Discontinuous Galerkin method in time combined with a stabilized finite element method in space for linear first-order PDEs

Alexandre Ern*, Friedhelm Schieweck†

February 17, 2014

Abstract

We analyze the discontinuous Galerkin method in time combined with a finite element method with symmetric stabilization in space to approximate evolution problems with a linear, first-order differential operator. A unified analysis is presented for space discretization, including the discontinuous Galerkin method and H^1-conforming finite elements with interior penalty on gradient jumps. Our main results are error estimates in various norms for smooth solutions. Two key ingredients are the post-processing of the fully discrete solution by lifting its jumps in time and a new time-interpolate of the exact solution. We first analyze the $L^\infty(L^2)$ and $L^2(L^2)$ errors and derive a super-convergent bound of order $(\tau^{k+2} + h^{r+1/2})$ in the case of static meshes for $k \geq 1$. Here, τ is the time step, k the polynomial order in time, h the size of the space mesh, and r the polynomial order in space. For the case of dynamically changing meshes, we derive a novel bound on the resulting projection error. Finally, we prove new optimal bounds on static meshes for the error in the time-derivative and in the discrete graph norm.

Keywords: discontinuous Galerkin in time, stabilized FEM, first-order PDEs, graph norm error estimates, superconvergence, dynamic meshes

2000 Mathematics Subject Classification (MSC): 65M12, 65M60, 65J10

1 Introduction

Our goal is to analyze the discontinuous Galerkin (dG) method in time combined with a finite element method (FEM) with symmetric stabilization in space to approximate the linear evolution problem

$$\partial_t u + Au = f \qquad \text{in } \Omega \times (0,T), \tag{1}$$

completed with suitable boundary and initial conditions, see Section 2. Here, Ω is a domain in $\mathbb{R}^d$, $d \geq 1$, T some positive final time, f a source term, and A some linear, time-independent,

*Universite Paris-Est, CERMICS (ENPC), 77455 Marne la Vallee Cedex 2, France, Email: ern@cermics.enpc.fr

†Institut für Analysis und Numerik, Otto-von-Guericke-Universität Magdeburg, Postfach 4120, D-39016 Magdeburg, Germany, Email: schiewec@ovgu.de

first-order differential operator in space. We assume for simplicity that A is an advection-reaction operator; more generally, A can be a Friedrichs' operator, whereby (1) is a system of first-order PDEs endowed with a symmetry and a positivity property [21, 19].

The dG method in time can be used, in the spirit of Rothe's method, to semi-discretize in time the evolution problem (1). This method uses piecewise polynomial ansatz and trial spaces of some order $k \geq 0$, whose elements are V-valued functions of time that can be discontinuous at the discrete nodes defining the time partition; here, V denotes the graph space associated with the space differential operator A. The time semi-discrete problem can then be discretized in space by a stabilized FEM. A relatively wide class of stabilized FEM is that based on symmetric stabilization, including on the one hand the dG method in space [30, 29, 16] and on the other hand H^1-conforming finite elements with various stabilization techniques, e.g., interior penalty on gradient jumps [6, 9], local projection [4, 34], subgrid viscosity [22, 23], or orthogonal subscales [14, 15]. When used to approximate the steady version of (1), all of these methods lead to quasi-optimal L^2-error estimates of order $h^{r+1/2}$, where h denotes the size of the space mesh and r the polynomial order in space, and optimal error estimates in the discrete graph norm. In what follows, we consider a unified analysis for a discrete differential operator A_h satisfying three design properties (consistency, stability, and boundedness). These properties cover, as main examples, dG methods and H^1-conforming finite elements with interior penalty on gradient jumps. Minor modifications are to be included for other (weakly) non-consistent stabilization techniques.

The dG method in time has been extensively studied in [37, 17], see also references therein, for linear parabolic problems, either as time semi-discretization method in the Hilbert space H^1 or combined with H^1-conforming finite elements for space discretization. One of the main results for linear parabolic problems with a symmetric coercive operator A is super-convergence in $L^\infty(L^2)$ (L^∞ in time and L^2 in space) and $L^2(L^2)$ norms of order $(\tau^{2k+1}+h^{r+1})$ where τ is the time step and h the space mesh size. Further results on the dG method in time, combined with dG methods in space, concern the convergence to entropy solutions for hyperbolic conservation laws [28], the hp-analysis for diffusion and incompressible flow problems [35, 39], and inviscid compressible flows [38]. $L^\infty(L^2)$ and $L^2(L^2)$ error estimates of order $(\tau^{k+1} + h^r)$ are derived in [20, 10] for nonlinear convection-diffusion problems on time-varying meshes under the assumption $h^2 \lesssim \tau$. Finally, we mention the work [2] for linear transient convection-diffusion-reaction problems on static meshes where the dG method in time is combined with local projection stabilization in space for H^1-conforming finite elements leading to an $L^\infty(L^2)$ and $L^2(L^2)$ error bound of order $(\tau^{k+1}+h^{r+1/2})$ if the diffusion parameter ε is less than h.

An alternative approach for discretizating the evolution problem (1) is the explicit Runge–Kutta (RK) method in time combined with the dG method in space and suitable limiters, see [13, 12, 11]. Since the time-stepping scheme is explicit, such methods are computationally effective, but are only conditionally stable, and the error bounds require the application of the Gronwall argument which implies that the error constant grows exponentially with respect to the final time T. The analysis of explicit RK methods combined with stabilized FEM entails some subtleties. Error estimates are available for second-order (RK2) and third-

order (RK3) methods; see [42, 43] for nonlinear conservation laws and dG methods in space and [8] for Friedrichs' systems and FEM with symmetric stabilization. The main results are $L^\infty(L^2)$ error estimates of order $(\tau^2 + h^{r+1/2})$ for RK2 under a tightened 4/3-CFL condition (except for piecewise affine polynomials in space where the usual CFL condition suffices) and of order $(\tau^3 + h^{r+1/2})$ for RK3 under the usual CFL condition. In contrast to explicit RK methods, the dG method in time is unconditionally stable, and leads, in some cases, to super-convergent error estimates. The prize to pay is obviously increased computational cost, although this drawback can be tamed using efficient multigrid solvers, as in [25, 27, 26] for the heat, Stokes, and Navier–Stokes equations. Antoher advantage of dG methods in time is that their analysis readily encompasses all polynomial orders $k \geq 1$ in time (the lowest-order case $k = 0$, corresponding to the implicit Euler scheme, being slightly different). In addition, the derived error bounds do not require the Gronwall argument, and the error constant grows only like $T^{1/2}$ with respect to the final time T. Moreover, since they hinge on the weak form of the evolution problem, dG methods in time can be cast more effectively into optimization problems constrained by time-dependent PDEs [36, 31, 33]. Finally, we mention the recent analysis of implicit RK methods in time combined with dG methods in space for linear Maxwell's equations, leading in particular to an $L^\infty(L^2)$ error estimate of order $(\tau^{s+1} + h^{r+1/2})$ where s is the number of distinct time nodes in the RK method [24].

The main contributions of this paper are as follows. Let u be the exact solution and let $u_{\tau h}$ be the fully discrete solution. We analyze the error between the exact solution and a post-processed discrete solution $\tilde{u}_{\tau h} = \mathcal{L}_\tau u_{\tau h}$ which is continuous in time and is a piecewise polynomial in time of order $(k+1)$. The operator $\mathcal{L}_\tau$, motivated by [32] where the link between the dG method in time of order k and a continuous Petrov–Galerkin method of order $(k+1)$ with reduced integration is explored, is the first key ingredient of our analysis. The second one is a new time-interpolate $R_\tau^{k+1}u$ of the exact solution u of order $(k+1)$, which, in particular, interpolates at the $(k+1)$ right-sided Gauss–Radau integration points the time-derivative of a C^1-interpolate of order $(k+2)$ of the exact solution. Our first main result, which improves on the current state-of-the-art, are estimates for the error $(u - \tilde{u}_{\tau h})$ in $L^\infty(L^2)$ at the discrete nodes defining the time partition and in $L^2(L^2)$, both of order $(\tau^{k+2} + h^{r+1/2})$. This result implies, in particular, a super-convergent estimate of the same order for the error $(u - u_{\tau h})$ at the discrete time nodes. The error estimates are presented first for static space meshes. For time-varying meshes, there is an additional projection error due to mesh changes, for which we derive a novel sharp estimate compared to existing results [40, 1, 20, 10]. Our second main result are estimates on the error derivatives; we focus on static space meshes for simplicity. To our knowledge, such estimates are not yet available even for static meshes. We emphasize that their derivation is more delicate than for parabolic PDEs since the differential operator A is neither symmetric nor coercive in the graph norm. Our idea is to measure the error in the time-derivative using the post-processed time-derivative of the post-processed discrete solution, yielding a bound on $(\partial_t u - \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h})$ of order $(\tau^{k+1} + h^{r+1/2})$ in $L^\infty(L^2)$ at the discrete time nodes and in $L^2(L^2)$. Finally, an optimal bound for the error $(u - \tilde{u}_{\tau h})$ in the discrete graph norm is derived using the inf-sup stability of the discrete operator A_h.

The paper is organized as follows. Section 2 presents the continuous setting and Section 3

the discrete setting in time and in space, including the lifting operator $\mathcal{L}_\tau$. Section 4 contains some preliminary results for the error analysis and, in particular, introduces the new time-interpolate $R_\tau^{k+1}u$. Section 5 is devoted to the $L^\infty(L^2)$ and $L^2(L^2)$ error estimates. Finally, Section 6 deals with the estimates on the error derivatives.

2 The continuous setting

Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a domain with Lipschitz boundary Γ, and let T be some positive final time. We consider the model problem: *Find* $u : \Omega \times [0,T] \to \mathbb{R}$ *such that*

$$\begin{aligned} \partial_t u + Au &= f \qquad \text{in } \Omega \times (0,T), \\ u &= 0 \qquad \text{on } \Gamma_- \times (0,T), \\ u &= u_0 \qquad \text{on } \Omega \times \{0\}, \end{aligned} \tag{2}$$

where the linear first-order differential operator A is of the form $Au := \beta \cdot \nabla u + \sigma u$, $\beta : \Omega \to \mathbb{R}^d$ is a given Lipschitz convection field, $\sigma : \Omega \to \mathbb{R}$ a bounded reaction function, $f : \Omega \times [0,T] \to \mathbb{R}$ a source term, $u_0 : \Omega \to \mathbb{R}$ a given initial value of u, and Γ_- (resp., Γ_+) the inflow (resp., outflow) part of the boundary defined as

$$\Gamma_\pm := \{x \in \Gamma : \ \pm\beta(x) \cdot n > 0\},$$

with n denoting the outer normal unit vector on Γ. We assume that the data β and σ are *time-independent*, and that

$$\sigma(x) - \frac{1}{2}\operatorname{div}\beta(x) \geq \mu_0 > 0 \qquad \forall\, x \in \Omega. \tag{3}$$

The inner product in $L := L^2(\Omega)$ is denoted $(\cdot,\cdot)_L$, and the norm $\|v\|_L^2 := (v,v)_L$. The graph space defined as $\{v \in L : \ \beta \cdot \nabla v \in L\}$ is a Hilbert space when equipped with the graph norm $\|v\|_V^2 := \mu_0\|v\|_L^2 + \|\beta \cdot \nabla v\|_L^2$ (and the corresponding inner product). Assuming that Γ_- and Γ_+ are well-separated, $V := \{v \in L : \ \beta \cdot \nabla v \in L,\ v|_{\Gamma_-} = 0\}$ is a closed subspace of the graph space, and the operator $A : V \to L$ is an isomorphism; see, e.g., [19].

To characterize the smoothness of functions $t \mapsto v(t)$, we introduce, for a subinterval $J \subset [0,T]$, the space $C^r(J,B)$ spanned by r times continuously differentiable, B-valued functions on J, where B denotes a Banach space with norm $\|\cdot\|_B$ (typically, $B \in \{L, V\}$). The space $C^r(J,B)$ is equipped with the following norm and semi-norm:

$$\|v\|_{C^r(J,B)} := \max_{0\leq k\leq r} \sup_{t\in J} \|\partial_t^k v(t)\|_B, \qquad |v|_{C^r(J,B)} := \sup_{t\in J} \|\partial_t^r v(t)\|_B.$$

For a measurable subset $J \subset [0,T]$, we also use the Bochner space $L^2(J,B)$ defined as

$$L^2(J,B) := \{v : J \to B : \ \|v\|_{L^2(J,B)} < \infty\} \qquad \text{with} \qquad \|v\|_{L^2(J,B)} := \left(\int_J \|v(t)\|_B^2\, dt\right)^{1/2}.$$

Now, assuming $f \in C^0([0,T],L)$ and $u_0 \in V$, the model problem (2) can be written as the following linear evolution problem: *Find* $u \in C^0([0,T],V) \cap C^1([0,T],L)$ *such that*

$$\begin{aligned} (\partial_t u(t), v)_L + (Au(t), v)_L &= (f(t), v)_L \qquad \forall\, v \in L, \quad \forall\, t \in (0,T), \\ u(0) &= u_0. \end{aligned} \tag{4}$$

The well-posedness of (4) results from the Hille–Yosida Theorem; see, e.g., [41, p. 248] or [18, p. 313].

3 The discrete setting

We proceed in the spirit of Rothe's method whereby the evolution problem (4) is first semi-discretized in time, leading to a sequence of discrete problems in a Hilbert space, which are then discretized in space. An important ingredient for the time semi-discretization is the lifting operator $\mathcal{L}_\tau$ introduced in Section 3.2. In what follows, for positive real numbers a and b, $a \lesssim b$ stands for the inequality $a \le Cb$ with generic constant C independent of the size of the space and time meshes, of the final time, and of the exact solution u; the value of C can depend on the regularity of the spatial mesh, the polynomial degrees used for space and time discretization, and the model parameters (including Ω and the constant μ_0 in (3)).

3.1 Time semi-discretization by the dG(k)-method

In order to semi-discretize problem (4) in time, we decompose the time interval $I := (0, T]$ into N disjoint subintervals $I_n := (t_{n-1}, t_n]$, where $n = 1, \ldots, N$ and $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$, so that $I = \cup_{n=1}^N I_n$. Observe that all the time intervals are conventionally open at the left endpoint and closed at the right endpoint. In what follows, the maximum time step size $\tau := \max_{1\le n\le N} \tau_n$ where $\tau_n := t_n - t_{n-1}$ is used to denote the time discretization parameter, and the set of time intervals $\mathcal{M}_\tau := \{I_1, \ldots, I_N\}$ is called the *time mesh*.

We approximate the exact solution $u : \bar{I} \to V$ by means of a function $u_\tau : \bar{I} \to V$ which is piecewise a polynomial of some order $k \ge 0$ with respect to time. For $B \in \{L, V\}$, let $\mathbb{P}_k(I_n, B) := \left\{ w : I_n \to B \ : \ w(t) = \sum_{j=0}^k W^j t^j \,,\ \forall\, t \in I_n,\ W^j \in B,\ \forall\, j \right\}$ be the space of B-valued polynomials in time of order k over I_n and let

$$X_\tau^k(B) := \{ w_\tau : \bar{I} \to B : \ w_\tau\big|_{I_n} \in \mathbb{P}_k(I_n, B) \quad \forall\, I_n \in \mathcal{M}_\tau \}. \tag{5}$$

Then, we seek the time semi-discrete solution u_τ in the space $X_\tau^k(V)$. It is possible to consider a polynomial degree k_n specific to each time interval I_n. All what follows extends to this more general setting.

Let w_τ be a function in $X_\tau^k(B)$. Then, w_τ can be discontinuous at the discrete times t_n, for all $n = 0, \ldots, N$, while w_τ is by definition continuous from the left at t_n, for all $n = 1, \ldots, N$, i.e., $w_\tau(t_n) = \lim_{t \uparrow t_n} w_\tau(t)$. Moreover, $w_\tau(0)$ has to be specified separately since $0 \notin I_1$. The space $X_\tau^k(B)$ is a subspace of

$$C^{-1}(\mathcal{M}_\tau, B) := \{ w_\tau : \bar{I} \to B : \ w_\tau\big|_{I_n} \in C^0(I_n, B),\ w_\tau(t_{n-1}^+) \text{ exists } \forall\, I_n \in \mathcal{M}_\tau \}, \tag{6}$$

with the notation $w_\tau(t_{n-1}^+) := \lim_{t \downarrow t_{n-1}} w_\tau(t)$. Functions in $C^{-1}(\mathcal{M}_\tau, B)$ are by definition continuous from the left at all t_n, $n = 1, \ldots, N$, and their value at 0 has to be specified separately. For all $w_\tau \in C^{-1}(\mathcal{M}_\tau, B)$, the jump of w_τ at t_n, for all $n = 0, \ldots, N-1$, is defined as

$$[w_\tau]_n := w_\tau(t_n^+) - w_\tau(t_n). \tag{7}$$

The discontinuous Galerkin method of order k (in short, dG(k)) applied as time semi-discretization of problem (4) reads: *Find* $u_\tau \in X_\tau^k(V)$ *such that* $u_\tau(0) = u_0$ *and*

$$\sum_{n=1}^{N} \int_{I_n} (\partial_t u_\tau + A u_\tau, v_\tau)_L \, dt + \sum_{n=0}^{N-1} \big([u_\tau]_n, v_\tau(t_n^+)\big)_L = \int_I (f, v_\tau)_L \, dt \qquad \forall\, v_\tau \in Y_\tau^k(L), \quad (8)$$

with test space $Y_\tau^k(L) := \{v : I \to L : \ v\big|_{I_n} \in \mathbb{P}_k(I_n, L) \quad \forall\, I_n \in \mathcal{M}_\tau\}$.

Problem (8) can be decoupled into a sequence of local problems by choosing test functions supported on a single time interval I_n. Then, the time semi-discrete solution u_τ can be determined by successively solving a local problem on I_n. Using the known value $u_\tau(t_{n-1})$ from the previous time interval (and u_0 for $n = 1$), the local problem on I_n reads: *Find* $u_\tau|_{I_n} \in \mathbb{P}_k(I_n, V)$ *such that*

$$\int_{I_n} (\partial_t u_\tau + A u_\tau, v_\tau)_L \, dt + \big([u_\tau]_{n-1}, v_\tau(t_{n-1}^+)\big)_L = \int_{I_n} (f, v_\tau)_L \, dt \qquad \forall\, v_\tau \in \mathbb{P}_k(I_n, L). \quad (9)$$

In practice, the right-hand side of (9) is evaluated by means of some numerical integration formula. In the context of the dG(k)-method in time, a natural choice is to consider the (k+1)-point right-sided Gauss–Radau quadrature formula on each time interval I_n. For a function $g \in C^{-1}(\mathcal{M}_\tau, \mathbb{R})$ and $I_n \in \mathcal{M}_\tau$, this formula takes the form

$$Q_n(g) := \frac{\tau_n}{2} \sum_{\mu=1}^{k+1} \widehat{w}_\mu g(t_{n,\mu}) \approx \int_{I_n} g(t)\, dt, \quad (10)$$

where $t_{n,\mu} \in I_n$ are the integration points and $\widehat{w}_\mu > 0$ the weights. Note that $t_{n,k+1} = t_n$ so that $g(t_{n,k+1})$ is equal to $g(t_n)$ (recall that g is continuous from the left at t_n). It is known that (10) is exact for all polynomials in $\mathbb{P}_{2k}(I_n)$. Then, using (10) on the right-hand side of (9) leads to the numerically-integrated version of (9):

$$\int_{I_n} (\partial_t u_\tau + A u_\tau, v_\tau)_L \, dt + \big([u_\tau]_{n-1}, v_\tau(t_{n-1}^+)\big)_L = Q_n((f, v_\tau)_L) \qquad \forall\, v_\tau \in \mathbb{P}_k(I_n, L). \quad (11)$$

To rewrite (11), we define the Lagrange interpolation operator $I_\tau^{\mathrm{GR}} : C^0(\bar{I}, L) \to X_\tau^k(L)$ by means of the conditions

$$I_\tau^{\mathrm{GR}} w(t_{n,\mu}) = w(t_{n,\mu}), \quad \mu = 1, \dots, k+1, \ n = 1, \dots, N, \quad I_\tau^{\mathrm{GR}} w(0) = w(0), \quad (12)$$

using the (k+1) right-sided Gauss–Radau integration points $t_{n,\mu} \in I_n$. Then, (11) is equivalent to

$$\int_{I_n} (\partial_t u_\tau + A u_\tau, v_\tau)_L \, dt + \big([u_\tau]_{n-1}, v_\tau(t_{n-1}^+)\big)_L = \int_{I_n} (I_\tau^{\mathrm{GR}} f, v_\tau)_L \, dt \qquad \forall\, v_\tau \in \mathbb{P}_k(I_n, L). \quad (13)$$

Let us briefly look, for example, at the dG(1)-method. Here, we apply the 2-point right-sided Gauss–Radau quadrature formula with points $t_{n,1} = t_{n-1} + \tau_n/3$, $t_{n,2} = t_n$ and reference weights $\hat{w}_1 = 3/2$, $\hat{w}_2 = 1/2$. On the time interval I_n, we have to solve for the two unknowns $U_n^j = u_\tau(t_{n,j})$ for $j = 1, 2$. The coupled (2×2)-block-system for $U_n^1, U_n^2 \in V$ reads:

$$\begin{aligned} \tfrac{3}{4} U_n^1 + \tfrac{\tau_n}{2} A U_n^1 \ &+\ \tfrac{1}{4} U_n^2 &&= \ u_\tau(t_{n-1}) + \tfrac{\tau_n}{2} f(t_{n,1}), \\ -\tfrac{9}{4} U_n^1 \ &+\ \tfrac{5}{4} U_n^2 + \tfrac{\tau_n}{2} A U_n^2 &&= -u_\tau(t_{n-1}) + \tfrac{\tau_n}{2} f(t_{n,2}). \end{aligned} \quad (14)$$

3.2 A lifting operator

As a key point of our analysis, we introduce the lifting operator

$$\mathcal{L}_\tau : X_\tau^k(B) \to X_\tau^{k+1}(B) \cap C^0(\bar{I}, B), \tag{15}$$

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

$$\mathcal{L}_\tau w_\tau(t) := w_\tau(t) - [w_\tau]_{n-1}\vartheta_n(t) \qquad \forall\, t \in I_n = (t_{n-1}, t_n], \tag{16}$$

and $\mathcal{L}_\tau w_\tau(0) = w_\tau(0)$. Here, $\vartheta_n \in \mathbb{P}_{k+1}(I_n, \mathbb{R})$ is defined by means of the integration points $t_{n,\mu}$ of the $(k+1)$-point right-sided Gauss-Radau quadrature formula on the interval I_n as

$$\vartheta_n(t) := \prod_{\mu=1}^{k+1} \frac{t - t_{n,\mu}}{t_{n-1} - t_{n,\mu}} \qquad \forall\, t \in I_n. \tag{17}$$

The continuity in time of $\mathcal{L}_\tau w_\tau$ follows from the properties $\vartheta_n(t_{n-1}^+) = 1$ and $\vartheta_n(t_n) = \vartheta_n(t_{n,k+1}) = 0$ together with the definition of the jump $[w_\tau]_{n-1}$. Since $\vartheta_n(t)$ vanishes at the integration points, we get the property

$$\mathcal{L}_\tau w_\tau(t_{n,\mu}) := w_\tau(t_{n,\mu}) \qquad \forall\, \mu = 1, \ldots, k+1,\ \forall\, n = 1, \ldots, N. \tag{18}$$

The lifting operator $\mathcal{L}_\tau$ can be more generally defined for functions in $C^{-1}(\mathcal{M}_\tau, B)$ and then maps onto $C^0(\bar{I}, B)$.

3.3 Space discretization by stabilized FEM

In this section, we briefly recall some basic elements on the discretization of the space differential operator A by means of a stabilized FEM. For clarity, we consider functions depending only on the space variable, and return to the full space-time setting in Section 3.4.

Let $\mathcal{T}_h$ be a shape-regular mesh of Ω and let V_h be a finite element space built on that mesh, where h denotes the mesh-size. For simplicity, we assume that the mesh $\mathcal{T}_h$ is affine and that Ω has a polygonal ($d = 2$) or polyhedral ($d = 3$) boundary. To fix the ideas, we assume that V_h contains at least piecewise polynomials of order r, yielding the following local approximation property: For all $w \in H^{r+1}(\Omega)$, there exists an interpolate $i_h w \in V_h$ such that, for $m \in \{0, 1\}$,

$$|w - i_h w|_{H^m(T)} \lesssim h_T^{r+1-m} |w|_{H^{r+1}(\Delta_T)} \qquad \forall\, T \in \mathcal{T}_h, \tag{19}$$

where h_T stands for the diameter of T and Δ_T is a set of mesh elements neighboring T.

The differential operator $A : V \to L$ is approximated by a discrete differential operator $A_h : W + V_h \to V_h$, where W is a dense subspace of V used to assert the consistency of A_h (see (20) below); typically, $W = H^s(\Omega)$, $s \geq 1$. For the operator A_h, we require the following properties:

- *Consistency*: Letting $P_h : L \to V_h$ denote the L^2-orthogonal projector onto V_h,

$$A_h w = P_h A w \qquad \forall\, w \in W. \tag{20}$$

- *Discrete coercivity and boundedness on orthogonal subscales*:

$$|||v_h|||^2 \lesssim (A_h v_h, v_h)_L \qquad \forall\, v_h \in V_h, \tag{21}$$

$$(A_h(w - P_h w), v_h)_L \lesssim |||w - P_h w|||_{h,\frac{1}{2}} |||v_h||| \qquad \forall\, v_h \in V_h,\ w \in W, \tag{22}$$

  where $|||\cdot|||$ and $|||\cdot|||_{h,\frac{1}{2}}$ are mesh-dependent norms on $(W + V_h)$ satisfying

$$\mu_0^{1/2} \|v\|_L \le |||v||| \le |||v|||_{h,\frac{1}{2}} \qquad \forall\, v \in (W + V_h). \tag{23}$$

- *Discrete inf-sup stability*: For all $v_h \in V_h$, there exists $w_h \in V_h$ such that

$$|||v_h|||_\sharp^2 \lesssim (A_h v_h, w_h)_L \qquad \text{and} \qquad |||w_h|||_\sharp \le |||v_h|||_\sharp, \tag{24}$$

  where $|||\cdot|||_\sharp$ is a mesh-dependent norm on V_h such that $|||v_h||| \le |||v_h|||_\sharp$ for all $v_h \in V_h$, and $|||w|||_\sharp \lesssim \|w\|_W$ for all $w \in W$. In what follows, $|||\cdot|||_\sharp$ is termed the discrete graph norm since it provides a control on the advective derivative.

We now present two examples for the discrete operator A_h matching the above framework, one obtained using the Continuous Interior Penalty (CIP) method based on H^1-conforming finite elements and gradient jump penalty at interfaces and the other obtained using the dG method (in space) with upwind fluxes. Some slight adaptations of the consistency assumption are needed to handle other stablizations for H^1-conforming finite elements (e.g., local projection, subgrid viscosity, orthogonal subscales). Mesh faces are collected in the set $\mathcal{F}_h$ split into the set of interior faces, $\mathcal{F}_h^{\rm int}$, and boundary faces, $\mathcal{F}_h^{\rm ext}$. For $F \in \mathcal{F}_h^{\rm int}$, there are T^-, T^+ in $\mathcal{T}_h$ such that $F = \partial T^- \cap \partial T^+$, n_F is the unit normal to F pointing from T^- to T^+, and for a piecewise smooth enough function v, we define its jump and mean value at F as $[v] := v|_{T^-} - v|_{T^+}$ and $\{\!\{v\}\!\} := \frac{1}{2}(v|_{T^-} + v|_{T^+})$, respectively. The arbitrariness in the sign of $[v]$ is irrelevant. Meshes can possess hanging nodes when using dG methods under the usual assumption that face diameters are comparable to local element diameters. In what follows, h_F denotes the diameter of $F \in \mathcal{F}_h$, and for $R \in \{T, \partial T, F\}$, $(\cdot,\cdot)_{L,R}$ the $L^2(R)$-inner product with associated norm $\|\cdot\|_{L,R}$.

In the CIP method, see [6, 9, 7], the discrete space V_h is H^1-conforming, and, letting $W = H^2(\Omega)$, the discrete operator $A_h^{\rm cip}$ is given, for all $v \in W + V_h$, $w_h \in V_h$, by

$$\begin{aligned}(A_h^{\rm cip} v, w_h)_L = (\sigma v + \beta\cdot\nabla v, w_h)_L &+ \sum_{F\in\mathcal{F}_h^{\rm ext}} ((\beta\cdot n_F)^\ominus v, w_h)_{L,F} \\ &+ \sum_{F\in\mathcal{F}_h^{\rm int}} (\gamma_F [\nabla v], [\nabla w_h])_{L,F},\end{aligned} \tag{25}$$

where $x^\ominus := \frac{1}{2}(|x| - x)$ denotes the negative part of a real number x and $\gamma_F = \varrho h_F^2 |\beta\cdot n_F|$ with $\varrho > 0$ a user-dependent parameter. Then, consistency holds. Discrete coercivity holds with the mesh-dependent norm on $(W + V_h)$

$$|||v|||^2 := \mu_0 \|v\|_L^2 + \sum_{F\in\mathcal{F}_h^{\rm ext}} \||\beta\cdot n_F|^{1/2} v\|_{L,F}^2 + \sum_{F\in\mathcal{F}_h^{\rm int}} \|\gamma_F^{1/2}[\nabla v]\|_{L,F}^2, \tag{26}$$

boundedness on orthogonal subscales with $|||v|||_{h,\frac{1}{2}}^2 = |||v|||^2 + \sum_{T\in\mathcal{T}_h}\{h_T^{-1}\|v\|_{L,T}^2 + \|v\|_{L,\partial T}^2\}$, and discrete inf-sup stability with $|||v_h|||_\sharp^2 := |||v_h|||^2 + \sum_{T\in\mathcal{T}_h} h_T\|\beta\cdot\nabla v_h\|_{L,T}^2$.

In the DG method, see [30, 29, 5, 19, 16], the discrete space V_h is spanned by piecewise polynomials with no continuity explicitly enforced at interfaces, and, letting $W = H^1(\Omega)$, the discrete operator $A_h^{\rm dg}$ is given, for all $v \in W + V_h$, $w_h \in V_h$, by

$$\begin{aligned}(A_h^{\rm dg}v, w_h)_L = &\sum_{T\in\mathcal{T}_h}(\sigma v + \beta\cdot\nabla v, w_h)_{L,T} + \sum_{F\in\mathcal{F}_h^{\rm ext}}((\beta\cdot n_F)^\ominus v, w_h)_{L,F}\\ &- \sum_{F\in\mathcal{F}_h^{\rm int}}((\beta\cdot n_F)[v], \{\!\{w_h\}\!\})_{L,F} + \sum_{F\in\mathcal{F}_h^{\rm int}}(\gamma_F[v],[w_h])_{L,F},\end{aligned} \tag{27}$$

where $\gamma_F = \varrho|\beta\cdot n_F|$ with $\varrho > 0$ a user-dependent parameter ($\varrho = \frac{1}{2}$ yields the classical upwind fluxes). Then, consistency holds. Discrete coercivity holds with the mesh-dependent norm on $(W + V_h)$

$$|||v|||^2 := \mu_0\|v\|_L^2 + \sum_{F\in\mathcal{F}_h^{\rm ext}}\||\beta\cdot n_F|^{1/2}v\|_{L,F}^2 + \sum_{F\in\mathcal{F}_h^{\rm int}}\|\gamma_F^{1/2}[v]\|_{L,F}^2, \tag{28}$$

boundedness on orthogonal subscales with $|||v|||_{h,\frac{1}{2}}^2 = |||v|||^2 + \sum_{T\in\mathcal{T}_h}\|v\|_{L,\partial T}^2$, and discrete inf-sup stability with $|||v_h|||_\sharp^2 := |||v_h|||^2 + \sum_{T\in\mathcal{T}_h} h_T\|\beta\cdot\nabla v_h\|_{L,T}^2$.

To derive error estimates, we need suitable approximation properties of the L^2-orthogonal projector P_h, namely

$$|||w - P_hw|||_{h,\frac{1}{2}} + |||w - P_hw|||_\sharp \lesssim h^{r+1/2}|w|_{H^{r+1}(\Omega)}. \tag{29}$$

Such a property is satisfied for shape-regular meshes and any polynomial degree r for the DG method since the projector P_h enjoys local approximation properties. For conforming FEM with CIP, we use the recent result of [3] focusing for simplicity on simplicial meshes. We assume that to each mesh element $T \in \mathcal{T}_h$, we can assign a nonnegative integer k_T, the level of T, such that $h_T \sim 2^{-k_T}$ and such that for any two elements sharing a vertex, their level differs at most by one. We refer to [3] for a discussion on the relevance of this hypothesis in the context of adaptive and more general nonuniform meshes. Following [3], we also assume that, for $d = 2$, $r \le 12$ and, for $d = 3$, $r \le 7$. In what follows, the above assumptions are referred to as **(H)**.

Lemma 1 *Under Assumption **(H)**, the approximation property* (29) *holds.*

Proof. We only prove the bound on the $|||\cdot|||_{h,\frac{1}{2}}$-norm; the proof for the other bound is similar. By the triangle inequality, we infer that

$$|||w - P_hw|||_{h,\frac{1}{2}} \le |||w - i_hw|||_{h,\frac{1}{2}} + |||P_h(w - i_hw)|||_{h,\frac{1}{2}},$$

since $i_hv \in V_h$. The first term on the right-hand side is bounded by $h^{r+1/2}|w|_{H^{r+1}(\Omega)}$ owing to (19). Concerning the second term, we first observe using inverse and trace inequalities that $|||v_h|||_{h,\frac{1}{2}} \lesssim \|\mathfrak{h}^{-1/2}v_h\|_L$ for all $v_h \in V_h$, where $\mathfrak{h}$ denotes the piecewise constant function

equal locally to the mesh element diameter. Moreover, the Cauchy–Schwarz inequality implies $\|\mathfrak{h}^{-1/2}v_h\|_L \le \|\mathfrak{h}^{-1}v_h\|_L^{1/2}\|v_h\|_L^{1/2}$. Since Lemma 4.1 in [3] shows that $\|\mathfrak{h}^{-1}P_h v\|_L \lesssim \|\mathfrak{h}^{-1}v\|_L$ for all $v \in L$, we infer with $v = w - i_h w$ and $v_h = P_h v$ that

$$\|\!|P_h(w - i_h w)|\!\|_{h,\frac{1}{2}} \lesssim \|\mathfrak{h}^{-1}(w - i_h w)\|_L^{1/2}\|w - i_h w\|_L^{1/2},$$

whence the assertion results from (19). □

3.4 Full space-time discretization

In the full space-time discretization, we approximate on each time interval $I_n = (t_{n-1}, t_n]$ the time semi-discrete solution u_τ by means of a fully discrete solution $u_{\tau h}$ using a finite element space $V_h^n \subset L$ resulting from a mesh $\mathcal{T}_h^n$ which can change from one time interval to the next. The corresponding discrete differential operator is $A_h^n : W + V_h^n \to V_h^n$ and satisfies the design conditions of Section 3.3 uniformly in n. Concerning mesh-dependent norms like $\|\!|\cdot|\!\|$, we use a subscript n as in $\|\!|\cdot|\!\|_n$ to indicate that this norm is defined using the mesh $\mathcal{T}_h^n$. The global solution space for the fully discrete solution $u_{\tau h}$ is

$$X_{\tau h}^k := \{v_{\tau h} : \bar{I} \to L;\ v_{\tau h}\big|_{I_n} \in \mathbb{P}_k(I_n, V_h^n) \quad \forall\, I_n \in \mathcal{M}_\tau\}. \tag{30}$$

Let $u_{0h} \in V_h^0$ be an approximation of the initial condition u_0. The initial mesh $\mathcal{T}_h^0$ used to build V_h^0 can differ from the mesh $\mathcal{T}_h^1$ used in the first time interval I_1.

For brevity, we only present the fully discrete problem on each time interval, i.e., resulting from the space discretization of (11). For all $n = 1, \dots, N$, using the known value $u_{\tau h}(t_{n-1}) \in V_h^{n-1}$ from the previous time interval (and u_{0h} for $n = 1$), the local, fully discrete problem on I_n reads: *Find $u_{\tau h}|_{I_n} \in \mathbb{P}_k(I_n, V_h^n)$ such that, for all $v_{\tau h} \in \mathbb{P}_k(I_n, V_h^n)$,*

$$\int_{I_n} (\partial_t u_{\tau h} + A_h^n u_{\tau h}, v_{\tau h})_L \, dt + \big([u_{\tau h}]_{n-1}, v_{\tau h}(t_{n-1}^+)\big)_L = Q_n((f, v_{\tau h})_L)\,. \tag{31}$$

We now derive a useful result allowing us to rewrite the fully discrete scheme (31) using the lifting operator $\mathcal{L}_\tau$ introduced in Section 3.2.

Lemma 2 *For all $n = 1, \dots, N$,* (31) *is equivalent to*

$$\int_{I_n} (\partial_t \mathcal{L}_\tau u_{\tau h} + A_h^n u_{\tau h}, v_{\tau h})_L \, dt = Q_n((f, v_{\tau h})_L) \qquad \forall v_{\tau h} \in \mathbb{P}_k(I_n, V_h^n). \tag{32}$$

Proof. For all $n = 1, \dots, N$, using integration by parts for the ϑ_n-term, we obtain

$$\begin{aligned}\int_{I_n} (\partial_t \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L \, dt = &\int_{I_n} (\partial_t u_{\tau h}, v_{\tau h})_L \, dt + \int_{I_n} ([u_{\tau h}]_{n-1}\vartheta_n, \partial_t v_{\tau h})_L \, dt \\ &+ \big([u_{\tau h}]_{n-1}, v_{\tau h}(t_{n-1}^+)\big)_L\,,\end{aligned}$$

since $\vartheta_n(t_{n-1}) = 1$ and $\vartheta_n(t_n) = 0$. The integrand of the second integral on the right-hand side is in $\mathbb{P}_{2k}(I_n, \mathbb{R})$. Then, the $(k+1)$-point right-sided Gauss-Radau quadrature formula is exact and the integral vanishes. □

4 Preparation for the error analysis

4.1 Basic stability result

To state the consistency properties of the fully discrete problem, it is convenient to define, in each time interval I_n, $n = 1, \ldots, N$, the bilinear form

$$\widetilde{B}_h^n(w, v) := Q_n((\partial_t w, v)_L) + Q_n((A_h^n w, v)_L)\,, \tag{33}$$

where $v \in C^{-1}(\mathcal{M}_\tau, L)$ and w must satisfy the following smoothness conditions, expressed as $w \in \tilde{X}$ with

$$\tilde{X} := \left\{ w : \bar{I} \to L;\ w(t_{n,\mu}) \in W + V_h^n,\ \partial_t w\big|_{I_n}(t_{n,\mu}) \in L,\ \forall \mu = 1, \ldots, k+1,\ n = 1, \ldots, N \right\}, \tag{34}$$

where $t_{n,\mu} \in I_n$ are the right-sided Gauss–Radau integration points. Note that, although the function $w = \mathcal{L}_\tau u_{\tau h}$ restricted to I_n is in $\mathbb{P}_{k+1}(I_n, V_h^{n-1} + V_h^n)$, this function satisfies $w(t_{n,\mu}) \in V_h^n$ owing to (18); hence, $w \in \tilde{X}$. The bilinear form $\widetilde{B}_h^n$ is closely related to the fully discrete problem, as we now show.

Lemma 3 *The fully discrete solution $u_{\tau h} \in X_{\tau h}^k$ is such that, for all $n = 1, \ldots, N$,*

$$\widetilde{B}_h^n(\mathcal{L}_\tau u_{\tau h}, v_{\tau h}) = Q_n((f, v_{\tau h})_L) \qquad \forall\, v_{\tau h} \in \mathbb{P}_k(I_n, V_h^n). \tag{35}$$

Proof. Using definition (33) yields

$$\widetilde{B}_h^n(\mathcal{L}_\tau u_{\tau h}, v_{\tau h}) = Q_n((\partial_t \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L) + Q_n((A_h^n \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L)\,.$$

Since $(\partial_t \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L$ is in $\mathbb{P}_{2k}(I_n)$, we obtain $Q_n((\partial_t \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L) = \int_{I_n} (\partial_t \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L\, dt$. Moreover, using (18) and since $(A_h^n u_{\tau h}, v_{\tau h})_L$ is in $\mathbb{P}_{2k}(I_n, \mathbb{R})$, we infer that

$$Q_n((A_h^n \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L) = Q_n((A_h^n u_{\tau h}, v_{\tau h})_L) = \int_{I_n} (A_h^n u_{\tau h}, v_{\tau h})_L\, dt.$$

We conclude using (32). □

The discrete bilinear form $\widetilde{B}_h^n$ satisfies a basic stability result which is the starting point of our error analysis.

Lemma 4 (Stability) *For given $w \in C^0(\bar{I}, L) \cap X_\tau^{k+1}(L) \cap \tilde{X}$, let $I_\tau^{\mathrm{GR}} w \in X_\tau^k(L)$ be defined in* (12)*. Then, for all $n = 1, \ldots, N$, the following bound holds:*

$$\widetilde{B}_h^n(w, I_\tau^{\mathrm{GR}} w) \geq \frac{1}{2}\|w(t_n)\|_L^2 - \frac{1}{2}\|w(t_{n-1})\|_L^2 + \frac{1}{2}\|[I_\tau^{\mathrm{GR}} w]_{n-1}\|_L^2 + \alpha Q_n\big(|\!|\!| w |\!|\!|_n^2\big)\,. \tag{36}$$

Proof. Using definition (33) yields

$$\widetilde{B}_h^n(w, I_\tau^{\mathrm{GR}} w) = Q_n\big((\partial_t w, I_\tau^{\mathrm{GR}} w)_L\big) + Q_n\big((A_h^n w, I_\tau^{\mathrm{GR}} w)_L\big)\,.$$

Concerning the first term on the right-hand side, since the integrand is in $\mathbb{P}_{2k}(I_n,\mathbb{R})$, we obtain $Q_n\big((\partial_t w, I_\tau^{\mathrm{GR}} w)_L\big) = \int_{I_n} \big(\partial_t w, I_\tau^{\mathrm{GR}} w\big)_L \, dt$. Moreover, since $w \in \mathbb{P}_{k+1}(I_n, L)$, we observe that

$$w(t) = I_\tau^{\mathrm{GR}} w(t) + d_{n-1}\vartheta_n(t) \qquad \text{with} \qquad d_{n-1} := w(t_{n-1}) - I_\tau^{\mathrm{GR}} w(t_{n-1}^+) = -[I_\tau^{\mathrm{GR}} w]_{n-1},$$

$t \in I_n$ and ϑ_n defined in (17). This implies

$$\begin{aligned}
Q_n\big((\partial_t w, I_\tau^{\mathrm{GR}} w)_L\big) &= \int_{I_n} (\partial_t w, w - d_{n-1}\vartheta_n)_L \, dt \\
&= \int_{I_n} (\partial_t w, w)_L \, dt - \int_{I_n} (\partial_t w, d_{n-1}\vartheta_n)_L \, dt \\
&= \int_{I_n} \frac{1}{2}\frac{d}{dt}\|w\|_L^2 dt - \int_{I_n} \big(\partial_t I_\tau^{\mathrm{GR}} w + d_{n-1}\partial_t\vartheta_n, d_{n-1}\vartheta_n\big)_L \, dt \\
&= \int_{I_n} \frac{1}{2}\frac{d}{dt}\|w\|_L^2 dt - \|d_{n-1}\|_L^2 \int_{I_n} \frac{1}{2}\frac{d}{dt}|\vartheta_n|^2 dt,
\end{aligned}$$

since $\int_{I_n} \big(\partial_t I_\tau^{\mathrm{GR}} w, d_{n-1}\vartheta_n\big)_L \, dt = 0$ being the integrand in $\mathbb{P}_{2k}(I_n,\mathbb{R})$ and vanishing at all right-sided Gauss–Radau integration points in I_n. As a result,

$$Q_n\big((\partial_t w, I_\tau^{\mathrm{GR}} w)_L\big) = \frac{1}{2}\|w(t_n)\|_L^2 - \frac{1}{2}\|w(t_{n-1})\|_L^2 + \frac{1}{2}\|d_{n-1}\|_L^2.$$

Furthermore, using (21) (discrete coercivity) and since w and $I_\tau^{\mathrm{GR}} w$ coincide at all $t_{n,\mu}$, $\mu = 1,\dots,k+1$, we infer that

$$Q_n\big((A_h^n w, I_\tau^{\mathrm{GR}} w)_L\big) = Q_n\big((A_h^n I_\tau^{\mathrm{GR}} w, I_\tau^{\mathrm{GR}} w)_L\big) \geq \alpha Q_n\big(|\!|\!| I_\tau^{\mathrm{GR}} w |\!|\!|_n^2\big) = \alpha Q_n\big(|\!|\!| w |\!|\!|_n^2\big),$$

whence the assertion follows. □

4.2 Construction of a special interpolate in time

In this section, we assume $k \geq 1$. Let B be a Banach space, typically $B \in \{L, V\}$. For a function $u \in C^1(\bar{I}, B)$,, we define a time-polynomial interpolate $R_\tau^{k+1} u \in C^0(\bar{I}, B)$ whose restriction to $I_n = (t_{n-1}, t_n]$ is in $\mathbb{P}_{k+1}(I_n, B)$. We first choose a Lagrange/Hermite interpolate $I_\tau^{k+2} u \in C^0(\bar{I}, B)$ such that, for all $n = 1,\dots,N$, $I_\tau^{k+2} u|_{I_n} \in \mathbb{P}_{k+2}(I_n, B)$ and

$$I_\tau^{k+2} u(t_n) = u(t_n) \qquad \text{and} \qquad \partial_t I_\tau^{k+2} u(t_n) = \partial_t u(t_n) \qquad \forall\, n = 0,\dots,N.$$

For $k = 1$, these conditions fully determine $I_\tau^{k+2} u$ (for $k = 0$, the above construction is not possible), while, for $k \geq 2$, values at, say, additional Lagrange nodes can be prescribed inside each I_n so that, assuming u smooth enough,

$$\|\partial_t u - \partial_t I_\tau^{k+2} u\|_{C^0(\bar{I}_n, B)} \lesssim \tau_n^{k+2} |u|_{C^{k+3}(\bar{I}_n, B)}, \tag{37}$$

$$\|\partial_t^2 u - \partial_t^2 I_\tau^{k+2} u\|_{C^0(\bar{I}_n, B)} \lesssim \tau_n^{k+1} |u|_{C^{k+3}(\bar{I}_n, B)}. \tag{38}$$

Then, we define $R_\tau^{k+1} u|_{I_n} \in \mathbb{P}_{k+1}(I_n, B)$ by means of the $(k+2)$ conditions

$$\partial_t R_\tau^{k+1} u(t_{n,\mu}) = \partial_t I_\tau^{k+2} u(t_{n,\mu}) \qquad \forall\, \mu = 1,\dots,k+1, \tag{39}$$

$$R_\tau^{k+1} u(t_{n-1}^+) = I_\tau^{k+2} u(t_{n-1}), \tag{40}$$

and finally we set $R_\tau^{k+1} u(0) = u(0) = u_0$.

Lemma 5 *Assume $k \geq 1$. The function $R_\tau^{k+1}u$ is continuous in time on $\bar{I}$ with $R_\tau^{k+1}u(t_n) = u(t_n)$ for all $n = 0, \ldots, N$.*

Proof. The function $R_\tau^{k+1}u$ is continuous at 0 since $R_\tau^{k+1}u(0^+) = I_\tau^{k+2}u(0) = u(0) = R_\tau^{k+1}u(0)$. Let now $n = 1, \ldots, N$. From (39) and (40), we obtain for an arbitrary time-independent test function $v \in L$,

$$\begin{aligned}\Big(R_\tau^{k+1}u(t_n), v\Big)_L &= \Big(R_\tau^{k+1}u(t_{n-1}^+), v\Big)_L + \int_{I_n}\Big(\partial_t R_\tau^{k+1}u, v\Big)_L dt \\ &= \Big(I_\tau^{k+2}u(t_{n-1}), v\Big)_L + Q_n\Big(\Big(\partial_t R_\tau^{k+1}u, v\Big)_L\Big) \\ &= \Big(I_\tau^{k+2}u(t_{n-1}), v\Big)_L + Q_n\Big(\Big(\partial_t I_\tau^{k+2}u, v\Big)_L\Big).\end{aligned}$$

Since $\partial_t I_\tau^{k+2}u$ is in $\mathbb{P}_{k+1}(I_n, B)$ and $k+1 \leq 2k$ for all $k \geq 1$, we obtain $Q_n\big(\big(\partial_t I_\tau^{k+2}u, v\big)_L\big) = \int_{I_n}\big(\partial_t I_\tau^{k+2}u, v\big)_L dt$. As a result,

$$\Big(R_\tau^{k+1}u(t_n), v\Big)_L = \Big(I_\tau^{k+2}u(t_{n-1}), v\Big)_L + \int_{I_n}\Big(\partial_t I_\tau^{k+2}u, v\Big)_L dt = \Big(I_\tau^{k+2}u(t_n), v\Big)_L,$$

proving the assertion. □

Lemma 6 *Assume $k \geq 1$. For all $n = 1, \ldots, N$ and all $u \in C^{k+2}(\bar{I}_n, B)$, the following bound holds:*

$$\|u - R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} \lesssim \tau_n^{k+2}|u|_{C^{k+2}(\bar{I}_n,B)}. \tag{41}$$

Moreover, the bound $\|R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} \lesssim \max(1, \tau_n)\|u\|_{C^1(\bar{I}_n,B)}$ holds for all $u \in C^1(\bar{I}_n, B)$.

Proof. See appendix. □

Corollary 7 *Assume $k \geq 1$. For all $n = 1, \ldots, N$ and all $u \in C^{k+2}(\bar{I}_n, B)$, the following bound holds:*

$$\|\partial_t u - \partial_t R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} \lesssim \tau_n^{k+1}|u|_{C^{k+2}(\bar{I}_n,B)}. \tag{42}$$

Moreover, the bound $\|\partial_t R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} \lesssim \|u\|_{C^1(\bar{I}_n,B)}$ holds for all $u \in C^1(\bar{I}_n, B)$.

Proof. Let $L_\tau^{k+1}u$ be the Lagrange interpolate of u in $\mathbb{P}_{k+1}(\bar{I}_n, B)$ based on the $(k+1)$ right-sided Gauss–Radau integration points on each $\bar{I}_n$ and the left endpoint t_{n-1}. Then,

$$\begin{aligned}\|\partial_t u - \partial_t R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} &\leq \|\partial_t u - \partial_t L_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} + \|\partial_t L_\tau^{k+1}u - \partial_t R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} \\ &\lesssim \tau_n^{k+1}|u|_{C^{k+2}(\bar{I}_n,B)} + \tau_n^{-1}\|L_\tau^{k+1}u - R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)},\end{aligned}$$

and using the triangle inequality, we infer that

$$\|L_\tau^{k+1}u - R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} \leq \|L_\tau^{k+1}u - u\|_{C^0(\bar{I}_n,B)} + \|u - R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} \lesssim \tau_n^{k+2}|u|_{C^{k+2}(\bar{I}_n,B)},$$

thereby proving (42). The proof of the stability bound is similar upon observing that $\|\partial_t R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} \lesssim \tau_n^{-1}\|R_\tau^{k+1}u - u(t_n)\|_{C^0(\bar{I}_n,B)} \le \tau_n^{-1}\|R_\tau^{k+1}u - u\|_{C^0(\bar{I}_n,B)} + \tau_n^{-1}\|u - u(t_n)\|_{C^0(\bar{I}_n,B)}$. The first term on the right-hand side is bounded by $\|u\|_{C^1(\bar{I}_n,B)}$ (see the proof of Lemma 6 in the Appendix using a constant Taylor polynomial) and the second one by $|u|_{C^1(\bar{I}_n,B)}$. This completes the proof. □

5 L^2-norm error estimates

This section is devoted to the $L^\infty(L^2)$ and $L^2(L^2)$ error estimates, first for static and then for time-varying meshes. Our main goal is to estimate the error defined as

$$\tilde{e}(t) := u(t) - \mathcal{L}_\tau u_{\tau h}(t) \qquad \forall\, t \in \bar{I}. \tag{43}$$

We observe that the error is evaluated using the post-processed solution $\mathcal{L}_\tau u_{\tau h}$ and that $\tilde{e}$ is continuous in time on $\bar{I}$ and, moreover, $\tilde{e} \in \tilde{X}$, see (34), if we assume for our analysis that the exact solution u has at least the regularity $u \in C^0(\bar{I}, W) \cap C^1(\bar{I}, L)$. Throughout this section, we assume $k \ge 1$. The case $k = 0$ corresponding to the implicit Euler scheme is briefly discussed in Remark 15. We start with the following consistency result.

Lemma 8 (Consistency) *Assume $u \in C^0(\bar{I}, W) \cap C^1(\bar{I}, L)$. For all $n = 1, \dots, N$, the following equality holds:*

$$\widetilde{B}_h^n(\tilde{e}, v_{\tau h}) = 0 \qquad \forall\, v_{\tau h} \in X_{\tau h}^k. \tag{44}$$

Proof. We recall from Lemma 3 that, for all $n = 1, \dots, N$ and all $v_{\tau h} \in \mathbb{P}_k(I_n, V_h)$, $\widetilde{B}_h^n(\mathcal{L}_\tau u_{\tau h}, v_{\tau h}) = Q_n((f, v_{\tau h})_L)$. Moreover, since the exact solution satisfies $\partial_t u(t_{n,\mu}) + Au(t_{n,\mu}) = f(t_{n,\mu})$ for all $\mu = 1, \dots, k+1$, we infer using the consistency of A_h, see (20), that

$$\widetilde{B}_h^n(u, v_{\tau h}) = Q_n((\partial_t u + A_h u, v_{\tau h})_L) = Q_n((\partial_t u + Au, v_{\tau h})_L) = Q_n((f, v_{\tau h})_L)\,,$$

whence the assertion follows. □

5.1 Static meshes

In the case of static meshes, we drop the superscript n on the mesh $\mathcal{T}_h$, the finite element space V_h, and the discrete differential operator A_h. Our analysis hinges on the following error decomposition:

$$\tilde{e}(t) = \underbrace{(u(t) - P_h R_\tau^{k+1} u(t))}_{=:\eta(t)} + \underbrace{(P_h R_\tau^{k+1} u(t) - \mathcal{L}_\tau u_{\tau h}(t))}_{=\tilde{e}_{\tau h}(t)} \qquad \forall\, t \in \bar{I}, \tag{45}$$

observing that both η and $\tilde{e}_{\tau h}$ are continuous in time on $\bar{I}$. The function η is referred to as the interpolation error. Note that both η as well as $\tilde{e}_{\tau h}$ are in the space $\tilde{X}$, see (34), so that they can be used as arguments in the bilinear form $\widetilde{B}_h^n$.

Lemma 9 (Boundedness) *For all $n = 1, \dots, N$, the following bound holds:*

$$|\widetilde{B}_h^n(\eta, v_{\tau h})| \lesssim \left\{(E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2\right\}^{1/2} \left\{Q_n\big(|\!|\!| v_{\tau h} |\!|\!|^2\big)\right\}^{1/2}, \tag{46}$$

with the time and space errors respectively given by

$$(E_n^{\mathrm{T}}(u))^2 = Q_n\Big(\|\partial_t(u - I_\tau^{k+2}u)\|_L^2 + \|u - R_\tau^{k+1}u\|_V^2\Big), \tag{47}$$

$$(E_n^{\mathrm{S}}(u))^2 = Q_n\Big(|\!|\!| R_\tau^{k+1}u - P_h R_\tau^{k+1}u |\!|\!|_{h,\frac{1}{2}}^2\Big). \tag{48}$$

Proof. We decompose $\widetilde{B}_h^n(\eta, v_{\tau h})$ as

$$\begin{aligned}\widetilde{B}_h^n(\eta, v_{\tau h}) &= Q_n\Big(\partial_t(u - P_h R_\tau^{k+1}u), v_{\tau h})_L\Big) + Q_n\Big((A_h(u - R_\tau^{k+1}u), v_{\tau h})_L\Big) \\ &\quad + Q_n\Big((A_h(R_\tau^{k+1}u - P_h R_\tau^{k+1}u), v_{\tau h})_L\Big) =: T_1 + T_2 + T_3.\end{aligned}$$

Concerning T_1, we can drop the projection P_h and use the property (39) of R_τ^{k+1} followed by the Cauchy–Schwarz inequality to infer that

$$|T_1| \le \Big\{Q_n\Big(\|\partial_t(u - I_\tau^{k+2}u)\|_L^2\Big)\Big\}^{1/2} \left\{Q_n\big(\|v_{\tau h}\|_L^2\big)\right\}^{1/2}.$$

Concerning T_2, since $(u - R_\tau^{k+1}u)$ is in W, the consistency of the discrete operator A_h together with the Cauchy–Schwarz inequality and the boundedness of A lead to

$$\begin{aligned}|T_2| &= \Big|Q_n\Big((A(u - R_\tau^{k+1}u), v_{\tau h})_L\Big)\Big| \\ &\lesssim \Big\{Q_n\Big(\|u - R_\tau^{k+1}u\|_V^2\Big)\Big\}^{1/2} \left\{Q_n\big(\|v_{\tau h}\|_L^2\big)\right\}^{1/2}.\end{aligned}$$

Concerning T_3, we use boundedness on orthogonal subscales, see (22), to infer that

$$|T_3| \lesssim \Big\{Q_n\Big(|\!|\!| R_\tau^{k+1}u - P_h R_\tau^{k+1}u |\!|\!|_{h,\frac{1}{2}}^2\Big)\Big\}^{1/2} \left\{Q_n\big(|\!|\!| v_{\tau h} |\!|\!|^2\big)\right\}^{1/2}.$$

Collecting the above bounds yields the assertion since $\|\cdot\|_L \lesssim |\!|\!|\cdot|\!|\!|$, see (23). □

Lemma 10 (Estimates on $\tilde{e}_{\tau h}$) *For all $m = 1, \dots, N$, the following bound holds:*

$$\|\tilde{e}_{\tau h}(t_m)\|_L^2 + \sum_{n=1}^{m} Q_n\big(|\!|\!| \tilde{e}_{\tau h} |\!|\!|^2\big) \lesssim (E_0)^2 + \sum_{n=1}^{m} \{(E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2\}, \tag{49}$$

with initial error $E_0 = \|P_h u_0 - u_{0h}\|_L$. Moreover, assuming $\tau_n \lesssim \tau_{n-1}$ for all $n = 2, \dots, N$ and $\tau_1 \lesssim 1$, the following bound holds:

$$\|\tilde{e}_{\tau h}\|_{L^2(I,L)}^2 \lesssim (E_0)^2 + \sum_{n=1}^{N} \{(E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2\}. \tag{50}$$

Proof. Owing to Lemma 8 (consistency) and the error decomposition (45), we infer that $\widetilde{B}_h^n(\tilde{e}_{\tau h}, v_{\tau h}) = -\widetilde{B}_h^n(\eta, v_{\tau h})$, so that using Lemma 9 leads to

$$\widetilde{B}_h^n(\tilde{e}_{\tau h}, v_{\tau h}) \lesssim \left\{(E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2\right\}^{1/2} \left\{Q_n\big(|\!|\!| v_{\tau h} |\!|\!|^2\big)\right\}^{1/2}.$$

Setting $v_{\tau h} = I_\tau^{\mathrm{GR}} \tilde{e}_{\tau h}$ and using the stability property of $\widetilde{B}_h^n$ stated in Lemma 4 for $w = \tilde{e}_{\tau h}$ together with the continuity in time of $\tilde{e}_{\tau h}$ and a Young inequality, we infer that

$$\frac{1}{2}\|\tilde{e}_{\tau h}(t_n)\|_L^2 - \frac{1}{2}\|\tilde{e}_{\tau h}(t_{n-1})\|_L^2 + Q_n\big(|\!|\!| \tilde{e}_{\tau h} |\!|\!|^2\big) \lesssim (E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2,$$

where we have dropped the nonnegative jump term at t_{n-1} from the stability property. Taking an arbitrary $m = 1, \ldots, N$ and summing the above inequality from $n = 1$ to m leads to (49) since $\tilde{e}_{\tau h}(0) = P_h u_0 - u_{0h}$. To prove (50), we start with the estimate

$$\|w_\tau\|_{L^2(I_n,L)}^2 \lesssim Q_n\big(\|w_\tau\|_L^2\big) + \tau_n \|w_\tau(t_{n-1}^+)\|_L^2 \qquad \forall\, w_\tau \in \mathbb{P}_{k+1}(I_n, L),$$

which follows by transformation from $\bar{I}_n$ to the reference element $\hat{I} = [-1, 1]$ and application of a norm equivalence on the time polynomial space $\mathbb{P}_{k+1}(\hat{I}, L)$. Applying this inequality to $\tilde{e}_{\tau h}$ which is continuous in time leads to

$$\|\tilde{e}_{\tau h}\|_{L^2(I_n,L)}^2 \lesssim Q_n\big(\|\tilde{e}_{\tau h}\|_L^2\big) + \tau_n \|\tilde{e}_{\tau h}(t_{n-1})\|_L^2.$$

Summing this bound from $n = 1$ to N and observing that, for all $n \geq 2$, $\tau_n \|\tilde{e}_{\tau h}(t_{n-1})\|_L^2 \lesssim \tau_{n-1}\|\tilde{e}_{\tau h}(t_{n-1})\|_L^2 \lesssim Q_{n-1}\big(\|\tilde{e}_{\tau h}\|_L^2\big)$ since $\tau_n \lesssim \tau_{n-1}$ and $t_{n-1,k+1} = t_{n-1}$, we infer that

$$\|\tilde{e}_{\tau h}\|_{L^2(I,L)}^2 \lesssim \sum_{n=1}^{N} Q_n\big(\|\tilde{e}_{\tau h}\|_L^2\big) + \tau_1 \|\tilde{e}_{\tau h}(0)\|_L^2,$$

whence (50) follows since $\tau_1 \lesssim 1$. □

Theorem 11 (L^2-error estimate) *Let u be the exact solution and let $u_{\tau h}$ be the fully discrete solution. Assume $k \geq 1$ and $\tau_n \lesssim 1$ for all $n = 1, \ldots, N$. Assume **(H)** in the case of conforming FEM. Then, for the error $\tilde{e}(t)$ defined in (43), the following bound holds for all $m = 1, \ldots, N$,*

$$\|\tilde{e}(t_m)\|_L^2 \lesssim (E_0)^2 + t_m \max_{1 \leq n \leq m} \left\{C_n^{\mathrm{T}}(u)\tau_n^{2(k+2)} + C_n^{\mathrm{S}}(u) h^{2r+1}\right\} + C'_m(u) h^{2(r+1)}, \tag{51}$$

with $C_n^{\mathrm{T}}(u) := |u|_{C^{k+3}(\bar{I}_n,L)}^2 + |u|_{C^{k+2}(\bar{I}_n,V)}^2$, $C_n^{\mathrm{S}}(u) := \|u\|_{C^1(\bar{I}_n,H^{r+1}(\Omega))}^2$, and $C'_m(u) = |u(t_m)|_{H^{r+1}(\Omega)}^2$. Moreover, assuming $\tau_n \lesssim \tau_{n-1}$ for all $n = 2, \ldots, N$, the following bound holds:

$$\|\tilde{e}\|_{L^2(I,L)}^2 \lesssim (E_0)^2 + T \max_{1 \leq n \leq N} \left\{C_n^{\mathrm{T}}(u)\tau_n^{2(k+2)} + C_n^{\mathrm{S}}(u) h^{2r+1}\right\}. \tag{52}$$

Proof. From definitions (47)-(48) of $E_n^{\mathrm{T}}(u)$ and $E_n^{\mathrm{S}}(u)$ for all $n = 1, \ldots, N$, we infer that

$$(E_n^{\mathrm{T}}(u))^2 \lesssim \tau_n \left(\tau_n^{2(k+2)} |u|_{C^{k+3}(\bar{I}_n,L)}^2 + \tau_n^{2(k+2)} |u|_{C^{k+2}(\bar{I}_n,V)}^2\right),$$
$$(E_n^{\mathrm{S}}(u))^2 \lesssim \tau_n h^{2r+1} \|u\|_{C^1(\bar{I}_n,H^{r+1}(\Omega))}^2,$$

where the bound on $E_n^{\mathrm{T}}(u)$ results from (37) with $B = L$ and (41) with $B = V$, and that on $E_n^{\mathrm{S}}(u)$ from the approximation property (29) of P_h combined with the stability of R_τ^{k+1} from Lemma 6 with $B = H^{r+1}(\Omega)$ and the assumption $\tau_n \lesssim 1$. Moreover, recalling the error decomposition (45), we observe that, for all $n = 1, \dots, N$,

$$\|\eta(t_n)\|_L = \|u(t_n) - P_h R_\tau^{k+1} u(t_n)\|_L = \|u(t_n) - P_h u(t_n)\|_L \lesssim h^{r+1} |u(t_n)|_{H^{r+1}(\Omega)},$$

where we used Lemma 5, and

$$\begin{aligned}
\|\eta\|_{L^2(I_n,L)}^2 &= \|u - P_h R_\tau^{k+1} u\|_{L^2(I_n,L)}^2 \le 2\|u - P_h u\|_{L^2(I_n,L)}^2 + 2\|P_h(u - R_\tau^{k+1} u)\|_{L^2(I_n,L)}^2 \\
&\le 2\|u - P_h u\|_{L^2(I_n,L)}^2 + 2\|u - R_\tau^{k+1} u\|_{L^2(I_n,L)}^2 \\
&\lesssim \tau_n \left(h^{2(r+1)} \|u\|_{C^0(\bar{I}_n, H^{r+1}(\Omega))}^2 + \tau_n^{2(k+2)} |u|_{C^{k+2}(\bar{I}_n,L)}^2 \right).
\end{aligned}$$

We conclude using Lemma 10 and the triangle inequality, as well as $h \le \mathrm{diam}(\Omega) \lesssim 1$ and $\|\cdot\|_L \lesssim \|\cdot\|_V$ for (52). □

Remark 12 (Assumption on the time steps) *The assumption $\tau_n \lesssim \tau_{n-1}$ is quite mild; it means that the time step can be increased at most by a uniformly bounded factor. The assumption $\tau_n \lesssim 1$ is also quite mild; it means that the time steps resolve the fastest time scale present in the governing equations which is here given by $\min(\mu_0, L_\beta)$ with μ_0 from (3) and L_β the Lipschitz constant of β.*

Remark 13 (Initial error) *The initial error E_0 vanishes when the discrete initial condition is chosen to be $u_{0h} = P_h u_0$. Otherwise, this error is typically of order h^{r+1} if u_0 is smooth enough.*

Remark 14 (Estimate on the error $(u - u_{\tau h})$) *Since $\tilde{e}(t_m) = u(t_m) - \mathcal{L}_\tau u_{\tau h}(t_m) = u(t_m) - u_{\tau h}(t_m)$ for all $m = 1, \dots, N$, the right-hand side of (51) also bounds the error $\|u(t_m) - u_{\tau h}(t_m)\|_L$, showing that a superconvergent L^2-error estimate of order $(\tau^{k+2} + h^{r+1/2})$ also holds under the assumptions of Theorem 11 for the original fully discrete solution $u_{\tau h}$ at the discrete nodes defining the time partition. Instead, the bound on $\|u - u_{\tau h}\|_{L^2(I,L)}$ is of order $(\tau^{k+1} + h^{r+1/2})$ which is optimal with respect to τ. This bound can be derived from*

$$\|I_\tau^{\mathrm{GR}} u - u_{\tau h}\|_{L^2(I_n,L)}^2 = Q_n\big(\|I_\tau^{\mathrm{GR}} u - u_{\tau h}\|_L^2\big) = Q_n\big(\|\tilde{e}\|_L^2\big) \lesssim Q_n\big(\|\tilde{e}_{\tau h}\|_L^2\big) + Q_n\big(\|\eta\|_L^2\big),$$

where $\tilde{e}$, $\tilde{e}_{\tau h}$, and η are defined in (45). The sum over all $n = 1, \dots, N$ yields as in the previous analysis the same upper bound as the right-hand side of (52). Then, invoking the triangle inequality $\|u - u_{\tau h}\|_{L^2(I,L)} \le \|u - I_\tau^{\mathrm{GR}} u\|_{L^2(I,L)} + \|I_\tau^{\mathrm{GR}} u - u_{\tau h}\|_{L^2(I,L)}$ yields the claim. Finally, since $[u_{\tau h}]_{n-1} \vartheta_n(t) = u_{\tau h}(t) - \mathcal{L}_\tau u_{\tau h}(t)$ for all $t \in I_n$, we infer by means of the triangle inequality the same bound for the jump seminorm $(\sum_{n=1}^N \|[u_{\tau h}]_{n-1}\|_L^2)^{1/2}$.

Remark 15 (Implicit Euler) *For $k = 0$, the interpolate $R_\tau^{k+1} u$ is not available since the construction of Section 4.2 requires $k \ge 1$. The analysis proceeds by replacing in the above proofs $R_\tau^{k+1} u$ by the piecewise affine Lagrange interpolate of u in time and leads to an error bound of order $(\tau + h^{r+1/2})$ for the error $\tilde{e}$ in the L norm at the discrete times defining the time partition and in the $L^2(I, L)$ norm.*

5.2 Time-varying meshes

Now, we allow that, on each time interval $I_n = (t_{n-1}, t_n]$, we can have a new mesh $\mathcal{T}_h^n$ which can be created from the previous mesh $\mathcal{T}_h^{n-1}$ by means of local refinements and derefinements. Therefore, it is necessary to use the superscript n also for the finite element space V_h^n and the discrete differential operator A_h^n. By h_n we denote the maximum of all diameters h_K of the mesh cells $K \in \mathcal{T}_h^n$. For each $n = 0, 1, \ldots, N$, let $P_h^n : L \to V_h^n$ denote the L^2-projector onto V_h^n. For a time-dependent function $w : \bar{I} \to L$, we define its space projection $P_h w : \bar{I} \to L$ as

$$(P_h w)(t) := \begin{cases} P_h^n\big(w(t)\big) & \text{if} \quad t \in I_n = (t_{n-1}, t_n], \\ P_h^0\big(w(0)\big) & \text{if} \quad t = 0. \end{cases} \tag{53}$$

Note that, even for a continuous function $w \in C^0(\bar{I}, L)$, its space projection $P_h w$ can be discontinuous in time at the discrete points t_{n-1}, i.e., $P_h w \in C^{-1}(\mathcal{M}_\tau, L)$ with the jump

$$[P_h w]_{n-1} = P_h^n w(t_{n-1}^+) - P_h^{n-1} w(t_{n-1}) \in V_h^n + V_h^{n-1}, \tag{54}$$

which is in general non-zero if $V_h^n \neq V_h^{n-1}$.

5.2.1 Error estimates with projection error

Recall the error $\tilde{e}(t)$ defined in (43). In the case of time-varying meshes, the decomposition (45) of the error $\tilde{e}(t)$ has to be modified as

$$\tilde{e}(t) = \underbrace{\big(u(t) - \mathcal{L}_\tau P_h R_\tau^{k+1} u(t)\big)}_{=:\eta(t)} + \underbrace{\big(\mathcal{L}_\tau P_h R_\tau^{k+1} u(t) - \mathcal{L}_\tau u_{\tau h}(t)\big)}_{=\tilde{e}_{\tau h}(t)} \qquad \forall\, t \in \bar{I}, \tag{55}$$

where the use of the lifting operator in the definition of η allows us to recover a continuous function in time. We can write $\eta(t)$ for $t \in I_n$ as

$$\eta(t) = \underbrace{\big(u(t) - P_h^n R_\tau^{k+1} u(t)\big)}_{=:\eta^{\mathrm{old}}(t)} + [P_h u]_{n-1} \vartheta_n(t),$$

where we have used that $R_\tau^{k+1} u(t_{n-1}) = u(t_{n-1})$ and $\eta^{\mathrm{old}}(t)$ denotes the interpolation error used for the L^2-analysis in the case of static meshes. The part $[P_h u]_{n-1}$ leads to an extra term in the error analysis for time-varying meshes and can be regarded as the *projection error* for the time interval I_n. Note that again η and $\tilde{e}_{\tau h}$ are contained in the space $\tilde{X}$ defined in (34). Let $\Pi_h^{n-1} : V_h^{n-1} + V_h^n \to V_h^{n-1}$ denote an L^2-stable, linear quasi-interpolation operator satisfying the following properties:

$$\Pi_h^{n-1} v_h = v_h \qquad \forall\, v_h \in V_h^{n-1}, \tag{56}$$

$$\|\Pi_h^{n-1} v_h\|_L \lesssim \|v_h\|_L \qquad \forall\, v_h \in V_h^{n-1} + V_h^n. \tag{57}$$

Lemma 16 (Boundedness) *For all $n = 1, \ldots, N$, the following bound holds:*

$$|\widetilde{B}_h^n(\eta, v_{\tau h})| \lesssim \big\{(E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2\big\}^{1/2} \big\{Q_n\big(|\!|\!| v_{\tau h} |\!|\!|_n^2\big)\big\}^{1/2} + E_n^{\mathrm{P}}(u) \|[v_{\tau h}]_{n-1}\|_L, \tag{58}$$

with $E_n^{\mathrm{T}}(u)$ *given by* (47), $(E_n^{\mathrm{S}}(u))^2 := Q_n\Big(\|R_\tau^{k+1}u - P_h^n R_\tau^{k+1}u\|^2_{h,\frac{1}{2},n}\Big)$, *and the local projection error defined by*

$$E_n^{\mathrm{P}}(u) := \sup_{v_h \in V_h^n} \frac{\big(u(t_{n-1}) - P_h^{n-1}u(t_{n-1}), v_h - \Pi_h^{n-1}v_h\big)_L}{\|v_h - \Pi_h^{n-1}v_h\|_L}, \tag{59}$$

with the convention that the ratio is zero if $v_h \in V_h^{n-1}$, *which means, in particular, that* $E_n^{\mathrm{P}}(u) = 0$ *in the case that* $V_h^n \subset V_h^{n-1}$.

Proof. Since ϑ_n vanishes at the $(k+1)$ right-sided Gauss–Radau integration points, we can decompose $\widetilde{B}_h^n(\eta, v_{\tau h})$ as

$$\begin{aligned}\widetilde{B}_h^n(\eta, v_{\tau h}) &= \widetilde{B}_h^n(\eta^{\mathrm{old}}, v_{\tau h}) + \widetilde{B}_h^n([P_h u]_{n-1}\vartheta_n, v_{\tau h}) \\ &= (T_1 + T_2 + T_3) + \underbrace{Q_n\big(([P_h u]_{n-1}\vartheta_n', v_{\tau h})_L\big)}_{=:T_4},\end{aligned}$$

where $T_{1,2,3}$ denote the terms introduced in the proof of Lemma 9. Concerning T_4, we exploit that the integrand is a polynomial in $\mathbb{P}_{2k}(I_n, \mathbb{R})$ to infer that

$$\begin{aligned}T_4 &= \int_{I_n} \big([P_h u]_{n-1}\vartheta_n', v_{\tau h}\big)_L \, dt = -\int_{I_n} ([P_h u]_{n-1}\vartheta_n, \partial_t v_{\tau h})_L \, dt - \big([P_h u]_{n-1}, v_{\tau h}(t_{n-1}^+)\big)_L \\ &= \big([u - P_h u]_{n-1}, v_{\tau h}(t_{n-1}^+)\big)_L \\ &= -\big(u(t_{n-1}) - P_h^{n-1}u(t_{n-1}), v_{\tau h}(t_{n-1}^+)\big)_L \\ &= -\big(u(t_{n-1}) - P_h^{n-1}u(t_{n-1}), [v_{\tau h}]_{n-1}\big)_L\end{aligned}$$

using that $\vartheta_n(t_{n,\mu}) = 0$ for $\mu = 1, \dots, k+1$, $\vartheta_n(t_{n-1}^+) = 1$, $[u]_{n-1} = 0$, $v_{\tau h}(t_{n-1}^+) \in V_h^n$, and $v_{\tau h}(t_{n-1}) \in V_h^{n-1}$. We use the notation $v_h^+ := v_{\tau h}(t_{n-1}^+) \in V_h^n$ and $v_h^{n-1} := v_{\tau h}(t_{n-1}) \in V_h^{n-1}$. If v_h^+ turns out to be in V_h^{n-1}, then $T_4 = 0$ which obviously satisfies $|T_4| \le E_n^{\mathrm{P}}(u)\|[v_{\tau h}]_{n-1}\|_L$. Otherwise, $v_h^+ - \Pi_h^{n-1}v_h^+ \neq 0$ and since, owing to (56),

$$[v_{\tau h}]_{n-1} - \Pi_h^{n-1}[v_{\tau h}]_{n-1} = (v_h^+ - \Pi_h^{n-1}v_h^+) - (v_h^{n-1} - \Pi_h^{n-1}v_h^{n-1}) = v_h^+ - \Pi_h^{n-1}v_h^+,$$

we infer that

$$\begin{aligned}|T_4| &= \big|\big(u(t_{n-1}) - P_h^{n-1}u(t_{n-1}), v_h^+ - \Pi_h^{n-1}v_h^+\big)_L\big| \\ &= \frac{\big|\big(u(t_{n-1}) - P_h^{n-1}u(t_{n-1}), v_h^+ - \Pi_h^{n-1}v_h^+\big)_L\big|}{\|v_h^+ - \Pi_h^{n-1}v_h^+\|_L}\|[v_{\tau h}]_{n-1} - \Pi_h^{n-1}[v_{\tau h}]_{n-1}\|_L \\ &\lesssim E_n^{\mathrm{P}}(u)\|[v_{\tau h}]_{n-1}\|_L,\end{aligned}$$

where we used (57) in the last bound. Collecting the bounds for $T_{1,2,3}$ from the proof of Lemma 9 and the above bound for T_4 yields the assertion. □

Theorem 17 (L^2-error estimate) *Let* u *be the exact solution and let* $u_{\tau h} \in X_{\tau h}^k$ *be the fully discrete solution. Assume* $k \ge 1$ *and* $\tau_n \lesssim 1$ *for all* $n = 1, \dots, N$. *Assume* ***(H)*** *in the*

case of conforming FEM. Then, for the error $\tilde{e}(t)$ defined in (43), the following bound holds for all $m = 1, \ldots, N$,

$$\|\tilde{e}(t_m)\|_L^2 \lesssim (E_0)^2 + t_m \max_{1\le n\le m} \left\{ C_n^{\mathrm{T}}(u)\tau_n^{2(k+2)} + C_n^{\mathrm{S}}(u)h_n^{2r+1} \right\} + C'_m(u)h_m^{2(r+1)} + (E_{P,m}(u))^2, \tag{60}$$

with $C_n^{\mathrm{T}}(u)$, $C_n^{\mathrm{S}}(u)$ and $C'_m(u)$ defined in Theorem 11 and $E_0 := \|P_h^0 u_0 - u_{0h}\|_L$, while the cumulated projection error $E_{P,m}(u)$ is defined as

$$(E_{P,m}(u))^2 := \sum_{n=1}^{m} (E_n^{\mathrm{P}}(u))^2 \qquad \text{with } E_n^{\mathrm{P}}(u) \text{ defined in (59).} \tag{61}$$

Moreover, assuming $\tau_n \lesssim \tau_{n-1}$ for all $n = 2, \ldots, N$, the following bound holds:

$$\|\tilde{e}\|_{L^2(I,L)}^2 \lesssim (E_0)^2 + T \max_{1\le n\le N} \left\{ C_n^{\mathrm{T}}(u)\tau_n^{2(k+2)} + C_n^{\mathrm{S}}(u)h_n^{2r+1} \right\} + (E_{P,N}(u))^2. \tag{62}$$

Proof. Applying Lemma 8, we again infer that $\widetilde{B}_h^n(\tilde{e}_{\tau h}, v_{\tau h}) = -\widetilde{B}_h^n(\eta, v_{\tau h})$, so that Lemma 16 leads to

$$\widetilde{B}_h^n(\tilde{e}_{\tau h}, v_{\tau h}) \lesssim \left\{ (E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2 \right\}^{1/2} \left\{ Q_n\big(|\!|\!| v_{\tau h} |\!|\!|_n^2\big) \right\}^{1/2} + E_n^{\mathrm{P}}(u) \|[v_{\tau h}]_{n-1}\|_L.$$

Since $\tilde{e}_{\tau h}(t_{n,\mu}) \in V_h^n$ for all right-sided Gauss–Radau integration points $t_{n,\mu} \in I_n$, the function $w = \tilde{e}_{\tau h}$ satisfies the asumption $w \in C^0(\bar{I}, L) \cap X_\tau^{k+1}(L) \cap \tilde{X}$ in Lemma 4 (stability property of $\widetilde{B}_h^n$) so that by setting $v_{\tau h} = I_\tau^{\mathrm{GR}} \tilde{e}_{\tau h} \in X_{\tau h}^k$, we infer the lower bound

$$\widetilde{B}_h^n(\tilde{e}_{\tau h}, v_{\tau h}) \ge \frac{1}{2}\|\tilde{e}_{\tau h}(t_n)\|_L^2 - \frac{1}{2}\|\tilde{e}_{\tau h}(t_{n-1})\|_L^2 + \frac{1}{2}\|[v_{\tau h}]_{n-1}\|_L^2 + \alpha Q_n\big(|\!|\!| \tilde{e}_{\tau h} |\!|\!|_n^2\big).$$

The rest of the proof is similar to that of Theorem 11 and is skipped for brevity. □

5.2.2 Bound on the projection error

Our goal is now to derive estimates on the cumulated projection error $E_{P,m}(u)$ for all $m = 1, \ldots, N$. Obviously, $E_{P,m}(u) = 0$ if $V_h^n \subseteq V_h^{n-1}$ for all $n = 1, \ldots, m$, i.e., if only mesh coarsenings occur from one time interval to the next one. To treat more general situations, we introduce some additional notation.

Let $n = 1, \ldots, N$. We denote by $\mathcal{T}_h^{n,\mathrm{coa}}$ the subset of the coarse mesh cells in $\mathcal{T}_h^n$ which are such that either they are in $\mathcal{T}_h^{n-1}$ or they can be decomposed by means of mesh cells from $\mathcal{T}_h^{n-1}$, and by $\mathcal{T}_h^{n,\mathrm{ref}}$ the subset of the finer mesh cells in $\mathcal{T}_h^n$, i.e.,

$$\mathcal{T}_h^{n,\mathrm{coa}} := \{K \in \mathcal{T}_h^n;\ \exists\, \mathcal{T} \subset \mathcal{T}_h^{n-1} :\ \bar{K} = \bigcup_{T\in\mathcal{T}} \bar{T}\}, \qquad \mathcal{T}_h^{n,\mathrm{ref}} := \mathcal{T}_h^n \setminus \mathcal{T}_h^{n,\mathrm{coa}}. \tag{63}$$

We denote by Ω_n^{ref} the subset of the domain Ω where the mesh $\mathcal{T}_h^n$ is finer than $\mathcal{T}_h^{n-1}$, i.e., $\overline{\Omega_n^{\mathrm{ref}}} := \bigcup_{K\in\mathcal{T}_h^{n,\mathrm{ref}}} \bar{K}$, and we set $h_n^{\mathrm{ref}} := \max_{K\in\mathcal{T}_h^{n,\mathrm{ref}}} h_K$, while $|\Omega_n^{\mathrm{ref}}|$ denotes the d-dimensional measure of Ω_n^{ref}.

We assume that the quasi-interpolation operator Π_h^{n-1} has the property that

$$\left(v_h - \Pi_h^{n-1} v_h\right)\big|_K = 0 \qquad \forall\, v_h \in V_h^n,\ \forall\, K \in \mathcal{T}_h^{n,\mathrm{coa}}. \tag{64}$$

All the assumptions (56), (57) and (64) are satisfied if we choose for $\Pi_h^{n-1} v_h$ the standard Lagrange finite element interpolate of $v_h \in V_h^{n-1} + V_h^n$ defined on the mesh $\mathcal{T}_h^{n-1}$ in the case of conforming FEM or the L^2-orthogonal projection of v_h onto the space V_h^{n-1} in the case of DG methods where piecewise polynomial spaces are used. We first estimate the local projection error $E_n^{\mathrm{P}}(u)$ defined in (59).

Lemma 18 (Local projection error) *The following bounds hold: In the case of DG spaces,*

$$E_n^{\mathrm{P}}(u) \lesssim |\Omega_n^{ref}|^{1/2} (h_n^{ref})^{1/2} \left\{ (h_n^{ref})^{r+1/2} C_n^{\mathrm{P,S}}(u) \right\}, \tag{65}$$

with $C_n^{\mathrm{P,S}}(u) := |u(t_{n-1})|_{W^{r+1,\infty}(\Omega_n^{ref})}$*, and in the case of conforming FEM, under assumption* ***(H)****,*

$$E_n^{\mathrm{P}}(u) \lesssim (h_n^{ref})^{1/2} \left\{ (h_n)^{r+1/2} C_n^{\mathrm{P,S}}(u) \right\}, \tag{66}$$

with $C_n^{\mathrm{P,S}}(u) := |u(t_{n-1})|_{H^{r+1}(\Omega)}$.

Proof. Owing to (64), we infer that $\|v_h - \Pi_h^{n-1} v_h\|_{L^2(\Omega \setminus \Omega_n^{\mathrm{ref}})} = 0$ for all $v_h \in V_h^n$. Hence,

$$E_n^{\mathrm{P}}(u) \le \|u(t_{n-1}) - P_h^{n-1} u(t_{n-1})\|_{L^2(\Omega_n^{\mathrm{ref}})}.$$

In the case of DG methods, we use the bound

$$\|u(t_{n-1}) - P_h^{n-1} u(t_{n-1})\|_{L^2(\Omega_n^{\mathrm{ref}})} \le |\Omega_n^{\mathrm{ref}}|^{1/2} \|u(t_{n-1}) - P_h^{n-1} u(t_{n-1})\|_{L^\infty(\Omega_n^{\mathrm{ref}})},$$

and the local approximation properties of the L^2-orthogonal projection in $W^{r+1,\infty}(\Omega_n^{\mathrm{ref}})$ to infer (65). In the case of conforming FEM, we observe that

$$\begin{aligned} E_n^{\mathrm{P}}(u) &\le (h_n^{\mathrm{ref}})^{1/2} \|\mathfrak{h}^{-1/2} (u(t_{n-1}) - P_h^{n-1} u(t_{n-1}))\|_{L^2(\Omega_n^{\mathrm{ref}})} \\ &\le (h_n^{\mathrm{ref}})^{1/2} \|\mathfrak{h}^{-1/2} (u(t_{n-1}) - P_h^{n-1} u(t_{n-1}))\|_{L^2(\Omega)}, \end{aligned}$$

and proceed as in the proof of Lemma 1 to conclude. □

Remark 19 (Comparison of (65) **and** (66)**)** *In both estimates, the term between braces has the same order as the other terms stemming from space errors. Note, however, that for DG methods, the constant* $C_n^{\mathrm{P,S}}(u)$ *and the mesh size depend on* Ω_n^{ref} *only, while this dependency concerns the whole domain* Ω *for conforming FEM. Furthermore, the bound on the local projection error in the case of DG spaces contains the additional factor* $|\Omega_n^{ref}|^{1/2}$*. In practical refinement regimes, a reasonable assumption is that*

$$\max_{n \in \mathcal{N}_m^{ref}} |\Omega_n^{ref}| \lesssim h_n^{ref}, \tag{67}$$

which is mostly due to the fact that singularities or critical parts of the exact solution are typically located in $(d-1)$-dimensional manifolds that are covered by a grid part with d-dimensional measure of order h_n^{ref}. In this situation, the bound on $E_n^{\mathrm{P}}(u)$ for DG methods is improved by a factor $(h_n^{ref})^{1/2}$ with respect to that for conforming FEM. The price to pay is a slightly more stringent regularity assumption on the exact solution. If assumption (67) *cannot be exploited, both bounds* (65) *and* (66) *exhibit the same asymptotic behavior.*

We now estimate the cumulated projection error. We define the index set $\mathcal{N}_m^{\mathrm{ref}}$ of those $n \le m$, where V_h^n contains some refinement with respect to V_h^{n-1}, and the number M_m of its elements, i.e.,

$$\mathcal{N}_m^{\mathrm{ref}} := \{n \in \{1,\dots,m\};\quad V_h^n \not\subseteq V_h^{n-1}\}, \qquad M_m := \mathrm{card}(\mathcal{N}_m^{\mathrm{ref}}). \tag{68}$$

A straightforward consequence of Lemma 18 is the following result.

Corollary 20 (Cumulated projection error) *The following bounds hold for all $m = 1,\dots,N$: In the case of DG methods,*

$$|E_{P,m}(u)| \lesssim M_m^{1/2} \max_{n\in\mathcal{N}_m^{ref}} \left\{ |\Omega_n^{ref}|^{1/2} (h_n^{ref})^{1/2} \left\{ (h_n^{ref})^{r+1/2} C_n^{\mathrm{P,S}}(u) \right\} \right\}, \tag{69}$$

*and in the case of conforming FEM under assumption **(H)**,*

$$|E_{P,m}(u)| \lesssim M_m^{1/2} \max_{n\in\mathcal{N}_m^{ref}} \left\{ (h_n^{ref})^{1/2} \left\{ (h_n)^{r+1/2} C_n^{\mathrm{P,S}}(u) \right\} \right\}. \tag{70}$$

Remark 21 (Interpretation of Corollary 20) *Let $\tau_{av,m} := \frac{t_m}{m}$ denote the averaged time step in the interval $[0,t_m]$. Then, owing to the obvious estimate $M_m \le m$, we infer in the case of DG methods that*

$$|E_{P,m}(u)| \lesssim t_m^{1/2} \left(\frac{(h_{(m)}^{ref})^{1+\alpha}}{\tau_{av,m}} \right)^{1/2} \max_{n\in\mathcal{N}_m^{ref}} \left\{ (h_n^{ref})^{r+1/2} C_n^{\mathrm{P,S}}(u) \right\}, \tag{71}$$

where $h_{(m)}^{ref} := \max_{n\in\mathcal{N}_m^{ref}} h_n^{ref}$ and $\alpha = 1$ if assumption (67) *is valid, while $\alpha = 0$ otherwise. This means that the projection error due to dynamic grids only causes a weakening of the overall order of accuracy if the averaged time step $\tau_{av,m}$ is strongly less than $(h_{(m)}^{ref})^{1+\alpha}$. This is a mild restriction to practical regimes for controlling the ratio between mesh and time steps. In the case of conforming FEM, the bound* (71) *holds with $\alpha = 0$ and h_n in place of h_n^{ref} in the term between braces.*

6 Estimates on error derivatives

This section is devoted to estimating the time-derivative error in the L-norm and the error in the discrete graph norm. Throughout this section, we consider static meshes for simplicity, so that we drop the superscript n on the mesh $\mathcal{T}_h$, the finite element space V_h, the discrete differential operator A_h, and the L^2-orthogonal projector P_h onto V_h. In what follows, we assume that the exact solution u has the regularity $u \in C^1(\bar{I},W) \cap C^2(\bar{I},L)$ (so that $f \in C^1(\bar{I},L)$). Moreover, we assume $k \ge 1$ as in Section 5.

6.1 Time-derivative error estimates

Our main idea is to derive a time-derivative error estimate by comparing the time-derivative of the exact solution with the post-processed time-derivative of the post-processed discrete solution. Since $\mathcal{L}_\tau u_{\tau h} \in X_\tau^{k+1}(V_h) \cap C^0(\bar{I}, V_h)$, the time-derivative of $\mathcal{L}_\tau u_{\tau h}$ is well-defined in the interior intervals (t_{n-1}, t_n) and can be continuously extended from the left at all t_n by setting

$$\partial_t \mathcal{L}_\tau u_{\tau h}(t_n) := \lim_{t \uparrow t_n} \partial_t \mathcal{L}_\tau u_{\tau h}(t) \qquad \forall\, n = 1, \ldots, N. \tag{72}$$

This defines the piecewise polynomial function $\partial_t \mathcal{L}_\tau u_{\tau h}$ over I, and at $t = 0$, we define

$$\partial_t \mathcal{L}_\tau u_{\tau h}(0) := P_h f(0) - A_h u_{\tau h}(0). \tag{73}$$

The function $\partial_t \mathcal{L}_\tau u_{\tau h}$ is now defined over $\bar{I}$, and $\partial_t \mathcal{L}_\tau u_{\tau h} \in X_{\tau h}^k$.

Lemma 22 *The fully discrete solution $u_{\tau h}$ is such that*

$$\partial_t \mathcal{L}_\tau u_{\tau h}(t) + A_h u_{\tau h}(t) = P_h I_\tau^{\mathrm{GR}} f(t) \qquad \forall\, t \in \bar{I}, \tag{74}$$

where the time-interpolate $I_\tau^{\mathrm{GR}} f$ of f is defined as in (12).

Proof. Recalling (32), we obtain, for all $n = 1, \ldots, N$ and all $v_{\tau h} \in \mathbb{P}_k(I_n, V_h)$,

$$\int_{I_n} (\partial_t \mathcal{L}_\tau u_{\tau h} + A_h u_{\tau h}, v_{\tau h})_L \, dt = Q_n((f, v_{\tau h})_L) = \int_{I_n} (I_\tau^{\mathrm{GR}} f, v_{\tau h})_L \, dt = \int_{I_n} (P_h I_\tau^{\mathrm{GR}} f, v_{\tau h})_L \, dt.$$

Hence, (74) holds in the interior of all time intervals. Moreover, (74) holds by definition at $t = 0$, see (73), and at all t_n for all $n = 1, \ldots, N$, owing to (72) since both $A_h u_{\tau h}$ and $P_h I_\tau^{\mathrm{GR}} f$ are continuous from the left at t_n. □

The error on the time-derivative can now be defined as

$$\widehat{e}(t) := \partial_t u(t) - \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h}(t) \qquad \forall\, t \in \bar{I}. \tag{75}$$

We observe that $\widehat{e}$ is continuous in time on $\bar{I}$.

Lemma 23 (Consistency) *Assume $u \in C^1(\bar{I}, W) \cap C^2(\bar{I}, L)$. For all $n = 1, \ldots, N$, the following identity holds:*

$$\widetilde{B}_h^n(\widehat{e}, v_{\tau h}) = Q_n\Big((\partial_t f - \partial_t L_\tau^{k+1} f, v_{\tau h})_L\Big) \qquad \forall\, v_{\tau h} \in X_{\tau h}^k, \tag{76}$$

where $L_\tau^{k+1} f\big|_{\bar{I}_n}$ is the Lagrange interpolate of f in $\mathbb{P}_{k+1}(\bar{I}_n, L)$ based on the $(k+1)$ right-sided Gauss–Radau integration points on each $\bar{I}_n$ and the left endpoint t_{n-1}.

Proof. Since $\partial_t^2 u(t) + A \partial_t u(t) = \partial_t f(t)$ for all $t \in I$, applying this equation at the $(k+1)$ right-sided Gauss–Radau integration points in I_n and using the consistency (20) of the discrete operator A_h, it is inferred that

$$\widetilde{B}_h^n(\partial_t u, v_{\tau h}) = Q_n((\partial_t f, v_{\tau h})_L)\,.$$

Moreover, applying the linear operator $\mathcal{L}_\tau$ to (74) yields

$$\mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h}(t) + A_h \mathcal{L}_\tau u_{\tau h}(t) = \mathcal{L}_\tau P_h I_\tau^{\mathrm{GR}} f(t) \qquad \forall\, t \in \bar{I},$$

where we have used that $\mathcal{L}_\tau A_h u_{\tau h}(t) = A_h \mathcal{L}_\tau u_{\tau h}(t)$. Taking the time-derivative, we infer that, for all $\mu = 1, \ldots, k+1$ and all $n = 1, \ldots, N$,

$$\partial_t \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h}(t_{n,\mu}) + \partial_t A_h \mathcal{L}_\tau u_{\tau h}(t_{n,\mu}) = \partial_t \mathcal{L}_\tau P_h I_\tau^{\mathrm{GR}} f(t_{n,\mu}),$$

where we have taken the limit from the left at $t_{n,k+1} = t_n$. Recalling the definition (33) of the bilinear form $\widetilde{B}_h^n$, we obtain

$$\begin{aligned}
\widetilde{B}_h^n(\mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h}, v_{\tau h}) &= Q_n((\partial_t \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h} + A_h \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L) \\
&= Q_n((\partial_t \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h} + A_h \partial_t \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L) \\
&= Q_n((\partial_t \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h} + \partial_t A_h \mathcal{L}_\tau u_{\tau h}, v_{\tau h})_L) \\
&= Q_n\big((\partial_t \mathcal{L}_\tau P_h I_\tau^{\mathrm{GR}} f, v_{\tau h})_L\big) = Q_n\big((\partial_t \mathcal{L}_\tau I_\tau^{\mathrm{GR}} f, v_{\tau h})_L\big)\,,
\end{aligned}$$

where we have used (18) on the second line to drop the $\mathcal{L}_\tau$ operator after A_h, the fact that A_h is time-independent on the third line, and dropped the projector P_h on the last line since P_h commutes with the operators $\mathcal{L}_\tau$ and ∂_t. To conclude, we observe that $L_\tau^{k+1} f = \mathcal{L}_\tau I_\tau^{\mathrm{GR}} f$ since the restrictions of both functions to I_n are in $\mathbb{P}_{k+1}(I_n, L)$ and coincide at $(k+2)$ distinct points of $\bar{I}_n$. $\square$

Our analysis hinges on the following error decomposition:

$$\widehat{e}(t) = \underbrace{\Big(\partial_t u(t) - \mathcal{L}_\tau P_h \partial_t R_\tau^{k+1} u(t)\Big)}_{=:\widehat{\eta}(t)} + \underbrace{\Big(\mathcal{L}_\tau P_h \partial_t R_\tau^{k+1} u(t) - \mathcal{L}_\tau \partial_t \mathcal{L}_\tau u_{\tau h}(t)\Big)}_{=:\widehat{e}_{\tau h}(t)} \qquad \forall\, t \in \bar{I}, \quad (77)$$

where to define $\mathcal{L}_\tau P_h \partial_t R_\tau^{k+1} u(t)$ on the first time interval, we define $\partial_t R_\tau^{k+1} u(0) := \partial_t u(0)$. We observe that both $\widehat{\eta}$ and $\widehat{e}_{\tau h}$ are continuous functions in time on $\bar{I}$.

Lemma 24 (Boundedness) *For all $n = 1, \ldots, N$, the following bound holds:*

$$|\widetilde{B}_h^n(\widehat{\eta}, v_{\tau h})| \lesssim \Big\{(\widehat{E}_n^{\mathrm{T}}(u))^2 + (\widehat{E}_n^{\mathrm{S}}(u))^2\Big\}^{1/2} \big\{Q_n\big(|\!|\!| v_{\tau h} |\!|\!|^2\big)\big\}^{1/2}, \quad (78)$$

with the time and space errors respectively given by

$$(\widehat{E}_n^{\mathrm{T}}(u))^2 := Q_n\Big(\|\partial_t(u - R_\tau^{k+1} u)\|_V^2 + \|\partial_t^2(u - I_\tau^{k+2} u)\|_L^2\Big)\,, \quad (79)$$

$$(\widehat{E}_n^{\mathrm{S}}(u))^2 := Q_n\Big(|\!|\!| \partial_t R_\tau^{k+1} u - P_h \partial_t R_\tau^{k+1} u |\!|\!|_{h,\frac{1}{2}}^2\Big)\,. \quad (80)$$

Proof. We decompose $\widetilde{B}_h^n(\widehat{\eta}, v_{\tau h})$ as

$$\begin{aligned}
\widetilde{B}_h^n(\widehat{\eta}, v_{\tau h}) &= Q_n\Big((\partial_t(\partial_t u - \mathcal{L}_\tau P_h \partial_t R_\tau^{k+1} u), v_{\tau h})_L\Big) \\
&\quad + Q_n\Big((A_h(\partial_t u - \partial_t R_\tau^{k+1} u), v_{\tau h})_L\Big) \\
&\quad + Q_n\Big((A_h(\partial_t R_\tau^{k+1} u - P_h \partial_t R_\tau^{k+1} u), v_{\tau h})_L\Big) =: T_1 + T_2 + T_3,
\end{aligned}$$

where we have used $Q_n\big((A_h\mathcal{L}_\tau P_h\partial_t R_\tau^{k+1}u, v_{\tau h})_L\big) = Q_n\big((A_h P_h\partial_t R_\tau^{k+1}u, v_{\tau h})_L\big)$ owing to (18). We first bound T_2 and T_3. Since $(\partial_t u - \partial_t R_\tau^{k+1}u)$ is in W, the consistency of the discrete operator A_h together with the Cauchy–Schwarz inequality lead to

$$\begin{aligned}|T_2| &= \Big|Q_n\Big((A(\partial_t u - \partial_t R_\tau^{k+1}u), v_{\tau h})_L\Big)\Big| \\ &\lesssim \Big\{Q_n\Big(\|\partial_t(u - R_\tau^{k+1}u)\|_V^2\Big)\Big\}^{1/2}\big\{Q_n\big(\|v_{\tau h}\|_L^2\big)\big\}^{1/2}.\end{aligned}$$

For T_3, we use boundedness on orthogonal subscales (22) to infer that

$$|T_3| \le \Big\{Q_n\Big(\|\!|\partial_t R_\tau^{k+1}u - P_h\partial_t R_\tau^{k+1}u|\!\|_{h,\frac12}^2\Big)\Big\}^{1/2}\big\{Q_n\big(\|\!|v_{\tau h}|\!\|^2\big)\big\}^{1/2}.$$

Finally, concerning T_1, we observe that

$$T_1 = Q_n\Big((\partial_t^2(u - I_\tau^{k+2}u), v_{\tau h})_L\Big) + Q_n\Big((\partial_t\{\partial_t I_\tau^{k+2}u - \mathcal{L}_\tau P_h\partial_t R_\tau^{k+1}u\}, v_{\tau h})_L\Big) =: T_{1,1} + T_{1,2}.$$

The term $T_{1,1}$ can be simply bounded as

$$|T_{1,1}| \le \Big\{Q_n\Big(\|\partial_t^2(u - I_\tau^{k+2}u)\|_L^2\Big)\Big\}^{1/2}\big\{Q_n\big(\|v_{\tau h}\|_L^2\big)\big\}^{1/2}.$$

Turning to $T_{1,2}$, we can drop the projection P_h, and since both arguments of the L^2-inner product are in $\mathbb{P}_k(I_n, L)$, we infer that

$$\begin{aligned}T_{1,2} &= \int_{I_n}(\partial_t(\partial_t I_\tau^{k+2}u - \mathcal{L}_\tau\partial_t R_\tau^{k+1}u), v_{\tau h})_L \\ &= -\int_{I_n}(\partial_t I_\tau^{k+2}u - \mathcal{L}_\tau\partial_t R_\tau^{k+1}u, \partial_t v_{\tau h})_L + \Big[(\partial_t I_\tau^{k+2}u - \mathcal{L}_\tau\partial_t R_\tau^{k+1}u, v_{\tau h})_L\Big]_{t_{n-1}^+}^{t_n}\end{aligned}$$

The first term on the right-hand side vanishes since the integrand is in $\mathbb{P}_{2k}(I_n)$ and vanishes at all the right-sided Gauss–Radau integration points by construction. The contribution of the second term at t_n vanishes for the same reason. Finally, the contribution of the second term at t_{n-1}^+ also vanishes since

$$\begin{aligned}\partial_t I_\tau^{k+2}u(t_{n-1}^+) - \mathcal{L}_\tau\partial_t R_\tau^{k+1}u(t_{n-1}^+) &= \partial_t I_\tau^{k+2}u(t_{n-1}^+) - \partial_t R_\tau^{k+1}u(t_{n-1}) \\ &= \partial_t I_\tau^{k+2}u(t_{n-1}^+) - \partial_t I_\tau^{k+2}u(t_{n-1}) = 0.\end{aligned}$$

Collecting the above bounds yields the assertion. □

Lemma 25 (Estimates on $\widehat{e}_{\tau h}$) *For all $m = 1, \dots, N$, the following bound holds:*

$$\|\widehat{e}_{\tau h}(t_m)\|_L^2 + \sum_{n=1}^m Q_n\big(\|\!|\widehat{e}_{\tau h}|\!\|^2\big) \lesssim (\widehat{E}_0)^2 + \sum_{n=1}^m\{(\widehat{E}_n^{\mathrm{T}}(u))^2 + (\widehat{E}_n^{\mathrm{S}}(u))^2 + (\widehat{E}_n^{\mathrm{R}}(f))^2\}, \tag{81}$$

with right-hand side error $(\widehat{E}_n^{\mathrm{R}}(f))^2 := Q_n\big(\|\partial_t f - \partial_t L_\tau^{k+1}f\|_L^2\big)$ and initial error $(\widehat{E}_0)^2 := \|P_h A u_0 - A_h u_{0h})\|_L^2$. Moreover, assuming $\tau_n \lesssim \tau_{n-1}$ and $\tau_1 \lesssim 1$, the following bound holds:

$$\|\widehat{e}_{\tau h}\|_{L^2(I,L)}^2 \lesssim (\widehat{E}_0)^2 + \sum_{n=1}^N\{(\widehat{E}_n^{\mathrm{T}}(u))^2 + (\widehat{E}_n^{\mathrm{S}}(u))^2 + (\widehat{E}_n^{\mathrm{R}}(f))^2\}. \tag{82}$$

Proof. Owing to Lemma 23 (consistency) and the error decomposition (77), we infer that

$$\widetilde{B}_h^n(\widehat{e}_{\tau h}, v_{\tau h}) = -\widetilde{B}_h^n(\widehat{\eta}, v_{\tau h}) + Q_n\Big((\partial_t f - \partial_t L_\tau^{k+1} f, v_{\tau h})_L\Big).$$

The first term on the right-hand side is bounded using Lemma 24 and the second one using the Cauchy–Schwarz inequality, yielding

$$\widetilde{B}_h^n(\widehat{e}_{\tau h}, v_{\tau h}) \lesssim \Big\{(\widehat{E}_n^{\mathrm{T}}(u))^2 + (\widehat{E}_n^{\mathrm{S}}(u))^2 + \widehat{E}_n^{\mathrm{R}}(f))^2\Big\}^{1/2} \big\{Q_n\big(|\!|\!| v_{\tau h} |\!|\!|^2\big)\big\}^{1/2}$$

Setting $v_{\tau h} = I_\tau^{\mathrm{GR}} \widehat{e}_{\tau h}$ and using the stability property of $\widetilde{B}_h^n$ stated in Lemma 4 for the time-continuous function $\widehat{e}_{\tau h}$ together with a Young inequality, we infer, as in the proof of Lemma 10, that

$$\frac{1}{2}\|\widehat{e}_{\tau h}(t_n)\|_L^2 - \frac{1}{2}\|\widehat{e}_{\tau h}(t_{n-1})\|_L^2 + Q_n\big(|\!|\!| \widehat{e}_{\tau h} |\!|\!|^2\big) \lesssim (\widehat{E}_n^{\mathrm{T}}(u))^2 + (\widehat{E}_n^{\mathrm{S}}(u))^2 + \widehat{E}_n^{\mathrm{R}}(f))^2.$$

Moreover, the initial discrete error is $\widehat{e}_{\tau h}(0) = P_h \partial_t u(0) - \partial_t \mathcal{L}_\tau u_{\tau h}(0) = -P_h A u_0 + A_h u_{0h}$. We conclude as in the proof of Lemma 10. □

Theorem 26 (Time-derivative error estimate) *Let u be the exact solution and let $u_{\tau h}$ be the fully discrete solution. Assume $k \geq 1$. Assume **(H)** in the case of conforming FEM. Then, for the error $\widehat{e}(t)$ defined in (75), the following bound holds for all $m = 1, \ldots, N$,*

$$\|\widehat{e}(t_m)\|_L^2 \lesssim (\widehat{E}_0)^2 + t_m \max_{1\leq n\leq m} \Big\{\widehat{C}_n^{\mathrm{T}}(u,f)\tau_n^{2(k+1)} + C_n^{\mathrm{S}}(u)h^{2r+1}\Big\} + \widehat{C}'_m(u)h^{2(r+1)}, \tag{83}$$

with $\widehat{C}_n^{\mathrm{T}}(u,f) := C_n^{\mathrm{T}}(u) + |f|^2_{C^{k+2}(\bar{I}_n,L)}$, $C_n^{\mathrm{T}}(u)$ and $C_n^{\mathrm{S}}(u)$ defined in Theorem 11, and $\widehat{C}'_m(u) = |\partial_t u(t_m)|^2_{H^{r+1}(\Omega)}$. Moreover, assuming $\tau_n \lesssim \tau_{n-1}$ for all $n = 2, \ldots, N$ and $\tau_1 \lesssim 1$, the following bound holds:

$$\|\widehat{e}\|^2_{L^2(I,L)} \lesssim (\widehat{E}_0)^2 + T \max_{1\leq n\leq N} \Big\{\widehat{C}_n^{\mathrm{T}}(u,f)\tau_n^{2(k+1)} + C_n^{\mathrm{S}}(u)h^{2r+1}\Big\}. \tag{84}$$

Proof. We infer that, for all $n = 1, \ldots, N$,

$$\begin{aligned}
|\widehat{E}_n^{\mathrm{T}}(u)|^2 &\lesssim \tau_n \left(\tau_n^{2(k+1)} |u|^2_{C^{k+3}(\bar{I}_n,L)} + \tau_n^{2(k+1)} |u|^2_{C^{k+2}(\bar{I}_n,V)}\right),\\
|\widehat{E}_n^{\mathrm{S}}(u)|^2 &\lesssim \tau_n h^{2r+1} \|u\|^2_{C^1(\bar{I}_n,H^{r+1}(\Omega))},\\
|\widehat{E}_n^{\mathrm{R}}(f)|^2 &\lesssim \tau_n \tau_n^{2(k+1)} |f|^2_{C^{k+2}(\bar{I}_n,L)}.
\end{aligned}$$

using (38) with $B = L$ and (42) with $B = V$ for the time error, the approximation property (29) of P_h combined with the stability of R_τ^{k+1} from Corollary 7 with $B = H^{r+1}(\Omega)$ for the space error, and the approximation properties of the Lagrange interpolate L_τ^{k+1} for the right-hand side error. Moreover, recalling the error decomposition (77), we observe that, for all $m = 1, \ldots, N$,

$$\|\widehat{\eta}(t_m)\|_L = \|\partial_t u(t_m) - P_h \partial_t u(t_m)\|_L \lesssim h^{r+1} |\partial_t u|_{H^{r+1}(\Omega)},$$

since $\vartheta_m(t_m) = 1$ and $\partial_t R_\tau^{k+1} u(t_m) = \partial_t u(t_m)$. In addition, using the definition of $\mathcal{L}_\tau$ and the triangle inequality, we infer that

$$\|\widehat{\eta}\|^2_{L^2(I_n,L)} \le 2\|\partial_t u - P_h \partial_t R_\tau^{k+1} u\|^2_{L^2(I_n,L)} + 2\|P_h[\partial_t R_\tau^{k+1} u]_{n-1}\vartheta_n\|^2_{L^2(I_n,L)} =: T_1 + T_2.$$

Using again the triangle inequality, the fact that P_h is a projection, and (42) with $B = L$ leads to

$$\begin{aligned} |T_1| &\lesssim \|\partial_t u - P_h \partial_t u\|^2_{L^2(I_n,L)} + \|\partial_t u - \partial_t R_\tau^{k+1} u\|^2_{L^2(I_n,L)} \\ &\lesssim \tau_n \left(h^{2(r+1)} |u|^2_{C^1(\bar{I}_n, H^{r+1}(\Omega))} + \tau_n^{2(k+1)} |u|^2_{C^{k+2}(\bar{I}_n,L)} \right), \end{aligned}$$

while using the regularity of u, the fact that $\partial_t R_\tau^{k+1} u(t_{n-1}) = \partial_t u(t_{n-1})$, and (42) with $B = L$ leads to

$$\begin{aligned} |T_2| &\lesssim \tau_n \|[\partial_t R_\tau^{k+1} u]_{n-1}\|^2_L = \tau_n \|\partial_t R_\tau^{k+1} u(t_{n-1}^+) - \partial_t u(t_{n-1})\|^2_L \\ &\le \tau_n \|\partial_t u - \partial_t R_\tau^{k+1} u\|^2_{C^0(\bar{I}_n,L)} \lesssim \tau_n \tau_n^{2(k+1)} |u|^2_{C^{k+2}(\bar{I}_n,L)}. \end{aligned}$$

We conclude using Lemma 25 and the triangle inequality, as well as $h \lesssim 1$ and $\|\cdot\|_L \lesssim \|\cdot\|_V$ for (84). □

Remark 27 (Initial error) *The initial error $\widehat{E}_0$ vanishes when the discrete initial condition is chosen as the solution of the steady transport problem $A_h u_{0h} = P_h A u_0$. Otherwise, this error is typically of order h^r if u_0 is smooth enough.*

6.2 Discrete graph norm error estimates

Lemma 28 (Graph norm estimate on $\tilde{e}_{\tau h}$) *For all $m = 1, \dots, N$, the following bound holds:*

$$\sum_{n=1}^m Q_n\big(|||\tilde{e}_{\tau h}|||^2_\sharp\big) \lesssim (\widehat{E}_0)^2 + \sum_{n=1}^m \{(E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2 + (\widehat{E}_n^{\mathrm{T}}(u))^2 + (\widehat{E}_n^{\mathrm{S}}(u))^2 + (\widehat{E}_n^{\mathrm{R}}(f))^2\}. \quad (85)$$

Proof. Owing to the discrete inf-sup condition satisfied by A_h, see (24), we know that, for all $n = 1, \dots, N$ and all $\mu = 1, \dots, k+1$, there is $w_{n,\mu} \in V_h$ such that

$$|||\tilde{e}_{\tau h}(t_{n,\mu})|||^2_\sharp \lesssim (A_h \tilde{e}_{\tau h}(t_{n,\mu}), w_{n,\mu})_L \qquad \text{and} \qquad |||w_{n,\mu}|||_\sharp \le |||\tilde{e}_{\tau h}(t_{n,\mu})|||_\sharp.$$

Let $w_{\tau h}$ be the function in $X_{\tau h}^k$ uniquely defined by $w_{\tau h}(t_{n,\mu}) := w_{n,\mu}$, for all $n = 1, \dots, N$ and all $\mu = 1, \dots, k+1$. Then, we obtain

$$Q_n\big(|||\tilde{e}_{\tau h}|||^2_\sharp\big) \lesssim Q_n((A_h \tilde{e}_{\tau h}, w_{\tau h})_L) \qquad \text{and} \qquad Q_n\big(|||w_{\tau h}|||^2_\sharp\big) \le Q_n\big(|||\tilde{e}_{\tau h}|||^2_\sharp\big).$$

Using the definition of $\widetilde{B}_h^n$ and consistency (Lemma 8), we obtain

$$\begin{aligned} Q_n((A_h \tilde{e}_{\tau h}, w_{\tau h})_L) &= \widetilde{B}_h^n(\tilde{e}_{\tau h}, w_{\tau h}) - Q_n((\partial_t \tilde{e}_{\tau h}, w_{\tau h})_L) \\ &= -\widetilde{B}_h^n(\eta, w_{\tau h}) - Q_n((\partial_t \tilde{e}_{\tau h}, w_{\tau h})_L) \\ &= -\widetilde{B}_h^n(\eta, w_{\tau h}) - Q_n((\widehat{e}_{\tau h}, w_{\tau h})_L), \end{aligned}$$

since it is readily deduced from (45) and (77) that $\widehat{e}_{\tau h}(t_{n,\mu}) = \mathcal{L}_\tau \partial_t \tilde{e}_{\tau h}(t_{n,\mu}) = \partial_t \tilde{e}_{\tau h}(t_{n,\mu})$ for all $n = 1, \dots, N$ and all $\mu = 1, \dots, k+1$. The first term on the right-hand side is bounded using Lemma 9 and the second term using the Cauchy–Schwarz inequality, yielding

$$Q_n\big(|\!|\!|\tilde{e}_{\tau h}|\!|\!|_\sharp^2\big) \lesssim \big\{(E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2 + Q_n\big(\|\widehat{e}_{\tau h}\|_L^2\big)\big\}^{1/2} \big\{Q_n\big(|\!|\!|w_{\tau h}|\!|\!|^2\big)\big\}^{1/2}.$$

Observing that $Q_n\big(|\!|\!|w_{\tau h}|\!|\!|^2\big) \le Q_n\Big(|\!|\!|w_{\tau h}|\!|\!|_\sharp^2\Big) \le Q_n\Big(|\!|\!|\tilde{e}_{\tau h}|\!|\!|_\sharp^2\Big)$, we obtain by Young's inequality

$$Q_n\big(|\!|\!|\tilde{e}_{\tau h}|\!|\!|_\sharp^2\big) \lesssim (E_n^{\mathrm{T}}(u))^2 + (E_n^{\mathrm{S}}(u))^2 + Q_n\big(\|\widehat{e}_{\tau h}\|_L^2\big).$$

We conclude summing from $n = 1$ to m, and using Lemma 25 to bound $Q_n\big(\|\widehat{e}_{\tau h}\|_L^2\big)$. □

Theorem 29 (Graph norm error estimate) *Let u be the exact solution and let $u_{\tau h}$ be the fully discrete solution. Assume $k \ge 1$ and $\tau_n \lesssim 1$ for all $n = 1, \dots, N$. Assume **(H)** in the case of conforming FEM. Then, for the error $\tilde{e}(t)$ defined in (43) and all $m = 1, \dots, N$, the following bound holds:*

$$\sum_{n=1}^m Q_n\big(|\!|\!|\tilde{e}|\!|\!|_\sharp^2\big) \lesssim (\widehat{E}_0)^2 + t_m \max_{1\le n\le m} \Big\{\widetilde{C}_n^{\mathrm{T}}(u,f)\tau_n^{2(k+1)} + C_n^{\mathrm{S}}(u)h^{2r+1}\Big\}, \tag{86}$$

where $\widetilde{C}_n^{\mathrm{T}}(u,f) := |u|^2_{C^{k+3}(\bar{I}_n,L)} + |u|^2_{C^{k+2}(\bar{I}_n,W)} + |f|^2_{C^{k+2}(\bar{I}_n,L)}$ and $C_n^{\mathrm{S}}(u)$ defined in Theorem 11.

Proof. Recalling the error decomposition (45) and using the triangle inequality, we need to bound $\sum_{n=1}^m Q_n\Big(|\!|\!|\tilde{e}_{\tau h}|\!|\!|_\sharp^2\Big)$ and $\sum_{n=1}^m Q_n\Big(|\!|\!|\eta|\!|\!|_\sharp^2\Big)$. For the first term, we use Lemma 28 together with the bounds on the various errors derived in the proofs of Theorems 11 and 26 to infer that $\sum_{n=1}^m Q_n\Big(|\!|\!|\tilde{e}_{\tau h}|\!|\!|_\sharp^2\Big)$ is bounded by the right-hand side of (86) (observe in particular for the time error that $\|\cdot\|_V \lesssim \|\cdot\|_W$ so that $\widehat{C}_n^{\mathrm{T}}(u,f) \lesssim \widetilde{C}_n^{\mathrm{T}}(u,f)$). Concerning the second term, we use the triangle inequality, Lemma 6 (with $B = W$), the fact that $|\!|\!|w|\!|\!|_\sharp \lesssim \|w\|_W$ for all $w \in W$, the approximation property of P_h in the $|\!|\!|\cdot|\!|\!|_\sharp$-norm, and the stability of R_τ^{k+1} (with $B = H^{r+1}(\Omega)$) to infer that

$$\begin{aligned} Q_n\big(|\!|\!|\eta|\!|\!|_\sharp^2\big) &\le 2Q_n\Big(|\!|\!|u - R_\tau^{k+1}u|\!|\!|_\sharp^2\Big) + 2Q_n\Big(|\!|\!|R_\tau^{k+1}u - P_hR_\tau^{k+1}u|\!|\!|_\sharp^2\Big) \\ &\lesssim \tau_n^{2(k+2)}|u|^2_{C^{k+2}(\bar{I}_n,W)} + h^{2r+1}\|u\|^2_{C^1(\bar{I}_n,H^{r+1}(\Omega))}. \end{aligned}$$

Summing over n and recalling that $\tau_n \lesssim 1$ yields the assertion. □

7 Appendix

Proof of Lemma 6. Let $\hat{I} := [-1,1]$ denote the reference time interval and $T_n : \hat{I} \to \bar{I}_n$ the affine reference mapping with $t = T_n(\hat{t}) := (t_{n-1} + t_n)/2 + \frac{\tau_n}{2}\hat{t}$. We assign to $u \in C^{k+2}(\bar{I}_n, B)$ a reference function $\hat{u} \in C^{k+2}(\hat{I}, B)$ defined by $\hat{u}(\hat{t}) := u|_{I_n}(T_n(\hat{t}))$ for all $\hat{t} \in \hat{I}$, where we take the right-sided limit $u(t_{n-1}^+)$ for $\hat{t} = -1$. Then, for $m = 1, \dots, k+2$, we infer that

$$\partial_{\hat{t}}^m \hat{u}(\hat{t}) = \partial_t^m u(T_n(\hat{t}))\left(\frac{\tau_n}{2}\right)^m \qquad \forall\, \hat{t} \in \hat{I},$$

where for boundary points $\hat{t} = \pm 1$ the one-sided derivatives of u are taken that come from interval I_n. The idea of the proof is to construct an interpolation operator $\hat{R}_\tau^{k+1} : C^1(\hat{I}, B) \to \mathbb{P}_{k+1}(\hat{I}, B)$, which is related to R_τ^{k+1} by

$$\hat{R}_\tau^{k+1}\hat{u}(\hat{t}) = R_\tau^{k+1}u(t) \qquad \forall\, \hat{t} \in \hat{I}, \quad t = T_n(\hat{t}), \tag{87}$$

and satisfies the properties

$$\hat{R}_\tau^{k+1}\hat{p} = \hat{p} \qquad \forall\, \hat{p} \in \mathbb{P}_{k+1}(\hat{I}, B). \tag{88}$$

$$\|\hat{R}_\tau^{k+1}\hat{w}\|_{C^0(\hat{I},B)} \le M\|\hat{w}\|_{C^1(\hat{I},B)} \qquad \forall\, \hat{w} \in C^1(\hat{I}, B). \tag{89}$$

Let $\hat{p} \in \mathbb{P}_{k+1}(\hat{I}, B)$ be the Taylor polynomial $\hat{p}(\hat{t}) := \sum_{j=0}^{k+1} \frac{1}{j!}\partial_{\hat{t}}^j \hat{u}(0)\hat{t}^j$. Since $\partial_{\hat{t}}\hat{p}$ is the Taylor polynomial for $\partial_{\hat{t}}\hat{u}$ of order k, we infer that

$$\|\hat{u} - \hat{p}\|_{C^1(\hat{I},B)} \le \frac{1}{(k+1)!}|\hat{u}|_{C^{k+2}(\hat{I},B)} = \frac{\tau_n^{k+2}}{2^{k+2}(k+1)!}|u|_{C^{k+2}(\bar{I}_n,B)}$$

Then, using (88) and (89) leads to

$$\begin{aligned}\|u - R_\tau^{k+1}u\|_{C^0(\bar{I}_n,B)} &= \|\hat{u} - \hat{R}_\tau^{k+1}\hat{u}\|_{C^0(\hat{I},B)} \le \|\hat{u} - \hat{p}\|_{C^0(\hat{I},B)} + \|\hat{R}_\tau^{k+1}(\hat{p} - \hat{u})\|_{C^0(\hat{I},B)} \\ &\le (1+M)\|\hat{u} - \hat{p}\|_{C^1(\hat{I},B)} \lesssim \tau_n^{k+2}|u|_{C^{k+2}(\bar{I}_n,B)},\end{aligned}$$

yielding (41), while the stability of R_τ^{k+1} results from (89) and the use of the reference mapping. It remains to construct the operator $\hat{R}_\tau^{k+1}$ and to verify (87), (88), and (89). Let $\hat{\varphi}_j \in \mathbb{P}_{k+2}(\hat{I})$, $j = 0, \dots, k+2$, denote the basis functions of the Lagrange/Hermite interpolation with respect to the nodal points $-1 = \hat{s}_0 < \cdots < \hat{s}_k = 1$ such that the interpolate $\hat{I}_\tau^{k+2}\hat{u} \in \mathbb{P}_{k+2}(\hat{I}, B)$ verifying $\hat{I}_\tau^{k+2}\hat{u}(\hat{s}_j) = \hat{u}(\hat{s}_j)$, $j = 0, \dots, k$, and $\partial_{\hat{t}}(\hat{I}_\tau^{k+2}\hat{u})(\pm 1) = \partial_{\hat{t}}\hat{u}(\pm 1)$ has the representation

$$\hat{I}_\tau^{k+2}\hat{u} = \sum_{j=0}^{k} \hat{u}(\hat{s}_j)\hat{\varphi}_j + \partial_{\hat{t}}\hat{u}(-1)\hat{\varphi}_{k+1} + \partial_{\hat{t}}\hat{u}(1)\hat{\varphi}_{k+2}.$$

This representation implies the stability estimate

$$\|\hat{I}_\tau^{k+2}\hat{u}\|_{C^1(\hat{I},B)} \le M_1\|\hat{u}\|_{C^1(\hat{I},B)} \qquad \text{with} \quad M_1 := \sum_{j=0}^{k+2} \|\hat{\varphi}_j\|_{C^1(\hat{I},\mathbb{R})}. \tag{90}$$

Moreover, $\hat{I}_\tau^{k+2}\hat{u}(\hat{t}) = I_\tau^{k+2}u(T_n(\hat{t}))$ for all $\hat{t} \in \hat{I}$. Now, let $\hat{\phi}_j \in \mathbb{P}_k(\hat{I})$, $j = 1, \dots, k+1$, denote the basis functions of the Lagrange interpolation with respect to the right-sided Gauss–Radau integration points $\hat{t}_\mu \in \hat{I}$, $\mu = 1, \dots, k+1$, satisfying the property $\hat{\phi}_j(\hat{t}_\mu) = \delta_{j,\mu}$ for all $j, \mu \in \{1, \dots, k+1\}$ where $\delta_{j,\mu}$ denotes Kronecker symbol. Using these basis functions, we define functions $\hat{\psi}_j \in \mathbb{P}_{k+1}(\hat{I})$, $j = 0, \dots, k+1$, by means of

$$\hat{\psi}_0(\hat{t}) := 1, \qquad \hat{\psi}_j(\hat{t}) := \int_{-1}^{\hat{t}} \hat{\phi}_j(s)\, ds \quad \forall\, j = 1, \dots, k+1.$$

These functions satisfy the properties $\partial_{\hat{t}}\hat{\psi}_j(\hat{t}_\mu) = \delta_{j,\mu}$ for all $j = 0, \ldots, k+1$ and $\mu = 1, \ldots, k+1$ as well as $\hat{\psi}_j(-1) = \delta_{j,0}$ for all $j = 0, \ldots, k+1$. Therefore, the functions $\hat{\psi}_j$ form a basis of $\mathbb{P}_{k+1}(\hat{I})$ and the interpolate

$$\hat{R}_\tau^{k+1}\hat{u} := \hat{I}_\tau^{k+2}\hat{u}(-1)\hat{\psi}_0 + \sum_{j=1}^{k+1} \partial_{\hat{t}}\hat{I}_\tau^{k+2}\hat{u}(\hat{t}_j)\hat{\psi}_j \tag{91}$$

satisfies the conditions

$$\hat{R}_\tau^{k+1}\hat{u}(-1) = \hat{I}_\tau^{k+2}\hat{u}(-1), \qquad \partial_{\hat{t}}\hat{R}_\tau^{k+1}\hat{u}(\hat{t}_\mu) = \partial_{\hat{t}}\hat{I}_\tau^{k+2}\hat{u}(\hat{t}_\mu) \quad \forall\, \mu = 1, \ldots, k+1.$$

This shows that (87) holds since the polynomial $R_\tau^{k+1}u(T_n(\hat{t}))$ is also in $\mathbb{P}_{k+1}(\hat{I}, B)$ and satisfies the above conditions owing to (39)-(40). Applying the definition (91) to an arbitrary $\hat{w} \in C^1(\hat{I}, B)$, we obtain by means of (90) the stability estimate

$$\|\hat{R}_\tau^{k+1}\hat{w}\|_{C^0(\hat{I},B)} \le \|\hat{I}_\tau^{k+2}\hat{w}\|_{C^1(\hat{I},B)} \sum_{j=0}^{k+1} \|\hat{\psi}_j\|_{C^0(\hat{I},\mathbb{R})} \le M\|\hat{w}\|_{C^1(\hat{I},B)},$$

where $M = M_1 \sum_{j=0}^{k+1} \|\hat{\psi}_j\|_{C^0(\hat{I},\mathbb{R})}$ which proves (89). Finally, to prove (88), we observe that an arbitrary polynomial $\hat{p} \in \mathbb{P}_{k+1}(\hat{I}, B)$ has a basis representation of the form

$$\hat{p}(\hat{t}) = \sum_{j=0}^{k+1} \hat{p}_j\hat{\psi}_j(\hat{t}) \qquad \forall\, \hat{t} \in \hat{I}, \quad \hat{p}_j \in B.$$

From the properties of the basis functions $\hat{\psi}_j$, we obtain $\hat{p}(-1) = \hat{p}_0$ and $\partial_{\hat{t}}\hat{p}(\hat{t}_j) = \hat{p}_j$ for all $j = 1, \ldots, k+1$. Since $\hat{I}_\tau^{k+2}$ leaves $\mathbb{P}_{k+1}(\hat{I}, B)$ invariant, we infer that $\hat{p} = \hat{p}(-1)\hat{\psi}_0 + \sum_{j=1}^{k+1} \partial_{\hat{t}}\hat{p}(\hat{t}_j)\hat{\psi}_j = \hat{R}_\tau^{k+1}\hat{p}$. $\square$

References

[1] E. Bänsch A. Brenner and M. Bause. A priori error analysis for finite element approximations of the stokes problem on dynamic meshes. *IMA J. Numer. Anal.*, 34(1):123–146, 2014.

[2] N. Ahmed, G. Matthies, L. Tobiska, and H. Xie. Discontinuous Galerkin time stepping with local projection stabilization for transient convection-diffusion-reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 200(21-22):1747–1756, 2011.

[3] R. E. Bank and H. Yserentant. On the H^1-stability of the L_2-projection onto finite element spaces. *Numer. Math.*, 126:361–381, 2014.

[4] M. Braack, E. Burman, V. John, and G. Lube. Stabilized finite element methods for the generalized Oseen problem. *Comput. Methods Appl. Mech. Engrg.*, 196(4-6):853–866, 2007.

[5] F. Brezzi, L. D. Marini, and E. Süli. Discontinuous Galerkin methods for first-order hyperbolic problems. *Math. Models Methods Appl. Sci.*, 14(12):1893–1903, 2004.

[6] E. Burman. A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.*, 43(5):2012–2033 (electronic), 2005.

[7] E. Burman and A. Ern. A continuous finite element method with face penalty to approximate Friedrichs' systems. *M2AN Math. Model. Numer. Anal.*, 41(1):55–76, 2007.

[8] E. Burman, A. Ern, and M.A. Fernández. Explicit Runge-Kutta schemes and finite elements with symmetric stabilization for first-order linear PDE systems. *SIAM J. Numer. Anal.*, 48(6):2019–2042, 2010.

[9] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection–diffusion–reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193:1437–1453, 2004.

[10] J. Česenek and M. Feistauer. Theory of the space-time discontinuous Galerkin method for nonstationary parabolic problems with nonlinear convection and diffusion. *SIAM J. Numer. Anal.*, 50(3):1181–1206, 2012.

[11] B. Cockburn, S. Hou, and C.-W. Shu. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV. The multidimensional case. *Math. Comp.*, 54(190):545–581, 1990.

[12] B. Cockburn, S. Lin, and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. III. One-dimensional systems. *J. Comput. Phys.*, 84(1):90–113, 1989.

[13] B. Cockburn and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. II. General framework. *Math. Comp.*, 52(186):411–435, 1989.

[14] R. Codina. Stabilization of incompressibility and convection through orthogonal sub-scales in finite element methods. *Comput. Methods Appl. Mech. Engrg.*, 190(13-14):1579–1599, 2000.

[15] R. Codina. Stabilized finite element approximation of transient incompressible flows using orthogonal subscales. *Comput. Methods Appl. Mech. Engrg.*, 191(39-40):4295–4321, 2002.

[16] D. A. Di Pietro and A. Ern. *Mathematical Aspects of Discontinuous Galerkin Methods*, volume 69 of *Mathématiques & Applications*. Springer, Berlin Heidelberg, 2012.

[17] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems. II. Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$. *SIAM J. Numer. Anal.*, 32(3):706–740, 1995.

[18] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[19] A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. I. General theory. *SIAM J. Numer. Anal.*, 44(2):753–778, 2006.

[20] M. Feistauer, V. Kučera, K. Najzar, and J. Prokopová. Analysis of space-time discontinuous Galerkin method for nonlinear convection-diffusion problems. *Numer. Math.*, 117(2):251–288, 2011.

[21] K. O. Friedrichs. Symmetric positive linear differential equations. *Comm. Pure and Appl. Math.*, 11:333–418, 1958.

[22] J.-L. Guermond. Stabilization of Galerkin approximations of transport equations by subgrid modeling. *M2AN Math. Model. Numer. Anal.*, 33(6):1293–1316, 1999.

[23] J.-L. Guermond. Subgrid stabilization of Galerkin approximations of linear monotone operators. *IMA J. Numer. Anal.*, 21:165–197, 2001.

[24] M. Hochbruck and T. Pažur. Implicit Runge–Kutta methods and discontinuous Galerkin discretizations for linear Maxwell's equations. Technical report, Karlsruhe Institute of Technology, 2013.

[25] S. Hussain, F. Schieweck, and S. Turek. Higher order Galerkin time discretizations and fast multigrid solvers for the heat equation. *J. Numer. Math.*, 19(1):41–61, 2011.

[26] S. Hussain, F. Schieweck, and S. Turek. Higher order Galerkin time discretization for nonstationary incompressible flow. Technical report, Fakultät für Mathematik, TU Dortmund, January 2012. Ergebnisberichte des Instituts für Angewandte Mathematik, Nummer 444.

[27] S. Hussain, F. Schieweck, and S. Turek. A note on accurate and efficient higher order Galerkin time stepping schemes for the nonstationary Stokes equations. *The Open Numerical Methods Journal*, 4:35–45, 2012.

[28] J. Jaffré, C. Johnson, and A. Szepessy. Convergence of the discontinuous Galerkin finite element method for hyperbolic conservation laws. *Math. Models Methods Appl. Sci.*, 5(3):367–386, 1995.

[29] C. Johnson and J. Pitkäranta. An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.*, 46(173):1–26, 1986.

[30] P. Lesaint and P.-A. Raviart. On a finite element method for solving the neutron transport equation. In *Mathematical Aspects of Finite Elements in Partial Differential Equations*, pages 89–123. Publication No. 33. Math. Res. Center, Univ. of Wisconsin-Madison, Academic Press, New York, 1974.

[31] D. Leykekhman and B. Vexler. Optimal a priori error estimates of parabolic optimal control problems with pointwise control. *SIAM J. Numer. Anal.*, 51(5):2797–2821, 2013.

[32] G. Matthies and F. Schieweck. Higher order variational time discretizations for nonlinear systems of ordinary differential equations. Preprint 23/2011, Otto-von-Guericke Universität Magdeburg, Fakultät für Mathematik, 2011. `http://www-ian.math.uni-magdeburg.de/home/schieweck`.

[33] T. Richter, A. Springer, and B. Vexler. Efficient numerical realization of discontinuous Galerkin methods for temporal discretization of parabolic problems. *Numer. Math.*, 124(1):151–182, 2013.

[34] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust numerical methods for singularly perturbed differential equations*, volume 24 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2008. Convection-diffusion-reaction and flow problems.

[35] D. Schötzau and C. Schwab. An *hp* a priori error analysis of the DG time-stepping method for initial value problems. *Calcolo*, 37(4):207–232, 2000.

[36] A. Springer and B. Vexler. Third order convergent time discretization for parabolic optimal control problems with control constraints. *Comput. Optim. Appl.*, 57(1):205–240, 2014.

[37] V. Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[38] J. J. W. van der Vegt and H. van der Ven. Space-time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows. I. General formulation. *J. Comput. Phys.*, 182(2):546–585, 2002.

[39] T. Werder, K. Gerdes, D. Schötzau, and C. Schwab. *hp*-discontinuous Galerkin time stepping for parabolic problems. *Comput. Methods Appl. Mech. Engrg.*, 190(49-50):6685–6708, 2001.

[40] D. Yang. Improved error estimation of dynamic finite element methods for second-order parabolic equations. *J. Comput. Appl. Math.*, 126(1-2):319–338, 2000.

[41] K. Yosida. *Functional Analysis*. Classics in Mathematics. Springer-Verlag, Berlin, Germany, 1995. Reprint of the sixth (1980) edition.

[42] Q. Zhang and C.-W. Shu. Error estimates to smooth solutions of Runge-Kutta discontinuous Galerkin methods for scalar conservation laws. *SIAM J. Numer. Anal.*, 42(2):641–666 (electronic), 2004.

[43] Q. Zhang and C.-W. Shu. Stability analysis and a priori error estimates of the third order explicit Runge-Kutta discontinuous Galerkin method for scalar conservation laws. *SIAM J. Numer. Anal.*, 48(3):1038–1063, 2010.