TECHNISCHE
UNIVERSITÄT
DARMSTADT

Finite Element Mass Lumping for H(div) and H(curl)

Vom Fachbereich Mathematik

der Technischen Universität Darmstadt

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte Dissertation

von

Bogdan Radu, M.Sc.

aus Cluj-Napoca, Rumänien

Referent : Prof. Dr. Herbert Egger

Korreferent : Prof. Dr. Guido Kanschat

Tag der Einreichung: 09.02.2022

Tag der mündlichen Prüfung: 31.03.2022

Darmstadt
D17

Finite element mass lumping in H(div) and H(curl)

Accepted doctoral thesis by Bogdan Radu

Referentin: Prof. Dr. Herbert Egger
Korreferent: Prof. Dr. Guido Kanschat

Date of submission: 09.02.2022
Date of thesis defense: 31.03.2022

Darmstadt – D17

Bitte zitieren Sie dieses Dokument als:
URN: urn:nbn:de:tuda-tuprints-219485
URL: https://tuprints.ulb.tu-darmstadt.de/21948

Dieses Dokument wird bereitgestellt von tuprints,
E-Publishing-Service der TU Darmstadt
http://tuprints.ulb.tu-darmstadt.de
tuprints@ulb.tu-darmstadt.de

To my parents
and my sister
for their love and support

Acknowledgments

First of all, I wish to thank my supervisor Herbert Egger for all the valuable advice and unwavering guidance through all the challenges. His door was always open and he always made ample time for discussing my questions and showed genuine interest in my work and opinion.

Special thanks go to my colleagues Elisa, Gabriel, Moritz, Nora, Seva, Tommy, Hrishikesh, and Richard from the numerics group. I will never forget the countless Doppelkopf sessions and fun conversations we shared.

Finally, I'd like to thank my parents and my sister Isabell for their support throughout the years.

Abstract

In this work, we consider the efficient implementation of finite element approximations for porous media, poroelasticity, and wave propagation problems. This is conducted by using *mass-lumping*, an approximation technique that enables a faster inversion of the mass matrix. This method has been primarily used for discretizing wave propagation problems in H^1, as it allows for the efficient application of explicit time-stepping schemes. Another example of the usage of *mass-lumping* is the mixed finite element discretization of porous medium flow. Here, *mass-lumping* is utilized to reduce the algebraic saddle-point structure to a symmetric positive definite system, which is a lot easier to solve.

In this work, we specifically examine *mass-lumping* strategies for the functional spaces $H(\mathrm{div})$ and $H(\mathrm{curl})$. In the first part of this thesis, we investigate the discretization of porous medium flow and poroelasticity. We will analyze already existing first-order convergent methods with *mass-lumping* and propose extensions that have better convergence properties and are optimal in the number of degrees of freedom. We will complement this with a rigorous error analysis, which confirms the accuracy of the method.

In the second part, we consider discretization techniques with *mass-lumping* for the acoustic wave equation in $H(\mathrm{div})$ and for Maxwell's equations in $H(\mathrm{curl})$. Specifically, we will propose first and second-order accurate methods. Moreover, we will propose a new technique for the first-order convergent method that will cut the number of degrees of freedom in half (or even more). For all the methods we propose, we also provide full error convergence analyses. Furthermore, all our results will be validated by numerical experiments.

Zusammenfassung

In dieser Arbeit beschäftigen wir uns mit der effizienten Implementierung von Finite-Elemente Verfahren für poröse Medien, Poroelastizität und Wellenausbreitung. Dies wird mittels *mass-lumping* realisiert, eine Approximationstechnik die das Invertieren von Massematrizen erheblich vereinfacht. Diese Methode taucht hauptsächlich in der Diskretisierung von Wellenpropagationsproblemen in H^1 auf und ermöglicht eine schnellere Anwendung von expliziten Zeitschrittverfahren. Ein weiteres Themengebiet in dem *mass-lumping* eine Rolle spielt ist die Diskretisierung der Strömung in porösen Medien mittels gemischten Finite-Elementen. In diesem Kontext wird durch mass-lumping das resultierende algebraische Sattelpunktproblem zu einer symmetrisch positiv definiten Gleichung reduziert, die erheblich leichter zu lösen ist.

In der vorliegenden Arbeit untersuchen wir speziell *mass-lumping* Strategien für die Diskretisierung in den Funktionenräumen $H(\mathrm{div})$ und $H(\mathrm{curl})$. Im ersten Teil dieser Arbeit widmen wir uns der Diskretisirung von porösen Medien sowie Poroelastizität. Wir werden in diesem Zusammenhang die bestehende Methode erster Ordnung mit *mass-lumping* untersuchen und Erweiterungen vorschlagen, die bessere Konvergenzeingeschaften besitzen und optimal in der Anzahl der Freiheitsgrade sind. Zudem werden wir eine rigurose Konvergenzanalyse präsentieren, die die Genauigkeit unserer Methoden belegt.

In der zweiten Hälfte untersuchen wir Diskretisierungen mit *mass-lumping* für die akustische Wellengleichung in $H(\mathrm{div})$ sowie für die Maxwell-Gleichungen in $H(\mathrm{curl})$. Wir werden Methoden vorschlagen, die erste und zweite Ordnung konvergent sind. Zudem entwickeln wir eine neues Verfahren für die Methode erster Ordnung, die die Anzahl der Freiheitsgrade um die Hälfte (oder sogar mehr) reduziert. All diese Methoden werden mit einer vollen Konvergenzanalyse vorgestellt. Anschliessend werden wir unsere Resultate auch durch numerische Untersuchungen bestätigen.

Contents

1. Introduction

In this thesis, we investigate *mass lumping* strategies for the efficient implementation of finite element methods for wave propagation problems and porous medium flow. The goal is to increase efficiency without sacrificing the versatility and accuracy of the finite element approximations. To ensure this, full convergence analyses for all the proposed methods will be given. Before we delve into the details of mass lumping, let us briefly recall some key historical facts regarding the finite element method.

The finite element method

The finite element method (FEM) is a well-established discretization approach for partial differential equations. In principle, its foundations were laid in the 17th and 18th centuries by Euler, Bernoulli, and Leibnitz, who first proposed the concept of variational calculus. The first notable application of variational arguments to solving engineering problems was conducted in 1909 by Walter Ritz [**98**] in his successful attempts to numerically compute eigenvalues for the linear elastic Kirchhoff plate. A few years later in 1915, Boris Galerkin [**59**] published his influential work on the analysis of variational methods, in which he showed that the residual error is orthogonal concerning the functions in the test space. Nowadays, this property is widely referred to as *Galerkin orthogonality*. A few decades later, in 1943, Courant [**39**] proposed to use piecewise polynomials over a finite number of subdomains for the approximation, so-called *finite elements*. In his paper, he used triangular and quadrilateral subdomains to produce an approximation to the torsion problem. The method did not gain immediate attention, since it required the solution of a large system of equations, which was not practical at the time. With the advent of modern computers, the FEM started gaining momentum in the mechanical engineering community around 1960, culminating in 1967 with the book by Zienkiewicz et. al. titled *"The finite element method"* [**116**], in which the authors examined the method for various applications, including heat conduction, fluid flow, and elastodynamics. Since then, the FEM has evolved into a versatile technique for approximating solutions of partial differential equations. Among its many fields of application, a notable one is the approximation of solutions to wave propagation problems in the time domain.

A very simple example for such applications is the scalar acoustic wave equation

$$\partial_{tt} p - \Delta p = 0, \tag{1.1}$$

describing variations of the pressure p and the propagation of the resulting sound waves. For ease of notation, we further assume dimensionless variables and a homogeneous medium speed of sound $c = 1$ here. For our further considerations, we consider a bounded

computational domain and complement the system with homogeneous boundary and appropriate initial conditions. The variational formulation and subsequent discretization by finite elements then leads to a finite-dimensional system of ordinary differential equations

$$\mathsf{M}\partial_{tt}\mathsf{p} + \mathsf{K}\mathsf{p} = 0, \tag{1.2}$$

where p is the coordinate vector of the basis representation $p_h(x,t) = \sum_i \mathsf{p}_i(t)\phi_i(x)$ of the finite element approximation p_h with respect to a given set of basis functions $\{\phi_i\}_i$. The mass and stiffness matrices M and K are defined by

$$\mathsf{M}_{ji} = \int_\Omega \phi_i(x)\phi_j(x)\,dx \qquad \text{and} \qquad \mathsf{K}_{ji} = \int_\Omega \nabla\phi_i(x)\cdot\nabla\phi_j(x)\,dx$$

are the discrete representations of the identity and the $-\Delta$ operator.

Using a FEM for approximating wave propagation problems comes with many advantages: First of all, it can be applied to general, unstructured grids and can accommodate anisotropic and inhomogeneous coefficients natively. Moreover, it provides us with comprehensive error analysis and allows for provable convergence rates. Significant contributions in this field were made in the 70s by Dupont [**45**] and Baker [**12**] and later extended by Wheeler et. al. [**42**]. When using piecewise linear basis functions, the error for the semi-discretization can be bounded in terms of the mesh size h by

$$\|\partial_t(p-p_h)\|_{L^\infty(0,T,L^2(\Omega))} + \|p-p_h\|_{L^\infty(0,T,L^2(\Omega))} \le Ch^2,$$

and this can be achieved under minimal regularity assumptions on the data. A simple second-order time discretization of (1.1) leads to the scheme

$$\mathsf{M}\frac{\mathsf{p}^{n+1} - 2\mathsf{p}^n + \mathsf{p}^{n-1}}{\tau^2} - \mathsf{K}\mathsf{p}^n = 0, \tag{1.3}$$

with $\tau > 0$ denoting the time step size. The stability of this scheme can be guaranteed under a mild restriction on the time step size, known as the CFL condition [**40**]. Under this condition, one can prove the estimate

$$\max_{t^n \le T} \|p(t^n) - p_h^n\|_{L^2(\Omega)} = O(h^2 + \tau^2);$$

see e.g. [**73**]. Thus, the method exhibits second-order convergence in space and time. Unconditionally stable implicit time stepping methods leading to the same convergence rates have been considered in [**12, 45**].

Compared to finite difference approximations, for which similar convergence results can be proven under rather restrictive assumptions on model parameters, the computational domain, and the regularity of the solution, the finite element method is certainly more general and flexible, but it also comes with some major drawbacks: Explicit time marching schemes such as (1.3) can only be applied efficiently if the associated mass matrix M can be inverted easily. Although finite element mass matrices are sparse, their inverse is, in general, not sparse. This essentially hinders the efficient realization of explicit time-stepping schemes and poses a severe disadvantage compared to efficient finite difference approximations, which typically have diagonal mass matrices.

The main goal of this thesis is to construct and analyze certain modifications of finite element methods which lead to diagonal or block-diagonal mass matrices and thus to methods with the efficiency of typical finite-difference methods while at the same time providing the flexibility and generality of finite element approximations. Besides wave–propagation, we also study applications in porous medium flow, where similar modifications again facilitate the numerical solution.

A brief history of mass lumping

One way to overcome the computational bottleneck of the FEM for wave propagation problems is *mass lumping*. The main idea is to modify the mass matrix in such a way that its inverse can be applied efficiently. In general, this modification should lead to a diagonal or block-diagonal mass matrix, but we will also encounter different constructions later on. Mass lumping thus allows explicit time-stepping methods to be applied efficiently and also to keep the other advantages of the FEM, like the ability to deal with rather general meshes and coefficients. The viability of such approaches of course hinges on how the modification of the mass matrix is realized in detail.

An early attempt in this direction was made already in the 60s by Zienkiewicz [**116**], who proposed the *row-sum lumping*. In this approach, the entries of each row of the mass matrix are moved to the diagonal. This heuristic scheme works surprisingly well in practice, but it can also fail for general meshes. In particular, one can give explicit examples for which at least one diagonal entry is zero, which leads to a singular lumped mass matrix. This can again be fixed heuristically by carefully weighing each row entry, see[**44**]; a rigorous mathematical error analysis however seems not available to date.

A more systematic approach is the *nodal mass lumping*, a technique where numerical quadrature is used in conjunction with carefully chosen shape functions to generate an approximation of the mass matrix. The approach was proposed in 1975 by Fried and Malkis [**58**] and investigated intensively around 2000 by Mulder et. al. [**31, 86**]. A thorough error analysis for finite element approximations of the wave equation including such mass lumping techniques can be found in Gary Cohen's book [**35**]. Under certain conditions on the finite element basis and the underlying quadrature rule, the resulting lumped mass matrix can be proven to be positive definite, and the resulting method can be shown to not lose accuracy in comparison to its standard finite element counterpart. Let us briefly highlight the main ingredients for nodal mass lumping for the simple acoustic wave equation considered before.

Mass lumping in H^1 for the wave equation

For the space discretization of (1.2), we consider a finite element method with continuous P_1-elements. The usual Lagrange basis on the reference triangle $\widehat{K}$ is given by

$$\widehat{\phi}_1(x,y) = 1 - x - y \qquad \widehat{\phi}_2(x,y) = x \qquad \widehat{\phi}_3(x,y) = y; \tag{1.4}$$

see Figure 1.1 for a sketch of the reference element and the location of the degrees of freedom associated with these basis functions. By using this basis and an appropriate

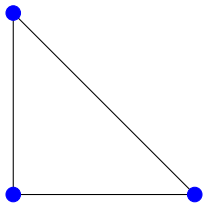

Figure 1.1.: Degrees of freedom of the space P_1 on $\widehat{K}$.

mapping from the reference element to the physical element, the local mass matrix on the physical element K can be shown to be

$$\mathsf{M}|_K = \frac{|K|}{12}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}. \tag{1.5}$$

It is not difficult to see that the global mass matrix will have non-zero entries for every pair of vertices connected by an edge. As a consequence, the global mass matrix will not be diagonal and not even block-diagonal.

We now seek to replace the local mass matrix with a diagonal approximation. To do so, we simply approximate the integrals appearing in the definition of the element mass matrix by an appropriate quadrature formula. For the particular setting, we choose

$$\mathsf{M}_{ij} \approx (\mathsf{M}_\mathsf{h})_{ij} := \sum\nolimits_{|K|} \frac{|K|}{3} \sum_{k=1}^{3} \phi_i(v_k^K)\phi_j(v_k^K), \tag{1.6}$$

where v_k^K is the k-th vertex of the element K. This integration rule is known as the *vertex rule*, a natural generalization of the trapezoidal rule to dimension $d > 1$. The proposed modification immediately leads to a "lumped" local mass matrix of the diagonal form

$$\mathsf{M}_\mathsf{h}|_K = \frac{|K|}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.7}$$

As a consequence, also the global mass matrix, which is constructed by assembling local mass matrices, will be diagonal and hence efficiently invertible. Moreover, the explicit time stepping scheme (1.3) can now be realized efficiently.

By an appropriate error analysis for non-conforming finite element discretizations, one can show that the method retains optimal convergence order [**36**], i.e.

$$\max_n \|p(t^n) - p_h^n\|_{L^2(\Omega)} + h(\sum\nolimits_{k=1}^{n} \tau \|p(t^k) - p_h^k\|^2_{H^1(\Omega)})^{1/2} = O(h^2 + \tau^2).$$

From this simple example, one might get the wrong impression that mass lumping is always for free, which is however not true in general, as we illustrate next.

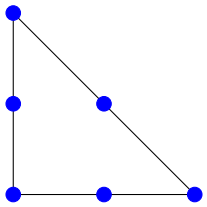

Figure 1.2.: Degrees of freedom of the space P_2 on $\widehat{K}$.

Second order mass lumping in H^1

To improve the convergence order in h, we have to increase the polynomial order of the spatial approximation and look for an appropriate quadrature rule having sufficient accuracy and giving rise to a diagonal mass matrix. For quadratic finite elements $P_2(K)$, we have a total of six shape functions for every element K; three are associated with the vertices, and three are associated with the three edges of K; see Figure 1.2 for a sketch. To obtain a diagonal mass matrix, it is then natural to choose the location of the degrees of freedom as the quadrature points, and to determine the quadrature weights to obtain the optimal accuracy, here exactness for (at least) quadratic polynomials; see [**36**]. Unfortunately, the only quadrature rule that fulfills these conditions has zero weights at the vertices and hence leads to a singular mass matrix.

A solution to overcome this problem is to extend the finite element space by an additional degree of freedom, which allows for more flexibility in finding an appropriate quadrature rule. Let us consider the extended space

$$\widetilde{P}_2 = P_2 \oplus \operatorname{span}\{b_3\},$$

where b_3 represents the cubic bubble function that vanishes at the boundary of the element. Using the natural location of the degrees of freedom as quadrature points, see

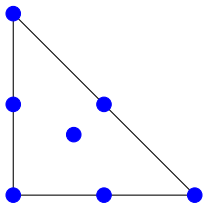

Figure 1.3.: Degrees of freedom of the space $\widetilde{P}_2$ on $\widehat{K}$.

Figure 1.3, one can find a corresponding quadrature rule which is exact for all cubic polynomials and has strictly positive weights. The lumped mass matrix is defined by

$$(\mathsf{M_h})_{ij} \coloneqq \sum\nolimits_{|K|} |K| \left(\frac{1}{20} \sum_{k=1}^{3} \phi_i(v_k)\phi_j(v_k) + \frac{2}{15} \sum_{k=1}^{3} \phi_i(e_k)\phi_j(e_k) + \frac{9}{20}\phi_i(m)) \right)$$

where v_k represents the k-th vertex and e_k the k-th edge midpoint and m the barycenter of the element K. If the basis functions are chosen as the Lagrangian basis concerning

the quadrature points, the local mass matrix is again diagonal. For the resulting finite element method, one can show that

$$\max_n \|p(t^n) - p^n\|_{L^2(\Omega)} + h(\sum\nolimits_{k=1}^{n} \tau \|p(t^k) - p^k\|^2_{H^1(\Omega)})^{1/2} = O(h^3 + \tau^2).$$

In particular, the method is second-order accurate in the energy norm, as desired.

For a comprehensive analysis of mass lumping schemes for the wave equation in H^1 and extensions to higher order approximations, we refer to [**35, 73**]. The error analysis has been extended recently by Geevers [**60**], who developed relaxed conditions on the exactness of the quadrature rule required for higher order approximations on tetrahedral grids. The available results indicate that the design of appropriate quadrature rules for simplicial elements requires some flexibility and additional degrees of freedom in the corresponding finite element spaces, which becomes disadvantageous for very high approximation orders. In contrast, almost optimal mass lumping strategies of arbitrary order can be obtained for quadrilateral and hexahedral grids by tensor-product Gauss- and Gauss-Lobatto quadrature rules and corresponding finite element spaces; see [**35**] for details.

Another strategy for obtaining finite element approximations with diagonal or block-diagonal mass matrices is offered by *discontinuous Galerkin* (dG) methods. We refer to the book of Warburton and Hesthaven [**66**] for a detailed introduction. The advantage of having easily invertible mass matrices, however, comes at the cost of a significant increase in the number of degrees of freedom and additional coupling terms required to obtain a consistent and stable approximation. For moderate approximation orders, continuous Galerkin schemes with mass lumping therefore seem to be more efficient than corresponding discontinuous Galerkin schemes; see [**60, 115**] for detailed comparisons.

Mixed finite element methods for the wave equation

Mixed finite element approximations have been analyzed thoroughly in the early 70s by Crouzeix and Raviart [**43**] and Brezzi [**24**] for the solution of problems in fluid dynamics and elasticity. Applications include the Darcy equations modeling flow in porous media, the Maxwell eigenvalue problem, incompressible flow, poroelasticity, and so on. For a comprehensive list of applications, we refer to [**19, Chapter 7**].

As one area of applications, we are interested here in mixed finite element approximations for wave propagation problems. By introducing the velocity $v = \partial_t p$ and the stress $\sigma = -\nabla p$ as new variables, the wave equation (1.1) can be written equivalently as a first order system

$$\partial_t v + \operatorname{div} \sigma = 0, \tag{1.8}$$

$$\partial_t \sigma + \nabla v = 0, \tag{1.9}$$

called a *velocity-stress* formulation in [**61**]. When considered on a bounded domain Ω, the system is again to be complemented with appropriate boundary and initial conditions. In the context of variational approximation methods, we now have two choices for spatial discretization. The first natural choice is to discretize v in $H^1(\Omega)$ and σ in $L^2(\Omega)^d$. For

this choice, it is possible to eliminate the stress on the algebraic level and essentially revert to the second order formulation (1.1) from before.

Another choice is to discretize v in $L^2(\Omega)$ and σ in $H(\mathrm{div}, \Omega)$, leading to a mixed finite element method. Corresponding schemes have been studied by Geveci [**61**], and later by Dupont and Wheeler [**41**], Makridakis [**79**] and Jenkins et al. [**71**]. In [**46**], we established super-convergence results and post-processing schemes for such mixed finite element approximations. After choosing appropriate basis functions, (1.8)–(1.9) leads to the algebraic system

$$\begin{aligned} \mathsf{D}\partial_t \mathsf{v} + \mathsf{B}\sigma &= 0, \\ \mathsf{M}\partial_t \sigma - \mathsf{B}^\top \mathsf{v} &= 0, \end{aligned}$$

where B is the discrete divergence matrix and M and D are the $H(\mathrm{div}, \Omega)$ and $L^2(\Omega)$ mass matrices, respectively. Since $L^2(\Omega)$ does not require any sort of continuity across element interfaces, it is clear that D can be made diagonal by a proper choice of the basis. With this in mind, we can eliminate the velocity v on the algebraic level, and obtain the system

$$\mathsf{M}\partial_{tt}\sigma + \mathsf{B}^\top \mathsf{D}^{-1} \mathsf{B}\sigma = 0$$

for the stress σ only, which is of the same structural form as (1.2). For the efficient time integration by explicit schemes, it is now essential that the mass matrix M for $H(\mathrm{div})$-finite elements is easily invertible, which again requires appropriate mass lumping schemes. In contrast to wave propagation problems in H^1, mass lumping strategies for $H(\mathrm{div})$ are however hardly available in the literature. In [**29**], mass lumping for the lowest order RT_0 space on rectangular Cartesian grids has been investigated by Becache et.al. for discretizations of the elastodynamic wave equation. Strategies for quadrilateral and hexahedral elements can again be found in [**35**]. In this work, we focus on mass lumping for finite element approximations in $H(\mathrm{div})$ on unstructured grids, which has also been investigated in the context of mixed finite element approximations of the Poisson problem.

Porous medium flow

A standard model for the subsurface flow of an incompressible fluid is given by the system

$$\mathcal{K}^{-1} u + \nabla p = 0, \tag{1.10}$$

$$\mathrm{div}\, u = f. \tag{1.11}$$

Here u is the flow velocity, p is the pressure, and $\mathcal{K}$ is the hydraulic conductivity of the medium. The equations are assumed to hold on a bounded domain Ω and complemented by homogeneous boundary conditions for the pressure. A mixed finite element approximation for u in $H(\mathrm{div}, \Omega)$ and p in $L^2(\Omega)$ then leads to the algebraic saddle point problem

$$\begin{pmatrix} \mathsf{M} & -\mathsf{B}^\top \\ \mathsf{B} & 0 \end{pmatrix} \begin{pmatrix} \mathsf{u} \\ \mathsf{p} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathsf{f} \end{pmatrix},$$

where B is the discrete divergence matrix and M is the $H(\mathrm{div})$ mass matrix. While this approximation is known to have many advantageous features, like local mass conservation and improved accuracy for rough coefficients, the increased dimension and saddle point structure of the algebraic system are certainly obstacles to computational efficiency.

A well-known strategy to overcome the saddle point structure is *hybridization*, originally introduced by Arnold and Brezzi in [**7**]. In this method, the continuity of the velocity variables is broken up and re-established via Lagrange multipliers. This allows for eliminating velocities and pressures by local static condensation and obtaining a global positive definite system for the Lagrange multipliers, which play the role of pressure variables located at the element interfaces. By equivalence to the standard mixed finite element method, all convergence results of the latter remain valid.

Direct elimination of the velocity degrees of freedom, on the other hand, would lead to a Schur complement problem for the original pressure variables, which is also positive definite. Unfortunately, the resulting system matrix $\mathsf{B}\mathsf{M}^{-1}\mathsf{B}^{\top}$ is not sparse and this approach is therefore not computationally efficient. In [**112**], Wheeler and Yotov proposed the *multipoint flux mixed finite element* (MFMFE) method for Darcy flow, in which the $H(\mathrm{div})$-conforming mass matrix M is replaced by a block-diagonal approximation. This is achieved by considering the lowest order BDM space on triangles and parallelograms, together with appropriate numerical quadrature by the vertex rule and a special choice of basis functions. As a consequence, the velocity degrees of freedom can be eliminated by a local process and the resulting Schur complement matrix for the pressure is sparse and can be interpreted as a certain finite volume of discontinuous Galerkin method. The approach was extended to hexahedral elements in [**70**], and quite recently to arbitrary order elements [**4**], albeit only for quadrilaterals and hexahedral grids. An extension to higher order approximations for unstructured grids under relaxed conditions on the quadrature rule has been proposed in our recent publication [**48**].

While the advantage of the MFMFE method compared to hybridization is not completely clear in the context of porous medium flow, the underlying approximation of the mass matrix has drastic benefits for the efficient solution of acoustic wave propagation problems in $H(\mathrm{div})$. Hybridization in this case would still require the solution of linear systems in every time step, and thus not bring the same advantage. For matters of presentation, we will nevertheless study mass lumping strategies for $H(\mathrm{div})$ first in the context of porous medium flow, and then extend the results to wave propagation problems.

Maxwell's equations

Similar problems regarding the efficient realization of mixed finite element methods arise also in the context of electromagnetic wave propagation. As a simple model problem, we consider the time domain Maxwell's equations

$$\epsilon \partial_t E - \operatorname{curl} H = 0, \tag{1.12}$$

$$\mu \partial_t H + \operatorname{curl} E = 0. \tag{1.13}$$

Here E and H describe the electric and magnetic field intensities, and ϵ, μ, the electromagnetic material parameters. The equations are again considered in a bounded domain

Ω and supplemented with appropriate initial and homogeneous boundary conditions.

The approximation by mixed finite elements in $H(\mathrm{curl}, \Omega)$ and $L^2(\Omega)$ leads to an algebraic system of equations

$$\begin{aligned} \mathsf{M}\partial_t \mathsf{e} + \mathsf{B}\mathsf{h} &= 0, \\ \mathsf{D}\partial_t \mathsf{h} - \mathsf{B}^\top \mathsf{e} &= 0. \end{aligned}$$

Here M and D are the mass matrices for $H(\mathrm{curl}, \Omega)$ and $L^2(\Omega)$, while B is the discrete curl matrix. Elimination of h on the algebraic level again yields a second-order form

$$\mathsf{M}\partial_{tt} \mathsf{e} + \mathsf{B}^\top \mathsf{D}^{-1} \mathsf{B}\mathsf{e} = 0.$$

The analysis of finite element methods for time-dependent Maxwell's equations has been conducted by Monk in the 90s, see e.g. [**80, 85**], based on the $H(\mathrm{curl})$-conforming finite element spaces proposed in the 80s by Nédélec [**92, 90**]; also see [**87**].

For the efficient solution by the explicit time-stepping scheme, it is again essential to have an easily invertible approximation for the $H(\mathrm{curl})$ mass matrix M. Corresponding mass lumping techniques for $H(\mathrm{curl})$ have been proposed in the literature: A method that strongly resembles row-sum lumping has been mentioned by Baranger [**13**]; also see [**78**] for a detailed investigation. The method is applicable to the lowest order Nédélec approximation on triangular grids and produces a fully diagonal lumped mass matrix. Unfortunately, the method is stable only if the underlying mesh is of the Voronoi type. An extension to tetrahedral grids has also been proposed [**78**], but the stability conditions are rather restrictive, depending heavily on the mesh quality. Mass lumping in $H(\mathrm{curl})$ has been studied more systematically in the late 90s by Cohen and Monk [**37**] for hexahedral grids and by Elmkies and Joly in [**54, 55**] for simplicial elements. These works focus on a dispersion error analysis and a rigorous convergence analysis was missing until recently [**50**], when we were able to close this gap.

Another, very efficient way of approximating Maxwell's equations in time domain is given by the finite difference time domain method, originally proposed by Yee [**114**] in the 60s, and later studied intensively, e.g. by Taflove and co-workers; see [**104**] and the survey [**105**]; see also, Weiland [**109**], who laid the foundations for the *finite integration technique* (FIT) for Maxwell's equations, which leads to similar algebraic systems as the FDTD method, but has much stronger roots in the field of discrete exterior calculus. On rectangular grids and for isotropic smoothly varying materials, both methods result in extremely efficient explicit time-stepping schemes with diagonal mass matrices and second-order accuracy in space and time. The systematic approximation of complex geometries and anisotropic and/or strongly varying coefficients, however, leads to severe complications, which require non-trivial modifications that negatively affect stability and accuracy. In the course of this thesis, we will provide a natural generalization of the FDTD/FIT methods to unstructured grids, which is backed by a systematic stability and convergence analysis. In contrast to related results in [**33**], our approach is again backed up by a strong relation to mixed finite element methods.

Outline of the main results

As announced in the previous discussion, we seek to construct and analyze new mass lumping strategies for problems in $H(\mathrm{curl})$ and $H(\mathrm{div})$ on unstructured grids and to investigate applications in the context of wave propagation and porous medium flow. The goal is to significantly improve the efficiency of mixed finite element methods without sacrificing their stability and accuracy. While our theoretical findings are in principle applicable to higher-order approximations, we will focus in particular on approximations of first and second order, for which the proposed methods seem to provide the highest gain in efficiency. For approximations of a very high order, discontinuous Galerkin methods seem to be rather competitive, and at least on unstructured grids, the advantage of the mass lumping strategies is not so clear. Before going into the details, let us briefly outline the contents of the thesis and sketch our main contributions.

Chapter 2: A multipoint flux mixed finite element method

We start with studying mass lumping in $H(\mathrm{div})$ in the context of mixed finite element approximations of porous medium flow; see (1.10)–(1.11). First order approximations with mass lumping on simplicial grids have already been considered in [**112**] based on the finite element pair $\mathrm{BDM}_1 - \mathrm{P}_0$. After briefly reviewing this method, we present a second-order accurate approximation with mass lumping based on $\mathrm{RT}_1 - \mathrm{P}_1$ finite elements. The error analysis we conduct is not just a pure extension of the theory presented in [**112**], but requires new ideas that allow us to reduce the assumptions on the quadrature rule. In contrast to the first order method of Wheeler and Yotov, the proposed mass lumping does not affect the convergence orders of the mixed finite element approximation, which is also verified by numerical tests. The main results of this chapter have been published in [**48**], where we also extended the analysis to arbitrary order approximations and different element types, such as quadrilaterals, hexahedra, and even prismatic elements.

Chapter 3: A second order finite element method for poroelasticity

We then apply the methodology developed in Chapter 2 to design and analyze an efficient mixed finite element method for poroelasticity. We consider the three-field formulation in a quasi-static setting, given by

$$\begin{aligned} -\mathrm{div}(2\mu\,\epsilon(u) + \lambda\,\mathrm{div}(u)I) + \alpha\nabla p &= f, \\ \alpha\,\mathrm{div}(\partial_t u) + \mathrm{div}(v) &= g, \\ \mathcal{K}^{-1}v + \nabla p &= 0. \end{aligned}$$

Note that the lower right part of the system amounts to the equations (1.10)–(1.11) governing porous medium flow, but now the coupled deformation of the elastic porous matrix is taken into account. Mass lumping here allows to eliminate the seepage velocity v from the system and to obtain a two-field formulation involving only the elastic deformation u and the pressure p. For our analysis, we consider here a second-order approximation based on the finite element triplet $P_2^+ - P_1 - \mathrm{RT}_1$ with mass lumping. Here P_2^+ represents

an extension to the quadratic polynomial space P_2 by appropriate bubble functions. Utilizing some non-trivial extensions of the theoretical results of the first chapter allows us to show that the resulting method is second-order accurate in all variables. Second-order convergence is also established for an appropriate implicit time-stepping scheme, which is similar to the trapezoidal rule. The results presented in this chapter are under preparation for publication.

Chapter 4: Mass lumping for wave propagation in H(div)

We continue with taking advantage of our theoretical findings concerning mass lumping in $H(\mathrm{div})$ for the application to acoustic wave propagation problems. In [**47**], we already investigated mass lumping for the discretization of (1.8)–(1.9) by the mixed finite element pair BDM_1–P_0. We showed that the method is only first-order convergent, but proved super-convergence for the pressure averages, which allowed us to analyze post-processing strategies that grant second-order accuracy in both variables. Unfortunately, the post-processing for the velocity requires the solution of a global system in space. After briefly summarizing these results, we also present our more recent work [**49**], in which we construct a method that can be interpreted as a generalization of the FDTD method to unstructured grids. After this, we present our results published in [**51**], which are concerned with second-order approximations for acoustic wave propagation based on the RT_1–P_1 finite element with mass lumping. In addition, we also provide a full convergence analysis for the explicit time stepping scheme (1.3) mentioned in the introduction. The resulting method is of second-order convergence in space and time.

Chapter 5: Mass lumping in H(curl) for electromagnetic waves

In the last chapter, we extend our considerations to electromagnetic wave propagation modeled by the time dependent Maxwell's equations (1.12)–(1.13). While the two-dimensional case can be treated by slightly modifying our results for wave propagation in $H(\mathrm{div})$, the extension to three-dimensional problems in $H(\mathrm{curl})$ requires some significant modifications. We again start with studying approximations of the first order and present a method that can be understood as a variational extension of the FDTD method to unstructured grids. We then turn to second order approximations and present our results published in [**50**], where we provide a detailed error analysis of a method proposed by Elmkies and Joly [**55**], and show that it is second order accurate if, and only if, the true solution satisfies a certain compatibility condition. Moreover, we presented a modification of this method that yields second-order accuracy without this restriction. A full convergence analysis of this method, including time discretization, is given and second order accuracy in space and time is established.

2. A multipoint flux mixed finite element method for porous medium flow

In this chapter, we investigate the efficient numerical approximation of fluid flow in porous media by mixed finite element methods with mass lumping. As a simple model for porous medium flow, we consider the Darcy problem

$$u + \mathcal{K}\nabla p = 0 \qquad \text{in } \Omega, \tag{2.1}$$

$$\operatorname{div} u = f \qquad \text{in } \Omega. \tag{2.2}$$

Here u is the flow velocity, also called flux in the sequel, p is the fluid pressure, f is the mass flux source term, and $\mathcal{K}$ denotes the hydraulic conductivity tensor.

The two equations describe *Darcy's law* and the conservation of mass, respectively. Equation (2.1) is a phenomenologically derived and experimentally validated constitutive equation describing a linear relationship between velocity and the pressure gradients. We refer to [**16**] for background on the modeling and various extensions. We consider a bounded domain Ω and assume, for simplicity, that the pressure satisfies

$$p = 0 \qquad \text{on } \partial\Omega. \tag{2.3}$$

More general boundary conditions can be treated with minor modifications to our arguments.

Inserting u according to (2.1) into (2.2) leads to the generalized Poisson equation

$$-\operatorname{div}(\mathcal{K}\nabla p) = f \qquad \text{in } \Omega \tag{2.4}$$

for the pressure only. Existence and uniqueness of weak solutions of (2.3)–(2.4) follows by the Lax-Milgram lemma, [**56, Chapter 6.2**]. From this, one can also deduce the unique solvability of (2.2)–(2.3) in an appropriate sense; see Lemma 2.1.1 below.

The discretization of the weak form of (2.3)–(2.4) can be done by H^1-conforming finite element methods. Convergence results are well-known and hold under rather general conditions for the problem data. Unfortunately, the method is known to suffer from rough coefficients $\mathcal{K}$ which typically appear in porous media. Further note that Darcy's law (2.1), which is of a phenomenological nature, is incorporated directly on the discrete level, while the essential conservation law (2.2) is only satisfied in a weak sense. From a physical point of view, it seems more natural to preserve (2.2) in a strong sense instead. Moreover, the flow velocity $u_h = -\mathcal{K}\nabla p_h$, which is in many cases the actual quantity of interest, is approximated with less accuracy than the pressure p_h.

An alternative approach that allows guaranteeing mass conservation explicitly also on the discrete level is offered by finite volume methods [**57**]. Their basic idea is to integrate (2.2) over small control volumes and subsequently apply the divergence theorem, which leads to integrals over the cell boundary involving the normal fluxes $n \cdot u$. In the *two-point flux* approximation [**57**], the flux term $n \cdot u$ is approximated by the difference in the values of two neighboring elements. This technique works well on regular orthogonal grids, as well as for Delaunay or Voronoi meshes satisfying certain angle conditions [**57**], but may lead to instabilities on general meshes and for anisotropic coefficients $\mathcal{K}$. To circumvent this issue, so-called *multipoint flux approximations* have been considered, which gather the flux information from multiple neighboring elements [**1**]. A related discretization approach is given by discontinuous Galerkin methods [**8, 9**], which offer a systematic way of interpreting and generalizing finite volume methods in a variational setting. This allows the extension to arbitrary order methods and enables a rigorous error analysis. All these approaches lead to algebraic systems of the form

$$\mathsf{K}\mathsf{p} = \mathsf{f},$$

where K is discrete approximation of $-\mathrm{div}(\mathcal{K}\nabla\cdot)$ and p is the coefficient vector of the discontinuous approximation of p_h of the pressure p in (2.4).

Another well-established discretization strategy that leads to mass conservative schemes is the *mixed finite element methods*, see [**19**] for a comprehensive survey. Rigorous convergence results for these methods are available under minimal assumptions on both, data and mesh. A downside is that their straightforward algebraic realization leads to saddle-point problems of the form

$$\begin{pmatrix} \mathsf{M} & -\mathsf{B}^\top \\ \mathsf{B} & 0 \end{pmatrix} \begin{pmatrix} \mathsf{u} \\ \mathsf{p} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathsf{f} \end{pmatrix}, \tag{2.5}$$

which involve substantially more degrees of freedom and yield indefinite systems. This complicates the efficient computation of the solution to some extent; see [**27, 53, 22**].

In [**112**], Wheeler and Yotov proposed a modification of a mixed finite element method that allows the reduction of (2.5) to a positive definite system of the form

$$\mathsf{B}^\top \mathsf{M}_\mathsf{h}^{-1} \mathsf{B}\mathsf{p} = \mathsf{B}^\top \mathsf{f}.$$

Here M_h is a block-diagonal approximation of the mass matrix M in (2.5) obtained through inexact numerical integration, which we refer to as *(nodal) mass lumping* in this thesis. Since the matrices B and $\mathsf{M}_\mathsf{h}^{-1}$ are sparse, the system matrix $\mathsf{K}_\mathsf{h} = \mathsf{B}^\top \mathsf{M}_\mathsf{h}^{-1} \mathsf{B}$ provides a sparse approximation of the differential operator $-\mathrm{div}(\mathcal{K}\nabla\cdot)$, similar to a finite volume approximation. The algebraic flux velocity variable u can then be computed by

$$\mathsf{u} = -\mathsf{M}_\mathsf{h}^{-1} \mathsf{B}\mathsf{p}. \tag{2.6}$$

Note that, since $\mathsf{M}_\mathsf{h}^{-1}\mathsf{B}$ is sparse, the flux u can be constructed locally from p. Because of its similarity to multipoint flux finite volume schemes, the method has been termed *"the multipoint flux mixed finite element"* (MFMFE) method in [**112**].

Full convergence analysis for first-order MFMFE approximations on structured and unstructured grids was given in [**112**] the framework of inexact Galerkin approximations. In [**70**], the extension to meshes consisting of (slightly perturbed) affine hexahedral meshes was considered, which required the construction of an extension of the BDDF finite element space [**25**]. The case of a general quadrilateral or hexahedral grid, resulting from non-affine transformations of corresponding reference elements, was treated in [**110**]. A similar analysis was developed previously for the investigation of related finite difference and finite volume methods [**2, 77**]. In a recent preprint [**4**], high order approximations on affine quadrilateral and hexahedral grids have been constructed by using tensor-product Gauss-Lobatto quadrature rules and appropriate extensions of RT_k elements.

While the construction of higher-order MFMFE methods on quadrilateral and hexahedral grids is rather straightforward, the formulation and analysis of higher-order methods on simplicial meshes are more difficult. In this chapter, we discuss such a higher order MFMFE method on unstructured triangular and tetrahedral meshes. For ease of presentation, we focus on the second-order approximation, but our analysis reveals the general ingredients that are required to obtain higher-order approximations on various types of elements. More details can be found in our publication [**48**] and the concluding discussion.

Outline

Let us briefly sketch the outline of the sections contained within this chapter.

Section 2.1. *Notation and preliminaries:* We give a brief overview regarding the existence, uniqueness, and regularity of solutions to the Darcy problem.

Section 2.2. *Mixed finite element approximation:* As a next step, we introduce the standard mixed finite element approximation, state basic convergence results, and also recall the first order MFMFE method from [**112**].

Section 2.3. *A second order multipoint flux mixed finite element method:* We introduce the second order approximation and make statements regarding the existence and uniqueness of solutions to the discrete problem. We then provide a comprehensive error analysis, establishing second-order convergence in both variables, as well as super-convergence of the pressure averages.

Section 2.4. *Implementation and numerical tests:* After the theoretical analysis, we specify appropriate basis functions that lead to the desired block-diagonal structure of the mass matrix. We do this for both the 2D and 3D simplicial cases. Moreover, we provide numerical results that verify and illustrate our analytical findings.

Section 2.5 *Discussion:* In closing this chapter, we give a short outlook on the construction of higher-order MFMFE methods and briefly comment on the comparison to other methods, such as hybridization and dG methods.

2.1. Notation and preliminaries

We denote by $L^p(\Omega)$ and $W^{k,p}(\Omega)$ the standard Lebesgue and Sobolev spaces which are equipped with their usual norms. We further write $H^k(\Omega) = W^{k,2}(\Omega)$ and define

$$H(\mathrm{div},\Omega) \coloneqq \{u \in L^2(\Omega)^d \,|\, \mathrm{div}\, u \in L^2(\Omega)\}$$

for $d = 2, 3$ with norm $\|u\|_{H(\mathrm{div},\Omega)} = (\|u\|^2_{L^2(\Omega)} + \|\mathrm{div}\, u\|^2_{L^2(\Omega)})^{1/2}$. Throughout our analytical considerations, we utilized the following assumptions:

(A1) $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a bounded Lipschitz domain and $\mathcal{K} \in L^\infty(\Omega)^{d\times d}$ such that $\mathcal{K}(x)$ is symmetric and uniformly positive definite for a.e. $x \in \Omega$, i.e.

$$\underline{k}|\xi|^2 \le \xi^\top \mathcal{K}(x)\xi \le \overline{k}|\xi|^2, \tag{2.7}$$

for some constants $0 < \underline{k}, \overline{k} < \infty$. Furthermore, let $f \in L^2(\Omega)$.

2.1.1. Well-posedness

We start by establishing the well-posedness of our model problem.

Lemma 2.1.1 (Well-posedness)**.**
Let assumption (A1) hold. Then (2.1)–(2.3) *has a unique solution* $u \in H(\mathrm{div},\Omega) \cap L^r(\Omega)$ *and* $p \in W^{1,r}(\Omega)$ *for some* $r > 2$ *and*

$$\|u\|_{H(\mathrm{div},\Omega)} + \|u\|_{L^r(\Omega)} + \|p\|_{W^{1,r}(\Omega)} \le C\|f\|_{L^2(\Omega)}$$

with positive constants r, C *depending only on* Ω *and the bounds in* (2.7)*.*

Proof. The system (2.1)–(2.3) is equivalent to the boundary value problem (2.3)–(2.4), for which existence of a unique weak solution $p \in H^1_0(\Omega)$ follows readily from the Lax-Milgram theorem; see [**56**]. From the results of [**63, 64**], we can further infer that $p \in W^{1,r}(\Omega)$ for some $r > 2$, depending only on the Lipschitz character of Ω and the bounds for $\mathcal{K}$. Setting $u = -\mathcal{K}\nabla p \in L^2(\Omega)^d$ yields $u \in L^2(\Omega)$ and also

$$\|u\|_{L^r(\Omega)} = \|\mathcal{K}\nabla p\|_{L^r(\Omega)} \le \|\mathcal{K}\|_{L^\infty(\Omega)}\|\nabla p\|_{L^r(\Omega)} \le c\|p\|_{W^{1,r}(\Omega)} \le C\|f\|_{L^2(\Omega)}.$$

From (2.4) one can also see that $\mathrm{div}\, u = \mathrm{div}(-\mathcal{K}\nabla p) = f \in L^2(\Omega)$. This immediately yields $u \in H(\mathrm{div},\Omega)$ and the corresponding bound for $\|u\|_{H(\mathrm{div},\Omega)}$. □

As a starting point for the construction of discretization methods, we consider the following variational characterization of solutions to our model problem.

Lemma 2.1.2 (Variational characterization)**.**
Let $(u, p) \in H(\mathrm{div},\Omega) \times H^1_0(\Omega)$ *denote a solution of* (2.1)–(2.3)*. Then*

$$(\mathcal{K}^{-1}u, v) - (p, \mathrm{div}\, v) = 0 \qquad \forall\, v \in H(\mathrm{div},\Omega), \tag{2.8}$$

$$(\mathrm{div}\, u, q) = (f, q) \qquad \forall\, q \in L^2(\Omega). \tag{2.9}$$

Proof. The identity (2.9) follows by testing (2.2) with a $q \in L^2(\Omega)$ and integrating over the domain Ω. For (2.8), we test (2.1) with a function $v \in H(\mathrm{div}, \Omega)$ and compute

$$0 = (\mathcal{K}^{-1}u, v) + (\nabla p, v) = (\mathcal{K}^{-1}u, v) - (p, \mathrm{div}\, v) + (p, n \cdot v)_{L^2(\partial\Omega)}.$$

The boundary term vanishes as a result of condition (2.3). □

Remark 2.1.3. The variational identities (2.8)–(2.9) make sense for functions $(u, p) \in H(\mathrm{div}, \Omega) \times L^2(\Omega)$. By Lemma 2.1.1, we already know that a solution with even a bit more regularity exists. Alternatively, the existence of a unique solution $(u, p) \in H(\mathrm{div}, \Omega) \times L^2(\Omega)$ could also be established by the Brezzi theorem; see [**24**]. The essential ingredient here is the inf-sup condition

$$\sup_{v \in H(\mathrm{div},\Omega)} \frac{(\mathrm{div}\, v, q)}{\|v\|_{H(\mathrm{div},\Omega)}} \geq \beta \|q\|_{L^2(\Omega)} \qquad \forall q \in L^2(\Omega), \tag{2.10}$$

which is equivalent to the surjectivity of the divergence operator $\mathrm{div} : H(\mathrm{div}, \Omega) \to L^2(\Omega)$. This property can easily be verified by integration and elementary estimates.

2.2. Mixed finite element approximation

In the following, we consider discretization schemes based on Galerkin approximations of the variational principle (2.8)–(2.9). Before considering specific examples of approximation spaces $V_h \subset H(\mathrm{div}, \Omega)$ and $Q_h \subset L^2(\Omega)$, let us briefly introduce some of our notation and recall some basic results for finite element approximations.

2.2.1. Preliminaries

By $\mathcal{T}_h = \{K_n : n = 1, \ldots, N\}$ we denote a geometrically conforming (regular) family of decompositions of the domain Ω into d–dimensional simplices K_n, in the sense of [**32**]. We write h_K and ρ_K for the diameter and inner circle radius of the element K and call $h = \max_K h_K$ the mesh size. For our subsequent analysis, we always assume that

(A2) $\mathcal{T}_h$ is a γ-shape regular simplicial mesh of Ω, i.e. there exists a constant $\gamma > 0$ such that $\gamma h_K \leq \rho_K \leq h_K$ for all $K \in \mathcal{T}_h$. Moreover, $\mathcal{K} \in W^{1,\infty}(\mathcal{T}_h)^{d\times d}$.

Here $W^{k,p}(\mathcal{T}_h)$ denotes the broken Sobolev space of piecewise regular functions over the mesh $\mathcal{T}_h$. The assumption on the hydraulic conductivity tensor $\mathcal{K} \in W^{1,\infty}(\mathcal{T}_h)$ is made to ensure that point evaluations of $\mathcal{K}$ are well defined on the given mesh, which is required for some of our results later on. We further introduce

$$H^k(\mathcal{T}_h) = \{q \in L^2(\Omega) : q|_K \in H^k(K), \quad \forall K \in \mathcal{T}_h\} \tag{2.11}$$

with norm $\|v\|^2_{H^k(\mathcal{T}_h)} = \sum_{K\in\mathcal{T}_h} \|u\|^2_{H^k(\mathcal{K})}$. In a similar manner, we denote by

$$P_k(\mathcal{T}_h) = \{v \in L^2(\Omega) : v|_T \in P_k(K)\} \tag{2.12}$$

the space of piecewise polynomials over the mesh $\mathcal{T}_h$, with $P_k(K)$ representing the space of polynomials of degree at most k over the element K.

2.2.2. Standard mixed finite element approximations in H(div)

For later reference, let us briefly recall some results about the standard mixed finite element approximation of (2.8)–(2.9). To approximate functions in $H(\mathrm{div},\Omega)$ and $L^2(\Omega)$, we consider the finite element spaces

$$V_h = \{v_h \in H(\mathrm{div},\Omega) \ : \ v_h|_K \in \mathcal{M}_k(K)\} \quad \text{and} \quad Q_h = P_k(\mathcal{T}_h). \tag{2.13}$$

As candidates for the local velocity spaces, we use

$$\mathcal{M}_k(K) = \mathrm{BDM}_{k+1}(K) \coloneqq P_{k+1}(K)^d \qquad \text{or} \tag{2.14}$$
$$\mathcal{M}_k(K) = \mathrm{RT}_k(K) \coloneqq P_k(K)^d \oplus \{x \cdot P_k^h(K)\}. \tag{2.15}$$

These spaces were introduced in [**26**] and [**97**]; also see [**19**] for a comprehensive survey. We then consider the following discretization scheme.

Problem 2.2.1 (Exact Galerkin approximation)**.**
Find $(u_h, p_h) \in V_h \times Q_h$ such that

$$(\mathcal{K}^{-1}u_h, v_h) - (p_h, \mathrm{div}\, v_h) = 0 \qquad \forall v_h \in V_h, \tag{2.16}$$
$$(\mathrm{div}\, u_h, q_h) = (f, q_h) \qquad \forall q_h \in Q_h. \tag{2.17}$$

Well-posedness and convergence results for this method are well established, see e.g. [**27, 18**]. The following assertions are collected from [**19, Proposition 7.1.2**] and [**27**].

Lemma 2.2.2 (Error estimates for the standard mixed finite element method)**.**
Let (A1)–(A2) hold and let (u,p) denote a sufficiently smooth solution of (2.1)–(2.2)*. Then for V_h and Q_h as defined in* (2.13)*, there exists a unique solution (u_h, p_h) to Problem 2.2.1 and the following error estimates are valid:*

$$\|u - u_h\|_{L^2(\Omega)} \le \begin{cases} Ch^{k+2}\|u\|_{H^{k+2}(\mathcal{T}_h)} & \text{if } \mathcal{M}_k(K) = BDM_{k+1}(K), \\ Ch^{k+1}\|u\|_{H^{k+1}(\mathcal{T}_h)} & \text{if } \mathcal{M}_k(K) = RT_k(K), \end{cases}$$
$$\|\mathrm{div}(u - u_h)\|_{L^2(\Omega)} \le Ch^{k+1}\|\mathrm{div}\, u\|_{H^{k+1}(\mathcal{T}_h)},$$
$$\|p - p_h\|_{L^2(\Omega)} \le Ch^{k+1}(\|u\|_{H^{k+1}(\mathcal{T}_h)} + \|p\|_{H^{k+1}(\mathcal{T}_h)}).$$

If Ω is convex and $\mathcal{K} \in W^{1,\infty}(\Omega) \cap W^{2,\infty}(\mathcal{T}_h)$, then further

$$\|\pi_h^0(p - p_h)\|_{L^2(\Omega)} \le Ch^{k+2}\|u\|_{H^{k+1}(\mathcal{T}_h)}.$$

These results will serve as a comparison to the mass lumped methods we introduce next. The last estimate shows that the pressure averages on every element converge faster than the pressure itself, which cannot converge faster than with first order. This behavior is called *super-convergence*, and post-processing strategies can be formulated to construct better approximations for the pressure; see [**103**] and Section 2.3.6 below.

2.2.3. The first order multipoint flux mixed finite element method

We now recall the approach introduced in [**112**]. It is based on modifying (2.16)–(2.17) by replacing the scalar product $(\mathcal{K}^{-1}u_h, v_h)$ by an approximation $(\mathcal{K}^{-1}u_h, v_h)_h$ which is obtained by appropriate numerical quadrature. The approximation spaces for the first order method are chosen as

$$V_h = \{v_h \in H(\mathrm{div}, \Omega) \ : \ v_h|_K \in \mathrm{BDM}_1(K)\} \quad \text{and} \quad Q_h = P_0(\mathcal{T}_h). \tag{2.18}$$

Note that the BDM_1 element has two degrees of freedom per edge on the triangle and three degrees of freedom per face on the tetrahedron, which can be shifted to the corresponding vertices; see Figure 2.1 for a sketch. This shows that on every element, exactly d degrees

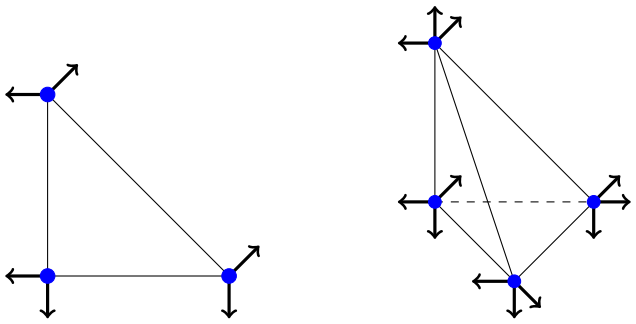

Figure 2.1.: Degrees of freedom for the space $\mathrm{BDM}_1(\widehat{K})$ on the reference triangle (left) and on the reference tetrahedron (right). The blue dots represent the locations of the quadrature points.

of freedom can be associated with each vertex. We then introduce the local quadrature rule

$$\mathcal{Q}_K(f) \coloneqq \frac{|K|}{d+1} \sum_{k=1}^{d+1} f(v_k^K)$$

also known as the vertex rule, and note that $Q_K(f)$ integrates all linear polynomials exactly. For piecewise smooth functions u and w, we define the inexact scalar product

$$(\mathcal{K}^{-1}u, w)_h \coloneqq \sum\nolimits_{K\in\mathcal{T}_h} \mathcal{Q}_K((\mathcal{K}^{-1}u)\cdot w). \tag{2.19}$$

For the numerical approximation of (2.8)–(2.9), we now consider the following method.

Problem 2.2.3 (First order MFMFE method)**.**
Let V_h, Q_h and $(\cdot,\cdot)_h$ be defined as in (2.18)–(2.19)*. Find $(u_h, p_h) \in V_h \times Q_h$ such that*

$$\begin{aligned} (\mathcal{K}^{-1}u_h, v_h)_h - (p_h, \mathrm{div}\, v_h) &= 0 && \forall v_h \in V_h, \\ (\mathrm{div}\, u_h, q_h) &= (f, q_h) && \forall q_h \in Q_h. \end{aligned}$$

For later reference, we summarize the main convergence results for this scheme [**112**].

Lemma 2.2.4 (Error estimates for the first order method)**.**
Let (A1)–(A2) hold. Then Problem 2.2.3 has a unique solution (u_h, p_h) *that satisfies*

$$\|u - u_h\|_{L^2(\Omega)} + \|p - p_h\|_{L^2(\Omega)} \leq Ch(\|u\|_{H^1(\mathcal{T}_h)} + \|p\|_{H^1(\mathcal{T}_h)}),$$
$$\|\mathrm{div}(u - u_h)\|_{L^2(\Omega)} \leq Ch\|\mathrm{div}\, u\|_{H^1(\mathcal{T}_h)}.$$

If Ω *is convex and* $\mathcal{K} \in W^{2,\infty}(\mathcal{T}_h) \cap W^{1,\infty}(\Omega)$*, then*

$$\|\pi_h^0(p - p_h)\|_{L^2(\Omega)} \leq Ch^2(\|u\|_{H^1(\mathcal{T}_h)} + \|\mathrm{div}\, u\|_{H^1(\mathcal{T}_h)}).$$

Remark 2.2.5. According to Lemma 2.2.2, the standard mixed finite element approximation yields second-order convergence in the velocity, i.e. $\|u - u_h\|_{L^2(\Omega)} = O(h^2)$. In contrast to that, only first-order convergence is attained for the MFMFE method. This loss in accuracy is also observed in numerical experiments. Surprisingly, the pressure averages still exhibit super-convergence. To the best of our knowledge, there seems no way to recover second-order accurate velocity approximations by local post-processing. Thus, mass lumping leads to a true reduction in the order of convergence.

2.3. A second-order multipoint flux mixed finite element method

By extending the ideas of [**112**], we now construct and analyze a second-order MFMFE method. The analysis of this scheme requires some non-trivial new arguments, which in principle also carry over to higher order approximations; see the discussion in [**48**] and at the end of this section. As global approximation spaces, we choose

$$V_h = \{v_h \in H(\mathrm{div}, \Omega) \quad : \quad v_h|_K \in \mathrm{RT}_1(K)\} \qquad \text{and} \qquad Q_h = P_1(\mathcal{T}_h). \tag{2.20}$$

Let us note that the local RT_1 space on triangles has two degrees of freedom per edge and two in the interior, and on tetrahedra, we have three degrees of freedom per face and three in the interior. By formally shifting some of the degrees of freedom towards the vertices, we obtain the particular distribution as depicted in Figure 2.2. Observe that exactly d degrees of freedom are associated with the $d + 2$ point depicted in blue. As a quadrature rule, we now choose

$$\mathcal{Q}_K(f) = |K| \left(\alpha f(m) + \sum_{i=1}^{d+1} \beta\, f(v_i) \right). \tag{2.21}$$

Here m and v_i represent the element midpoint and its vertices, which are the integration points, while α and β are the corresponding weights defined in the caption of Figure 2.2. This quadrature rule is exact for quadratic polynomials on both triangles and tetrahedra. General cubic polynomials are, however, not integrated exactly. We again denote by

$$(\mathcal{K}^{-1}u, w)_h \coloneqq \sum\nolimits_{K \in \mathcal{T}_h} Q_K(\mathcal{K}^{-1}u \cdot w), \tag{2.22}$$

the corresponding inexact scalar product for our problem. For the numerical approximation of the variational problem (2.8)–(2.9), we now consider the following scheme.

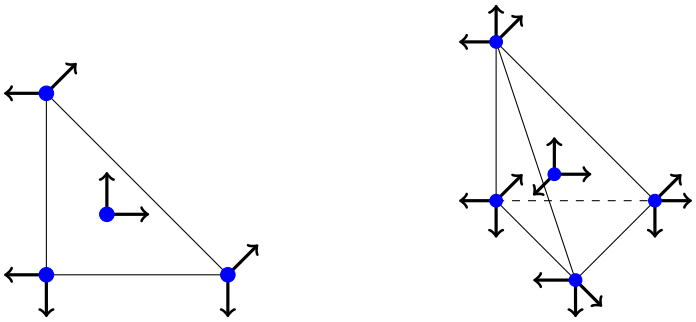

Figure 2.2.: Degrees of freedom for the space $\mathrm{RT}_1(\widehat{K})$ on the triangle (left) and on the tetrahedron (right). The blue dots represent the locations of the quadrature points. The associated weights are $\alpha = \frac{3}{4}$ and $\beta = \frac{1}{12}$ on the triangle and $\alpha = \frac{4}{5}$ and $\beta = \frac{1}{20}$ on the tetrahedron.

Problem 2.3.1 (Second order MFMFE method)**.**
Let V_h, Q_h and $(\cdot,\cdot)_h$ be defined as in (2.20)–(2.22). Find $(u_h, p_h) \in V_h \times Q_h$ such that

$$(\mathcal{K}^{-1}u_h, v_h)_h - (p_h, \operatorname{div} v_h) = 0 \qquad \forall v_h \in V_h, \tag{2.23}$$

$$(\operatorname{div} u_h, q_h) = (f, q_h) \qquad \forall q_h \in Q_h. \tag{2.24}$$

To establish the well-posedness of this inexact Galerkin method, let us mention two auxiliary results. As an immediate consequence of the definition of the inexact scalar product (2.22), one can deduce the following assertion.

Lemma 2.3.2 (Norm equivalence)**.**
Let (A1)–(A2) hold. Then there exist $c_1, c_2 > 0$, independent of the mesh $\mathcal{T}_h$, such that

$$c_1(v_h, v_h) \le (\mathcal{K}^{-1}v_h, v_h)_h \le c_2(v_h, v_h), \qquad \forall v_h \in V_h. \tag{2.25}$$

Proof. It is not difficult to verify that for every $K \in \mathcal{T}_h$, the term $Q_K(\mathcal{K}^{-1}u_h \cdot v_h)$ defines a scalar product on the finite-dimensional space $\mathrm{RT}_1(K)$. The assertion of the lemma is then obtained by scaling arguments and summation over all elements. □

In addition, we require the discrete inf-sup condition

$$\sup_{v_h \in V_h} \frac{(\operatorname{div} v_h, q_h)}{\|v_h\|_{H(\operatorname{div},\Omega)}} \ge \beta_h \|q_h\|_{L^2(\Omega)} \qquad \forall q_h \in Q_h. \tag{2.26}$$

with some constant $\beta_h > 0$. Under assumption (A2), the constant β_h can be chosen independent of h; we refer to [**19, Chapter 7.1.2**] for details. By combining these observations, we can immediately prove the following claim.

Lemma 2.3.3 (Well-posedness)**.**
Let (A1)–(A2) hold. Then Problem 2.3.1 admits a unique solution.

Proof. Since V_h and Q_h are finite-dimensional, it suffices to show uniqueness. We set $f = 0$ and test (2.23) and (2.24) with $v_h = u_h$ and $q_h = p_h$ and add the two, which yields

$$c_1\|u_h\|^2_{L^2(\Omega)} \le (\mathcal{K}^{-1}u_h, u_h) = (p_h, \operatorname{div} u_h) = 0,$$

where we used Lemma 2.3.2 in the first step. This implies that $u_h = 0$. Using the discrete inf-sup condition (2.26), equation (2.23) and Lemma 2.3.2 also leads to

$$\beta_h \|p_h\|_{L^2(\Omega)} \leq \sup_{v_h \in V_h} \frac{(\operatorname{div} v_h, p_h)}{\|v_h\|_{H(\operatorname{div},\Omega)}} = \sup_{v_h \in V_h} \frac{(\mathcal{K}^{-1} u_h, v_h)_h}{\|v_h\|_{H(\operatorname{div},\Omega)}} \leq c \|u_h\|_{L^2(\Omega)} = 0,$$

which implies $p_h = 0$. This already completes the proof of the Lemma. □

From the approximation properties of the involved finite element spaces, we would expect the method to be second-order accurate in all components; compare with Lemma 2.2.2. Our main focus in this section is to prove that this is in fact the case.

Theorem 2.3.4 (Error estimates)**.**
Let (A1)–(A2) hold and (u, p) and (u_h, p_h) be solutions to (2.8)–(2.9) and (2.23)–(2.24). Further assume that $\mathcal{K} \in W^{2,\infty}(\mathcal{T}_h)$ and that u, p are sufficiently smooth. Then

$$\|u - u_h\|_{L^2(\Omega)} + \|p - p_h\|_{L^2(\Omega)} \leq C h^2 (\|u\|_{H^2(\mathcal{T}_h)} + \|p\|_{H^2(\mathcal{T}_h)})$$

and $\|\operatorname{div}(u - u_h)\|_{L^2(\Omega)} \leq C h^2 \|\operatorname{div} u\|_{H^2(\mathcal{T}_h)}$. If Ω is convex and $\mathcal{K} \in W^{1,\infty}(\Omega)$, then also

$$\|\pi_h^0 p - p_h\|_{L^2(\Omega)} \leq C h^3 \|u\|_{H^2(\mathcal{T}_h)}.$$

Remark 2.3.5. In contrast to the first order method discussed in the previous section, the inexact numerical approximation does not lead to a loss of accuracy in our second order MFMFE method. To be more precise: Super-convergence in $\|\pi_h^1 p - p_h\|_{L^2(\Omega)}$ is lost, but this is of little relevance since this can be overcome by local post-processing; see Section 2.3.6 below.

The remainder of this section is devoted to a detailed proof of Theorem 2.3.4. We start by introducing some notation and auxiliary results.

2.3.1. Projection operators

In our analysis, we will make repeated use of the following projection operators:

$$\begin{aligned} \Pi_h &: L^p(\Omega) \cap H(\operatorname{div},\Omega) \to \mathrm{RT}_1(\mathcal{T}_h) \cap H(\operatorname{div},\Omega), \\ \Pi_h^* &: L^p(\Omega) \cap H(\operatorname{div},\Omega) \to \mathrm{BDM}_1(\mathcal{T}_h) \cap H(\operatorname{div},\Omega), \end{aligned}$$

with some $p > 2$, which are chosen as the standard interpolation operators for the RT_1 and BDM_1 spaces; see [**19, Section 2.5**]. We further denote by $\pi_h^k : L^2(\Omega) \to P_k(\mathcal{T}_h)$, the L^2-orthogonal projection onto $P_k(\mathcal{T}_h)$ defined by

$$(\pi_h^k v, v_h) = (v, v_h) \qquad \forall v_h \in P_k(\mathcal{T}_h). \tag{2.27}$$

The same symbol will be used for the projection of vector-valued functions. The well-known error estimates for these projection operators can be summarized as

$$\begin{aligned} \|u - \Pi_h^* u\|_{H^\ell(\mathcal{T}_h)} &\leq C h^{r-\ell} \|u\|_{H^{r+1}(\mathcal{T}_h)}, \\ \|u - \Pi_h u\|_{H^\ell(\mathcal{T}_h)} &\leq C h^{r+1-\ell} \|u\|_{H^{r+1}(\mathcal{T}_h)}, \\ \|\mathrm{div}(u - \Pi_h u)\|_{L^2(\mathcal{T}_h)} &\leq C h^r \|\mathrm{div}\, u\|_{H^r(\mathcal{T}_h)}, \end{aligned} \tag{2.28}$$

for $1 \leq \ell \leq r \leq 2$ as well as

$$\|u - \pi_h^k u\|_{L^2(\Omega)} \leq C h_K^s \|u\|_{H^s(\Omega)}, \tag{2.29}$$

for $0 \leq s \leq k+1$. Furthermore, the operators satisfy the commuting diagram properties

$$\mathrm{div}\, \Pi_h u = \pi_h^1 \mathrm{div}\, u \qquad \text{and} \qquad \mathrm{div}\, \Pi_h^* u = \pi_h^0 \mathrm{div}\, u. \tag{2.30}$$

Auxiliary results

Our error analysis hinges on some particular further properties of the RT_1 finite element spaces, which we summarize for convenience. First, note that

$$\mathrm{div}(\mathrm{BDM}_1(K)) = P_0(K) \qquad \text{and} \qquad \mathrm{div}(\mathrm{RT}_1(K)) = P_1(K). \tag{2.31}$$

Further recall from the definition of RT_k that

$$\mathrm{RT}_1(K) = \mathrm{BDM}_1(K) \oplus \mathcal{B}(K) \qquad \text{with} \qquad \mathcal{B}(K) = x \cdot P_1^h(K).$$

The following results follow readily from this definition and scaling arguments.

Lemma 2.3.6 (Stable splitting)**.**
Let (A2) hold and let $\mathcal{B}(K) = x \cdot P_1^h(K)$ as above. Then

(i) $\dim(\mathcal{B}(K)) = \dim(\mathrm{div}(\mathcal{B}(K))) = d$.

(ii) The expression $\|\mathrm{div}(\cdot)\|_{L^2(K)}$ *defines a norm on* $\mathcal{B}(K)$ *and*

$$\|\nabla v_K^{\mathcal{B}}\|_{L^2(K)} \leq C \|\mathrm{div}(v_K^{\mathcal{B}})\|_{L^2(K)} \qquad \forall v_K^{\mathcal{B}} \in \mathcal{B}(K).$$

(iii) $\|\mathrm{div}(v_K^{\mathcal{B}})\|_{L^2(K)} \leq C \|\mathrm{div}\,(v_K^{B} + v_K^1)\|_{L^2(K)}$, *for all* $v_K^1 \in P_1(K)$.

The constants C in these estimates only depend on the shape regularity of K.

Proof. (i) From (2.31), we know that $\mathrm{div}(\mathrm{BDM}_1(K)) = P_0(K)$ and $\mathrm{div}(\mathrm{RT}_1(K)) = P_1(K)$. The first assumption now follows from the counting argument.

(ii) Let $\widehat{K}$ denote the reference simplex. As a consequence of the dimension count in (i), we know that $\|\mathrm{div}\,(\cdot)\|_{L^2(\widehat{K})}$ defines a norm on $\mathcal{B}(\widehat{K})$. This implies

$$\|\widehat{\nabla} \widehat{v}_K^{\mathcal{B}}\|_{L^2(\widehat{K})} \leq c \|\widehat{\mathrm{div}}(\widehat{v}_K^{\mathcal{B}})\|_{L^2(\widehat{K})}.$$

Further, by mapping to the reference using the Piola transformation (A.1), we obtain

$$\|\nabla v_K^{\mathcal{B}}\|_{L^2(K)} \leq ch^{-d/2}\|\widehat{\nabla}\widehat{v}_K^{\mathcal{B}}\|_{L^2(\widehat{K})} \leq c'h^{-d/2}\|\widehat{\operatorname{div}}(\widehat{v}_K^{\mathcal{B}})\|_{L^2(\widehat{K})} \leq c_\gamma\|\operatorname{div}(v_K^{\mathcal{B}})\|_{L^2(K)},$$

where c_γ depends only on the shape-regularity of K.

(iii) Let $\widehat{K}$ denote the reference simplex and let $\widehat{q}_0 \in \operatorname{div}(\mathrm{BDM}_1(\widehat{K})) = P_0(\widehat{K})$ and let $\widehat{q}_{\mathcal{B}} \in \operatorname{div}(\mathcal{B}(\widehat{K}))$. Moreover, let π_h^0 denote the L^2-projection onto $P_0(\widehat{K})$.

Assume now that $\widehat{q}_{\mathcal{B}} - \pi_h^0\widehat{q}_{\mathcal{B}} = 0$. This implies that $\widehat{q}_{\mathcal{B}} = \pi_h^0\widehat{q}_{\mathcal{B}} \in P_0(\widehat{K})$. Since $P_0(\widehat{K}) \not\subseteq \widehat{\operatorname{div}}\,\mathcal{B}(\widehat{K})$, we infer that $\widehat{q}_{\mathcal{B}} = 0$. Hence $(\mathrm{Id} - \pi_h^0)$ induces a norm on $\widehat{\operatorname{div}}(\mathcal{B}(\widehat{K}))$, i.e. there exists a constant $c > 0$ such that

$$\|\widehat{q}_{\mathcal{B}} - \pi_h^0\widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})}^2 \geq c\|\widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})}^2.$$

By definition of the L^2-projection π_h^0, we know that

$$\|\widehat{q}_{\mathcal{B}} - \pi_h^0\widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})} = \min_{\widehat{q}_0 \in P_0(\widehat{K})} \|\widehat{q}_0 + \widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})}.$$

We can then write

$$\|\widehat{q}_0 + \widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})}^2 \geq \|\widehat{q}_{\mathcal{B}} - \pi_h^0\widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})} \geq c\|\widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})}^2.$$

By a scaling argument, together with assumption (A2), we further obtain

$$\|q_0 + q_{\mathcal{B}}\|_{L^2(K)}^2 \geq c_\gamma\|q_{\mathcal{B}}\|_{L^2(K)}^2,$$

for any $q_0 \in P_0(K)$ and $q_{\mathcal{B}} \in \operatorname{div}(\mathcal{B}(K))$ with a constant c_γ depending only on the shape-regularity of $\mathcal{T}_h$. This completes the proof of the last assertion. □

2.3.2. Estimates for the quadrature error

The error analysis of Problem 2.3.1 requires a detailed analysis of the quadrature. To simplify notation, we introduce the local and global quadrature errors

$$\sigma_{h,K}(u,v) \coloneqq (u,v)_{h,K} - (u,v)_K \qquad \text{and} \qquad \sigma_h(u,v) = \sum\nolimits_{K\in T_h} \sigma_{h,K}(u,v) \tag{2.32}$$

with $(u,v)_{h,\mathcal{K}} = Q_K(u\cdot v)$ and numerical quadrature Q_K as defined in (2.21).

Lemma 2.3.7 (Quadrature error)**.**
The quadrature rule (2.21) *is exact for polynomials of degree $k \leq 2$ on both triangles and tetrahedra. Moreover, if (A1)–(A2) hold, then*

$$\begin{aligned} |\sigma_h(\mathcal{K}^{-1}\Pi_h u, v_h)| &\leq Ch^2\|\mathcal{K}^{-1}\|_{W^{2,\infty}(\mathcal{T}_h)}\|u\|_{H^2(\mathcal{T}_h)}\|v_h\|_{H(\operatorname{div},\Omega)} \qquad \textit{and} \\ |\sigma_h(\mathcal{K}^{-1}u_h, v_h^*)| &\leq Ch^3\|\mathcal{K}^{-1}\|_{W^{3,\infty}(\mathcal{T}_h)}\|u_h\|_{H^2(\mathcal{T}_h)}\|v_h^*\|_{H^1(\mathcal{T}_h)}. \end{aligned}$$

for any $u \in H(\operatorname{div},\Omega)\cap H^2(\mathcal{T}_h)^d$, $u_h, v_h \in V_h$, $v_h^* \in P_1(\mathcal{T}_h)^d$ and for sufficiently regular $\mathcal{K}$. The constant C depends only on the shape-regularity constant γ and the domain Ω.

Proof. The exactness of the quadrature rule can be verified directly. For the error on every element $K \in \mathcal{T}_h$, we then split the error in

$$\begin{aligned}\sigma_{h,K}(\mathcal{K}^{-1}\Pi_h u, v_h) &= \sigma_{h,K}(\mathcal{K}^{-1}\Pi_h u - \pi_h^1(\mathcal{K}^{-1}\Pi_h u), v_h) + \sigma_{h,K}(\pi_h^1(\mathcal{K}^{-1}\Pi_h u), v_h)\\ &= (i) + (ii).\end{aligned}$$

The first term can be estimated by

$$\begin{aligned}(i) &= \sigma_{h,K}(\mathcal{K}^{-1}\Pi_h u - \pi_h^1(\mathcal{K}^{-1}\Pi_h u), v_h)\\ &\leq c h_K^d \|\mathcal{K}^{-1}\Pi_h u - \pi_h^1(\mathcal{K}^{-1}\Pi_h u)\|_{L^\infty(K)}\|v_h\|_{L^\infty(K)}\\ &\leq C h_K^{d+2}\|\nabla^2(\mathcal{K}^{-1}\Pi_h u)\|_{L^\infty(K)}\|v_h\|_{L^\infty(K)},\end{aligned}$$

where we used the exactness of the quadrature rule in the first identity, a scaling argument in the second step, and the approximation properties of the projection π_h^1 in the last. We proceed by further estimating

$$\|\nabla^2(\mathcal{K}^{-1}\Pi_h u)\|_{L^\infty(K)} \leq C'\|\nabla^2\mathcal{K}^{-1}\|_{L^\infty(K)}\|\nabla^2\Pi_h u\|_{L^\infty(K)}.$$

By the usual scaling arguments [**23**], one can show $h_T^{d/2}\|w_h\|_{L^\infty(K)} \leq c'\|w_h\|_{L^2(K)}$ for $w_h = v_h$ and $w_h = \nabla^2\Pi_h u$. Moreover, from (2.28) with $\ell = 2$ and $r = 1$, we deduce

$$\|\Pi_h u\|_{H^2(K)} \leq \|u\|_{H^2(K)} + \|\Pi_h u - u\|_{H^2(K)} \leq (1+c)\|u\|_{H^2(K)},$$

with a uniform constant c for all $K \in \mathcal{T}_h$. This yields

$$(i) \leq c h^2\|\mathcal{K}^{-1}\|_{W^{2,\infty}(K)}\|\Pi_h u\|_{H^2(K)}\|v_h\|_{L^2(K)} \leq c' h^2\|\mathcal{K}^{-1}\|_{W^{2,\infty}(K)}\|u\|_{H^2(K)}\|v_h\|_{L^2(K)}.$$

For the second term, we proceed in a similar fashion and obtain

$$\begin{aligned}(ii) &= \sigma_{h,K}(\pi_h^1(\mathcal{K}^{-1}\Pi_h u) - \pi_h^0(\mathcal{K}^{-1}\Pi_h u), v_h - \pi_h^0 v_h)\\ &\leq c h_K^d\|\pi_h^1(\mathcal{K}^{-1}\Pi_h u) - \pi_h^0(\mathcal{K}^{-1}\Pi_h u)\|_{L^\infty(K)}\|v_h - \pi_h^1 v_h\|_{L^\infty(K)}\\ &\leq C h_K^{d+3}\|\nabla(\mathcal{K}^{-1}\Pi_h u)\|_{L^\infty(K)}\|\nabla^2 v_h\|_{L^\infty(K)}.\end{aligned}$$

By the same arguments as before, we get

$$(ii) \leq C h^3\|\mathcal{K}^{-1}\|_{W^{1,\infty}(K)}\|u\|_{H^1(K)}\|\nabla^2 v_h\|_{L^2(K)}.$$

Using assertion (i) of Lemma 2.3.6, we split $v_h = v_h^1 \oplus v_h^{\mathcal{B}}$ and further estimate

$$\|\nabla^2 v_h\|_{L^2(K)} = \|\nabla^2 v_h^B\|_{L^2(K)} \leq C' h^{-1}\|\nabla v_h^{\mathcal{B}}\|_{L^2(K)} \leq C'' h^{-1}\|\mathrm{div}(v_h^{\mathcal{B}})\|_{L^2(K)}.$$

In the last step, we made use of assertion (ii) of Lemma 2.3.6. By the claim (iii) of Lemma 2.3.6, we can bound $\|\mathrm{div}(v_h^B)\|_{L^2(K)} \leq C\|\mathrm{div}(v_h)\|_{L^2(K)}$. A combination of these results and summation over all elements now already yields the first estimate of the lemma. For the second, we proceed in a similar fashion. On every element $K \in T_h$, we split the quadrature error into

$$\begin{aligned}\sigma_{h,K}(\mathcal{K}^{-1}u_h, v_h^*) &= \sigma_{h,K}(\mathcal{K}^{-1}u_h - \pi_h^1(\mathcal{K}^{-1}u_h), v_h^* - \pi_h^0 v_h^*) + \sigma_{h,K}(\mathcal{K}^{-1}u_h, \pi_h^0 v_h^*)\\ &= \sigma_{h,K}(\mathcal{K}^{-1}u_h - \pi_h^1(\mathcal{K}^{-1}u_h), v_h^* - \pi_h^0 v_h^*) + \sigma_{h,K}(\mathcal{K}^{-1}u_h - \pi_h^2(\mathcal{K}^{-1}u_h), \pi_h^0 v_h^*)\\ &= (i) + (ii).\end{aligned}$$

Here, the exactness of the quadrature rule (2.22) for polynomials of degree $k \leq 2$ was employed several times. The remaining terms can now be estimated just as before. □

2.3.3. Second order convergence for the velocity

We have now collected all the technical tools necessary to conduct the error analysis of our method. As a first step, we seek to establish the estimates

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch^2\|u\|_{H^2(\mathcal{T}_h)},$$
$$\|\mathrm{div}(u - u_h)\|_{L^2(\Omega)} \leq Ch^2\|\mathrm{div}\, u\|_{H^2(\mathcal{T}_h)}$$

of Theorem 2.3.4. In the usual manner, we start by splitting the error into an interpolation error and a discrete error component, i.e.

$$u - u_h = (u - \Pi_h u) + (\Pi_h u - u_h).$$

Using (2.28), we can readily estimate the interpolation error by

$$\|u - \Pi_h u\|_{L^2(\Omega)} \leq Ch^2\|\nabla^2 u\|_{L^2(\Omega)} \quad \text{and} \quad \|\mathrm{div}(u - \Pi_h u)\|_{L^2(\Omega)} \leq Ch^2\|\nabla^2 \mathrm{div}\, u\|_{L^2(\Omega)}.$$

We now turn to the discrete error component $w_h = \Pi_h u - u_h$. Using (2.31), we deduce from equations (2.9) and (2.24) that

$$(\mathrm{div}\,(\Pi_h u - u_h), q_h) = (\pi_h^1 \mathrm{div}\, u - \mathrm{div}\, u_h, q_h) = (\mathrm{div}\, u - \mathrm{div}\, u_h, q_h) = 0, \tag{2.33}$$

for all $q_h \in Q_h$. Since $\mathrm{div}\, V_h = Q_h$, see (2.31), we see that

$$\mathrm{div}\, w_h = \mathrm{div}(\Pi_h u - u_h) = 0. \tag{2.34}$$

From the variational characterizations (2.8)–(2.9) and (2.23)–(2.24) of the continuous and the discrete solutions, and the norm equivalence result from Lemma 2.3.2, we get

$$\begin{aligned} c_1\|w_h\|^2_{L^2(\Omega)} &\leq (\mathcal{K}^{-1}(\Pi_h u - u_h), w_h)_h = (\mathcal{K}^{-1}(\Pi_h u - u_h), w_h)_h + (\pi_h^1 p - p_h, \mathrm{div}\, w_h) \\ &= (\mathcal{K}^{-1}(\Pi_h u - u), w_h) + \sigma_h(\mathcal{K}^{-1}\Pi_h u, w_h) = (i) + (ii). \end{aligned}$$

The first term can be estimated by the Cauchy-Schwarz inequality, leading to

$$(i) \leq C\|\Pi_h u - u\|_{L^2(\Omega)}\|w_h\|_{L^2(\Omega)} \leq Ch^2\|u\|_{H^2(\mathcal{T}_h)}\|w_h\|_{L^2(\Omega)}.$$

From Lemma 2.3.7, we obtain for the second term

$$\begin{aligned} (ii) &\leq Ch^2\|\mathcal{K}^{-1}\|_{W^{2,\infty}(\mathcal{T}_h)}\|u\|_{H^2(\mathcal{T}_h)}\|w_h\|_{H(\mathrm{div},\Omega)} \\ &= Ch^2\|\mathcal{K}^{-1}\|_{W^{2,\infty}(\mathcal{T}_h)}\|u\|_{H^2(\mathcal{T}_h)}\|w_h\|_{L^2(\Omega)}, \end{aligned}$$

where we used $\mathrm{div}\, w_h = 0$ in the last step; see (2.34). A combination of the above results shows that

$$\begin{aligned} \|u - u_h\|_{L^2(\Omega)} &\leq \|u - \Pi_h u\|_{L^2(\Omega)} + \|\Pi_h u - u_h\|_{L^2(\Omega)} \\ &\leq Ch^2(2 + \|\mathcal{K}^{-1}\|_{W^{2,\infty}(\mathcal{T}_h)})\|u\|_{H^2(\mathcal{T}_h)}, \end{aligned}$$

which is the estimate for L^2-error in the velocity. From equation (2.33), we obtain

$$\|\mathrm{div}(u - u_h)\|_{L^2(\Omega)} = \|\mathrm{div}\, u - \pi_h \mathrm{div}\, u\|_{L^2(\Omega)} \leq Ch^2\|\mathrm{div}\, u\|_{H^2(\mathcal{T}_h)},$$

which yields the estimate for the error in the divergence. This complete the proof of the error estimates for the velocity. □

2.3.4. Second-order convergence for the pressure

As a next step in the verification of Theorem 2.3.4, we now prove that

$$\|p - p_h\|_{L^2(\Omega)} \leq Ch^2(\|u\|_{H^2(\mathcal{T}_h)} + \|p\|_{H^2(\mathcal{T}_h)}). \tag{2.35}$$

Just as in the estimates for the velocity, we begin by splitting the error into a projection error and a discrete error component, i.e.

$$\|p - p_h\|_{L^2(\Omega)} \leq \|p - \pi_h^1 p\|_{L^2(\Omega)} + \|\pi_h^1 p - p_h\|_{L^2(\Omega)}.$$

The projection error can be estimated by $\|p-\pi_h^1 p\|_{L^2(\Omega)} \leq Ch^2\|p\|_{H^2(\mathcal{T}_h)}$, using the bounds in (2.29). We continue with analyzing the discrete error. For any $v_h \in V_h$, we have

$$\begin{aligned}(\operatorname{div} v_h, \pi_h^1 p - p_h) &= (\operatorname{div} v_h, p - p_h) \\ &= (\mathcal{K}^{-1}u, v_h) - (\mathcal{K}^{-1}u_h, v_h)_h \\ &= (\mathcal{K}^{-1}(u - \Pi_h u), v_h) - (\mathcal{K}^{-1}(u_h - \Pi_h u), v_h)_h + \sigma(\mathcal{K}^{-1}\Pi_h u, v_h) \\ &= (i) + (ii) + (iii).\end{aligned}$$

The first term can be estimated by the interpolation error results of (2.28). The second error component was analyzed in the previous section, and the third error component is again estimated by Lemma 2.3.7. In summary, we obtain

$$(\operatorname{div} v_h, \pi_h^1 p - p_h) \leq Ch^2\|u\|_{H^2(\mathcal{T}_h)}\|v_h\|_{H(\operatorname{div},\Omega)}.$$

By invoking the discrete inf-sup condition (2.26), we finally obtain

$$\beta_h\|\pi_h^1 p - p_h\|_{L^2(\Omega)} \leq \sup_{v\in V_h} \frac{(\operatorname{div} v_h, \pi_h^1 p - p_h)}{\|v_h\|_{H(\operatorname{div},\Omega)}} \leq Ch^2\|u\|_{H^2(\mathcal{T}_h)}.$$

Together with the estimates for the interpolation error, we obtain (2.35). This establishes the first part of the estimates for the pressure. □

2.3.5. Super-convergence for the projected pressure

As the last step in the proof of Theorem 2.3.4, we now show that

$$\|\pi_h^0(p - p_h)\|_{L^2(\Omega)} \leq Ch^3\|u\|_{H^2(\mathcal{T}_h)}. \tag{2.36}$$

Here we use a duality argument and require that Ω is convex and that $\mathcal{K}$ is in $W^{1,\infty}(\Omega)$ globally. Next, let ϕ denote the solution to the auxiliary problem

$$\begin{aligned}-\operatorname{div}(\mathcal{K}\nabla\phi) &= \pi_h^0(p - p_h) && \text{in } \Omega, \\ \phi &= 0 && \text{on } \partial\Omega.\end{aligned}$$

By the convexity of Ω and regularity of $\mathcal{K}$, we know that $\|\phi\|_{H^2(\Omega)} \leq C\|\pi_h^0(p - p_h)\|_{L^2(\Omega)}$; see [**56**]. Further, recall that Π_h^* denotes the interpolation operator for the space BDM_1 and that $\mathrm{BDM}_1(K) = \mathrm{P}_1(K)^d$. Then

$$\begin{aligned}\|\pi_h^0(p - p_h)\|_{L^2(\Omega)}^2 &= -(\pi_h^0(p - p_h), \pi_h^0\operatorname{div}(\mathcal{K}\nabla\phi)) = -(\pi_h^0(p - p_h), \operatorname{div}(\Pi_h^*(\mathcal{K}\nabla\phi))) \\ &= -(\mathcal{K}^{-1}(u - u_h), \Pi_h^*(\mathcal{K}\nabla\phi)) + \sigma_h(\mathcal{K}^{-1}u_h, \Pi_h^*(\mathcal{K}\nabla\phi)) = (i) + (ii).\end{aligned}$$

For dealing with the term (i), we first compute

$$\begin{aligned}-(\mathcal{K}^{-1}(u-u_h),\mathcal{K}\nabla\phi) &= -(u-u_h,\nabla\phi)\\ &= (\mathrm{div}(u-u_h),\phi) = (\mathrm{div}(u-u_h),\phi-\pi_h^1\phi),\end{aligned}$$

which follows from (2.9) and (2.23), the properties of the projections in Section 2.3.1, and the bounds of Section 2.3.3. We can now estimate

$$\begin{aligned}(i) &= -(\mathcal{K}^{-1}(u-u_h),\Pi_h^*(\mathcal{K}\nabla\phi)-\mathcal{K}\nabla\phi) - (\mathcal{K}^{-1}(u-u_h),\mathcal{K}\nabla\phi)\\ &= -(\mathcal{K}^{-1}(u-u_h),\Pi_h^*(\mathcal{K}\nabla\phi)-\mathcal{K}\nabla\phi) + (\mathrm{div}(u-u_h),\phi-\pi_h^1\phi)\\ &\le C(\|u-u_h\|_{L^2(\Omega)}\|\Pi_h^*(\mathcal{K}\nabla\phi)-\mathcal{K}\nabla\phi\|_{L^2(\Omega)} + \|\mathrm{div}(u-u_h)\|_{L^2(\Omega)}\|\phi-\pi_h^1\phi\|_{L^2(\Omega)})\\ &\le C'h^3(\|u\|_{H^2(\mathcal{T}_h)} + \|\mathrm{div}\,u\|_{H^1(\mathcal{T}_h)})\|\phi\|_{H^2(\mathcal{T}_h)} \le C''h^3\|u\|_{H^2(\mathcal{T}_h)}\|\pi_h^0(p-p_h)\|_{L^2(\Omega)}.\end{aligned}$$

Here we used that $\|\Pi_h^*(\mathcal{K}\nabla\phi)-\mathcal{K}\nabla\phi\|_{L^2(\Omega)} \le Ch\|\Pi_h^*\nabla\phi\|_{H^1(\Omega)} \le Ch\|\phi\|_{H^2(\Omega)}$ in the last step. For the second term, we use Lemma 2.3.7, the regularity of $\mathcal{K}$ and $\mathcal{K}^{-1}$, and the shape regularity of the mesh to show that on every element

$$\begin{aligned}(ii)_K &= \sigma_{h,K}(\mathcal{K}^{-1}u_h,\Pi_h^*(\mathcal{K}\nabla\phi))\\ &\le Ch_K^3\|\mathcal{K}^{-1}\|_{W^{3,\infty}(K)}\|u_h\|_{H^2(K)}\|\Pi_h^*(\mathcal{K}\nabla\phi)\|_{H^1(K)}\\ &\le C'h_K^3\big(\|\Pi_Tu-u_h\|_{H^2(K)} + \|\Pi_Ku\|_{H^2(K)}\big)\|\Pi_h^*(\mathcal{K}\nabla\phi)\|_{H^1(K)}\\ &\le C''\big(h_K\|\Pi_Tu-u_h\|_{L^2(K)} + h_K^3\|\Pi_Ku\|_{H^2(K)}\big)\|\Pi_h^*(\mathcal{K}\nabla\phi)\|_{H^1(K)}.\end{aligned}$$

By using the properties of the projection operator Π_h^* from Section 2.3.1, the estimate for the velocity from Section 2.3.3, and summation over all elements, we obtain

$$(ii) \le Ch^3\|u\|_{H^2(\mathcal{T}_h)}\|\pi_h^0(p-p_h)\|_{L^2(\Omega)}.$$

By combination with the previous bounds, we obtain the estimate (2.36). This concludes the proof of Theorem 2.3.4. □

2.3.6. Post-processing and third-order convergence for the pressure

Based on the super-convergence of the pressure averages, we can now define a local post-processing scheme in the spirit of Stenberg [**103**]. We do so by constructing local approximations $\widetilde{p}_h \in P_2(K)$ in the following way.

Problem 2.3.8 (Post-processing for the pressure)**.**
Find $\widetilde{p}_h \in P_2(\mathcal{T}_h)$ such that for all $K \in \mathcal{T}_h$ there holds

$$(\nabla\widetilde{p}_h,\nabla\widetilde{q}_h)_K = -(\mathcal{K}^{-1}u_h,\nabla\widetilde{q}_h)_K \qquad \forall\,\widetilde{q}_h \in P_2(K), \tag{2.37}$$

$$(\widetilde{p}_h,q_h^0)_K = (p_h,q_h^0)_K \qquad \forall\,q_h^0 \in P_0(K). \tag{2.38}$$

Note that $\widetilde{p}_h$ can be computed separately on each element K, rendering the method computationally efficient. Following the analysis of [**103**], we obtain the following bounds.

Lemma 2.3.9 (Error splitting)**.**
Let $\widetilde{p}_h$ be the solution of Problem 2.3.8 and let (u,p) solve (2.8)–(2.9)*. Then*

$$\|p-\widetilde{p}_h\|_{L^2(K)} \leq \|p-\pi_h^2 p\|_{L^2(K)} + \|\pi_h^0(p-p_h)\|_{L^2(K)} + Ch_K(\|u-u_h\|_{L^2(K)} + \|\nabla(\pi_h^2 p - p)\|_{L^2(K)}).$$

Proof. By applying the triangle inequality, we obtain

$$\begin{aligned}\|p-\widetilde{p}_h\|_{L^2(K)} &\leq \|p-\pi_h^2 p\|_{L^2(K)} + \|\pi_h^2 p - \widetilde{p}_h\|_{L^2(K)} \\ &\leq \|p-\pi_h^2 p\|_{L^2(K)} + \|\pi_h^0(\pi_h^2 p - \widetilde{p}_h)\|_{L^2(K)} + \|(id-\pi_h^0)(\pi_h^2 p - \widetilde{p}_h)\|_{L^2(K)} \\ &= (i) + (ii) + (iii).\end{aligned}$$

The first term already appears in the final estimate. For the second, observe that

$$(ii) = \|\pi_h^0(\pi_h^2 p - \widetilde{p}_h)\|_{L^2(K)} = \|\pi_h^0(p-p_h)\|_{L^2(K)},$$

where we used that $\pi_h^0\pi_h^2 p = \pi_h^0 p$ and (2.38). By the optimality of the L^2–projection and the Poincaré inequality, the third term can be estimated by

$$(iii) = \|(\pi_h^2 p - \widetilde{p}_h) - \pi_h^0(\pi_h^2 p - \widetilde{p}_h)\|_{L^2(K)} \leq Ch_K\|\nabla(\pi_h^2 p - \widetilde{p}_h)\|_{L^2(K)}.$$

Using equation (2.8), we obtain for $\widetilde{q}_h = \pi_h^2 p - \widetilde{p}_h$ that

$$\begin{aligned}\|\nabla\widetilde{q}_h\|_{L^2(K)}^2 &= (\nabla(\pi_h^2 p - \widetilde{p}_h), \nabla\widetilde{q}_h)_K \\ &= (\nabla(\pi_h^2 p - p), \nabla\widetilde{q}_h)_K + (\mathcal{K}^{-1}(u-u_h), \nabla\widetilde{q}_h)_K \\ &\leq \big(\|\nabla(\pi_h^2 p - p)\|_{L^2(K)} + C\|u-u_h\|_{L^2(K)}\big)\|\nabla\widetilde{q}_h\|_{L^2(K)}.\end{aligned}$$

Dividing by $\|\nabla\widetilde{q}_h\|_{L^2(K)}$ yields the bound for the term (iii). The assertion of the Lemma now follows by combination with the previous estimates. □

Together with the estimates of Theorem 2.3.4, we obtain the following assertion.

Lemma 2.3.10 (Estimate for the post-processed pressure)**.**
Let (A1)–(A2) hold, assume that Ω is convex and $\mathcal{K} \in W^{1,\infty}(\Omega) \cap W^{3,\infty}(\mathcal{T}_h)$. Then

$$\|p-\widetilde{p}_h\|_{L^2(\Omega)} \leq Ch^3(\|u\|_{H^2(\mathcal{T}_h)} + \|p\|_{H^3(\mathcal{T}_h)}).$$

The constant C in this estimate is independent of the mesh size.

Remark 2.3.11. For the standard mixed finite element method without mass lumping, one can further show that $\|\pi_h^1 p - p_h\|_{L^2(\Omega)} \leq Ch^3$, which is again a super-convergence result. This estimate does not hold for the method with mass lumping, as we will see in our numerical tests. This however has little impact, since only super-convergence of the pressure averages is required. We may thus conclude that the proposed second-order mass lumping strategy does not affect the accuracy of the mixed finite element approximation.

2.4. Implementation and numerical tests

We now discuss some aspects regarding the implementation of the proposed second-order mass lumping strategy. We first specify appropriate basis functions that, together with the proposed quadrature rule, leads to a block-diagonal mass matrix. In the second part of the section, we will illustrate the validity of our theoretical results by numerical tests and compare them with the standard mixed finite element approximation.

2.4.1. Choice of basis functions

We first specify the basis functions for RT_1 to be used for mass lumping. It suffices to do this on the reference element K. We can then apply the Piola transformation (A.1) to map the basis functions to arbitrary elements obtained by an affine transformation. In the following, we will use the notation $(a;b) = (a,b)^\top$ to describe column vectors.

Triangular basis functions

Let $\widehat{K} = \mathrm{span}\{(0;0),(1;0),(0;1)\}$ denote the reference triangle and define

$$\begin{aligned} &\widehat{\varphi}_1 = (0;x) && \widehat{\varphi}_2 = (-y;0) && \widehat{\varphi}_3 = (-y;-y) \\ &\widehat{\varphi}_4 = (0;x+y-1) && \widehat{\varphi}_5 = (-x-y+1;0) && \widehat{\varphi}_6 = (x;x). \end{aligned}$$

It is easy to verify that $\mathrm{BDM}_1(\widehat{K}) = P_1(\widehat{K})^2 = \mathrm{span}\{\varphi_i,\, 1 \le i \le 6\}$, which shows that that these functions form a basis for $\mathrm{BDM}_1(K)$. Moreover, we can check that exactly two of these functions are non-zero at each of the three vertices. Further define

$$\widehat{\Phi}_7 = (-y^2+y;xy) \qquad \text{and} \qquad \widehat{\Phi}_8 = (-xy;x^2-x)$$

and note that these functions vanish at all vertices of K. We use these to modify the previous basis functions according to

$$\begin{aligned} &\widehat{\Phi}_1 = \widehat{\varphi}_1 + \widehat{\Phi}_7 + 2\widehat{\Phi}_8 && \widehat{\Phi}_2 = \widehat{\varphi}_2 + 2\widehat{\Phi}_7 + \widehat{\Phi}_8 && \widehat{\Phi}_3 = \widehat{\varphi}_3 + \widehat{\Phi}_7 - \widehat{\Phi}_8 \\ &\widehat{\Phi}_4 = \widehat{\varphi}_4 - \widehat{\Phi}_7 - 2\widehat{\Phi}_8 && \widehat{\Phi}_5 = \widehat{\varphi}_5 - 2\widehat{\Phi}_7 - \widehat{\Phi}_8 && \widehat{\Phi}_6 = \widehat{\varphi}_6 - \widehat{\Phi}_7 + \widehat{\Phi}_8. \end{aligned}$$

One can again directly check that $\mathrm{RT}_1(\widehat{K}) = \mathrm{span}\{\widehat{\Phi}_i,\, 1 \le i \le 8\}$. Moreover, exactly two basis functions are non-zero on each of the quadrature points; see Figure 2.2 for a sketch. This property remains valid for arbitrary elements, which are obtained by affine transformations of the reference triangle. The basis functions are just transformed by Piola mappings.

Tetrahedral basis functions

Let $\widehat{K} = \mathrm{span}\{(0;0;0),(1;0;0),(0;1;0),(0;0;1)\}$ be the reference tetrahedron and define

$$\begin{aligned} &\widehat{\varphi}_1 = (x;0;0) && \widehat{\varphi}_2 = (0;y;0) && \widehat{\varphi}_3 = (0;0;z) \\ &\widehat{\varphi}_4 = (-y;y;0) && \widehat{\varphi}_5 = (-z;0;z) && \widehat{\varphi}_6 = (x+y+z-1;0;0) \\ &\widehat{\varphi}_7 = (0;-z;z) && \widehat{\varphi}_8 = (0;x+y+z-1;0) && \widehat{\varphi}_9 = (x;-x;0) \\ &\widehat{\varphi}_{10} = (0;0;x+y+z-1) && \widehat{\varphi}_{11} = (x;0;-x) && \widehat{\varphi}_{12} = (0;y;-y). \end{aligned}$$

Again, these functions form a basis for $\mathrm{BDM}_1(K)$ and exactly three of these functions are non-zero at each of the four vertices. Further, let

$$\widehat{\Phi}_{13} = (x^2 - x; xy; xz) \qquad \widehat{\Phi}_{14} = (yx; y^2 - y; yz) \qquad \widehat{\Phi}_{15} = (zx; zy; z^2 - z).$$

which again vanish identically on all vertices. We use this to modify the above basis functions according to

$$\begin{aligned}
\widehat{\Phi}_1 &= \widehat{\varphi}_1 + 2\widehat{\Phi}_{13} + \widehat{\Phi}_{14} + \widehat{\Phi}_{15} & \widehat{\Phi}_2 &= \widehat{\varphi}_2 + \widehat{\Phi}_{13} + 2\widehat{\Phi}_{14} + \widehat{\Phi}_{15} \\
\widehat{\Phi}_3 &= \widehat{\varphi}_3 + \widehat{\Phi}_{13} + \widehat{\Phi}_{14} + 2\widehat{\Phi}_{15} & \widehat{\Phi}_4 &= \widehat{\varphi}_4 - \widehat{\Phi}_{13} + \widehat{\Phi}_{14} \\
\widehat{\Phi}_5 &= \widehat{\varphi}_5 - \widehat{\Phi}_{13} + \widehat{\Phi}_{15} & \widehat{\Phi}_6 &= \widehat{\varphi}_6 - 2\widehat{\Phi}_{13} - \widehat{\Phi}_{14} - \widehat{\Phi}_{15} \\
\widehat{\Phi}_7 &= \widehat{\varphi}_7 - \widehat{\Phi}_{14} + \widehat{\Phi}_{15} & \widehat{\Phi}_8 &= \widehat{\varphi}_8 - \widehat{\Phi}_{13} - 2\widehat{\Phi}_{14} - \widehat{\Phi}_{15} \\
\widehat{\Phi}_9 &= \widehat{\varphi}_9 + \widehat{\Phi}_{13} - \widehat{\Phi}_{14} & \widehat{\Phi}_{10} &= \widehat{\varphi}_{10} - \widehat{\Phi}_{13} - \widehat{\Phi}_{14} - 2\widehat{\Phi}_{15} \\
\widehat{\Phi}_{11} &= \widehat{\varphi}_{11} + \widehat{\Phi}_{13} - \widehat{\Phi}_{15} & \widehat{\Phi}_{12} &= \widehat{\varphi}_{12} + \widehat{\Phi}_{14} - \widehat{\Phi}_{15}.
\end{aligned}$$

By the same argument as for the triangular construction, the local mass matrix can be seen to consist of five 3×3 blocks, one associated with each vertex and one associated with the element midpoint; see Figure 2.2. This property again transfers to general tetrahedrons obtained by affine transformations.

2.4.2. Structure of the mass matrix and elimination of the flux

In the previous two sections, we proposed specific sets of basis functions that led to local block-diagonal mass matrices. When assembling the global mass matrix, the blocks that are associated with the same vertex couple together, generating slightly larger blocks. The size of the resulting global blocks corresponds to the degree of the vertex, i.e., the number of adjacent edges plus one. The blocks associated with the element midpoints continue to remain $d \times d$ blocks also in the global matrix since no additional coupling takes place. In Figure 2.3, we depict a typical situation.

Let us emphasize that the size of the blocks does *not* depend on the mesh size but only on the mesh topology, which is controlled by the shape regularity assumption (A2). The lumped mass matrix $\mathsf{M_h}$ is obviously block diagonal. From Lemma 2.3.2, we further deduce that the lumped mass matrix is symmetric and positive definite, and hence also invertible. The inverse $\mathsf{M_h^{-1}}$ then is again block diagonal.

Elimination of the velocity

After choosing the aforementioned set of basis functions for V_h on each element and a standard basis for Q_h, we obtain an algebraic system of the form

$$\begin{pmatrix} \mathsf{M_h} & -\mathsf{B}^\top \\ \mathsf{B} & 0 \end{pmatrix} \begin{pmatrix} \mathsf{u} \\ \mathsf{p} \end{pmatrix} = \begin{pmatrix} \mathsf{g} \\ \mathsf{f} \end{pmatrix}. \tag{2.39}$$

The Schur complement system for the pressure attains the form

$$\mathsf{B}\mathsf{M_h^{-1}}\mathsf{B}^\top \mathsf{p} = \mathsf{f} - \mathsf{B}\mathsf{M_h^{-1}}\mathsf{g}. \tag{2.40}$$

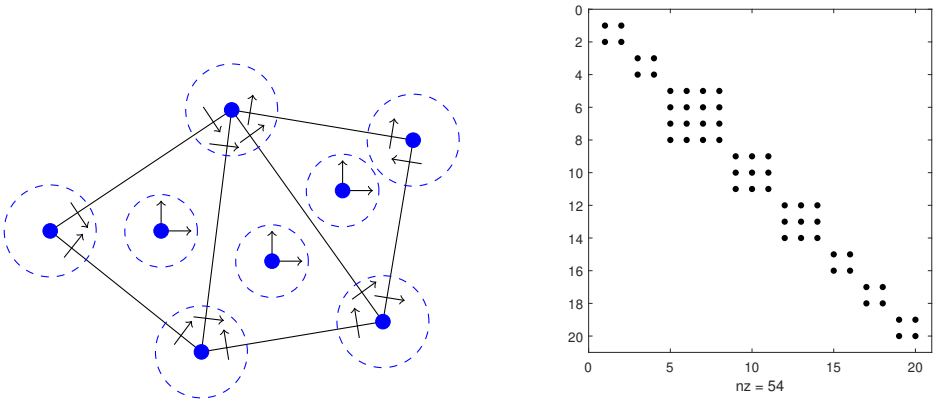

Figure 2.3.: Example of a 2D mesh (left). The blue dots represent the quadrature points, the arrows represent degrees of freedom. The dashed blue circle shows the coupling degrees of freedom. On the right, we see the block-diagonal structure of the resulting lumped mass matrix.

Since B is sparse and $\mathsf{M}_\mathsf{h}^{-1}$ is block diagonal, the system matrix $\mathsf{K}_\mathsf{h} = \mathsf{B}\mathsf{M}_\mathsf{h}^{-1}\mathsf{B}^\top$ is sparse as well. Moreover, since the matrix B has full rank and M_h is symmetric and positive, K_h is a symmetric and positive definite approximation for $-\mathrm{div}(\mathcal{K}\nabla\cdot)$ and (2.40) can be solved efficiently by direct or iterative solvers.

2.4.3. Numerical validation

For an illustration of our theoretical results, we now report on some numerical tests, in which we also compare with the standard mixed finite element method without mass lumping. Our test problem is taken from [**70**], where the authors considered the extension of their method in [**112**] to hexahedral meshes. Let $\Omega = (0,1)^3$ be the computational domain and let $\mathcal{K}$ be a full tensor with variable entries

$$\mathcal{K}(x,y,z) = \begin{pmatrix} x^2+(y+2)^2 & 0 & \cos(xy) \\ 0 & z^2+2 & \sin(yz) \\ \cos(xy) & \sin(yz) & \sin(y+3)^2 \end{pmatrix}.$$

We assume that the exact solution is given by

$$\begin{aligned} p(x,y,z) &= x^4y^3 + x^2 + yz^2 + \cos(xy) + \sin(z), \\ u(x,y,z) &= -\mathcal{K}(x,y,z)\nabla p(x,y,z), \end{aligned}$$

which is used to compute the right-hand side f and the boundary values for p. Since the solution (u,p) as well as the parameter $\mathcal{K}$ are smooth functions and the domain Ω is convex, we expect our numerical experiments to produce the rates we predicted in our numerical analysis.

Setup of convergence tests

For our computations, we consider a family of tetrahedral quasi-uniform meshes $\{\mathcal{T}_h\}_h$, which are generated by MATLAB's `initmesh` routine available in the PDE Toolbox v3.4.

We note that the meshes $\mathcal{T}_h$ for different mesh sizes are non-nested.

Convergence results for the exact mixed finite element method

In Table 2.1, we display the errors and estimated convergence rates obtained for the standard mixed finite element method with exact integration. The corresponding solutions are denoted by $\widehat{u}_h$ and $\widehat{p}_h$. As expected from the well-known convergence results which were summarized in Lemma 2.2.2, we obtain order optimal convergence rates for the velocity error and super-convergence for the pressure averages. In addition, we also observe super-convergence for $\|\pi_h^1 p - p_h\|_{L^2(\Omega)}$, which can also be proven rigorously; we refer to [**19, Chapter 7**] for some results this direction.

h	$\dim(V_h)$	$\dim(Q_h)$	$\|\Pi_h u - u_h\|$	eoc	$\|\pi_h^1 p - \widehat{p}_h\|$	eoc	$\|\pi_h^0(p - \widehat{p}_h)\|$	eoc
2^{-1}	1152	456	0.145059	—	0.007015	—	0.005799	—
2^{-2}	4779	1964	0.048244	1.59	0.001410	2.31	0.001114	2.37
2^{-3}	37755	16180	0.013651	1.82	0.000205	2.78	0.000157	2.82
2^{-4}	297522	129920	0.003727	1.87	0.000028	2.85	0.000023	2.74

Table 2.1.: System dimension, discrete errors, and estimated orders of convergence (eoc) for the standard second-order mixed finite element method on tetrahedra.

Convergence results for the second order MFMFE method

In Table 2.2 we report on the corresponding results obtained with the second-order MFMFE method studied in the previous section. The numerical solution here is denoted by u_h and p_h. All convergence rates predicted by Theorem 2.3.4 are again observed

h	$\dim(V_h)$	$\dim(Q_h)$	$\|\Pi_h u - u_h\|$	eoc	$\|\pi_h^1 p - p_h\|$	eoc	$\|\pi_h^0(p - p_h)\|$	eoc
2^{-1}	1152	456	0.252144	—	0.065159	—	0.007702	—
2^{-2}	4779	1964	0.073773	1.77	0.020516	1.67	0.001290	2.58
2^{-3}	37755	16180	0.018418	2.00	0.005434	1.92	0.000183	2.81
2^{-4}	297522	129920	0.004550	2.02	0.001350	2.01	0.000024	2.90
2^{-5}	2430747	1070996	0.001052	2.11	0.000317	2.09	0.000003	3.09

Table 2.2.: System dimension, discrete errors, and estimated orders of convergence (eoc) of the second order MFMFE method on tetrahedra for Problem 2.3.1.

also in our tests. In particular, we observe second-order convergence for both velocity and pressure, and third-order convergence for the pressure averages. In comparison to the standard method without mass lumping, see Table 2.1, we however do not observe super-convergence for $\|\pi_h^1 p - p_h\|_{L^2(\Omega)}$ here. As mentioned in our remarks of the previous section, this can be resolved by local post-processing, for which we only require super-convergence of $\pi_h^0(p - p_h)$; see Lemma 2.3.9. We only mention that for the post-processed pressure $\tilde{p}_h \in P_2(\mathcal{T}_h)$, we again observe third-order convergence in the L^2-norm.

2.5. Discussion

In this chapter, we proposed a mass lumping strategy for the RT_1 space leading to a MFMFE method. We showed that the scheme is second-order accurate and superconvergent in the pressure averages, thus providing the full accuracy of a standard mixed finite element approximation in RT_1-P_1. Although our analysis was tailored to the second-order approximation on simplices, our arguments can be generalized to arbitrary orders and other element types. In [**48**], we provided a general error analysis that also covers quadrilateral, hexahedral and prismatic elements, and higher order approximations. The main ingredients for the construction of the mass lumping schemes are

- an appropriate numerical quadrature with a certain number of quadrature points on the boundary;
- a suitable ansatz space for V_h that contains at least either RT_k or BDM_k, in order to guarantee the necessary approximation properties;
- appropriate basis functions such that exactly d degrees of freedom can be associated with each quadrature point; in addition, these functions need to have linearly independent divergences and have to be integrated appropriately by the quadrature rule.

An example of a third-order MFMFE method is given in [**48**]. The main difficulty for the extension to higher orders on unstructured grids lies in finding appropriate quadrature formulas Q_K that have the appropriate number of quadrature points on the boundary of K and are, at the same time, sufficiently accurate. What we observed is that increasingly more interior bubble functions are required when going to higher orders, in order to achieve the block-diagonal structure of the mass matrix. This also results in larger spaces $Q_h = \operatorname{div} V_h$ for the pressure and counteracts the efficiency of the methods.

In [**4**], multipoint-flux mixed finite elements of arbitrary order have been constructed for quadrilaterals and hexahedra based on tensor product Gauss-Lobatto quadrature. The spaces there are extensions of the standard RT_k-P_k elements. Using our theory from [**48**], we are able to provide mixed finite element pairs that have dramatically fewer degrees of freedom, but we are unable to do this systematically for arbitrary orders.

Let us finally also compare the MFMFE method briefly to some of its direct competitors. One of them is *hybridization* [**7, 19**], in which the normal continuity of the velocity space V_h is broken up and re-established via Lagrange multipliers. This allows to eliminate the original velocity and pressure variables completely and to obtain a globally symmetric and positive definite system for the Lagrange multipliers. In comparison to the reduced form of the corresponding MFMFE method, hybridization usually results in a slightly smaller global algebraic system with smaller stencils.

Another competitor are the *discontinuous Galerkin* (dG) methods. Many variants have been proposed over the years, from the interior penalty method by Arnold [**8**], to specific formulations for single-phase flow; see [**100, 28, 9, 14, 15**]. Typically, the dG methods result in algebraic systems for the pressure with similar stencil sizes as the MFMFE method. While the error analysis of the MFMFE seems somewhat nicer and closer to

the continuous analysis, the extension of the methods to higher orders is certainly more simple in the dG context. In the end, all these methods lead to global symmetric positive definite algebraic systems with very similar properties. The computational complexity of their solution is rather comparable.

Another area of application for the mass lumping techniques developed for the MFMFE methods arises in the numerical solution of acoustic or electromagnetic wave propagation problems, where mass lumping allows obtaining efficient explicit time-stepping schemes; see Chapter 4 and 5 for details. Hybridization does not offer this advantage, but always requires the solution of a global linear system in every time step. For moderate approximation orders, corresponding dG methods, on the other hand, have substantially more degrees of freedom. We will return to these topics later in this thesis.

3. A multipoint flux mixed finite element method for poroelasticity

We now extend our considerations to simulate flow through a porous medium accompanied by elastic deformation of the porous matrix. This can be observed, for instance, in geology, more specifically in soil mechanics. For a comprehensive list of applications, we refer to the introduction of [**93**] and to [**30, 106**]. In this chapter, we make the simplifying assumption that the elastic deformation occurs instantaneously, which is known as the *quasi-static* case. The governing equations, called the *two-field formulation*, were developed by Biot [**17**] and are given by

$$-\operatorname{div}(2\mu\,\epsilon(u) + \lambda \operatorname{div}(u) I) + \alpha \nabla p = f \tag{3.1}$$

$$c_s \partial_t p + \operatorname{div}(\alpha \partial_t u) - \operatorname{div}(\mathcal{K} \nabla p) = g. \tag{3.2}$$

Here u describes the displacement of the porous medium while p denotes the pressure of the fluid inside the porous medium. The model parameters are the Lamé coefficients μ, λ, the specific storage coefficient c_s of the fluid, the hydraulic conductivity tensor $\mathcal{K}$, as well as the Biot-Willis coupling constant α. Moreover, f and g represent volume forces and the fluid sources, respectively. The first equation (3.1) describes the equilibrium of forces in the solid while the second equation (3.2) models the mass balance and flow of the fluid. The coupling works as follows: pressure gradients of the fluid act as forces onto the porous matrix, whereas the deformation of the structure acts as a source or sink in the flow equation.

Well-posedness of the system (3.1)–(3.2) together with appropriate initial and boundary conditions has been established by Zenisek [**107, 108**], who constructed a sequence of finite element approximations and showed their convergence to a weak solution of the problem. Later, Showalter [**101**] used semi-group theory to prove the existence of solutions. We also refer to [**52**] for a different approach based on variational arguments.

The discretization of the two-field formulation by H^1-conforming finite elements for both u and p has been discussed in [**88, 89**] for the special case $c_s = 0$ with Dirichlet boundary conditions and initial condition $\operatorname{div} u(0) = 0$. In these works, the authors recognized the importance of using inf-sup stable finite element pairs for the approximation of u and p, in order to avoid consistency errors in the initial layer. A thorough error analysis, including decay estimates, was conducted for P_k-P_{k-1} approximations, i.e., Taylor-Hood elements. The structure-preserving time discretization based on the Petrov-Galerkin approximation has been studied in [**52**].

Following the arguments already used for the Darcy problem, it seems natural to replace

the flow equation (3.2) with the system

$$c_s \partial_t p + \operatorname{div}(\alpha \partial_t u) + \operatorname{div}(w) = g, \tag{3.3}$$

$$\mathcal{K}^{-1} w + \nabla p = 0. \tag{3.4}$$

The new variable w here denotes the *seepage velocity*. The problem consisting of (3.1) and (3.3)–(3.4) is known as the *three-field formulation* of poroelasticity.

Let us review some results about corresponding discretization schemes. In [**93, 94, 95**], the authors study continuous (cG) and discontinuous (dG) Galerkin approximations for the solid displacement coupled to a mixed finite element discretization for the pressure and seepage velocity. Non-conforming discretizations of the three-field formulation have also been proposed in [**69, 67**] and some of the references therein. In recent works [**67, 68, 75**], the authors consider specific dG approximations for the elastic displacement with $H(\operatorname{div})$ elements coupled to mixed finite element approximations for the flow problem. As a consequence, the divergence constraint (3.3) can be satisfied exactly on the discrete level. Moreover, the schemes turn out to be stable in the incompressible limit $\lambda \to \infty$.

In principle, dG approximations for (3.2) can be used to obtain conservative approximations for the two-field formulations due to the intrinsic underlying inf-sup condition of the method; these however lead to a poor approximation of the seepage velocity; see [**28**]. In [**75**], the time discretization for the three field formulation by the backward Euler method has also been investigated, and numerical tests with the Crank-Nicolson scheme have been performed, showing second-order convergence in practice.

Contributions

In this chapter, we consider inexact Galerkin approximations of the three-field formulation based on an H^1-conforming approximation of the elasticity equation (3.1) and a mixed finite element approximation for (3.3)–(3.4). Mass lumping allows eliminating the seepage velocity on the algebraic level, leading to a reduced algebraic system that can be interpreted as a non-conforming but locally mass conservative approximation of the two-field formulation. A first-order accurate method in this spirit has already been proposed in [**111**]. In this chapter, we consider a second-order method that makes use of the finite element spaces $\mathrm{P}_2^+ - \mathrm{P}_1 - \mathrm{RT}_1$ and of numerical quadrature as in the previous chapter; see Figure 3.1 later for a depiction of degrees of freedom and location of quadrature points. We show that the method is second-order convergent in all three variables, and even exhibits super-convergence in the pore pressure. Moreover, we study the time-discretization by a variant of the trapezoidal rule and also rigorously prove second-order convergence in time.

Outline

Let us give a brief overview of the contents of this chapter.

Section 3.1. *The model problem:* We introduce the mathematical problem in a rigorous way and prove the existence and uniqueness of solutions. We also discuss compatibility conditions for initial values and right-hand sides and introduce a corresponding variational characterization of solutions.

Section 3.2. *A mass lumped mixed finite element method:* We then introduce the finite element spaces for the three field formulation of our method with mass lumping and state existence and uniqueness results for the semi-discrete problem.

Section 3.3. *Error estimates for the semi-discretization:* This section provides a complete error analysis for the proposed semi-discrete method with mass lumping.

Section 3.4. *Time discretization:* To complete the discretization process, we study time discretization by a variant of the trapezoidal rule and prove that the resulting method is also second-order convergent in time.

Section 3.5. *Numerical experiments and discussion*: To support our theoretical findings, we provide numerical results for a simple test problem, similar to the one used in [**75**]. We then conclude by discussing possible extensions of our approach, including non-conforming approximations for the elasticity problem considered in [**67, 75**].

3.1. The model problem

For ease of presentation, we assume that the medium under consideration is homogeneous and isotropic, i.e., the model parameters are constants. By rescaling, we may then choose $\alpha = 1$ and $\mathcal{K} = \kappa I$ in (3.1),(3.3)–(3.4) and obtain

$$-\mathrm{div}(2\mu\epsilon(u) + \lambda \mathrm{div}(u) I) + \nabla p = f \qquad \text{on } \Omega,\ t > 0, \tag{3.5}$$

$$\mathrm{div}\, \partial_t u + c_s \partial_t p + \mathrm{div}\, w = g \qquad \text{on } \Omega,\ t > 0, \tag{3.6}$$

$$\kappa^{-1} w + \nabla p = 0 \qquad \text{on } \Omega,\ t > 0. \tag{3.7}$$

For ease of presentation, we complement (3.5)–(3.7) by homogeneous boundary conditions

$$u = 0 \qquad \text{and} \qquad n \cdot w = 0 \qquad \text{on } \partial\Omega,\ t > 0. \tag{3.8}$$

With minor modifications, our analysis can be generalized to non-constant coefficients, and a general symmetric positive tensor $\mathcal{K}$. We further pose initial conditions

$$u(0) = u_0, \qquad p(0) = p_0, \qquad \text{and} \qquad w(0) = w_0 \qquad \text{on } \Omega. \tag{3.9}$$

Note that (3.5)–(3.7) is a partial-differential algebraic equation and therefore, the initial values cannot be chosen arbitrarily.

Remark 3.1.1 (Compatibility conditions for initial values)**.**
Any sufficiently smooth solution of (3.5)–(3.9) also satisfies

$$-\mathrm{div}(2\epsilon(u_0) + \lambda \mathrm{div}(u_0 I) + \nabla p_0 = f(0), \tag{3.10}$$

$$\kappa^{-1} w_0 + \nabla p_0 = 0. \tag{3.11}$$

These conditions for the initial values are necessary to ensure the existence of solutions with sufficient regularity. Initial values satisfying (3.10)–(3.11) are called *consistent* in the language of differential algebraic equations. Let us further note that for given p_0, the values u_0 and w_0 are uniquely determined by (3.10)–(3.11). This means that our problem has only one truly independent initial value.

Remark 3.1.2 (Compatibility conditions for right hand side)**.**
Without loss of generality, one can assume that $p(t)$ and $g(t)$ have zero mean. To see this, we integrate equation (3.6) over Ω, apply the divergence theorem, and subsequently use the boundary conditions from (3.8), which immediately leads to

$$c_s \int_\Omega \partial_t p(x,t)\,dx = \int_\Omega g(x,t)\,dx.$$

If $c_s = 0$, then necessarily

$$\int_\Omega g(x,t)\,dx = 0 \tag{3.12}$$

for all $t > 0$. In the case when $c_s > 0$, then

$$\int_\Omega p(x,t)\,dx = \int_\Omega p(x,0)\,dx + \frac{1}{c_s}\int_0^t \int_\Omega g(x,s)\,dx\,ds.$$

Hence the average of the pressure is completely determined by the average of the initial value and the right-hand side. Since this value is not relevant for (3.5) and (3.7), we can simply modify the pressure by a time varying constant, such that $\int_\Omega p(x,t)\,dx = \int_\Omega g(x,t)\,dx = 0$ for all $t > 0$, and we will use these assumptions for all further considerations.

Notation

By $L^p(\Omega)$ and $W^{k,p}(\Omega)$ we denote the usual Lesbegue and Sobolev spaces equipped with their standard norms, and we set $H^k(\Omega) = W^{k,2}(\Omega)$. For our analysis, we will make extensive use of the function spaces

$$\begin{aligned} H_0^1(\Omega) &= \left\{u \in H^1(\Omega),\ u = 0 \ \text{ on } \ \partial\Omega\right\}, \\ L_0^2(\Omega) &= \left\{q \in L^2(\Omega),\ \int_\Omega q = 0\right\}, \quad \text{and} \\ H_0(\mathrm{div},\Omega) &= \left\{w \in H(\mathrm{div},\Omega),\ n\cdot w = 0 \ \text{ on } \ \partial\Omega\right\}, \end{aligned}$$

in dimension $d = 2, 3$. The norm of the latter is defined by

$$\|u\|_{H(\mathrm{div},\Omega)} = (\|u\|^2_{L^2(\Omega)} + \|\mathrm{div}\, u\|^2_{L^2(\Omega)})^{1/2}.$$

We further denote by $H^{-1}(\Omega) = H_0^1(\Omega)'$ the dual space of $H_0^1(\Omega)$ and by $H_0(\mathrm{div},\Omega)'$ the dual of $H_0(\mathrm{div},\Omega)$. The space $L_0^2(\Omega)$ and its dual are identified with each other. For a Hilbert space H, the norm of the dual space is given by

$$\|u\|_{H'} = \sup_{v\in H} \frac{\langle u, v\rangle_{H'\times H}}{\|v\|_H},$$

where $\langle u, v\rangle_{H'\times H}$ denotes the duality product. We write (u, v) for the scalar product of $L^2(\Omega)$ and use the same symbol for the duality product if there is no danger of confusion. Finally recall the Bochner spaces $C^k([0,T];X)$ and $W^{k,p}(0,T;X)$ consisting of functions of time with values in a Hilbert space X; let us refer to [**56**] for details on the notation and further results.

Weak formulation

For establishing the existence and uniqueness of solutions and as a starting point for the discretization of the problem to be discussed in the following sections, we will use the following weak characterizations of solutions.

Lemma 3.1.3 (Variational characterization)**.**
Let (u, p, w) denote a sufficiently smooth solution of (3.5)–(3.9). Then

$$\begin{aligned} a(u(t),v) \;\; -b(v,p(t)) \qquad\qquad &= (f(t),v) \qquad \forall\, v \in H_0^1(\Omega), && (3.13)\\ b(\partial_t u(t),q)+c(\partial_t p(t),q)+d(w(t),q) &= (g(t),q) \qquad \forall\, q \in L_0^2(\Omega), && (3.14)\\ -d(z,p(t)) \;\; +m(w(t),z) &= 0 \qquad \forall\, z \in H_0(\mathrm{div},\Omega), && (3.15) \end{aligned}$$

for almost all $t > 0$, with bilinear forms defined by

$$\begin{aligned} a(u,v) &= (\epsilon(u),\epsilon(v)) + (\lambda \,\mathrm{div}(u),\mathrm{div}(v)), \qquad b(v,p) = (\mathrm{div}(v),p),\\ c(p,q) &= (c_s p,q), \qquad d(w,q) = (\mathrm{div}(w),q), \quad m(w,z) = (\kappa^{-1}w,z). \end{aligned} \tag{3.16}$$

Moreover, the initial values satisfy

$$\begin{aligned} a(u(0),v) \;\; &= (f(0),v) + b(v,p(0)) \qquad \forall\, v \in H_0^1(\Omega), && (3.17)\\ m(w(0),z) &= d(z,p(0)) \qquad \forall\, z \in H_0(\mathrm{div},\Omega). && (3.18) \end{aligned}$$

The validity of the assertions is verified in the usual manner, by multiplying (3.5)–(3.7) with appropriate test functions, integrating over the domain Ω, using integration-by-parts here and there, and using the boundary conditions to get rid of the boundary integrals.

3.1.1. Auxiliary results

Let us start by collecting the basic assumptions that we use in our analysis.

(A1) $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a bounded Lipschitz domain and $\lambda \geq 1$, $c_s \geq 0$, $\kappa > 0$ constant.

As a next preparatory step, we gather the most important properties of the bilinear forms appearing in the weak form of the equations.

Lemma 3.1.4 (Properties of the bilinear forms)**.**
Let (A1) hold. Then the mappings

$$\begin{aligned} &a : H_0^1(\Omega)^d \times H_0^1(\Omega)^d \to \mathbb{R}, && b : H_0^1(\Omega)^d \times L_0^2(\Omega) \to \mathbb{R},\\ &c : L_0^2(\Omega) \times L_0^2(\Omega) \to \mathbb{R}, && d : H_0(\mathrm{div},\Omega) \times L_0^2(\Omega) \to \mathbb{R},\\ &m : H_0(\mathrm{div},\Omega) \times H_0(\mathrm{div},\Omega) \to \mathbb{R} \end{aligned}$$

are continuous bilinear forms and $a(\cdot,\cdot)$, $c(\cdot,\cdot)$ and $m(\cdot,\cdot)$ are symmetric. Moreover

(i) a is H_0^1-elliptic, i.e. there exists $\alpha > 0$ with $a(u,u) \geq \alpha \|u\|_{H^1(\Omega)}$ $\forall u \in H_0^1(\Omega)^d$.

(ii) c is non-negative, i.e. $c(p,p) \geq 0$ for all $p \in L_0^2(\Omega)$.

(iii) m *is* L^2*-elliptic, i.e. there exists* $\delta > 0$ *with* $m(w,w) \geq \delta \|w\|_{L^2(\Omega)}$ $\forall w \in L^2(\Omega)^d$.

(iv) There exists a $\beta > 0$ *such that for* $V = H_0^1(\Omega)^d$ *and* $V = H_0(\mathrm{div}, \Omega)$ *one has*

$$\inf_{q \in L_0^2(\Omega)} \sup_{v \in V} \frac{b(v,q)}{\|v\|_{H^1(\Omega)} \|q\|_{L^2(\Omega)}} \geq \beta.$$

Proof. Continuity and symmetry of $a(\cdot,\cdot)$, $c(\cdot,\cdot)$ and $m(\cdot,\cdot)$ follows directly from their definition and the Cauchy-Schwarz inequality. To verify (i), we compute

$$a(u,u) = (\epsilon(u), \epsilon(u)) + (\lambda \,\mathrm{div}(u), \mathrm{div}(u)) \geq \|\epsilon(u)\|^2_{L^2(\Omega)} \geq \alpha \|u\|_{H^1(\Omega)},$$

where we used Korn's first inequality in the last step; see [**91**]. Properties (ii) and (iii) follow from $c_s \geq 0$ and $\kappa > 0$ and the Friedrichs' inequality. For $V = H_0^1(\Omega)^d$, property (iv) is the celebrated inf-sup condition for the Stokes problem; see [**62**]. Since $H_0(\mathrm{div}, \Omega) \subset H_0^1(\Omega)$, the second case follows from the first one. □

3.1.2. Well-posedness

Next, we establish the existence and uniqueness of weak solutions for which we use the particular algebraic structure of the equations and Galerkin approximations.

Theorem 3.1.5 (Well-posedness)**.**
Let (A1) hold. Then for any $p_0 \in H^1(\Omega) \cap L_0^2(\Omega)$ *and* $f(0) \in H^{-1}(\Omega)$ *there exist unique* $u_0 \in H_0^1(\Omega)^d$ *and* $w_0 \in L^2(\Omega)^d$ *satisfying the compatibility conditions* (3.17)–(3.18)*. Moreover, for any* $f \in H^1(0,T,H^{-1}(\Omega)^d)$, $g \in L^2(0,T,L_0^2(\Omega))$ *there exists a unique weak solution*

$$\begin{aligned} &u \in H^1(0,T,H_0^1(\Omega)^d), \\ &p \in H^1(0,T,L_0^2(\Omega)), \quad \text{and} \\ &w \in \big(L^2(0,T,H_0(\mathrm{div},\Omega)) \cap H^1(0,T,H_0(\mathrm{div},\Omega)')\big) \hookrightarrow C(0,T,L^2(\Omega)^d) \end{aligned}$$

of (3.5)–(3.9) *which is characterized by the variational identities* (3.13)–(3.15)*.*

Proof. For any $p_0 \in H^1(\Omega) \cap L_0^2(\Omega)$ and $f(0) \in H^{-1}(\Omega)$, the values $u_0 \in H_0^1(\Omega)^d$ and $w_0 \in H_0(\mathrm{div}, \Omega)$ are uniquely determined by solving (3.17)–(3.18). The solvability follows from the Lax-Milgram Lemma [**11**] and Lemma 3.1.4, which guarantees continuity and ellipticity for $a(\cdot,\cdot)$ and $m(\cdot,\cdot)$. The rest of the proof follows with similar arguments as in [**52, Theorem 2**], which we modify to fit our problem: Using Lemma 3.1.4, the bilinear forms introduced in (3.16) induce bounded linear operators

$$\begin{aligned} &A : H_0^1(\Omega) \to H^{-1}(\Omega), && B : H_0^1(\Omega) \to L_0^2(\Omega)', && B^* : L_0^2(\Omega) \to H^{-1}(\Omega), \\ &D : H_0(\mathrm{div},\Omega) \to L_0^2(\Omega)', && D^* : L_0^2(\Omega) \to H_0(\mathrm{div},\Omega)', \\ &M : H_0(\mathrm{div},\Omega) \to H_0(\mathrm{div},\Omega)', && C : L_0^2(\Omega) \to L_0^2(\Omega)', \end{aligned}$$

which are formally defined by

$$\begin{aligned}
&\langle Au, v\rangle = a(u, v), \qquad &&\langle Bv, q\rangle = b(v, q) = \langle B^*q, v\rangle, \\
&\langle Mw, z\rangle = m(w, z), \qquad &&\langle Dw, q\rangle = d(w, q) = \langle D^*q, w\rangle, \qquad \langle Cp, q\rangle = c(p, q).
\end{aligned}$$

This allows us to rewrite (3.13)–(3.15) in the equivalent operator form

$$\begin{aligned}
Au(t) \quad -B^*p(t) \qquad\qquad &= f(t), &(3.19)\\
B\partial_t u(t) + C\partial_t p(t) + Dw(t) &= g(t), &(3.20)\\
-D^*p(t) \; + Mw(t) &= 0. &(3.21)
\end{aligned}$$

From the results of Lemma 3.1.4, we can infer that the operators A and M are bounded and elliptic, and hence continuously invertible. The two divergence operators B and D are bounded and surjective.

We can then continue by differentiating and rewriting equation (3.19) in explicit form

$$\partial_t u(t) = A^{-1}(B^*\partial_t p(t) + \partial_t f(t)). \tag{3.22}$$

Due to the compatibility condition (3.17), this equation is equivalent to (3.19). Plugging the expression for $\partial_t u$ into (3.20) leads to the simplified system

$$\begin{aligned}
(BA^{-1}B^* + C)\partial_t p(t) + Dw(t) &= g(t) - BA^{-1}\partial_t f(t) &(3.23)\\
-D^*p(t) \quad + Mw(t) &= 0 &(3.24)
\end{aligned}$$

which only involves the pressure and seepage velocity. By differentiating (3.24), we obtain

$$-D^*\partial_t p(t) + M\partial_t w(t) = 0 \tag{3.25}$$

which is again equivalent to (3.21) due to the compatibility condition (3.18). From the properties of the operators A, B, and C, we further deduce that

$$\left(BA^{-1}B^* + C\right) : L_0^2(\Omega) \to L_0^2(\Omega) \tag{3.26}$$

is an isomorphism, which can be seen as follows: Continuity of $a(\cdot,\cdot)$ and $c(\cdot,\cdot)$ and the definition of the corresponding operators imply that

$$\begin{aligned}
\langle BA^{-1}B^* + Cp, q\rangle &= a(A^{-1}B^*p, A^{-1}B^*q) + c(p, q)\\
&\le c_a\|A^{-1}B^*p\|_{H^1(\Omega)}\|A^{-1}B^*q\|_{H^1(\Omega)} + c_s\|p\|_{L^2(\Omega)}\|q\|_{L^2(\Omega)}.
\end{aligned}$$

From Lemma 3.1.4 and the corresponding estimates for the associated operators, we then obtain

$$\begin{aligned}
\|A^{-1}B^*p\|^2_{H^1(\Omega)} &\le \frac{1}{\alpha}\|B^*p\|^2_{H^{-1}(\Omega)} \le \frac{c_b^2}{\alpha}\|p\|^2_{L^2(\Omega)},\\
\|A^{-1}B^*p\|^2_{H^1(\Omega)} &\ge \frac{\alpha}{c_a^2}\|B^*p\|^2_{H^{-1}(\Omega)} \ge \frac{\alpha\beta^2}{c_a^2}\|p\|_{L^2(\Omega)},
\end{aligned}$$

which proves the desired properties. This allows to eliminate $\partial_t p(t)$ in (3.25) by employing (3.23), which leads to an abstract evolution equation

$$M\partial_t w(t) + Lw(t) = h(t). \tag{3.27}$$

Here $L \coloneqq D^*(BA^{-1}B^* + C)^{-1}D$ and $h(t) \coloneqq (BA^{-1}B^* + C)^{-1}(g(t) - BA^{-1}\partial_t f(t))$. By similar arguments as above, we deduce that $L : H_0(\mathrm{div}, \Omega) \to (H_0(\mathrm{div}, \Omega))'$ is self-adjoint and satisfies a Garding-type inequality

$$\begin{aligned}(Lw, w)_{L^2(\Omega)} + \|w\|^2_{L^2(\Omega)} &= (D^*(BA^{-1}B^* + C)^{-1}Dw, w)_{L^2(\Omega)} + \|w\|^2_{L^2(\Omega)} \\ &= ((BA^{-1}B^* + C)^{-1}Dw, Dw)_{L^2(\Omega)} + \|w\|^2_{L^2(\Omega)} \\ &\geq \alpha_L \|Dw\|^2_{L^2(\Omega)} + \|w\|^2_{L^2(\Omega)} \geq c\|w\|^2_{H(\mathrm{div},\Omega)}.\end{aligned}$$

This inequality holds for all $w \in H_0(\mathrm{div}, \Omega)$ with a uniform constant $c > 0$. Moreover, we also have $h \in L^2(0, T, L^2(\Omega))$ with its norm bounded by the norm of the data f and g. For any initial value $w(0) = w_0 \in L^2(\Omega)^d$, existence of a unique solution

$$w \in L^2(0, T, H_0(\mathrm{div}, \Omega)) \cap H^1(0, T, H_0(\mathrm{div}, \Omega)')$$

can be established via Galerkin approximation, just as in [**56, Chapter 7.1.2**]. Moreover, the solution can be bounded in terms of the data by

$$\begin{aligned}\|w\|_{L^2(0,T,H_0(\mathrm{div},\Omega))} + \|\partial_t w\|_{L^2(0,T,H_0(\mathrm{div},\Omega)'} &\leq C(\|w_0\|_{L^2(\Omega)} + \|h\|_{L^2(0,T,L^2(\Omega))}) \\ &\leq C'(\|w_0\|_{L^2(\Omega)} + \|g\|_{L^2(0,T,L^2(\Omega))} + \|\partial_t f\|_{L^2(0,T,H^{-1}(\Omega))}).\end{aligned}$$

Plugging the function w back into (3.23) yields

$$\partial_t p = (BA^{-1}B^* + C)^{-1}(-Dw + h) \in H^1(0, T, L^2_0(\Omega)).$$

The pressure component of the solution is then found by integration in time and using the initial condition $p(0) = p_0$. From equation (3.19), we finally get $Au = B^*p + f \in H^1(0, T, H^{-1}(\Omega))$, and thus $u \in H^1(0, T, H^1_0(\Omega)^d$. This shows the existence of a solution (u, w, p) together with the corresponding regularity assumptions; furthermore, w is unique. The uniqueness of u and p can be deduced from the linearity of the problem. □

3.2. A mass lumped mixed finite element method

We now turn to the space discretization of (3.13)–(3.15), for which we use a conforming discretization of the elasticity problem combined with a mixed finite element approximation with mass lumping for the flow problem. We consider a geometrically conforming partition $\mathcal{T}_h = \{K\}$ of the domain Ω into elements K and denote by h_K and ρ_K the diameter and inner circle radius of the element K, and call $h = \max_K h_K$ the mesh size. Furthermore, we assume that

(A2) $\mathcal{T}_h$ is γ-shape regular simplicial mesh, i.e. there exists a uniform constant $\gamma > 0$ such that $\gamma h_K \leq \rho_K \leq h_K$ for all $K \in \mathcal{T}_h$.

Like in the previous chapter, we denote by $P_k(\mathcal{T}_h)$ and $H^k(\mathcal{T}_h)$ the spaces of piecewise smooth functions that are either polynomials or H^k-regular on every element $K \in \mathcal{T}_h$.

3.2.1. The finite element triple $P_2^+ - P_1 - RT_1$

For the approximation of the individual components of the elastic deformation on every element K, we utilize the extended P^2-finite element space

$$P_2(K)^+ = \begin{cases} P_2(K) \oplus b_K \cdot P_0(K) \coloneqq P_2^+(K), & d = 2. \\ P_2(K) \oplus b_K \cdot P_0(K) \oplus \operatorname{span}\{b_{F_i},\, i = 1, \dots, 4\} \coloneqq P_2^+(K), & d = 3. \end{cases}$$

Here $b_K \in P_{d+1}(K) \cap H_0^1(K)$ denotes the element bubble function and b_{F_i} denotes the face bubble in 3D associated to the i-th face; see [**19, Example 8.7.2.**] for details. For the local approximation of the seepage velocity and the pressure, we further use the spaces

$$\mathrm{RT}_1(K) \coloneqq P_1(K)^d \oplus \{x \cdot P_1^h(K)\} \qquad \text{and} \qquad P_1(K).$$

In Figure 3.1 we depict the location of the degrees of freedom for these spaces on the reference element.

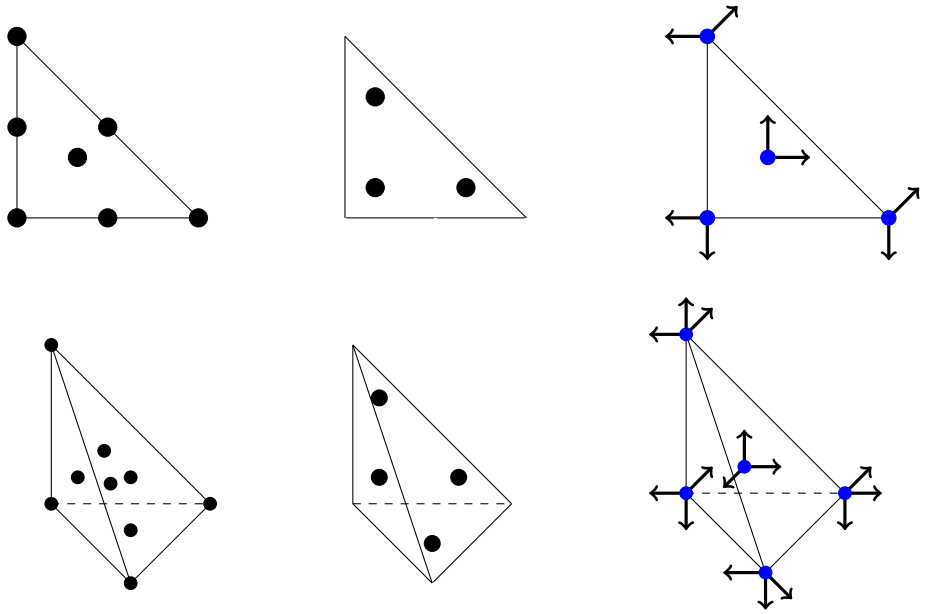

Figure 3.1.: Depiction of the degrees of freedom for the triplet $\mathrm{P}_2^+ - \mathrm{P}_1 - \mathrm{RT}_1$ on the reference triangle (top) and reference tetrahedron (bottom). The blue dots represent the quadrature points used for mass lumping. The associated weights are $\alpha = \frac{3}{4}$ and $\beta = \frac{1}{12}$ on the triangle and $\alpha = \frac{4}{5}$ and $\beta = \frac{1}{20}$ on the tetrahedron.

The global approximation spaces for the finite element method are then defined as

$$\begin{aligned} V_h &= \{v_h \in H_0^1(\Omega)^d \;\; : \;\; v_h|_K \in P_2(K)^+, \quad \forall K \in \mathcal{T}_h\}, \\ Q_h &= \{q_h \in L_0^2(\Omega) \;\; : \;\; q_h|_K \in P_1(K), \quad \forall K \in \mathcal{T}_h\}, \\ Z_h &= \{z_h \in H_0(\mathrm{div}, \Omega) \;\; : \;\; z_h|_K \in \mathrm{RT}_1(K), \quad \forall K \in \mathcal{T}_h\}. \end{aligned} \tag{3.28}$$

For later reference, let us recall the stability conditions

$$\inf_{p_h \in Q_h} \sup_{v_h \in V_h} \frac{b(v_h, p_h)}{\|v_h\|_{H^1(\Omega)} \|q_h\|_{L^2(\Omega)}} \geq \beta_h \tag{3.29}$$

and

$$\inf_{p_h \in Q_h} \sup_{w_h \in Z_h} \frac{d(w_h, p_h)}{\|w_h\|_{H(\mathrm{div},\Omega)} \|q_h\|_{L^2(\Omega)}} \geq \beta_h. \tag{3.30}$$

Under assumption (A2), these conditions hold with a constant β_h independent of the mesh size. The first condition can be found in [**19, Proposition 8.6.2.**] for the 2D case and [**19, Example 8.7.2.**] for the 3D case; for the second see [**19, Chapter 7.1.2**].

3.2.2. Choice of the quadrature formula

Let us recall the quadrature formula

$$\mathcal{Q}_K(f) = |K| \left(\alpha f(m) + \sum_{i=1}^{d+1} \beta f(v_i) \right) \tag{3.31}$$

that will be used for mass lumping of the RT_1 space. Here m and v_i represent the element midpoint and its vertices, which together comprise the integration points, while α and β are the corresponding weights defined in the caption of Figure 3.1. Let us recall that this quadrature formula is exact for polynomials of degree $p \leq 2$. We further set $(\kappa^{-1} w, z)_{h,K} = \mathcal{Q}_K((\kappa^{-1} w) \cdot z)$ and define by

$$m_h(w, z) \coloneqq \sum\nolimits_{K \in \mathcal{T}_h} (\kappa^{-1} w, z)_{h,K} \tag{3.32}$$

the corresponding inexact realization of the scalar product $m(\cdot, \cdot)$. From our considerations of the previous chapter, we immediately infer that

$$c_1(z_h, z_h) \leq m_h(z_h, z_h) \leq c_2(z_h, z_h) \qquad \forall z_h \in Z_h, \tag{3.33}$$

with uniform constants c_1, c_2 independent of the mesh $\mathcal{T}_h$.

3.2.3. Projection Operators

As a next step, we introduce two projection operators, which will be required for the choice of appropriate initial conditions and our error analysis later on.

Lemma 3.2.1 (Elliptic projection, part one)**.**
Let (A1) and (A2) hold. Then for any $(w, p) \in H_0(\mathrm{div}, \Omega) \times L^2_0(\Omega)$ there exists a unique $(\widetilde{w}_h, \widetilde{p}_h) \in Z_h \times Q_h$ with

$$\begin{aligned} m_h(\widetilde{w}_h, z_h) - d(z_h, \widetilde{p}_h) &= m(w, z_h) - d(z_h, p) && \forall z_h \in Z_h, \\ d(\widetilde{w}_h, q_h) &= d(w, q_h) && \forall q_h \in Q_h. \end{aligned}$$

Moreover, if (w, p) is sufficiently smooth, then

$$\begin{aligned} \|\pi_h^1 p - \widetilde{p}_h\|_{L^2(\Omega)} + \|w - \widetilde{w}_h\|_{L^2(\Omega)} &\leq c h^2 \|w\|_{H^2(\mathcal{T}_h)}, \\ \|p - \widetilde{p}_h\|_{L^2(\Omega)} &\leq c h^2 (\|w\|_{H^2(\mathcal{T}_h)} + \|p\|_{H^2(\mathcal{T}_h)}), \\ \|\mathrm{div}(w - \widetilde{w}_h)\|_{L^2(\Omega)} &\leq C h^2 \|\mathrm{div}\, w\|_{H^2(\mathcal{T}_h)}, \end{aligned}$$

where π_h^1 is the L^2-projection onto $P_1(\mathcal{T}_h)$.

Up to minor modifications concerning the boundary conditions, the assertions follow from Theorem 2.3.4. For approximating elastic displacements, we use

Lemma 3.2.2 (Elliptic projection, part two)**.**
Let (A1) and (A2) hold and let $u \in H_0^1(\Omega)^d$. Then there exists a unique $\widetilde{u}_h \in V_h$ such that

$$a(\widetilde{u}_h, v_h) = a(u, v_h) \qquad \forall v_h \in V_h. \tag{3.34}$$

Moreover, there exists a constant $C > 0$ such that

$$\|u - \widetilde{u}_h\|_{H^1(\Omega)} \leq Ch^2 \|u\|_{H^3(\mathcal{T}_h)}. \tag{3.35}$$

From Lemma 3.1.4 we know that $a(\cdot, \cdot)$ is elliptic and continuous. The existence and uniqueness of solutions to (3.34) then follow from the Lax-Milgram lemma, Cea's lemma, and the standard interpolation error estimate; see [**32, Chapter 3**].

3.2.4. Galerkin semi-discretization

We are now in the position to formulate and analyze the following discretization scheme.

Problem 3.2.3 (Semi-discretization)**.**
Let (p_0, w_0, u_0) denote compatible initial values satisfying (3.17)–(3.18)*. Define*

$$(w_{0,h}, p_{0,h}) = (\widetilde{w}_h(0), \widetilde{p}_h(0)) \in Z_h \times Q_h \tag{3.36}$$

by the elliptic projection of Lemma 3.2.1 and let $u_{0,h} \in V_h$ be the solution of

$$a(u_{0,h}, v_h) = (f(0), v_h) + b(v_h, p_{0,h}) \qquad \forall v_h \in V_h. \tag{3.37}$$

Then find $(u_h, p_h, w_h) \in H^1(0, T; V_h \times Q_h \times Z_h)$ such that for a.e. $0 < t < T$

$$\begin{aligned}
a(u_h(t), v_h) \;\; -b(v_h, p_h(t)) \qquad\qquad\qquad &= (f(t), v_h) \qquad \forall\, v_h \in V_h, && (3.38)\\
b(\partial_t u_h(t), q_h) + c(\partial_t p_h(t), q_h) + d(w_h(t), q_h) \;\; &= (g(t), q_h) \qquad \forall\, q_h \in Q_h, && (3.39)\\
-d(z_h, p_h(t)) \;\; + m_h(w_h(t), z_h) &= 0 \qquad \forall\, z_h \in Z_h. && (3.40)
\end{aligned}$$

Remark 3.2.4. The system (3.38)–(3.40) is a differential-algebraic set of equations. By definition (3.37), we see that (3.38) is satisfied at time $t = 0$. From Lemma 3.2.1 and the continuous compatibility conditions for the initial values, we obtain

$$\begin{aligned}
-d(z_h, p_{0,h}) + m_h(w_{0,h}, z_h) &= -d(z_h, \widetilde{p}_h(0)) + m_h(\widetilde{w}_h(0), z_h) \\
&= -d(z_h, p(0)) + m(w(0), z_h) = 0 \qquad \forall z_h \in Z_h.
\end{aligned} \tag{3.41}$$

This shows that (3.40) is also satisfied at $t = 0$. Hence our choice of initial values is consistent with the differential-algebraic equation. This is a necessary condition for the existence of a solution.

Based on the consistency of initial values and the discrete inf-sup stability of our discretization spaces, we can establish the existence of a unique discrete solution.

Lemma 3.2.5 (Discrete well-posedness)**.**
Let (A1) and (A2) hold and f, g be as in Theorem 3.1.5. Then Problem 3.2.3 has a unique solution $(u_h, p_h, w_h) \in H^1(0,T; V_h \times Q_h \times Z_h)$.

Proof. Since the initial values are consistent, see Remark 3.2.4, we can differentiate (3.38) and (3.40) with respect to time, which yields the equivalent system

$$\begin{aligned} a(\partial_t u_h(t), v_h) - b(v_h, \partial_t p_h(t)) &= (\partial_t f(t), v_h) && \forall\, v_h \in V_h, \\ b(\partial_t u_h(t), q_h) + c(\partial_t p_h(t), q_h) + d(w_h(t), q_h) &= (g(t), q_h) && \forall\, q_h \in Q_h, \\ -d(z_h, \partial_t p_h(t)) + m_h(\partial_t w_h(t), z_h) &= 0 && \forall\, z_h \in Z_h. \end{aligned}$$

After a choice of a basis for the approximation spaces, we arrive at a system

$$\begin{pmatrix} \mathsf{A} & -\mathsf{B}^\top & 0 \\ \mathsf{B} & \mathsf{C} & 0 \\ 0 & -\mathsf{D}^\top & \mathsf{M_h} \end{pmatrix} \begin{pmatrix} \partial_t \mathsf{u} \\ \partial_t \mathsf{p} \\ \partial_t \mathsf{w} \end{pmatrix} = \begin{pmatrix} \partial_t \mathsf{f} \\ \mathsf{g} - \mathsf{D}\mathsf{w} \\ 0 \end{pmatrix} \tag{3.42}$$

of ordinary differential equations. The finite element matrices A, B, C, D and $\mathsf{M_h}$ are the discrete versions of the corresponding operators. To show solvability, it suffices to verify the regularity of the matrix $\mathsf{M_h}$ and of the block matrix

$$\mathsf{E} \coloneqq \begin{pmatrix} \mathsf{A} & -\mathsf{B}^\top \\ \mathsf{B} & \mathsf{C} \end{pmatrix}.$$

The regularity of $\mathsf{M_h}$ follows from the norm equivalence property (3.33). For the block matrix E, we show the uniqueness of the corresponding homogeneous system. Let $\mathsf{x} = (\mathsf{u}, \mathsf{p})^\top$ with $\mathsf{E}\mathsf{x} = 0$. Then

$$0 = \mathsf{x}^\top \mathsf{E}\mathsf{x} = \begin{pmatrix} \mathsf{u} & \mathsf{p} \end{pmatrix} \begin{pmatrix} \mathsf{A} & -\mathsf{B}^\top \\ \mathsf{B} & \mathsf{C} \end{pmatrix} \begin{pmatrix} \mathsf{u} \\ \mathsf{p} \end{pmatrix} = \mathsf{u}^\top \mathsf{A}\mathsf{u} + \mathsf{p}^\top \mathsf{C}\mathsf{p} \geq \mathsf{u}^\top \mathsf{A}\mathsf{u}.$$

Since A was obtained from the elliptic bilinear form $a(\cdot,\cdot)$, we know that it is positive definite, and thus $\mathsf{u} = 0$. From the first row of $\mathsf{E}\mathsf{x} = 0$, we obtain $-\mathsf{B}^\top \mathsf{p} = 0$. The discrete inf-sup condition (3.29) now implies that $\mathsf{B}^\top$ is injective, and thus $\mathsf{p} = 0$. Since we are in finite dimensions and E is square, injectivity implies regularity. □

Before we begin with the error analysis, let us first state two auxiliary results concerning the discrete spaces and suitable projection operators.

3.3. Error estimates for the semi-discretization

We again start by splitting the error into projection and discrete errors by

$$\begin{aligned} u(t) - u_h(t) &= (u(t) - \widetilde{u}_h(t)) + (\widetilde{u}_h(t) - u_h(t)), \\ p(t) - p_h(t) &= (p(t) - \widetilde{p}_h(t)) + (\widetilde{p}_h(t) - p_h(t)), \\ w(t) - w_h(t) &= (w(t) - \widetilde{w}_h(t)) + (\widetilde{w}_h(t) - w_h(t)). \end{aligned} \tag{3.43}$$

The auxiliary approximations $\widetilde{u}_h$, $\widetilde{p}_h$, and $\widetilde{w}_h$ are chosen as the elliptic projections of the exact solutions, introduced in Lemma 3.2.1 and Lemma 3.2.2. The projection errors are then readily estimated by the bounds in these results.

3.3.1. Estimates for the discrete error

For ease of notation, we introduce the following abbreviation to the discrete errors:

$$\chi_u(t) \coloneqq \widetilde{u}_h(t) - u_h(t), \qquad \chi_p(t) \coloneqq \widetilde{p}_h(t) - p_h(t), \qquad \text{and} \qquad \chi_w(t) \coloneqq \widetilde{w}_h(t) - w_h(t).$$

From the definition of the discrete solution, the elliptic projections, and the variational characterization of (u, p, w) in Lemma 3.1.3, we obtain the following result.

Lemma 3.3.1 (Discrete error equation)**.**
Let (u, p, w) be the solution of (3.5)–(3.9) and let (u_h, p_h, w_h) denote the solution of Problem 3.2.3. Then the discrete errors satisfy

$$\begin{aligned}
a(\chi_u(t), v_h) \;& -b(v_h, \chi_p(t)) && = (R_f(t), v_h) && \forall\, v_h \in V_h, && (3.44)\\
b(\partial_t \chi_u(t), q_h) & + c(\partial_t \chi_p(t), q_h) + d(\chi_w(t), q_h) && = (R_g(t), q_h) && \forall\, q_h \in Q_h, && (3.45)\\
& -d(z_h, \chi_p(t)) \;\; + m_h(\chi_w(t), z_h) && = 0 && \forall\, z_h \in Z_h, && (3.46)
\end{aligned}$$

for a.a. $t > 0$, with residuals defined by

$$\begin{aligned}
(R_f(t), v_h) &:= -b(v_h, \widetilde{p}_h(t) - p(t)), && (3.47)\\
(R_g(t), q_h) &:= b(\partial_t \widetilde{u}_h(t) - \partial_t u(t), q_h) + c(\partial_t \widetilde{p}_h(t) - \pi_h^1 \partial_t p(t), q_h). && (3.48)
\end{aligned}$$

Proof. Using equation (3.38) and equation (3.13) with $v = v_h$ yields (3.44) with

$$(R_f(t), v_h) = a(\widetilde{u}_h(t) - u(t), v_h) - b(v_h, \widetilde{p}_h(t) - p(t)).$$

By the definition of $\widetilde{u}_h(t)$, the first term on the right-hand side vanishes, and we get (3.47). For the second residual, we use (3.39) and (3.14) with $q = q_h$, which gives

$$\begin{aligned}
(R_g(t), q_h) &= b(\partial_t \widetilde{u}_h(t) - \partial_t u(t), q_h) + c(\partial_t \widetilde{p}_h(t) - \partial_t p(t), q_h) + d(\widetilde{w}_h(t) - w(t), q_h)\\
&= (i) + (ii) + (iii).
\end{aligned}$$

For (ii), we use the properties of the L^2-projection and obtain

$$(ii) = c(\partial_t \widetilde{p}_h(t) - \pi_h^1 \partial_t p(t), q_h).$$

The term (iii) vanishes due to the definition of $\widetilde{w}_h(t)$, and the remaining term (i) leads to (3.48). In order to verify (3.46), we use (3.40) and (3.15) with $z = z_h$, which gives

$$\begin{aligned}
-d(z_h, \chi_p(t)) + m_h(\chi_w(t), z_h) &= -d(z_h, \widetilde{p}_h(t) - p(t)) + m_h(\widetilde{w}_h(t), z_h) - m(w(t), z_h)\\
&= -d(z_h, \widetilde{p}_h(t)) + m_h(\widetilde{w}_h(t), z_h) = 0.
\end{aligned}$$

Here we again utilized the definitions of the elliptic projection $(\widetilde{w}_h(t), \widetilde{p}_h(t))$ from Lemma 3.2.1. This already concludes the proof of the assertions. □

We next derive bounds for the discrete error components in terms of the residuals.

Lemma 3.3.2 (Discrete energy estimate)**.**
Let (χ_u, χ_p, χ_w) be defined as above. Then $\chi_p(0) = 0$, $\chi_w(0) = 0$ and

$$\|\chi_w(t)\|^2_{L^2(\Omega)} + \int_0^t \|\partial_t \chi_u(s)\|^2_{H^1(\Omega)}\, ds + \int_0^t \|\partial_t \chi_p(s)\|^2_{L^2(\Omega)}\, ds \\ \leq C\left(\int_0^t \|\partial_t R_f(s)\|^2_{H^{-1}(\Omega)} + \int_0^t \|R_g(s)\|^2_{L^2(\Omega)}\right)$$

with a constant $C > 0$ independent of the mesh size.

Proof. The properties of the initial values follow directly from their definition in (3.36). Moreover, from (3.37), we derive that

$$\begin{aligned} a(\chi_u(0), v_h) - b(v_h, \chi_p(0)) &= a(\widetilde{u}_h(0) - u_h(0), v_h) - b(v_h, \widetilde{p}_h(0) - p_h(0)) \\ &= -b(v_h, \widetilde{p}_h(0) - p_h(0)) = (R_f(0), v_h). \end{aligned}$$

In addition, from (3.36), we immediately get

$$-d(z_h, \chi_p(0)) + m_h(\chi_w(0), z_h) = 0.$$

By now formally differentiating (3.44) and (3.46) in time, we obtain the system

$$\begin{aligned} a(\partial_t \chi_u(t), v_h) - b(v_h, \partial_t \chi_p(t)) &= (\partial_t R_f(t), v_h) && \forall\, v_h \in V_h, && (3.49)\\ b(\partial_t \chi_u(t), q_h) + c(\partial_t \chi_p(t), q_h) + d(\chi_w(t), q_h) &= (R_g(t), q_h) && \forall\, q_h \in Q_h, && (3.50)\\ -d(z_h, \partial_t \chi_p(t)) + m_h(\partial_t \chi_w(t), z_h) &= 0 && \forall\, z_h \in Z_h. && (3.51) \end{aligned}$$

We proceed by testing the equations (3.49)–(3.51) with $v_h = \partial_t \chi_u(t)$, $q_h = \partial_t \chi_p(t)$ and $z_h = \chi_w(t)$ and adding the three resulting equations. This yields

$$a(\partial_t \chi_u(t), \partial_t \chi_u(t)) + c(\partial_t \chi_p(t), \partial_t \chi_p(t)) + m_h(\partial_t \chi_w(t), \chi_w(t)) \\ = (\partial_t R_f(t), \partial_t \chi_u(t)) + (R_g(t), \partial_t \chi_p(t)).$$

Using the properties of the bilinear form stated in Lemma 3.1.4, we see that

$$\alpha \|\partial_t \chi_u(t)\|^2_{H^1(\Omega)} + c_s \|\partial_t \chi_p(t)\|^2_{L^2(\Omega)} + \frac{1}{2}\frac{d}{dt}\|\kappa^{-1/2}\chi_w(t)\|^2_h \\ = (\partial_t R_f(t), \partial_t \chi_u(t)) + (R_g(t), \partial_t \chi_p(t)).$$

Using Young's inequality, integrating from 0 to t, and also using (3.33) leads to

$$\begin{aligned} c_1 \|\chi_w(t)\|^2_{L^2(\Omega)} + \int_0^t \alpha \|\partial_t \chi_u(s)\|^2_{H^1(\Omega)}\, ds + \int_0^t c_s \|\partial_t \chi_p(s)\|^2_{L^2(\Omega)}\, ds \\ \leq \frac{1}{\alpha}\int_0^t \|\partial_t R_f(s)\|^2_{H^{-1}(\Omega)}\, ds + \frac{\alpha}{4}\int_0^t \|\partial_t \chi_u(s)\|^2_{H^1(\Omega)}\, ds \\ + \frac{1}{\varepsilon}\int_0^t \|R_g(s)\|^2_{L^2(\Omega)}\, ds + \frac{\varepsilon}{4}\int_0^t \|\partial_t \chi_p(s)\|^2_{L^2(\Omega)}\, ds, \end{aligned} \tag{3.52}$$

for any $\varepsilon > 0$. Here we also used that $\chi_w(0) = 0$. To estimate $\|\partial_t \chi_p\|_{L^2(\Omega)}$ independently of the value of c_s, we proceed as follows: By the first inf-sup condition (3.29), we obtain an element $\widehat{v}_h \in V_h$ such that $\operatorname{div} \widehat{v}_h(t) = \partial_t \chi_p(t)$ and $\|\widehat{v}_h(t)\|_{H^1(\Omega)} \leq \frac{1}{\beta_h} \|\partial_t \chi_p(t)\|_{L^2(\Omega)}$. We then choose $v_h = \widehat{v}_h(t)$ in (3.49) and obtain

$$\begin{aligned}\|\partial_t \chi_p(t)\|^2_{L^2(\Omega)} &= a(\partial_t \chi_u(t), \widehat{v}_h(t)) - (\partial_t R_f(t), \widehat{v}_h(t)) \\ &\leq (c_a \|\partial_t \chi_u(t)\|_{H^1(\Omega)} + \|\partial_t R_f(t)\|_{H^{-1}(\Omega)}) \|\widehat{v}_h(t)\|_{H^1(\Omega)} \\ &\leq \frac{1}{\beta_h} (c_a \|\partial_t \chi_u(t)\|_{H^1(\Omega)} + \|\partial_t R_f(t)\|_{H^{-1}(\Omega)}) \|\partial_t \chi_p(t)\|_{L^2(\Omega)}.\end{aligned}$$

Dividing by $\|\partial_t \chi_p(t)\|_{L^2(\Omega)}$ and setting $c = \max(c_a/\beta_h, 1)$ finally yields

$$\|\partial_t \chi_p(t)\|_{L^2(\Omega)} \leq c(\|\partial_t \chi_u(t)\|_{H^1(\Omega)} + \|R_f(t)\|_{H^{-1}(\Omega)}).$$

With this, we can estimate the last term in (3.52) by

$$\frac{\varepsilon}{4} \int_0^t \|\partial_t \chi_p(s)\|^2_{L^2(\Omega)} \, ds \leq \frac{\varepsilon \cdot c}{2} \left(\int_0^t \|\partial_t \chi_u(s)\|_{H^1(\Omega)} \, ds + \int_0^t \|R_f(s)\|_{H^{-1}(\Omega)} \, ds \right).$$

Further choosing $\varepsilon = \frac{\alpha}{2c}$ in (3.52) allows to absorb the last term in (3.52) into the left hand side of the estimate, which concludes the proof. □

3.3.2. Error estimates

By a combination of the estimates for the elliptic projections and for the discrete error component, we are now in the position to state and prove the main result of this section.

Theorem 3.3.3 (Second order convergence for the semi-discrete scheme)**.**
Let (A1) and (A2) hold, (u, p, w) be a sufficiently smooth solution of (3.5)–(3.11) and (u_h, p_h, w_h) be the corresponding solution to Problem 3.2.3. Then

$$\|u - u_h\|_{H^1(0,T,H^1(\Omega))} + \|p - p_h\|_{H^1(0,T,L^2(\Omega))} + \|w - w_h\|_{L^\infty(0,T,L^2(\Omega))} \leq C(u,p,w) h^2,$$

where $C(u,p,w) = c\,(\|u\|_{H^1(0,T;H^3(\mathcal{T}_h))} + \|p\|_{H^1(0,T;H^2(\mathcal{T}_h))} + \|w\|_{H^1(0,T;H^2(\mathcal{T}_h))})$. Moreover

$$\|\operatorname{div}(w - w_h)\|_{L^2(0,T,L^2(\Omega))} \leq \left(c' \|\operatorname{div} w\|_{L^2(0,T,H^2(\mathcal{T}_h))} + C(u,p,w) \right) h^2.$$

The constants c and c' in these estimates are independent of h.

Proof. We begin by splitting the error as described in (3.43). For the projection error, using the estimates of Lemma 3.2.1 and Lemma 3.2.2, we obtain

$$\begin{aligned}\|u - \widetilde{u}_h\|_{H^1(0,T,H^1(\Omega))} &+ \|p - \widetilde{p}_h\|_{H^1(0,T,L^2(\Omega))} + \|w - \widetilde{w}_h\|_{L^\infty(0,T,L^2(\Omega))} \\ &\leq Ch^2 \left(\|u\|_{H^1(0,T,H^3(\mathcal{T}_h))} + \|p\|_{H^1(0,T,H^2(\mathcal{T}_h))} + \|w\|_{H^1(0,T,H^2(\mathcal{T}_h))} \right).\end{aligned}$$

From the estimate of Lemma 3.3.2, we further know that

$$\begin{aligned}\|\chi_w(t)\|^2_{L^2(\Omega)} &+ \int_0^t \|\partial_t \chi_u(s)\|^2_{H^1(\Omega)} \, ds + \int_0^t \|\partial_t \chi_p(s)\|^2_{L^2(\Omega)} \, ds \\ &\leq C \left(\int_0^t \|\partial_t R_f(s)\|^2_{H^{-1}(\Omega)} + \int_0^t \|R_g(s)\|^2_{L^2(\Omega)} \right) = C\Big((i) + (ii)\Big).\end{aligned}$$

The residuals are further estimated as follows: By definition of the dual norm, we get

$$\|\partial_t R_f(t)\|_{H^{-1}(\Omega)} = \sup_{v\in H_0^1(\Omega)} \frac{-(\partial_t \widetilde{p}_h(t) - \partial_t p(t), \operatorname{div} v)}{\|v\|_{H^1(\Omega)}} \leq \|\partial_t \widetilde{p}_h(t) - \partial_t p(t)\|_{L^2(\Omega)}.$$

Using Lemma 3.2.1 allows us to estimate

$$(i) \leq \int_0^t \|\partial_t \widetilde{p}_h(s) - \partial_t p(s)\|^2_{L^2(\Omega)} \leq ch^4 \left(\|\partial_t w\|^2_{L^2(0,t,H^2(\mathcal{T}_h))} + \|\partial_t p\|^2_{L^2(0,t,H^2(\mathcal{T}_h))} \right).$$

For the second residual, we use (3.48) and estimate

$$\begin{aligned}(ii) &\leq 2\int_0^t \|\operatorname{div}(\partial_t \widetilde{u}_h(s) - \partial_t u(s))\|^2_{L^2(\Omega)}\, ds \; + 2c_s \int_0^t \|\partial_t \widetilde{p}_h(s) - \pi_h^1 \partial_t p(s)\|^2_{L^2(\Omega)}\, ds \\ &\leq ch^4 \left(\|\partial_t u\|^2_{L^2(0,t,H^3(\mathcal{T}_h))} + \|\partial_t w\|^2_{L^2(0,t,H^2(\mathcal{T}_h))} \right).\end{aligned}$$

Here we employed the estimates of Lemma 3.2.1 and Lemma 3.2.2 for the elliptic projections. By integration in time and using $\chi_p(0) = 0$, we then obtain

$$\begin{aligned}&\|\chi_u\|_{L^2(0,t,L^2(\Omega))} + \|\chi_p\|_{L^2(0,t,L^2(\Omega))} \\ &\qquad \leq C(t) \left(\|\chi_u(0)\|_{L^2(\Omega)} + \|\partial_t \chi_u\|_{L^2(0,t,L^2(\Omega))} + \|\partial_t \chi_p\|_{L^2(0,t,L^2(\Omega))} \right).\end{aligned}$$

It remains to bound $\|\chi_u(0)\|_{H^1}$. Using (3.13), (3.38) with $v = v_h$, and (3.34) yields

$$\begin{aligned}a(\chi_u(0), v_h) &= a(\widetilde{u}_h(0) - u_h(0), v_h) = a(\widetilde{u}_h(0), v_h) - (f(0), v_h\rangle - (p_h(0), \operatorname{div} v_h) \\ &= (p(0) - \widetilde{p}_h(0), \operatorname{div} v_h).\end{aligned}$$

By choosing $v_h = \chi_u(0)$ and using the ellipticity of the bilinear form $a(\cdot,\cdot)$, we get

$$\begin{aligned}\alpha \|\chi_u(0)\|^2_{H^1(\Omega)} &\leq \|p(0) - \widetilde{p}_h(0)\|_{L^2(\Omega)} \|\operatorname{div} \chi_u(0)\|_{L^2(\Omega)} \\ &\leq \sqrt{d} \|p(0) - \widetilde{p}_h(0)\|_{L^2(\Omega)} \|\chi_u(0)\|_{H^1(\Omega)}.\end{aligned}$$

Dividing by $\|\chi_u(0)\|_{H^1(\Omega)}$ and using the estimate of Lemma 3.2.1 for the error of the elliptic projection, we find

$$\|\chi_u(0)\|_{H^1(\Omega)} \leq ch^2 (\|w(0)\|_{H^2(\mathcal{T}_h)} + \|p(0)\|_{H^2(\mathcal{T}_h)}). \tag{3.53}$$

Since the estimates hold uniformly in $t > 0$, we obtain the first statement of the theorem. For the projection error in the divergence, we see from Lemma 3.2.1 that

$$\|\operatorname{div}(w - \widetilde{w}_h)\|_{L^2(0,T,L^2(\Omega))} \leq ch^2 \|\operatorname{div} w\|_{L^2(0,T,H^2(\mathcal{T}_h))}.$$

Since $\operatorname{div} Z_h \subseteq Q_h$, we may choose $q_h = \operatorname{div} \chi_w(t)$ as a test function in (3.45), which gives

$$\|\operatorname{div} \chi_w(t)\|^2_{L^2(\Omega)} \leq (R_g(t), \operatorname{div} \chi_w(t)) + b(\partial_t \chi_u(t), \operatorname{div} \chi_w(t)) + c(\partial_t \chi_p(t), \operatorname{div} \chi_w(t)).$$

By using Cauchy-Schwarz and elementary computations, we obtain

$$\begin{aligned}&\int_0^t \|\operatorname{div} \chi_w(t)\|^2_{L^2(\Omega)}\, ds \\ &\leq C \left(\int_0^t \|R_g(t)\|^2_{L^2(\Omega)}\, ds + \int_0^t \|\partial_t \chi_u(t)\|^2_{H^1(\Omega)}\, ds + c_s \int_0^t \|\partial_t \chi_p(t)\|^2_{L^2(\Omega)}\, ds \right).\end{aligned}$$

The terms on the right-hand side have already been estimated before, which leads to the second estimate of the lemma. □

Concluding remarks

The proposed inexact Galerkin method with mass lumping is indeed second-order convergent in all three variables, i.e., mass lumping has no negative effect on the accuracy of the method. Based on the results of Theorem 2.3.4 and duality arguments, we expect third-order convergence for the pressure averages, if the domain Ω is convex and the solution is sufficiently smooth. This will also be observed in our numerical tests in Section 3.5.

3.4. Time discretization

Let $t^n = n\tau$, $0 \leq n \leq N$ be a given sequence of uniformly spaced time points with time step $\tau = T/N$. We denote by $f^n = f(t^n)$ the evaluation of a function f on the resulting time grid $I_\tau = \{0 = t^0 < t^1 < \ldots < t^N = T\}$ and write

$$f^{n-1/2} = \frac{1}{2}(f^n + f^{n-1}) \qquad \text{and} \qquad d_\tau f^{n-1/2} = \frac{1}{\tau}(f^n - f^{n-1})$$

the usual average and difference quotient. For functions $g \in L^2(0,T)$, we denote by

$$\widehat{g}^{n-1/2} = \frac{1}{\tau}\int_{t^{n-1}}^{t^n} g(t)dt$$

the average over the time interval $[t^{n-1}, t^n]$. Let us note that $g^{n-1/2} = \widehat{g}^{n-1/2}$ if g is affine linear on the interval $[t^{n-1}, t^n]$. For the convenience of notation, we further introduce discrete variants of the corresponding Bochner norms by

$$\|u\|^p_{\ell^p(I_\tau,V)} \coloneqq \sum_{n=0}^{N} \tau \|u^n\|^p_V \qquad \text{and} \qquad \|u\|_{\ell^\infty(I_\tau,V)} \coloneqq \max_{0\leq n\leq N} \|u^n\|_V,$$

and in a similar manner, we define the norms $\|\widehat{u}\|^p_{\ell^p(I_\tau,V)} \coloneqq \sum_{n=1}^{N} \tau \|u^{n-1/2}\|^p_V$. The fully discrete approximation for problem (3.5)–(3.9) is now described as follows.

Problem 3.4.1 (Full discretization)**.**
Let the initial values (w_h^0, p_h^0, u_h^0) be chosen as in Problem 3.2.3. Then for $1 \leq n \leq N$, find $(u_h^n, p_h^n, w_h^n)_{1\leq n\leq N} \subset V_h \times Q_h \times Z_h$ such that

$$a(u_h^n, v_h) \quad -b(v_h, p_h^n) \quad = (f^n, v_h) \qquad \forall\, v_h \in V_h, \tag{3.54}$$

$$b(d_\tau u_h^{n-1/2}, q_h) + c(d_\tau p_h^{n-1/2}, q_h) + d(w_h^{n-1/2}, q_h) = (\widehat{g}^{n-1/2}, q_h) \qquad \forall\, q_h \in Q_h, \tag{3.55}$$

$$-d(z_h, p_h^n) \quad + m_h(w_h^n, z_h) \quad = 0 \qquad \forall\, z_h \in Z_h. \tag{3.56}$$

Remark 3.4.2. The underlying time stepping scheme is similar to the implicit trapezoidal rule. In particular, if the right-hand side g is linear in time on I_τ, then the two methods coincide. With minor modifications, our analysis also covers the trapezoidal rule, which was used in [**75**] for numerical tests.

For completeness, let us again establish the well-posedness of the fully discrete scheme.

Lemma 3.4.3 (Discrete well-posedness)**.**
Let (A1) and (A2) hold and $f \in H^1(0,T,H^{-1}(\Omega))$, $g \in L^2(0,T,L_0^2(\Omega))$. Then Problem 3.4.1 attains a unique solution.

Proof. Since the initial values are chosen like in the semi-discrete case, we can see that (3.54) and (3.56) are also satisfied at $t = 0$. After discrete differentiation of (3.54) and (3.56), the algebraic realization of Problem 3.4.1 leads to a system of the form

$$\begin{aligned} \mathsf{A}d_\tau\mathsf{u}^{n-1/2} - \mathsf{B}^\top d_\tau\mathsf{p}^{n-1/2} &= d_\tau\mathsf{f}^{n-1/2}, \\ \mathsf{B}d_\tau\mathsf{u}^{n-1/2} + \mathsf{C}d_\tau\mathsf{p}^{n-1/2} + \mathsf{D}\mathsf{w}^{n-1/2} &= \widehat{\mathsf{g}}^{n-1/2}, \\ -\mathsf{D}^\top d_\tau\mathsf{p}^{n-1/2} + \mathsf{M}_\mathsf{h} d_\tau\mathsf{w}^{n-1/2} &= 0. \end{aligned}$$

The system of equations can be rewritten in recursive form, and the regularity of the update matrix can be shown with similar arguments as on the semi-discrete level. From this, it follows that each step of the method is well-defined. □

3.4.1. Error decomposition and discrete stability

We again split the error into a projection error and a discrete error by

$$u(t^n) - u_h^n = (u(t^n) - \widetilde{u}_h(t^n)) + (\widetilde{u}_h(t^n) - u_h^n), \tag{3.57}$$
$$p(t^n) - p_h^n = (p(t^n) - \widetilde{p}_h(t^n)) + (\widetilde{p}_h(t^n) - p_h^n), \tag{3.58}$$
$$w(t^n) - w_h^n = (w(t^n) - \widetilde{w}_h(t^n)) + (\widetilde{w}_h(t^n) - w_h^n). \tag{3.59}$$

Like in the previous section, $\widetilde{u}_h(t^n), \widetilde{p}_h(t^n), \widetilde{w}_h(t^n)$ denote the elliptic projections defined in Lemma 3.2.1 and Lemma 3.2.2, which allows estimating the projection error immediately. For ease of notation, we introduce abbreviations

$$\chi_u^n = \widetilde{u}_h(t^n) - u_h^n, \qquad \chi_p^n = \widetilde{p}_h(t^n) - p_h^n, \qquad \text{and} \qquad \chi_w^n = \widetilde{w}_h(t^n) - w_h^n.$$

By the particular definitions of the projections, the discrete solution, and the variational characterization of the continuous solution, we now obtain the following result.

Lemma 3.4.4 (Discrete error equation)**.**
Let (u,p,w) be the solution of (3.5)–(3.9) and let $(u_h^n, p_h^n, w_h^n)_{0\le n\le N}$ denote the corresponding solution of Problem 3.4.1. Then the discrete errors satisfy the equations

$$a(\chi_u^n, v_h) - b(v_h, \chi_p^n) = (R_f^n, v_h) \qquad \forall\, v_h \in V_h, \tag{3.60}$$
$$b(d_\tau\chi_u^{n-1/2}, q_h) + c(d_\tau\chi_p^{n-1/2}, q_h) + d(\chi_w^{n-1/2}, q_h) = (R_g^{n-1/2}, q_h) \qquad \forall\, q_h \in Q_h, \tag{3.61}$$
$$-d(z_h, \chi_p^n) + m_h(\chi_w^n, z_h) = 0 \qquad \forall\, z_h \in Z_h, \tag{3.62}$$

for all $1 \le n \le N$ with discrete residuals defined by

$$(R_f^n, v_h) := -b(v_h, \widetilde{p}_h(t^n) - p(t^n)), \tag{3.63}$$
$$\begin{aligned} (R_g^{n-1/2}, q_h) := {} & b(d_\tau\widetilde{u}_h^{n-1/2} - d_\tau u^{n-1/2}, q_h) + c(d_\tau\widetilde{p}_h^{n-1/2} - \pi_h^1 d_\tau p^{n-1/2}, q_h) \\ & + d(w^{n-1/2} - \widehat{w}^{n-1/2}, q_h). \end{aligned} \tag{3.64}$$

Proof. The identities (3.60) and (3.62) follow just as in the semi-discrete case; see the proof of Lemma 3.3.1. Let us take a closer look at the second identity (3.61), since it is the only term containing discrete time derivatives. We take equation (3.14), divide by τ, integrate from t^{n-1} to t^n, and use the definition of d_τ, which gives

$$b(d_\tau u^{n-1/2}, q) + c(d_\tau p^{n-1/2}, q) + d(\widehat{w}^{\,n-1/2}, q) = (\widehat{g}^{\,n-1/2}, q).$$

Choosing $q = q_h$ and subtracting (3.55) then already leads to the desired identity (3.64) for the residual $R_g^{n-1/2}$. This completes the proof of the lemma. □

By discrete energy estimates, we further obtain the following stability result.

Lemma 3.4.5 (Discrete energy identity)**.**
Let $(\chi_u^n, \chi_p^n, \chi_w^n)$ be defined as above. Then $\chi_p^0 = 0$, $\chi_w^0 = 0$ and

$$\|\chi_w^n\|_{L^2(\Omega)}^2 + \sum_{i=1}^n \tau \|d_\tau \chi_u^{i-1/2}\|_{H^1(\Omega)}^2 + \sum_{i=1}^n \tau \|d_\tau \chi_p^{i-1/2}\|_{L^2(\Omega)}^2 \\ \leq C \left(\sum_{i=1}^n \tau \|d_\tau R_f^{i-1/2}\|_{H^{-1}(\Omega)}^2 + \sum_{i=1}^n \tau \|R_g^{i-1/2}\|_{L^2(\Omega)}^2 \right)$$

with a constant $C > 0$ which is independent of the mesh size h.

Proof. By the same arguments as used in the proof of Lemma 3.3.2, we can show that (3.65) and (3.67) are also satisfied for $t = 0$. Next, by subtracting (3.60) from (3.62) at t^{n-1} and t^n and then dividing by τ, we obtain

$$\begin{aligned}
a(d_\tau \chi_u^{n-1/2}, v_h) - b(v_h, d_\tau \chi_p^{n-1/2}) &= (d_\tau R_f^{n-1/2}, v_h) && \forall\, v_h \in V_h, \quad (3.65)\\
b(d_\tau \chi_u^{n-1/2}, q_h) + c(d_\tau \chi_p^{n-1/2}, q_h) + d(\chi_w^{n-1/2}, q_h) &= (R_g^{n-1/2}, q_h) && \forall\, q_h \in Q_h, \quad (3.66)\\
-d(z_h, d_\tau \chi_p^{n-1/2}) + m_h(d_\tau \chi_w^{n-1/2}, z_h) &= 0 && \forall\, z_h \in Z_h. \quad (3.67)
\end{aligned}$$

Testing these equations with $v_h = d_\tau \chi_u^{n-1/2}$, $q_h = d_\tau \chi_p^{n-1/2}$, and $z_h = \chi_w^{n-1/2}$ and then adding them up results in

$$a(d_\tau \chi_u^{n-1/2}, d_\tau \chi_u^{n-1/2}) + c(d_\tau \chi_p^{n-1/2}, d_\tau \chi_p^{n-1/2}) + m_h(d_\tau \chi_w^{n-1/2}, \chi_w^{n-1/2}) \\ = (d_\tau R_f^{n-1/2}, d_\tau \chi_u^{n-1/2}) + (R_g^{n-1/2}, d_\tau \chi_p^{n-1/2}).$$

By the definition and properties of the bilinear forms stated in Lemma 3.1.4, we obtain

$$\begin{aligned}
&\alpha \|d_\tau \chi_u^{n-1/2}\|_{H^1(\Omega)}^2 + c_s \|d_\tau \chi_p^{n-1/2}\|_{L^2(\Omega)}^2 + \frac{1}{2\tau} \left(\|\kappa^{-1/2} \chi_w^n\|_h^2 - \|\kappa^{-1/2} \chi_w^{n-1}\|_h^2 \right) \\
&= (d_\tau R_f^{n-1/2}, d_\tau \chi_u^{n-1/2}) + (R_g^{n-1/2}, d_\tau \chi_p^{n-1/2}) \\
&\leq \frac{1}{\varepsilon} \|d_\tau R_f^{n-1/2}\|_{H^{-1}(\Omega)}^2 + \frac{\varepsilon}{4} \|d_\tau \chi_u^{n-1/2}\|_{H^1(\Omega)}^2 + \frac{1}{\eta} \|R_g^{n-1/2}\|_{L^2(\Omega)}^2 + \frac{\eta}{4} \|d_\tau \chi_p^{n-1/2}\|_{L^2(\Omega)}^2.
\end{aligned}$$

Multiplying by τ, summing from 1 to N, and using $\chi_w^0 = 0$ then leads to

$$\begin{aligned}\frac{1}{2}\|\kappa^{-1/2}\chi_w^N\|_h^2 + \sum_{n=1}^{N}\tau c_k\|d_\tau\chi_u^{n-1/2}\|_{H^1(\Omega)}^2 + \sum_{n=1}^{N}\tau c_s\|d_\tau\chi_p^{n-1/2}\|_{L^2(\Omega)}^2 \\ \leq \frac{1}{\varepsilon}\sum_{n=1}^{N}\tau\|d_\tau R_f^{n-1/2}\|_{H^{-1}(\Omega)}^2 + \frac{\varepsilon}{4}\sum_{n=1}^{N}\tau\|d_\tau\chi_u^{n-1/2}\|_{H^1(\Omega)}^2 + \\ + \frac{1}{\eta}\sum_{n=1}^{N}\tau\|R_g^{n-1/2}\|_{L^2(\Omega)}^2 + \frac{\eta}{4}\sum_{n=1}^{N}\tau\|d_\tau\chi_p^{n-1/2}\|_{L^2(\Omega)}^2.\end{aligned} \tag{3.68}$$

Just as in the proof of Lemma 3.3.2, we can show that

$$\|d_\tau\chi_p^{n-1/2}\|_{L^2(\Omega)} \leq c'\left(\|d_\tau\chi_u^{n-1/2}\|_{H^1(\Omega)} + \|d_\tau R_f^{n-1/2}\|_{H^{-1}(\Omega)}\right), \tag{3.69}$$

which allows us to control the norm of $d_\tau\chi_p^{n-1/2}$ independent of c_s. Choosing ε and η appropriately in (3.68) allows us to absorb the necessary terms to the left-hand side, which then completes the proof.

□

3.4.2. Error estimates

By a combination of the estimates for the discrete error and the elliptic projection, we can now prove the second main result of this chapter.

Theorem 3.4.6 (Second order convergence for the fully-discrete scheme)**.**
Let (A1) and (A2) hold, (u, p, w) be a sufficiently smooth solution of (3.5)–(3.11) and $(u_h^n, p_h^n, w_h^n)_{0\leq n\leq N}$ be the corresponding solution of Problem 3.4.1. Then

$$\begin{aligned}\|u - u_h\|_{\ell^2(I_\tau,H^1(\Omega))} + \|p - p_h\|_{\ell^2(I_\tau,L^2(\Omega))} \\ + \|w - w_h\|_{\ell^\infty(I_\tau,L^2(\Omega))} \leq C_1(u,p,w)h^2 + C_2(w)\tau^2\end{aligned}$$

with $C_1(u,p,w) = c'(\|u\|_{H^1(0,T;H^3(\mathcal{T}_h))} + \|p\|_{H^1(0,T;H^2(\mathcal{T}_h))} + \|w\|_{H^1(0,T;H^2(\mathcal{T}_h))})$ and further $C_2(w) = c''\|\operatorname{div}\partial_{tt}w\|_{L^\infty(0,T,L^2(\Omega))}$. Moreover

$$\|\operatorname{div}(\widehat{w} - \widehat{w}_h)\|_{\ell^2(L^2(\Omega))} \leq \left(c'''\|\operatorname{div} w\|_{\ell^2(H^2(\mathcal{T}_h))} + C_1(u,p,w)\right)h^2 + C_2(w)\tau^2.$$

The constants c', c'' and c''' in these bounds are independent of h and τ.

Proof. We begin by splitting the error according to (3.57)–(3.59). For the projection error, we directly see that

$$\begin{aligned}\|u - \widetilde{u}_h\|_{\ell^2(I_\tau,H^1(\Omega))} + \|p - \widetilde{p}_h\|_{\ell^2(I_\tau,L^2(\Omega))} + \|w - \widetilde{w}_h\|_{\ell^\infty(I_\tau,L^2(\Omega))} \\ \leq Ch^2\left(\|u\|_{\ell^2(I_\tau,H^3(\mathcal{T}_h))} + \|p\|_{\ell^2(I_\tau,H^2(\mathcal{T}_h))} + \|w\|_{\ell^\infty(I_\tau,H^2(\mathcal{T}_h))}\right),\end{aligned}$$

where we simply used the estimates of Lemma 3.2.1 and Lemma 3.2.2. From Lemma 3.4.5, we further know that

$$\|\chi_w^n\|_{L^2(\Omega)}^2 + \sum_{i=1}^n \tau \|d_\tau \chi_u^{i-1/2}\|_{H^1(\Omega)}^2 + \sum_{i=1}^n \tau \|d_\tau \chi_p^{i-1/2}\|_{L^2(\Omega)}^2 \\ \leq C \left(\sum_{i=1}^n \tau \|d_\tau R_f^{i-1/2}\|_{H^{-1}(\Omega)}^2 + \sum_{i=1}^n \tau \|R_g^{i-1/2}\|_{L^2(\Omega)}^2 \right) = C\Big((i) + (ii)\Big).$$

The first term on the right-hand side can be further estimated by

$$\begin{aligned}(i) &= \sum_{i=1}^n \tau \|d_\tau(\widetilde{p}_h^{i-1/2} - p^{i-1/2})\|_{L^2(\Omega)}^2 = \sum_{i=1}^n \tau \left\| \frac{1}{\tau} \int_{t^{i-1}}^{t^i} \partial_t \widetilde{p}_h(s) - \partial_t p(s)\, ds \right\|_{L^2(\Omega)}^2 \\ &\leq c \sum_{i=1}^n \int_{t^{i-1}}^{t^i} \|\partial_t \widetilde{p}_h(s) - \partial_t p(s)\, ds\|_{L^2(\Omega)}^2 = c\|\partial_t \widetilde{p}_h - \partial_t p\|_{L^2(0,t^n,L^2(\Omega))}^2 \\ &\leq c' h^4 \left(\|\partial_t w\|_{L^2(0,t^n,H^2(\mathcal{T}_h))}^2 + \|\partial_t p\|_{L^2(0,t^n,H^2(\mathcal{T}_h))}^2 \right).\end{aligned}$$

Here we used Jensen's inequality in the second step and the projection error estimates of Lemma 3.2.1 in the last step. For the second term, we further split

$$\begin{aligned}(ii) &\leq 3 \sum_{i=1}^n \tau \|\mathrm{div}\,(d_\tau \widetilde{u}_h^{i-1/2} - d_\tau u^{i-1/2})\|_{L^2(\Omega)}^2 \\ &\quad + 3 \sum_{i=1}^n \tau c_s \|d_\tau \widetilde{p}_h^{i-1/2} - \pi_h^1 d_\tau p^{i-1/2}\|_{L^2(\Omega)}^2 + 3 \sum_{i=1}^n \tau \|\mathrm{div}\,(w^{i-1/2} - \widehat{w}^{\,i-1/2})\|_{L^2(\Omega)}^2 \\ &= 3(iii) + 3(iv) + 3(v).\end{aligned}$$

For (iii), we use the same arguments as for (i), and obtain

$$\begin{aligned}(iii) &= \sum_{i=1}^n \tau \left\| \frac{1}{\tau} \int_{t^{i-1}}^{t^i} \mathrm{div}\,(\partial_t \widetilde{u}_h(s) - \partial_t u(s))\, ds \right\|_{L^2(\Omega)}^2 \\ &\leq c' \|\mathrm{div}\,(\partial_t \widetilde{u}_h - \partial_t u)\|_{L^2(0,t^n,L^2(\Omega))}^2 \leq c'' h^4 \|\partial_t u\|_{L^2(0,t^n;H^3(\mathcal{T}_h))}^2.\end{aligned}$$

Here we used the projection error estimate of Lemma 3.2.2. For the next term, we compute in a similar fashion that

$$(iv) \leq c h^4 \left(\|\partial_t w\|_{L^2(0,t^n,H^2(\mathcal{T}_h))}^2 + \|\partial_t p\|_{L^2(0,t^n,H^2(\mathcal{T}_h))}^2 \right).$$

To bound the remaining term (v), which is the error that accounts for the time discretization, we note that

$$|w^{n-1/2} - \widehat{w}^{\,n-1/2}| = \left| \frac{w^n + w^{n-1}}{2} - \frac{1}{\tau} \int_{t^{n-1}}^{t^n} w(s)\, ds \right| \leq C\tau^2 \max_{t \in [t^{n-1};t^n]} |\partial_{tt} u(t)|.$$

With this, we can then directly estimate

$$(v) \leq c \max_{1\leq n\leq N} \left\|\operatorname{div}(w^{n-1/2} - \widehat{w}^{n-1/2})\right\|^2_{L^2(\Omega)} \leq c'\tau^4 \|\operatorname{div} \partial_{tt} w\|^2_{L^\infty(0,t^n,L^2(\Omega))}.$$

Summing over n, using Cauchy-Schwarz inequalities, and $\chi_p = 0$, we see that

$$\|\chi_u^n\|_{L^2(\Omega)} + \|\chi_p^n\|_{L^2(\Omega)} \leq \|\chi_u^0\|_{L^2(\Omega)} + \sum_{i=1}^{n} \tau \|d_\tau \chi_u^{i-1/2}\|^2_{H^1(\Omega)} + \sum_{i=1}^{n} \tau \|d_\tau \chi_p^{i-1/2}\|^2_{L^2(\Omega)}.$$

Just like in (3.53), we further get

$$\|\chi_u^0\|_{H^1(\Omega)} \leq ch^2(\|w(0)\|_{H^2(\mathcal{T}_h)} + \|p(0)\|_{H^2(\mathcal{T}_h)}).$$

By collecting all results derived so far, we already obtain the first estimate of the theorem. From the results of Lemma 3.2.1, we also know that

$$\|\operatorname{div}(w^{n-1/2} - \widetilde{w}_h^{n-1/2})\|_{L^2(\Omega)} \leq ch^2 \|\operatorname{div} w^{n-1/2}\|_{H^2(\mathcal{T}_h)}.$$

We then choose $q_h = \operatorname{div} \chi_w^{n-1/2}$ in (3.61) and obtain

$$\|\operatorname{div} \chi_w^{n-1/2}\|_{L^2(\Omega)} \leq \|R_g^{n-1/2}\|_{L^2(\Omega)} + \|d_\tau \chi_u^{n-1/2}\|_{H^1(\Omega)} + c_s \|d_\tau \chi_p^{n-1/2}\|_{L^2(\Omega)}.$$

Together with the previous estimates, this yields the second bound of the theorem. □

Concluding remark

The proposed fully discrete method is second-order convergent in both space and time. In particular, mass lumping does not lead to a reduction of the expected optimal convergence orders. Similar to the Darcy problem, see Section 2.3.4, not even super-convergence effects are lost, as we will see in the numerical tests below. The advantage of using mass lumping will be explained in a bit more detail in the following section.

3.5. Numerical experiments and discussion

Before concluding this chapter, we would like to illustrate our theoretical findings with some numerical tests and comment on possible generalizations of our approach.

3.5.1. Implementation

The algebraic form of (3.54)–(3.56) reads

$$\begin{aligned} \mathsf{A}\mathsf{u}^n \quad & -\mathsf{B}^\top \mathsf{p}^n & & = \mathsf{f}^n, \\ \mathsf{B} d_\tau \mathsf{u}^{n-1/2} & + \mathsf{C} d_\tau \mathsf{p}^{n-1/2} & + \mathsf{D}\mathsf{w}^{n-1/2} & = \widehat{\mathsf{g}}^{n-1/2}, \\ & -\mathsf{D}^\top \mathsf{p}^n & + \mathsf{M}_\mathsf{h} \mathsf{w}^n & = 0. \end{aligned}$$

As explained in Section 2.4, the mass matrix $\mathsf{M_h}$ can be made block-diagonal, if appropriate basis functions for the space Z_h are chosen. This allows for the efficient elimination of the seepage velocity w^n from the system and to obtain a reduced system

$$\begin{aligned} \mathsf{A}\mathsf{u}^n \qquad & -\mathsf{B}^\top \mathsf{p}^n && = \mathsf{f}^n, \\ \mathsf{B} d_\tau \mathsf{u}^{n-1/2} & + \mathsf{C} d_\tau \mathsf{p}^{n-1/2} + \mathsf{D}\mathsf{M}_\mathsf{h}^{-1}\mathsf{D}^\top \mathsf{p}^{n-1/2} && = \widehat{\mathsf{g}}^{n-1/2}, \end{aligned}$$

which can be interpreted as a non-conforming approximation of the two-field formulation (3.1)–(3.2). Like for the Darcy problem, the matrix $\mathsf{D}\mathsf{M}_\mathsf{h}^{-1}\mathsf{D}^\top$ can be interpreted as a non-conforming approximation for the differential operator $-\mathrm{div}(\kappa\nabla\cdot)$. By construction, the matrix $\mathsf{K_h}$ is sparse. Moreover, the seepage velocity can be recovered via $\mathsf{w}^n = \mathsf{M}_\mathsf{h}^{-1}\mathsf{D}^\top\mathsf{p}^n$, which only involves local operators. Solving for u^n and p^n leads to the implicit time-stepping scheme

$$\begin{pmatrix} \mathsf{A} & -\mathsf{B}^\top \\ \mathsf{B} & \mathsf{C} + \frac{\tau}{2}\mathsf{D}\mathsf{M}_\mathsf{h}^{-1}\mathsf{D}^\top \end{pmatrix} \begin{pmatrix} \mathsf{u}^n \\ \mathsf{p}^n \end{pmatrix} = \begin{pmatrix} \mathsf{A} & -\mathsf{B}^\top \\ \mathsf{B} & \mathsf{C} - \frac{\tau}{2}\mathsf{D}\mathsf{M}_\mathsf{h}^{-1}\mathsf{D}^\top \end{pmatrix} \begin{pmatrix} \mathsf{u}^{n-1} \\ \mathsf{p}^{n-1} \end{pmatrix} + \begin{pmatrix} \mathsf{f}^n - \mathsf{f}^{n-1} \\ \tau\,\widehat{\mathsf{g}}^{n-1/2} \end{pmatrix}, \tag{3.70}$$

which we also use for our numerical tests. Following [**117**], the system (3.70) can be solved efficiently by operator preconditioning in discrete H^1 norms. Robustness of such preconditioners w.r.t. the Lamé parameter λ and the time step τ have been investigated in more detail in [**67, 68**]. Let us note that the results of this paper can be transferred and substantially simplified after the elimination of the seepage velocity.

3.5.2. Numerical test

Our analysis was carried out for the case of homogeneous Dirichlet conditions for the elastic deformation u and the seepage velocity w. With minor modifications, our results can however be extended to a wide range of boundary conditions, as long as the resulting spaces satisfy both the inf-sup conditions (3.29) and (3.30), as well as a version of Korn's inequality that guarantees the ellipticity of the bilinear form $a(\cdot,\cdot)$.

Such a generalization is required for our test problem, which is similar to the one used in [**75, Section 6**] and in [**68**]. For our computations, we consider the three-field formulation (3.5)–(3.9) with constant parameters $\alpha = \mu = \lambda = \kappa = 1$ on the two-dimensional domain $\Omega = (0,1)^2$. We first define two auxiliary functions

$$\begin{aligned} \psi(t) &= \frac{1}{64\pi^4 - 4\pi^2}(8\pi^2 \sin(2\pi t) - 2\pi\cos(2\pi t) + 2\pi e^{-8\pi^2 t}) \qquad \text{and} \\ \phi(x,y) &= \sin(2\pi x)\sin(2\pi y). \end{aligned}$$

Then the derived functions

$$u(x,y,t) = \frac{\psi(t)}{8\pi^2}\nabla\phi(x,y), \qquad p(x,y,t) = \psi(t)\phi(x,y), \qquad w(x,y,t) = -\psi(t)\nabla\phi(x,y)$$

can be seen to satisfy (3.5)–(3.7) with right hand sides

$$f(x,y,t) = (8\pi^2\psi(t) - \partial_t\psi(t))\phi(x,y) \qquad \text{and} \qquad g(x,y,t) = 4\psi(t)\nabla\phi(x,y).$$

Furthermore, the solution satisfies the boundary conditions

$$n \times u = 0, \qquad \partial_n(u\cdot n) = 0, \qquad p = 0 \qquad \text{on } \partial\Omega.$$

As initial conditions, we choose the evaluation of the true solutions at $t = 0$.

Simulation setup

We set the time horizon to $T = 1$ and define the norm $|||e|||_X := \|e(t^n)\|_{\ell^\infty(I_\tau,X)}$. In our simulation, we use $\tau = 0.5h$ on each discretization level. By doing this, we test the convergence in both h and τ simultaneously. In Table 3.1, we depict the errors covered by the theoretical results of Theorem 3.4.6.

h	$\|\|\|\widetilde{u}_h - u_h\|\|\|_{H^1(\Omega)}$	eoc	$\|\|\|\widetilde{p}_h - p_h\|\|\|_{L^2(\Omega)}$	eoc	$\|\|\|\widetilde{w}_h - w_h\|\|\|_{L^2(\Omega)}$	eoc
2^{-2}	0.000306	—	0.000314	—	0.001819	—
2^{-3}	0.000086	1.88	0.000079	2.00	0.000456	2.00
2^{-4}	0.000019	2.16	0.000020	2.00	0.000113	2.01
2^{-5}	0.000004	2.21	0.000005	2.00	0.000028	2.01

Table 3.1.: Discrete errors and estimated orders of convergence (eoc) for the second order method with mass lumping, part one.

As expected, we see second-order convergence in all solution components. In Table 3.2, we depict further error quantities, for which one might expect improved convergence rates.

h	$\|\|\|\widetilde{u}_h - u_h\|\|\|_{L^2(\Omega)}$	eoc	$\|\|\|\pi_h^0(p - p_h)\|\|\|_{L^2(\Omega)}$	eoc
2^{-2}	0.001197	—	0.005117	—
2^{-3}	0.000139	3.11	0.000627	3.03
2^{-4}	0.000014	3.25	0.000077	3.02
2^{-5}	0.000001	3.22	0.000010	3.01

Table 3.2.: Discrete errors (scaled by a factor of 100) and estimated orders of convergence (eoc) for the second order method with mass lumping, part two.

The results clearly indicate that super-convergence in the pressure averages and optimal convergence for the L^2-norm of the displacement can be expected. Both results can probably be proven rigorously using duality arguments; compare with the proof of Theorem 2.3.4. Let us note that optimal convergence in the L^2-norm of the displacement has also been reported in [**75**] for a slightly different discretization scheme.

3.5.3. Discussion

In this chapter, we extended the second order multipoint flux method for the Darcy problem, discussed in the previous chapter, to poroelasticity. We presented a second-order accurate finite element discretization with mass lumping for the three-field formulation and explained its possible reduction to a two-field formulation on the algebraic level. This method is conservative and second-order accurate in all three variables. In addition, we analyzed an implicit time-stepping scheme and proved second-order convergence in time.

With minor modifications, our mass lumping strategy and the corresponding analysis can also be extended to the methods discussed in [**67, 75**], which are based on a non-conforming approximation of the elastic displacement. In this case, we would choose the

local finite element spaces as

$$V(K) = \mathrm{BDM}_2(K),$$
$$Q(K) = P_1(K), \qquad \text{and} \qquad Z(K) = P_1(K)^d \oplus \boldsymbol{x} \cdot P_1^h(K) \coloneqq \mathrm{RT}_1(K).$$

For a graphical representation of the degrees of freedom, we refer to Figure 3.2. The

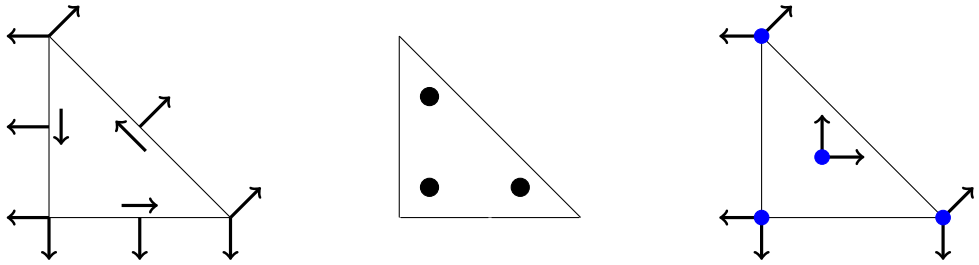

Figure 3.2.: Depiction of the degrees of freedom for the triplet $\mathrm{BDM}_2 - \mathrm{P}_1 - \mathrm{RT}_1$ on the reference triangle. The blue dots represent the quadrature points used for mass lumping.

global spaces Q_h and Z_h are then defined as in (3.28). For the displacement, we now have

$$V_h = \{v_h \in H_0(\mathrm{div}, \Omega) \quad : \quad v_h|_K \in V(K)\} \not\subseteq H_0^1(\Omega).$$

The bilinear form $a(\cdot, \cdot)$ is then replaced by a stabilized variant $a_h(\cdot, \cdot)$ to ensure stability of the method; see [**67, 75**] for details. We implemented this method and repeated our numerical tests. The corresponding results are listed in Table 3.3. Here the projections

h	$\|\|\pi_h^2 u - u_h\|\|_{H^1(\Omega)}$	eoc	$\|\|\pi_h^1 p - p_h\|\|_{L^2(\Omega)}$	eoc	$\|\|\Pi_h w - w_h\|\|_{L^2(\Omega)}$	eoc
2^{-1}	0.000315	—	0.000315	—	0.001819	—
2^{-2}	0.000090	1.80	0.000079	1.99	0.000456	1.99
2^{-3}	0.000023	1.95	0.000020	2.00	0.000113	2.01
2^{-4}	0.000006	1.98	0.000005	2.00	0.000028	2.01

Table 3.3.: Discrete errors and estimated orders of convergence (eoc) for the second order fully conservative method with mass lumping.

π_h^1 and π_h^2 are L^2-orthogonal projections on the spaces Q_h and V_h, respectively, while Π_h is the standard interpolation operator onto RT_1, see Lemma 3.2.1. As expected, we again obtain second-order convergence in all solution components.

The multipoint flux method can be employed to solve other problems where the inversion of the $H(\mathrm{div})$ mass matrix is relevant. One example that requires further research is the study of mixed finite element methods for linear elasticity with weakly imposed symmetry; see [**10**]. In this formulation, the symmetry of the $H(\mathrm{div})$ matrix field is realized with the help of a Lagrange function. Quite recently, in [**6, 5**], the authors managed to use an extended version of the first order multipoint flux method from [**112**] and showed

linear convergence in all variables. As a result of the mass lumping, the algebraic structure reduces from a three variable saddle point problem to a symmetric positive definite system for the pressure only, while also benefiting from the advantages that come from the mixed formulation. The extension of the approach to a second-order formulation with mass lumping is an open topic. In our attempts, we managed to successfully eliminate the $H(\mathrm{div})$ variables, i.e., the stresses, which results in a symmetric positive definite system for the pressure and the Lagrange multiplier. Further investigations in this direction would certainly be of interest.

In the next two chapters, we turn to the application of mass lumping for the efficient simulation of time dependent wave propagation problems.

4. Mass lumping for the wave equation in H(div)

Problems involving the propagation of waves have been a topic of continuing interest in science and engineering. A few of the many areas of application are radar design, the interaction of electromagnetic fields with biological tissues, weather prediction, wireless communications, and even earthquake prediction. In this chapter, we study numerical approximations based on $H(\mathrm{div})$ finite elements and their efficient realization via mass lumping and explicit time-stepping methods. As a simple model problem, we consider acoustic wave propagation, modeled by

$$\rho_0\, \partial_t u(t) + \nabla p(t) = f(t) \qquad \text{in } \Omega,\ t > 0, \tag{4.1}$$

$$\frac{1}{c^2}\partial_t p(t) + \rho_0 \operatorname{div} u(t) = 0 \qquad \text{in } \Omega,\ t > 0. \tag{4.2}$$

Here u and p are the pressure and velocity fields, ρ_0 is the density of the medium, c denotes the speed of sound, and f is a source term. The two equations describe the conservation of mass and momentum, and they can be derived by linearization of the Euler equations; see [**74, Chapter 5**]. For simplicity, we set $\rho_0 = 1$ and $c = 1$ from here on.

A standard and widely used scheme for discretizing (4.1)–(4.2) is the *finite difference time domain* (FDTD) method, first introduced in [**114**] for approximating Maxwell's equations. For a thorough analysis in the context of acoustic wave propagation, we refer to [**35, Part II.**] and [**105**]. The method is based on staggered grids in space and time and leads to efficient and explicit time-stepping schemes. For constant, isotropic coefficients and orthogonal grids, one can show second-order accuracy in both space and time. Generalizations to more complex geometries, unstructured grids, as well as anisotropic coefficients are however nontrivial and often lead to loss in accuracy [**21**].

A more flexible and reliable discretization framework is offered by finite element methods, which come in two flavors: In the primal formulation, one approximates $p \in H^1(\Omega)$ and $u \in L^2(\Omega)^d$. For this choice, the velocity u can be eliminated from (4.1)–(4.2), on the continuous and discrete level, which leads to the second order wave equation

$$\partial_{tt} p(t) - \Delta p(t) = f(t) \qquad \text{in } \Omega,\ t > 0. \tag{4.3}$$

In [**35, 60**], the authors proposed mass lumping techniques for the finite element approximation of (4.3) which, together with appropriate time-marching schemes, lead to highly efficient and flexible methods that can be also used on unstructured grids and with anisotropic coefficients.

If, on the other hand, the velocity u is of main interest, then one can instead consider a dual formulation, for which approximations are sought for $p \in L^2(\Omega)$ and $u \in H(\mathrm{div}, \Omega)$.

In this case, one can locally eliminate the pressure p from (4.1)–(4.2), again on the continuous and discrete level, which yields a second-order equation

$$\partial_{tt} u(t) - \nabla \mathrm{div}\, u(t) = \partial_t f(t) \qquad \text{in } \Omega,\ t > 0 \tag{4.4}$$

involving only the velocity variable. Related dual finite element approximations for the acoustic wave equation have been investigated in [**61**] for the equivalent mixed form (4.1)–(4.2); also see [**41, 42, 72**] for later results. In [**79**], the method has been investigated in the context of the elastic wave equation. Clearly, an important ingredient for the efficient time-domain simulation here is the availability of an appropriate mass lumping strategy for $H(\mathrm{div})$. Corresponding techniques can be found in [**35, 38**] for quadrilateral and hexahedral elements, as well as for lowest order approximations on simplicial elements.

Another discretization approach that allows for efficient time integration is offered by *discontinuous Galerkin* (dG) schemes; see [**99, 65, 66, 102**]. Due to the discontinuity of the basis functions, these methods automatically lead to diagonal or block-diagonal mass matrices. On the other hand, this leads to a substantial increase in the number of degrees of freedom, which certainly is a downside of this approach. In [**60**], a comparison between finite-element approximations with mass lumping and discontinuous Galerkin methods is given for wave propagation problems in H^1. The conclusion is that, at least for moderate approximations orders, finite-element methods with mass lumping are more efficient.

In this chapter, we study the approximation of (4.4) by $H(\mathrm{div})$-conforming finite elements with mass lumping. In view of the above considerations, we focus on first and second-order approximations. Our first method is based on the mass lumping strategy of [**112**] and its convergence analysis is closely related to our results in [**47**] for the discretization of (4.1)–(4.2). Here we stay with the second order form of the problem and develop the corresponding analysis. Based on these results, we propose another first-order method, which requires fewer degrees of freedom while providing the same approximation properties. This method can be interpreted as a natural generalization of the FDTD method to unstructured grids. On regular quadrilateral meshes, the scheme reduces to the FDTD method, and second-order accuracy is attained; see [**49, Lemma 7**]. This method also seems closely related to the recent work in [**96**], where a sparse representation of the inverse mass matrix is constructed by geometric considerations. In comparison to that, our method has the advantage of being backed up by the finite element framework, which allows us to conduct a full error analysis. By utilizing the mass lumping strategy presented in Section 2.3, we also derive a method with second-order accuracy. For this method, we also discuss in detail the time-discretization and present a full convergence analysis establishing second-order convergence in space and time.

Outline

Before going into the details, let us briefly review the contents of this chapter.

Section 4.1. *Preliminaries:* We introduce the mathematical problem in a rigorous way and give a brief overview regarding the existence and uniqueness of solutions, as well as conforming Galerkin discretizations in $H(\mathrm{div})$.

Section 4.2. *A first-order method:* Similar to the Darcy problem, we start by introducing the first-order method with mass lumping, for which we provide a complete error analysis. Again, convergence in the velocity is limited here by numerical integration errors.

Section 4.3. *A first-order variational Yee-like method:* We then propose another first-order method that requires fewer degrees of freedom, which can be understood as a generalization of the finite difference time domain method to unstructured grids. The close relation of BDM_1 and RT_0 methods is used for the construction and analysis of this method.

Section 4.4. *A second-order method:* We then introduce the second-order scheme with mass lumping and again provide a full error analysis. This method yields optimal error estimates for the involved approximation spaces and mass lumping does not affect accuracy here.

Section 4.5. *Time discretization:* For time integration, we consider an explicit time stepping scheme, which can fully exploit the advantages brought by mass lumping. Second-order convergence in time is again proven rigorously.

Section 4.6. *Numerical validation*: We report on numerical tests, in which the theoretical results are illustrated and a comparison between the proposed methods is established.

Section 4.7. *Discussion*: We close with a short summary of our results and highlight some directions for possible extensions and future research.

4.1. Preliminaries

Throughout this chapter, we assume that

(A1) $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a bounded Lipschitz domain and $T > 0$ is the time horizon.

We consider the second-order wave equation

$$\partial_{tt} u(t) - \nabla \mathrm{div}\, u(t) = f(t) \qquad \text{in } \Omega \times (0, T), \tag{4.5}$$

which results from setting $c = \rho_0 = 1$ in (4.1)–(4.2) and eliminating p. For ease of presentation, we consider homogeneous Neumann boundary conditions

$$n \cdot u(t) = 0 \qquad \text{on } \partial\Omega \times (0, T), \tag{4.6}$$

and we prescribe initial values by

$$u(0) = u_0, \quad \partial_t u(0) = v_0 \qquad \text{in } \Omega. \tag{4.7}$$

More general problems, including non-constant and matrix-valued coefficients or damping terms as well as mixed and non-homogeneous boundary conditions, could be considered with minor modification of our arguments.

We denote by $L^p(\Omega)$ and $W^{k,p}(\Omega)$ the standard Lebesgue and Sobolev spaces and abbreviate $H^k(\Omega) = W^{k,2}(\Omega)$. Of particular interest for our analysis is the space

$$H_0(\mathrm{div}, \Omega) = \{u \in L^2(\Omega)^d \;\mid\; \mathrm{div}\, u \in L^2(\Omega),\; n \cdot u = 0 \text{ on } \partial\Omega\}$$

in dimension $d = 2, 3$ with norm $\|u\|_{H(\mathrm{div},\Omega)} = (\|u\|^2_{L^2(\Omega)} + \|\mathrm{div}\, u\|^2_{L^2(\Omega)})^{1/2}$.

4.1.1. Well-posedness

Let us briefly comment on the well-posedness of our model problem. Any smooth solution u of (4.5)–(4.7) also satisfies the variational identity

$$(\partial_{tt} u(t), v) + (\mathrm{div}\, u(t), \mathrm{div}\, v) = (f(t), v) \qquad \forall v \in H_0(\mathrm{div}, \Omega),\;\; t > 0. \tag{4.8}$$

This follows directly by multiplying (4.5) with a test function, integrating over the domain, using integration-by-parts, as well as the boundary conditions. Based on this weak form of the equations, we can deduce the following result.

Lemma 4.1.1 (Existence of weak solutions)**.**
Let (A1) hold. Then for any $u_0 \in H_0(\mathrm{div}, \Omega)$, $v_0 \in L^2(\Omega)$ and $f \in L^2(0, T, L^2(\Omega))$, the system (4.5)–(4.7) admits a unique weak solution

$$u \in L^2(0, T, H_0(\mathrm{div}, \Omega)) \cap H^1(0, T, L^2(\Omega)) \cap H^2(0, T, (H_0(\mathrm{div}, \Omega))')$$

depending continuously on the problem data.

The assertion is proven by Galerkin approximation; see [**113, Section 26**] for details. In the next sections, we consider discretization techniques based on the weak form (4.8).

4.1.2. Finite element approximations

Let us briefly introduce our notation and, for later comparison, recall some available results from the literature. We denote by $\mathcal{T}_h = \{K\}$ a regular partition of the domain Ω in the sense of [**32**]. We write h_K and ρ_K for the diameter and inner circle radius of the element K, and call $h = \max_K h_K$ the mesh size. We always assume that

(A2) $\mathcal{T}_h$ is a γ-shape regular simplicial mesh, i.e. there exists a uniform constant $\gamma > 0$ such that $\gamma h_K \le \rho_K \le h_K$ for all $K \in \mathcal{T}_h$.

We will use broken Sobolev spaces $H^k(\mathcal{T}_h) = \{q \in L^2(\Omega) \,:\, q|_K \in H^k(K), \quad \forall K \in \mathcal{T}_h\}$ with norm $\|v\|^2_{H^k(\mathcal{T}_h)} = \sum_{K\in\mathcal{T}_h} \|u\|^2_{H^k(K)}$. By $P_k(K)$ and $P^h_k(K)$, we denote the space of polynomials of degree at most or exactly k. Let us recall the definitions

$$\mathrm{BDM}_{k+1}(K) = P_{k+1}(K)^d \qquad \text{or} \tag{4.9}$$

$$\mathrm{RT}_k(K) = P_k(K)^d \oplus \{x \cdot P^h_k(K)\}, \tag{4.10}$$

which are standard local finite element spaces for $H(\mathrm{div})$; see [**19**] for details. As global approximation spaces for the velocity u in $H_0(\mathrm{div}, \Omega)$, we consider

$$V_h = \{v_h \in H_0(\mathrm{div}, \Omega) \,:\, v_h|_K \in \mathcal{M}_k(K), \quad \forall K \in \mathcal{T}_h\},$$

where either $\mathcal{M}_k(K) = \mathrm{BDM}_{k+1}(K)$ or $\mathcal{M}_k(K) = \mathrm{RT}_k(K)$. By Π_h we again denote the canonical interpolation operator for this space; see Section 2.3.1 for an overview of its properties.

As a benchmark, we consider standard finite element approximations of our problem.

Problem 4.1.2 (Exact Galerkin approximation)**.**
Set $u_h(0) = \Pi_h u_0$, $\partial_t u_h(0) = \Pi_h v_0$, and then find $u_h \in C^2([0,T], V_h)$ such that

$$(\partial_{tt} u_h(t), v_h) + (\operatorname{div} u_h(t), \operatorname{div} v_h) = (f(t), v_h) \qquad \forall v_h \in V_h, \;\; t > 0. \tag{4.11}$$

Let us briefly comment on the well-posedness of this method: The realization of (4.11) leads to an ordinary differential equation $\mathsf{M}\partial_{tt}\mathsf{u} + \mathsf{K}\mathsf{u} = \mathsf{f}$. By introducing $\mathsf{v} \coloneqq \partial_t \mathsf{u}$, this can be written equivalently as

$$\begin{pmatrix} \mathsf{Id} & 0 \\ 0 & \mathsf{M} \end{pmatrix} \begin{pmatrix} \partial_t \mathsf{u} \\ \partial_t \mathsf{v} \end{pmatrix} + \begin{pmatrix} 0 & -\mathsf{Id} \\ \mathsf{K} & 0 \end{pmatrix} \begin{pmatrix} \mathsf{u} \\ \mathsf{v} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathsf{f} \end{pmatrix}.$$

For any choice of initial values, the existence of a unique solution then follows from the Picard-Lindelöff theorem. In addition, one obtains the following convergence results.

Lemma 4.1.3 (Exact Galerkin discretization)**.**
Let (A1) and (A2) hold and let u be a sufficiently smooth solution of (4.5)–(4.7). Further, let u_h be the corresponding solution of Problem 4.1.2. Then

$$\|\partial_t(u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} + h\|\operatorname{div}(u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} \leq C_1(u) h^{k+1}.$$

If $\mathcal{M}_k(K) = RT_k(K)$, then also $\|\operatorname{div}(u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} \leq C_2(u) h^{k+1}$.

For details and proofs, we refer to [**61**] and, in particular, to [**46, Theorem 3.6**]. Related results for problems in H^1 and $H(\mathrm{curl})$ can be found in [**45**] and [**82**]. As mentioned in the introduction, the algebraic realization of the standard finite element method involves mass matrices that are not easily invertible. This hinders the efficient application of explicit time-stepping schemes. In the following sections, we therefore propose and analyze first and second-order mass lumping schemes that overcome this inefficiency.

4.2. A first-order method

Inspired by the mixed finite element approximation for Darcy flow in [**112**], we choose

$$V_h = \{v_h \in H_0(\operatorname{div}, \Omega) \;:\; v_h|_K \in \mathrm{BDM}_1(K), \quad \forall K \in \mathcal{T}_h\} \tag{4.12}$$

and denote by Π_h the canonical interpolation operator for V_h; see Section 2.3.1 for its approximation properties. We then introduce the quadrature formula

$$Q_K(f) \coloneqq \frac{|K|}{d+1} \sum_{k=1}^{d+1} f(v_k^K), \tag{4.13}$$

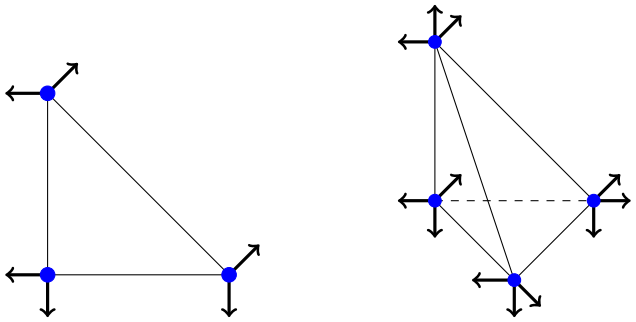

Figure 4.1.: Degrees of freedom for the space BDM_1 on the unit triangle (left) and the unit tetrahedron (right), as well as the quadrature points of (4.13).

with v_k^K denoting the k-th vertex of K. Let us note that exactly d degrees of freedom can be associated with every quadrature point; see Figure 4.1 for a sketch. Just as in Chapter 2, we further introduce the global inexact scalar product

$$(u, w)_h \coloneqq \sum\nolimits_{K \in \mathcal{T}_h} Q_K(u \cdot w). \tag{4.14}$$

These definitions now lead to the following inexact Galerkin approximation.

Problem 4.2.1 (First order method)**.**
For $u_h(0) = \Pi_h u(0)$ and $\partial_t u_h(0) = \Pi_h \partial_t u(0)$ find $u_h \in C^2([0,T], V_h)$ such that

$$(\partial_{tt} u_h(t), v_h)_h + (\operatorname{div} u_h(t), \operatorname{div} v_h) = (f(t), v_h) \qquad \forall v_h \in V_h, \quad t > 0. \tag{4.15}$$

Let us recall from [**112**] that

$$c_1(v_h, v_h) \le (v_h, v_h)_h \le c_2(v_h, v_h) \qquad \forall v_h \in V_h, \tag{4.16}$$

with some positive constants c_1, c_2 independent of the mesh size. As a consequence, the mass matrix $\mathsf{M_h}$ in the equivalent algebraic system $\mathsf{M_h}\partial_{tt}\mathsf{u} + \mathsf{Ku} = \mathsf{f}$ is regular, which implies the well-posedness of the discretization scheme, using the same arguments as for the exact Galerkin approximation. The following result is similar to [**47, Theorem 3.8**], where we analyzed the corresponding discretization of the first order system (4.1)-(4.2).

Theorem 4.2.2 (First order convergence)**.**
Let (A1) and (A2) hold, let u be a sufficiently smooth solution of (4.5)–(4.7), and let u_h be the corresponding solution of Problem 4.2.1. Then

$$\|\partial_t(u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} + \|\operatorname{div}(u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} \le Ch\|u\|_{W^{2,1}(0,T;H^1(\Omega))},$$

with a constant C that is independent of the mesh size.

Proof. Let

$$\sigma_K(u, v) \coloneqq Q_K(u \cdot v) - (u, v)_K \qquad \text{and} \qquad \sigma_h(u, v) = \sum\nolimits_{K \in T_h} \sigma_K(u, v) \tag{4.17}$$

define the local and global quadrature errors. We begin by splitting the error in interpolation and discrete error components in the following way:

$$u(t) - u_h(t) = (u(t) - \Pi_h u(t)) + (\Pi_h u(t) - u_h(t)) = -\eta(t) + \psi_h(t). \tag{4.18}$$

From (2.28), we derive the interpolation error estimate

$$\|\partial_t \eta(t)\|_{L^2(\Omega)} + \|\operatorname{div} \eta(t)\|_{L^2(\Omega)} \leq ch\|u(t)\|_{H^2(\mathcal{T}_h)}.$$

Let π_h^1 denote the L^2-orthogonal projection onto $P_1(\mathcal{T}_h)$ as defined in (2.27). Then the discrete error $\psi_h(t)$ satisfies

$$\begin{aligned}(\partial_{tt}\psi_h(t), v_h)_h &+ (\operatorname{div} \psi_h(t), \operatorname{div} v_h) \\ &= (\partial_{tt}\eta(t), v_h) + (\operatorname{div} \eta(t), \operatorname{div} v_h) + \sigma_h(\Pi_h \partial_{tt} u(t), v_h) \\ &= (\partial_{tt}\eta(t), v_h) + \sigma_h(\Pi_h \partial_{tt} u(t), v_h).\end{aligned} \tag{4.19}$$

Here we simply subtracted (4.15) from (4.8) with $v = v_h$ and added the corresponding terms with $\Pi_h u(t)$ on both sides. Note that the second term in the first line vanishes due to the commuting diagram property (2.30). We then note that

$$(\partial_{tt}\psi_h(t), \partial_t \psi_h(t)) + (\operatorname{div} \psi_h(t), \partial_t \operatorname{div} \psi_h(t)) = \frac{1}{2}\frac{d}{dt}\left(\|\partial_t \psi_h(t)\|_h^2 + \|\operatorname{div} \psi_h(t)\|_{L^2(\Omega)}^2\right),$$

and continue choosing $v_h \coloneqq \partial_t \psi_h(t)$ in (4.19), integrating from 0 to t, and using the definition of the initial values $\psi_h(0) = \partial_t \psi_h(0) = 0$. This leads to

$$\begin{aligned}\frac{1}{2}\left(\|\partial_t \psi_h(t)\|_h^2 + \|\operatorname{div} \psi_h(t)\|_{L^2(\Omega)}^2\right) &= \int_0^t (\partial_{tt}\eta(s), \partial_t \psi_h(s)) + \sigma_h(\pi_h^1 \partial_{tt} u(s), \partial_t \psi_h(s)) \\ &= (i) + (ii).\end{aligned}$$

The first term can be estimated via Cauchy-Schwarz and Young's inequality by

$$\begin{aligned}(i) &\leq \int_0^t \|\partial_{tt}\eta(s)\|_{L^2(\Omega)} \|\partial_t \psi_h(s)\|_{L^2(\Omega)}\, ds \leq \|\partial_{tt}\eta\|_{L^1(0,t,L^2(\Omega))}^2 + \frac{1}{4c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2 \\ &\leq Ch^2 \|\partial_{tt} u\|_{L^1(0,t,H^1(\mathcal{T}_h))}^2 + \frac{1}{4c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.\end{aligned}$$

For the second term, we use the exactness of the quadrature rule for piecewise linear functions and conclude that

$$\begin{aligned}(ii) &= \int_0^t \sigma_h(\Pi_h \partial_{tt} u(s), \partial_t \psi_h(s))\, ds = \int_0^t \sigma_h(\Pi_h \partial_{tt} u(s) - \pi_h^0 \partial_{tt} u(s), \partial_t \psi_h(s))\, ds \\ &\leq \int_0^t \|\Pi_h \partial_{tt} u(s) - \pi_h^0 \partial_{tt} u(s)\|_{L^2(\Omega)} \|\partial_t \psi_h(s)\|_{L^2(\Omega)}\, ds \\ &\leq \|\Pi_h \partial_{tt} u - \pi_h^0 \partial_{tt} u\|_{L^1(0,t,L^2(\Omega))}^2 + \frac{1}{4c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2 \\ &\leq ch^2 \|\partial_{tt} u\|_{L^1(0,t,H^1(\mathcal{T}_h))}^2 + \frac{1}{4c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.\end{aligned}$$

We refer to Section 2.3.1 for the interpolation error estimates used in the above derivations. Summing all the terms and using the norm equivalence from (4.16), we obtain

$$\frac{1}{c_1}\|\partial_t\psi_h(t)\|^2_{L^2(\Omega)} + \|\operatorname{div}\psi_h(t)\|^2_{L^2(\Omega)} \leq ch^2\|\partial_{tt}u\|^2_{L^1(0,t,H^1(\mathcal{T}_h))} + \frac{1}{2c_1}\|\partial_t\psi_h\|^2_{L^\infty(0,t,L^2(\Omega))}.$$

Taking the maximum over all t and subsequently absorbing the last term by the left-hand side yields the L^∞-estimate. The assertion of the theorem now follows by combining the two bounds for the discrete and the interpolation error. □

Remark 4.2.3. By comparing this result to Lemma 4.1.3, we see that one order of accuracy in the term $\|\partial_t(u-u_h)\|_{L^2(\Omega)}$ is lost. This is not an artifact of our analysis, but can also be observed in our numerical tests. Let us note that, just as in [**47, Theorem 4.5**] for the equivalent first order system (4.1)–(4.2), one can design a global post-processing scheme in space that generates a second order accurate approximation $\widetilde{u}_h \in V_h$. Thus, the second-order convergence lost by mass lumping can be restored, in principle. While the post-processing is local in time and can therefore be conducted somehow independently of the time stepping, it seems unlikely that the localization in spaces is possible. In [**47**], we also established super-convergence for the pressure averages for the corresponding approximation of (4.1)–(4.2), and we were able to propose a post-processing scheme, which is similar to what we proposed in Section 2.3.6 in the context of porous media flow.

4.3. A first-order variational Yee-like method

For ease of notation, we set $f = 0$ in this section. Lemma 4.1.3 indicates that the convergence rates of Theorem 4.2.2 should already be achievable by approximation in the smaller finite element space

$$\widetilde{V}_h = \{\widetilde{v}_h \in H_0(\operatorname{div},\Omega) \ : \ \widetilde{v}_h|_K \in \mathrm{RT}_0(K), \quad \forall K \in \mathcal{T}_h\}. \tag{4.20}$$

Unfortunately, no general mass lumping technique is available for $\widetilde{V}_h$ on simplices, although some attempts exist. In [**13**], the author established a connection between the finite volume method and the lowest order mixed method based on a discretization for RT_0 in 2D. The author gave an explicit formula for the construction of the lumped mass matrix, which has only diagonal entries. He concluded that for his construction, the quadrature error decays linearly in the mesh size h whenever the mesh is acute, meaning no angles measure more than $\pi/2$. Later, in [**20, 78**], this condition was relaxed to allow for Delaunay meshes. In [**78**], the author extended this approach to tetrahedral grids, which proved to be even more restrictive. The method has also been used in the context of Maxwell's equations; see e.g. [**20**]. By using the results of the previous section, we will show that it is possible to achieve mass lumping for $\widetilde{V}_h$ on general simplicial meshes. Our approach is based on the close relationship between the spaces $\mathrm{RT}_0(K)$ and $\mathrm{BDM}_1(K)$ and the solutions obtained by the corresponding approximation methods.

4.3.1. Preliminaries

The space $\widetilde{V}_h$ defined above is naturally embedded in the larger space

$$V_h = \{v_h \in H_0(\mathrm{div}, \Omega) \ : \ v_h|_K \in \mathrm{BDM}_1(K), \quad \forall K \in \mathcal{T}_h\} \tag{4.21}$$

which was employed in the previous section. Let us denote by $E_h : \widetilde{V}_h \to V_h$ the natural embedding and by $\widetilde{\Pi}_h : V_h \to \widetilde{V}_h$ the restriction of the canonical interpolation operator for $\widetilde{V}_h$. It is not difficult to verify that $\widetilde{\Pi}_h \circ E_h = id_{\widetilde{V}_h}$. From the local definition of the embedding and the projection, it is clear that these relations also hold on the element level.

Now let $\{\Phi_k\} \subset V_h$ denote the basis constructed in Section 2.4, which was proven to be suitable for mass lumping. Further, let P denote the matrix representation of the projection $\widetilde{\Pi}_h$. In Section 4.3.4, we define a corresponding basis $\{\widetilde{\Phi}_k\} \subset \widetilde{V}_h$, again constructed locally, with the following properties:

(P1) $\mathsf{P}^\top$ is the matrix representation of the embedding E_h and $\mathsf{P}\mathsf{P}^\top = id$.

(P2) Let n be the number of edges ($d = 2$) or the number of faces ($d = 3$). Then $\mathsf{P} \in \mathbb{R}^{n \times d \cdot n}$ is block diagonal with $1 \times d$ blocks.

These properties immediately follow from the explicit construction in Section 4.3.4, and we take them for granted in the sequel.

4.3.2. Construction on the algebraic level

As outlined in the previous section, the algebraic form of the first order approximation based on V_h and stated in Problem 4.2.1 is given by

$$\mathsf{M_h}\partial_{tt}\mathsf{u}(t) + \mathsf{K}\mathsf{u}(t) = 0, \tag{4.22}$$

with block-diagonal lumped mass matrix $\mathsf{M_h}$. We write M for the exact mass matrix in the sequel. From the definition of P, we immediately deduce that

$$\widetilde{\mathsf{M}} = \mathsf{P}\mathsf{M}\mathsf{P}^\top \qquad \text{and} \qquad \widetilde{\mathsf{K}} = \mathsf{P}\mathsf{K}\mathsf{P}^\top,$$

which are the exact mass and stiffness matrices for the Galerkin approximation in $\widetilde{V}_h$. Furthermore, we have

$$\mathsf{K} = \mathsf{P}^\top\widetilde{\mathsf{K}}\mathsf{P}, \tag{4.23}$$

which can be proven by the commuting diagram property (2.30) of the projection operators and the specific properties of the finite element spaces. More precisely, we have

$$\mathrm{div}\, v_h = \pi_h^0 \mathrm{div}\, v_h = \mathrm{div}\, \widetilde{\Pi}_h v_h \qquad \forall v_h \in V_h,$$

which implies $\mathsf{K}_{ij} = (\mathrm{div}\, \Phi_j, \mathrm{div}\, \Phi_i) = (\mathrm{div}\, \widetilde{\Pi}_h \Phi_j, \mathrm{div}\, \widetilde{\Pi}_h \Phi_i) = (\mathsf{P}^\top\widetilde{\mathsf{K}}\mathsf{P})_{ij}$. With this final observation, we can now introduce our new scheme.

Lemma 4.3.1 (Yee-like scheme)**.**
Let (A1)–(A2) and (P1)–(P2) hold and let u *denote the solution to* (4.22)*. Then* $\widetilde{\mathsf{u}}(t) = \mathsf{P}\mathsf{u}(t)$ *solves the algebraic system*

$$\widetilde{\mathsf{M}}_\mathsf{h}\partial_{tt}\widetilde{\mathsf{u}}(t) + \widetilde{\mathsf{K}}\widetilde{\mathsf{u}}(t) = 0, \tag{4.24}$$

with modified mass matrix $\widetilde{\mathsf{M}}_\mathsf{h} \coloneqq (\mathsf{P}\mathsf{M}_\mathsf{h}^{-1}\mathsf{P}^\top)^{-1}$*.*

Proof. Using the properties stated above, we compute

$$\begin{aligned}\partial_{tt}\widetilde{\mathsf{u}}(t) &= \mathsf{P}\partial_{tt}\mathsf{u}(t) = -\mathsf{P}\mathsf{M}_\mathsf{h}^{-1}\mathsf{K}\mathsf{u}(t) = -\mathsf{P}\mathsf{M}_\mathsf{h}^{-1}\mathsf{P}^\top\widetilde{\mathsf{K}}\mathsf{P}\mathsf{u}(t)\\ &= -(\mathsf{P}\mathsf{M}_\mathsf{h}^{-1}\mathsf{P}^\top)\widetilde{\mathsf{K}}\widetilde{\mathsf{u}}(t),\end{aligned}$$

where we used the system (4.22) and (4.23). This already concludes the proof. □

Remark 4.3.2. The system (4.24) can be understood as a projected version of (4.22). Also, consider the system

$$\widetilde{\mathsf{M}}\partial_{tt}\widehat{\mathsf{u}}(t) + \widetilde{\mathsf{K}}\widehat{\mathsf{u}}(t) = 0,$$

which represents the algebraic realization of the exact Galerkin approximation in $\widetilde{V}_h$. By comparison, we see that (4.24) has the same stiffness matrix, but a different representation of the mass matrix. In contrast to $\widetilde{\mathsf{M}}$, we have that $\widetilde{\mathsf{M}}_\mathsf{h}$ is full, in general, while its inverse $\widetilde{\mathsf{M}}_\mathsf{h}^{-1}$ is sparse, if P is sparse. For this reason, (4.24) can not be understood as a realization of a mixed finite element method, but rather just as an algebraic trick to reduce the number of degrees of freedom of (4.22).

4.3.3. Interpretation in function space

The solution $\widetilde{\mathsf{u}}(t)$ from (4.24) can be translated back to an element $\widetilde{u}_h \in \widetilde{V}_h$ by assigning

$$\widetilde{u}_h(t) = \sum\nolimits_k \widetilde{\mathsf{u}}_k(t)\widetilde{\Phi}_k.$$

We now show that the method is indeed first-order convergent and that the convergence result essentially follows from Theorem 4.2.2.

Theorem 4.3.3. *Let (A1)–(A2) and (P1)–(P2) hold. Further let* u *be a sufficiently smooth solution of* (4.5)–(4.7) *and* $\widetilde{u}_h(t) \in \widetilde{V}_h$ *denote the finite element solution associated to* $\widetilde{\mathsf{u}}(t)$ *constructed according to* (4.24)*. Then*

$$\|\partial_t(u - \widetilde{u}_h)\|_{L^\infty(0,T,L^2(\Omega))} + \|\mathrm{div}(u - \widetilde{u}_h)\|_{L^\infty(0,T,L^2(\Omega))} \leq Ch\|u\|_{W^{2,1}(0,T,H^1(\Omega))},$$

with constant C *independent of the mesh size.*

Proof. We proceed in a similar fashion as in the last section. We again split the error in an interpolation and a discrete error component

$$u(t) - \widetilde{u}_h(t) = (u(t) - \widetilde{\Pi}_h u(t)) + (\widetilde{\Pi}_h u(t) - \widetilde{u}_h(t)) = -\eta(t) + \psi_h(t), \tag{4.25}$$

where $\widetilde{\Pi}_h$ is again the interpolation operator onto $\widetilde{V}_h$. According to (2.28), the interpolation error can be estimated by

$$\|\partial_t \eta\|_{L^\infty(0,T,L^2(\Omega))} + \|\operatorname{div} \eta\|_{L^\infty(0,T,L^2(\Omega))} \leq C'(u,T)h.$$

For the discrete error, we can compute

$$\begin{aligned}
\|\widetilde{\Pi}_h \partial_t u(t) - \partial_t \widetilde{u}_h(t)\|_{L^2(\Omega)} &= \|\widetilde{\Pi}_h \partial_t u(t) - \widetilde{\Pi}_h \partial_t u_h(t)\|_{L^2(\Omega)} \\
&= \|\widetilde{\Pi}_h(\widetilde{\Pi}_h \partial_t u(t) - \partial_t u_h(t))\|_{L^2(\Omega)} \\
&\leq c\|\widetilde{\Pi}_h \partial_t u(t) - \partial_t u_h(t)\|_{L^2(\Omega)}.
\end{aligned}$$

The L^2-estimate is allowed since $\widetilde{\Pi}_h$ is applied to discrete functions. Similarly, we get

$$\begin{aligned}
(\operatorname{div}(\widetilde{\Pi}_h u(t) - \widetilde{u}_h(t)), q_h) &= (\operatorname{div}(\widetilde{\Pi}_h u(t) - \widetilde{\Pi}_h u_h(t), q_h) \\
&= (\operatorname{div} \widetilde{\Pi}_h u(t) - \pi_h^1 \operatorname{div} u_h(t), q_h) \\
&= (\operatorname{div} \widetilde{\Pi}_h u(t) - \operatorname{div} u_h(t)), q_h),
\end{aligned}$$

which implies that $\operatorname{div}(\widetilde{\Pi}_h u(t) - \widetilde{u}_h(t)) = \operatorname{div}(\widetilde{\Pi}_h u(t) - u_h(t))$. Combining these estimates with the results of Theorem 4.2.2 completes the proof. □

Remark 4.3.4. Note that in the proof above, the estimate for the divergence holds with equality. In particular, this means that the divergence produced by the first order method coincides with the divergence produced by the Yee-like method. For the mixed system, this property also translates to the pressure component p. For details, we refer to [**47**].

4.3.4. Construction of basis functions

In this section, we construct explicit basis functions for $\widetilde{V}_h$ on triangles and tetrahedra. For completeness, we also briefly restate the basis functions for V_h and we introduce the structure of the projection matrix P on both triangular and tetrahedral meshes. In the following, we will use the notation $(a;b) = (a,b)^\top$ to describe column vectors.

Construction on triangular meshes

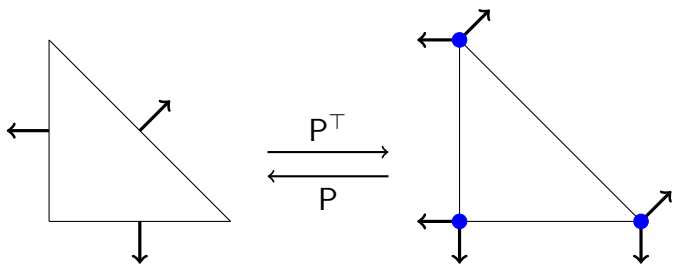

Figure 4.2.: Visual representation of the action of the projection matrix P mapping basis functions from $\mathrm{BDM}_1(\widehat{K})$ to $\mathrm{RT}_0(\widehat{K})$ and back.

Let $\widehat{K} = \mathrm{span}\{(0;0),(1;0),(0;1)\}$ denote the reference triangle. On $\widehat{K}$, we then consider the local basis functions

$$\begin{aligned} \widehat{\varphi}_1 &= (0;x) & \widehat{\varphi}_2 &= (-y;0) & \widehat{\varphi}_3 &= (-y;-y) \\ \widehat{\varphi}_4 &= (0;x+y-1) & \widehat{\varphi}_5 &= (-x-y+1;0) & \widehat{\varphi}_6 &= (x;x). \end{aligned}$$

Recall that $\mathrm{BDM}_1(\widehat{K}) = P_1(\widehat{K})^2 = \mathrm{span}\{\widehat{\varphi}_i,\, 1 \le i \le 6\}$. Moreover, together with the vertex rule, we obtain a block-diagonal local mass matrix consisting of 2×2 blocks. Furthermore, the functions

$$\begin{aligned} \widehat{\Phi}_1 &= \frac{\sqrt{2}}{2}\,(\widehat{\varphi}_1 + \widehat{\varphi}_2) = \frac{\sqrt{2}}{2}(x;y), \\ \widehat{\Phi}_2 &= \frac{\sqrt{2}}{2}\,(\widehat{\varphi}_3 + \widehat{\varphi}_4) = \frac{\sqrt{2}}{2}(x-1;y), \\ \widehat{\Phi}_3 &= \frac{\sqrt{2}}{2}\,(\widehat{\varphi}_5 + \widehat{\varphi}_6) = \frac{\sqrt{2}}{2}(x;y-1) \end{aligned}$$

define basis functions for RT_0, namely

$$\mathrm{RT}_0(\widehat{K}) = P_0(\widehat{K})^2 \oplus x \cdot P_0(\widehat{K}) = \mathrm{span}\{\widehat{\Phi}_i,\, 1 \le i \le 3\}.$$

Then, locally, the projection P takes the form

$$\mathsf{P}|_{\widehat{K}} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}.$$

Globally, P has the same structure as above, with exactly two entries $\frac{\sqrt{2}}{2}$ per row, with their exact location depending on the numbering of the degrees of freedom. We can also check that $\mathsf{P}\mathsf{P}^\top = \mathsf{Id}$. In conclusion, (P1)–(P2) are satisfied.

Construction on tetrahedral meshes

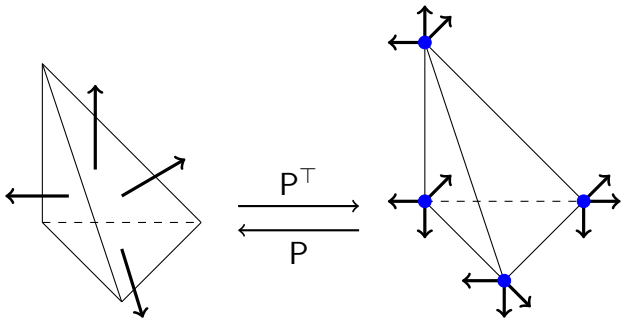

Figure 4.3.: Visual representation of the action of the projection matrix P mapping basis functions from $\mathrm{BDM}_1(\widehat{K})$ to $\mathrm{RT}_0(\widehat{K})$ and back.

Let $\widehat{K} = \text{span}\{(0;0;0),(1;0;0),(0;1;0),(0;0;1)\}$ denote the reference tetrahedron. We then consider the functions

$$\begin{aligned}
&\widehat{\varphi}_1 = (x;0;0) && \widehat{\varphi}_2 = (0;y;0) && \widehat{\varphi}_3 = (0;0;z)\\
&\widehat{\varphi}_4 = (-y;y;0) && \widehat{\varphi}_5 = (-z;0;z) && \widehat{\varphi}_6 = (x+y+z-1;0;0)\\
&\widehat{\varphi}_7 = (0;-z;z) && \widehat{\varphi}_8 = (0;x+y+z-1;0) && \widehat{\varphi}_9 = (x;-x;0)\\
&\widehat{\varphi}_{10} = (0;0;x+y+z-1) && \widehat{\varphi}_{11} = (x;0;-x) && \widehat{\varphi}_{12} = (0;y;-y),
\end{aligned}$$

which define a basis for $\text{BDM}_1(\widehat{K})$. Furthermore, the functions

$$\begin{aligned}
\widehat{\Phi}_1 &= \frac{\sqrt{3}}{3}(\widehat{\varphi}_1+\widehat{\varphi}_2+\widehat{\varphi}_3) &&= \frac{\sqrt{3}}{3}(x;y;z),\\
\widehat{\Phi}_2 &= \frac{\sqrt{3}}{3}(\widehat{\varphi}_4+\widehat{\varphi}_5+\widehat{\varphi}_6) &&= \frac{\sqrt{3}}{3}(x-1;y;z),\\
\widehat{\Phi}_3 &= \frac{\sqrt{3}}{3}(\widehat{\varphi}_7+\widehat{\varphi}_8+\widehat{\varphi}_9) &&= \frac{\sqrt{3}}{3}(x;y-1;z),\\
\widehat{\Phi}_4 &= \frac{\sqrt{3}}{3}(\widehat{\varphi}_{10}+\widehat{\varphi}_{11}+\widehat{\varphi}_{12}) &&= \frac{\sqrt{3}}{3}(x;y;z-1)
\end{aligned}$$

define basis functions for $RT_0(\widehat{K})$, namely

$$\text{RT}_0(\widehat{K}) = P_0(\widehat{K})^2 \oplus x\cdot P_0(\widehat{K}) = \text{span}\{\widehat{\Phi}_i,\ 1\le i\le 4\}.$$

Then, locally, the projection matrix P takes the form

$$\mathsf{P}|_{\widehat{K}} = \begin{pmatrix}
\frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3}
\end{pmatrix}.$$

Globally, P has the same structure as above, with exactly three entries $\frac{\sqrt{3}}{3}$ per row, with their location depending on the numbering of the degrees of freedom. We can check that $\mathsf{P}\mathsf{P}^\top = \mathsf{Id}$. In conclusion, (P1)–(P2) are satisfied.

4.3.5. Remarks

Our proposed Yee-like scheme can also be extended on quadrilaterals and hexahedra by utilizing the corresponding RT_0 and BDM_1 spaces, see Figure 4.4 for a representation of the action of the projection matrix P in 2D. In [**49**], we observed that for the mixed formulation (4.1)–(4.2), our Yee-like method results in the same algebraic system as the finite difference time domain method. For this reason, our approach can be seen as a natural variational generalization of the FDTD method to unstructured grids.

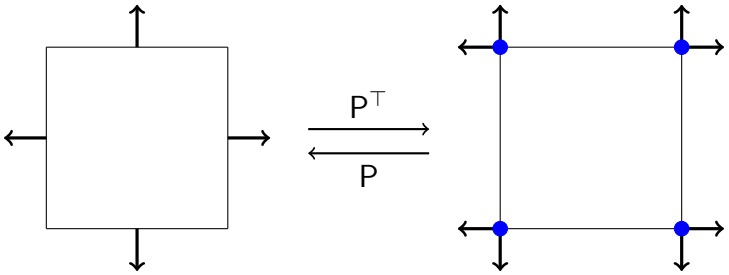

Figure 4.4.: Visual representation of the action of the projection matrix P mapping basis functions from $\mathrm{BDM}_1(\widehat{K})$ to $\mathrm{RT}_0(\widehat{K})$ and back.

4.4. A second-order method

We now consider the extension of the first order method from Section 4.2 to second order. Inspired by the results of Section 2.3, we consider the approximation space

$$V_h = \{v_h \in H_0(\mathrm{div}, \Omega) \ : \ v_h|_K \in \mathrm{RT}_1(K), \quad \forall K \in \mathcal{T}_h\} \tag{4.26}$$

and denote by Π_h its canonical interpolation operator. Let us further recall the associated quadrature formula

$$Q_K(f) = |K| \left(\alpha f(m) + \sum_{i=1}^{d+1} \beta \, f(v_i) \right), \tag{4.27}$$

with m and v_i representing the element midpoint and its vertices, which are used as quadrature points, while α and β are the corresponding weights; see Figure 4.5.

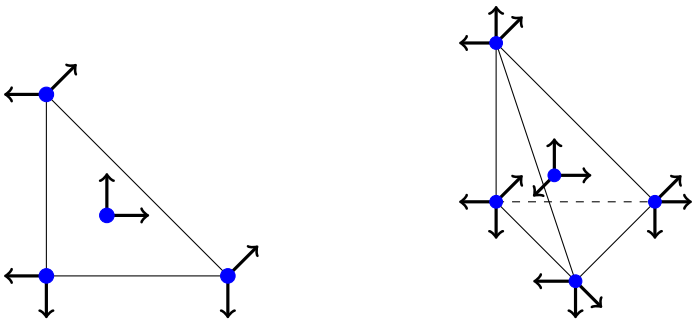

Figure 4.5.: Degrees of freedom for the space RT_1 on the unit triangle (left) and the unit tetrahedron (right), as well as the quadrature points of (4.27). The blue dots represent the locations of the quadrature points. The associated weights are $\alpha = \frac{3}{4}$ and $\beta = \frac{1}{12}$ on the triangle and $\alpha = \frac{4}{5}$ and $\beta = \frac{1}{20}$ on the tetrahedron.

With the help of the local quadrature rule, we define the global inexact scalar product

$$(u, w)_h \coloneqq \sum\nolimits_{K \in \mathcal{T}_h} Q_K(u \cdot w). \tag{4.28}$$

We then consider the following second-order finite element scheme.

Problem 4.4.1 (Second order inexact Galerkin approximation)**.**
For $u_h(0) = \Pi_h u(0)$ and $\partial_t u_h(0) = \Pi_h \partial_t u(0)$ find $u_h \in C^2([0,T], V_h)$ such that

$$(\partial_{tt} u_h(t), v_h)_h + (\operatorname{div} u_h(t), \operatorname{div} v_h) = (f(t), v_h), \qquad \forall v_h \in V_h \;\; t > 0. \tag{4.29}$$

The well-posedness of this method is again a direct consequence of the norm equivalence

$$c_1(v_h, v_h) \leq (v_h, v_h)_h \leq c_2(v_h, v_h), \qquad \forall v_h \in V_h, \tag{4.30}$$

which holds with uniform constants c_1, c_2; see Chapter 3 for details. As a next step, we gather some information regarding the error due to inexact numerical integration.

4.4.1. Quadrature error

For any element $K \in \mathcal{T}_h$, we introduce the local and global quadrature errors

$$\sigma_K(u, v) \coloneqq Q_K(u \cdot v) - (u, v)_K \qquad \text{and} \qquad \sigma_h(u, v) = \sum\nolimits_{K \in T_h} \sigma_K(u, v). \tag{4.31}$$

We will make use of the following result, which is a small modification of Lemma 2.3.7.

Lemma 4.4.2 (Quadrature error)**.**
The quadrature rule (4.27) *is exact for polynomials of degree $k \leq 2$ on both triangles and tetrahedra. Further, let (A1)–(A2) hold and let π_h^1 denote the L^2-orthogonal projection onto $P_1(\mathcal{T}_h)^d$. Then the quadrature error* (4.31) *satisfies*

$$|\sigma_h(\pi_h^1 u, v_h)| \leq C h^2 \|u\|_{H^1(\mathcal{T}_h)} \|\operatorname{div} v_h\|_{L^2(\Omega)} \qquad \forall v_h \in V_h, \;\; u \in H^1(\mathcal{T}_h)^d. \tag{4.32}$$

Proof. The exactness property can be verified easily. In view of the definition of the global quadrature error, it suffices to show the inequality locally for individual elements $K \in \mathcal{T}_h$. We begin by splitting

$$\begin{aligned} \sigma_K(\pi_h^1 u, v_h) &= \sigma_K(\pi_h^1 u - \pi_h^0 u, v_h - \pi_h^1 v_h) \\ &\leq c \|\pi_h^1 u - \pi_h^0 u\|_{L^2(K)} \|v_h - \pi_h^1 v_h\|_{L^2(K)} \\ &\leq c' h_T^3 \|u\|_{H^1(K)} \|\nabla^2 v_h\|_{L^2(K)}. \end{aligned}$$

Here, we used the exactness of the quadrature rule first, the norm equivalence (4.30) in the second line, and lastly the approximation properties of the projection π_h^1, see (2.29). Using assertion (*ii*) of Lemma 2.3.6, we now split $v_h = v_h^1 \oplus v_h^{\mathcal{B}}$ and further estimate

$$\|\nabla^2 v_h\|_{L^2(K)} = \|\nabla^2 v_h^B\|_{L^2(K)} \leq C' h^{-1} \|\nabla v_h^{\mathcal{B}}\|_{L^2(K)} \leq C'' h^{-1} \|\operatorname{div}(v_h^{\mathcal{B}})\|_{L^2(K)}.$$

Here, we used estimate (*ii*) of Lemma 2.3.6 in the last step. By the estimate (*iii*) of Lemma 2.3.6, we can further bound $\|\operatorname{div}(v_h^B)\|_{L^2(K)} \leq C \|\operatorname{div}(v_h)\|_{L^2(K)}$. A combination of these results and summation over elements yields the first estimate of the lemma. □

4.4.2. Error analysis

We begin by splitting the error in interpolation and discrete error components, i.e.

$$u(t) - u_h(t) = (u(t) - \Pi_h u(t)) + (\Pi_h u(t) - u_h(t)) = -\eta(t) + \psi_h(t). \tag{4.33}$$

Then we obtain the following result regarding the discrete error component $\psi_h(t)$.

Lemma 4.4.3 (Discrete error estimate)**.**
Let (A1) and (A2) hold, let u be a sufficiently smooth solution of (4.5)–(4.7), and u_h be the corresponding solution of Problem 4.4.1. Then

$$\|\partial_t(\Pi_h u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} + \|\mathrm{div}\,(\Pi_h u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} \leq C(u,T)h^2$$

with $C(u,T) = c(\|\partial_{tt}u\|_{L^1(0,T,H^2(\mathcal{T}_h))} + \|\partial_{tt}u\|_{L^\infty(0,T,H^1(\mathcal{T}_h))} + \|\partial_{ttt}u\|_{L^1(0,T,H^1(\mathcal{T}_h))})$ and a constant c which is independent of the mesh size.

Proof. The discrete error $\psi_h(t) = \Pi_h u(t) - u_h(t)$ satisfies the discrete system

$$\begin{aligned}(\partial_{tt}\psi_h(t), v_h)_h + (\mathrm{div}\,\psi_h(t), \mathrm{div}\,v_h)& \\ = (\partial_{tt}\eta(t), v_h) + (\mathrm{div}\,\eta(t), \mathrm{div}\,v_h) + \sigma_h(\Pi_h\partial_{tt}u(t), v_h)&\end{aligned}$$

with initial values $\psi_h(0) = \partial_t\psi_h(0) = 0$. Recall that

$$(\partial_{tt}\psi_h(t), \partial_t\psi_h(t)) + (\mathrm{div}\,\psi_h(t), \partial_t\mathrm{div}\,\psi_h(t)) = \frac{1}{2}\frac{d}{dt}\left(\|\partial_t\psi_h(t)\|_h^2 + \|\mathrm{div}\,\psi_h(t)\|_{L^2(\Omega)}^2\right).$$

Then, choosing $v_h := \partial_t\psi_h(t)$ as a test function, integrating from 0 to t and using the definition of the initial conditions yields the identity

$$\begin{aligned}&\frac{1}{2}\left(\|\partial_t\psi_h(t)\|_h^2 + \|\mathrm{div}\,\psi_h(t)\|_{L^2(\Omega)}^2\right)\\ &= \int_0^t (\partial_{tt}\eta(s), \partial_t\psi_h(s)) + (\mathrm{div}\,\eta(s), \mathrm{div}\,\partial_t\psi_h(s)) + \sigma_h(\Pi_h\partial_{tt}u(s), \partial_t\psi_h(s))\\ &= (i) + (ii) + (iii).\end{aligned} \tag{4.34}$$

The first term can be estimated by Cauchy-Schwarz and Young's inequality by

$$\begin{aligned}(i) &\leq \|\partial_{tt}(\Pi_h u - u)\|_{L^1(0,t,L^2(\Omega))} + \frac{1}{4c_1}\|\partial_t\psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2\\ &\leq Ch^4\|\partial_{tt}u\|_{L^1(0,t,H^2(\mathcal{T}_h))}^2 + \frac{1}{4c_1}\|\partial_t\psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.\end{aligned}$$

Using the commuting diagram property (2.30) of Π_h, we conclude that $(ii) = 0$. The third term can be split into

$$\begin{aligned}(iii) &= \int_0^t \sigma_h(\Pi_h\partial_{tt}u(s) - \pi_h^1\partial_{tt}u(s), \partial_t\psi_h(s)) + \int_0^t \sigma_h(\pi_h^1\partial_{tt}u(s), \partial_t\psi_h(s))\\ &= (iv) + (v),\end{aligned}$$

where π_h^1 denotes the L^2-projection onto space $P_1(\mathcal{T}_h)^d$. With similar arguments as used in the estimate of (i), we get

$$\begin{aligned}(iv) &\le \|\partial_{tt}(\Pi_h u - \pi_h^1 u)\|^2_{L^1(0,t,L^2(\Omega))} + \frac{1}{4c_1}\|\partial_t \psi_h\|^2_{L^\infty(0,t,L^2(\Omega))} \\ &\le \|\partial_{tt}(\Pi_h u - u)\|^2_{L^1(0,t,L^2(\Omega))} + \|\partial_{tt}(u - \pi_h^1 u)\|^2_{L^1(0,t,L^2(\Omega))} + \frac{1}{4c_1}\|\partial_t \psi_h\|^2_{L^\infty(0,t,L^2(\Omega))} \\ &\le Ch^4 \|\partial_{tt} u\|^2_{L^1(0,t,H^2(\mathcal{T}_h))} + \frac{1}{4c_1}\|\partial_t \psi_h\|^2_{L^\infty(0,t,L^2(\Omega))}.\end{aligned}$$

For (v), we use integration by parts in time and $\psi_h(0) = 0$, which yields

$$(v) = \sigma_h(\pi_h^1 \partial_{tt} u(t), \psi_h(t)) - \int_0^t \sigma_h(\pi_h^1 \partial_{ttt} u(s), \psi_h(s))\, ds.$$

By the properties of the quadrature error stated in Lemma 4.4.2, we then obtain

$$\begin{aligned}(v) &\le ch^4 \|\partial_{tt} u(t)\|^2_{H^1(\mathcal{T}_h)} + \frac{1}{4}\|\operatorname{div} \psi_h(t)\|^2_{L^2(\Omega)} \\ &\qquad + c' h^4 \|\partial_{ttt} u\|^2_{L^1(0,t,H^1(\mathcal{T}_h))} + \frac{1}{4}\|\operatorname{div} \psi_h\|^2_{L^\infty(0,t,L^2(\Omega))} \\ &\le Ch^4 \left(\|\partial_{tt} u\|^2_{L^\infty(0,t;H^1(\mathcal{T}_h))} + \|\partial_{ttt} u\|^2_{L^1(0,t,H^1(\mathcal{T}_h))} \right) + \frac{1}{2}\|\operatorname{div} \psi_h\|^2_{L^\infty(0,t,L^2(\Omega))}\end{aligned}$$

with $C_1''(u,t) = C(\|\partial_{tt} u\|_{L^\infty(0,t,H^1(\mathcal{T}_h))} + \|\partial_{ttt} u\|_{L^1(0,t,H^1(\mathcal{T}_h))})$. Summing all the terms and using the norm equivalence from (4.16), we obtain

$$\begin{aligned}&\frac{1}{c_1}\|\partial_t \psi_h(t)\|^2_{L^2(\Omega)} + \|\operatorname{div} \psi_h(t)\|^2_{L^2(\Omega)} \\ &\qquad\qquad \le C(u,t)h^4 + \frac{1}{2c_1}\|\partial_t \psi_h\|^2_{L^\infty(0,t,L^2(\Omega))} + \frac{1}{2}\|\operatorname{div} \psi_h(t)\|^2_{L^2(\Omega)}.\end{aligned}$$

Taking the maximum over all t and subsequently absorbing the last two terms by the left-hand side yields the desired estimate for the discrete error. □

Combining the bounds for the interpolation error in Section 2.3.1 and for the discrete error from Lemma 4.4.3 immediately yields the following main result of this section.

Theorem 4.4.4 (Second order convergence)**.**
Let (A1) and (A2) hold, u be a sufficiently smooth solution of (4.5)–(4.7), and u_h be the corresponding solution of Problem 4.4.1. Then

$$\|\partial_t(u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} + \|\operatorname{div}(u - u_h)\|_{L^\infty(0,T;L^2(\Omega))} \le C(u,T)h^2 \tag{4.35}$$

with

$$\begin{aligned}C(u,T) = c\Big(&\|\partial_{tt} u\|_{L^1(0,T,H^2(\mathcal{T}_h))} + \|\partial_{tt} u\|_{L^\infty(0,T,H^1(\mathcal{T}_h))} + \|\partial_{ttt} u\|_{L^1(0,T,H^1(\mathcal{T}_h))} \\ &+ \|\partial_t u\|_{L^\infty(0,T;H^2(\mathcal{T}_h))} + \|\operatorname{div} u\|_{L^\infty(0,T;H^2(\mathcal{T}_h))}) \Big)\end{aligned}$$

and a constant c that is independent of the mesh size h.

Proof. According to (2.28), we can estimate the interpolation error by

$$\|\partial_t \eta\|_{L^\infty(0,T;L^2(\Omega))} + \|\operatorname{div} \eta\|_{L^\infty(0,T;L^2(\Omega))} \leq C'(u,T)h^2 \tag{4.36}$$

with $C'(u,t) = C(\|\partial_t u\|_{L^\infty(0,T;H^2(\mathcal{T}_h))} + \|\operatorname{div} u\|_{L^\infty(0,T;H^2(\mathcal{T}_h))})$. The proof is completed by adding the discrete error components from Lemma 4.4.3. □

Remark 4.4.5. Comparing this result to the error estimates for the exact Galerkin approximation from Lemma 4.1.3 shows that mass lumping does not affect the overall accuracy of the method. After choosing the set of basis functions from Section 2.4.1, the algebraic realization of (4.29) leads to the system $\mathsf{M_h}\partial_{tt}\mathsf{u}(t) + \mathsf{Ku}(t) = 0$ with a block-diagonal mass matrix $\mathsf{M_h}$. We can now propose explicit time-stepping schemes, which can be applied efficiently as a result of mass lumping.

4.5. Time discretization

In this section, we discuss the time discretization for the second order method with mass lumping from Section 4.4. We utilize a central difference approximation on a uniform grid in time, which leads to an explicit time-stepping method with formally second-order accuracy. Let $t^n = n\tau$, $0 \leq n \leq N$ be a given sequence of time points with $\tau = T/N$. We denote by $g^n = g(t^n)$ the evaluation of a function g defined on the time grid

$$I_\tau = \{0 = t^0 < t^1 < \ldots < t^N = T\}.$$

Furthermore, for a continuous function g in time, we define

$$g^{n+\frac{1}{2}} \coloneqq \frac{1}{2}(g^{n+1} + g^n),$$

which sets the intermediate evaluations to equal the average of two consecutive integer time evaluations. Furthermore, we consider the first order (central) difference quotients

$$d_\tau g^{n+\frac{1}{2}} \coloneqq \frac{1}{\tau}(g^{n+1} - g^n). \qquad \text{and} \qquad \widehat{d_\tau g}^{\,n} \coloneqq \frac{d_\tau g^{n+\frac{1}{2}} + d_\tau g^{n-\frac{1}{2}}}{2} = \frac{g^{n+1} - g^{n-1}}{2\tau},$$

as well as the second order central difference quotient

$$d_{\tau\tau} g^n \coloneqq \frac{1}{\tau^2}(g^{n+1} - 2g^n + g^{n-1}).$$

In the following, let V_h be defined as in (4.26), while $(\cdot,\cdot)_h$ is given by (4.28). We then consider the following fully discrete scheme.

Problem 4.5.1 (Second order fully discrete method)**.**
Set $u_h^0 = \Pi_h u_0$ *and* $u_h^1 = \Pi_h \left(u_0 + \tau v_0 + \frac{\tau^2}{2}(f(0) + \nabla \operatorname{div} u_0)\right)$. *Then for* $n \geq 1$ *find* u_h^n *such that*

$$(d_{\tau\tau} u_h^n, v_h)_h + (\operatorname{div} u_h^n, \operatorname{div} v_h) = (f(t^n), v_h) \qquad \forall v_h \in V_h. \tag{4.37}$$

Let us again very briefly discuss the well-posedness of the scheme. On the algebraic level, the system (4.37) takes the form

$$\mathsf{M_h u}^{n+1} = 2\mathsf{M_h u}^n - \mathsf{M_h u}^{n-1} - \tau \mathsf{K u}^n + \tau \mathsf{f}^n.$$

Since $\mathsf{M_h}$ is regular, see (4.30), the iteration scheme is well-defined, and by extension, so is (4.37). For the stability of the fully discrete scheme, we further assume that

(A3) the time step τ is chosen such that

$$(\operatorname{div} v_h, \operatorname{div} v_h) \leq \frac{4}{\tau^2}(v_h, v_h)_h \qquad \forall v_h \in V_h,$$

which can be interpreted as an abstract CFL condition; see [**73**] for details. We proceed as in Section 4.4.2 by splitting the error into discrete and projection error components

$$u(t^n) - u_h^n = -(\Pi_h u(t^n) - u(t^n)) + (\Pi_h u(t^n) - u_h^n) =: -\eta^n + \xi_h^n.$$

4.5.1. Discrete error estimates

The discrete error can be again interpreted as a solution of (4.37) with a special right-hand side. First, we make the following statement regarding the stability of the discrete problem.

Lemma 4.5.2 (Discrete stability estimate)**.**
Let $\{\xi_h^n\}, \{r_h^n\} \subset V_h$ be given sequences such that

$$(d_{\tau\tau}\xi_h^n, v_h)_h + (\operatorname{div} \xi_h^n, \operatorname{div} v_h) = (r_h^n, v_h) \qquad \forall v_h \in V_h, \;\; n \geq 1. \tag{4.38}$$

Furthermore, assume that (A1)–(A3) hold. Then for all $0 \leq n \leq N-1$, we have

$$\begin{aligned} &\|d_\tau \xi_h^{n+\frac{1}{2}}\|^2_{L^2(\Omega)} + \|\operatorname{div} \widehat{\xi}_h^{n+\frac{1}{2}}\|^2_{L^2(\Omega)} \\ &\qquad \leq C\left(\|d_\tau \xi_h^{\frac{1}{2}}\|^2_{L^2(\Omega)} + \|\operatorname{div} \widehat{\xi}_h^{\frac{1}{2}}\|^2_{L^2(\Omega)} + \sum_{i=1}^{n} \tau (r_h^i, \widehat{d_\tau \xi_h}^{\,n}) \right). \end{aligned} \tag{4.39}$$

Proof. We follow the arguments of [**73**]. Testing (4.38) with $v_h = \widehat{d_\tau \xi_h}^{\,n}$ leads to

$$(r_h^n, \widehat{d_\tau \xi_h}^{\,n}) = (d_{\tau\tau}\xi_h^n, \widehat{d_\tau \xi_h}^{\,n})_h + (\operatorname{div} \xi_h^n, \operatorname{div} \widehat{d_\tau \xi_h}^{\,n}) = (i) + (ii). \tag{4.40}$$

For the first term, we obtain by elementary computations that

$$(i) = \frac{1}{2\tau}\left(\|d_\tau \xi_h^{n+\frac{1}{2}}\|_h^2 - \|d_\tau \xi_h^{n-\frac{1}{2}}\|_h^2 \right),$$

and the second term can be expanded into

$$\begin{aligned} (ii) = \frac{1}{2\tau}\Big(&\|\operatorname{div} \widehat{\xi}_h^{n+\frac{1}{2}}\|^2_{L^2(\Omega)} - \|\operatorname{div} \widehat{\xi}_h^{n-\frac{1}{2}}\|^2_{L^2(\Omega)} \\ &- \frac{\tau^2}{4}\|\operatorname{div} d_\tau \xi_h^{n+\frac{1}{2}}\|^2_{L^2(\Omega)} + \frac{\tau^2}{4}\|\operatorname{div} d_\tau \xi_h^{n-\frac{1}{2}}\|^2_{L^2(\Omega)} \Big). \end{aligned}$$

We now define the discrete energy as

$$\mathcal{E}_h^{n+\frac{1}{2}} = \|d_\tau \xi_h^{n+\frac{1}{2}}\|_h^2 + \|\operatorname{div} \widehat{\xi}_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 - \frac{\tau^2}{4}\|\operatorname{div} d_\tau \xi_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2.$$

Plugging this definition back into (4.40) yields

$$\mathcal{E}_h^{n+\frac{1}{2}} = \mathcal{E}_h^{n-\frac{1}{2}} + 2\tau(r_h^n, \widehat{d_\tau \xi_h}^n).$$

A recursive application of this identity finally leads to $\mathcal{E}_h^{n+\frac{1}{2}} = \mathcal{E}_h^{\frac{1}{2}} + 2\sum_{i=1}^n \tau(r_h^i, \widehat{d_\tau \xi_h}^i)$. From (A3), we can see that $\mathcal{E}_h^{n+\frac{1}{2}}$ is positive and

$$\mathcal{E}_h^{n+\frac{1}{2}} \leq \|d_\tau \xi_h^{n+\frac{1}{2}}\|_h^2 + \|\operatorname{div} \widehat{\xi}_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 \leq 2\mathcal{E}_h^{n+\frac{1}{2}}.$$

The assertion now follows by using the norm equivalence (4.30). □

Now that stability of the scheme has been established, we can complete the analysis of the discrete error by the following estimate.

Lemma 4.5.3 (Discrete error estimate)**.**
Let (A1)–(A3) hold, u be a sufficiently smooth solution of (4.5)–(4.7), and u_h be the corresponding solution of Problem 4.4.1. Then the discrete error satisfies

$$\|d_\tau \xi_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 + \|\operatorname{div} \widehat{\xi}_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 \leq C(u,T)h^2 + C'(u,T)\tau^2,$$

for all $t^n \leq T$ and with constants $C(u,T)$ and $C'(u,T)$ as in the proof below.

Proof. We have to estimate the terms on the right-hand side of (4.39). First, note that $\xi_h^0 = 0$ by definition of the first initial value. By Taylor estimates, we further obtain

$$\begin{aligned}\xi_h^1 = \Pi_h u(t^1) - u_h^1 &= \Pi_h \left(u(t^1) - u(0) - \tau \partial_t u(0) - \frac{\tau^2}{2}(f(0) + \nabla \operatorname{div} u(0)) \right) \\ &= \Pi_h \left(u(t^1) - u(0) - \tau \partial_t u(0) - \frac{\tau^2}{2}\partial_{tt} u(0) \right) = \Pi_h \left(\frac{\tau^3}{6} \partial_{ttt} u(s) \right)\end{aligned}$$

for an $s \in [0, t^1]$. With this in mind, we compute

$$\begin{aligned}\|d_\tau \xi_h^{\frac{1}{2}}\|_{L^2(\Omega)}^2 + \|\operatorname{div} \widehat{\xi}_h^{\frac{1}{2}}\|_{L^2(\Omega)}^2 &= \frac{1}{\tau^2}\|d_\tau \xi_h^1\|_{L^2(\Omega)}^2 + \frac{1}{4}\|\operatorname{div} \xi_h^1\|_{L^2(\Omega)}^2 \\ &\leq C\tau^2 \left(\|\Pi_h \partial_{ttt} u\|_{L^\infty(0,t^1,L^2(\Omega))}^2 + \|\operatorname{div} \Pi_h \partial_{tt} u\|_{L^\infty(0,t^1,L^2(\Omega))}^2 \right) \\ &\leq C\tau^2 \left(\|\partial_{ttt} u\|_{L^\infty(0,t^1,H^1(\mathcal{T}_h))}^2 + \|\pi_h^1 \operatorname{div} \partial_{tt} u\|_{L^\infty(0,t^1,L^2(\Omega))}^2 \right) \\ &\leq C\tau^2 \left(\|\partial_{ttt} u\|_{L^\infty(0,t^1,H^1(\mathcal{T}_h))}^2 + \|\operatorname{div} \partial_{tt} u\|_{L^\infty(0,t^1,L^2(\Omega))}^2 \right),\end{aligned}$$

where we used the H^1 stability of the interpolation Π_h and (2.30). For the third term, we test equations (4.8) and (4.37) with $v_h = \widehat{d_\tau \xi_h}^i$ and then add from 1 to n, which yields

$$\begin{aligned}\sum_{i=1}^{n}\tau(r_h^i,\widehat{d_\tau\xi_h}^{\,i}) &= \tau\sum_{i=1}^{n}(d_{\tau\tau}u(t^i)-\partial_{tt}u(t^i),\widehat{d_\tau\xi_h}^{\,i}) + \tau\sum_{i=1}^{n}(d_{\tau\tau}\eta^i,\widehat{d_\tau\xi_h}^{\,i})\\ &\quad+\tau\sum_{i=1}^{n}(\operatorname{div}\eta^i,\operatorname{div}\widehat{d_\tau\xi_h}^{\,i}) - \tau\sum_{i=1}^{n}\sigma_h(d_{\tau\tau}\Pi_h u(t^i),\widehat{d_\tau\xi_h}^{\,i})\\ &= (i)+(ii)+(iii)+(iv).\end{aligned}$$

Using Taylor expansions and Cauchy-Schwarz inequalities, we get

$$\begin{aligned}|(i)| &\le \tau\sum_{i=1}^{n}C\tau\|\partial_{tttt}u\|_{L^1(t^{i-1},t^{i+1},L^2(\Omega))}\|\widehat{d_\tau\xi_h}^{\,i}\|_{L^2(\Omega)}\\ &\le \sum_{i=1}^{n}C\tau^2\|\partial_{tttt}u\|_{L^1(t^{i-1},t^{i+1};L^2(\Omega))}\cdot\max_{1\le i\le n}\|\widehat{d_\tau\xi_h}^{\,i}\|_{L^2(\Omega)}\\ &\le C\tau^4\|\partial_{tttt}u\|^2_{L^1(0,t^{n+1};L^2(\Omega))}+\frac{1}{6c_1}\max_{1\le i\le n}\|d_\tau\xi_h^{i+\frac12}\|^2_{L^2(\Omega)}.\end{aligned}$$

In a similar fashion, we get for the second term

$$|(ii)|\le 2Ch^4\|\partial_{tt}u\|^2_{L^1(0,t^{n+1};H^2(\mathcal{T}_h))}+\frac{1}{6c_1}\max_{0\le i\le n}\|d_\tau\xi_h^{i+\frac12}\|^2_{L^2(\Omega)}.$$

For the third term, we use summation by parts to obtain

$$\begin{aligned}(iii) &= (\operatorname{div}\eta^n,\operatorname{div}\widehat{\xi}_h^{n+\frac12}) - (\operatorname{div}\eta^1,\operatorname{div}\widehat{\xi}_h^{\frac12}) - \tau\sum_{i=2}^{n}(\operatorname{div}d_\tau\eta^{i-\frac12},\operatorname{div}\widehat{\xi}_h^{i-\frac12})\\ &= (iii)_a+(iii)_b+(iii)_c.\end{aligned}$$

The first two terms here can be further estimated by

$$|(iii)_a|+|(iii)_b|\le 2Ch^4\|\operatorname{div}u\|^2_{L^\infty(0,t^n;H^2(\mathcal{T}_h))}+\frac16\max_{1\le i\le n}\|\operatorname{div}\widehat{\xi}_h^{i+\frac12}\|^2_{L^2(\Omega)}.$$

For the third term, we obtain

$$|(iii)_c|\le 2Ch^4\|\operatorname{div}\partial_t u\|^2_{L^1(0,t^{n+1};H^1(\mathcal{T}_h))}+\frac16\max_{1\le i\le n}\|\operatorname{div}\widehat{\xi}_h^{i+\frac12}\|^2_{L^2(\Omega)}.$$

Using Lemma 4.4.2 and again summation by parts, we get

$$|(iv)|\le\widetilde{C}(u,T)^2h^4+\frac{1}{6c_1}\max_{0\le i\le n}\|d_\tau\xi_h^{i+\frac12}\|^2_{L^2(\Omega)}+\frac16\max_{0\le i\le n}\|\operatorname{div}\widehat{\xi}_h^{i+\frac12}\|^2_{L^2(\Omega)}$$

with

$$\widetilde{C}(u,T)=\left(\|\partial_{tt}u\|_{L^1(0,t^{n+1},H^2(\mathcal{T}_h))}+\|\partial_{tt}u\|_{L^\infty(0,t^{n+1};H^1(\mathcal{T}_h))}+\|\partial_{ttt}u\|_{L^1(0,t^{n+1},H^1(\mathcal{T}_h))}\right).$$

Summing up all the terms, using the norm equivalence from (4.30), and subsequently applying Lemma 4.5.2, we finally obtain

$$\begin{aligned}&\frac{1}{c_1}\|d_\tau \xi_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 + \|\operatorname{div}\widehat{\xi}_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 \\ &\leq C(u,T)^2 h^4 + C'(u,T)^2\tau^4 + \frac{1}{2c_1}\max_{0\leq i\leq n}\|d_\tau \xi_h^{i+\frac{1}{2}}\|_{L^2(\Omega)}^2 + \frac{1}{2}\max_{0\leq i\leq n}\|\operatorname{div}\widehat{\xi}_h^{i+\frac{1}{2}}\|_{L^2(\Omega)}^2, \quad (4.41)\end{aligned}$$

with $C(u,T) = \widetilde{C}(u,T) + C\Big(\|\operatorname{div} u\|_{L^\infty(0,t^n;H^2(\mathcal{T}_h))} + \|\operatorname{div}\partial_t u\|_{L^1(0,t^{n+1};H^1(\mathcal{T}_h))}\Big)$ and

$$C'(u,T) = C\left(\|\partial_{tttt}u\|_{L^1(0,t^{n+1},H^1(\mathcal{T}_h))} + \|\partial_{ttt}u\|_{L^\infty(0,t^1,H^1(\mathcal{T}_h))} + \|\operatorname{div}\partial_{tt}u\|_{L^\infty(0,t^1,L^2(\Omega))}\right).$$

Taking the maximum over all t^n in (4.41) and absorbing the last two terms by the left-hand side yields the estimate. The assertion of the theorem now follows by adding the two estimates for the projection and discrete error. □

4.5.2. Error estimates

Together with the estimates for the interpolation operator from Section 2.3.1 and the discrete error from Lemma 4.5.2, we now obtain the following main result.

Theorem 4.5.4 (Second order convergence of the fully discrete scheme)**.**
Let (A1)–(A3) hold and let u denote a sufficiently smooth solution of (4.5)–(4.7). Then the corresponding solution $(u_h^n)_n$ of Problem 4.5.1 satisfies

$$\begin{aligned}\max_{0\leq n<N}\Big(\|\partial_t u(t^{n+\frac{1}{2}}) - d_\tau u_h^{n+\frac{1}{2}}\|_{L^2(\Omega)} + \|\operatorname{div}(u(t^{n+\frac{1}{2}}) - \widehat{u}_h^{n+\frac{1}{2}}))\|_{L^2(\Omega)}\Big)& \\ \leq C(u,T)h^2 + C'(u,T)\tau^2,&\end{aligned}$$

for all $t^n \leq T$ with constant $C(u,T)$ from Theorem 4.4.4 and

$$C'(u,T) = C\left(\|\partial_{tttt}u\|_{L^1(0,t^{n+1},H^1(\mathcal{T}_h))} + \|\partial_{ttt}u\|_{L^\infty(0,t^1,H^1(\mathcal{T}_h))} + \|\operatorname{div}\partial_{tt}u\|_{L^\infty(0,t^1,L^2(\Omega))}\right).$$

Proof. Note that the continuous errors satisfy

$$\begin{aligned}\max_{0\leq n<N}\Big(\|d_\tau \eta^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 + \|\operatorname{div}\widehat{\eta}^{\,n+\frac{1}{2}}\|_{L^2(\Omega)}^2\Big)& \\ \leq ch^2\left(\|\partial_t u\|_{L^\infty(0,T,H^2(\mathcal{T}_h))} + \|\operatorname{div} u\|_{L^\infty(0,T,H^2(\mathcal{T}_h))}\right).&\end{aligned}$$

Aggregating this result with Lemma 4.5.3 already yields the result. □

4.6. Numerical validation

In this section, we compare the three methods proposed in this chapter and verify that the expected convergence rates hold. As a test example, we consider a plane wave traveling through a square domain in 2D. For this example, the exact solution is known, and

convergence rates can be easily computed by comparing the numerical solution to the real one.

For our computation, we consider the exact solution

$$u_{ex}(x,y,t) = g(x+y-2t)\begin{pmatrix}1\\1\end{pmatrix} \quad \text{with} \quad g(x) = 2\exp(-50(x+1)^2)$$

which is a plane wave traveling through $\Omega = (0,1)^2$ from the bottom left to top right corner; see Figure 4.6. One can quickly check that u_{ex} satisfies problem (4.5)–(4.7) with

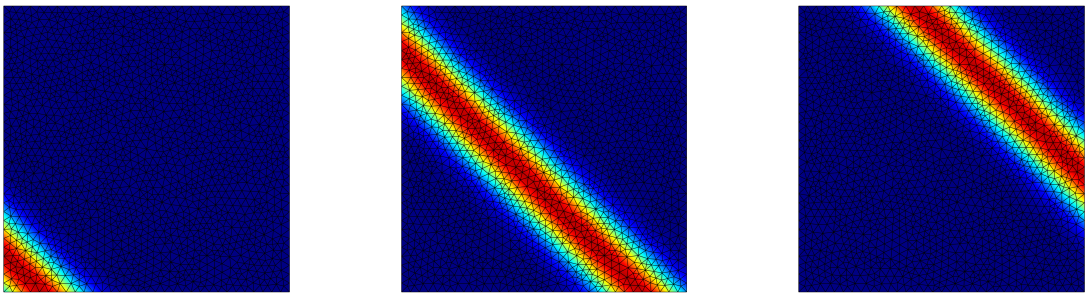

Figure 4.6.: Snapshots of the first component of u at time steps $t \in \{0.8, 1.3, 1.8\}$ computed on the discrete level $h = 2^{-3}$.

$f = 0$ for initial and boundary conditions chosen to match u_{ex}. Let

$$|\!|\!| e_{h,\tau} |\!|\!| := \max_{1\le n\le N} \left(\|\partial_t u(t^{n+\frac{1}{2}}) - d_\tau u_h^{n+\frac{1}{2}}\|_{L^2(\Omega)} + \|\mathrm{div}(u(t^{n+\frac{1}{2}}) - \widehat{u}_h^{n+\frac{1}{2}}))\|_{L^2(\Omega)} \right)$$

denote the discrete error. We set the time horizon to $T = 2$ and choose $\tau = 0.1h$ for our simulations. Since the solution is smooth, we expect that all three methods from Sections 4.2, 4.3 and 4.4 deliver the predicted convergence rates.

In Table 4.1, we display the errors obtained by our finite-element methods with mass lumping on a sequence of non-nesting meshes $\{\mathcal{T}_h\}_h$. The columns denoted by BDM_1, RT_0 and RT_1 contain the error $|\!|\!| e_{h,\tau} |\!|\!|$ for the first order, the Yee-like, and the second order method, respectively. We observe that each of the methods performs as predicted by theory. Note that the error produced by the Yee-like method is a bit larger than the first order method based on BDM_1, which is to be expected from construction.

Let us now further elaborate on the choice of the time step τ, which is restricted by the CFL condition (A3). From the proof of Lemma 4.5.2, we see that a sufficient condition for discrete stability is

$$\frac{\tau_{\max}^2}{4}\|\mathsf{M}_h^{-1}\mathsf{K}\| = 1.$$

By setting $c_{\max} = \frac{\tau_{\max}}{h}$ for the CFL constant, we reshuffle to obtain

$$c_{\max} = \frac{4}{h\sqrt{\lambda_{\max}(\mathsf{M}_h^{-1}\mathsf{K})}}. \tag{4.42}$$

h	τ	BDM_1	eoc	dofs	RT_0	eoc	dofs	RT_1	eoc	dofs
2^{-1}	2^{-2}	0.635807	—	56	0.612817	—	28	0.783171	—	88
2^{-2}	2^{-3}	0.565408	1.12	160	0.577903	0.98	80	0.279806	1.93	336
2^{-3}	2^{-4}	0.265606	1.03	668	0.271699	1.02	334	0.060933	2.10	1312
2^{-4}	2^{-5}	0.080122	1.01	2680	0.083710	1.00	1340	0.016570	1.88	5184
2^{-5}	2^{-6}	0.025764	0.99	10580	0.029837	1.01	5260	0.004380	1.92	20608

Table 4.1.: Errors, mesh size h, time step τ, the estimated order of convergence (eoc) as well as the number of dofs for a plane wave solution on a rectangular domain.

The value of $c_{\max}$ can be computed by solving generalized eigenvalue problems $\mathsf{Kx} = \lambda \mathsf{M_h x}$. We do this for the methods with mass lumping and also for the corresponding exact Galerkin methods without mass lumping. The results are depicted in Table 4.2 below. Since the stability of the RT_0 method is directly inherited from BDM_1, we only display the values for BDM_1. Let us emphasize that CFL constants for the first order methods

h	RT_0	BDM_1	$\mathrm{RT}_0/\mathrm{BDM}_1$ (lump)	RT_1	RT_1 (lump)
2^{-2}	0.375602	0.361745	0.644864	0.282508	0.228658
2^{-3}	0.423002	0.393924	0.714711	0.315001	0.251065
2^{-4}	0.378344	0.363319	0.641621	0.283880	0.238001
2^{-5}	0.364451	0.347337	0.615268	0.269956	0.218350
2^{-6}	0.345823	0.328456	0.597351	0.264649	0.212532

Table 4.2.: Values of the CFL-constant $c_{\max}$ for the exact discretizations by RT_0, BDM_1 and RT_1 *without* mass lumping, and the BDM_1 and RT_1 elements *with* mass lumping; larger is better.

with mass lumping are about twice as large as those for the corresponding methods with exact integration. For the second order method with mass lumping, it is a bit smaller than that of the exact method.

4.7. Discussion

In this chapter, we investigated three finite element methods with mass lumping for acoustic wave propagation in $H(\mathrm{div})$. For each of the methods, a full convergence analysis of the space discretization was given and, additionally, the time discretization by an explicit scheme of second-order was analyzed. In principle, the concepts and results developed in this chapter can be extended to higher order approximations and other element types; see the discussion in Section 2.5. In [**48**], we proposed second-order extensions on quadrilaterals, hexahedra, and prismatic elements.

Let us briefly compare our methods to the direct competitors: As explained in Section 4.3.5, the proposed Yee-like scheme can be extended to regular rectangular grids and

is then equivalent to the FDTD method. In this case, super-convergence can be deduced from the results available for the FDTD scheme. This property is lost, however, on irregular grids. In any case, our approach can be regarded as a natural variational extension of the FDTD schemes to general meshes. Another competitor for the simulation of wave propagation problems is the discontinuous Galerkin (dG) scheme; see [**99, 102, 38**] for some works in this direction. While diagonal or block-diagonal mass matrices are naturally obtained by dG methods, the overall number of DOFs is higher in comparison to conforming approximations with mass lumping. For moderate approximation orders, the proposed methods, therefore, seem very competitive.

In the next chapter, we extend our investigations to Maxwell's equations, which are formulated in $H(\mathrm{curl})$. While the basic idea of mass lumping is similar, the finite element spaces are conceptually different and they require a more delicate analysis.

is

5. Mass lumping for Maxwell's equations in H(curl)

In this last chapter, we extend our mass lumping techniques to Maxwell's equations in time domain in a linear, isotropic, and non-dispersive medium. The governing equations are given by the Ampère and Faraday laws

$$\varepsilon \partial_t E + \sigma E - \operatorname{curl} H = -J \qquad \text{in } \Omega,\ t > 0, \tag{5.1}$$

$$\mu \partial_t H + \operatorname{curl} E = 0 \qquad \text{in } \Omega,\ t > 0, \tag{5.2}$$

where E and H denote the electric and magnetic field intensities, ε and μ denote the permittivity and permeability of the material, σ denotes the electric conductivity, while J models the electric current density.

One of the first and yet to this day most popular discretization technique for (5.1)–(5.2) is the *finite difference time domain* (FDTD) method. It was first proposed in 1966 by Yee [**114**] and extended to various applications, which can be found in [**104**]. The method has several advantages. It has a simple, matrix-free structure, is computationally efficient, and also offers second-order accuracy in both space and time. A closely related method is the *finite integration technique* (FIT) developed by Weiland [**109**], which can be interpreted as a generalized FDTD scheme derived from Whitney forms in the context of discrete exterior calculus. The main difference from classic FDTD is the treatment of the discrete Hodge operator. As it pertains to performance and accuracy, the methods are very similar.

To provably work, the FDTD method requires a pair of orthogonal Cartesian dual grids. This is highly restrictive, as it impedes the discretization of curved boundaries. Moreover, it also leads to complications when it comes to inhomogeneous materials and anisotropic coefficients. Various adaptations to accommodate these shortcomings do exist, but usually result in a loss of accuracy, see [**34, 33**].

As a more flexible alternative, one can consider $H(\operatorname{curl})$-conforming finite element approximations instead. The semi-discretization in space using $H(\operatorname{curl})$ conforming finite elements has been thoroughly analyzed by Monk in [**80, 81, 82, 85, 83**]. In order to guarantee efficiency after time discretization, several mass lumping strategies have been proposed in [**37, 35, 55, 54**], with no statements regarding convergence. Another way around the limitations of FDTD is to use dG schemes, for which we refer to [**66**].

In this chapter, we seek to develop first and second-order $H(\operatorname{curl})$-conforming finite element methods with mass lumping that allow for efficient time integration compatible with the FDTD method. We will only consider the 3D case of tetrahedral elements. This is because in two dimensions, $H(\operatorname{div})$ and $H(\operatorname{curl})$ are isomorphic. As a result, the same

methods proposed in Chapter 4 for triangles also carry over to $H(\mathrm{curl})$ by a simple $\pi/2$ rotation of the basis functions; see [**19, Remark 2.1.5**] and the discussion at the end of the chapter.

For the second order method, we will consider a modification of the element proposed by Elmkies and Joly in [**54**]. In [**51**], we have shown that the method is second-order accurate only under certain conditions, and we manage to correct this by a minor alteration of the space.

Outline

Let us give a brief overview of the contents of this chapter.

Section 5.1. *Preliminaries:* In this section, we introduce the mathematical problem in a rigorous way and give a brief overview regarding the existence and uniqueness of solutions, as well as conforming Galerkin discretizations in $H(\mathrm{curl})$.

Section 5.2. *A first-order method:* We introduce a first-order accurate method with mass lumping for which we also show the desired convergence rates.

Section 5.3. *A first-order Yee-like method:* We propose another first-order method that requires fewer degrees of freedom, which can be understood as a generalization of the finite difference time domain method.

Section 5.4. *A second-order method:* Here, we introduce a novel second-order accurate scheme with mass lumping and provide a full error analysis that confirms its second-order accuracy.

Section 5.5. *Time discretization:* We propose a second-order accurate, explicit time discretization that exploits the efficiency brought by mass lumping.

Section 5.6. *Numerical validation*: We conduct a numerical experiment of a traveling plane wave being scattered on a spherical boundary and compare the three errors of the three methods we proposed.

Section 5.7. *Discussion*: In the end, we also give a brief overview and propose further extensions.

5.1. Preliminaries

First, we assume that

(A1) $\Omega \subset \mathbb{R}^3$, is a bounded Lipschitz domain and $T > 0$ denotes the time horizon.

For the convenience of notation, we consider the second-order formulation of Maxwell's equations

$$\partial_{tt} E + \mathrm{curl}(\mathrm{curl}\, E) = 0 \qquad \text{in } \Omega \times (0, T), \tag{5.3}$$

which is obtained first by differentiating (5.1) with respect to t and substituting in (5.2), and further setting $\varepsilon = \mu = 1$ and $\sigma = 0$. For ease of presentation, we set the perfectly conducting boundary condition

$$E \times n = 0, \qquad \text{on } \partial\Omega \times (0,T), \tag{5.4}$$

as well as initial values defined by

$$E(0) = E_0, \quad \partial_t E(0) = F_0 \qquad \text{in } \Omega. \tag{5.5}$$

For the analytical setting, we again consider the standard Sobolev spaces $L^p(\Omega)$ and $W^{k,p}(\Omega)$ with their usual norms, as well as the Hilbert spaces $H^k(\Omega) = W^{k,2}(\Omega)$, see [**3, 56**]. For our analysis, we also require the Sobolev space

$$H_0(\mathrm{curl},\Omega) = \{v \in L^2(\Omega)^3 \;\mid\; \mathrm{curl}\, v \in L^2(\Omega)^3,\ n \times v = 0 \text{ on } \partial\Omega\}$$

with norm $\|v\|_{H(\mathrm{curl},\Omega)} = (\|v\|^2_{L^2(\Omega)} + \|\mathrm{curl}\, v\|^2_{L^2(\Omega)})^{1/2}$.

5.1.1. Well-posedness

Before we continue, let us make a few remarks regarding the well-posedness of weak solutions. By the variational principle, we have that any sufficiently smooth solution E of (5.3)–(5.5) also satisfies the identity

$$(\partial_{tt} E(t), \phi) + (\mathrm{curl}\, E(t), \mathrm{curl}\, \phi) = 0 \qquad \forall \phi \in H_0(\mathrm{curl},\Omega), \;\; t > 0. \tag{5.6}$$

This follows by multiplying (5.3) by a test function ϕ, integrating over Ω, and finally using Stokes' theorem and the choice of initial values. Based on this formulation, we obtain the following result

Lemma 5.1.1 (Existence of weak solutions)**.**
Let (A1) hold. Then for any $E_0 \in H_0(\mathrm{curl},\Omega)$, $F_0 \in L^2(\Omega)$ and $f \in L^2(0,T,L^2(\Omega))$, the system (5.3)–(5.5) *admits a unique weak solution*

$$E \in L^2(0,T,H_0(\mathrm{curl},\Omega)) \cap H^1(0,T,L^2(\Omega)) \cap H^2(0,T,(H_0(\mathrm{curl},\Omega))')$$

depending continuously on the data.

In the next sections, we develop discretization techniques based on the weak form (5.6).

5.1.2. Finite element approximations

Let us briefly introduce our notation and then recall some available results from the literature for later comparison. Let $\mathcal{T}_h = \{K\}$ denote a geometrically conforming family of decompositions of the domain Ω into *tetrahedra*. For the rest of the chapter, we assume that

(A2) $\mathcal{T}_h$ is γ-shape regular partition of Ω into simplices, i.e. there exists a $\gamma > 0$ such that $\gamma h_K \le \rho_K \le h_K$ for all $K \in \mathcal{T}_h$, where h_K and ρ_K denote the diameter and inner circle radius of the element K, respectively.

Again, we define the broken Sobolev spaces of piecewise smooth functions

$$H^k(\mathcal{T}_h) = \{q \in L^2(\Omega) \,:\, q|_K \in H^k(K), \quad \forall K \in \mathcal{T}_h\}$$

with norm $\|v\|^2_{H^k(\mathcal{T}_h)} = \sum_{K\in\mathcal{T}_h} \|u\|^2_{H^k(K)}$. In addition, we write $P_k(K)$ for the space of polynomials of degree at most k over an element K and $P_k^h(K)$ for the space of homogeneous polynomials of degree k. Further, we define the local spaces

$$\mathcal{NC}_{k+1}(K) = P_{k+1}(K)^3 \qquad \text{or} \tag{5.7}$$

$$\mathcal{N}_k(K) = P_k(K)^3 \oplus \{x \times P_k^h(K)^3\}. \tag{5.8}$$

The family of spaces $\mathcal{N}_{k+1}$ was first introduced in [**92**], while the $\mathcal{NC}_k$ spaces originate from [**90**]. The notations we used adhere to the definition in [**19, Chapter 2.3.2**]. As global approximation spaces, we now consider

$$V_h = \{v_h \in H_0(\mathrm{curl}, \Omega) \,:\, E_h|_K \in \mathcal{M}_k(K), \quad \forall K \in \mathcal{T}_h\},$$

where either $\mathcal{M}_k(K) = \mathcal{NC}_{k+1}(K)$ or $\mathcal{M}_k(K) = \mathcal{N}_k(K)$. In addition, we introduce the canonical interpolation operators

$$\Pi_h : L^p(\Omega) \cap H(\mathrm{curl}, \Omega) \to \mathcal{M}_k(\mathcal{T}_h) \cap H(\mathrm{curl}, \Omega), \tag{5.9}$$

which satisfies the estimates

$$\begin{aligned} \|u - \Pi_h u\|_{H^\ell(\mathcal{T}_h)} &\leq \begin{cases} Ch_K^{r+1-\ell}\|u\|_{H^{r+1}(\mathcal{T}_h)} & \text{if } \mathcal{M}_k(\mathcal{T}_h) = \mathcal{NC}_{k+1}(\mathcal{T}_h), \\ Ch_K^{r-\ell}\|u\|_{H^{r+1}(\mathcal{T}_h)} & \text{if } \mathcal{M}_k(\mathcal{T}_h) = \mathcal{N}_k(\mathcal{T}_h), \end{cases} \\ \|\mathrm{curl}(u - \Pi_K u)\|_{L^2(\mathcal{T}_h)} &\leq Ch_K^r \|\mathrm{curl}\, u\|_{H^r(\mathcal{T}_h)}, \end{aligned} \tag{5.10}$$

for $1 \leq \ell \leq r \leq k$. Furthermore, we have the commuting diagram property

$$\mathrm{curl}\, \Pi_h u = \Pi_h^{\#} \mathrm{curl}\, u, \tag{5.11}$$

where $\Pi_h^{\#}$ is the $H(\mathrm{div})$ interpolation operator for RT_k if $\mathcal{M}_k(\mathcal{T}_h) = \mathcal{N}_k(\mathcal{T}_h)$ or BDM_{k+1} if $\mathcal{M}_k(\mathcal{T}_h) = \mathcal{NC}_{k+1}(\mathcal{T}_h)$. We refer to [**19, Section 2.5**] for a more in-depth discussion. With this, we can now introduce the following discrete scheme

Problem 5.1.2 (Exact Galerkin approximation)**.**
For $E_h(0) = \Pi_h E(0)$ and $\partial_t E_h(0) = \Pi_h \partial_t E(0)$ find $E_h \in C^2([0,T], V_h)$ such that

$$(\partial_{tt} E_h(t), \phi_h) + (\mathrm{curl}\, E_h(t), \mathrm{curl}\, \phi_h) = 0 \qquad \forall \phi_h \in V_h, \;\; t > 0. \tag{5.12}$$

After choosing a set of basis functions, the algebraic realization of (5.12) leads to the ordinary differential equation $\mathsf{M}\partial_{tt}\mathsf{e} + \mathsf{K}\mathsf{e} = 0$. By rewriting this as a first-order system, we can verify the existence of a unique solution by Picard-Lindelöff, see the discussion after Problem 4.1.2 in the previous chapter. As a result, we have that Problem 4.1.2 is well-posed. Furthermore, the solution satisfies the following convergence result.

Lemma 5.1.3 (Exact Galerkin discretization)**.**
Let (A1) and (A2) hold and let E be a sufficiently smooth solution of (5.3)–(5.5)*. Further, let E_h be the corresponding solution of Problem 5.1.2. Then*

$$\|\partial_t(E - E_h)\|_{L^\infty(0,T;L^2(\Omega))} + h\|\mathrm{curl}(E - E_h)\|_{L^\infty(0,T;L^2(\Omega))} \leq C_1(E)h^{k+1}.$$

If $\mathcal{M}_k(K) = \mathcal{N}_k(K)$, then also $\|\mathrm{curl}(E - E_h)\|_{L^\infty(0,T;L^2(\Omega))} \leq C_2(E)h^{k+1}$.

This result and further analysis of (mixed) finite element methods for Maxwell's equations were conducted by Monk in a series of publications [**80, 81, 84, 85, 83**]. As for the wave equation in $H(\mathrm{div})$, we observe that the convergence result is quasi-optimal for our choice of spaces.

In the next sections, we introduce mass lumping techniques that will enable the efficient realization of explicit schemes in time.

5.2. A first-order method

Using the same idea to develop the method in Section 4.2, we choose the space

$$V_h = \{\phi_h \in H_0(\mathrm{curl}, \Omega) \ : \ \phi_h|_K \in \mathcal{NC}_1(K), \quad \forall K \in \mathcal{T}_h\} \tag{5.13}$$

in conjunction with the quadrature formula

$$Q_K(f) \coloneqq \frac{|K|}{4} \sum_{k=1}^{4} f(v_k^K), \tag{5.14}$$

where v_k^K denotes the k-th vertex of K, see Figure 5.1 for details. This quadrature formula

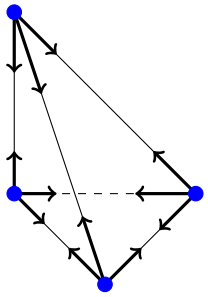

Figure 5.1.: Depiction of the degrees of freedom of the first order element $\mathcal{NC}_1(\widehat{K})$ on the unit tetrahedron as well as the location of the quadrature points (blue).

then induces the global inexact scalar product

$$(u, w)_h \coloneqq \sum\nolimits_{K \in \mathcal{T}_h} Q_K(u \cdot w). \tag{5.15}$$

Furthermore, we will write Π_h for the standard interpolation operator on V_h introduced in (5.9)–(5.10). With this, we can now formulate our inexact Galerkin approximation

Problem 5.2.1 (First order inexact Galerkin approximation)**.**
For $E_h(0) = \Pi_h E(0)$ and $\partial_t E_h(0) = \Pi_h \partial_t E(0)$ find $E_h \in C^2([0,T], V_h)$ such that

$$(\partial_{tt} E_h(t), \phi_h)_h + (\operatorname{curl} E_h(t), \operatorname{curl} \phi_h) = 0 \qquad \forall \phi_h \in V_h, \;\; t > 0. \tag{5.16}$$

By elementary computations, we can verify that $Q_K(\mathcal{K}^{-1} u_h \cdot v_h)$ defines a scalar product on $\mathcal{NC}_1(K)$. By scaling arguments, we obtain the equivalence property

$$c_1(\phi_h, \phi_h) \leq (\phi_h, \phi_h)_h \leq c_2(\phi_h, \phi_h) \qquad \forall \phi_h \in V_h. \tag{5.17}$$

This shows that for any choice of basis functions, the mass matrix M is regular, and by extension, this proves the well-posedness of Problem 5.2.1; see the discussion for the exact method. We then obtain the following convergence result.

Theorem 5.2.2 (First order convergence)**.**
Let (A1) and (A2) hold and let E be a sufficiently smooth solution of (5.3)–(5.5). Further, let E_h be the corresponding solution of Problem 5.1.2. Then

$$\|\partial_t (E - E_h)\|_{L^\infty(0,T;L^2(\Omega))} + \|\operatorname{curl}(E - E_h)\|_{L^\infty(0,T;L^2(\Omega))} \leq C(E,T) h,$$

where

$$C(E,T) = c\Big(\|\operatorname{curl} E\|^2_{L^\infty(0,T,H^1(\mathcal{T}_h))} + \|\operatorname{curl} \partial_t E\|^2_{L^1(0,T,H^1(\mathcal{T}_h))} + \|\partial_{tt} E\|^2_{L^1(0,T,H^1(\mathcal{T}_h))} \Big)$$

for a constant c independent of the mesh size.

Proof. First, we define the local and global quadrature errors

$$\sigma_K(\phi, \psi) \coloneqq Q_K(\phi \cdot \psi) - (\phi, \psi)_K \qquad \text{and} \qquad \sigma_h(\phi, \psi) = \sum\nolimits_{K \in T_h} \sigma_K(\phi, \psi). \tag{5.18}$$

As always, we split the error into interpolation and discrete error parts.

$$E(t) - E_h(t) = (E(t) - \Pi_h E(t)) + (\Pi_h E(t) - E_h(t)) = -\eta(t) + \psi_h(t). \tag{5.19}$$

From (5.10), we derive the interpolation error estimate

$$\begin{aligned} &\|\partial_t \eta\|_{L^\infty(0,T,L^2(\Omega))} + \|\operatorname{curl} \eta(t)\|_{L^\infty(0,T,L^2(\Omega))} \\ &\qquad \leq ch \left(\|\partial_t E\|_{L^\infty(0,T,H^1(\mathcal{T}_h))} + \|\operatorname{curl} E\|_{L^\infty(0,T,H^1(\mathcal{T}_h))} \right). \end{aligned}$$

Let π_h^1 denote the L^2-projection defined in (2.27). Then the discrete error $\psi_h(t)$ satisfies

$$\begin{aligned} &(\partial_{tt} \psi_h(t), \phi_h) + (\operatorname{curl} \psi_h(t), \operatorname{curl} \phi_h) \\ &\qquad = (\partial_{tt} \eta(t), \phi_h) + (\operatorname{curl} \eta(t), \operatorname{curl} \phi_h) + \sigma_h(\Pi_h \partial_{tt} u(t), \phi_h), \end{aligned} \tag{5.20}$$

where subtracted (5.16) from (5.6) with $\phi = \phi_h$. Next, choose $\phi_h \coloneqq \partial_t \psi_h(t)$. Then

$$(\partial_{tt} \psi_h(t), \partial_t \psi_h(t)) + (\operatorname{curl} \psi_h(t), \partial_t \operatorname{curl} \psi_h(t)) = \frac{1}{2} \frac{d}{dt} \left(\|\partial_t \psi_h(t)\|_h^2 + \|\operatorname{curl} \psi_h(t)\|^2_{L^2(\Omega)} \right).$$

Using this identity and choosing $\phi_h \coloneqq \partial_t \psi_h(t)$ in (5.20), integrating from 0 to t and using the definition of the initial values $\psi_h(0) = \partial_t \psi_h(0) = 0$ yields

$$\begin{aligned}
&\frac{1}{2}\left(\|\partial_t \psi_h(t)\|_h^2 + \|\operatorname{curl} \psi_h(t)\|_{L^2(\Omega)}^2\right) \\
&\quad = \int_0^t (\partial_{tt}\eta(s), \partial_t \psi_h(s)) + (\operatorname{curl} \eta(s), \operatorname{curl} \partial_t \psi_h(s)) + \sigma_h(\pi_h^1 \partial_{tt} u(s), \partial_t \psi_h(s)) \\
&\quad = (i) + (ii) + (iii).
\end{aligned}$$

The first term can be estimated via Cauchy-Schwarz and Young's inequality by

$$\begin{aligned}
(i) &\le \int_0^t \|\partial_{tt}\eta(s)\|_{L^2(\Omega)} \|\partial_t \psi_h(s)\|_{L^2(\Omega)}\, ds \le \|\partial_{tt}\eta\|_{L^1(0,t,L^2(\Omega))}^2 + \frac{1}{4c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2 \\
&\le Ch^2 \|\partial_{tt} u\|_{L^1(0,t,H^1(\mathcal{T}_h))}^2 + \frac{1}{4c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.
\end{aligned}$$

For the second term, we apply integration by parts and obtain

$$(ii) = \int_0^t \frac{d}{dt}(\operatorname{curl} \eta(s), \operatorname{curl} \psi_h(s))\, ds - \int_0^t (\operatorname{curl} \partial_t \eta(s), \operatorname{curl} \psi_h(s))\, ds = (ii_a) + (ii_b).$$

For (ii_a), since curl is linear we have $\operatorname{curl} \psi_h(0) = 0$ and

$$\begin{aligned}
(ii_a) &= (\operatorname{curl} \eta(t), \operatorname{curl} \psi_h(t)) \le \|\operatorname{curl} \eta(t)\|_{L^2(\Omega)}^2 + \frac{1}{4}\|\operatorname{curl} \psi_h(t)\|_{L^2(\Omega)}^2 \\
&\le Ch^2 \|\operatorname{curl} u\|_{L^\infty(0,t,H^1(\mathcal{T}_h))}^2 + \frac{1}{4}\|\operatorname{curl} \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.
\end{aligned}$$

The term (ii_b) can again be estimated by (5.10) to

$$(ii_b) \le Ch^2 \|\operatorname{curl} \partial_t u\|_{L^1(0,t,H^1(\mathcal{T}_h))}^2 + \frac{1}{4}\|\operatorname{curl} \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.$$

For (iii), we use the exactness of the quadrature rule for piecewise linear functions and conclude that

$$\begin{aligned}
(ii) &= \int_0^t \sigma_h(\Pi_h \partial_{tt} u(s), \partial_t \psi_h(s))\, ds = \int_0^t \sigma_h(\Pi_h \partial_{tt} u(s) - \pi_h^0 \partial_{tt} u(s), \partial_t \psi_h(s))\, ds \\
&\le \int_0^t \|\Pi_h \partial_{tt} u(s) - \pi_h^0 \partial_{tt} u(s)\|_{L^2(\Omega)} \|\partial_t \psi_h(s)\|_{L^2(\Omega)}\, ds \\
&\le \|\Pi_h \partial_{tt} u - \pi_h^0 \partial_{tt} u\|_{L^1(0,t,L^2(\Omega))}^2 + \frac{1}{4c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2 \\
&\le ch^2 \|\partial_{tt} u\|_{L^1(0,t,H^1(\mathcal{T}_h))}^2 + \frac{1}{4c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.
\end{aligned}$$

Summing all the terms and using the norm equivalence from (5.17), we obtain

$$\begin{aligned}
&\frac{1}{c_1}\|\partial_t \psi_h(t)\|_{L^2(\Omega)}^2 + \|\operatorname{curl} \psi_h(t)\|_{L^2(\Omega)}^2 \\
&\qquad\qquad \le C(u,t)h^2 + \frac{1}{2c_1}\|\partial_t \psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2 + \frac{1}{2}\|\operatorname{curl} \psi_h(t)\|_{L^2(\Omega)}^2.
\end{aligned}$$

Taking the maximum over all t and subsequently absorbing the last two terms by the left-hand side yields the L^∞-estimate. The main result of the Theorem now follows by combining the two bounds for the discrete and interpolation error. □

Remark 5.2.3. Note that we observe a loss in convergence in the term $\|\partial_t(E - E_h)\|_{L^2(\Omega)}$ in comparison to the exact method, see Lemma 5.1.3, which is due to the quadrature error. This is the same behavior as we already observed for the wave equation in $H(\mathrm{div})$. In addition, note that the error analysis is almost identical to the proof of Theorem 4.2.2, with the exception of the term $(\operatorname{curl}\eta(t), \operatorname{curl}\phi_h)$, which is not equal to zero here, in comparison to the $H(\mathrm{div})$ case. This is because unlike $H(\mathrm{div})$, the $H(\mathrm{curl})$ interpolation Π_h does not commute with the L^2-orthogonal projection as in (2.30). As a consequence, we had an additional term to estimate.

5.3. A first-order Yee-like method

The method is motivated by the fact that the first order method with mass lumping based on $\mathcal{NC}_1$ in Section 5.2 is efficient, but not optimal in terms of degrees of freedom. For optimality, only one degree of freedom per edge should be utilized. This would then correspond to a discretization in $\mathcal{N}_0$. In [**54**], the authors indeed proposed such a discretization method with mass lumping but had to ultimately add three additional degrees of freedom per element to guarantee the block-diagonal structure. Unfortunately, no optimal mass lumping procedures are known for this space on simplicial grids.

In [**20**] and later in [**78**], the authors proposed a geometric construction of a diagonal pseudo mass-matrix. Unfortunately, this construction is constrained to acute meshes and can, in principle, at most be extended to Delaunay grids.

Another idea is to give up the pursuit of a (block) diagonal mass matrix, and instead look for a sparse representation of the inverse of the mass matrix. This idea gained a lot of momentum in the last years, especially with the increased interest in discrete exterior calculus, see [**33**] and more recently [**96, 76**] for constructions on tetrahedra. The disadvantage here is that the construction is purely algebraic and because of this, convergence results are few and far between.

In this section, we pursue the construction of such a sparse inverse method that is closely related to a finite element approximation. We show how to derive the method algebraically and prove first-order convergence for general meshes. In comparison to the methods proposed in Section 5.2, we manage to reduce the number of degrees of freedom by half.

5.3.1. Preliminaries

For the construction, we consider the spaces

$$V_h = \{v_h \in H_0(\operatorname{curl}, \Omega) \ : \ v_h|_K \in \mathcal{NC}_1(K), \quad \forall K \in \mathcal{T}_h\},$$
$$\widetilde{V}_h = \{\widetilde{v}_h \in H_0(\operatorname{curl}, \Omega) \ : \ \widetilde{v}_h|_K \in \mathcal{N}_0(K), \quad \forall K \in \mathcal{T}_h\}.$$

Let us denote by $E_h : \widetilde{V}_h \to V_h$ the natural embedding and by $\widetilde{\Pi}_h : V_h \to \widetilde{V}_h$ the restriction of the canonical interpolation operator for $\widetilde{V}_h$. We can verify that $\widetilde{\Pi}_h \circ E_h = id_{\widetilde{V}_h}$ holds. From the local definition of the embedding and the projection, it is clear that these relations also hold on the element level.

Now let $\{\Phi_k\} \subset V_h$ denote the basis we will construct in Section 5.3.4 below, which is suitable for mass lumping. Further, let P denote the matrix representation of the projection $\widetilde{\Pi}_h$. In Section 5.3.4, we will also define a corresponding basis $\{\widetilde{\Phi}_k\} \subset \widetilde{V}_h$, again constructed locally, with the following properties:

(P1) $\mathsf{P}^\top$ is the matrix representation of the embedding E_h and $\mathsf{P}\mathsf{P}^\top = id$.

(P2) Let n be the number of edges. Then $\mathsf{P} \in \mathbb{R}^{n \times 2 \cdot n}$ is block diagonal with 1×2 blocks.

These properties immediately follow from the explicit construction in Section 4.3.4, and we assume them to be true in the sequel.

5.3.2. Construction on the algebraic level

By choosing the basis functions in accordance with Section 5.3.4 below, Problem 5.2.1 reduces to an ordinary differential equation of the form

$$\mathsf{M}_\mathsf{h} \partial_{tt} \mathsf{e}(t) + \mathsf{K}\mathsf{e}(t) = 0, \tag{5.21}$$

where M_h is block-diagonal. We also write M for the exact mass matrix. From the definition of P, we can directly infer that

$$\widetilde{\mathsf{M}} = \mathsf{P}\mathsf{M}\mathsf{P}^\top \qquad \text{and} \qquad \widetilde{\mathsf{K}} = \mathsf{P}\mathsf{K}\mathsf{P}^\top$$

are the corresponding mass and stiffness matrices on $\widetilde{V}_h$. In addition, the projection $\widetilde{\Pi}_h$ satisfies the identity $\operatorname{curl}(\phi_h - \widetilde{\Pi}_h \phi_h) = 0$ for all $\phi_h \in V_h$, since $\operatorname{curl} V_h = \operatorname{curl} \widetilde{V}_h$. Then $\mathsf{K}_{ij} = (\operatorname{curl} \Phi_j, \operatorname{curl} \Phi_i) = (\operatorname{curl} \widetilde{\Pi}_h \Phi_j, \operatorname{curl} \widetilde{\Pi}_h \Phi_i) = (\mathsf{P}^\top \widetilde{\mathsf{K}} \mathsf{P})_{ij}$, which implies

$$\mathsf{K} = \mathsf{P}^\top \widetilde{\mathsf{K}} \mathsf{P}. \tag{5.22}$$

Lemma 5.3.1 (Yee-like scheme)**.**
Let (A1) and (A2) hold and let e *denote the solution to* (5.21)*. Then* $\widetilde{\mathsf{e}}(t) = \mathsf{P}\mathsf{e}(t)$ *solves the algebraic system*

$$\widetilde{\mathsf{M}}_\mathsf{h} \partial_{tt} \widetilde{\mathsf{e}}(t) + \widetilde{\mathsf{K}} \widetilde{\mathsf{e}}(t) = 0, \tag{5.23}$$

where $\widetilde{\mathsf{M}}_\mathsf{h} \coloneqq (\mathsf{P}\mathsf{M}_\mathsf{h}^{-1}\mathsf{P}^\top)^{-1}$.

Proof. We compute

$$\begin{aligned} \partial_{tt} \widetilde{\mathsf{e}}(t) &= \mathsf{P} \partial_{tt} \mathsf{e}(t) = -\mathsf{P}\mathsf{M}_\mathsf{h}^{-1} \mathsf{K}\mathsf{e}(t) = -\mathsf{P}\mathsf{M}_\mathsf{h}^{-1}\mathsf{P}^\top \widetilde{\mathsf{K}} \mathsf{P}\mathsf{e}(t) \\ &= -(\mathsf{P}\mathsf{M}_\mathsf{h}^{-1}\mathsf{P}^\top) \widetilde{\mathsf{K}} \widetilde{\mathsf{e}}(t), \end{aligned}$$

where we used the system (5.21) and (5.22). This immediately concludes the proof. □

Remark 5.3.2. At its core, (5.23) is essentially a projected version of (5.21). By conducting a comparison to the exact implementation of the $\mathcal{N}_0$ method using the same basis functions, we attain the algebraic system

$$\widetilde{\mathsf{M}}\partial_{tt}\widehat{\mathsf{e}}(t) + \widetilde{\mathsf{K}}\widehat{\mathsf{e}}(t) = 0.$$

We infer that (5.23) has the same stiffness matrix, but a different mass matrix. In comparison to $\widetilde{\mathsf{M}}$, we see that $\widetilde{\mathsf{M}}_\mathsf{h}$ is, in general, a full matrix, while its inverse $\widetilde{\mathsf{M}}_\mathsf{h}^{-1}$ is sparse, if the projection matrix P is sparse.

5.3.3. Interpretation in function space

The vector $\widetilde{\mathsf{e}}(t)$ from solving (5.23) can be reinterpreted as an element $\widetilde{E}_h(t) \in \widetilde{V}_h$ by

$$\widetilde{E}_h(t) = \sum\nolimits_k \widetilde{\mathsf{e}}_k(t)\widetilde{\Phi}_k.$$

In remains to check how well $\widetilde{E}_h(t)$ approximates the true solution $E(t)$. As we will see, the convergence is inherited in full from the first order method, see Theorem 5.2.2.

Theorem 5.3.3. *Let (A1)–(A2) and (P1)–(P1) hold. Moreover, let $\widetilde{E}_h(t) \in \widetilde{V}_h$ denote the finite element solution associated to $\widetilde{\mathsf{e}}(t)$ from* (5.23) *and let $E(t)$ be the solution to* (5.3)–(5.5)*. Then*

$$\|\partial_t(E - \widetilde{E}_h)\|_{L^\infty(0,TL^2(\Omega))} + \|\operatorname{curl}(E - \widetilde{E}_h)\|_{L^\infty(0,T,L^2(\Omega))} \leq C(E)h,$$

where

$$C(E) = c\Big(\|\operatorname{curl} E\|^2_{L^\infty(0,T,H^1(\mathcal{T}_h))} + \|\operatorname{curl}\partial_t E\|^2_{L^1(0,T,H^1(\mathcal{T}_h))} + \|\partial_{tt} E\|^2_{L^1(0,T,H^1(\mathcal{T}_h))}\Big)$$

for a constant c independent of the mesh size.

Proof. We proceed in a similar fashion as in the last section. We again split the error into interpolation and a discrete error component.

$$E(t) - \widetilde{E}_h(t) = (E(t) - \widetilde{\Pi}_h E(t)) + (\widetilde{\Pi}_h E(t) - \widetilde{E}_h(t)) = -\eta(t) + \psi_h(t), \tag{5.24}$$

where $\widetilde{\Pi}_h$ is again the interpolation operator onto $\widetilde{V}_h$. According to (2.28), the interpolation error can be estimated by

$$\|\partial_t\eta\|_{L^\infty(0,T,L^2(\Omega))} + \|\operatorname{curl}\eta\|_{L^\infty(0,T,L^2(\Omega))} \leq C'(E,T)h.$$

For the discrete error component, we can compute

$$\begin{aligned}\|\widetilde{\Pi}_h\partial_t E(t) - \partial_t\widetilde{E}_h(t)\|_{L^2(\Omega)} &= \|\widetilde{\Pi}_h\partial_t E(t) - \widetilde{\Pi}_h\partial_t E_h(t)\|_{L^2(\Omega)} \\ &= \|\widetilde{\Pi}_h(\widetilde{\Pi}_h\partial_t E(t) - \partial_t E_h(t))\|_{L^2(\Omega)} \\ &\leq c\|\widetilde{\Pi}_h\partial_t E(t) - \partial_t E_h(t)\|_{L^2(\Omega)},\end{aligned}$$

where the L^2 estimate is due to the discrete input. Similarly, we also evaluate

$$\begin{aligned}\|\mathrm{curl}(\widetilde{\Pi}_h E(t) - \widetilde{E}_h(t))\|_{L^2(\Omega)} &= \|\mathrm{curl}(\widetilde{\Pi}_h E(t) - \widetilde{\Pi}_h E_h(t))\|_{L^2(\Omega)} \\ &= \|\Pi_h^{\#}(\mathrm{curl}\,\Pi_h E(t) - \mathrm{curl}\,E_h(t))\|_{L^2(\Omega)} \\ &\leq c\|\mathrm{curl}\,(\Pi_h E(t) - E_h(t))\|_{L^2(\Omega)},\end{aligned}$$

where $\Pi_h^{\#}$ denotes the $H(\mathrm{div})$ interpolation operator, see (5.11). Combining these estimates with the results of Theorem 5.2.2 completes the proof. □

Remark 5.3.4. Unlike for the wave equation in $H(\mathrm{div})$, we do not obtain equality in the estimate for the curl error component. This is also observed in numerical experiments.

5.3.4. Construction on tetrahedral meshes

In this section, we construct explicit basis functions for $\widetilde{V}_h$ and V_h on tetrahedra. In the following, we will use the notation $(a;b) = (a,b)^\top$ to describe column vectors. Let

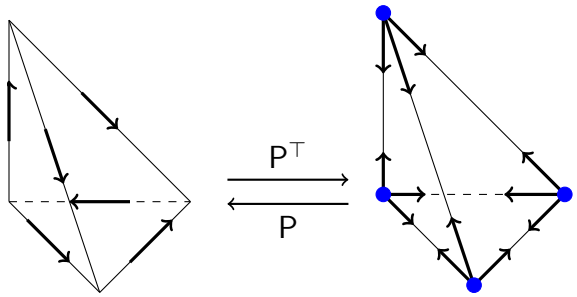

Figure 5.2.: Visual representation of the action of the projection matrix P mapping basis functions from $\mathcal{NC}_1(\widehat{K})$ to $\mathcal{N}_0(\widehat{K})$ and back.

$$\widehat{K} = \mathrm{span}\{(0;0;0),(1;0;0),(0;1;0),(0;0;1)\}.$$

denote the reference tetrahedron. On $\widehat{K}$, we then consider the functions

$$\begin{aligned}
&\widehat{\varphi}_1 = (1-x-y-z;0;0) && \widehat{\varphi}_2 = (x;x;x) && \widehat{\varphi}_3 = (0;1-x-y-z;0)\\
&\widehat{\varphi}_4 = (y;y;y) && \widehat{\varphi}_5 = (0;0;1-x-y-z) && \widehat{\varphi}_6 = (z;z;z)\\
&\widehat{\varphi}_7 = (0;x;0) && \widehat{\varphi}_8 = (-y;0;0) && \widehat{\varphi}_9 = (0;0;y)\\
&\widehat{\varphi}_{10} = (0;-z;0) && \widehat{\varphi}_{11} = (z;0;0) && \widehat{\varphi}_{12} = (0;0;-x),
\end{aligned}$$

which define a basis for $\mathcal{NC}_1(\widehat{K})$ on the reference tetrahedron. Moreover, we can check that exactly three functions are non-zero at each of the four vertices. By employing the vertex rule for the assembly, we observe that the local mass matrix is comprised of four 3×3 blocks. This structure remains intact even after transformation to the physical element by the covariant Piola mapping, see Section A.1. As a consequence, the global mass matrix is also block diagonal, with each block corresponding to a mesh vertex, and size determined by the degree of the vertex.

Furthermore, the functions

$$\begin{aligned}
\widehat{\Phi}_1 &= \frac{\sqrt{2}}{2}(\widehat{\varphi}_1 + \widehat{\varphi}_2) &&= \frac{\sqrt{2}}{2}(1-y-z; x; x),\\
\widehat{\Phi}_2 &= \frac{\sqrt{2}}{2}(\widehat{\varphi}_3 + \widehat{\varphi}_4) &&= \frac{\sqrt{2}}{2}(y; 1-x-z; y),\\
\widehat{\Phi}_3 &= \frac{\sqrt{2}}{2}(\widehat{\varphi}_5 + \widehat{\varphi}_6) &&= \frac{\sqrt{2}}{2}(z; z; 1-x-y),\\
\widehat{\Phi}_4 &= \frac{\sqrt{2}}{2}(\widehat{\varphi}_7 + \widehat{\varphi}_8) &&= \frac{\sqrt{2}}{2}(-y; x; 0),\\
\widehat{\Phi}_5 &= \frac{\sqrt{2}}{2}(\widehat{\varphi}_9 + \widehat{\varphi}_{10}) &&= \frac{\sqrt{2}}{2}(0; -z; y),\\
\widehat{\Phi}_6 &= \frac{\sqrt{2}}{2}(\widehat{\varphi}_{11} + \widehat{\varphi}_{12}) &&= \frac{\sqrt{2}}{2}(z; 0; -x)
\end{aligned}$$

define basis functions for $\mathcal{N}_0$, namely

$$\mathcal{N}_0(\widehat{K}) = P_0(\widehat{K})^3 \oplus x \times P_0(\widehat{K})^3 = \operatorname{span}\{\widehat{\Phi}_i,\ 1 \le i \le 6\}.$$

Locally, the projection P admits the structure

$$\mathsf{P}|_{\widehat{K}} = \begin{pmatrix}
\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}
\end{pmatrix}.$$

It is easy to check that $\mathsf{P}\mathsf{P}^\top = \mathsf{Id}$. Globally, P has the same structure as above, with exactly two entries $\frac{\sqrt{2}}{2}$ per row, with their location depending on the numbering of the degrees of freedom. In conclusion, (P1)–(P2) are satisfied.

5.3.5. Remarks

The extension of the method in 2D to triangles and quadrilaterals follows from the discussion in Section 4.3.5 by a simple rotation of the basis functions by $\pi/2$. On hexahedra, a similar technique can be devised with one and two degrees of freedom per edge.

For now, the method is not directly applicable when a non-trivial electric conductivity σ is present. Let us briefly explain why this is the case. Consider the following non-trivial form of Maxwell's equations:

$$\varepsilon \partial_{tt} E + \sigma \partial_t E + \operatorname{curl}(\mu^{-1} \operatorname{curl} E) = 0 \qquad \text{in } \Omega,\ t > 0. \tag{5.25}$$

After semi-discretization of (5.25) using the first order method from Section 5.2, we obtain the algebraic system

$$\widetilde{\mathsf{M}}_{\mathsf{h},\varepsilon} \partial_{tt} \widetilde{\mathsf{e}}(t) + \widetilde{\mathsf{M}}_{\mathsf{h},\sigma} \partial_t \widetilde{\mathsf{e}}(t) + \widetilde{\mathsf{K}}_\mu \widetilde{\mathsf{e}}(t) = 0.$$

Using the same construction as in the proof of Lemma 5.3.1, we can compute

$$\begin{aligned}\partial_{tt}\mathsf{e}(t) = \mathsf{P}\partial_{tt}\widetilde{\mathsf{e}}(t) &= -\mathsf{P}\left(\widetilde{\mathsf{M}}_{\mathsf{h},\varepsilon}^{-1}\widetilde{\mathsf{K}}_{\mu}\widetilde{\mathsf{e}}(t) + \widetilde{\mathsf{M}}_{\mathsf{h},\varepsilon}^{-1}\widetilde{\mathsf{M}}_{\mathsf{h},\sigma}\widetilde{\partial}_t\mathsf{e}(t)\right)\\ &= -\mathsf{P}\widetilde{\mathsf{M}}_{\mathsf{h},\varepsilon}^{-1}\mathsf{P}^{\top}\mathsf{K}_{\mu}\mathsf{P}\widetilde{\mathsf{e}}(t) - \mathsf{P}\widetilde{\mathsf{M}}_{\mathsf{h},\varepsilon}^{-1}\widetilde{\mathsf{M}}_{\mathsf{h},\sigma}\widetilde{\partial}_t\mathsf{e}(t)\\ &= -(\mathsf{P}\widetilde{\mathsf{M}}_{\mathsf{h},\varepsilon}^{-1}\mathsf{P}^{\top})\mathsf{K}_{\mu}\mathsf{e}(t) - \mathsf{P}\widetilde{\mathsf{M}}_{\mathsf{h},\varepsilon}^{-1}\widetilde{\mathsf{M}}_{\mathsf{h},\sigma}\widetilde{\partial}_t\mathsf{e}(t).\end{aligned}$$

For now, it is unclear how to project the second term in such a way that $\partial_t\widetilde{\mathsf{e}}$ can be replaced by $\partial_t\mathsf{e}$. This impasse seems to be in line with results from [**20**], where authors also cite difficulties when dealing with the case $\sigma \neq 0$ in their "Yee-like" scheme.

5.4. A second-order method

In this section, we consider the second order extension of our first order method from Section 5.2. Inspired by the element proposed by Elmkies and Joly in [**55, Section 4**], we consider the following construction. Let $\lambda_{i,K}$, $i = 1,\dots,4$ denote the barycentric coordinates of a tetrahedral element K, and consider the four functions

$$\begin{aligned}w_{1,K} &= \lambda_{2,K}\lambda_{3,K}\lambda_{4,K}\nabla\lambda_{1,K},\\ w_{2,K} &= \lambda_{1,K}\lambda_{3,K}\lambda_{4,K}\nabla\lambda_{2,K},\\ w_{3,K} &= \lambda_{1,K}\lambda_{2,K}\lambda_{4,K}\nabla\lambda_{3,K},\\ w_{4,K} &= \lambda_{1,K}\lambda_{2,K}\lambda_{3,K}\nabla\lambda_{4,K}.\end{aligned} \tag{5.26}$$

Each of these functions $w_{i,K}$ can be associated to the face $f_{i,K}$ opposite to the vertex $v_{i,K}$; see Figure 5.3 for an illustration. Note that $w_{i,K}$ has zero tangential trace on ∂K and,

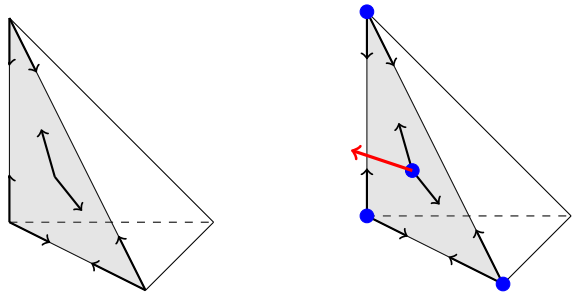

Figure 5.3.: Representation of the degrees of freedom for $\mathcal{N}_1(K)$ (left), and the degrees of freedom for $\mathcal{EJ}_1^*(K)$ (right), restricted to a single face of the tetrahedron. Blue dots are quadrature points. The associated weights are $\alpha = \frac{1}{40}, \beta = \frac{9}{40}$.

therefore, its extension by zero also lies in $H(\mathrm{curl},\Omega)$. Then the element proposed in [**55**] is defined as

$$\mathcal{EJ}_1(K) \coloneqq \mathcal{N}_1(K) \oplus \mathrm{span}\{w_{1,K}, w_{2,K}, w_{3,K}, w_{4,K}\} \subset P_3(K)^3.$$

In [**50**], we extensively analyzed the element $\mathcal{EJ}_1$ and showed that the method is only first-order convergent, in general. More precisely, the method is second-order convergent

whenever the solution is divergence-free. This condition is not only theoretical, and we can infer its strict requirement from numerical experiments. As a fix, we proposed a modification of this element, which leads to a second-order convergent method without any restrictions. For this, we consider the altered basis function

$$w^*_{4,K} = \lambda_{1,K}\lambda_{2,K}\lambda_{3,K}(1 + \lambda_{2,K} - \lambda_{1,K})\nabla\lambda_{4,K}. \tag{5.27}$$

Following [**35, 55**], we can now define

$$\mathcal{ES}^*_1(K) = \mathcal{N}_1(K) \oplus \mathcal{B}(K) = \mathcal{N}_1(K) \oplus \operatorname{span}\{w_{1,K}, w_{2,K}, w_{3,K}, w^*_{4,K}\}. \tag{5.28}$$

We then consider the $H(\mathrm{curl})$-conforming finite element subspace

$$V_h = \{\phi_h \in H_0(\mathrm{curl}, \Omega) \ : \ \phi_h|_K \in \mathcal{ES}^*_1(K), \quad \forall K \in \mathcal{T}_h\}. \tag{5.29}$$

In addition, we also consider the quadrature formula

$$Q_K(f) = |K| \left(\sum_{i=1}^{4} \alpha\, f(v_i) + \sum_{i=1}^{4} \beta\, f(m_i) \right). \tag{5.30}$$

where v_i and m_i represent the vertices and the face midpoints, respectively, while α and β denote the corresponding weights; we refer to the caption of Figure 5.3 for their exact values. This quadrature formula then induces a global inexact scalar product by

$$(E, \phi)_h \coloneqq \sum\nolimits_{K\in\mathcal{T}_h} Q_K(E \cdot \phi). \tag{5.31}$$

With this, we can now define the following discrete scheme. For the analysis, we will also use the standard interpolation operator Π_h on $\mathcal{N}_1$ defined in (5.9)–(5.10).

Problem 5.4.1 (Second order inexact Galerkin approximation)**.**
For $E_h(0) = \Pi_h E(0)$ and $\partial_t E_h(0) = \Pi_h \partial_t E(0)$ find $E_h \in C^2([0,T], V_h)$ such that

$$(\partial_{tt} E_h(t), \phi_h)_h + (\operatorname{curl} E_h(t), \operatorname{curl} \phi_h) = 0 \qquad \forall \phi_h \in V_h, \ \ t > 0. \tag{5.32}$$

By elementary computations, we can again verify that $Q_K(\mathcal{K}^{-1}\phi_h \cdot \psi_h)$ defines a scalar product on $\mathcal{ES}^*_1(K)$. By scaling arguments, we obtain the equivalence property

$$c_1(\phi_h, \phi_h) \le (\phi_h, \phi_h)_h \le c_2(\phi_h, \phi_h) \qquad \forall \phi_h \in V_h. \tag{5.33}$$

Well-posedness of Problem 5.1.2 follows by the same arguments used for the first order method. Before we show that the method is indeed second-order convergent, we need to first gather a few auxiliary results.

5.4.1. Auxiliary result

First, recall that

$$\mathcal{ES}^*_1(K) = \mathcal{N}_1(K) \oplus \mathcal{B}(K) \qquad \text{with} \qquad \mathcal{B}(K) = \operatorname{span}\{w_{1,K}, w_{2,K}, w_{3,K}, w^*_{4,K}\}.$$

For our analysis, we gather the following properties

Lemma 5.4.2 (Stable splitting)**.**
Let (A2) hold. Then

(i) $\dim(\mathcal{B}(K)) = \dim(\mathrm{curl}(\mathcal{B}(K))) = 4$.

(ii) The expression $\|\mathrm{curl}(\cdot)\|_{L^2(K)}$ *defines a norm on* $\mathcal{B}(K)$ *and*

$$\|\nabla \phi_K^{\mathcal{B}}\|_{L^2(K)} \leq C\|\mathrm{curl}(\phi_K^{\mathcal{B}})\|_{L^2(K)}, \qquad \forall \phi_K^{\mathcal{B}} \in \mathcal{B}(K).$$

(iii) $\|\mathrm{curl}(\phi_K^{\mathcal{B}})\|_{L^2(K)} \leq C\|\mathrm{curl}\,(\phi_K^{B} + \phi_K^{\mathcal{N}})\|_{L^2(K)}$, *for all* $\phi_K^{\mathcal{N}} \in \mathcal{N}_1(K)$.

The constants C *in these estimates only depend on the shape regularity of* K.

Proof. The first assumption follows by elementary computations. For (ii), we see from the dimension count in (i) that $\mathrm{curl}\,\phi_K^{\mathcal{B}} = 0$ iff $\phi_K^{\mathcal{B}} = 0$ for any $\phi_K^{\mathcal{B}} \in \mathcal{B}(K)$. This automatically yields the estimate as well. For (iii), let $\widehat{K}$ denote the reference tetrahedron and let $\widehat{q}_1 \in P_1(\widehat{K})^3$ and let $\widehat{q}_{\mathcal{B}} \in \mathrm{curl}(\mathcal{B}(\widehat{K}))$. In addition, let π_h^1 denote the L^2-projection onto $P_1(\widehat{K})^3$.

Assume now that $\widehat{q}_{\mathcal{B}} - \pi_h^1 \widehat{q}_{\mathcal{B}} = 0$. This implies that $\widehat{q}_{\mathcal{B}} = \pi_h^1 \widehat{q}_{\mathcal{B}} \in P_1(\widehat{K})$. Since $P_1(\widehat{K}) \not\subseteq \widehat{\mathrm{curl}}\,\mathcal{B}(\widehat{K})$, we infer that $\widehat{q}_{\mathcal{B}} = 0$. This implies that $\mathrm{Id} - \pi_h^1$ induces a norm on $\widehat{\mathrm{curl}}(\mathcal{B}(\widehat{K}))$, i.e. there exists a constant $c > 0$ such that

$$\|\widehat{q}_{\mathcal{B}} - \pi_h^1 \widehat{q}_{\mathcal{B}}\|^2_{L^2(\widehat{K})} \geq c\|\widehat{q}_{\mathcal{B}}\|^2_{L^2(\widehat{K})}.$$

By definition, we also know that $\|\widehat{q}_{\mathcal{B}} - \pi_h^1 \widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})} = \min_{\widehat{q}_1 \in P_1(\widehat{K})^3} \|\widehat{q}_1 + \widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})}$. We can then write

$$\|\widehat{q}_1 + \widehat{q}_{\mathcal{B}}\|^2_{L^2(\widehat{K})} \geq \|\widehat{q}_{\mathcal{B}} - \pi_h^1 \widehat{q}_{\mathcal{B}}\|_{L^2(\widehat{K})} \geq c\|\widehat{q}_{\mathcal{B}}\|^2_{L^2(\widehat{K})}.$$

By a scaling argument, together with assumption (A2), we also obtain

$$\|q_1 + q_{\mathcal{B}}\|^2_{L^2(K)} \geq c_\gamma \|q_{\mathcal{B}}\|^2_{L^2(K)}$$

for any $q_1 \in P_1(K)^3$ and $q_{\mathcal{B}} \in \mathrm{curl}(\mathcal{B}(K))$ with a constant c_γ depending only on the shape-regularity of $\mathcal{T}_h$. Lastly, note that $\mathrm{curl}(\mathcal{N}_1(K)) \subseteq P_1(K)^3$. □

5.4.2. Quadrature error

The error analysis of Problem 5.4.1 requires a detailed analysis of the quadrature. To simplify notation later on, we introduce the local and global quadrature errors

$$\sigma_{h,K}(u, v) \coloneqq Q_K(u \cdot v) - (u, v)_K \qquad \text{and} \qquad \sigma_h(u, v) = \sum\nolimits_{K \in T_h} \sigma_{h,K}(u, v), \tag{5.34}$$

where Q_K is defined as in (5.30). In the next two lemmas, we cover estimates on the quadrature error.

Lemma 5.4.3 (Quadrature error, part one)**.**
The quadrature rule (5.30) *integrates all polynomials in* $P_3(K) \cup \mathcal{B}(K)$ *exactly. Further, let (A1) and (A2) hold and let* π_h^1 *denote the* L^2*-orthogonal projection onto* $P_1(\mathcal{T}_h)^d$*. Then the quadrature error* (5.34) *satisfies*

$$\sigma_h(\pi_h^1 E, \phi_h) \leq Ch^2 \|E\|_{H^1(\mathcal{T}_h)} \|\operatorname{curl} \phi_h\|_{L^2(\Omega)} \qquad \forall \phi_h \in V_h.$$

Proof. The exactness property can be verified easily. Next, it suffices to show the inequality locally for every element $K \in \mathcal{T}_h$. We begin by splitting

$$\begin{aligned} \sigma_K(\pi_h^1 E, \phi_h) &= \sigma_T(\pi_h^1 E - \pi_h^0 E, \phi_h - \pi_h^2 \phi_h) \\ &\leq C \|\pi_h^1 E - \pi_h^0 E\|_{L^2(K)} \|\phi_h - \pi_h^2 \phi_h\|_{L^2(K)} \\ &\leq C' h_T^4 \|E\|_{H^1(K)} \|\nabla^3 \phi_h\|_{L^2(K)}, \end{aligned} \tag{5.35}$$

where we used the exactness of the quadrature formula in the first step, the norm equivalence property (5.33), as well as estimates for the projection π_h^k from (2.28). From the definition of $\mathcal{B}(K)$, we know that for any $\phi_h \in \mathcal{EJ}_1^*(K)$, we can write $\phi_h = \phi_h^{\mathcal{N}} \oplus \phi_h^{\mathcal{B}}$, where $\phi_h^{\mathcal{N}} \in \mathcal{N}_1(K) \subseteq P_2(K)^3$ and $\phi_h^{\mathcal{B}} \in \mathcal{B}(K)$. We then have

$$\begin{aligned} \|\nabla^3 \phi_h\|_{L^2(K)} &= \|\nabla^3 \phi_h^{\mathcal{B}}\|_{L^2(K)} \leq C_\gamma h_K^{-2} \|\nabla \phi_h^{\mathcal{B}}\|_{L^2(K)} \\ &\leq C_\gamma C'_\gamma h_K^{-2} \|\operatorname{curl} \phi_h^{\mathcal{B}}\|_{L^2(K)} \leq C_\gamma C'_\gamma C''_\gamma h_K^{-2} \|\operatorname{curl} (\phi_h^{\mathcal{B}} + \phi_h^{\mathcal{N}})\|_{L^2(K)} \\ &= C_\gamma C'_\gamma C''_\gamma h_K^{-2} \|\operatorname{curl} \phi_h\|_{L^2(K)}, \end{aligned} \tag{5.36}$$

where we used that $\nabla^3 \phi_h^{\mathcal{N}} = 0$ in the first step, the inverse inequality in the second, and identities (*ii*) and (*iii*) of Lemma 5.4.2. Combining (5.35) and (5.36) and summing over all elements yields the assertion. □

Corollary 5.4.4 (Quadrature error, part two)**.**
Let (A1) and (A2) hold and let E *be sufficiently smooth that*

$$C(E) = C_\gamma(\|E\|^2_{L^\infty(0,t,H^1(\mathcal{T}_h))} + \|E\|^2_{L^1(0,t,H^2(\mathcal{T}_h))} + \|\partial_t E\|^2_{L^1(0,t,H^1(\mathcal{T}_h))})$$

exits and let $\phi_h \in W^{1,\infty}([0,t], V_h)$ *with* $\phi_h(0) = 0$*. Then*

$$\int_0^t \sigma_h(\Pi_h E(s), \partial_t \phi_h(s))\, ds \leq C(E)h^4 + \frac{1}{4c_1} \|\partial_t \phi_h\|^2_{L^\infty(0,t,L^2(\Omega))} + \frac{1}{6} \|\operatorname{curl} \phi_h\|^2_{L^\infty(0,t,L^2(\Omega))}.$$

Proof. For ease of presentation, we suppress the time dependency. Then

$$\begin{aligned} \int_0^t \sigma_h(\Pi_h E, \partial_t \phi_h) &= \int_0^t \sigma_h(\Pi_h E - \pi_h^1 E, \partial_t \phi_h) + \int_0^t \sigma_h(\pi_h^1 E, \partial_t \phi_h) \\ &= \int_0^t \sigma_h(\Pi_h E - \pi_h^1 E, \partial_t \phi_h) + \int_0^t \frac{d}{dt} \sigma_h(\pi_h^1 E, \phi_h) - \int_0^t \sigma_h(\pi_h^1 \partial_t E, \phi_h) \\ &= (i) + (ii) + (iii). \end{aligned}$$

For the first term, we have

$$\begin{aligned}(i) &\leq \int_0^t \|\Pi_h E - \pi_h^1 E\|_{L^2(\Omega)} \|\partial_t \phi_h\|_{L^2(\Omega)} \leq \int_0^t Ch^2 \|E\|_{H^2(\mathcal{T}_h)} \|\partial_t \phi_h\|_{L^2(\Omega)} \\ &\leq C' h^4 \|E\|^2_{L^1(0,t,H^2(\mathcal{T}_h))} + \frac{1}{4c_1} \|\partial_t \phi_h\|^2_{L^\infty(0,t,L^2(\Omega))}.\end{aligned}$$

For the second term, we use $\phi_h(0) = 0$ and Lemma 5.4.3 and estimate

$$\begin{aligned}(ii) = \sigma_h(\pi_h^1 E(t), \phi_h(t)) &\leq Ch^2 \|E(t)\|_{H^1(\mathcal{T}_h)} \|\operatorname{curl} \phi_h(t)\|_{L^2(\Omega)} \\ &\leq C' h^4 \|E(t)\|^2_{H^1(\mathcal{T}_h)} + \frac{1}{12} \|\operatorname{curl} \phi_h(t)\|^2_{L^2(\Omega)} \\ &\leq C'' h^4 \|E\|^2_{L^\infty(0,t,H^1(\mathcal{T}_h))} + \frac{1}{12} \|\operatorname{curl} \phi_h\|^2_{L^\infty(0,t,L^2(\Omega))}.\end{aligned}$$

For the third term, we write

$$\begin{aligned}(iii) &= \int_0^t Ch^2 \|\partial_t E\|_{H^1(\mathcal{T}_h)} \|\operatorname{curl} \phi_h(t)\|_{L^2(\Omega)} \\ &\leq C' h^4 \|\partial_t E\|^2_{L^1(0,t,H^1(\mathcal{T}_h))} + \frac{1}{12} \|\operatorname{curl} \phi_h(t)\|^2_{L^\infty(0,t,L^2(\Omega))}.\end{aligned}$$

Summing all the terms yields the result. □

Remark 5.4.5. Unlike what we have seen before, the quadrature rule is not only exact for cubic polynomials, but it also integrates the functions from $\mathcal{B}(K)$ exactly. More precisely, the modified basis function from (5.27) was chosen specifically such that on the one hand it is integrated exactly by the quadrature rule, but also such that the assertions of Lemma 5.4.2 hold.

5.4.3. Error analysis

Having introduced all the necessary tools, we can now prove the following main result.

Theorem 5.4.6 (Second order convergence)**.**
Let (A1)–(A3) hold and let E be a sufficiently smooth solution of (5.3)–(5.5)*. Further, let E_h be the corresponding solution of Problem 5.4.1. Then*

$$\|\partial_t (E - E_h)\|_{L^\infty(0,T,L^2(\Omega))} + \|\operatorname{curl}(E - E_h)\|_{L^\infty(0,T,L^2(\Omega))} \leq C(E) h^2,$$

where

$$\begin{aligned}C(E) = c \Big(&\|\partial_{tt} E\|_{L^\infty(0,t,H^1(\mathcal{T}_h))} + \|\partial_{tt} E\|_{L^1(0,t,H^2(\mathcal{T}_h))} + \|\partial_{ttt} E\|_{L^1(0,t,H^1(\mathcal{T}_h))} \\ &+ \|\operatorname{curl} E\|_{L^\infty(0,T,H^2(\mathcal{T}_h))} + \|\operatorname{curl} \partial_t E\|_{L^1(0,T,H^2(\mathcal{T}_h))} \Big)\end{aligned}$$

for a constant c independent of the mesh size.

Proof. Let Π_h denote the interpolation operator onto $\mathcal{N}_1$ introduced in (5.9)–(5.10). We then split the error into an interpolation and a projection error component by

$$E(t) - E_h(t) = -(\Pi_h E(t) - E(t)) + (\Pi_h E(t) - E_h(t)) =: -\eta(t) + \psi_h(t).$$

The projection error $\eta(t)$ can be bounded using the bounds from (5.10) by

$$\begin{aligned}\|\partial_t \eta\|_{L^\infty(0,T,L^2(\Omega))} + \|\operatorname{curl}\eta\|_{L^\infty(0,T,L^2(\Omega))} \\ \leq Ch^2(\|\partial_t E\|_{L^\infty(0,T,H^2(\mathcal{T}_h))} + \|\operatorname{curl} E\|_{L^\infty(0,T,H^2(\mathcal{T}_h))}).\end{aligned}$$

We now turn to the interpolation error ψ_h. Due to the choice of initial values, we have $\psi_h(0) = 0$ and $\partial_t\psi_h(0) = 0$ and consequently, also $\operatorname{curl}\psi_h(0) = 0$. Evaluating (5.6) with $\phi = \phi_h$ and subtracting it from (5.32), we obtain

$$\begin{aligned}(\partial_{tt}\psi_h(t), \phi_h)_h + (\operatorname{curl}\psi_h(t), \operatorname{curl}\phi_h) \\ = (\partial_{tt}\eta(t), \phi_h) + (\operatorname{curl}\eta(t), \operatorname{curl}\phi_h) + \sigma_h(\Pi_h\partial_{tt}E(t), \phi_h).\end{aligned}$$

Choosing $\phi_h = \partial_t\psi_h(t)$ as test function and integrating from 0 to t further yields

$$\begin{aligned}&\frac{1}{2}\Big(\|\partial_t\psi_h(t)\|_h^2 + \|\operatorname{curl}\psi_h(t)\|_{L^2(\Omega)}^2\Big) \\ &= \int_0^t (\partial_{tt}\eta(s), \partial_t\psi_h(s)) + \int_0^t (\operatorname{curl}\eta(s), \operatorname{curl}\partial_t\psi_h(s)) - \int_0^t \sigma_h(\Pi_h\partial_{tt}E(s), \partial_t\psi_h(s)) \\ &= (i) + (ii) + (iii).\end{aligned}$$

The first term can be estimated via Cauchy-Schwarz and Young inequalities by

$$\begin{aligned}(i) &\leq \int_0^t \|\partial_{tt}\eta(s)\|_{L^2(\Omega)}\|\partial_t\psi_h(s)\|_{L^2(\Omega)}\,ds \leq \|\partial_{tt}\eta\|_{L^1(0,t,L^2(\Omega))}^2 + \frac{1}{4c_1}\|\partial_t\psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2 \\ &\leq Ch^4\|\partial_{tt}E\|_{L^1(0,t,H^2(\mathcal{T}_h))}^2 + \frac{1}{4c_1}\|\partial_t\psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.\end{aligned}$$

For the second term, we apply integration by parts and obtain

$$(ii) = \int_0^t \frac{d}{dt}(\operatorname{curl}\eta(s), \operatorname{curl}\psi_h(s))\,ds - \int_0^t (\operatorname{curl}\partial_t\eta(s), \operatorname{curl}\psi_h(s))\,ds = (ii_a) + (ii_b).$$

For (ii_a), since $\operatorname{curl}\psi_h(0) = 0$, we have

$$\begin{aligned}(ii_a) = (\operatorname{curl}\eta(t), \operatorname{curl}\psi_h(t)) &\leq c\|\operatorname{curl}\eta(t)\|_{L^2(\Omega)}^2 + \frac{1}{6}\|\operatorname{curl}\psi_h(t)\|_{L^2(\Omega)}^2 \\ &\leq Ch^4\|\operatorname{curl} E\|_{L^\infty(0,t,H^2(\mathcal{T}_h))}^2 + \frac{1}{6}\|\operatorname{curl}\psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.\end{aligned}$$

The term (ii_b) is estimated again in a standard way by

$$(ii_b) \leq Ch^4\|\operatorname{curl}\partial_t E\|_{L^1(0,t,H^2(\mathcal{T}_h))}^2 + \frac{1}{6}\|\operatorname{curl}\psi_h\|_{L^\infty(0,t,L^2(\Omega))}^2.$$

Using Lemma 5.4.4, we can estimate the third term by

$$(iii) \leq C(\partial_{tt}E)^2 h^4 + \frac{1}{4c_1}\|\partial_t \psi_h\|^2_{L^\infty(0,t,L^2(\Omega))} + \frac{1}{6}\|\mathrm{curl}\,\psi_h\|^2_{L^\infty(0,t,L^2(\Omega))},$$

with

$$C(\partial_{tt}E) = C\left(\|\partial_{tt}E\|^2_{L^\infty(0,t,H^1(\mathcal{T}_h))} + \|\partial_{tt}E\|^2_{L^1(0,t,H^2(\mathcal{T}_h))} + \|\partial_{ttt}E\|^2_{L^1(0,t,H^1(\mathcal{T}_h))}\right).$$

Summing all the terms and using the norm equivalence (5.33), we obtain

$$\begin{aligned} &\frac{1}{c_1}\|\partial_t \psi_h(t)\|^2_{L^2(\Omega)} + \|\mathrm{curl}\,\psi_h(t)\|^2_{L^2(\Omega)} \\ &\leq C(E)^2 h^4 + \frac{1}{2}\|\partial_t \psi_h\|^2_{L^\infty(0,t,L^2(\Omega))} + \frac{1}{2}\|\mathrm{curl}\,\psi_h\|^2_{L^\infty(0,t,L^2(\Omega))}. \end{aligned} \tag{5.37}$$

Taking the maximum over all t in (5.37) and subsequently absorbing the last two terms by the left-hand side yields the L^∞-estimate. The main result follows by adding the two results for the interpolation and discrete error component. □

5.4.4. Choice of basis functions

To complete the description of the mass-lumping method, let us briefly explain how to choose appropriate basis functions for the second order method such that the resulting mass matrix has the desired block-diagonal structure. For ease of presentation, we introduce the basis functions on the reference element $\widehat{K}$, mapped by covariant Piola transformation to the physical element K, see (A.1). In the following, we will use the notation $(a;b) = (a,b)^\top$ to describe column vectors. Let

$$\widehat{K} = \mathrm{span}\{(0;0;0),(1;0;0),(0;1;0),(0;0;1)\}$$

denote the reference tetrahedron. On $\widehat{K}$, we then consider the functions

$$\begin{array}{lll} \widehat{\varphi}_1 = (1-x-y-z;0;0) & \widehat{\varphi}_2 = (x;x;x) & \widehat{\varphi}_3 = (0;1-x-y-z;0) \\ \widehat{\varphi}_4 = (y;y;y) & \widehat{\varphi}_5 = (0;0;1-x-y-z) & \widehat{\varphi}_6 = (z;z;z) \\ \widehat{\varphi}_7 = (0;x;0) & \widehat{\varphi}_8 = (-y;0;0) & \widehat{\varphi}_9 = (0;0;y) \\ \widehat{\varphi}_{10} = (0;-z;0) & \widehat{\varphi}_{11} = (z;0;0) & \widehat{\varphi}_{12} = (0;0;-x). \end{array}$$

It is easy to verify that $\mathcal{NC}_1(\widehat{K}) = P_1(\widehat{K})^2 = \mathrm{span}\{\varphi_i,\ 1 \leq i \leq 12\}$, meaning that these functions form a basis for $\mathcal{NC}_1$. Moreover, we can check that together with the vertex rule (5.15), the local mass matrix is comprised of four 3×3 blocks, each of which can be associated with a vertex. As a result, for any given tetrahedral mesh, the global mass matrix is also block-diagonal, where the size of the blocks is independent of the mesh size. This concludes the choice of basis functions for the first order method with mass lumping.

Next, let us consider

$$\begin{aligned}
&\widehat{\phi}_{13} = (0; xz; -xy) && \widehat{\phi}_{14} = (-yz; 0; -xz)\\
&\widehat{\phi}_{15} = (-yz; -yz; -y(1-x-y)) && \widehat{\phi}_{16} = (yz; z(1-x-z); yz)\\
&\widehat{\phi}_{17} = (-z(1-y-z); -xz; -xz) && \widehat{\phi}_{18} = (xz; xz; x(1-x-y))\\
&\widehat{\phi}_{19} = (-xy; -x(1-x-z)); -xy) && \widehat{\phi}_{20} = (y(1-y-z); xy; xy)\\
&\widehat{\Phi}_{21} = (xyz; xyz; xyz) && \widehat{\Phi}_{22} = (yz(1-x-y-z); 0; 0)\\
&\widehat{\Phi}_{23} = (0; xy(1-x-y-z); 0) && \widehat{\Phi}_{24} = (0; 0; xy(1+x-y)(1-x-y-z)).
\end{aligned}$$

With the help of these functions, we introduce the following modified basis functions

$$\begin{aligned}
&\widehat{\Phi}_{13} = \widehat{\varphi}_{13} + \widehat{\Phi}_{23} - \widehat{\Phi}_{22} && \widehat{\Phi}_{14} = \widehat{\varphi}_{14} + \widehat{\Phi}_{21} - \widehat{\Phi}_{23} && \widehat{\Phi}_{15} = \widehat{\varphi}_{15} + \widehat{\Phi}_{23} - \widehat{\Phi}_{24}\\
&\widehat{\Phi}_{16} = \widehat{\varphi}_{16} + \widehat{\Phi}_{24} - \widehat{\Phi}_{22} && \widehat{\Phi}_{17} = \widehat{\varphi}_{17} + \widehat{\Phi}_{21} - \widehat{\Phi}_{24} && \widehat{\Phi}_{18} = \widehat{\varphi}_{18} + \widehat{\Phi}_{24} - \widehat{\Phi}_{23}\\
&\widehat{\Phi}_{19} = \widehat{\varphi}_{19} + \widehat{\Phi}_{22} - \widehat{\Phi}_{24} && \widehat{\Phi}_{20} = \widehat{\varphi}_{20} + \widehat{\Phi}_{24} - \widehat{\Phi}_{21}
\end{aligned}$$

and

$$\begin{aligned}
&\widehat{\Phi}_{1} = \widehat{\varphi}_{1} + \widehat{\Phi}_{23} - \widehat{\Phi}_{22} && \widehat{\Phi}_{2} = \widehat{\varphi}_{2} + \widehat{\Phi}_{21} - \widehat{\Phi}_{23} && \widehat{\Phi}_{3} = \widehat{\varphi}_{3} + \widehat{\Phi}_{23} - \widehat{\Phi}_{24}\\
&\Phi_{4} = \widehat{\varphi}_{4} + \widehat{\Phi}_{24} - \widehat{\Phi}_{22} && \widehat{\Phi}_{5} = \widehat{\varphi}_{5} + \widehat{\Phi}_{21} - \widehat{\Phi}_{24} && \widehat{\Phi}_{6} = \widehat{\varphi}_{6} + \widehat{\Phi}_{24} - \widehat{\Phi}_{23}\\
&\widehat{\Phi}_{7} = \widehat{\varphi}_{7} + \widehat{\Phi}_{22} - \widehat{\Phi}_{24} && \widehat{\Phi}_{8} = \widehat{\varphi}_{8} + \widehat{\Phi}_{24} - \widehat{\Phi}_{21} && \widehat{\Phi}_{9} = \widehat{\varphi}_{9} + \widehat{\Phi}_{22} - \widehat{\Phi}_{23}\\
&\widehat{\Phi}_{10} = \widehat{\varphi}_{10} + \widehat{\Phi}_{22} - \widehat{\Phi}_{24} && \widehat{\Phi}_{11} = \widehat{\varphi}_{11} + \widehat{\Phi}_{24} - \widehat{\Phi}_{21} && \widehat{\Phi}_{12} = \widehat{\varphi}_{18} + \widehat{\Phi}_{22} - \widehat{\Phi}_{23}\\
&\widehat{\Phi}_{13} = \widehat{\varphi}_{13} + \widehat{\Phi}_{22} - \widehat{\Phi}_{24} && \widehat{\Phi}_{14} = \widehat{\varphi}_{14} + \widehat{\Phi}_{24} - \widehat{\Phi}_{21} && \widehat{\Phi}_{15} = \widehat{\varphi}_{15} + \widehat{\Phi}_{22} - \widehat{\Phi}_{23}\\
&\widehat{\Phi}_{16} = \widehat{\varphi}_{16} + \widehat{\Phi}_{22} - \widehat{\Phi}_{24}.
\end{aligned}$$

For this modified choice, we have that $\mathcal{N}_1(\widehat{K}) = \operatorname{span}\{\widehat{\Phi}_i,\, 1 \le i \le 20\}$. Moreover, we have $\mathcal{EJ}_1^*(\widehat{K}) = \operatorname{span}\{\widehat{\Phi}_i,\, 1 \le i \le 24\}$. The functions $\widehat{\Phi}_1, \dots, \widehat{\Phi}_{12}$ satisfy the same properties as described for $\widehat{\varphi}_1, \dots, \widehat{\varphi}_{12}$ above, but also $\widehat{\Phi}_i$ vanish at all face midpoints for all $1 \le i \le 12$. Furthermore, we have that $\widehat{\Phi}_{13}, \dots, \widehat{\Phi}_{24}$ vanish at all vertices and at all but one face midpoint. Moreover, $\widehat{\Phi}_{21}, \dots, \widehat{\Phi}_{24}$ are $H(\mathrm{curl})$-bubble functions. Following these properties and the quadrature rule (5.31), we find that the local mass matrix is comprised of eight 3×3 blocks, one associated with each vertex and one associated with each face midpoint. As a result, the global mass matrix is comprised on the one hand of the same blocks of the first order lumped mass matrix and additional 4×4 blocks for each face. For the same basis functions given in terms of barycentric coordinates, we refer to [**50**].

Now that we can guarantee a block diagonal structure for the mass matrix, we can propose explicit schemes in time.

5.5. Time discretization

For the methods proposed in Sections 5.2, 5.3, and 5.4, the algebraic realization of the mass matrix results in a block-diagonal structure, which enables the efficient application

of its inverse. As a result, we can propose explicit time-stepping schemes. In this section, we discuss the time discretization for the second order method with mass lumping from Section 5.4. We utilize a central difference approximation on a uniform grid in time, which yields an explicit time-stepping method that is second-order accurate. This scheme can also be coupled with any of the first order methods from Sections 5.2 and 5.3.

Let $t^n = n\tau$, $0 \leq n \leq N$ be a given sequence of time points with $\tau = T/N$. We denote by $g^n = g(t^n)$ the evaluation of a function g defined on the time grid $I_\tau = \{0 = t^0 < t^1 < \ldots < t^N = T\}$. Furthermore, we set

$$g^{n+\frac{1}{2}} \coloneqq \frac{1}{2}(g^{n+1} + g^n),$$

which sets the intermediate evaluations to equal the average of two consecutive integer time evaluations. Furthermore, we consider the first order (central) difference quotients

$$d_\tau g^{n+\frac{1}{2}} \coloneqq \frac{1}{\tau}(g^{n+1} - g^n). \qquad \text{and} \qquad \widehat{d_\tau g}^{\,n} \coloneqq \frac{d_\tau g^{n+\frac{1}{2}} + d_\tau g^{n-\frac{1}{2}}}{2} = \frac{g^{n+1} - g^{n-1}}{2\tau},$$

as well as the second order central difference quotient

$$d_{\tau\tau} g^n \coloneqq \frac{1}{\tau^2}(g^{n+1} - 2g^n + g^{n-1}).$$

These definitions also extend to any sequences defined on the time grid. In the following, we consider the second order method from Section 4.4, where V_h is defined as in (4.26), while $(\cdot,\cdot)_h$ is given by (4.28). We can consider the following fully discrete scheme.

Problem 5.5.1 (Fully discrete scheme)**.**
Set $E_h^0 = \Pi_h E_0$ *and* $E_h^1 = \Pi_h \left(E_0 + \tau F_0 - \frac{\tau^2}{2}\operatorname{curl}\operatorname{curl} E_0\right)$. *For* $n \geq 1$ *find* E_h^n *such that*

$$(d_{\tau\tau} E_h^n, \phi_h)_h + (\operatorname{curl} E_h^n, \operatorname{curl} \phi_h) = 0 \qquad \forall \phi_h \in V_h. \tag{5.38}$$

Note that the algebraic realization of (5.38) generates the iteration procedure

$$\mathsf{M_h u}^{n+1} = 2\mathsf{M_h e}^n - \mathsf{M_h e}^{n-1} - \tau \mathsf{K e}^n.$$

By (5.33), we know that $\mathsf{M_h}$ is invertible, and as a result, the iteration step and Problem 5.5.1 are well-defined. For the stability of the fully discrete scheme, we further assume

(A3) the time step τ is chosen such that

$$(\operatorname{curl} \phi_h, \operatorname{curl} \phi_h) \leq \frac{4}{\tau^2}(\phi_h, \phi_h)_h \qquad \forall \phi_h \in V_h,$$

which can be interpreted as an abstract CFL condition; see [**73**] for details. We can now derive the following estimate for the discrete error. In order to derive error estimates, we split the error into interpolation and projection components, namely

$$u(t^n) - u_h^n = -(\Pi_h u(t^n) - u(t^n)) + (\Pi_h u(t^n) - u_h^n) =: -\eta^n + \xi_h^n.$$

For the discrete error component, we have the following estimate

Lemma 5.5.2 (Discrete stability estimate)**.**
Let $\{\xi_h^n\}, \{r_h^n\} \subset V_h$ be given sequences such that

$$(d_{\tau\tau}\xi_h^n, \phi_h)_h + (\operatorname{curl}\xi_h^n, \operatorname{curl}\phi_h) = (r_h^n, \phi_h) \qquad \forall \phi_h \in V_h, n \geq 1.$$

Furthermore, assume that (A1)–(A3) hold. Then for all $0 \leq n \leq N-1$, we have

$$\begin{aligned} &\|d_\tau \xi_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 + \|\operatorname{curl}\widehat{\xi}_h^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 \\ &\qquad \leq C\left(\|d_\tau \xi_h^{\frac{1}{2}}\|_{L^2(\Omega)}^2 + \|\operatorname{curl}\widehat{\xi}_h^{\frac{1}{2}}\|_{L^2(\Omega)}^2 + \sum_{i=1}^{n}\tau(r_h^i, \widehat{d_\tau \xi_h}^n)\right). \end{aligned} \tag{5.39}$$

Proof. The proof follows directly from Lemma 4.5.2 by replacing "div" with "curl". □

Theorem 5.5.3 (Second order convergence of the fully discrete scheme)**.** *Let (A1)–(A3) hold and let E denote a sufficiently smooth solution of (4.5)–(4.7). Then the corresponding solution $(E_h^n)_n$ of Problem 5.5.1 satisfies*

$$\begin{aligned} \max_{0\leq n<N}\left(\|\partial_t E(t^{n+\frac{1}{2}}) - d_\tau E_h^{n+\frac{1}{2}}\|_{L^2(\Omega)} + \|\operatorname{curl}(E(t^{n+\frac{1}{2}}) - \widehat{E}_h^{n+\frac{1}{2}}))\|_{L^2(\Omega)}\right) \\ \leq C(E,T)h^2 + C'(E,T)\tau^2, \end{aligned}$$

where $T = N\tau$ and constant $C(E,T)$ from Theorem 5.4.6 and

$$C'(E,T) = C\left(\|\partial_{tttt}E\|_{L^1(0,t^{n+1},H^1(\mathcal{T}_h))} + \|\partial_{ttt}E\|_{L^\infty(0,t^1,H^1(\mathcal{T}_h))} + \|\operatorname{curl}\partial_{tt}E\|_{L^\infty(0,t^1,L^2(\Omega))}\right).$$

Proof. Note that the continuous errors satisfy

$$\begin{aligned} \max_{0\leq n<N}\left(\|d_\tau \eta^{n+\frac{1}{2}}\|_{L^2(\Omega)}^2 + \|\operatorname{curl}\widehat{\eta}^{\,n+\frac{1}{2}}\|_{L^2(\Omega)}^2\right) \\ \leq ch^2\left(\|\partial_t E\|_{L^\infty(0,T,H^2(\mathcal{T}_h))} + \|\operatorname{curl}E\|_{L^\infty(0,T,H^2(\mathcal{T}_h))}\right). \end{aligned}$$

It remains to estimate the discrete error in Lemma 5.5.2, more precisely the right-hand side of (5.39). Let us first tend to the initial values. By definition, we have $\xi_h^0 = 0$. By Taylor estimates, we further obtain

$$\begin{aligned} \xi_h^1 = \Pi_h E(t^1) - E_h^1 &= \Pi_h\left(E(t^1) - E(0) - \tau\partial_t E(0) - \frac{\tau^2}{2}\operatorname{curl}\operatorname{curl}E(0)\right) \\ &= \Pi_h\left(E(t^1) - E(0) - \tau\partial_t E(0) - \frac{\tau^2}{2}\partial_{tt}E(0)\right) = \Pi_h\left(\frac{\tau^3}{6}\partial_{ttt}E(s)\right) \end{aligned}$$

for an $s \in [0, t^1]$. With this in mind, we compute

$$\begin{aligned} \|d_\tau \xi_h^{\frac{1}{2}}\|_{L^2(\Omega)}^2 + \|\operatorname{curl}\widehat{\xi}_h^{\frac{1}{2}}\|_{L^2(\Omega)}^2 &= \frac{1}{\tau^2}\|d_\tau \xi_h^1\|_{L^2(\Omega)}^2 + \frac{1}{4}\|\operatorname{curl}\xi_h^1\|_{L^2(\Omega)}^2 \\ &\leq C\tau^2\left(\|\Pi_h\partial_{ttt}E\|_{L^\infty(0,t^1,L^2(\Omega))}^2 + \|\operatorname{curl}\Pi_h\partial_{tt}E\|_{L^\infty(0,t^1,L^2(\Omega))}^2\right) \\ &\leq C\tau^2\left(\|\partial_{ttt}E\|_{L^\infty(0,t^1,H^1(\mathcal{T}_h))}^2 + \|\Pi_h^{\#}\operatorname{curl}\partial_{tt}E\|_{L^\infty(0,t^1,L^2(\Omega))}^2\right) \\ &\leq C\tau^2\left(\|\partial_{ttt}E\|_{L^\infty(0,t^1,H^1(\mathcal{T}_h))}^2 + \|\operatorname{curl}\partial_{tt}E\|_{L^\infty(0,t^1,H^1(\mathcal{T}_h))}^2\right), \end{aligned}$$

where we used the H^1 stability of the interpolations Π_h and $\Pi_h^{\#}$. For the discrete error, we now use Lemma 5.5.2. From this lemma, we find that we just have to estimate the residual on the right-hand side of (5.39). Testing equations (5.6) and (5.38) with $\phi_h = \widehat{d_\tau \xi_h}^{\,i}$ and subsequently adding from 1 to n yields

$$\begin{aligned}\sum_{i=1}^{n} \tau(r_h^i, \widehat{d_\tau \xi_h}^{\,i}) &= \tau \sum_{i=1}^{n} (d_{\tau\tau} E(t^i) - \partial_{tt} E(t^i), \widehat{d_\tau \xi_h}^{\,i}) + \tau \sum_{i=1}^{n} (d_{\tau\tau} \eta^i, \widehat{d_\tau \xi_h}^{\,i}) \\ &\quad + \tau \sum_{i=1}^{n} (\operatorname{curl} \eta^i, \operatorname{curl} \widehat{d_\tau \xi_h}^{\,i}) - \tau \sum_{i=1}^{n} \sigma_h (d_{\tau\tau} \Pi_h E(t^i), \widehat{d_\tau \xi_h}^{\,i}) \\ &= (i) + (ii) + (iii) + (iv).\end{aligned}$$

The first term can be estimated by Taylor expansions and Cauchy-Schwarz inequalities, leading to

$$\begin{aligned}|(i)| &\leq \tau \sum_{i=1}^{n} C\tau \|\partial_{tttt} E\|_{L^1(t^{i-1}, t^{i+1}, L^2(\Omega))} \|\widehat{d_\tau \xi_h}^{\,i}\|_{L^2(\Omega)} \\ &\leq \sum_{i=1}^{n} C\tau^2 \|\partial_{tttt} E\|_{L^1(t^{i-1}, t^{i+1}; L^2(\Omega))} \cdot \max_{1 \leq i \leq n} \|\widehat{d_\tau \xi_h}^{\,i}\|_{L^2(\Omega)} \\ &\leq 2C\tau^4 \|\partial_{tttt} E\|^2_{L^1(0, t^{n+1}; L^2(\Omega))} + \frac{1}{6c_1} \max_{1 \leq i \leq n} \|d_\tau \xi_h^{i+\frac{1}{2}}\|^2_{L^2(\Omega)}.\end{aligned}$$

In a similar fashion, we get for the second term

$$|(ii)| \leq 2Ch^4 \|\partial_{tt} E\|^2_{L^1(0, t^{n+1}; H^2(\mathcal{T}_h))} + \frac{1}{6c_1} \max_{0 \leq i \leq n} \|d_\tau \xi_h^{i+\frac{1}{2}}\|^2_{L^2(\Omega)}.$$

For the third term, we use summation by parts to obtain

$$\begin{aligned}(iii) &= (\operatorname{curl} \eta^n, \operatorname{curl} \widehat{\xi}_h^{n+\frac{1}{2}}) - (\operatorname{curl} \eta^1, \operatorname{curl} \widehat{\xi}_h^{\frac{1}{2}}) - \tau \sum_{i=2}^{n} (\operatorname{curl} d_\tau \eta^{i-\frac{1}{2}}, \operatorname{curl} \widehat{\xi}_h^{i-\frac{1}{2}}) \\ &= (iii)_a + (iii)_b + (iii)_c.\end{aligned}$$

For first two terms can be estimated by

$$|(iii)_a| + |(iii)_b| \leq 2Ch^4 \|\operatorname{curl} E\|^2_{L^\infty(0, t^n, H^2(\mathcal{T}_h))} + \frac{1}{6} \max_{1 \leq i \leq n} \|\operatorname{curl} \widehat{\xi}_h^{i+\frac{1}{2}}\|^2_{L^2(\Omega)}.$$

For the third term, we obtain

$$|(iii)_c| \leq 2Ch^4 \|\operatorname{curl} \partial_t E\|^2_{L^1(0, t^{n+1}, H^2(\mathcal{T}_h))} + \frac{1}{6} \max_{1 \leq i \leq n} \|\operatorname{curl} \widehat{\xi}_h^{i+\frac{1}{2}}\|^2_{L^2(\Omega)}.$$

Using a discrete version of Lemma 5.4.3, we get

$$|(iv)| \leq C(E, T)^2 h^4 + \frac{1}{6c_1} \max_{0 \leq i \leq n} \|d_\tau \xi_h^{i+\frac{1}{2}}\|^2_{L^2(\Omega)} + \frac{1}{6} \max_{0 \leq i \leq n} \|\operatorname{curl} \widehat{\xi}_h^{i+\frac{1}{2}}\|^2_{L^2(\Omega)},$$

with

$$C(E,T) = C\left(\|\partial_{tt}E\|_{L^\infty(0,t^{n+1};H^1(\mathcal{T}_h))} + \|\partial_{tt}E\|_{L^1(0,t^{n+1};H^2(\mathcal{T}_h))} + \|\partial_{ttt}E\|_{L^1(0,t^{n+1};H^1(\mathcal{T}_h))}\right).$$

Summing all the terms, using the norm equivalence (5.33) and subsequently applying Lemma 5.5.2, we obtain

$$\begin{aligned}&\frac{1}{c_1}\|d_\tau \xi_h^{n+\frac{1}{2}}\|^2_{L^2(\Omega)} + \|\operatorname{curl}\widehat{\xi}_h^{n+\frac{1}{2}}\|^2_{L^2(\Omega)} \\ &\leq C(E,T)^2h^4 + C'(E,T)^2\tau^4 + \frac{1}{2c_1}\max_{0\leq i\leq n}\|d_\tau \xi_h^{i+\frac{1}{2}}\|^2_{L^2(\Omega)} + \frac{1}{2}\max_{0\leq i\leq n}\|\operatorname{curl}\widehat{\xi}_h^{i+\frac{1}{2}}\|^2_{L^2(\Omega)}.\end{aligned} \tag{5.40}$$

The assertion follows by taking the maximum over all t^n in (5.40) and absorbing the last two terms by the left-hand side yields the estimate. The assertion of the theorem now follows by adding the two estimates for the projection and discrete error. □

5.6. Numerical validation

In this section, we compare the three proposed methods and show that the expected convergence rates do in fact hold. We consider a plane wave traveling through a cube domain Ω that is scattered at a sphere in the interior. Let

$$\begin{aligned}\Omega_1 &= (0,1)^3,\\ \Omega_2 &= \{(x,y,z)\,|\,\|(x-0.5, y-0.5, z-0.5)\|_2 \leq 0.33\},\\ \Omega &= \Omega_1 \setminus \Omega_2\end{aligned}$$

define the relevant computational domains; see Figure 5.4 for a sketch. We consider the

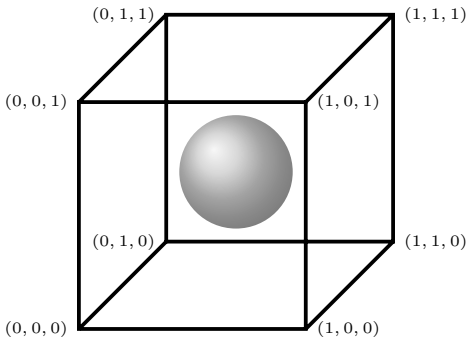

Figure 5.4.: Representation of the computational domain for the wave scattering problem.

system (5.3)–(5.4) on Ω for $0 \leq t \leq T = 2$ with boundary conditions

$$\begin{aligned}E(x,y,z,t) &= \begin{pmatrix}0\\0\\1\end{pmatrix} g(x-t) && \text{on } \partial\Omega_1,\\ n \times E(x,y,z,t) &= 0 && \text{on } \partial\Omega_2,\end{aligned}$$

where $g(s) = 3e^{-100(s+0.5)^2}$. As initial conditions, we choose

$$\begin{aligned} E(x, y, z, 0) &= 0 \qquad \text{on } \partial\Omega, \\ \partial_t E(x, y, z, 0) &= 0 \qquad \text{on } \partial\Omega. \end{aligned}$$

This test case models a plane wave entering the domain on the left boundary which is then scattered at the spherical boundary $\partial\Omega_2$, see Figure 5.5. We use this scenario to

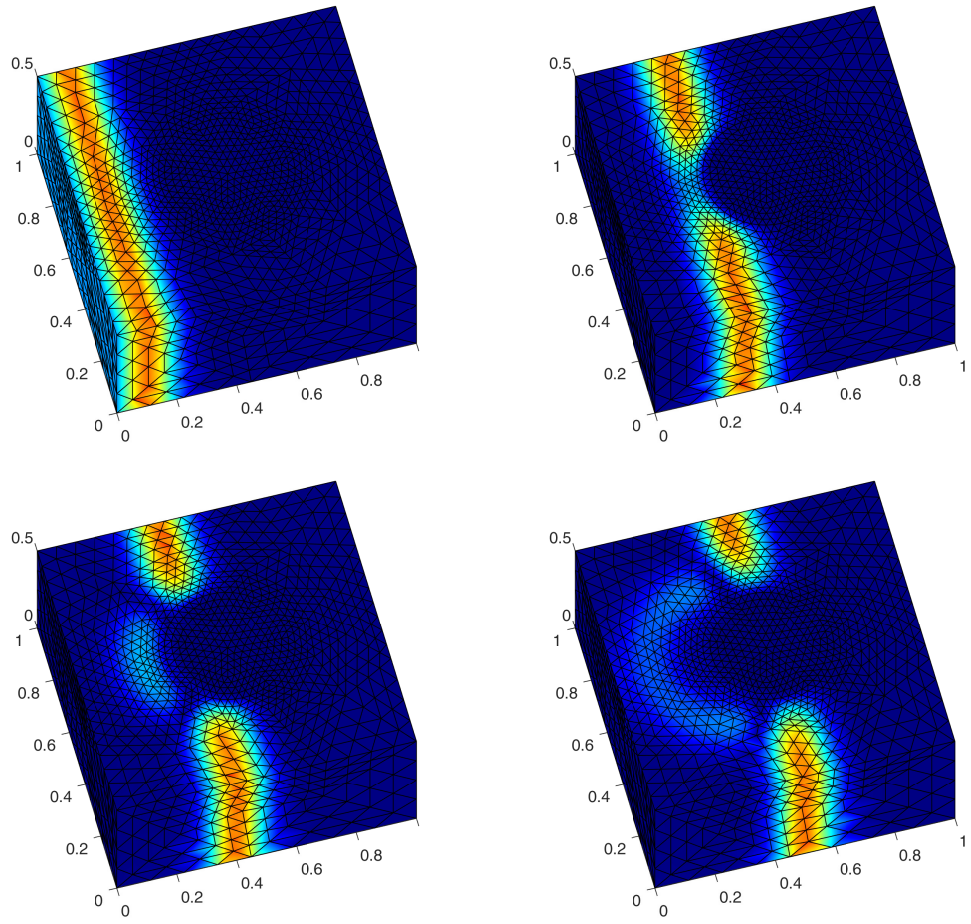

Figure 5.5.: Bird's-eye view of the magnitude of E at time steps $t \in \{0.6, 0.8, 0.9, 1\}$ computed on the discrete level $h = 2^{-3}$ with the $\mathcal{EJ}_1^*$ method. The top part of the domain is hidden to reveal the scattering pattern.

test our three proposed methods. For the evaluation of the convergence of the proposed discretization schemes, we employ a sequence of meshes $\mathcal{T}_h$ obtained by uniform refinement of a quasi-uniform coarse mesh with $h = 2^{-1}$. Since the coarse mesh does not approximate the boundary of Ω_2 very well, with each refinement step, we project the newly generated vertices to the boundary of Ω_2. This mitigates the error introduced by the geometry. However, this introduces the issue of correctly comparing functions defined on non-nested meshes, which can be lifted by defining an appropriate projection operator from the

coarse to the fine grid, which we call $\pi_{h/2}$ in the following. For more details, we refer to the numerical discussion from [**47**]. Then, we define

$$|||e_{h,\tau}||| := \max_{1\leq n\leq N}\Big(\|d_\tau E_{h/2}^{n+\frac{1}{2}} - \pi_{h/2} d_\tau E_h^{n+\frac{1}{2}}\|_{L^2(\Omega)} + \|\mathrm{curl}(E_{h/2}^{n+\frac{1}{2}} - \pi_{h/2} E_h^{n+\frac{1}{2}})\|_{L^2(\Omega)}\Big).$$

In Table 5.1, we list the errors obtained in our computations after the post-processing was applied. Furthermore, we choose $\tau \approx h$ for our simulations. Since the time-stepping procedure is second-order convergent, this choice should not impact the overall convergence of any of the methods. In Table 5.1, we display the errors obtained by our finite-element approximations with mass lumping on a sequence of non-nesting meshes $\{\mathcal{T}_h\}_h$. The columns denoted by $\mathcal{NC}_1$, $\mathcal{N}_0$ and $\mathcal{EJ}_1^*$ contain the error $|||e_{h,\tau}|||$ for the first, the second, and the Yee-like method, respectively.

h	τ	$\mathcal{NC}_1$	eoc	dofs	$\mathcal{N}_0$	eoc	dofs
2^{-1}	2^{-2}	0.635807	—	2682	0.612817	—	1341
2^{-2}	2^{-3}	0.565408	1.12	20504	0.577903	0.98	10252
2^{-3}	2^{-4}	0.265606	1.03	160592	0.271699	1.02	80296

h	τ	$\mathcal{EJ}_1^*$	eoc	dofs
2^{-1}	2^{-2}	0.783171	—	9134
2^{-2}	2^{-3}	0.279806	1.93	71608
2^{-3}	2^{-4}	0.060933	2.10	567376

Table 5.1.: Errors, mesh size h, time step τ, the estimated order of convergence (eoc) as well as the number of dofs for the scattered wave on a sphere.

For a discussion on the behavior of the CFL constant, we refer to the discussion in [**51**].

5.7. Discussion

In this chapter, we presented three finite element methods with mass lumping for the second-order form of Maxwell's equations. A detailed convergence analysis was presented for each of the schemes in space, which we also complemented by a second-order explicit time stepping method. We also observed that the Yee-like and the second order method deliver quasi-best approximation results in the optimal number of degrees of freedom.

As already alluded to in the introduction, the 2D construction can be transferred over from Chapter 4. Figure 5.6 depicts the degrees of freedom necessary for the first and second-order methods in $H(\mathrm{curl})$. As for the basis functions necessary to achieve mass lumping, we define

$$\widehat{\Psi}_i = \mathsf{R}\widehat{\Phi}_i$$

with $\widehat{\Phi}_i$ defined as in Section 2.4.1 and

$$\mathsf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

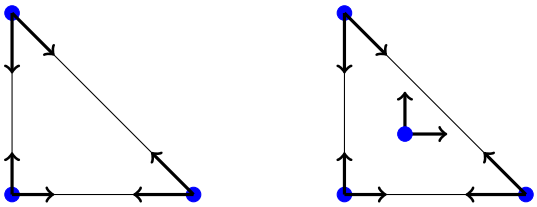

Figure 5.6.: Depiction of the degrees of freedom of the first ($\mathcal{NC}_1(\widehat{K})$) and second order element ($\mathcal{N}_1(\widehat{K})$) as well as the location of the quadrature points depicted by blue dots.

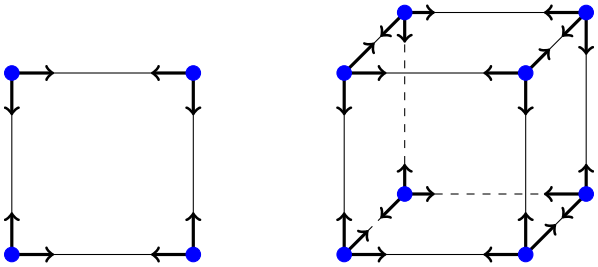

Figure 5.7.: Degrees of freedom for the $H(\mathrm{curl})$ conforming finite element space corresponding to the first order approximation with two degrees of freedom per edge on the reference square (left) and cube (right) represented by arrows. Blue dots depict the quadrature points.

being the $\pi/2$-rotation matrix.

In [**51**], we also established that if the solution is divergence-free, then the finite element proposed by Elmkies and Joly in [**55**] produces second-order accurate discretizations. In general however, non divergence-free solutions lead to a loss in the order of accuracy.

In principle, the extension of our arguments to higher order is also possible. However, finding corresponding quadrature rules is not trivial. Moreover, the additional number of degrees of freedom needed for mass lumping increases strongly with the approximation order. In that case, discontinuous Galerkin methods seem more advantageous, as suggested in [**60**] for problems in H^1.

In addition, other element types can also be considered, see Figure 5.7 for the quadrilateral and hexahedral case.

6. Outlook

In this thesis, we introduced several schemes with mass lumping for problems in porous media, poroelasticity, acoustic wave propagation, and Maxwell's equations. We restricted ourselves to first and second-order accurate methods, but the techniques we employed generalize easily to higher orders as well. Let us briefly touch on a few points that warrant further investigation

- When it comes to the higher order generalization for the MFMFE method for discretizing Darcy flow, we predict that the number of degrees of freedom will rise faster than the corresponding dG scheme when it comes to the discretization order. As a further topic of research, it would be interesting to investigate whether it is possible to generalize the MFMFE to higher orders while still maintaining a moderate number of degrees of freedom.

- The MFMFE can be used to solve other problems where the fast inversion of the $H(\mathrm{div})$ mass matrix is needed. One example is the study of mixed finite element methods for linear elasticity with weakly imposed symmetry; see [**10**]. Here, the symmetry of the $H(\mathrm{div})$ matrix field is imposed weakly with the help of a Lagrange function. In recent publications [**6, 5**], the authors managed to use an extended version of the first order multipoint flux method from [**112**] and showed linear convergence in all variables. As a result of mass lumping, the algebraic structure reduces from a three variable saddle point problem to a symmetric positive definite system for the pressure only, while also benefiting from the advantages that come from the mixed formulation. The extension of the approach to a second-order formulation with mass lumping is an open topic. In our attempts, we managed to successfully eliminate the $H(\mathrm{div})$ variables, i.e., the stresses, which results in a symmetric positive definite system for the pressure and the Lagrange multiplier. However, our attempts failed when trying to also eliminate the Lagrange multiplier from the system. This is subject to further analysis.

- As already discussed in Chapter 3, the three-field MFMFE formulation we designed for poroelasticity can be extended to a fully mass conservative scheme. This is achieved by instead using $H(\mathrm{div})$ conforming finite elements to approximate the displacement filed u and an SIPG correction which also ensures the continuity of tangential traces. The analysis itself is a bit more involved, but it should be possible to prove the convergence of the method by employing standard techniques used for SIPG methods. The numerical tests we conducted in Section 3.5.3 seem to also validate this. We will tackle this analysis in future work.

- The Yee-like scheme introduced in and 5 for Maxwell's equations only seems to work for $\sigma \neq 0$, i.e. whenever the electric conductivity vanishes. This is, unfortunately, only seldom the case. The extension to the $\sigma \neq 0$ case is not trivial and will be tackled in future work.

A. Appendix

A.1. The Piola transformations

For the implementation of $H(\mathrm{div})$ and $H(\mathrm{curl})$-conforming finite elements, it is important to understand how to preserve conformity under a change of basis, namely how to properly map these vector-valued functions from a reference element $\widehat{K}$. Any simplex $K \in \mathcal{T}_h$ is the image $K = F_K(\widehat{K})$ of a reference element $\widehat{K}$ under an affine mapping

$$F_K(\widehat{x}) = a_K + B_K\widehat{x} \qquad \text{with } a_K \in \mathbb{R}^d,\ B_K \in \mathbb{R}^{d\times d}.$$

Under the *contravariant* Piola transformation, v is defined by a vector field $\widehat{v}$ on the reference element under the mapping

$$v(x) = \frac{1}{|\det(B_K)|} B_K \widehat{v}(F_K^{-1}(x)). \tag{A.1}$$

The crucial property of this transformation is that it preserves normal traces of vector fields. In particular, this transformation maps tangent vectors on the reference element to tangent vectors on the physical element.

Under the *covariant* Piola transformation, v is induced by a vector field $\widehat{v}$ on the reference element under the mapping

$$v(x) = B_K^{-T} \widehat{v}(F_K^{-1}(x)). \tag{A.2}$$

The property of this transformation is that it preserves tangential traces of vector fields. In particular, this transformation maps normal vectors on the reference element to normal vectors on the physical element. For more details, we refer to [**19, Chapter 2.1.3**].

Bibliography

[**1**] I. Aavatsmark. An introduction to multipoint flux approximations for quadrilateral grids. volume 6, pages 405–432. 2002. Locally conservative numerical methods for flow in porous media.

[**2**] I. Aavatsmark, G. T. Eigestad, R. A. Klausen, M. F. Wheeler, and I. Yotov. Convergence of a symmetric MPFA method on quadrilateral grids. *Comput. Geosci.*, 11(4):333–345, 2007.

[**3**] R. A. Adams. *Sobolev spaces*. Pure and Applied Mathematics, Vol. 65. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

[**4**] I. Ambartsumyan, E. Khattatov, J. J. Lee, and I. Yotov. Higher order multipoint flux mixed finite element methods on quadrilaterals and hexahedra. *Math. Models Methods Appl. Sci.*, 29(6):1037–1077, 2019.

[**5**] I. Ambartsumyan, E. Khattatov, J. M. Nordbotten, and I. Yotov. A multipoint stress mixed finite element method for elasticity on quadrilateral grids, 2019.

[**6**] I. Ambartsumyan, E. Khattatov, J. M. Nordbotten, and I. Yotov. A multipoint stress mixed finite element method for elasticity on simplicial grids. *SIAM J. Numer. Anal.*, 58(1):630–656, 2020.

[**7**] D. Arnold and F. Brezzi. Mixed and nonconforming finite element methods: Implementation, postprocessing and error estimates. *RAIRO Model. Math. Anal. Numer.*, 19:7–32, 1985.

[**8**] D. N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19(4):742–760, 1982.

[**9**] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39(5):1749–1779, 2001/02.

[**10**] D. N. Arnold, R. S. Falk, and R. Winther. Mixed finite element methods for linear elasticity with weakly imposed symmetry. *Math. Comp.*, 76(260):1699–1723, 2007.

[**11**] I. Babuvska. Error-bounds for finite element method. *Numer. Math.*, 16:322–333, 1970/71.

[**12**] G. A. Baker. Error estimates for finite element approximations for second order hyperbolic equations. *SIAM J. Numer. Anal.*, 13:564–576, 1976.

[**13**] J. Baranger, J.-F. Maitre, and F. Oudin. Connection between finite volume and mixed finite element methods. *RAIRO Modél. Math. Anal. Numér.*, 30(4):445–465, 1996.

[**14**] P. Bastian. Higher order discontinuous Galerkin methods for flow and transport in porous media. In E. Bänsch, editor, *Challenges in Scientific Computing - CISC 2002*, pages 1–22, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[**15**] P. Bastian and B. Rivière. Superconvergence and H(div) projection for discontinuous Galerkin methods. *International Journal for Numerical Methods in Fluids*, 42(10):1043–1057, 2003.

[**16**] J. Bear. *Dynamics of fluids in porous media*. Courier Corporation, 1988.

[**17**] M. A. Biot. General theory of three-dimensional consolidation. *J. Appl. Phys.*, 12:155–164, 1941.

[**18**] D. Boffi, F. Brezzi, L. F. Demkowicz, R. G. Duràn, R. S. Falk, and M. Fortin. *Mixed finite elements, compatibility conditions, and applications*, volume 1939 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin; Fondazione C.I.M.E., Florence, 2008.

[**19**] D. Boffi, F. Brezzi, and M. Fortin. *Mixed finite element methods and applications*, volume 44 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2013.

[**20**] A. Bossavit and L. Kettunen. Yee-like schemes on a tetrahedral mesh, with diagonal lumping. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 12(1-2):129–142, 1999.

[**21**] D. Botteldooren. Acoustical finite-difference time-domain simulation in a quasi-cartesian grid. *The Journal of the Acoustical Society of America*, 95(5):2313–2319, 1994.

[**22**] S. C. Brenner, D.-S. Oh, and L.-Y. Sung. Multigrid methods for saddle point problems: Darcy systems. *Numer. Math.*, 138(2):437–471, 2018.

[**23**] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[**24**] F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge*, 8(no. , no. R-2):129–151, 1974.

[**25**] F. Brezzi, J. Douglas, R. Duràn, and M. Fortin. Mixed finite elements for second order elliptic problems in three variables. *Numer. Math.*, 47:237–250, 1987.

[**26**] F. Brezzi, J. Douglas, and L. D. Marini. Two families of mixed elements for second order elliptic problems. *Numer. Math.*, 88:217–235, 1985.

[**27**] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*, volume 15 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, 1991.

[**28**] F. Brezzi, T. J. R. Hughes, L. D. Marini, and A. Masud. Mixed discontinuous Galerkin methods for Darcy flow. *J. Sci. Comput.*, 22/23:119–145, 2005.

[**29**] E. Bécache, P. Joly, and C. Tsogka. An analysis of new mixed finite elements for the approximation of wave propagation problems. *SIAM Journal on Numerical Analysis*, 37(4):1053–1084, 2000.

[**30**] M. Carlson. *Practical Reservoir Simulation*. PennWell Corporation, 2006.

[**31**] M. J. S. Chin-Joe-Kong, W. A. Mulder, and M. Van Veldhuizen. Higher-order triangular and tetrahedral finite elements with mass lumping for solving the wave equation. *Journal of Engineering Mathematics*, 35(4):405–426, May 1999.

[**32**] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. Studies in Mathematics and its Applications, Vol. 4.

[**33**] L. Codecasa, B. Kapidani, R. Specogna, and F. Trevisan. Novel FDTD technique over tetrahedral grids for conductive media. *IEEE Transactions on Antennas and Propagation*, 66(10):5387–5396, 2018.

[**34**] L. Codecasa and M. Politi. Explicit, consistent, and conditionally stable extension of FDTD to tetrahedral grids by FIT. *IEEE Transactions on Magnetics*, 44(6):1258–1261, 2008.

[**35**] G. Cohen. *Higher-Order Numerical Methods for Transient Wave Equations*. Springer, Heidelberg, 2002.

[**36**] G. Cohen, O. Joly, J. E. Roberts, and N. Tordjman. Higher order triangular finite elements with mass lumping for the wave equation. *SIAM J. Numer. Anal.*, 38:2047–2078, 2001.

[**37**] G. Cohen and P. Monk. Mass lumping edge elements in three dimensions. *ICOSAHOM95 Proceedings*, 1995.

[**38**] G. Cohen and S. Pernet. *Finite Element and Discontinuous Galerkin Methods for Transient Wave Equations*. Scientific Computation. Springer Netherlands, 2016.

[**39**] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math. Soc.*, 49:1–23, 1943.

[**40**] R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.*, 100(1):32–74, 1928.

[**41**] L. C. Cowsar, T. F. Dupont, and M. F. Wheeler. A priori estimates for mixed finite element methods for the wave equation. *Comput. Methods Appl. Mech. Eng.*, 82(1-3):205–222, Sept. 1990.

[**42**] L. C. Cowsar, T. F. Dupont, and M. F. Wheeler. A priori estimates for mixed finite element approximations of second-order hyperbolic equations with absorbing boundary conditions. *SIAM J. Numer. Anal.*, 33:492–504, 1996.

[**43**] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations I. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 7(R3):33–75, 1973.

[**44**] S. Duczek and H. Gravenkamp. Mass lumping techniques in the spectral element method: on the equivalence of the row-sum, nodal quadrature, and diagonal scaling methods. *Comput. Methods Appl. Mech. Engrg.*, 353:516–569, 2019.

[**45**] T. Dupont. L^2 estimates for Galerkin methods for second-order hyperbolic equations. *SIAM J. Numer. Anal.*, 10:880–889, 1973.

[**46**] H. Egger and B. Radu. Super-convergence and post-processing for mixed finite element approximations of the wave equation. *Numer. Math.*, 140(2):427–447, Oct 2018.

[**47**] H. Egger and B. Radu. A mass-lumped mixed finite element method for acoustic wave propagation. *Numer. Math.*, 145(2):239–269, 2020.

[**48**] H. Egger and B. Radu. On a second-order multipoint flux mixed finite element methods on hybrid meshes. *SIAM J. Numer. Anal.*, 58(3):1822–1844, 2020.

[**49**] H. Egger and B. Radu. A mass-lumped mixed finite element method for Maxwell's equations. In *Scientific computing in electrical engineering*, volume 32 of *Math. Ind.*, pages 15–24. Springer, Cham, [2020] ©2020.

[**50**] H. Egger and B. Radu. A second-order finite element method with mass lumping for Maxwell's equations on tetrahedra. *SIAM J. Numer. Anal.*, 59(2):864–885, 2021.

[**51**] H. Egger and B. Radu. A second order finite element method with mass lumping for wave equations in $H(\mathrm{div})$. In *Numerical mathematics and advanced applications—ENUMATH 2019*, volume 139 of *Lect. Notes Comput. Sci. Eng.*, pages 793–801. Springer, Cham, [2021] ©2021.

[**52**] H. Egger and M. Sabouri. On the structure preserving high-order approximation of quasistatic poroelasticity, 2019.

[**53**] H. C. Elman and G. H. Golub. Inexact and preconditioned Uzawa algorithms for saddle point problems. *SIAM J. Numer. Anal.*, 31(6):1645–1661, 1994.

[**54**] A. Elmkies and P. Joly. Éléments finis d'arête et condensation de masse pour les équations de Maxwell: le cas 2D. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 324(11):1287 – 1293, 1997.

[**55**] A. Elmkies and P. Joly. Éléments finis d'arête et condensation de masse pour les équations de Maxwell: le cas de dimension 3. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 325(11):1217 – 1222, 1997.

[**56**] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 1998.

[**57**] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Solution of Equation in Rn (Part 3), Techniques of Scientific Computing (Part 3)*, volume 7 of *Handbook of Numerical Analysis*, pages 713–1018. Elsevier, 2000.

[**58**] I. Fried and D. S. Malkus. Finite element mass matrix lumping by numerical integration with no convergence rate loss. *International Journal of Solids and Structures*, 11(4):461–466, 1975.

[**59**] B. G. Galerkin. Rods and plates. Series occurring in various questions concerning the elastic equilibrium of rods and plates. *Engineers Bulletin (Vestnik Inzhenerov)*, 19:897–908, 1915.

[**60**] S. Geevers, W. Mulder, and J. van der Vegt. New higher-order mass-lumped tetrahedral elements for wave propagation modelling. *SIAM Journal on Scientific Computing*, 40(5):A2830–A2857, 2018.

[**61**] T. Geveci. On the application of mixed finite element methods to the wave equations. *RAIRO Model. Math. Anal. Numer.*, 22:243–250, 1988.

[**62**] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations*, volume 5 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1986. Theory and algorithms.

[**63**] K. Gröger. A $W^{1,p}$-estimate for solutions to mixed boundary value problems for second order elliptic differential equations. *Math. Ann.*, 283(4):679–687, 1989.

[**64**] K. Gröger and J. Rehberg. Resolvent estimates in $W^{-1,p}$ for second order elliptic differential operators in case of mixed boundary conditions. *Math. Ann.*, 285(1):105–113, 1989.

[**65**] M. Grote, A. Schneebeli, and D. Schötzau. Discontinuous Galerkin finite element method for the wave equation. *SIAM Journal on Numerical Analysis*, 44(6):2408–2431, 2006.

[**66**] J. S. Hesthaven and T. Warburton. *Nodal discontinuous Galerkin methods*, volume 54 of *Texts in Applied Mathematics*. Springer, New York, 2008. Algorithms, analysis, and applications.

[**67**] Q. Hong and J. Kraus. Parameter-robust stability of classical three-field formulation of Biot's consolidation model. *Electron. Trans. Numer. Anal.*, 48:202–226, 2018.

[**68**] Q. Hong, J. Kraus, M. Lymbery, and F. Philo. Conservative discretizations and parameter-robust preconditioners for Biot and multiple-network flux-based poroelasticity models. *Numer. Linear Algebra Appl.*, 26(4):e2242, 25, 2019.

[**69**] X. Hu, C. Rodrigo, F. J. Gaspar, and L. T. Zikatanov. A nonconforming finite element method for the Biot's consolidation model in poroelasticity. *J. Comput. Appl. Math.*, 310:143–154, 2017.

[**70**] R. Ingram, M. F. Wheeler, and I. Yotov. A multipoint flux mixed finite element method on hexahedra. *SIAM J. Numer. Anal.*, 48(4):1281–1312, 2010.

[**71**] E. W. Jenkins. Numerical solution of the acoustic wave equation using Raviart–Thomas elements. *Journal of Computational and Applied Mathematics*, 206(1):420 – 431, 2007.

[**72**] E. W. Jenkins, T. Rivière, and M. F. Wheeler. A priori error estimates for mixed finite element approximations of the acoustic wave equation. *SIAM J. Numer. Anal.*, 40:1698–1715, 2002.

[**73**] P. Joly. Variational methods for time-dependent wave propagation problems. In *Topics in Computational Wave Propagation*, volume 31 of *LNCSE*, pages 201–264. Springer, 2003.

[**74**] M. Kaltenbacher. *Numerical Simulation of Mechatronic Sensors and Actuators: Finite Elements for Computational Multiphysics*. Springer Berlin Heidelberg, 2015.

[**75**] G. Kanschat and B. Riviere. A finite element method with strong mass conservation for Biot's linear consolidation model. *J. Sci. Comput.*, 77:1762–1779, 2018.

[**76**] B. Kapidani, L. Codecasa, and J. Schöberl. An arbitrary-order Cell Method with block-diagonal mass-matrices for the time-dependent 2D Maxwell equations. *J. Comput. Phys.*, 433:Paper No. 110184, 20, 2021.

[**77**] R. A. Klausen and A. F. Stephansen. Convergence of multi-point flux approximations on general grids and media. *Int. J. Numer. Anal. Model.*, 9(3):584–606, 2012.

[**78**] P. Lacoste. Mass-lumping for the first order Raviart-Thomas-Nédélec finite elements. *Comptes Rendus. Mathématique. Académie des Sciences, Paris*, 10, 01 2004.

[**79**] C. G. Makridakis. On mixed finite element methods for linear elastodynamics. *Numer. Math.*, 61:235–260, 1992.

[**80**] P. Monk. Analysis of a finite element methods for Maxwell's equations. *SIAM J. Numer. Anal.*, 29:714–729, 1992.

[**81**] P. Monk. A finite element method for approximating the time-harmonic Maxwell equations. *Numer. Math.*, 63(1):243–261, Dec 1992.

[**82**] P. Monk. An analysis of Nédélec's method for the spatial discretization of Maxwell's equations. *Journal of Computational and Applied Mathematics*, 47(1):101 – 121, 1993.

[**83**] P. Monk. Superconvergence of finite element approximations to Maxwell's equations. *Numerical Methods for Partial Differential Equations*, 10(6):793–812, 1994.

[**84**] P. Monk. *Finite element methods for Maxwell's equations*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 2003.

[**85**] P. Monk and A. Parrott. A dispersion analysis of finite element methods for Maxwell's equations. *SIAM Journal on Scientific Computing*, 15(4):916–937, 1994.

[**86**] W. A. Mulder. Higher-order mass-lumped finite elements for the wave equation. *Journal of Computational Acoustics*, 09(02):671–680, 2001.

[**87**] G. Mur and A. de Hoop. A finite-element method for computing three-dimensional electromagnetic fields in inhomogeneous media. *IEEE Transactions on Magnetics*, 21(6):2188–2191, November 1985.

[**88**] M. A. Murad and A. F. D. Loula. Improved accuracy in finite element analysis of Biot's consolidation problem. *Comput. Methods Appl. Mech. Engrg.*, 95:359–382, 1992.

[**89**] M. A. Murad and A. F. D. Loula. On stability and convergence of finite element approximations of Biot's consolidation problem. *Internat. J. Numer. Methods Engrg.*, 37:645–667, 1994.

[**90**] J. C. Nédélec. A new family of mixed finite elements in $\mathbb{R}^3$. *Numer. Math.*, 50(1):57–81, Jan 1986.

[**91**] J. A. Nitsche. On Korn's second inequality. *RAIRO Anal. Numér.*, 15(3):237–248, 1981.

[**92**] J. C. Nédélec. Mixed finite elements in $\mathbb{R}^3$. *Numer. Math.*, 35:315–341, 1980.

[**93**] P. J. Phillips and M. F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelasticity. I. The continuous in time case. *Comput. Geosci.*, 11:131–144, 2007.

[**94**] P. J. Phillips and M. F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelasticity. II. The discrete-in-time case. *Comput. Geosci.*, 11:145–158, 2007.

[**95**] P. J. Phillips and M. F. Wheeler. A coupling of mixed and discontinuous Galerkin finite-element methods for poroelasticity. *Comput. Geosci.*, 12:417–435, 2008.

[**96**] S. Pitassi, F. Trevisan, and R. Specogna. Explicit geometric construction of sparse inverse mass matrices for arbitrary tetrahedral grids. *Computer Methods in Applied Mechanics and Engineering*, 377:113699, 2021.

[**97**] P. A. Raviart and J. M. Thomas. A mixed finite element method for 2nd order elliptic problems. In I. Galligani and E. Magenes, editors, *Mathematical Aspects of Finite Element Methods*, pages 292–315, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg.

[**98**] W. Ritz. Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. *J. Reine Angew. Math.*, 135:1–61, 1909.

[**99**] B. Rivière and M. F. Wheeler. Discontinuous finite element methods for acoustic and elastic wave problems. In *Current trends in scientific computing (Xi'an, 2002)*, volume 329 of *Contemp. Math.*, pages 271–282. Amer. Math. Soc., Providence, RI, 2003.

[**100**] B. Rivière, M. F. Wheeler, C. Baumann, and W. C. Baumann. Part II. Discontinuous Galerkin method applied to a single phase flow in porous media. *Comput. Geosci*, 4:337–349, 1999.

[**101**] R. E. Showalter. Diffusion in poro-elastic media. *J. Math. Anal. Appl.*, 251:310–340, 2000.

[**102**] M. Stanglmeier, N. C. Nguyen, J. Peraire, and B. Cockburn. An explicit hybridizable discontinuous Galerkin method for the acoustic wave equation. *Comput. Methods Appl. Mech. Engrg.*, 300:748–769, 2016.

[**103**] R. Stenberg. Postprocessing schemes for some mixed finite elements. *RAIRO Model. Math. Anal. Numer.*, 25:151–167, 1991.

[**104**] A. Taflove. Application of the finite-difference time-domain method to sinusoidal steady-state electromagnetic-penetration problems. *IEEE Transactions on Electromagnetic Compatibility*, EMC-22(3):191–202, Aug 1980.

[**105**] A. Taflove and S. C. Hagness. *Computational electrodynamics: the finite-difference time-domain method*. Artech House, Norwood, 3rd edition, 2005.

[**106**] D. K. Todd and L. W. Mays. *Groundwater Hydrology*. Wiley, 2004.

[**107**] A. vZenívsek. The existence and uniqueness theorem in Biot's consolidation theory. *Apl. Mat.*, 29:194–211, 1984.

[**108**] A. vZenívsek. Finite element methods for coupled thermoelasticity and coupled consolidation of clay. *RAIRO Anal. Numér.*, 18:183–205, 1984.

[**109**] T. Weiland. A discretization model for the solution of Maxwell's equations for six-component fields. *Archiv Elektronik und Uebertragungstechnik*, 31:116–120, Mar. 1977.

[**110**] M. Wheeler, G. Xue, and I. Yotov. A multipoint flux mixed finite element method on distorted quadrilaterals and hexahedra. *Numer. Math.*, 121(1):165–204, 2012.

[**111**] M. Wheeler, G. Xue, and I. Yotov. Coupling multipoint flux mixed finite element methods with continuous Galerkin methods for poroelasticity. *Comput. Geosci.*, 18(1):57–75, 2014.

[**112**] M. F. Wheeler and I. Yotov. A multipoint flux mixed finite element method. *SIAM J. Numer. Anal.*, 44(5):2082–2106, 2006.

[**113**] J. Wloka. *Partielle Differentialgleichungen*. Mathematische Leitfäden. [Mathematical Textbooks]. B. G. Teubner, Stuttgart, 1982. Sobolevräume und Randwertaufgaben. [Sobolev spaces and boundary value problems].

[**114**] K. Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14:302–307, May 1966.

[**115**] E. Zhebel, S. Minisini, A. Kononov, and W. A. Mulder. A comparison of continuous mass-lumped finite elements with finite differences for 3-d wave propagation. *Geophysical Prospecting*, 62(5):1111–1125, 2014.

[**116**] O. C. Zienkiewicz. *The finite element method in engineering science*. McGraw-Hill, London-New York-Düsseldorf, 1967.

[**117**] W. Zulehner. Nonstandard norms and robust estimates for saddle point problems. *SIAM J. Matrix Anal. Appl.*, 32(2):536–560, 2011.

Wissenschaftlicher Werdegang

Bogdan Radu

2022	Promotion Mathematik
2019-2022	Wissenschaftlicher Mitarbeiter an der TU Darmstadt in der AG Numerik
2016-2019	Stipendiat der Graduiertenschule Computational Engineering in Darmstadt
2013-2015	Studium der Mathematik an der TU Darmstadt, Abschluss M.Sc.
2010-2013	Studium der Mathematik an der TU Darmstadt, Abschluss B.Sc.
2010	Abitur am Friedrich-von-Alberti Gymnasium in Bad Friedrichshall