Discontinuous Galerkin Time Discretization Methods for Parabolic Problems with Linear Constraints

Igor Voulis[*] and Arnold Reusken[†]

Institut für Geometrie und Praktische Mathematik
Templergraben 55, 52062 Aachen, Germany

[*] Institut für Geometrie und praktische Mathematik, RWTH Aachen University, Templergraben 55, D-52062 Aachen, Germany. (voulis@igpm.rwth-aachen.de)

[†] Institut für Geometrie und praktische Mathematik, RWTH Aachen University, Templergraben 55, D-52062 Aachen, Germany. (reusken@igpm.rwth-aachen.de)

DISCONTINUOUS GALERKIN TIME DISCRETIZATION METHODS FOR PARABOLIC PROBLEMS WITH LINEAR CONSTRAINTS

IGOR VOULIS* AND ARNOLD REUSKEN†

Abstract. We consider time discretization methods for abstract parabolic problems with inhomogeneous linear constraints. Prototype examples that fit into the general framework are the heat equation with inhomogeneous (time-dependent) Dirichlet boundary conditions and the time-dependent Stokes equation with an inhomogeneous divergence constraint. Two common ways of treating such linear constraints, namely explicit or implicit (via Lagrange multipliers) are studied. These different treatments lead to different variational formulations of the parabolic problem. For these formulations we introduce a modification of the standard discontinuous Galerkin (DG) time discretization method in which an appropriate projection is used in the discretization of the constraint. For these discretizations (optimal) error bounds, including superconvergence results, are derived. Discretization error bounds for the Lagrange multiplier are presented. Results of experiments confirm the theoretically predicted optimal convergence rates and show that without the modification the (standard) DG method has sub-optimal convergence behavior.

Key words. abstract parabolic problem, discontinuous Galerkin methods, discretization of linear constraints, optimal discretization error bounds

AMS subject classifications. 65M60, 65J10

1. Introduction. Nowadays the discontinuous Galerkin (DG) finite element method is a popular discretization technique for many classes of ordinary and partial differential equations, cf. the overviews in [12, 20, 7]. For ordinary differential equations DG finite element methods have been analyzed in e.g. [6, 13, 22] and the references therein. For parabolic partial differential equations the DG finite element for time discretization has been introduced in [21] and was further studied in [9, 10, 11, 23, 32, 28, 1]. In this paper we study DG finite element *time* discretization methods for *parabolic problems with constraints*. As far as we know, in the DG error analyses for parabolic problems available in the literature only the case with *homogeneous* constraints has been considered. In practice one often has to deal with *in*homogeneous constraints. Typical examples are a heat equation with nonzero time-dependent Dirichlet boundary conditions, an instationary Stokes flow with a time-dependent inflow boundary condition and a Stokes flow with an inhomogeneous time-dependent divergence constraint. Runge-Kutta methods for parabolic equations with a non-homogenous constraint have been analyzed in the recent paper [3]. A DG time discretization method for the wave equation with an inhomogenous boundary condition is analyzed in [33]. In this paper we consider a DG time discretization of a parabolic problem with non-homogeneous linear constraints and present an error analysis resulting in optimal error bounds. It turns out that for obtaining such optimal bounds we have to *modify the standard formulation of a DG Galerkin finite element method* as given in [32]. Results of numerical experiments will show that without an appropriate modification the standard DG Galerkin finite element method applied to a parabolic problem with an inhomogeneous constraint does *not* yield optimal convergence rates.

In this paper the parabolic problem and the error analysis of the DG time dis-

cretization are presented in an abstract setting (as in e.g. [28]). The linear constraints that are part of the problem can be treated in different ways: explicitly, by eliminating variables, implicitly, by means of a Lagrange-multiplier, or by a combination of these. The analysis in the paper covers all three cases. Prototype examples that fit into the abstract framework are the standard scalar heat equation with inhomogeneous (time-dependent) Dirichlet boundary conditions (usually treated explicitly) and the instationary Stokes equations with inhomogeneous (time-dependent) Dirichlet boundary conditions (usually treated explicitly) and/or an inhomogeneous divergence constraint (usually treated implicitly by the pressure Lagrange multiplier). These examples will be discussed in detail in Subsection 2.3.

The main results of the paper are the following. Firstly, we present a *modification of the standard formulation* of a DG time discretization method that yields optimal order discretization errors also for the case of *in*homogeneous constraints. Secondly, for this modified method we *prove optimal discretization error bounds*. Both optimal global energy norm bounds and optimal superconvergence results are derived. Finally, for the case that the constraints are treated implicitly (e.g. divergence constraint in Stokes problem) we derive *discretization error bounds for the corresponding Lagrange multiplier*. In the example of the Stokes problem this yields error bounds for the DG time discretization of the pressure variable. We note that even for the case of homogeneous constraints we are not aware of error analyses of DG time discretization methods for Stokes equations that yield (optimal order) error bounds for the pressure variable. A more detailed discussion of how the results of this paper are related to other literature is given in Remark 2.1.

The paper is organized as follows. In Section 2 we introduce the class of abstract parabolic problems that we consider. Different formulations, related to the different treatments of the constraints, are presented. Furthermore, specific concrete examples that fit in the abstract setting are discussed. In Section 3 we introduce a DG scheme for the time discretization of the abstract parabolic problem. In Section 4 we give an error analysis of the DG scheme. We derive optimal order error bounds in the energy norm and prove a nodal superconvergence result. In Section 5, for the case of an implicit treatment of the constraint, we analyze the discretization error in the Lagrange multiplier. We derive, under very mild assumptions, a sub-optimal result for the convergence of the Lagrange multiplier and using an additional regularity condition we are able to prove an optimal discretization error bound. In Section 6 we introduce a fully discrete numerical scheme. In Section 7 this scheme is used to perform numerical experiments and results of a few experiments are presented that illustrate the convergence behavior of the DG method. Finally, we give an outlook in Section 8.

2. Parabolic problem with linear constraints. In this section we introduce the class of parabolic problems with linear constraints that we treat in this paper. For this we first introduce some notation and recall relevant well-known results on well-posedness of (abstract) parabolic problems e.g. [32, 28]. Let $\mathcal{U}, \mathcal{H}$ be real separable Hilbert spaces with a dense continuous embedding $\mathcal{U} \hookrightarrow \mathcal{H}$. The norms are denoted by $\|\cdot\|_{\mathcal{U}}, \|\cdot\|_{\mathcal{H}}$, respectively. These spaces induce a Gelfand triple $\mathcal{U} \hookrightarrow \mathcal{H} \cong \mathcal{H}' \hookrightarrow \mathcal{U}'$. Let $a : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ be a symmetric continuous coercive bilinear form on $\mathcal{U}$:

$$|a(u,v)| \leq \Gamma \|u\|_{\mathcal{U}} \|v\|_{\mathcal{U}} \quad \text{for all} \quad u, v \in \mathcal{U}, \tag{2.1}$$

$$a(v,v) \geq \gamma \|v\|_{\mathcal{U}}^2 \quad \text{for all} \quad v \in \mathcal{U}, \tag{2.2}$$

with $\gamma > 0$. The corresponding operator is denoted by $A : \mathcal{U} \to \mathcal{U}'$, $Au(v) = a(u, v)$. Let $I = (0, T)$ be a given time interval and for Hilbert spaces Z_1, Z_2 we define $W^1(Z_1; Z_2) := \{\, u \in L^2(I; Z_1) \mid u' \in L^2(I; Z_2) \,\}$, $H^m(I; Z_1) := H^m(I) \otimes Z_1$, cf. [34]. A standard weak formulation of an abstract parabolic problem is as follows: find $u \in W^1(\mathcal{U}; \mathcal{U}')$ such that $u(0) = u_0$ and

$$u' + Au = f \quad \text{in} \ \ L^2(I; \mathcal{U}'). \tag{2.3}$$

We assume $u_0 \in \mathcal{H}$, $f \in L^2(I; \mathcal{U}')$. This problem is well-posed [34, 28]. Standard examples of problems that fit in this abstract framework are the heat equation with *homogeneous* BC ($\mathcal{U} = H_0^1(\Omega)$) and the Stokes equation with *homogeneous* BC ($\mathcal{U} = \{\, u \in H_0^1(\Omega)^d \mid \operatorname{div} u = 0 \,\}$), cf. Examples 2.1 and 2.2 below. In the literature one can find analyses of DG time discretizations applied to these parabolic problems [32, 28]. In this paper we present an extension of the time discretization error analysis that applies to such problems with *in*homogeneous boundary conditions or an *in*homogeneous divergence constraint $\operatorname{div} u = g$. For this we introduce an abstract (variatiational) formulation of a problem as in (2.3) with linear constraints. For a Hilbert space $\mathcal{Q}$ let $b(\cdot, \cdot) : \mathcal{U} \times \mathcal{Q} \to \mathbb{R}$ be a continuous bilinear form and $B : \mathcal{U} \to \mathcal{Q}'$ the corresponding operator representation: $Bu(q) = b(u, q)$ for all $u \in \mathcal{U}$, $q \in \mathcal{Q}$. This bilinear form and linear operator B are called *constraint operators*. We define

$$\mathcal{V} := \ker(B) = \{\, u \in \mathcal{U} \mid Bu = 0 \,\},$$

which is a Hilbert space. We assume that

$$B : \mathcal{U} \to \mathcal{Q}' \quad \text{is surjective.} \tag{2.4}$$

This is equivalent ([12], Lemma A.40) to the inf-sup property:

$$\exists \ \alpha > 0 : \quad \inf_{q \in \mathcal{Q}} \sup_{u \in \mathcal{U}} \frac{b(u, q)}{\|u\|_{\mathcal{U}} \|q\|_{\mathcal{Q}}} \geq \alpha. \tag{2.5}$$

In the subsections below we give different formulations for the abstract parabolic problem with constraints which are relevant in particular applications, cf. the examples in Subsection 2.3.

2.1. Constrained formulation. For given $f \in L^2(I; \mathcal{V}')$, $g \in L^2(I; \mathcal{Q}')$ we consider the following abstract parabolic problem: find $u \in W^1(\mathcal{U}; \mathcal{V}')$ such that $u(0) = u_0$ and

$$\begin{aligned} u' + Au &= f \quad \text{in} \ \ L^2(I; \mathcal{V}') \\ Bu &= g \quad \text{in} \ \ L^2(I; \mathcal{Q}'). \end{aligned} \tag{2.6}$$

We relax the ellipticity condition (2.2) and only require ellipticity of $a(\cdot, \cdot)$ to hold on the subspace $\mathcal{V} \subset \mathcal{U}$. We thus have

$$|a(u, v)| \leq \Gamma \|u\|_{\mathcal{U}} \|v\|_{\mathcal{U}} \quad \text{for all} \ \ u, v \in \mathcal{U}, \tag{2.7}$$

$$a(v, v) \geq \gamma \|v\|_{\mathcal{U}}^2 \quad \text{for all} \ \ v \in \mathcal{V}, \tag{2.8}$$

with $\gamma, \Gamma > 0$. Concerning existence of a unique solution of (2.6) we note the following. If $g = 0$ then the constraint in (2.6) implies $u \in \mathcal{V}$, and thus we are in the standard setting described above, with $\mathcal{U}$ replaced by $\mathcal{V}$ (and the Gelfand triple $\mathcal{V} \hookrightarrow \overline{\mathcal{V}}^{\mathcal{H}} \hookrightarrow \mathcal{V}'$).

3

Hence we have well-posedness. For the inhomogeneous case we apply a standard translation argument. Note that from the inf-sup property it follows that there exists a unique $v \in L^2(I; \mathcal{V}^\perp)$ such that $Bv = g$, where $\mathcal{V}^\perp$ is the orthogonal complement of $\mathcal{V}$ in $\mathcal{U}$. Assume that g has sufficient regularity, e.g. $g \in H^1(I; \mathcal{Q}')$, such that there exists a $v \in W^1(\mathcal{U}; \mathcal{V}')$ that satisfies $Bv = g$. Application of the standard analysis yields that there exists (a unique) $w \in W^1(\mathcal{V}; \mathcal{V}')$ with $w(0) = u_0 - v(0)$ and

$$w' + Aw = f - Av - v' \quad \text{in} \ \ L^2(I; \mathcal{V}').$$

Hence $u := w + v \in W^1(\mathcal{U}; \mathcal{V}')$ satisfies $u(0) = u_0$ and the variational equation (2.6). Uniqueness follows from the fact that (2.6) with $g = 0$, $f = 0$ and $u_0 = 0$ has only the trivial solution.

2.2. Mixed formulation. In practice, in particular in the context of Galerkin finite element discretization methods, it may be not convenient to use the kernel space $\mathcal{V}$ in the first equation in (2.6). One then introduces a suitable Lagrange multiplier to enlarge the constrained space $\mathcal{V}$. We introduce such a mixed formulation. The structure of this mixed formulation is inspired by the examples considered in Subsection 2.3, in which only for a part of the constraint equation $Bu = g$ a corresponding Lagrange multiplier is used. For this we decompose the constraint bilinear form as

$$b(u, q) = b_1(u, q_1) + b_2(u, q_2), \quad q = (q_1, q_2) \in \mathcal{Q}_1 \times \mathcal{Q}_2 = \mathcal{Q},$$

with $\mathcal{Q}_i$, $i = 1, 2$, Hilbert spaces, and $b_i : \mathcal{U} \times \mathcal{Q}_i \to \mathbb{R}$ continuous bilinear forms. This splitting is such that for the $b_1(\cdot, \cdot)$ part a Lagrange multiplier will be introduced, whereas for the $b_2(\cdot, \cdot)$ part no Lagrange multiplier is used. We assume that $\mathcal{Q}_1 \neq 0$. The bilinear forms $b_i(\cdot, \cdot)$ induce corresponding operators $B_i : \mathcal{U} \to \mathcal{Q}_i'$. From (2.5) it follows that

$$\inf_{q_i \in \mathcal{Q}_i} \sup_{u \in \mathcal{U}} \frac{b_i(u, q_i)}{\|u\|_{\mathcal{U}} \|q_i\|_{\mathcal{Q}_i}} \geq \alpha, \quad i = 1, 2, \tag{2.9}$$

holds, provided $b_i(\cdot, \cdot)$ is not identically zero. Define

$$\mathcal{V}_i := \ker(B_i) = \{ u \in \mathcal{U} \mid b_i(u, q_i) = 0 \ \ \forall \, q_i \in \mathcal{Q}_i \}, \quad i = 1, 2.$$

Hence, $\mathcal{V} = \mathcal{V}_1 \cap \mathcal{V}_2$. We also have

$$\mathcal{V} = \{ u \in \mathcal{V}_2 \mid b_1(u, q_1) = 0 \ \ \forall \, q_1 \in \mathcal{Q}_1 \}. \tag{2.10}$$

We allow $\mathcal{Q}_2 = \{0\}$, $\mathcal{Q}_1 = \mathcal{Q}$, in which case we have $\mathcal{V}_2 = \mathcal{U}$, $\mathcal{V} = \mathcal{V}_1$. Note that $B : \mathcal{V}^\perp \to \mathcal{Q}_1' \times \mathcal{Q}_2'$ is a bijection and the preimage of $\mathcal{Q}_1' \times \{0\}$ is $\mathcal{V}^\perp \cap \mathcal{V}_2$. It follows that $B_1 : \mathcal{V}^\perp \cap \mathcal{V}_2 \to \mathcal{Q}_1'$ is a bijection, from which we conclude that $B_1 : \mathcal{V}_2 \to \mathcal{Q}_1'$ is surjective. This is equivalent ([12], Lemma A.40) to the inf-sup property:

$$\exists \ \beta > 0 : \quad \inf_{q_1 \in \mathcal{Q}_1} \sup_{u \in \mathcal{V}_2} \frac{b_1(u, q_1)}{\|u\|_{\mathcal{U}} \|q_1\|_{\mathcal{Q}_1}} \geq \beta. \tag{2.11}$$

Note that if $\mathcal{Q}_2 = \{0\}$ then (2.11) is the same as (2.9) with $i = 1$. From (2.11) and (2.10) it follows that for the adjoint B_1' of $B_1 : \mathcal{V}_2 \to \mathcal{Q}_1'$ we have that

$$B_1' : Q_1 \to \mathcal{V}_2^0 := \{ g \in \mathcal{V}_2' \mid g(v) = 0 \quad \text{for all} \ \ v \in \mathcal{V} \} \tag{2.12}$$

4

is an isomorphism and

$$\|B_1'p\|_{\mathcal{V}_2'} \geq \beta\|p\|_{Q_1} \quad \text{for all} \ \ p \in Q_1, \tag{2.13}$$

cf. Lemma 4.1 in [15]. We now introduce a mixed formulation of (2.6). For $f \in L^2(I; \mathcal{V}_2')$ determine $u \in W^1(\mathcal{U}; \mathcal{V}_2')$, $p \in L^2(I; Q_1)$ such that

$$\begin{aligned} u' + Au + B_1'p &= f \quad \text{in} \ \ L^2(I; \mathcal{V}_2') \\ Bu &= g \quad \text{in} \ \ L^2(I; Q'). \end{aligned} \tag{2.14}$$

Here $B_1' : Q_1 \to \mathcal{V}_2'$ is the adjoint of $B_1 : \mathcal{V}_2 \to Q_1'$. Note that for $v \in \mathcal{V}_2$ we have $b_2(v, q_2) = 0$ for all $q_2 \in Q_2$, hence, $(B_1'p)(v) = b_1(v, p) = b(v, (p, 0))$ holds for $v \in \mathcal{V}_2$, $p \in Q_1$.

We now discuss the relation between the constrained formulation (2.6) and the mixed formulation (2.14). Let $u \in W^1(\mathcal{U}; \mathcal{V}')$ be the unique solution of (2.6). We *assume* that u has additional regularity such that $u \in W^1(\mathcal{U}; \mathcal{V}_2')$. Then $\ell(t) := f(t) - u'(t) - Au(t) \in \mathcal{V}_2'$ (a.e. in t) and $\ell(t)(v) = 0$ for all $v \in \mathcal{V}$, i.e., $\ell(t) \in \mathcal{V}_2^0$ holds. Hence, there exists a unique $p(t) \in Q_1$ such that $B_1'p(t) = \ell(t)$. Using (2.13) it follows that $p \in L^2(I; Q_1)$. Thus, if the unique solution u of the constrained formulation (2.6) has regularity $u \in W^1(\mathcal{U}; \mathcal{V}_2')$, there is a unique $p \in L^2(I; Q_1)$ such that (u, p) is the unique solution of the mixed formulation (2.14).

2.3. Examples. We describe three examples that fit into the general abstract framework presented above. In these examples we consider different types of constraints, which are allowed to be inhomogeneous.

EXAMPLE 2.1 (Heat equation). The heat equation in $\Omega \subset \mathbb{R}^d$, with *(time-dependent) inhomogeneous Dirichlet boundary conditions* $h \in L^2(I; H^{\frac{1}{2}}(\partial\Omega))$ in constrained formulation is as follows:

$$\begin{aligned} u' - \Delta u &= f \quad \text{in} \ \ L^2(I; H^{-1}(\Omega)) \\ u|_{\partial\Omega} &= h. \end{aligned}$$

For this problem we have $\mathcal{U} = H^1(\Omega)$, $\mathcal{H} = L^2(\Omega)$, $\mathcal{Q} = H^{\frac{1}{2}}(\partial\Omega)$, $a(u, v) := \int_\Omega \nabla u \cdot \nabla v$, $b(u, q) = (\text{tr}\,(u), q)_{H^{\frac{1}{2}}(\partial\Omega)}$, with $\text{tr} : H^1(\Omega) \to H^{\frac{1}{2}}(\partial\Omega)$ the trace operator, which is surjective. This yields $\mathcal{V} = \ker(B) = H_0^1(\Omega)$. The assumptions (2.4), (2.7), (2.8) are satisfied. We do not consider a mixed formulation, because in the constrained formulation (2.6) we use the space $\mathcal{V} = H_0^1(\Omega)$, which is the natural one for the bilinear form $a(\cdot, \cdot)$.

EXAMPLE 2.2 (Stokes equations). We consider the Stokes equation in $\Omega \subset \mathbb{R}^d$ with homogeneous boundary conditions and a *non-zero (time-dependent) divergence* $g \in L^2(I; L_0^2(\Omega))$. In this setting we have $\mathcal{U} = H_0^1(\Omega)^d$, $\mathcal{H} = L^2(\Omega)^d$, $\mathcal{Q} = L_0^2(\Omega) := L^2(\Omega)/\mathbb{R}$, $a(u, v) := \int_\Omega \nabla u : \nabla v$, $b(u, q) = \int_\Omega \text{div}\, u\, q$, $\mathcal{V} = \{\, u \in H_0^1(\Omega)^d \mid \text{div}\, u = 0\,\}$. Recall that $u \in H_0^1(\Omega)^d \to \text{div}\, u \in L_0^2(\Omega)$ is surjective. The Stokes equation in constrained formulation is as follows:

$$\begin{aligned} u' - \Delta u &= f \quad \text{in} \ \ L^2(I; \mathcal{V}') \\ \text{div}\, u &= g \quad \text{in} \ \ L^2(I; L_0^2(\Omega)). \end{aligned}$$

The assumptions (2.4), (2.7), (2.8) are satisfied. For the mixed formulation we take $b_2 \equiv 0$, i.e., $b_1(u, q) = \int_\Omega \operatorname{div} u \, q$, $q \in \mathcal{Q}_1 = \mathcal{Q}$, and $\mathcal{V}_2 = \mathcal{U} = H_0^1(\Omega)^d$. Hence, the formulation (2.14) is the standard mixed formulation of a time-dependent Stokes problem with homogeneous boundary conditions, cf. [12] Sect. 4.1.

EXAMPLE 2.3. We consider the Stokes equation in $\Omega \subset \mathbb{R}^d$ with *non-zero (time-dependent) boundary data* $h \in L^2(I; H^{\frac{1}{2}}(\partial\Omega)^d)$ *and a non-zero (time-dependent) divergence* $g \in L^2(I; L_0^2(\Omega))$. We then have $\mathcal{U} = H^1(\Omega)^d$, $\mathcal{H} = L^2(\Omega)^d$, $\mathcal{Q} = L_0^2(\Omega) \times H^{\frac{1}{2}}(\partial\Omega)^d$, $a(u, v) := \int_\Omega \nabla u : \nabla v$, $b(u, (q_1, q_2)) = \int_\Omega \operatorname{div} u \, q_1 + (\operatorname{tr} u, q_2)_{H^{\frac{1}{2}}(\partial\Omega)}$, $\mathcal{V} = \{\, u \in H_0^1(\Omega)^d \mid \operatorname{div} u = 0 \,\}$. Note that $u \in H^1(\Omega)^d \to (\operatorname{div} u, \operatorname{tr} u) \in L_0^2(\Omega) \times H^{\frac{1}{2}}(\Omega)^d$ is surjective. The Stokes equation in constrained formulation is given by

$$
\begin{aligned}
u' - \Delta u &= f &&\text{in } L^2(I; \mathcal{V}') \\
\operatorname{div} u &= g &&\text{in } L^2(I; L_0^2(\Omega)) \\
\operatorname{tr} u &= h &&\text{in } L^2(I; H^{\frac{1}{2}}(\partial\Omega)^d).
\end{aligned}
$$

The assumptions (2.4), (2.7), (2.8) are satisfied. For the mixed formulation we take $\mathcal{Q}_1 = L_0^2(\Omega)$, $\mathcal{Q}_2 = H^{\frac{1}{2}}(\Omega)^d$, $b_1(u, q_1) = \int_\Omega \operatorname{div} u \, q_1$, $b_2(u, q_2) = (\operatorname{tr} u, q_2)_{H^{\frac{1}{2}}(\partial\Omega)}$. Note that $\mathcal{V}_2 = H_0^1(\Omega)^d$.

REMARK 2.1. In the numerical approximation of these parabolic problems one has to consider discretization in space and in time. In this paper we focus on the time discretization. We now briefly discuss the space discretization and continue this discussion in Section 6. The two formulations (2.6) and (2.14) lead to two different numerical schemes. In the former case one has to replace the space $\mathcal{V}$ by a suitable discrete (e.g., finite element) space. In the mixed formulation the spaces $\mathcal{V}_2$ and $\mathcal{Q}_1$ have to be replaced by a suitable pair of discrete (finite element) spaces.

In the specific case of the Stokes problem this leads to two different well-known finite element approaches. Numerous conforming and non-conforming finite element (or wavelet) schemes for space discretization of the Stokes problem exist. By far most of these schemes use the mixed formulation (2.14), cf. [15, 12, 27, 4, 8]. In that setting we need a pair of discrete spaces $\mathcal{V}_2^h, \mathcal{Q}_1^h$ for approximation of functions from the spaces $\mathcal{V}_2 = H_0^1(\Omega)^d$ and $\mathcal{Q}_1 = L_0^2(\Omega)$. Extensive literature concerning the choice of such a pair of spaces is available, cf. [12, 15] and the references therein. Recently so-called pressure robust mixed methods have been studied, e.g. [24, 5, 26]. To treat the Stokes problem by means of the constrained formulation (2.14) a discrete (finite element or wavelet) space $\mathcal{V}^h$ with divergence-free functions is required. Examples of such discrete spaces are treated in, e.g., [31, 35, 14, 19, 25]. By far most of the error analyses of these methods (for both approaches) known in the liteature result in bounds for the *spatial* discretization error in case of a stationary problem and for the discretization error in the time-dependent *semidiscrete* (i.e., discrete in space only) problem. In only very few papers, e.g., [1, 2], the error in the fully discrete problem is treated. In these papers, however, only the case with homogeneous constraints (i.e., $\operatorname{div} u = 0$ and homogeneous Dirichlet boundary conditions) is treated and the error in the discretization of the Lagrange multiplier (pressure) is not considered or only suboptimal error bounds are derived.

We also note that instead of these popular Rothe or method of lines techniques one can apply a direct full space-time approach, e.g. the recent papers [29, 30].

In this paper we do not compare all these different approaches. We present and analyze a finite element method that is of Rothe type. For discretization in time we apply a DG finite element method. An important aspect is that this method, which is a modification of the standard DG time discretization method ([32]) allows the treatment of time-dependent inhomogeneous constraints with an optimal order of accuracy. For the semidiscrete problem (i.e., discrete in time only) optimal discretization error bounds for both approaches (constrained and mixed) are derived. In the mixed formulation we derive optimal bounds for both the primal variable u and the Lagrange multiplier p.

REMARK 2.2. In the analysis it is essential that the constraint operator B does not depend on t. Above we also assumed that A is independent of t. This assumption, however, is only essential for superconvergence and Subsection 5.1, all other results remain valid for a time-dependent A. Furthermore, the analysis also applies to a generalization of the problem (2.6) where the term $u'(t)$ is replaced by $Mu'(t)$ with M a symmetric, elliptic bounded linear operator on $\mathcal{H}$. Without loss of generality we can assume that $Mu(v) = (u, v)_{\mathcal{H}}$. An example of the latter generalization is given in Subsection 7.2.

3. Discontinuous Galerkin discretization. In this section we present DG time discretization methods for the variational problems (2.6) and (2.14). First some notation is introduced. We take a fixed $q \in \mathbb{N}$, $q \geq 1$. In the discretization we will use polynomials of degree $q - 1$ (q degrees of freedom) in time. The space of polynomials of degree $q - 1$ is denoted by $\mathcal{P}_{q-1}$. For $N \in \mathbb{N}$, introduce $0 = t_0 < \cdots < t_N = T$, $I_n = (t_{n-1}, t_n]$, and $k_n = |I_n|$ for $n = 1, \ldots, N$. For simplicity, we assume that $k = \max_n k_n \leq 1$. We define the broken spaces

$$\mathcal{P}^b(I) = \bigoplus_{n=1}^{N} \mathcal{P}_{q-1}(I_n) \subset L^2(I), \quad H^{1,b}(I) := \bigoplus_{n=1}^{N} H^1(I_n)$$

$$\mathcal{P}^b(I; \hat{\mathcal{H}}) := \mathcal{P}^b(I) \otimes \hat{\mathcal{H}}, \quad H^{1,b}(I; \hat{\mathcal{H}}) := H^{1,b} \otimes \hat{\mathcal{H}},$$

with $\hat{\mathcal{H}}$ a given Hilbert space. For $U \in H^{1,b}(I; \hat{\mathcal{H}})$ and $n = 1, \ldots, N$ we will write $U^n = U|_{I_n}(t_n)$, $U_+^{n-1} = \lim_{t \downarrow t_{n-1}} U|_{I_n}(t)$. We also define $[U]^n = U_+^n - U^n$ for $n = 1, \ldots, N - 1$. We define U' by taking the derivative on each interval I_n:

$$U' = \sum_{n=1}^{N} U|'_{I_n} \chi_{I_n},$$

with χ_{I_n} the characteristic function for I_n. We define the following bilinear form on $H^{1,b}(I; \hat{\mathcal{H}})$ which corresponds to a discrete time derivative:

$$(Y, X) \mapsto D_{\hat{\mathcal{H}}}(Y, X) := \sum_{n=1}^{N} \int_{I_n} (Y', X)_{\hat{\mathcal{H}}} + \sum_{n=1}^{N-1} ([Y]^n, X_+^n)_{\hat{\mathcal{H}}} + (Y_+^0, X_+^0)_{\hat{\mathcal{H}}}$$

and we define

$$(Y, X) \mapsto D_{\hat{\mathcal{H}}}^*(Y, X) := \sum_{n=1}^{N} \int_{I_n} (Y, X')_{\hat{\mathcal{H}}} + \sum_{n=1}^{N-1} (Y^n, [X]^n)_{\hat{\mathcal{H}}} - (Y^N, X^N)_{\hat{\mathcal{H}}}.$$

We will need some basic properties for $D_{\hat{\mathcal{H}}}$ and $D_{\hat{\mathcal{H}}}^*$. These results are standard and can be found in [32, Chapter 12].

LEMMA 3.1. *The following holds for any Hilbert space $\hat{\mathcal{H}}$:*

$$D_{\hat{\mathcal{H}}}(u,v) = -D_{\hat{\mathcal{H}}}^*(u,v) \quad \text{for all} \ \ u,v \in H^{1,b}(I;\hat{\mathcal{H}}) \tag{3.1}$$

$$D_{\hat{\mathcal{H}}}(u,u) \geq \frac{1}{2}\|u^N\|_{\hat{\mathcal{H}}}^2 \quad \text{for all} \ \ u \in H^{1,b}(I;\hat{\mathcal{H}}). \tag{3.2}$$

Proof. Equation (3.1) follows by integration by parts on each interval I_n. The second part follows from

$$2D_{\hat{\mathcal{H}}}(u,u) = \sum_{n=1}^{N} 2\int_{I_n} (u',u)_{\hat{\mathcal{H}}} + \sum_{n=1}^{N-1} 2([u]^n,u_+^n)_{\hat{\mathcal{H}}} + 2(u_+^0,u_+^0)_{\hat{\mathcal{H}}}$$

$$= \sum_{n=1}^{N} (\|u^n\|_{\hat{\mathcal{H}}}^2 - \|u_+^{n-1}\|_{\hat{\mathcal{H}}}^2) + \sum_{n=1}^{N-1} 2([u]^n,u_+^n)_{\hat{\mathcal{H}}} + 2\|u_+^0\|_{\hat{\mathcal{H}}}^2$$

$$= \|u^N\|_{\hat{\mathcal{H}}}^2 + \|u_+^0\|_{\hat{\mathcal{H}}}^2 + \sum_{n=1}^{N-1} \left(\|u^n\|_{\hat{\mathcal{H}}}^2 - \|u_+^n\|_{\hat{\mathcal{H}}}^2 + 2\|u_+^n\|_{\hat{\mathcal{H}}}^2 - 2(u^n,u_+^n)_{\hat{\mathcal{H}}} \right)$$

$$= \|u^N\|_{\hat{\mathcal{H}}}^2 + \|u_+^0\|_{\hat{\mathcal{H}}}^2 + \sum_{n=0}^{N-1} \left(\|u^n\|_{\hat{\mathcal{H}}}^2 + \|u_+^n\|_{\hat{\mathcal{H}}}^2 - 2(u^n,u_+^n)_{\hat{\mathcal{H}}} \right)$$

$$\geq \|u^N\|_{\hat{\mathcal{H}}}^2,$$

where we have used $(u^n - u_+^n, u^n - u_+^n)_{\hat{\mathcal{H}}} \geq 0$ in the last step. $\square$

We now recall a projection operator that plays a key role in the error analyses presented in [32, 28]. This operator is *not* used explicitly in a standard DG time discretization method applied to a parabolic problem without constraints. It will, however, be used in the DG method for the discretization of the problem with constraints (2.6), that is introduced below.

DEFINITION 3.1. *(as in [28]) Let $J = (a,b)$. For a scalar function $\phi \in L^2(J)$ which is continuous at $t = b$ we define its projection $\Pi_J^{q-1}\phi \in \mathcal{P}_{q-1}(J)$ by the q conditions*

$$\Pi_J^{q-1}\phi(b) = \phi(b), \quad \int_a^b (\Pi_J^{q-1}\phi - \phi) \cdot \psi = 0 \quad \forall \ \psi \in \mathcal{P}_{q-2}(J).$$

(for $q = 1$ only the first condition is used).

For $\phi \in L_c^2(I) := \{ \psi \in L^2(I) \mid \psi_{|I_n} \ \text{is continuous at} \ t_n, \ 1 \leq n \leq N \}$ we define the projection $\mathcal{I}_{\mathcal{P}^b} : L_c^2(I) \to \mathcal{P}^b(I)$ by $(\mathcal{I}_{\mathcal{P}^b}\phi)_{|I_n} := \Pi_{I_n}^{q-1}(\phi_{|I_n})$. Finally for any Hilbert space $\hat{\mathcal{H}}$ and $u \in L_c^2(I;\hat{\mathcal{H}}) = L_c^2(I) \otimes \hat{\mathcal{H}}$, we define $\mathcal{I}_q u := (\mathcal{I}_{\mathcal{P}^b} \otimes \text{id}_{\hat{\mathcal{H}}})u$.

For the derivation of properties of this projection operator we refer to [28, 32]. A useful characterization of the projection is given in the following lemma.

LEMMA 3.2. *[32, p. 207-208] Let $w \in H^{1,b}(I;\hat{\mathcal{H}})$ for any separable Hilbert space $\hat{\mathcal{H}}$. The solution $W \in \mathcal{P}^b(I;\hat{\mathcal{H}})$ of*

$$D_{\hat{\mathcal{H}}}(W,X) = D_{\hat{\mathcal{H}}}(w,X) \quad \text{for all} \ \ X \in \mathcal{P}^b(I;\hat{\mathcal{H}}) \tag{3.3}$$

fulfills $W = \mathcal{I}_q w$.

Proof. By taking an orthogonal basis of $\hat{\mathcal{H}}$ it suffices to establish this for the scalar version of (3.3), which reads as follows: for given $f \in H^{1,b}(I)$ find $F \in \mathcal{P}^b(I)$

such that for all $n = 1, \ldots, N$:

$$\int_{I_n} (F' - f') \cdot 1 + (F - f)_+^{n-1} - (F - f)^{n-1} = 0 \quad \text{with} \quad (F - f)^0 := 0 \qquad (3.4)$$

$$\text{and} \quad \int_{I_n} (F' - f')(t - t_{n-1})^{k+1} = 0, \quad \text{for} \quad k = 0, \ldots, q - 2. \quad (3.5)$$

After integration (3.4) becomes

$$(F - f)^n = (F - f)^{n-1} \quad \text{with} \quad (F - f)^0 := 0.$$

From this we obtain $F(t_n) = f(t_n)$ for all $n = 1, \ldots, N$ by induction. Integration by parts in (3.5), together with $F(t_n) = f(t_n)$ shows that (3.5) is equivalent to

$$\int_{I_n} (F - f) \cdot p = 0 \quad \text{for all } p \in \mathcal{P}_{q-2}(I_n).$$

Hence $F = \mathcal{I}_{\mathcal{P}^b} f$, cf. Definition 3.1. $\square$

We now describe the DG time discretizations of the constrained and mixed formulations (2.6) and (2.14), respectively. We introduce the notation

$$K(u, v) := D_{\mathcal{H}}(u, v) + \int_I a(u, v), \qquad u, v \in H^{1,b}(I; \mathcal{H}) \cap L^2(I; \mathcal{U}).$$

Time-discrete constrained formulation. The discrete (in time) version of (2.6) reads: find $U \in \mathcal{P}^b(I; \mathcal{U})$ such that

$$K(U, X) = \int_I f(X) + (u_0, X_+^0)_{\mathcal{H}} \quad \text{for all} \quad X \in \mathcal{P}^b(I; \mathcal{V}) \qquad (3.6)$$

$$\int_I b(U, R) = \int_I (\mathcal{I}_q g)(R) \quad \text{for all} \quad R \in \mathcal{P}^b(I; \mathcal{Q}). \qquad (3.7)$$

Time-discrete mixed formulation. The discrete (in time) version of (2.14) reads: find $U \in \mathcal{P}^b(I; \mathcal{U})$, $P \in \mathcal{P}^b(I; \mathcal{Q}_1)$ such that

$$K(U, X) + \int_I b_1(X, P) = \int_I f(X) + (u_0, X_+^0)_{\mathcal{H}} \quad \text{for all} \quad X \in \mathcal{P}^b(I; \mathcal{V}_2) \qquad (3.8)$$

$$\int_I b(U, R) = \int_I (\mathcal{I}_q g)(R) \quad \text{for all} \quad R \in \mathcal{P}^b(I; \mathcal{Q}). \qquad (3.9)$$

Note that in the discretization of the constraints the projection operator $\mathcal{I}_q$ is used, cf. Remark 3.1 below. If B corresponds to a trace operator (boundary condition) this is the usual treatment of an inhomogeneous Dirichlet boundary condition, *but with the Dirichlet data projected in a suitable manner in time by means of $\mathcal{I}_q$*. In case of a Stokes problem with a homogeneous Dirichlet boundary condition, the condition in (3.9) is the divergence constraint with *projected data $\mathcal{I}_q g$*.

REMARK 3.1. The formulations of the constraint equations in (3.7) and (3.9) are given in variational form because this is closest to the actual inplementation of the method, cf. Subsection 6.2. For the analysis below it is useful to introduce an equivalent operator formulation. In operator form the equations (3.7) and (3.9) are equivalent to $BU = \mathcal{I}_q g$ in $\mathcal{P}^b(I; \mathcal{Q}')$. Due to the tensor product structure of $\mathcal{I}_q$ and the facts that B does not depend on t and $U \in \mathcal{P}^b(I; \mathcal{U})$ we get $BU = B\mathcal{I}_q U =$

$\mathcal{I}_q BU$. Hence the constraint conditions (3.7) and (3.9) have the equivalent operator formulation

$$\mathcal{I}_q(BU - g) = 0 \quad \text{in} \quad \mathcal{P}^b(I; \mathcal{Q}').$$

In the error analysis we use this more compact representation of the discrete constraints. In the mixed formulation, for the case $b_2 \neq 0$, cf. Example 2.3, it is natural to split the constraint equation (3.9) in two parts:

$$\int_I b_1(U, R) = b(U, (R, 0)) = \int_I (\mathcal{I}_q g_1)(R) \quad \text{for all} \quad R \in \mathcal{P}^b(I; \mathcal{Q}_1), \qquad (3.10)$$

$$\int_I b_2(U, R) = b(U, (0, R)) = \int_I (\mathcal{I}_q g_2)(R) \quad \text{for all} \quad R \in \mathcal{P}^b(I; \mathcal{Q}_2), \qquad (3.11)$$

where $g = (g_1, g_2)$. Note that in (3.10) the same bilinear form $b_1(\cdot, \cdot)$ and Lagrange multiplier space $\mathcal{P}^b(I; \mathcal{Q}_1)$ as in (3.8) are used.

In the following lemma we derive consistency and well-posedness of these discretizations.

LEMMA 3.3. *Assume that the solution u of (2.6) has regularity $u \in H^1(I; \mathcal{H})$ and assume that $g \in L_c^2(I; \mathcal{Q}')$. The time-discrete constrained and mixed formulations (3.6)-(3.7) and (3.8)-(3.9) are consistent, i.e.:*

$$K(u, X) = \int_I f(X) + (u_0, X_+^0)_{\mathcal{H}} \quad \text{for all} \quad X \in \mathcal{P}^b(I; \mathcal{V}) \qquad (3.12)$$

$$\mathcal{I}_q(Bu - g) = 0, \qquad (3.13)$$

and

$$K(u, X) + \int_I b_1(X, p) = \int_I f(X) + (u_0, X_+^0)_{\mathcal{H}} \quad \text{for all} \quad X \in \mathcal{P}^b(I; \mathcal{V}_2) \qquad (3.14)$$

$$\mathcal{I}_q(Bu - g) = 0 \qquad (3.15)$$

hold. For $f \in L^2(I; \mathcal{V}')$ the time-discrete problem (3.6)-(3.7) has a unique solution U. If f has regularity $f \in L^2(I; \mathcal{V}_2')$ then there is a unique $P \in \mathcal{P}^b(I; \mathcal{Q}_1)$ such that with the unique solution U of (3.6)-(3.7) the pair (U, P) is the unique solution of the time-discrete mixed problem (3.8)-(3.9).

Proof. The consistency properties are evident (note that due to continuity in time of u the jump terms in $D_{\mathcal{H}}(\cdot, \cdot)$ vanish). Existence of a discrete solution of (3.6)-(3.7) is proved by a standard shift argument as follows. Using that $B : \mathcal{U} \to \mathcal{Q}'$ is surjective, we find a $v \in L_c^2(I; \mathcal{U})$ which satisfies the constraint $Bv = g$. Take $V := \mathcal{I}_q v \in \mathcal{P}^b(I; \mathcal{U})$. Due to ellipticity of $a(\cdot, \cdot)$ on $\mathcal{V}$ and $D_{\mathcal{H}}(X, X) \geq 0$ for all $X \in \mathcal{P}^b(I; \mathcal{U})$ the variational problem: find $W \in \mathcal{P}^b(I; \mathcal{V})$ such that

$$K(W, X) = \int_I f(X) + (u_0, X_+^0)_{\mathcal{H}} - K(V, X) \quad \text{for all} \quad X \in \mathcal{P}^b(I; \mathcal{V})$$

has a unique solution. Take $U := W + V \in \mathcal{P}^b(I; \mathcal{U})$. Then U satisfies (3.6) and we have

$$\mathcal{I}_q(BU - g) = \mathcal{I}_q(BV - g) = \mathcal{I}_q(B\mathcal{I}_q v - g) = \mathcal{I}_q(\mathcal{I}_q Bv - g) = \mathcal{I}_q(Bv - g) = 0,$$

hence also the discrete constraint is satisfied. Suppose that there are two solutions U_1, U_2 of (3.6)-(3.7). Then $BU_i = B(\mathcal{I}_q U_i) = \mathcal{I}_q BU_i = \mathcal{I}_q g$ holds, hence $U_1 - U_2 \in L^2(I; \mathcal{V})$. From the ellipticity of $K(\cdot, \cdot)$ on $\mathcal{P}^b(I; \mathcal{V})$ we get $U_1 = U_2$.

It remains to prove that (3.8)-(3.9) has a unique solution. If (U, P) solves (3.8)-(3.9), then U solves (3.6)-(3.7). Conversely, if U solves (3.6)-(3.7), then (U, P) solves (3.8)-(3.9) if and only if

$$B_1' P(X) = \int_I f(X) + (u_0, X_+^0)_{\mathcal{H}} - K(U, X) \quad \text{for all} \ \ X \in \mathcal{P}^b(I; \mathcal{V}_2)$$

holds. Since the right hand side vanishes for all $X \in \mathcal{P}^b(I; \mathcal{V})$ and $B_1' : \mathcal{Q}_1 \mapsto \mathcal{V}_2^0$ is an isomorphism, there exists a unique $P \in \mathcal{P}^b(I; \mathcal{Q}_1)$ which satisfies this relationship. □

REMARK 3.2. In the analysis below we will see that the projection operator $\mathcal{I}_q$ used in the time-discrete constraints (3.7) and (3.9) is essential for optimal order convergence results. Applying Lemma 3.2 to (3.7) (or (3.9)) shows that (3.7) is equivalent to $D_{\mathcal{Q}'} BU = D_{\mathcal{Q}'} g$, which is the discrete analogon of $Bu' = g'$ with $Bu(0) = g(0)$. The discretization of the constraint by $D_{\mathcal{Q}'} BU = D_{\mathcal{Q}'} g$ is very similar to the index reduction approach used in [3] for obtaining accurate Runge-Kutta discretizations of DAEs with differentiation index 2 (such as the Stokes equation), cf. equation (3.3c) in [3]. The projection operator $\mathcal{I}_q$ is also used in [33] to obtain optimal error bounds for the wave equation with non-homogeneous boundary conditions.

4. Error analysis for time-discrete constrained formulation. In this section we derive optimal discretization error bounds for the discrete constrained formulation (3.6)-(3.7). We first derive global bounds in the energy norm and then give a superconvergence result.

4.1. Optimal global error bounds. We apply a standard argument as in [28] and show that the discretization error can be bounded by the projection error. For this projection there are error bounds available.

THEOREM 4.1. *Assume that the solution u of (2.6) has regularity $u \in L_c^2(I; \mathcal{U}) \cap H^1(I; \mathcal{H})$ and let $U \in \mathcal{P}^b(I; \mathcal{U})$ be the solution of (3.6)-(3.7). The following holds:*

$$\|u(T) - U(T)\|_{\mathcal{H}} + \|u - U\|_{L^2(I;\mathcal{U})} \leq (1 + \sqrt{2}c_\gamma \Gamma)\|u - \mathcal{I}_q u\|_{L^2(I;\mathcal{U})}, \qquad (4.1)$$

with $c_\gamma := \max\{\frac{1}{\gamma}, 1\}$.

Proof. Define $E := U - \mathcal{I}_q u$. Note that

$$\begin{aligned} BE &= BU - B(\mathcal{I}_q u) = B(\mathcal{I}_q U) - B(\mathcal{I}_q u) \\ &= \mathcal{I}_q(BU - Bu) = \mathcal{I}_q(BU - g) = 0. \end{aligned} \qquad (4.2)$$

Hence $E \in \mathcal{P}^b(I; \mathcal{V})$ can be used as a test function in (3.6). This then yields, using the ellipticity of $a(\cdot, \cdot)$ on $\mathcal{V}$ and $D_{\mathcal{H}}(E, E) \geq \frac{1}{2}\|E(T)\|_{\mathcal{H}}^2$, with $c_\gamma = \max\{\frac{1}{\gamma}, 1\}$:

$$\frac{1}{2}\|E(T)\|_{\mathcal{H}}^2 + \|E\|_{L^2(I;\mathcal{U})}^2 \leq c_\gamma K(E, E) = c_\gamma K(u - \mathcal{I}_q u, E)$$

$$= c_\gamma \Big(D_{\mathcal{H}}(u - \mathcal{I}_q u, E) + \int_I a(u - \mathcal{I}_q u, E)\Big).$$

Using (3.3) we get

$$D_{\mathcal{H}}(u - \mathcal{I}_q u, E) = 0, \qquad (4.3)$$

11

and thus

$$\frac{1}{2}\|E(T)\|_{\mathcal{H}}^2 + \|E\|_{L^2(I;\mathcal{U})}^2 \leq c_\gamma \Gamma \|E\|_{L^2(I;\mathcal{U})} \|u - \mathcal{I}_q u\|_{L^2(I;\mathcal{U})}$$

$$\leq \frac{1}{2} c_\gamma^2 \Gamma^2 \|u - \mathcal{I}_q u\|_{L^2(I;\mathcal{U})}^2 + \frac{1}{2} \|E\|_{L^2(I;\mathcal{U})}^2$$

holds. This yields $\|E(T)\|_{\mathcal{H}} + \|E\|_{L^2(I;\mathcal{U})} \leq \sqrt{2} c_\gamma \Gamma \|u - \mathcal{I}_q u\|_{L^2(I;\mathcal{U})}$. Using this, a triangle inequality and $u(T) - (\mathcal{I}_q u)(T) = 0$ we obtain the result (4.1). $\square$

REMARK 4.1. For this analysis to work and the projection error bound (4.1) to hold it is crucial, cf. (4.2) that in the discretization of the constraint we use $BU = \mathcal{I}_q g$ and not $BU = g$ or some other interpolation (in time) of the data g. In Subsection 7.1 we will present a numerical experiment where the use of $BU = g$ instead of $BU = \mathcal{I}_q g$ leads to sub-optimal results.

Results for the projection error $\|u - \mathcal{I}_q u\|_{L^2(I;\mathcal{U})}$ are known in the literature [28, Theorem 3.10]. Using these results we obtain the following optimal discretization error bound.

THEOREM 4.2. Let $U \in \mathcal{P}^b(I;\mathcal{U})$ be the solution of (3.6)-(3.7), and assume that the solution u of (2.6) has smoothness $u \in H^m(I;\mathcal{U})$ for an m with $1 \leq m \leq q$. The following holds:

$$\|u(T) - U(T)\|_{\mathcal{H}} + \|u - U\|_{L^2(I;\mathcal{U})} \leq c \left(\sum_{n=1}^N k_n^{2m} \|u^{(m)}\|_{L^2(I_n;\mathcal{U})}^2 \right)^{\frac{1}{2}} \tag{4.4}$$

$$\leq c k^m \|u^{(m)}\|_{L^2(I;\mathcal{U})} \tag{4.5}$$

for some $c > 0$ which only depends on q, γ and Γ.

REMARK 4.2. Typically one is interested in the case $q = m$. The other cases will however also be of some use in what follows. For the homogeneous case $Bu = 0$ the result (4.4) is the same as the one in [32, Theorem 12.1].

In applications the constraint $Bu = g$ may be incompatible this the initial condition $u(0) = u_0$. This leads to low regularity of the solution u at $t = 0$, i.e. $\|u^{(m)}\|_{L^2(I_n;\mathcal{U})}^2$ is unbounded for larger m and small n. In such a setting the step size k_n in (4.4) needs to be taken sufficiently small in the initial steps. In the specific case of Example 2.1 and Example 2.3 this is the case if the Dirichlet boundary condition is incompatible with the initial condition, cf. [28].

We also derive a bound for the error in the time derivative, which will be used in the analysis of superconvergence.

THEOREM 4.3. Assume that the solution u of (2.6) has regularity $u \in H^1(I;\mathcal{H})$ and assume that $g \in L_c^2(I;\mathcal{Q}')$. Let $U \in \mathcal{P}^b(I;\mathcal{U})$ be the solution of (3.6)-(3.7). The following holds:

$$\|u' - U'\|_{L^2(I;\mathcal{V}')} \leq c \left(\|u' - (\mathcal{I}_q u)'\|_{L^2(I;\mathcal{V}')} + \|u - U\|_{L^2(I;\mathcal{U})} \right) \tag{4.6}$$

for some $c > 0$ which only depends on q and Γ. If we additionally assume that u has smoothness $u \in H^{m-1}(I;\mathcal{U}) \cap H^m(I;\mathcal{V}')$ for an m with $2 \leq m \leq q$, then the following

12

holds:

$$\|u' - U'\|_{L^2(I;\mathcal{V}')} \le c' \left(\sum_{n=1}^{N} k_n^{2m-2} (\|u^{(m)}\|^2_{L^2(I_n;\mathcal{V}')} + \|u^{(m-1)}\|^2_{L^2(I_n;\mathcal{U})}) \right)^{\frac{1}{2}}$$

$$\le c' k^{m-1} (\|u^{(m)}\|_{L^2(I;\mathcal{V}')} + \|u^{(m-1)}\|_{L^2(I;\mathcal{U})}), \qquad (4.7)$$

for some $c' > 0$ which only depends on q, γ and Γ.

Proof. We will use $E = U - \mathcal{I}_q u \in \mathcal{P}^b(I;\mathcal{H})$. Using the Riesz representation theorem in $\mathcal{V}$ (with $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_{\mathcal{U}}$), we find $\tilde{E} \in \oplus_{n=1}^{N} \mathcal{P}_{q-2}(I_n;\mathcal{V})$ such that

$$\|E'(t)\|^2_{\mathcal{V}'} = (E'(t), \tilde{E}(t))_{\mathcal{H}} = \|\tilde{E}(t)\|^2_{\mathcal{U}} \quad \text{for almost all } t \in I.$$

We define $X = \sum_{n=1}^{N} k_n^{-1}(t - t_{n-1})\tilde{E}(t)\chi_{I_n} \in \mathcal{P}^b(I;\mathcal{V})$ and using Lemma 3.2 and the Galerkin orthogonality, cf. Lemma 3.3, we get

$$D_{\mathcal{H}}(E, X) = D_{\mathcal{H}}(U - \mathcal{I}_q u, X) = D_{\mathcal{H}}(U - u, X)$$

$$= K(U - u, X) - \int_I a(U - u, X) = \int_I a(u - U, X).$$

This yields

$$D_{\mathcal{H}}(E, X) \le \sum_{n=1}^{N} \int_{I_n} k_n^{-1}(t - t_{n-1})\Gamma\|u - U\|_{\mathcal{U}}\|\tilde{E}\|_{\mathcal{U}}$$

$$\le \frac{1}{2} \sum_{n=1}^{N} \int_{I_n} k_n^{-1}(t - t_{n-1}) \left(\Gamma^2\|u - U\|^2_{\mathcal{U}} + \|E'\|^2_{\mathcal{V}'} \right).$$

Noting that the left-hand side is

$$D_{\mathcal{H}}(E, X) = \sum_{n=1}^{N} \int_{I_n} k_n^{-1}(t - t_{n-1})(E', \tilde{E})_{\mathcal{H}} = \sum_{n=1}^{N} \int_{I_n} k_n^{-1}(t - t_{n-1})\|E'\|^2_{\mathcal{V}'},$$

we get

$$\sum_{n=1}^{N} \int_{I_n} k_n^{-1}(t - t_{n-1})\|E'\|^2_{\mathcal{V}'} \le \sum_{n=1}^{N} \int_{I_n} k_n^{-1}(t - t_{n-1})\Gamma^2\|u - U\|^2_{\mathcal{U}} \le \Gamma^2\|U - u\|^2_{L^2(I;\mathcal{U})}.$$

Using that the L^1-norm on $\mathcal{P}_{2q-4}([0,1])$ is equivalent to the norm $\phi \mapsto \int_0^1 t|\phi(t)|dt$, we get

$$\int_I \|E'\|^2_{\mathcal{V}'} = \sum_{n=1}^{N} \int_{I_n} \|E'\|^2_{\mathcal{V}'} \le \sum_{n=1}^{N} \int_{I_n} C k_n^{-1}(t - t_{n-1})\|E'\|^2_{\mathcal{V}'} \le C\Gamma^2\|U - u\|^2_{L^2(I;\mathcal{U})},$$

for some constant $C > 0$ which depends only on q. Using a triangle inequality completes the proof of (4.6). Results for the projection error $\|u' - (\mathcal{I}_q u)'\|_{L^2(I;\mathcal{V}')}$ are known in the literature [32, p. 214]. Using these results and Theorem 4.1 we obtain the optimal discretization error bound (4.7). □

13

4.2. Optimal superconvergence result. As is known from the literature (see [32, Theorem 12.3]), for the case with a *homogeneous* constraint in (2.6) we get a superconvergence result provided the solution u is sufficiently smooth. The analysis in [32, Theorem 12.3] is not directly applicable to the case with an inhomogeneous constraint. Note that in the latter case the projection operator $\mathcal{I}_q$ is used in the discretization method, cf. (3.7). In this subsection we derive an optimal superconvergence result for the discretization (3.6)-(3.7). For this we first introduce an abstract notion of regularity, which is very similar to the one used in [32]. We define $\mathcal{V}^0 := \overline{\mathcal{V}}^{\mathcal{H}}$ (with norm $\|\cdot\|_{\mathcal{H}}$), $\mathcal{V}^1 := \mathcal{V}$ (with norm $\|\cdot\|_{\mathcal{U}}$), which induces the Gelfand triple $\mathcal{V} \hookrightarrow \mathcal{V}^0 \cong (\mathcal{V}^0)' \hookrightarrow \mathcal{V}'$. Furthermore

$$\mathcal{V}^2 := \{\, u \in \mathcal{V} \mid Au \in (\mathcal{V}^0)' \,\}.$$

Based on the Riesz representation theorem in $\mathcal{V}^0$, for any $u \in \mathcal{U}$ with $Au \in (\mathcal{V}^0)'$ we define $\tilde{L}u \in \mathcal{V}^0$ by $(\tilde{L}u, v)_{\mathcal{H}} = a(u, v)$ for all $v \in \mathcal{V}$. Let us consider the restriction of $L := \tilde{L}|_{\mathcal{V}^2}$, i.e., $L : \mathcal{V}^2 \to \mathcal{V}^0$ with

$$(Lu, v)_{\mathcal{H}} = (Au)(v) = a(u, v) \quad \text{for all} \ \ u \in \mathcal{V}^2, v \in \mathcal{V}. \tag{4.8}$$

For $u \in \mathcal{V}^2$ with $Lu = 0$ it follows that $a(u, v) = 0$ for all $v \in \mathcal{V}$, hence, $u = 0$. Take $f \in \mathcal{V}^0$. There exists a unique $u \in \mathcal{V}$ such that $a(u, v) = (f, v)_{\mathcal{H}}$ for all $v \in \mathcal{V}$. This implies $\|Au\|_{(\mathcal{V}^0)'} = \|f\|_{\mathcal{H}}$, hence $u \in \mathcal{V}^2$. We conclude that $L : \mathcal{V}^2 \to \mathcal{V}^0$ is a bijection. Its inverse $L^{-1} : \mathcal{V}^0 \to \mathcal{V}^2 \subset \mathcal{V}^0$ is symmetric, bounded and positive. The unique square root operator of L^{-1} (that commutes with L^{-1}) is denoted by $L^{-\frac{1}{2}} : \mathcal{V}^0 \to \mathcal{V}^0$. We define $L^{\frac{1}{2}} := L^{-\frac{1}{2}}L : \mathcal{V}^2 \to \mathcal{V}^0$, hence, $(L^{\frac{1}{2}}v, L^{\frac{1}{2}}v)_{\mathcal{H}} = (Lv, v)_{\mathcal{H}} = a(v, v)$ for all $v \in \mathcal{V}^2$. The definition of $L^{\frac{1}{2}}$ can be extended to $\mathcal{V}$. First note that $A : \mathcal{V} \to \mathcal{V}'$ is a homeomorphism and $\mathcal{V}^2$ is defined as the preimage of $(\mathcal{V}^0)'$. Since $(\mathcal{V}^0)' \subset \mathcal{V}'$ is dense, it follows that $\mathcal{V}^2 \subset \mathcal{V}$ is dense. Using the ellipticity and continuity of $a(\cdot, \cdot)$ we see that

$$\gamma^{\frac{1}{2}}\|v\|_{\mathcal{U}} \le a(v, v)^{\frac{1}{2}} = \|L^{\frac{1}{2}}v\|_{\mathcal{H}} \le \Gamma^{\frac{1}{2}}\|v\|_{\mathcal{U}} \quad \text{for all} \ \ v \in \mathcal{V}^2.$$

Thus, using the density of $\mathcal{V}^2$ in $\mathcal{V}$, we conclude that $L^{\frac{1}{2}} : \mathcal{V} \to \mathcal{V}^0$ is a homeomorphism. Since $L^{\frac{1}{2}}$ is symmetric (due to $L^{-\frac{1}{2}}L = LL^{-\frac{1}{2}}$), the dual norm equivalence reads $\gamma^{\frac{1}{2}}\|v\|_{\mathcal{H}} \le \|L^{\frac{1}{2}}v\|_{\mathcal{V}'} \le \Gamma^{\frac{1}{2}}\|v\|_{\mathcal{H}}$ for all $v \in \mathcal{V}$.

For $m \in \mathbb{N}$, $m > 2$, we define

$$\mathcal{V}^m = \{\, u \in \mathcal{V}^2 \mid Lu \in \mathcal{V}^{m-2} \,\}.$$

Note that $L^{\frac{1}{2}}Lv = LL^{\frac{1}{2}}$ for all $v \in \mathcal{V}^3$ and $\mathcal{D}(L^{\frac{m}{2}}) = \mathcal{V}^m$, $m \in \mathbb{N}$. As a measure for regularity we define

$$|u|_m := \sup_{v \in \mathcal{V}^m} \frac{(u, L^{\frac{m}{2}}v)_{\mathcal{H}}}{\|v\|_{\mathcal{H}}}, \quad u \in \mathcal{H}, \quad m \in \mathbb{N}$$

(note that $|u|_m = \infty$ is allowed).

Below we use the notation $L^{\frac{m}{2}}u := (\mathrm{id}_{L^2(I)} \otimes L^{\frac{m}{2}})u$, $\tilde{L}u = (\mathrm{id}_{L^2(I)} \otimes \tilde{L})u$, for $m \in \mathbb{Z}$ and for functions u for which the right hand-side expressions are well-defined. For the analysis we introduce a standard dual *homogeneous* problem, i.e., source term $f = 0$ and with a homogeneous constraint $Bu = 0$; due to duality t is replaced by

$T - t$ and we prescribe an initial value at $t = T$. This leads to the following problem formulation: find $z \in W^1(\mathcal{V}; \mathcal{V}')$ with $z(T) = \phi \in \mathcal{V}^0$ and

$$-z'(t) + Az(t) = 0 \quad \text{in} \ \ L^2(I; \mathcal{V}'). \tag{4.9}$$

This problem is well posed. In the analysis below we need that the solution z has regularity $z \in H^1(I; \mathcal{H})$. It is easy to see that z has this regularity property if $\phi \in \mathcal{V}$. Indeed, for $\phi \in \mathcal{V}$ let $y \in W^1(\mathcal{V}; \mathcal{V}')$ be the solution of the parabolic problem (4.9) with $y(T) = L^{\frac{1}{2}}\phi \in \mathcal{V}^0$, then $z = L^{-\frac{1}{2}}y$ solves (4.9) with $z(T) = \phi$ and $z' = L^{-\frac{1}{2}}y' \in L^2(I; \mathcal{H})$. Hence, for $\phi \in \mathcal{V}$ we have $z \in H^1(I; \mathcal{H})$. This implies that (4.9) holds in $L^2(I, (\mathcal{V}^0)')$.

The corresponding discrete problem is as follows: determine $Z \in \mathcal{P}^b(I; \mathcal{V})$ such that

$$K(X, Z) = -D_{\mathcal{H}}^*(X, Z) + \int_I a(Z, X) = (\phi, X^N)_{\mathcal{H}} \quad \text{for all} \ \ X \in \mathcal{P}^b(I; \mathcal{V}). \tag{4.10}$$

This is the analogue of (3.6) for (4.9). This *homogeneous* backward problem has been discussed in detail in [32, p. 212-213]. If we replace t by $T - t$, then we can apply the results from the previous subsection to z and Z. In particular we can derive the following error estimate.

LEMMA 4.4. *Let z and Z be the solution of (4.9) and (4.10), respectively and assume that $z \in H^1(I; \mathcal{H})$. The following error estimate holds for $1 \le m \le q$:*

$$\|L^{-m+1+\frac{1}{2}}(z - Z)\|_{L^2(I;\mathcal{H})} + \|L^{-m+\frac{1}{2}}(z' - Z')\|_{L^2(I;\mathcal{H})} \le Ck^{m-1}\|\phi\|_{\mathcal{H}} \tag{4.11}$$

for some $C > 0$ which only depends on q, γ and Γ.

Proof. First consider the case $m = 1$. The homogeneous problem (4.9) can be formulated as $z' = Lz$ in $L^2(I; \mathcal{V}^0)$. From $(z', z)_{L^2(I;\mathcal{H})} = (L^{\frac{1}{2}}z, L^{\frac{1}{2}}z)_{L^2(I;\mathcal{H})}$ and integration by parts, we get $\|L^{\frac{1}{2}}z\|_{L^2(I;\mathcal{H})}^2 \le \frac{1}{2}\|\phi\|_{\mathcal{H}}^2$. Similarly, taking $X = Z$ in (4.10) and applying Lemma 3.1 shows that

$$\frac{1}{2}\|Z^N\|_{\mathcal{H}}^2 + \|L^{\frac{1}{2}}Z\|_{L^2(I;\mathcal{H})}^2 \le -D_{\mathcal{H}}^*(Z, Z) + \int_I a(Z, Z) \le \frac{1}{2}\|\phi\|_{\mathcal{H}}^2 + \frac{1}{2}\|Z^N\|_{\mathcal{H}}^2.$$

We conclude that

$$\|L^{\frac{1}{2}}z\|_{L^2(I;\mathcal{H})}^2 + \|L^{\frac{1}{2}}Z\|_{L^2(I;\mathcal{H})}^2 \le \|\phi\|_{\mathcal{H}}^2. \tag{4.12}$$

From (4.6) and the standard bound $\|z' - (\mathcal{I}_q z)'\|_{L^2(I;\mathcal{V}')} \le \tilde{c}\|z'\|_{L^2(I;\mathcal{V}')}$ (for some $\tilde{c}$ which depends only on q, cf. [32, p. 214]), we know that

$$\begin{aligned}
\|L^{-\frac{1}{2}}(z' - Z')\|_{L^2(I;\mathcal{H})} &\le \gamma^{-\frac{1}{2}}\|z' - Z'\|_{L^2(I;\mathcal{V}')} \\
&\le c\big(\|z'\|_{L^2(I;\mathcal{V}')} + \|z\|_{L^2(I;\mathcal{U})} + \|Z\|_{L^2(I;\mathcal{U})}\big) \\
&\le c\big(\|z'\|_{L^2(I;\mathcal{V}')} + \gamma^{-\frac{1}{2}}\|L^{\frac{1}{2}}z\|_{L^2(I;\mathcal{H})} + \gamma^{-\frac{1}{2}}\|L^{\frac{1}{2}}Z\|_{L^2(I;\mathcal{H})}\big),
\end{aligned}$$

with a constant c which only depends on q, γ and Γ. Noting that $\|z'\|_{L^2(I;\mathcal{V}')} = \|Lz\|_{L^2(I;\mathcal{V}')} \le \Gamma^{\frac{1}{2}}\|L^{\frac{1}{2}}z\|_{L^2(I;\mathcal{H})}$ and using (4.12), completes the proof for $m = 1$. We now consider the case $m \ge 2$. Let $y = L^{1-m}z$, $Y = L^{1-m}Z$ and $\psi = L^{1-m}\phi$. Note that the solution operators $\phi \to z$ and $\phi \to Z$ in (4.9) and (4.10) commute with L.

15

Thus y solves (4.9) with initial condition $y(T) = \psi$ and Y solves the corresponding discrete problem

$$K(X, Y) = -D_{\mathcal{H}}^*(X, Y) + \int_I a(Y, X) = (\psi, X^N)_{\mathcal{H}} \quad \text{for all} \ \ X \in \mathcal{P}^b(I; \mathcal{V}).$$

Note that $y^{(m-1)} = L^{m-1}y = z \in W^1(\mathcal{V}; \mathcal{V}')$. Using the estimates (4.4), (4.12) and $y^{(m-1)} = z$ we get

$$\|y - Y\|_{L^2(I;\mathcal{U})} \le ck^{m-1}\|y^{(m-1)}\|_{L^2(I;\mathcal{U})} \le ck^{m-1}\gamma^{-\frac{1}{2}}\|L^{\frac{1}{2}}z\|_{L^2(I;\mathcal{H})} \le c\gamma^{-\frac{1}{2}}k^{m-1}\|\phi\|_{\mathcal{H}}$$

and

$$\|L^{-m+1+\frac{1}{2}}(z - Z)\|_{L^2(I;\mathcal{H})} = \|L^{\frac{1}{2}}(y - Y)\|_{L^2(I;\mathcal{H})} \le ck^{m-1}\Gamma^{\frac{1}{2}}\gamma^{-\frac{1}{2}}\|\phi\|_{\mathcal{H}},$$

with a constant c which only depends on q, γ and Γ. From this we obtain the estimate for the first term in (4.11). In order to bound the second term we use (4.7), (4.12) and $y^{(m-1)} = z$:

$$\begin{aligned}
\|L^{-m+\frac{1}{2}}(z' - Z')\|_{L^2(I;\mathcal{H})} = \|L^{-\frac{1}{2}}(y' - Y')\|_{L^2(I;\mathcal{H})} &\le \gamma^{-\frac{1}{2}}\|y' - Y'\|_{L^2(I;\mathcal{V}')} \\
&\le ck^{m-1}(\|y^{(m)}\|_{L^2(I;\mathcal{V}')} + \|y^{(m-1)}\|_{L^2(I;\mathcal{U})}) \\
&\le ck^{m-1}(\Gamma^{\frac{1}{2}}\|L^{-\frac{1}{2}}z'\|_{L^2(I;\mathcal{H})} + \gamma^{-\frac{1}{2}}\|L^{\frac{1}{2}}z\|_{L^2(I;\mathcal{H})}) \\
&\le ck^{m-1}(\Gamma^{\frac{1}{2}} + \gamma^{-\frac{1}{2}})\|L^{\frac{1}{2}}z\|_{L^2(I;\mathcal{H})} \\
&\le ck^{m-1}(\Gamma^{\frac{1}{2}} + \gamma^{-\frac{1}{2}})\|\phi\|_{\mathcal{H}},
\end{aligned}$$

with a constant c which only depends on q, γ and Γ. This yields the bound for the second term in (4.11). $\square$

Using this we can derive the following result, which has a superconvergence result as an easy corollary.

THEOREM 4.5. *Assume that the solution u of (2.6) has regularity $u \in L_c^2(I;\mathcal{U}) \cap H^1(I;\mathcal{H})$ and $Au \in L_c^2(I; (\mathcal{V}^0)')$. Let $U \in \mathcal{P}^b(I;\mathcal{U})$ be the solution of (3.6)-(3.7). Let $2 \le \ell \le 2q - 1$. The following holds:*

$$\|U^N - u(T)\|_{\mathcal{H}} \le ck^{\frac{\ell-1}{2}}\left(\int_I |\tilde{L}u - \mathcal{I}_q\tilde{L}u|_{\ell-2}^2\right)^{\frac{1}{2}}$$

for some $c > 0$ which only depends on q, γ and Γ.

Proof. We introduce $E := U - u$, $\phi := E(T) \in \mathcal{V}$, $\hat{E} = U - \mathcal{I}_qu \in \mathcal{P}^b(I;\mathcal{V})$. $\theta = \mathcal{I}_qu - u = E - \hat{E}$. We take ϕ in the homogeneous backward problem (4.9). The corresponding solutions of (4.9) and (4.10) are denoted by z and Z, respectively. We define $\zeta = z - Z \in H^{1,b}(I;\mathcal{H})$. One can not apply a straightforward duality argument for the following reason. We (only) have $\mathcal{I}_qB(U-u) = 0$ and thus in general $E \notin L^2(I;\mathcal{V})$. Hence, we can *not* use E as a test function in (4.9). To overcome this problem, below we consider a transformed dual problem with endpoint condition $L^{-1}\phi$ (instead of ϕ) and use a test function $\tilde{L}E$. The precise arguments are as follows. Due the regularity assumption $Au \in L_c^2(I; (\mathcal{V}^0)')$ we have that $\tilde{L}u \in L_c^2(I;\mathcal{V}^0)$ is well-defined. Using the Galerkin property $K(U - u, \cdot) = 0$ in $\mathcal{P}^b(I, \mathcal{V}')$ we have

$$AU = D_{\mathcal{H}}(u - U, \cdot) + Au \quad \text{in} \ \mathcal{P}^b(I, \mathcal{V}').$$

Due to the density of $\mathcal{V}$ in $\mathcal{V}^0$ it follows that $AU \in \mathcal{P}^b(I;(\mathcal{V}^0)')$ and thus $\tilde{L}U \in \mathcal{P}^b(I;\mathcal{V}^0) \subset L_c^2(I;\mathcal{V}^0)$ is well-defined. It follows that $\tilde{L}E \in L_c^2(I;\mathcal{V}^0)$. Using this and that $\tilde{L}\theta \in L_c^2(I;\mathcal{V}^0)$ is well defined, we find that $L\hat{E} = \tilde{L}\hat{E} = \tilde{L}\theta - \tilde{L}E$ is well defined. Therefore, we have $\hat{E} \in L_c^2(I;\mathcal{V}^2)$. Note that due to $E(T) = \hat{E}(T) \in \mathcal{V}^2$ we have $(\tilde{L}E)(T) = \tilde{L}E(T) = LE(T) = L\phi$. The dual problem with endpoint condition $L^{-1}\phi$ has corresponding solutions of (4.9) and (4.10) given by $y := L^{-1}z$ and $Y := L^{-1}Z$. Hence,

$$(L^{-1}\phi, v)_{\mathcal{H}} = K(v,y) = D_{\mathcal{H}}(v,y) + \int_I a(v,y) = D_{\mathcal{H}}(v,y) + \int_I (v, Ly)_{\mathcal{H}} \; \forall \, v \in L^2(I;\mathcal{V}).$$

Using the density of $\mathcal{V}$ in $\mathcal{V}^0$ we obtain

$$(L^{-1}\phi, v)_{\mathcal{H}} = D_{\mathcal{H}}(v, L^{-1}z) + \int_I (v, z)_{\mathcal{H}} \quad \text{for all} \;\; v \in L^2(I;\mathcal{V}^0).$$

We use $v = \tilde{L}E$ as test function, which yields

$$\|\phi\|_{\mathcal{H}}^2 = (L^{-1}\phi, L\phi)_{\mathcal{H}} = (L^{-1}\phi, \tilde{L}E(T))_{\mathcal{H}} = D_{\mathcal{H}}(\tilde{L}E, L^{-1}z) + \int_I (\tilde{L}E, z)_{\mathcal{H}}. \quad (4.13)$$

Note that from Lemma 3.2, the Galerkin property $0 = K(E, X)$ for all $X \in \mathcal{P}^b(I;\mathcal{V})$ and $\hat{E} \in \mathcal{P}^b(I;\mathcal{V}^2)$ we get

$$0 = K(E, Z) = D_{\mathcal{H}}(E, Z) + \int_I a(E, Z) = D_{\mathcal{H}}(\hat{E}, Z) + \int_I a(E, Z)$$

$$= D_{\mathcal{H}}(L\hat{E}, L^{-1}Z) + \int_I (\tilde{L}E, Z)_{\mathcal{H}} = D_{\mathcal{H}}(\tilde{L}\hat{E}, L^{-1}Z) + \int_I (\tilde{L}E, Z)_{\mathcal{H}}$$

$$= D_{\mathcal{H}}(\tilde{L}E, L^{-1}Z) + \int_I (\tilde{L}E, Z)_{\mathcal{H}}.$$

Combing this with (4.13) and using $K(\hat{E}, \zeta) = 0$ and (3.1), we obtain:

$$\|\phi\|_{\mathcal{H}}^2 = D_{\mathcal{H}}(\tilde{L}E, L^{-1}\zeta) + \int_I (\tilde{L}E, \zeta)_{\mathcal{H}}$$

$$= D_{\mathcal{H}}(\tilde{L}\theta, L^{-1}\zeta) + \int_I (\tilde{L}\theta, \zeta)_{\mathcal{H}} + D_{\mathcal{H}}(\tilde{L}\hat{E}, L^{-1}\zeta) + \int_I (\tilde{L}\hat{E}, \zeta)_{\mathcal{H}}$$

$$= D_{\mathcal{H}}(\tilde{L}\theta, L^{-1}\zeta) + \int_I (\tilde{L}\theta, \zeta)_{\mathcal{H}} + D_{\mathcal{H}}(L\hat{E}, L^{-1}\zeta) + \int_I a(\hat{E}, \zeta)$$

$$= D_{\mathcal{H}}(\tilde{L}\theta, L^{-1}\zeta) + \int_I (\tilde{L}\theta, \zeta)_{\mathcal{H}} + D_{\mathcal{H}}(\hat{E}, \zeta) + \int_I a(\hat{E}, \zeta)$$

$$= D_{\mathcal{H}}(\tilde{L}\theta, L^{-1}\zeta) + \int_I (\tilde{L}\theta, \zeta)_{\mathcal{H}} + K(\hat{E}, \zeta) = -D_{\mathcal{H}}^*(\tilde{L}\theta, L^{-1}\zeta) + \int_I (\tilde{L}\theta, \zeta)_{\mathcal{H}}.$$

Noting that $\theta^n = 0$ for all $n = 1, \ldots, N$, we get

$$\|\phi\|_{\mathcal{H}}^2 = -\sum_{n=1}^{N} \int_{I_n} (L^{-1}\zeta', \tilde{L}\theta)_{\mathcal{H}} + \int_I (\zeta, \tilde{L}\theta)_{\mathcal{H}}. \quad (4.14)$$

By the definition of $|\cdot|_{\ell-2}$, we have

$$\|\phi\|_{\mathcal{H}}^2 \le \sum_{n=1}^{N} \int_{I_n} \|L^{-\frac{\ell}{2}}\zeta'\|_{\mathcal{H}} |\tilde{L}\theta|_{\ell-2} + \int_I \|L^{-\frac{\ell}{2}+1}\zeta\|_{\mathcal{H}} |\tilde{L}\theta|_{\ell-2}.$$

17

Using the Cauchy-Schwarz inequality, we get

$$\|\phi\|_{\mathcal{H}}^2 \leq \left(\int_I |\tilde{L}\theta|_{\ell-2}^2 \right)^{1/2} \left(\|L^{-\frac{\ell}{2}}\zeta'\|_{L^2(I;\mathcal{H})} + \|L^{-\frac{\ell}{2}+1}\zeta\|_{L^2(I;\mathcal{H})} \right). \tag{4.15}$$

We will estimate the second factor in (4.15) by using bounds from Lemma 4.4. First we consider the case that ℓ is odd. Then $m := \frac{\ell}{2} + \frac{1}{2}$ is a natural number with $1 \leq m \leq q$. We then obtain

$$\begin{aligned}
&\|L^{-\frac{\ell}{2}+1}\zeta\|_{L^2(I;\mathcal{H})} + \|L^{-\frac{\ell}{2}}\zeta'\|_{L^2(I;\mathcal{H})} \\
&= \|L^{-(m-1)+\frac{1}{2}}\zeta\|_{L^2(I;\mathcal{H})} + \|L^{-m+\frac{1}{2}}\zeta'\|_{L^2(I;\mathcal{H})} \\
&\leq Ck^{m-1}\|\phi\|_{\mathcal{H}} = Ck^{\frac{\ell-1}{2}}\|\phi\|_{\mathcal{H}},
\end{aligned} \tag{4.16}$$

and in combination with (4.15) this yields the desired result. For ℓ even we use the following property

$$\forall x \in \mathcal{V}^2 : \|L^{\frac{1}{2}}x\|_{\mathcal{H}}^2 = (Lx, x)_{\mathcal{H}} \leq \|Lx\|_{\mathcal{H}}\|x\|_{\mathcal{H}}.$$

The term $\|L^{-\frac{\ell}{2}}\zeta'\|_{L^2(I;\mathcal{H})}$ can be treated as follows, where we use the bound derived in (4.16), Lemma 4.4,

$$\begin{aligned}
\|L^{-\frac{\ell}{2}}\zeta'\|_{L^2(I;\mathcal{H})} &\leq \|L^{-\frac{\ell-1}{2}}\zeta'\|_{L^2(I;\mathcal{H})}^{\frac{1}{2}}\|L^{-\frac{\ell+1}{2}}\zeta'\|_{L^2(I;\mathcal{H})}^{\frac{1}{2}} \\
&\leq C\left(k^{\frac{\ell-2}{2}}\right)^{\frac{1}{2}}\left(k^{\frac{\ell}{2}}\right)^{\frac{1}{2}}\|\phi\|_{\mathcal{H}} = Ck^{\frac{\ell-1}{2}}\|\phi\|_{\mathcal{H}}.
\end{aligned}$$

The term $\|L^{-\frac{\ell}{2}+1}\zeta\|_{L^2(I;\mathcal{H})}$ can be treated in the same way. Thus also for ℓ even we get the bound as in (4.16). □

Results for the projection error $\int_I |\tilde{L}u - \mathcal{I}_q\tilde{L}u|_{\ell-2}^2$ are known in the literature [28, Theorem 3.10]. Using these results we obtain the following optimal discretization error bound.

THEOREM 4.6. *Assume that the solution u of (2.6) has regularity $u \in L_c^2(I;\mathcal{U}) \cap H^m(I;\mathcal{H})$ and $Au \in H^m(I;(\mathcal{V}^0)')$. Let $U \in \mathcal{P}^b(I;\mathcal{U})$ be the solution of (3.6)-(3.7). For any $2 \leq \ell \leq 2q-1$, $1 \leq m \leq q$, we have*

$$\|u(T) - U(T)\|_{\mathcal{H}} \leq ck^{\frac{\ell-1}{2}}\left(\sum_{n=1}^N k_n^{2m}\int_{I_n}|\tilde{L}u^{(m)}|_{\ell-2}^2\right)^{\frac{1}{2}} \leq ck^{\frac{\ell-1}{2}+m}\left(\int_I|\tilde{L}u^{(m)}|_{\ell-2}^2\right)^{\frac{1}{2}} \tag{4.17}$$

for some $c > 0$ which depends only on q, γ and Γ.

Note that for the case of "full regularity", i.e, $\ell = 2q-1$, $m = q$ we obtain the optimal superconvergence bound of order k^{2q-1}. For $\mathcal{U} = \mathcal{V}$ this corresponds to the result in [32, Theorem 12.3]. Our estimate is more general because it states which convergence order one has, depending on the regularity of the solution, without assuming "full regularity" $\ell = 2q-1$. For the general case $\mathcal{V} \neq \mathcal{U}$, we have introduced the appropriate semi-norms $|\cdot|_\ell$ which express regularity.

5. Error analysis for time-discrete mixed formulation. In this section we derive discretization error bounds for the discrete mixed formulation (3.8)-(3.9). Let (u,p) be the solution of (2.14) and (U,P) the solution of (3.8)-(3.9), then u and U solve (2.6) and (3.6)-(3.7), respectively, cf. Lemma 3.3. Thus, *the optimal error bounds*

18

derived in Section 4 also hold for the solution U of the discrete mixed formulation. It remains to bound the error for the Lagrange multiplier. We present two results. The first result gives a sub-optimal bound of order $\mathcal{O}(k^{q-\frac{1}{2}})$, without making any further specific assumptions. In the second part we introduce a certain regularity assumption under which an optimal error bound is derived.

5.1. Sub-optimal result. In this subsection we derive a sub-optimal discretization error bound for the Lagrange multiplier. Below we use the notation introduced in Subsection 2.2, in particular the bilinear form $b_1(\cdot, \cdot)$ and the corresponding operator $B_1 : \mathcal{V}_2 \to Q_1'$ with $\ker(B_1) = \mathcal{V}_1$. To simplify the presentation we assume that we have a uniform step size $k_n = k$. We start with two lemmas that provide useful bounds for the error term $\mathcal{I}_q u - U$.

LEMMA 5.1. *Let $k_n = k$ for all $n = 1, \ldots, N$. Assume that the solution u of (2.6) has regularity $u \in H^1(I; \mathcal{U})$. Let $U \in \mathcal{P}^b(I; \mathcal{U})$ be the solution of (3.6)-(3.7). The following holds:*

$$\|(\mathcal{I}_q u - U)'\|_{L^2(I;\mathcal{H})} \leq c(k^{\frac{1}{2}} \|(u - \mathcal{I}_q u)'\|_{L^2(I;\mathcal{U})} + k^{-\frac{1}{2}} \|\mathcal{I}_q u - u\|_{L^2(I;\mathcal{U})})$$

for some $c > 0$ which depends only on q, γ and Γ.

Proof. Let $E = \mathcal{I}_q u - U \in \mathcal{P}^b(I; \mathcal{V})$. From the consistency property (3.12) and Lemma 3.2 we obtain the Galerkin relation

$$D_{\mathcal{H}}(E, X) + \int_I a(u - U, X) = 0 \quad \text{for all} \quad X \in \mathcal{P}^b(I; \mathcal{V}).$$

For $X := \sum_{n=1}^N (t - t_{n-1}) \chi_{I_n} E' \in \mathcal{P}^b(I; \mathcal{V})$ we get

$$\int_I (E', X)_{\mathcal{H}} + \int_I a(E, X) + \int_I a(u - \mathcal{I}_q u, X) = 0. \tag{5.1}$$

The L^1-norm on $P^{2q-4}([0, 1])$ is equivalent to

$$\phi \mapsto \int_0^1 t|\phi(t)| \, dt.$$

Thus, by scaling, we get the following inverse inequality

$$k \int_{I_n} \|E'\|_{\mathcal{H}}^2 \leq C \int_{I_n} (t - t_{n-1}) \|E'\|_{\mathcal{H}}^2$$

for each interval I_n, for some C which depends only on q. Using this and (5.1) we get

$$k\|E'\|_{L^2(I;\mathcal{H})}^2 \leq C \int_I (E', X)_{\mathcal{H}} = C \left(-\int_I a(E, X) - \int_I a(u - \mathcal{I}_q u, X) \right). \tag{5.2}$$

Using integration by parts on each interval I_n we obtain

$$
\begin{aligned}
-\int_I a(E, X) &= -\sum_{n=1}^N \int_{I_n} a(E, (t - t_{n-1}) E') \\
&= -\sum_{n=1}^N \frac{k}{2} a(E^n, E^n) + \frac{1}{2} \int_I a(E, E) \leq \frac{1}{2} \Gamma \|E\|_{L^2(I;\mathcal{U})}^2.
\end{aligned}
\tag{5.3}
$$

19

Integration by parts on each interval I_n yields

$$
\begin{aligned}
- \int_I a(u - \mathcal{I}_q u, X) &\leq \left| \sum_{n=1}^N \int_{I_n} a(u - \mathcal{I}_q u, (t - t_{n-1})E') \right| \\
&= \left| \sum_{n=1}^N \int_{I_n} a((u - \mathcal{I}_q u)', (t - t_{n-1})E) - \int_I a(u - \mathcal{I}_q u, E) \right| \\
&\leq \sum_{n=1}^N \int_{I_n} |t - t_{n-1}| |a((u - \mathcal{I}_q u)', E)| + \int_I |a(u - \mathcal{I}_q u, E)| \\
&\leq k\Gamma \|(u - \mathcal{I}_q u)'\|_{L^2(I;\mathcal{U})} \|E\|_{L^2(I;\mathcal{U})} + \Gamma \|u - \mathcal{I}_q u\|_{L^2(I;\mathcal{U})} \|E\|_{L^2(I;\mathcal{U})}.
\end{aligned} \tag{5.4}
$$

Using (5.3) and (5.4) in (5.2) we obtain

$$
k\|E'\|_{L^2(I;\mathcal{H})}^2 \leq \tilde{C}\left(\|E\|_{L^2(I;\mathcal{U})}^2 + \|u - \mathcal{I}_q u\|_{L^2(I;\mathcal{U})}^2 + k^2 \|(u - \mathcal{I}_q u)'\|_{L^2(I;\mathcal{U})}^2 \right),
$$

with a constant $\tilde{C}$ which depends only on q and Γ. Using the triangle inequality and Theorem 4.1 completes the proof. $\square$

LEMMA 5.2. *Let $k_n = k$ for all $n = 1, \ldots, N$. Assume that the solution u of (2.6) has regularity $u \in H^1(I;\mathcal{U})$. Let $U \in \mathcal{P}^b(I;\mathcal{U})$ be the solution of (3.6)-(3.7). The following holds:*

$$
k^{-\frac{1}{2}} \left(\|(\mathcal{I}_q u - U)_+^0\|_{\mathcal{H}}^2 + \sum_{n=1}^{N-1} \|[\mathcal{I}_q u - U]^n\|_{\mathcal{H}}^2 \right)^{\frac{1}{2}}
$$

$$
\leq c(\|(u - \mathcal{I}_q u)'\|_{L^2(I;\mathcal{U})} + k^{-\frac{1}{2}} \|\mathcal{I}_q u - u\|_{L^2(I;\mathcal{U})})
$$

for some $c > 0$ which depends only on q, γ and Γ.

Proof. Let $E := \mathcal{I}_q u - U \in \mathcal{P}^b(I;\mathcal{V})$. Let $E_n = \chi_{I_n} E$ for all $n = 1, \ldots, N$. Take for $2 \leq n \leq N$, $\tilde{E}_n \in \mathcal{P}_{q-1}(I_n;\mathcal{U})$ such that

$$
\tilde{E}_n(t_{n-1} + t) = E_{n-1}(t_{n-1} - t) \quad \text{for } t \in (0, k).
$$

Take $\tilde{E}_1 = 0$ and define $V_n := E_n - \tilde{E}_n$ for all n. Note that $(V_1)_+^0 = E_+^0$ and $(V_n)_+^{n-1} = [E]^{n-1}$ for $n = 2, \ldots, N$. For all $n = 1, \ldots, N$ and for $t \in I_n$ we have

$$
\begin{aligned}
\|V_n(t)\|_{\mathcal{H}} &\leq \|(V_n)_+^{n-1}\|_{\mathcal{H}} + \int_{t_{n-1}}^t \|V_n'\|_{\mathcal{H}} \leq \|(V_n)_+^{n-1}\|_{\mathcal{H}} + \int_{I_n} \|V_n'\|_{\mathcal{H}} \\
&\leq \|(V_n)_+^{n-1}\|_{\mathcal{H}} + \int_{I_n} \|E_n'\|_{\mathcal{H}} + \int_{I_n} \|\tilde{E}_n'\|_{\mathcal{H}} \\
&\leq \|(V_n)_+^{n-1}\|_{\mathcal{H}} + k^{\frac{1}{2}} \left(\|E'\|_{L^2(I_n;\mathcal{H})} + \|E'\|_{L^2(I_{n-1};\mathcal{H})} \right),
\end{aligned} \tag{5.5}
$$

with $I_0 := \emptyset$. From the consistency property (3.12) and Lemma 3.2 we obtain the Galerkin relation

$$
D_{\mathcal{H}}(E, X) + \int_I a(u - U, X) = 0 \quad \text{for all } X \in \mathcal{P}^b(I;\mathcal{V}).
$$

If we take $X = \sum_{n=1}^{N} V_n$, we obtain

$$\|E_+^0\|_{\mathcal{H}}^2 + \sum_{n=1}^{N-1} \|[E]^n\|_{\mathcal{H}}^2 = -\int_I (E', X)_{\mathcal{H}} - \int_I a(u - U, X). \qquad (5.6)$$

Using (5.5) we obtain for the first term on the right-hand side:

$$\left| \int_I (E', X)_{\mathcal{H}} \right| \leq \sum_{n=1}^{N} \int_{I_n} |(E', V_n)_{\mathcal{H}}| \leq \sum_{n=1}^{N} \|E'\|_{L^1(I_n; \mathcal{H})} \|V_n\|_{L^\infty(I_n; \mathcal{H})}$$

$$\leq k^{\frac{1}{2}} \sum_{n=1}^{N} \|E'\|_{L^2(I_n; \mathcal{H})} \left(\|(V_n)_+^{n-1}\|_{\mathcal{H}} + k^{\frac{1}{2}} (\|E'\|_{L^2(I_n; \mathcal{H})} + \|E'\|_{L^2(I_{n-1}; \mathcal{H})}) \right)$$

$$\leq \frac{k}{2} \|E'\|_{L^2(I; \mathcal{H})}^2 + \frac{1}{2} \|E_+^0\|_{\mathcal{H}}^2 + \frac{1}{2} \sum_{n=1}^{N-1} \|[E]^n\|_{\mathcal{H}}^2 + k \|E'\|_{L^2(I; \mathcal{H})}^2.$$

The second term can be bounded as follows:

$$\left| \int_I a(u - U, X) \right| \leq \sum_{n=1}^{N} \Gamma \|u - U\|_{L^2(I_n; \mathcal{U})} \|V_n\|_{L^2(I_n; \mathcal{U})}$$

$$\leq \sum_{n=1}^{N} \Gamma \|u - U\|_{L^2(I_n; \mathcal{U})} (\|E_n\|_{L^2(I_n; \mathcal{U})} + \|\tilde{E}_n\|_{L^2(I_n; \mathcal{U})})$$

$$\leq \frac{\Gamma}{2} \|u - U\|_{L^2(I; \mathcal{U})}^2 + \Gamma \|E\|_{L^2(I; \mathcal{U})}^2.$$

Using these bounds in (5.6) we get

$$k^{-1} \left(\|(\mathcal{I}_q u)_+^0 - U_+^0\|_{\mathcal{H}}^2 + \sum_{n=1}^{N-1} \|[E]^n\|_{\mathcal{H}}^2 \right)$$

$$\leq C \left(\|E'\|_{L^2(I; \mathcal{H})}^2 + k^{-1} (\|E\|_{L^2(I; \mathcal{U})}^2 + \|u - U\|_{L^2(I; \mathcal{U})}^2) \right)$$

with a constant C which depends only on Γ. Using a triangle inequality and the results in Lemma 5.1 and Theorem 4.1, we obtain the desired result. $\square$

THEOREM 5.3. *Let $k_n = k$ for all $n = 1, \ldots, N$. Assume that the solution (u, p) of (2.14) has regularity $u \in H^1(I; \mathcal{U})$. Let $(U, P) \in \mathcal{P}^b(I; \mathcal{U})$ be the solution of (3.8)-(3.9). Let $\mathcal{I}_q^{L^2}$ be the L^2-orthogonal projection from $L^2(I)$ to $\mathcal{P}^b(I)$. The following holds:*

$$\|p - P\|_{L^2(I; \mathcal{Q}_1)} \leq c \left(k^{\frac{1}{2}} \|(\mathcal{I}_q u - u)'\|_{L^2(I; \mathcal{U})} + k^{-\frac{1}{2}} \|\mathcal{I}_q u - u\|_{L^2(I; \mathcal{U})} \right)$$
$$+ \|p - \mathcal{I}_q^{L^2} p\|_{L^2(I; \mathcal{Q}_1)}$$

with a constant c which depends only on q, γ, Γ, β from (2.11) and $c_{\mathcal{H}} := \sup_{x \in \mathcal{U}} \frac{\|x\|_{\mathcal{H}}}{\|x\|_{\mathcal{U}}}$.

Proof. From the consistency property (3.14) we obtain the Galerkin relation

$$D_{\mathcal{H}}(u - U, X) + \int_I a(u - U, X) + \int_I b_1(X, p - P) = 0 \quad \text{for all } X \in \mathcal{P}^b(I; \mathcal{V}_2).$$

Let $\tilde{P} = \mathcal{I}_q^{L^2} p$ be the orthogonal projection of $p \in L^2(I; \mathcal{Q}_1)$ to $\mathcal{P}^b(I, \mathcal{Q}_1)$. From the inf-sup property (2.11) it follows that for all $t \in I$ there exists a unique $\hat{V}(t) \in \mathcal{V}^{\perp \mathcal{V}_2} \subset$

$\mathcal{V}_2$ such that:

$$b_1(\hat{V}(t), q) = (\tilde{P}(t) - P(t), q)_{\mathcal{Q}_1} \quad \text{for all} \ \ q \in \mathcal{Q}_1$$

$$\text{and} \ \ \|\hat{V}(t)\|_{\mathcal{U}} \le \frac{1}{\beta}\|\tilde{P}(t) - P(t)\|_{\mathcal{Q}_1}.$$

This, $\tilde{P} - P \in \mathcal{P}^b(I; \mathcal{Q}_1)$ and a tensor-product argument imply $\hat{V} \in \mathcal{P}^b(I; \mathcal{V}_2)$. Let $E = \mathcal{I}_q u - U$. By Lemma 3.2 we have

$$D_{\mathcal{H}}(E, \hat{V}) + \int_I a(u - U, \hat{V}) + \int_I b_1(\hat{V}, p - P) = 0. \tag{5.7}$$

We first consider the first term of (5.7). Using the continuous embedding $\mathcal{U} \hookrightarrow \mathcal{H}$ we get

$$\left| \int_I (E', \hat{V})_{\mathcal{H}} \right| \le \|E'\|_{L^2(I;\mathcal{H})}\|\hat{V}\|_{L^2(I;\mathcal{H})} \le c_{\mathcal{H}}\|E'\|_{L^2(I;\mathcal{H})}\|\hat{V}\|_{L^2(I;\mathcal{U})}.$$

and using the inverse estimate $\|\hat{V}_+^n\|_{\mathcal{U}} \le Ck^{-\frac{1}{2}}\|\hat{V}\|_{L^2(I_{n+1};\mathcal{U})}$, with a constant C which depends only on q, we get

$$|(E_+^0, V_+^0)_{\mathcal{H}}| + \sum_{n=1}^{N-1} |([E]^n, \hat{V}_+^n)_{\mathcal{H}}| \le c_{\mathcal{H}}\|E_+^0\|_{\mathcal{H}}\|\hat{V}_+^0\|_{\mathcal{U}} + c_{\mathcal{H}} \sum_{n=1}^{N-1} \|[E]^n\|_{\mathcal{H}}\|\hat{V}_+^n\|_{\mathcal{U}}$$

$$\le Cc_{\mathcal{H}}k^{-\frac{1}{2}}\|E_+^0\|_{\mathcal{H}}\|\hat{V}\|_{L^2(I_1;\mathcal{U})} + Cc_{\mathcal{H}}k^{-\frac{1}{2}} \sum_{n=1}^{N-1} \|[E]^n\|_{\mathcal{H}}\|\hat{V}\|_{L^2(I_{n+1};\mathcal{U})}$$

$$\le Cc_{\mathcal{H}}k^{-\frac{1}{2}}\Big(\|E_+^0\|_{\mathcal{H}}^2 + \sum_{n=1}^{N-1} \|[E]^n\|_{\mathcal{H}}^2\Big)^{\frac{1}{2}}\|\hat{V}\|_{L^2(I;\mathcal{U})}.$$

For the second term, we have

$$\left| \int_I a(u - U, \hat{V}) \right| \le \Gamma\|u - U\|_{L^2(I;\mathcal{U})}\|\hat{V}\|_{L^2(I;\mathcal{U})}.$$

For the third term we have

$$\|\tilde{P} - P\|_{L^2(I;\mathcal{Q}_1)}^2 = \int_I (\tilde{P} - P, \tilde{P} - P)_{\mathcal{Q}_1} = \int_I (\tilde{P} - P, p - P)_{\mathcal{Q}_1} = \int_I b_1(\hat{V}, p - P).$$

Combing these results with the definition of $\hat{V}$, we get

$$\|\tilde{P} - P\|_{L^2(I;\mathcal{Q}_1)}^2 = \int_I b_1(\hat{V}, p - P) = -D_{\mathcal{H}}(E, \hat{V}) - \int_I a(u - U, \hat{V})$$

$$\le c\Big(\|E'\|_{L^2(I;\mathcal{H})} + k^{-\frac{1}{2}}\big(\|E_+^0\|_{\mathcal{H}}^2 + \sum_{n=1}^{N-1} \|[E]^n\|_{\mathcal{H}}^2\big)^{\frac{1}{2}} + \|u - U\|_{L^2(I;\mathcal{U})}\Big)\|\hat{V}\|_{L^2(I;\mathcal{U})}$$

$$\le \frac{c}{\beta}\Big(\|E'\|_{L^2(I;\mathcal{H})} + k^{-\frac{1}{2}}\big(\|E_+^0\|_{\mathcal{H}}^2 + \sum_{n=1}^{N-1} \|[E]^n\|_{\mathcal{H}}^2\big)^{\frac{1}{2}} + \|u - U\|_{L^2(I;\mathcal{U})}\Big)\|\tilde{P} - P\|_{L^2(I;\mathcal{Q}_1)}.$$

Combining this with the previous two lemmas, Theorem 4.1 and the triangle inequality completes the proof. □

Results for the projection error $\|p - \mathcal{I}_q^{L^2} p\|_{L^2(I;\mathcal{Q}_1)}$ with respect to the orthogonal projection are standard. Results for the projection errors $\|(\mathcal{I}_q u - u)'\|_{L^2(I;\mathcal{U})}$ and $\|\mathcal{I}_q u - u\|_{L^2(I;\mathcal{U})}$ are also standard, cf. Subsection 4.1. Using these we obtain the following sub-optimal discretization error bound.

THEOREM 5.4. *Let* (u, p) *be the solution of* (2.14) *and* (U, P) *the solution of* (3.8)-(3.9). *Assume that the solution* (u, p) *has smoothness properties* $u \in H^m(I;\mathcal{U})$, $p \in H^m(I;\mathcal{Q}_1)$, *for an* m *with* $1 \leq m \leq q$. *The following holds:*

$$\|p - P\|_{L^2(I;\mathcal{Q}_1)} \leq ck^{m-\frac{1}{2}} \big(\|u^{(m)}\|_{L^2(I;\mathcal{U})} + k^{\frac{1}{2}} \|p^{(m)}\|_{L^2(I;\mathcal{Q}_1)} \big) \tag{5.8}$$

with a constant c *which depends only on* q, γ, Γ, β.

5.2. Optimal result. Below we again use the notation introduced in Subsection 2.2, in particular the bilinear form $b_1(\cdot, \cdot)$ and the corresponding operator $B_1 : \mathcal{U} \to \mathcal{Q}'_1$, with $\ker(B_1) = \mathcal{V}_1$. We introduce the $\mathcal{H}$-orthogonal projection $P_{\mathcal{H}} : \mathcal{H} \to \overline{\mathcal{V}}^{\mathcal{H}} =: \mathcal{V}_{\mathcal{H}}$ onto the closed subspace $\mathcal{V}_{\mathcal{H}}$ of $\mathcal{H}$, and define $P_{\mathcal{H}}^{\perp} := \mathrm{id}_{\mathcal{H}} - P_{\mathcal{H}} : \mathcal{H} \to \mathcal{H}$. The following (regularity type) assumption on $P_{\mathcal{H}}$ is crucial for our analysis.

ASSUMPTION 5.1. $P_{\mathcal{H}} \mathcal{V}_2 \subset \mathcal{V}_1$ *and there exists a constant* c_P *such that*

$$\|P_{\mathcal{H}} v\|_{\mathcal{U}} \leq c_P \|v\|_{\mathcal{U}} \quad \text{for all} \ \ v \in \mathcal{V}_2.$$

In the remainder of this subsection we assume that Assumption 5.1 is satisfied. Using this assumption we derive an optimal discretization error bound for the Lagrange multiplier.

EXAMPLE 5.1. We consider this assumption for Example 2.3. In this case we have $\mathcal{U} = H^1(\Omega)^d$, $\mathcal{H} = L^2(\Omega)^d$, $\mathcal{V}_1 = \{\, u \in H^1(\Omega)^d \mid \mathrm{div}\, u = 0 \,\}$, $\mathcal{V}_2 = H_0^1(\Omega)^d$ and $\mathcal{V} = \mathcal{V}_1 \cap \mathcal{V}_2 = \{\, u \in H_0^1(\Omega)^d \mid \mathrm{div}\, u = 0 \,\}$. The space $\mathcal{V}_{\mathcal{H}}$ is given by

$$\mathcal{V}_{\mathcal{H}} = \overline{\mathcal{V}}^{\mathcal{H}} = \{\, u \in L^2(\Omega)^d \mid \mathrm{div}\, u = 0, u \cdot n = 0 \text{ on } \partial\Omega \,\},$$

where n denotes the normal of $\partial\Omega$, cf. [15, p. 29]. The projection $P_{\mathcal{H}}$ (also known as the Leray projector) can be characterized as follows. For given $u \in L^2(\Omega)^d$, $P_{\mathcal{H}} u = u + \nabla q$, with q such that $(u + \nabla q, \nabla w)_{L^2(\Omega)^d} = 0$ for all $w \in H^1(\Omega)/\mathbb{R}$. When $u \in H_0^1(\Omega)^d$, this q solves the Poisson problem with homogeneous Neumann boundary conditions:

$$-\Delta q = \mathrm{div}\, u \quad \text{on} \ \ \Omega$$
$$\frac{\partial q}{\partial n} = 0 \quad \text{on} \ \ \partial\Omega$$
$$\int_{\Omega} q = 0.$$

We *assume* H^2-*regularity of this problem*, hence,

$$\|\nabla q\|_{H^1(\Omega)^d} \leq \|q\|_{H^2(\Omega)} \leq c\|\mathrm{div}\, u\|_{L^2(\Omega)} \leq c\|u\|_{H^1(\Omega)^d},$$

which implies that $P_{\mathcal{H}} : H_0^1(\Omega)^d \to H^1(\Omega)^d$ is continuous. This and $\mathrm{div}\, P_{\mathcal{H}} u = \mathrm{div}\, u + \Delta q = 0$ imply that Assumption 5.1 holds. This H^2-regularity assumption, which holds when Ω is convex or has a C^2-boundary, is exactly the same as in [18, Assumption 3.1]. There it is used in the well-posedness analysis of time-dependent

Stokes equations, cf. [18, Theorem 3.5].

THEOREM 5.5. *Assume that the solution (u, p) of (2.14) has regularity $u \in H^1(I; \mathcal{H})$ and $Au \in L^2(I; \mathcal{V}_{\mathcal{H}})'$. Let (U, P) be the solution of (3.8)-(3.9). The following holds:*

$$\|p - P\|_{L^2(I; \mathcal{Q}_1)} \leq \frac{\Gamma(1 + c_P)}{\beta} \|u - U\|_{L^2(I; \mathcal{U})} + \|p - \mathcal{I}_q^{L^2} p\|_{L^2(I; \mathcal{Q}_1)},$$

where $\mathcal{I}_q^{L^2}$ denotes the L^2-orthogonal projection from $L^2(I)$ to $\mathcal{P}^b(I)$.

Proof. From the consistency property (3.14) we obtain the Galerkin relation

$$D_{\mathcal{H}}(u - U, X) + \int_I a(u - U, X) = 0 \quad \text{for all} \ \ X \in \mathcal{P}^b(I; \mathcal{V}).$$

The regularity assumptions $u', Au \in L^2(I; \mathcal{V}_{\mathcal{H}})'$ and the density of $\mathcal{V}$ in $\mathcal{V}_{\mathcal{H}}$ imply $K(U, \cdot) \in \mathcal{P}^b(I; \mathcal{V}_{\mathcal{H}})'$ and

$$D_{\mathcal{H}}(u - U, X) + \int_I a(u - U, X) = 0 \quad \text{for all} \ \ X \in \mathcal{P}^b(I; \mathcal{V}_{\mathcal{H}}).$$

From $P_{\mathcal{H}} \mathcal{V}_2 \subset \mathcal{V}_{\mathcal{H}}$, $P_{\mathcal{H}} \mathcal{V}_2 \subset \ker(B_1)$ and the consistency property (3.14) we obtain the Galerkin relation

$$D_{\mathcal{H}}(u - U, X) + \int_I a(u - U, X) + \int_I b_1(X, p - P) = 0 \quad \forall \, X \in \mathcal{P}^b(I; \mathcal{V}_2 + P_{\mathcal{H}} \mathcal{V}_2). \quad (5.9)$$

Let $\tilde{P} = \mathcal{I}_q^{L^2} p$ be the orthogonal projection of $p \in L^2(I; \mathcal{Q}_1)$ to $\mathcal{P}^b(I, \mathcal{Q}_1)$. From the inf-sup property (2.11) it follows that for all $t \in I$ there exists a unique $\hat{V}(t) \in \mathcal{V}^{\perp_{\mathcal{V}_2}} \subset \mathcal{V}_2$ such that:

$$b_1(\hat{V}(t), q) = (\tilde{P}(t) - P(t), q)_{\mathcal{Q}_1} \quad \text{for all} \ \ q \in \mathcal{Q}_1$$

$$\text{and} \ \ \|\hat{V}(t)\|_{\mathcal{U}} \leq \frac{1}{\beta} \|\tilde{P}(t) - P(t)\|_{\mathcal{Q}_1}. \quad (5.10)$$

This, $\tilde{P} - P \in \mathcal{P}^b(I; \mathcal{Q}_1)$ and a tensor-product argument imply $\hat{V} \in \mathcal{P}^b(I; \mathcal{V}_2)$. From $P_{\mathcal{H}} \mathcal{V}_2 \subset \ker(B_1)$ it follows that for $v \in \mathcal{V}_2$ we have

$$b_1(v, p) = b_1(P_{\mathcal{H}}^{\perp} v, p) \quad \text{for all} \ \ p \in \mathcal{Q}_1.$$

Using this and (5.9), (5.10) we get

$$\|\tilde{P} - P\|^2_{L^2(I; \mathcal{Q}_1)} = \int_I (\tilde{P} - P, \tilde{P} - P)_{\mathcal{Q}_1} = \int_I (\tilde{P}(t) - P(t), p(t) - P(t))_{\mathcal{Q}_1} \, dt$$

$$= \int_I b_1(\hat{V}(t), p(t) - P(t)) \, dt = \int_I b_1(P_{\mathcal{H}}^{\perp} \hat{V}(t), p(t) - P(t)) \, dt$$

$$= \int_I a(U - u, P_{\mathcal{H}}^{\perp} \hat{V}) + D_{\mathcal{H}}(U - u, P_{\mathcal{H}}^{\perp} \hat{V}). \quad (5.11)$$

For the first term we get

$$\int_I a(U - u, P_{\mathcal{H}}^{\perp} \hat{V}) \leq \Gamma \|U - u\|_{L^2(I; \mathcal{U})} (1 + c_P) \|\hat{V}\|_{L^2(I; \mathcal{U})}$$

$$\leq \frac{\Gamma(1 + c_P)}{\beta} \|U - u\|_{L^2(I; \mathcal{U})} \|\tilde{P} - P\|_{L^2(I; \mathcal{Q}_1)}.$$

For the second term we use Lemma 3.2 and the fact that $D_\mathcal{H}(\cdot, \cdot)$ consists of sums and integrals of $\mathcal{H}$ scalar products, hence

$$D_\mathcal{H}(U - u, P_\mathcal{H}^\perp \hat{V}) = D_\mathcal{H}(U - \mathcal{I}_q u, P_\mathcal{H}^\perp \hat{V}) = D_\mathcal{H}(P_\mathcal{H}^\perp(U - \mathcal{I}_q u), \hat{V}) = 0, \qquad (5.12)$$

where the last equality follows from $U(t) - \mathcal{I}_q u(t) \in \ker(B) = \mathcal{V}$ and $P_\mathcal{H}^\perp(U'(t) - u'(t)) = \left(P_\mathcal{H}^\perp(U - \mathcal{I}_q u)\right)'(t)$. Thus we obtain

$$\|\tilde{P} - P\|_{L^2(I;\mathcal{Q}_1)} \leq \frac{\Gamma(1 + c_P)}{\beta} \|U - u\|_{L^2(I;\mathcal{U})}.$$

This, in combination with the triangle inequality concludes the proof. □

Results for the projection error $\|p - \mathcal{I}_q^{L^2} p\|_{L^2(I;\mathcal{Q}_1)}$ with respect to the orthogonal projection are standard. Using these and (4.4) we obtain the following optimal discretization error bound.

THEOREM 5.6. *Let (u, p) be the solution of (2.14) and (U, P) the solution of (3.8)-(3.9). Assume that the solution (u, p) has smoothness properties $Au \in L^2(I;\mathcal{V}_\mathcal{H})'$, $u \in H^m(I;\mathcal{U})$, $p \in H^m(I;\mathcal{Q}_1)$, for an m with $1 \leq m \leq q$. The following holds:*

$$\|p - P\|_{L^2(I;\mathcal{Q}_1)} \leq c \left(\sum_{n=1}^{N} k_n^{2m} \left(\left(\frac{1 + c_P}{\beta}\right)^2 \|u^{(m)}\|_{L^2(I_n;\mathcal{U})}^2 + \|p^{(m)}\|_{L^2(I_n;\mathcal{Q}_1)}^2 \right) \right)^{\frac{1}{2}}$$

$$\leq c k^m \left(\frac{1 + c_P}{\beta} \|u^{(m)}\|_{L^2(I;\mathcal{U})} + \|p^{(m)}\|_{L^2(I;\mathcal{Q}_1)} \right) \qquad (5.13)$$

for some $c > 0$, which depends only on q, γ and Γ.

REMARK 5.1. For the derivation of an optimal error bound for the Lagrange multiplier, as in Theorem 5.6, for the general abstract (saddle point) problem (2.14) Assumption 5.1 is sufficient. For the specific case of the Stokes equations (Example 2.3) such optimal error bounds can be derived under weaker assumptions than the H^2-regularity discussed in Example 5.1. The analysis, which is a topic of current research, uses additional specific properties of the Stokes problem.

6. Fully discrete problem. In applications of the abstract setting one typically has infinite dimensional spaces $\mathcal{U}, \mathcal{V}, \mathcal{Q}, \mathcal{Q}_1, \mathcal{V}_2$ in the time-discrete problems (3.6)-(3.7) (constrained formulation) and (3.8)-(3.9) (mixed formulation). These spaces are then replaced by finite dimensional (e.g. finite element) spaces to obtain a feasible method. Below we describe this fully discrete setting.

6.1. Fully discrete constrained formulation. In this subsection we introduce the spacial discretization for the constrained problem (3.6)-(3.7). We assume that a conforming finite element method is used, i.e. we have a finite dimensional subspace $\mathcal{U}^h \subset \mathcal{U}$. We define $\mathcal{V}^h = \mathcal{U}^h \cap \mathcal{V}$ and $\mathcal{Q}^h = B\mathcal{U}^h \subset \mathcal{Q}'$. We assume that we have *spatial* projection operators $\mathbf{I}_{\mathcal{Q}^h} : \mathcal{Q}' \to \mathcal{Q}^h$ and $\mathbf{I}_{\mathcal{U}^h} : \mathcal{U} \to \mathcal{U}^h$, which satisfy $\mathbf{I}_{\mathcal{Q}^h} B = B\mathbf{I}_{\mathcal{U}^h}$. If we assume some additional spatial regularity on g and u_0, then it is sufficient if $\mathbf{I}_{\mathcal{U}^h}$ and $\mathbf{I}_{\mathcal{Q}^h}$ are only defined on subsets of $\mathcal{U}$ and $\mathcal{Q}'$ and $u_0 \in \mathcal{D}(\mathbf{I}_{\mathcal{U}^h}) \subset \mathcal{U}$, $g(t) \in \mathcal{D}(\mathbf{I}_{\mathcal{Q}^h}) \subset \mathcal{Q}'$ for almost all $t \in I$. In applications, these spatial projections typically are nodal interpolations, cf. Example 6.1 below.

The fully discrete version of (2.6) reads: find $U_h \in \mathcal{P}^b(I;\mathcal{U}^h)$ such that

$$K(U_h, X_h) = \int_I f(X_h) + (\mathbf{I}_{\mathcal{U}^h} u_0, (X_h)_+^0)_{\mathcal{H}} \quad \text{for all} \quad X_h \in \mathcal{P}^b(I;\mathcal{V}^h) \qquad (6.1)$$

$$BU_h = \mathcal{I}_q \mathbf{I}_{\mathcal{Q}^h} g. \qquad (6.2)$$

The commutation property $\mathbf{I}_{\mathcal{Q}^h} B = B\mathbf{I}_{\mathcal{U}^h}$ implies the important consistency property $\mathcal{I}_q B(U - \mathbf{I}_{\mathcal{U}^h} u) = 0$ for the constraint discretization (in space and time). The fully discrete problem (6.1)-(6.2) gives rise to a system of linear equations. For practical purposes it is useful to have a (cheap) way to compute for $g_h \in \mathcal{Q}^h$ a $G_h \in \mathcal{U}^h$ with $BG_h = g_h$. We will denote this operator by $\mathbf{L}_{\mathcal{Q}^h}$. Provided one has such an operator, (6.1)-(6.2) is equivalent to finding $W_h \in \mathcal{P}^b(I;\mathcal{V}^h)$ such that

$$K(W_h, X_h) = \int_I f(X_h) + (\mathbf{I}_{\mathcal{U}^h} u_0, (X_h)_+^0)_{\mathcal{H}} - K(\mathcal{I}_q G_h, X_h) \quad \forall X_h \in \mathcal{P}^b(I;\mathcal{V}^h) \quad (6.3)$$

with $G_h = \mathbf{L}_{\mathcal{Q}^h} \mathbf{I}_{\mathcal{Q}^h} g$. We then have $U_h = W_h + \mathcal{I}_q G_h$. Note that in (6.3) both the trial and test space are $\mathcal{P}^b(I;\mathcal{V}^h)$. The global in time equation (6.3) does not need to be solved in this form. It has a lower block triangular structure, cf. definition of $K(\cdot, \cdot)$ and $D_{\mathcal{H}}(\cdot, \cdot)$. Thus (6.3) can be solved sequentially by solving for each $n = 1, \ldots, N$

$$\int_{I_n} (W'_{h,n}, X_h)_{\mathcal{H}} + ((W_{h,n})_+^{n-1}, (X_h)_+^{n-1})_{\mathcal{H}} + \int_{I_n} a(W_{h,n}, X_h)$$

$$= \int_{I_n} f(X_h) + (W_{h,n-1}^{n-1}, (X_h)_+^{n-1})_{\mathcal{H}} - \int_{I_n} (G'_h, X_h)_{\mathcal{H}} - \int_{I_n} a(\mathcal{I}_q G_h, X_h) \qquad (6.4)$$

for all $X_h \in \mathcal{P}_{q-1}(I_n;\mathcal{V}^h)$, where $W_{h,n} = W_h|_{I_n}$ and $W_{h,0}^0 = \mathbf{I}_{\mathcal{U}^h} u_0 - G_h(0)$. Typically g is given explicitly, but g' is not, thus G'_h is also not known explicitly. Hence, it is often convenient to rewrite the term $\int_{I_n} (G'_h, X_h)_{\mathcal{H}}$ using integration by parts. Note that solving (6.4) requires explicit computations in $\mathcal{V}^h$ which may or may not be easy in a practical setting.

EXAMPLE 6.1. We outline the fully discrete problem for the specific case given in Example 2.1. Let Ω be a polygonal domain and, for a given simplicial triangulation $\mathbb{T}_h$ and $r \in \mathbb{N}$,

$$\mathcal{U}^h = \{u \in H^1(\Omega) \mid u|_{T_s} \in \mathcal{P}_r(T_s) \text{ for all } T_s \in \mathbb{T}_h \}$$

a standard H^1-conforming finite element space. In order to express (6.4) as a linear system, we need an explicit basis of

$$\mathcal{V}^h = \mathcal{U}^h \cap H_0^1(\Omega = \{u \in H_0^1(\Omega) \mid u|_{T_s} \in \mathcal{P}_r(T_s) \text{ for all } T_s \in \mathbb{T}_h \}.$$

Taking a nodal basis for $\mathcal{U}^h$ and omitting the nodes on the boundary $\partial\Omega$ yields such a basis. The operator $\mathbf{L}_{\mathcal{Q}^h} \mathbf{I}_{\mathcal{Q}^h}|_{H^{\frac{1}{2}}(\partial\Omega)\cap C(\partial\Omega)}$ can be taken to be a Lagrange interpolation on operator $\partial\Omega$, cf. [12, Section 3.2.2]. The operator $\mathbf{I}_{\mathcal{U}^h}|_{H^1(\Omega)\cap C(\Omega)}$, which is used for the initial data, can be taken to be a nodal interpolation operator. For these choices the commutation property $\mathbf{I}_{\mathcal{Q}^h} B = B\mathbf{I}_{\mathcal{U}^h}$, with B the trace operator on $\partial\Omega$, holds.

For the Stokes equation (Examples 2.2, 2.3) one would need an explicit feasible basis for the space

$$\mathcal{V}^h = \{u \in H_0^1(\Omega)^d | \operatorname{div} u = 0, \quad u|_{T_s} \in \mathcal{P}_r(T_s)^d \text{ for all } T_s \in \mathbb{T}_h \},$$

i.e., of a space of divergence free finite elements. In many cases this is rather cumbersome. Furthermore, computing $\mathbf{L}_{\mathcal{Q}^h}\mathbf{I}_{\mathcal{Q}^h}$ would involve finding G_h such that $\operatorname{div} G_h = \mathbf{I}_{\mathcal{Q}^h}g$, which also has certain disadvantages. These difficulties can be avoided by using the mixed method which will be presented in the next subsection.

6.2. Fully discrete mixed formulation. In this subsection we introduce the spatial discretization for the mixed problem (3.8)-(3.9). As in the previous section, we assume that a conforming finite element method is used, i.e. we have finite dimensional subspaces $\mathcal{U}^h \subset \mathcal{U}$ and $\mathcal{Q}_1^h \subset \mathcal{Q}_1$. We define $\mathcal{Q}_2^h = B_2\mathcal{U}^h \subset \mathcal{Q}_2'$ and $\mathcal{V}_2^h = \mathcal{U}^h \cap \mathcal{V}_2$. For existence of a unique solution we assume (h-dependent) inf-sup stability of the pair $(\mathcal{V}_2^h, \mathcal{Q}_1^h)$ with respect to the constraint operator that corresponds to the implicitly treated constraints, i.e, we assume that there is a $\tilde{\beta}_h > 0$ such that

$$\inf_{q_1 \in \mathcal{Q}_1^h} \sup_{u_h \in \mathcal{V}_2^h} \frac{b_1(u_h, q_1)}{\|u_h\|_{\mathcal{U}}\|q_1\|_{\mathcal{Q}_1}} \geq \tilde{\beta}_h. \tag{6.5}$$

Due to the different ways of treating b_1 and b_2 we will use (3.10)-(3.11) instead of (3.9). Note that $\mathcal{Q}_1^h$ is taken to be a subspace of $\mathcal{Q}_1$, which is the solution space for the Lagrange multiplier, whereas $\mathcal{Q}_2^h$ is taken to be a subspace of $\mathcal{Q}_2'$ (as in the previous subsection). We assume that we have a projection $\mathbf{I}_{\mathcal{Q}_2^h} : \mathcal{Q}_2' \to \mathcal{Q}_2^h$ and a projection $\mathbf{I}_{\mathcal{U}^h} : \mathcal{U} \to \mathcal{U}^h$, which satisfy $\mathbf{I}_{\mathcal{Q}_2^h}B = B\mathbf{I}_{\mathcal{U}^h}$.

The fully discrete version of (2.6) reads: find $U_h \in \mathcal{P}^b(I; \mathcal{U}^h)$, $P_h \in \mathcal{P}^b(I; \mathcal{Q}_1^h)$ such that

$$K(U_h, X_h) + \int_I b_1(X_h, P_h) = \int_I f(X_h) + (\mathbf{I}_{\mathcal{U}^h}u_0, (X_h)_+^0)_{\mathcal{H}} \quad \text{for all } X_h \in \mathcal{P}^b(I; \mathcal{V}_2^h)$$

$$\int_I b_1(U_h, R_h) = \int_I (\mathcal{I}_q g_1)(R_h) \quad \text{for all } R_h \in \mathcal{P}^b(I; \mathcal{Q}_1^h) \tag{6.6}$$

$$B_2 U_h = \mathcal{I}_q \mathbf{I}_{\mathcal{Q}_2^h} g_2$$

The fully discrete problem (6.6) gives rise to a system of linear equations. The explicit constraints are treated in the same way as in Subsection 6.1, i.e., we assume a (cheap) way to compute for $g_h \in \mathcal{Q}_2^h$ a $G_h \in \mathcal{U}^h$ with $B_2 G_h = g_h$. We denote this operator by $\mathbf{L}_{\mathcal{Q}_2^h}$. Provided one has such an operator, (6.6) is equivalent to finding $W_h \in \mathcal{P}^b(I; \mathcal{V}_2^h)$, $P_h \in \mathcal{P}^b(I; \mathcal{Q}_1^h)$ such that for all $X_h \in \mathcal{P}^b(I; \mathcal{V}_2^h)$:

$$K(W_h, X_h) + \int_I b_1(X_h, P_h) = \int_I f(X_h) + (\mathbf{I}_{\mathcal{U}^h}u_0, (X_h)_+^0)_{\mathcal{H}} - K(\mathcal{I}_q G_h, X_h) \tag{6.7}$$

$$\int_I b_1(W_h, R_h) = \int_I (\mathcal{I}_q g_1)(R_h) - \int_I b_1(\mathcal{I}_q G_h, R_h) \quad \text{for all } R_h \in \mathcal{P}^b(I; \mathcal{Q}_1^h), \tag{6.8}$$

with $G_h = \mathbf{L}_{\mathcal{Q}_2^h}\mathbf{I}_{\mathcal{Q}_2^h}g_2$. Assumption (6.5) ensures that this system has a unique solution. We have $U_h = W_h + \mathcal{I}_q G_h$. The global in time equation (6.7)-(6.8) can be solved sequentially by solving for each $n = 1, \ldots, N$

$$\int_{I_n} (W_{h,n}', X_h)_{\mathcal{H}} + ((W_{h,n})_+^{n-1}, (X_h)_+^{n-1})_{\mathcal{H}} + \int_{I_n} a(W_{h,n}, X_h) + \int_{I_n} b_1(X_h, P_{h,n})$$

$$= \int_{I_n} f(X_h) + (W_{h,n-1}^{n-1}, (X_h)_+^{n-1})_{\mathcal{H}} - \int_{I_n} (G_h', X_h)_{\mathcal{H}} - \int_{I_n} a(\mathcal{I}_q G_h, X_h) \tag{6.9}$$

$$\int_{I_n} b_1(W_{h,n}, R_h) = \int_{I_n} (\mathcal{I}_q g_1)(R_h) - \int_{I_n} b_1(\mathcal{I}_q G_h, R_h) \tag{6.10}$$

27

for all $X_h \in \mathcal{P}_{q-1}(I_n; \mathcal{V}_2^h)$, $R_h \in \mathcal{P}_{q-1}(I_n; \mathcal{Q}_1^h)$, where $W_{h,n} = W_h|_{I_n}$, $P_{h,n} = P_h|_{I_n}$ and $W_{h,0}^0 = \mathbf{I}_{\mathcal{U}^h} u_0 - G_h(0)$. Solving the saddle point problem (6.9)-(6.10) requires explicit computations in $\mathcal{V}_2^h$ and $\mathcal{Q}_1^h$ but not in $\mathcal{V}^h$. For the term involving G_h' integration by parts is convenient, cf. the discussion in the previous subsection.

In the specific case of the Stokes problem (on a polygonal domain Ω) as discussed in Example 2.3, an obvious choice is the Hood-Taylor pair:

$$\mathcal{Q}_1^h = \{ p \in H^1(\Omega) \cap L_0^2(\Omega) \mid p|_{T_s} \in \mathcal{P}_{r-1}(T_s) \text{ for all } T_s \in \mathbb{T}_h \},$$
$$\mathcal{U}^h = \{ u \in H^1(\Omega)^d \mid u|_{T_s} \in \mathcal{P}_r(T_s)^d \text{ for all } T_s \in \mathbb{T}_h \},$$

where $2 \leq r \in \mathbb{N}$. In order to express (6.9)-(6.10) as a linear system, we need an explicit basis of $\mathcal{Q}_1^h$ and of

$$\mathcal{V}_2^h = \mathcal{U}^h \cap \mathcal{V}_2 = \{ u \in H_0^1(\Omega)^d \mid u|_{T_s} \in \mathcal{P}_r(T_s)^d \text{ for all } T_s \in \mathbb{T}_h \}.$$

In the same way as in the previous subsection, a basis for $\mathcal{V}_2^h$ can be obtained by taking a nodal basis for $\mathcal{U}^h$ and omitting the nodes on the boundary $\partial\Omega$. The operators $\mathbf{L}_{\mathcal{Q}_2^h} \mathbf{I}_{\mathcal{Q}_2^h}$ and $\mathbf{I}_{\mathcal{U}^h}$ can also be taken as in the previous subsection. For these choices of finite element spaces the stability Assumption (6.5) is satisfied, even with $\tilde{\beta}_h$ independent of h.

7. Numerical experiments. We consider two Stokes problems with a known analytic solution to validate the results from the error analysis numerically and to illustrate certain phenomena. The first example is a Stokes problem with a non-zero divergence condition and a non-zero Dirichlet boundary condition. Despite the fact that the results of the error analysis are derived (only) in a semi-discrete setting, we see the predicted convergence rates (w.r.t. time discretization) also in the fully discrete setting. We use this Stokes example not only to validate our theoretical results but also to show that one does *not* obtain optimal results if the projection $\mathcal{I}_q$ in (6.6) is omitted, cf. Remark 4.1. The second example is a Stokes interface problem with a stationary interface. The density and diffusion coefficients are piecewise constant and discontinuous across the interface. We take a homogeneous divergence condition and a non-zero Dirichlet boundary condition (in/out-flow). This less smooth example is also used to illustrate a certain effect related to the semi-norms $|\cdot|_\ell$, which were introduced in Section 4.2.

The methods are implemented in the software package DROPS, cf. [16].

7.1. Stokes Problem. We consider a problem as in Example 2.3. We take $\Omega = (-1,1)^3$ and a time interval $(0,1)$. Different from Example 2.3, in the diffusion part we use the symmetrized gradient $Du = \nabla u + \nabla u^T$, i.e., $a(u,v) = \int_\Omega Du : Dv$. This difference is not essential. We use the symmetrized gradient because in view of the interface problem considered below it is more natural. The constraint operators are $B_1 = \operatorname{div}$ and $B_2 = \operatorname{tr}_{|\partial\Omega}$. We take

$$u = \begin{pmatrix} (x^2 + 1)(z + y)\sin(4t) \\ (y^2 + 1)(z + x)\sin(4t) \\ (z^2 + 1)(x + y)\sin(4t) \end{pmatrix}, \quad p = e^t(x^2 + y^2 + z^2)$$

and we discretize the problem

$$u' - \operatorname{div} Du + \nabla p = f \tag{7.1}$$
$$\operatorname{div} u = g_1 \tag{7.2}$$
$$u|_{\partial\Omega} = g_2 \tag{7.3}$$

for the appropriate right hand sides. Note that $g_1 \neq 0$, $g_2 \neq 0$ and both are time-dependent. We take a uniform step size $k = \frac{1}{N}$ in time and $q = 2$. For the discretization in space we take a triangulation of Ω. To obtain the triangulation $\mathbb{T}_h$ the domain Ω is divided into cubes with side length $h := \frac{1}{N_S}$ and each of the cubes is divided into six tetrahedra. We use the $\mathcal{P}_2$-$\mathcal{P}_1$ Hood-Taylor pair

$$
\begin{aligned}
\mathcal{Q}_1^h &= \{p \in H^1(\Omega) \mid p|_{T_s} \in \mathcal{P}_1(T_s) \text{ for all } T_s \in \mathbb{T}_h\}, \\
\mathcal{U}^h &= \{u \in H^1(\Omega)^3 \mid u|_{T_s} \in \mathcal{P}_2(T_s)^3 \text{ for all } T_s \in \mathbb{T}_h\}.
\end{aligned}
\tag{7.4}
$$

To assemble the linear system (6.9)-(6.10) we use the method described in the previous section. We denote the standard nodal basis of $\mathcal{U}^h$ by $\{\psi_1, \ldots \psi_N\} = \Psi$. From the we extract a nodal basis of $\mathcal{V}_2^h$: $\Psi_0 = \{\psi \mid \psi \in \Psi \cap H_0^1(\Omega)^3\}$ and we take $\Psi_{\partial\Omega} = \Psi \setminus \Psi_0$. We treat the explicit constraint $u|_{\partial\Omega} = g_2$ by taking $G_h = \sum_{\psi \in \Psi_{\partial\Omega}} c(\psi)\psi$ such that $G_h|_{\partial\Omega}(x) = g_2(x)$ for all nodal points $x \in \partial\Omega$. To compute the matrices and the right-hand sides in (6.7)-(6.8) we exploit the tensor-product structure of the mesh. A fifth order spatial quadrature rule is used and combined with a four point Gauss quadrature in time.

Table 7.1: Error in $L^2 \otimes H^1$-norm between u and the solution of (6.6). The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	0.47350	0.13324	0.06644	0.05980	0.05936	0.05934	
8	0.46973	0.11950	0.03076	0.01038	0.00744	0.00721	3.04024
16	0.46966	0.11929	0.02991	0.00753	0.00207	0.00100	2.84747
32	0.46966	0.11928	0.02990	0.00748	0.00187	0.00048	1.06226
EOC_T		1.97726	1.99607	1.99831	1.99965	1.96332	

Table 7.2: Error in $L^2 \otimes H^1$-norm between u and the solution of (6.6), if we omit the projection $\mathcal{I}_q$. The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	0.82210	0.21769	0.07950	0.06081	0.05943	0.05934	
8	1.22252	0.31238	0.07942	0.02133	0.00883	0.00732	3.01999
16	1.78108	0.45328	0.11427	0.02887	0.00737	0.00206	1.82496
32	2.55929	0.64970	0.16318	0.04099	0.01032	0.00261	-0.33854
EOC_T		1.97791	1.99326	1.99318	1.99008	1.98258	

The Stokes problem (7.1) - (7.3) is discretized by (6.7)-(6.8) as described in the previous section. The error $\|u - U_h\|_{L^2 \otimes H^1}$ is given in Table 7.1. In this table we see that the error is of order $\mathcal{O}(k^2 + h^2)$ (cf. diagonals in the table) which is optimal in time, as predicted by Theorem 4.2, and in space. If we omit the projection $\mathcal{I}_q$ in (6.6), then we obtain a sub-optimal results of order $\mathcal{O}(k^2 h^{-\frac{1}{2}} + h^2)$ in Table 7.2. Note that for a fixed temporal discretization refinement in space results in divergence of order $\mathcal{O}(h^{-1/2})$. This behavior can be observed rather clearly in the first two columns ($N = 4, 8$) of Table 7.2, where the time discretization error dominates. The $h^{-\frac{1}{2}}$ behavior can be explained using some further analysis, which we do not present here

(since it involves the analysis of the spatial discretization). We also see that along the diagonal we only have a space-time convergence order of 1.5.

We are also interested in superconvergence. Theorem 4.5 can be applied not only on $[0, T]$, but also on any subinterval $[0, t_n]$. We thus are interested in the convergence order for the maximal nodal error $\max_{n=1,...,N} \|u^n - U_h^n\|_{L^2}$. We expect a $\mathcal{O}(k^3 + h^3)$ convergence order: optimal in time, due to Theorem 4.5 and optimal in space. This is consistent with the numerical results in Table 7.3, however, the spatial error dominates. If we omit the projection operator $\mathcal{I}_q$, we again obtain results that are sub-optimal (Table 7.4). In particular we do not observe superconvergence. Related to this, the nodal (in time) discretization error can be several orders of magnitude smaller due to the use of the projection operator $\mathcal{I}_q$, e.g., the results for $N = 4, N_S = 32$ in the Tables 7.3, 7.4.

Table 7.3: Maximal nodal error $\|u^n - U_h^n\|_{L^2}$, where U_h is the solution of (6.6). The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	0.02137	0.02343	0.02343	0.02346	0.02349	0.02349	
8	0.00252	0.00276	0.00277	0.00277	0.00277	0.00277	3.0834
16	3.09e-04	3.39e-04	3.40e-04	3.40e-04	3.40e-04	3.40e-04	3.0265
32	3.94e-05	4.34e-05	4.38e-05	4.43e-05	4.23e-05	4.23e-05	3.0077
EOC_T		-0.1384	-0.0154	-0.0150	0.0666	-5.8e-05	

Table 7.4: Maximal nodal error $\|u^n - U_h^n\|_{L^2}$, where U_h is the solution of (6.6), if we omit the projection $\mathcal{I}_q$. The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	0.28772	0.06873	0.02488	0.02247	0.02317	0.02341	
8	0.37160	0.09146	0.02327	0.00611	0.00293	0.00274	3.0948
16	0.41012	0.10133	0.02600	0.00651	0.00164	5.10e-03	2.4255
32	0.42844	0.10590	0.02719	0.00683	0.00170	4.27e-04	0.2570
EOC_T		2.0164	1.9615	1.9932	2.0028	1.9974	

We are also interested in the $L^2 \otimes L^2$ error for the pressure p. Recall that we have two results. Theorem 5.8 provides a general, but sub-optimal result. Theorem 5.13 provides an optimal result, if Assumption 5.1 holds. In the setting of this experiment this assumption is satisfied, cf. Example 5.1. Due to the convexity of Ω we have H^2-regularity for the Poisson problem. We observe an optimal convergence order $\mathcal{O}(k^2 + h^2)$ in Table 7.5. Again, if we do not use the projection operator $\mathcal{I}_q$, we lose the optimal convergence order, see Table 7.6.

7.2. A Stokes interface problem. We consider the same setting as in the previous experiment. However, we now use discontinuous piecewise constant density and diffusion coefficients ρ and μ and a spatial domain with a re-entrant corner $\Omega = (-1, 1)^3 \setminus [-\frac{1}{2}, 1) \times [0, 1)^2$. In the previous experiment we have $\rho = \mu = 1$ in Ω. A stationary interface separates two subdomains (phases) $\Omega_+ = \{(x, y, z) \in \Omega \mid x \geq 1/7\}$ and $\Omega_- = \Omega \setminus \Omega_+$. The density and diffusion coefficients are $\rho = \mu = 5\chi_{\Omega_-} + \chi_{\Omega_+}$.

Table 7.5: Error in $L^2 \otimes L^2$-norm between p and the solution of (6.6). The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	0.42505	0.30505	0.29582	0.29525	0.29522	0.29522	
8	0.31218	0.09285	0.05403	0.05064	0.05040	0.05039	2.5507
16	0.30845	0.07883	0.02237	0.01204	0.01089	0.01083	2.2185
32	0.30828	0.07813	0.01984	0.00589	0.00286	0.00261	2.0540
EOC_T		1.9804	1.9771	1.7534	1.0400	0.1347	

Table 7.6: Error in $L^2 \otimes L^2$-norm between p and the solution of (6.6), if we omit the projection $\mathcal{I}_q$. The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	1.20438	0.69114	0.43399	0.33535	0.30565	0.29784	
8	1.32184	0.71175	0.36637	0.18878	0.10372	0.06762	2.1390
16	1.35591	0.73062	0.37476	0.18879	0.09479	0.04816	0.4896
32	1.36231	0.73455	0.37695	0.19011	0.09513	0.04755	0.0184
EOC_T		0.8911	0.9625	0.9875	0.9989	1.0004	

The corresponding equations read

$$\rho u' - \operatorname{div}(\mu Du) + \nabla p = f \quad \text{in} \quad L^2(I; H^{-1}(\Omega)) \tag{7.5}$$

$$\operatorname{div} u = 0 \tag{7.6}$$

$$u|_{\partial\Omega} = g. \tag{7.7}$$

In this problem we have $a(u,v) = \int_\Omega \mu Du : Dv$ and the constraint operators B_1, B_2 are defined as in the previous experiment. In order to apply the theory from the previous sections, we replace the standard L^2 scalar product by $(u,v)_\mathcal{H} = \int_\Omega \rho u \cdot v$, see Remark 2.2. Note that the jump in μ strongly influences the behavior of the semi-norms $|\cdot|_\ell$ which were introduced in Subsection 4.2. We take

$$p = \left(\frac{32}{7}t^2 z\right)\chi_{\Omega_+}(x,y,z),$$

$$u = \begin{pmatrix} 2t^2 z(x^2 + y^2) \\ -4t^2 xyz \\ -2/3t^2 x(x^2 + 3y^2) \end{pmatrix}$$

and the corresponding f and g. We use a similar tetrahedral triangulation of the domain and obtain the space $\mathcal{U}^h$ in the same way. To account for the discontinuity in p we use the following XFEM space, cf. [17]:

$$\mathcal{Q}_1^{h,X} := \mathcal{R}_+ \mathcal{Q}_1^h \oplus \mathcal{R}_- \mathcal{Q}_1^h,$$

where $\mathcal{Q}_1^h$ is the standard finite element space of piecewise linears, cf. (7.4), and $\mathcal{R}_\pm : v \mapsto v\chi_{\Omega_\pm}$ is the restriction operator to the subdomain $\Omega_\pm$. The basis for $\mathcal{V}_2^h$ and the interpolation operator for the boundary data g are as in the previous experiment. For

our particular choice of u, p and μ, we have that $f \in C^\infty(I; L^2(\Omega)^d)$. From this, $\rho u' \in C^\infty(I; L^2(\Omega))$ and $p \in C^\infty(I; L^2(\Omega))$ we can conclude that $\operatorname{div}(\mu Du) \in C^\infty(I; \mathcal{V}_{\mathcal{H}}')$, where $\mathcal{V}_{\mathcal{H}}$ is defined as in Example 5.1. It follows that $\int_0^T |\operatorname{div}(\mu Du^{(2)})|_0^2 < \infty$. We can expect at least a temporal superconvergence rate of 2.5 in Theorem 4.6. This is what we indeed observe in Table 7.8, namely a convergence of (approximately) order $\mathcal{O}(k^{2.5} + h^3)$, where after a few temporal refinements the spatial error dominates. In the global $L^2 \otimes H^1$-norm we observe an optimal temporal convergence order in Table 7.7.

Concerning the pressure discretization we note the following. For given $t \in I$ the pressure $p(t)$ can be expressed *exactly* in the XFEM space. Hence, the spatial convergence order of the pressure discretization is not relevant. Due to Theorem 5.4 we expect at least a temporal convergence order $\mathcal{O}(k^{1.5})$. The actual temporal convergence order, which we see in Table 7.9, is $\mathcal{O}(k^2)$. This is despite the fact that the projection $P_{\mathcal{H}}$ form Assumption 5.1 involves the discontinuous coefficient ρ and the domain Ω has a re-entrant corner. The latter implies that Assumption 5.1 does *not* hold. For an explanation of this we refer to Remark 5.1.

Table 7.7: Error in $L^2 \otimes H^1$-norm between u and the solution of (6.6). The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	0.12001	0.09779	0.11857	0.12821	0.12821	0.12954	
8	0.08439	0.02558	0.01569	0.01503	0.01497	0.01497	3.11318
16	0.08483	0.02148	0.00612	0.00333	0.00307	0.00306	2.29090
32	0.08388	0.02101	0.00528	0.00138	5.40e-04	4.35e-04	2.81364
EOC_T		1.99728	1.99302	1.93063	1.35955	0.31024	

Table 7.8: Maximal nodal error $\|u^n - U_h^n\|_{L^2}$, where U_h is the solution of (6.6). The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	0.03099	0.03173	0.03384	0.03479	0.03479	0.03491	
8	0.00420	0.00362	0.00361	0.00361	0.00361	0.00361	3.27398
16	0.00243	6.05e-04	4.47e-04	4.42e-04	4.42e-04	4.42e-04	3.02991
32	0.00234	3.77e-04	7.98e-05	5.34e-05	5.25e-05	5.25e-05	3.07321
EOC_T		2.63560	2.24033	0.58021	0.02296	0.00057	

Table 7.9: Error in $L^2 \otimes L^2$-norm between p and the solution of (6.6). The estimated temporal (spacial) order of convergence EOC_T (EOC_S) is computed using the last row (column).

$N_S \backslash N$	4	8	16	32	64	128	EOC_S
4	0.60072	0.85615	1.48345	1.67915	1.67916	1.75676	
8	0.26843	0.22768	0.22215	0.24958	0.24960	0.24962	2.81511
16	0.21029	0.06563	0.03246	0.03015	0.02991	0.03032	3.04123
32	0.21296	0.05310	0.01362	0.00413	0.00257	0.00245	3.63160
EOC_T		2.00379	1.96269	1.72067	0.68326	0.07327	

8. Conclusion and outlook. We have studied, in an abstract framework, DG time discretization methods for parabolic problems with non-homogeneous linear constraints. Two common ways of treating such linear constraints, namely explicit or implicit (via Lagrange multpliers), were discussed. These different treatments lead to different variational formulations of the parabolic problem. For these formulations we introduced a *modification* of the standard DG time discretization method in which an appropriate projection is used in the discretization of the constraint; see (3.6)-(3.7) and (3.8)-(3.9). For these discretizations (optimal) discretization error bounds are derived; see Theorem 4.2, Theorem 4.5, Theorem 5.4, Theorem 5.6. In the latter two theorems error bounds for the Lagrange multiplier are presented. As far as we know, even for the case with homogeneous constraints, such bounds are not available in the literature, yet. Numerical experiments confirm the predicted optimal convergence. Furthermore, experiments show that without the modification the (standard) DG method yields sub-optimal results.

We consider the following topics to be of interest for future research. The fully discrete scheme has merely been introduced in this paper and no error bounds for the fully discrete (in time and space) schemes have been derived. It seems not straightforward how to obtain satisfactory error bounds for the fully discrete schemes. In particular it is not clear how to derive a superconvergence result and optimal convergence results for the Lagrange multiplier. Another topic is related to problems where the occurring operators, both in the parabolic equation (A and $(\cdot, \cdot)_{\mathcal{H}}$) and in the constraints (B_1 and B_2), depend on time. It is not clear which convergence results hold (and can be proved) for problems with time-dependent operators, especially ones with low regularity. This includes, for example, operators that arise from multi-phase flow problems.

REFERENCES

[1] N. AHMED, S. BECHER, AND G. MATTHIES, *Higher-order discontinuous Galerkin time stepping and local projection stabilization techniques for the transient Stokes problem*, Computer Methods in Applied Mechanics and Engineering, 313 (2017), pp. 28–52.

[2] N. AHMED, A. LINKE, AND C. MERDON, *On Really Locking-Free Mixed Finite Element Methods for the Transient Incompressible Stokes Equations*, SIAM Journal on Numerical Analysis, 56 (2018), pp. 185–209.

[3] R. ALTMANN AND C. ZIMMER, *Runge-Kutta methods for linear semi-explicit operator differential-algebraic equations*, Mathematics of Computation, 87 (2017), pp. 149–174.

[4] D. BRAESS, *Finite elements: theory, fast solvers, and applications in elasticity theory*, Cambridge University Press, 3 ed., 2007.

[5] C. BRENNECKE, A. LINKE, C. MERDON, AND J. SCHÖBERL, *Optimal and pressure-independent L^2 velocity error estimates for a modified Crouzeix-Raviart Stokes element with BDM reconstructions*, J. Comput. Math., 33 (2015), pp. 191–208.

[6] M. DELFOUR, W. HAGER, AND F. TROCHU, *Discontinuous Galerkin methods for ordinary differential equations*, Mathematics of Computation, 36 (1981), pp. 455–473.

[7] D. A. DI PIETRO AND A. ERN, *Mathematical Aspects of Discontinous Galerkin Methods*, Springer-Verlag New York, Inc., 2012.

[8] H. ELMAN, D. SILVESTER, AND A. WATHEN, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, Numerical mathematics and scientific computation, Oxford University Press, 2014.

[9] K. ERIKSSON AND C. JOHNSON, *Adaptive Finite Element Methods for Parabolic Problems I: A Linear Model Problem*, SIAM Journal on Numerical Analysis, 28 (1991), pp. 43–77.

[10] ———, *Adaptive Finite Element Methods for Parabolic Problems II: Optimal Error Estimates in $L_\infty L_2$ and $L_\infty L_\infty$*, SIAM Journal on Numerical Analysis, 32 (1995), pp. 706–740.

[11] K. ERIKSSON, C. JOHNSON, AND S. LARSSON, *Adaptive Finite Element Methods for Parabolic Problems VI: Analytic Semigroups*, SIAM Journal on Numerical Analysis, 35 (1998), pp. 1315–1325.

[12] A. ERN AND J. GUERMOND, *Theory and Practice of Finite Elements*, vol. 159 of Applied Mathematical Sciences, Springer New York, 2013.

[13] D. ESTEP, *A Posteriori Error Bounds and Global Error Control for Approximation of Ordinary Differential Equations*, SIAM Journal on Numerical Analysis, 32 (1995), pp. 1–48.

[14] R. FALK AND M. NEILAN, *Stokes complexes and the construction of stable finite element methods with pointwise mass conservation*, SIAM J. Numer. Anal., 51 (2013), pp. 1308–1326.

[15] V. GIRAULT AND P. RAVIART, *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*, Springer Series in Computational Mathematics, Springer Berlin Heidelberg, 2012.

[16] S. GROSS, J. PETERS, V. REICHELT, AND A. REUSKEN, *The DROPS Package for Numerical Simulations of Incompressible Flows Using Parallel Adaptive Multigrid Techniques*, Preprint 227, IGPM, RWTH Aachen University, 2002.

[17] S. GROSS AND A. REUSKEN, *Numerical Methods for Two-phase Incompressible Flows*, Springer Series in Computational Mathematics, Springer Berlin Heidelberg, 2011.

[18] R. GUBEROVIC, C. SCHWAB, AND R. STEVENSON, *Space-time variational saddle point formulations of Stokes and Navier–Stokes equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 48 (2014), pp. 875–894.

[19] J. GUZMAN AND M. NEILAN, *Conforming and divergence-free Stokes elements in three dimensions*, IMA J. Numer. Anal., 34 (2013), pp. 1489–1508.

[20] J. S. HESTHAVEN AND T. WARBURTON, *Nodal Discontinuous Galerkin Methods*, Springer-Verlag New York, Inc., 2008.

[21] P. JAMET, *Galerkin-Type Approximations which are Discontinuous in Time for Parabolic Equations in a Variable Domain*, SIAM Journal on Numerical Analysis, 15 (1978), pp. 912–928.

[22] C. JOHNSON, *Error Estimates and Adaptive Time-Step Control for a Class of One-Step Methods for Stiff Ordinary Differential Equations*, SIAM Journal on Numerical Analysis, 25 (1988), pp. 908–926.

[23] S. LARSSON, V. THOMÉE, AND L. B. WAHLBIN, *Numerical solution of parabolic integrodifferential equations by the discontinuous Galerkin method*, Mathematics of Computation of the American Mathematical Society, 67 (1998), pp. 45–71.

[24] P. L. LEDERER, C. LEHRENFELD, AND J. SCHÖBERL, *Hybrid Discontinuous Galerkin methods with relaxed H(div)-conformity for incompressible flows. Part I*, arXiv preprint arXiv:1707.02782, (2017).

[25] C. LEHRENFELD AND J. SCHÖBERL, *High order exactly divergence-free Hybrid Discontinuous Galerkin Methods for unsteady incompressible flows*, Computer Methods in Applied Mechanics and Engineering, 307 (2016), pp. 339–361.

[26] A. LINKE, G. MATTHIES, AND L. TOBISKA, *Robust arbitrary order mixed finite element methods for the incompressible Stokes equations with pressure independent velocity errors*, ESAIM:M2AN, 50 (2016), pp. 289–309.

[27] A. QUARTERONI AND A. VALLI, *Numerical Approximation of Partial Differential Equations*, Springer-Verlag New York, Inc., 1994.

[28] D. SCHÖTZAU AND C. SCHWAB, *Time Discretization of Parabolic Problems by the hp-Version of the Discontinuous Galerkin Finite Element Method*, SIAM Journal on Numerical Analysis, 38 (2000), pp. 837–875.

[29] C. SCHWAB AND R. STEVENSON, *Space-Time Adaptive Wavelet Methods for Parabolic Evolution Problems*, Mathematics of Computation, 78 (2009), pp. 1293–1318.

[30] O. STEINBACH AND H. YANG, *Comparison of algebraic multigrid methods for an adaptive space–time finite-element discretization of the heat equation in 3D and 4D*, Numerical Linear Algebra with Applications, (2018). http://dx.doi.org/10.1002/nla.2143.

[31] R. STEVENSON, *Divergence-Free Wavelets on the Hypercube: General Boundary Conditions*, Constructive Approximation, 44 (2016), pp. 233–267.

[32] V. THOMÉE, *Galerkin Finite Element Methods for Parabolic Problems (Springer Series in Computational Mathematics)*, Springer-Verlag New York, Inc., 2006.

[33] N. J. WALKINGTON, *Combined DG–CG time stepping for wave equations*, SIAM Journal on Numerical Analysis, 52 (2014), pp. 1398–1417.

[34] J. WLOKA, *Partial Differential Equations*, Cambridge University Press, 1987.

[35] S. ZHANG, *A new family of stable mixed finite elements for the 3D Stokes equations*, Math. Computation, 74 (2005), pp. 543–554.