boldsymbol{u}_h^n\|_{\partial\Omega^n}^2\right)^{1/2}$$

Lemma 3.5 (Coercivity and Continuity in the discrete setting) *For the bilinear form $\mathcal{A}_h$ defined in (16) and $\boldsymbol{u}_h \in \mathcal{V}_h^n$ and $p_h \in \mathcal{L}_h^n$, it holds for γ_D sufficiently large*

$$\mathcal{A}_h^n(\boldsymbol{u}_h, p_h; \boldsymbol{u}_h, p_h) \geq \frac{1}{2}\left(|||\boldsymbol{u}_h|||_{h,n}^2 + \gamma_p s_h^n(p_h, p_h)\right). \tag{26}$$

Moreover, we have for $\boldsymbol{u}_h, \boldsymbol{v}_h \in \mathcal{V}_h^n$ and $p_h, q_h \in \mathcal{L}_h^n$

$$\begin{aligned}\mathcal{A}_h^n(\boldsymbol{u}_h, p_h; \boldsymbol{v}_h, q_h) \leq\ & \left(|||\boldsymbol{u}_h|||_{h,n} + \|p_h\|_{\Omega^n} + s_h^n(p_h, p_h)^{1/2}\right) \\ & \left(|||\boldsymbol{v}_h|||_{h,n} + \|q_h\|_{\Omega^n} + s_h^n(q_h, q_h)^{1/2}\right). \end{aligned} \tag{27}$$

Proof To show coercivity (26), we note that

$$\mathcal{A}_h^n(\boldsymbol{u}_h, p_h; \boldsymbol{u}_h, p_h) = |||u_h|||_{h,n}^2 + \gamma_p s_h^n(p_h, p_h) - 2(\boldsymbol{u}_h^n, \partial_n \boldsymbol{u}_h^n)_{\partial\Omega^n}.$$

To estimate the term $-2(\boldsymbol{u}_h^n, \partial_n \boldsymbol{u}_h^n)_{\partial\Omega^n}$, we apply a Cauchy-Schwarz and Young's inequality for $\epsilon > 0$, followed by an inverse inequality on Ω_h^n

$$-2(\boldsymbol{u}_h, \partial_n \boldsymbol{u}_h)_{\partial\Omega^n} \geq -\frac{1}{\epsilon h}\|\boldsymbol{u}_h\|_{\partial\Omega^n}^2 - \epsilon h\|\nabla \boldsymbol{u}_h\|_{\partial\Omega^n}^2 \geq -\frac{1}{\epsilon h}\|\boldsymbol{u}_h\|_{\partial\Omega^n}^2 - c\epsilon\|\nabla \boldsymbol{u}_h\|_{\Omega_h^n}^2.$$

Using Lemma 3.3, we obtain

$$\begin{aligned} 2(\boldsymbol{u}_h, \partial_n \boldsymbol{u}_h)_{\partial\Omega^n} &\geq -\frac{1}{\epsilon h}\|\boldsymbol{u}_h\|^2_{\partial\Omega^n} - c\epsilon\left(\|\nabla \boldsymbol{u}_h\|^2_{\Omega^n} + \mathcal{K} g_h^n(\boldsymbol{u}_h, \boldsymbol{u}_h)\right) \\ &\geq -\frac{\gamma_D}{2h}\|\boldsymbol{u}_h\|^2_{\partial\Omega^n} - \frac{1}{2}\left(\|\nabla \boldsymbol{u}_h\|^2_{\Omega^n} + \gamma_g g_h^n(\boldsymbol{u}_h, \boldsymbol{u}_h)\right) \end{aligned}$$

for γ_D sufficiently large. Concerning continuity, we estimate

$$\mathcal{A}_S^n(\boldsymbol{u}_h, p_h; \boldsymbol{v}_h, q_h) \leq c\left(\|\nabla \boldsymbol{u}_h\|_{\Omega^n} + \|p_h\|_{\Omega^n}\right)\left(\|\nabla \boldsymbol{v}_h\|_{\Omega^n} + \|q_h\|_{\Omega^n}\right) \tag{28}$$

For the Nitsche terms, we have using inverse inequalities and Lemma 3.3

$$\begin{aligned} a_D^n(\boldsymbol{u}_h, p_h; \boldsymbol{v}_h, q_h) &= \frac{\gamma_D}{h}(\boldsymbol{u}_h, \boldsymbol{v}_h)_{\partial\Omega^n} - (\partial_n \boldsymbol{u}_h - p_h n, \boldsymbol{v}_h)_{\partial\Omega^n} - (\boldsymbol{u}_h, \partial_n \boldsymbol{v}_h + q_h \boldsymbol{n})_{\partial\Omega^n} \\ &\leq c\left(\frac{\gamma_D^{1/2}}{h^{1/2}}\|\boldsymbol{u}_h\|_{\partial\Omega^n} + \|\nabla \boldsymbol{u}_h\|^2_{\Omega_h^n} + \|p_h\|_{\Omega^n}\right)\left(\frac{\gamma_D^{1/2}}{h^{1/2}}\|\boldsymbol{v}_h\|_{\partial\Omega^n} + \|\nabla \boldsymbol{v}_h\|^2_{\Omega_h^n} + \|q_h\|_{\Omega^n}\right) \\ &\leq c\left(|||\boldsymbol{u}_h|||_{h,n} + \|p_h\|_{\Omega^n}\right)\left(|||\boldsymbol{v}_h|||_{h,n} + \|q_h\|_{\Omega^n}\right). \end{aligned} \tag{29}$$

Finally, Lemma 3.3 and the assumption (10) for the pressure stabilisation yield

$$\begin{aligned} g_h^n(\boldsymbol{u}_h, \boldsymbol{v}_h) &\leq g_h^n(\boldsymbol{u}_h, \boldsymbol{u}_h)^{1/2} g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h)^{1/2}, \\ s_h^n(p_h, q_h) &\leq s_h^n(p_h, p_h)^{1/2} s_h^n(q_h, q_h)^{1/2}. \end{aligned}$$

□

Moreover, we have the following modified *inf-sup* condition for the discrete spaces.

Lemma 3.6 *Let $p_h^n \in \mathcal{L}_h^n$. There exists a constant $\beta > 0$, such that*

$$\beta\|p_h^n\|_{\Omega^n} \leq \sup_{\boldsymbol{v}_h^n \in \mathcal{V}_h^n} \frac{(\operatorname{div} \boldsymbol{v}_h^n, p_h^n)_{\Omega^n} - (\boldsymbol{v}_h^n \cdot n, p_h^n)_{\partial\Omega^n}}{|||\boldsymbol{v}_h^n|||_{h,n}} + h\|\nabla p_h^n\|_{\Omega^n}. \tag{30}$$

Proof We follow Burman and Hansbo [14] and define $\boldsymbol{v}_p^n \in H_0^1(\Omega^n)^d$ as solution to

$$\operatorname{div} \boldsymbol{v}_p^n = -\frac{p_h^n}{\|p_h^n\|_{\Omega^n}} \quad \text{on } \Omega^n. \tag{31}$$

Such a solution exists, see Temam [58], and fulfils $\|\boldsymbol{v}_p^n\|_{H^1(\Omega^n)} \leq c$. We introduce an L^2-stable interpolation $i_h^n \boldsymbol{v}_p^n$ (for example the Clément interpolation) to get

$$\|p_h^n\|_{\Omega^n} = -(p_h^n, \operatorname{div} \boldsymbol{v}_p^n)_{\Omega^n} = -(p_h^n, \operatorname{div}(\boldsymbol{v}_p^n - i_h^n \boldsymbol{v}_p^n))_{\Omega^n} - (p_h^n, \operatorname{div}(i_h^n \boldsymbol{v}_p^n))_{\Omega^n}. \tag{32}$$

We apply integration by parts in the first term and use that $\boldsymbol{v}_p^n$ vanishes in $\partial\Omega^n$

$$\begin{aligned} -(p_h^n, \operatorname{div}(\boldsymbol{v}_p^n - i_h^n \boldsymbol{v}_p^n))_{\Omega^n} &= (\nabla p_h^n, \boldsymbol{v}_p^n - i_h^n \boldsymbol{v}_p^n)_{\Omega^n} - (p_h^n \boldsymbol{n}, \boldsymbol{v}_p^n - i_h^n \boldsymbol{v}_p^n)_{\partial\Omega^n} \\ &\leq ch \|\nabla p_h^n\|_{\Omega^n} + (p_h^n \boldsymbol{n}, i_h^n \boldsymbol{v}_p^n)_{\partial\Omega^n} \end{aligned} \tag{33}$$

The statement follows by noting that

$$\begin{aligned} |||i_h^n \boldsymbol{v}_p^n|||_{h,n}^2 &= \|i_h^n \nabla \boldsymbol{v}_p^n\|_{\Omega^n}^2 + \frac{\gamma_D}{h} \|i_h^n \boldsymbol{v}_p^n\|_{\partial\Omega^n}^2 + g_h^n(i_h^n \boldsymbol{v}_p^n, i_h^n \boldsymbol{v}_p^n) \\ &\leq c \left(\|\nabla \boldsymbol{v}_p^n\|_{\Omega^n}^2 + \frac{\gamma_D}{h} \|i_h^n \boldsymbol{v}_p^n - \boldsymbol{v}_p^n\|_{\partial\Omega^n}^2 + \|\nabla i_h^n \boldsymbol{v}_p^n\|_{\Omega_{h,\delta}^n}^2 \right) \leq c \|\nabla \boldsymbol{v}_p^n\|_{\Omega^n}^2 \leq c. \end{aligned}$$

□

The well-posedness of the discrete system (15) for sufficiently large γ_p, γ_g, γ_D and given $\boldsymbol{u}_h^{n-1}$ (and $\boldsymbol{u}_h^{n-2}$ for BDF(2)) follows by standard arguments, see for example [14].

4 Stability analysis

In order to simplify the analysis, we restrict ourselves in this and the next section to the case $s = 1$ first, i.e., the backward Euler variant of the time discretisation and comment on the case $s = 2$ in remarks. In order to abbreviate the notation, we write for the space-time Bochner norms

$$\|\boldsymbol{u}\|_{\infty,m,I_k} := \|\boldsymbol{u}\|_{L^\infty(I_k, H^m(\Omega(t)))}, \qquad \|\boldsymbol{u}\|_{\infty,m} := \|\boldsymbol{u}\|_{\infty,m,I},$$

where $m \in \mathbb{Z}$ and $H^0(\Omega(t)) := L^2(\Omega(t))$.

We start with a preliminary result concerning the extension of discrete functions to Ω_δ^n.

Lemma 4.1 *Let $\boldsymbol{v} \in \mathcal{V}$, $\delta \leq c_\delta s w_{\max} \Delta t$ and $S_\delta^n := \Omega_\delta^n \backslash \Omega^n$. It holds for arbitrary $\epsilon > 0$*

$$\|\boldsymbol{v}\|_{S_\delta^n}^2 \leq c\delta \left((\epsilon + \epsilon^{-1}) \|\boldsymbol{v}\|_{\Omega_\delta^n}^2 + \epsilon \|\nabla \boldsymbol{v}\|_{\Omega_\delta^n}^2 \right). \tag{34}$$

For $\boldsymbol{v}_h \in \mathcal{V}_h^n$, we have further for h sufficiently small

$$\|\boldsymbol{v}_h\|_{\Omega_\delta^n}^2 \leq (1 + c_1(w_{\max})\Delta t) \|\boldsymbol{v}_h\|_{\Omega^n}^2 + \frac{\Delta t}{2} \|\nabla \boldsymbol{v}_h\|_{\Omega^n}^2 + c_2(w_{\max}) \Delta t \mathcal{K} g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h) \tag{35}$$

with constants $c_1(w_{\max}) := 1/2 + cs^2 w_{\max}^2$, $c_2(w_{\max}) := c w_{\max}^2 h^2 + 1$ and $c > 0$.

Proof These results follow similarly to Lemmas 3.4 and 5.3 in [42]. Nevertheless, we give here a sketch of the proof due to the importance of the Lemma in the following estimates. We define

$$\Omega_r^n := \Omega^n \cup \{x \in S_\delta^n, \operatorname{dist}(x, \partial\Omega^n) < r\}, \quad \Gamma_r^n := \{x \in S_\delta^n, \operatorname{dist}(x, \partial\Omega^n) = r\} = \partial\Omega_r^n.$$

We apply a multiplicative trace inequality and Young's inequality for arbitrary $\epsilon > 0$

$$\begin{aligned}\|v\|_{\Gamma_r^n}^2 &\leq c\|\boldsymbol{v}\|_{\Omega_r^n}\|\boldsymbol{v}\|_{H^1(\Omega_r^n)} \leq c_0\left(\epsilon^{-1}\|\boldsymbol{v}\|_{\Omega_r^n}^2 + \epsilon\|\boldsymbol{v}\|_{H^1(\Omega_r^n)}^2\right)\\ &= c_0\left((\epsilon+\epsilon^{-1})\|\boldsymbol{v}\|_{\Omega_r^n}^2 + \epsilon\|\nabla\boldsymbol{v}\|_{\Omega_r^n}^2\right)\end{aligned} \tag{36}$$

with a constant c_0 depending on the curvature of $\partial\Omega^n$. Integration over $r \in (0, \delta)$ yields (34). For a discrete function $\boldsymbol{v}_h \in \mathcal{V}_h^n$ we use Lemma 3.3 to obtain

$$\begin{aligned}\|\boldsymbol{v}_h\|_{S_\delta^n}^2 &\leq c_0\delta(\epsilon+\epsilon^{-1})\|\boldsymbol{v}_h\|_{\Omega_\delta^n}^2 + c_0\delta\epsilon\|\nabla\boldsymbol{v}_h\|_{\Omega_\delta^n}^2\\ &\leq c_0\delta(\epsilon+\epsilon^{-1})\|\boldsymbol{v}_h\|_{\Omega^n}^2 + c_0\delta\epsilon\|\nabla\boldsymbol{v}_h\|_{\Omega^n}^2 + c_0\delta\mathcal{K}\left((\epsilon+\epsilon^{-1})h^2+\epsilon\right)g_h^n(\boldsymbol{v}_h,\boldsymbol{v}_h).\end{aligned}$$

Using (9) and choosing $\epsilon = \frac{1}{2c_0c_\delta s w_{\max}}$, we have $c_0\delta\epsilon \leq \frac{\Delta t}{2}$ and

$$\|\boldsymbol{v}_h\|_{S_\delta^n}^2 \leq c_1(w_{\max})\Delta t\|\boldsymbol{v}_h\|_{\Omega^n}^2 + \frac{\Delta t}{2}\|\nabla\boldsymbol{v}_h\|_{\Omega^n}^2 + c_2(w_{\max})\Delta t\mathcal{K}g_h^n(\boldsymbol{v}_h,\boldsymbol{v}_h) \tag{37}$$

for $h < 1$ with the constants $c_1(w_{\max})$, $c_2(w_{\max})$ given in the statement. The inequality (35) follows by combining (37) with the equality

$$\|\boldsymbol{v}_h\|_{\Omega_\delta^n}^2 = \|\boldsymbol{v}_h\|_{\Omega^n}^2 + \|\boldsymbol{v}_h\|_{S_\delta^n}^2. \tag{38}$$

□

Now we are ready to show a stability result for the discrete formulation (15).

Theorem 4.2 *Let $\boldsymbol{u}_h = (\boldsymbol{u}_h^k)_{k=1}^N$, $p_h = (p_h^k)_{k=1}^N$ be the solution of (15) for $s = 1$, $\gamma_g \geq c_2(w_{\max})\mathcal{K}$, where $c_2(w_{\max})$ denotes the constant from Lemma 4.1 and γ_D sufficiently large.*

Under the regularity assumptions stated above, it holds for $n \geq 1$ that

$$\begin{aligned}&\|\boldsymbol{u}_h^n\|_{\Omega^n}^2 + \sum_{k=1}^n \|\boldsymbol{u}_h^k - \boldsymbol{u}_h^{k-1}\|_{\Omega^k}^2 + \Delta t\left(|||\boldsymbol{u}_h^k|||_{h,k}^2 + \gamma_p s_h^k(p_h^k, p_h^k)\right)\\ &\quad\leq c\exp(c_1(w_{\max})t_n)\left(\|\boldsymbol{u}^0\|_{\Omega^0}^2 + t_n\|\boldsymbol{f}\|_{\infty,0}^2\right),\end{aligned} \tag{39}$$

with $c_1(w_{\max})$ given in Lemma 4.1 and $\boldsymbol{u}_h^0 := \pi_h^1 E\boldsymbol{u}^0$.

Proof Testing (15) with $\boldsymbol{v}_h = 2\Delta t \boldsymbol{u}_h^n$, $q_h = 2\Delta t p_h^n$, using the coercivity (26) and the relation

$$-2(\boldsymbol{u}_h^n, \boldsymbol{u}_h^{n-1})_{\Omega^n} = \|\boldsymbol{u}_h^n - \boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2 - \|\boldsymbol{u}_h^n\|_{\Omega^n}^2 - \|\boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2 \tag{40}$$

yields for $n > 1$

$$\begin{aligned}&\|\boldsymbol{u}_h^n\|_{\Omega^n}^2 + \|\boldsymbol{u}_h^n - \boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2 - \|\boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2 + \Delta t |||\boldsymbol{u}_h^n|||_{h,n}^2 + \Delta t \gamma_p s_h^n(p_h^n, p_h^n) \\ &\quad \le 2\Delta t (\boldsymbol{f}, \boldsymbol{u}_h^n)_{\Omega^n}.\end{aligned} \tag{41}$$

We bring the term $\|\boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2$ to Ω^{n-1} by using Lemma 4.1

$$\begin{aligned}\|\boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2 \le (1 + c_1(w_{\max})\Delta t)\, \|\boldsymbol{u}_h^{n-1}\|_{\Omega^{n-1}}^2 + \frac{\Delta t}{2}\|\nabla \boldsymbol{u}_h^{n-1}\|_{\Omega^{n-1}}^2 \\ + c_2(w_{\max})\mathcal{K}\Delta t g_h^{n-1}(\boldsymbol{u}_h^{n-1}, \boldsymbol{u}_h^{n-1}).\end{aligned} \tag{42}$$

Inserting (42) into (41) we have

$$\begin{aligned}&\|\boldsymbol{u}_h^n\|_{\Omega^n}^2 + \|\boldsymbol{u}_h^n - \boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2 + \Delta t \left(|||\boldsymbol{u}_h^n|||_{h,n}^2 + \gamma_p s_h^n(p_h^n, p_h^n)\right) \\ &\quad \le 2\Delta t (\boldsymbol{f}, \boldsymbol{u}_h^n)_{\Omega^n} + (1 + c_1(w_{\max})\Delta t)\, \|\boldsymbol{u}_h^{n-1}\|_{\Omega^{n-1}}^2 + \frac{\Delta t}{2}\|\nabla \boldsymbol{u}_h^{n-1}\|_{\Omega^{n-1}}^2 \\ &\qquad + \Delta t \gamma_g g_h^{n-1}(\boldsymbol{u}_h^{n-1}, \boldsymbol{u}_h^{n-1})\end{aligned} \tag{43}$$

for $\gamma_g \ge c_2(w_{\max})K$ and γ_D sufficiently large. For $n = 1$, we have instead of (41)

$$\begin{aligned}&\|\boldsymbol{u}_h^1\|_{\Omega^1}^2 + \|\boldsymbol{u}_h^1 - E\boldsymbol{u}^0\|_{\Omega^1}^2 - \|E\boldsymbol{u}^0\|_{\Omega^1}^2 + \Delta t \left(|||\boldsymbol{u}_h^1|||_{h,1}^2 + \gamma_p s_h^1(p_h^1, p_h^1)\right) \\ &\quad = 2\Delta t (\boldsymbol{f}, \boldsymbol{u}_h^1)_{\Omega^1}.\end{aligned} \tag{44}$$

In both cases ($n \ge 1$) we use the Cauchy–Schwarz and Young's inequality for the first term on the right-hand side to get

$$2\Delta t (\boldsymbol{f}, \boldsymbol{u}_h^n)_{\Omega^n} \le \Delta t \|\boldsymbol{u}_h^n\|_{\Omega^n}^2 + \Delta t \|\boldsymbol{f}\|_{\Omega^n}^2.$$

Summing over $k = 0, \ldots, n$ in (43) and using the L^2-stability of the extension of the initial value yields

$$\begin{aligned}&\|\boldsymbol{u}_h^n\|_{\Omega^n}^2 + \sum_{k=1}^{n} \|\boldsymbol{u}_h^k - \boldsymbol{u}_h^{k-1}\|_{\Omega^k}^2 + \frac{\Delta t}{2}\left(|||\boldsymbol{u}_h^k|||_{h,k}^2 + 2\gamma_p s_h^k(p_h^k, p_h^k)\right) \\ &\qquad \le c\|\boldsymbol{u}^0\|_{\Omega^0} + 2t_n \|\boldsymbol{f}\|_{\infty,0}^2 + c_1(w_{\max})\Delta t \sum_{k=0}^{n-1} \|\boldsymbol{u}_h^k\|_{\Omega^k}^2.\end{aligned} \tag{45}$$

Application of a discrete Gronwall lemma yields the statement. □

Remark 4.3 (BDF(2)) For the BDF(2) variant, we get the stability estimate (39) with the weaker dissipation $\|\boldsymbol{u}_h^k - 2\boldsymbol{u}_h^{k-1} + \boldsymbol{u}_h^{k-2}\|_{\Omega^k}^2$ instead of $\|\boldsymbol{u}_h^k - \boldsymbol{u}_h^{k-1}\|_{\Omega^k}^2$. To this end, one uses the relation

$$(3\boldsymbol{u}_h^n - 4\boldsymbol{u}_h^{n-1} + \boldsymbol{u}_h^{n-2}, \boldsymbol{u}_h^n)_{\Omega^n} = \frac{1}{2}\big(\|\boldsymbol{u}_h^n\|_{\Omega^n}^2 - \|\boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2 + \|2\boldsymbol{u}_h^n - \boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2 - \|2\boldsymbol{u}_h^{n-1} - \boldsymbol{u}_h^{n-2}\|_{\Omega^n}^2 + \|\boldsymbol{u}_h^n - 2\boldsymbol{u}_h^{n-1} + \boldsymbol{u}_h^{n-1}\|_{\Omega^n}^2\big)$$

instead of (40).

4.1 Stability estimate for the pressure

We show the following stability estimates for the L^2- and H^1-semi-norm of pressure.

Lemma 4.4 *Let $(\boldsymbol{u}_h^n, p_h^n)$ be the discrete solution of (15). For $n \geq 1$ it holds that*

$$\|p_h^n\|_{\Omega^n}^2 \leq c\left(\|D_t^{(s)}\boldsymbol{u}_h^n\|_{\Omega^n}^2 + |||\boldsymbol{u}_h^n|||_{h,n}^2 + s_h^n(p_h^n, p_h^n) + \|\boldsymbol{f}\|_{\Omega^n}^2\right), \tag{46}$$

$$h^2\|\nabla p_h^n\|_{\Omega^n}^2 \leq c\left(h^2\|D_t^{(s)}\boldsymbol{u}_h^n\|_{\Omega^n}^2 + |||\boldsymbol{u}_h^n|||_{h,n}^2 + s_h^n(p_h^n, p_h^n) + h^2\|\boldsymbol{f}\|_{\Omega^n}^2\right), \tag{47}$$

where $\boldsymbol{u}_h^0 = \pi_h^1 E\boldsymbol{u}^0$.

Proof First, we derive a bound for $h^2\|\nabla p_h^n\|_{\Omega^n}^2$. To this end, we extend ∇p_h^n by zero to $\Omega_{\delta,h}^n \backslash \Omega_h^n$, using the same notation for the extended function. We insert $\pm C_h^n \nabla p_h^n$, where $C_h^n : L^2(\Omega_{\delta,h}^n)^d \to \mathcal{V}_h^n$ is the interpolation operator used in (10)-(13), and integrate by parts

$$\begin{aligned} h^2\|\nabla p_h^n\|_{\Omega^n}^2 &= h^2(\nabla p_h^n - C_h^n\nabla p_h^n, \nabla p_h^n)_{\Omega^n} + h^2(C_h^n\nabla p_h^n, \nabla p_h^n)_{\Omega^n} \\ &= h^2(\nabla p_h^n - C_h^n\nabla p_h^n, \nabla p_h^n)_{\Omega^n} - h^2(\operatorname{div}(C_h^n\nabla p_h^n), p_h^n)_{\Omega^n} \\ &\quad + h^2(C_h^n\nabla p_h^n, p_h^n\boldsymbol{n})_{\partial\Omega^n}. \end{aligned} \tag{48}$$

For the first term, we have by means of (12) and Young's inequality

$$\begin{aligned} h^2(\nabla p_h^n - C_h^n\nabla p_h^n, \nabla p_h^n)_{\Omega^n} &\leq h^2\|\nabla p_h^n - C_h^n\nabla p_h^n\|_{\Omega_h^n}\|\nabla p_h^n\|_{\Omega^n} \\ &\leq cs_h^n(p_h^n, p_h^n) + \frac{h^2}{4}\|\nabla p_h^n\|_{\Omega^n}^2. \end{aligned} \tag{49}$$

The last term in (49) will be absorbed into the left-hand side of (48). For the second term on the right-hand side of (48), we use that $(\boldsymbol{u}_h^n, p_h^n)$ solves the discrete system (15)

$$\begin{aligned} -h^2(\operatorname{div}(C_h^n\nabla p_h^n), p_h^n)_{\Omega^n} = -h^2\Big(&\left(D_t^{(s)}\boldsymbol{u}_h^n, C_h^n\nabla p_h^n\right)_{\Omega^n} + \left(\nabla\boldsymbol{u}_h^n, \nabla(C_h^n\nabla p_h^n)\right)_{\Omega^n} \\ &+ a_D^n(\boldsymbol{u}_h^n, p_h^n; C_h^n\nabla p_h^n, 0) + \gamma_g g_h^n(\boldsymbol{u}_h^n, C_h^n\nabla p_h^n) \\ &- (\boldsymbol{f}, C_h^n\nabla p_h^n)_{\Omega^n}\Big). \end{aligned} \tag{50}$$

To estimate the first term on the right-hand side of (50), we use the Cauchy–Schwarz inequality and (12) to get

$$-h^2\left(D_t^{(s)}\boldsymbol{u}_h^n, C_h^n\nabla p_h^n\right)_{\Omega^n} \leq ch\|D_t^{(s)}\boldsymbol{u}_h^n\|_{\Omega^n}\left(h\|\nabla p_h^n\|_{\Omega^n} + s_h^n(p_h^n, p_h^n)^{1/2}\right). \quad (51)$$

Similarly, we get for the last term on the right-hand side of (50)

$$h^2\left(\boldsymbol{f}, C_h^n\nabla p_h^n\right)_{\Omega^n} \leq ch\|\boldsymbol{f}\|_{\Omega^n}\left(h\|\nabla p_h^n\|_{\Omega^n} + s_h^n(p_h^n, p_h^n)^{1/2}\right). \quad (52)$$

For the second-term on the right-hand side of (50), we use an inverse inequality on Ω_h^n and (12)

$$\begin{aligned}-h^2\left(\nabla\boldsymbol{u}_h^n, \nabla(C_h^n\nabla p_h^n)\right)_{\Omega^n} &\leq ch^2\|\nabla\boldsymbol{u}_h^n\|_{\Omega^n}\|\nabla(C_h^n\nabla p_h^n)\|_{\Omega_h^n}\\ &\leq ch\|\nabla\boldsymbol{u}_h^n\|_{\Omega^n}\|C_h^n\nabla p_h^n\|_{\Omega_h^n}\\ &\leq c\|\nabla\boldsymbol{u}_h^n\|_{\Omega^n}\left(h\|\nabla p_h^n\|_{\Omega_h^n} + s_h^n(p_h^n, p_h^n)^{\frac{1}{2}}\right).\end{aligned}$$

For the Nitsche term a_D^n, we have as in (29)

$$\begin{aligned}&-h^2a_D^n(\boldsymbol{u}_h^n, p_h^n; C_h^n\nabla p_h^n, 0)\\ &\quad= -h^2\left(\frac{\gamma_D}{h}\left(\boldsymbol{u}_h^n, C_h^n\nabla p_h^n\right)_{\partial\Omega^n} - \left(\partial_n\boldsymbol{u}_h^n - p_h^n\boldsymbol{n}, C_h^n\nabla p_h^n\right)_{\partial\Omega^n} + \left(\boldsymbol{u}_h^n, \partial_n(C_h^n\nabla p_h^n)\right)_{\partial\Omega^n}\right)\\ &\quad\leq ch|||u_h^n|||_{h,n}\|C_h^n\nabla p_h^n\|_{\Omega_h^n} - h^2\left(p_h^n\boldsymbol{n}, C_h^n\nabla p_h^n\right)_{\partial\Omega^n}\\ &\quad\leq c|||u_h^n|||_{h,n}\left(h\|\nabla p_h^n\|_{\Omega^n} + s_h^n(p_h^n, p_h^n)^{1/2}\right) - h^2\left(p_h^n\boldsymbol{n}, C_h^n\nabla p_h^n\right)_{\partial\Omega^n}. \quad (53)\end{aligned}$$

In the last step (12) has been used. Note that the boundary term on the right-hand side will cancel out with the third term in (48). For the ghost penalty we have by means of an inverse inequality and (12)

$$\begin{aligned}h^2g_h^n(\boldsymbol{u}_h^n, C_h^n\nabla p_h^n) &\leq chg_h^n(\boldsymbol{u}_h^n, \boldsymbol{u}_h^n)^{1/2}\|C_h^n\nabla p_h^n\|_{\Omega_{h,\delta}^n}\\ &\leq cg_h^n(\boldsymbol{u}_h^n, \boldsymbol{u}_h^n)^{1/2}\left(h\|\nabla p_h^n\|_{\Omega^n} + s_h^n(p_h^n, p_h^n)^{1/2}\right).\end{aligned} \quad (54)$$

Altogether, (50)-(51) and (53)-(54) result in

$$\begin{aligned}&-h^2(\mathrm{div}(C_h^n\nabla p_h^n), p_h^n)_{\Omega^n} + h^2\left(p_h^n\boldsymbol{n}, C_h^n\nabla p_h^n\right)_{\partial\Omega^n}\\ &\quad\leq c\left(h\|\nabla p_h^n\|_{\Omega^n} + s_h^n(p_h^n, p_h^n)^{1/2}\right)\left(h\left\|D_t^{(s)}\boldsymbol{u}_h^n\right\|_{\Omega^n} + |||\boldsymbol{u}_h^n|||_{h,n} + h\|\boldsymbol{f}\|_{\Omega^n}\right)\\ &\quad\leq \frac{h^2}{4}\|\nabla p_h^n\|_{\Omega^n}^2 + c\left(h^2\left\|D_t^{(s)}\boldsymbol{u}_h^n\right\|_{\Omega^n}^2 + |||\boldsymbol{u}_h^n|||_{h,n}^2 + h^2\|\boldsymbol{f}\|_{\Omega^n}^2 + s_h^n(p_h^n, p_h^n)\right). \quad (55)\end{aligned}$$

In the last step we have applied Young's inequality. Combination of (48), (49) an (55) yields (47). To show (46) we start using the modified inf-sup condition (Lemma 3.6)

$$\beta \|p_h^n\|_{\Omega^n} \leq \sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h^n} \frac{(\operatorname{div} \boldsymbol{v}_h, p_h^n)_{\Omega^n} - (\boldsymbol{v}_h \cdot \boldsymbol{n}, p_h^n)_{\partial\Omega^n}}{|||\boldsymbol{v}_h|||_{h,n}} + h\|\nabla p_h^n\|_{\Omega^n}. \tag{56}$$

By (15), we have

$$(\operatorname{div} \boldsymbol{v}_h, p_h^n)_{\Omega^n} - (\boldsymbol{v}_h \cdot n, p_h^n)_{\partial\Omega^n} = (D_t^{(s)} \boldsymbol{u}_h^n, \boldsymbol{v}_h)_{\Omega^n} + \mathcal{A}_h^n(\boldsymbol{u}_h^n, 0; \boldsymbol{v}_h, 0) - (\boldsymbol{f}, \boldsymbol{v}_h)_{\Omega^n} \tag{57}$$

To estimate the right-hand side of (57), we use the continuity of the bilinear form $\mathcal{A}_h^n$ (27) and the Cauchy–Schwarz inequality

$$\begin{aligned}
&(D_t^{(s)} \boldsymbol{u}_h^n, \boldsymbol{v}_h)_{\Omega^n} + \mathcal{A}_h^n(\boldsymbol{u}_h^n, 0; \boldsymbol{v}_h, 0) - (\boldsymbol{f}, \boldsymbol{v}_h)_{\Omega^n} \\
&\quad \leq c \left(\|D_t^{(s)} \boldsymbol{u}_h^n\|_{\Omega^n} + \|\boldsymbol{f}\|_{\Omega^n} \right) \|\boldsymbol{v}_h\|_{\Omega^n} + c|||\boldsymbol{u}_h^n|||_{h,n} |||\boldsymbol{v}_h|||_{h,n} \\
&\quad \leq c \left(\|D_t^{(s)} \boldsymbol{u}_h^n\|_{\Omega^n} + |||\boldsymbol{u}_h^n|||_{h,n} + \|\boldsymbol{f}\|_{\Omega^n} \right) |||\boldsymbol{v}_h|||_{h,n}.
\end{aligned} \tag{58}$$

In the last step, we have used that $\|\boldsymbol{v}_h^n\|_{\Omega^n} \leq c\,(\|\nabla \boldsymbol{v}_h\|_{\Omega^n} + \|\boldsymbol{v}_h\|_{\partial\Omega^n}) \leq c|||\boldsymbol{v}_h|||_{h,n}$ by a Poincaré- type estimate. Combination of (56)–(58) and (47) yields (46). □

Lemma 4.4 gives a stability result for $\|\nabla p_h^k\|_{\Omega^k}$, which results in the following corollary:

Corollary 4.5 *Under the assumptions of Theorem (4.2), it holds for $s = 1$ that*

$$\begin{aligned}
&\|\boldsymbol{u}_h^n\|_{\Omega^n}^2 + \Delta t \sum_{k=1}^{n} \left(|||\boldsymbol{u}_h^k|||_{h,k}^2 + \frac{1}{\Delta t} \|\boldsymbol{u}_h^k - \boldsymbol{u}_h^{k-1}\|_{\Omega^k}^2 + \gamma_p s_h^k(p_h^k, p_h^k) + \min\{h^2, \Delta t\} \|\nabla p_h^k\|_{\Omega^k}^2 \right) \\
&\quad \leq \exp(c_1(w_{\max}) t_n) \left(c\|\boldsymbol{u}^0\|_{\Omega^0} + 2t_n \|\boldsymbol{f}\|_{\infty,0}^2 \right).
\end{aligned} \tag{59}$$

For $s = 2$, we have

$$\begin{aligned}
&\|\boldsymbol{u}_h^n\|_{\Omega^n}^2 + \Delta t \sum_{k=1}^{n} \left(|||\boldsymbol{u}_h^k|||_{h,k}^2 + \gamma_p s_h^k(p_h^k, p_h^k) + \min\{h^2, \Delta t^2\} \|\nabla p_h^k\|_{\Omega^k}^2 \right) \\
&\quad \leq \exp(c_1(w_{\max}) t_n) \left(c\|\boldsymbol{u}^0\|_{\Omega^0} + 2t_n \|\boldsymbol{f}\|_{\infty,0}^2 \right).
\end{aligned} \tag{60}$$

Proof We start by proving (59) for $s = 1$. To this end, we distinguish between the cases $\Delta t \geq h^2$ and $\Delta t < h^2$. In the first case, we note that, by (47)

$$\begin{aligned}
\Delta t h^2 \|\nabla p_h^n\|_{\Omega^n}^2 &\leq c\Delta t h^2 \|D_t^{(1)} \boldsymbol{u}_h^n\|_{\Omega^n}^2 + c\Delta t \left(|||\boldsymbol{u}_h^n|||_{h,n}^2 + s_h^n(p_h^n, p_h^n) + \|\boldsymbol{f}\|_{\Omega^n}^2 \right) \\
&\leq c\Delta t^2 \|D_t^{(1)} \boldsymbol{u}_h^n\|_{\Omega^n}^2 + c\Delta t \left(|||\boldsymbol{u}_h^n|||_{h,n}^2 + s_h^n(p_h^n, p_h^n) + \|\boldsymbol{f}\|_{\Omega^n}^2 \right).
\end{aligned} \tag{61}$$

As $\Delta t^2 \| D_t^{(1)} \boldsymbol{u}_h^n \|^2_{\Omega^n} = \| \boldsymbol{u}_h^n - \boldsymbol{u}_h^{n-1} \|^2_{\Omega^n}$ (59) follows from Theorem 4.2. For $\Delta t < h^2$, we multiply (61) by $\frac{\Delta t}{h^2}$ to get

$$\Delta t^2 \| \nabla p_h^n \|^2_{\Omega^n} \leq c \Delta t^2 \| D_t^{(1)} \boldsymbol{u}_h^n \|^2_{\Omega^n} + c \Delta t \frac{\Delta t}{h^2} \left(|||\boldsymbol{u}_h^n|||^2_{h,n} + s_h^n(p_h^n, p_h^n) + \| \boldsymbol{f} \|^2_{\Omega^n} \right)$$

and use the same argumentation. For $s = 2$, we do not have control over $\Delta t^2 \| D_t^{(2)} \boldsymbol{u}_h^n \|^2_{\Omega^n}$. Instead, we use the estimate

$$\Delta t^3 \| D_t^{(2)} \boldsymbol{u}_h^n \|^2_{\Omega^n} \leq c \Delta t \sum_{k=0}^{2} \left(\| \boldsymbol{u}_h^{n-k} \|^2_{\Omega^{n-k}} + \Delta t \| \nabla \boldsymbol{u}_h^{n-k} \|^2_{\Omega^{n-k}} + \mathcal{K} \Delta t g_h^{n-k}(\boldsymbol{u}_h^{n-k}, \boldsymbol{u}_h^{n-k}) \right)$$

that follows from the triangle inequality and (35). The estimate (60) follows by a similar argumentation by distinguishing between the cases $\Delta t \lesseqgtr h$. □

Concerning the L^2-norm of the pressure, Lemma 4.4 gives a stability result only for

$$\Delta t^2 \sum_{k=1}^{n} \| p_h^k \|^2_{\Omega^k},$$

even in the case $s = 1$. In the case of fixed domains and fixed discretisations, a stability estimate for $\| p_h^k \|$ can be derived by showing an upper bound for the right-hand side in (46), including the term $\frac{1}{\Delta t} \| \boldsymbol{u}_h^n - \boldsymbol{u}_h^{n-1} \|^2_{\Omega^n}$, see for example Besier and Wollner [4]. The argumentation requires, however, that the term $(\operatorname{div} \boldsymbol{u}_h^{n-1}, \xi_h^n)_{\Omega^n}$ vanishes for $\xi_h^n \in \mathcal{L}_h^n$. This is not true in the case of time-dependent domains, as $\boldsymbol{u}_h^{n-1}$ is not discrete divergence-free with respect to Ω^n

$$(\operatorname{div} \boldsymbol{u}_h^{n-1}, \xi_h^n)_{\Omega^n} \neq 0$$

for certain $\xi_h^n \in \mathcal{L}_h^n$. Moreover, the domain mismatch $\Omega^{n-1} \neq \Omega^n$ causes additional problems in the transfer of the term $|||\boldsymbol{u}_h^{n-1}|||_{h,n} \neq |||\boldsymbol{u}_h^{n-1}|||_{h,n-1}$ from one time level to the previous one. In the error analysis developed in the following section, we will therefore use the H^1-stability results in Corollary 4.5 for the pressure variable.

5 Error analysis

The energy error analysis for the velocities follows largely the argumentation of Lehrenfeld and Olshanskii [42] and is based on Galerkin orthogonality and the stability result of Theorem 4.2. We write $\boldsymbol{u}^n := \boldsymbol{u}(t_n)$, $p^n := p(t_n)$ and introduce the notation

$$\begin{aligned} \boldsymbol{e}_u^n &:= \boldsymbol{u}^n - \boldsymbol{u}_h^n, \quad \eta_u^n := \boldsymbol{u}^n - I_h^n \boldsymbol{u}^n, \quad \xi_{h,u}^n := I_h^n \boldsymbol{u}^n - \boldsymbol{u}_h^n, \\ e_p^n &:= p^n - p_h^n, \quad \eta_p^n := p^n - i_h^n p^n, \quad \xi_{h,p}^n := i_h^n p^n - p_h^n \end{aligned}$$

for $n \geq 1$, where I_h^n denotes the standard Lagrangian nodal interpolation to $\mathcal{T}_{h,\delta}^n$ and i_h^n a generalised L^2-stable interpolation (for example the Clément interpolation) to $\mathcal{T}_h^n$. Moreover, we set

$$\boldsymbol{e}_u^0 = \eta_u^0 = \xi_{h,u}^0 = 0.$$

This is possible, as $\boldsymbol{u}_h^0$ cancels out in the summed space-time system (19). The following estimates for the interpolation errors are well-known

$$\|\eta_u^n\|_{H^l(\Omega)} \leq ch^{k-l}\|\boldsymbol{u}^n\|_{H^k(\Omega)} \qquad \text{for } 0 \leq l \leq 1, \quad 2 \leq k \leq m+1, \tag{62}$$

$$\|\eta_p^n\|_{H^l(\Omega)} \leq ch^{k-l}\|p^n\|_{H^k(\Omega)} \qquad \text{for } 0 \leq l \leq 1, \quad 1 \leq k \leq m+1, \tag{63}$$

$$\|\eta_p^n\|_{H^l(\partial\Omega)} \leq ch^{k-l-1/2}\|p^n\|_{H^k(\Omega)} \qquad \text{for } 0 \leq l \leq 1, \quad 1 < k \leq m+1. \tag{64}$$

We will again make use of the extension operators E^n introduced in Sect. 3.3. For better readability, we will sometimes skip the operators E^n assuming that quantities that would be undefined on the domains of integration are extended smoothly.

For the error analysis, we assume that the solution $(\boldsymbol{u}, p)$ to (5) lies in $L^2(I, H^{m+1}(\Omega(t))^d) \times L^2(I, H^m(\Omega(t)))$ for $m \geq 1$. Then, we can incorporate the Nitsche terms in the variational formulation on the continuous level and see that $(\boldsymbol{u}, p)$ is the solution to

$$\begin{aligned}(\partial_t \boldsymbol{u}, \boldsymbol{v})_{\Omega(t)} + \mathcal{A}_S(\boldsymbol{u}, p; \boldsymbol{v}, q) + a_D^n(\boldsymbol{u}, p; \boldsymbol{v}, q) \\ = (\boldsymbol{f}, \boldsymbol{v})_{\Omega(t)} \quad \forall \boldsymbol{v} \in \tilde{\mathcal{V}}(t), \ q \in \mathcal{L}(t) \quad \text{a.e. in } t \in I,\end{aligned} \tag{65}$$

where

$$\tilde{\mathcal{V}}(t) := H^1(\Omega(t))^d.$$

5.1 Energy error

As a starting point for the error estimation, we subtract (15) from (65) to obtain the orthogonality relation

$$\begin{aligned}&\left(D_t^{(s)} \boldsymbol{e}_u^n, \boldsymbol{v}_h\right)_{\Omega^n} + (\mathcal{A}_S^n + a_D^n)(\boldsymbol{e}_u^n, e_p^n; \boldsymbol{v}_h, q_h) + \gamma_g g_h^n(\boldsymbol{e}_u^n, \boldsymbol{v}_h) + \gamma_p s_h^n(\xi_{h,p}^n, q_h) \\ &= \underbrace{(D_t^{(s)} \boldsymbol{u}^n - \partial_t \boldsymbol{u}(t_n), \boldsymbol{v}_h)_{\Omega^n} + \gamma_g g_h^n(\boldsymbol{u}^n, \boldsymbol{v}_h) + \gamma_p s_h^n(i_h^n p^n, q_h)}_{=:\mathcal{E}_c^n(\boldsymbol{v}_h, q_h)} \qquad \forall \boldsymbol{v}_h \in \mathcal{V}_h^n, q_h \in \mathcal{L}_h^n,\end{aligned} \tag{66}$$

for $n \geq s$ with the consistency error $\mathcal{E}_c^n(\boldsymbol{v}_h, q_h)$. Note that this relation holds in particular also for $n = s$, as we have defined $\boldsymbol{e}_u^0 = 0$. We have used a different splitting in the pressure stabilisation s_h^n compared to the other terms, in order to include the case $p \in H^1(\Omega)$ $(m = 1)$, where $s_h^n(p^n, q_h)$ would not be well-defined.

We further split (66) into interpolation and discrete error parts

$$\begin{aligned}\left(D_t^{(s)}\xi_{h,u}^n, \boldsymbol{v}_h\right)_{\Omega^n} + \left(\mathcal{A}_S^n + a_D^n\right)(\xi_{h,u}^n, \xi_{h,p}^n; \boldsymbol{v}_h, q_h) + \gamma_g g_h^n(\xi_{h,u}^n, \boldsymbol{v}_h) + \gamma_p s_h^n(\xi_{h,p}^n, q_h) \\ = -\mathcal{E}_i^n(\boldsymbol{v}_h, q_h) + \mathcal{E}_c^n(\boldsymbol{v}_h, q_h) \quad \forall \boldsymbol{v}_h \in \mathcal{V}_h^n, q_h \in \mathcal{L}_h^n,\end{aligned} \tag{67}$$

where the interpolation error is defined by

$$\mathcal{E}_i^n(\boldsymbol{v}_h, q_h) := (D_t^{(s)}\eta_u^n, \boldsymbol{v}_h)_{\Omega^n} + \left(\mathcal{A}_S^n + a_D^n\right)(\eta_u^n, \eta_p^n; \boldsymbol{v}_h, q_h) + \gamma_g g_h^n(\eta_u^n, \boldsymbol{v}_h). \tag{68}$$

We will apply the stability result of Theorem 4.2 to (67), which will be the basis of the error estimate. For better readability, we will restrict restrict ourselves again to the case $s = 1$ first. Let us first estimate the consistency and interpolation errors.

Lemma 5.1 (Consistency error) *Let $\boldsymbol{u} \in W^{2,\infty}(I_n, L^2(\Omega^n)^d) \cap L^\infty(I_n, H^{m+1}(\Omega^n)^d)$ and $p \in L^\infty(I_n, H^m(\Omega^n))$. Under the assumptions made in Sect. 3, including Assumption 3.2, it holds for $s = 1$, $\boldsymbol{v}_h \in \mathcal{V}_h^n$, $q_h \in \mathcal{L}_h^n$ and $n \geq 1$ that*

$$\begin{aligned}|\mathcal{E}_c^n(\boldsymbol{v}_h, q_h)| \leq c\Delta t^{\frac{1}{2}} \|\partial_t^2 \boldsymbol{u}\|_{Q^n} \|\boldsymbol{v}_h\|_{\Omega^n} \\ + ch^m \left(\|\boldsymbol{u}\|_{H^{m+1}(\Omega^n)} + \|p\|_{H^m(\Omega^n)}\right) \left(g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h)^{1/2} + s_h^n(q_h, q_h)^{1/2}\right).\end{aligned}$$

Proof For the first part of the consistency error, we have using integration by parts and a Cauchy–Schwarz inequality in time

$$\begin{aligned}\frac{1}{\Delta t}\left(\boldsymbol{u}^n - E^n\boldsymbol{u}^{n-1}\right) - \partial_t \boldsymbol{u}(t_n) &= -\frac{1}{\Delta t}\int_{t_{n-1}}^{t_n} \partial_t(E^n\boldsymbol{u}(t)) - \partial_t(E^n\boldsymbol{u}(t_n))\,dt \\ &= -\frac{1}{\Delta t}\int_{t_{n-1}}^{t_n} (t - t_{n-1})\partial_t^2(E^n\boldsymbol{u}(t))\,dt \\ &\leq \frac{1}{\Delta t}\left(\int_{t_{n-1}}^{t_n} (t - t_{n-1})^2 dt\right)^{1/2}\left(\int_{t_{n-1}}^{t_n} \partial_t^2(E^n\boldsymbol{u}(t))^2 dt\right)^{1/2} \\ &\leq \Delta t^{1/2}\left(\int_{t_{n-1}}^{t_n} \partial_t^2(E^n\boldsymbol{u}(t))^2 dt\right)^{1/2}.\end{aligned}$$

Using (22) this implies

$$\begin{aligned}\left|\frac{1}{\Delta t}\left(\boldsymbol{u}^n - E^n\boldsymbol{u}^{n-1}, \boldsymbol{v}_h\right)_{\Omega^n} - (\partial_t \boldsymbol{u}(t_n), \boldsymbol{v}_h)_{\Omega^n}\right| &\leq c\Delta t^{1/2}\|\partial_t^2(E^n\boldsymbol{u})\|_{Q_\delta^n}\|\boldsymbol{v}_h\|_{\Omega^n} \\ &\leq c\Delta t^{1/2}\|\partial_t^2\boldsymbol{u}\|_{Q^n}\|\boldsymbol{v}_h\|_{\Omega^n}.\end{aligned} \tag{69}$$

The extension operator E^n is needed, as the integration domain in the left-hand side of (69) includes parts, that lie outside the physical domain Q^n. For the ghost penalty part, we have with Lemma 3.3 and the H^{m+1}-stability of the extension (20)

$$g_h^n(\boldsymbol{u}^n, \boldsymbol{v}_h) \leq g_h^n(\boldsymbol{u}^n, \boldsymbol{u}^n)^{1/2} g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h)^{1/2} \leq ch^m \|\boldsymbol{u}^n\|_{H^{m+1}(\Omega^n)} g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h)^{1/2}$$

Concerning the pressure stabilisation, we note that for $p^n \in H^1(\Omega^n)$ the term $s_h^n(p^n, p^n)$ is not well-defined. For this reason we distinguish between the cases $m = 1$ and $m \geq 2$. In the first case, we estimate using (10) and the H^1-stability of the interpolation

$$s_h^n(i_h^n p^n, q_h) \leq ch \|i_h^n p^n\|_{H^1(\Omega^n)} s_h^n(q_h, q_h)^{1/2} \leq ch \|p^n\|_{H^1(\Omega^n)} s_h^n(q_h, q_h)^{1/2}.$$

For $m \geq 2$, we insert $\pm p^n$ and use (10), (14) and the interpolation error estimate (64)

$$\begin{aligned} s_h^n(i_h^n p^n, q_h) &\leq \left(s_h^n(\eta_p^n, \eta_p^n)^{1/2} + s_h^n(p^n, p^n)^{1/2} \right) s_h^n(q_h, q_h)^{1/2} \\ &\leq ch^m \|p^n\|_{H^m(\Omega^n)} s_h^n(q_h, q_h)^{1/2}. \end{aligned}$$

□

Lemma 5.2 (Interpolation error) *Let* $\boldsymbol{u} \in L^\infty(I_n, H^{m+1}(\Omega(t))^d)$, $\partial_t \boldsymbol{u} \in L^\infty(I_n, H^m(\Omega(t))^d)$, $p \in L^\infty(I_n, H^m(\Omega(t))$. *Under the assumptions made in Sect. 3, including Assumption 3.2, it holds for* $\boldsymbol{v}_h \in \mathcal{V}_h^n$ *and* $q_h \in \mathcal{L}_h^n$ *that*

$$\begin{aligned} |\mathcal{E}_i^n(\boldsymbol{v}_h, q_h)| \leq ch^m \big(&\|\boldsymbol{u}\|_{\infty,m+1,I_n} + \|\partial_t \boldsymbol{u}\|_{\infty,m,I_n} \\ &+ \|p\|_{H^m(\Omega^n)} \big) \left(|||\boldsymbol{v}_h|||_{h,n} + h\|\nabla q_h\|_{\Omega^n} \right). \end{aligned}$$

Proof We estimate the interpolation error (68) term by term. For the first term we use that we can exchange time derivative and interpolation operator $\partial_t I_h \boldsymbol{u}^n = I_h \partial_t \boldsymbol{u}(t_n)$

$$\begin{aligned} \left| \frac{1}{\Delta t} (\eta_u^n - \eta_u^{n-1}, \boldsymbol{v}_h)_{\Omega^n} \right| &\leq \frac{1}{\Delta t} \|\eta_u^n - \eta_u^{n-1}\|_{\Omega^n} \|\boldsymbol{v}_h\|_{\Omega^n} \\ &= \frac{1}{\Delta t} \left\| \int_{t_{n-1}}^{t_n} \partial_t (\boldsymbol{u}(t) - I_h \boldsymbol{u}(t))\, dt \right\|_{\Omega^n} \|\boldsymbol{v}_h\|_{\Omega^n} \\ &\leq h^m \|\partial_t (E^n \boldsymbol{u})\|_{\infty,m,I_n} \|\boldsymbol{v}_h\|_{\Omega^n}. \end{aligned} \tag{70}$$

We note again that the integration domain in the first norm on the right-hand side includes parts, that might lie outside the physical domain Q^n. By means of (21) we conclude

$$\left| \frac{1}{\Delta t} (\eta_u^n - \eta_u^{n-1}, \boldsymbol{v}_h)_{\Omega^n} \right| \leq ch^m \left(\|\partial_t \boldsymbol{u}\|_{\infty,m} + \|\boldsymbol{u}\|_{\infty,m+1} \right) \|\boldsymbol{v}_h\|_{\Omega^n}$$

For the second term in (68), we use Lemma 3.4

$$\begin{aligned} &(\mathcal{A}_S^n + a_D^n)(\eta_u^n, \eta_p^n; \boldsymbol{v}_h, q_h) \\ &\quad \leq c \left(\|\nabla \eta_u^n\|_{\Omega^n} + h^{-1} \|\eta_u^n\|_{\Omega^n} + h^{-1/2} \|\eta_u^n\|_{\partial\Omega^n} + h^{1/2} \left(\|\eta_u^n\|_{\partial\Omega^n} + \|\eta_p^n\|_{\partial\Omega^n} \right) + \|\eta_p^n\|_{\Omega^n} \right) \\ &\qquad \cdot \left(\|\nabla \boldsymbol{v}_h\|_{\Omega^n} + h^{-1/2} \|\boldsymbol{v}_h\|_{\partial\Omega^n} + h^{1/2} \|\partial_n \boldsymbol{v}_h\|_{\partial\Omega^n} + h\|\nabla q_h\|_{\Omega^n} \right) \\ &\quad \leq ch^m \left(\|\boldsymbol{u}^n\|_{H^{m+1}(\Omega^n)} + \|p^n\|_{H^m(\Omega^n)} \right) \left(|||\boldsymbol{v}_h|||_{h,n} + h\|\nabla q_h\|_{\Omega^n} \right) \end{aligned} \tag{71}$$

Finally, we get for the ghost penalty part from (24) and the H^{m+1}-stability of the extension

$$g_h^n(\eta_u^n, \boldsymbol{v}_h) \leq ch^m \|\boldsymbol{u}^n\|_{H^{m+1}(\Omega_\delta^n)} g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h)^{1/2} \leq ch^m \|\boldsymbol{u}^n\|_{H^{m+1}(\Omega^n)} g_h^n(\boldsymbol{v}_h, \boldsymbol{v}_h)^{1/2}.$$

□

Now, we are ready to show an error estimate for the velocities.

Theorem 5.3 *Let $\boldsymbol{u}_h = (\boldsymbol{u}_h^k)_{k=1}^n$, $p_h = (p_h^k)_{k=1}^n$ be the discrete solution of (15) for $s = 1$ and $(\boldsymbol{u}, p)$ the continuous solution of (5). Further, let $\gamma_g \geq c_2(w_{\max})K$ with $c_2(w_{\max})$ defined in Lemma 4.1, γ_D, γ_p sufficiently large and $\Delta t \geq ch^2$ for some $c > 0$. Under the assumptions stated in Sect. 3, including Assumption 3.2, it holds for the error $\boldsymbol{e}_u^k = \boldsymbol{u}^k - \boldsymbol{u}_h^k$, $e_p^k = p^k - p_h^k$ for $n \geq 1$*

$$\begin{aligned}
&\|\boldsymbol{e}_u^n\|_{\Omega^n}^2 + \sum_{k=1}^n \left\{ \|\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}\|_{\Omega^k}^2 + \Delta t \left(|||\boldsymbol{e}_u^k|||_{h,k}^2 + h^2 \|\nabla e_p^k\|_{\Omega^k}^2 \right) \right\} \\
&\qquad \leq c \exp(c_1(w_{\max}) t_n) \Big(\Delta t^2 \|\partial_t^2 \boldsymbol{u}\|_Q^2 + h^{2m} \left(\|\boldsymbol{u}\|_{\infty,m+1}^2 + \|\partial_t \boldsymbol{u}\|_{\infty,m}^2 + \|p\|_{\infty,m}^2 \right) \Big),
\end{aligned}$$

where $\boldsymbol{e}_u^0 := 0$ and $c_1(w_{\max})$ is defined in Lemma 4.1.

Proof As in the stability proof (Theorem 4.2, (43)), we obtain from (67) for $n \geq 1$

$$\begin{aligned}
&\|\xi_{h,u}^n\|_{\Omega^n}^2 + \|\xi_{h,u}^n - \xi_{h,u}^{n-1}\|_{\Omega^n}^2 + \Delta t \left(|||\xi_{h,u}^n|||_{h,n}^2 + \gamma_p s_h^n(\xi_{h,p}^n, \xi_{h,p}^n) \right) \\
&\qquad \leq (1 + c_1(w_{\max})\Delta t) \|\xi_{h,u}^{n-1}\|_{\Omega^{n-1}}^2 + \frac{\Delta t}{2} \|\nabla \xi_{h,u}^{n-1}\|_{\Omega^{n-1}}^2 + \Delta t \gamma_g g_h^{n-1}(\xi_{h,u}^{n-1}, \xi_{h,u}^{n-1}) \\
&\qquad\quad + 2\Delta t \left(\left| \mathcal{E}_c^n(\xi_{h,u}^n, \xi_{h,p}^n) \right| + \left| \mathcal{E}_i^n(\xi_{h,u}^n, \xi_{h,p}^n) \right| \right).
\end{aligned} \tag{72}$$

for $\gamma_g \geq c_2(w_{\max})\mathcal{K}$. A bound for $\|\nabla \xi_{h,p}^n\|_{\Omega^n}$ can be obtained from (66) as in the proof of Lemma 4.4 (compare (47))

$$\begin{aligned}
\Delta t h^2 \|\nabla \xi_{h,p}^n\|_{\Omega^n}^2 \leq c \Delta t \Big(&h^2 \|D_t^{(1)} \xi_{h,u}^n\|_{\Omega^n}^2 + |||\xi_{h,u}^n|||_{h,n}^2 + s_h^n(\xi_{h,p}^n, \xi_{h,p}^n) \\
&+ \left| \mathcal{E}_c^n(\xi_{h,u}^n, \xi_{h,p}^n) \right| + \left| \mathcal{E}_i^n(\xi_{h,u}^n, \xi_{h,p}^n) \right| \Big).
\end{aligned} \tag{73}$$

We multiply (73) by $\epsilon > 0$ and add it to (72). Due to the assumption $\Delta t \geq ch^2$ the first three terms on the right-hand side of (73) can be absorbed into the left-hand side of (72) for sufficiently small ϵ

$$\begin{aligned}
&\|\xi_{h,u}^n\|_{\Omega^n}^2 + \frac{3}{4} \|\xi_{h,u}^n - \xi_{h,u}^{n-1}\|_{\Omega^n}^2 + \frac{3\Delta t}{4} \left(|||\xi_{h,u}^n|||_{h,n}^2 + \gamma_p s_h^n(\xi_{h,p}^n, \xi_{h,p}^n) + \epsilon h^2 \|\nabla \xi_{h,p}^n\|_{\Omega^n}^2 \right) \\
&\qquad \leq (1 + c_1(w_{\max})\Delta t) \|\xi_{h,u}^{n-1}\|_{\Omega^{n-1}}^2 + \frac{\Delta t}{2} \|\nabla \xi_{h,u}^{n-1}\|_{\Omega^{n-1}}^2 + \Delta t \gamma_g g_h^{n-1}(\xi_{h,u}^{n-1}, \xi_{h,u}^{n-1}) \\
&\qquad\quad + 2\Delta t \left(\left| \mathcal{E}_c^n(\xi_{h,u}^n, \xi_{h,p}^n) \right| + \left| \mathcal{E}_i^n(\xi_{h,u}^n, \xi_{h,p}^n) \right| \right).
\end{aligned} \tag{74}$$

Next, we use Lemmata 5.1 and 5.2 in combination with Young's inequality to estimate $\mathcal{E}_c^n$ and $\mathcal{E}_i^n$

$$
\begin{aligned}
&(1-\Delta t)\|\xi_{h,u}^n\|_{\Omega^n}^2 \\
&\quad + \frac{1}{2}\|\xi_{h,u}^n - \xi_{h,u}^{n-1}\|_{\Omega^n}^2 + \frac{\Delta t}{2}\left(|||\xi_{h,u}^n|||_{h,n} + \gamma_p s_h^n(\xi_{h,p}^n, \xi_{h,p}^n) + \epsilon h^2\|\nabla\xi_{h,p}^n\|_{\Omega^n}^2\right) \\
&\quad \leq (1 + c_1(w_{\max})\Delta t)\,\|\xi_{h,u}^{n-1}\|_{\Omega^{n-1}}^2 + \frac{\Delta t}{2}\|\nabla\xi_{h,u}^{n-1}\|_{\Omega^{n-1}}^2 + \Delta t\gamma_g g_h^{n-1}(\xi_{h,u}^{n-1}, \xi_{h,u}^{n-1}) \\
&\quad + c\Delta t\Big(\Delta t\|\partial_t^2\boldsymbol{u}\|_{Q^n}^2 + h^{2m}\left(\|\boldsymbol{u}\|_{\infty,m+1,I_n}^2 + \|\partial_t\boldsymbol{u}\|_{\infty,m,I_n}^2 + \|p\|_{H^m(\Omega^n)}^2\right)\Big).
\end{aligned}
\tag{75}
$$

We sum over $k = 1, \ldots, n$ and apply a discrete Gronwall lemma to find

$$
\begin{aligned}
\|\xi_{h,u}^n\|_{\Omega^n}^2 + \sum_{k=1}^n \left(\|\xi_{h,u}^k - \xi_{h,u}^{k-1}\|_{\Omega^k} + \Delta t\left(|||\xi_{h,u}^k|||_{h,k}^2 + \gamma_p s_h^k(\xi_{h,p}^k, \xi_{h,p}^k) + h^2\|\nabla\xi_{h,p}^k\|_{\Omega^k}^2\right)\right) \\
\leq c\exp(c_1(w_{\max})t_n)\Big(\Delta t^2\|\partial_t^2\boldsymbol{u}\|_Q^2 + h^{2m}\left(\|\boldsymbol{u}\|_{\infty,m+1}^2 + \|\partial_t\boldsymbol{u}\|_{\infty,m}^2 + \|p\|_{\infty,m}^2\right)\Big).
\end{aligned}
\tag{76}
$$

Finally, the interpolation estimates (62)–(64) and the argumentation used in (70) yield

$$
\begin{aligned}
\|\eta_u^n\|_{\Omega^n}^2 + \sum_{k=1}^n \left(\|\eta_u^k - \eta_u^{k-1}\|_{\Omega^k} + \Delta t\left(|||\eta_u^k|||_{h,k}^2 + h^2\|\nabla\eta_p^k\|_{\Omega^k}^2\right)\right) \\
\leq ch^{2m}\left(\|\boldsymbol{u}\|_{\infty,m+1}^2 + \|\partial_t\boldsymbol{u}\|_{\infty,m}^2 + \|p\|_{\infty,m}^2\right).
\end{aligned}
\tag{77}
$$

Addition of (76) and (77) proves the statement. □

Remark 5.4 *(Optimality)* The energy norm estimate is optimal under the inverse CFL condition $\Delta t \geq ch^2$. This condition is needed to control the pressure error $h\|\nabla\xi_{h,p}^n\|_{\Omega^n}$ using Lemma 4.4, see Corollary 4.5. If the Brezzi-Pitkäranta stabilisation would be used instead of the CIP pressure stabilisation, this term would be controlled by the pressure stabilisation in Theorem 4.2, as $h\|\nabla\xi_{h,p}^n\|_{\Omega^n} = s_h^n(\xi_{h,p}^n, \xi_{h,p}^n)^{1/2}$. Hence, an unconditional error estimate of first order in space would result.

Remark 5.5 *(BDF(2))* For $s = 2$ we obtain a similar result under the stronger condition $\Delta t \geq ch$ This is needed to get control over $h\|\nabla\xi_{h,p}^k\|_{\Omega^k}$, see Corollary 4.5 (60). Under this assumption, we can show the following result for $n \geq 2$:

$$
\begin{aligned}
&\|\boldsymbol{e}_u^n\|_{\Omega^n}^2 + \Delta t \sum_{k=1}^{n} \left(|||\boldsymbol{e}_u^k|||_{h,k}^2 + h^2 \|\nabla e_p^k\|_{\Omega^k}^2 \right) \\
&\quad \le c \exp(c_1(w_{\max}) t_n) \\
&\quad \times \left(\Delta t^4 \|\boldsymbol{u}\|_{H^3(I,L^2(\Omega^n))}^2 + h^{2m} \left(\|\boldsymbol{u}\|_{\infty,m+1}^2 + \|\partial_t \boldsymbol{u}\|_{\infty,m}^2 + \|p\|_{\infty,m}^2 \right) \right) \\
&\quad + c \left(\|\boldsymbol{e}_u^1\|_{\Omega^1}^2 + \Delta t |||\boldsymbol{e}_u^1|||_{h,1}^2 \right).
\end{aligned}
$$

which is of second order in time Δt, if we assume that the initial error is bounded by

$$
\begin{aligned}
&\|\boldsymbol{e}_u^1\|_{\Omega^1}^2 + \Delta t |||\boldsymbol{e}_u^1|||_{h,1}^2 \\
&\qquad \le c \left(\Delta t^4 \|\boldsymbol{u}\|_{H^3(I,L^2(\Omega^n))}^2 + h^{2m} \left(\|\boldsymbol{u}\|_{\infty,m+1,I_1}^2 + \|\partial_t \boldsymbol{u}\|_{\infty,m,I_1}^2 + \|p\|_{H^m(\Omega^1)}^2 \right) \right).
\end{aligned} \tag{78}
$$

The initialisation will be discussed in the following remark. The main modifications in the proof concern the approximation of the time derivative in Lemmas 5.1 and 5.2. In (69) we estimate

$$
\left(D_t^{(2)} \boldsymbol{u}^n - \partial_t \boldsymbol{u}(t_n), \boldsymbol{v}_h \right)_{\Omega^n} \le c \Delta t^{3/2} \|\partial_t^3 \boldsymbol{u}\|_{Q^n} \|\boldsymbol{v}_h\|_{\Omega^n},
$$

see [12,34]. In order to estimate the analogue of (70), we use

$$
D_t^{(2)} \boldsymbol{u}_h^n = \frac{3}{2} D_t^{(1)} \boldsymbol{u}_h^n + \frac{1}{2} D_t^{(1)} \boldsymbol{u}_h^{n-1}. \tag{79}
$$

Then the argumentation used in (70) can be applied to both terms on the right-hand side of (79).

Remark 5.6 *(Initialisation of BDF(2))* To initialise the BDF(2) scheme, the function $\boldsymbol{u}_h^1$ needs to be computed with sufficient accuracy. The simplest possibility is to use one BDF(1) step by solving

$$
\begin{aligned}
\frac{1}{\Delta t} \left(\boldsymbol{u}_h^1, \boldsymbol{v}_h^1 \right)_{\Omega^k} + \mathcal{A}_h^1(\boldsymbol{u}_h^1, p_h^1; \boldsymbol{v}_h^1, q_h^1) &= \frac{1}{\Delta t} \left(E\boldsymbol{u}^0, \boldsymbol{v}_h^1 \right)_{\Omega^1} \\
&\quad + (\boldsymbol{f}, \boldsymbol{v}_h^1)_{\Omega^1} \quad \forall \boldsymbol{v}_h^1 \in \mathcal{V}_h^1, q_h^1 \in \mathcal{L}_h^1
\end{aligned}
$$

for $(\boldsymbol{u}_h^1, p_h^1) \in (\mathcal{V}_h^1 \times \mathcal{L}_{h,0}^1)$. Similar to the proof of Theorem 5.3, the error after one BDF(1) step can be estimated by

$$
\begin{aligned}
&\|\boldsymbol{e}_u^1\|_{\Omega^1}^2 + \Delta t |||\boldsymbol{e}_u^1|||_{h,1}^2 \\
&\quad \le c \left(\Delta t^3 \|\partial_t^2 \boldsymbol{u}\|_{Q^1}^2 + \Delta t\, h^{2m} \left(\|\boldsymbol{u}\|_{\infty,m+1,I_1}^2 + \|\partial_t \boldsymbol{u}\|_{\infty,m,I_1}^2 + \|p\|_{H^m(\Omega^1)}^2 \right) \right) \\
&\quad \le c \left(\Delta t^4 \|\boldsymbol{u}\|_{H^3(I,L^2(\Omega^n))}^2 + \Delta t\, h^{2m} \left(\|\boldsymbol{u}\|_{\infty,m+1,I_1}^2 + \|\partial_t \boldsymbol{u}\|_{\infty,m,I_1}^2 + \|p\|_{H^m(\Omega^1)}^2 \right) \right),
\end{aligned}
$$

where in the last step a Sobolev inequality has been applied in time to show $\|\partial_t^2 \boldsymbol{u}\|_{Q^1}^2 \leq \Delta t \|\partial_t^2 \boldsymbol{u}\|_{\infty,0,I_1}^2 \leq c\Delta t \|\boldsymbol{u}\|_{H^3(I,L^2(\Omega^n))}^2$.

5.1.1 $L^2(L^2)$-norm error of pressure

The energy estimate in Theorem 5.3 includes an optimal bound for the H^1-norm of the pressure. To show an optimal bound in the L^2-norm seems to be non-trivial, due to the fact that $\boldsymbol{u}_h^{n-1}$ is not discrete divergence-free with respect to Ω^n and $\mathcal{V}_h^n$, see the discussion in Sect. 4.1. We show here only a sub-optimal bound for $s = 1$. An optimal estimate is subject to future work.

Lemma 5.7 *Under the assumptions of Theorem 5.3 it holds for* $s = 1$

$$\left(\Delta t \sum_{k=1}^{n} \|e_p^k\|_{\Omega^k}^2\right)^{1/2} \leq c \exp(c_1(w_{\max})t_n) \left(\Delta t^{1/2} \|\partial_t^2 \boldsymbol{u}\|_Q + \frac{h^m}{\Delta t^{1/2}} \left(\|\boldsymbol{u}\|_{\infty,m+1} + \|\partial_t \boldsymbol{u}\|_{\infty,m} + \|p\|_{\infty,m}\right)\right),$$

where $\boldsymbol{e}_u^0 := 0$.

Proof We use the modified inf-sup condition for the discrete part $\xi_{h,p}^n = i_h^n p^n - p_h^n$ and standard interpolation estimates

$$\begin{aligned}
\beta \|\xi_{h,p}^n\|_{\Omega^n} &\leq \sup_{\boldsymbol{v}_h^n \in \mathcal{V}_h^n} \frac{(\operatorname{div} \boldsymbol{v}_h^n, \xi_{h,p}^n)_{\Omega^n} - (\boldsymbol{v}_h^n \cdot n, \xi_{h,p}^n)_{\partial\Omega^n}}{|||\boldsymbol{v}_h^n|||_{h,n}} + h\|\nabla \xi_{h,p}^n\|_{\Omega^n} \\
&\leq \sup_{\boldsymbol{v}_h^n \in \mathcal{V}_h^n} \frac{(\operatorname{div} \boldsymbol{v}_h^n, e_p^n)_{\Omega^n} - (\boldsymbol{v}_h^n \cdot n, e_p^n)_{\partial\Omega^n}}{|||\boldsymbol{v}_h^n|||_{h,n}} \\
&\quad + \sup_{\boldsymbol{v}_h^n \in \mathcal{V}_h^n} \frac{(\operatorname{div} \boldsymbol{v}_h^n, \eta_p^n)_{\Omega^n} - (\boldsymbol{v}_h^n \cdot n, \eta_p^n)_{\partial\Omega^n}}{|||\boldsymbol{v}_h^n|||_{h,n}} \\
&\quad + h\left(\|\nabla e_p^n\|_{\Omega^n} + \|\nabla \eta_p^n\|_{\Omega^n}\right) \\
&\leq \sup_{\boldsymbol{v}_h^n \in \mathcal{V}_h^n} \frac{(\operatorname{div} \boldsymbol{v}_h^n, e_p^n)_{\Omega^n} - (\boldsymbol{v}_h^n \cdot n, e_p^n)_{\partial\Omega^n}}{|||\boldsymbol{v}_h^n|||_{h,n}} \\
&\quad + h\|\nabla e_p^n\|_{\Omega^n} + ch^m \|p^n\|_{H^m(\Omega^n)}. \qquad (80)
\end{aligned}$$

The second term on the right-hand side is bounded by the energy estimate. For the first term, we use Galerkin orthogonality (66), followed by Cauchy–Schwarz and Poincaré inequalities

$$
\begin{aligned}
&(\operatorname{div} \boldsymbol{v}_h^n, e_p^n)_{\Omega^n} - (\boldsymbol{v}_h^n \cdot \boldsymbol{n}, e_p^n)_{\partial\Omega^n} \\
&\quad = -(D_t^{(1)} \boldsymbol{e}_u^n, \boldsymbol{v}_h)_{\Omega^n} - (\mathcal{A}_S^n + a_D^n)(\boldsymbol{e}_u^n, 0; \boldsymbol{v}_h, 0) - \gamma_g g_h^n(\boldsymbol{e}_u^n, \boldsymbol{v}_h) \\
&\qquad + \gamma_g g_h^n(\boldsymbol{u}^n, \boldsymbol{v}_h) + (D_t^{(1)} \boldsymbol{u}^n - \partial_t \boldsymbol{u}(t_n), \boldsymbol{v}_h)_{\Omega^n} \\
&\quad \leq c \left\{ \|D_t^{(1)} \boldsymbol{e}_u^n\|_{\Omega^n} + |||\boldsymbol{e}_u^n|||_{h,n} + h^m \|\boldsymbol{u}^n\|_{H^{m+1}(\Omega^n)} + \Delta t \|\partial_t^2 \boldsymbol{u}^n\|_{\Omega^n} \right\} |||\boldsymbol{v}_h|||_{h,n}.
\end{aligned}
$$

After summation in (80), we obtain

$$
\begin{aligned}
\Delta t \sum_{k=1}^{n} \|\xi_{h,p}^k\|_{\Omega^k}^2 \leq c \sum_{k=1}^{n} \Big\{ &\frac{1}{\Delta t} \|\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}\|_{\Omega^k}^2 + \Delta t \left(|||\boldsymbol{e}_u^k|||_{h,k}^2 + h^2 \|\nabla e_p^k\|_{\Omega^k}^2 \right) \\
&+ \Delta t h^{2m} \left(\|\boldsymbol{u}^k\|_{H^{m+1}(\Omega^k)}^2 + \|p^k\|_{H^m(\Omega^k)}^2 \right) + \Delta t^3 \|\partial_t^2 \boldsymbol{u}^k\|_{\Omega^k}^2 \Big\}.
\end{aligned} \tag{81}
$$

Using the standard interpolation estimate $\|\eta_p^k\|_{\Omega^k}^2 \leq c h^{2m} \|p^k\|_{H^m(\Omega^k)}^2$, we see that (81) holds for $\xi_{h,p}^k$ replaced by e_p^k. Finally, Theorem 5.3 yields the statement. Unfortunately, the factor $\frac{1}{\Delta t}$ in front of the first term on the right-hand side of (81) leads to a loss of $\Delta t^{-1/2}$ in the final estimate. □

Remark 5.8 *(BDF(2))* For $s = 2$ we can only control $\Delta t^3 \|D_t^{(2)} \boldsymbol{e}_u^n\|_{\Omega^n}^2 = \frac{\Delta t}{2} \|3\boldsymbol{e}_u^n - 4\boldsymbol{e}_u^{n-1} + \boldsymbol{e}_u^{n-2}\|_{\Omega^n}^2$ (compared to $\Delta t^2 \|D_t^{(1)} \boldsymbol{e}_u^n\|_{\Omega^n}^2$ for $s = 1$), which leads to a further loss of Δt^{-1} in the above estimate:

$$
\begin{aligned}
\left(\Delta t \sum_{k=1}^{n} \|e_p^k\|_{\Omega^k}^2 \right)^{1/2} &\leq c \exp(c_1 (w_{\max}) t_n) \Big(\Delta t \|\partial_t^2 \boldsymbol{u}\|_Q \\
&+ \frac{h^m}{\Delta t} \left(\|\boldsymbol{u}\|_{\infty,m+1} + \|\partial_t \boldsymbol{u}\|_{\infty,m} + \|p\|_{\infty,m} \right) \Big).
\end{aligned}
$$

Remark 5.9 The estimate in Lemma 5.7 is balanced, if we choose $\Delta t \sim h^m$, which yields a convergence order of $\mathcal{O}(\Delta t^{1/2}) = \mathcal{O}(h^{m/2})$. This means that the convergence order is reduced by $\mathcal{O}(h^{m/2})$ compared to the situation on a fixed domain $\Omega(t) = \Omega$. For BDF(2) the estimate is balanced for $\Delta t^2 \sim h^m$ and we obtain a convergence order of $\mathcal{O}(\Delta t) = \mathcal{O}(h^{m/2})$. The inverse CFL conditions in Theorem 5.3 and Remark 5.5 are automatically fulfilled for these choices, if $m \geq 2$ or $m = s = 1$.

5.2 $L^2(L^2)$-norm error of velocity

To obtain an optimal bound for the velocity error in the L^2-norm, we introduce a dual problem. The argumentation of Burman and Fernández [12], that does not require a dual problem, but is based on a Stokes projection $P_h(\boldsymbol{u}, p)$ of the continuous solution, can not be transferred in a straight-forward way to the case of moving domains, as it requires an estimate for the time derivative $\partial_t(u - P_h^u u)$. Time derivative and Stokes projection do, however, not commute in the case of moving domains, as $P_h^u \boldsymbol{u}(t)$

depends on the domain $\Omega(t)$. For this reason an estimate for the time derivative is non-trivial.

We focus again on the case $s = 1$ first and remark on how to transfer the argumentation to the case $s > 1$ afterwards. The argumentation will be based on a semi-discretised (in time) dual problem. Before we introduce the dual problem, let us note that the semi-discretised primal problem is given by: *Find* $(\boldsymbol{u}^k, p^k)_{k=1}^n$ *with* $\boldsymbol{u}^k \in H_0^1(\Omega^k)$, $p^k \in L_0^2(\Omega^k)$ *such that*

$$\sum_{k=1}^{n} \left\{ (\boldsymbol{u}^k - E^{k-1}\boldsymbol{u}^{k-1}, \psi_u^k)_{\Omega^k} + \Delta t \mathcal{A}_S^k(\boldsymbol{u}^k, p^k; \psi_u^k, \psi_p^k) \right\} + (E^0\boldsymbol{u}^0, \psi_u^1)_{\Omega^1}$$
$$= \left(E^0\boldsymbol{u}^0, \psi_u^1\right)_{\Omega^1} + \Delta t \sum_{k=1}^{n} \left(\boldsymbol{f}, \psi_u^k\right)_{\Omega^k} \quad \forall \psi_u^k \in H_0^1(\Omega^k)^d,$$
$$\psi_p^k \in L^2(\Omega^k), \ k = 1, \ldots, n, \tag{82}$$

where E^k denotes the smooth extension operator to Ω_δ^k introduced in Sect. 3.3.

The corresponding semi-discretised dual problem, which will be needed in the following, reads: *Find* $(\boldsymbol{z}_u^k, z_p^k)_{k=1}^n$ *with* $\boldsymbol{z}_u^k \in H_0^1(\Omega^k)$, $z_p^k \in L_0^2(\Omega^k)$ *such that*

$$\Delta t \sum_{k=1}^{n} \left(\boldsymbol{e}_u^k, \phi_u^k\right)_{\Omega^k} = \sum_{k=1}^{n} \left\{ (\phi_u^k - E^{k-1}\phi_u^{k-1}, \boldsymbol{z}_u^k)_{\Omega^k} + \Delta t \mathcal{A}_S^k(\phi_u^k, \phi_p^k; \boldsymbol{z}_u^k, z_p^k) \right\}$$
$$+ (E^0\phi_u^0, \boldsymbol{z}_u^1)_{\Omega^1} \quad \forall \phi_u^k \in H_0^1(\Omega^k)^d, \ \phi_p^k \in L_0^2(\Omega^k), \quad k = 1, \ldots, n. \tag{83}$$

Note that the Dirichlet conditions are imposed strongly in this formulation and the bilinear form $\mathcal{A}_S^k$ does not include the Nitsche terms.

We start by showing the well-posedness of the problem (83).

Lemma 5.10 *Let* $s = 1$, $\boldsymbol{e}_u^k \in L^2(\Omega^k)$ *for* $k = 1, \ldots, n$ *and assume Assumption 3.2. The semi-discrete dual problem (83) defines unique solutions* $(\boldsymbol{z}_u^k, z_p^k)_{k=1}^n$ *with regularity* $\boldsymbol{z}_u^k \in H^2(\Omega^k)$, $z_p^k \in H^1(\Omega^k)$. *Moreover, the following regularity estimates are valid, where* $S_\delta^k := \Omega_\delta^k \setminus \Omega^k$ *and* $D_t^{(1)} z_u^{k+1} = \frac{1}{\Delta t}(E^k z_u^{k+1} - z_u^k)$

$$\|\boldsymbol{z}_u^k\|_{H^2(\Omega^k)} + \|z_p^k\|_{H^1(\Omega^k)} \leq c \left(\|D_t^{(1)} z_u^{k+1}\|_{\Omega^k} + \frac{1}{\Delta t} \|E^k \boldsymbol{z}_u^{k+1}\|_{S_\delta^k} + \|\boldsymbol{e}_u^k\|_{\Omega^k} \right) \quad \text{for } k < n, \tag{84}$$

$$\|\boldsymbol{z}_u^n\|_{H^2(\Omega^n)} + \|z_p^n\|_{H^1(\Omega^n)} \leq c \left(\frac{1}{\Delta t} \|\boldsymbol{z}_u^n\|_{\Omega^n} + \|\boldsymbol{e}_u^n\|_{\Omega^n} \right). \tag{85}$$

Proof By testing (83) with $\tilde{\phi}_u^l = \delta_{kl}\, \phi_u^k$, $\tilde{\phi}_p^l = \delta_{kl} \phi_p^k$, $l = 1, \ldots, n$, where δ_{kl} is the Kronecker delta, we observe that the system splits into n separate time steps, where each step corresponds to a stationary Stokes system with an additional L^2-term coming

from the discretisation of the time derivative. For $k < n$ we have

$$\frac{1}{\Delta t}(\phi_u^k, z_u^k)_{\Omega^k} + \mathcal{A}_S^k(\phi_u^k, \phi_p^k; z_u^k, z_p^k) = \frac{1}{\Delta t}(E^k\phi_u^k, z_u^{k+1})_{\Omega^{k+1}} + \left(\boldsymbol{e}_u^k, \phi_u^k\right)_{\Omega^k}$$
$$\forall \phi_u^k \in H_0^1(\Omega^k)^d, \phi_p^k \in L_0^2(\Omega^k), \tag{86}$$

and for $k = n$

$$\frac{1}{\Delta t}\left(\phi_u^n, z_u^n\right)_{\Omega^n} + \mathcal{A}_S^n(\phi_u^n, \phi_p^n; z_u^n, z_p^n) = \left(\boldsymbol{e}_u^n, \phi_u^n\right)_{\Omega^n} \quad \forall \phi_u^n \in H_0^1(\Omega^n)^d, \phi_p^k \in L_0^2(\Omega^n). \tag{87}$$

As the corresponding reduced problems are coercive in the velocity space $\mathcal{V}_0(t_k)$ (cf. Sect. 2.1), existence and uniqueness of solutions $\boldsymbol{z}_u^k \in H_0^1(\Omega^k)$, $z_p^k \in L_0^2(\Omega^k)$ follow inductively by standard arguments for $k = n, \ldots, 1$, see e.g. Temam [58], Section I.2.

To show the regularity estimates (84) and (85), let us re-formulate the problems (86) and (87) in the following way: For $k < n$ we have

$$\mathcal{A}_S^k(\phi_u^k, \phi_p^k; z_u^k, z_p^k) = \underbrace{\frac{1}{\Delta t}\left((E^k\phi_u^k, z_u^{k+1})_{\Omega^{k+1}} - (\phi_u^k, z_u^k)_{\Omega^k}\right) + \left(\boldsymbol{e}_u^k, \phi_u^k\right)_{\Omega^k}}_{=: F_k(\phi_u^k)} \tag{88}$$
$$\forall \phi_u^k \in H_0^1(\Omega^k)^d, \phi_p^k \in L_0^2(\Omega^k),$$

and for $k = n$

$$\mathcal{A}_S^n(\phi_u^n, \phi_p^n; z_u^n, z_p^n) = \underbrace{\left(\boldsymbol{e}_u^n, \phi_u^n\right)_{\Omega^n} - \frac{1}{\Delta t}\left(\phi_u^n, z_u^n\right)_{\Omega^n}}_{=: F_n(\phi_u^n)} \quad \forall \phi_u^n \in H_0^1(\Omega^n)^d, \phi_p^k \in L_0^2(\Omega^n).$$

If we can prove that F_k lies in the dual space $[L^2(\Omega^k)^d]^*$, Proposition I.2.2 in Temam's book [58] guarantees the regularity estimate

$$\|\boldsymbol{z}_u^k\|_{H^2(\Omega^k)} + \|z_p^k\|_{H^1(\Omega^k)} \leq c \sup_{\phi_u^k \in L^2(\Omega^k)} \frac{F_k(\phi_u^k)}{\|\phi_u^k\|_{\Omega^k}}. \tag{89}$$

We need to show that the right-hand side is bounded. Splitting the first integral on the right-hand side into an integral over Ω^k and S_δ^k, we have for $k < n$

$$F_k(\phi_u^k) \leq \|\phi_u^k\|_{\Omega^k}\|D_t^{(1)} z_u^{k+1}\|_{\Omega^k} + \frac{1}{\Delta t}\|E^k\phi_u^k\|_{S_\delta^k}\|E^k\boldsymbol{z}_u^{k+1}\|_{S_\delta^k} + \|\boldsymbol{e}_u^k\|_{\Omega^k}\|\phi_u^k\|_{\Omega^k}.$$

and thus,

$$F_k(\phi_u^k) \leq c\left(\|D_t^{(1)} z_u^{k+1}\|_{\Omega^k} + \frac{1}{\Delta t}\|E^k\boldsymbol{z}_u^{k+1}\|_{S_\delta^k} + \|\boldsymbol{e}_u^k\|_{\Omega^k}\right)\|\phi_u^k\|_{\Omega^k}. \tag{90}$$

For $k = n$, we obtain

$$F_n(\phi_u^n) \leq c \left(\frac{1}{\Delta t} \|z_u^n\|_{\Omega^n} + \|\boldsymbol{e}_u^n\|_{\Omega^n} \right) \|\phi_u^n\|_{\Omega^n}. \tag{91}$$

The boundedness of F_k follows by induction for $k = n, \ldots, 1$ and by using the stability of the extension operator E^k. Combination of (89) and (90), resp. (91), yield the regularity estimates (86) and (87). □

Next, we derive a stability estimate for the semi-discretised dual problem (83). We remark that a stability estimate for the continuous dual problem, including the first time derivative $\partial_t z$, could be obtained as well. This is however not enough to bound the consistency error of the time derivative in a sufficient way for an optimal L^2-norm error estimate.

Lemma 5.11 *Let the assumptions made in Sect. 3 be valid, including Assumption 3.2. For sufficiently small $\Delta t < \xi$, where ξ depends only on c_δ, $w_{\max}$ and the domains Ω^k, $k = 1, \ldots, n$, the solution $(z_u^k, z_p^k)_{k=1}^n$ to the semi-discretised dual problem (83) for $s = 1$ fulfils the stability estimate*

$$\|\nabla z_u^1\|_{\Omega^1}^2 + \frac{1}{\Delta t} \|\nabla z_u^n\|_{\Omega^n}^2 + \sum_{k=1}^{n-1} \left\{ \|\nabla (z_u^k - E^k z_u^{k+1})\|_{\Omega^k}^2 + \|D_t^{(1)} z_u^{k+1}\|_{\Omega^k}^2 \right\}$$
$$+ \Delta t \sum_{k=1}^{n} \left\{ \|z_u^k\|_{H^2(\Omega^k)}^2 + \|z_p^k\|_{H^1(\Omega^k)}^2 \right\} \leq c w_{\max}^2 \Delta t \sum_{k=1}^{n} \|\boldsymbol{e}_u^k\|_{\Omega^k}^2.$$

Proof We show a stability estimate for the first derivatives ∇z_u^k first. For better readability we will in the following skip the extension operators E^k and denote the extension $E^k z_u^{k+1}$ still by z_u^{k+1} and similarly for other variables.

Diagonal testing in (83) with $\phi_u^k = z_u^k$, $\phi_p^k = z_p^k$ results in

$$\sum_{k=1}^{n} \left\{ (z_u^k - z_u^{k-1}, z_u^k)_{\Omega^k} + \Delta t \|\nabla z_u^k\|_{\Omega^k}^2 \right\} + (z_u^0, z_u^1)_{\Omega^1} = \Delta t \sum_{k=1}^{n} \left(\boldsymbol{e}_u^k, z_u^k \right)_{\Omega^k},$$

or equivalently

$$\sum_{k=1}^{n-1} \left\{ \|z_u^k\|_{\Omega^k}^2 - (z_u^k, z_u^{k+1})_{\Omega^{k+1}} + \Delta t \|\nabla z_u^k\|_{\Omega^k}^2 \right\} + \|z_u^n\|_{\Omega^n}^2 + \Delta t \|\nabla z_u^n\|_{\Omega^n}^2$$
$$= \Delta t \sum_{k=1}^{n} \left(\boldsymbol{e}_u^k, z_u^k \right)_{\Omega^k}. \tag{92}$$

As z_u^k vanishes on $\partial\Omega^k$, a Poincaré-like estimate gives in combination with (9) and the stability of the extension operator

$$\|z_u^k\|_{S_\delta^k} \leq c_p \left(\delta^{1/2}\|z_u^k\|_{\partial\Omega^k} + \delta\|\nabla z_u^k\|_{S_\delta^k}\right) \leq c_p c_\delta w_{\max}\Delta t\|\nabla z_u^k\|_{\Omega^k}, \tag{93}$$

where c_p denotes a constant depending on the domain Ω^k and $c_\delta > 1$ is the constant in (9). Using Young's inequality, this implies for $\Delta t \leq (2c_p^2 c_\delta^2 w_{\max}^2)^{-1}$

$$\begin{aligned}\|z_u^k\|_{S_\delta^k}\|z_u^{k+1}\|_{S_\delta^k} &\leq \frac{1}{2}c_p^2 c_\delta^2 w_{\max}^2\Delta t^2\left(\|\nabla z_u^k\|_{\Omega^k}^2 + \|\nabla z_u^{k+1}\|_{\Omega^{k+1}}^2\right)\\ &\leq \frac{\Delta t}{4}\left(\|\nabla z_u^k\|_{\Omega^k}^2 + \|\nabla z_u^{k+1}\|_{\Omega^{k+1}}^2\right).\end{aligned}$$

We obtain

$$\begin{aligned}\|z_u^k\|_{\Omega^k}^2 - (z_u^k, z_u^{k+1})_{\Omega^{k+1}} &\geq (z_u^k, z_u^k - z_u^{k+1})_{\Omega^k} - \|z_u^k\|_{S_\delta^k}\|z_u^{k+1}\|_{S_\delta^k}\\ &\geq \frac{1}{2}\left(\|z_u^k\|_{\Omega^k}^2 + \|z_u^k - z_u^{k+1}\|_{\Omega^k}^2 - \|z_u^{k+1}\|_{\Omega^{k+1}}^2\right)\\ &\quad - \frac{\Delta t}{4}\left(\|\nabla z_u^k\|_{\Omega^k}^2 + \|\nabla z_u^{k+1}\|_{\Omega^{k+1}}^2\right).\end{aligned} \tag{94}$$

For the right-hand side in (92), we apply the Cauchy-Schwarz, a Poincaré and Young's inequality to get

$$\Delta t\sum_{k=1}^n \left(\boldsymbol{e}_u^k, z_u^k\right)_{\Omega^k} \leq \sum_{k=1}^n \frac{\Delta t}{4}\|\nabla z_u^k\|_{\Omega^k}^2 + c\Delta t\|\boldsymbol{e}_u^k\|_{\Omega^k}^2. \tag{95}$$

Using (94) and (95), (92) writes

$$\|z_u^1\|_{\Omega^1}^2 + \|z_u^n\|_{\Omega^n}^2 + \sum_{k=1}^{n-1}\|z_u^k - z_u^{k+1}\|_{\Omega^k}^2 + \sum_{k=1}^n \Delta t\|\nabla z_u^k\|_{\Omega^k}^2 \leq c\Delta t\sum_{k=1}^n \|\boldsymbol{e}_u^k\|_{\Omega^k}^2. \tag{96}$$

Next, we use the regularity estimates in Lemma 5.10 to get a bound for the second derivatives of z_u^k. For $k = n$ we have

$$\|z_u^n\|_{H^2(\Omega^n)} + \|z_p^n\|_{H^1(\Omega^n)} \leq c\left(\frac{1}{\Delta t}\|z_u^n\|_{\Omega^n} + \|\boldsymbol{e}_u^n\|_{\Omega^n}\right).$$

For $k < n$ Lemma 5.10 gives us

$$\|z_u^k\|_{H^2(\Omega^k)} + \|z_p^k\|_{H^1(\Omega^k)} \leq c\left(\|D_t^{(1)} z_u^{k+1}\|_{\Omega^k} + \frac{1}{\Delta t}\|z_u^{k+1}\|_{S_\delta^k} + \|\boldsymbol{e}_u^k\|_{\Omega^k}\right).$$

We estimate the term on S_δ^k by using a Poincaré-type inequality with a domain-dependent constant $c_p > 0$ as in (93), followed by (34) for $\epsilon = 1$ and the stability of the extension

$$c\|z_u^{k+1}\|_{S_\delta^k} \leq cc_p\delta\|\nabla z_u^{k+1}\|_{S_\delta^k} \leq cc_p\delta^{3/2}\|z_u^{k+1}\|_{H^2(\Omega_\delta^k)} \leq cc_p\delta^{3/2}\|z_u^{k+1}\|_{H^2(\Omega^{k+1})}.$$

Using (9) we get for $\Delta t < (4c^2c_p^2c_\delta^3 w_{\max}^3)^{-1}$

$$c\|z_u^{k+1}\|_{S_\delta^k} \leq cc_p(c_\delta w_{\max}\Delta t)^{3/2}\|z_u^{k+1}\|_{H^2(\Omega^{k+1})} \leq \frac{\Delta t}{2}\|z_u^{k+1}\|_{H^2(\Omega^{k+1})}, \tag{97}$$

and hence

$$\|z_u^k\|_{H^2(\Omega^k)} + \|z_p^k\|_{H^1(\Omega^k)} \leq c\left(\|D_t^{(1)}z_u^{k+1}\|_{\Omega^k} + \|e_u^k\|_{\Omega^k}\right) + \frac{1}{2}\|z_u^{k+1}\|_{H^2(\Omega^{k+1})}.$$

Summation over $k = 1, \ldots, n$ results in

$$\begin{aligned}\Delta t\sum_{k=1}^{n}\left\{\|z_u^k\|_{H^2(\Omega^k)}^2 + \|z_p^k\|_{H^1(\Omega^k)}^2\right\}\\ \leq c_0\left(\frac{1}{\Delta t}\|z_u^n\|_{\Omega^n}^2 + \sum_{k=1}^{n-1}\|D_t^{(1)}z_u^{k+1}\|_{\Omega^k}^2\right) + c\Delta t\sum_{k=1}^{n}\|e_u^k\|_{\Omega^k},\end{aligned} \tag{98}$$

where c_0 denotes a constant. It remains to derive a bound for the discrete time derivative on the right-hand side. Therefore, note that for $k < n$ we can write (88) equivalently by using the density of $H^1(\Omega^k)$ in $L^2(\Omega^k)$ as

$$\begin{aligned}-(\Delta z_u^k,\phi_u^k)_{\Omega^k} + (\nabla z_p^k, \phi_u^k)_{\Omega^k} - (z_u^k, \nabla\phi_p^k)_{\Omega^k}\\ = \frac{1}{\Delta t}\left((E^{k+1}\phi_u^k, z_u^{k+1})_{\Omega^{k+1}} - (\phi_u^k, z_u^k)_{\Omega^k}\right)\\ + \left(e_u^k, \phi_u^k\right)_{\Omega^k} \quad \forall\phi_u^k \in L^2(\Omega^k)^d, \phi_p^k \in H^1(\Omega^k).\end{aligned} \tag{99}$$

For $k = n$ we have

$$\begin{aligned}-(\Delta z_u^n, \phi_u^n)_{\Omega^n} + (\nabla z_p^n, \phi_u^n)_{\Omega^n} - (z_u^n, \nabla\phi_p^n)_{\Omega^n}\\ = \left(e_u^n, \phi_u^n\right)_{\Omega^n} - \frac{1}{\Delta t}\left(\phi_u^n, z_u^n\right)_{\Omega^n} \quad \forall\phi_u^n \in L^2(\Omega^n)^d, \phi_p^n \in H^1(\Omega^n).\end{aligned} \tag{100}$$

For $k < n$ we test (99) with $\phi_u^k = z_u^k - z_u^{k+1}$, $\phi_p^k = 0$

$$\begin{aligned}-(\Delta z_u^k, z_u^k - z_u^{k+1})_{\Omega^k} + (\nabla z_p^k, z_u^k - z_u^{k+1})_{\Omega^k} + \frac{1}{\Delta t}\left(\|z_u^k - z_u^{k+1}\|_{\Omega^{k+1}}^2\right.\\ \left. + (z_u^k - z_u^{k+1}, z_u^{k+1})_{\Omega^k\setminus\Omega^{k+1}} - (z_u^k - z_u^{k+1}, z_u^{k+1})_{\Omega^{k+1}\setminus\Omega^k}\right) = \left(e_u^k, z_u^k - z_u^{k+1}\right)_{\Omega^k}.\end{aligned} \tag{101}$$

Using integration by parts and the fact that $z_u^k|_{\partial\Omega^k} = 0$, the first term in (101) writes

$$-(\Delta z_u^k, z_u^k - z_u^{k+1})_{\Omega^k} = \left(\nabla z_u^k,\ \nabla(z_u^k - z_u^{k+1})\right)_{\Omega^k} + \left(\partial_n z_u^k, z_u^{k+1}\right)_{\partial\Omega^k}.$$

For the second term in (101) we note that $(\nabla z_p^k, z_u^l)_{\Omega^l} = 0$ for $l = k, k+1$

$$\begin{aligned}(\nabla z_p^k, z_u^k - z_u^{k+1})_{\Omega^k} = &\underbrace{(\nabla z_p^k, z_u^k)_{\Omega^k}}_{=0} - \underbrace{(\nabla z_p^k, z_u^{k+1})_{\Omega^{k+1}}}_{=0} \\ &+(\nabla z_p^k, z_u^{k+1})_{\Omega^{k+1}\setminus\Omega^k} - (\nabla z_p^k, z_u^{k+1})_{\Omega^k\setminus\Omega^{k+1}}.\end{aligned}$$

Setting $B_\delta^k := (\Omega^{k+1}\setminus\Omega^k) \cup (\Omega^k\setminus\Omega^{k+1})$ we obtain further

$$(\nabla z_p^k, z_u^k - z_u^{k+1})_{\Omega^k} \geq -\|\nabla z_p^k\|_{B_\delta^k}\|z_u^{k+1}\|_{B_\delta^k}.$$

Using the Cauchy–Schwarz and Young's inequality, we obtain from (101)

$$\begin{aligned}\frac{1}{2\Delta t}\left(\left\|z_u^k - z_u^{k+1}\right\|_{\Omega^k}^2 - \left\|z_u^k - z_u^{k+1}\right\|_{B_\delta^k}\left\|z_u^{k+1}\right\|_{B_\delta^k}\right) + \left(\nabla z_u^k, \nabla(z_u^k - z_u^{k+1})\right)_{\Omega^k} \\ - (\partial_n z_u^k, z_u^{k+1})_{\partial\Omega^k} - \|\nabla z_p^k\|_{B_\delta^k}\|z_u^{k+1}\|_{B_\delta^k} \leq c\Delta t\|\boldsymbol{e}_u^k\|_{\Omega^k}^2.\end{aligned} \tag{102}$$

To estimate the second term on the left-hand side, we apply the triangle inequality and Young's inequality to get

$$\frac{1}{\Delta t}\left\|z_u^k - z_u^{k+1}\right\|_{B_\delta^k}\left\|z_u^{k+1}\right\|_{B_\delta^k} \leq \frac{c}{\Delta t}\left(\left\|z_u^k\right\|_{B_\delta^k}^2 + \left\|z_u^{k+1}\right\|_{B_\delta^k}^2\right) \tag{103}$$

Next, we note that, due to (8) for $n = k$ and $n = k+1$, the maximum width of the strip B_δ^k is of size $\mathcal{O}(\delta)$ as for S_δ^k. Thus, we can use a Poincaré-type estimate with a constant $c_p > 0$ as in (93). In combination with the fact that $z_u^l = 0$ on $\partial\Omega^l$ for $l = k, k+1$, we obtain

$$\frac{c}{\Delta t}\left\|z_u^l\right\|_{B_\delta^k}^2 \leq cc_p^2\frac{\delta}{\Delta t}\left(\left\|z_u^l\right\|_{\partial\Omega}^2 + \delta\left\|\nabla z_u^l\right\|_{B_\delta^k}^2\right) = cc_p^2\frac{\delta^2}{\Delta t}\left\|\nabla z_u^l\right\|_{B_\delta^k}^2 \tag{104}$$

Using (34) followed by (9) and the stability of the extensions, we obtain further for $\epsilon < 1$

$$\begin{aligned}cc_p^2\frac{\delta^2}{\Delta t}\left\|\nabla z_u^l\right\|_{B_\delta^k}^2 &\leq cc_p^2\frac{\delta^3}{\Delta t}\left(\epsilon^{-1}\left\|\nabla z_u^l\right\|_{\Omega_\delta^k}^2 + \epsilon\left\|\nabla^2 z_u^l\right\|_{\Omega_\delta^k}^2\right) \\ &\leq cc_p^2c_\delta^3 w_{\max}^3\Delta t^2\left(\epsilon^{-1}\left\|\nabla z_u^l\right\|_{\Omega^l}^2 + \epsilon\left\|\nabla^2 z_u^l\right\|_{\Omega^l}^2\right).\end{aligned} \tag{105}$$

For sufficiently small $\Delta t < \frac{1}{cc_P^2 c_\delta^3 w_{\max}^3}$ we obtain from (103)-(105)

$$\frac{1}{\Delta t}\left\|z_u^k - z_u^{k+1}\right\|_{B_\delta^k}\left\|z_u^{k+1}\right\|_{B_\delta^k} \leq \Delta t \sum_{l=k}^{k+1}\left(\epsilon^{-1}\left\|\nabla z_u^l\right\|_{\Omega^l}^2 + \epsilon\left\|\nabla^2 z_u^l\right\|_{\Omega^l}^2\right). \quad (106)$$

For the third term in (102), we use a telescope argument

$$\left(\nabla z_u^k, \nabla(z_u^k - z_u^{k+1})\right)_{\Omega^k} = \frac{1}{2}\left(\|\nabla z_u^k\|_{\Omega^k}^2 + \|\nabla(z_u^k - z_u^{k+1})\|_{\Omega^k}^2 - \|\nabla z_u^{k+1}\|_{\Omega^k}^2\right).$$

To bring the last term to Ω^{k+1}, we estimate using (34)

$$\|\nabla z_u^{k+1}\|_{\Omega^k}^2 \leq \|\nabla z_u^{k+1}\|_{\Omega^{k+1}}^2 + \|\nabla z_u^{k+1}\|_{S_\delta^k}^2 \leq (1 + cc_\delta^2 w_{\max}^2 \epsilon^{-1}\Delta t)\|\nabla z_u^{k+1}\|_{\Omega^{k+1}}^2 + \epsilon\Delta t\|\nabla^2 z_u^{k+1}\|_{\Omega^{k+1}}^2$$

For the boundary term in (102), we use Green's theorem on S_δ^k

$$(\partial_n z_u^k, z_u^{k+1})_{\partial\Omega^k} \leq -(\partial_n z_u^k, \underbrace{z_u^{k+1}}_{=0})_{\partial\Omega^{k+1}} + \|\nabla z_u^k\|_{S_\delta^k}\|\nabla z_u^{k+1}\|_{S_\delta^k} + \|\Delta z_u^k\|_{S_\delta^k}\|z_u^{k+1}\|_{S_\delta^k} \quad (107)$$

For the second term on the right-hand side in (107) we use (34) twice with $\epsilon < 1$, followed by (9), the stability of the extensions and Young's inequality

$$\begin{aligned}
&\|\nabla z_u^k\|_{S_\delta^k}\|\nabla z_u^{k+1}\|_{S_\delta^k} \\
&\quad\leq c\delta\left(\epsilon^{-1/2}\|\nabla z_u^k\|_{\Omega_\delta^k} + \epsilon^{1/2}\|\nabla^2 z_u^k\|_{\Omega_\delta^k}\right) \\
&\qquad\times\left(\epsilon^{-1/2}\|\nabla z_u^{k+1}\|_{\Omega_\delta^{k+1}} + \epsilon^{1/2}\|\nabla^2 z_u^{k+1}\|_{\Omega_\delta^{k+1}}\right) \\
&\quad\leq cc_\delta w_{\max}\Delta t\left(\epsilon^{-1/2}\|\nabla z_u^k\|_{\Omega^k} + \epsilon^{1/2}\|\nabla^2 z_u^k\|_{\Omega^k}\right) \\
&\qquad\times\left(\epsilon^{-1/2}\|\nabla z_u^{k+1}\|_{\Omega^{k+1}} + \epsilon^{1/2}\|\nabla^2 z_u^{k+1}\|_{\Omega^{k+1}}\right) \\
&\quad\leq \Delta t\sum_{l=k}^{k+1}\left(cc_\delta^2 w_{\max}^2\epsilon^{-1}\|\nabla z_u^l\|_{\Omega^l}^2 + \epsilon\|\nabla^2 z_u^l\|_{\Omega^l}^2\right).
\end{aligned}$$

For the last term in (107) we obtain as in (104)

$$\begin{aligned}
\|\Delta z_u^k\|_{S_\delta^k}\|z_u^{k+1}\|_{S_\delta^k} &\leq c\delta\|\Delta z_u^k\|_{S_\delta^k}\|\nabla z_u^{k+1}\|_{S_\delta^k} \\
&\leq \epsilon\Delta t\|\Delta z_u^k\|_{\Omega^k}^2 + cc_\delta^2 w_{\max}^2\epsilon^{-1}\Delta t\|\nabla z_u^{k+1}\|_{\Omega^{k+1}}^2. \quad (108)
\end{aligned}$$

In the last inequality, we have used (9) and Young's inequality. Together, (107)–(108) yield the estimate

$$(\partial_n \boldsymbol{z}_u^k, \boldsymbol{z}_u^{k+1})_{\partial\Omega^k} \leq \Delta t \sum_{l=k}^{k+1} \left(c c_\delta^2 w_{\max}^2 \epsilon^{-1} \|\nabla \boldsymbol{z}_u^l\|_{\Omega^l}^2 + \epsilon \|\nabla^2 \boldsymbol{z}_u^l\|_{\Omega^l}^2 \right).$$

To estimate the pressure term in (102), we obtain as in (108)

$$\|\nabla z_p^k\|_{B_\delta^k} \|\boldsymbol{z}_u^{k+1}\|_{B_\delta^k} \leq \Delta t \left(\epsilon \|\nabla z_p^k\|_{\Omega^k}^2 + c c_\delta^2 w_{\max}^2 \epsilon^{-1} \|\nabla \boldsymbol{z}_u^{k+1}\|_{\Omega^{k+1}}^2 \right).$$

To summarise we have shown that

$$\begin{aligned} \frac{1}{2\Delta t} \left\| \boldsymbol{z}_u^k - \boldsymbol{z}_u^{k+1} \right\|_{\Omega^k}^2 + \frac{1}{2} \left((1 - c\epsilon^{-1} w_{\max}^2 \Delta t) \|\nabla \boldsymbol{z}_u^k\|_{\Omega^k}^2 + \|\nabla (\boldsymbol{z}_u^k - \boldsymbol{z}_u^{k+1})\|_{\Omega^k}^2 \right) \\ \leq \frac{1}{2} (1 + c\epsilon^{-1} w_{\max}^2 \Delta t) \|\nabla \boldsymbol{z}_u^{k+1}\|_{\Omega^{k+1}}^2 \\ + \epsilon \Delta t \left(\|\nabla z_p^k\|_{\Omega^k}^2 + \|\nabla^2 \boldsymbol{z}_u^k\|_{\Omega^k}^2 + \|\nabla^2 \boldsymbol{z}_u^{k+1}\|_{\Omega^{k+1}}^2 \right) + c \Delta t \|\boldsymbol{e}_u^k\|_{\Omega^k}^2. \end{aligned} \tag{109}$$

For $k = n$ we obtain from (100) tested with $\boldsymbol{\phi}_u^n = \boldsymbol{z}_u^n$ and $\phi_p^n = z_p^n$ that

$$\frac{1}{2\Delta t} \|\boldsymbol{z}_u^n\|_{\Omega^n}^2 + \frac{1}{2} \|\nabla \boldsymbol{z}_u^n\|_{\Omega^n}^2 \leq \frac{\Delta t}{2} \|\boldsymbol{e}_u^n\|_{\Omega^n}^2. \tag{110}$$

Summation in (109) over $k = 1, \ldots, n-1$ and addition of (110) and (98) multiplied by a factor of 3ϵ yields for $\epsilon < \frac{1}{12 c_0}$

$$\begin{aligned} \|\nabla \boldsymbol{z}_u^1\|_{\Omega^1}^2 + \|\nabla \boldsymbol{z}_u^n\|_{\Omega^n}^2 + \frac{1}{\Delta t} \|\boldsymbol{z}_u^n\|_{\Omega^n}^2 + \sum_{k=1}^{n-1} \left\{ \frac{1}{\Delta t} \left\| \boldsymbol{z}_u^k - \boldsymbol{z}_u^{k+1} \right\|_{\Omega^k}^2 + \|\nabla (\boldsymbol{z}_u^k - \boldsymbol{z}_u^{k+1})\|_{\Omega^k}^2 \right\} \\ + \epsilon \Delta t \sum_{k=1}^{n} \left\{ \|\boldsymbol{z}_u^k\|_{H^2(\Omega^k)}^2 + \|z_p^k\|_{H^1(\Omega^k)}^2 \right\} \leq c \Delta t \sum_{k=1}^{n} \left\{ \|\boldsymbol{e}_u^k\|_{\Omega^k}^2 + w_{\max}^2 \|\nabla \boldsymbol{z}_u^k\|_{\Omega^k}^2 \right\}. \end{aligned} \tag{111}$$

Using (96) we can estimate the last term by

$$w_{\max}^2 \Delta t \sum_{k=1}^{n} \|\nabla \boldsymbol{z}_u^k\|_{\Omega^k}^2 \leq c w_{\max}^2 \Delta t \sum_{k=1}^{n} \|\boldsymbol{e}_u^k\|_{\Omega^k}^2,$$

which completes the proof. □

Now we are ready to prove an error estimate for the $L^2(L^2)$-norm of the velocities. First, we note that, due to the regularity proven in Lemma 5.10, the solution $(\boldsymbol{z}_u^k, z_p^k)_{k=1}^n$

of (83) is also the unique solution to the Nitsche formulation: *Find* $(\boldsymbol{z}_u^k, z_p^k)_{k=1}^n$, *where* $\boldsymbol{z}_u^k \in H^2(\Omega^k)^d$, $z_p^k \in H^1(\Omega) \cap L_0^2(\Omega^k)$ *such that*

$$\Delta t \sum_{k=1}^{n} \left(\boldsymbol{e}_u^k, \boldsymbol{\phi}_u^k\right)_{\Omega^k} = \sum_{k=1}^{n} \left\{ (\boldsymbol{\phi}_u^k - E^k \boldsymbol{\phi}_u^{k-1}, \boldsymbol{z}_u^k)_{\Omega^k} + \Delta t \left(\mathcal{A}_S^k + a_D^k\right) (\boldsymbol{\phi}_u^k, \phi_p^k; \boldsymbol{z}_u^k, z_p^k) \right\} + (\boldsymbol{\phi}_u^0, \boldsymbol{z}_u^1)_{\Omega^1} \quad \forall \boldsymbol{\phi}_u^k \in H^1(\Omega^k), \phi_p^k \in L^2(\Omega^k), \quad k = 1, \ldots, n. \tag{112}$$

Theorem 5.12 *We assume that the solution* $(\boldsymbol{u}, p)$ *of* (5) *fulfils the regularity assumptions* $\boldsymbol{u}(t_k) \in H^{m+1}(\Omega^k)^d$ *and* $p(t_k) \in H^m(\Omega^k)$ *for* $k = 1, \ldots n$ *and* $s = 1$. *Under the assumptions of Theorem* 5.3 *and the inverse CFL condition* $\Delta t \geq ch^2$ *for some* $c > 0$, *it holds that*

$$\left(\Delta t \sum_{k=1}^{n} \|\boldsymbol{e}_u^k\|_{\Omega^k}^2\right)^{1/2} \leq c w_{\max} \exp(c_1(w_{\max}) t_n) \left(\Delta t \|\partial_t^2 \boldsymbol{u}\|_Q + h^{m+1} \left(\|\boldsymbol{u}\|_{\infty,m+1} + \|\partial_t \boldsymbol{u}\|_{\infty,m} + \|p\|_{\infty,m}\right)\right),$$

with $c_1(w_{\max})$ *specified in Lemma* 4.1.

Proof We test (112) with $\boldsymbol{\phi}_u^k = \boldsymbol{e}_u^k$, $\phi_p^k = e_p^k$, $k = 0, \ldots, n$ to get

$$\Delta t \sum_{k=1}^{n} \|\boldsymbol{e}_u^k\|_{\Omega^k}^2 = \sum_{k=1}^{n} \left\{ (\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}), \boldsymbol{z}_u^k)_{\Omega^k} + \Delta t \left(\mathcal{A}_S^k + a_D^k\right) (\boldsymbol{e}_u^k, e_p^k; \boldsymbol{z}_u^k, z_p^k) \right\} + (\boldsymbol{e}_u^0, \boldsymbol{z}_u^1)_{\Omega^1}.$$

We define

$$\eta_{z,u}^k := \boldsymbol{z}_u^k - I_h^k \boldsymbol{z}_u^k, \quad \eta_{z,p}^k := z_p^k - i_h^k z_p^k$$

and use Galerkin orthogonality to insert the interpolants $I_h^k \boldsymbol{z}_u^k$ and $i_h^k z_p^k$

$$\Delta t \sum_{k=1}^{n} \|\boldsymbol{e}_u^k\|_{\Omega^k}^2 = \sum_{k=1}^{n} \left\{ (\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}), \eta_{z,u}^k)_{\Omega^k} + \Delta t \left(\mathcal{A}_S^k + a_D^k\right) (\boldsymbol{e}_u^k, e_p^k; \eta_{z,u}^k, \eta_{z,p}^k) \right\} + \Delta t \sum_{k=1}^{n} \left\{ D_t^{(1)} \boldsymbol{u}(t_k) - \partial_t \boldsymbol{u}(t_k), i_h^k \boldsymbol{z}_u^k)_{\Omega^k} + \gamma_g g_h^k(\boldsymbol{u}_h^k, i_h^k \boldsymbol{z}_u^k) + \gamma_p s_h^k(p_h^k, i_h^k z_p^k) \right\} \tag{113}$$

We use the continuity of the bilinear form $\mathcal{A}_S^k + a_D^k$ (Lemma 3.4) and standard interpolation estimates

$$\begin{aligned}(\mathcal{A}_S^k + a_D^k)(\boldsymbol{e}_u^k, e_p^k; \eta_{z,u}^k, \eta_{z,p}^k) &\le c\left(\|\nabla \boldsymbol{e}_u^k\|_{\Omega^k} + h^{-1/2}\|\boldsymbol{e}_u^k\|_{\partial\Omega^k} + h^{1/2}\|\partial_n \boldsymbol{e}_u^k\|_{\partial\Omega^k} + h\|\nabla e_p^k\|_{\Omega^k}\right)\\ &\quad\cdot\Big(\|\nabla\eta_{z,u}^k\|_{\Omega^k} + h^{-1}\|\eta_{z,u}^k\|_{\Omega^k} + \|\eta_{z,p}^k\|_{\Omega^k} + h^{-1/2}\|\eta_{z,u}^k\|_{\partial\Omega^k} + h^{1/2}\\ &\qquad\left(\|\partial_n\eta_{z,u}^k\|_{\partial\Omega^k} + \|\eta_{z,p}^k\|_{\partial\Omega^k}\right)\Big)\\ &\le ch\left(\|\nabla \boldsymbol{e}_u^k\|_{\Omega^k} + h^{-1/2}\|\boldsymbol{e}_u^k\|_{\partial\Omega^k} + h^{1/2}\|\nabla \boldsymbol{e}_u^k\|_{\partial\Omega^k} + h\|\nabla e_p^k\|_{\Omega^k}\right)\left(\|\nabla^2 \boldsymbol{z}_u^k\|_{\Omega^k} + \|\nabla z_p^k\|_{\Omega^k}\right).\end{aligned}$$

To estimate $h^{1/2}\|\nabla \boldsymbol{e}_u^k\|_{\partial\Omega^k}$ we split into a discrete and an interpolatory part and use an inverse inequality and Lemma 3.3

$$\begin{aligned}h^{1/2}\|\nabla \boldsymbol{e}_u^k\|_{\partial\Omega^k} \le h^{1/2}\left(\|\nabla\eta_u^k\|_{\partial\Omega^k} + \|\nabla\xi_{h,u}^k\|_{\partial\Omega^k}\right) &\le ch^m\|\boldsymbol{u}\|_{H^{m+1}(\Omega^k)} + c\|\nabla\xi_{h,u}^k\|_{\Omega_h^k}\\ &\le ch^m\|\boldsymbol{u}\|_{H^{m+1}(\Omega^k)} + c\|\nabla \boldsymbol{e}_u^k\|_{\Omega_h^k}\\ &\le ch^m\|\boldsymbol{u}\|_{H^{m+1}(\Omega^k)} + c|||\boldsymbol{e}_u^k|||_{h,k}.\end{aligned} \tag{114}$$

This yields

$$\begin{aligned}&(\mathcal{A}_S^k + a_D^k)(\boldsymbol{e}_u^k, e_p^k; \eta_{z,u}^k, \eta_{z,p}^k)\\ &\quad\le ch\left(|||\boldsymbol{e}_u^k|||_{h,k} + h\|\nabla e_p^k\|_{\Omega^k} + h^m\|\boldsymbol{u}^k\|_{H^{m+1}(\Omega^k)}\right)\left(\|\nabla^2 \boldsymbol{z}_u^k\|_{\Omega^k} + \|\nabla z_p^k\|_{\Omega^k}\right).\end{aligned}$$

For the consistency error of the time derivative on the right-hand side of (113), we obtain as in Lemma 5.1

$$\left|\left(D_t^{(1)}\boldsymbol{u}(t_k) - \partial_t\boldsymbol{u}(t_k), i_h^k \boldsymbol{z}_u^k\right)_{\Omega^n}\right| \le c\Delta t\|\partial_t^2\boldsymbol{u}\|_{Q^k}\|\boldsymbol{z}_u^k\|_{\Omega^k}.$$

For the ghost penalty we insert $\pm \boldsymbol{z}_u^k$ and $\pm \boldsymbol{u}(t_k)$ and use Lemma 3.3 as well as standard estimates for the interpolation

$$\begin{aligned}g_h^k(\boldsymbol{u}_h^k, i_h^k \boldsymbol{z}_u^k) = g_h^k(\boldsymbol{e}_u^k, \eta_{z,u}^k) - g_h^k(\boldsymbol{u}(t_k), \eta_{z,u}^k) - g_h^k(\boldsymbol{e}_u^k, \boldsymbol{z}_u^k) + g_h^k(u(t_k), \boldsymbol{z}_u^k)\\ \le ch\left(g_h^k(\boldsymbol{e}_u^k, \boldsymbol{e}_u^k)^{1/2} + h^m\|\boldsymbol{u}\|_{H^{m+1}(\Omega^k)}\right)\|\boldsymbol{z}_u^k\|_{H^2(\Omega^k)}.\end{aligned}$$

For the pressure stabilisation we distinguish between the cases $m = 1$ and $m > 1$, the latter implying by assumption that $p^k \in H^2(\Omega^k)$. For $m = 1$, the following estimate is optimal

$$s_h^k(p_h^k, i_h^k z_p^k) \le ch^2\|\nabla p_h^k\|_{\Omega^k}\|\nabla z_p^k\|_{\Omega^k} \le ch^2\left(\|\nabla p^k\|_{\Omega^k} + \|\nabla e_p^k\|\right)\|\nabla z_p^k\|_{\Omega^k}.$$

For $m > 1$ we insert $\pm p^k$ and use (10), (11) and (14)

$$
\begin{aligned}
s_h^k(p_h^k, i_h^k z_p^k) &= -s_h^k(e_p^k, i_h^k z_p^k) + s_h^k(p^k, i_h^k z_p^k) \\
&\leq s_h^k(e_p^k, e_p^k)^{1/2} s_h^k(i_h^k z_p^k, i_h z_p^k)^{1/2}) + \underbrace{s_h^k(p^k, p^k)}_{=0}{}^{1/2} s_h^k(i_h^k z_p^k, i_h^k z_p^k)^{1/2} \\
&\leq ch^2 \|\nabla e_p^k\|_{\Omega^k} \|\nabla z_p^k\|_{\Omega^k}.
\end{aligned}
$$

It remains to estimate the terms corresponding to the discrete time derivative in (113). We use a standard interpolation estimate and the inverse CFL condition $h^2 \leq c\Delta t$ to get

$$
\begin{aligned}
(\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}, \eta_{z,u}^k)_{\Omega^k} &\leq ch^2 \|\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}\|_{\Omega^k} \|\nabla^2 \boldsymbol{z}_u^k\|_{\Omega^k} \\
&\leq ch \|\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}\|_{\Omega^k} \Delta t^{1/2} \|\nabla^2 \boldsymbol{z}_u^k\|_{\Omega^k}. \qquad (115)
\end{aligned}
$$

By combining the above estimates, we have from (113)

$$
\begin{aligned}
\Delta t \sum_{k=1}^n \|\boldsymbol{e}_u^k\|_{\Omega^k}^2 &\leq c\Delta t h \sum_{k=1}^n \Big\{ \Big(|||\boldsymbol{e}_u^k|||_{k,h} + h\|\nabla e_p^k\|_{\Omega^k} + h^m \left(\|\boldsymbol{u}^k\|_{H^{m+1}(\Omega^k)} + \|p^k\|_{H^m(\Omega^k)} \right) \Big) \\
&\quad \cdot \left(\|\boldsymbol{z}_u^k\|_{H^2(\Omega^k)} + \|\nabla z_p^k\|_{\Omega^k} \right) \Big\} \\
&\quad + c\Delta t^2 \sum_{k=1}^n \left\{ \|\partial_t^2 \boldsymbol{u}\|_{Q^k} \|\boldsymbol{z}_u^k\|_{\Omega^k} \right\} + ch \sum_{k=1}^{n-1} \left\{ \|\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}\|_{\Omega^k} \Delta t^{1/2} \|\nabla^2 \boldsymbol{z}_u^k\|_{\Omega^k} \right\} \\
&\leq ch^2 \Bigg(\sum_{k=1}^n \|\boldsymbol{e}_u^k - \boldsymbol{e}_u^{k-1}\|_{\Omega^k}^2 + \Delta t \Big\{ |||\boldsymbol{e}_u^k|||_{k,h}^2 + h^2 \|\nabla e_p^k\|_{\Omega^k}^2 + \Delta t^2 \|\partial_t^2 \boldsymbol{u}\|_{Q^k}^2 \\
&\quad + h^{2m} \left(\|\boldsymbol{u}^k\|_{H^{m+1}(\Omega^k)}^2 + \|p^k\|_{H^m(\Omega^k)}^2 \right) \Big\} \Bigg)^{1/2} \left(\Delta t \sum_{k=1}^n \|\boldsymbol{z}_u^k\|_{H^2(\Omega^k)}^2 + \|\nabla z_p^k\|_{\Omega^k}^2 \right)^{1/2}. \qquad (116)
\end{aligned}
$$

The last inequality follows by the Cauchy–Schwarz inequality. Now, the statement follows from Theorem 5.3 and Lemma 5.11.

□

Remark 5.13 An analogous result can be shown for the BDF(2) variant under slightly stronger conditions. For $\Delta t \geq ch$, which is needed for the energy estimate, the following estimate can be shown

$$
\begin{aligned}
&\left(\Delta t \sum_{k=1}^n \|\boldsymbol{e}_u^k\|_{\Omega^k}^2 \right)^{1/2} \\
&\quad \leq c w_{\max} \exp(c_1 (w_{\max}) t_n) \Big(\Delta t^2 \|\partial_t^3 \boldsymbol{u}\|_Q \\
&\qquad + h^{m+1} \left(\|\boldsymbol{u}\|_{\infty,m+1} + \|\partial_t \boldsymbol{u}\|_{\infty,m} + \|p\|_{\infty,m} \right) \Big). \qquad (117)
\end{aligned}
$$

The main difference in the proof is that the energy norm estimate does not give a bound for $\Delta t \| D_t^{(2)} \boldsymbol{e}_u^k \|_{\Omega^k}$, see Remark 5.5. We have using (34) with $\epsilon = 1$

$$\begin{aligned}
\Delta t (D_t^{(2)} \boldsymbol{e}_u^k, \eta_{z,u}^k)_{\Omega^k} &\leq c h^2 \sum_{i=0}^{2} \| \boldsymbol{e}_u^{k+i} \|_{\Omega^k} \| \nabla^2 \boldsymbol{z}_u^k \|_{\Omega^k} \\
&\leq c h^2 \sum_{i=0}^{2} \left(\| \boldsymbol{e}_u^{k+i} \|_{\Omega^{k+i}} + \| \boldsymbol{e}_u^{k+i} \|_{S_\delta^k} \right) \| \nabla^2 \boldsymbol{z}_u^k \|_{\Omega^k} \\
&\leq c \Delta t^2 \sum_{i=0}^{2} \left(\| \boldsymbol{e}_u^{k+i} \|_{\Omega^{k+i}} + h \| \nabla \boldsymbol{e}_u^{k+i} \|_{\Omega^{k+i}} \right) \| \nabla^2 \boldsymbol{z}_u^k \|_{\Omega^k}.
\end{aligned}$$

The L^2-term on the right-hand side can then be absorbed into the left-hand side of (116) to obtain (117).

6 Numerical example

To substantiate the theoretical findings, we present numerical results for polynomial degrees $m = 1, 2$ and BDF formulas of order $s = 1, 2$. The results have been obtained using the CutFEM library [16], which is based on FeNiCS [1].

We consider flow through a 3-dimensional rectangular channel with a moving upper and lower wall in the time interval is $I = [0, 2]$. The moving domain is given by

$$\Omega(t) = (0, 4) \times \left(-1 + \frac{\sin(t)}{10}, 1 - \frac{\sin(t)}{10} \right) \times (-1, 1).$$

Due to the simple polygonal structure of the domain $\Omega(t)$, the integrals in (19) are evaluated exactly within the CutFEM library [16] and we can expect higher-order convergence in space for $m \geq 2$.

The data $\boldsymbol{f}$ and $\boldsymbol{u}^D$ is chosen in such a way that the manufactured solution

$$\begin{aligned}
\boldsymbol{u}(x, y, z; t) &= \left(\sin(t) \cdot \left((1 - \frac{\sin(t)}{10})^2 - y^2 \right)(1 - z^2), 0, 0 \right), \\
p(x, y, z; t) &= \sin(t) \cdot (8 - 2x)
\end{aligned}$$

solves the system (5). We impose the corresponding Dirichlet boundary conditions $\boldsymbol{u}^D$ on the left *inflow* boundary (given by $x = 0$), a *do-nothing* boundary condition $\partial_n \boldsymbol{u} - p\boldsymbol{n} = 0$ on the right *outflow* boundary (given by $x = 4$) and no-slip boundary conditions on the remaining boundary parts, including the moving upper and lower boundary. The initial value is homogeneous $\boldsymbol{u}^0(x) = 0$. We choose a Nitsche parameter $\gamma_D = 500$, stabilisation parameters $\gamma_g = \gamma_p = 10^{-3}$ and $\delta = w_{\max} s \Delta t$, where $w_{\max} = \max_{t \in I, x \in \partial\Omega(t)} \| \partial_t \boldsymbol{T} \cdot \boldsymbol{n} \| = 0.1$. The background triangulations $\mathcal{T}_h$ are constructed from a uniform subdivision of the box $[0, 4] \times [-1.1, 1.1] \times [-1, 1]$ into hexahedra

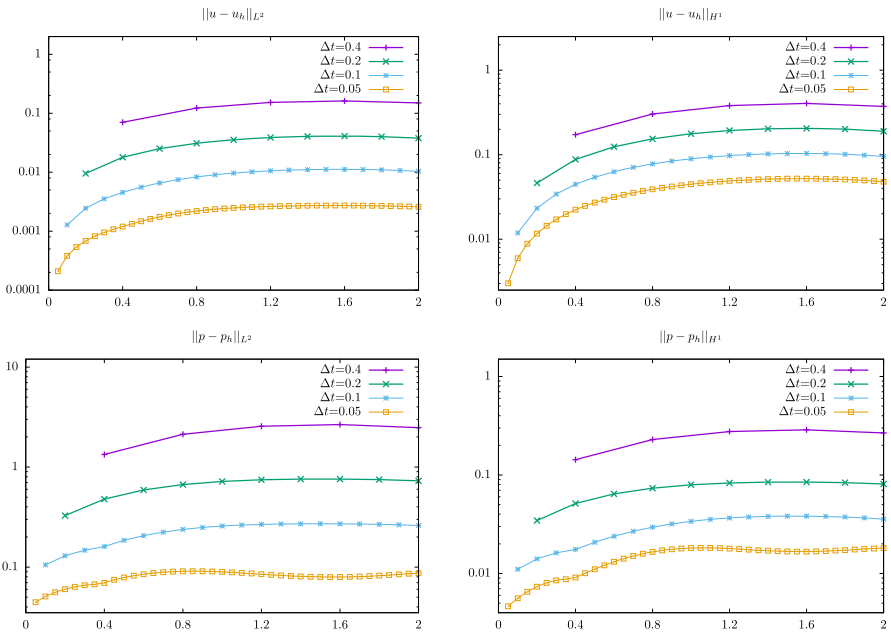


Fig. 2 Top: L^2- and H^1-norm error of velocity, bottom: L^2-and H^1 norm error of pressure over time for different time-step sizes and mesh levels, where $\Delta t = 0.8h$. All the norms are normalised by the maximum (in time) of the respective norm of the continuous functions

and a subsequent split of each of the hexahedral elements into 6 tetrahedra. These background triangulations are then reduced in each time-step by eliminating those elements that lie outside of Ω_δ^n.

6.1 P_1 - BDF(1)

First, we use P_1 finite elements ($m = 1$) and the BDF(1) variant ($s = 1$). The computed errors $\|\boldsymbol{u}-\boldsymbol{u}_h\|_\Omega$, $\|\nabla(\boldsymbol{u}-\boldsymbol{u}_h)\|_\Omega$, $\|p^k-p_h^k\|_\Omega$ and $\|\nabla(p^k-p_h^k)\|_\Omega$ are plotted over time in Fig. 2 for $\Delta t = 0.8h$, where each of the norms has been normalised by the $L^\infty(L^2)$-norm of the respective continuous functions, e.g. $\|\boldsymbol{u}-\boldsymbol{u}_h\|_\Omega/\|\boldsymbol{u}\|_{\infty,0,I}$. We observe convergence in all norms for all times as $\Delta t = 0.8h \to 0$. Moreover, no oscillations are visible in any of the norms. While the error bounds shown in the previous sections include an exponential growth in time, coming from the application of Gronwall's lemma, the error does not accumulate significantly over time in the numerical results presented here.

To study the convergence orders in space and time, we show values for four different time-step and four different mesh sizes in Table 1. For P_1 finite elements, the finest mesh contains approximately 143.000 degrees of freedom. We observe that the temporal error is barely visible in the $L^2(L^2)$-norm and $L^2(H^1)$-semi-norm of velocities, as the spatial error is dominant. The spatial component of the velocities converges as expected by the theory (Theorems 5.3, 5.12) with orders 2 and 1.

Table 1 Errors for the fully discrete solutions for P_1 finite elements and BDF(1) for different mesh and time-step sizes

$h \downarrow \backslash \Delta t \rightarrow$	0.4	0.2	0.1	0.05	g_h	$\mathrm{eoc}_{\Delta t}$
$\|\boldsymbol{u}-\boldsymbol{u}_{kh}\|_Q/\|\boldsymbol{u}\|_Q$						
1/2	1.60×10^{-1}	1.60×10^{-1}	1.60×10^{-1}	1.57×10^{-1}	1.57×10^{-1}	–
1/4	4.07×10^{-2}	4.05×10^{-2}	4.05×10^{-2}	4.05×10^{-2}	4.04×10^{-2}	1.16
1/8	1.14×10^{-2}	1.11×10^{-2}	1.10×10^{-2}	1.10×10^{-2}	1.09×10^{-2}	1.57
1/16	4.25×10^{-3}	3.24×10^{-3}	2.88×10^{-3}	2.76×10^{-3}	2.69×10^{-3}	1.52
$g_{\Delta t}$	1.90×10^{-3}	9.52×10^{-4}	5.87×10^{-4}	3.02×10^{-4}	0	0.90
eoc_h	2.03	2.00	1.99	1.96	1.95	
$\|\nabla(\boldsymbol{u}-\boldsymbol{u}_{kh})\|_Q/\|\nabla\boldsymbol{u}\|_Q$						
1/2	4.13×10^{-1}	4.14×10^{-1}	4.13×10^{-1}	4.05×10^{-1}	4.05×10^{-1}	–
1/4	2.05×10^{-1}	2.05×10^{-1}	2.05×10^{-1}	2.05×10^{-1}	2.05×10^{-1}	1.07
1/8	1.03×10^{-1}	1.03×10^{-1}	1.03×10^{-1}	1.03×10^{-1}	1.03×10^{-1}	1.75
1/16	5.20×10^{-2}	5.17×10^{-2}	5.16×10^{-2}	5.16×10^{-2}	5.16×10^{-2}	1.90
$g_{\Delta t}$	2.95×10^{-3}	2.94×10^{-3}	-1.88×10^{-3}	-1.88×10^{-3}	0	–
eoc_h	1.02	1.02	0.98	0.98	0.99	

Table 1 continued

$h \downarrow \backslash \Delta t \rightarrow$	0.4	0.2	0.1	0.05	g_h	$\text{eoc}_{\Delta t}$
$\|p - p_{kh}\|_Q / \|p\|_Q$						
1/2	2.83×10^{-1}	2.61×10^{-1}	2.49×10^{-1}	2.46×10^{-1}	2.42×10^{-1}	1.14
1/4	1.08×10^{-1}	8.36×10^{-2}	7.14×10^{-2}	6.53×10^{-2}	5.92×10^{-2}	1.00
1/8	6.65×10^{-2}	4.22×10^{-2}	2.98×10^{-2}	2.36×10^{-2}	1.71×10^{-2}	0.98
1/16	5.36×10^{-2}	2.93×10^{-2}	1.67×10^{-2}	1.02×10^{-2}	3.20×10^{-3}	0.95
$g_{\Delta t}$	5.09×10^{-2}	2.08×10^{-2}	1.42×10^{-2}	6.83×10^{-3}	0	1.05
eoc_h	2.01	2.03	2.02	2.18	2.01	
$\|\nabla(p - p_{kh})\|_Q / \|\nabla p\|_Q$						
1/2	3.05×10^{-1}	2.82×10^{-1}	2.71×10^{-1}	2.68×10^{-1}	2.65×10^{-1}	1.25
1/4	1.15×10^{-1}	9.25×10^{-2}	8.13×10^{-2}	7.58×10^{-2}	7.03×10^{-2}	1.01
1/8	7.16×10^{-2}	4.97×10^{-2}	3.96×10^{-2}	3.51×10^{-2}	3.12×10^{-2}	1.13
1/16	5.66×10^{-2}	3.39×10^{-2}	2.36×10^{-2}	1.94×10^{-2}	1.58×10^{-2}	1.18
$g_{\Delta t}$	5.44×10^{-2}	3.19×10^{-2}	2.20×10^{-2}	1.82×10^{-2}	0	0.61
eoc_h	2.03	2.02	2.04	2.23	1.77	

The mesh size h stands here for the minimal edge of the triangles and $Q = \{(x,t) \in \mathbb{R}^{d+1}, t \in I, x \in \Omega(t)\}$. Moreover, we show extrapolated values g_h and $g_{\Delta t}$ and *experimental order of convergences* eoc_h and $\text{eoc}_{\Delta t}$ based on a least squares fit of the four available values for fixed h or Δt against the three parameters of the function $g(h) = g_{\Delta t} + c_h h^{\text{eoc}_h}$ and $g(\Delta t) = g_h + c_{\Delta t} h^{\text{eoc}_{\Delta t}}$, respectively. On the bottom right of each table, we show estimated convergence orders of the extrapolated values g_h and $g_{\Delta} t$ by fitting the two parameters of the functions $g(\Delta t) = 0 + c_{\Delta t} \Delta t^{\text{eoc}_{\Delta t}}$ and $g(h) = 0 + c_h h^{\text{eoc}_h}$, respectively. A "–" in the eoc column means that the asymptotic standard error of the fit (computed by gnuplot) was $> 20\%$, which means that the parameter could not be determined with a reasonable accuracy from the given values

On the other hand, the temporal error shows up clearly in the pressure norms. To compute an *estimated order of convergence* (eoc), let us assume that the overall error can be separated into a temporal and a spatial component

$$g(\Delta t, h) = g_{\Delta t}(\Delta t) + g_h(h) = c_h h^{\mathrm{eoc_h}} + c_{\Delta t} \Delta t^{\mathrm{eoc}_{\Delta t}}. \tag{118}$$

To estimate for instance the temporal order of convergence $\mathrm{eoc}_{\Delta t}$, we fit the three parameters g_h, $c_{\Delta t}$ and $\mathrm{eoc}_{\Delta t}$ of the function

$$g(\Delta t, \cdot) = g_h + c_{\Delta t} \Delta t^{\mathrm{eoc_{1t}}}$$

for a fixed mesh size $h \in \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}\}$ against the computed values. This is done by means of a least-squares fit using gnuplot [40]. The values for g_h and $\mathrm{eoc}_{\Delta t}$ in the first row are for example computed by fitting the previous values in the same row (i.e. those obtained with $h = \frac{1}{2}$ for different time-step sizes). A spatial order of convergence eoc_h is estimated similarly using the values for a fixed time-step size $\Delta t \in \{0.4, 0.2, 0.1, 0.05\}$, i.e. those in the same column.

For the pressure norms the estimated temporal order of convergence is very close to 1 in both the L^2- and the H^1 semi-norm. This is expected for the $L^2(H^1)$-semi-norm by Theorem 5.3, but better than proven in Lemma 5.7 for the $L^2(L^2)$-norm. The spatial component of the error converges much faster than expected with eoc_h around 2 for both norms (compared to $\mathcal{O}(1)$, which has been shown for the H^1-semi-norm, and $\mathcal{O}(h)$ for the L^2-norm). This might be due to superconvergence effects, as frequently observed for CIP stabilisations (see e.g. [26]), and possibly due to the sub-optimality of the pressure estimates.

The convergence orders of both pressure norms are very similar, especially for larger h and Δt. Here it seems that due to the superconvergence of the $L^2(H^1)$-semi-norm the simple Poincaré estimate

$$\|e_p^k\|_{\Omega^k} \leq c_P \|\nabla e_p^k\|_{\Omega^k}$$

is optimal for the $L^2(L^2)$-norm. Only for smaller Δt and h, the convergence of the $L^2(L^2)$-norm seems to be slightly faster compared to the $L^2(H^1)$-semi-norm.

6.2 P_2-BDF(1)

In order to increase the visibility of the temporal error component, we increase the order of the spatial discretisation first. In Table 2 we show results for P_2 finite elements and BDF(1) ($m = 2, s = 1$) on three different mesh levels. For P_2 the finest mesh level has again around 143.000 degrees of freedom, which is similar to P_1 elements on the next-finer mesh level. Again the spatial error is dominant in the velocity norms on coarser meshes and shows convergence orders of approximately 3 in the $L^2(L^2)$-norm and 2 in the $L^2(H^1)$-semi-norm, as shown in Theorems 5.12 and 5.3. In contrast to P_1 elements, the temporal error is however visible on the finest mesh level, where $\mathrm{eoc}_{\Delta t}$ is close to 1, as expected.

In the $L^2(L^2)$-norm of pressure, the temporal error is dominant and shows again a convergence order of $\mathcal{O}(\Delta t)$. Due to the dominance of the temporal component, it is less clear to deduce the spatial error contribution. From the values and the eoc_h it seems to converge again faster as predicted. Concerning the $L^2(H^1)$-norm of pressure, the assumption (118) that the spatial and temporal error are separated, which was assumed in order to compute $\text{eoc}_{\Delta t}$ and eoc_h, is not valid, as the extrapolated values g_h and $g_{\Delta t}$ do not or converge only very slowly towards zero. For this reason, the computed convergence orders $\text{eoc}_{\Delta t}$ and eoc_h are not meaningful in this case. This does not contradict the theory, as Theorem 5.3 guarantees only the bound

$$\left(\sum_{k=1}^{n} \Delta t \| \nabla e_p^k \|_{\Omega^k}^2 \right)^{1/2} \leq \mathcal{O}\left(\frac{\Delta t}{h} \right) + \mathcal{O}(h).$$

6.3 P_2-BDF(2)

Finally, we show results for $m = 2$ and $s = 2$ in Tables 3 and 4 . In Table 3, we use an extension of the analytically given solution $\boldsymbol{u}(x, t)$ to $t < 0$ for initialisation, i.e. we use the starting values $\boldsymbol{u}^0 = 0$ and $\boldsymbol{u}^{-1} := \boldsymbol{u}(-\Delta t)$ in the first time step. Due to the (expected) second-order convergence in time, the temporal error is barely visible in the velocity norms on the finer mesh levels, in contrast to the results for BDF(1). The estimated order of convergence of the spatial component lies slightly below the orders 3 and 2 in the $L^2(L^2)$-norm and $L^2(H^1)$-semi-norm, respectively, that have been shown analytically.

In the $L^2(L^2)$-norm of pressure both temporal and spatial errors are visible. Both eoc_h and $\text{eoc}_{\Delta t}$ are around 2, which has been shown in Sect. 5.1.1 for the spatial part. For the temporal part only a reduced order of convergence of $\mathcal{O}(\Delta t)$ has been shown theoretically. This bound seems not to be sharp in the numerical example studied here. In the $L^2(H^1)$-semi-norm of pressure the spatial error is dominant, which is in contrast to the BDF(1) results. However, the assumption (118) that the error allows for a separation into spatial and temporal error components is again not valid, which makes the computed values of $\text{eoc}_{\Delta t}$ and eoc_h meaningless.

In Table 4 we show results, where -instead of analytical values- one BDF(1) step has been used for initialisation, according to the discussion in Remark 5.6. We see that for large Δt the errors are slightly larger, due to the additional initial error. In fact the velocity norm errors are still relatively close, in particular for smaller Δt, while a stronger impact is visible in the pressure norms. These deviations get, however, smaller for $\Delta t \to 0$. Moreover and most importantly, the estimated orders of convergence are very similar in the velocity norms and lie still significantly above the theoretical predictions in the pressure norms. This confirms numerically that the initialisation with BDF(1) is indeed sufficient to preserve the theoretically predicted convergence orders.

Table 2 Errors for the fully discrete solutions for P_2 finite elements and BDF(1) for different mesh and time-step sizes

$h \downarrow \backslash \Delta t \rightarrow$	0.4	0.2	0.1	0.05	g_h	$\mathrm{eoc}_{\Delta t}$
$\|\boldsymbol{u}-\boldsymbol{u}_{kh}\|_Q/\|\boldsymbol{u}\|_Q$						
1/2	9.15×10^{-3}	8.92×10^{-3}	8.88×10^{-3}	8.88×10^{-3}	8.87×10^{-3}	2.40
1/4	3.20×10^{-3}	1.97×10^{-3}	1.50×10^{-3}	1.38×10^{-3}	1.28×10^{-3}	1.50
1/8	3.12×10^{-3}	1.62×10^{-3}	8.52×10^{-4}	4.78×10^{-4}	8.59×10^{-5}	0.99
$g_{\Delta t}$	3.12×10^{-3}	1.60×10^{-3}	7.90×10^{-4}	3.55×10^{-4}	0	0.99
eoc_h	6.08	4.31	3.51	3.06	2.82	
$\|\nabla(\boldsymbol{u}-\boldsymbol{u}_{kh})\|_Q/\|\nabla\boldsymbol{u}\|_Q$						
1/2	5.55×10^{-2}	5.55×10^{-2}	5.55×10^{-2}	5.55×10^{-2}	5.55×10^{-2}	–
1/4	1.67×10^{-2}	1.58×10^{-2}	1.56×10^{-2}	1.56×10^{-2}	1.56×10^{-2}	1.91
1/8	7.51×10^{-3}	5.18×10^{-3}	4.36×10^{-3}	4.12×10^{-3}	3.97×10^{-3}	1.56
$g_{\Delta t}$	4.66×10^{-3}	1.30×10^{-3}	-4.82×10^{-5}	-5.17×10^{-4}	0	2.13
eoc_h	2.08	1.90	1.83	1.80	1.85	
$\|p-p_{kh}\|_Q/\|p\|_Q$						
1/2	5.35×10^{-2}	2.87×10^{-2}	1.57×10^{-2}	9.20×10^{-3}	2.06×10^{-3}	0.95
1/4	5.08×10^{-2}	2.64×10^{-2}	1.37×10^{-2}	7.18×10^{-3}	1.10×10^{-3}	0.95
1/8	5.00×10^{-2}	2.57×10^{-2}	1.30×10^{-2}	6.53×10^{-3}	-5.40×10^{-4}	0.95
$g_{\Delta t}$	4.97×10^{-2}	2.54×10^{-2}	1.26×10^{-2}	6.22×10^{-3}	0	0.99
eoc_h	1.75	1.72	1.51	1.64	1.71	
$\|\nabla(p-p_{kh})\|_Q/\|\nabla p\|_Q$						
1/2	6.29×10^{-2}	4.13×10^{-2}	3.20×10^{-2}	2.83×10^{-2}	2.54×10^{-2}	1.24
1/4	5.43×10^{-2}	3.18×10^{-2}	2.17×10^{-2}	1.78×10^{-2}	1.45×10^{-2}	1.21
1/8	5.25×10^{-2}	2.99×10^{-2}	1.97×10^{-2}	1.60×10^{-2}	1.27×10^{-2}	1.23
$g_{\Delta t}$	5.20×10^{-2}	2.94×10^{-2}	1.92×10^{-2}	1.56×10^{-2}	0	0.67
eoc_h	2.26	2.32	2.36	2.54	-	

The *experimental orders of convergence* (eoc) have been computed as in Table 1

7 Conclusion

We have derived a detailed a priori error analysis for two Eulerian time-stepping schemes based on backward difference formulas applied to the non-stationary Stokes equations on time-dependent domains. Following Schott [56] and Lehrenfeld and Olshanskii [42] discrete quantities are extended implicitly by means of ghost penalty terms to a larger domain, which is needed in the following step of the time-stepping scheme.

In particular, we have shown optimal-order error estimates for the $L^2(H^1)$-semi-norm and the $L^2(L^2)$-norm error for the velocities. The main difficulties herein consisted in the transfer of quantities between domains Ω^n and Ω^{n-1} at different

Table 3 Errors for the fully discrete solutions for P_2 finite elements and BDF(2) for different mesh and time-step sizes with analytical values for the initialisation

$h \downarrow \backslash \Delta t \rightarrow$	0.4	0.2	0.1	0.05	g_h	$\text{eoc}_{\Delta t}$
$\|\boldsymbol{u} - \boldsymbol{u}_{kh}\|_Q / \|\boldsymbol{u}\|_Q$						
1/2	8.91×10^{-3}	8.89×10^{-3}	8.89×10^{-3}	8.89×10^{-3}	8.89×10^{-3}	3.47
1/4	1.44×10^{-3}	1.35×10^{-3}	1.35×10^{-3}	1.35×10^{-3}	1.35×10^{-3}	5.04
1/8	6.38×10^{-4}	2.77×10^{-4}	2.33×10^{-4}	2.29×10^{-4}	2.28×10^{-4}	3.06
$g_{\Delta t}$	5.41×10^{-4}	9.90×10^{-5}	3.87×10^{-5}	3.32×10^{-5}	0	2.31
eoc_h	3.22	2.81	2.75	2.75	2.71	
$\|\nabla(\boldsymbol{u} - \boldsymbol{u}_{kh})\|_Q / \|\nabla \boldsymbol{u}\|_Q$						
1/2	5.60×10^{-2}	5.60×10^{-2}	5.60×10^{-2}	5.60×10^{-2}	5.60×10^{-2}	2.74
1/4	1.56×10^{-2}	1.56×10^{-2}	1.56×10^{-2}	1.56×10^{-2}	1.56×10^{-2}	3.60
1/8	4.23×10^{-3}	4.06×10^{-3}	4.05×10^{-3}	4.05×10^{-3}	4.05×10^{-3}	3.91
$g_{\Delta t}$	-2.23×10^{-4}	-5.54×10^{-4}	-5.74×10^{-4}	-5.74×10^{-4}	0	–
eoc_h	1.83	1.81	1.81	1.81	1.86	
$\|p - p_{kh}\|_Q / \|p\|_Q$						
1/2	7.78×10^{-3}	3.32×10^{-3}	2.96×10^{-3}	2.95×10^{-3}	2.94×10^{-3}	3.67
1/4	7.27×10^{-3}	1.86×10^{-3}	9.41×10^{-4}	8.91×10^{-4}	8.29×10^{-4}	2.66
1/8	7.27×10^{-3}	1.79×10^{-3}	5.34×10^{-4}	3.28×10^{-4}	2.28×10^{-4}	2.18
$g_{\Delta t}$	7.27×10^{-3}	1.79×10^{-3}	4.31×10^{-4}	1.16×10^{-4}	0	1.97
eoc_h	10.01	4.38	2.31	1.87	1.83	
$\|\nabla(p - p_{kh})\|_Q / \|\nabla p\|_Q$						
1/2	3.43×10^{-2}	3.31×10^{-2}	3.29×10^{-2}	3.29×10^{-2}	3.29×10^{-2}	3.35
1/4	1.96×10^{-2}	1.81×10^{-2}	1.80×10^{-2}	1.80×10^{-2}	1.80×10^{-2}	4.04
1/8	1.86×10^{-2}	1.71×10^{-2}	1.70×10^{-2}	1.70×10^{-2}	1.70×10^{-2}	4.05
$g_{\Delta t}$	1.85×10^{-2}	1.70×10^{-2}	1.69×10^{-2}	1.69×10^{-2}	0	–
eoc_h	3.88	3.91	3.90	3.90	–	

The *experimental orders of convergence* (eoc) have been computed as in Table 1

time-steps and in the estimation of the pressure error. Optimal $L^2(H^1)$-norm errors for the pressure can be derived under the inverse CFL conditions $\Delta t \geq ch^2$ for the CIP pressure stabilisation and BDF(1) ($\Delta t \geq ch$ for BDF(2)), or unconditionally, when the Brezzi-Pitkäranta pressure stabilisation is used. Fortunately, these estimates are sufficient to show optimal bounds for the velocities in both the $L^2(H^1)$- and the $L^2(L^2)$-norms. All these estimates are in good agreement with the numerical results presented.

For the $L^2(L^2)$-norm error of the pressure, we have shown suboptimal bounds in terms of the time step Δt. The derivation of optimal bounds seems to be non-trivial and

Table 4 Errors for the fully discrete solutions for P_2 and BDF(2), when one BDF(1) step is used for initialisation

$h \downarrow \backslash \Delta t \rightarrow$	0.4	0.2	0.1	0.05	g_h	$\text{eoc}_{\Delta t}$
$\|\boldsymbol{u} - \boldsymbol{u}_{kh}\|_Q / \|\boldsymbol{u}\|_Q$						
1/2	8.95×10^{-3}	8.90×10^{-3}	8.89×10^{-3}	8.89×10^{-3}	8.89×10^{-3}	3.23
1/4	1.75×10^{-3}	1.39×10^{-3}	1.36×10^{-3}	1.35×10^{-3}	1.35×10^{-3}	3.27
1/8	1.20×10^{-3}	4.29×10^{-4}	2.54×10^{-4}	2.32×10^{-4}	2.17×10^{-4}	2.22
$g_{\Delta t}$	1.15×10^{-3}	2.87×10^{-4}	6.56×10^{-5}	3.76×10^{-5}	0	2.00
eoc_h	3.69	2.96	2.77	2.76	2.71	
$\|\nabla(\boldsymbol{u} - \boldsymbol{u}_{kh})\|_Q / \|\nabla \boldsymbol{u}\|_Q$						
1/2	5.60×10^{-2}	5.60×10^{-2}	5.60×10^{-2}	5.60×10^{-2}	5.60×10^{-2}	–
1/4	1.57×10^{-2}	1.56×10^{-2}	1.56×10^{-2}	1.56×10^{-2}	1.56×10^{-2}	3.31
1/8	4.71×10^{-3}	4.12×10^{-3}	4.06×10^{-3}	4.05×10^{-3}	4.05×10^{-3}	3.27
$g_{\Delta t}$	5.51×10^{-4}	-4.11×10^{-4}	-5.13×10^{-4}	-5.21×10^{-4}	0	–
eoc_h	1.87	1.82	1.81	1.81	1.86	
$\|p - p_{kh}\|_Q / \|p\|_Q$						
1/2	2.40×10^{-2}	8.33×10^{-3}	3.90×10^{-3}	3.07×10^{-3}	2.55×10^{-3}	1.91
1/4	2.36×10^{-2}	7.79×10^{-3}	2.68×10^{-3}	1.23×10^{-3}	4.52×10^{-4}	1.66
1/8	2.36×10^{-2}	7.79×10^{-3}	2.58×10^{-3}	9.09×10^{-4}	6.89×10^{-5}	1.61
$g_{\Delta t}$	2.36×10^{-2}	7.79×10^{-3}	2.57×10^{-3}	9.04×10^{-4}	0	1.60
eoc_h	4.17	6.29	3.53	6.16	2.51	
$\|\nabla(p - p_{kh})\|_Q / \|\nabla p\|_Q$						
1/2	4.14×10^{-2}	3.40×10^{-2}	3.31×10^{-2}	3.29×10^{-2}	3.29×10^{-2}	2.98
1/4	2.99×10^{-2}	1.96×10^{-2}	1.82×10^{-2}	1.80×10^{-2}	1.80×10^{-2}	2.83
1/8	2.92×10^{-2}	1.87×10^{-2}	1.72×10^{-2}	1.71×10^{-2}	1.70×10^{-2}	2.80
$g_{\Delta t}$	2.91×10^{-2}	1.87×10^{-2}	1.72×10^{-2}	1.70×10^{-2}	0	–
eoc_h	3.96	4.01	3.97	4.61	–	

The *experimental orders of convergence* (eoc) have been computed as in Table 1

needs to be investigated in future work. Moreover, it would be interesting to further investigate if the exponential growth in the stability and error estimates can indeed be observed in numerical computations, for example by considering more complex domain motions.

Further directions of research are the application of the approach to the non-linear Navier–Stokes equations, multi-phase flows and fluid-structure interactions, as well as the investigation of different time-stepping schemes, such as Crank–Nicolson or the fractional-step θ scheme within the framework presented and investigated in the present work.

Acknowledgements The first author acknowledges support by the EPSRC Grant EP/P01576X/1. The second author was supported by the DFG Research Scholarship FR3935/1-1. The work of the third author was funded by the Swedish Research Council under Starting Grant 2017-05038.

Funding Open Access funding enabled and organized by Projekt DEAL.

Appendix: Proof of Lemma 2.2

Proof Our proof is similar to the one given in [6] for the non-linear Navier–Stokes equations. As usual, we start by showing existence and uniqueness for the velocities $\boldsymbol{u}$ by considering a reduced problem in the space of divergence-free trial and test functions

$$\begin{aligned} \mathcal{V}_0(t) &:= \{\boldsymbol{u} \in \mathcal{V}(t),\ \operatorname{div} \boldsymbol{u} = 0 \text{ a.e. in } \Omega(t)\}, \\ \mathcal{V}_{0,I} &:= \{\boldsymbol{u} \in L^2(I, \mathcal{V}_0(t)),\ \partial_t \boldsymbol{u} \in L^2(I, \mathcal{L}(t)^d)\}. \end{aligned} \tag{119}$$

The reduced problem is given by: *Find* $\boldsymbol{u} \in \mathcal{V}_{0,I}$ such that

$$(\partial_t \boldsymbol{u}, \boldsymbol{v})_{\Omega(t)} + (\nabla \boldsymbol{u}, \nabla \boldsymbol{v})_{\Omega(t)} = (\boldsymbol{f}, \boldsymbol{v})_{\Omega(t)} \quad \forall \boldsymbol{v} \in \mathcal{V}_0(t) \qquad \text{a.e. in } t \in I, \tag{120}$$

$$\boldsymbol{u}(x, 0) = \boldsymbol{u}^0(x) \qquad \text{a.e. in } \Omega(0). \tag{121}$$

It can be easily seen that $\boldsymbol{u} \in \mathcal{V}_{0,I}$ is a solution to (120) if and only if it is the velocity part of a solution to (5).

(i) *Transformation* By means of the map $\boldsymbol{T}$ in (2), we can transform the system of equations to an equivalent system on $\Omega(0)$: *Find* $\hat{\boldsymbol{u}} \in \hat{\mathcal{V}}_{0,I}$ such that

$$\begin{aligned} \left(J(t)(\partial_t \hat{\boldsymbol{u}} - F^{-1}(t)\partial_t \boldsymbol{T}(t) \cdot \hat{\nabla}\hat{\boldsymbol{u}}), \hat{v}\right)_{\Omega(0)} &+ (J(t)\hat{\nabla}\hat{\boldsymbol{u}}F(t)^{-1}, \hat{\nabla}\hat{\boldsymbol{v}}F(t)^{-1})_{\Omega(0)} \\ &= (J(t)\hat{\boldsymbol{f}}, \hat{\boldsymbol{v}})_{\Omega(0)} \quad \forall \hat{\boldsymbol{v}} \in \hat{V}_0 \qquad \text{a.e. in } t \in I, \\ \hat{\boldsymbol{u}}(x, 0) &= \hat{\boldsymbol{u}}^0(x) \qquad \text{a.e. in } \Omega(0), \end{aligned} \tag{122}$$

where $F = \hat{\nabla}\boldsymbol{T}$, $J = \det F$, $\hat{\nabla}$ denotes derivatives with respect to $\Omega(0)$ and quantities with a "hat" correspond to their counterparts without a hat by the relation

$$\hat{\boldsymbol{u}}(x, t) = \boldsymbol{u}(\boldsymbol{T}(x, t), t) \quad \text{for } x \in \Omega(0).$$

Test and trial spaces are defined as

$$\hat{\mathcal{V}}_0(t) := \{\hat{\boldsymbol{u}} \in \mathcal{V}(0),\ \hat{\mathrm{div}}(J(t)F(t)^{-1}\hat{\boldsymbol{u}}) = 0 \text{ a.e. in } \Omega(0)\},$$
$$\hat{\mathcal{V}}_{0,I} := \{\hat{\boldsymbol{u}} \in L^2(I, \hat{\mathcal{V}}_0(t)),\ \partial_t\hat{\boldsymbol{u}} \in L^2(I, \mathcal{L}(0)^d)\}.$$

Given that $\boldsymbol{T}$ is a $W^{1,\infty}$-diffeomorphism, it can be shown that [23]

$$\boldsymbol{u} \in \mathcal{V}_0(t) \ \Leftrightarrow\ \hat{\boldsymbol{u}} \in \hat{\mathcal{V}}_0(t), \quad \boldsymbol{u} \in \mathcal{V}_{0,I} \ \Leftrightarrow\ \hat{\boldsymbol{u}} \in \hat{\mathcal{V}}_{0,I}.$$

We will show the well-posedness of (122) by a Galerkin argumentation. A basis $\{\hat{\boldsymbol{w}}_j\}_{j\in\mathbb{N}}$ of the time-dependent space $\hat{\mathcal{V}}_0(t)$ is given by the inverse Piola transform of an L^2-orthonormal basis $\{\hat{\phi}_j\}_{j\in\mathbb{N}}$ of the space $\hat{\mathcal{V}}_0(0)$

$$\hat{\boldsymbol{w}}_j(t) = J(t)^{-1}F(t)\hat{\phi}_j, \quad j \in \mathbb{N}.$$

Under the given regularity assumptions on the domain movement T, the basis functions lie in $W^{1,\infty}(I, H^1(\Omega(0))^d)$.

(ii) *Galerkin approximation* The ansatz

$$\hat{\boldsymbol{u}}_l = \sum_{j=1}^{l} \alpha_j(t)\hat{\boldsymbol{w}}_j(t)$$

with coefficients $\alpha_j(t) \in \mathbb{R}$ leads to the Galerkin problem

$$\begin{aligned}\Big(J(t)(\partial_t\hat{\boldsymbol{u}}_l - F^{-1}(t)\partial_t\boldsymbol{T}(t)\cdot\hat{\nabla}\hat{\boldsymbol{u}}_l), \hat{\boldsymbol{w}}_k\Big)_{\Omega(0)} &+ (J(t)\hat{\nabla}\hat{\boldsymbol{u}}_l F(t)^{-1}, \hat{\nabla}\hat{\boldsymbol{w}}_k F(t)^{-1})_{\Omega(0)} \\ &= (J(t)\hat{\boldsymbol{f}}, \hat{\boldsymbol{w}}_k)_{\Omega(0)} \quad k = 1, \ldots, l, \\ \hat{\boldsymbol{u}}_l(x, 0) &= \hat{\boldsymbol{u}}_l^0(x) \quad \text{a.e. in } \Omega(0),\end{aligned} \tag{123}$$

where $\hat{\boldsymbol{u}}_l^0$ is an L^2-orthogonal projection of $\hat{\boldsymbol{u}}^0$ onto $\mathrm{span}\{\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_l\}$. This is a system of ordinary differential equation for the coefficients $\alpha_j(t)$, $j = 1, \ldots, l$

$$\begin{aligned}\sum_{j=1}^{l} \alpha_j'(t) \underbrace{\big(J(t)\hat{\boldsymbol{w}}_j, \hat{\boldsymbol{w}}_k\big)_{\Omega(0)}}_{M(t)} &+ \alpha_j(t) \underbrace{\Big(J(t)(\partial_t\hat{\boldsymbol{w}}_j - F^{-1}(t)\partial_t\boldsymbol{T}(t)\cdot\hat{\nabla}\hat{\boldsymbol{w}}_j), \hat{\boldsymbol{w}}_k\Big)_{\Omega(0)}}_{B(t)} \\ &+ \alpha_j(t) \underbrace{(J(t)\hat{\nabla}\hat{\boldsymbol{w}}_j F(t)^{-1}, \hat{\nabla}\hat{\boldsymbol{w}}_k F(t)^{-1})_{\Omega(0)}}_{A(t)} = \underbrace{(J(t)\hat{\boldsymbol{f}}, \hat{\boldsymbol{w}}_k)_{\Omega(0)}}_{b(t)} \quad k = 1, \ldots, l.\end{aligned} \tag{124}$$

The assumption that T describes a $W^{1,\infty}(\Omega(0))$ diffeomorphism implies that

$$0 < J_{\min} < J(t) < J_{\max} < \infty, \qquad J_{\min}, J_{\max} \in \mathbb{R}.$$

It follows that the matrix $M(t)$ is invertible for all $t \in I$ and we can write (124) as

$$\alpha' = -M(t)^{-1}(A(t) + B(t))\alpha + M(t)^{-1}b(t). \tag{125}$$

Due to the time regularity of the basis functions $\partial_t \hat{\boldsymbol{w}}_j = \partial_t(J^{-1}F)\hat{\phi}_j \in L^\infty(I, H^1(\Omega(0))^d)$ the right-hand side in (125) is Lipschitz. Hence, the Picard-Lindelöf theorem guarantees a unique solution to (123).

(iii) *A priori estimate* We test (122) with $\hat{\boldsymbol{w}} = \hat{\boldsymbol{u}}_l$. After some basic calculus, we obtain the system

$$\begin{aligned}\left(\partial_t(J^{1/2}\hat{\boldsymbol{u}}_l), J^{1/2}\hat{\boldsymbol{u}}_l\right)_{\Omega(0)} - \left(\partial_t(J^{1/2})\hat{\boldsymbol{u}}_l + J^{1/2}F^{-1}\partial_t\boldsymbol{T}\cdot\hat{\nabla}\hat{\boldsymbol{u}}_l, J^{1/2}\hat{\boldsymbol{u}}_l\right)_{\Omega(0)} \\ + (J^{1/2}\hat{\nabla}\hat{\boldsymbol{u}}_l F^{-1}, J^{1/2}\hat{\nabla}\hat{\boldsymbol{u}}_l F^{-1})_{\Omega(0)} = (J^{1/2}\hat{\boldsymbol{f}}, J^{1/2}\hat{\boldsymbol{u}}_l)_{\Omega(0)},\end{aligned}$$

where we have skipped the dependencies of J and F on time for better readability. Integration in time gives the estimate

$$\begin{aligned}\left\|J^{1/2}(t_{\text{fin}})\hat{\boldsymbol{u}}_l(t_{\text{fin}})\right\|^2_{\Omega(0)} + \int_0^{t_{\text{fin}}} \left\|J^{1/2}\hat{\nabla}\hat{\boldsymbol{u}}_l F^{-1}\right\|^2_{\Omega(0)} dt \\ \leq \left\|\hat{\boldsymbol{u}}_l(0)\right\|^2_{\Omega(0)} + c\|\partial_t\boldsymbol{T}\|_{W^{1,\infty}(\Omega(0))}\int_0^{t_{\text{fin}}} \|J^{1/2}\hat{\boldsymbol{u}}_l\|^2_{\Omega(0)}\, dt \\ + c\int_0^{t_{\text{fin}}} \left\|J^{1/2}\hat{\boldsymbol{f}}\right\|^2_{\Omega(0)} dt.\end{aligned}$$

Using Gronwall's lemma, we obtain the first a priori estimate

$$\begin{aligned}\left\|J^{1/2}(t_{\text{fin}})\hat{\boldsymbol{u}}_l(t_{\text{fin}})\right\|^2_{\Omega(0)} + \int_0^{t_{\text{fin}}} \left\|J^{1/2}\hat{\nabla}\hat{\boldsymbol{u}}_l F^{-1}\right\|^2_{\Omega(0)} dt \\ \leq c\exp\left(c\|\partial_t\boldsymbol{T}\|_{W^{1,\infty}(\Omega(0))}t_{\text{fin}}\right)\left(\left\|\hat{\boldsymbol{u}}_l(0)\right\|^2_{\Omega(0)} + c\int_0^{t_{\text{fin}}} \left\|J^{1/2}\hat{\boldsymbol{f}}\right\|^2_{\Omega(0)} dt.\right)\end{aligned} \tag{126}$$

This implies that $\hat{\boldsymbol{u}}_l$ is bounded in $L^\infty(I, L^2(\Omega(0))^d)$ and $L^2(I, \hat{\mathcal{V}}_0)$. This implies the existence of convergent subsequences and limit functions $\hat{u}, \hat{u}^*$ in the following sense

$$\begin{aligned}&\hat{\boldsymbol{u}}_{l'} \to \hat{\boldsymbol{u}}^* \quad \text{weak star in } L^\infty(I, L^2(\Omega(0))^d), \\ &\hat{\boldsymbol{u}}_{l'} \to \hat{\boldsymbol{u}} \quad \text{weakly in } L^2(I, \hat{\mathcal{V}}_0)) \text{ and strongly in } L^2(I, L^2(\Omega(0))^d).\end{aligned} \tag{127}$$

It is not difficult to prove that $\hat{\boldsymbol{u}}^* = \hat{\boldsymbol{u}}$, see [58], Section III.1.3.

(iv) *A priori estimate for the time derivative* In principle, we would like to test (123) with $\partial_t \hat{\boldsymbol{u}}_l$. Unfortunately, this is not possible, as in general $\partial_t \hat{\boldsymbol{u}}_l \notin \operatorname{span}(\hat{\boldsymbol{w}}_1, \dots, \hat{\boldsymbol{w}}_l)$ due to the time-dependence of the basis functions. Instead, we can test with $J^{-1} F^T \partial_t (J F^{-T} \hat{\boldsymbol{u}}_l)$, as

$$\begin{aligned} J^{-1} F \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l) &= \sum_{j=1}^{l} J^{-1} F \partial_t (J F^{-1} \alpha_j \hat{\boldsymbol{w}}_j) = \sum_{j=1}^{l} J^{-1} F \partial_t (\alpha_j \hat{\boldsymbol{\phi}}_j) \\ &= \sum_{j=1}^{l} \alpha_j' J^{-1} F \hat{\boldsymbol{\phi}}_j = \sum_{j=1}^{l} \alpha_j' \hat{\boldsymbol{w}}_j. \end{aligned}$$

We obtain

$$\begin{aligned} &\left(\partial_t \hat{\boldsymbol{u}}_l, F \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l)\right)_{\Omega(0)} - \left(\partial_t \boldsymbol{T} \cdot \hat{\nabla} \hat{\boldsymbol{u}}_l, \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l)\right)_{\Omega(0)} \\ &\qquad + (J \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1}, \hat{\nabla} (J^{-1} F \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l) F^{-1})_{\Omega(0)} = (\hat{\boldsymbol{f}}, F \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l))_{\Omega(0)}. \end{aligned} \tag{128}$$

The third term on the left-hand side is well-defined under the regularity assumptions stated, as $J F^{-1}$ is the cofactor matrix to F, which can be written in terms of T. Using the product rule, we see that the first term on the left-hand side is bounded below by

$$\begin{aligned} \left(\partial_t \hat{\boldsymbol{u}}_l, F \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l)\right)_{\Omega(0)} &= \left(\partial_t \hat{\boldsymbol{u}}_l, J \partial_t \hat{\boldsymbol{u}}_l\right)_{\Omega(0)} + \left(\partial_t \hat{\boldsymbol{u}}_l, F \partial_t (J F^{-1}) \hat{\boldsymbol{u}}_l\right)_{\Omega(0)} \\ &\geq \left\| J^{1/2} \partial_t \hat{\boldsymbol{u}}_l \right\|_{\Omega(0)}^2 - c(\boldsymbol{T}) \left\| J^{1/2} \partial_t \hat{\boldsymbol{u}}_l \right\|_{\Omega(0)} \left\| J^{1/2} \hat{\boldsymbol{u}}_l \right\|_{\Omega(0)} \end{aligned}$$

For the third term on the left-hand side, we have

$$\begin{aligned} &(J \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1}, \hat{\nabla} (J^{-1} F \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l) F^{-1})_{\Omega(0)} \\ &\quad = (J \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1}, \hat{\nabla} \partial_t \hat{\boldsymbol{u}}_l F^{-1})_{\Omega(0)} + (J \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1}, \hat{\nabla} \left(J^{-1} F \partial_t (J F^{-1}) \hat{\boldsymbol{u}}_l\right) F^{-1})_{\Omega(0)} \\ &\quad = \left(J^{1/2} \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1}, \hat{\partial}_t \left(J^{1/2} \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1}\right)\right)_{\Omega(0)} - \left(J^{1/2} \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1}, \hat{\nabla} \left(\hat{\boldsymbol{u}}_l \partial_t (J^{1/2} F^{-1})\right)\right)_{\Omega(0)} \\ &\qquad + (J \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1}, \hat{\nabla} \left(J^{-1} F \partial_t (J F^{-1}) \hat{\boldsymbol{u}}_l\right) F^{-1})_{\Omega(0)} \\ &\quad \geq \frac{1}{2} \partial_t \left\| J^{1/2} \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1} \right\|_{\Omega(0)}^2 - c(\boldsymbol{T}) \left\| J^{1/2} \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1} \right\|_{\Omega(0)}^2 \end{aligned}$$

Using a similar argumentation and Young's inequality, we can show the bounds

$$\begin{aligned} \left(\partial_t \boldsymbol{T} \cdot \hat{\nabla} \hat{\boldsymbol{u}}_l, \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l)\right)_{\Omega(0)} &\leq c(\boldsymbol{T}) \left\| J^{1/2} \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1} \right\|_{\Omega(0)}^2 + \frac{1}{4} \left\| J^{1/2} \partial_t \hat{\boldsymbol{u}}_l \right\|_{\Omega(0)}^2 \\ \left(\hat{\boldsymbol{f}}, F \partial_t (J F^{-1} \hat{\boldsymbol{u}}_l)\right)_{\Omega(0)} &\leq c(\boldsymbol{T}) \left(\left\| J^{1/2} \hat{\boldsymbol{f}} \right\|_{\Omega(0)}^2 + \left\| J^{1/2} \hat{\nabla} \hat{\boldsymbol{u}}_l F^{-1} \right\|_{\Omega(0)}^2 \right) \\ &\quad + \frac{1}{4} \left\| J^{1/2} \partial_t \hat{\boldsymbol{u}}_l \right\|_{\Omega(0)}^2. \end{aligned}$$

Integration over $t \in I$ in (123) gives the estimate

$$\left\| J^{1/2}(t_{\text{fin}})\hat{\nabla}\hat{\boldsymbol{u}}_l(t_{\text{fin}})F^{-1}(t_{\text{fin}}) \right\|^2_{\Omega(0)} + \int_0^{t_{\text{fin}}} \left\| J^{1/2}\partial_t\hat{\boldsymbol{u}}_l \right\|^2_{\Omega(0)} dt$$
$$\leq \left\| \hat{\nabla}\hat{\boldsymbol{u}}_l(0) \right\|^2_{\Omega(0)} + c(\boldsymbol{T}) \int_0^{t_{\text{fin}}} \left\| J^{1/2}\hat{\boldsymbol{f}} \right\|^2_{\Omega(0)} + \left\| J^{1/2}\hat{\nabla}\hat{\boldsymbol{u}}_l F^{-1} \right\|^2_{\Omega(0)} dt.$$

Using Gronwall's lemma we obtain

$$\left\| J^{1/2}(t_{\text{fin}})\hat{\nabla}\hat{\boldsymbol{u}}_l(t_{\text{fin}})F^{-1}(t_{\text{fin}}) \right\|^2_{\Omega(0)} + \int_0^{t_{\text{fin}}} \left\| J^{1/2}\partial_t\hat{\boldsymbol{u}}_l \right\|^2_{\Omega(0)} dt$$
$$\leq c\exp(c(\boldsymbol{T})t_{\text{fin}}) \left(\left\| \hat{\nabla}\hat{\boldsymbol{u}}_l(0) \right\|^2_{\Omega(0)} + \int_0^{t_{\text{fin}}} \left\| J^{1/2}\hat{\boldsymbol{f}} \right\|^2_{\Omega(0)} dt \right).$$

This shows the boundedness of $\partial_t\hat{\boldsymbol{u}}_l$ in $L^2(I, L^2(\Omega(0))^d)$ and the convergence of a subsequence (see Temam [58], Proposition III.1.2, for the details)

$$\partial_t\hat{\boldsymbol{u}}_{l'} \to \partial_t\hat{\boldsymbol{u}} \qquad \text{weakly in } L^2(I, L^2(\Omega(0))^d). \tag{129}$$

(v) *Conclusion* The a priori bounds shown in (ii) and (iii) and the resulting convergence behaviour allows us to pass to the limit $l \to \infty$ in (123). The convergences (127) and (129) imply that $\hat{\boldsymbol{u}}_l(x, 0) \to \hat{\boldsymbol{u}}^0$ $(l \to \infty)$. We find that the limit $\hat{u}$ is a solution to (122). Uniqueness is easily proven by testing (122) with $\hat{v} = \hat{u}$ and the a priori estimate (126). Due to the equivalence of (122) and (120), the pullback $u = \hat{u} \circ T^{-1} \in \mathcal{V}_{0,I}$ is the unique solution to (120).

(vi) *Pressure* Finally, the unique existence of a pressure for a.e. $t \in I$ follows by showing the existence of a weak pressure gradient that fulfils

$$\text{grad } p(t) = \boldsymbol{f}(t) + \Delta\boldsymbol{u}(t) - \partial_t\boldsymbol{u}(t) \quad \text{in } \Omega(t)$$

using the de Rham theorem. We refer to [58], Proposition III.1.2, for the details. □

References

1. Alnæs, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, ME., Wells, G.N.: The FEniCS project version 1.5. Arch. Numer. Soft, **3**(100) (2015)
2. Baker, G.A., Dougalis, V.A., Karakashian, O.A.: On a higher order accurate fully discrete Galerkin approximation to the Navier–Stokes equations. Math. Comput. **39**(160), 339–375 (1982)
3. Bazilevs, Y., Takizawa, K., Tezduyar, T.E.: Computational Fluid–Structure Interaction: Methods and Applications. Wiley, Hoboken (2013)
4. Besier, M., Wollner, W.: On the pressure approximation in nonstationary incompressible flow simulations on dynamically varying spatial meshes. Int. J. Numer. Methods Fluids **69**(6), 1045–1064 (2012)
5. Bochev, P.B., Gunzburger, M.D., Lehoucq, R.B.: On stabilized finite element methods for the stokes problem in the small time step limit. Int. J. Numer. Methods Fluids **53**(4), 573–597 (2007)
6. Bock, D.N.: On the Navier–Stokes equations in noncylindrical domains. J. Differ. Equ. **25**(2), 151–162 (1977)

7. Boffi, D., Gastaldi, L.: Stability and geometric conservation laws for ALE formulations. Comput. Methods Appl. Mech. Eng. **193**(42–44), 4717–4739 (2004)
8. Braack, M., Lang, J., Taschenberger, N.: Stabilized finite elements for transient flow problems on varying spatial meshes. Comput. Methods Appl. Mech. Eng. **253**, 106–116 (2013)
9. Brezzi, F., Pitkäranta, J.: On the Stabilization of Finite Element Approximations of the Stokes Equations, pp. 11–19. Vieweg+Teubner (1984)
10. Burman, E.: Ghost penalty. C. R. Math. **348**(21–22), 1217–1220 (2010)
11. Burman, E., Fernández, M.A.: Continuous interior penalty finite element method for the time-dependent Navier–Stokes equations: space discretization and convergence. Numer. Math. **107**(1), 39–77 (2007)
12. Burman, E., Fernández, M.A.: Galerkin finite element methods with symmetric pressure stabilization for the transient Stokes equations: stability and convergence analysis. SIAM J. Numer. Anal. **47**(1), 409–439 (2008)
13. Burman, E., Fernández, M.A.: An unfitted Nitsche method for incompressible fluid–structure interaction using overlapping meshes. Comput. Methods Appl. Mech. Eng. **279**, 497–514 (2014)
14. Burman, E., Hansbo, P.: Edge stabilization for the generalized Stokes problem: a continuous interior penalty method. Comput. Methods Appl. Mech. Eng. **195**(19), 2393–2410 (2006)
15. Burman, E., Hansbo, P.: Fictitious domain methods using cut elements: III. A stabilized Nitsche method for Stokes' problem. ESAIM: M2AN, **48**(3), 859–874 (2014)
16. Burman, E., Claus, S., Hansbo, P., Larson, M.G., Massing, A.: CutFEM: discretizing geometry and partial differential equations. Int. J. Numer. Methods Eng. **104**(7), 472–501 (2015)
17. Burman, E., Claus, S., Massing, A.: A stabilized cut finite element method for the three field Stokes problem. SIAM J. Sci. Comput. **37**(4), A1705–A1726 (2015)
18. Burman, E., Fernández, M.A., Frei, S.: A Nitsche-based formulation for fluid–structure interactions with contact. arXiv:1808.08758 (2018)
19. Caucha, L.J., Frei, S., Rubio, O.: Finite element simulation of fluid dynamics and CO_2 gas exchange in the alveolar sacs of the human lung. Comput. Appl. Math. **37**(5), 6410–6432 (2018)
20. Chessa, J., Belytschko, T.: An extended finite element method for two-phase fluids. J. Appl. Mech. **70**(1), 10–17 (2003)
21. Codina, R., Houzeaux, G., Coppola-Owen, H., Baiges, J.: The fixed-mesh ALE approach for the numerical approximation of flows in moving domains. J. Comput. Phys. **228**(5), 1591–1611 (2009)
22. Donea, J., Huerta, A., Ponthot, J.P., Rodríguez-Ferran, A.: Arbitrary Lagrangian–Eulerian Methods. Wiley, Hoboken (2004)
23. Failer, L.: Optimal Control of Time-Dependent Nonlinear Fluid–Structure Interaction. Ph.D. thesis, Technische Universität München (2017)
24. Frank, R.M., Lazarus, R.B.: Mixed Eulerian–Lagrangian method. In: Methods in Computational Physics, vol. 3: Fundamental Methods in Hydrodynamics, pp. 47–67. Acad Press (1964)
25. Frei, S.: Eulerian Finite Element Methods for Interface Problems and Fluid–Structure Interactions. Ph.D. thesis, Heidelberg University (2016). http://www.ub.uni-heidelberg.de/archiv/21590
26. Frei, S.: An edge-based pressure stabilization technique for finite elements on arbitrarily anisotropic meshes. Int. J. Numer. Methods Fluids **89**(10), 407–429 (2019)
27. Frei, S., Richter, T.: A locally modified parametric finite element method for interface problems. SIAM J. Numer. Anal. **52**(5), 2315–2334 (2014)
28. Frei, S., Richter, T.: A second order time-stepping scheme for parabolic interface problems with moving interfaces. ESAIM: M2AN **51**(4), 1539–1560 (2017)
29. Frei, S., Holm, B., Richter, T., Wick, T., Yang, H. (eds.): Fluid–Structure Interaction, Modeling, Adaptive Discretisations and Solvers, Rad Ser Comput Appl Math, 2017. De Gruyter
30. Girault, V., Raviart, P.A.: Finite element approximation of the Navier–Stokes equations. Lect Notes Math, p. 749. Springer (1979)
31. Groß, S., Reusken, A.: An extended pressure finite element space for two-phase incompressible flows with surface tension. J. Comput. Phys. **224**(1), 40–58 (2007)
32. Gürkan, C., Massing, A.: A stabilized cut discontinuous Galerkin framework for elliptic boundary value and interface problems. Comput. Methods Appl. Mech. Eng. **348**, 466–499 (2019)
33. Guzmán, J., Olshanskii, M.: Inf-sup stability of geometrically unfitted stokes finite elements. Math. Comput. **87**(313), 2091–2112 (2018)
34. Hairer, E., Nørsett, S.P., Wanner, G.: Solving Ordinary differential equations. 1, Nonstiff problems. Springer, Berlin (1991)

35. Hansbo, A., Hansbo, P.: An unfitted finite element method, based on Nitsche's method, for elliptic interface problems. Comput. Methods Appl. Mech. Eng. **191**(47–48), 5537–5552 (2002)
36. Hansbo, P., Larson, M.G., Zahedi, S.: A cut finite element method for a Stokes interface problem. Appl. Numer. Math. **85**, 90–114 (2014)
37. Hansbo, P., Larson, M.G., Zahedi, S.: A cut finite element method for coupled bulk-surface problems on time-dependent domains. Comput. Methods Appl. Mech. Eng. **307**, 96–116 (2016)
38. Hecht, F., Pironneau, O.: An energy stable monolithic Eulerian fluid–structure finite element method. Int. J. Numer. Methods Fluids **85**(7), 430–446 (2017)
39. Heywood, J., Rannacher, R.: Finite-element approximation of the nonstationary Navier–Stokes problem. Part IV: error analysis for second-order time discretization. SIAM J. Numer. Anal. **27**(2), 353–384 (1990)
40. Kelley, C, Williams, T.: many others. Gnuplot 4.6: an interactive plotting program. http://gnuplot.sourceforge.net/ (2013)
41. Lehrenfeld, C.: The Nitsche XFEM-DG space-time method and its implementation in three space dimensions. SIAM J. Sci. Comput. **37**(1), A245–A270 (2015)
42. Lehrenfeld, C., Olshanskii, M.: An Eulerian finite element method for PDEs in time-dependent domains. ESAIM: M2AN **53**(2), 585–614 (2019)
43. Lehrenfeld, C., Olshanskii, M.A., Xu, X.: A stabilized trace finite element method for partial differential equations on evolving surfaces. SIAM J. Numer. Anal. **56**(3), 1643–1672 (2018)
44. Massing, A.: Analysis and Implementation of Finite Element Methods on Overlapping and Fictitious Domains. Ph.D. thesis, Department of Informatics, University of Oslo (2012)
45. Massing, A., Larson, M.G., Logg, A., Rognes, M.E.: A stabilized Nitsche fictitious domain method for the Stokes problem. J. Sci. Comput. **61**(3), 604–628 (2014)
46. Massing, A., Schott, B., Wall, W.A.: A stabilized Nitsche cut finite element method for the Oseen problem. Comput. Methods Appl. Mech. Eng. **328**, 262–300 (2018)
47. Moës, N., Dolbow, J., Belytschko, T.: A finite element method for crack growth without remeshing. Int. J. Numer. Methods Eng. **46**, 131–150 (1999)
48. Nitsche, J.A.: Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. Abh Math. Univ. Hamburg **36**, 9–15 (1970)
49. Nobile, F., Formaggia, L.: A stability analysis for the arbitrary Lagrangian Eulerian formulation with finite elements. East-West J. Numer. Math. **7**, 105–132 (1999)
50. Noh, W.F.: CEL: a time-dependent, two-space-dimensional, coupled Eulerian-Lagrange code. In: Methods Comput Phys, vol.3: Fundamental Methods in Hydrodynamics, pp. 117–179. Acad Press (1964)
51. Olshanskii, M.A., Xu, X.: A trace finite element method for PDEs on evolving surfaces. SIAM J. Sci. Comput. **39**(4), A1301–A1319 (2017)
52. Preuß, J.: Higher order unfitted isoparametric space-time FEM on moving domains (2018)
53. Richter, T.: Fluid–Structure Interactions: Models, Analysis and Finite Elements, vol. 118. Springer, Berlin (2017)
54. Richter, T., Wick, T.: On time discretizations of fluid–structure interactions. In: Multiple Shooting and Time Domain Decomposition Methods, pp. 377–400. Springer (2015)
55. Salvi, R.: On the Navier–Stokes equations in non-cylindrical domains: on the existence and regularity. Math. Z. **199**(2), 153–170 (1988)
56. Schott, B.: Stabilized Cut Finite Element Methods for Complex Interface Coupled Flow Problems. Ph.D. thesis, Technische Universität München (2017)
57. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions, vol. 2. Princeton University Press, Princeton (1970)
58. Temam, R.: Navier–Stokes Equations: Theory and Numerical Analysis. Amer Math Soc, Providence (2000)

An implicitly extended Crank-Nicolson scheme for the heat equation on a time-dependent domain

Stefan Frei [*], Maneesh Kumar Singh [†]

Abstract

We consider a time-stepping scheme of Crank-Nicolson type for the heat equation on a moving domain in Eulerian coordinates. As the spatial domain varies between subsequent time steps, an extension of the solution from the previous time step is required. Following Lehrenfeld & Olskanskii [ESAIM: M2AN, 53(2): 585-614, 2019], we apply an implicit extension based on so-called ghost-penalty terms. For spatial discretisation, a cut finite element method is used. We derive a complete a priori error analysis in space and time, which shows in particular second-order convergence in time under a parabolic CFL condition. Finally, we present numerical results in two and three space dimensions that confirm the analytical estimates, even for much larger time steps.

1 Introduction

Partial differential equations (PDEs) posed on moving domains are significant in many areas of science and engineering. They arise for example in flow problems around moving structures, such as pumps [4], wind or water turbines [54], within moving objects [15], or as sub-problems in fluid-structure interactions or multiphase flows. Fluid-structure interactions arise in aerodynamical applications like flow around airplanes or parachutes [58], in biomedical problems such as blood flow through the cardiovascular system [53, 60, 25] or the airflow within the respiratory system [62] and even in tribological applications [46]. Multiphase problems include for instance gas-liquid and particle-laden gas flows [36, 19, 45], rising bubbles [43], droplets in microfluidic devices [17] or the simulation of tumor growth [34]. For further details and applications we refer to the textbooks [55, 3] and [35], respectively.

In this article we consider the time discretisation of a parabolic model problem (namely the heat equation) which is posed on a moving domain $\Omega(t) \subset \mathbb{R}^d\,(d = 2, 3)$ that evolves smoothly in time for $t \in I = [0, t_{max}]$:

$$u_t - \Delta u = f \quad \text{in } \Omega(t), \qquad u = 0 \quad \text{on } \partial\Omega(t), \qquad u(x, 0) = u_0(x) \quad \text{in } \Omega(0). \tag{1.1}$$

In literature, two major numerical approaches can be found for the simulation of partial differential equations on moving domains: the Arbitrary Lagrangian Eulerian (ALE) [21, 22] approach, where the equations are transformed to an arbitrary reference domain which is independent of time, and Eulerian approaches, where the equations are solved in the time-dependent Eulerian framework [23, 27, 48].

[*]Department of Mathematics and Statistics, University of Konstanz, 78457 Konstanz, Germany, stefan.frei@uni-konstanz.de (Corresponding author)

[†]Department of Mathematics, Imperial College London, SW7 2AZ, London, UK, maneesh-kumar.singh@imperial.ac.uk

The ALE approach is a popular technique for the numerical simulation of PDEs on moving domains, in particular for flow problems [42, 41]. For details, we refer to the textbooks [3, 55] and reference cited therein. Convection-diffusion problems on moving domains were, for example, solved in [33, 57] using a stabilised ALE method. The ALE approach is very attractive in the case of moderate domain movements, but shows problems when the shape of the domain changes significantly in time. In particular, topology changes of $\Omega(t)$, as occurring for example in contact problems or considering the separation or union of bubbles can not be modelled by means of an ALE approach [55, 17, 8, 9]. Another example of extreme variations of $\Omega(t)$ are so-called fingering phenomena, which can be frequently observed in multi-phase flows or even for tumor growth [34].

In such cases, a numerical approach that discretises the equations directly in the moving Eulerian coordinate framework is preferable. The Eulerian framework is also the coordinate framework, which is typically used to model flow problems and consequently, in multi-phase flows [35] and fluid-structure interactions with large displacements [23, 10, 27]. However, as the domains $\Omega(t)$ to be discretised vary with time t, additional difficulties arise concerning a proper and accurate discretisation, both in space and in time.

In recent decades, a great amount of works have been contributed concerning the spatial discretisation of curved or moving boundaries by means of finite elements. The techniques can be categorised in *fitted* and *unfitted* finite element methods. In *fitted* methods, the boundary $\partial\Omega(t)$ is resolved in each time step by the finite element mesh [24, 5, 30]. If the domain is time-dependent, this means that new meshes need to be created in each time step. Several approaches have been proposed to alleviate this issue, such as the locally fitted finite element method [30, 29], which is based on a fixed coarse and a variable fine mesh. However, different issues might arise, such as anisotropic fine cells that complicate the numerical discretisation [28].

The idea in *unfitted* finite element methods, on the other hand, is to use the same finite element mesh for all times t, independently of the position of the boundary $\partial\Omega(t)$. A popular approach is the cut finite element method (CutFEM) [7, 13, 38, 51, 37, 64], where cells of the finite element mesh are cut into parts that lie inside $\Omega(t)$ and parts outside for numerical integration. Boundary values are then incorporated weakly by means of Nitsche's method [52]. The method shows similarities to the extended finite element method [20, 16, 32] and the generalised finite element method [2], where the finite element spaces are enriched by suitable functions to account for the position of the boundary.

Much less works can be found in literature concerning a proper time discretisation on moving domains. In the case of moving domains, standard time discretisation based on the method of lines is not directly applicable. The reason is that the domain of definition of the variables changes from time step to time step. As an example consider the finite difference discretisation of the time derivative within a variational formulation ($\Delta t = t_n - t_{n-1}$)

$$(\partial_t u_h(t_n), \phi_h^n)_{\Omega(t_n)} \approx \frac{1}{\Delta t}(u_h(t_n) - u_h(t_{n-1}), \phi_h^n)_{\Omega(t_n)}.$$

The function $u_h(t_{n-1})$ is only well-defined on $\Omega(t_{n-1})$, but is needed on $\Omega(t_n)$.

A possible remedy is to use characteristic-based approaches based on trajectories that follow the motion of the domain, see e.g. [40]. Similar time-stepping schemes result when applying the ALE method only locally within one time step and projecting back to the original reference frame after each step [18], or based on Galerkin time discretisations with modified Galerkin spaces [31]. The disadvantage of these approaches is the necessity for a projection that needs to be computed within each or after a certain number of steps.

A further alternative are space-time approaches [39, 47], where a $d+1$-dimensional domain is discretised. These are, however, computationally demanding, in particular within complex three-

dimensional applications. The implementation of higher-dimensional discretisations and accurate quadrature formulas pose additional challenges. If a discontinuous Galerkin approach is applied in time for the test functions, the formulation decouples in certain time intervals and can be seen as an Eulerian time-stepping scheme [39, 64, 26].

In this work, we follow a slightly different approach first used by Schott [56] and later analysed by Lehrenfeld & Olshanskii [48]. Here, the idea is to define extensions of the solution $u(t_{n-1})$ from previous time steps to a domain $\Omega_\delta(t_{n-1})$ that spans at least $\Omega(t_n)$. On the finite element level these extensions can be incorporated implicitly in the time-stepping scheme by so-called *ghost penalty* stabilisations [6] to a sufficiently large domain. These techniques have originally been proposed to extend the coercivity of elliptic bilinear forms from the physical to the computational domain in the context of CutFEM or fictitious domain approaches [6].

Lehrenfeld & Olshanskii [48] analysed the so extended Backward Euler method in detail for a convection-diffusion problem and gave hints on how to transfer the argumentation to the second-order backward difference scheme (BDF2). Recently, the analysis has been extended to higher order in space and time using an isoparametric finite element approach [49]. In [11, 61], extended BDF time-stepping schemes were applied and analysed for the non-stationary Stokes equations on moving domains.

The reason why only BDF-type time-stepping schemes have been considered in previous works, is that in these schemes spatial derivatives appear only on the "new" time step, i.e. $\nabla u(t_n)$. We will see below that the appearance of additional derivatives on $u(t_{n-1})$ will complicate the error analysis severely. This paper gives a first step towards the analysis of time-stepping schemes that require derivatives at different time instants, such as the Crank-Nicolson method, the Fractional-step-θ-, implicit Runge-Kutta- or Adams-Bashforth schemes.

As a first step, we focus in this work on the popular Crank-Nicolson time-stepping scheme. Up to now, it has been largely open, if and under what conditions a Crank-Nicolson-type scheme can be used within an Eulerian time discretisation on moving domains. We give a detailed stability and convergence analysis. While the analysis requires a strong parabolic CFL condition of type $\Delta t \leq ch^2$, our numerical results indicate that the scheme is stable also for much larger time steps.

The article is organised as follows: In Section 2, we introduce the discretisation of the model problem (1.1) in time and space. Section 3 presents a stability analysis for the fully discrete scheme using a CFL condition. In Section 4, we show a detailed a priori convergence analysis. Numerical experiments in two and three space dimensions are presented in Section 5. Section 6 summarises this article with some concluding remarks.

2 Discretisation

In this section, we present the numerical approximation of the model problem (1.1). We start with discretisation in time, and continue with the spatial discretisation of the resulting time-discrete formulation.

2.1 Temporal discretisation

For time discretisation, we divide the time interval of interest $I = [0, t_{max}]$ in intervals $I_n = (t_{n-1}, t_n]$. For simplicity, we take a uniform time step $\Delta t = \frac{t_{max}}{N}$ and define $t_n = n\Delta t$. We define the domain $\Omega^n := \Omega(t_n)$ with boundary $\Gamma^n := \Gamma(t_n)$ and write $u^n = u(t_n)$ for the exact solution of the continuous problem (1.1) at time t_n.

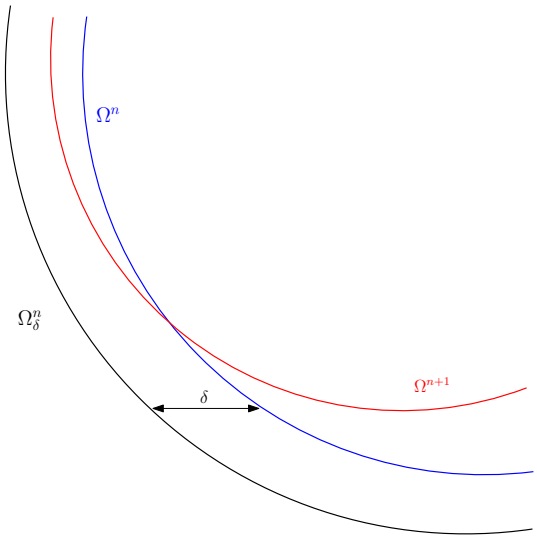


Figure 1: Illustration of the domains Ω^n, Ω^{n+1} and the extension Ω^n_δ.

A δ-neighborhood of $\Omega(t)$ at time step n is chosen large enough such that $(\Omega^n \cup \Omega^{n+1}) \subset \Omega^n_\delta$, see Figure 1. Therefore we choose

$$\delta \geq w_{max}\Delta t, \qquad w_{max} = \sup_{t\in I,\, x\in\partial\Omega(0)} \|\partial_t T(x,t)\cdot n\|.$$

The required regularity of the domain mapping T will be ensured in Assumption 1 below. For the error analysis we will also assume the upper bound

$$\delta \leq c_\delta w_{max}\Delta t \tag{2.1}$$

with a constant $c_\delta > 1$. Finally, we introduce the following notations for some space-time domains

$$Q := \bigcup_{t\in I}\{t\}\times\Omega(t), \quad Q^n := \bigcup_{t\in I_n}\{t\}\times\Omega(t) \quad Q^n_\delta := \bigcup_{t\in I_n}\{t\}\times\Omega_\delta(t), \quad \hat{Q} = \Omega(0)\times[0,t_{max}].$$

Now, the Crank-Nicolson method applied to (1.1) writes formally

$$\frac{u^n - u^{n-1}}{\Delta t} - \frac{1}{2}(\Delta u^n + \Delta u^{n-1}) = \frac{1}{2}(f^n + f^{n-1}), \quad x\in\Omega^n. \tag{2.2}$$

The main issue of this formulation is that u^{n-1} is needed on Ω^n, while it is defined on Ω^{n-1}. Thus, we will add implicit extension operators below to define u^n on $\Omega^n_\delta \supset \Omega^{n+1}$, where it is needed in the following time step. Similarly, f^{n-1} might be undefined on $\Omega^n \setminus \Omega^{n-1}$. If f^{n-1} is given analytically, it can typically be extended in a canonical way to Ω^n. To cover different scenarios, we do not want to restrict the analysis in this work to a particular extension, but assume only that f^{n-1} is smoothly extended to Ω^n.

In this article, we will use the abbreviation c to refer to a generic positive constant, which is independent of discretisation parameters $(\Delta t,\, h)$ and the relative positions of the boundary with respect to the mesh.

2.1.1 Extension operator

In this part, we introduce an extension operator of the exact solution u to extend variables to larger domains, as the spatial domain evolves. We make the following assumption (see also [11, Assumption 3.2]) for the analysis of this article.

Assumption 1 *The boundary of the initial domain $\Omega(0)$ is assumed to be piecewise smooth and Lipschitz, and the domain motion $T(t)$ is a $W^{1,\infty}$-diffeomorphism for each t, that fulfills $T \in W^{r,\infty}(\hat{Q})$, where $r = \max\{3, m+1\}$ and m is the polynomial degree of the finite element space defined in the following subsection.*

By using assumption 1, there exist $W^{r,\infty}$-stable extension operators E^n from Ω^n to Ω^n_δ that satisfy the following analytical properties:

$$\|E^n u - u\|_{W^{m+1,p}(\Omega)} = 0, \quad \|E^n u\|_{W^{m+1,p}(\Omega^n_\delta)} \leq c\|u\|_{W^{m+1,p}(\Omega^n)}, \tag{2.3}$$

$$\|\partial_t E^n u\|_{H^m(\Omega^n_\delta)} \leq c\left(\|u\|_{H^{m+1}(\Omega^n)} + \|\partial_t u\|_{H^m(\Omega^n)}\right), \tag{2.4}$$

$$\|\partial_t^3 E^n u\|_{L^\infty(Q^n_\delta)} \leq c\|u\|_{W^{3,\infty}(Q)}. \tag{2.5}$$

The properties (2.3) and (2.4) are discussed in [11]. In an analogous way, one can derive the estimate for the third-order time derivative in (2.5).

2.2 Spatial discretisation

For spatial discretisation, we introduce a polygonal domain D, which is chosen large enough, such that $\Omega^n_\delta \subset D$ for all n. We introduce a quasi-uniform family of triangulations $(\mathcal{T}_h)_{h>0}$ of D with maximum cell size h, which will serve as background meshes.

In each time step, we extract from $\mathcal{T}_h$ all cells of non-empty intersection with Ω^n_δ and define

$$\mathcal{T}^n_{h,\delta} := \{K \in \mathcal{T}_h : K \cap \Omega^n_\delta \neq \emptyset\}.$$

We write $\Omega^n_{h,\delta}$ for the domain spanned by all cells $K \in \mathcal{T}^n_{h,\delta}$ and define the following finite element space:

$$V_h^{n,m} := \{v \in C(\Omega^n_{h,\delta}), v|_K \in P_m(K)\ \forall K \in \mathcal{T}^n_{h,\delta}\}, \quad m \geq 1.$$

The set of elements that lie (at least partially) outside of Ω^{n-1}, but in Ω^n, will be of particular interest in the analysis. The domain spanned by them will be denoted by

$$\mathcal{S}_h^{n,n-1} := \bigcup_{K \in \mathcal{T}_h^{n,n-1}} K, \qquad \text{where } \mathcal{T}_h^{n,n-1} := \left\{K \in \mathcal{T}^n_{h,\delta},\, K \cap (\Omega^n \setminus \Omega^{n-1}) \neq \emptyset\right\}.$$

Moreover, we introduce the following notations for the facets of $\mathcal{T}^n_{h,\delta}$, see Figure 2 for an illustration:

- $\mathcal{F}^n_{h,\delta}$: the set of interior facets of $\mathcal{T}^n_{h,\delta}$.
- $\mathcal{F}^{n,int}_{h,\delta}$: the set of facets that belong exclusively to elements $K \in \mathcal{T}^n_{h,\delta}$ that lie completely in the interior of Ω^n.
- $\mathcal{F}^{n,cut}_{h,\delta}$: the set of facets that belong to some element $K \in \mathcal{T}^n_{h,\delta}$ with $K \cap \partial\Omega^n \neq \emptyset$.

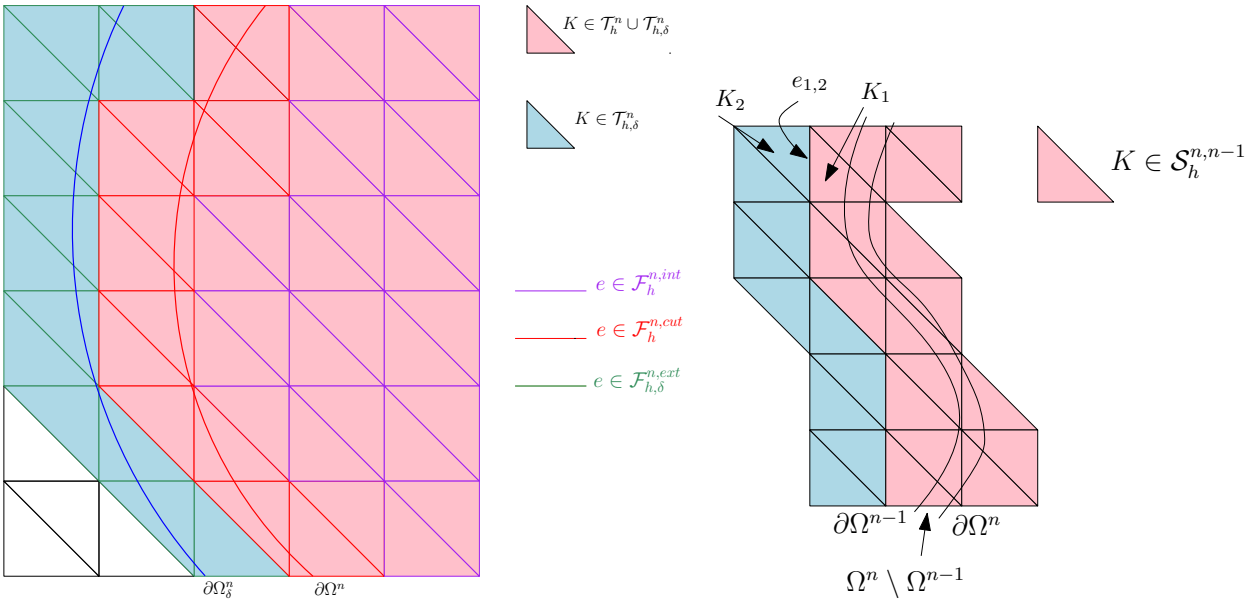


Figure 2: **Left**: Illustration of the triangulations $\mathcal{T}_h^n$ and $\mathcal{T}_{h,\delta}^n$ and the sets of facets $\mathcal{F}_{h,\delta}^n = \mathcal{F}_{h,\delta}^{n,int} \cup \mathcal{F}_{h,\delta}^{n,cut} \cup \mathcal{F}_{h,\delta}^{n,ext}$. **Right**: Set of boundary cells $\mathcal{S}_h^{n,n-1}$.

- $\mathcal{F}_{h,\delta}^{n,ext}$: the set of remaining facets of $\mathcal{F}_{h,\delta}^n$, i.e. $\mathcal{F}_{h,\delta}^{n,ext} := \mathcal{F}_{h,\delta}^n \setminus \left(\mathcal{F}_{h,\delta}^{n,int} \cup \mathcal{F}_{h,\delta}^{n,cut}\right)$
- $\mathcal{F}_{h,\delta}^{n,g} := \mathcal{F}_{h,\delta}^{n,cut} \cup \mathcal{F}_{h,\delta}^{n,ext}$.

Assumption 2 (CFL condition) *We assume the parabolic CFL condition $\Delta t \leq c_{CFL} h^2$, where c_{CFL} is an arbitrary constant for $m = 1$, while we assume c_{CFL} sufficiently small for $m > 1$.*

The inequality (2.1) and the CFL condition (Assumption 2) lead to

$$\delta \leq c_\delta w_{max} \Delta t \leq c w_{max} h^2. \tag{2.6}$$

Remark 2.1 *The inequality* (2.6) *implies that the distance between $\partial\Omega^n$ and $\partial\Omega^{n-1}$ is bounded by $\mathcal{O}(h^2)$. This implies the following property, which will be needed in the analysis below: For each cell $K \in S_h^{n,n-1}$, there exists a path of cells K_i, $i = 1, \ldots, M$, such that $\overline{K}_i \cap \overline{K}_{i+1}$ is a facet in $\mathcal{F}_{h,\delta}^{g,n} \cap \mathcal{F}_{h,\delta}^{g,n-1}$ and the final cell K_M lies fully in the interior of Ω^n. Furthermore, the number of cases, in which an element $K_M \subset \Omega^n$ is utilised as a final element among all paths, can be bounded independently of h and Δt.*

2.2.1 Discrete variational formulation

In the numerical approximation, the boundary condition of the discrete problem (2.7) is implemented weakly by means of Nitsche's method. Moreover, the function u_h^n is extended by means of a ghost penalty term $g_h^n(\cdot,\cdot)$. In each time step $n = 1, 2, \ldots, N$, we consider the following discrete variational formulation: Find $u_h^n \in V_h^{n,m}$ such that

$$\mathcal{A}(u_h^n, u_h^{n-1}; v_h) = (f_h^{n-\frac{1}{2}}, v_h), \quad \forall v_h \in V_h^{n,m}, \tag{2.7}$$

where

$$\mathcal{A}(u_h^n, u_h^{n-1}; v_h) := \left(D_{\Delta t}^- u_h^n, v_h\right)_{\Omega^n} + \frac{1}{2} a_h^n(u_h^n, v_h) + \frac{1}{2} a_h^n(u_h^{n-1}, v_h) + \frac{\gamma_D}{h}(u_h^n, v_h)_{\partial\Omega^n} + \gamma_g g_h^n(u_h^n, v_h) \tag{2.8}$$

and

$$D_{\Delta t}^- u_h^n = \frac{u_h^n - u_h^{n-1}}{\Delta t}, \quad a_h^n(u_h^k, v_h) = (\nabla u_h^k, \nabla v_h)_{\Omega^n} - (\partial_n u_h^k, v_h)_{\partial\Omega^n}, \quad f^{n-\frac{1}{2}} = \frac{f^n + f^{n-1}}{2}.$$

We assume that all integrals in (2.8) are evaluated exactly. For a consideration of additional quadrature errors, that result when cut cells are approximated linearly in the computation of the integrals, we refer to [48].

The ghost penalty stabilization is defined by

$$g_h^n(w_h, v_h) = \sum_{e \in \mathcal{F}_{h,\delta}^{n,g}} \sum_{k=1}^{m} \frac{h^{2k-1}}{k!^2} \int_e [[\partial_n^k w_h]] \cdot [[\partial_n^k v_h]] \, \mathrm{d}s, \tag{2.9}$$

where $[[\cdot]]$ is the jump operator and ∂_n the exterior normal derivative. For further possibilities for the extension g_h^n, we refer to [48]. The variant chosen here based on the jump of derivatives over edges has the advantage that it is fully consistent, in the sense that $g_h^n(u, v)$ vanishes for $u \in H^{m+1}(\Omega_\delta^n)$. The purpose of the ghost penalty is twofold: First, it serves to extend the solution u_h^n implicitly to $\Omega_{h,\delta}^n$. Secondly, it ensures the discrete coercivity of the formulation (2.7) on $\mathcal{T}_{h,\delta}^n$.

To incorporate the initial condition, we set $u_h^0 := E^1 u^0$ in (2.8) for $n = 1$, where $E^1 u^0$ is a smooth, e.g. a canonical extension, of the initial value u_0. This corresponds to the following Ritz projection of the initial value u^0

$$\Delta t^{-1}(u_h^0, v_h)_{\Omega^1} + a^1(u_h^0, v_h) = \Delta t^{-1}(E^1 u^0, v_h)_{\Omega^1} + a^1(E^1 u^0, v_h) \quad \forall v_h \in V_h^1,$$

The following lemma is the key to extend the discrete coercivity to Ω_δ^n:

Lemma 2.2 *Given Assumption 2, any discrete function $v_h^n \in V_h^{n,m}$ satisfies*

$$\|v_h^n\|_{\Omega_{h,\delta}^n}^2 \leq c\|v_h^n\|_{\Omega^n}^2 + ch^2 g_h^n(v_h, v_h), \qquad \|\nabla v_h^n\|_{\Omega_{h,\delta}^n}^2 \leq c\|\nabla v_h^n\|_{\Omega^n}^2 + c g_h^n(v_h, v_h).$$

In addition, for $v, w \in H^{m+1}(\Omega_\delta^n)$, $m \geq 1$, it holds

$$g_h^n(v, w) \leq g_h^n(v, v)^{1/2} g_h^n(w, w)^{1/2}, \quad g_h^n(v, v) \leq ch^{2m}\|v\|_{H^{m+1}(\Omega_\delta^n)}^2.$$

Proof. A proof of this lemma is given in [48]. ∎

At the end of this section, we briefly show that the variational formulation (2.7) is well-posed for each n. We define the discrete energy as

$$\mathfrak{E}^n(u_h^n, u_h^{n-1}) = \left(\frac{1}{2}\|\nabla u_h^n + \nabla u_h^{n-1}\|_{\Omega^n}^2 + \frac{1}{\Delta t}\|u_h^n - u_h^{n-1}\|_{\Omega^n}^2 + \frac{\gamma_D}{h}\|u_h^n\|_{\partial\Omega^n}^2 + \gamma_g g_h^n(u_h^n, u_h^n)\right)^{1/2}. \tag{2.10}$$

and the energy norm as

$$|||u_h^n|||_n := \mathfrak{E}^n(u_h^n, 0).$$

We will show the coercivity relation

$$|||u_h^n|||_n^2 \leq \frac{1}{4}\mathcal{A}(u_h^n, 0; u_h^n), \qquad u_h^n \in V_h^{m,n}. \tag{2.11}$$

for sufficiently large γ_g, γ_D. The well-posedness of (2.7) follows then by standard arguments.

From the definition of the bilinear form $\mathcal{A}(u_h^n, 0; u_h^n)$, we have

$$\mathcal{A}(u_h^n, 0; u_h^n) = \frac{1}{\Delta t}\|u_h^n\|_{\Omega^n}^2 + \frac{1}{2}\|\nabla u_h^n\|_{\Omega^n}^2 - \frac{1}{2}(\partial_n u_h^n, u_h^n)_{\partial\Omega^n} + \frac{\gamma_D}{h}\|u_h^n\|_{\partial\Omega^n}^2 + \gamma_g g_h^n(u_h^n, u_h^n).$$

The term $-\frac{1}{2}(\partial_n u_h^n, u_h^n)_{\partial\Omega^n}$ is estimated by means of Young's inequality, an inverse inequality and Lemma 2.2 as follows for sufficiently small $\epsilon > 0$:

$$\begin{aligned} -\frac{1}{2}(\partial_n u_h^n, u_h^n)_{\partial\Omega^n} &\geq -\frac{\epsilon h}{4}\|\nabla u_h^n\|_{\partial\Omega^n}^2 - \frac{1}{4\epsilon h}\|u_h^n\|_{\partial\Omega^n}^2 \\ &\geq -c\epsilon\|\nabla u_h^n\|_{\Omega_h^n}^2 - \frac{1}{4\epsilon h}\|u_h^n\|_{\partial\Omega^n}^2. \end{aligned}$$

For sufficiently large parameters γ_D, γ_g, we have

$$-\frac{1}{2}(\partial_n u_h^n, u_h^n)_{\partial\Omega^n} \geq -\frac{1}{4}\left(\|\nabla u_h^n\|_{\Omega^n}^2 + \gamma_g g_h^n(u_h^n, u_h^n)\right) - \frac{\gamma_D}{4h}\|u_h^n\|_{\partial\Omega^n}^2.$$

This proves the coercivity (2.11).

3 Stability analysis

In this section, a detailed stability analysis of the discrete problem (2.7) is developed. One of the main issues in the analysis is that the discrete functions u_h^{n-1} and ∇u_h^{n-1} appear on Ω^n in the n-th time step, whereas bounds are only available for $\|u_h^{n-1}\|_{\Omega^{n-1}}$ and $\|\nabla u_h^{n-1}\|_{\Omega^{n-1}}$ from the previous time step. We start with some technical lemmas that will enable us to deal with this issue.

3.1 Auxiliary estimates

Lemma 3.1 *Let Assumptions 2 be valid. Any discrete functions $v_h^n \in V_h^{n,m}$, $v_h^{n-1} \in V_h^{n-1,m}$ satisfy the following inequality:*

$$\|\nabla v_h^n - \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 \leq \frac{2}{\Delta t}\|v_h^n - v_h^{n-1}\|_{\Omega^n}^2 + c g_h^n(v_h^n, v_h^n) + c g_h^{n-1}(v_h^{n-1}, v_h^{n-1}). \tag{3.1}$$

Proof. For $m = 1$,$\nabla v_h^n - \nabla v_h^{n-1}$ is constant on a cell $T \in \mathcal{T}_h^n$. Due to the quasi-uniformity of the background mesh, we have $|T| \geq ch^d$ and from (2.1) $|T \cap (\Omega^n \setminus \Omega^{n-1})| \leq ch^{d-1}\Delta t$. It follows that

$$\|\nabla v_h^n - \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 \leq c\frac{\Delta t}{h}\|\nabla v_h^n - \nabla v_h^{n-1}\|_{\mathcal{S}_h^{n,n-1}}^2. \tag{3.2}$$

In the general case ($m > 1$), we have still

$$\|\nabla v_h^n - \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 \leq \|\nabla v_h^n - \nabla v_h^{n-1}\|_{\mathcal{S}_h^{n,n-1}}^2. \tag{3.3}$$

Now let $K_1 \in \mathcal{S}_h^{n,n-1}$. By Remark 2.1, there is a set of neighbouring cells $K_2, ..., K_M$, such that $(\overline{K}_i \cap \overline{K}_{i+1}) \in (\mathcal{F}_{h,\delta}^{g,n} \cap \mathcal{F}_{h,\delta}^{g,n-1})$ and K_M lies fully in the interior of Ω^n. Let $e_{1,2}$ be the edge that

separates the cells K_1 and K_2, see Figure 2. Then, using arguments from [50, Lemma 5.1] and [48, Lemma 5.2], we can deduce that

$$\begin{aligned}\|\nabla v_h^n - \nabla v_h^{n-1}\|_{K_1}^2 &\leq c\|\nabla v_h^n - \nabla v_h^{n-1}\|_{K_2}^2 + c\sum_{k=1}^{m}\int_{e_{1,2}} h^{2k-1}[[\nabla^k v_h^n - \nabla^k v_h^{n-1}]]^2 ds \\ &\leq c\|\nabla v_h^n - \nabla v_h^{n-1}\|_{K_2}^2 + c\sum_{k=1}^{m}\int_{e_{1,2}} h^{2k-1}[[\nabla^k v_h^n]]^2 ds + c\int_{e_{1,2}} h^{2k-1}[[\nabla^k v_h^{n-1}]]^2 ds.\end{aligned} \tag{3.4}$$

We follow this process from K_2 to K_M by crossing edges $e_{2,3}$ to $e_{M-1,M}$ to obtain

$$\begin{aligned}\|\nabla v_h^n - \nabla v_h^{n-1}\|_{K_1}^2 \leq& c\|\nabla v_h^n - \nabla v_h^{n-1}\|_{K_M}^2 \\ &+ \sum_{k=1}^{m}\sum_{j=2}^{M}\left(\int_{e_{j-1,j}} h^{2k-1}[[\nabla^k v_h^n]]^2 ds + c\int_{e_{j-1,j}} h^{2k-1}[[\nabla^k v_h^{n-1}]]^2 ds\right).\end{aligned} \tag{3.5}$$

For the first term on the right-hand side of (3.5), we use an inverse inequality and the CFL condition with a sufficiently small constant c_{CFL}

$$\|\nabla v_h^n - \nabla v_h^{n-1}\|_{K_M}^2 \leq \frac{c}{h^2}\|v_h^n - v_h^{n-1}\|_{K_M}^2 \leq \frac{2}{\Delta t}\|v_h^n - v_h^{n-1}\|_{K_M}^2. \tag{3.6}$$

As all edges $e_{j-1,j}$ belong to both $\mathcal{F}_{h,\delta}^{g,n}$ and $\mathcal{F}_{h,\delta}^{g,n-1}$, we obtain after summation over all cells in $\mathcal{S}_h^{n,n-1}$ from (3.3), (3.5) and (3.6)

$$\|\nabla v_h^n - \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 \leq \frac{2}{\Delta t}\|v_h^n - v_h^{n-1}\|_{\Omega^n}^2 + c\left(g_h^n(v_h^n, v_h^n) + g_h^{n-1}(v_h^{n-1}, v_h^{n-1})\right).$$

Using (3.2) instead of (3.3) for $m = 1$, the same result follows under the CFL condition $\Delta t \leq c_{CFL}h^2$ for an arbitrary constant c_{CFL}. ■

We note that the CFL condition is required to estimate the "mismatch" $\|\nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2$ by means of the discrete time derivative $\frac{1}{\Delta t}\|v_h^n - v_h^{n-1}\|_{\Omega^n}^2$ (see the following lemma).

Next, we discuss how the term $\|\nabla u_h^{n-1}\|_{\Omega^n}$ can be bounded by $\|\nabla u_h^{n-1}\|_{\Omega^{n-1}}$ plus further terms that can be controlled in the following stability analysis.

Lemma 3.2 *Under the assumptions of Lemma 3.1 it holds for $v_h^n \in V_h^{n,m}$, $v_h^{n-1} \in V_h^{n-1,m}$ that*

$$\begin{aligned}\Delta t\|\nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 \leq& \frac{\Delta t}{2}\|\nabla v_h^n + \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 + \|v_h^n - v_h^{n-1}\|_{\Omega^n}^2 \\ &+ c\Delta t g_h^n(v_h^n, v_h^n) + c\Delta t g_h^{n-1}(v_h^{n-1}, v_h^{n-1}).\end{aligned} \tag{3.7}$$

Proof. By the triangle inequality, we have

$$\|\nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}} \leq \frac{1}{2}\|\nabla v_h^n + \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}} + \frac{1}{2}\|\nabla v_h^n - \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}.$$

By means of the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ this implies

$$\Delta t\|\nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 \leq \frac{\Delta t}{2}\|\nabla v_h^n + \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 + \frac{\Delta t}{2}\|\nabla v_h^n - \nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2. \tag{3.8}$$

The statement follows by using Lemma 3.1 for the second term in (3.8)

$$\frac{\Delta t}{2}\|\nabla v_h^n - \nabla v_h^{n-1}\|^2_{\Omega^n\setminus\Omega^{n-1}} \leq \|v_h^n - v_h^{n-1}\|^2_{\Omega^n} + c\Delta t g_h^n(v_h^n, v_h^n) + c\Delta t g_h^{n-1}(v_h^{n-1}, v_h^{n-1}). \tag{3.9}$$

■

Next, we provide the following Poincaré-type estimate:

Lemma 3.3 *Let $u \in W^{1,p}(\Omega)$ for $1 \leq p < \infty$ and let the CFL condition (Assumption 2) be valid. It holds for $l \in \{n-1, n\}$*

$$\|u\|^p_{L^p(\Omega^n\setminus\Omega^{n-1})} \leq c\Delta t\|u\|^p_{L^p(\partial\Omega^l)} + c\Delta t^2\|\nabla u\|^p_{L^p(\Omega^n\setminus\Omega^{n-1})}. \tag{3.10}$$

Proof. The proof follows the lines of [55, Lemma 4.34] and uses the fact that $\operatorname{dist}(\Omega^n, \Omega^{n-1}) \leq c\Delta t$. ■

Using this, we can derive bounds for $\|v_h^{n-1}\|^2_{\Omega^n\setminus\Omega^{n-1}}$ and $\|v_h^n\|^2_{\Omega^n\setminus\Omega^{n-1}}$:

Lemma 3.4 *Under the assumptions of Lemma 3.1 it holds for $v_h^l \in V_h^{l,m}$, $l \in \{n-1, n\}$*

$$\begin{aligned}\|v_h^l\|^2_{\Omega^n\setminus\Omega^{n-1}} \quad &\leq c\Big(\Delta t\|v_h^l\|^2_{\partial\Omega^l} + \Delta t\|v_h^n - v_h^{n-1}\|^2_{\Omega^n} + \Delta t^2\|\nabla v_h^n + \nabla v_h^{n-1}\|^2_{\Omega^n} \\ &\qquad + \Delta t^2 g_h^{n-1}(v_h^{n-1}, v_h^{n-1}) + \Delta t^2 g_h^n(v_h^n, v_h^n)\Big)\end{aligned} \tag{3.11}$$

Proof. By means of Lemma 3.3 for $p = 2$ we have for $l \in \{n-1, n\}$

$$\|v_h^l\|^2_{\Omega^n\setminus\Omega^{n-1}} \leq c\Delta t\|v_h^l\|^2_{\partial\Omega^l} + c\Delta t^2\|\nabla v_h^l\|^2_{\Omega^n\setminus\Omega^{n-1}} \tag{3.12}$$

The statement follows by applying Lemma 3.2 to the second term in (3.12). ■

3.2 Stability result

Before discussing the stability result, we introduce some abbreviations for the space-time Bochner norms to simplify the mathematical expressions

$$\|u\|_{\infty,m,I_k} := \|u\|_{L^\infty(I_k, H^m(\Omega(t)))}, \qquad \|u\|_{\infty,m} := \|u\|_{\infty,m,I},$$

where $m \in \mathbb{N} \cup \{0\}$ and $H^0(\Omega(t)) := L^2(\Omega(t))$.

Now we are ready to prove the following stability result.

Theorem 3.5 (Stability) *Let Assumptions 2 be valid and let $f \in L^\infty(I, L^2(\Omega(t)))$, $u^0 \in H^1(\Omega^0)$ and let the mapping T be a $W^{1,\infty}$-diffeomorphism. For sufficiently large γ_g, γ_D the solution $\{u_h^k\}_{k=1}^n$ of the discrete problem (2.7) fulfills*

$$\|u_h^n\|^2_{\Omega^n} + \Delta t\|u_h^n\|^2_{\Omega^n} + \Delta t\sum_{k=1}^n \mathfrak{E}^k(u_h^k, u_h^{k-1}) \leq c\exp(c\,t_n)\left(\|u^0\|^2_{\Omega^0} + t_n\|\nabla u^0\|^2_{\Omega^0} + t_n\|f\|^2_{\infty,0}\right). \tag{3.13}$$

Proof. We test (2.7) with $v_h = 2\Delta t u_h^n$ to obtain

$$\begin{aligned}&2\left(u_h^n - u_h^{n-1}, u_h^n\right)_{\Omega^n} + \Delta t\left(\nabla u_h^n + \nabla u_h^{n-1}, \nabla u_h^n\right)_{\Omega^n} + \frac{2\gamma_D\Delta t}{h}(u_h^n, u_h^n)_{\partial\Omega^n} \\ &-\Delta t(\partial_n u_h^n + \partial_n u_h^{n-1}, u_h^n)_{\partial\Omega^n} + 2\Delta t\gamma_g g_h^n(u_h^n, u_h^n) = 2\Delta t(f^{n-\frac{1}{2}}, u_h^n).\end{aligned} \tag{3.14}$$

We estimate the fourth term in (3.14) by means of Young's inequality with a sufficiently small $\epsilon > 0$ followed by an inverse inequality

$$\begin{aligned}\Delta t(\partial_n u_h^n + \partial_n u_h^{n-1}, u_h^n)_{\partial\Omega^n} &\leq \frac{\Delta t \epsilon h}{16}\|\partial_n u_h^n + \partial_n u_h^{n-1}\|_{\partial\Omega^n}^2 + \frac{4\Delta t}{\epsilon h}\|u_h^n\|_{\partial\Omega^n}^2 \\ &\leq \frac{\Delta t}{16}\|\nabla u_h^n + \nabla u_h^{n-1}\|_{\Omega^n}^2 + \frac{\gamma_D \Delta t}{2h}\|u_h^n\|_{\partial\Omega^n}^2,\end{aligned}$$

where $\gamma_D \geq 8/\epsilon$. By using the relation $2(a+b,a) = (a+b)^2 + a^2 - b^2$ for the first two terms in (3.14), we obtain that

$$\begin{aligned}\|u_h^n\|_{\Omega^n}^2 + \|u_h^n - u_h^{n-1}\|_{\Omega^n}^2 + \frac{\Delta t}{2}\|\nabla u_h^n\|_{\Omega^n}^2 + \frac{7\Delta t}{16}\|\nabla u_h^n + \nabla u_h^{n-1}\|_{\Omega^n}^2 + \frac{3\gamma_D \Delta t}{2h}\|u_h^n\|_{\partial\Omega^n}^2 \\ + 2\Delta t \gamma_g g_h^n(u_h^n, u_h^n) \leq \|u_h^{n-1}\|_{\Omega^n}^2 + \frac{\Delta t}{2}\|\nabla u_h^{n-1}\|_{\Omega^n}^2 + 2\Delta t(f^{n-1/2}, u_h^n).\end{aligned} \tag{3.15}$$

For $n > 1$, we bring the terms $\|u_h^{n-1}\|_{\Omega^n}$ and $\|\nabla u_h^{n-1}\|_{\Omega^n}$ to the domain Ω^{n-1}. By employing Lemmas 3.2 and 3.4, we have

$$\begin{aligned}\|u_h^{n-1}\|_{\Omega^n}^2 + \frac{\Delta t}{2}\|\nabla u_h^{n-1}\|_{\Omega^n}^2 \leq \|u_h^{n-1}\|_{\Omega^{n-1}}^2 + c\Delta t\|u_h^{n-1}\|_{\partial\Omega^{n-1}}^2 + \frac{\Delta t}{2}\|\nabla u_h^{n-1}\|_{\Omega^{n-1}}^2 \\ + \left(\frac{\Delta t}{4} + c\Delta t^2\right)\|\nabla u_h^n + \nabla u_h^{n-1}\|_{\Omega^n}^2 + \left(\frac{1}{2} + c\Delta t\right)\|u_h^n - u_h^{n-1}\|_{\Omega^n}^2 \\ + c\Delta t g_h^n(u_h^n, u_h^n) + c\Delta t g_h^{n-1}(u_h^{n-1}, u_h^{n-1}).\end{aligned} \tag{3.16}$$

Inserting (3.16) into (3.15) and using $2\Delta t(f^{n-1/2}, u_h^n)_{\Omega^n} \leq \Delta t\left(\|f^{n-1/2}\|_{\Omega^n}^2 + \|u_h^n\|_{\Omega^n}^2\right)$ gives for sufficiently large γ_g

$$\begin{aligned}(1-\Delta t)\|u_h^n\|_{\Omega^n}^2 + \frac{\Delta t}{2}\|\nabla u_h^n\|_{\Omega^n}^2 + \frac{1}{4}\|u_h^n - u_h^{n-1}\|_{\Omega^n}^2 + \frac{\Delta t}{8}\|\nabla u_h^n + \nabla u_h^{n-1}\|_{\Omega^n}^2 \\ + \frac{3\gamma_D \Delta t}{2h}\|u_h^n\|_{\partial\Omega^n}^2 + \Delta t\gamma_g g_h^n(u_h^n, u_h^n) \\ \leq \|u_h^{n-1}\|_{\Omega^{n-1}}^2 + \frac{\Delta t}{2}\|\nabla u_h^{n-1}\|_{\Omega^{n-1}}^2 + c\Delta t\|u_h^{n-1}\|_{\partial\Omega^{n-1}}^2 \\ + c\Delta t g_h^{n-1}(u_h^{n-1}, u_h^{n-1}) + \Delta t\|f_h^{n-\frac{1}{2}}\|_{\Omega^n}^2.\end{aligned} \tag{3.17}$$

For $n = 1$, we obtain from (3.15) and the stability of the extension E^1

$$\begin{aligned}(1-\Delta t)\|u_h^1\|_{\Omega^1}^2 + \|u_h^1 - u_h^0\|_{\Omega^1}^2 + \frac{\Delta t}{2}\|\nabla u_h^1\|_{\Omega^1}^2 + \frac{3\Delta t}{8}\|\nabla u_h^1 + \nabla u_h^0\|_{\Omega^1}^2 \\ + \frac{3\gamma_D \Delta t}{2h}\|u_h^1\|_{\partial\Omega^1}^2 + 2\Delta t\gamma_g g_h^1(u_h^1, u_h^1) \\ \leq \|E^1 u^0\|_{\Omega^1}^2 + \frac{\Delta t}{2}\|\nabla(E^1 u^0)\|_{\Omega^1}^2 + \Delta t\|f_h^{\frac{1}{2}}\|_{\Omega^1}^2 \\ \leq c\|u^0\|_{\Omega^0}^2 + c\Delta t\|\nabla u^0\|_{\Omega^0}^2 + \Delta t\|f_h^{\frac{1}{2}}\|_{\Omega^1}^2.\end{aligned} \tag{3.18}$$

Taking the sum over $k = 1, 2, \ldots, n$, this yields for sufficiently large γ_g and γ_D

$$\begin{aligned}\|u_h^n\|_{\Omega^n}^2 + \Delta t\|\nabla u_h^n\|_{\Omega^n}^2 + \frac{\Delta t}{4}\sum_{k=1}^{n}\mathfrak{E}^k(u_h^k, u_h^{k-1}) \\ \leq c\|u_h^0\|_{\Omega^0}^2 + ct_n\|\nabla u_h^0\|_{\Omega^1}^2 + 2t_n\|f\|_{\infty,0}^2 + c\Delta t\sum_{k=2}^{n}\|u_h^{k-1}\|_{\Omega^{k-1}}^2.\end{aligned} \tag{3.19}$$

The statement follows by means of the discrete Gronwall lemma. ■

Remark 3.6 (CFL condition) *For $m = 1$, the stability result in Theorem 3.5 could be shown under the weaker CFL condition $\Delta t \leq c_{CFL}h^{3/2}$ for sufficiently small c_{CFL}. This is due to the possibility to use the estimate*

$$\|\nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2 \leq c\frac{\Delta t}{h}\|\nabla v_h^{n-1}\|_{\Omega_{h,\delta}^{n-1}}^2 \tag{3.20}$$

in (3.2) *(Note that, for $m = 1$, ∇v_h^{n-1} is constant in each cell T). However, in the following section the stronger CFL condition $\Delta t \leq c_{CFL}h^2$ with arbitrary c_{CFL} (see Assumption 2) will be used in order to show optimal-order convergence estimates. The CFL condition is needed to estimate the term $\|\nabla v_h^{n-1}\|_{\Omega^n\setminus\Omega^{n-1}}^2$ by the discrete time derivative $\Delta t^{-1}\|v_h^n - v_h^{n-1}\|_{\Omega^n}^2$, see* (3.6) *and Lemma 3.2.*

4 A priori error analysis

In this section, we show an *a priori* error estimate for the discrete problem (2.7). We define the discretisation error as

$$e^n := u^n - u_h^n, \quad n \geq 1, \tag{4.1}$$

where $u^n := u(t_n)$ is assumed to be at least in $H^2(\Omega^n)$. Further regularity assumptions on u will be made below. The error is decomposed into an interpolation error η^n and a discrete error ξ_h^n terms defined by

$$\eta^n := u^n - I_h^n u^n, \qquad \xi_h^n := I_h^n u^n - u_h^n, \tag{4.2}$$

where $I_h^n u^n$ denotes the standard Lagrangian nodal interpolation of u^n on $\mathcal{T}_{h,\delta}^n$. For $n = 0$ we have, by definition of u_h^0, that $e^0 = u^0 - u_h^0 = 0$ in Ω^0 and thus, we also set $\eta^0 = \xi_h^0 = 0$. We will make use of the following standard interpolation estimates for $n \geq 1$

$$\|\eta^n\|_{H^l(\Omega)} \leq ch^{k-l}\|u^n\|_{H^k(\Omega)}, \quad \text{for } 0 \leq l \leq 1,\ 2 \leq k \leq m+1, \tag{4.3}$$

$$\|\eta^n\|_{H^l(\partial\Omega)} \leq ch^{k-l-1/2}\|u^n\|_{H^k(\Omega)}, \quad \text{for } 0 \leq l \leq 1,\ 2 \leq k \leq m+1. \tag{4.4}$$

4.1 Consistency error

The exact solution $u \in H^1(\Omega(t))$ of the continuous problem (1.1) satisfies the following weak formulation:

$$(u_t, v)_{\Omega(t)} + a(u, v) = (f, v)_{\Omega(t)}, \quad t \in I_n \tag{4.5}$$

for $v \in H^1(\Omega(t))$ and the bilinear form

$$a(u, v) = (\nabla u, \nabla v)_{\Omega(t)} - (\partial_n u, v)_{\partial\Omega(t)}.$$

At time t_{n-1}, we have

$$(u_t(t_{n-1}), v)_{\Omega^{n-1}} + a^{n-1}(u^{n-1}, v) = (f^{n-1}, v)_{\Omega^{n-1}} \tag{4.6}$$

where

$$a^{n-1}(u^{n-1}, v) = (\nabla u^{n-1}, \nabla v)_{\Omega^{n-1}} - (\partial_n u^{n-1}, v)_{\partial\Omega^{n-1}}.$$

To estimate the consistency error, we will need an analogous equality for u^{n-1} on Ω^n instead of Ω^{n-1}. Therefore, we define $\tilde{u}$ as a smooth extension of the exact solution u to Q_δ^n. Moreover, we define a smooth extension of the source term f as follows:

$$\tilde{f}(t_{n-1}) = \tilde{u}_t(t_{n-1}) - \Delta\tilde{u}(t_{n-1}), \quad \text{on } \Omega_\delta^n. \tag{4.7}$$

It holds

$$\tilde{f}(t_{n-1}) = f(t_{n-1}) \quad \text{on } \Omega^{n-1}$$

However, as we have allowed an arbitrary smooth extension of f to Ω_δ^{n-1} in Section 2, this does not necessarily hold in $\Omega_\delta^{n-1} \setminus \Omega^{n-1}$. By using a test function $v_h \in V_h^{n,m}$, we get

$$(\tilde{u}_t^{n-1}, v_h)_{\Omega^n} + (\nabla\tilde{u}^{n-1}, \nabla v_h)_{\Omega^n} - (\partial_n\tilde{u}_h^{n-1}, v_h)_{\partial\Omega^n} = (\tilde{f}^{n-1}, v_h)_{\Omega^n}. \tag{4.8}$$

By adding the equations (4.5) and (4.8), we obtain

$$(u_t^n + \tilde{u}_t^{n-1}, v_h)_{\Omega^n} + (\nabla u^n + \nabla\tilde{u}^{n-1}, \nabla v_h)_{\Omega^n} - (\partial_n u_h^n + \partial_n\tilde{u}_h^{n-1}, v_h)_{\partial\Omega^n} = (f^n + \tilde{f}^{n-1}, v_h)_{\Omega^n}. \tag{4.9}$$

We note that the right-hand side in (4.9) differs from the discrete formulation (2.7) by $(\tilde{f}^{n-1} - f^{n-1}, v_h)_{\Omega^n\setminus\Omega^{n-1}}$. Hence, (4.9) can be rewritten as

$$\left(\frac{u_t^n + \tilde{u}_t^{n-1}}{2}, v_h\right)_{\Omega^n} + \frac{1}{2}a_h^n(u^n, v_h) + \frac{1}{2}a_h^{n-1}(\tilde{u}^{n-1}, v_h) = \frac{1}{2}(f^n + f^{n-1}, v_h)_{\Omega^n} + \frac{1}{2}\mathcal{E}_f^{n-1}(v_h), \tag{4.10}$$

where $\mathcal{E}_f^{n-1}(v_h)$ is given by

$$\mathcal{E}_f^{n-1}(v_h) = (\tilde{f}^{n-1} - f^{n-1}, v_h)_{\Omega^n\setminus\Omega^{n-1}}. \tag{4.11}$$

In the next lines, we will discuss a bound for the term $\mathcal{E}_f^{n-1}$.

Lemma 4.1 *Under the assumptions of Lemma 3.1, the error term $\mathcal{E}_f^{n-1}$ defined in* (4.11) *satisfies the following estimate for $v_h^n \in V_h^{n,m}$ and sufficiently regular u:*

$$\Delta t\big|\mathcal{E}_f^{n-1}(v_h^n)\big| \le c\Delta t^5\mathcal{R}_C(u)^2 + \frac{1}{32}\Big(\Delta t\|\nabla v_h^n + \nabla v_h^{n-1}\|_{\Omega^n}^2 + \|v_h^n - v_h^{n-1}\|_{\Omega^n}^2 \tag{4.12}$$

$$+ \Delta t^2\|v_h^n\|_{\partial\Omega^n}^2 + \Delta t g_h^n(v_h^n, v_h^n) + \Delta t g_h^{n-1}(v_h^{n-1}, v_h^{n-1})\Big),$$

where $\mathcal{R}_C(u) = \|u\|_{\infty,3,I} + \|\partial_t u\|_{\infty,1,I}$.

Proof. We apply Lemma 3.3 to $u = (f^{n-1} - \tilde{f}^{n-1})v_h^n$ for $p = 1$

$$\begin{aligned}\Delta t\big|(f^{n-1} - \tilde{f}^{n-1}, v_h^n)_{\Omega^n\setminus\Omega^{n-1}}\big| &\le \Delta t\|(f^{n-1} - \tilde{f}^{n-1})v_h^n\|_{L^1(\Omega^n\setminus\Omega^{n-1})} \\ &\le c\Delta t^2\|(f^{n-1} - \tilde{f}^{n-1})v_h^n\|_{L^1(\partial\Omega^{n-1})} \\ &\quad + c\Delta t^3\left\|\nabla\big[(f^{n-1} - \tilde{f}^{n-1})v_h^n\big]\right\|_{L^1(\Omega^n\setminus\Omega^{n-1})}.\end{aligned} \tag{4.13}$$

As $f^{n-1} = \tilde{f}^{n-1}$ on $\partial\Omega^{n-1}$, the first term in (4.13) vanishes. By definition of $\tilde{f}$ in (4.7) we can estimate further

$$\begin{aligned}\Delta t\big|(f^{n-1}-\tilde{f}^{n-1}, v_h^n)_{\Omega^n\setminus\Omega^{n-1}}\big| &\leq c\Delta t^3\left[\|(f^{n-1}-\tilde{f}^{n-1})\|_{\Omega^n\setminus\Omega^{n-1}}\|\nabla v_h^n\|_{\Omega^n\setminus\Omega^{n-1}}\right]\\ &\quad +c\Delta t^3\left[\|\nabla(f^{n-1}-\tilde{f}^{n-1})\|_{\Omega^n\setminus\Omega^{n-1}}\|v_h^n\|_{\Omega^n\setminus\Omega^{n-1}}\right]\\ &\leq c\Delta t^3\mathcal{R}_C(u)\left[\|\nabla v_h^n\|_{\Omega^n\setminus\Omega^{n-1}}+\|v_h^n\|_{\Omega^n\setminus\Omega^{n-1}}\right].\end{aligned} \tag{4.14}$$

Now, by employing Lemmas 3.2 and 3.4, we obtain

$$\begin{aligned}\|\nabla v_h^n\|^2_{\Omega^n\setminus\Omega^{n-1}}+\|v_h^n\|^2_{\Omega^n\setminus\Omega^{n-1}} &\leq \left(\frac{1}{2}+c\Delta t^2\right)\|\nabla v_h^n+\nabla v_h^{n-1}\|^2_{\Omega^n\setminus\Omega^{n-1}}+\frac{c}{\Delta t}\|v_h^n-v_h^{n-1}\|^2_{\Omega^n}\\ &\quad +c\Delta t\|v_h^n\|^2_{\partial\Omega^n}+cg_h^n(v_h^n,v_h^n)+cg_h^{n-1}(v_h^{n-1},v_h^{n-1}).\end{aligned} \tag{4.15}$$

Inserting these estimates into (4.14) and using Young's inequality and Assumption 2, we obtain

$$\begin{aligned}\Delta t\big|(f^{n-1}-\tilde{f}^{n-1}, v_h^n)_{\Omega^n\setminus\Omega^{n-1}}\big| &\leq c\Delta t^3\mathcal{R}_C(u)\left[\|\nabla v_h^n\|_{\Omega^n\setminus\Omega^{n-1}}+\|v_h^n\|_{\Omega^n\setminus\Omega^{n-1}}\right]\\ &\leq c\Delta t^5\mathcal{R}_C(u)^2+\frac{1}{32}\Big(\Delta t\|\nabla v_h^n+\nabla v_h^{n-1}\|^2_{\Omega^n}+\|v_h^n-v_h^{n-1}\|^2_{\Omega^n}+\Delta t^2\|v_h^n\|^2_{\partial\Omega^n}\\ &\quad +\Delta t\gamma_g g_h^n(v_h^n,v_h^n)+\Delta t\gamma_g g_h^{n-1}(v_h^{n-1},v_h^{n-1})\Big).\end{aligned} \tag{4.16}$$

■

Now we are ready to estimate the consistency error related to the discrete problem (2.7). By subtracting (2.7) from (4.10), the global error term e^n satisfies the equality

$$\left(D^-_{\Delta t}e^n, v_h\right)_{\Omega^n}+\frac{1}{2}a_h^n(e^n,v_h)+\frac{1}{2}a_h^{n-1}(e^{n-1},v_h)+\frac{\gamma_D}{h}(e^n,v_h)_{\partial\Omega^n}+\gamma_g g_h^n(e^n,v_h)=\frac{1}{2}\mathcal{E}_f^{n-1}(v_h)+\mathcal{E}_c^n(v_h), \tag{4.17}$$

where the consistency error $\mathcal{E}_c^n(v_h)$ is given by

$$\begin{aligned}\mathcal{E}_c^n(v_h) &= \underbrace{\left(D^-_{\Delta t}u^n-\partial_t E^n u(t_{n-1/2}), v_h\right)_{\Omega^n}}_{I_1}-\underbrace{\left(\frac{u_t(t_n)+\partial_t E^n u(t_{n-1})}{2}-\partial_t E^n u(t_{n-1/2}), v_h\right)_{\Omega^n}}_{I_2}\\ &\quad +\underbrace{\frac{\gamma_D}{h}(E^n u^n, v_h)_{\partial\Omega^n}}_{I_3}+\underbrace{\gamma_g g_h^n(E^n u^n, v_h)}_{I_4}.\end{aligned} \tag{4.18}$$

The terms I_3 and I_4 vanish due to the homogeneous boundary condition and the regularity assumption on the exact solution $u^n \in H^2(\Omega^n)$ and its extension $E^n u^n \in H^2(\Omega^n_\delta)$. The remaining terms are estimated in the following lemma.

Lemma 4.2 *Let $u \in W^{3,\infty}(Q^n)$. Under Assumption 1, the consistency error for $v_h \in V_h^{n,m}$ is bounded by*

$$\Delta t|\mathcal{E}_c^n(v_h^n)| \leq c\Delta t^5\|u\|^2_{W^{3,\infty}(Q^n)}+\frac{\Delta t}{16}\|v_h^n\|^2_{\Omega^n}. \tag{4.19}$$

Proof. First, we will show a bound for the term I_1. By following the argumentation in [59, Chapter 1], we have

$$\begin{aligned}\left|\left(\frac{u^n - E^n u^{n-1}}{\Delta t} - \partial_t E^n u(t_{n-1/2})\right)\right| &\leq \frac{1}{\Delta t}\left|\int_{t_{n-1/2}}^{t^n} \frac{(t_n - s)^2}{3!}\partial_t^3 E^n u(s) ds\right| \\ &\quad + \frac{1}{\Delta t}\left|\int_{t^{n-1}}^{t_{n-1/2}} \frac{(s - t_{n-1})^2}{3!}\partial_t^3 E^n u(s) ds\right| \\ &\leq c\Delta t^2 \sup_{t\in[t_{n-1},t_n]} |\partial_t^3 E^n u(t)|.\end{aligned}$$

Using the stability of the extension operator E^n given in (2.5) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\left|\left(\frac{u^n - E^n u^{n-1}}{\Delta t} - \partial_t E^n u(t_{n-1/2}), v_h^n\right)_{\Omega^n}\right| &\leq c\Delta t^2 \|\partial_t^3 E^n u^n\|_{L^\infty(Q_\delta^n)} \|v_h^n\|_{\Omega^n} \\ &\leq c\Delta t^2 \|u\|_{W^{3,\infty}(Q^n)} \|v_h^n\|_{\Omega^n}.\end{aligned} \tag{4.20}$$

A bound for the second term I_2 follows in a similar way, see also [59, Chapter 1]. The statement follows using Young's inequality. ∎

4.2 Interpolation error

To derive an interpolation error estimate, we devise a discrete problem associated with the discrete error ξ_h^n. By definition of ξ_h^n (4.2) and using (4.17), we have for $v_h \in V_h^{n,m}$

$$\begin{aligned}\left(D_{\Delta t}^- \xi_h^n, v_h\right)_{\Omega^n} + \frac{1}{2}a_h^n(\xi_h^n, v_h) + \frac{1}{2}a_h^n(\xi_h^{n-1}, v_h) + \frac{\gamma_D}{h}(\xi_h^n, v_h)_{\partial\Omega^n} \\ + \gamma_g g_h^n(\xi_h^n, v_h) = \frac{1}{2}\mathcal{E}_f^{n-1}(v_h) + \mathcal{E}_c^n(v_h) - \mathcal{E}_I^n(v_h),\end{aligned} \tag{4.21}$$

where the interpolation error $\mathcal{E}_I^n(v_h)$ is given by

$$\mathcal{E}_I^n(v_h) = \left(D_{\Delta t}^- \eta^n, v_h\right)_{\Omega^n} + \frac{1}{2}a_h^n(\eta^n, v_h) + \frac{1}{2}a_h^n(\eta^{n-1}, v_h) + \frac{\gamma_D}{h}(\eta^n, v_h)_{\partial\Omega^n} + \gamma_g g_h^n(\eta^n, v_h). \tag{4.22}$$

We remark that we are using smooth extensions of, e.g., η^{n-1}, in the following without further notice, whenever variables would be undefined on parts of Ω^n.

Lemma 4.3 (Interpolation error) *Let $u \in L^\infty(I_n, H^{m+1}(\Omega(t))$, $u_t \in L^\infty(I_n, H^m(\Omega(t))$ and $v_h \in V_h^{n,m}$. Based on the assumptions 1 and 2, the interpolation error is bounded by*

$$\begin{aligned}\Delta t|\mathcal{E}_I^n(v_h^n)| &\leq c\Delta t\, h^{2m}\mathcal{R}_I(u)^2 + \frac{\Delta t}{8}\|v_h^n\|_{\Omega^n}^2 + \frac{\Delta t}{32}\|\nabla v_h^n + \nabla v_h^{n-1}\|_{\Omega^n}^2 + \frac{1}{32}\|v_h^n - v_h^{n-1}\|_{\Omega^n}^2 \\ &\quad + \frac{\Delta t}{16h}\|v_h^n\|_{\partial\Omega^n}^2 + \frac{\Delta t}{8}g_h^n(v_h^n, v_h^n),\end{aligned} \tag{4.23}$$

where $\mathcal{R}_I(u) = \|u_t\|_{\infty,m} + \|u\|_{\infty,m+1}$.

Proof. The first term in the interpolation error $\mathcal{E}_I^n(v_h)$ from (4.22) is estimated as follows:

$$\begin{aligned}\left|\left(\eta^n-\eta^{n-1},v_h^n\right)_{\Omega^n}\right| &\leq \|\eta^n-\eta^{n-1}\|_{\Omega^n}\|v_h^n\|_{\Omega^n}\\ &= \left\|\int_{t_{n-1}}^{t_n}\eta_t(s)ds\right\|_{\Omega^n}\|v_h^n\|_{\Omega^n} \leq \Delta t h^m\|\partial_t E^n u\|_{\infty,m,I_n}\|v_h^n\|_{\Omega^n}.\end{aligned} \tag{4.24}$$

Using the stability of the extension (2.4), we obtain

$$\left|\left(\eta^n-\eta^{n-1},v_h^n\right)_{\Omega^n}\right| \leq c\Delta t\, h^m\left(\|u_t\|_{\infty,m}+\|u\|_{\infty,m+1}\right)\|v_h^n\|_{\Omega^n}. \tag{4.25}$$

Next, we use interpolation estimates (4.3) and (4.4) in combination with (2.3) to deduce for $k\in\{n-1,n\}$ that

$$\begin{aligned}\frac{\Delta t}{2}\left|a_h^n(\eta^k,v_h^n)\right| &\leq \frac{\Delta t}{2}\|\nabla\eta^k\|_{\Omega^n}\|\nabla v_h^n\|_{\Omega^n}+\Delta t\|\partial_n\eta^k\|_{\partial\Omega^n}\|v_h^n\|_{\partial\Omega^n}\\ &\leq c\Delta t\, h^m\|u^k\|_{H^{m+1}(\Omega^k)}\|\nabla v_h^n\|_{\Omega^n}+c\Delta t\, h^{m-1/2}\|u^k\|_{H^{m+1}(\Omega^k)}\|v_h^n\|_{\partial\Omega^n}.\end{aligned} \tag{4.26}$$

Using that, by an inverse inequality and the CFL condition $\Delta t\leq c_{CFL}h^2$,

$$\begin{aligned}\|\nabla v_h^n\|_{\Omega^n} &\leq \frac{1}{2}\|\nabla v_h^n+\nabla v_h^{n-1}\|_{\Omega^n}+\frac{1}{2}\|\nabla v_h^n-\nabla v_h^{n-1}\|_{\Omega^n}\\ &\leq \frac{1}{2}\|\nabla v_h^n+\nabla v_h^{n-1}\|_{\Omega^n}+\frac{c}{h}\|v_h^n-v_h^{n-1}\|_{\Omega^n}\\ &\leq \frac{1}{2}\|\nabla v_h^n+\nabla v_h^{n-1}\|_{\Omega^n}+\frac{c}{\Delta t^{1/2}}\|v_h^n-v_h^{n-1}\|_{\Omega^n}\end{aligned}$$

we obtain further

$$\begin{aligned}\frac{\Delta t}{2}\left|a_h^n(\eta^k,v_h^n)\right| &\leq c\Delta t^{1/2}\, h^m\|u^k\|_{H^{m+1}(\Omega^n)}\left(\frac{\Delta t^{1/2}}{2}\|\nabla v_h^n+\nabla v_h^{n-1}\|_{\Omega^n}+\|v_h^n-v_h^{n-1}\|_{\Omega^n}\right)\\ &\quad +c\Delta t^{1/2}\, h^m\|u^k\|_{H^{m+1}(\Omega^n)}\frac{\Delta t^{1/2}}{h^{1/2}}\|v_h^n\|_{\partial\Omega^n}.\end{aligned} \tag{4.27}$$

For the Nitsche penalty term, we have

$$\Delta t\left|\frac{\gamma_D}{h}(\eta^n,v_h^n)_{\partial\Omega^n}\right| \leq c\Delta t\frac{\gamma_D}{h}\|\eta^n\|_{\partial\Omega^n}\|v_h^n\|_{\partial\Omega^n} \leq c\Delta t^{1/2}\, h^m\|u^n\|_{H^{m+1}(\Omega^n)}\frac{\Delta t^{1/2}}{h^{1/2}}\|v_h^n\|_{\partial\Omega^n}. \tag{4.28}$$

Finally, the ghost-penalty term is approximated by using $g_h^n(\eta^n,v_h^n)\leq g_h^n(\eta^n,\eta^n)^{1/2}g_h^n(v_h^n,v_h^n)^{1/2}$ and an interpolation estimate

$$\Delta t\left|\gamma_g g_h^n(\eta^n,v_h^n)\right| \leq c\Delta t\, h^m\|u^n\|_{H^{m+1}(\Omega^n)}g_h^n(v_h^n,v_h^n)^{1/2}. \tag{4.29}$$

The statement (4.23) follows by combining estimates (4.25)-(4.29) and using Young's inequality. ■

4.3 Convergence estimate

Lemma 4.4 (Discrete error) *Let $u\in L^\infty(I_n,H^{m+1}(\Omega(t)))\cap W^{1,\infty}(I_n,H^m(\Omega(t))\cap W^{3,\infty}(Q)$ be the solution of (1.1) and $\{u_h^k\}_{k=1}^n$ the discrete solution of (2.7), respectively. Under Assumptions 1 and 2, the discrete error term ξ^n satisfies for γ_g,γ_D sufficiently large*

$$\|\xi^n\|_{\Omega^n}^2+\Delta t\|\nabla\xi^n\|_{\Omega^n}^2+\Delta t\sum_{k=1}^n\mathfrak{E}^k(\xi_h^k,\xi_h^{k-1}) \leq \exp(c_{T4.4}t_n)\left(\Delta t^4+h^{2m}\right)\mathcal{R}(u)^2, \tag{4.30}$$

where $\mathcal{R}(u) = \mathcal{R}_C(u) + \mathcal{R}_I(u) + \|u(t_n)\|_{W^{3,\infty}(Q)}$, with $\mathcal{R}_C$ and $\mathcal{R}_I$ specified in Lemma 4.1 and 4.3, respectively.

Proof. By taking $v_h = 2\Delta t \xi_h^n$ in (4.21) and using the argumentation from the stability proof (Theorem 3.5), see (3.17), we obtain that

$$\begin{aligned}
&\|\xi_h^n\|_{\Omega^n}^2 + \frac{1}{4}\|\xi_h^n - \xi_h^{n-1}\|_{\Omega^n}^2 + \frac{\Delta t}{2}\|\nabla \xi_h^n\|_{\Omega^n}^2 + \frac{\Delta t}{8}\|\nabla \xi_h^n + \nabla \xi_h^{n-1}\|_{\Omega^n}^2 + \frac{3\gamma_D \Delta t}{2h}\|\xi_h^n\|_{\partial\Omega^n}^2 \\
&+2\Delta t \gamma_g g_h(\xi_h^n, \xi_h^n) \leq \|\xi_h^{n-1}\|_{\Omega^{n-1}}^2 + \frac{\Delta t}{2}\|\nabla \xi_h^{n-1}\|_{\Omega^{n-1}}^2 + c\Delta t \|\xi_h^{n-1}\|_{\partial\Omega^{n-1}}^2 \\
&\qquad +c\Delta t g_h^{n-1}(\xi_h^{n-1}, \xi_h^{n-1}) + 2\Delta t \left(|\mathcal{E}_f^{n-1}(\xi_h^n)| + |\mathcal{E}_c^n(\xi_h^n)| + |\mathcal{E}_I^n(\xi_h^n)| \right).
\end{aligned} \tag{4.31}$$

By combining results from Lemmas 4.1, 4.2 and 4.3 we have

$$\begin{aligned}
&2\Delta t \left(|\mathcal{E}_f^{n-1}(\xi_h^n)| + |\mathcal{E}_c^n(\xi_h^n)| + |\mathcal{E}_I^n(\xi_h^n)| \right) \\
&\qquad \leq c\Delta t \left(\Delta t^4 + h^{2m} \right) \mathcal{R}(u) + \frac{\Delta t}{4}\|\xi_h^n\|_{\Omega^n}^2 + \frac{\Delta t}{16}\|\nabla \xi_h^n + \nabla \xi_h^{n-1}\|_{\Omega^n}^2 + \frac{1}{16}\|\xi_h^n - \xi_h^{n-1}\|_{\Omega^n}^2 \\
&\qquad + \frac{\Delta t}{8h}\|\xi_h^n\|_{\partial\Omega^n}^2 + c\Delta t g_h^n(\xi_h^n, \xi_h^n) + c\Delta t g_h^{n-1}(\xi_h^{n-1}, \xi_h^{n-1}).
\end{aligned} \tag{4.32}$$

Inserting (4.32) into (4.31) and absorbing terms into the left-hand side yields for $n > 1$ and γ_D, γ_g sufficiently large

$$\begin{aligned}
&\left(1 - \frac{\Delta t}{4}\right)\|\xi_h^n\|_{\Omega^n}^2 + \frac{1}{4}\|\xi_h^n - \xi_h^{n-1}\|_{\Omega^n}^2 + \frac{\Delta t}{4}\|\nabla \xi_h^n\|_{\Omega^n}^2 + \frac{\Delta t}{16}\|\nabla \xi_h^n + \nabla \xi_h^{n-1}\|_{\Omega^n}^2 \\
&\qquad + \frac{\gamma_D \Delta t}{h}\|\xi_h^n\|_{\partial\Omega^n}^2 + \Delta t \gamma_g g_h(\xi_h^n, \xi_h^n) \leq \|\xi_h^{n-1}\|_{\Omega^{n-1}}^2 + \frac{\Delta t}{2}\|\nabla \xi_h^{n-1}\|_{\Omega^{n-1}}^2 + c\Delta t \|\xi_h^{n-1}\|_{\partial\Omega^{n-1}}^2 \\
&\qquad\qquad + c\Delta t g_h^{n-1}(\xi_h^{n-1}, \xi_h^{n-1}) + c\Delta t \left(\Delta t^4 + h^{2m} \right) \mathcal{R}(u).
\end{aligned} \tag{4.33}$$

As $\xi_h^0 = 0$, we obtain for $n = 1$

$$\begin{aligned}
&\left(1 - \frac{\Delta t}{4}\right)\|\xi_h^1\|_{\Omega^1}^2 + \frac{1}{4}\|\xi_h^1 - \xi_h^0\|_{\Omega^1}^2 + \frac{\Delta t}{4}\|\nabla \xi_h^1\|_{\Omega^1}^2 + \frac{\Delta t}{16}\|\nabla \xi_h^1 + \nabla \xi_h^0\|_{\Omega^1}^2 \\
&\qquad + \frac{\gamma_D \Delta t}{h}\|\xi_h^1\|_{\partial\Omega^1}^2 + \Delta t \gamma_g g_h^1(\xi_h^1, \xi_h^1) \leq c\Delta t \left(\Delta t^4 + h^{2m} \right) \mathcal{R}(u).
\end{aligned}$$

Now, summing over $k = 1, 2, \ldots, n$, we deduce that

$$\|\xi_h^n\|_{\Omega^n}^2 + \Delta t \|\nabla \xi_h^n\|_{\Omega^n}^2 + \frac{\Delta t}{8}\sum_{k=1}^{n} \mathfrak{E}^k(\xi_h^k, \xi_h^{k-1}) \leq c\, t_n \left(\Delta t^4 + h^{2m} \right) \mathcal{R}(u) + c\Delta t \sum_{k=2}^{n} \|\xi_h^{k-1}\|_{\Omega^{k-1}}^2. \tag{4.34}$$

Finally, we use the discrete Gronwall lemma to conclude the result. This completes the proof. ■

Theorem 4.5 (Global error) *Based on the assumption made in Lemma 4.4, the global error* $e^k = u^k - u_h^k, k = 1, ..., n$ *satisfies*

$$\|e^n\|_{\Omega^n}^2 + \Delta t \|\nabla e^n\|_{\Omega^n}^2 + \Delta t \sum_{k=1}^{n} \mathfrak{E}^k(e^k, e^{k-1}) \leq \exp(c_{T4.5} t_n) \left(\Delta t^4 + h^{2m}\right) \mathcal{R}(u)^2, \tag{4.35}$$

where $\mathcal{R}(u)$ *is defined in Lemma 4.4.*

Proof. Using the interpolation error estimates (4.3), (4.4) and (4.25), we deduce that

$$\|\eta_h^n\|_{\Omega^n}^2 + \Delta t \|\nabla \eta_h^n\|_{\Omega^n}^2 + \Delta t \sum_{k=1}^{n} \mathfrak{E}^k(\eta_h^k, \eta_h^{k-1}) \leq c\, h^{2m} \mathcal{R}_I(u)^2. \tag{4.36}$$

In combination with Lemma 4.4, this proves (4.35). ■

5 Numerical results

In this section, we show numerical results in two and three space dimensions to verify the theoretical findings and the practical behaviour of the numerical method. All the numerical experiments have been obtained using the CutFEM library [7], which is based on FEniCS [1].

To verify the theoretical results, we will analyse the error terms $e^k = u^k - u_h^k, k = 1, ..., n$ in the following norms

$$\|e^n\|_{L^2(\Omega^n)},\ \|e\|_{L^2(L^2)} = \left(\Delta t \sum_{k=1}^{n} \|e^k\|_{L^2(\Omega^k)}^2\right)^{1/2},\ \|e\|_{L^2(H^1_{av})} = \left(\Delta t \sum_{k=1}^{n} \|\nabla e^k + \nabla e^{k-1}\|_{\Omega^k}^2\right)^{1/2}.$$

Given the respective CFL condition (Assumption 2), Theorem 4.5 guarantees second-order convergence in time and convergence of order m in space in the L^2-norm at the end time $\|e^n\|_{L^2(\Omega^n)}$ and in the averaged $L^2(H^1)$-norm $\|e\|_{L^2(H^1_{av})}$.

5.1 2d example

Example 5.1 *We consider a circle traveling with constant velocity* $w = (1, 0)$ *towards the right. The domain is given by*

$$\Omega(t) = \{(x, y) : (x - 0.5 - t)^2 + (y - 0.5)^2 \leq 0.9\}$$

in the time interval $I = [0, 0.1]$*. The data of the model example is chosen in such a way that the exact solution is*

$$u(x, y, t) = \exp(-4\pi^2 t) \cos(2\pi x) \cos(2\pi y).$$

An illustration of the numerical solution is given in Figure 3. As background domain D, we use the unit square, i.e. $D = [0, 1]^2$. The background triangulations $\mathcal{T}_h$ are created by a uniform subdivision of the unit square into triangles and successive refinement. For each time step n, the active triangulation $\mathcal{T}_{h,\delta}^n$ is then extracted from $\mathcal{T}_h$, as described in Section 2. We use $\delta = 4\Delta t$.

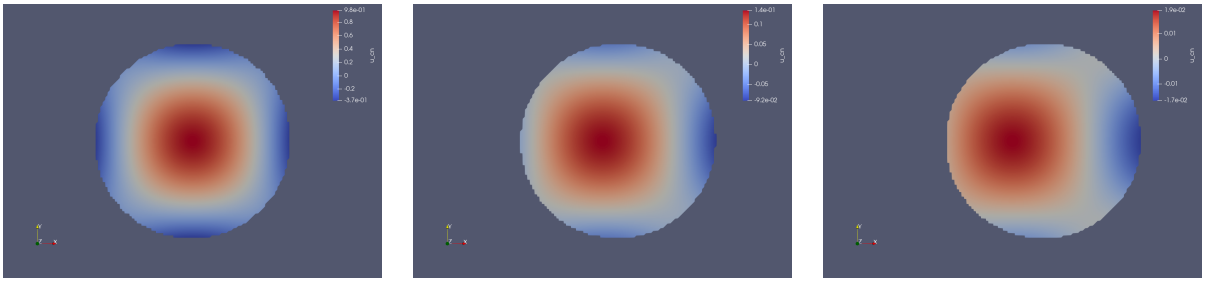


Figure 3: Illustration of the numerical solution u at $t = 0$ (left), $t = 0.05$ (center) and $t = 0.1$ (right).

Estimated orders of convergence We will show results for different time-step sizes $\Delta t_i = 1/50 \cdot 2^{-i}, i = 0, ..., 4$ and mesh sizes $h_j = 1/32 \cdot 2^{-j}, j = 0, ..., 3$. From the computed errors, we will estimate the temporal and spatial order of convergence. Therefore, we assume that the total error can be decomposed into a temporal and a spatial component as follows

$$g(\Delta t, h) = g_{\Delta t}(\Delta t) + g_h(h) = c_{\Delta t} \Delta t^{eoc_{\Delta t}} + c_h h^{eoc_h}$$

with constants $c_{\Delta t}, c_h$ and estimated orders of convergence $eoc_{\Delta t}$, eoc_h. For a fixed mesh size h_j, this relation becomes

$$g(\Delta t, \cdot) = g_{h_j} + c_{\Delta t} \Delta t^{eoc_{\Delta t}} \tag{5.1}$$

with a fixed spatial error part g_{h_j}. We will use (5.1) to estimate the order of convergence in time $eoc_{\Delta t}$ by means of a least-squares fit of all available error values for a fixed h_j to find the three parameters g_{h_j}, $c_{\Delta t}$ and $eoc_{\Delta t}$. Analogously, we estimate the spatial order of convergence eoc_h by a least-squares fit of the function

$$g(\cdot, h) = g_{\Delta t_i} + c_h h^{eoc_h} \tag{5.2}$$

against all available error values for a fixed time-step size Δt_i.

Finally, we will also compute estimated orders of convergence for the "diagonal values", which correspond to fixing $\Delta t = \bar{c} h$ for $\bar{c} \in \{32/50, 32/100\}$. Here we fit the two parameters $c_{\Delta t,h}$ and $eoc_{\Delta t,h}$ of the function

$$g(\bar{c} h, h) = c_{\Delta t,h} h^{eoc_{\Delta t,h}}. \tag{5.3}$$

against the computed error values.

5.1.1 P_1 finite elements

Firstly, we consider P_1 finite elements. We choose the numerical parameters as $\gamma_D = 1$ and $\gamma_g = 10^{-3}$. The errors in the L^2-norm at the end time, the $L^2(L^2)$-norm and the $L^2(H^1_{av})$-norm are shown in Table 1 for different Δt and h. The estimated orders of convergence are shown, if the asymptotic standard error (computed by gnuplot [63]) was below 20%; otherwise we draw a '-'.

We observe estimated spatial convergence orders close to two in the L^2-norms and close to one in the $L^2(H^1_{\text{av}})$-norm. Note that Theorem 4.5 guarantees only first-order convergence in space. We expect, however, that using a duality argument second-order convergence in space could be shown in the $L^2(L^2)$-norm, as in [11].

End-time error $\|e^n\|_{L^2(\Omega^n)}$

$h\downarrow/\Delta t\rightarrow$	$1/50$	$1/100$	$1/200$	$1/400$	$1/800$	$\mathbf{eoc}_{\Delta t}$
$1/32$	1.93e-03	5.40e-04	3.14e-04	2.78e-04	2.68e-04	**2.62**
$1/64$	1.21e-03	2.77e-04	1.08e-04	7.84e-05	7.17e-05	**2.46**
$1/128$	8.88e-04	1.96e-04	4.95e-05	2.54e-05	1.96e-05	**2.29**
$1/256$	7.78e-04	1.57e-04	3.18e-05	1.15e-05	6.18e-06	**2.35**
$\mathbf{eoc}_h$	**1.26**	**1.55**	**1.80**	**1.91**	**1.92**	
$\mathbf{eoc}_{\Delta t,h}$				**2.79**	**2.30**	

Error $\|e\|_{L^2(L^2)}$

$h\downarrow/\Delta t\rightarrow$	$1/50$	$1/100$	$1/200$	$1/400$	$1/800$	$\mathbf{eoc}_{\Delta t}$
$1/32$	1.44e-03	8.42e-04	7.12e-04	6.87e-04	6.83e-04	**2.25**
$1/64$	9.69e-04	3.62e-04	2.21e-04	1.89e-04	1.82e-04	**2.11**
$1/128$	8.35e-04	2.37e-04	9.13e-05	5.72e-05	4.89e-05	**2.04**
$1/256$	7.97e-04	2.01e-04	5.75e-05	2.30e-05	1.45e-05	**2.05**
$\mathbf{eoc}_h$	**1.81**	**1.91**	**1.92**	**1.92**	**1.92**	
$\mathbf{eoc}_{\Delta t,h}$				**1.99**	**1.93**	

Error $\|e\|_{L^2(H^1_{av})}$

$h\downarrow/\Delta t\rightarrow$	$1/50$	$1/100$	$1/200$	$1/400$	$1/800$	$\mathbf{eoc}_{\Delta t}$
$1/32$	4.77e-02	4.53e-02	4.50e-02	4.49e-02	4.49e-02	**2.83**
$1/64$	2.64e-02	2.36e-02	2.33e-02	2.33e-02	2.33e-02	**3.33**
$1/128$	1.65e-02	1.23e-02	1.19e-02	1.19e-02	1.19e-02	**3.32**
$1/256$	1.26e-02	6.72e-03	6.06e-03	6.01e-03	6.00e-03	**3.21**
$\mathbf{eoc}_h$	**1.17**	**0.96**	**0.93**	**0.93**	**0.93**	
$\mathbf{eoc}_{\Delta t,h}$				**1.00**	**0.96**	

Table 1: $L^2(T), L^2(L^2)$ and $L^2(H^1_{\text{av}})$-norm errors for P_1 finite elements applied to Example 5.1. The estimated orders of convergence are computed according to (5.1)-(5.3). The diagonal orders are computed from the underlined error values.

The estimated temporal orders of convergence are close to two or larger in the L^2-norms. This is in agreement with Theorem 4.5. In the $L^2(H^1_{\text{av}})$-norm the estimated $\text{eoc}_{\Delta t}$ seems to be even larger than three. This has to be read carefully, however, as the spatial error part clearly dominates the overall error in this case.

Finally, the diagonal orders are around two in the $L^2(L^2)$-norm and even slightly higher in the L^2-norm at the end time. In the $L^2(H_{\text{av}})$-norm the spatial part is dominant and we obtain $\text{eoc}_{\Delta t,h}$ close the one, in agreement with Theorem 4.5.

5.1.2 P_2 finite elements

Next, we consider Example 5.1 with P_2 finite elements. We increase the Nitsche parameter to $\gamma_D = 10$, as for higher polynomial degree a larger Nitsche parameter is required, see e.g. [44]. The ghost-penalty parameter γ_g is still chosen as 10^{-3}, but now, according to (2.9), second derivatives are included in the ghost-penalty term. The L^2-norm errors at the end time, the $L^2(L^2)$- and the $L^2(H^1_{av})$-norm errors are shown in Table 2.

Firstly, we observe that the absolute values of the errors are significantly smaller compared to

End-time error $\|e^n\|_{L^2(\Omega^n)}$

$h\downarrow/\Delta t\rightarrow$	$1/100$	$1/200$	$1/400$	$1/800$	$1/1\,600$	$\mathbf{eoc}_{\Delta t}$
$1/32$	1.16e-04	3.09e-05	1.08e-05	5.80e-06	4.66e-06	**2.06**
$1/64$	1.17e-04	2.66e-05	7.33e-06	2.78e-06	1.44e-06	**2.18**
$1/128$	1.11e-04	2.43e-05	6.23e-06	1.76e-06	6.78e-07	**2.22**
$1/256$	1.12e-04	2.31e-05	5.78e-06	1.54e-06	4.36e-07	**2.29**
$\mathbf{eoc}_h$	–	**0.91**	**1.60**	**1.68**	**2.01**	
$\mathbf{eoc}_{\Delta t,h}$				**2.12**	**2.08**	

Error $\|e\|_{L^2(L^2)}$

$h\downarrow/\Delta t\rightarrow$	$1/100$	$1/200$	$1/400$	$1/800$	$1/1\,600$	$\mathbf{eoc}_{\Delta t}$
$1/32$	1.89e-04	5.35e-05	2.13e-05	1.27e-05	1.10e-05	**2.04**
$1/64$	1.84e-04	4.66e-05	1.36e-05	5.09e-06	3.16e-06	**2.03**
$1/128$	1.82e-04	4.50e-05	1.16e-05	3.32e-06	1.28e-06	**2.03**
$1/256$	1.82e-04	4.45e-05	1.12e-05	2.88e-06	8.32e-07	**2.03**
$\mathbf{eoc}_h$	**1.72**	**2.00**	**1.99**	**2.09**	**2.05**	
$\mathbf{eoc}_{\Delta t,h}$				**2.02**	**1.98**	

Error $\|e\|_{L^2(H^1_{av})}$

$h\downarrow/\Delta t\rightarrow$	$1/100$	$1/200$	$1/400$	$1/800$	$1/1\,600$	$\mathbf{eoc}_{\Delta t}$
$1/32$	3.42e-03	2.25e-03	2.17e-03	2.17e-03	2.17e-03	**3.88**
$1/64$	2.81e-03	9.11e-04	6.22e-04	6.03e-04	6.03e-04	**2.82**
$1/128$	2.80e-03	7.26e-04	2.45e-04	1.77e-04	1.72e-04	**2.22**
$1/256$	2.81e-03	7.14e-04	1.87e-04	6.84e-05	5.30e-05	**2.03**
$\mathbf{eoc}_h$	–	**2.91**	**2.11**	**1.89**	**1.86**	
$\mathbf{eoc}_{\Delta t,h}$				**1.90**	**1.85**	

Table 2: $L^2(T), L^2(L^2)$ and $L^2(H^1_{\text{av}})$-norm errors for P_2 finite elements applied to Example 5.1.

P_1 finite elements. The spatial orders of convergence are close to two in all norms for smaller Δt. We note that in Theorem 4.5 second order in space has been shown for the end-time L^2-norm and the $L^2(H^1_{av})$-norm. Using a duality argument, one could even hope for convergence order three in the $L^2(L^2)$-norm. We need to consider, however, that for these results the quadrature error related to a curved boundary has not been taken into account. In the CutFEM library used here, the geometry is approximated linearly in the set-up of the quadrature rule, see [7]. This can lead to a reduced order of convergence, namely order 1.5 in the H^1-norms and 2 in the L^2-norms, see [12] for results for a CutFEM approach applied to an elliptic problem on curved domains. This reduction can also be observed in the L^2-norm errors in the example considered here.

The estimated temporal orders of convergence are again close to two in the L^2-norms, which confirm the estimates in Theorem 4.5. In the $L^2(H^1_{\text{av}})$-norm the error is still clearly dominated by the spatial part for $h \geq \frac{1}{64}$. This changes, however, on the finer levels, where the temporal part gets dominant and the $\text{eoc}_{\Delta t}$ is very close to two, in agreement with Theorem 4.5.

The spatial and temporal convergence orders are confirmed by the "diagonal" orders, which are close to two in all cases.

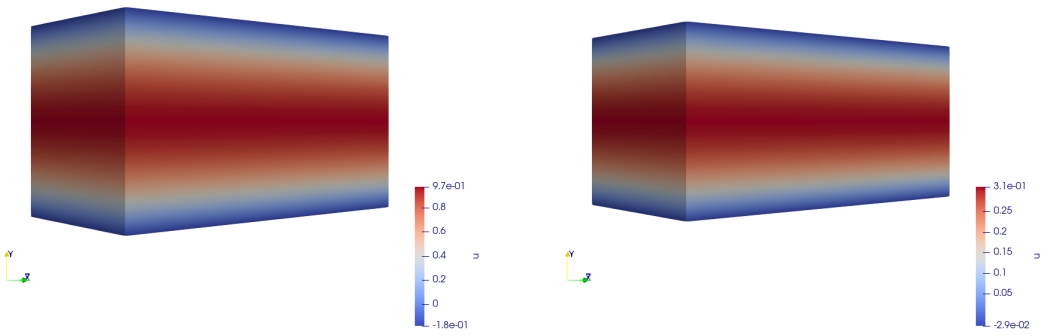


Figure 4: Illustration of the numerical solution of Example 5.2 at time $t = 0$ (left) and $t = 1$ (right).

5.2 3d example

Example 5.2 *We consider a 3-dimensional rectangular channel with a moving upper and lower wall in the time interval $I = [0, 1]$, inspired by a pump. The moving domain is given by*

$$\Omega(t) = (0, 4) \times (-1 + 0.1 \sin t, 1 - 0.1 \sin t) \times (-1, 1).$$

The source term and boundary data is chosen in such a way that the exact solution of the model problem (1.1) is

$$u(x, y, z, t) = \exp(-t) \left((1 - 0.1 \sin t)^2 - y^2 \right).$$

As background domain D, we use the box $[0, 4] \times [-1.1, 1.1] \times [-1, 1]$. The background triangulations $\mathcal{T}_h$ are created by uniform subdivisions of D into tetrahedra and successive refinement. We use again $\delta = 4\Delta t$ and choose $\gamma_g = 0.1$ for P_1 and $\gamma_g = 1$ for P_2 finite elements, respectively and $\gamma_D = 10$ in both cases. We note that in this example the quadrature error is zero, as the boundary $\partial\Omega^k$ consists of plane surfaces for all k. An illustration of the numerical solution at times $t = 0$ and $t = 1$ is given in Figure 4.

5.2.1 P_1 finite elements

As the numerical experiments in three space dimensions are much more time-consuming compared to two dimensions, we focus on simultaneous refinement in space and time by choosing $h_i = \frac{\Delta t_i}{10} = 2^{-i-1}$ for $i = 0, ..., 3$. The resulting errors in the four norms introduced above are plotted in Figure 5 over the mesh size (blue curves) and compared to linear (red) and quadratic convergence (pink). We observe second-order convergence in the L^2-norms and first-order convergence in the $L^2(H^1_{\text{av}})$-norm. We note that for $m = 1$ first-order convergence in space has been shown in Theorem 4.5 in the L^2-norm at the end time and in the $L^2(H^1_{av})$-norm. The numerical results indicate again that second-order convergence in space could be shown in the $L^2(L^2)$-norm using a duality argument. The first-order convergence in the $L^2(H^1_{\text{av}})$-norm is optimal, as the spatial error part dominates the overall error.

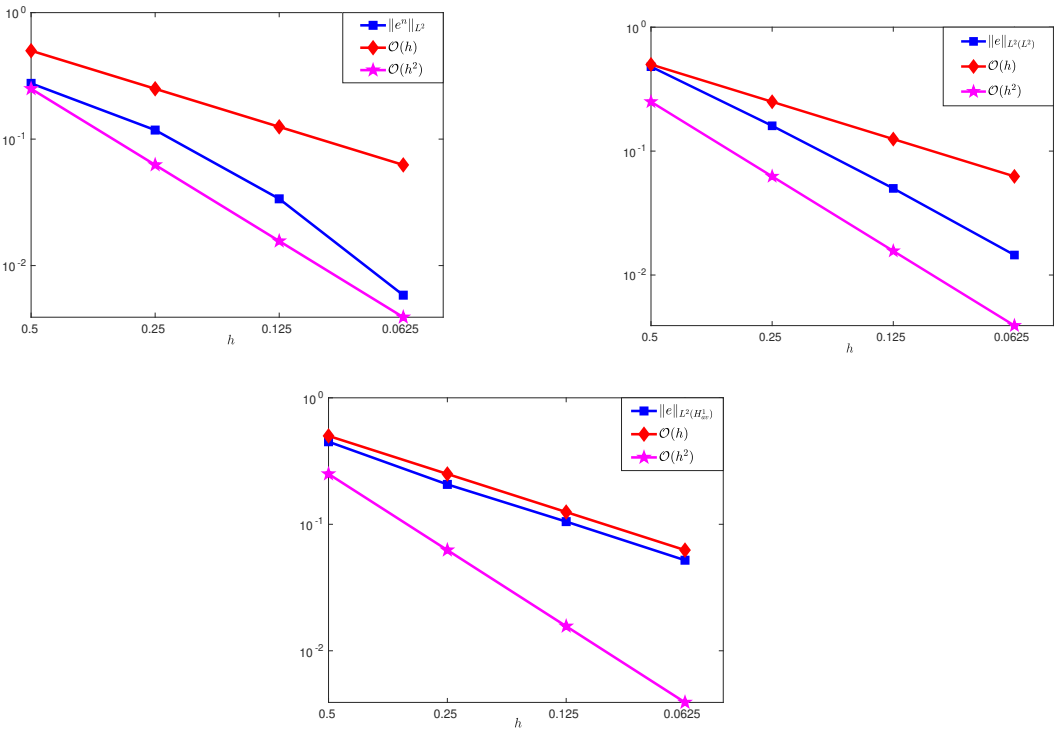


Figure 5: Errors for example 5.2 (3d) for P_1 finite elements under simultaneous refinement in space and time ($\Delta t = h/10$) for P_1 finite elements. **Top left**: L^2-norm at the end time. **Top right**: $L^2(L^2)$-norm. **Bottom**: $L^2(H^1_{\text{av}})$-norm.

5.2.2 P_2 finite elements

In Figure 6, we illustrate the errors under simultaneous refinement ($h_i = \frac{\Delta t_i}{10}$) for P_2 finite elements. We observe again (at least) second-order convergence in the L^2-norms, in agreement with Theorem 4.5. The convergence in the $L^2(H^1_{\text{av}})$-norm lies between linear and quadratic convergence and decreases slightly for finer mesh sizes. This reduction in the convergence rate could be related to a violation of the CFL condition, which has been used in the error analysis. On the other hand, we did not find any stability issues in our computations.

6 Concluding remarks

We have analysed a Crank-Nicolson variant of the implicitly extended Eulerian time-stepping scheme for the heat equation on time-dependent domains. Theoretically, stability and optimal-order convergence estimates were derived in the energy norm under the assumption of a parabolic CFL condition. In the numerical results, on the other hand, we did not observe any stability issue related to a violated CFL condition. The three-dimensional results for second-order polynomials indicate that a violated CFL condition could result in a slightly reduced convergence order in the $L^2(H^1_{\text{av}})$-norm.

To our knowledge this is the first work, in which an implicitly extended Eulerian time-stepping scheme is applied with a scheme that requires derivative information at different time steps. As mentioned in the introduction, this could be the basis for an analysis of a whole zoo of time-stepping

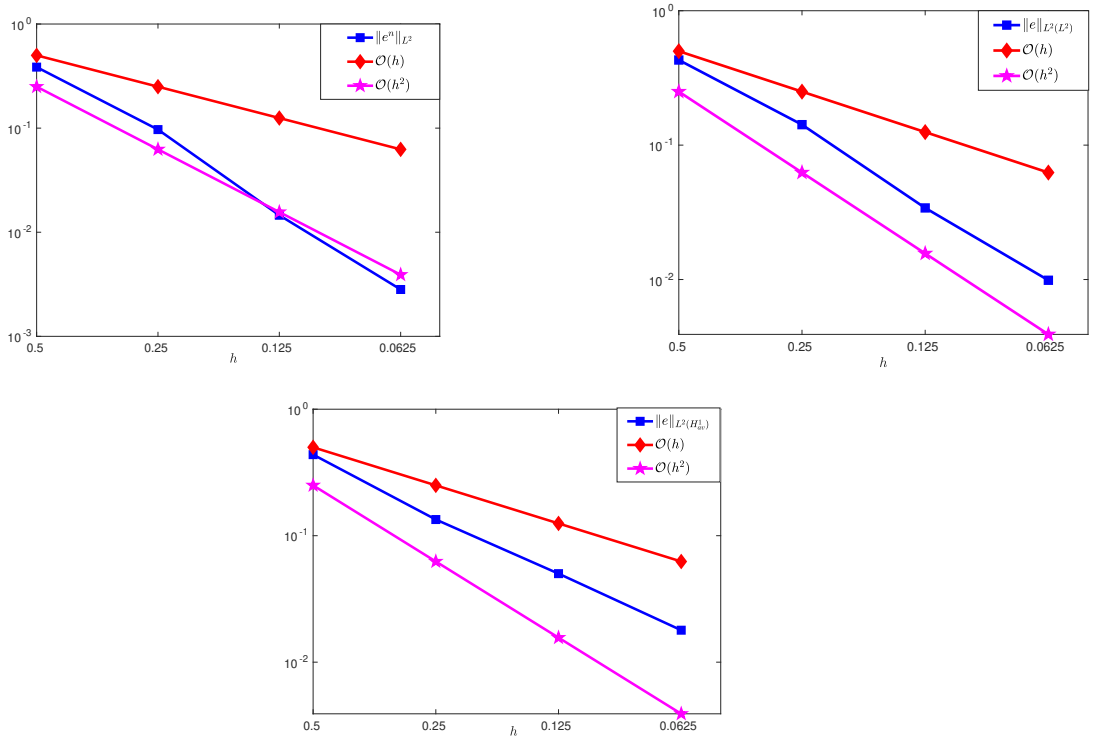


Figure 6: Errors for example 5.2 (3d) for P_2 finite elements under simultaneous refinement in space and time ($\Delta t = h/10$) for P_1 finite elements. **Top left**: L^2-norm at the end time. **Top right**: $L^2(L^2)$-norm. **Bottom**: $L^2(H^1_{av})$-norm.

schemes, such as the Fractional-step-θ method, implicit Runge-Kutta- or Adams-Bashforth schemes. Moreover, we plan to apply the developed time-stepping scheme to flow problems on time-dependent domains and to fluid-structure interactions with large displacements, see e.g. [27, 14].

References

[1] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, and G.N. Wells. The fenics project version 1.5. Arch. Numer. Soft., 3(100), 2015.

[2] I. Babuška, U. Banarjee, and J.E. Osborn. Generalized finite element methods: Main ideas, results, and perspective. Int. J. Comput. Methods, 1:67–103, 2004.

[3] Y. Bazilevs, K. Takizawa, and T. E. Tezduyar. Computational Fluid-Structure Interaction: Methods and Applications. John Wiley & Sons, 2013.

[4] R. Becker. An Adaptive Finite Element Method for the Incompressible Navier-Stokes Equations on Time-dependent Domains. PhD thesis, Habilitation Thesis, University of Heidelberg, Germany, 1995.

[5] J.H. Bramble and J.T. King. A finite element method for interface problems in domains with smooth boundaries and interfaces. Adv. Comput. Math., 6:109–138, 1996.

[6] E. Burman. Ghost penalty. CR Math, 348(21-22):1217–1220, 2010.

[7] E. Burman, S. Claus, P. Hansbo, M.G. Larson, and A. Massing. Cutfem: Discretizing geometry and partial differential equations. Int. J. Numer. Methods Eng., 104(7):472–501, 2015.

[8] E. Burman, M. A. Fernández, and S. Frei. A Nitsche-based formulation for fluid-structure interactions with contact. ESAIM: M2AN, 54(2):531–564, 2020.

[9] E. Burman, M. A. Fernández, S. Frei, and F.M. Gerosa. A mechanically consistent model for fluid–structure interactions with contact including seepage. Comput. Methods Appl. Mech. Eng., 392:114637, 2022.

[10] E. Burman and M.A. Fernández. An unfitted Nitsche method for incompressible fluid–structure interaction using overlapping meshes. Comput. Methods Appl. Mech. Eng., 279:497–514, 2014.

[11] E. Burman, S. Frei, and A. Massing. Eulerian time-stepping schemes for the non-stationary Stokes equations on time-dependent domains. Numer. Math., pages 1–56, 2022.

[12] E. Burman, P. Hansbo, and M.G. Larson. A cut finite element method with boundary value correction. Math. Comput., 87(310):633–657, 2018.

[13] E. Burman, P. Hansbo, M.G. Larson, and S. Zahedi. Cut finite element methods for coupled bulk–surface problems. Numer. Math., 133(2):203–231, 2016.

[14] Erik Burman, Miguel A Fernández, and Stefan Frei. A Nitsche-based formulation for fluid-structure interactions with contact. ESAIM: M2AN, 54(2):531–564, 2020.

[15] L.J. Caucha, S. Frei, and O. Rubio. Finite element simulation of fluid dynamics and CO_2 gas exchange in the alveolar sacs of the human lung. Comput. Appl. Math., 37(5):6410–6432, 2018.

[16] J. Chessa and T. Belytschko. An extended finite element method for two-phase fluids. J. Appl. Mech., 70(1):10–17, 2003.

[17] S. Claus and P. Kerfriden. A CutFEM method for two-phase flow problems. Comput. Methods Appl. Mech. Eng., 348:185–206, 2019.

[18] R. Codina, G. Houzeaux, H. Coppola-Owen, and J. Baiges. The Fixed-Mesh ALE approach for the numerical approximation of flows in moving domains. J. Comput. Phys., 228(5):1591–1611, 2009.

[19] C.T. Crowe, J.D. Schwarzkopf, M. Sommerfeld, and Y. Tsuji. Multiphase Flows with Droplets and Particles. Taylor & Francis, 1997.

[20] C. Daux, N. Moës, J. Dolbow, N. Sukumar, and T. Belytschko. Arbitrary branched and intersecting cracks with the extended finite element method. Int. J. Numer. Methods Eng., 48(12):1741–1760, 2000.

[21] J. Donea, S. Giuliani, and J.-P. Halleux. An Arbitrary Lagrangian-Eulerian finite element method for transient dynamic fluid-structure interactions. Comput. Methods Appl. Mech. Eng., 33(1-3):689–723, 1982.

[22] J. Donea, A. Huerta, J.-P. Ponthot, and A. Rodríguez-Ferran. Arbitrary Lagrangian–Eulerian Methods. Wiley Online Library, 2004.

[23] Th. Dunne. An Eulerian approach to fluid–structure interaction and goal-oriented mesh adaptation. Int. J. Numer. Methods Fluids, 51(9-10):1017–1039, 2006.

[24] M. Feistauer and V. Sobotíková. Finite element approximation of nonlinear problems with discontinuous coefficients. ESAIM: M2AN, 24:457–500, 1990.

[25] L. Formaggia, A. Quarteroni, and A. Veneziani. Cardiovascular Mathematics: Modeling and simulation of the circulatory system, volume 1. Springer Science & Business Media, 2010.

[26] T. Frachon and S. Zahedi. A cut finite element method for incompressible two-phase Navier–Stokes flows. J. Comput. Phys., 384:77–98, 2019.

[27] S. Frei. Eulerian Finite Element Methods for Interface Problems and Fluid-Structure Interactions. PhD thesis, University of Heidelberg, Germany, 2016.

[28] S. Frei. An edge-based pressure stabilization technique for finite elements on arbitrarily anisotropic meshes. Int. J. Numer. Methods Fluids, 89(10):407–429, 2019.

[29] S. Frei, G. Judakova, and T. Richter. A locally modified second-order finite element method for interface problems. accepted for publication in ESAIm M2AN, 2023.

[30] S. Frei and T. Richter. A locally modified parametric finite element method for interface problems. SIAM J. Numer. Anal., 52(5):2315–2334, 2014.

[31] S. Frei and T. Richter. A second order time-stepping scheme for parabolic interface problems with moving interfaces. ESAIM: M2AN, 51(4):1539–1560, 2017.

[32] T.-P. Fries and T. Belytschko. The extended/generalized finite element method: an overview of the method and its applications. Int. J. Numer. Methods Eng., 84(3):253–304, 2010.

[33] S. Ganesan and S. Srivastava. ALE-SUPG finite element method for convection–diffusion problems in time-dependent domains: Conservative form. Appl. Math. Comput., 303:128–145, 2017.

[34] H. Garcke, K.F. Lam, R. Nürnberg, and E. Sitka. A multiphase Cahn–Hilliard–Darcy model for tumour growth with necrosis. Math. Models Methods Appl. Sci., 28(03):525–577, 2018.

[35] S. Gross and A. Reusken. Numerical Methods for Two-phase Incompressible Flows, volume 40. Springer Science & Business Media, 2011.

[36] M. Gurris, D. Kuzmin, and S. Turek. Finite element simulation of compressible particle-laden gas flows. J. Comput. Appl. Math., 233(12):3121–3129, 2010.

[37] A. Hansbo and P. Hansbo. An unfitted finite element method, based on Nitsche's method, for elliptic interface problems. Comput. Methods Appl. Mech. Eng., 191(47-48):5537–5552, 2002.

[38] P. Hansbo, M.G. Larson, and S. Zahedi. A cut finite element method for a Stokes interface problem. Appl. Numer. Math., 85:90–114, 2014.

[39] P. Hansbo, M.G. Larson, and S. Zahedi. A cut finite element method for coupled bulk-surface problems on time-dependent domains. Comput. Methods Appl. Mech. Eng., 307:96–116, 2016.

[40] F. Hecht and O. Pironneau. An energy stable monolithic Eulerian fluid-structure finite element method. Int. J. Numer. Methods Fluids, 85(7):430–446, 2017.

[41] C.W. Hirt, A.A. Amsden, and J.L. Cook. An Arbitrary Lagrangian–Eulerian computing method for all flow speeds. J. Comput. Phys., 135(2):203–216, 1997.

[42] T.J. Hughes, W.K. Liu, and T.K. Zimmermann. Lagrangian-Eulerian finite element formulation for incompressible viscous flows. Comput. Methods Appl. Mech. Eng., 29(3):329–349, 1981.

[43] S.-R. Hysing, S. Turek, D. Kuzmin, N. Parolini, E. Burman, S. Ganesan, and L. Tobiska. Quantitative benchmark computations of two-dimensional bubble dynamics. Int. J. Numer. Methods Fluids, 60(11):1259–1288, 2009.

[44] A. Johansson, M.G. Larson, and A. Logg. High order cut finite element methods for the Stokes problem. Adv. Model. Simul. Eng. Sci., 2(1):1–23, 2015.

[45] G. Judakova and M. Bause. Numerical investigation of multiphase flow in pipelines. Int. J. Mech. Mechatron, 11(9):1540 – 1546, 2017.

[46] S. Knauf, S. Frei, T. Richter, and R. Rannacher. Towards a complete numerical description of lubricant film dynamics in ball bearings. Comput. Mech., 53(2):239–255, 2014.

[47] C. Lehrenfeld. The Nitsche XFEM-DG space-time method and its implementation in three space dimensions. SIAM J. Sci. Comput., 37(1):A245–A270, 2015.

[48] C. Lehrenfeld and M.A. Olshanskii. An finite element method for PDEs in time-dependent domains. ESAIM: M2AN, 53(2):585–614, 2019.

[49] Y Lou and C Lehrenfeld. Isoparametric unfitted BDF–finite element method for PDEs on evolving domains. SIAM J Numer Anal, 60(4):2069–2098, 2022.

[50] A. Massing, M. G. Larson, A. Logg, and M. E. Rognes. A stabilized Nitsche fictitious domain method for the Stokes problem. J Sci Comput, 61(3):604–628, 2014.

[51] A. Massing, B. Schott, and W.A. Wall. A stabilized Nitsche cut finite element method for the Oseen problem. Comput. Methods Appl. Mech. Eng., 328:262–300, 2018.

[52] J.A. Nitsche. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. Abh Math Univ Hamburg, 36:9–15, 1970.

[53] C.S. Peskin. Flow patterns around heart valves: a numerical method. J. Comput. Phys., 10(2):252–271, 1972.

[54] F. Porté-Agel, M. Bastankhah, and S. Shamsoddin. Wind-turbine and wind-farm flows: a review. Boundary-Layer Meteorol., 174(1):1–59, 2020.

[55] T. Richter. Fluid-Structure Interactions: Models, Analysis and Finite Elements, volume 118. Springer, 2017.

[56] B. Schott. Stabilized Cut Finite Element Methods for Complex Interface Coupled flow Problems. PhD thesis, Technische Universität München, 2017.

[57] S. Srivastava and S. Ganesan. Local projection stabilization with discontinuous Galerkin method in time applied to convection dominated problems in time-dependent domains. BIT Numer. Math., 60(2):481–507, 2020.

[58] K. Stein, R. Benney, V. Kalro, T.E. Tezduyar, J. Leonard, and M. Accorsi. Parachute fluid–structure interactions: 3-d computation. Comput. Methods Appl. Mech. Eng., 190(3-4):373–386, 2000.

[59] V. Thomée. Galerkin Finite Element Methods for Parabolic Problems, volume 25. Springer Science & Business Media, 2007.

[60] F.N. Van de Vosse, J. De Hart, C.H.G.A. Van Oijen, D. Bessems, T.W.M. Gunther, A. Segal, B.J.B.M. Wolters, J.M.A. Stijnen, and F.P.T. Baaijens. Finite-element-based computational methods for cardiovascular fluid-structure interaction. J. Eng. Math, 47(3):335–368, 2003.

[61] H. von Wahl, T. Richter, and C. Lehrenfeld. An unfitted finite element method for the time-dependent Stokes problem on moving domains. IMA J. Numer. Anal., pages 1–40, 2021.

[62] W.A. Wall and T. Rabczuk. Fluid–structure interaction in lower airways of CT-based lung geometries. Int. J. Numer. Methods Fluids, 57(5):653–675, 2008.

[63] T. Williams, C. Kelley, C. Bersch, H.-B. Bröker, J. Campbell, R. Cunningham, D. Denholm, G. Elber, R. Fearick, C. Grammes, et al. gnuplot 5.2. An interactive plotting program. Available online: http://www. gnuplot. info/docs_5, 2, 2017.

[64] S. Zahedi. A space-time cut finite element method with quadrature in time. In Geometrically Unfitted Finite Element Methods and Applications, pages 281–306. Springer, 2017.

SIAM J. NUMER. ANAL.
Vol. 58, No. 4, pp. 2334–2350

WEAK IMPOSITION OF SIGNORINI BOUNDARY CONDITIONS ON THE BOUNDARY ELEMENT METHOD*

ERIK BURMAN†, STEFAN FREI‡, AND MATTHEW W. SCROGGS§

Abstract. We derive and analyze a boundary element formulation for boundary conditions involving inequalities. In particular, we focus on Signorini contact conditions. The Calderón projector is used for the system matrix, and boundary conditions are weakly imposed using a particular variational boundary operator designed using techniques from augmented Lagrangian methods. We present a complete numerical a priori error analysis and present some numerical examples to illustrate the theory.

Key words. boundary element methods, Nitsche's method, Signorini problem, Calderón projector

AMS subject classifications. 65N38, 65R20, 74M15

DOI. 10.1137/19M1281721

1. Introduction.

The application of Nitsche techniques to deal with variational inequalities has received increasing interest recently, starting from a series of works by Chouly, Hild, and Renard for elasticity problems with contact [7]. Their approach goes back to an augmented Lagrangian formulation, that was first introduced by Alart and Curnier [1].

In a previous paper [2], we have shown how Nitsche techniques can be used to impose Dirichlet, Neumann, mixed Dirichlet–Neumann, or Robin conditions weakly within boundary element methods. By using the Calderón projector, we were able to derive a unified framework that can be used for different boundary conditions.

The purpose of this article is to extend these techniques to boundary conditions involving inequalities, such as Signorini contact conditions. In particular, we consider the Laplace equation with mixed Dirichlet and Signorini boundary conditions: Find u such that

$$-\Delta u = 0 \qquad \text{in } \Omega, \tag{1.1a}$$

$$u = g_{\mathrm{D}} \qquad \text{on } \Gamma_{\mathrm{D}}, \tag{1.1b}$$

$$u \leqslant g_{\mathrm{C}} \quad \text{and} \quad \frac{\partial u}{\partial \boldsymbol{\nu}} \leqslant \psi_{\mathrm{C}} \qquad \text{on } \Gamma_{\mathrm{C}}, \tag{1.1c}$$

$$\left(\frac{\partial u}{\partial \boldsymbol{\nu}} - \psi_{\mathrm{C}}\right)\left(u - g_{\mathrm{C}}\right) = 0 \qquad \text{on } \Gamma_{\mathrm{C}}. \tag{1.1d}$$

Here $\Omega \subset \mathbb{R}^3$ denotes a polyhedral domain with outward pointing normal $\boldsymbol{\nu}$ and boundary $\Gamma := \Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{C}}$. We assume for simplicity that the boundary between Γ_{D}

*Received by the editors August 15, 2019; accepted for publication (in revised form) May 11, 2020; published electronically August 13, 2020.
https://doi.org/10.1137/19M1281721
Funding: The work of the first author was supported by the EPSRC grant EP/P01576X/1. The work of the second author was supported by the DFG Research Scholarship 3935/1-1.
†Department of Mathematics, University College London, London, WC1E 6BT, UK (e.burman@ucl.ac.uk).
‡Department of Mathematics and Statistics, University of Konstanz, Konstanz D-78457, Germany (stefan.frei@uni-konstanz.de).
§Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK (mws48@cam.ac.uk, http://www.mscroggs.co.uk).

and Γ_C coincides with edges between the faces of Γ. Whenever it is ambiguous, we will write $\boldsymbol{\nu}_{\boldsymbol{x}}$ for the outward pointing normal at the point $\boldsymbol{x}$. We assume that

$$g = \begin{cases} g_D & \text{in } \Gamma_D \\ g_C & \text{in } \Gamma_C \end{cases} \in L^2(\Gamma) \text{ and } \psi_C \in H^{1/2}(\Gamma_C).$$

Observe that when $\Gamma_C = \varnothing$, there exists a unique solution to (1.1) by the Lax–Milgram lemma. In the case that $\text{meas}(\Gamma_C) > 0$, the theory of Lions and Stampacchia [12] for variational inequalities yields existence and uniqueness of solutions. We assume that $u \in H^{3/2+\epsilon}(\Omega)$ for some $\epsilon > 0$.

Boundary element methods for Signorini problems were first studied by Han [11]. A variational formulation involving the Calderón projector was presented in [10]. An alternative formulation is based on Steklov–Poincaré operators [20, 22]. The numerical approaches to solve such formulations include a penalty formulation [15], operator splitting techniques [17, 23], or semi-smooth Newton methods [20, 22]. Besides the usual energy norm estimates, the latter reference includes an $L^2(\Gamma)$-error estimate based on a duality argument. Maischak and Stephan [13] presented a posteriori error estimates and an *hp*-adaptive algorithm for the Signorini problem. A priori error estimates for a penalty-based *hp* algorithm were shown by Chernov, Maischak, and Stephan [6]. Recently, an augmented Lagrangian approach has been presented in combination with a semi-smooth Newton method [22], and variational inequalities have been successfully used for time-dependent contact problems [9].

We will consider an approach where the full Calderón projector is used and the boundary conditions are included by properly adding scaled penalty terms to the two equations. This results in formulations similar to the ones obtained for weak imposition of boundary conditions using Nitsche's method [14]. The proposed framework is flexible and allows for the design of a range of different methods depending on the choice of weights and residuals.

An outline of this paper is as follows. In section 2, we introduce the basic boundary operators that will be needed and review some of their properties. Then, in section 3, we introduce the variational framework and review the results from [2] for the pure Dirichlet problem. In section 4, we show how the framework can be applied to Signorini boundary conditions and the mixed problem (1.1). The method is analyzed in section 5. We conclude by showing some numerical experiments in section 6.

2. Boundary operators. We define Green's function for the Laplace operator in $\mathbb{R}^3$ by

$$G(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{4\pi|\boldsymbol{x} - \boldsymbol{y}|}. \tag{2.1}$$

In this paper, we focus on the problem in $\mathbb{R}^3$. Similar analysis can be used for problems in $\mathbb{R}^2$, in which case this definition should be replaced by $G(\boldsymbol{x}, \boldsymbol{y}) = -\log|\boldsymbol{x} - \boldsymbol{y}|/2\pi$.

In the standard fashion (see, e.g., [19, Chapter 6]), we define the single layer potential operator, $\mathcal{V} : H^{-1/2}(\Gamma) \to H^1(\Omega)$, and the double layer potential, $\mathcal{K} : H^{1/2}(\Gamma) \to H^1(\Omega)$ for $v \in H^{1/2}(\Gamma)$, $\mu \in H^{-1/2}(\Gamma)$, and $\boldsymbol{x} \in \Omega \setminus \Gamma$ by

$$(\mathcal{V}\mu)(\boldsymbol{x}) := \int_\Gamma G(\boldsymbol{x}, \boldsymbol{y})\mu(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}, \tag{2.2}$$

$$(\mathcal{K}v)(\boldsymbol{x}) := \int_\Gamma \frac{\partial G(\boldsymbol{x}, \boldsymbol{y})}{\partial \boldsymbol{\nu}_{\boldsymbol{y}}} v(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{2.3}$$

We define the space $H^1(\Delta,\Omega) := \{v \in H^1(\Omega) : \Delta v \in L^2(\Omega)\}$, and the Dirichlet and Neumann traces, $\gamma_{\mathrm{D}} : H^1(\Omega) \to H^{1/2}(\Gamma)$ and $\gamma_{\mathrm{N}} : H^1(\Delta,\Omega) \to H^{-1/2}(\Gamma)$, by

$$\gamma_{\mathrm{D}} f(\boldsymbol{x}) := \lim_{\Omega \ni \boldsymbol{y} \to \boldsymbol{x} \in \Gamma} f(\boldsymbol{y}), \tag{2.4}$$

$$\gamma_{\mathrm{N}} f(\boldsymbol{x}) := \lim_{\Omega \ni \boldsymbol{y} \to \boldsymbol{x} \in \Gamma} \boldsymbol{\nu}_{\boldsymbol{x}} \cdot \nabla f(\boldsymbol{y}). \tag{2.5}$$

We recall that if the Dirichlet and Neumann traces of a harmonic function are known, then the potentials (2.2) and (2.3) may be used to reconstruct the function in Ω using the following relation:

$$u = -\mathcal{K}(\gamma_{\mathrm{D}} u) + \mathcal{V}(\gamma_{\mathrm{N}} u). \tag{2.6}$$

It is also known [19, Lemma 6.6] that for all $\mu \in H^{-1/2}(\Gamma)$, the function

$$u_\mu^{\mathcal{V}} := \mathcal{V}\mu \tag{2.7}$$

satisfies $-\Delta u_\mu^{\mathcal{V}} = 0$ and

$$\|u_\mu^{\mathcal{V}}\|_{H^1(\Omega)} \leqslant c\|\mu\|_{H^{-1/2}(\Gamma)}. \tag{2.8}$$

Similarly, in [19, Lemma 6.10] the function

$$u_v^{\mathcal{K}} := \mathcal{K}v \tag{2.9}$$

satisfies $-\Delta u_v^{\mathcal{K}} = 0$ for all $v \in H^{1/2}(\Gamma)$ and

$$\|u_v^{\mathcal{K}}\|_{H^1(\Omega)} \leqslant c\|v\|_{H^{1/2}(\Gamma)}. \tag{2.10}$$

We define $\{\gamma_{\mathrm{D}} f\}_\Gamma$ and $\{\gamma_{\mathrm{N}} f\}_\Gamma$ to be the averages of the interior and exterior Dirichlet and Neumann traces of f. We define the single layer, double layer, adjoint double layer, and hypersingular boundary integral operators, $\mathsf{V} : H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)$, $\mathsf{K} : H^{1/2}(\Gamma) \to H^{1/2}(\Gamma)$, $\mathsf{K}' : H^{-1/2}(\Gamma) \to H^{-1/2}(\Gamma)$, and $\mathsf{W} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$, by

$$(\mathsf{K}v)(\boldsymbol{x}) := \{\gamma_{\mathrm{D}}\mathcal{K}v\}_\Gamma(\boldsymbol{x}), \qquad (\mathsf{V}\mu)(\boldsymbol{x}) := \{\gamma_{\mathrm{D}}\mathcal{V}\mu\}_\Gamma(\boldsymbol{x}), \tag{2.11a}$$

$$(\mathsf{W}v)(\boldsymbol{x}) := -\{\gamma_{\mathrm{N}}\mathcal{K}v\}_\Gamma(\boldsymbol{x}), \qquad (\mathsf{K}'\mu)(\boldsymbol{x}) := \{\gamma_{\mathrm{N}}\mathcal{V}\mu\}_\Gamma(\boldsymbol{x}), \tag{2.11b}$$

where $\boldsymbol{x} \in \Gamma$, $v \in H^{1/2}(\Gamma)$, and $\mu \in H^{-1/2}(\Gamma)$ [19, Chapter 6].

Next, we define the Calderón projector by

$$\mathsf{C} := \begin{pmatrix} (1-\sigma)\mathsf{Id} - \mathsf{K} & \mathsf{V} \\ \mathsf{W} & \sigma\mathsf{Id} + \mathsf{K}' \end{pmatrix}, \tag{2.12}$$

where σ is defined for $\boldsymbol{x} \in \Gamma$ by [19, equation 6.11],

$$\sigma(\boldsymbol{x}) = \lim_{\epsilon \to 0} \frac{1}{4\pi\epsilon^2} \int_{\boldsymbol{y} \in \Omega : |\boldsymbol{y}-\boldsymbol{x}| = \epsilon} \mathrm{d}\boldsymbol{y}. \tag{2.13}$$

Recall that if u is a solution of (1.1), then it satisfies

$$\mathsf{C}\begin{pmatrix} \gamma_{\mathrm{D}} u \\ \gamma_{\mathrm{N}} u \end{pmatrix} = \begin{pmatrix} \gamma_{\mathrm{D}} u \\ \gamma_{\mathrm{N}} u \end{pmatrix}. \tag{2.14}$$

Taking the product of (2.14) with two test functions, and using the fact that $\sigma = \frac{1}{2}$ almost everywhere, we arrive at the following equations:

$$\langle \gamma_{\mathrm{D}} u, \mu \rangle_\Gamma = \left\langle (\tfrac{1}{2}\mathsf{Id} - \mathsf{K}) \gamma_{\mathrm{D}} u, \mu \right\rangle_\Gamma + \langle \mathsf{V} \gamma_{\mathrm{N}} u, \mu \rangle_\Gamma \qquad \forall \mu \in H^{-1/2}(\Gamma), \tag{2.15}$$

$$\langle \gamma_{\mathrm{N}} u, v \rangle_\Gamma = \left\langle (\tfrac{1}{2}\mathsf{Id} + \mathsf{K}') \gamma_{\mathrm{N}} u, v \right\rangle_\Gamma + \langle \mathsf{W} \gamma_{\mathrm{D}} u, v \rangle_\Gamma \qquad \forall v \in H^{1/2}(\Gamma). \tag{2.16}$$

For a more compact notation, we introduce $\lambda = \gamma_{\mathrm{N}} u$, $u = \gamma_{\mathrm{D}} u$, and the Calderón form

$$\begin{aligned} \mathcal{C}[(u,\lambda),(v,\mu)] := \left\langle (\tfrac{1}{2}\mathsf{Id} - \mathsf{K}) u, \mu \right\rangle_\Gamma + \langle \mathsf{V}\lambda, \mu \rangle_\Gamma \\ + \left\langle (\tfrac{1}{2}\mathsf{Id} + \mathsf{K}')\lambda, v \right\rangle_\Gamma + \langle \mathsf{W} u, v \rangle_\Gamma . \end{aligned} \tag{2.17}$$

We may then rewrite (2.15) and (2.16) as

$$\mathcal{C}[(u,\lambda),(v,\mu)] = \langle u, \mu \rangle_\Gamma + \langle \lambda, v \rangle_\Gamma . \tag{2.18}$$

We will also frequently use the multitrace form, defined by

$$\mathcal{A}[(u,\lambda),(v,\mu)] := - \langle \mathsf{K} u, \mu \rangle_\Gamma + \langle \mathsf{V}\lambda, \mu \rangle_\Gamma + \langle \mathsf{K}'\lambda, v \rangle_\Gamma + \langle \mathsf{W} u, v \rangle_\Gamma . \tag{2.19}$$

Using this, we may rewrite (2.18) as

$$\mathcal{A}[(u,\lambda),(v,\mu)] = \tfrac{1}{2} \langle u, \mu \rangle_\Gamma + \tfrac{1}{2} \langle \lambda, v \rangle_\Gamma . \tag{2.20}$$

To quantify the two traces, we introduce the product space

$$\mathbb{V} := H^{1/2}(\Gamma) \times H^{-1/2}(\Gamma)$$

and the associated norm

$$\|(v,\mu)\|_{\mathbb{V}} := \|v\|_{H^{1/2}(\Gamma)} + \|\mu\|_{H^{-1/2}(\Gamma)}.$$

The continuity and coercivity of $\mathcal{A}$ are immediate consequences of the properties of the operators V, K, K', and W:

LEMMA 2.1 (continuity & coercivity). *There exists $C > 0$ such that*

$$|\mathcal{A}[(w,\eta),(v,\mu)]| \leqslant C \|(w,\eta)\|_{\mathbb{V}} \|(v,\mu)\|_{\mathbb{V}} \qquad \forall (w,\eta),(v,\mu) \in \mathbb{V}.$$

There exists $\alpha > 0$ such that

$$\alpha \left(|v|^2_{H^{1/2}_*(\Gamma)} + \|\mu\|^2_{H^{-1/2}(\Gamma)} \right) \leqslant \mathcal{A}[(v,\mu),(v,\mu)] \qquad \forall (v,\mu) \in \mathbb{V}.$$

Proof. See [2] for the proof. ∎

3. Discretization and weak imposition of Dirichlet boundary conditions. In this section, we introduce the discrete spaces and review briefly how (nonhomogeneous) Dirichlet boundary conditions can be imposed weakly within the variational formulations introduced above. For a detailed derivation, and for different boundary conditions, we refer to [2].

To reduce the number of constants that appear, we introduce the following notation:

- If $\exists\, C > 0$ such that $a \leqslant Cb$, then we write $a \lesssim b$.

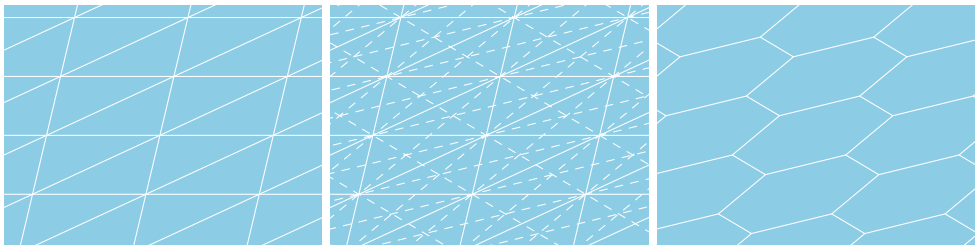


FIG. 1. *A grid (left), the barycentric refinement of the grid (center), and the dual grid (right). In a typical example, the initial grid will not be flat, and so the elements of the dual grid will not necessarily be flat.*

- If $a \lesssim b$ and $b \lesssim a$, then we write $a \eqsim b$.

We assume that Ω is a polygonal domain with faces denoted by $\{\Gamma_i\}_{i=1}^M$. We introduce a family of conforming, shape regular triangulations of Γ, $\{\mathcal{T}_h\}_{h>0}$, indexed by the largest element diameter of the mesh, h. We let $T_1, \dots, T_m \in \mathcal{T}_h$ be the triangles of a triangulation.

We consider the following finite element spaces:

$$\begin{aligned}
\mathrm{P}_h^k(\Gamma) &:= \{v_h \in C^0(\Gamma) : v_h|_{T_i} \in \mathbb{P}_k(T_i) \text{ for every } T_i \in \mathcal{T}_h\},\\
\mathrm{DP}_h^l(\Gamma) &:= \{v_h \in L^2(\Gamma) : v_h|_{T_i} \in \mathbb{P}_l(T_i) \text{ for every } T_i \in \mathcal{T}_h\},\\
\widetilde{\mathrm{DP}}_h^l(\Gamma) &:= \{v_h \in \mathrm{DP}_h^l(\Gamma) : v_h|_{\Gamma_i} \in C^0(\Gamma_i) \text{ for } i = 1, \dots, M\},
\end{aligned}$$

where $\mathbb{P}_k(T_i)$ denotes the space of polynomials of order less than or equal to k on the triangle T_i.

In addition, we consider the space $\mathrm{DUAL}_h^0(\Gamma)$ of piecewise constant functions on the barycentric dual grid, as shown in Figure 1. On nonsmooth domains, these spaces have lower order approximation properties than the standard space $\mathrm{DP}_h^0(\Gamma)$, as given in the following lemma.

LEMMA 3.1. *Let $\mu \in H^s(\Gamma)$. If Γ consists of a finite number of smooth faces meeting at edges, then*

$$\inf_{\eta_h \in \mathrm{DUAL}_h^0(\Gamma)} \|\mu - \eta_h\|_{H^{-1/2}(\Gamma)} \lesssim h^{\xi+1/2} \|\mu\|_{H^\xi(\Gamma)},$$

where $\xi = \min(\frac{1}{2}, s)$. If Γ is smooth, then the same result holds with $\xi = \min(1, s)$.

Proof. See [16, Appendix 2] for the proof. ∎

We observe that $\mathrm{P}_h^k(\Gamma) \subset H^{1/2}(\Gamma)$, $\mathrm{DP}_h^l(\Gamma) \subset L^2(\Gamma)$, $\widetilde{\mathrm{DP}}_h^l(\Gamma) \subset L^2(\Gamma)$, and $\mathrm{DUAL}_h^0(\Gamma) \subset L^2(\Gamma)$. We define the discrete product space

$$\mathbb{V}_h := \mathrm{P}_h^k(\Gamma) \times \Lambda_h^l,$$

where Λ_h^l can be any of the spaces $\mathrm{DP}_h^l(\Gamma), \widetilde{\mathrm{DP}}_h^l(\Gamma)$, or $\mathrm{DUAL}_h^0(\Gamma)$.

3.1. Dirichlet boundary conditions. Let us, for the moment, assume that $\Gamma \equiv \Gamma_{\mathrm{D}}$. Then, the basic idea is to add the following suitably weighted boundary residual to the weak formulation:

$$R_{\Gamma_{\mathrm{D}}}(u_h, \lambda_h) := \beta_{\mathrm{D}}^{1/2}(g_{\mathrm{D}} - u_h). \tag{3.1}$$

This is defined such that $R_{\Gamma_D}(u_h, \lambda_h) = 0$ is equivalent to the boundary condition (1.1b). We obtain an expression of the form

$$\mathcal{C}[(u_h, \lambda_h), (v_h, \mu_h)] = \langle u_h, \mu_h \rangle_\Gamma + \langle \lambda_h, v_h \rangle_\Gamma + \langle R_{\Gamma_D}(u_h, \lambda_h), \beta_1 v_h + \beta_2 \mu_h \rangle_\Gamma , \tag{3.2}$$

or equivalently,

$$\mathcal{A}[(u_h, \lambda_h), (v_h, \mu_h)] = \tfrac{1}{2} \langle u_h, \mu_h \rangle_\Gamma + \tfrac{1}{2} \langle \lambda_h, v_h \rangle_\Gamma + \langle R_{\Gamma_D}(u_h, \lambda_h), \beta_1 v_h + \beta_2 \mu_h \rangle_\Gamma , \tag{3.3}$$

where β_1 and β_2 are problem dependent scaling operators that can be chosen as a function of the physical parameters in order to obtain robustness of the method.

For the Dirichlet problem, we choose $\beta_1 = \beta_D^{1/2}$, $\beta_2 = \beta_D^{-1/2}$, where different choices for β_D in the range $0 \leqslant \beta_D \lesssim h^{-1}$ are possible. Inserting this into (3.3), we obtain the formulation

$$\begin{aligned} \mathcal{A}[(u, \lambda), (v_h, \mu_h)] - \tfrac{1}{2} \langle \lambda_h, v_h \rangle_{\Gamma_D} + \tfrac{1}{2} \langle u_h, \mu_h \rangle_{\Gamma_D} + \langle \beta_D u_h, v_h \rangle_{\Gamma_D} \\ = \langle g_D, \beta_D v_h + \mu_h \rangle_{\Gamma_D} . \end{aligned} \tag{3.4}$$

By formally identifying λ_h with $\partial_\nu u_h$ and μ_h with $\partial_\nu v_h$, we obtain the classical (non-symmetric) Nitsche's method (up to the multiplicative factor $\frac{1}{2}$).

For a more compact notation, we introduce the boundary operator associated with the nonhomogeneous Dirichlet condition

$$\mathcal{B}_D[(u_h, \lambda_h), (v_h, \mu_h)] := -\tfrac{1}{2} \langle \lambda_h, v_h \rangle_{\Gamma_D} + \tfrac{1}{2} \langle u_h, \mu_h \rangle_{\Gamma_D} + \langle \beta_D u_h, v_h \rangle_{\Gamma_D} , \tag{3.5}$$

the operator corresponding to the left-hand side

$$\mathcal{A}_D[(u_h, \lambda_h), (v_h, \mu_h)] := \mathcal{A}[(u_h, \lambda_h), (v_h, \mu_h)] + \mathcal{B}_D[(u_h, \lambda_h), (v_h, \mu_h)] \tag{3.6}$$

and the operator associated with the right-hand side

$$\mathcal{L}_D(v_h, \mu_h) := \langle g_D, \beta_D v_h + \mu_h \rangle_{\Gamma_D} . \tag{3.7}$$

Using these and (3.4), we arrive at the following boundary element formulation: Find $(u_h, \lambda_h) \in \mathbb{V}_h$ such that

$$\mathcal{A}_D[(u_h, \lambda_h), (v_h, \mu_h)] = \mathcal{L}_D(v_h, \mu_h) \qquad \forall (v_h, \mu_h) \in \mathbb{V}_h. \tag{3.8}$$

We introduce the following $\mathcal{B}_D$-norm:

$$\|(v, \mu)\|_{\mathcal{B}_D} := \|(v, \mu)\|_{\mathbb{V}} + \beta_D^{1/2} \|v\|_{\Gamma_D},$$

and summarize the properties of the bilinear form $\mathcal{A}_D$ in the following lemma.

LEMMA 3.2 (properties of the bilinear form). *Let $\mathbb{W}$ be a product Hilbert space for the primal and flux variables, such that $\mathbb{V} \subset \mathbb{W}$. The bilinear form has the following properties:*

PROPERTY 1 (coercivity). *If $\beta_D = 0$ or if there exists $\beta_{\min} > 0$ (independent of h) such that $\beta_D > \beta_{\min}$, then there exists $\alpha > 0$ such that $\forall (v, \mu) \in \mathbb{W}$,*

$$\alpha \|(v, \mu)\|_{\mathcal{B}_D} \leqslant \mathcal{A}_D[(v, \mu), (v, \mu)].$$

PROPERTY 2 (continuity). *There exists $M > 0$ such that $\forall (w, \eta), (v, \mu) \in \mathbb{W}$,*

$$|\mathcal{A}_D[(v, \mu), (w, \eta)]| \leqslant M \|(v, \mu)\|_{\mathcal{B}_D} \|(w, \eta)\|_{\mathcal{B}_D}.$$

Proof. See [2, section 4.1] for the proof. ∎

4. Weak imposition of Signorini boundary conditions. Recently, Chouly, Hild, and Renard [7, 8] showed how contact problems can be treated in the context of Nitsche's method. We will show here how we may use arguments similar to theirs in the present framework to integrate unilateral contact seamlessly. The result is a nonlinear system to which one may apply Newton's method or a fixed-point iteration in a straightforward manner. We prove existence and uniqueness of solutions to the nonlinear system and optimal order error estimates.

For the derivation of the formulation on the contact boundary we will first omit the Dirichlet part, letting $\Gamma = \Gamma_{\mathrm{C}}$. To impose the contact conditions, we recall the following relations, introduced by Alart and Curnier [1], with $[x]_{\pm} := \pm\max(0, \pm x)$:

$$(u - g_{\mathrm{C}}) = \left[(u - g_{\mathrm{C}}) - \tau^{-1}(\lambda - \psi_{\mathrm{C}})\right]_{-} \qquad \text{on } \Gamma_{\mathrm{C}}, \tag{4.1}$$

$$(\lambda - \psi_{\mathrm{C}}) = -\left[\tau(u - g_{\mathrm{C}}) - (\lambda - \psi_{\mathrm{C}})\right]_{+} \qquad \text{on } \Gamma_{\mathrm{C}} \tag{4.2}$$

for all $\tau > 0$. It is straightforward [7] to show that each of these two conditions is equivalent to the contact boundary conditions (1.1c) and (1.1d).

To simplify the notation, we introduce the operators

$$P^{\tau}(u_h, \lambda_h) := \tau(u_h - g_{\mathrm{C}}) - (\lambda_h - \psi_{\mathrm{C}}) \qquad \text{and} \qquad P_0^{\tau}(u_h, \lambda_h) := \tau u_h - \lambda_h.$$

Using (4.1), we arrive at the following boundary term for the contact conditions:

$$R^1_{\Gamma_{\mathrm{C}}}(u_h, \lambda_h) = (g_{\mathrm{C}} - u_h) + \tau^{-1}\left[P^{\tau}(u_h, \lambda_h)\right]_{-}. \tag{4.3}$$

Alternatively, by using (4.2), we arrive at the following boundary term:

$$R^2_{\Gamma_{\mathrm{C}}}(u_h, \lambda_h) = \tau^{-1}\left((\psi_{\mathrm{C}} - \lambda_h) - \left[P^{\tau}(u_h, \lambda_h)\right]_{+}\right). \tag{4.4}$$

By using the fact that $x = [x]_{+} + [x]_{-}$, it can be shown that (4.3) and (4.4) are equal.

Substituting (4.3) into (3.3) and using the weights $\beta_1 = \tau$ and $\beta_2 = 1$, we obtain

$$\begin{aligned}\mathcal{A}[(u_h, \lambda_h), (v_h, \mu_h)] &+ \tfrac{1}{2}\left\langle \mu_h, u_h\right\rangle_{\Gamma_{\mathrm{C}}} + \left\langle \tau u_h - \tfrac{1}{2}\lambda_h, v_h\right\rangle_{\Gamma_{\mathrm{C}}} \\ &- \left\langle \left[P^{\tau}(u_h, \lambda_h)\right]_{-}, v_h + \tau^{-1}\mu_h\right\rangle_{\Gamma_{\mathrm{C}}} = \left\langle g_{\mathrm{C}}, \tau v_h + \mu_h\right\rangle_{\Gamma_{\mathrm{C}}}.\end{aligned} \tag{4.5}$$

Using (4.4), we have

$$\begin{aligned}\mathcal{A}[(u_h, \lambda_h), (v_h, \mu_h)] &+ \tfrac{1}{2}\left\langle \lambda_h, v_h\right\rangle_{\Gamma_{\mathrm{C}}} + \left\langle \tau^{-1}\lambda_h - \tfrac{1}{2}u_h, \mu_h\right\rangle_{\Gamma_{\mathrm{C}}} \\ &+ \left\langle \left[P^{\tau}(u_h, \lambda_h)\right]_{+}, v_h + \tau^{-1}\mu_h\right\rangle_{\Gamma_{\mathrm{C}}} = \left\langle \psi_{\mathrm{C}}, v_h + \tau^{-1}\mu_h\right\rangle_{\Gamma_{\mathrm{C}}}.\end{aligned} \tag{4.6}$$

We see that (4.6) is similar to the nonsymmetric version of the method proposed in [8] and (4.5) is similar to the nonsymmetric Nitsche formulation for contact discussed in [5]. As pointed out in the latter reference, the two formulations are equivalent, with the same solutions. In what follows, we focus exclusively on the variant (4.6). Defining

$$\begin{aligned}\mathcal{B}_{\mathrm{C}}[(u_h, \lambda_h), (v_h, \mu_h)] &:= \tfrac{1}{2}\left\langle \lambda_h, v_h\right\rangle_{\Gamma_{\mathrm{C}}} + \left\langle \tau^{-1}\lambda_h - \tfrac{1}{2}u_h, \mu_h\right\rangle_{\Gamma_{\mathrm{C}}} \\ &\quad + \left\langle \left[P^{\tau}(u_h, \lambda_h)\right]_{+}, v_h + \tau^{-1}\mu_h\right\rangle_{\Gamma_{\mathrm{C}}},\end{aligned} \tag{4.7}$$

$$\mathcal{L}_{\mathrm{C}}(v_h, \mu_h) := \left\langle \psi_{\mathrm{C}}, v_h + \tau^{-1}\mu_h\right\rangle_{\Gamma_{\mathrm{C}}}, \tag{4.8}$$

$$\mathcal{A}_{\mathrm{C}}[(u_h, \lambda_h), (v_h, \mu_h)] := \mathcal{A}[(u_h, \lambda_h), (v_h, \mu_h)] + \mathcal{B}_{\mathrm{C}}[(u_h, \lambda_h), (v_h, \mu_h)], \tag{4.9}$$

we arrive at the boundary element method formulation: Find $(u_h, \lambda_h) \in \mathbb{V}_h$ such that

$$\mathcal{A}_{\mathrm{C}}[(u_h, \lambda_h), (v_h, \mu_h)] = \mathcal{L}_{\mathrm{C}}(v_h, \mu_h) \qquad \forall\, (v_h, \mu_h) \in \mathbb{V}_h. \tag{4.10}$$

4.1. Mixed Dirichlet and contact boundary conditions. Combining the formulations for the Dirichlet and contact conditions, we arrive at the following boundary element method for the problem (1.1): Find $(u_h, \lambda_h) \in \mathbb{V}_h$ such that

$$\mathcal{A}_{\mathrm{D}}[(u_h,\lambda_h),(v_h,\mu_h)] + \mathcal{B}_{\mathrm{C}}[(u_h,\lambda_h),(v_h,\mu_h)] = \mathcal{L}_{\mathrm{D}}(v_h,\mu_h) + \mathcal{L}_{\mathrm{C}}(v_h,\mu_h) \quad \forall\, (v_h,\mu_h) \in \mathbb{V}_h, \tag{4.11}$$

where $\mathcal{A}_{\mathrm{D}}$, $\mathcal{L}_{\mathrm{D}}$, $\mathcal{B}_{\mathrm{C}}$, and $\mathcal{L}_{\mathrm{C}}$ are defined in (3.6), (3.7), (4.7), and (4.8). For discretization, we use the assumptions and spaces introduced in section 3. Note that the formulation (4.11) is consistent, i.e., the continuous solution (u, λ) to (1.1) fulfills (4.11) for all $(v_h, \mu_h) \in \mathbb{V}_h$.

5. Analysis. In this section, we prove the existence of unique solutions to the nonlinear system of equations (4.11) as well as optimal error estimates.

We assume that the solution (u, λ) of (1.1) lies in $\mathbb{W} := H^{1+\epsilon}(\Gamma) \times H^{\epsilon}(\tilde{\Gamma})$ for some $\epsilon \in (0, 1/2]$, where $\tilde{\Gamma} = \cup_{i=1}^{M} \Gamma_i \setminus \partial\Gamma_i$ is the set of boundary points that lie in the interior of the faces Γ_i. As the normal vectors $\boldsymbol{\nu_x}$ are discontinuous between faces, we cannot expect a higher global regularity for λ.

We define the distance function d_{C} and norm $\|\cdot\|_*$ for $(v,\mu),(w,\eta) \in \mathbb{W}$, by

$$\begin{aligned} d_{\mathrm{C}}\left((v,\mu),(w,\eta)\right) &:= \|(v-w,\mu-\eta)\|_{\mathcal{B}_{\mathrm{D}}} \\ &\quad + \|\tau^{-\frac{1}{2}}\left(\mu-\eta+[P^{\tau}(v,\mu)]_{+}-[P^{\tau}(w,\eta)]_{+}\right)\|_{\Gamma_{\mathrm{C}}}, \end{aligned} \tag{5.1}$$

$$\|(v,\mu)\|_* := \|(v,\mu)\|_{\mathcal{B}_{\mathrm{D}}} + \|\tau^{\frac{1}{2}} v\|_{\Gamma_{\mathrm{C}}} + \|\tau^{-\frac{1}{2}}\mu\|_{\Gamma_{\mathrm{C}}}. \tag{5.2}$$

We note that due to the appearance of $[\cdot]_+$ in its second term, d_{C} is not a norm. d_{C} does provide a bound on the error; however, as for all $(v,\mu) \in \mathbb{W}$, $d_{\mathrm{C}}\left((v,\mu),(0,0)\right) \geqslant \|(v,\mu)\|_{\mathcal{B}_{\mathrm{D}}} \geqslant \|(v,\mu)\|_{\mathbb{V}}$.

When proving this section's results, we will use properties of the $[\cdot]_+$ function that are given in the following lemma.

LEMMA 5.1. *For all $a, b \in \mathbb{R}$,*

$$\left([a]_+ - [b]_+\right)^2 \leqslant \left([a]_+ - [b]_+\right)(a-b), \tag{5.3}$$

$$|\,[a]_+ - [b]_+\,| \leqslant |a-b|. \tag{5.4}$$

Proof. For a proof of these well-known properties, see, e.g., [7]. ∎

We now prove a result analogous to the coercivity assumption in [2].

LEMMA 5.2. *If there is $\beta_{\min} > 0$, independent of h, such that $\beta_{\mathrm{D}} > \beta_{\min}$, then there is $\alpha > 0$ such that for all $(v,\mu),(w,\eta) \in \mathbb{W}$,*

$$\begin{aligned} \alpha\left(d_{\mathrm{C}}\left((v,\mu),(w,\eta)\right)\right)^2 \leqslant\ & (\mathcal{A}+\mathcal{B}_{\mathrm{D}})[(v-w,\mu-\eta),(v-w,\mu-\eta)] \\ & + \mathcal{B}_{\mathrm{C}}[(v,\mu),(v-w,\mu-\eta)] - \mathcal{B}_{\mathrm{C}}[(w,\eta),(v-w,\mu-\eta)]. \end{aligned}$$

Proof. From the analysis of the Dirichlet problem (Lemma 3.2) we know that when $\beta_{\mathrm{D}} > \beta_{\min} > 0$,

$$\alpha\|(v-w,\mu-\eta)\|^2_{\mathcal{B}_{\mathrm{D}}} \leqslant (\mathcal{A}+\mathcal{B}_{\mathrm{D}})[(v-w,\mu-\eta),(v-w,\mu-\eta)]. \tag{5.5}$$

Introducing the notation $\delta P := [P^{\tau}(v,\mu)]_+ - [P^{\tau}(w,\eta)]_+$, we have

$$\begin{aligned} \mathcal{B}_{\mathrm{C}}[(v,\mu),(v-w,\mu-\eta)] - \mathcal{B}_{\mathrm{C}}[(w,\eta),(v-w,\mu-\eta)] \\ = \tau^{-1}\|\mu-\eta\|^2_{\Gamma_{\mathrm{C}}} + \left\langle \delta P, v-w+\tau^{-1}(\mu-\eta)\right\rangle_{\Gamma_{\mathrm{C}}}. \end{aligned} \tag{5.6}$$

To estimate the expression on the right-hand side, we use

$$\tau^{-1}\|\mu-\eta+\delta P\|_{\Gamma_C}^2 = \tau^{-1}\left(\|\mu-\eta\|_{\Gamma_C}^2 + \|\delta P\|_{\Gamma_C}^2 + 2\left\langle \mu-\eta, \delta P\right\rangle_{\Gamma_C}\right).$$

Using (5.3), this implies the bound

$$\begin{aligned}\tau^{-1}\|\mu-\eta+\delta P\|_{\Gamma_C}^2 &\\ &\leqslant \tau^{-1}\left(\|\mu-\eta\|_{\Gamma_C}^2 + \left\langle \delta P, P_0^\tau(v-w,\mu-\eta)\right\rangle_{\Gamma_C} + 2\left\langle \mu-\eta,\delta P\right\rangle_{\Gamma_C}\right).\end{aligned}$$

Observing that $P_0^\tau(v-w,\mu-\eta)+2(\mu-\eta) = \tau(v-w)+\mu-\eta$, we infer that

$$\tau^{-1}\|\mu-\eta+\delta P\|_{\Gamma_C}^2 \leqslant \mathcal{B}_C[(v,\mu),(v-w,\mu-\eta)] - \mathcal{B}_C[(w,\eta),(v-w,\mu-\eta)]. \tag{5.7}$$

We conclude the proof by noting that

$$\begin{aligned}\left(d_C\left((v,\mu),(w,\eta)\right)\right)^2 \lesssim \|(v-w,\mu-\eta)\|_{\mathcal{B}_D}^2 &\\ + \tau^{-1}\|\mu-\eta+[P^\tau(v,\mu)]_+ - [P^\tau(w,\eta)]_+\|_{\Gamma_C}^2,&\end{aligned}$$

and applying (5.5) and (5.7). ∎

Next, we prove a result analogous to the discrete coercivity assumption in [2].

LEMMA 5.3. *If there is $\beta_{\min} > 0$, independent of h, such that $\beta_D > \beta_{\min}$, then there is $\alpha > 0$ such that for all $(v_h,\mu_h) \in \mathbb{V}_h$,*

$$\begin{aligned}\alpha\left(\|(v_h,\mu_h)\|_{\mathcal{B}_D} + \|\tau^{-\frac{1}{2}}\left(\mu_h + [P^\tau(v_h,\mu_h)]_+\right)\|_{\Gamma_C}\right)^2 &\\ \leqslant (\mathcal{A}+\mathcal{B}_D+\mathcal{B}_C)[(v_h,\mu_h),(v_h,\mu_h)] - \left\langle [P^\tau(v_h,\mu_h)]_+, g_C - \tau^{-1}\psi_C\right\rangle_{\Gamma_C}.&\end{aligned}$$

Proof. The proof is similar to that of Lemma 5.2, but with μ_h and v_h instead of $\mu-\eta$ and $v-w$. The appearance of the data term in the right-hand side is due to the relation

$$\begin{aligned}\tau^{-1}\|\,[P^\tau(v_h,\mu_h)]_+\,\|_{\Gamma_C}^2 + 2\tau^{-1}\left\langle \mu_h, [P^\tau(v_h,\mu_h)]_+\right\rangle_{\Gamma_C} + \tau^{-1}\|\mu_h\|_{\Gamma_C}^2 &\\ = \tau^{-1}\left\langle [P^\tau(v_h,\mu_h)]_+, P^\tau(v_h,\mu_h)\right\rangle_{\Gamma_C} + \tau^{-1}\|\mu_h\|_{\Gamma_C}^2 &\\ = \left\langle [P^\tau(v_h,\mu_h)]_+, u_h + \tau^{-1}\mu_h\right\rangle_{\Gamma_C} &\\ \quad - \left\langle [P^\tau(v_h,\mu_h)]_+, g_C - \tau^{-1}\psi_C\right\rangle_{\Gamma_C} + \tau^{-1}\|\mu_h\|_{\Gamma_C}^2 &\\ = \mathcal{B}_C[(v_h,\mu_h),(v_h,\mu_h)] - \left\langle [P^\tau(v_h,\mu_h)]_+, g_C - \tau^{-1}\psi_C\right\rangle_{\Gamma_C}.&\end{aligned}$$ ∎

Using Lemmas 5.2 and 5.3, we may now prove that (4.11) is well-posed.

THEOREM 5.4. *The finite dimensional nonlinear system* (4.11) *admits a unique solution.*

Proof. To prove the existence of a solution, we show the continuity and the positivity of the nonlinear operator $\mathcal{A}+\mathcal{B}_D+\mathcal{B}_C$. This allows us to apply Brouwer's fixed point theorem; see, e.g., [21, Chapter 2, Lemma 1.4].

We define $\mathsf{F}: \mathbb{V}_h \to \mathbb{V}_h$ for $(v_h,\mu_h) \in \mathbb{V}_h$, by

$$\begin{aligned}\langle \mathsf{F}(v_h,\mu_h),(w_h,\eta_h)\rangle_\Gamma = (\mathcal{A}+\mathcal{B}_D+\mathcal{B}_C)[(v_h,\mu_h),(w_h,\eta_h)] &\\ - \mathcal{L}_D(w_h,\eta_h) - \mathcal{L}_C(w_h,\eta_h)&\end{aligned}$$

for all $(w_h, \eta_h) \in \mathbb{V}_h$. We may write the nonlinear system (4.11) as

$$\langle \mathsf{F}(v_h, \mu_h), (w_h, \eta_h)\rangle_\Gamma = 0 \qquad \forall\, (w_h, \eta_h) \in \mathbb{V}_h. \tag{5.8}$$

For fixed h, by the equivalence of norms on discrete spaces, there exist $c_1, c_2 > 0$ such that for all $(v_h, \mu_h) \in \mathbb{V}_h$,

$$c_1 \|(v_h, \mu_h)\|_\Gamma \leqslant \|(v_h, \mu_h)\|_{\mathcal{B}_\mathrm{D}} \leqslant c_2 \|(v_h, \mu_h)\|_\Gamma.$$

To show positivity, we let $(v_h, \mu_h) \in \mathbb{V}_h$. Using Lemma 5.3, we see that

$$\begin{aligned}\langle \mathsf{F}(v_h, \mu_h), (v_h, \mu_h)\rangle_\Gamma \geqslant{}& \alpha \|(v_h, \mu_h)\|^2_{\mathcal{B}_\mathrm{D}} + \alpha\tau^{-1} \|\mu_h + [P^\tau(v_h, \mu_h)]_+ \|^2_{\Gamma_\mathrm{C}} \\ &+ \left\langle [P^\tau(v_h, \mu_h)]_+ , g_\mathrm{C} - \tau^{-1}\psi_\mathrm{C}\right\rangle_{\Gamma_\mathrm{C}} - \mathcal{L}_\mathrm{D}(v_h, \mu_h) - \mathcal{L}_\mathrm{C}(v_h, \mu_h).\end{aligned}$$

Using the Cauchy–Schwarz inequality and an arithmetic-geometric inequality, we see that there exists $C_{g_\mathrm{C}\psi_\mathrm{C}} > 0$ such that

$$\begin{aligned}&\left\langle [P^\tau(v_h, \mu_h)]_+ , g_\mathrm{C} - \tau^{-1}\psi_\mathrm{C}\right\rangle_{\Gamma_\mathrm{C}} - \mathcal{L}_\mathrm{D}(v_h, \mu_h) - \mathcal{L}_\mathrm{C}(v_h, \mu_h) \\ &\quad= \left\langle [P^\tau(v_h, \mu_h)]_+ + \mu_h, g_\mathrm{C} - \tau^{-1}\psi_\mathrm{C}\right\rangle_{\Gamma_\mathrm{C}} - \left\langle \mu_h, g_\mathrm{C} - \tau^{-1}\psi_\mathrm{C}\right\rangle_{\Gamma_\mathrm{C}} \\ &\qquad - \langle g_\mathrm{D}, \beta_\mathrm{D} v_h + \mu_h\rangle_{\Gamma_\mathrm{D}} - \left\langle \psi_\mathrm{C}, v_h + \tau^{-1}\mu_h\right\rangle_{\Gamma_\mathrm{C}} \\ &\quad\geqslant -C^2_{g_\mathrm{C}\psi_\mathrm{C}} - \tfrac{\alpha}{2}\left(\|(v_h, \mu_h)\|^2_{\mathcal{B}_\mathrm{D}} + \tau^{-1}\|\mu_h + [P^\tau(v_h, \mu_h)]_+ \|^2_{\Gamma_\mathrm{C}}\right).\end{aligned}$$

Using norm equivalence, we obtain

$$\begin{aligned}&\langle \mathsf{F}(v_h, \mu_h), (v_h, \mu_h)\rangle_\Gamma \\ &\qquad\geqslant \tfrac{\alpha}{2}\left(\|(v_h, \mu_h)\|^2_{\mathcal{B}_\mathrm{D}} + \tau^{-1}\|\mu_h + [P^\tau(v_h, \mu_h)]_+ \|^2_{\Gamma_\mathrm{C}}\right) - C^2_{g_\mathrm{C}\psi_\mathrm{C}} \\ &\qquad\geqslant C'\|(v_h, \mu_h)\|^2_\Gamma - C^2_{g_\mathrm{C}\psi_\mathrm{C}}\end{aligned}$$

for some $C' > 0$. We conclude that for all $(v_h, \mu_h) \in \mathbb{V}_h$ with

$$\|(v_h, \mu_h)\|^2_\Gamma > \frac{C^2_{g_\mathrm{C}\psi_\mathrm{C}}}{C'} + 1,$$

there holds $\langle \mathsf{F}(v_h, \mu_h), (v_h, \mu_h)\rangle_\Gamma > 0$.

To show continuity, let $(v_h^1, \mu_h^1), (v_h^2, \mu_h^2) \in \mathbb{V}_h$. We have for all $(w_h, \eta_h) \in \mathbb{V}_h$,

$$\begin{aligned}&\left\langle \mathsf{F}(v_h^1, \mu_h^1) - \mathsf{F}(v_h^2, \mu_h^2), (w_h, \eta_h)\right\rangle_\Gamma \\ &\qquad= \left\langle \left[P^\tau(v_h^1, \mu_h^1)\right]_+ - \left[P^\tau(v_h^2, \mu_h^2)\right]_+ , w_h + \tau^{-1}\eta_h\right\rangle_{\Gamma_\mathrm{C}} \\ &\qquad\quad + \tfrac{1}{2}\left\langle \mu_h^1 - \mu_h^2, w_h + \tau^{-1}\eta_h\right\rangle_\Gamma - \tfrac{1}{2}\left\langle v_h^1 - v_h^2, \mu_h^1 - \mu_h^2\right\rangle_{\Gamma_\mathrm{C}} \\ &\qquad\quad + (\mathcal{A} + \mathcal{B}_\mathrm{D})[(v_h^1 - v_h^2, \mu_h^1 - \mu_h^2), (w_h, \eta_h)] \\ &\qquad\leqslant \left(\tau\|v_h^1 - v_h^2\|_{\Gamma_\mathrm{C}} + \|\mu_h^1 - \mu_h^2\|_{\Gamma_\mathrm{C}}\right)\left(\|w_h\|_{\Gamma_\mathrm{C}} + \tau^{-1}\|\eta_h\|_{\Gamma_\mathrm{C}}\right),\end{aligned}$$

where we have used (5.4). By norm equivalence, this means that

$$\frac{\left\langle \mathsf{F}(v_h^1, \mu_h^1) - \mathsf{F}(v_h^2, \mu_h^2), (w_h, \eta_h)\right\rangle_\Gamma}{\|(w_h, \eta_h)\|_\Gamma} \leqslant C\|(v_h^1 - v_h^2, \mu_h^1 - \mu_h^2)\|_\Gamma$$

showing that F is continuous.

It then follows by Brouwer's fixed point theorem [21, Chapter 2, Lemma 1.4] that there exists a solution to (5.8) and hence also to (4.11).

Uniqueness is an immediate consequence of Lemma 5.2. Assume that (u_h^1, λ_h^1) and (u_h^2, λ_h^2) are solutions to (4.11). We immediately see that

$$\alpha \left(d_{\mathrm{C}} \left((u_h^1, \lambda_h^1), (u_h^2, \lambda_h^2) \right) \right)^2 = 0,$$

and we conclude that the solution is unique. ∎

We now proceed to prove the following best approximation result.

LEMMA 5.5. *Let $(u, \lambda) \in \mathbb{W}$ be the solution of (1.1) and $(u_h, \lambda_h) \in \mathbb{V}_h$ the solution of (4.11). Then there holds*

$$d_{\mathrm{C}} \left((u, \lambda), (u_h, \lambda_h) \right) \leqslant C \inf_{(v_h, \mu_h) \in \mathbb{V}_h} \| (u - v_h, \lambda - \mu_h) \|_*.$$

Proof. Using Lemma 5.2 and Galerkin orthogonality, we see that for arbitrary $(v_h, \mu_h) \in \mathbb{V}_h$,

$$\begin{aligned}
&\alpha \left(d_{\mathrm{C}} \left((u, \lambda), (u_h, \lambda_h) \right) \right)^2 \\
&\quad \leqslant (\mathcal{A} + \mathcal{B}_{\mathrm{D}})[(u - u_h, \lambda - \lambda_h), (u - u_h, \lambda - \lambda_h)] \\
&\qquad + \mathcal{B}_{\mathrm{C}}[(u, \lambda), (u - u_h, \lambda - \lambda_h)] - \mathcal{B}_{\mathrm{C}}[(u_h, \lambda_h), (u - u_h, \lambda - \lambda_h)] \\
&\quad = (\mathcal{A} + \mathcal{B}_{\mathrm{D}})[(u - u_h, \lambda - \lambda_h), (u - v_h, \lambda - \mu_h)] \\
&\qquad + \mathcal{B}_{\mathrm{C}}[(u, \lambda), (u - v_h, \lambda - \mu_h)] - \mathcal{B}_{\mathrm{C}}[(u_h, \lambda_h), (u - v_h, \lambda - \mu_h)].
\end{aligned}$$

Next, we use

$$\begin{aligned}
&\mathcal{B}_{\mathrm{C}}[(u, \lambda), (u - v_h, \lambda - \mu_h)] - \mathcal{B}_{\mathrm{C}}[(u_h, \lambda_h), (u - v_h, \lambda - \mu_h)] \\
&\quad = \left\langle \lambda - \lambda_h + [P^\tau(u, \lambda)]_+ - [P^\tau(u_h, \lambda_h)]_+ , (u - v_h) + \tau^{-1}(\lambda - \mu_h) \right\rangle_{\Gamma_{\mathrm{C}}} \\
&\qquad\qquad - \tfrac{1}{2} \langle u - u_h, \lambda - \mu_h \rangle_{\Gamma_{\mathrm{C}}} - \tfrac{1}{2} \langle \lambda - \lambda_h, u - v_h \rangle_{\Gamma_{\mathrm{C}}}
\end{aligned}$$

to show that

$$\begin{aligned}
&(\mathcal{A} + \mathcal{B}_{\mathrm{D}})[(u - u_h, \lambda - \lambda_h), (u - u_h, \lambda - \lambda_h)] \\
&\qquad + \mathcal{B}_{\mathrm{C}}[(u, \lambda), (u - u_h, \lambda - \lambda_h)] - \mathcal{B}_{\mathrm{C}}[(u_h, \lambda_h), (u - u_h, \lambda - \lambda_h)] \\
&\qquad\quad = \underbrace{(\mathcal{A} + \mathcal{B}_{\mathrm{D}})[(u - u_h, \lambda - \lambda_h), (u - v_h, \lambda - \mu_h)]}_{\text{(I)}} \\
&\qquad\qquad \underbrace{- \tfrac{1}{2} \langle u - u_h, \lambda - \mu_h \rangle_{\Gamma_{\mathrm{C}}} - \tfrac{1}{2} \langle \lambda - \lambda_h, u - v_h \rangle_{\Gamma_{\mathrm{C}}}}_{\text{(II)}} \\
&\qquad + \underbrace{\left\langle \lambda - \lambda_h + [P^\tau(u, \lambda)]_+ - [P^\tau(u_h, \lambda_h)]_+ , (u - v_h) + \tau^{-1}(\lambda - \mu_h) \right\rangle_{\Gamma_{\mathrm{C}}}}_{\text{(III)}}.
\end{aligned}$$

We estimate the three parts of the right-hand side separately. For the first term, we use the continuity of $\mathcal{A} + \mathcal{B}_{\mathrm{D}}$ (Lemma 3.2) to obtain

$$\text{(I)} \leqslant M \| (u - u_h, \lambda - \lambda_h) \|_{\mathcal{B}_{\mathrm{D}}} \| (u - v_h, \lambda - \mu_h) \|_{\mathcal{B}_{\mathrm{D}}}.$$

For the second line, we use $H^{1/2}(\Gamma)$–$H^{-1/2}(\Gamma)$ duality and the Cauchy–Schwarz inequality to obtain

$$\text{(II)} \leqslant \| (u - u_h, \lambda - \lambda_h) \|_{\mathcal{B}_{\mathrm{D}}} \| (u - v_h, \lambda - \mu_h) \|_{\mathcal{B}_{\mathrm{D}}}.$$

For the last term, we use the Cauchy–Schwarz inequality to get

$$\text{(III)} \leqslant \|\tau^{-1/2}\left(\lambda-\lambda_h+[P^\tau(u,\lambda)]_+-[P^\tau(u_h,\lambda_h)]_+\right)\|_{\Gamma_C} \cdot\left(\|\tau^{1/2}(u-v_h)\|_{\Gamma_C}+\|\tau^{-1/2}(\lambda-\mu_h)\|_{\Gamma_C}\right).$$

Collecting these bounds, we see that

$$d_C\left((u,\lambda),(u_h,\lambda_h)\right)^2 \lesssim d_C\left((u,\lambda),(u_h,\lambda_h)\right)\|(u-v_h,\lambda-\mu_h)\|_*.$$

Dividing through by $d_C\left((u,\lambda),(u_h,\lambda_h)\right)$ and taking the infimum yields the desired result. ∎

We now prove the main result of this section, an a priori bound on the error of the solution of (4.11).

THEOREM 5.6. *Let $(u,\lambda)\in H^s(\Gamma)\times H^r(\tilde\Gamma)$ for some $s\geqslant 1, r\geqslant 0$, and let $(u_h,\lambda_h)\in \mathrm{P}_h^k(\Gamma)\times\Lambda_h^l$ be the solutions of (1.1) and the discrete problem (4.11), respectively. If there is $\beta_{\min}>0$ such that $\beta_{\min}<\beta_D\lesssim h^{-1}$ and $\tau\eqsim h^{-1}$, then*

$$\|(u-u_h,\lambda-\lambda_h)\|_{\mathbb{V}} \leqslant d_C\left((u,\lambda),(u_h,\lambda_h)\right) \lesssim h^{\zeta-1/2}|u|_{H^\zeta(\Gamma)}+h^{\xi+1/2}|\lambda|_{H^\xi(\tilde\Gamma)},$$

where $\zeta=\min(k+1,s)$ and $\xi=\min(l+1,r)$ for $\Lambda_h^l\in\{\mathrm{DP}_h^l(\Gamma),\widetilde{\mathrm{DP}}_h^l(\Gamma)\}$ and $\zeta=\min(2,s)$ and $\xi=\min(\frac12,r)$ for $\Lambda_h^l=\mathrm{DUAL}_h^0(\Gamma)$. Additionally,

$$\|\tilde u-\tilde u_h\|_{H^1(\Omega)}\lesssim h^{\zeta-1/2}|u|_{H^\zeta(\Gamma)}+h^{\xi+1/2}|\lambda|_{H^\xi(\tilde\Gamma)},$$

where $\tilde u$ and $\tilde u_h$ are the solutions in Ω defined by (2.6).

Proof. First, we observe that for all (v,μ) and (w,η) in $\mathbb{W}$,

$$\|(v-w,\mu-\eta)\|_{\mathbb{V}}\leqslant d_C\left((v,\mu),(w,\eta)\right).$$

Using standard approximation results for $\Lambda_h^l\in\{\mathrm{DP}_h^l(\Gamma),\widetilde{\mathrm{DP}}_h^l(\Gamma)\}$ (see, e.g., [19, Chapter 10]) and Lemma 3.1 for $\Lambda_h^l=\mathrm{DUAL}_h^0(\Gamma)$, we see that

$$\begin{aligned}\inf_{(v_h,\mu_h)\in\mathbb{V}_h}\|(u-v_h,\lambda-\mu_h)\|_{\mathbb{V}} &= \inf_{v_h\in\mathrm{P}_h^k(\Gamma)}\|u-v_h\|_{H^{1/2}(\Gamma)}+\inf_{\mu_h\in\Lambda_h^l(\Gamma)}\|\lambda-\mu_h\|_{H^{-1/2}(\Gamma)}\\ &\lesssim h^{\zeta-1/2}|u|_{H^\zeta(\Gamma)}+h^{\xi+1/2}|\lambda|_{H^\xi(\tilde\Gamma)},\\ \inf_{v_h\in\mathrm{P}_h^k(\Gamma)}\|u-v_h\|_\Gamma\lesssim h^\zeta|u|_{H^\zeta(\Gamma)},\quad \inf_{\mu_h\in\Lambda_h^l}\|\lambda-\mu_h\|_\Gamma &\lesssim h^\xi|\lambda|_{H^\xi(\tilde\Gamma)}.\end{aligned}$$

Applying these to the definition of $\|\cdot\|_*$ gives

$$\inf_{(v_h,\mu_h)\in\mathbb{V}_h}\|(u-v_h,\lambda-\mu_h)\|_*\lesssim h^{\zeta-1/2}|u|_{H^\zeta(\Gamma)}+h^{\xi+1/2}|\lambda|_{H^\xi(\tilde\Gamma)} +\beta_D^{1/2}h^\zeta|u|_{H^\zeta(\Gamma)}+\tau^{1/2}h^\zeta|u|_{H^\zeta(\Gamma)}+\tau^{-1/2}h^\xi|\lambda|_{H^\xi(\tilde\Gamma)}.$$

By means of Lemma 5.5 and the given choice of the parameters τ and β_D, we prove the first assertion. The estimate in the domain Ω follows by using the relations (2.8) and (2.10). ∎

If λ is smooth enough and $k=l$, the bounds on τ can be replaced with $h\lesssim\tau\lesssim h^{-1}$ without reducing the order of convergence.

6. Numerical results. We now demonstrate the theory with a series of numerical examples. In this section, we consider the following test problem. Let $\Omega = [0,1] \times [0,1] \times [0,1]$ be the unit cube, $\Gamma_{\mathrm{C}} := \{(x,y,z) \in \Gamma : z = 1\}$, and $\Gamma_{\mathrm{D}} := \Gamma \setminus \Gamma_{\mathrm{C}}$. Let

$$g_D = 0, \tag{6.1a}$$

$$g_{\mathrm{C}} = \begin{cases} \sin(\pi x)\sin(\pi y)\sinh(\sqrt{2}\pi), & x \leqslant \frac{1}{2}, \\ \sin(\pi y)\sinh(\sqrt{2}\pi), & x > \frac{1}{2}, \end{cases} \tag{6.1b}$$

$$\psi_{\mathrm{C}} = \begin{cases} \sqrt{2}\pi\sin(\pi x)\sin(\pi y)\cosh(\sqrt{2}\pi), & x \geqslant \frac{1}{2}, \\ \sqrt{2}\pi\sin(\pi y)\cosh(\sqrt{2}\pi), & x < \frac{1}{2}. \end{cases} \tag{6.1c}$$

It can be shown that

$$u(x,y,z) = \sin(\pi x)\sin(\pi y)\sinh(\sqrt{2}\pi z)$$

is the solution to (1.1) with these boundary conditions.

To solve the nonlinear system (4.10), we will treat the nonlinear term explicitly. Therefore, we define

$$\mathcal{B}'_{\mathrm{C}}[(u,\lambda),(v,\mu)] := \tfrac{1}{2}\langle \lambda, v\rangle_{\Gamma_{\mathrm{C}}} + \left\langle \tau^{-1}\lambda - \tfrac{1}{2}u, \mu\right\rangle_{\Gamma_{\mathrm{C}}}. \tag{6.2}$$

Note that $\mathcal{B}'_{\mathrm{C}}$ differs from $\mathcal{B}_{\mathrm{C}}$ only by the missing nonlinear term.

We pick initial guesses $(u_0,\lambda_0) \in \mathbb{V}_h$ and define $(u_{n+1},\lambda_{n+1}) \in \mathbb{V}_h$, for $n \in \mathbb{N}$, to be the solution of

$$\begin{aligned}(\mathcal{A} + \mathcal{B}_{\mathrm{D}} + \mathcal{B}'_{\mathrm{C}})[(u_{n+1},\lambda_{n+1}),(v_h,\mu_h)] \\ = \mathcal{L}_{\mathrm{C}}(v_h,\mu_h) - \left\langle [P^\tau(u_n,\lambda_n)]_+, v_h + \tau^{-1}\mu_h\right\rangle_{\Gamma_{\mathrm{C}}} \quad \forall\, (v_h,\mu_h) \in \mathbb{V}_h.\end{aligned} \tag{6.3}$$

This leads us to Algorithm 6.1, an iterative method for solving the contact problem.

In all of the computations in this section, we preconditioned the GMRES solver using a mass matrix preconditioner applied blockwise from the left, as described in [3].

Algorithm 6.1. Iterative algorithm for solving the contact problem.

```
Input (u_0, λ_0), TOL, MAXITER
for n ← 0 to MAXITER do
  (u_{n+1}, λ_{n+1}) ← solution of (6.3), calculated using GMRES
  if ‖(u_{n+1}, λ_{n+1}) − (u_n, λ_n)‖_𝕍 < TOL then
    return (u_{n+1}, λ_{n+1})
  end if
end for
```

Inspired by the parameter choices in [2], we fix $\beta_{\mathrm{D}} = 0.01$ and look for suitable values of the parameter τ. Figure 2 shows how the error, number of outer iterations, and the average number of GMRES iterations inside each outer iteration change as the parameter τ is varied for both $\mathbb{V}_h = \mathrm{P}_h^1(\Gamma) \times \mathrm{DUAL}_h^0(\Gamma)$ (left, blue) and $\mathbb{V}_h = \mathrm{P}_h^1(\Gamma) \times \mathrm{DP}_h^0(\Gamma)$ (right, orange). Here, we see that the error and number of outer iterations are lowest when τ is between around 1 and 10.

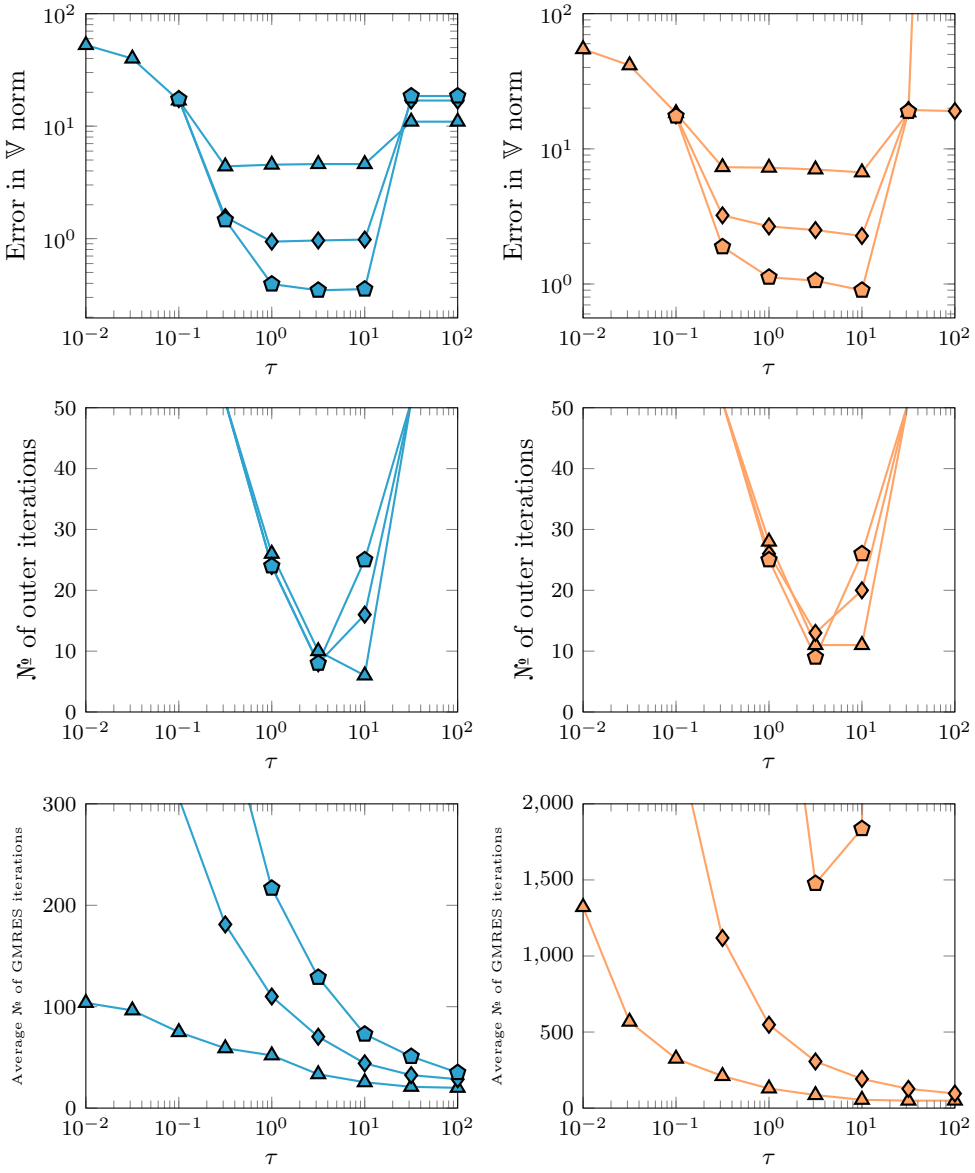


FIG. 2. *The dependence of the error, number of outer iterations, and the average number of GMRES iterations on* τ *for the problem* (1.1) *with boundary conditions* (6.1) *on the unit cube with* $h = 2^{-2}$ *(triangles),* $h = 2^{-3.5}$ *(diamonds), and* $h = 2^{-5}$ *(pentagons). Here we take* $u_0 = \lambda_0 = 0$, $\beta_{\mathrm{D}} = 0.01$, TOL $= 0.05$, *and* MAXITER $= 50$. *On the left (blue), we take* $(u_n, \lambda_n), (v_h, \mu_h) \in \mathrm{P}_h^1(\Gamma) \times \mathrm{DUAL}_h^0(\Gamma)$*; on the right (orange), we take* $(u_n, \lambda_n), (v_h, \mu_h) \in \mathrm{P}_h^1(\Gamma) \times \mathrm{DP}_h^0(\Gamma)$. *(Figure in color online.)*

Motivated by Figure 2 and the bounds in Theorem 5.6, we take $\tau = 0.5/h$, and look at the convergence as h is decreased. Figure 3 shows how the error and iteration counts vary as h is decreased when $\mathbb{V}_h = \mathrm{P}_h^1(\Gamma) \times \mathrm{DUAL}_h^0(\Gamma)$ (left, blue circles) and $\mathbb{V}_h = \mathrm{P}_h^1(\Gamma) \times \mathrm{DP}_h^0(\Gamma)$ (right, orange squares).

For $\mathbb{V}_h = \mathrm{P}_h^1(\Gamma) \times \mathrm{DUAL}_h^0(\Gamma)$, we observe slightly higher than the order 1 con-

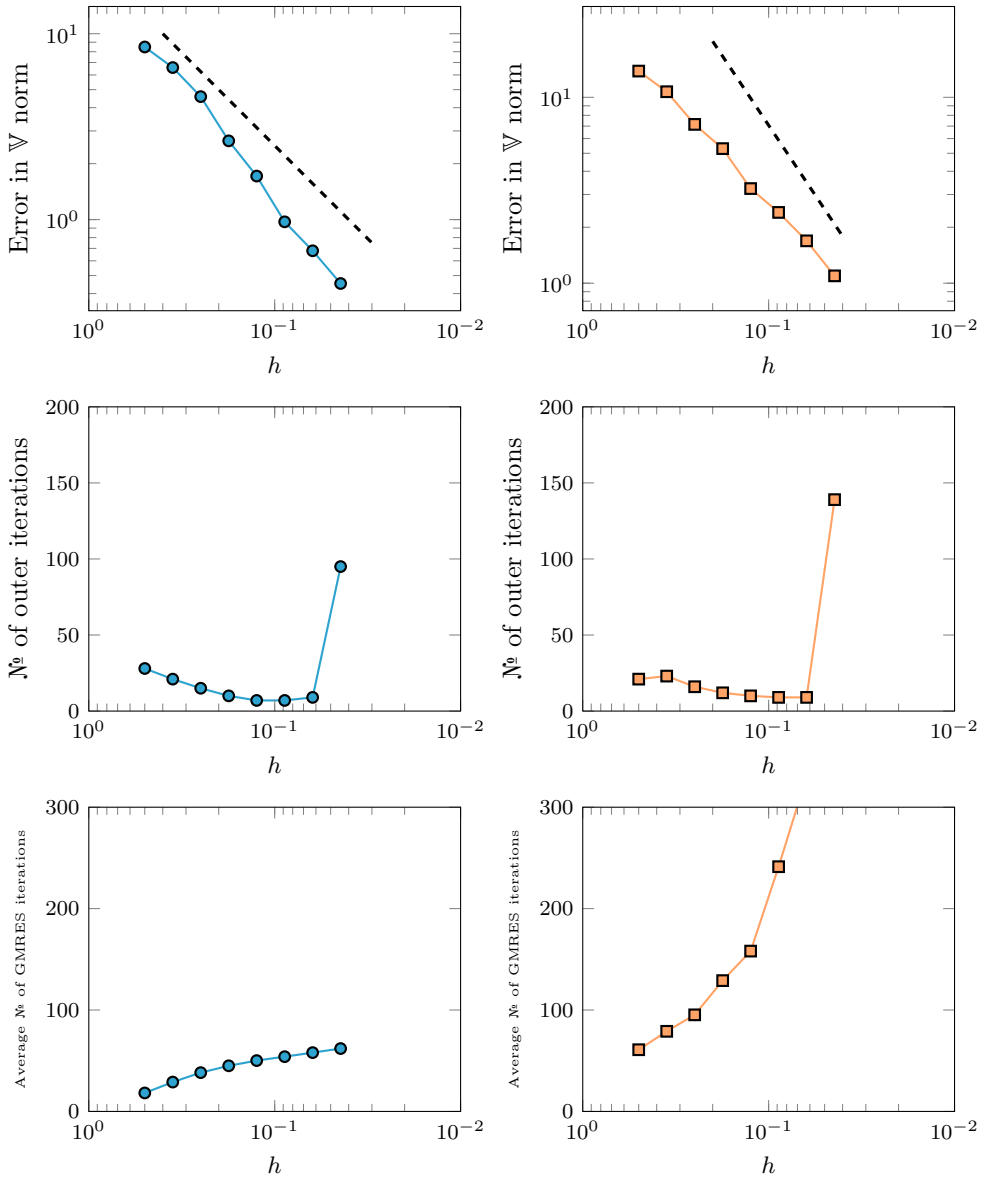


FIG. 3. *The error, number of outer iterations and average number of inner GMRES iteration for the problem* (1.1) *with boundary conditions* (6.1) *on the unit cube as h is reduced. Here we take $u_0 = \lambda_0 = 0$, $\beta_D = 0.01$,* TOL $= 0.05$, MAXITER $= 200$, *and $\tau = 0.5/h$. On the left (blue circles), we take $(u_n, \lambda_n), (v_h, \mu_h) \in \mathrm{P}_h^1(\Gamma) \times \mathrm{DUAL}_h^0(\Gamma)$; on the right (orange squares), we take $(u_n, \lambda_n), (v_h, \mu_h) \in \mathrm{P}_h^1(\Gamma) \times \mathrm{DP}_h^0(\Gamma)$. The dashed lines show order* 1 *convergence (left) and order* 1.5 *convergence (right). (Figure in color online.)*

vergence predicted by Theorem 5.6. In this case, the mass matrix preconditioner is effective, as the number of GMRES iterations required inside each outer iteration is reasonably low, and only grows slowly as h is decreased. We believe that the effectiveness of the preconditioner for this choice of spaces is due to the spaces $\mathrm{P}_h^1(\Gamma)$ and $\mathrm{DUAL}_h^0(\Gamma)$ forming an inf-sup stable pair [18, Lemma 3.1].

When $\mathbb{V}_h = \mathrm{P}_h^1(\Gamma) \times \mathrm{DP}_h^0(\Gamma)$, Theorem 5.6 tells us to expect order 1.5 convergence. However, we observe a slightly lower order. This appears to be due to the ill-conditioning of this system, and the mass matrix preconditioner being ineffective, leading to an inaccurate solution when using GMRES. In this case, the spaces $\mathrm{P}_h^1(\Gamma)$ and $\mathrm{DP}_h^0(\Gamma)$ do not form an inf-sup stable pair, and so the mass-matrix between them is not guaranteed to be invertible leading to a less effective preconditioner.

In order to obtain order 1.5 convergence with a well-conditioned system, we could look for $(u_h, \lambda_h) \in \mathrm{P}_h^1(\Gamma) \times \mathrm{DP}_h^0(\Gamma)$ and test with $(v_h, \mu_h) \in \mathrm{DUAL}_h^1(\Gamma) \times \mathrm{DUAL}_h^0(\Gamma)$, where $\mathrm{DUAL}_h^1(\Gamma)$ is the space of piecewise linear functions on the dual grid that forms an inf-sup stable pair with the space $\mathrm{DP}_h^0(\Gamma)$, as defined in [4]. With this choice of spaces, we obtain the higher order convergence as in Theorem 5.6, while having stable dual pairings and hence more effective mass matrix preconditioning.

For the problems discussed in [2], we have run numerical experiments using this space pairing and observe the full order $\frac{3}{2}$ convergence in a low number of iterations. A deeper investigation of this method using these dual spaces, and the adaption of the theory to this case, warrants future work.

7. Conclusions. Based on our work in [2], we have analyzed and demonstrated the effectiveness of Nitsche type coupling methods for boundary element formulations of contact problems.

An open problem is preconditioning. While the iteration counts in the presented examples were already practically useful, for large and complex structures preconditioning is still essential. The hope is to use the properties of the Calderón projector to build effective operator preconditioning techniques for the presented Nitsche type frameworks.

Avenues of future research include looking at how this approach could be applied to problems in linear elasticity, and an extension of this method to problems involving friction.

REFERENCES

[1] P. Alart and A. Curnier, *A mixed formulation for frictional contact problems prone to Newton like solution methods*, Comput. Methods Appl. Mech. Engrg., 92 (1991), pp. 353–375.

[2] T. Betcke, E. Burman, and M. W. Scroggs, *Boundary element methods with weakly imposed boundary conditions*, SIAM J. Sci. Comput., 41 (2019), pp. A1357–A1384, https://doi.org/10.1137/18M119625X.

[3] T. Betcke, M. W. Scroggs, and W. Śmigaj, *Product algebras for Galerkin discretizations of boundary integral operators and their applications*, ACM Trans. Math. Software, 46 (2020), 4.

[4] A. Buffa and S. H. Christiansen, *A dual finite element complex on the barycentric refinement*, Math. Comp., 76 (2007), pp. 1743–1769.

[5] E. Burman, P. Hansbo, and M. G. Larson, *The penalty-free Nitsche method and nonconforming finite elements for the Signorini problem*, SIAM J. Numer. Anal., 55 (2017), pp. 2523–2539, https://doi.org/10.1137/16M107846X.

[6] A. Chernov, M. Maischak, and E. Stephan, *A priori error estimates for hp penalty BEM for contact problems in elasticity*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 3871–3880.

[7] F. Chouly and P. Hild, *A Nitsche-based method for unilateral contact problems: Numerical analysis*, SIAM J. Numer. Anal., 51 (2013), pp. 1295–1307, https://doi.org/10.1137/12088344X.

[8] F. Chouly, P. Hild, and Y. Renard, *Symmetric and non-symmetric variants of Nitsche's method for contact problems in elasticity: theory and numerical experiments*, Math. Comp., 84 (2015), pp. 1089–1112.

[9] H. Gimperlein, F. Meyer, C. Özdemird, and E. P. Stephan, *Time domain boundary ele-*

ments for dynamic contact problems, Comput. Methods Appl. Mech. Engrg., 333 (2018), pp. 147–175.

[10] H. Han, *A direct boundary element method for Signorini problems*, Math. Comp., 55 (1990), pp. 115–128.

[11] H.-D. Han, *The boundary finite element methods for Signorini problems*, in Numerical Methods for Partial Differential Equations, Y.-I. Zhu and B.-Y. Guo, eds., Lecture Notes in Math. 1297, Springer, Berlin, Heidelberg, 1987, pp. 38–49.

[12] J. L. Lions and G. Stampacchia, *Variational inequalities*, Comm. Pure Appl. Math., 20 (1967), pp. 493–519.

[13] M. Maischak and E. P. Stephan, *Adaptive hp-versions of BEM for Signorini problems*, Appl. Numer. Math., 54 (2005), pp. 425–449.

[14] J. Nitsche, *Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind*, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg, 36 (1971), pp. 9–15.

[15] H. Schmit and G. Schneider, *Boundary element solution of the Dirichlet-Signorini problem by a penalty method*, Appl. Anal., 51 (1993), pp. 175–186.

[16] M. W. Scroggs, *Efficient Computation and Applications of the Calderón Projector*, Ph.D. thesis, University College London, London, UK, 2019.

[17] W. Spann, *On the boundary element method for the Signorini problem of the Laplacian*, Numer. Math., 65 (1993), pp. 337–356.

[18] O. Steinbach, *On a generalized L_2 projection and some related stability estimates in Sobolev spaces*, Numer. Math., 90 (2002), pp. 775–786.

[19] O. Steinbach, *Numerical Approximation Methods for Elliptic Boundary Value Problems*, Springer-Verlag, New York, 2008.

[20] O. Steinbach, *Boundary element methods for variational inequalities*, Numer. Math., 126 (2014), pp. 173–197.

[21] R. Temam, *Navier-Stokes Equations: Theory and Numerical Analysis*, Studies in Mathematics and its Applications, Vol. 2, North-Holland, Amsterdam, New York, Oxford, 1977.

[22] S. Zhang and X. Li, *An augmented Lagrangian method for the Signorini boundary value problem with BEM*, Bound. Value Probl., 2016 (2016), 62.

[23] S. Zhang and J. Zhu, *A projection iterative algorithm boundary element method for the Signorini problem*, Eng. Anal. Bound. Elem., 37 (2013), pp. 176–181.

MULTISCALE MODEL. SIMUL.
Vol. 18, No. 2, pp. 942–969

EFFICIENT APPROXIMATION OF FLOW PROBLEMS WITH MULTIPLE SCALES IN TIME*

S. FREI† AND T. RICHTER‡

Abstract. In this article we address flow problems that carry a multiscale character in time. In particular we consider the Navier–Stokes flow in a channel on a fast scale that influences the movement of the boundary which undergoes a deformation on a slow scale in time. We derive an averaging scheme that is of first order with respect to the ratio of time scales ϵ. In order to cope with the problem of unknown initial data for the fast-scale problem, we assume near-periodicity in time. Moreover, we construct a second-order accurate time discretization scheme and derive a complete error analysis for a corresponding simplified ODE system. The resulting multiscale scheme does not ask for the continuous simulation of the fast-scale variable and shows powerful speedups up to 1:10 000 compared to a resolved simulation. Finally, we present some numerical examples for the full Navier–Stokes system to illustrate the convergence and performance of the approach.

Key words. temporal multiscale, Navier–Stokes, finite elements, time discretization, periodic solutions, a priori analysis

AMS subject classifications. 65N12, 65L20, 76D05

DOI. 10.1137/19M1258396

1. Introduction.

We are interested in the numerical approximation and long term simulation of flow problems that carry a multiscale character in time. Such problems appear for example in the formation of atherosclerotic plaque in arteries, where flow dynamics acting on a scale of milliseconds to seconds have an effect on plaque growth in the vessel, which typically takes place within a range of several months. Another application is the investigation of chemical flows in pipelines, where long-time effects of weathering, accelerated by the transported substances, cause material alteration.

These examples have in common, that it is computationally infeasible to resolve the fast scale over the whole time interval of interest. In the case of atherosclerotic plaque growth, a suitable time step of $\frac{1}{20}$ s would require nearly 10^9 steps to cover the period of interest, which is at least 6 months.

Inspired by the temporal dynamics of atherosclerotic plaque growth, we will consider the flow in a channel whose boundary is deformed over a long time scale. This deformation is controlled by the concentration variable $u(t)$ that is governed by a simple reaction equation and that depends on the fluid forces:

$$\begin{aligned} &\mathbf{v}(0) = \mathbf{v}_0, \qquad \operatorname{div} \mathbf{v} = 0, \quad \rho(\partial_t \mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v}) - \operatorname{div} \boldsymbol{\sigma}(\mathbf{v}, p) = \mathbf{f} \qquad \text{in } \Omega(u(t)), \\ &u(0) = 0, \qquad\qquad u' = \epsilon R(u, \mathbf{v}). \end{aligned} \tag{1}$$

Here, ρ is the density of the fluid, $\boldsymbol{\sigma} = \rho\nu(\nabla\mathbf{v} + \nabla\mathbf{v}^T) - pI$ the Cauchy stress with the kinematic viscosity ν, and $R(\mathbf{v}, u) \geq 0$ a reaction term describing the influence of the fluid forces (namely, the wall shear stress) on the boundary growth. The growth

*Received by the editors April 26, 2019; accepted for publication (in revised form) March 30, 2020; published electronically May 26, 2020.
https://doi.org/10.1137/19M1258396

†Department of Mathematics and Statistics, University of Konstanz, Konstanz, 78457, Germany (stefan.frei@uni-konstanz.de).

‡Otto-von-Guericke-Universität Magdeburg, 39106 Magdeburg, and Interdisciplinary Center for Scientific Computing, Heidelberg, Germany (thomas.richter@ovgu.de).

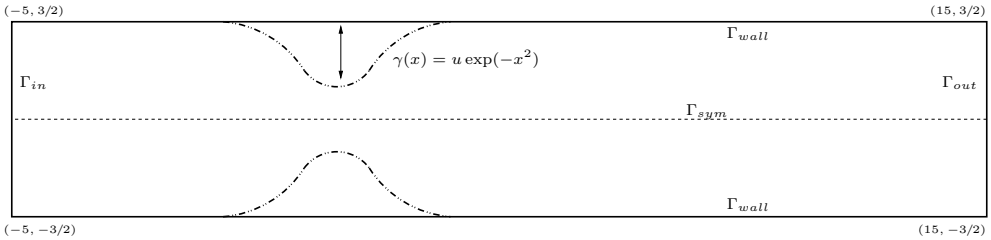


FIG. 1. *Configuration of the test case. We study flow in a channel with a boundary* Γ *that depends on a concentration variable* u*. This* u *follows a simple reaction law with a right-hand side depending on the wall shear stress on* Γ_{wall}*.*

term $R(\cdot,\cdot)$ is modeled such that $|R(\cdot,\cdot)| = \mathcal{O}(1)$:

$$
\begin{aligned}
R(u,\mathbf{v}) &\coloneqq \big(1+u\big)^{-1}\big(1+|\sigma_{WSS}(\mathbf{v})|^2\big)^{-1},\\
\sigma_{WSS}(\mathbf{v}) &\coloneqq \sigma_0^{-1}\int_\Gamma \rho\nu\big(I_d-\vec{n}\vec{n}^T\big)(\nabla\mathbf{v}+\nabla\mathbf{v}^T)\vec{n}\,\mathrm{d}o,
\end{aligned}
\tag{2}
$$

where $\vec{n}$ denotes the outward facing unit normal vector at the boundary Γ. The parameter $\sigma_0 > 0$ will be tuned to give $|\sigma_{WSS}(v)| = \mathcal{O}(1)$; see section 5. The domain $\Omega = \Omega(u(t))$ depends explicitly on the concentration $u(t)$. We show a sketch of the configuration in Figure 1. The flow problem is driven by a periodic oscillating inflow profile of period 1 s:

$$
\mathbf{v} = \mathbf{v}^D \text{ on } \Gamma_{in} \text{ with } \mathbf{v}^D(t) = \mathbf{v}^D(t+1s).
$$

This period describes the *fast scale* of the problem. By $\epsilon \ll 1$ we denote a small parameter that controls the time scale of the (slow) growth of the concentration, i.e., $|u'| = \mathcal{O}(\epsilon)$ and $T = \mathcal{O}(\epsilon^{-1})$ is the expected long term horizon. While the problem itself is strongly simplified compared to the detailed nonlinear mechano-chemical fluid-structure interaction model of plaque growth [8, 50, 17], we choose the parameters in such a way that the temporal dynamics are very similar.

The structure of this article is as follows: In section 2 we introduce a simple model problem consisting of two coupled ODEs, that are related to (1), and for which we will be able to conduct a complete error analysis. Moreover, we discuss some of the available approaches in literature and outline the multiscale algorithm developed in this article. In section 3 we derive the effective long term equations and give an error analysis on the continuous level. In section 4 we describe the temporal discretization of the multiscale scheme and show optimal order convergence in all discretization parameters: mesh size h, time step size k for the fast problem and time step size K for the slowly evolving variable. In section 5 we apply the multiscale scheme to the complex problem introduced in section 2, which is based on the Navier–Stokes equations. We show numerically optimal order convergence in agreement with the theoretical findings for the simplified system. We conclude with a short summary and a discussion of some open problems.

2. Time scales. In this section we start by analyzing the temporal multiscale character of the plaque formation problem. We simplify the coupled problem and introduce a model problem coupling two ODEs. Then, we present various approaches for the numerical treatment of temporal multiscale problems that are discussed in literature. Finally we sketch the idea of the multiscale scheme that is considered in this work, which fits into the framework of the heterogeneous multiscale method.

2.1. A model problem. To start with, we introduce a simple model problem, a system of two ODEs that shows the same coupling and temporal multiscale characteristics as the full plaque growth system:

$$u(0) = u_0, \qquad u'(t) = \epsilon R(u(t), v(t)), \tag{3a}$$

$$v(0) = v_0, \qquad v'(t) + \lambda(u(t))v(t) = f(t), \tag{3b}$$

where $f(t) = f(t+1)$ is periodic and $R(\cdot,\cdot)$ is given by

$$R(u, v) \coloneqq \big(1 + u\big)^{-1}\big(1 + v^2\big)^{-1}. \tag{4}$$

For the reaction term it holds $|R(u, v)| \leq 1$. The parameter $\lambda(u) \geq \lambda_0 > 0$ depends on the concentration u. We will assume that the relation $u \to \lambda(u)$ is differentiable and that the derivative $\frac{d\lambda(u)}{du}$ remains bounded.

In section 5.2 we will argue that this system can indeed be considered as a simplification of the full plaque growth system by neglecting the nonlinearity and by diagonalizing the resulting Stokes equation with respect to an orthonormal eigenfunction basis. Moreover, if we introduce $\tau \coloneqq \epsilon t$, $v_\tau(\tau) \coloneqq v(t)$, $u_\tau(\tau) \coloneqq u(t)$, as well as $f_\tau(\tau) \coloneqq f(t)$ we can scale this system to

$$v_\tau(0) = v_0,\ u_\tau(0) = u_0, \quad v'_\tau + \epsilon^{-1}\lambda(u_\tau)v_\tau = \epsilon^{-1} f_\tau,\ u'_\tau = R(u_\tau, v_\tau), \tag{5}$$

which reveals the typical character of ODE systems with multiple scales in time as discussed in [14, 1]. In the language of the heterogeneous multiscale method (HMM) (see also [15]), such a problem is called a type B problem and it is characterized by the acting of fast and slow scales throughout the whole (long) time span $[0, T]$ in contrast to problems with localized singular behavior. Since $|R| = \mathcal{O}(1)$ it holds $|u'_\tau| = \mathcal{O}(1)$ and u_τ describes the slow variable while $|v'_\tau| = \mathcal{O}(\epsilon^{-1})$ indicates the fast and oscillatory variable.

2.2. Numerical approaches for temporal multiscale problems. While multiscale problems in space are extensively studied in literature (see, e.g. [10, 38]), fewer works are found on problems with multiscale character in time. Some literature exists that uses a homogenization approach based on asymptotic expansions in time for viscoelastic, viscoplastic, or elasto-viscoplastic solids [22, 52, 3, 24]. Under suitable assumptions, the short-scale part of the multiscale algorithm becomes stationary for this class of equations, such that difficulties to define initial values on the short scale are avoided.

Multirate time-stepping methods [20] split the system into slow and large components and use different time step sizes according to the dominant scales. All scales are still resolved on the complete time interval. Since the fast scale of problem (1) which requires a small time step is the computationally intensive Navier–Stokes equations and since the scales are vastly separated, such multirate methods are not appropriate for the problem under investigation.

In the context of continuum damage mechanics, processes with high frequent oscillatory impact can be approached by block cycle jumping techniques [33], where a large number of cycles are skipped and replaced by a linear approximation of the damage effect. An overview of different techniques is given in the first two introductionary sections of [7]. These approaches do not resolve the complete system on a full temporal interval but reside on local solutions. This gives rise to the problem of finding initial values.

If the time scales are close enough that the short-scale dynamics can be resolved within one time step of the long-scale discretization, the *variational multiscale method* [28, 6] or approaches that construct long-scale basis functions from the short-scale information [39, 2] are applicable. Similar algorithms are also used to construct parallel-in-time integrators, for example, the *parareal* method [35]. In this work, we are interested in problems with a stronger *scale separation*, where the resolution of the short scale within a long-scale interval is very costly up to being computationally unfeasible.

Only very few numerical works can be found concerning flow problems with multiple scales in time. An exception are the works of Masud and Khurram [36, 31], where the variational multiscale method is applied, assuming again that the time scales are sufficiently close. On the other hand, several theoretical works exist that show convergence towards averaged equations for specific flow configurations in the situation that the ratio of time scales $\epsilon = \frac{t_{\text{fast}}}{T_{\text{long}}}$ tends to zero (see, e.g., [29, 9, 34]), however, without considering practical numerical algorithms or discretization.

A common numerical approach is to replace the fast problem by an averaged one using a fixed-in-time inflow profile [50, 8]. It is however widely accepted and also confirmed in numerical studies [17] that such a simple averaging does not necessarily reproduce the correct dynamics. In [17] we presented a first multiscale scheme for the approximation of such a problem, however, with a focus on the modeling of a full closure of the channel and without any analysis on the robustness and accuracy. Similar algorithms can be found in Sanders, Verhulst, and Murdock [44] and by Crouch and Oskay [11] in different applications. In this work, we will derive an improved algorithm in a mathematically rigorous way, including a detailed error analysis for both modeling and discretization errors. To our knowledge this is the first time that the interplay between modeling errors of the temporal multiscale scheme and temporal discretization errors on both scales is analyzed.

One of the most prominent class of techniques is the *HMM* [15, 14, 1, 16] that aims at an efficient decoupling of macroscale and microscale, where the latter one enters the macroscale problem in terms of temporal averages. Typically, the procedure is as follows: one determines the fast and the slow variables of the coupled problem. For the slow variables an integrator with a long time step $\Delta T := T_{n+1} - T_n$ and good stability properties is used. At each of these macro-time-steps, the fast-scale problem is initialized based on the current slow variable and solved on the interval $I_n^\eta := [T_n, T_n + \eta]$. Finally, the fast variable output on I_n^η is averaged to yield the effective operator for the slow-scale problem.

The efficiency of the resulting HMM scheme depends on the choice of η which indicates the scale to allow for equilibration and adjustment of the micro model. Too large values will reduce the efficiency, too small values will limit the accuracy. The underlying problem is the lack of initial values at the new macrostep for the fast scale, which in our case (1) or (3), is the oscillatory velocity $\mathbf{v}(t)$ and $v(t)$, respectively. The realization presented in this article is based on time-periodic solutions to the micro problem. Instead of solving the microscale problem on an interval I_n^η at macrostep T_n we aim at a localized solution of the fluid problem that satisfies a periodicity condition in time. This approach allows us to conduct a complete error analysis of the resulting scheme when applied to the simplified model problem (3).

2.3. Outline of the multiscale scheme. We conclude this section by briefly describing the multiscale algorithm that is considered in this article. The derivation given here is based on problem (3). We start by defining the slow variable as an

average of the concentration $u(t)$,

$$U(t) \coloneqq \int_t^{t+1\,\mathrm{s}} u(s)\,\mathrm{d}s. \tag{6}$$

This gives rise to the averaged equation for the long term dynamics

$$U'(t) = \int_t^{t+1} \epsilon R(u(s), v(s))\,\mathrm{d}s. \tag{7}$$

A time integration formula with a macro-time-step is used to approximate this equation. Two approximation steps are performed to reach an effective equation. First, the reaction term in (7) is evaluated in $U(t)$ instead of $u(s)$ and, second, the fast component $v(s)$ will be replaced by the localized solution of the time-periodic problem obtained for a fixed value of $U(t)$:

$$v'_{U(t)} + \lambda(U(t))v_{U(t)} = f \text{ in } [0,1] \text{ with } v_{U(t)}(1) = v_{U(t)}(0). \tag{8}$$

These approximations will be discussed and analyzed in the following section. Assuming that (7), approximated in these two steps, is integrated with the forward Euler method, a macro-time-step is given by

$$U_n = U_{n-1} + (T_n - T_{n-1}) \int_{T_{n-1}}^{T_n} \epsilon R(U_{n-1}, v_{U_{n-1}}(s))\,\mathrm{d}s. \tag{9}$$

2.3.1. Motivation for the locally periodic approximations. Due to the nonlinearity of the reaction term $R(\cdot,\cdot)$, the microscale variations in $v(s)$ cannot simply be averaged. Instead the velocities $v(s)$ need to be computed on the fast scale in each macrostep $T_{n-1} \to T_n$ in order to obtain a good approximation of the integral on the right-hand side of (7). A computation of $v(s)$ over the complete interval $[T_{n-1}, T_n]$ is, however, unfeasible for small ϵ. For this reason, the imposition of accurate initial values $v(T_n)$ for the fast-scale problem is not straightforward. Neither the short-scale velocity $v(T_{n-1})$ from the previous macro-time-step nor an averaged quantity $V(T_n)$ can guarantee a sufficiently good approximation for $v(T_n)$. In practice, a relaxation time η is frequently introduced (see, for example, [16, 1]), in order to improve the initial values by means of a few forward iterations.

As an alternative, we propose to introduce the time-periodic fast-scale problem (8). This has the advantage that in principle only one period of the fast-scale problem needs to be resolved per macrostep. We can show theoretically (Lemma 8) that the approximation error introduced by the periodic problem is of order ϵ. Efficient approximations of these time-periodic problems will be discussed in section 4.3.

2.3.2. Abstract multiscale scheme. We conclude this section by formulating the abstract multiscale scheme that can be applied to both problems, the plaque growth system and the simplified model problem.

ALGORITHM 1 (abstract multiscale scheme). *Let $0 = T_0 < T_1 < \cdots < T_N = T$ be a partition of the macrointerval with uniform step size $K \coloneqq T_n - T_{n-1}$. Further, let $U_0 \coloneqq u_0$ be the initial value of the slow variable. Iterate for $n = 1, 2, \ldots$*

1. *Solve the time-periodic problem* (12) *or* (13) *for* $\mathbf{v}_{U_{n-1}}$.
2. *Evaluate the reaction term*

$$R_{n-1} \coloneqq \int_0^1 R(U_{n-1}, \mathbf{v}_{U_{n-1}}(s))\,ds.$$

3. *Forward the slow variable with an (explicit) one-step scheme*

$$U_n = \mathcal{F}(K; U_{n-1}; R_{n-1}).$$

The structure of the time integrator $\mathcal{F}$ depends on the system of equations and the time-stepping method. For simplicity, we have formulated the algorithm for an explicit time integrator in the slow variable. The use of an implicit time-stepping scheme would require an iterated evaluation of the periodic problem for updated values of U_n. In the case of r-step schemes periodic solutions $\mathbf{v}_{U_{n-k}}$ would be required for $k = 1, \ldots, r$.

Remark 2. We note that the proposed algorithm does not compute an average of the fast variable v. An approximation to v is only computed on the short periodic interval of the fast scale as $\mathbf{v}_{U_{n-1}}$ (step 1). The slow variable u, on the other hand, is only computed on the slow scale as the average U (step 3).

The typical behavior of the slow and fast variables is illustrated in Figure 2 (top). On the bottom of Figure 2, the multiscale algorithm is visualized, including the transfer of quantities between the slow and the fast scales. For ease of presentation a large ϵ has been chosen for the purpose of visualization.

To conclude this section we anticipate the main result of the analysis given below. For the combination of the second-order Adams–Bashforth rule for the discretization of the slow problem and the Crank–Nicolson scheme for the fast problem we will show optimal convergence of the resulting multiscale scheme:

$$|U_N - u(T)| = \mathcal{O}(\epsilon) + \mathcal{O}(\epsilon^2 K^2) + \mathcal{O}(k^2) + \mathcal{O}(\mathrm{tol}_P).$$

By K we denote the step size of the macro solver, k is the step size of the micro solver, and by tol_P we denote the tolerance of the periodicity constraint: $\max_n |v_{U_n}(1) - v_{U_n}(0)| < \mathrm{tol}_P$.

3. Derivation and analysis of the effective equations. In this section we derive the temporal multiscale scheme that has been outlined in the previous section. We will discuss the coupled Navier–Stokes problem on the evolving domain $\Omega(u)$, problem (1), and the reduced ODE system (3) side by side. Whenever results apply to the ODE system only, we will clearly mention this. We start by collecting some preliminary assumptions on the underlying problems.

3.1. Preliminaries.

Assumption 3 (reaction term). Let $u_{\max} < \infty$ be a maximum concentration. Let $0 \le u \le u_{\max}$ and $\mathbf{v} \in X$ ($X = \mathbb{R}$ for the model problem, $X = H^2(\Omega)$ for the plaque growth problem). The reaction term is bounded

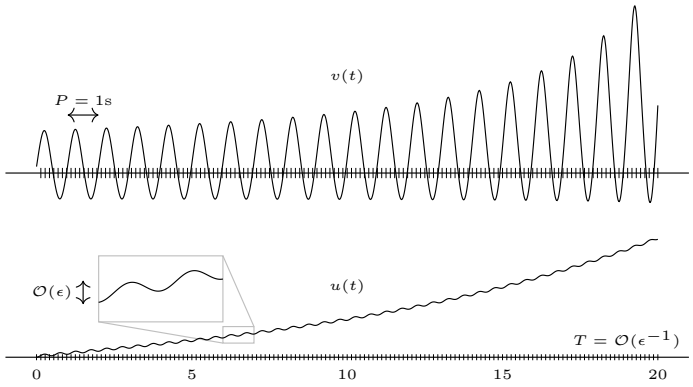
$$|R(u, \mathbf{v})| \le C_{A3a}, \tag{10}$$

and Lipschitz continuous in both arguments

$$|R(u_1, \mathbf{v}) - R(u_2, \mathbf{v})| \le C_{A3b}|u_1 - u_2|, \quad |R(u, \mathbf{v}_1) - R(u, \mathbf{v}_2)| \le C_{A3b}\|\mathbf{v}_1 - \mathbf{v}_2\|_X, \tag{11}$$

where the constant C does not depend on ϵ.

Assumption 3 is easily verified for the simplified reaction term (4). A proof for the Navier–Stokes case will be given in section 5.2.



Layout of the multiscale problem. The fast variable $v(t)$ (top) and the slow variable $u(t)$ (bottom) couple on the complete long time interval $I = [0, T]$. The slow variable is also oscillating, but the oscillations are small, of size $\mathcal{O}(\epsilon)$. The fast variable is locally nearly periodic.

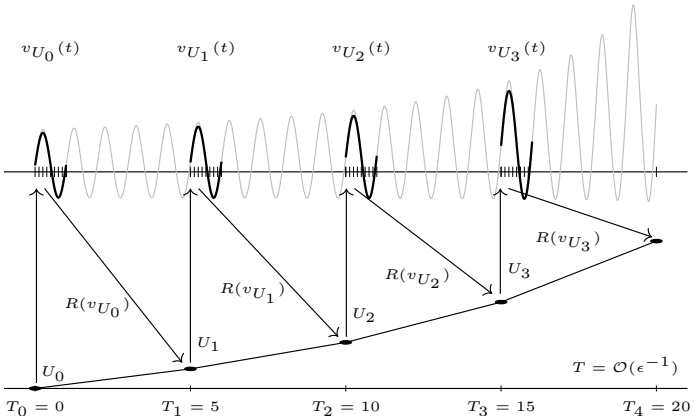


FIG. 2. *Construction of the multiscale scheme:* 0. *The slow variable $U(t)$ is discretized with a time-stepping scheme with macro-step-size $K \gg 1$.* 1. *In each step the current slow state U_{n-1} is transferred to the fast problem (top) and a periodic solution $\mathbf{v}_{U_{n-1}}$ is computed on $[T_{n-1}, T_{n-1}+1]$ as an approximation for $\mathbf{v}(t)$.* 2. *The averaged reaction term $R_{n-1} = R(\mathbf{v}_{U_{n-1}})$ is computed from $\mathbf{v}_{U_{n-1}}$ and transferred to the slow problem (bottom).* 3. *The slow variable U_n is updated by the macrostep $T_{n-1} \to T_n$.*

Remark 4 (generic constants). Throughout this manuscript we use generic constants C. These constants may depend on the domain Ω, the maximum concentration $u_{\max}$, the right-hand side $\mathbf{f}$, and the Dirichlet data. They do not, however, depend on the solution, the scale parameter ϵ, or the discretization parameters that will be introduced in the remainder of this article.

We further assume that the isolated micro problems allow for a unique periodic solution.

Assumption 5 (periodic solution). Let $u \in \mathbb{R}$ with $0 \le u \le u_{\max}$. We assume that there exist unique periodic solutions $v_u \in C([0,1])$ to

$$\partial_t v_u + \lambda(u) v_u = f \text{ in } [0,1], \quad v_u(1) = v_u(0), \tag{12}$$

as well as solutions $\mathbf{v}_u \in H^2(\Omega(u), p_u \in H^1(\Omega(u))$ to the incompressible Navier–Stokes

equations

$$\nabla \cdot \mathbf{v}_u = 0, \quad \rho(\partial_t \mathbf{v}_u + (\mathbf{v}_u \cdot \nabla)\mathbf{v}_u) - \operatorname{div} \boldsymbol{\sigma}(\mathbf{v}_u, p_u) = \mathbf{f} \qquad \text{in } [0,1] \times \Omega(u), \tag{13a}$$

$$\mathbf{v}_u = \mathbf{v}_u^D \qquad \text{on } [0,1] \times \partial\Omega(u), \tag{13b}$$

$$\mathbf{v}_u(1) = \mathbf{v}_u(0) \qquad \text{in } \Omega(u). \tag{13c}$$

Both solutions are uniformly bounded in time

$$\sup_{t\in[0,T]} |v_u(t)| \le C, \quad \sup_{t\in[0,T]} \left(\|\mathbf{v}_u(t)\|_{H^2(\Omega)} + \|p_u(t)\|_{H^1(\Omega)} \right) \le C. \tag{14}$$

For the ODE problem (12), the existence of a unique periodic solutions follows from the evolution of $w(t) = v(t+1) - v(t)$ which fulfills $w' + \lambda(u)w = 0$ and thus vanishes; see [42]. For a discussion on the Navier–Stokes equations we refer to section 5.2.

3.2. Derivation of an effective equation. We introduce the averaged concentration

$$U(t) \coloneqq \int_t^{t+1} u(s)\,\mathrm{d}s.$$

Using (1) and (3), respectively, and inserting $\pm R(U(t), \mathbf{v}(s))$, we have

$$\begin{aligned} U'(t) &= \int_t^{t+1} \epsilon R\big(u(s), \mathbf{v}(s)\big)\,\mathrm{d}s \\ &= \int_t^{t+1} \epsilon R\big(U(t), \mathbf{v}(s)\big)\,\mathrm{d}s - \int_t^{t+1} \epsilon \left(R\big(U(t), \mathbf{v}(s)\big) - R\big(u(s), \mathbf{v}(s)\big)\right)\,\mathrm{d}s. \end{aligned}$$

LEMMA 6 (averaging error). *Let $u \in C^1([0,T])$, $\mathbf{v} \in C([0,T]; X)$, and let Assumption 3 be satisfied. Then, it holds*

$$\max_t \left| \int_t^{t+1} \left(R(U(t), \mathbf{v}(s)) - R(u(s), \mathbf{v}(s)) \right)\, ds \right| \le C\epsilon$$

with a constant $C > 0$ that depends on Assumption 3.

Proof. By Lipschitz continuity of $R(\cdot,\cdot)$ in the first argument (11) it holds

$$|R(U(t), \mathbf{v}(s)) - R(u(s), \mathbf{v}(s))| \le C|u(s) - U(t)|. \tag{15}$$

We estimate

$$\begin{aligned} \int_t^{t+1} |u(s) - U(t)|\,\mathrm{d}s &= \int_t^{t+1} \left| \int_t^{t+1} (u(s) - u(r))\,\mathrm{d}r \right| \mathrm{d}s \\ &= \int_t^{t+1} \left| \int_t^{t+1} \int_r^s u'(x)\,\mathrm{d}x\,\mathrm{d}r \right| \mathrm{d}s \le C\epsilon, \end{aligned} \tag{16}$$

where we used (10), such that a combination of (15) and (16) shows the assertion. ∎

We can thus approximate the averaged evolution equation for U by

$$U'(t) = \int_t^{t+1} \epsilon R(U(t), \mathbf{v}(s))\,\mathrm{d}s + \mathcal{O}\left(\epsilon^2\right). \tag{17}$$

The benefit of introducing the periodic solution lies in a localization of the fast-scale influences. Given an approximation U_n at time T_n, the microscale influence $(\mathbf{v}_{U_n}, p_{U_n})$ can be determined independently of the last approximation $(\mathbf{v}_{U_{n-1}}, p_{U_{n-1}})$. We approximate the averaged equation (17) by inserting the periodic solution $\mathbf{v}_{U(t)}(s)$ for a fixed value $U(t)$:

$$\begin{aligned} U'(t) &= \int_t^{t+1} \epsilon R(U(t), \mathbf{v}_{U(t)}(s))\,\mathrm{d}s \\ &\quad + \int_t^{t+1} \epsilon\big(R(U(t), \mathbf{v}(s)) - R(U(t), \mathbf{v}_{U(t)}(s))\big)\,\mathrm{d}s + \mathcal{O}\left(\epsilon^2\right). \end{aligned} \tag{18}$$

We will show that the second remainder

$$\max_t \left| \int_t^{t+1} \epsilon\big(R(U(t), \mathbf{v}(s)) - R(U(t), \mathbf{v}_{U(t)}(s))\big)\,\mathrm{d}s \right| \tag{19}$$

is also of order $\mathcal{O}(\epsilon^2)$. The analysis is presented for the ODE system (3) in the following section (Lemmas 8, 9, and 10). Extensions to the full system (1) are discussed in section 5. Having shown that the average $U(t)$ satisfies the equation

$$U'(t) = \int_t^{t+1} \epsilon R\big(U(t), \mathbf{v}_{U(t)}(s)\big)\,\mathrm{d}s + \mathcal{O}\left(\epsilon^2\right), \tag{20}$$

we define the effective equation for the approximation of $U(t)$ by neglecting the remainder of order $\mathcal{O}(\epsilon^2)$, i.e., by the equation

$$U'(t) = \int_t^{t+1} \epsilon R\big(U(t), \mathbf{v}_{U(t)}(s)\big)\,\mathrm{d}s, \quad U(0) = u_0. \tag{21}$$

In Lemma 10, we will estimate the error resulting from skipping the remainder $\mathcal{O}(\epsilon^2)$ in (20). We further note that the initial values $U(0) = u_0$ and the averaged initial $\int_0^1 u(s)\,\mathrm{d}s$ do not necessarily coincide. Instead, (21) deals with an offset of order $\mathcal{O}(\epsilon)$:

$$\int_0^1 u(t)\,\mathrm{d}t = \int_0^1 u(0) + \int_0^s u'(s)\,\mathrm{d}s\,\mathrm{d}t = u_0 + \mathcal{O}\left(\epsilon\right). \tag{22}$$

3.3. Analysis of the averaging error for the model problem. In this section, we outline the ideas for showing convergence of the multiscale scheme, Algorithm 1. As mentioned above the analysis for the Navier–Stokes/ODE system is beyond the scope of this work. Instead we consider problem (3). One reason is the lack of unique periodic solutions $(\mathbf{v}_{U(t)}(s), p_{U(t)}(s))$ for larger Reynolds numbers. Second, the following analysis is based on the linearity of the model problem. In addition to Assumptions 3 and 5 we assume the following.

Assumption 7. We assume that the map $u \mapsto \lambda(u)$ is differentiable with a bounded derivative

$$\left| \frac{d\lambda(u)}{du} \right| \leq C_{A7}, \quad u \in [0, u_{\max}]. \tag{23}$$

Algorithm 1 applied to the model problem (3) calls for the solution of the following averaged slow problem

$$U'(t) = \int_t^{t+1} \epsilon R\big(U(t), v_{U(t)}(s)\big)\,\mathrm{d}s, \quad U(0) = u_0, \tag{24}$$

and the corresponding time-periodic micro problems

$$v_U'(t) + \lambda(U) v_U(t) = f(t), \quad v_U(1) = v_U(0), \tag{25}$$

for each fixed parameter $0 \le U \le u_{\max}$.

LEMMA 8 (periodic solutions). *Let $0 \le u \le u_{\max}$ be fixed and let Assumption 7 hold. For the solution to the periodic problem (25) it holds*

$$|v_u(t)| \le C_{L8a}. \tag{26}$$

Further, for $0 \le u, \eta \le u_{\max}$ let $v_u(t), v_\eta(t)$ be two such periodic solutions. It holds

$$|v_u(t) - v_\eta(t)| \le C_{L8b} |\lambda(u) - \lambda(\eta)|, \tag{27}$$

where C depends on f and $\lambda_0 > 0$.

Proof. (i) To show (26) we skip the index u for better readability. The general solution to the ODE is given by

$$v(t) = \mathrm{e}^{-\lambda t} \left(v(0) + \int_0^t f(s) \mathrm{e}^{\lambda s} \, \mathrm{d}s \right), \tag{28}$$

which we estimate by

$$|v(t)| \le \mathrm{e}^{-\lambda t} |v_0| + \frac{1}{\lambda} \|f\|_{L^\infty([0,1])}. \tag{29}$$

Since $v(t)$ is periodic, $v(1) = v(0)$, we obtain by (28)

$$v(0) = \frac{\mathrm{e}^{-\lambda}}{1 - \mathrm{e}^{-\lambda}} \int_0^1 f(s) \mathrm{e}^{\lambda s} \, \mathrm{d}s \quad \Rightarrow \quad |v(0)| \le \frac{1}{\lambda} \|f\|_{L^\infty([0,1])}. \tag{30}$$

Inserting (30) into (29) we get for all $t \in [0,1]$

$$|v(t)| \le \frac{1 + \mathrm{e}^{-\lambda t}}{\lambda} \|f\|_{L^\infty([0,1])} \le \frac{2}{\lambda} \|f\|_{L^\infty([0,1])}, \tag{31}$$

which gives (26) since $\lambda \ge \lambda_0$.

(ii) Let $w(t) := v_u(t) - v_\eta(t)$. It holds

$$w'(t) + \lambda(u) w(t) = \big(\lambda(\eta) - \lambda(u)\big) v_\eta(t), \quad w(1) = w(0) = v_u(0) - v_\eta(0).$$

Note that the right-hand side of this ODE is time periodic. Hence, we use (31) twice and obtain the estimate

$$|w(t)| \le \frac{2}{\lambda(u)} |\lambda(\eta) - \lambda(u)| \max_{t \in [0,1]} |v_\eta(t)| \le \frac{4}{\lambda(u)\lambda(\eta)} |\lambda(\eta) - \lambda(u)| \|f\|_{L^\infty([0,1])}. \qquad \square$$

The following essential lemma sets the foundation for replacing the dynamic fast component $v(t)$ by localized periodic-in-time solutions. For a given slow function $u(t)$ we will compare the corresponding dynamic fast scale $v(t)$ with the family of periodic solutions $v_{u(t)}(s)$.

LEMMA 9. *Let $u \in C([0,T])$ be given with*

$$u(0) = 0, \quad 0 \le u'(t) \le C_{A3a}\epsilon, \quad t \in [0,T]. \tag{32}$$

Then, let $v(t)$ be the dynamic solution to (3b)*, i.e.,*

$$v(0) = v_0, \quad \partial_t v(t) + \lambda\big(u(t)\big)v(t) = f(t) \text{ for } t \in [0,T], \tag{33}$$

and let $v_{u(t)}(s)$ be the family of time-periodic solutions to

$$\begin{aligned} v_{u(t)}(0) &= v_{u(t)}(1) \\ \partial_s v_{u(t)}(s) + \lambda\big(u(t)\big)v_{u(t)}(s) &= f(s) \text{ for } s \in [0,1] \text{ and for all } t \in [0,T]. \end{aligned} \tag{34}$$

Finally, let $v_0 = v_{u(0)}(0)$, i.e., the initial values to (33) *and* (34) *at time $t = 0$ agree. Let $\lambda(\cdot)$ satisfy Assumption* 7*. Then it holds*

$$|v(t) - v_{u(t)}(t)| \le C_{L9}\epsilon$$

with a constant $C > 0$ that depends on f, λ_0, and on Assumptions 3 *and* 7*.*

Proof. For $v_{u(t)}(t)$ it holds by the chain rule

$$\frac{d}{d_t} v_{u(t)}(t) = v'_{u(t)}(t) + \frac{dv_{u(t)}}{du(t)}(t)u'(t),$$

such that $v_{u(t)}(t)$ is governed by

$$\partial_t v_{u(t)}(t) + \frac{dv_{u(t)}}{du(t)}(t)u'(t) + \lambda\big(u(t)\big)v_{u(t)}(t) = 0, \quad v_{u(0)}(0) = v_0.$$

Thus, it holds for the difference $w(t) := v(t) - v_{u(t)}(t)$,

$$\partial_t w(t) + \lambda\big(u(t)\big)w(t) = -\frac{dv_{u(t)}}{du(t)}(t)u'(t), \quad w(0) = 0$$

with the solution

$$w(t) = -\int_0^t \frac{dv_{u(s)}}{du(s)}(s)u'(s) \exp\left(-\int_s^t \lambda\big(u(r)\big)\,\mathrm{d}r\right)\,\mathrm{d}s. \tag{35}$$

To estimate the derivative $\frac{dv_{u(s)}}{du(s)}$ we consider two such time-periodic solutions $v_{u(t)}$ and $v_{\eta(t)}$ for fixed $0 \le u, \eta \le u_{\max}$. We estimate their distance by (27) in Lemma 8:

$$\frac{|v_u - v_\eta|}{|u - \eta|} \le \frac{4}{\lambda_0^2}\|f\|_{L^\infty([0,1])}\left|\frac{\lambda(u) - \lambda(\eta)}{u - \eta}\right|. \tag{36}$$

This bound is uniform in u, η, and t, such that differentiability of $\lambda(u)$, (23), gives

$$\left|\frac{dv_{u(t)}}{du(t)}(t)\right| = \lim_{\eta \to u} \frac{|v_{u(t)} - v_{\eta(t)}|}{|u - \eta|} \le \frac{4C_{A7}}{\lambda_0^2}\|f\|_{L^\infty([0,1])}.$$

This allows us to estimate (35) by

$$|w(t)| = |v(t) - v_{u(t)}(t)| \le \frac{4C_{A7}C_{A3a}}{\lambda_0^2}\|f\|_{L^\infty([0,1])}\epsilon.$$

∎

In the previous lemma we investigated the coupling from a fixed slow variable $u(t)$ to the fast components $v(t)$ and $v_{u(t)}(t)$. This last lemma will study the different evolutions of the slow variable $u(t)$ governed by (3a), (3b) and of the averaged variable $U(t)$ that is determined by (24) with periodic micro influences (25).

LEMMA 10. *Let* $(u(t), v(t))$ *and* $(U(t), v_{U(t)}(t))$ *be defined by* (3a), (3b) *and* (24), (25), *respectively, with the initial values* $u(0) = U(0) = 0$ *and* $v(0) = v_{U(0)}(0)$. *For* $0 \le t \le T = \mathcal{O}(\epsilon^{-1})$ *it holds that*

$$|U(t) - u(t)| \le C\epsilon$$

with a constant $C > 0$ *that depends on the constants from Lemma* 8, 9, *and* 10, *as well as Assumption* 3.

Proof. We introduce

$$w(t) \coloneqq U(t) - \int_t^{t+1} u(s)\,\mathrm{d}s,$$

which is governed by

$$\begin{aligned} w'(t) &= \int_t^{t+1} \epsilon\Big(R\big(U(t), v_{U(t)}(s)\big) - R\big(u(s), v(s)\big)\Big)\,\mathrm{d}s, \\ w(0) &= U(0) - \int_0^1 u(s)\,\mathrm{d}s =: w_0. \end{aligned}$$

The initial error is small, $|w_0| = \mathcal{O}(\epsilon)$; compare (22). We insert $\pm R(U(t), v(s))$:

$$\begin{aligned} w'(t) = \int_t^{t+1} &\epsilon\Big(R\big(U(t), v_{U(t)}(s)\big) - R\big(U(t), v(s)\big)\Big) \\ &+ \epsilon\Big(R\big(U(t), v(s)\big) - R\big(u(s), v(s)\big)\Big)\,\mathrm{d}s. \end{aligned} \tag{37}$$

Lipschitz continuity of $R(\cdot,\cdot)$, Assumption 3, gives

$$|w'(t)| \le C_{A3b}\epsilon \int_t^{t+1} |v_{U(t)}(s) - v(s)|\,\mathrm{d}s + C_{A3b}\epsilon \int_t^{t+1} |U(t) - u(s)|\,\mathrm{d}s. \tag{38}$$

The second term is estimated by inserting $\pm \int_t^{t+1} u(r)\,\mathrm{d}r$ and by using $|u'| \le C_{A3a}\epsilon$:

$$\begin{aligned} &\int_t^{t+1} |U(t) - u(s)|\,\mathrm{d}s \\ &\quad\le \int_t^{t+1} \left|U(t) - \int_t^{t+1} u(r)\,\mathrm{d}r\right|\,\mathrm{d}s + \int_t^{t+1} \left|\int_t^{t+1} \big(u(r) - u(s)\big)\,\mathrm{d}r\right|\,\mathrm{d}s \\ &\quad= \int_t^{t+1} |w(t)|\,\mathrm{d}s + \int_t^{t+1} \left|\int_t^{t+1} \int_s^r u'(x)\,\mathrm{d}x\,\mathrm{d}r\right|\,\mathrm{d}s \le |w(t)| + C_{A3a}\epsilon. \end{aligned} \tag{39}$$

To estimate the first term in (38) we introduce $\pm v_{u(s)}(s)$ and use Lemmas 8 and 9, Assumption 7 (differentiability of $\lambda(u)$), and finally (39):

$$\begin{aligned} \int_t^{t+1} |v_{U(t)}(s) - v(s)|\,\mathrm{d}s &\le \int_t^{t+1} \Big(|v_{U(t)}(s) - v_{u(s)}(s)| + |v_{u(s)}(s) - v(s)|\Big)\,\mathrm{d}s \\ &\le C_{L8b}C_{A7} \int_t^{t+1} |U(t) - u(s)|\,\mathrm{d}s + C_{L9}\epsilon \\ &\le C_{L8b}C_{A7}\big(|w(t)| + C_{A3a}\epsilon\big) + C_{L9}\epsilon. \end{aligned} \tag{40}$$

With $C = C(C_{A3a}, C_{A3b}, C_{A7}, C_{L8a}, C_{L8b}, C_{L9})$ we combine (37)–(40) to find the relation

$$-C\,(\epsilon + |w(t)|)\,\epsilon \le w'(t) \le C\,(\epsilon + |w(t)|)\,\epsilon.$$

An estimate for $|w(t)|$ follows by the a bound of the solution to the corresponding ODE with initial value $w(0) = w_0$, where $|w_0| \le C_{3a}\epsilon$:

$$|w(t)| \le C\epsilon \mathrm{e}^{C\epsilon t}, \tag{41}$$

which satisfies $|w(t)| = \mathcal{O}(\epsilon)$ for $t \le T = \mathcal{O}(\epsilon^{-1})$. Finally,

$$|U(t) - u(t)| \le |w(t)| + \left| \int_t^{t+1} u(s) - u(t)\,\mathrm{d}s \right| \le C\epsilon. \qquad \square$$

4. Time discretization. In this section we introduce second-order time-stepping schemes to approximate the coupled problem. As in the previous section, where we derived the multiscale algorithm, we start with the full plaque growth problem (1). Then, the error analysis for estimating the discretization error is based on the simplified model equations (3).

The discretization is based on the second-order Adams–Bashforth scheme for the slow scale and a Crank–Nicolson scheme for the fast scale. Both choices are exemplary and can in principle be substituted for by any suitable time-stepping scheme. We choose an explicit scheme for the slow scale in order to avoid that several fast-scale problems have to be solved in each time step; see Remark 13 below.

4.1. Second-order multiscale schemes. First, we split the time interval $I = [0, T]$ into subintervals of equal size

$$0 = T_0 < T_1 < \cdots < T_N, \quad K \coloneqq T_n - T_{n-1}. \tag{42}$$

We define approximations $U_n := U(T_n)$ based on the second-order Adams–Bashforth multistep method

$$(AB) \quad \frac{U_{n+1} - U_n}{K} = \frac{3}{2} \int_0^1 \epsilon R(U_n, \mathbf{v}_{U_n;k})\,\mathrm{d}s - \frac{1}{2} \int_0^1 \epsilon R(U_{n-1}, \mathbf{v}_{U_{n-1};k})\,\mathrm{d}s. \tag{43}$$

In order to compute the required starting value U_1 for the Adams–Bashforth scheme, we put one forward Euler step at the beginning of the iteration, which is sufficient to obtain second-order convergence.

These schemes are formally explicit; they depend, however, on the averaged fast-scale influences $R(U_n, \mathbf{v}_{U_n;k})$ To compute these terms, we introduce a (for simplicity again, uniform) temporal subdivision of the fast periodic interval $I_P = [0, 1]$ of step size k,

$$0 = t_0 < t_1 < \cdots < t_M = 1, \quad k \coloneqq t_m - t_{m-1}. \tag{44}$$

Given a fixed value $0 \le U \le u_{\max}$, we introduce the notation $\mathbf{v}_{U,m} \coloneqq \mathbf{v}_{U;k}(t_m)$ and we approximate the periodic solution on the fast scale with the Crank–Nicolson

time-stepping scheme

(45)
$$\nabla \cdot \mathbf{v}_{U,m} = 0, \quad \rho k^{-1}\big(\mathbf{v}_{U,m} - \mathbf{v}_{U,m-1}\big) + \frac{\rho}{2}\big((\mathbf{v}_{U,m-1}\cdot\nabla)\mathbf{v}_{U,m-1} + (\mathbf{v}_{U,m}\cdot\nabla)\mathbf{v}_{U,m}\big) - \frac{1}{2}\operatorname{div}\big(\boldsymbol{\sigma}(\mathbf{v}_{U,m-1}) + \boldsymbol{\sigma}(\mathbf{v}_{U,m})\big) = \frac{1}{2}\big(\mathbf{f}(t_{m-1}) + \mathbf{f}(t_m)\big) \quad m = 1,\dots,M \text{ such that } |\mathbf{v}_M^u - \mathbf{v}_0^u| \le \mathrm{tol}_P.$$

Based on the approximations made in the previous section, we introduce the following multiscale method.

ALGORITHM 11 (explicit temporal multiscale method). *Given subdivisions* (42) *and* (44) *of* $I = [0,T]$ *and* $I_P = [0,1]$. *Let* $U_0 = 0$. *Iterate for* $n = 1,\dots,N$

1. *For* $U \coloneqq U_{n-1}$ *solve the periodic problem* (45) *to obtain* $(\mathbf{v}_{U_{n-1},m}, p_{U_{n-1},m})$ *for* $m = 1,2,\dots,M$.
2. *Compute the averaged feedback*
$$R_{n-1} \coloneqq \frac{k}{M}\sum_{m=1}^{M} R(U_{n-1}, \mathbf{v}_{U_{n-1},m}). \tag{46}$$
3. *Step forward* $U_{n-1} \to U_n$ *with the Adams–Bashforth method* (43)
$$U_n \coloneqq U_{n-1} + \frac{3K}{2}\epsilon R_{n-1} - \frac{K}{2}\epsilon R_{n-2} \tag{47}$$
or, in the first step, with the forward Euler method
$$U_1 \coloneqq U_0 + K\epsilon R_0. \tag{48}$$

Remark 12. In practice we ensure in step 1 that the solution is periodic up to a certain threshold $\|\mathbf{v}_M^u - \mathbf{v}_0^u\| \le \mathrm{tol}_P$. The box rule used to compute the averaged wall shear stress in step 2 of the algorithm is therefore equivalent to the second-order trapezoidal rule (up to the small error $\mathcal{O}(k\,\mathrm{tol}_P)$).

The main computational cost comes from the approximation of the periodic solutions $(\mathbf{v}_{U,m}, p_{u,m})$ for a fixed value of U. The efficient computation of these periodic problems is discussed below.

Remark 13 (implicit multiscale schemes). We are considering the rather simple interaction of the Navier–Stokes equations with a scalar ODE. For more detailed models, for example, a boundary PDE to model the spatially diverse accumulation of $u(x,t)$ along the boundary $\Gamma(u)$ or even a full PDE/PDE model considering dynamical fluid-structure interactions and a detailed modeling of the biochemical processes causing plaque growth, as introduced by Yang [49] and Yang et al. [50], stiffness issues may call for implicit discretizations of the equation for u. To realize an implicit multiscale method, e.g., based on the Crank–Nicolson scheme for both temporal scales, an outer iteration must be introduced. We refer to [37] for a first application of the multiscale scheme to a PDE/PDE coupled flow problem.

4.2. Error analysis for the model problem. We consider the system of equations (3a)–(3b), its multiscale approximation (24)–(25), and the discrete problem (45)–(48). Concerning the short-scale problem, we make the following assumption.

Assumption 14 (approximation of the periodic problem). Let $\mathrm{tol}_P > 0$ be the tolerance for reaching periodicity. We assume that there exists a constant $C > 0$ such

that the Crank–Nicolson discretization (45) to the flow problem satisfies the bound

$$\|\mathbf{v}_{U,M} - \mathbf{v}_{U,0}\| + \max_{m=1,\dots,M} \|\mathbf{v}_U(t_m) - \mathbf{v}_{U,m}\| \leq Ck^2 + \mathrm{tol}_P,$$

where the constant C in particular does not depend on ϵ and U.

Remark 15 (approximation of the periodic problem). Considering ODEs, the error estimate for the trapezoidal scheme is standard and can be found in many textbooks. Applied to the Navier–Stokes equations optimal order estimates under realistic regularity assumptions are given in [26]. Similar estimates that also include second-order-in-time estimates for the pressure (which might be required for a stress-based feedback) are given in [45]. For algorithms to control the periodicity error, we refer to section 4.3 below.

LEMMA 16 (regularity of the solution). *Let $U(t)$ be the solution to* (24) *for $v_U \in C(0,T)$ and $0 \leq U \leq u_{\max}$. Moreover, let the map $U \mapsto \lambda(U)$ be twice differentiable with bounded second derivatives. It holds that*

$$U \in C^3(I), \quad \max_{[0,T]} |U''| = \mathcal{O}\left(\epsilon^2\right), \quad \max_{[0,T]} |U'''| = \mathcal{O}\left(\epsilon^3\right).$$

Proof. Let us first note that U' is bounded due to the continuity of the right-hand side $R(U(t), v_{U(t)}(s))$ of (24). Next, we consider the (total) temporal derivative of the right-hand side. The chain rule gives

$$d_t \int_0^1 R\big(U(t), v_{U(t)}(s)\big)\,\mathrm{d}s = R\big(U(t), v_{U(t)}(1)\big) - R\big(U(t), v_{U(t)}(0)\big) + \int_0^1 d_t R\big(U(t), v_{U(t)}(s)\big)\,\mathrm{d}s.$$

The first part vanishes due to the periodicity of $v_{U(t)}$. For the second part we have with (4)

$$d_t R\big(U(t), v_{U(t)}(s)\big) = -\frac{U'(t)}{1+U(t)^2} \int_0^1 \frac{1}{1+(v_{U(t)}(s))^2}\,\mathrm{d}s - \frac{1}{1+U(t)} \int_0^1 \frac{2 v_{U(t)}(s) d_t v_{U(t)}(s)}{\big(1+(v_{U(t)}(s))^2\big)^2}\,\mathrm{d}s.$$

As in the proof of Lemma 9 we show

$$\left|d_t v_{U(t)}\right| = \left|\frac{dv_{U(t)}}{dU} U'\right| = \mathcal{O}(\epsilon).$$

In combination with the bound $|U'(t)| \leq c\epsilon$, we obtain

$$|U''(t)| = \left| d_t \int_0^1 \epsilon R\big(U(t), v_{U(t)}(s)\big)\,\mathrm{d}s \right| \leq C\epsilon^2.$$

A similar argument yields for the third derivative

$$|U'''(t)| = \left| d_t^2 \int_0^1 \epsilon R\big(U(t), v_{U(t)}(s)\big)\,\mathrm{d}s \right| = \left| \int_0^1 d_t^2 \epsilon R\big(U(t), v_{U(t)}(s)\big)\,\mathrm{d}s \right| \leq C\epsilon^3,$$

where we have used that

$$d_t^2 v_{U(t)} = \frac{dv_{U(t)}}{dU} U'' + \frac{d^2 v_{U(t)}}{dU^2} U'^2 = \mathcal{O}(\epsilon^2),$$

given that $\lambda(U)$ is twice differentiable in U. ∎

LEMMA 17 (local approximation error of the effective equation). *Let $U \in C^3(I)$ be the solution to (24)–(25), $U_{K;k} \in \mathbb{R}^{N+1}$ the approximation given by Algorithm 11. For $T_n = \mathcal{O}(\epsilon^{-1})$, the error $E_n := U(T_n) - U_{n;k}$ is bounded by*

$$|E_n| \leq C\big(\epsilon^2 K^2 + k^2 + tol_P\big)$$

with a constant $C > 0$ that does not depend on ϵ, K, k, and tol_P.

Proof. Combination of Taylor expansions around T_n and T_{n-1} of the continuous solution U gives

$$\begin{aligned} U(T_{n+1}) = U(T_n) &+ \frac{3K}{2} \int_0^1 \epsilon R\big(U(T_n), v_{U(T_n)}(s)\big)\, \mathrm{d}s \\ &- \frac{K}{2} \int_0^1 \epsilon R\big(U(T_{n-1}), v_{U(T_{n-1})}(s)\big)\, \mathrm{d}s + \mathcal{O}(K^3) \max_{\xi \in (T_{n-1}, T_{n+1})} |U'''(\xi)|, \end{aligned}$$

where $\xi \in [T_{n-1}, T_n]$. In combination with (43), we obtain the error representation

$$\begin{aligned} E_{n+1} = E_n &+ \frac{3K}{2} \int_0^1 \epsilon \Big(R\big(U(T_n), v_{U(T_n)}(s)\big) - R\big(U_n, v_{U_n;k}(s)\big) \Big)\, \mathrm{d}s \\ &- \frac{K}{2} \int_0^1 \epsilon \Big(R\big(U(T_{n-1}), v_{U(T_{n-1})}(s)\big) - R\big(U_{n-1}, v_{U_{n-1},k}(s)\big) \Big)\, \mathrm{d}s \\ &+ \mathcal{O}(K^3) \max_{\xi \in (T_{n-1}, T_{n+1})} |U'''(\xi)|. \end{aligned}$$

With the Lipschitz continuity of $R(\cdot)$, Assumption 3, we estimate

$$\begin{aligned} \int_0^1 \Big| R\big(U(T_n), v_{U(T_n)}(s)\big) - R\big(U_n, v_{U_n;k}(s)\big) \Big|\, \mathrm{d}s \\ \leq C \left(\int_0^1 \big| v_{U(T_n)}(s) - v_{U_n;k}(s) \big|\, \mathrm{d}s + \big| U(T_n) - U_n \big| \right). \end{aligned}$$

For the first part, we use the estimate (27) of Lemma 8 and Assumption 14:

$$\begin{aligned} \int_0^1 \big| v_{U(T_n)}(s) - v_{U_n;k}(s) \big|\, \mathrm{d}s &\leq \int_0^1 \big| v_{U(T_n)} - v_{U_n} \big|\, \mathrm{d}s + \int_0^1 \big| v_{U_n} - v_{U_n;k} \big|\, \mathrm{d}s \\ &\leq C_{L8b} C_{A14} \left(|U(T_n) - U_n| + k^2 \|v_{U_n}\|_{C^3([0,1])} + \mathrm{tol}_P \right). \end{aligned}$$

In combination with Lemma 16 this yields

$$|E_{n+1}| \leq |E_n| + C\epsilon K \left(|E_n| + |E_{n-1}| + k^2 + \mathrm{tol}_P + \epsilon^2 K^2 \right)$$

with $C = C(C_{L8b}, C_{A14}, C_{16})$. Summing over $n = 1, \dots, N-1$ and using $E_0 = 0$, we obtain

$$|E_N| \leq |E_1| + C\epsilon T_N\big(k^2 + \mathrm{tol}_P + \epsilon^2 K^2\big) + C\sum_{n=1}^{N-1} \epsilon K |E_n|.$$

The term $|E_1|$ depends on the forward Euler method, which is used to compute $U_{K,1}$:

$$|E_1| \leq CK^2 |U''|_\infty \leq C\epsilon^2 K^2,$$

where we have used Lemma 16 and $|E_0| = 0$. Finally, a discrete Gronwall inequality yields

$$|E_N| \leq C\epsilon T_N \Big(k^2 + \epsilon^2 K^2 + \mathrm{tol}_P\Big) \exp\Big(\epsilon T_N\Big).$$

The postulated result follows for $T_N = \mathcal{O}(\epsilon^{-1})$. ∎

Finally, we can estimate the error between the multiscale algorithm and the solution $u(t)$ to the original coupled problem.

THEOREM 18 (a priori estimate for the multiscale algorithm). *Let $I = [0, T]$ with $T = \mathcal{O}(\epsilon^{-1})$ and let $u \in C(I)$ and U_K be the solutions to the original problem* (1) *and the discrete effective equations* (43)*, respectively. It holds*

$$|u(T_n) - U_n| = C\Big(k^2 + \epsilon^2 K^2 + tol_P + \epsilon\Big),$$

where $C > 0$ does not depend on ϵ, K, k, and tol_P.

Proof. We introduce $\pm U(T_n)$ and estimate

$$|u(T_n) - U_n| \leq |u(T_n) - U(T_n)| + |U(T_n) - U_n|.$$

The two terms on the right-hand side are estimated with Lemmas 10 and 17. ∎

4.3. Approximation of the periodic flow problem. The temporal multiscale schemes are based on periodic solutions $\mathbf{v}_{U_n}(s)$ for $s \in [0, 1]$, where the variable $U_n = U(t_n)$ is fixed such that no feedback between fluid problem and reaction equation takes place within this short interval. A numerical difficulty lies in the determination of the correct initial value $\mathbf{v}_{U_n,0}$ that yields periodicity $\mathbf{v}_{U_n}(0) = \mathbf{v}_{U_n}(1)$. Let us consider again the full Navier–Stokes problem

$$\nabla \cdot \mathbf{v}_{U_n} = 0, \quad \rho(\partial_t \mathbf{v}_{U_n} + (\mathbf{v}_{U_n} \cdot \nabla)\mathbf{v}_{U_n}) - \operatorname{div} \boldsymbol{\sigma}(\mathbf{v}_{U_n}, p_{U_n}) = \mathbf{f}, \quad \mathbf{v}_{U_n}(1) = \mathbf{v}_{U_n}(0) \text{ in } \Omega(U_n). \tag{49}$$

We assume that such a periodic solution of the Navier–Stokes equations exists. Some results are given by Kyed and Galdi [18, 19, 32] that require, however, severe restrictions on the problem data. Depending on the transient dynamics, the decay of the nonstationary solution to this periodic solution can be very slow. It depends basically on $\exp(-\nu\lambda_0)$, where ν is the viscosity and $\lambda_0 > 0$ the smallest eigenvalue of the Stokes operator, which depends on the inverse of the domain size. Several acceleration techniques that are based on shooting methods [21, 53], Newton schemes [46, 30], or gradient-based optimization techniques [23, 43] have been proposed to quickly identify the initial value $\mathbf{v}_{U_n,0}$. We apply a simple acceleration scheme that is based on decomposing the periodic solution into its average and the fluctuations; see [42]. Here, we shortly recapitulate the algorithm that has been introduced in [42].

ALGORITHM 19 (averaging scheme for the identification of periodic solutions). *Given the initial value* $\mathbf{v}_{U_n,0}^{(1)}$, *usually* $\mathbf{v}_{U_n,0}^{(1)} \coloneqq \mathbf{v}_{U_{n-1},0}$ *and let* $tol_P > 0$ *be a given tolerance. Iterate for* $l = 1, 2, \ldots$.

1. *Solve one cycle of* (49) *for* $(\mathbf{v}_{U_n}^{(l)}, p_{U_n}^{(l)})$ *with the initial* $\mathbf{v}_{U_n}^{(l)}(0) = \mathbf{v}_{U_n,0}^{(l)}$.
2. *Compute the error in periodicity*

$$err_P^{(l)} \coloneqq \|\mathbf{v}_{U_n}^{(l)}(1) - \mathbf{v}_{U_n}^{(l)}(0)\|.$$

3. *Stop, if* $err_P^{(l)} < tol_P$.
4. *Compute the velocity average over the cycle*

$$\bar{\mathbf{v}}_{U_n}^{(l)} \coloneqq \int_0^1 \mathbf{v}_{U_n}^{(l)}(s)\, ds.$$

5. *Compute the stationary update problem for* $(\bar{\mathbf{w}}_{U_n}^{(l)}, \bar{q}_{U_n}^{(l)})$*:*

$$\nabla \cdot \bar{\mathbf{w}}_{U_n}^{(l)} = 0, \quad \rho\big((\bar{\mathbf{v}}_{U_n}^{(l)} \cdot \nabla)\bar{\mathbf{w}}_{U_n}^{(l)} + (\bar{\mathbf{w}}_{U_n}^{(l)} \cdot \nabla)\bar{\mathbf{v}}_{U_n}^{(l)}\big) - \operatorname{div} \boldsymbol{\sigma}(\bar{\mathbf{w}}_{U_n}^{(l)}, \bar{q}_{U_n}^{(l)}) = \mathbf{v}_{U_n}^{(l)}(1) - \mathbf{v}_{U_n}^{(l)}(0). \tag{50}$$

6. *Update the initial*

$$\mathbf{v}_{U_n,0}^{(l+1)} \coloneqq \mathbf{v}_{U_n}^{(l)}(1) + \bar{\mathbf{w}}_{U_n}^{(l)}$$

and go to step 1.

The basic idea of introducing the averaged update problem (50) in step 5 of the algorithm is to quickly predict the correct average of the periodic solution. The computational effort for each iteration lies mostly in step 1, where one complete nonstationary cycle of the periodic problem over the period $[0, 1]$ is computed. Given the step size k this means solving k^{-1} time steps of the discrete Navier–Stokes problem. In addition, step 5 calls for the solution of one additional stationary problem.

Using this scheme we are able to reduce the periodicity error to $\mathrm{tol}_P < 10^{-4}$ in less than 5 cycles of the algorithm. In the context of the usual HMM approaches this would correspond to choosing the relaxation time as $\eta = 5\,\mathrm{s}$ in terms of computational effort, i.e., 5 times the period length; see [1].

5. Numerical examples. We consider the full problem described in the introduction, namely, the incompressible Navier–Stokes equations coupled to a scalar ODE model. In order to transfer the proofs from the simplified setting to this more relevant case, several open questions regarding the existence and regularity theory of the Navier–Stokes equations in the periodic setting would have to be addressed. The numerical results presented in this section will, however, reveal convergence rates and error constants that are in full agreement with the theoretical findings for the simplified model problem.

5.1. Configuration of the plaque formation problem. A sketch of the plaque growth problem is given in Figure 1; the governing equations have been outlined in the introduction, section 1. On the fast scale we consider a Navier–Stokes flow in a channel whose width depends on the *slowly evolving variable* $u(t)$. The variable domain describing the channel is given by

$$\begin{aligned} &\Omega(u) = \left\{(x, y) \in \mathbb{R}^2 \,:\, -5\,\mathrm{cm} < x < 10\,\mathrm{cm},\ |y| < (1.5 - u\gamma(x))\,\mathrm{cm}\right\}, \\ &\gamma(x) = \exp\left(-x^2\right). \end{aligned} \tag{51}$$

Instead of a complex growth model for the plaque formation as introduced in [50] we use this explicit dependence of the domain on the scalar $u(t)$. The periodic Navier–Stokes problem is driven by a time periodic Dirichlet condition on the inflow boundary Γ_{in}:

$$\mathbf{v}_{in}(y,t) = 25\sin(\pi t)^2\left(1-\frac{y^2}{1.5^2}\right)\mathrm{cm/s} \quad \text{on } \Gamma_{in}\times[0,T]. \tag{52}$$

On the outflow boundary Γ_{out} we specify the *do-nothing* outflow condition

$$\rho\nu\partial_{\vec{n}}\mathbf{v} - p\vec{n} = 0 \tag{53}$$

that includes a pressure normalizing condition $\int_{\Gamma_{out}} p\,\mathrm{d}s = 0$; see [27]. Kinematic viscosity and density resemble blood and the parameters in the reaction term (2) are tuned to obtain a realistic behavior concerning the different temporal scales of atherosclerotic plaque growth:

$$\rho = 1\,\mathrm{g/cm^3}, \quad \nu = 0.04\,\mathrm{cm^2\cdot s^{-1}}, \quad \sigma_0 = 30. \tag{54}$$

The constant σ_0 is such that the concentration u reaches the value 1 at approximately $T = \mathcal{O}(\epsilon^{-1})$.

We exploit the symmetry of the problem and compute on the upper half of the domain only. On the symmetry boundary at $y = 0$ we prescribe the condition

$$\mathbf{v}\cdot\vec{n} = 0, \quad \boldsymbol{\sigma}(\mathbf{v},p)\vec{n}\cdot\vec{\tau} = 0.$$

Problem (1) can be formulated on a reference domain $\Omega := \Omega(0)$ by means of an *arbitrary Lagrangian Eulerian* approach (see [13] or [41, Chapter 5]), using the map

$$T:\Omega(0)\to\Omega(u), \quad T(u(t);x,y) = \begin{pmatrix} x \\ \frac{1.5-u(t)\gamma(x)}{1.5}y\end{pmatrix}, \quad \gamma(x) = \exp(-x^2), \tag{55}$$

with derivative and determinant given by

$$\mathbf{F} := \hat{\nabla}T = \begin{pmatrix} 1 & 0 \\ -\frac{u\gamma'(x)}{1.5}y & \frac{1.5-u\gamma(x)}{1.5}\end{pmatrix}, \quad J := \det(\mathbf{F}) = \frac{1.5-u\gamma(x)}{1.5}. \tag{56}$$

The Navier–Stokes equations mapped to the reference domain take the form

$$\begin{aligned} \operatorname{div}\left(J\mathbf{F}^{-1}\mathbf{v}\right) &= 0, \\ \rho J\left(\partial_t\mathbf{v} + (\mathbf{F}^{-1}(\mathbf{v}-\partial_t T)\cdot\nabla)\mathbf{v}\right) - \operatorname{div}\left(J\mathbf{F}^{-T}\hat{\boldsymbol{\sigma}}(\mathbf{v},p)\mathbf{F}^{-1}\right) &= 0 \quad \text{in } \Omega, \\ \mathbf{v}(0) = \mathbf{v}_0, \qquad \hat{\boldsymbol{\sigma}}(\mathbf{v},p) &:= -pI + \rho_f\nu_f\nabla\mathbf{v}\mathbf{F}^{-1}. \end{aligned} \tag{57}$$

Formulations (57) and (1) are equivalent as long as $J > 0$, which we can guarantee if we limit the maximum deformation by $u_{\max} := 1$. The resulting Reynolds number

$$Re = \frac{\bar{\mathbf{v}}L}{\nu}$$

with the channel diameter $L = 3\,\mathrm{cm}$, the kinematic viscosity $\nu = 0.04\,\mathrm{cm^2\cdot s^{-1}}$, and the flow rate $\bar{v} = (3-2U)^{-1}\mathrm{cm\cdot s^{-1}}$ in the remaining gap of width $3-2U(t)$ is in

the range of 0 and about 3 750 as long as $U \leq u_{\max} = 1$. Such high values are only reached at peak inflow; compare (52).

The correct model and a full comprehension of shear effects on plaque formation and growth are still under active discussion. It is, however, understood that regions of (relatively) low shear stress that exhibit an oscillatory character are more prone to plaque growth [47, 12]. The reaction term (2) mimics this behavior. Its dependence on the flow problem by means of the wall shear stress is nonlinear and cannot be considered by a simple averaging as done in [49, 50, 51].

5.2. Significance of the model problem and application to the plaque formation model. The analysis of the temporal multiscale scheme was based on Assumptions 3, 5, 7, and and 14. Further, we have used the linearity of the model problem and the availability of analytical solutions to the ODEs appearing in the model problem. Here, we will shortly motivate the relevance of the simplified model problem and discuss the application of the multiscale scheme to the full plaque formation model.

If we linearize the Navier–Stokes equations (57) by omitting the convective terms $((\mathbf{v} - \partial_t T) \cdot \nabla)\mathbf{v}$, we obtain the Stokes equations on $\Omega(u)$. These have a system of L^2-orthonormal eigenfunctions with eigenvalues $0 < \lambda_0(u) \leq \lambda_1(u) \leq \cdots$. As long as the domain does not deteriorate, i.e., for $0 \leq u \leq u_{\max}$ it holds that $\lambda_0(u) \geq \lambda_0 > 0$. Due to the regularity of the reference map (55), the mapping of the equations to the reference domain is also differentiable. Its derivative (23) is bounded as long as contact with the boundary walls is prevented, i.e., as long as $u \leq u_{\max}$ is bound away from 1.5; compare (51) and (55) (Assumption 7). By diagonalization of the Stokes problem with respect to the system of eigenfunctions we reduce the problem to a system of ODEs of type (3b). Lemma 8 can be applied to each component of the diagonalized system. However, since the eigenvalues of the Stokes operator are not bounded, a formal extension to the full Stokes problem requires further steps.

The essential assumptions for the application of the multiscale method is the boundedness and the Lipschitz continuity of the reaction term $R(\cdot, \cdot)$ with respect to slow and fast variables (Assumption 3) as well as the existence of time-periodic solutions to the isolated fast-scale problem (Assumption 5). Given a fixed value of u, the fast-scale problem (13) is given by the Navier–Stokes equations on the domain $\Omega(u)$. The unique existence of periodic solutions to the Navier–Stokes equations is only guaranteed for small problem data; see [18, 19, 32]. These results will most likely not apply to the higher Reynolds number regime of typical blood flow configurations, such that Assumption 5 cannot be verified in our setting. However, given a periodic solution, since $\mathbf{f} = 0$, the Dirichlet data $\mathbf{v}_{in}$ are smooth and since the domain allows for a piecewise C^∞ parametrization with a finite number of convex corners, we expect the regularity

$$\sup \left(\|\mathbf{v}(t)\|_{H^2(\Omega)} + \|p(t)\|_{H^1(\Omega)} \right) \leq C_{58}; \tag{58}$$

see [25]. Under this assumption we can show Lipschitz continuity and boundedness of the reaction term.

LEMMA 20 (Lipschitz continuity). *Let $u \in C(I)$ with $0 \leq u(t) \leq u_{\max}$. Assume that for fixed u, the time-periodic Navier–Stokes problem allows for a unique solution $\mathbf{v}_u$ satisfying* (58) *with $C_{58} = C(u_{\max})$. Then, the reaction term* (2) *is bounded,*

$$|R(u, \mathbf{v}_u)| \leq 1, \tag{59}$$

and Lipschitz continuous with respect to both arguments,

$$|R(u,\mathbf{v}) - R(\eta,\mathbf{v})| \le |u-\eta| \qquad \forall u,\eta \in [0,u_{\max}] \quad \forall \mathbf{v} \in H^2(\Omega), \tag{60a}$$

$$|R(u,\mathbf{v}) - R(u,\mathbf{u})| \le L|\mathbf{u}-\mathbf{v}| \qquad \forall \mathbf{v},\mathbf{u} \in H^2(\Omega) \quad \forall u \in [0,u_{\max}], \tag{60b}$$

with a constant $L > 0$.

Proof. Given (58), the wall shear stress is well defined:

$$|\sigma_{WSS}(\mathbf{v})| \le 2\rho\nu\sigma_0^{-1} c_{tr} \|\mathbf{v}\|_{H^2(\Omega)}, \tag{61}$$

where c_{tr} is the constant of the trace inequality $\|\nabla \mathbf{v}\|_\Gamma \le c_{tr}\|\mathbf{v}\|_{H^2(\Omega)}$. Then (59) directly follows by the construction of $R(\cdot,\cdot)$; see (2). Further, it holds that

$$\big|R(u,\mathbf{v}) - R(\eta,\mathbf{v})\big| = \big(1+|\sigma_{WSS}(\mathbf{v})|^2\big)^{-1}\big(1+u\big)^{-1}\big(1+\eta\big)^{-1}|u-\eta|,$$

which shows (60a). Likewise

$$\begin{aligned}\big|R(u,\mathbf{v}) - R(u,\mathbf{u})\big| &\\ = \big(1+u\big)^{-1} &\frac{\big|\sigma_{WSS}(\mathbf{v}) + \sigma_{WSS}(\mathbf{u})\big|}{\big(1+|\sigma_{WSS}(\mathbf{v})|^2\big)\big(1+|\sigma_{WSS}(\mathbf{u})|^2\big)}\big|\sigma_{WSS}(\mathbf{v}) - \sigma_{WSS}(\mathbf{u})\big|.\end{aligned}$$

Since the wall shear stress is a linear functional $\sigma_{WSS}(\mathbf{v}) + \sigma_{WSS}(\mathbf{u}) = \sigma_{WSS}(\mathbf{v}+\mathbf{u})$ and due to the relation $2\sigma_{WSS}(\mathbf{v}) \le 1 + \sigma_{WSS}(\mathbf{v})^2$, we estimate

$$\big|R(u,\mathbf{v}) - R(u,\mathbf{u})\big| \le \|\mathbf{v}-\mathbf{u}\|_{H^2(\Omega)};$$

see (61). ∎

Finally, the validity of Assumption 14 has been discussed in Remark 15.

5.3. Discretization. We briefly sketch the discretization in space and time. All numerical experiments have been realized in the software library Gascoigne 3D [5]. We use uniform time steps k and K on both scales and the time-stepping schemes presented in (43)–(45).

For spatial discretization we triangulate the reference domain $\hat{\Omega}$ into open quadrilaterals, allowing for local refinement based on hanging nodes; see [40] for details on the realization in the software Gascoigne 3D. Equal-order biquadratic finite elements are used for velocity and pressure degrees of freedom. Pressure stabilization is accomplished with the local projection stabilization scheme [4]. Stabilization of the convective terms is not required due to the moderate Reynolds numbers.

Direct simulation. The PDE/ODE system is a multiscale coupled problem. We will compare the presented multiscale scheme with a direct forward simulation. As we do not expect any stiffness-related problems in the ODE we decouple the PDE/ODE system by an implicit/explicit approach where, as in the multiscale approach, the discretization of the Navier–Stokes equation (57) is based on the second-order Crank–Nicolson scheme and the discretization of the ODE on the second-order explicit Adams–Bashforth formula resulting in a multiscale method that splits naturally into one explicit ODE step and an implicit Navier–Stokes step.

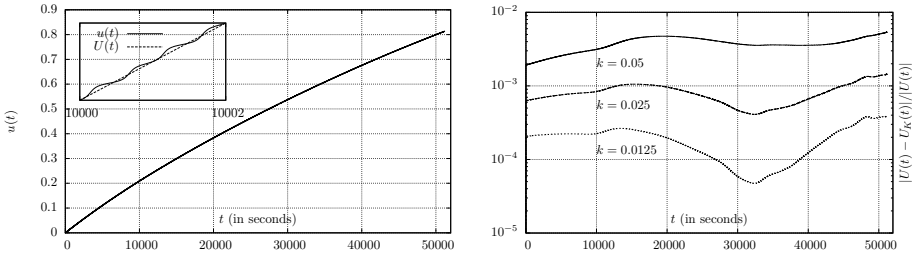


FIG. 3. Left*: Evolution of the concentration variable $U(t)$ as function over time (forward simulation with $k = 0.05\,\mathrm{s}$). In the small subplot we show both $U(t)$ and the resolved variable $u(t)$. The maximum deviation is bound by* $\max|u(t) - U(t)| \leq 3 \cdot 10^6$. Right*: Relative error in $U(t)$ under refinement of the time step k (compared to extrapolated values).*

5.4. Numerical analysis of the plaque formation problem.

5.4.1. Configuration with resolvable time scales. In a first test we take the value $\epsilon = 5 \cdot 10^{-5}$ in (54). By this choice, the concentration u reaches approximately 1 after about $50\,000\,\mathrm{s} \approx 1\,\mathrm{day}$ such that we can still resolve the coupled problem in all temporal scales (although the direct simulation still takes a substantial effort). To keep the computational effort within bounds we use a rather coarse spatial discretization with 320 elements, resulting in 4 131 unknowns of a biquadratic equal-order discretization for velocities and pressure.

In Figure 3 (left) we give an overview of the temporal evolution of the concentration variable $u_k(t)$ using a full resolution of the fast scale. The simulations break down at $T \approx 55\,000\,\mathrm{s}$ due to the deterioration of the ALE map and the high Reynolds number. For the small interval $[10\,000\,\mathrm{s}, 10\,002\,\mathrm{s}]$ we show a close-up view of the resolved solution $u_k(t)$ and the averaged value $\int_t^{t+1} u_k(s)\,\mathrm{d}s$. The deviation is bound by $3{\cdot}10^{-6} = \mathcal{O}(\epsilon)$, in agreement with Lemma 10. We determine reference values $u_{ref}(T_n)$ by extrapolating numerical results for $k = 0.05\,\mathrm{s}$, $k = 0.025\,\mathrm{s}$, and $k = 0.0125\,\mathrm{s}$. The relative errors $|u_k(t_n) - u_{ref}(t_n)|/|u_{ref}(t_n)|$ (based on these extrapolated errors) are given in Figure 3 (right). The convergence rate in terms of k is approximately quadratic. Further, there is no significant accumulation of simulation errors over time.

By the evolution of the concentration u over time, the computational domain undergoes substantial deformations with a strong narrowing of the flow domain. In Figure 4 we show snapshots of the solution at different time steps, $t \approx 6\,250\,\mathrm{s}, 13\,500\,\mathrm{s}, 18\,750\,\mathrm{s}, \ldots, 50\,000\,\mathrm{s}$. The narrowing of the gap causes an acceleration of the fluid resulting in a higher Reynolds number flow with a substantial variation in the feedback functional $R(u, \mathbf{v})$ which depends on the wall shear stress.

Multiscale approach. Next, in Figure 5 we show the results obtained with the multiscale method for this relaxed problem with $\epsilon = 5 \cdot 10^{-5}$. The tolerance for approximating the periodic flow problems is set to

$$\|\mathbf{v}_U(1) - \mathbf{v}_U(0)\|^2_{L^2(\Omega)} + \|p_U(1) - p_U(0)\|^2_{L^2(\Omega)} < \mathrm{tol}_P^2 := 10^{-8}.$$

For each of the three short time step sizes $k = 0.05\,\mathrm{s}$, $k = 0.025\,\mathrm{s}$ and $k = 0.0125\,\mathrm{s}$ we use long time step sizes ranging from $K = 6\,400\,\mathrm{s}$ to $K = 400\,\mathrm{s}$. In the left plot we compare the solutions for different values of the long time step size K. In this (nonlogarithmic) plot we see convergence of the results to the corresponding resolved simulation with the same short step size $k = 0.025\,\mathrm{s}$. The lower plot shows the corresponding results for a variation of the small step size k, while the long-scale

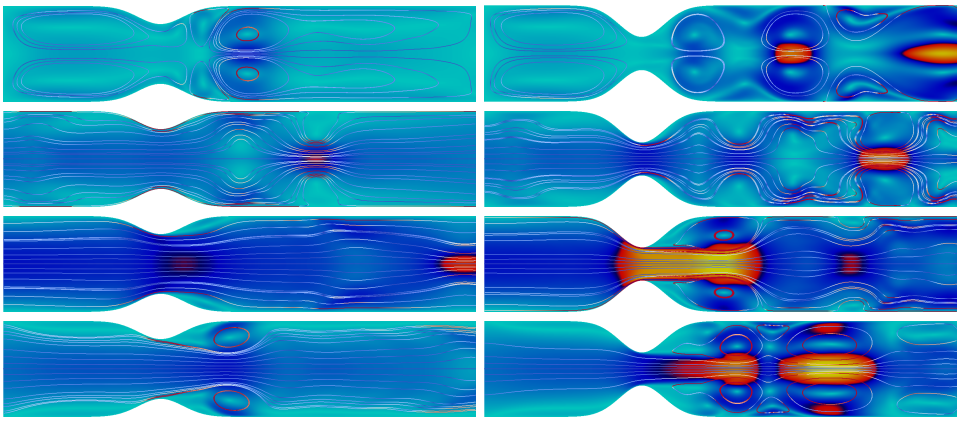


FIG. 4. *Velocity magnitude at times* $t = n \cdot 6\,250\,\mathrm{s}$ *for* $n = 1, 2, \ldots, 8$ *on domains with different growth. As the inflow profile is periodic, the narrowing of the domain causes a significant change of the flow pattern.*

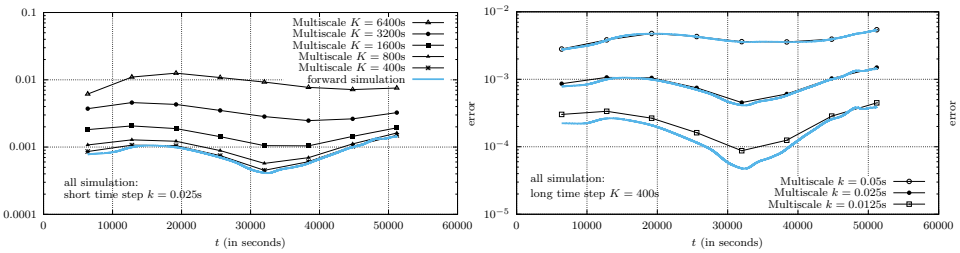


FIG. 5. *Convergence of the temporal multiscale method for the relaxed problem* $\epsilon = 5 \cdot 10^{-5}$. Left*: Effect of the long time step* K *using the small time step* $k = 0.025\,\mathrm{s}$. Right*: Effect of the short time step* k *using the long time step size* $K = 400\,\mathrm{s}$. *For comparison we plot the error of the fully resolved simulation using these time step sizes.*

step size is fixed to $K = 400\,\mathrm{s}$. For comparison we show the results obtained with the resolved simulation for these small step sizes. Again we see convergence of the multiscale scheme towards the resolved scheme.

The effect of the small step size k is dominant. This is highlighted by a closer analysis of the convergence at time $t = 51\,200\,\mathrm{s}$, the results being shown in Table 1. We indicate the concentration $U(t)$ and the errors for the different multiscale approaches as well as for the resolved forward simulation. We fit all these values to the postulated relation

$$U(k, K) = U + C_k k^{q_k} + C_K K^{q_K} \tag{62}$$

to get a better understanding of the convergence rates. We estimate all parameters u, C_k, C_K, q_k, q_K (obtained with gnuplot fit [48]) and find

$$U(k, K) = U - 1.12 \cdot k^{1.85} - 6.61 \cdot 10^{-10} \cdot K^{1.80};$$

see also Table 1. Convergence is close to the expected second order, both in k and K. The most striking result is the good estimation of the error constant that shows the proper scaling in ϵ^2. This result is in good correspondence with the error estimate derived in Theorem 18 where the constant in front of the K^2-term depends on ϵ^2. Balanced discretization errors are given for $\epsilon^2 K^2 \approx k^2$, i.e., for $K \approx \epsilon^{-1} k$.

TABLE 1

Convergence of the multiscale method at time $T = 51\,200$ s. We show the values of U_K and the error (w.r.t. the extrapolation in $k \to 0$ and $K \to 0$). We compare the results of the multiscale method with the fully resolved forward computation. Finally, we fit the numerical results to the expected convergence behavior.

	U_K			Error (w.r.t. extrapolation)		
k	0.05 s	0.025 s	0.0125 s	0.05 s	0.025 s	0.0125 s
$K = 6\,400$ s	0.8089225	0.8118418	0.8126760	$1.11 \cdot 10^{-2}$	$7.55 \cdot 10^{-3}$	$6.53 \cdot 10^{-3}$
$K = 3\,200$ s	0.8122441	0.8153539	0.8162696	$7.05 \cdot 10^{-3}$	$3.25 \cdot 10^{-3}$	$2.13 \cdot 10^{-3}$
$K = 1\,600$ s	0.8132730	0.8164325	0.8173319	$5.80 \cdot 10^{-3}$	$1.93 \cdot 10^{-3}$	$8.35 \cdot 10^{-4}$
$K = 800$ s	0.8135139	0.8166886	0.8175426	$5.50 \cdot 10^{-3}$	$1.62 \cdot 10^{-3}$	$5.77 \cdot 10^{-4}$
$K = 400$ s	0.8135782	0.8167926	0.8176490	$5.42 \cdot 10^{-3}$	$1.49 \cdot 10^{-3}$	$4.47 \cdot 10^{-4}$
Resolved	0.8135999	0.8168226	0.8176928	$5.42 \cdot 10^{-3}$	$1.46 \cdot 10^{-3}$	$3.93 \cdot 10^{-4}$

$$\text{Fit to } U(k,K) = U + C_k k^{q_k} + C_K K^{q_K}, \qquad U = 0.818006 \pm 10^{-3}\%,$$
$$C_k = -1.12 \pm 17\%, \quad C_k = -6.61 \cdot 10^{-10} \pm 34\%, \quad q_k = 1.85 \pm 3.39\%, \quad q_K = 1.80 \pm 2.14\%.$$

Based on the time step relation we can compute the possible speedup of the multiscale approach which we measure in the overall number of Navier–Stokes time steps to be performed. The forward algorithm requires $E_{fwd} = \frac{T}{k}$ solution steps, while the multiscale approach has an effort of $E_{ms} = {}^{T}/_{K} \cdot n_{period} {}^{1}/_{k} = {}^{T n_{period}}/_{kK}$ steps, where n_{period} is the number of cycles that are necessary to compute a periodic solution. Given $K \approx \epsilon^{-1} k$ we approximate $E_{ms} \approx {}^{\epsilon T n_{period}}/_{k^2}$, and the speedup is estimated by

$$\frac{E_{fwd}}{E_{ms}} = \frac{k}{\epsilon n_{period}}.$$

In our numerical example we identify $n_{period} \leq 5$ and with $\epsilon = 5 \cdot 10^{-5}$ we expect a speedup of $4\,000k$. In Figure 6 we plot the error over the required number of Navier–Stokes time steps. By circles we indicate the multiscale results with a balanced error contribution, which we define as the state where the error of the multiscale approach is within 10% of the error of a fully resolved simulation for the same k. We observe speedups of 1:250 for $k = 0.05$ s, 1:180 for $k = 0.025$ s, and 1:90 for $k = 0.0125$ s, slightly better values than the predicted ones based on $4\,000k$. The overall computational time for the forward simulation with $k = 0.0125$ s was about 13 days, while the multiscale simulation with $k = 0.0125$ s and $K = 800$ s, giving a comparable accuracy, was about 45 min.

5.4.2. Configuration with realistic time scales. Finally, we consider the coupled problem with the time scale parameter $\epsilon = 10^{-6}$, which is close to the temporal dynamics of atherosclerotic plaque growth and 50 times smaller than in the first example. Here, a resolved forward simulation is not feasible. The concentration $u(t)$ will reach a value of approximately 0.8 at $T \approx 2.5 \cdot 10^6$ s ≈ 30 days.

Assuming the validity of estimate (62) and in addition that $C_K \approx \epsilon^2$ we expect balanced error contributions for $K \approx \epsilon^{-1} k = 10^6 k$. The character of the short-scale problem does not depend on ϵ. Hence we consider again the step sizes $k = 0.05$ s, $k = 0.025$ s, and $k = 0.0125$ s. The large time step, however, can be significantly increased. We present results for $T \approx 21$ days in Table 2. For this second example, we vary also the mesh size h to discuss the impact of all relevant discretization parameters. While a smaller value of ϵ makes the time scale challenge more severe, the multiscale approach will profit, as the potential speedup will benefit from the relation $K \approx \epsilon^{-1} k$.

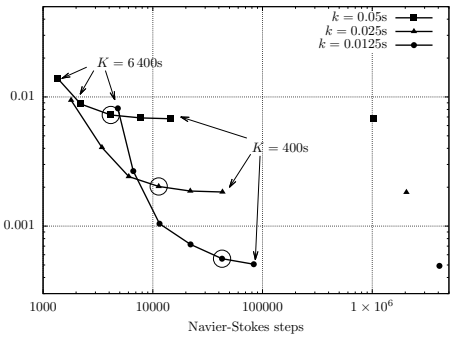


FIG. 6. *Computational effort (measured in Navier–Stokes time steps) for the multiscale approach (lines) and the resolved forward simulation (points). We use three small time steps k from* 0.05 s *down to* 0.0125 s *and vary the long time step K from* 6 400 s *to* 400 s. *Circles indicate multiscale solutions of a quality comparable to the resolved forward simulation (at most* 10% *additional error). The computational time for one Navier–Stokes step is about* 0.2 s *(Core i7-7700,* 3.60 *GHz,* 1370 *spatial unknowns, biquadratic finite elements).*

TABLE 2

Convergence of the multiscale approach for $\epsilon = 10^{-6}$. On three mesh levels we indicate the errors in the concentration $U(T)$ at $T = 1\,843\,200\,\mathrm{s} \approx 21\,\mathrm{days}$. In each block, the errors are given w.r.t. the extrapolation $k, K \to 0$. In the last line we indicate the (dominating) spatial error for each block.

	$h = 0.16\,\mathrm{cm}$			$h = 0.08\,\mathrm{cm}$			$h = 0.04\,\mathrm{cm}$		
	$k = 0.05\,\mathrm{s}$	$k = 0.025\,\mathrm{s}$	$k = 0.0125\,\mathrm{s}$	$k = 0.05\,\mathrm{s}$	$k = 0.025\,\mathrm{s}$	$k = 0.0125\,\mathrm{s}$	$k = 0.05\,\mathrm{s}$	$k = 0.025\,\mathrm{s}$	$k = 0.0125\,\mathrm{s}$
$K = 204\,800\,\mathrm{s}$	$4.45 \cdot 10^{-3}$	$2.46 \cdot 10^{-3}$	$2.20 \cdot 10^{-3}$	$4.06 \cdot 10^{-3}$	$2.42 \cdot 10^{-3}$	$2.04 \cdot 10^{-3}$	$4.06 \cdot 10^{-3}$	$2.42 \cdot 10^{-3}$	$2.04 \cdot 10^{-3}$
$K = 102\,400\,\mathrm{s}$	$2.86 \cdot 10^{-3}$	$8.94 \cdot 10^{-4}$	$6.17 \cdot 10^{-4}$	$2.75 \cdot 10^{-3}$	$1.06 \cdot 10^{-3}$	$6.28 \cdot 10^{-4}$	$2.75 \cdot 10^{-3}$	$1.06 \cdot 10^{-3}$	$6.28 \cdot 10^{-4}$
$K = 51\,200\,\mathrm{s}$	$2.43 \cdot 10^{-3}$	$4.76 \cdot 10^{-4}$	$1.96 \cdot 10^{-4}$	$2.39 \cdot 10^{-3}$	$7.04 \cdot 10^{-4}$	$2.58 \cdot 10^{-4}$	$2.39 \cdot 10^{-3}$	$7.04 \cdot 10^{-4}$	$2.58 \cdot 10^{-4}$
Extrapolated $(k, K \to 0)$	$\lvert U(T) - U_h(T)\rvert \approx 4.55 \cdot 10^{-2}$			$\lvert U(T) - U_h(T)\rvert \approx 8.61 \cdot 10^{-3}$			$\lvert U(T) - U_h(T)\rvert \approx 1.63 \cdot 10^{-3}$		

Combining all 27 computations based on three values for h, k, and K we find the relation

$$U(h, k, K) \approx 0.59076 + 7.6h^{2.4} - 1.7k^{2.2} - 0.04\epsilon^2 K^{1.9},$$

which shows approximately second-order convergence in both time step sizes and the mesh size and also the proper scaling of the constants in the $\mathcal{O}(k^2)$ and $\mathcal{O}(K^2)$ terms. Spatial and temporal errors show a different sign which is also seen in Figure 7, where we plot the errors for all computations. In the right sketch of this figure we compare the computational times of the multiscale approach with a hypothetical resolved simulation. Here, the errors are predicted by extrapolation. The computational times are based on the number of Navier–Stokes steps, namely, $k^{-1}T$ and the average computational time for each Navier–Stokes step, which is 0.135 s on the $h = 0.16\,\mathrm{cm}$ mesh, 0.62 s for $h = 0.08\,\mathrm{cm}$, and 2.3 s for $h = 0.04\,\mathrm{cm}$. The results are very similar to those shown in Figure 6 for the first example. The best multiscale results are close to the hypothetical resolved results. Here, however, the savings are substantially larger with 10 minutes versus 2 months (factor 1:8 000) for $h = 0.16\,\mathrm{cm}$, 1 hour versus nearly 2 years for $h = 0.08\,\mathrm{cm}$ (factor 1:12 000), and 15 hours versus more than 10 years for $h = 0.04\,\mathrm{cm}$ (factor 1:6 000).

Finally, we also evaluate the effect of the parameter tol_P used to control the periodicity of the Navier–Stokes solution; compare Theorem 18. In Table 3 we show the errors at $T = 1\,843\,200\,\mathrm{s} \approx 1\,\mathrm{month}$ for computations based on $K = 25\,600\,\mathrm{s} \approx 7\,\mathrm{h}$, $k = 0.0125\,\mathrm{s}$, and $h = 0.08\,\mathrm{cm}$. The effect of tol_P is very small.

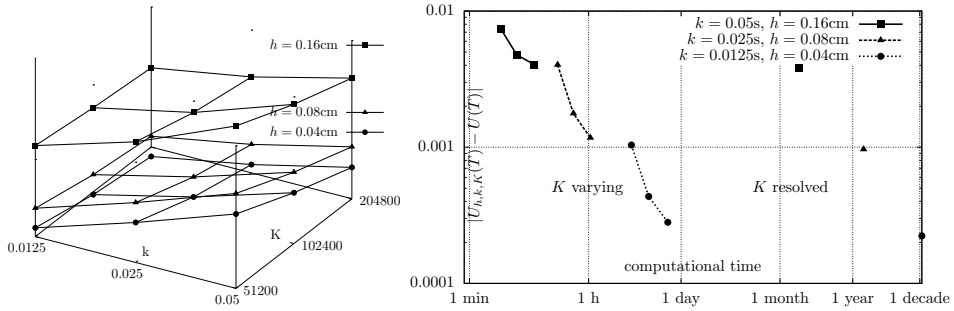


FIG. 7. Left*: Error of the multiscale method under refinement in h, k, and K versus the extrapolated reference value.* Right*: Comparison of the computational times of the multiscale method with the corresponding results for a resolved forward simulation. These results are based on an extrapolation of the error and a prediction of the computational times by multiplying the number of required time steps with the average computing times for each step.*

TABLE 3

Impact of the periodicity parameter tol_P on the error in concentration U in $T = 1\,843\,200\,\mathrm{s}$. Computed w.r.t. $\mathrm{tol}_P = 10^{-8}$. The discretization is chosen as $h = 0.08\,\mathrm{cm}$, $k = 0.0125\,\mathrm{s}$, and $K = 25\,600\,\mathrm{s}$.

tol_P	10^{-1}	10^{-2}	10^{-3}	10^{-4}
$\left\lvert U_{h,k,K}(T)_{\lvert_{\mathrm{tol}_P}} - U_{h,k,K}(T)_{\lvert_{\mathrm{tol}_P=10^{-8}}} \right\rvert$	$1.99 \cdot 10^{-5}$	$4.88 \cdot 10^{-7}$	$3.19 \cdot 10^{-7}$	$1.22 \cdot 10^{-7}$

6. Conclusion. We have presented a framework for the simulation of temporal multiscale problems, where we are interested in the evolution of a slow variable which depends on an oscillating fast variable. The numerical schemes are designed for models that are given by PDEs. The most important assumption is a local (in time) proximity of the fast-scale variable to the solution of a periodic problem. An effective scheme for the slow variable is derived by replacing the fast variable with the periodic solution which can be computed locally, as no initial values must be transferred. The only overhead of the multiscale scheme comes from the identification of initial values required for approximating the periodic problems. Nevertheless, we gain huge speedups compared to a simulation with resolved time scales. The efficiency of the multiscale approach increases when the time scale separation gets larger.

The resulting scheme depends on several numerical parameters, small and large time steps k and K, and the spatial mesh size h. It remains a topic for a future work to design an automatic and adaptive algorithm to control all these parameters in order to balance all contributing error terms.

REFERENCES

[1] A. ABDULLE, W. E, B. ENGQUIST, AND E. VANDEN-EIJNDEN, *The heterogeneous multiscale method*, Acta Numer., 21 (2012), pp. 1–87.

[2] A. AMMAR, F. CHINESTA, E. CUETO, AND M. DOBLARÉ, *Proper generalized decomposition of time-multiscale models*, Internat. J. Numer. Methods Engrg., 90 (2012), pp. 569–596.

[3] D. AUBRY AND G. PUEL, *Two-timescale homogenization method for the modeling of material fatigue*, in World Congress on Computational Mechanics, IOP Conf. Ser.: Mater. Sci. Eng., 10, IOP, Bristol, England, 2010, 012113.

[4] R. BECKER AND M. BRAACK, *A finite element pressure gradient stabilization for the Stokes equations based on local projections*, Calcolo, 38 (2001), pp. 173–199.

[5] R. BECKER, M. BRAACK, D. MEIDNER, T. RICHTER, AND B. VEXLER, *The finite element toolkit* GASCOIGNE, https://www.uni-kiel.de/gascoigne/.

[6] C.L. Bottasso, *Multiscale temporal integration*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 2815–2830.
[7] P. Chakraborty and S. Ghosh, *Accelerating cyclic plasticity simulations using an adaptive wavelet transformation based multitime scaling method*, Internat. J. Numer. Methods Engrg., 93 (2013), pp. 1425–1454.
[8] C.X. Chen, Y. Ding, and J.A. Gear, *Numerical simulation of atherosclerotic plaque growth using two-way fluid-structural interaction*, ANZIAM J., 53 (2012), pp. 278–291.
[9] V.V. Chepyzhov, V. Pata, and M.I. Vishik, *Averaging of 2D Navier–Stokes equations with singularly oscillating forces*, Nonlinearity, 22 (2008), pp. 351–370.
[10] D. Cioranescu and P. Donato, *An Introduction to Homogenization*, Oxford Lecture Ser. Math. Appl., Oxford University Press, Oxford, 1999.
[11] R.D. Crouch and C. Oskay, *Accelerated time integrator for multiple time scale homogenization*, Internat. J. Numer. Methods Engrg., 101 (2015), pp. 1019–1042.
[12] S.S. Dhawan, R.P.A. Nanjundappa, J.R. Branch, W.R. Taylor, A.A. Quyyumi, H. Jo, M.C. McDaniel, J. Suo, D. Giddens, and H. Samady, *Shear stress and plaque development*, Expert Rev. Cardiovasc. Ther., 8 (2010), pp. 545–556.
[13] J. Donea, *An arbitrary Lagrangian-Eulerian finite element method for transient dynamic fluid-structure interactions*, Comput. Methods Appl. Mech. Engrg., 33 (1982), pp. 689–723.
[14] W. E, *Principles of Multiscale Modeling*, Cambridge University Press, Cambridge, 2011.
[15] W. E and B. Engquist, *The heterogenous multiscale method*, Commun. Math. Sci., 1 (2003), pp. 87–132.
[16] B. Engquist and Y.-H. Tsai, *Heterogeneous multiscale methods for stiff ordinary differential equations*, Math. Comp., 74 (2005), pp. 1707–1742.
[17] S. Frei, T. Richter, and T. Wick, *Long-term simulation of large deformation, mechano-chemical fluid-structure interactions in ALE and fully Eulerian coordinates*, J. Comput. Phys., 321 (2016), pp. 874–891.
[18] G.P. Galdi and M. Kyed, *Time-Periodic Solutions to the Navier-Stokes Equations in Three-Dimensional Whole-Space with a Non-Zero Drift Term: Asymptotic Profile at Spatial Infinity*, preprint arXiv:1610.00677v1, 2016.
[19] G.P. Galdi and M. Kyed, *Time-periodic solutions to the Navier-Stokes equations*, in Handbook of Mathematical Analysis in Mechanics of Viscous Fluids, Y. Giga and A. Novotny, eds., Springer, Cham, Switzerland, 2016, pp. 1–70.
[20] M. Gander and L. Halpern, *Techniques for locally adaptive time stepping developed over the last two decades*, in Domain Decomposition Methods in Science and Engineering XX, Lect. Notes Comput. Sci. Eng. 91, Springer, Berlin, 2013, pp. 377–385.
[21] S. Govindjee, T. Potter, and J. Wilkening, *Cyclic steady states of treaded rolling bodies*, Internat. J. Numer. Methods Engrg., 99 (2014), pp. 203–220.
[22] T. Guennouni, *Sur une méthode de calcul de structures soumises à des chargements cycliques: l'Homogénéisation en temps*, ESAIM Math. Model. Numer. Anal., 22 (1988), pp. 417–455.
[23] F.M. Hante, M.S. Mommer, and A. Potschka, *Newton–Picard preconditioners for time-periodic, parabolic optimal control problems*, SIAM J. Numer. Anal., 53 (2015), pp. 2206–2225.
[24] S. Haouala and I. Doghri, *Modeling and algorithms for two-scale time homogenization of viscoelastic-viscoplastic solids under large numbers of cycles*, Internat. J. Plasticity, 70 (2015), pp. 98–125.
[25] J. Heywood and R. Rannacher, *Finite element approximation of the nonstationary Navier-Stokes problem.* I. *Regularity of solutions and second order error estimation for spatial discretizations*, IAM J. Numer. Anal., 19 (1982), pp. 275–311.
[26] J.G. Heywood and R. Rannacher, *Finite element approximation of the nonstationary Navier–Stokes problem. Part* IV*: Error analysis for second-order time discretization*, SIAM J. Numer. Anal., 27 (1990), pp. 353–384.
[27] J.G. Heywood, R. Rannacher, and S. Turek, *Artificial boundaries and flux and pressure conditions for the incompressible Navier-Stokes equations*, Internat. J. Numer. Methods Fluids., 22 (1992), pp. 325–352.
[28] T.J.R. Hughes and J.R. Stewart, *A space-time formulation for multiscale phenomena*, J. Comput. Appl. Math., 74 (1996), pp. 217–229.
[29] A.A. Ilyin, *Global averaging of dissipative dynamical systems*, Rendiconti Academia Nazionale delle Scidetta dli XL. Memorie di Matematica e Applicazioni, 116 (1998), pp. 165–191.
[30] L. Jiang, L.T. Biegler, and V.G. Fox, *Simulation and optimization of pressure-swing adsorption systems for air separation*, AIChE J., 49 (2003), pp. 1140–1157.
[31] R.A. Khurram and A. Masud, *A multiscale/stabilized formulation of the incompressible Navier–Stokes equations for moving boundary flows and fluid–structure interaction*, Comput. Mech., 38 (2006), pp. 403–416.

[32] M. Kyed, *Time-Periodic Solutions to the Navier-Stokes Equations*, Habilitationsschrift Technische Universität Darmstadt, Darmstadt, Germany, 2012.

[33] J. Lemaitre and I. Doghri, *Damage* 90: *A post processor for crack initiation*, Comput. Methods Appl. Mech. Engrg., 115 (1994), pp. 197–232.

[34] V. Levenshtam, *Justification of the averaging method for a system of equations with the Navier–Stokes operator in the principal part*, St. Petersburg Math. J., 26 (2015), pp. 69–90.

[35] Y. Maday and G. Turinici, *The parareal in time iterative solver: A further direction to parallel implementation*, in Domain Decomposition Methods in Science and Engineering, Springer, Berlin, 2005, pp. 441–448.

[36] A. Masud and R.A. Khurram, *A multiscale finite element method for the incompressible Navier–Stokes equations*, Comput. Methods Appl. Mech. Engrg., 195 (2006), pp. 1750–1777.

[37] J. Mizerski and T. Richter, *The candy wrapper problem - a temporal multiscale approach for PDE/PDE systems*, in Numerical Mathematics and Advanced Applications - Enumath 2019, Lect. Notes Comput. Sci. Eng., Springer, Cham, Switzerland, to appear, https://arxiv.org/abs/2002.09705.

[38] O.A. Oleinik, A.S. Shamaev, and G.A. Yosifian, *Mathematical Problems in Elasticity and Homogenization*, Stud. Math. Appl. 26, North-Holland, Amsterdam, 1992.

[39] J.-C. Passieux, P. Ladevèze, and D. Néron, *A scalable time–space multiscale domain decomposition method: Adaptive time scale separation*, Comput. Mech., 46 (2010), pp. 621–633.

[40] T. Richter, *Goal oriented error estimation for fluid-structure interaction problems*, Comput. Methods Appl. Mech. Eng., 223-224 (2012), pp. 28–42.

[41] T. Richter, *Fluid-Structure Interactions. Models, Analysis and Finite Elements*, Lect. Notes Comput. Sci. Eng. 118, Springer, Cham, Switzerland, 2017.

[42] T. Richter, *An Averaging Scheme for the Efficient Approximation of Time-Periodic Flow Problems*, https://arxiv.org/abs/1806.00906, 2019.

[43] T. Richter and W. Wollner, *Efficient computation of time-periodic solutions of partial differential equations*, Vietnam J. Math., 46 (2018), pp. 949–966.

[44] J.A. Sanders, F. Verhulst, and J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, Appl. Math. Sci. 59, Springer, New York, 2007.

[45] F. Sonner and T. Richter, *Second order pressure estimates for the Crank–Nicolson discretization of the incompressible Navier–Stokes equations*, SIAM J. Numer. Anal., 58 (2020), pp. 375–409.

[46] M. Steuerwalt, *The existence, computation, and number of solutions of periodic parabolic problems*, SIAM J. Numer. Anal., 16 (1979), pp. 402–420.

[47] A. Whale, J.J. Lopez, M.E. Oszewski, S.C. Vigmostad, K.B. Chandran, J.D. Rossen, and M. Sonka, *Plaque development, vessel curvature, and wall shear stress in coronary arteries assessed by x-ray angiography and intravascular ultrasound*, Med. Image Anal., 10 (2006), pp. 615–631.

[48] T. Williams, et al., *Gnuplot* 4.6: *An interactive plotting program*, http://gnuplot.sourceforge.net/ (2013).

[49] Y. Yang, *Mathematical Modeling and Simulation of the Evolution of Plaques in Blood Vessels*, Ph.D. thesis, Universität Heidelberg, Heidelberg, 2014, https://doi.org/10.11588/heidok.00016425.

[50] Y. Yang, W. Jäger, M. Neuss-Radu, and T. Richter, *Mathematical modeling and simulation of the evolution of plaques in blood vessels*, J. Math. Biol., 72 (2016), pp. 973–996.

[51] Y. Yang, T. Richter, W. Jäger, and M. Neuss-Radu, *An ALE approach to mechanochemical processes in fluid-structure interactions*, Internat. J. Numer. Methods Fluids., 84 (2017), pp. 199–220.

[52] Q. Yu and J. Fish, *Temporal homogenization of viscoelastic and viscoplastic solids subjected to locally periodic loading*, Comput. Mech., 29 (2002), pp. 199–211.

[53] M.J. Zahr, P.-O. Persson, and J. Wilkening, *A fully discrete adjoint method for optimization of flow problems on deforming domains with time-periodicity constraints*, Comput. & Fluids, 139 (2016), pp. 130–147.

Towards parallel time-stepping for the numerical simulation of atherosclerotic plaque growth

Stefan Frei*, Alexander Heinlein†

Abstract

The numerical simulation of atherosclerotic plaque growth is computationally prohibitive, since it involves a complex cardiovascular fluid-structure interaction (FSI) problem with a characteristic time scale of milliseconds to seconds, as well as a plaque growth process governed by reaction-diffusion equations, which takes place over several months. In this work we combine a temporal homogenization approach, which separates the problem in computationally expensive FSI problems on a micro scale and a reaction-diffusion problem on the macro scale, with parallel time-stepping algorithms. It has been found in the literature that parallel time-stepping algorithms do not perform well when applied directly to the FSI problem. To circumvent this problem, a parareal algorithm is applied on the macro-scale reaction-diffusion problem instead of the micro-scale FSI problem. We investigate modifications in the coarse propagator of the parareal algorithm, in order to further reduce the number of costly micro problems to be solved. The approaches are tested in detailed numerical investigations based on serial simulations.

1 Introduction

Cardiovascular diseases are by far the most common cause of death in industrialized nations today. One of the most common cardiovascular diseases is the growth of plaque (atherogenesis) in coronary arteries or the pathological accumulation of plaque (atherosclerosis), which can result in heart attacks or strokes that are often fatal. Since the formation of plaque ranges over a long period of time, from months to several years, early diagnosis and treatment to prevent plaque growth can have a good chance of success.

An important driving force for plaque growth is the wall shear stress distribution, which varies significantly within each heartbeat, i.e., every second; see, for instance, [20,57] for a discussion on the dependence of plaque growth on the wall shear stress. Using three-dimensional fluid-structure interaction (FSI) simulations with realistic material models, the heartbeat scale has to be resolved by time steps in the order of milliseconds. We refer to [5,6] for the first studies on using complex wall models that take into account the effect of the reinforcing fibers within three-dimensional FSI simulations. For further references on FSI for cardiovascular applications, see, e.g., [9,10,26,38,40]; for FSI simulations in the context of plaque growth, see also [25,28,30,46,56,60].

Hence, realistic numerical simulations of plaque growth over several months that resolve each individual heartbeat would easily require $\mathcal{O}(10^9)$ sequential time steps. Consequently, even using today's fastest supercomputers, such a simulation is clearly infeasible.

Therefore, in [29] a temporal homogenization approach for separating the plaque growth time scale (macro scale) in the order of days and the FSI time scale (micro scale) in the order of milliseconds has been introduced. The approach is based on the assumption that the FSI is approximately periodic in time on the micro scale, which makes it possible to upscale the fluid dynamics to the macro scale. Using this approach it is possible to simulate only a moderate number of FSI time steps for each time step of the plaque growth problem. This means that we have to simulate only a few heartbeats (seconds) of the full FSI problem

*Department of Mathematics and Statistics, University of Konstanz, 78457 Konstanz, Germany, stefan.frei@uni-konstanz.de (Corresponding author)

†Delft University of Technology, Delft Institute of Applied Mathematics, Mekelweg 4, 2628CD Delft, Netherlands, a.heinlein@tudelft.nl

instead of the whole 24 hours of each day; the total number of time steps is reduced by a factor of $\mathcal{O}(10^4)$. Note that, completely neglecting the fine scale may introduce a significant error; see [28, 30]. See also the recent PhD thesis of Florian Sonner for further details on temporal homogenization for plaque growth [55].

However, considering realistic three-dimensional simulations the number of time steps corresponding to fine scale FSI problems remains still infeasibly high, even after temporal homogenization. In order to further reduce the computational times, we introduce a new approach based on parallel time-stepping. We focus on a classical parallel time-stepping method, the parareal algorithm, which has been introduced by Lions et al. in [42]; for an overview over the vast literature on parallel-in-time integration methods, we refer to the review papers [31, 45] and the references therein.

Due to a phase-shift in a coarse solution for hyperbolic partial differential equations (PDEs), such as the structural problem in FSI, it is generally challenging to apply parallel time-stepping methods to FSI problems; see, e.g., [43, 51]. Thus, instead of applying the parareal algorithm to the micro-scale FSI problem, we apply it to the homogenized plaque growth problem on the macro scale. Plaque growth is typically modeled by a system of reaction-diffusion equations; see, e.g., [54, 59]. Instead of being hyperbolic, they have a parabolic character and hence can be solved more efficiently by parallel time-stepping methods.

The focus of this paper is not the computation of accurate plaque-growth predictions in patient-specific geometries. Instead, our objective is to introduce a numerical framework for making the simulation times feasible and to investigate the methods numerically. Therefore, we make several simplifications: We focus on two-dimensional FSI simulations on a simple geometry. Furthermore, we consider two simplified models for the plaque growth, that is, the ordinary differential equation (ODE) model already considered in [28, 29] as well as a more complex partial differential equation (PDE) of reaction-diffusion type. We formulate the parareal algorithm for the two-scale problem of plaque growth and also propose variants to increase the efficiency of the coarse-scale propagators. Finally, we investigate the parallel time-stepping methods numerically based on serial simulations. As, even in two dimensions, the micro-scale FSI problems are generally much more expensive than the coarse plaque growth problem, we are able to give some good estimates for the speedup that can be expected in fully parallel simulations.

The paper is structured as follows: In section 2, we introduce our fluid-structure interaction problem as well as the two solid growth models considered for modeling the plaque growth: an ODE-based model in section 2.2.1 and a reaction-diffusion equation-based model in section 2.2.2. Next, in section 3, we briefly introduce the variational formulation of the FSI problem and then describe the temporal homogenization approach including some first numerical results. In section 4, we recap the parareal algorithm and discuss how we estimate the computational costs and the possible speed up in parallel simulations. We also discuss numerical results for the parallel time-stepping approach using the simple ODE plaque growth model. Furthermore, we introduce some ideas for reducing the costs of the coarse-scale propagators. In section 5, we show results for the reaction-diffusion plaque growth model, including a theoretical discussion on the computational costs of the algorithms. We conclude in section 6, where we also give a brief outlook on topics for future work.

2 Model equations

We consider a time-dependent fluid-structure interaction system, where the fluid is modeled by the Navier–Stokes equations and the solid by the Saint Venant–Kirchhoff model. In order to account for the solid growth, we add a multiplicative growth term to the deformation gradient, which is motivated by typical plaque growth models [30, 44, 54, 59].

2.1 Fluid-structure interaction

Consider a partition of an overall domain $\Omega(t) = \mathcal{F}(t) \cup \Gamma(t) \cup \mathcal{S}(t)$ into a fluid part $\mathcal{F}(t)$, an interface $\Gamma(t)$ and a solid part $\mathcal{S}(t)$. The blood flow and its interaction with the surrounding vessel wall is modelled

by the following FSI system:

$$\begin{aligned} \rho_f(\partial_t \mathbf{v}_f + \mathbf{v}_f \cdot \nabla \mathbf{v}_f) - \operatorname{div} \boldsymbol{\sigma}_f = 0, \qquad & \operatorname{div} \mathbf{v}_f = 0 \qquad && \text{in } \mathcal{F}(t), \\ \rho_s \partial_t \hat{\mathbf{v}}_s - \operatorname{div}(\hat{\mathbf{F}}_e \hat{\boldsymbol{\Sigma}}_e) = 0, \qquad & \partial_t \hat{\mathbf{u}}_s - \hat{\mathbf{v}}_s = 0 \qquad && \text{in } \hat{S}, \\ \boldsymbol{\sigma}_f \vec{n}_f + \boldsymbol{\sigma}_s \vec{n}_s = 0, \qquad & \mathbf{v}_f = \mathbf{v}_s \qquad && \text{on } \Gamma(t). \end{aligned} \tag{1}$$

Here, $\mathbf{v}_f$ and $\hat{\mathbf{v}}_s$ denote the fluid and solid velocity, respectively, and $\hat{\mathbf{u}}_s$ the solid displacement. Quantities with a "hat" are defined in Lagrangian coordinates, while quantities without a "hat" are defined in the current Eulerian coordinate framework. Two quantities $\hat{f}(\hat{x})$ and $f(x)$ correspond to each other by a $C^{1,1}$-diffeomorphism $\hat{\xi} : \hat{\Omega} \to \Omega(t)$ and the relation $\hat{f} = f \circ \hat{\xi}$. Later, we will also need the solid deformation gradient $\hat{F}_s = I + \hat{\nabla} \hat{u}_s$, which is the derivative of $\hat{\xi}$ in the solid part. The constants ρ_f and ρ_s are the densities of blood and vessel wall, and $\vec{n}_f$ and $\vec{n}_s$ are outward pointing normal vectors of the fluid and solid domain, respectively.

By $\boldsymbol{\sigma}_f$ and $\boldsymbol{\sigma}_s$ we denote the Cauchy stress tensors of fluid and solid. Using the well-known Piola transformation between the Eulerian and the Lagrangian coordinate systems, we can relate $\boldsymbol{\sigma}_s$ and the second Piola–Kirchhoff stress $\hat{\Sigma}_e$ as follows:

$$\boldsymbol{\sigma}_s(x) = \hat{\boldsymbol{\sigma}}_s(\hat{x}) = \hat{J}_e^{-1} \hat{\mathbf{F}}_e \hat{\boldsymbol{\Sigma}}_e(\hat{x}) \hat{\mathbf{F}}_e^T,$$

where $\hat{F}_e$ is the elastic part of the deformation gradient $\hat{F}_s$, For modeling the material behavior of the vessel wall, different approaches have been proposed and investigated in literature; see, e.g., [5,7,8,13,37] for more sophisticated material models, for instance, incorporating an anisotropic behavior due to the reinforcing fibers. For the sake of simplicity, we use in this work the relatively simple Saint Venant-Kirchhoff model with the Lamé material parameters μ_s and λ_s

$$\hat{\boldsymbol{\Sigma}}_e = 2\mu_s \hat{\mathbf{E}}_e + \lambda_s \operatorname{tr}(\hat{\mathbf{E}}_e) I, \quad \hat{\mathbf{E}}_e := \frac{1}{2}(\hat{\mathbf{F}}_e^T \hat{\mathbf{F}}_e - I). \tag{2}$$

The Saint Venant-Kirchhoff model is based on Hooke's linear material law in a large strain formulation, resulting in a weakly nonlinear material model.

The blood flow is modeled as an incompressible Newtonian fluid, such that the Cauchy stresses are given by

$$\boldsymbol{\sigma}_f = \rho_f \nu_f (\nabla \mathbf{v}_f + \nabla \mathbf{v}_f^T) - p_f I, \tag{3}$$

where ν_f is the kinematic viscosity of blood. A sketch of the computational domain is given in fig. 1. We split the outer boundary of Ω into a solid part Γ_s with homogeneous Dirichlet conditions, a fluid part Γ_f^{in} with an inflow Dirichlet condition and an outflow part Γ_f^{out}, where a *do-nothing condition* is imposed.

The boundary data is given by

$$\mathbf{v}_f = \mathbf{v}^{\text{in}} \text{ on } \Gamma_f^{\text{in}}, \quad \rho_f \nu_f (\vec{n}_f \cdot \nabla) \mathbf{v}_f - p_f \vec{n}_f = 0 \text{ on } \Gamma_f^{\text{out}}, \quad \hat{\mathbf{u}}_s = 0 \text{ on } \hat{\Gamma}_s, \tag{4}$$

where $\mathbf{v}^{\text{in}}$ is the inflow velocity on Γ_f^{in}.

2.2 Modelling of solid growth

Developing a realistic model of plaque growth at the vessel wall is a complex task that involves the interaction of many different molecules and species, see for example [54]. Furthermore, the plaque growth will also strongly depend on the geometry and material model of the arterial wall. In this contribution, our focus does not lie on a realistic modeling of plaque growth, and thus, our model is greatly simplified. However, the two models considered here are chosen to behave similarly (from a numerical viewpoint) to more sophisticated models. We consider a simple ODE-based model in section 2.2.1 and a more complex but still relatively simple PDE model of reaction-diffusion type in section 2.2.2. Both models focus on the influence of the concentration of foam cells c_s on the growth. Numerical results can be found in sections 3.2, 4.1.4 and 4.2 as well as in section 5, respectively.

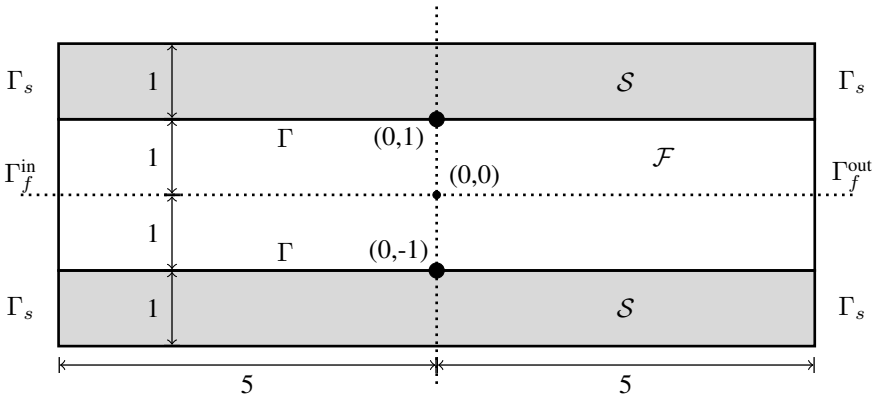


Figure 1: Sketch of the computational domain centered at the origin $(0,0)$; $\mathcal{F}$ and $\mathcal{S}$ are the fluid and solid parts, respectively, and the solid lines correspond to the fluid-solid interface Γ. The plaque growth is initiated at $(0,\pm 1)$.

2.2.1 A simple ODE model for solid growth

In the first model, which is taken from [28,30], the evolution of the foam cell concentration c_s depends only on its current value but not on its spatial distribution. Hence, it can be described by a simple ODE. The rate of formation of these cells depends on the distribution of the wall shear stress $\boldsymbol{\sigma}_f^{WS}$ at the vessel wall. As a result, we obtain the following simplified ODE model:

$$\begin{aligned} \partial_t c_s &= \gamma(\boldsymbol{\sigma}_f^{WS}, c_s) \coloneqq \alpha\big(1+c_s\big)^{-1}\left(1+\frac{\left\|\boldsymbol{\sigma}_f^{WS}\right\|^2_{L^2(\Gamma)}}{\sigma_0^2}\right)^{-1}, \\ \boldsymbol{\sigma}_f^{WS} &\coloneqq \rho\nu\big(I_d-\vec{n}_f\vec{n}_f^T\big)(\nabla\mathbf{v}+\nabla\mathbf{v}^T)\vec{n}_f. \end{aligned} \tag{5}$$

The reference wall shear stress σ_0 and the scale separation parameter α are parameters of the growth model. For cardiovascular plaque growth, we have typically $\alpha = \mathcal{O}(10^{-7})\,\mathrm{s}^{-1}$; see also [29].

We model the solid growth by a multiplicative splitting of the deformation gradient $\hat{\mathbf{F}}_s$ into an elastic part $\hat{\mathbf{F}}_e$ and a growth function $\hat{\mathbf{F}}_g$

$$\hat{\mathbf{F}}_s = \hat{\mathbf{F}}_e\hat{\mathbf{F}}_g \quad \Leftrightarrow \quad \hat{\mathbf{F}}_e = \hat{\mathbf{F}}_s\hat{\mathbf{F}}_g^{-1} = [I+\hat{\nabla}\hat{\mathbf{u}}_s]\hat{\mathbf{F}}_g^{-1}; \tag{6}$$

cf. [30, 50, 59]. In the ODE model, we use the following growth function depending on c_s

$$\hat{g}(\hat{x},\hat{y},t) = 1 + c_s\exp\left(-\hat{x}^2\right)(2-|\hat{y}|), \quad \hat{\mathbf{F}}_g(\hat{x},\hat{y},t) \coloneqq \hat{g}(\hat{x},\hat{y},t)\,I. \tag{7}$$

This means that the shape and position of the plaque growth is prescribed, but the growth rate depends on the variable c_s. As the simulation domain is centered around the origin, see fig. 1, growth is concentrated at $(0,\pm 1)$, in the center of the domain. It follows that

$$\hat{\mathbf{F}}_g = \hat{g}I \quad \Rightarrow \quad \hat{\mathbf{F}}_e \coloneqq \hat{g}^{-1}\hat{\mathbf{F}}_s, \tag{8}$$

and the elastic Green–Lagrange strain is given by

$$\hat{\boldsymbol{E}}_e = \frac{1}{2}(\hat{\boldsymbol{F}}_e^T\hat{\boldsymbol{F}}_e - I) = \frac{1}{2}(\hat{g}^{-2}\hat{\boldsymbol{F}}_s^T\hat{\boldsymbol{F}}_s - I) \tag{9}$$

resulting in the Piola–Kirchhoff stresses

$$\hat{\mathbf{F}}_e\hat{\boldsymbol{\Sigma}}_e = 2\mu_s\hat{\boldsymbol{F}}_e\hat{\boldsymbol{E}}_e + \lambda_s\,\mathrm{tr}(\hat{\boldsymbol{E}}_e)\hat{\boldsymbol{F}}_e = 2\mu_s\hat{g}^{-1}\hat{\boldsymbol{F}}_s\hat{\boldsymbol{E}}_e + \lambda_s\hat{g}^{-1}\,\mathrm{tr}(\hat{\boldsymbol{E}}_e)\hat{\boldsymbol{F}}_s. \tag{10}$$

2.2.2 A PDE reaction-diffusion model

Secondly, we consider a slightly more complex growth model, where the concentration of foam cells $\hat{c}_s = \hat{c}_s(\hat{x}, t)$ ($\hat{x} \in \hat{\mathcal{S}}$) is now governed by a PDE model, namely a non-stationary reaction-diffusion equation

$$\begin{aligned} \partial_t \hat{c}_s - D_s \hat{\Delta} \hat{c}_s + R_s \hat{c}_s (1 - \hat{c}_s) = 0 \quad \text{on } \hat{\mathcal{S}}, \qquad D_s \hat{\partial}_n \hat{c}_s = \gamma(\boldsymbol{\sigma}_f^{WS}) = \alpha \delta(x) \left(1 + \frac{\left\| \boldsymbol{\sigma}_f^{WS} \right\|^2}{\sigma_0^2} \right)^{-1} \quad \text{on } \hat{\Gamma}, \\ \hat{c}_s = 0 \quad \text{on } \hat{\Gamma}_s, \qquad \hat{c}_s(\cdot, 0) = 0 \quad \text{on } \hat{\mathcal{S}}, \end{aligned} \tag{11}$$

where $\delta(x) = \min\{0, (x-1)(x+1)\}^2$ and $\|\cdot\|$ is the Euclidean norm. Note that the norm $\|\boldsymbol{\sigma}_f^{WS}\|$ on the right-hand side of eq. (11) is a function in space and time, whereas the right-hand side of eq. (5), including $\|\boldsymbol{\sigma}_f^{WS}\|_{L^2(\Gamma)}$, depends only on time and is a constant in space. Furthermore, we have σ_0 and α as in the ODE model eq. (6), as well as positive diffusion and reaction coefficients D_s and R_s, respectively. The underlying idea is that the vessel wall is initially damaged in the central part of the interface around $\hat{x}_1 = 0$ and monocytes can penetrate into $\hat{\mathcal{S}}$ in the part of the interface corresponding to $\hat{x}_1 \in (-1, 1)$, where δ is actually positive.

Moreover, in contrast to the first numerical example, we do not prescribe the shape of the plaque growth by a growth function $\hat{g}$. Instead, the growth part F_g depends on the spatial distribution of the foam cell concentration via the relation

$$\hat{g}(\hat{x}, t) = 1 + \hat{c}_s(\hat{x}, t), \quad \hat{\mathbf{F}}_g(\hat{x}, t) := \hat{g}(\hat{x}, t)\, I. \tag{12}$$

The Green Lagrange strain $\hat{\boldsymbol{E}}_e$ and the Piola–Kirchhoff stresses $\hat{\boldsymbol{\Sigma}}_e$ are then defined as above in eqs. (9) and (10), with $\hat{g}$ given in eq. (12).

Note that, solving the PDE model has higher computational cost compared to the ODE model described in section 2.2.1. On the other hand, a sophisticated reaction-diffusion problem may yield more realistic results for the plaque growth. Nonetheless, such a model remains significantly cheaper compared to a fully coupled FSI problem; this discrepancy will only increase when moving to three-dimensional simulations with complex material models for the arterial wall. This observation is essential for the efficiency of our parallel time-stepping approach, as we will discuss in more detail in the following sections.

3 Numerical framework

In this work, we use an Arbitrary Lagrangian Eulerian (ALE) approach to solve the FSI problem in eq. (1). The ALE approach is the standard approach for FSI with small to moderate structural deformations; see, e.g., [10, 16, 18, 21, 26, 48, 58]. This assumption holds generally true for the simulation of plaque growth, unless the interest is to simulate a complete closure of the artery. To simulate a full closure, a Fully Eulerian formalism [22,27,47] has been used in [27,30]; see also [2,14,15] for further works on FSI-contact problems in a Fully Eulerian or Lagrange-Eulerian formalism.

Given a suitable ALE map $\hat{\xi}_f = \hat{x} + \hat{\mathbf{u}}_f$, its gradient $\hat{\mathbf{F}}_f = I + \hat{\nabla} \hat{\mathbf{u}}_f$ and determinant $\hat{J}_f = \det \hat{\mathbf{F}}_f$, the ALE formulation of the FSI problem is given by:

Variational formulation 1 (Non-stationary FSI problem)**.** *Find velocity $\hat{\mathbf{v}} \in \mathbf{v}^{in} + \mathcal{V}$, deformation $\hat{\mathbf{u}} \in \mathcal{W}$ and pressure $\hat{p}_f \in \mathcal{L}_f$, such that*

$$\begin{aligned} &\left(\rho_f \hat{J}_f \hat{\partial}_t \hat{\mathbf{v}}_f, \hat{\phi}\right)_{\hat{\mathcal{F}}} + \left(\rho_f \hat{J}_f \hat{\nabla} \hat{\mathbf{v}}_f \hat{\mathbf{F}}_f^{-1} (\hat{\mathbf{v}}_f - \partial_t \hat{\mathbf{u}}_f), \hat{\phi}\right)_{\hat{\mathcal{F}}} + \left(\hat{\rho}_s^0 \hat{\partial}_t \hat{\mathbf{u}}_s, \hat{\phi}\right)_{\hat{\mathcal{S}}} \\ &+ \left(\hat{J}_f \hat{\boldsymbol{\sigma}}_f \hat{\mathbf{F}}_f^{-T}, \hat{\nabla} \hat{\phi}\right)_{\hat{\mathcal{F}}} - \left(\hat{J}_f \nu_f \hat{\mathbf{F}}_f^{-1} \hat{\nabla} \hat{\mathbf{v}}_f \hat{\mathbf{F}}_f^{-1} \hat{n}_f, \hat{\phi}\right)_{\hat{\Gamma}^{out}} + \left(\hat{\mathbf{F}}_e(\hat{g}) \hat{\boldsymbol{\Sigma}}_e(\hat{g}), \hat{\phi}\right)_{\hat{\mathcal{S}}} = 0 \ \forall \hat{\phi} \in \mathcal{V}, \\ &\left(\widehat{\operatorname{div}}(\hat{J}_f \hat{\mathbf{F}}_f^{-1} \hat{\mathbf{v}}_f), \hat{\xi}_f\right)_{\hat{\mathcal{F}}} = 0 \ \forall \hat{\xi}_f \in \mathcal{L}_f, \\ &\left(d_t \hat{\mathbf{u}}_s - \hat{\mathbf{v}}_s, \hat{\psi}_s\right)_{\hat{\mathcal{S}}} = 0 \ \forall \hat{\psi}_s \in \mathcal{L}_s, \end{aligned}$$

where $\hat{\sigma}_f := \rho_f \nu_f (\hat{\nabla}\hat{\mathbf{v}}_f \hat{\mathbf{F}}_f^{-1} + \hat{\mathbf{F}}_f^{-T} \nabla \hat{\mathbf{v}}_f^T) - \hat{p}_f I$.

There are different ways to compute the ALE map depending on the displacement $\hat{\mathbf{u}}_s$ of the structural domain $\hat{\mathcal{S}}$. In our simulations, we extend $\hat{\mathbf{u}}_s$ into the fluid domain using harmonic extensions (called $\hat{\mathbf{u}}_f$); that is, we solve a Laplacian problem with right hand side zero and $\hat{\mathbf{u}}_s$ as Dirichlet boundary data on the fluid domain.

Note that $\hat{\mathbf{\Sigma}}_e$ and $\hat{\mathbf{F}}_e$ depend on $\hat{g}$ (and hence on c_s), as specified in eq. (10). The function spaces are given by

$$\mathcal{V} = [H_0^1(\hat{\Omega}; \hat{\Gamma}_f^{\text{in}} \cup \hat{\Gamma}_s)]^d, \quad \mathcal{W} = [H_0^1(\hat{\mathcal{S}}; \hat{\Gamma}_s)]^d, \quad \mathcal{L}_s := L^2(\hat{\mathcal{S}}), \quad \mathcal{L}_f = L^2(\hat{\mathcal{F}}).$$

Furthermore, we assume that the solution $\hat{\mathbf{v}}_f$ has higher regularity, such that the trace of $\hat{\nabla}\hat{\mathbf{v}}_f$ is well-defined on $\hat{\Gamma}^{\text{out}}$, as needed in Variational formulation 1.

3.1 Temporal two-scale approach

Even for the simplified two-dimensional configuration considered in this work, a resolution of the micro-scale dynamics with a scale of milliseconds to seconds is unfeasible over the complete time interval of interest $[0, T_{\text{end}}]$, with T_{end} being several months up to a year. For instance, when considering a relatively coarse micro-scale time step of $\delta\tau = 0.02$s, the number of time steps required to simulate a time frame of a whole year would be $365 \cdot 86\,400 \cdot \frac{1s}{\delta\tau} \approx 1.58 \cdot 10^9$, each step corresponding to the solution of a mechano-chemical FSI problem.

This dilemma is frequently solved by considering a heuristic averaging: as the micro-scale is much smaller than the macro scale one considers a stationary limit of the FSI and solves for the stationary FSI problem on the macro scale (e.g., $\delta t \approx 1$ day); see, e.g., [17, 30, 59]. The wall-shear stress $\overline{\boldsymbol{\sigma}}_f^{WS}$ of the solution of this stationary FSI problem is then used to advance the foam cell concentration in eq. (5) or eq. (11).

Variational formulation 2 (Stationary FSI problem)**.** *Find velocity $\bar{\mathbf{v}} \in \mathbf{v}^{in} + \mathcal{V}$, deformation $\bar{\mathbf{u}} \in \mathcal{W}$ and pressure $\bar{p}_f \in \mathcal{L}_f$, such that*

$$\begin{aligned} &\left(\rho_f \hat{J}_f \hat{\nabla}\bar{\mathbf{v}}_f \bar{\mathbf{F}}_f^{-1} \bar{\mathbf{v}}_f, \hat{\phi}\right)_{\hat{\mathcal{F}}} + \left(\bar{J}_f \bar{\boldsymbol{\sigma}}_f \bar{\mathbf{F}}_f^{-T}, \hat{\nabla}\hat{\phi}\right)_{\hat{\mathcal{F}}} \\ -&\left(\hat{J}_f \nu_f \hat{\mathbf{F}}_f^{-1} \hat{\nabla}\hat{\mathbf{v}}_f \hat{\mathbf{F}}_f^{-1} \hat{n}_f, \hat{\phi}\right)_{\hat{\Gamma}^{out}} + \left(\bar{\mathbf{F}}_e(\hat{g}) \bar{\mathbf{\Sigma}}_e(\hat{g}), \hat{\phi}\right)_{\hat{\mathcal{S}}} = 0 \;\; \forall \hat{\phi} \in \mathcal{W}, \\ &\left(\widehat{\text{div}}(\bar{J}_f \bar{\mathbf{F}}_f^{-1} \bar{\mathbf{v}}_f), \hat{\xi}_f\right)_{\hat{\mathcal{F}}} = 0 \;\; \forall \hat{\xi}_f \in \mathcal{L}_f \end{aligned}$$

The foam cell concentration is then advanced by the ODE in eq. (5) resp. the PDE in eq. (11), with $\boldsymbol{\sigma}_f^{WS}$ replaced by $\overline{\boldsymbol{\sigma}}_f^{WS}$. It has, however, been shown that $\gamma(\overline{\boldsymbol{\sigma}}_f^{WS})$ is not necessarily a good approximation of $\gamma(\boldsymbol{\sigma}_f^{WS})$, which depends on the pulsating blood flow; see the numerical results in [28, 30] and the analysis in [29].

A more accurate two-scale approach has been presented by Frei and Richter in [29]. The numerical approach can be cast in the framework of the Heterogeneous Multiscale Method (HMM); see, e.g., [1, 23, 24]. In [29], a periodic-in-time micro-scale problem is solved in each time step of the macro scale, for instance each day. The growth function $\gamma(\boldsymbol{\sigma}_f^{WS})$ is then averaged by integrating over one period of the heart beat, its average will be denoted by $\overline{\gamma}(\boldsymbol{\sigma}_f^{WS})$. This averaged growth function is applied to advance the foam cell concentration by eq. (5) or eq. (11), respectively. A schematic illustration of the two-scale algorithm is given in fig. 2.

To be precise, we divide the macro-scale time interval $[0, T_{\text{end}}]$ into N_f time steps of size δt

$$0 = t_0 < t_1 < \ldots < t_{N_f} = T_{\text{end}}, \qquad N_f = \frac{T_{\text{end}}}{\delta t}. \tag{13}$$

As c_s varies significantly on the macro scale only, using $c_s(t_m)$ as a fixed value for the growth variable on the micro scale results in a sufficiently good approximation in the time interval $[t_m, t_{m+1}]$. Then, one cycle

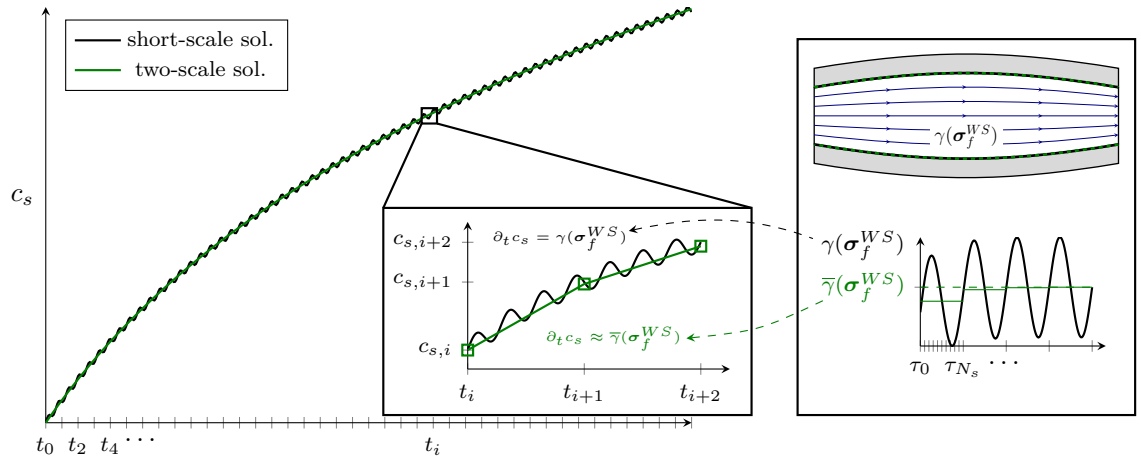


Figure 2: Schematic representation of the two-scale algorithm as described in section 3.1 and summarized in algorithm 1: instead of solving the FSI problem and updating the foam cell concentration c_s on the micro time scale ($\delta\tau$, black), we compute an averaged growth function $\overline{\gamma}(\boldsymbol{\sigma}_f^{WS})$ over a few periods (heartbeats) of the micro-scale problem (right); this average is used to update c_s on the macro scale (δt, green in the left part).

of the pulsating blood flow problem (around 1s) is to be resolved on the micro scale $\delta\tau$

$$0 = \tau_0 < \tau_1 < .... < \tau_{N_s} = 1s, \qquad N_s = \frac{1s}{\delta\tau}. \tag{14}$$

It has been shown in [29] (for a simplified flow configuration) that this approach leads to a model error $\mathcal{O}(\epsilon)$ compared to a full resolution of the micro scale, where $\epsilon = \frac{1s}{T_{\text{end}}}$ denotes the ratio between macro- and micro time scale and is in the range of $\mathcal{O}(10^{-7})$ for a typical cardiovascular plaque growth problem. The relations in eqs. (13) and (14) imply $\delta\tau = \epsilon\delta t$. In the model problems formulated above, the scale separation is induced by the parameter $\alpha = \mathcal{O}(\epsilon)$.

A difficulty lies in solving the periodic micro-scale problem. If accurate initial conditions $\mathbf{w}^0 := (\mathbf{v}^0, \mathbf{u}^0)$ are available on the micro-scale, a periodic solution can be computed using a time-stepping procedure for one cycle. If the initial conditions are known approximately, convergence to the periodic solution may still be obtained after simulating a few cycles of the micro-scale problem due to the dissipation of the flow problem; see [28, 29]. Numerically, it can be checked after each cycle if the solution is sufficiently close to a periodic state. In this work, we apply a stopping criterion based on the computed averaged growth value:

$$|\overline{\gamma}(\boldsymbol{\sigma}_f^{WS,r}) - \overline{\gamma}(\boldsymbol{\sigma}_f^{WS,r-1})| < \epsilon_p,$$

where $r = 1, 2, ...$ denotes the iteration index with respect to the number of cycles of the micro-scale problem. The algorithm is summarized as algorithm 1, where we use the abbreviation $\mathbf{w}^{r,s} := (\mathbf{v}^{r,s}, \mathbf{u}^{r,s})$.

As starting values $\mathbf{w}^{r,0} = (\mathbf{v}^{r,0}, \mathbf{u}^{r,0})$ in step 1. of the algorithm, we use the variables $\mathbf{w}^{r-1,N_s} = (\mathbf{v}^{r-1,N_s}, \mathbf{u}^{r-1,N_s})$ from the quasi-periodic state of the previous macro step. It has been observed in [28] that these are usually closer to the starting values of the periodic state than the solution of an averaged stationary problem on the macro scale.

3.2 Numerical example

Before we present the different approaches for parallel time-stepping, let us illustrate the two-scale approach by a first numerical example. Therefore, we use the simple ODE growth model introduced in section 2.2.1, eq. (5). The test configuration introduced here will be used in section 4 as well.

Algorithm 1: Two-Scale Algorithm

Set suitable starting values $\mathbf{w}^{0,0} = (\mathbf{v}^{0,0}, \mathbf{u}^{0,0})$ and time-step sizes $\delta t, \delta T$.

for $n = 1, 2, \ldots, N_f$ **do**

1.) **Micro problem**: Set $r \leftarrow 0$

while $|\overline{\gamma}(\boldsymbol{\sigma}_f^{WS,r}) - \overline{\gamma}(\boldsymbol{\sigma}_f^{WS,r-1})| > \epsilon_p$ **do**

1.a) Solve micro-scale problem in eq. (1) in $I_n = (t_n, t_n + 1s)$

$$\{\mathbf{w}^{r,0}, c_s^{n-1}\} \mapsto \{\mathbf{w}^{r,m}\}_{m=1}^{N_s}$$

1.b) Compute the averaged growth function

$$\overline{\gamma}(\boldsymbol{\sigma}_f^{WS,r+1}) = \frac{1}{N_s} \sum_{m=1}^{N_s} \gamma(\boldsymbol{\sigma}_f^{WS,m}(\mathbf{v}^{r+1,m}), c_s^{n-1})$$

and set $\mathbf{w}^{r+1,0} = \mathbf{w}^{r,N_s}$, $r \leftarrow r + 1$.

2.) **Macro problem**: Update the foam cell concentration c_s^n by eq. (5) or eq. (11).

Concerning the geometry we use a two-dimensional channel of length 10 cm and an initially constant width $\omega(0)$ of 2 cm as illustrated in fig. 1. The solid parts on the top and bottom corresponding to the arterial wall have an initial thickness of 1 cm each. Fluid density and viscosity are given by $\rho_f = 1\,\mathrm{g/cm^3}$ and $\nu_f = 0.04\,\mathrm{cm^2/s}$, respectively. The growth parameters are set to $\sigma_0 = 30\frac{\mathrm{g\,cm}}{\mathrm{s^2}}$ and $\alpha = 5 \cdot 10^{-7}\frac{1}{\mathrm{s}}$, which yields a realistic time-scale for the arterial plaque growth, see [29]. The solid parameters are $\rho_s = 1\,\mathrm{g/cm^3}$, $\mu_s = 10^4\,\mathrm{dyne/cm^2}$, and $\lambda_s = 4 \cdot 10^4\,\mathrm{dyne/cm^2}$. As an inflow boundary condition, we prescribe a pulsating velocity inflow profile on Γ_f^{in} given by

$$\mathbf{v}_f^{\mathrm{in}}(t, x) = 30 \begin{pmatrix} \sin(\pi \frac{t}{\mathcal{P}})^2 (1 - x_2^2) \\ 0 \end{pmatrix} \mathrm{cm/s}. \tag{15}$$

Here, $\mathcal{P} = 1\,\mathrm{s}$ is the period of a single heartbeat. The symmetry of the configuration can be exploited in order to reduce computational cost by simulating only on the lower half of the computational domain and imposing the symmetry condition $\mathbf{v}_f \cdot \vec{n} = 0$, $\tau^T \boldsymbol{\sigma}_f \vec{n} = 0$ on the symmetry boundary Γ_{sym}, where τ denotes a tangential vector.

We discretize the FSI problem (variational formulation 1) in time using the backward Euler method. For discretizing the ODE growth model, we use the forward Euler method, which results in

$$c_s^n = c_s^{n-1} + \delta t\, \overline{\gamma}(\boldsymbol{\sigma}_f^{WS}, c_s^{n-1}), \qquad n = 1, ..., N_f. \tag{16}$$

For spatial discretization, we use biquadratic (Q_2) equal-order finite elements for all variables and LPS stabilization [11] for the fluid problem. Our mesh, containing both fluid and solid, consists of 160 rectangular grid cells; this corresponds to a total of 3 157 degrees of freedom. The time-step sizes are chosen as $\delta\tau = 0.02\,\mathrm{s}$ and $\delta t = 0.3$ days; the tolerance for periodicity of the micro-scale problem as $\epsilon_p = 10^{-3}$. All the computational results have been obtained with the finite element library Gascoigne3d [12]. We use a fully monolithic approach for the FSI problem following Frei, Richter & Wick [30, 48].

In fig. 3, we compare results obtained with the two-scale approach for different macro time-step sizes δt with the heuristic averaging approach outlined above; see variational formulation 2. We see that the pure averaging approach underestimates the growth significantly. Even a very coarse discretization of the macro-scale time interval $\delta t = 15$ days in the two-scale approach gives a much better approximation. We note, however, that too coarse time steps might introduce different issues, as the starting values $\mathbf{v}^{0,0}$ and $\mathbf{u}^{0,0}$ might not be good approximations for the periodic state on the micro scale anymore. In combination with a significant mesh deformation for $t \approx 300$ days, this led to divergence of algorithm 1 for an even coarser

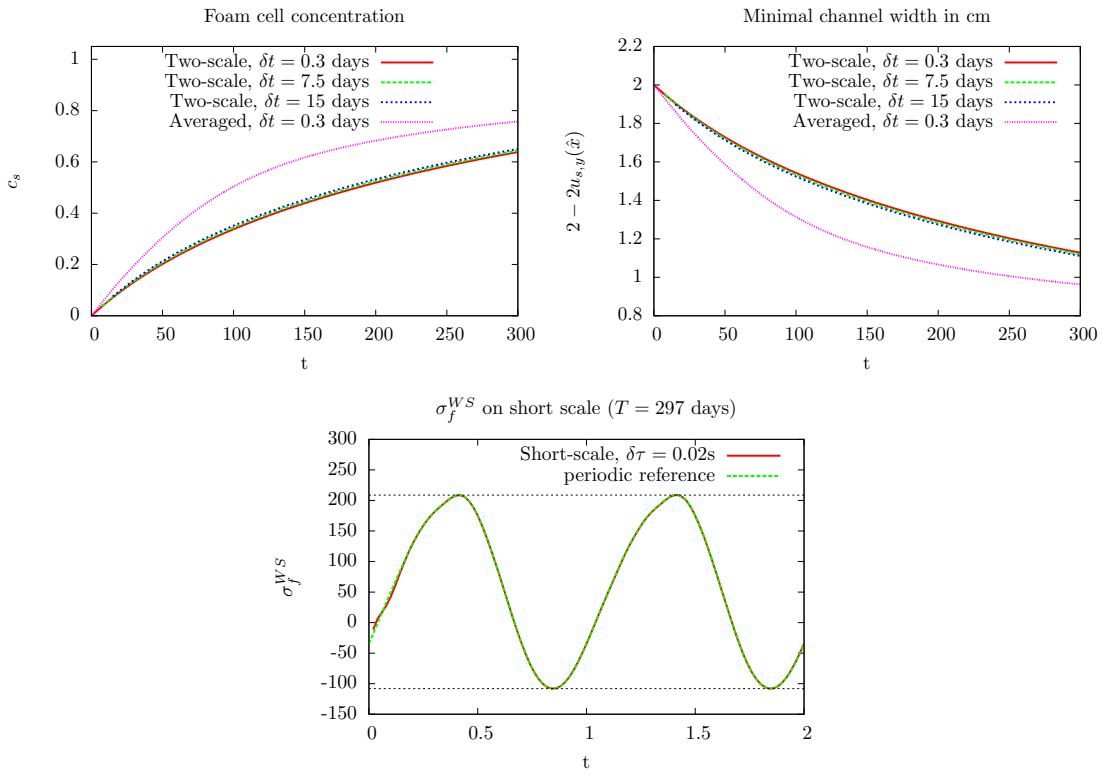


Figure 3: **Top:** Comparison of a pure averaging approach with the two-scale approaches with different macro-scale time steps δT. **Top left**: Concentration c_s over time. **Top right**: Channel width over time. **Bottom**: Wall shear stress $\boldsymbol{\sigma}_f^{WS}$ over two periods of the micro scale at $t = 297$ days for $\delta t = 9$ days. The initial values are taken from the previous micro problem at $t = 288$ days.

macro time-step size $\delta t = 30$ days. For $\delta t \leq 15$ days, a near-periodic state was reached in 2-3 iterations in each macro-step. A detailed convergence study for a very similar problem has been given in [28].

At the bottom of fig. 3, we show the $L^2(\Gamma)$-norm of the wall shear stress over two periods, i.e., 2 seconds, of the micro problem for $\delta t = 0.02$s and $\delta T = 9$ days. The initial values are taken from the periodic state of the micro problem at the previous macro time-step, i.e., 9 days before. We see that the wall shear stresses converge very quickly to the periodic state. An initial deviation of $6.28 \frac{\text{g cm}}{\text{s}^2}$ to the periodic reference solution at time $\tau = 0.02$ is reduced to $6 \cdot 10^{-4} \frac{\text{g cm}}{\text{s}^2}$ at time $\tau = 1$ within only one period.

4 Parallel time-stepping

The main cost in algorithm 1 lies in the solution of the non-stationary micro problem in step 1.a, which needs to be solved in each time step of the macro problem. Considering a relatively coarse micro-scale discretization of $\delta t = 0.02\,\text{s}$, as used in the previous section, 50 time steps are necessary to compute a single period of the heart beat. The simulation of two or more cycles might be necessary to obtain a near-periodic state in step 1.a of algorithm 1; cf. the discussions in [28, 29]. In a realistic scenario, each time step of the micro problem corresponds to the solution of a complex three-dimensional FSI problem, which makes already the solution of one micro problem very costly.

For this reason, parallelization needs to be exploited in different ways: First, one can make use of spatial parallelization using a scalable solver, for instance, based on domain decomposition or multigrid methods; see, e.g., [9,19,33,35,36,39,58]. Since the focus of this contribution is on parallelization in time, we will not

discuss this aspect here. In particular, we expect that even when the speedup due to spatial parallelization saturates, the computing times for the whole plaque growth simulations will remain unfeasible. Therefore, we have to make use of an additional level of parallelization, that is, temporal parallelization of the macro-scale problem. In particular, we will use the parareal algorithm, which we will recap in the next subsection.

In order to motivate the algorithmical developments in this section, we will already present some first numerical results for the ODE growth model in eq. (5) within the section.

4.1 The parareal algorithm

First, the time interval of interest $[0, T_{\text{end}}]$ on the macro scale is divided into P sub-intervals $I_p = [T_{p-1}, T_p]$ of equal size, where

$$0 = T_0 < T_1 < ... < T_P = T_{\text{end}}. \tag{17}$$

In order to define the parareal algorithm, suitable fine and coarse problems need to be introduced. Note again that we will apply the parareal algorithm only on the macro scale as introduced in section 3.1; hence, both the fine and the coarse scale of parareal correspond to the macro scale of the homogenization approach. The fine problem advances the growth variable c_s from time T_p to T_{p+1} by solving algorithm 1 with a smaller time-step size δt (e.g., 0.3 days) on the corresponding fine time discretization of $[T_p, T_{p+1}]$:

$$T_p = t_{p,0} < t_{p,1} < ... < t_{p,n_p} = T_{p+1}, \qquad n_p = \frac{T_{p+1} - T_p}{\delta t}, \qquad t_{p,q} := t_{p \cdot n_p + q} = t_{p,q-1} + \delta t.$$

The fine propagator on a process p consists of a time-stepping procedure to advance $c_s(T_p)$ to $c_s(T_{p+1})$ with the fine time step δt. We write

$$c_s^{\text{fine}}(T_{p+1}) = \mathcal{F}(c_s^{\text{fine}}(T_p)).$$

The efficiency of the parareal algorithm depends strongly on the computational cost of the coarse propagator. In particular, it needs to be much cheaper than the fine problems since it is defined globally on $[0, T_{\text{end}}]$ and, hence, introduces synchronization. Thus, we use a large time-step δT and

$$T_p = \bar{T}_{p,0} < ... < \bar{T}_{p,N_p} = T_{p+1}, \qquad N_p = \frac{T_{p+1} - T_p}{\delta T}, \qquad \bar{T}_{p,q} := \bar{T}_{p \cdot N_p + q} = \bar{T}_{p,q-1} + \delta T.$$

For simplicity, we assume that both time-step sizes δt and δT are uniform throughout $[0, T_{\text{end}}]$. In order to keep the cost of the coarse propagator as low as possible, we will mainly focus on the case that the coarse time steps coincide with the P sub-intervals I_p in the numerical results, i.e., $N_p = 1$, such that the total number of coarse time steps

$$N_c := P \cdot N_p$$

is equal to P. We denote the coarse propagation from T_p to T_{p+1} by

$$c_s^{\text{coarse}}(T_{p+1}) = \mathcal{C}(c_s^{\text{coarse}}(T_p)).$$

We use capital letters T_p to denote the coarse discretization of $[0, T_{\text{end}}]$ into P parts and for the time-steps $T_{p,q}$ on the coarse level. By small letters $t_{p,q}$ we denote the finer discretization on each sub-interval; the two discretizations yield the first level (fine problem) and second level (coarse problem) of the parareal algorithm. Of course, it is also possible to employ coarse time steps which differ from the sub-intervals on the fine level, but for the sake of simplicity, we will not discuss this case in this work. In the example given above with $T_{\text{end}} = 300$ days, δT would be 30 days for $P = N_c = 10$, while δt is 0.3 days. On the micro scale, the times τ_i and time-step size $\delta\tau$ are defined locally in $[t_i, t_i + 1s]$; in the example above, we had $\delta\tau = 0.02$s. Note that the micro scale influences the parareal algorithm only indirectly due to the temporal homogenization approach.

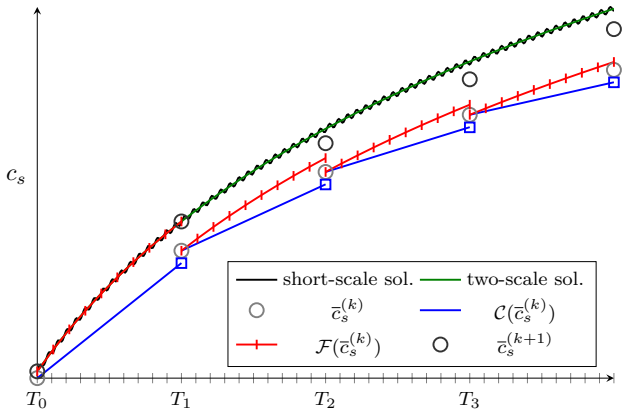


Figure 4: Schematic representation of the parareal algorithm applied to the homogenized (two-scale) problem as shown in fig. 2: micro-scale solution (black), two-scale solution (green), parareal iterates in iteration k (light grey) and $k+1$ (dark grey), fine scale solution (red), and coarse scale solution (blue).

Then, given an iterate $\overline{c}_s^{(k)}$ for some $k \leq 0$, the parareal algorithm computes $\overline{c}_s^{k+1}$ by setting

$$\overline{c}_s^{(k+1)}(T_{p+1}) = \mathcal{C}(\overline{c}_s^{(k+1)}(T_p)) + \mathcal{F}(\overline{c}_s^{(k)}(T_p)) - \mathcal{C}(\overline{c}_s^{(k)}(T_p)) \quad \text{for } p = 0, ..., P-1. \tag{18}$$

This can be seen as a predictor-corrector scheme, where the coarse predictor $\mathcal{C}(\overline{c}_s^{(k+1)}(T_p))$ is corrected by fine-scale contributions that depend only on the previous iterate $\overline{c}_s^{(k)}$; thus, the fine problems can be computed fully in parallel. A schematic illustration of the parareal algorithm is given in fig. 4.

Let us analyze the application of eq. (18) to the two-scale problem (algorithm 1) in more detail. The first term in eq. (18) requires the solution of one micro problem and an update of the foam cell concentration (by eq. (5) or eq. (11)) in each coarse time step $T_p \to T_{p+1}$. Within the fine-scale propagator (second term in eq. (18)) $n_p = \lceil N_f/P \rceil$ time-steps need to be computed per process, where $\lceil g \rceil$ denotes the next-biggest natural number to g ($\widehat{=}$ `ceil(g)`) and N_f is the number of macro time steps; see eq. (13). Each time step requires the solution of one micro problem and an update of the foam cell concentration. The last term in eq. (18) has already been computed in the previous iteration (compare the first term on the right-hand side of the same equation) and thus introduces no additional computational cost. The algorithm is summarized in algorithm 2.

We use the FSI variables $w^{(k)}(T_p)$ from the previous parareal iteration as initial values in step (II.a)(i). The initial values $\overline{c}_s^{(k)}(T_p)$ in step (II.a)(i) are taken from step (II.b) of the coarse problem. Moreover, to initialize the variables $\{\overline{c}_s^{(0)}(T_p), \mathbf{w}_s^{(0)}(T_p)\}_{p=1}^{P}$ before the first parareal iteration, we apply the coarse propagator once with a large time-step size $\delta T = (T_p - T_{p-1})/N_p$.

For the ODE growth model, we use the value of c_s^{fine} at the end time T_{end}

$$|c_s^{(k+1),\text{fine}}(T_{\text{end}}) - c_s^{(k),\text{fine}}(T_{\text{end}})| \leq \epsilon_{\text{par}}. \tag{19}$$

as the stopping criterion for the parareal algorithm.

4.1.1 Parallelization approaches

As mentioned earlier, in this work, we focus on studying the effectivity and efficiency of our approach by investigating the convergence of the parareal algorithm based on a serial implementation. Since the computational cost of the micro FSI problems is very high, we expect the communication cost to be negligible; cf. section 4.1.2 as well as sections 5.2 and 5.3 for a more detailed discussion of the computational cost for the ODE and the PDE growth model, respectively. Even though our implementation used in the numerical

Algorithm 2: Parareal algorithm

(I) **Initialization**: Compute $\left\{(\overline{c}_s^{(0)}(T_p), \mathbf{w}^{(0)}(T_p)\right\}_{p=1}^{P}$ by means of algorithm 1 with a coarse macro time-step size $\delta T := (T_p - T_{p-1})/N_p$. Set $k \leftarrow 0$

(II) **while** $\underline{|c_s^{(k+1),\text{fine}}(T_{\text{end}}) - c_s^{(k),\text{fine}}(T_{\text{end}})| > \epsilon_{\text{par}}}$ **do**

(II.a) **Fine problem**:
for $\underline{p = 1, ..., P}$ **do**

(i) Initialize $c_s^{(k+1),\text{fine}}(T_p) = \overline{c}_s^{(k)}(T_p)$, $\mathbf{w}^{(k+1)}(T_p) = \mathbf{w}^{(k)}(T_p)$

(ii) Compute $\left\{(c_s^{(k+1),\text{fine}}(t_{p,q}), \mathbf{w}^{(k+1)}(t_{p,q})\right\}_{q=1}^{n_p}$ by algorithm 1 with fine time-step size δt and set $\mathcal{F}(\overline{c}_s^{(k)}(T_p)) = c_s^{(k+1),\text{fine}}(t_{p,n_p})$

(II.b) **Coarse problem**
for $\underline{p = 1, ..., P}$ **do**

(i) Compute $\mathcal{C}(\overline{c}_s^{(k+1)}(T_p))$ by solving one time step of algorithm 1 with time-step size $\delta T = (T_p - T_{p-1})/N_p$

(ii) Parareal update

$$\overline{c}_s^{(k+1)}(T_{p+1}) = \mathcal{C}(\overline{c}_s^{(k+1)}(T_p)) + \mathcal{F}(\overline{c}_s^{(k)}(T_p)) - \mathcal{C}(\overline{c}_s^{(k)}(T_p)).$$

$k \leftarrow k + 1$

examples below is only serial, we want to briefly discuss two potential parallelization approaches for a parallel implementation of algorithm 2: *master-slave parallelization* and *distributed parallelization*; cf. fig. 5 for a sketch of both approaches for our parareal algorithm. Note that, as also indicated by fig. 5, we always assume a one-to-one correspondence of fine problems and processes in this paper.

In the master-slave parallelization approach, the coarse problem is assigned to a single process and the local problems on the fine level are distributed among the remaining processes. Each slave process p has to communicate $\mathcal{F}(\overline{c}_s^{(k)}(T_p))$ to the master after computation of the fine problem for the parareal update (step II.a in algorithm 2), and the master process has to communicate $\overline{c}_s^{(k+1)}(T_p)$ back after the parareal update (step II.b in algorithm 2); this corresponds to all-to-one and one-to-all communication patterns.

In the distributed parallelization approach, the parallelization of the coarse problem is different. In particular, time intervals are assigned to the different processes, and each process computes both the fine problem and the part of the coarse problem on those time intervals; for simplicity, we do not discuss the case that the partition of the coarse problem onto the processes is different from the local (fine) problems, which would also be possible. As can be seen in fig. 5, a major benefit of this approach is the different communication pattern: instead of all-to-one and one-to-all communication, each process sends the parareal update $\overline{c}_s^{(k+1)}(T_p)$ to the next process in line. Note that, in terms of the coarse problem, the processes can only perform their computations in serial, whereas the local problems can, again, be computed concurrently.

In some implementations of two-level methods, the coarse problem is assigned to one of the processes dealing with the local problems instead of an additional process; if the memory and computational cost of the coarse problem is negligible, this can be beneficial since all available processes can be employed for the local problems. It can be seen as a mixture of the master-slave and distributed parallelization approaches. In our case, this might not be feasible since the coarse problem is defined on the same spatial mesh as the local problems, leading to significant memory cost. Furthermore, more sophisticated approaches such as task-based scheduling [4] can be employed for the parallelization.

An investigation of the differences in performance of the different approaches can only be performed

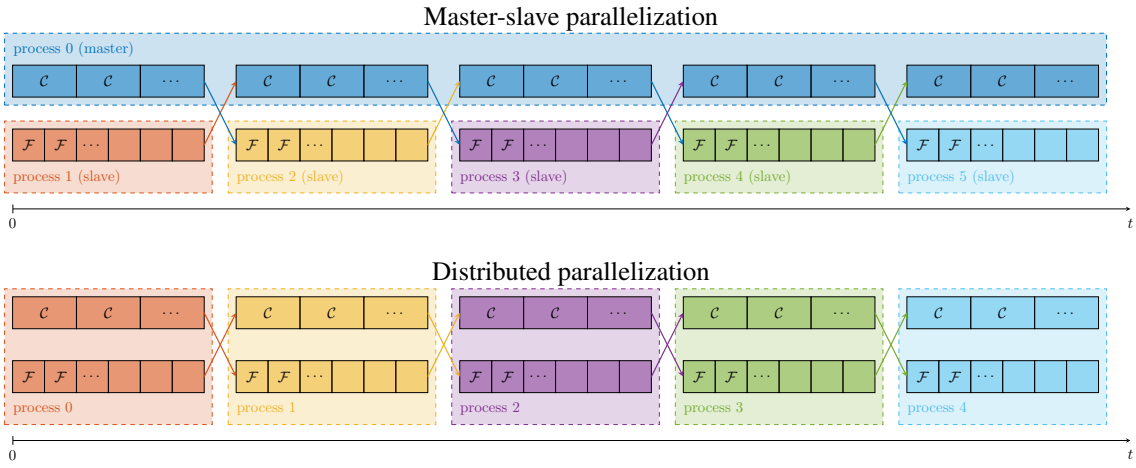


Figure 5: Master-slave and distributed parallelization schemes for the parareal approach. In the master-slave case, the coarse propagation is performed on a separate process (blue) and information is communicated between master and slave. In the distributed case, the coarse propagation and the fine propagation is performed on the same processes. Communication is indicated by an arrow in color of the sender.

using an actual parallel implementation. This is out of the scope of this paper, and therefore, we will leave it to future work.

4.1.2 Computational costs

Even in the simplified two-dimensional configuration considered in this work, it suffices to count the number of micro problems to be solved to estimate the computational cost of corresponding parallel computations and to compare different algorithms. Remember that, in each micro problem, $r_p \cdot \frac{1}{\delta t}$ FSI systems need to be solved, where r_p is the number of cycles required to reach a near-periodic state. In the example considered above, these were at least $2 \cdot \frac{1}{0.02} = 100$ costly FSI problems. Hence, our discussion in the following sections will be based on the assumption that the computational cost for all other steps of the parareal algorithm as well as the communication between the processes can be neglected. This is particularly obvious for the ODE growth model in eq. (5). For the PDE growth model in eq. (11), we will discuss the computational and communication cost in more detail in sections 5.2 and 5.3.

To analyze the computational cost of algorithm 2 under this assumption, let us denote the total number of iterations of the parareal algorithm by k_{par}. As discussed in sections 4.1 and 4.1.1, we assume that the number of coarse time steps is the same as the number of fine problems and processes, respectively. Therefore, we need to solve N_c micro problems in step I, $k_{\text{par}} \cdot \lceil N_f/P \rceil$ micro problems on each of the P processes in step II.a (ii), and $k_{\text{par}} \cdot N_c$ micro problems within the coarse propagation in step II.b (i). This corresponds to the solution of

$$\underbrace{k_{\text{par}} \cdot \lceil N_f/P \rceil}_{\text{fine level } (P \text{ parallel processes})} + \underbrace{(k_{\text{par}} + 1) \cdot N_c}_{\text{coarse level (1 serial process)}} \tag{20}$$

serial micro problems; this is the case for both parallelization approaches discussed in section 4.1.1. If the coarse time step δT is chosen independently of T, N_c is fixed and the computational cost tends to saturate for large P (at least if we assume that the number of required parareal iterations k_{par} is independent of P). The cheapest possible coarse propagator, on the other hand, is to use one coarse time step per coarse propagation, i.e., $N_p = 1$ and $N_c = P$. In this case the computational time increases for large P, if we assume that the number of parareal iterations k_{par} is independent of P. The minimum computational time is attained for $P \approx \sqrt{N_f}$.

As a serial implementation requires the solution of N_f micro problems, the speed-up of the proposed parareal algorithm is given by

$$\text{speedup}(P) = \frac{N_f}{k_{\text{par}} \cdot \lceil N_f/P \rceil + (k_{\text{par}}+1) \cdot N_c} \approx \frac{1}{\frac{k_{\text{par}}}{P} + (k_{\text{par}}+1)\frac{N_c}{N_f}}, \tag{21}$$

if we assume a perfect load balancing on the fine processes. This is a special case of the standard analysis for the speed-up of the parareal algorithm; see, e.g., [52].

The aim of this work is to show potential for the parareal algorithm as a parallel time-stepping method for plaque growth simulations. A final assessment of the parallelization capabilities can, of course, only be made based on computing times of an actual parallel implementation. We will leave this to future work.

4.1.3 Convergence theory

Let us briefly recap the standard convergence theory for the parareal algorithm; see, e.g., [32]. For this purpose, let us consider the simpler ODE model

$$\partial_t c_s = \gamma(c_s), \qquad c_s(0) = 0. \tag{22}$$

and its discretization by the explicit Euler method

$$c_s^n = c_s^{n-1} + \delta T\, \gamma(c_s^{n-1}), \qquad c_s^0 = 0. \tag{23}$$

We assume that the function γ and its derivative are Lipschitz continuous in c_s, i.e., we assume

$$|\gamma(c_s^1) - \gamma(c_s^2)| \le L|c_s^1 - c_s^2|, \quad \big|\frac{d}{dt}\gamma(c_s^1) - \frac{d}{dt}\gamma(c_s^2)\big| \le L|c_s^1 - c_s^2| \quad \text{for all } c_s^1, c_s^2 \in \mathbb{R} \tag{24}$$

for some $L > 0$. As in [32], we also assume for simplicity that the fine-scale propagator advances the ODE exactly, i.e.,

$$c_s(T_n) = \mathcal{F}(c_s(T_{n-1})) = c_s(T_{n-1}) + \int_{T_{n-1}}^{T_n} \gamma(c_s(s))\,\mathrm{d}s. \tag{25}$$

This is motivated by the fact that the time discretization error in the fine propagator is typically small compared to the coarse propagator. The following convergence result is shown by Gander & Hairer in [32]:

Lemma 1. *Let $c_s^* \in C^1(0, T_{end})$ be the exact solution of eq.* (22)*, and let $\{\bar{c}_s^{(k)}(T_n)\}_{n=1}^P$ be the k-th iterate of the parareal algorithm in eq.* (18) *using the forward Euler method in eq.* (23) *as the time discretization. Under the assumptions made above, it holds for the error $e_n^{(k)} = |\bar{c}_s^{(k)}(T_n) - c_s^*(T_n)|$ and $k \in \mathbb{N}_0, n \in \{1, \dots, P\}$ that*

$$e_n^{(k)} \le (cT_n)^{k+1}\, \beta^{n-k-1} \delta T^{k+1} \max_{t\in[0,T_{end}]} |\partial_t c_s^*(t)|. \tag{26}$$

with a constant $c > 0$ and $\beta = 1 + L\delta T$.

Remark 1. *The assumptions in eq.* (24) *cannot be easily verified for the ODE model in eq.* (5)*, as the fluid forces and $\bar{\gamma}(\sigma_f^{WS})$ depend in a highly nonlinear way on c_s. In the numerical examples given below, we will, however, observe that the convergence behavior is similar to the one shown in lemma 1.*

4.1.4 Numerical results

We use, again, the simple example described in section 3.2 with the ODE growth model in eq. (5) and set $\epsilon_{\text{par}} = 10^{-3}$. Results for $P = N_c = 10$ and $\delta t = 0.3$ days are shown in fig. 6, where the first 3 iterates are compared against a reference solution $(c_s^*, \mathbf{u}_s^*, \mathbf{v}_f^*)$, which was computed by a standard serial time-stepping

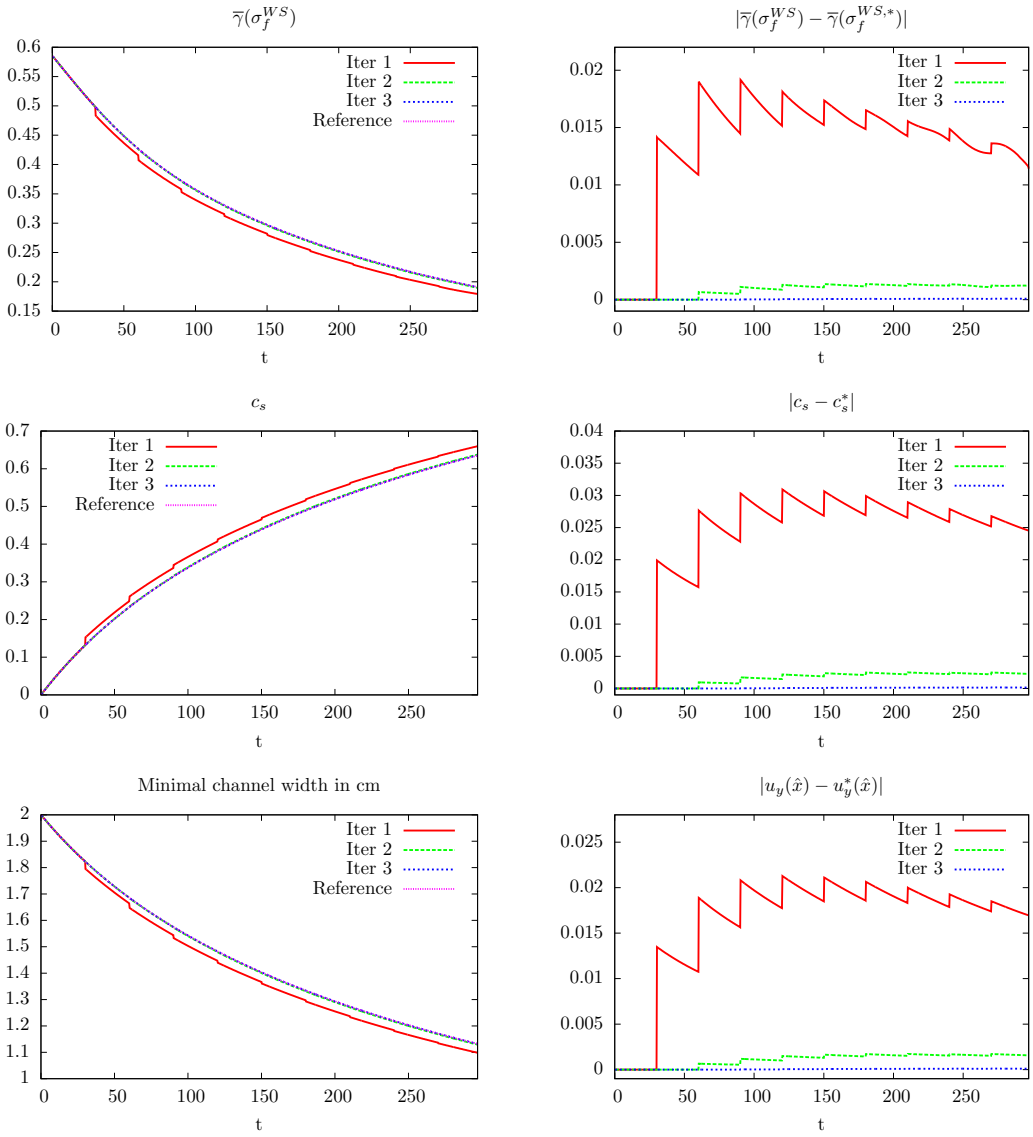


Figure 6: Behavior of the first three iterates of the parareal algorithm in the first numerical example (ODE growth) for $P = 10$. A description of the problem is given in section 4.1.4, numerical values in table 1. The iterates converge very quickly towards the reference solution. **Top**: Growth function and error w.r.t the reference solution $\boldsymbol{\sigma}_f^{WS}(\mathbf{v}_f^*)$ over time t. **Center**: $c_s^{(k),\text{fine}}$ and error w.r.t the reference solution c_s^* over time. **Bottom**: Channel width over time and error $|\mathbf{u}_{s,y}(\hat{x}) - \mathbf{u}_{s,y}^*(\hat{x})|$ w.r.t. the vertical component of the reference solution $\mathbf{u}_s^*$ at the narrowest point $\hat{x}$ over time .

scheme, as in section 3.2. We observe fast convergence towards the reference curve in all three quantities of interest. The stopping criterion in eq. (19) was satisfied after 4 iterations of the parareal algorithm.

In table 1, we show the deviations in the foam cell concentration $|c_s^{(k),\text{fine}}(T_\text{end}) - c_s^*(T_\text{end})|$ at final time T_end after each iteration of the parareal algorithm for $P = N_c = 10, 20, 30, 40$ and 50 processes. We observe that the number of iterations k_par decreases from four to three for $P \geq 20$. This is due to the fact that the coarse problem is solved with the smaller time-step size $\delta T = \lceil \frac{T_\text{end}}{P} \rceil$, which makes the coarse problem more expensive but also more accurate. In particular, we obtain a better approximation for the

k	$P = 10$	$P = 20$	$P = 30$	$P = 40$	$P = 50$	ref. (serial)
1	$2.21 \cdot 10^{-2}$	$1.19 \cdot 10^{-2}$	$8.12 \cdot 10^{-3}$	$5.83 \cdot 10^{-3}$	$5.46 \cdot 10^{-3}$	0
2	$2.24 \cdot 10^{-3}$	$5.63 \cdot 10^{-4}$	$2.62 \cdot 10^{-4}$	$1.37 \cdot 10^{-4}$	$8.61 \cdot 10^{-5}$	-
3	$1.42 \cdot 10^{-4}$	$2.02 \cdot 10^{-5}$	$6.53 \cdot 10^{-6}$	$2.32 \cdot 10^{-6}$	$8.30 \cdot 10^{-7}$	-
4	$5.76 \cdot 10^{-6}$	-	-	-	-	-
# mp	450	230	**222**	235	260	1 000
speedup	2.2	4.3	**4.5**	4.3	3.8	1.0
efficiency	**22 %**	**22 %**	15 %	11 %	8 %	100 %

Table 1: Errors $|c_s^{(k),\text{fine}}(T_\text{end}) - c_s^*(T_\text{end})|$ for $P = 10, \ldots, 50$ for the parareal algorithm (algorithm 2) in the first numerical example (ODE growth). For comparison, the reference value $c_s^*(T_\text{end}) = 0.63831273...$ resulting from a serial time-stepping with the same fine-scale time-step size δt is taken. The time measure in terms of the number of serial micro problems (# mp) as well as speedup and efficiency compared to the reference computation (right column) are shown; the stopping criterion $|c_s^{(k+1),\text{fine}}(T_\text{end}) - c_s^{(k),\text{fine}}(T_\text{end})| < \epsilon_\text{par} = 10^{-3}$ is used; best numbers marked in **bold face**.

k	$P = 10$	$P = 20$	$P = 30$	$P = 40$	$P = 50$	ref. (serial)
1	$2.59 \cdot 10^{-3}$	$5.97 \cdot 10^{-4}$	$3.29 \cdot 10^{-4}$	$1.42 \cdot 10^{-4}$	$8.95 \cdot 10^{-5}$	0
2	$2.81 \cdot 10^{-3}$	$2.04 \cdot 10^{-5}$	$6.59 \cdot 10^{-6}$	$2.36 \cdot 10^{-6}$	$8.43 \cdot 10^{-7}$	-
3	$6.37 \cdot 10^{-6}$	-	-	-	-	-
4	$1.73 \cdot 10^{-7}$	-	-	-	-	-
# mp	450	160	**158**	170	190	1 000
speedup	2.2	6.25	**6.3**	5.9	5.3	1.0
efficiency	22 %	**31 %**	21 %	15 %	11 %	100 %

Table 2: Errors $|\bar{c}_s^{(k)}(T_\text{end}) - c_s^*(T_\text{end})|$ for $P = 10, \ldots, 50$ for the parareal algorithm (algorithm 2) in the first numerical example (ODE growth). The time measure in terms of the number of serial micro problems (# mp) as well as speedup and efficiency compared to the reference computation (right column) are shown; the stopping criterion $|\bar{c}_s^{(k+1)}(T_\text{end}) - \bar{c}_s^{(k)}(T_\text{end})| < \epsilon_\text{par} = 10^{-3}$ is used; best numbers marked in **bold face**.

coarse values $\bar{c}_s^{(k)}(T_p)$ that are used as initial values in the fine problems in the next iterate. We can also observe that, for fixed P, the error decreases at least by a factor $\frac{1}{P}$ in each parareal iteration. This is in agreement with lemma 1, which predicts (for a simpler model problem) a reduction factor $c \cdot \delta T = c \cdot \lceil \frac{T_\text{end}}{P} \rceil$ for some constant $c > 0$, compared to the previous iterate.

Since the number of parareal iterations k_par is constant for all cases with $P \geq 20$, we get the lowest computational cost for $P = 30$, which is close to $\sqrt{N_f} \approx 31.6$; cf. the discussion in section 4.1.2. For $P = 30$, $3 \cdot 34 = 102$ micro problems need to solved on each process (step II.a) and $4 \cdot 30 = 120$ micro problems within the coarse propagators (steps I and II.b), i.e., 222 micro problems in total. Compared to a serial time-stepping scheme with the same fine-scale time-step size, which requires the solution of $N_f = 1\,000$ serial micro problems, this corresponds to a speed-up of $\frac{1\,000}{222} \approx 4.5$; see also eq. (21). As is usual for two-level methods, for larger numbers of processes (P), the cost of the coarse problems gets dominant. For $P = 40$, we solve, for example, $3 \cdot 25 = 75$ micro problems in the fine propagator compared to $4 \cdot 40 = 160$ micro problems for the coarse propagator; this results in a total of 235 micro problems and a speedup of 4.3.

In order to relate the speed-up to the computational resources used, we define the computational efficiency by the ratio $\dfrac{\text{speedup}}{P}$ as a second measure to quantify the results. By definition, the serial computation always has the best efficiency, as it converges within one sweep through all time steps. Even if the parareal algorithm would also converge within one iteration, the parallel computations come with additional costs since they require the solution of the coarse problem.

In fact, using the serial computation as a reference in the comparison in table 1, the efficiency is best

(22 %) for $P = 10$ or $P = 20$; it deteriorates for larger numbers of processes due to the increasing cost of the coarse problems. The low efficiency is a typical observation for parallel time integration methods, such as the parareal method and can be explained as follows: For $P = 10$, for instance, 4 parareal iterations are required and each iteration requires the solution of all 1 000 micro problems on the fine scale on the full time interval $[0, T_{\text{end}}]$. Hence, a total of 4 000 micro problems has to be solved. Assuming that the time to solve a micro problem is approximately constant, we can not expect efficiencies above 25%. Due to the additional cost of the coarse propagator, the efficiency is reduced to 22%. However, since we can always compute 10 micro problems on the fine scale at the same time, we can still obtain a speedup of 2.2.

To increase the efficiency, we could take the value $\overline{c}_s^{(k)}(T_{\text{end}})$ computed in eq. (18) instead of $c_s^{(k),\text{fine}}(T_{\text{end}})$ in the stopping criterion

$$|\overline{c}_s^{(k+1)}(T_{\text{end}}) - \overline{c}_s^{(k)}(T_{\text{end}})| < \epsilon_{\text{par}}. \tag{27}$$

In particular, at the end time T_{end} we expect that this has a higher accuracy compared to the fine-scale variable $c_s^{(k),\text{fine}}(T_{\text{end}})$. However, the values $\overline{c}_s^{(k+1)}(t_i)$ are only available at the coarse grid points $t_i \in \{T_p\}_{p=1}^{P}$. If one is interested in foam cell concentrations at intermediate time steps, a stopping criterion based on $c_s^{(k),\text{fine}}(t_i)$ needs to be used, as in eq. (19).

The errors concerning $\overline{c}_s^{(k)}(T)$ and the number of micro problems to be solved using the stopping criterion eq. (27) are given in table 2. We observe that for $P \geq 20$ this stopping criterion was satisfied already after $k_{\text{par}} = 2$ iterations. For $P = 30$ we get again the lowest computational cost with a speed-up of around 6.33 compared to a serial computation. Furthermore, the best efficiency is obtained here with $P = 20$, which is due to the lower iteration count compared to the case $P = 10$.

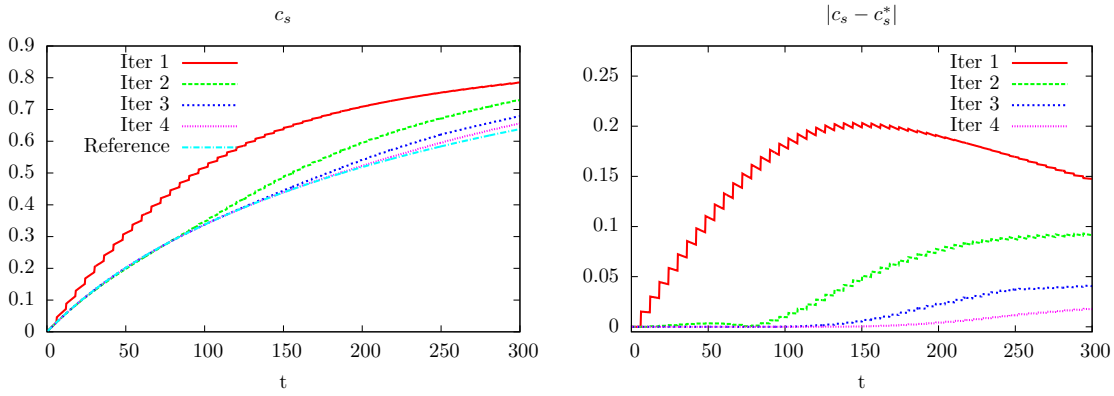
4.2 Variants with cheaper coarse-scale computations

In the parareal algorithm introduced above, a further improvement of the computational cost is not possible for $P \gtrsim \sqrt{N_f}$ processes in the case $P = N_c$ due to the (increasing) cost of the coarse-scale propagators. In fact, these get dominant compared to the fine-scale contributions for $P \gtrsim \sqrt{N_f}$. In this section, we will discuss approximate coarse-scale propagators, where no additional computations of micro problems are needed; hence, they are computationally cheaper.

4.2.1 Heuristic averaging of the FSI problems

As a first variant, we use the heuristic averaging approach mentioned in the beginning of section 3.1 for the coarse-scale propagation in steps (I) and (II.b)(i) of algorithm 2. The solution of the stationary FSI problem in variational formulation 2 is much cheaper compared to ≥ 100 time steps of a non-stationary FSI problem (approx. by a factor 100), and thus, its computational cost is neglected in the following discussion. Hence, the computational cost is reduced from $k_{\text{par}} \cdot \lceil N_f/P \rceil + (k_{\text{par}} + 1) \cdot P$ to $k_{\text{par}} \cdot \lceil N_f/P \rceil$ micro problems. In fig. 7, we illustrate the error in the foam cell concentration for $P = 50$ (left), and give the error $|c_s^{(k),\text{fine}}(T_{\text{end}}) - c_s^*(T_{\text{end}})|$ in each iteration k for $P = 30$ and $P = 50$ (right). We observe a much slower convergence compared to the standard parareal algorithm used above. This could already be expected from the results in fig. 3, where we saw that the heuristic averaging is not a good approximation. Using the stationary problem for the coarse propagator, 8 parareal iterations are necessary until the stopping criterion is satisfied for $P = 30$ and $P = 50$ compared to 3 iterations in the standard parareal algorithm.

In terms of the computating time, $8 \cdot 34 = 272$ and $8 \cdot 20 = 160$ micro problems are necessary for the cases $P = 30$ and $P = 50$, respectively. While for $P = 30$ this is worse compared to the standard parareal algorithm, this is an improvement of 67% for $P = 50$; cf. table 1. As for the standard parareal algorithm, convergence is obtained faster if we consider the stopping criterion in eq. (27) based on the parareal iterates $\overline{c}_s^{(k)}(T_{\text{end}})$; however, 7 iterations are still needed. The resulting computational cost is $7 \cdot 34 = 238$ micro problems for $P = 30$ and $7 \cdot 20 = 140$ micro problems for $P = 50$. Compared to 158 micro problems for $P = 30$ and 190 micro problems for $P = 50$ in the standard parareal algorithm (cf. table 2), we see only a small but no significant improvement. We will thus consider another, more promising approach to compute the required growth values $\gamma(\boldsymbol{\sigma}_f^{WS})$ for the coarse propagator in the following subsection.



k	$P = 30$	$P = 50$	ref. (serial)
1	$1.66 \cdot 10^{-1}$	$1.42 \cdot 10^{-1}$	0
2	$9.34 \cdot 10^{-2}$	$9.34 \cdot 10^{-2}$	-
3	$4.86 \cdot 10^{-2}$	$4.06 \cdot 10^{-2}$	-
4	$1.84 \cdot 10^{-2}$	$1.76 \cdot 10^{-2}$	-
5	$5.41 \cdot 10^{-3}$	$5.71 \cdot 10^{-3}$	-
6	$1.28 \cdot 10^{-3}$	$1.47 \cdot 10^{-3}$	-
7	$2.47 \cdot 10^{-4}$	$3.14 \cdot 10^{-4}$	-
8	$4.79 \cdot 10^{-5}$	$6.48 \cdot 10^{-5}$	-
# mp	272	**160**	1 000
speedup	3.7	**6.3**	1.0
eff.	12 %	**13 %**	100 %

Figure 7: Behavior of the iterates of the parareal algorithm, when the stationary FSI problem (variational formulation 2) is used in Step (I) and (II.b)(i) of algorithm 2. A description of the computation is given in section 4.2. Convergence to the reference solution is much slower compared to standard parareal (see fig. 6). **Top**: Evolution of $c_s^{(k),\text{fine}}(t)$ and error $|c_s^{(k),\text{fine}}(t) - c_s^*(t)|$ of the foam cell concentration over time t for $P = 50$ and $k = 1, 2, 3, 4$. **Right**: Errors $|c_s^{(k),\text{fine}}(T_{\text{end}}) - c_s^*(T_{\text{end}})|$ after each iteration for $P = 30$ and $P = 50$; total number of micro problems (# mp) solved in parallel as well as speedup and efficiency compared to the reference computation (right column); best numbers marked in **bold face**.

Remark 2. *As mentioned before, parallelization in space is generally more efficient than parallelization in time. Hence, complex three-dimensional FSI simulations already have a significant potential for parallelization. Additionally using temporal parallelization with large values of P, such as $P > 50$, may only be reasonable on very large supercomputers.*

4.2.2 Re-usage of computed growth values

As a second approach we will consider the re-use of growth values $\overline{\gamma}(\boldsymbol{\sigma}_f^{WS}(t_{p,i}))$ computed in the fine-scale propagator on the coarse scale. Therefore, we store all values $\overline{\gamma}_{p \cdot n_p + i} := \overline{\gamma}(\boldsymbol{\sigma}_f^{WS}(t_{p,i}))$ computed on the fine scale on all processes $p = 1, ..., P$ for all time-steps $i = 1, ..., n_p$; see step 1.b of algorithm 1. These can be used in the coarse propagator (step II.b) in the same parareal iteration instead of computing new growth values there.

We introduce the following notations for the coarse resp. fine propagators that start from $c_{s,n-1}$ using certain growth values $\gamma(T_{n-1})$ (that might now differ from $\overline{\gamma}(\sigma_f^{\text{WS}}(c_{s,n-1}))$):

$$c_{s,n}^{\text{coarse}} = \mathcal{C}(I_n, c_{s,n-1}, \gamma(T_{n-1})), \quad c_{s,n}^{\text{fine}} = \mathcal{F}(I_n, c_{s,n-1}, \gamma(T_{n-1})),$$

where $I_n = [t_{n-1}, t_n]$. After an initialization step, the modified parareal iteration is defined by the following

Algorithm 3: Parallel Time-Stepping with Re-Usage of Growth Values

(I) **Initialization**: Compute $\left\{(\overline{c}_s^{(0)}(T_p), \overline{\mathbf{w}}^{(0)}(T_p)\right\}_{p=1}^{P}$ by means of algorithm 1 with a coarse macro time-step size $\delta T = T_{p+1} - T_p$ on the master process. Set $k \leftarrow 0$

(II) **while** $|c_s^{(k+1),\text{fine}}(T_{\text{end}}) - c_s^{(k),\text{fine}}(T_{\text{end}})| > \epsilon_{\text{par}}$ **do**

(II.a) **Fine problem**:
for $p = 1, ..., P$ **do**

(i) Initialize $c_s^{(k+1),\text{fine}}(T_p) = \overline{c}_s^{(k)}(T_p)$, $\mathbf{w}^{(k+1)}(T_p) = \mathbf{w}^{(k)}(t_{p-1,N_p}))$

(ii) Compute $\left\{(c_s^{(k+1),\text{fine}}(t_{p,q}), \mathbf{w}^{(k+1)}(t_{p,q})\right\}_{q=1}^{n_p}$ by algorithm 1 with fine time step δt

(iii) Store the resulting growth functions $\overline{\gamma}_{p \cdot n_p + q} := \overline{\gamma}(\boldsymbol{\sigma}_f^{WS})(t_{p,q})$ at all fine points $t_{p,q}$, $q = 1, ..., n_p$ as well as $\mathbf{w}(t_{p,n_p})$ at the last time-step.

(II.b) **Coarse problem**:

(i) **for** $j = 1, ..., N_f$ **do**
Compute $\overline{c}_s^{(k+1)}(t_j)$ by advancing the ODE eq. (5) resp. solving the PDE in eq. (11) using the growth values $\overline{\gamma}_j$ computed in (II.a).

formula for $n = 1, ..., P$ and $k \geq 1$:

$$\overline{c}_{s,n}^{(k)} = \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k)}, \bar{\gamma}(c_{s,n-1}^{(k),\text{fine}})) + \mathcal{F}(I_n, \overline{c}_{s,n-1}^{(k-1)}, \bar{\gamma}(\overline{c}_{s,n-1}^{(k-1)})) - \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k-1)}, \bar{\gamma}(c_{s,n-1}^{(k-1),\text{fine}})), \tag{28}$$

where

$$c_{s,n-1}^{(k),\text{fine}} := \mathcal{F}(I_{n-1}, \overline{c}_{s,n-2}^{(k-1)}, \bar{\gamma}(\overline{c}_{s,n-2}^{(k-1)})), \tag{29}$$

and $c_{s,n}^{(0),\text{fine}} = \overline{c}_{s,n}^{(0)}$. Moreover, we set $c_{s,0}^{(k),\text{fine}} = \overline{c}_{s,0}^{(k)} = 0$ for all k.

As no new micro problems need to be solved and approximations of the growth values $\overline{\gamma}_j$ are available for all fine time steps $j = 1, ..., N_f$, the coarse propagator can now even use the fine-scale time-step δt. The only additional cost is to advance the foam cell concentration by eq. (5) resp. eq. (18) on the coarse level. This cost is clearly negligible for the ODE growth model in eq. (5). The additional cost in case of the PDE model eq. (18) will be discussed in section 5. The resulting algorithm is given as algorithm 3.

The growth values employed for the re-usage are available for all fine-scale time steps, and hence, we are able to perform the coarse propagation efficiently on the fine scale; cf. the discussion on the computational cost later in this section as well as for the PDE growth model in section 5.3. Nonetheless, the expected accuracy of the re-usage approach is still lower compared to the standard parareal algorithm. This is because the growth values on the fine scale have been computed using $c_{s,n-1}^{(k),\text{fine}}$, which depends on the previous iterate $\overline{c}_{s,n-2}^{(k-1)}$ (see eq. (29)), whereas the coarse propagator in the standard parareal algorithm already uses the more accurate $\overline{c}_{s,n-1}^{(k)}$ from the current iteration.

Computational costs Only the coarse step in the initialization (I) has to be carried out without re-usage and comes with a computational cost of N_c micro problems. On the other hand, step (II.b) does not require the solution of any micro problems. In terms of micro problems to be solved, the computational cost of step (II.a) is the same as in the standard parareal algorithm (algorithm 2). Altogether, the number of micro problems to be solved in k_{par} iterations of algorithm 3 is

$$k_{\text{par}} \cdot \lceil N_f / P \rceil + N_c.$$

Again, we obtain a saturation of the cost for large P, but the cost for large P is by a factor k_{par} smaller. Moreover, for $N_c = P$, the choice $P \approx \sqrt{k_{\text{par}} N_f}$ would be optimal if the number of iterations k_{par} was independent of the number of processes P.

The speed-up compared to a serial computation is given by

$$\text{speedup}(P) = \frac{N_f}{k_{\text{par}} \cdot \lceil N_f/P \rceil + N_c} \approx \frac{1}{\frac{k_{\text{par}}}{P} + \frac{N_c}{N_f}}, \tag{30}$$

assuming, again, a perfect load balancing of the micro problems.

The communication cost of the approach strongly depends on the employed parallelization scheme; cf. section 4.1.1. In the distributed parallelization approach, the growth values $\overline{\gamma}(\boldsymbol{\sigma}_f^{WS}(t_{p,i}))$, which are needed in the coarse propagator, do not have to be communicated. This is because the same intervals in fine and coarse problems are computed on the processes. In the master-slave approach, however, the values have to be communicated from each slave process to the master. Using the ODE growth model, these are N_f scalar values in total (in our example $N_f = 1\,000$). Even for the master-slave communication scheme, it is thus reasonable to assume that the computational cost of communication is still negligible compared to the solutions of the micro problems. In the case of a PDE growth model, the additional cost for communication will be discussed in section 5.

Theoretical convergence analysis We extend the convergence results discussed in section 4.1.3 for the model problem in eq. (22) for the re-usage algorithm in eq. (28) under the assumptions made in section 4.1.3.

The following result is a direct consequence of lemma 3, which is shown in the appendix.

Theorem 1. *Let $c_s^* \in C^1(0, T_{end})$ be the exact solution of eq.* (22) *and let $\{\overline{c}_s^{(k)}(T_n)\}_{n=1}^P$ be the k-th iterate of the re-usage algorithm in eq.* (28) *using the forward Euler method in eq.* (23)*. Under the assumptions made in section 4.1.3, it holds for the error $e_n^{(k)} = |\overline{c}_s^{(k)}(T_n) - c_s^*(T_n)|$ and $k \in \mathbb{N}_0, n \in \{1, \dots, P\}$ that*

$$e_n^{(k)} \leq L\delta T \max\left\{1, T_{n-\frac{k}{2}}\right\}^k \frac{c^k \beta^{n-k}}{\lceil k/2 \rceil!} \max_{t \in [0, T_{end}]} |\partial_t c_s^*(t)|. \tag{31}$$

with a constant $c > 0$ (see lemma 3) and $\beta = 1 + L\delta T$. This implies $e_n^{(k)} \to 0$ for $k \to \infty$.

The convergence $e_n^{(k)} \to 0$ for $k \to \infty$ follows due to the faculty $\lceil k/2 \rceil!$ in the denominator. Compared to lemma 1, the estimated convergence in k is much slower. However, we will observe in the numerical examples below that a good accuracy might still be reached within few iterations.

Numerical results In fig. 8, the evolution of the error in the concentration variable $c_s^{(k)}$ is plotted over time for $P = N_c = 10$ and $P = N_c = 60$. We observe convergence to the reference values c_s^* in both cases. Compared to the results in fig. 6 for the classical parareal algorithm (algorithm 2), the convergence is significantly slower. Nevertheless, the stopping criterion is already fulfilled after 4-5 iterations (see table 3), which is only 1-2 iterations more than in the standard parareal algorithm (see table 1).

While we had observed reduction factors in the order of $\delta T = \lceil \frac{T_{\text{end}}}{P} \rceil$ between two consecutive parareal iterations for the standard parareal algorithm in table 1, the reduction factors are slightly worse in table 3. They are, however, in all cases (besides the last value for $P = 20$) bounded above by $\frac{1}{k}$, where k is the parareal iterate. This indicates a convergence of order $\mathcal{O}(\frac{1}{k!})$, which is related to the factor $\mathcal{O}(\frac{1}{\lceil \frac{k}{2} \rceil!})$ in eq. (31) in theorem 1. Moreover, the absolute values of the error in each parareal iteration are significantly smaller for $P = 60$ compared to the case of $P = 10$ (Note the different scaling in the vertical axis).

In table 3, we compare the number of micro problems needed for $P = N_c = 10$ to 70. The optimal number of processes is $P = \sqrt{k_{\text{par}} N_f} \approx 63.25$, which is twice as many processes compared to the parareal algorithm in the previous section. The minimal cost in table 3 is 128 micro problems for $P = 60$, compared to 222 for standard parareal. Compared to a serial time-stepping scheme, we obtain a maximum estimated speed-up of 7.8. Moreover, the efficiency is also significantly improved for larger numbers of processes compared to the standard approach; cf. tables 1 and 3.

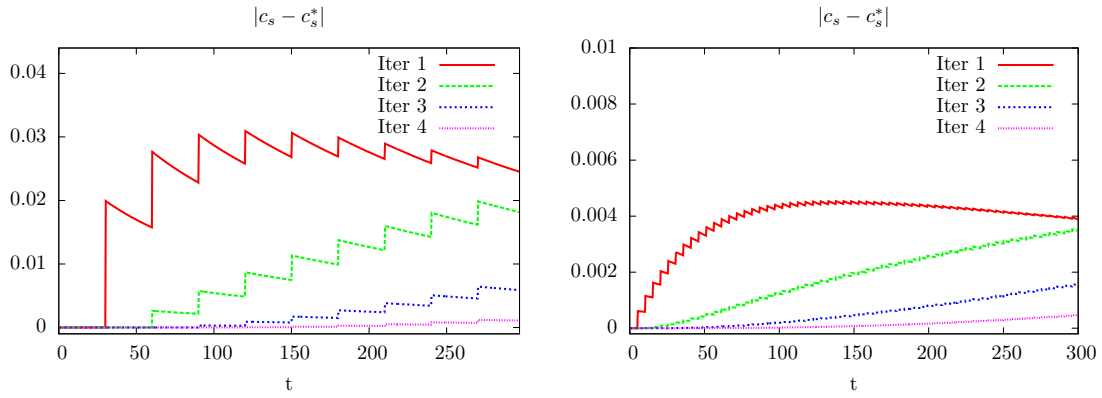


Figure 8: Error in the concentration $|c_s^{(k),\text{fine}}(t) - c_s^*(t)|$ over time t in the first 3 iterates of the parareal algorithm with re-usage of growth values (algorithm 3) for $P = 10$ (left) and $P = 60$ (right) and the first numerical example (ODE growth model).

k	$P = 10$	$P = 20$	$P = 30$	$P = 40$	$P = 50$	$P = 60$	$P = 70$	ref.
1	$2.56 \cdot 10^{-2}$	$1.10 \cdot 10^{-2}$	$7.48 \cdot 10^{-3}$	$5.46 \cdot 10^{-3}$	$4.23 \cdot 10^{-3}$	$3.63 \cdot 10^{-3}$	$3.12 \cdot 10^{-3}$	0
2	$7.78 \cdot 10^{-3}$	$4.48 \cdot 10^{-3}$	$3.16 \cdot 10^{-3}$	$2.36 \cdot 10^{-3}$	$1.99 \cdot 10^{-3}$	$1.61 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$	-
3	$1.73 \cdot 10^{-3}$	$1.20 \cdot 10^{-3}$	$9.02 \cdot 10^{-4}$	$6.92 \cdot 10^{-4}$	$6.34 \cdot 10^{-4}$	$5.14 \cdot 10^{-4}$	$4.32 \cdot 10^{-4}$	-
4	$2.32 \cdot 10^{-4}$	$2.28 \cdot 10^{-4}$	$1.87 \cdot 10^{-4}$	$1.31 \cdot 10^{-4}$	$1.26 \cdot 10^{-4}$	$1.10 \cdot 10^{-4}$	$9.98 \cdot 10^{-5}$	-
5	$2.19 \cdot 10^{-5}$	$4.71 \cdot 10^{-5}$	$3.05 \cdot 10^{-5}$	-	-	-	-	-
# mp	510	270	200	140	130	**128**	130	1 000
speedup	2.0	3.7	5.0	7.1	7.7	**7.8**	7.7	1.0
eff.	**20 %**	19 %	17 %	18 %	15 %	13 %	11 %	100 %

Table 3: Errors $|c_s^{(k),\text{fine}}(T_\text{end}) - c_s^*(T_\text{end})|$ for $P = 10, \ldots, 70$ for the parareal algorithm 3 (re-use of growth values) and the first numerical example (ODE growth model). Value $c_s^*(T_\text{end})$ of a serial reference computation for comparison. The time measure in terms of the number of micro problems (# mp) as well as speedup and efficiency compared to the reference computation (right column) are shown; the stopping criterion $|c_s^{(k+1),\text{fine}}(T_\text{end}) - c_s^{(k),\text{fine}}(T_\text{end})| < \epsilon_\text{par} = 10^{-3}$ is used; best numbers marked in **bold face**.

5 Plaque growth problem with a distributed foam cell concentration

We consider the PDE growth model introduced in section 2.2.2, where the concentration of foam cells $\hat{c}_s = \hat{c}_s(\hat{x}, t)$ is now governed by the non-stationary reaction-diffusion problem in eq. (11). For time discretization, we use a linearized variant of the backward Euler scheme. Starting from $c_{s,0} = 0$, the standard backward Euler scheme writes for $l = 1, ..., N_f$:
Find $c_{s,l+1} \in \mathcal{C} := H_0^1(\hat{\mathcal{S}}, \hat{\Gamma}_s)$ *such that*

$$\frac{1}{\delta t}\left(\hat{c}_{s,l+1} - \hat{c}_{s,l}, \hat{\varphi}\right)_{\hat{\mathcal{S}}} - \left(D_s \hat{\nabla} \hat{c}_{s,l+1}, \hat{\nabla}\varphi\right)_{\hat{\mathcal{S}}} + R_s\left(\hat{c}_{s,l+1}(1 - \hat{c}_{s,l+1}), \hat{\varphi}\right)_{\hat{\mathcal{S}}} + \left(\overline{\gamma}(\boldsymbol{\sigma}_f^{WS}), \hat{\varphi}\right)_{\hat{\Gamma}} = 0 \quad \forall \hat{\varphi} \in \mathcal{C}. \tag{32}$$

Since eq. (32) is a nonlinear system of equations, several iterations of a nonlinear solver (e.g., Newton's method) would be necessary to solve it. Thus, we consider the following linearized variant, which can be

k	$P = 10$	$P = 20$	$P = 30$	$P = 40$	$P = 50$	ref. (serial)
1	$2.80 \cdot 10^{-3}$	$1.42 \cdot 10^{-3}$	$7.49 \cdot 10^{-4}$	$6.53 \cdot 10^{-4}$	$5.04 \cdot 10^{-4}$	0
2	$8.73 \cdot 10^{-5}$	$9.46 \cdot 10^{-5}$	$1.27 \cdot 10^{-4}$	$6.22 \cdot 10^{-5}$	$4.78 \cdot 10^{-5}$	
3	$3.45 \cdot 10^{-5}$	$2.90 \cdot 10^{-5}$	$2.97 \cdot 10^{-5}$	$1.36 \cdot 10^{-5}$	$9.19 \cdot 10^{-6}$	
4	-	$1.93 \cdot 10^{-6}$	$4.45 \cdot 10^{-6}$	-	-	
Est. par.	11 347 s	9 661 s	8 692 s	**6 914 s**	7 491 s	26 840 s
speedup	2.4	2.8	3.1	**3.9**	3.6	1.0
efficiency	**24** %	14 %	10 %	10 %	7 %	100 %

Table 4: Errors $|\hat{c}_s^{(k),\text{fine}}(T_{\text{end}}) - \hat{c}_s^*(T_{\text{end}})|$ at the midpoint $\hat{x}_m$ of $\hat{\Gamma}$ for $P = 10$ to 50 for the parareal algorithm (algorithm 2) and a serial reference computation applied to the second numerical example (PDE growth model, $\hat{c}_s^*(T_{\text{end}}) = 0.5186632$). Estimated parallel runtimes are shown, as well as speedup and efficiency compared to the reference computation (right column). Details on the estimation of the runtimes are given in section 5.3. Best numbers are marked in **bold face**.

seen as an implicit-explicit (IMEX) scheme (see, e.g., [3])

$$\begin{aligned}\frac{1}{\delta t}\left(\hat{c}_{s,l+1} - \hat{c}_{s,l}, \hat{\varphi}\right)_{\hat{\mathcal{S}}} + \left(D_s \hat{\nabla} \hat{c}_{s,l+1}, \hat{\nabla}\varphi\right)_{\hat{\mathcal{S}}} + R_s \theta \left(\hat{c}_{s,l+1}(1-\hat{c}_{s,l}), \hat{\varphi}\right)_{\hat{\mathcal{S}}} \\ + R_s(1-\theta)\left(\hat{c}_{s,l}(1-\hat{c}_{s,l+1}), \hat{\varphi}\right)_{\hat{\mathcal{S}}} + \left(\overline{\gamma}(\boldsymbol{\sigma}_f^{WS}), \hat{\varphi}\right)_{\hat{\Gamma}} = 0,\end{aligned} \tag{33}$$

where $\theta \in [0,1]$ is a weighting parameter. In our numerical experiments, we choose $\theta = 0.7$. For spatial discretization, we use again Q_2 finite elements on the mesh described in section 3.2. If we assume that evolution of the wall shear stress $\boldsymbol{\sigma}_f^{WS}$ was given exactly, the following bound could be shown for the discretization error

$$\|\hat{c}_s(T_{\text{end}}) - \hat{c}_s^{N_f}\|_{\hat{\mathcal{S}}} + \left(\sum_{l=1}^{N_f} \delta T \|\hat{\nabla}(\hat{c}_s(T_l) - \hat{c}_{s,l})\|_{\hat{\mathcal{S}}}^2\right)^{1/2} \leq c_1 h^2 + c_2 \delta T. \tag{34}$$

Due to the nonlinear interaction between FSI and reaction-diffusion equation, an analysis of the discretization error of the coupled problem is more involved and not within the scope of this paper. The chosen time-step δT is $\frac{1}{1000}$ times the macro time interval length T_{end}, while the horizontal cell size h is a factor $\frac{1}{20}$ of the length of the channel in horizontal direction. Thus, the errors in eq. (34) should be roughly equilibrated.

The material parameters in the fluid and solid problems are chosen as in the first example. The parameters of the PDE growth model are set to $D_s = 1.2 \cdot 10^{-7} \frac{\text{cm}^2}{s}, R_s = 5 \cdot 10^{-7} \frac{1}{\text{s}}, \alpha = 5 \cdot 10^{-8} \frac{\text{cm}}{\text{s}}$, and $\sigma_0 = 30 \frac{\text{g}}{\text{cm s}^2}$. The inflow profile of velocity is chosen as

$$\mathbf{v}_f^{\text{in}}(t,x) = 30 \begin{pmatrix} (1 + \sin(\pi \frac{t}{\mathcal{P}})^2)(1 - x_2^2) \\ 0 \end{pmatrix} \text{cm/s},$$

where $\mathcal{P} = 1\text{s}$. The end time is $T_{\text{end}} = 200$ days, and for the fine time steps, we use $\delta t = 0.2$ days, such that again $N_f = 1\,000$. In fig. 9, we visualize the results for horizontal velocity v_x and vertical displacement u_y in the deformed fluid and solid domains in three different time instants, respectively. In fig. 10 the foam cell concentration c_s on the FSI interface is shown for different times t. First, we observe a dominant growth in the center of the domain due to the penetration of monocytes and the reaction term in eq. (11). After $t > 100$ the diffusion gets more significant, such that foam cells are distributed over the whole interface.

In the following section 5.1 we investigate numerically the convergence behavior of the standard parareal algorithm (algorithm 2) and the re-usage variant (algorithm 3). Then, we discuss the computational costs of the algorithms applied to the PDE growth model in section 5.2. Finally, we compare the algorithms in section 5.3 based on estimated runtimes of a parallel implementation.

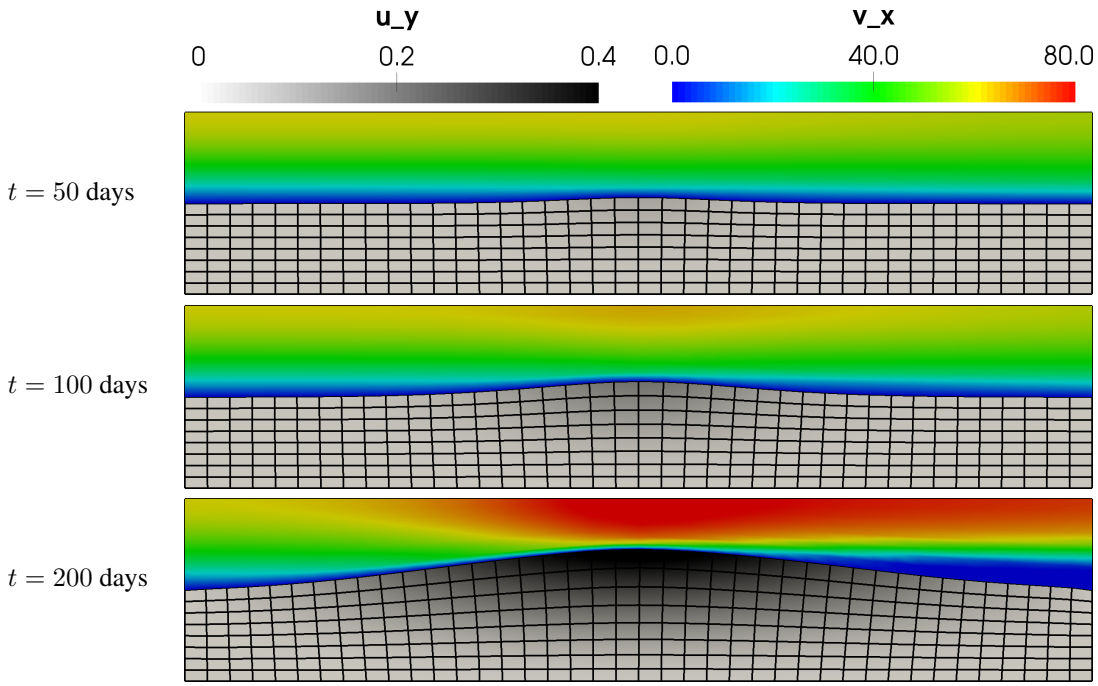


Figure 9: Visualization of the plaque growth at times $t = 50$ days, $t = 100$ days and $t = 200$ days for the second numerical example (PDE growth model). The horizontal velocity (in cm/s) and the vertical displacement (in cm) are shown on the deformed domain at micro time $\tau = 0.5$ s, i.e., the time of maximum inflow velocity. As the plaque growth evolves, significantly higher velocities arise in the central part.

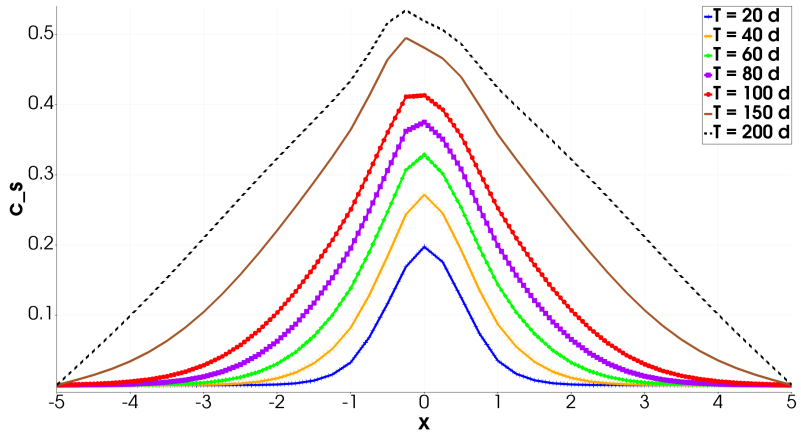


Figure 10: Foam cell concentration $\hat{c}_s$ at the FSI interface $\hat{\Gamma}$ at different times t in the second numerical example (PDE growth model). In contrast to the ODE model, the concentration is not symmetric around the center anymore for larger times t; this is due to the reaction-diffusion equation.

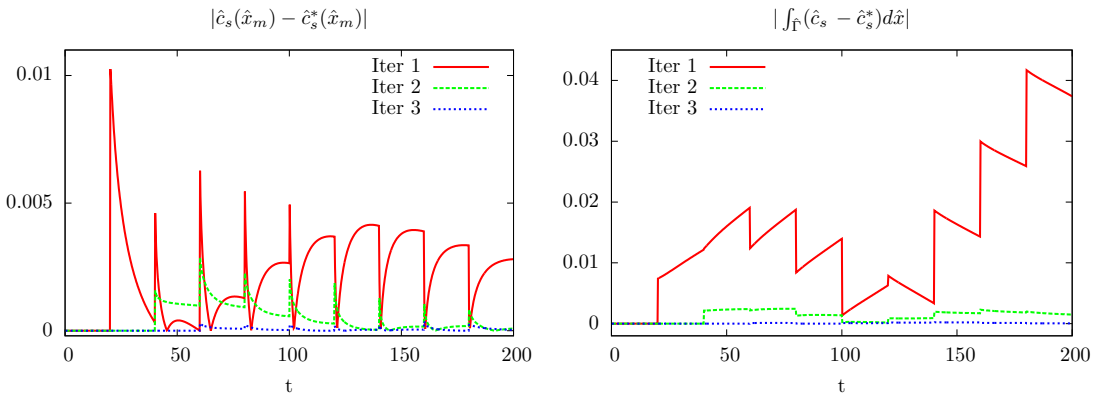


Figure 11: Errors for the second numerical example (PDE growth model) in different iterations of the parareal algorithm for $P = 10$; the setting is described in section 5. **Left**: Error in the concentration $|\hat{c}_s^{(k),\text{fine}}(\hat{x}_m, t) - \hat{c}_s^*(\hat{x}_m, t)|$ at the midpoint of $\hat{\Gamma}$ over time; see also fig. 10 for the temporal evolution of $\hat{c}_s$ at the interface. **Right**: Error in the mean value $\left| \int_{\hat{\Gamma}} \hat{c}_s^{(k),\text{fine}}(\hat{x}, t) - \hat{c}_s^*(\hat{x}, t)\, \mathrm{d}\hat{x} \right|$ over time.

5.1 Convergence behavior of the parallel time-stepping algorithms

We consider again the standard parareal algorithm (algorithm 2) and the modification with re-usage of growth values (algorithm 3). As stopping criterion, we now choose

$$|\hat{c}_s^{(k+1),\text{fine}}(\hat{x}_m, T_{\text{end}}) - \hat{c}_s^{(k),\text{fine}}(\hat{x}_m, T_{\text{end}})| < \epsilon_{\text{par}} = 10^{-4},$$

where $\hat{x}_m = (0, -1)$ is the center of the FSI interface $\hat{\Gamma}$, that is, a slightly more strict tolerance compared to the ODE model.

Standard parareal In fig. 11, we show the evolution of the error in the first three iterates of the parareal algorithm over time for $P = N_c = 10$. More precisely, we plot the error in the point functional $\hat{c}_s^{(k),\text{fine}}(\hat{x}_m, t)$, that is, the foam cell concentration at the center of the FSI interface, and the average $\int_{\hat{\Gamma}} \hat{c}_s^{(k),\text{fine}}(\hat{x}, t)\, \mathrm{d}x$ of the foam cell concentration over time. As in the first numerical example, we observe fast convergence towards the reference values. Again, the foam cell concentrations $\hat{c}_s(\hat{x}_m)$ are significantly overestimated in the initialization step, due to the large time-step δT in step (I). These are the starting values for the fine problems in the first parareal iteration.

In table 4, we compare the convergence of the function values of $\hat{c}_s^{(k),\text{fine}}(\hat{x}_m, T_{\text{end}})$ depending on the number of processes P. The findings are similar to the first numerical example (table 1). Again, depending on P, three to four parareal iterations were sufficient to reach the stopping criterion, with a slightly faster convergence behavior for larger P.

In table 5, we show numerical results for a fixed coarse time step $\delta T = \frac{T_{\text{end}}}{40}$, i.e., $N_c = 40$ and varying $P \in \{10, 20, 40\}$. This means that the cost of the coarse propagator (in terms of the number of micro problems to be solved) is equal in all three cases. While the convergence behavior is slightly faster for smaller P, the stopping criterion was satisfied after 3 parareal iterations in all cases. As the fine propagator is cheaper for larger P, the fastest computation is $P = N_c = 40$, with a speed-up of 4.3 in terms of the number of micro problems to be solved.

Parareal with re-usage of growth values In table 6, we show the results for the parareal algorithm with re-usage of growth values; cf. section 4.2.2. For all tested P, we need $k_{\text{par}} = 5$ iterations to satisfy the stopping criterion. Similar to the first numerical example, these are 1-2 additional iterations compared to the standard parareal algorithm.

k	$P = 10$	$P = 20$	$P = 40$	ref. (serial)
1	$8.51 \cdot 10^{-4}$	$7.60 \cdot 10^{-4}$	$6.53 \cdot 10^{-4}$	0
2	$7.64 \cdot 10^{-6}$	$2.25 \cdot 10^{-5}$	$6.22 \cdot 10^{-5}$	-
3	$4.36 \cdot 10^{-7}$	$3.66 \cdot 10^{-6}$	$1.36 \cdot 10^{-5}$	-
# mp	460	310	**235**	1 000
speedup	2.2	3.2	**4.3**	1.0
efficiency	**22** %	16 %	11 %	100 %

Table 5: Errors $|\hat{c}_s^{(k),\text{fine}}(T_{\text{end}}) - \hat{c}_s^*(T_{\text{end}})|$ at the midpoint $\hat{x}_m$ of $\hat{\Gamma}$ for a fixed coarse time step ($N_c = 40$), $P \in \{10, 20, 40\}$ and the parareal algorithm (algorithm 2) applied to the second numerical example (PDE growth model). Estimated parallel runtimes are shown, as well as speedup and efficiency compared to the reference computation (right column). Best numbers are marked in **bold face**.

5.2 Theoretical discussion of the computational cost

While using the ODE growth model, it was obvious that the solution of the growth model and the communication could be neglected, this requires some discussion for the PDE growth model, as the reaction-diffusion equation eq. (33) needs to be solved to advance the foam cell concentration $\hat{c}_s$.

Standard parareal In the standard parareal algorithm (algorithm 2), a time step of the PDE growth model always follows the solution of a micro problem, where the growth values $\gamma(\boldsymbol{\sigma}_f^{WS})$ are computed. Thus, the numbers of time steps of the PDE growth model and micro problems to be solved coincide. Considering that the latter consists of ≥ 100 time steps, each involving the solution of a nonlinear, coupled FSI problem, while the growth model only requires the solution of a single scalar PDE of reaction-diffusion type, it is clear that the computational cost of the growth model is negligible.

Concerning communication, each process $p = 1, ..., P$ needs to communicate its final foam cell concentration $\mathcal{F}(\overline{c}_s^{(k)}(T_p)) = \hat{c}_s^{(k),\text{fine}}(t_{p,n_p})$, which is a scalar-valued finite element function defined in the solid domain $\hat{\mathcal{S}}$, to the master process if the master-slave approach is used. No communication of $\mathcal{F}(\overline{c}_s^k(T_p))$ is required in a distributed approach; cf. section 4.1.1. Then, after the coarse propagator, the variables $\overline{c}_s^{(k+1)}(t_{p\cdot n_p})$ have to be communicated: in the master-slave approach, the data is transferred back to the slaves, and in the distributed case, it is communicated to the next process in line. Using the discretization outlined above, this corresponds to 369 degrees of freedom that need to be communicated each time.

As already discussed in section 4.1.1, in the master-slave case, this communication step is unfavorable because it involves all-to-one and one-to-all communication, whereas the communication pattern for the distributed parallelization only involves neighbor communication, which is more beneficial. However, in both cases, the two communication steps are only performed once during each parareal iteration. In order to investigate if the communication cost is still negligible for the PDE growth model, parallel numerical experiments based on an actual parallel implementation are needed, especially for realistic three-dimensional problems; see also the discussion in [34]. Due to the large computational cost of the micro problems in realistic three-dimensional problems, the communication cost might still be comparably small, which motivates the focus on the number of micro problems to be solved in our discussion; recall that, in our case, the micro problems consist of ≥ 100 time steps of a fully coupled FSI problem. Obviously, this argumentation needs to be confirmed based on an actual parallel implementation; we will investigate this further in our future work.

Parareal with re-usage of growth values The objective of the re-usage algorithm is to reduce the number of micro problems to be solved in the coarse-scale propagator. As this allows for a smaller time-step size δt in the coarse problem, the number of reaction-diffusion equations to be solved, might increase significantly. Instead of N_c of such equations in step (II.b)(i) of algorithm 2, algorithm 3 requires N_f reaction-diffusion steps in step (II.b)(ii), where $N_f = 1\,000$ in the example used in this section. Thus, in k_{par} iterations of the

k	$P = 10$	$P = 20$	$P = 30$	$P = 40$	$P = 50$	$P = 60$	$P = 70$	ref.
1	$2.80 \cdot 10^{-3}$	$1.42 \cdot 10^{-3}$	$7.49 \cdot 10^{-4}$	$6.53 \cdot 10^{-4}$	$5.04 \cdot 10^{-4}$	$4.00 \cdot 10^{-4}$	$3.32 \cdot 10^{-4}$	0
2	$6.37 \cdot 10^{-4}$	$6.96 \cdot 10^{-4}$	$5.62 \cdot 10^{-4}$	$4.08 \cdot 10^{-4}$	$3.45 \cdot 10^{-4}$	$3.09 \cdot 10^{-4}$	$2.82 \cdot 10^{-4}$	
3	$1.33 \cdot 10^{-4}$	$1.87 \cdot 10^{-4}$	$2.06 \cdot 10^{-4}$	$1.56 \cdot 10^{-4}$	$1.39 \cdot 10^{-4}$	$1.30 \cdot 10^{-4}$	$1.25 \cdot 10^{-4}$	
4	$2.87 \cdot 10^{-5}$	$5.31 \cdot 10^{-5}$	$6.14 \cdot 10^{-5}$	$5.08 \cdot 10^{-5}$	$4.61 \cdot 10^{-5}$	$4.29 \cdot 10^{-5}$	$4.02 \cdot 10^{-5}$	
5	$5.05 \cdot 10^{-6}$	$1.47 \cdot 10^{-5}$	$1.67 \cdot 10^{-5}$	$1.45 \cdot 10^{-5}$	$1.20 \cdot 10^{-5}$	$1.07 \cdot 10^{-5}$	$1.05 \cdot 10^{-5}$	
Est. par.	17 733 s	9 685 s	7 105 s	5 902 s	5 277 s	4 925 s	**4 804 s**	26 840 s
speedup	1.5	2.8	3.8	4.5	5.1	5.4	**5.6**	1.0
efficiency	**15** %	14 %	13 %	11 %	10 %	9 %	8 %	100 %

Table 6: Errors $|c_s^{(k),\text{fine}}(\hat{x}_m, T_{\text{end}}) - \hat{c}_s^*(\hat{x}_m, T_{\text{end}})|$ at the midpoint $\hat{x}_m$ of $\hat{\Gamma}$ for $P = 10$ to 70 for algorithm 3 (Re-usage of growth values) applied to the second numerical example (PDE growth model). We show estimated parallel runtimes, as well as speedup and efficiency compared to the reference computation (right column). Details on the estimation of runtimes are given in section 5.3. Best numbers are marked in **bold face**.

parareal algorithm, the number of such equations to be solved is $k_{\text{par}} \cdot N_f$ in step (II.b)(ii) plus N_c equations in step (I) and $k_{\text{par}} \cdot \lceil \frac{N_f}{P} \rceil$ in step (II.a)(ii) of algorithm 3. In total, algorithm 3 requires the solution of

$$k_{\text{par}} \cdot (N_f + \lceil N_f/P \rceil) + N_c$$

reaction-diffusion equations. For a comparison, we note that the number of micro problems to be solved for P processes was $(k_{\text{par}} + 1)N_c + k_{\text{par}} \cdot \lceil N_f/P \rceil$ (see section 4.1.2). This means that the number of reaction-diffusion equations to be solved is by a factor

$$\frac{k_{\text{par}} \cdot (N_f + \lceil N_f/P \rceil) + N_c}{(k_{\text{par}} + 1)N_c + k_{\text{par}} \cdot \lceil N_f/P \rceil}$$

larger compared to the number of micro-problems. A simple calculation yields that

$$\frac{k_{\text{par}} \cdot (N_f + \lceil N_f/P \rceil) + N_c}{(k_{\text{par}} + 1)N_c + k_{\text{par}} \cdot \lceil N_f/P \rceil} \leq \frac{k_{\text{par}}\,(N_f + \lceil N_f/P \rceil + N_c)}{k_{\text{par}}\,(N_c + \lceil N_f/P \rceil)} = \frac{N_f}{N_c + \lceil N_f/P \rceil} + 1 \leq \frac{\sqrt{N_f}}{2} + 1.$$

In the last inequality, we have used that $N_c \geq P$. For $N_f = 1\,000$, the bound on the right-hand side is approximately 16.8.

Of course, another option would be to use coarser time steps for the coarse propagator. However, noting again that a micro problem consists of ≥ 100 FSI steps, the computational cost for ≤ 16.8 scalar reaction-diffusion equations is still much cheaper. This will be confirmed in the following section, where we show estimated runtimes of a parallel implementation and the respective contributions from the micro problems and the solves of reaction-diffusion equations.

For the re-usage of growth values and the distributed parallelization scheme, the communication cost does not change compared to the standard parareal algorithm. This is because the growth values to be re-used are already available on the process; this follows directly from the discussion in section 4.2.2. In the case, of master-slave communication, the communication cost is increased. In particular, N_f growth functions $\overline{\gamma}_i, i = 1, ..., N_f$ need to be transferred from the processes $p = 1, ..., P$ to the master process. Each $\overline{\gamma}_i$ is a spatially discretized function which is defined on the FSI interface $\hat{\Gamma}$. In the example considered here, it is non-zero only in the area $\hat{x}_1 \in [-1, 1]$ which corresponds to 7 degrees of freedom in our discretization; again, for a realistic three-dimensional problem, the number of degrees of freedom will increase drastically but will still remain small compared to the full problem size. As mentioned before, we assume that the communication cost is negligible compared to the solution of the micro problems in our discussion. Again, this assumption has to be tested in the future based on parallel results.

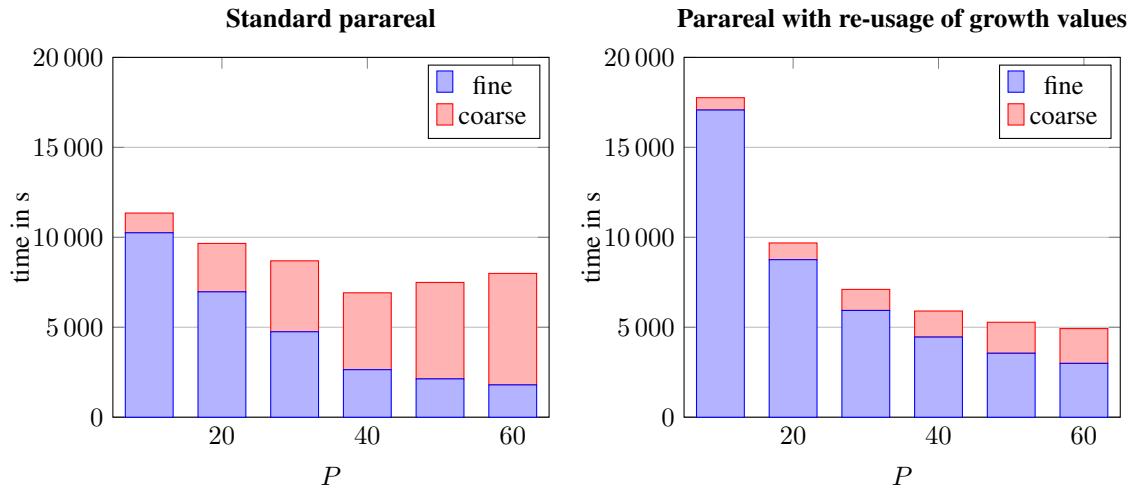


Figure 12: Illustration of the computational times spent within the coarse- (in serial) and fine-scale (in parallel) problems in algorithm 2 and algorithm 3 for the second numerical example (PDE growth problem)

Standard parareal (algorithm 2)

P	**coarse**	**fine**: max.	(aver.)	est. par.
10	1 096 s	10 251 s	(8 009 s)	11 347 s
20	2 691 s	6 968 s	(5 318 s)	9 661 s
30	3 943 s	4 749 s	(3 581 s)	8 692 s
40	4 273 s	2 641 s	(1 988 s)	6 914 s
50	5 361 s	2 130 s	(1 601 s)	7 491 s
60	6 197 s	1 796 s	(1 331 s)	7 993 s

Parareal with re-usage of growth values (algorithm 3)

P	**coarse**	**fine**: max.	(aver.)	est. par.
10	**658 s**	17 075 s	(13 342 s)	17 733 s
20	930 s	8 755 s	(6 669 s)	9 685 s
30	1 171 s	5 934 s	(4 455 s)	7 105 s
40	1 448 s	4 454 s	(3 339 s)	5 902 s
50	1 722 s	3 555 s	(2 670 s)	5 277 s
60	1 933 s	2 992 s	(2 219 s)	4 925 s

Table 7: Estimated parallel runtimes (in s) of the parareal algorithm and the variant "Re-usage of growth values" for $P = 10, ..., 60$ and a PDE growth model. We show the time in seconds spent on the master process and the maximum and average time spent on the slave processes. The estimated parallel runtime is the sum of the time spent in serial (master) and the maximum time needed among the slaves; best numbers marked in **bold face**. A visualization of the runtimes is given in fig. 13.

5.3 Comparison of runtimes

The computational results given in this paper serve as a proof of concept to test the presented algorithms. From the discussion in the previous subsection, we assume that the only relevant computational cost comes from the solution of the micro problems. As mentioned before in our current implementation, we do not solve the fine-scale problems in parallel on different processes. Instead, they are solved sequentially one after the other on a single process. Since the cost for the micro problems dominates the computation times for any serial or parallel simulation of the plaque growth, we can still discuss the parallelization potential of the methods. The parallel implementation of the algorithms itself is subject to future work.

In this section, we will give an estimate of the runtimes that would be required in a parallel implementation. For this purpose, we list the serial part and parallel contributions of the computing times in table 7. The serial part corresponds to the coarse propagation, which is performed on a single master process (steps (I) and (II.b) in algorithm 2 or algorithm 3). The parallel part corresponds to the solution of the micro problems (step (II.a)). It can be performed concurrently on all processes $p = 1, ..., P$. The computing times vary slightly across the processes. It increases for larger processes p, as more Newton iterations are required for the FSI problem, when the channel is already significantly narrowed due to the advancing plaque growth.

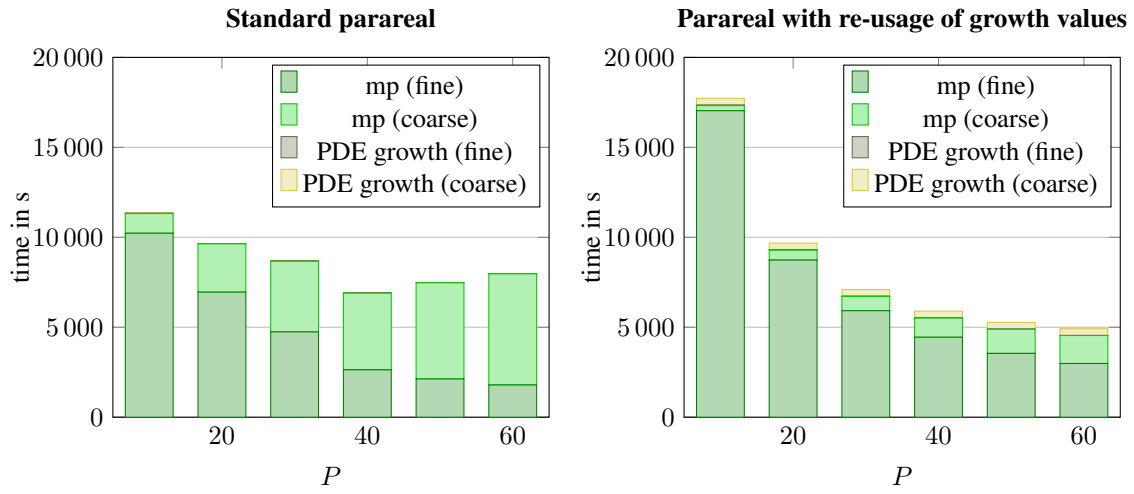


Figure 13: Illustration of the computational times spent within micro-scale problems and PDE growth problems on the coarse (in serial) and fine scale (in parallel) in algorithm 2 and algorithm 3 for the second numerical example (PDE growth problem). Note that the time needed for the PDE growth model within the fine-scale problems is so small that it is not visible in both plots. The corresponding times can be found in table 7.

The estimated parallel runtimes given in tables 4, 6 and 7 are the sum of the serial part, which corresponds to the coarse propagator, and the maximum time needed by one of the processes $p = 1, ..., P$ in the parallel part.

In fig. 12, we illustrate the computational times needed to solve the fine problems (in parallel) and coarse-scale propagators (in serial) for different processes and with algorithm 2 and algorithm 3. We see that the total computational times decrease for increasing P until $P = 40$ for standard parareal and until $P = 60$ for the re-usage variant. In the standard parareal algorithm (algorithm 2) the cost of the coarse propagators becomes significant already from $P = 20$, while it is much smaller for the re-usage variant. While for $P = 10$, algorithm 2 is still faster, this changes for $P > 20$.

The corresponding speed-ups and efficiencies compared to a serial time-stepping are given in the last rows of tables 4 and 6 for standard parareal algorithm (algorithm 2) and the re-usage variant (algorithm 3), respectively. The lowest computing time for standard parareal is 6 914 s for $P = 40$, which corresponds to a speed-up of 3.9. For the re-usage variant, we obtain a maximum speed-up of 5.6 for $P = 40$, which corresponds to an estimated parallel runtime of 4 804 s. The speed-ups are slightly lower than the speed-ups in section 4 (cf. tables 1 and 3), that were computed based on the number of micro problems to be solved. This is mostly due to the load imbalancing among the slave processes, as - depending on the state of the plaque growth - some micro problems are more costly than others; for instance, due to a higher number of Newton iterations. The differences can be inferred best by comparing the average and the maximum time spent on the slave processes in table 7. The efficiencies decrease again monotonically for increasing P in both tables 4 and 6. For the re-usage variant, the efficiencies are again much more stable, due to the reduced cost of the coarse propagator.

Finally, we compare in fig. 13 the computing times needed for the solution of the FSI micro problems and those needed for the reaction-diffusion equations, for both algorithms and $P = 10, ..., 60$. We observe that the times needed for the latter are negligible in all cases, which confirms the discussion in section 5.2.

6 Conclusion

We have derived a parareal algorithm for the time parallelization of the macro scale in a two-scale formulation for the simulation of atherosclerotic plaque growth. To reduce the computational cost of the coarse-scale propagators, we have introduced a variant which re-uses growth values that were computed within the fine problems and avoids additional costly micro-scale computations in serial.

The approaches have been tested on two different numerical examples of increasing complexity: first, by means of a simple ODE growth model, and secondly, by a PDE model of reaction-diffusion type. In this proof-of-concept, we analyze the approaches based on results and timings of a serial implementation. Since the number of communication steps is low compared to the computational work, we are still able to provide meaningful results. In the first case, we obtain estimated speedups up to 6.3 (standard parareal) respectively 7.8 (re-usage variant) in terms of the number of micro problems to be solved. In the PDE model, the maximum estimated speedups, now based on estimated parallel runtimes, were 3.9 respectively 5.6. In future work, these results will have to confirmed using a parallel implementation. Therefore, we have discussed theoretically two parallelization schemes, master-slave and distributed parallelization. The discussion indicated that the distributed scheme might be beneficial, which also needs to be tested in future work.

Additional research should also be invested into further improving the efficiency of the coarse propagator. Therefore, the idea of algorithm 3 could be extended, for example, by storing and interpolating the computed growth values, instead of simply re-using them. For the ODE case, such an interpolation approach has already been applied in our proceedings paper [53] with promising results. The extension to the PDE model is, however, not straight-forward, as an operator $\mathcal{M} : c_s \to \overline{\gamma}(\sigma_f^{WS}(c_s))$ between spatially distributed functions needs to be approximated. Further future investigations include the application of the algorithms in complex three-dimensional geometries, more complex plaque growth, and arterial wall models. Finally, the approaches presented here can be combined with a spatial parallelization of the FSI problems or with adaptive time-stepping on micro and macro scale, as presented by Richter & Lautsch [41].

References

[1] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden. The heterogeneous multiscale method. Acta Numerica, pages 1–87, 2012.

[2] C Ager, B Schott, AT Vuong, A Popp, and WA Wall. A consistent approach for fluid-structure-contact interaction based on a porous flow model for rough surface contact. Int J Numer Methods Eng, 119(13):1345–1378, 2019.

[3] UM Ascher, SJ Ruuth, and BTR Wetton. Implicit-explicit methods for time-dependent partial differential equations. SIAM J Numer Anal, 32(3):797–823, 1995.

[4] Eric Aubanel. Scheduling of tasks in the parareal algorithm. Parallel Computing, 37(3):172–182, March 2011.

[5] D. Balzani, S. Deparis, S. Fausten, D. Forti, A. Heinlein, A. Klawonn, A. Quarteroni, O. Rheinbach, and J. Schröder. Numerical modeling of fluid-structure interaction in arteries with anisotropic polyconvex hyperelastic and anisotropic viscoelastic material models at finite strains. Int J Numer Method Biomed Eng, 32(10):e02756, 2016.

[6] D. Balzani, S. Deparis, S. Fausten, D. Forti, A. Heinlein, A. Klawonn, A. Quarteroni, O. Rheinbach, and Jörg Schröder. Aspects of arterial wall simulations: Nonlinear anisotropic material models and fluid structure interaction. In Joint 11th WCCM, 5th ECCM and 6th ECFD 2014, pages 947–958, 2014.

[7] D. Balzani, A. Heinlein, A. Klawonn, O. Rheinbach, and J. Schröder. Comparison of arterial wall models in fluid–structure interaction simulations. Accepted for publication in Computational Mechnanics. March 2023.

[8] D. Balzani, P. Neff, J. Schröder, and G. A. Holzapfel. A polyconvex framework for soft biological tissues. adjustment to experimental data. Int J Solids Struct, 43(20):6052–6070, 2006.

[9] A. T Barker and X.-C. Cai. Scalable parallel methods for monolithic coupling in fluid–structure interaction with application to blood flow modeling. J Comp Phys, 229(3):642–659, 2010.

[10] Y. Bazilevs, K. Takizawa, and T. E. Tezduyar. Computational Fluid-Structure Interaction: Methods and Applications. Wiley Series in Computational Mechanics. Wiley, 2013.

[11] R. Becker and M. Braack. A finite element pressure gradient stabilization for the Stokes equations based on local projections. Calcolo, 38(4):173–199, 2001.

[12] R. Becker, M. Braack, D. Meidner, T. Richter, and B. Vexler. The finite element toolkit Gascoigne3d. http://www.gascoigne.de.

[13] Dominik Brands, A. Klawonn, O. Rheinbach, and Jörg Schröder. Modelling and convergence in arterial wall simulations using a parallel feti solution strategy. Comput Methods Biomech Biomed Engin, 11(5):569–583, 2008.

[14] E. Burman, M.A. Fernández, and S. Frei. A Nitsche-based formulation for fluid-structure interactions with contact. ESAIM: M2AN, 54(2):531–564, 2020.

[15] E. Burman, M.A. Fernández, S. Frei, and F.M. Gerosa. A mechanically consistent model for fluid-structure interactions with contact including seepage. Comput Methods Appl Mech Eng, 2022. accepted for publication. arXiv pre-print: https://arxiv.org/abs/2103.11243.

[16] LJ Caucha, S Frei, and O Rubio. Finite element simulation of fluid dynamics and CO_2 gas exchange in the alveolar sacs of the human lung. Comput Appl Math, 37(5):6410–6432, 2018.

[17] C.X. Chen, Y. Ding, and J.A. Gear. Numerical simulation of atherosclerotic plaque growth using two-way fluid-structural interaction. ANZIAM J., 53:278–291, 2012.

[18] P. Crosetto, S. Deparis, G. Fourestey, and A. Quarteroni. Parallel algorithms for fluid-structure interaction problems in haemodynamics. SIAM J Sci Comp, 33(4):1598–1622, 2011.

[19] S. Deparis, D. Forti, G. Grandperrin, and A. Quarteroni. Facsi: A block parallel preconditioner for fluid–structure interaction in hemodynamics. J Comp Phys, 327:700–718, 2016.

[20] S. S. Dhawan, R. P. Avati Nanjundappa, J. R. Branch, W. R. Taylor, A. A. Quyyumi, H. Jo, M. C. McDaniel, J. Suo, D. Giddens, and H. Samady. Shear stress and plaque development. Expert review of cardiovascular therapy, 8(4):545–556, 2010.

[21] J. Donea, A. Huerta, J.-P. Ponthot, and A. Rodríguez-Ferran. Arbitrary Lagrangian–Eulerian Methods, chapter 14. John Wiley & Sons, Ltd, 2004.

[22] T. Dunne. An Eulerian approach to fluid-structure interaction and goal-oriented mesh refinement. Int J Numer Methods Fluids, 51:1017–1039, 2006.

[23] W. E. Principles of Multiscale Modeling. Cambridge University Press, 2011.

[24] W. E and B. Engquist. The heterogenous multiscale method. Comm. Math. Sci., 1(1):87–132, 2003.

[25] CA Figueroa, S Baek, CA Taylor, and JD Humphrey. A computational framework for fluid–solid-growth modeling in cardiovascular simulations. Comput Methods Appl Mech Eng, 198(45-46):3583–3602, 2009.

[26] L. Formaggia, A. Quarteroni, and A. Veneziani. Cardiovascular Mathematics: Modeling and simulation of the circulatory system, volume 1. Springer Science & Business Media, 2010.

[27] S Frei. Eulerian finite element methods for interface problems and fluid-structure interactions. PhD thesis, Heidelberg University, 2016. http://www.ub.uni-heidelberg.de/archiv/21590.

[28] S Frei, A Heinlein, and T Richter. On temporal homogenization in the numerical simulation of atherosclerotic plaque growth. PAMM, 21(1):e202100055, 2021.

[29] S. Frei and T. Richter. Efficient approximation of flow problems with multiple scales in time. SIAM J Multiscale Model Simul, 18(2):942–969, 2020.

[30] S. Frei, T. Richter, and T. Wick. Long-term simulation of large deformation, mechano-chemical fluid-structure interactions in ALE and fully Eulerian coordinates. J Comput Phys, 321:874 – 891, 2016.

[31] M. J. Gander. 50 years of Time Parallel Time Integration. In Multiple Shooting and Time Domain Decomposition. Springer, 2015.

[32] Martin J. Gander and Ernst Hairer. Nonlinear convergence analysis for the parareal algorithm. In Ulrich Langer, Marco Discacciati, David E. Keyes, Olof B. Widlund, and Walter Zulehner, editors, Domain Decomposition Methods in Science and Engineering XVII, pages 45–56. Springer Berlin Heidelberg, 2008.

[33] M. W Gee, U. Küttler, and W. A. Wall. Truly monolithic algebraic multigrid for fluid–structure interaction. Int J Numer Methods Eng, 85(8):987–1016, 2011.

[34] Sebastian Götschel, Michael Minion, Daniel Ruprecht, and Robert Speck. Twelve Ways to Fool the Masses When Giving Parallel-in-Time Results. In Benjamin Ong, Jacob Schroder, Jemma Shipton, and Stephanie Friedhoff, editors, Parallel-in-Time Integration Methods, Springer Proceedings in Mathematics & Statistics, pages 81–94, Cham, 2021. Springer International Publishing.

[35] A. Heinlein, A. Klawonn, and O. Rheinbach. A parallel implementation of a two-level overlapping Schwarz method with energy-minimizing coarse space based on Trilinos. SIAM J Sci Comp, 38(6):C713–C747, 2016.

[36] A. Heinlein, A. Klawonn, and O. Rheinbach. Parallel two-level overlapping schwarz methods in fluid-structure interaction. In Proceedings of ENUMATH 2015, pages 521–530. Springer, 2016.

[37] G. A. Holzapfel, T. C. Gasser, and R. W. Ogden. A new constitutive framework for arterial wall mechanics and a comparative study of material models. J Elast, 61(1):1–48, 2000.

[38] J. Hron and S. Turek. A monolithic FEM/multigrid solver for an ALE formulation of fluid-structure interaction with applications in biomechanics. In Fluid-structure interaction, pages 146–170. Springer, 2006.

[39] F. Kong and X.-C. Cai. Scalability study of an implicit solver for coupled fluid-structure interaction problems on unstructured meshes in 3d. Int J High Perform Comput Appl, 32(2):207–219, 2018.

[40] U. Küttler, M. Gee, C. Förster, A. Comerford, and W. A. Wall. Coupling strategies for biomedical fluid–structure interaction problems. Int J Numer Methods Biomed Eng, 26(3-4):305–321, 2010.

[41] L. Lautsch and T. Richter. Error estimation and adaptivity for differential equations with multiple scales in time. Comput Methods Appl Math, 2021.

[42] J.-L. Lions, Y. Maday, and G. Turinici. A "parareal" in time discretization of PDE's. Comptes Rendus de l'Académie des Sciences - Series I - Mathematics, 332:661–668, 2001.

[43] Margenberg, N. and Richter, T. Parallel time-stepping for fluid-structure interactions. Math. Model. Nat. Phenom., 16:20, 2021.

[44] J Mizerski and T Richter. The candy wrapper problem-a temporal multiscale approach for pde/pde systems. In Proceedings of ENUMATH 2019, 2019.

[45] Benjamin W. Ong and Jacob B. Schroder. Applications of time parallelization. Computing and Visualization in Science, 23(1-4):Paper No. 11, 15, 2020.

[46] S Pozzi, A Redaelli, C Vergara, E Votta, and P Zunino. Mathematical modeling and numerical simulation of atherosclerotic plaque progression based on fluid-structure interaction. J Math Fluid Mech, 23(3):1–21, 2021.

[47] T. Richter. A fully Eulerian formulation for fluid-structure interactions. J Comput Phys, 233:227–240, 2013.

[48] T. Richter. Finite Elements for Fluid-Structure Interactions. Models, Analysis and Finite Elements., volume 118 of Lecture Notes in Computational Science and Engineering. Springer, 2017.

[49] Thomas Richter and Thomas Wick. Einführung in die Numerische Mathematik: Begriffe, Konzepte und zahlreiche Anwendungsbeispiele. Springer-Verlag, 2017.

[50] E.K. Rodriguez, A. Hoger, and A. D. McCulloch. Stress-dependent finite growth in soft elastic tissues. J. Biomech., 4:455–467, 1994.

[51] D Ruprecht. Wave propagation characteristics of parareal. Comput Vis Sci, 19(1):1–17, 2018.

[52] Daniel Ruprecht. Shared memory pipelined parareal. In Euro-Par 2017: Parallel Processing: 23rd International Conference on Parallel and Distributed Computing, Santiago de Compostela, Spain, August 28–September 1, 2017, Proceedings 23, pages 669–681. Springer, 2017.

[53] A. Heinlein S. Frei. Efficient coarse correction for parallel time-stepping in plaque growth simulations. In Proceedings of the 8th ECCOMAS Congress, 2022. https://www.scipedia.com/public/Frei_et_al_2022a.

[54] T. Silva, W. Jäger, M. Neuss-Radu, and A. Sequeira. Modeling of the early stage of atherosclerosis with emphasis on the regulation of the endothelial permeability. J Theor Biol, 496:110229, 2020.

[55] F. Sonner. Temporal Multiscale Methods for a Model of Atherosclerosis. PhD thesis, FAU Erlangen-Nürnberg, 2021.

[56] MP Thon, A Hemmler, A Glinzer, M Mayr, M Wildgruber, A Zernecke-Madsen, and MW Gee. A multiphysics approach for modeling early atherosclerosis. Biomech Model Mechanobiol, 17(3):617–644, 2018.

[57] A. Wahle, J. J. Lopez, M. E. Olszewski, S. C. Vigmostad, K. B. Chandran, J. D. Rossen, and M. Sonka. Plaque development, vessel curvature, and wall shear stress in coronary arteries assessed by x-r. angiography and intravascular ultrasound. Medical image analysis, 10(4):615–631, 2006.

[58] Y. Wu and X.-C. Cai. A fully implicit domain decomposition based ale framework for three-dimensional fluid–structure interaction with application in blood flow computation. J Comp Phys, 258:524–537, 2014.

[59] Y. Yang, W. Jäger, M. Neuss-Radu, and T. Richter. Mathematical modeling and simulation of the evolution of plaques in blood vessels. J Math Biol, pages 1–24, 2014.

[60] Y. Yang, T. Richter, W. Jäger, and M. Neuss-Radu. An ALE approach to mechano-chemical processes in fluid-structure interactions. Int J Numer Methods Fluids, 84(4):199–220, 2017.

Appendix

In this section, we show convergence of the re-usage parareal algorithm, algorithm 3 resp. eq. (28), applied to the ODE model in eq. (22) and its approximation by the explicit Euler method in eq. (23). First, we obtain from eqs. (23) and (24) directly for $c_s^1, c_s^2 \in \mathbb{R}$

$$|\mathcal{C}(I_n, c_s^1(T_{n-1}), \gamma) - \mathcal{C}(I_n, c_s^2(T_{n-1}), \gamma)| = |c_s^1(T_{n-1}) - c_s^2(T_{n-1})|, \tag{35}$$

$$|\mathcal{C}(I_n, c_s, \gamma(c_s^1)) - \mathcal{C}(I_n, c_s, \gamma(c_s^2))| = \delta T |\gamma(c_s^1) - \gamma(c_s^2)| \leq L\delta T |c_s^1 - c_s^2|. \tag{36}$$

By $c_{s,n-1}(T)$ we denote in the following the function that solves the (continuous) ODE eq. (22) with initial value $c_{s,n-1}(T_{n-1}) = c_{s,n-1}$. We have using eqs. (23) to (25)

$$\begin{aligned}
&\Big| \left(\mathcal{F}(I_n, c_s(T_{n-1}), \gamma(c_s(T_{n-1}))) - \mathcal{C}(I_n, c_s(T_{n-1}), \gamma(c_s(T_{n-1})))\right) \\
&\qquad - \left(\mathcal{F}(I_n, c_{s,n-1}, \gamma(c_{s,n-1})) - \mathcal{C}(I_n, c_{s,n-1}, \gamma(c_{s,n-1}))\right) \Big| \\
&= \Big| \int_{T_{n-1}}^{T_n} \gamma(c_s(T))) - \gamma(c_{s,n-1}(T))\, \mathrm{d}T - \delta T \left(\gamma(c_s(T_{n-1})) - \gamma(c_{s,n-1})\right) \Big| \\
&\leq c\delta T^2 \left| \frac{d}{dt}\gamma(c_s(T_{n-1}) - \frac{d}{dt}\gamma(c_{s,n-1}) \right| \leq \alpha_0 \delta T^2 \left| c_s(T_{n-1}) - c_{s,n-1} \right|
\end{aligned} \tag{37}$$

for some constant $\alpha_0 = cL > 0$; see also [32].

The following recursion builds the basis for the error estimate:

Lemma 2. *Let $c_s^* \in C^1(0, T_{end})$ be the exact solution of eq.* (22)*. $\{\overline{c}_s^{(k)}(T_n)\}_{n=1}^P$ be the k-th iterate of the re-usage algorithm in eq.* (28) *using the forward Euler method in eq.* (23) *and let $e_n^{(k)} = |c_s^*(T_n) - \overline{c}_{s,n}^{(k)}|$ be the error in the k-the iteration of the re-usage algorithm. Under the assumptions made in section 4.1.3, it holds for $k = 1$ that*

$$e_n^{(1)} \leq \alpha_0 \delta T^2 e_{n-1}^{(0)} + \alpha_1 \delta T e_{n-2}^{(0)} + e_{n-1}^{(1)}, \tag{38}$$

and for $k \geq 2$,

$$e_n^{(k)} \leq \alpha_1 \delta T \left(e_{n-1}^{(k-1)} + e_{n-2}^{(k-1)} + e_{n-2}^{(k-2)} \right) + e_{n-1}^{(k)}, \tag{39}$$

with a constant $\alpha_1 \geq \max\{L + \alpha_0 \delta T, L(1 + L\delta T)\}$.

Proof. Consider first the case $k \geq 2$. Using eqs. (25) and (28), we have

$$\begin{aligned}
e_n^{(k)} &= \big| \mathcal{F}(I_n, c_s^*(T_{n-1}), \gamma(c_s^*(T_{n-1})) - \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k)}, \gamma(c_{s,n-1}^{(k),\text{fine}})) \\
&\qquad - \mathcal{F}(I_n, \overline{c}_{s,n-1}^{(k-1)}, \gamma(\overline{c}_{s,n-1}^{(k-1)})) + \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k-1)}, \gamma(c_{s,n-1}^{(k-1),\text{fine}})) \big| \\
&\leq \big| \mathcal{F}(I_n, c_s^*(T_{n-1}), \gamma(c_s^*(T_{n-1}))) - \mathcal{C}(I_n, c_s^*(T_{n-1}), \gamma(c_s^*(T_{n-1}))) \\
&\qquad - \mathcal{F}(I_n, \overline{c}_{s,n-1}^{(k-1)}, \gamma(\overline{c}_{s,n-1}^{(k-1)})) + \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k-1)}, \gamma(\overline{c}_{s,n-1}^{(k-1)})) \big| \\
&\quad + \big| \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k-1)}, \gamma(\overline{c}_{s,n-1}^{(k-1)})) - \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k-1)}, \gamma(c_{s,n-1}^{(k-1),\text{fine}})) \big| \\
&\quad + \big| \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k)}, \gamma(c_{s,n-1}^{(k),\text{fine}})) - \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k)}, \gamma(c_s^*(T_{n-1}))) \big| \\
&\quad + \big| \mathcal{C}(I_n, \overline{c}_{s,n-1}^{(k)}, \gamma(c_s^*(T_{n-1}))) - \mathcal{C}(I_n, c_s^*(T_{n-1}), \gamma(c_s^*(T_{n-1}))) \big|
\end{aligned}$$

From eqs. (35) to (37), we have

$$\begin{aligned}
e_n^{(k)} &\leq \alpha_0 \delta T^2 \left| c_s^*(T_{n-1}) - \overline{c}_{s,n-1}^{(k-1)} \right| + L\delta T \left| \overline{c}_{s,n-1}^{(k-1)} - c_{s,n-1}^{(k-1),\text{fine}} \right| \\
&\quad + L\delta T \left| c_{s,n-1}^{(k),\text{fine}} - c_s^*(T_{n-1}) \right| + \left| \overline{c}_{s,n-1}^{(k)} - c_s^*(T_{n-1}) \right| \\
&= \alpha_0 \delta T^2 e_{n-1}^{(k-1)} + L\delta T (e_{n-1}^{(k-1),\text{fine}} + e_{n-1}^{(k-1)}) + L\delta T e_{n-1}^{(k),\text{fine}} + e_{n-1}^{(k)},
\end{aligned} \tag{40}$$

where $e_{n-1}^{(k),\text{fine}} := |c_{s,n-1}^{(k),\text{fine}} - c_s^*(T_n)|$. This error contribution is estimated further by using eqs. (25), (29) and (35)

$$\begin{aligned} e_{n-1}^{(k),\text{fine}} := |c_{s,n-1}^{(k),\text{fine}} - c_s^*(T_n)| &= |\mathcal{F}(I_{n-1}, c_{s,n-2}^{(k-1)}, \gamma(c_{s,n-2}^{(k-1)})) - \mathcal{F}(I_{n-1}, c_s^*(T_{n-2}), \gamma(c_s^*(T_{n-2}))| \\ &= \big|c_{s,n-2}^{(k-1)} - c_s^*(T_{n-2}) + \int_{T_{n-2}}^{T_{n-1}} \gamma(c_{s,n-2}^{(k-1)}(t))) - \gamma(c_s^*(t))\,\mathrm{d}t\big| \\ &\leq (1 + L\delta T) e_{n-2}^{(k-1)}. \end{aligned} \tag{41}$$

Inserting this into eq. (40) yields eq. (39). For $k = 1$, we have $c_{s,n-1}^{(0),\text{fine}} = c_{s,n-1}^{(0)}$. An analogous argumentation yields

$$\begin{aligned} e_n^{(1)} &= \big|\mathcal{F}(I_n, c_s^*(T_{n-1}), \gamma(c_s^*(T_{n-1}))) - \mathcal{C}(I_n, \bar{c}_{s,n-1}^{(1)}, \gamma(c_{s,n-1}^{(1),\text{fine}})) \\ &\quad - \mathcal{F}(I_n, \bar{c}_{s,n-1}^{(0)}, \gamma(\bar{c}_{s,n-1}^{(0)})) + \mathcal{C}(I_n, \bar{c}_{s,n-1}^{(0)}, \gamma(c_{s,n-1}^{(0)}))\big| \\ &\leq \big|\mathcal{F}(I_n, c_s^*(T_{n-1}), \gamma(c_s^*(T_{n-1}))) - \mathcal{C}(I_n, c_s^*(T_{n-1}), \gamma(c_s^*(T_{n-1}))) \\ &\quad - \mathcal{F}(I_n, \bar{c}_{s,n-1}^{(0)}, \gamma(\bar{c}_{s,n-1}^{(0)})) + \mathcal{C}(I_n, \bar{c}_{s,n-1}^{(0)}, \gamma(\bar{c}_{s,n-1}^{(0)}))\big| \\ &\quad + \big|\mathcal{C}(I_n, \bar{c}_{s,n-1}^{(1)}, \gamma(c_{s,n-1}^{(1),\text{fine}})) - \mathcal{C}(I_n, \bar{c}_{s,n-1}^{(1)}, \gamma(c_s^*(T_{n-1})))\big| \\ &\quad + \big|\mathcal{C}(I_n, \bar{c}_{s,n-1}^{(1)}, \gamma(c_s^*(T_{n-1}))) - \mathcal{C}(I_n, c_s^*(T_{n-1}), \gamma(c_s^*(T_{n-1})))\big| \\ &\leq \alpha_0 \delta T^2 e_{n-1}^{(0)} + L\delta T e_{n-1}^{(1),\text{fine}} + e_{n-1}^{(1)}. \end{aligned}$$

Then, eq. (38) follows by using eq. (41). □

Lemma 3. *Let $e_n^{(k)} = |\bar{c}_s^{(k)}(T_n) - c_s^*(T_n)|$ be the error in the k-the iteration of the re-usage algorithm. Under the assumptions made above, it holds for $k \in \mathbb{N}_0$ and $n \in \mathbb{N}$ that*

$$e_n^{(k)} \leq L\tilde{\alpha}^k \beta^{n-k} \cdot \left(\sum_{l=\lceil \frac{k}{2}+1 \rceil}^{k+1} \frac{3^{l-1}}{l!} \delta T^{l+1} \left(n - \frac{k}{2}\right)^l \right) \max_{t\in[0,T_{end}]} |\partial_t c_s^*(t)| \tag{42}$$

$$\leq L\tilde{\alpha}^k \beta^{n-k} \delta T \max\left\{1, T_{n-\frac{k}{2}}\right\}^k \frac{3^{\lceil \frac{k}{2} \rceil}}{\lceil k/2 \rceil!} \max_{t\in[0,T_{end}]} |\partial_t c_s^*(t)|, \tag{43}$$

where $\tilde{\alpha} = \max\{\alpha_0 \delta T, \alpha_1, 1\}$ and $\beta = 1 + L\delta T$.

Proof. We prove the lemma by induction over $k \in \mathbb{N}$. For $k = 0$, a standard estimate of the forward Euler methods gives, using eq. (24),

$$\begin{aligned} e_n^{(0)} := |c_{s,n}^0 - c_s^*(T_n)| &\leq |c_{s,n-1}^0 - c_s^*(T_{n-1})| + \int_{T_{n-1}}^{T_n} |\gamma(c_{s,n-1}^0) - \gamma(c_s^*(t))|\,\mathrm{d}t \\ &\leq e_{n-1}^{(0)} + L \int_{T_{n-1}}^{T_n} |c_{s,n-1}^0 - \underbrace{c_s^*(t)}_{=c_s^*(T_{n-1}) + \delta T \partial_t c_s^*(\xi)|}|\,\mathrm{d}t \\ &\leq (\underbrace{1 + L\delta T}_{=\beta}) e_{n-1}^{(0)} + L\delta T^2 \max_{t\in[0,T_{\text{end}}]} |\partial_t c_s^*(t)|. \end{aligned} \tag{44}$$

To abbreviate the notation, we set $c_0 := \max_{t\in[0,T_{\text{end}}]} |\partial_t c_s^*(t)|$. We apply eq. (44) recursively

$$e_n^{(0)} \leq \beta e_{n-1}^{(0)} + c_0 L\delta T^2 \leq c_0 L\delta T^2 \sum_{l=0}^{n-1} \beta^l = c_0 L\delta T^2 \underbrace{\frac{\beta^n - 1}{\beta - 1}}_{=L\delta T} = c_0 \delta T(\beta^n - 1) \leq c_0 L n \delta T^2 \beta^n. \tag{45}$$

In the last inequality, we have used that $\beta^n - 1 \leq nL\delta T\beta^n$, which can be shown by induction over n. The estimate in eq. (45) is exactly eq. (42) for $k = 0$.

For $k = 1$, we have using eqs. (38) and (45), the fact that $e_k^{(k)} = 0$, and by the definition of the parareal algorithm

$$\begin{aligned}
e_n^{(1)} &\leq \alpha_0 \delta T^2 e_{n-1}^{(0)} + \alpha_1 \delta T e_{n-2}^{(0)} + e_{n-1}^{(1)} \\
&\leq \alpha_0 c_0 L \delta T^4 (n-1)\beta^{n-1} + \alpha_1 c_0 L \delta T^3 (n-2)\beta^{n-2} + e_{n-1}^{(1)} \\
&\leq 2\tilde{\alpha} c_0 L \delta T^3 (n-1)\beta^{n-1} + e_{n-1}^{(1)} \\
&\leq 2\tilde{\alpha} c_0 L \beta^{n-1} \delta T^3 \sum_{l=1}^{n-1} l + \underbrace{e_1^{(1)}}_{=0} \\
&= 2\tilde{\alpha} c_0 L \beta^{n-1} \delta T^3 \frac{n(n-1)}{2} \leq \tilde{\alpha} c_0 L \beta^{n-1} \delta T^3 \left(n - 1/2\right)^2 .
\end{aligned}$$

This is by a factor $\frac{3}{2}$ smaller compared to eq. (42).

Now, let $k \geq 2$. We assume that the estimate in eq. (42) is true for $k-1$ and $k-2$. By eq. (39) and the assumption of the induction, we have

$$\begin{aligned}
e_n^{(k)} &\leq \alpha_1 \delta T \left(e_{n-1}^{(k-1)} + e_{n-2}^{(k-1)} + e_{n-2}^{(k-2)} \right) + e_{n-1}^{(k)} \\
&\leq \alpha_1 \delta T \Bigg(c_0 L \tilde{\alpha}^{k-1} \beta^{n-k} \cdot \left(\sum_{l=\lceil \frac{k+1}{2} \rceil}^{k} \frac{3^{l-1}}{l!} \delta T^{l+1} \left(n - \frac{k+1}{2} \right)^l \right) \\
&\qquad + c_0 L \tilde{\alpha}^{k-1} \beta^{n-k-1} \cdot \left(\sum_{l=\lceil \frac{k+1}{2} \rceil}^{k} \frac{3^{l-1}}{l!} \delta T^{l+1} \left(n - 1 - \frac{k+1}{2} \right)^l \right) \\
&\qquad + c_0 L \tilde{\alpha}^{k-2} \beta^{n-k} \cdot \left(\sum_{l=\lceil \frac{k}{2} \rceil}^{k-1} \frac{3^{l-1}}{l!} \delta T^{l+1} \left(n - 1 - \frac{k}{2} \right)^l \right) \Bigg) + e_{n-1}^{(k)} \\
&\leq 3 c_0 L \tilde{\alpha}^k \beta^{n-k} \cdot \left(\sum_{l=\lceil \frac{k}{2} \rceil}^{k} \frac{3^{l-1}}{l!} \delta T^{l+2} \left(n - \frac{k+1}{2} \right)^l \right) + e_{n-1}^{(k)}.
\end{aligned} \tag{46}$$

We apply this estimate recursively for the last term $e_{n-1}^{(k)}$ in eq. (46) to get

$$\begin{aligned}
e_n^{(k)} &\leq c_0 L \tilde{\alpha}^k \beta^{n-k} \cdot \sum_{m=k+1}^{n} \left(\sum_{l=\lceil \frac{k}{2} \rceil}^{k} \frac{3^l}{l!} \delta T^{l+2} \left(m - \frac{k+1}{2} \right)^l \right) + \underbrace{e_k^{(k)}}_{=0} \\
&= c_0 L \tilde{\alpha}^k \beta^{n-k} \cdot \left(\sum_{l=\lceil \frac{k}{2} \rceil}^{k} \frac{3^l}{l!} \delta T^{l+2} \cdot \underbrace{\left(\sum_{m=\frac{k+1}{2}}^{n-\frac{k+1}{2}} m^l \right)}_{=:s} \right).
\end{aligned}$$

In the case that $\frac{k+1}{2}$ and $n - \frac{k+1}{2}$ are no natural number, the sum s is to be understood in such a way that the index m advances iteratively by 1 until reaching the upper limit.

The sum s is an approximation of the integral

$$\int_{k/2}^{n-k/2} x^l \, \mathrm{d}x \leq \frac{1}{l+1} \left(n - \frac{k}{2} \right)^{l+1}$$

with the midpoint rule. As $f(x) = x^l$ is a convex function ($f''(x) > 0$) for a positive x, the integral is an upper bound for s; see the error representation of the midpoint rule, e.g., in Theorem 8.41 in [49].

We have thus shown that

$$e_n^{(k)} \leq c_0 L \tilde{\alpha}^k \beta^{n-k} \cdot \left(\sum_{l=\lceil \frac{k}{2} \rceil}^{k} \frac{3^l}{(l+1)!} \delta T^{l+2} \left(n - \frac{k}{2} \right)^{l+1} \right).$$

Shifting the index l by 1 gives eq. (42). The estimate in eq. (43) follows from eq. (42) by noting that

$$(n - \frac{k}{2})\delta T = T_{n-\frac{k}{2}} \leq \max \left\{ 1, T_{n-\frac{k}{2}} \right\}$$

and the fact that the term $\frac{3^{l-1}}{l!}$ is decreasing for $l \geq 2$:

$$\begin{aligned}
e_n^{(k)} &\leq c_0 L \tilde{\alpha}^k \beta^{n-k} \cdot \left(\sum_{l=\lceil \frac{k}{2}+1 \rceil}^{k+1} \frac{3^{l-1}}{l!} \delta T^{l+1} \left(n - \frac{k}{2} \right)^{l} \right) \\
&\leq c_0 L \tilde{\alpha}^k \beta^{n-k} \delta T \left(\max\{1, T_{n-\frac{k}{2}}\} \right)^{k+1} \cdot \left(\sum_{l=\lceil \frac{k}{2}+1 \rceil}^{k+1} \frac{3^{l-1}}{l!} \right) \\
&\leq c_0 L \tilde{\alpha}^k \beta^{n-k} \delta T \left(\max \left\{ 1, T_{n-\frac{k}{2}} \right\} \right)^{k+1} \lfloor \frac{k}{2} + 1 \rfloor \frac{3^{\lceil \frac{k}{2} \rceil}}{\lceil \frac{k}{2} + 1 \rceil !} \\
&\leq c_0 L \tilde{\alpha}^k \beta^{n-k} \delta T \left(\max \left\{ 1, T_{n-\frac{k}{2}} \right\} \right)^{k+1} \frac{3^{\lceil \frac{k}{2} \rceil}}{\lceil \frac{k}{2} \rceil !}.
\end{aligned}$$

□