Equilibrated flux a posteriori error estimates in $L^2(H^1)$-norms for high-order discretizations of parabolic problems

Alexandre Ern, Iain Smears, Martin Vohralík

Equilibrated flux a posteriori error estimates in $L^2(H^1)$-norms for high-order discretizations of parabolic problems*

Alexandre Ern† Iain Smears† Martin Vohralík†

June 15, 2018

Abstract

We consider the a posteriori error analysis of fully discrete approximations of parabolic problems based on conforming hp-finite element methods in space and an arbitrary order discontinuous Galerkin method in time. Using an equilibrated flux reconstruction, we present a posteriori error estimates yielding guaranteed upper bounds on the $L^2(H^1)$-norm of the error, without unknown constants and without restrictions on the spatial and temporal meshes. It is known from the literature that the analysis of the efficiency of the estimators represents a significant challenge for $L^2(H^1)$-norm estimates. Here we show that the estimator is bounded by the $L^2(H^1)$-norm of the error plus the temporal jumps under the one-sided parabolic condition $h^2 \lesssim \tau$. This result improves on earlier works that required stronger two-sided hypotheses such as $h \simeq \tau$ or $h^2 \simeq \tau$; instead our result now encompasses practically relevant cases for computations and allows for locally refined spatial meshes. The constants in our bounds are robust with respect to the mesh and time-step sizes, the spatial polynomial degrees, and also with respect to refinement and coarsening between time-steps, thereby removing any transition condition.

Key words: Parabolic partial differential equations, a posteriori error estimates, guaranteed upper bound, polynomial-degree robustness, high-order methods

1 Introduction

We consider the heat equation

$$\begin{aligned} \partial_t u - \Delta u &= f && \text{in } \Omega \times (0,T), \\ u &= 0 && \text{on } \partial\Omega \times (0,T), \\ u(0) &= u_0 && \text{in } \Omega, \end{aligned} \tag{1.1}$$

†Université Paris-Est, CERMICS (ENPC), 77455 Marne-la-Vallée cedex 2, France & Inria Paris, 2 rue Simone Iff, 75589 Paris, France, (alexandre.ern@enpc.fr, iain.smears@inria.fr, martin.vohralik@inria.fr).

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 647134 GATIPOR).

where $\Omega \subset \mathbb{R}^d$, $1 \leq d \leq 3$, is a bounded, connected, polytopal open set with Lipschitz boundary, and $T > 0$ is the final time. We assume that $f \in L^2(0,T;L^2(\Omega))$, and that $u_0 \in L^2(\Omega)$. We are interested here in the a posteriori error analysis in the $L^2(H^1)$-norm of fully discrete numerical methods for (1.1). In particular, we consider an arbitrary-order discontinuous Galerkin finite element method (DGFEM) in time, coupled with a conforming hp-FEM in space. A posteriori error estimates should ideally provide guaranteed upper bounds on the error, without unknown constants. Otherwise, if the estimators constitute an upper bound on the error up to an unknown constant, then we say instead that the estimators are reliable. Furthermore, the estimators should be locally efficient, meaning that the local estimators should lie below the error measured in a local space-time neighbourhood, up to a generic constant. Finally, the estimators should ideally be robust, with all constants in the bounds being independent of all discretization parameters. Furthermore, on a practical side it is highly desirable that the estimators be locally computable. We refer the reader to [Verfürth(2013)] for an introduction to these concepts. Our motivation for considering the heat equation (1.1) as a model problem is that the a posteriori error estimates developed in this context serve as a starting point for extensions to diverse applications, for example nonlinear problems (see [Amrein & Wihler(2016), Di Pietro et al.(2015), Dolejší et al.(2013), Kreuzer(2013)]), as well as playing a central role in adaptive algorithms (see [Chen & Feng(2004), Gaspoz et al.(2016), Kreuzer et al.(2012)]). For nonconforming discretization methods in space, we refer to [Ern & Vohralík(2010)] as well as [Georgoulis et al.(2011), Nicaise & Soualem(2005)].

The literature shows that the structure of parabolic problems leads to several outstanding challenges facing the central goals in a posteriori error estimation. In particular, several difficulties arise in the analysis of the efficiency and robustness of the estimators. To explain some of the challenges, first recall that the a posteriori error analysis of parabolic problems admits a range of norms in which to measure the error: for instance, these include the $L^2(H^1)$-norm (see [Picasso(1998), Verfürth(1998)]), $L^2(L^2)$-norm (see [Verfürth(1998)]), $L^\infty(L^2)$-norms and $L^\infty(L^\infty)$-norms (see [Eriksson & Johnson(1995)]), $L^\infty(L^2)\cap L^2(H^1)$-norms (see [Lakkis & Makridakis(2006), Makridakis & Nochetto(2003), Schötzau & Wihler(2010)] and more recently [Georgoulis et al.(2017)]), and $L^2(H^1) \cap H^1(H^{-1})$-norms (see [Bergam et al.(2005), Ern & Vohralík(2010), Gaspoz et al.(2016), Nicaise & Soualem(2005), Repin(2002), Verfürth(2003)]). To our knowledge, efficiency results have so far only been proved in the case of the $L^2(H^1)$ norm and of the $L^2(H^1)\cap H^1(H^{-1})$ norm. Although no analysis of efficiency is yet available in the setting of other norms, the optimal order of convergence of the estimators has nonetheless been observed in [Lakkis & Makridakis(2006), Makridakis & Nochetto(2003)] for instance. The efficiency results in the $L^2(H^1)$ norm have been attained under restrictions linking mesh and time-step sizes, whereas in the $L^2(H^1)\cap H^1(H^{-1})$ norm, such restrictions have been removed. It is important to observe that these two functional settings admit an inf-sup theory for the continuous problem that establishes an equivalence between appropriate norms of the error and of the residual. However, a difference between these two settings is that for $L^2(H^1)\cap H^1(H^{-1})$-estimates, the dual norm on the residual localizes straightforwardly with respect to time, whereas this is not the case for the $L^2(H^1)$-estimates.

A posteriori error estimators in the $L^2(H^1)$-norm for a class of nonlinear parabolic problem have been studied in [Verfürth(1998)]. In particular, [Verfürth(1998)] found that the ratio between the constants in the upper and lower bounds for the error by the estimators depends on $1 + \tau h^{-2} + \tau^{-1} h^2 + |\log h|$, see [Verfürth(1998), Prop. 4.1], where h denotes the spatial mesh size and τ denotes the time-step size, and thus the efficiency of the estimators is subject to the assumption that $\tau \simeq h^2$. [Picasso(1998)] studied implicit Euler discretizations of the heat equation: under the assumption that $\tau \simeq h$, he showed that the spatial estimator can be bounded from above by the $L^2(H^1)$-norm of the error plus the temporal jump estimator; in particular, the temporal jump estimator, denoted by ε_K^n in [Picasso(1998), eq. (2.11)], appears on the right-hand side of the lower bound in [Picasso(1998), eq. (2.24)]. In both [Picasso(1998), Verfürth(1998)], the two-sided restrictions between the time-step and mesh sizes have the disadvantage of necessarily requiring that the meshes must be quasi-uniform, and thus theoretically prohibiting adaptive refinement.

Starting with [Verfürth(2003)], one approach to removing these two-sided restrictions has been to consider a different functional framework for the a posteriori error analysis, namely by estimating the $L^2(H^1) \cap H^1(H^{-1})$-norm of the error. Part of the justification of this approach is to be found in the observation in [Verfürth(2003), p. 198, Par. (5)], showing that the estimators of [Picasso(1998), Verfürth(1998)] are upper bounds to not only the $L^2(H^1)$-norm of the error, but also the $L^2(H^1) \cap H^1(H^{-1})$-norm of the error, up to data oscillation. It was then shown in [Verfürth(2003)] that these estimators are efficient, locally-in-time yet only globally-in-space, with respect to the $L^2(H^1) \cap H^1(H^{-1})$-norm of the error, without requiring conditions between mesh and time-step sizes; see also [Bergam et al.(2005)]. Given that the estimators used in both frameworks are the same up to data oscillation, it is of course natural that more general efficiency results are obtainable when including the $H^1(H^{-1})$ part of the norm, since it allows for the appearance of additional terms on the right-hand side in the efficiency bounds.

Recently, we developed in [Ern et al.(2017b)] a posteriori error estimators, based on equilibrated fluxes, for arbitrary order discretizations of parabolic problems within the $L^2(H^1) \cap H^1(H^{-1})$-norm setting, that are guaranteed, locally efficient, and robust. In particular, the analysis does not require any coupling between mesh and time-step sizes, and overcomes the problem of obtaining local-in-space and local-in-time efficiency by considering a natural extension of the $L^2(H^1) \cap H^1(H^{-1})$-norm to the time-nonconforming approximation space. The estimators are robust not only with respect to the mesh and time-step sizes, but also with respect to the polynomial degrees in space and time, and also with respect to mesh coarsening and refinement, thereby removing the so-called transition conditions previously encountered in [Verfürth(2003)]. These results are built upon the analysis for elliptic problems in [Braess et al.(2009), Ern & Vohralík(2010), Ern & Vohralík(2015), Ern & Vohralík(2016)]. We refer to [Dolejší et al.(2016)] for numerical experiments showing the performance of these estimators in practice.

In this work, we present a posteriori error estimates for the $L^2(H^1)$-norm of the error, which are based on the same locally computable equilibrated flux as in [Ern et al.(2017b)], thereby showing that the same methodology can be used in the $L^2(H^1)$-norm estimates as for the $L^2(H^1) \cap H^1(H^{-1})$-norm. Our main contributions, presented in Theorem 5.1 in section 5 below, include guaranteed upper bounds for the $L^2(H^1)$-norm of the error, and local-in-space-and-time lower bounds for the spatial estimator under the one-sided condition $h^2 \lesssim \tau$. We therefore remove the need for the two-sided conditions encountered previously, and we note that the assumptions in [Picasso(1998), Verfürth(1998)] were stronger than our assumption. We emphasize that the regime where $h^2 \lesssim \tau$ is of practical interest in computations, since implicit methods offer the possibility for large time-steps. This condition connecting spatial and temporal resolutions is apparently related to the localisation of the dual norm: the error in the $L^2(H^1)$-norm is connected to the dual norm of the residual for a space of test functions in $L^2(H^1) \cap H^1(H^{-1})$ featuring regularity across both time and space, as shown in Theorem 2.1 below. Therefore the dual norm of the residual is not trivially localized in time. In comparison, for estimates of the error in the $L^2(H^1) \cap H^1(H^{-1})$-norm, the corresponding dual norm of the residual does localize in time because the test space there is $L^2(H^1)$, see, e.g., [Ern et al.(2017b), Theorem 2.1]. It is still unclear whether the condition $h^2 \lesssim \tau$ is really necessary or just technical. It is, however, worth noting that the recent adaptive algorithm from [Gaspoz et al.(2016)] guarantees a uniform lower-bound for the time-step sizes and subordinates the spatial approximation to the temporal indicators; therefore our assumption is not necessarily restrictive in an adaptive context. Our lower bound is similar to [Picasso(1998)] in at least one respect, namely that the right-hand side of our lower bound includes the temporal jump estimator, since it does not appear possible to show in general that this estimator is locally bounded from above by the $L^2(H^1)$-norm of the error. Furthermore, we show that the constant of the lower bound is robust with respect to the spatial polynomial degree, and is also robust with respect to refinement and coarsening of the meshes, thereby allowing us to remove the so-called transition conditions. We also show that our results imply local-in-space and local-in-time efficiency when considered in the framework of the augmented norms that were proposed in [Akrivis et al.(2009), Makridakis & Nochetto(2006), Schötzau & Wihler(2010)].

Our analysis rests upon the following key ingredients. First, in section 2, we present the inf-sup identity which relates the $L^2(H^1)$-norm of the error to an appropriate dual norm of the residual on test functions in a subspace of $L^2(H^1) \cap H^1(H^{-1})$. After setting the notation for the class of finite element methods in section 3, we recall the construction of the equilibrated flux from [Ern et al.(2017b)] in section 4. We state the main results in section 5. In section 6 we use the inf-sup framework to prove the guaranteed upper bounds and the proof of the lower bounds is the subject of section 7. It is based on the combination of two key ideas. The first is to take advantage of the semi-discreteness in time of the test functions appearing in the fundamental efficiency result of [Ern et al.(2017b), Lemma 8.2] in order to gain control over a negative norm on the time derivatives of the test functions; see Lemma 7.2 below. The second idea is to appeal to a specific pointwise-in-time identity for the discontinuous Galerkin time-stepping method, see Lemma 7.3 below. Thus, we employ the definition of the numerical scheme for proving the lower bounds, which is somewhat unusual for a posteriori error analysis.

The combination of these two ideas then yields the lower bounds stated in section 5 under the relaxed hypothesis that $h^2 \lesssim \tau$ only.

Throughout this paper, the notation $a \lesssim b$ means that $a \leq Cb$, with a generic constant C that depends possibly on the shape-regularity of the spatial meshes and the space dimension d, but is otherwise independent of the mesh-size, time-step size, as well as the spatial and temporal polynomial degrees, or on refinement and coarsening between time-steps.

2 Inf-sup theory

Recall that $\Omega \subset \mathbb{R}^d$, $1 \leq d \leq 3$ is a bounded, connected, polyhedral open set with Lipschitz boundary. For an arbitrary open subset $\omega \subset \Omega$, we use $(\cdot,\cdot)_\omega$ to denote the L^2-inner product for scalar- or vector-valued functions on ω, with associated norm $\|\cdot\|_\omega$. In the special case where $\omega = \Omega$, we drop the subscript notation, i.e. $\|\cdot\| \coloneqq \|\cdot\|_\Omega$.

The starting point of the analysis is the weak formulation of problem (1.1) where the time derivative has been cast onto a test function, using integration by parts in time. In particular, the solution space X and test space Y_T are defined by

$$\begin{aligned} X &\coloneqq L^2(0,T;H^1_0(\Omega)),\\ Y_T &\coloneqq \{\varphi \in L^2(0,T;H^1_0(\Omega)) \cap H^1(0,T;H^{-1}(\Omega)), \varphi(T) = 0\}. \end{aligned} \tag{2.1}$$

The spaces X and Y_T are equipped with the norms

$$\begin{aligned} \|v\|_X^2 &\coloneqq \int_0^T \|\nabla v\|^2 \,\mathrm{d}t && \forall\, v \in X,\\ \|\varphi\|_{Y_T}^2 &\coloneqq \int_0^T \|\partial_t \varphi\|^2_{H^{-1}(\Omega)} + \|\nabla \varphi\|^2 \,\mathrm{d}t + \|\varphi(0)\|^2 && \forall\, \varphi \in Y_T. \end{aligned} \tag{2.2}$$

Let the bilinear form $\mathcal{B}_X \colon X \times Y_T \to \mathbb{R}$ be defined by

$$\mathcal{B}_X(v,\varphi) \coloneqq \int_0^T -\langle \partial_t \varphi, v\rangle + (\nabla v, \nabla \varphi)\,\mathrm{d}t \quad \forall\, v \in X,\ \varphi \in Y_T, \tag{2.3}$$

where $\langle\cdot,\cdot\rangle$ denotes the duality pairing between $H^{-1}(\Omega)$ and $H^1_0(\Omega)$. Then, problem (1.1) admits the following weak formulation: find $u \in X$ such that

$$\mathcal{B}_X(u,\varphi) = \int_0^T (f,\varphi)\,\mathrm{d}t + (u_0, \varphi(0)) \quad \forall\, \varphi \in Y_T. \tag{2.4}$$

The well-posedness of (2.4) is well-known and can be shown by Galerkin's method, see for instance the textbook by [Wloka(1987)]. Note that in this weak formulation, the initial condition $u(0) = u_0$ is expressed as a natural condition, appearing in (2.4), rather than as an essential condition imposed by the choice of solution space.

The following result states an inf–sup stability result for the bilinear form $\mathcal{B}_X$. This inf–sup stability result has the interesting and important property of taking the form of an identity, which is advantageous for the sharpness of a posteriori error analysis, and shows that the choice of norms for the spaces X and Y_T in (2.2) above are optimal. We refer the reader to [Tantardini & Veeser(2016)] for further results on the inf-sup theory of parabolic problems.

Theorem 2.1 (Inf–sup identity)**.** *For every $v \in X$, we have*

$$\|v\|_X = \sup_{\varphi \in Y_T \setminus \{0\}} \frac{\mathcal{B}_X(v,\varphi)}{\|\varphi\|_{Y_T}}. \tag{2.5}$$

Proof. The arguments in the proof of [Ern et al.(2017b), Theorem 2.1] can be used to show the following inf-sup identity: for any $\varphi \in Y_T$, we have

$$\|\varphi\|_{Y_T} = \sup_{v \in X \setminus \{0\}} \frac{\mathcal{B}_X(v,\varphi)}{\|v\|_X}. \tag{2.6}$$

So, (2.6) immediately implies the lower bound $\|v\|_X \geq \sup_{\varphi \in Y_T \setminus \{0\}} \mathcal{B}_X(v,\varphi)/\|\varphi\|_{Y_T}$ for any fixed $v \in X$. To obtain the converse bound, let $\varphi_* \in Y_T$ denote the solution of $\mathcal{B}_X(w,\varphi_*) = \int_0^T (\nabla w, \nabla v)\,\mathrm{d}t$ for all $w \in X$. This problem can simply be seen as a backward-in-time parabolic problem with final time condition $\varphi_*(T) = 0$. Hence, we have $\|v\|_X^2 = \mathcal{B}_X(v,\varphi_*)$ and (2.6) implies that $\|\varphi_*\|_{Y_T} = \|v\|_X$. This immediately shows that $\|v\|_X \leq \sup_{\varphi \in Y_T \setminus \{0\}} \mathcal{B}_X(v,\varphi)/\|\varphi\|_{Y_T}$, and completes the proof of (2.5). □

In order to estimate the error between the solution u of (1.1) and its approximation, we define the residual functional $\mathcal{R}_X \colon X \to [Y_T]'$ by

$$\langle \mathcal{R}_X(v), \varphi \rangle_{[Y_T]' \times Y_T} \coloneqq \mathcal{B}_X(u - v, \varphi) = \int_0^T (f,\varphi) + \langle \partial_t \varphi, v \rangle - (\nabla v, \nabla \varphi)\,\mathrm{d}t + (u_0, \varphi(0)), \tag{2.7}$$

where $v \in X$ and $\varphi \in Y_T$, and where the equality follows simply from (2.4). The dual norm of the residual $\|\mathcal{R}_X(v)\|_{[Y_T]'}$ is naturally defined by

$$\|\mathcal{R}_X(v)\|_{[Y_T]'} \coloneqq \sup_{\varphi \in Y_T \setminus \{0\}} \frac{\langle \mathcal{R}_X(v), \varphi \rangle}{\|\varphi\|_{Y_T}}. \tag{2.8}$$

Theorem 2.1 implies the following equivalence between the error and dual norm of the residual:

$$\|u - v\|_X = \|\mathcal{R}_X(v)\|_{[Y_T]'} \quad \forall\, v \in X. \tag{2.9}$$

Remark 2.1. Problem (1.1) admits an alternative weak formulation where the test space is X and the trial space is $L^2(0,T;H_0^1(\Omega)) \cap H^1(0,T;H^{-1}(\Omega))$. We refer the reader to [Ern et al.(2017b)] and the references therein for further details. The solution of the problem remains the same for the two weak formulations, although each weak formulation is tied to a different inf-sup condition that relates the norm of the error and of the residual, exactly in the way (2.5) and (2.6) interplay.

3 Finite element approximation

The time interval $(0, T)$ is partitioned into sub-intervals $I_n := (t_{n-1}, t_n)$, with $1 \leq n \leq N$, where it is assumed that $[0, T] = \bigcup_{n=1}^{N} \overline{I_n}$, and that $\{t_n\}_{n=0}^{N}$ is strictly increasing with $t_0 = 0$ and $t_N = T$. For each interval I_n, we let $\tau_n := t_n - t_{n-1}$ denote the local time-step size. No special assumptions are made about the relative sizes of the time-steps to each other. A temporal polynomial degree $q_n \geq 0$ is associated with each time-step I_n, and we gather all the polynomial degrees in the vector $\boldsymbol{q} = (q_n)_{n=1}^{N}$. For a general vector space V, we shall write $\mathcal{Q}_{q_n}(I_n; V)$ to denote the space of V-valued univariate polynomials of degree at most q_n over the time-step interval I_n.

3.1 Meshes

For each $0 \leq n \leq N$, let $\mathcal{T}^n$ denote a matching simplicial mesh of the domain Ω, where we assume shape-regularity of the meshes uniformly with respect to n. We consider here only matching simplicial meshes for simplicity, although we indicate that mixed simplicial–parallelepipedal meshes, possibly containing hanging nodes, can also be treated: see [Dolejší et al.(2016)] for instance. The mesh $\mathcal{T}^0$ will be used to approximate the initial datum u_0. For each element $K \in \mathcal{T}^n$, let $h_K := \operatorname{diam} K$ denote the diameter of K. We associate a local spatial polynomial degree $p_K \geq 1$ with each $K \in \mathcal{T}^n$, and we gather all spatial polynomial degrees in the vector $\boldsymbol{p}_n = (p_K)_{K \in \mathcal{T}^n}$. In order to keep our notation sufficiently simple, the dependence of the local spatial polynomial degrees p_K on the time-step is kept implicit, although we bear in mind that the polynomial degrees may change between time-steps.

3.2 Approximation spaces

Given a general matching simplicial mesh $\mathcal{T}$ and given a vector of polynomial degrees $\boldsymbol{p} = (p_K)_{K \in \mathcal{T}}$, $p_K \geq 1$ for all $K \in \mathcal{T}$, we define the $H_0^1(\Omega)$-conforming hp-finite element space $V_h(\mathcal{T}, \boldsymbol{p})$ by

$$V_h(\mathcal{T}, \boldsymbol{p}) := \left\{ v_h \in H_0^1(\Omega),\ v_h|_K \in \mathcal{P}_{p_K}(K) \quad \forall K \in \mathcal{T} \right\}, \tag{3.1}$$

where $\mathcal{P}_{p_K}(K)$ denotes the space of polynomials of total degree at most p_K on K. To shorten the notation, let $V^n := V_h(\mathcal{T}^n, \boldsymbol{p}_n)$ for each $0 \leq n \leq N$. Let $\Pi_h u_0 \in V^0$ denote an approximation to the initial datum u_0, a typical choice being the L^2-orthogonal projection of u_0 onto V^0. Given the collection of time intervals $\{I_n\}_{n=1}^{N}$, the vector $\boldsymbol{q}$ of temporal polynomial degrees, and the hp-finite element spaces $\{V^n\}_{n=0}^{N}$, the finite element space $V_{h\tau}$ is defined by

$$V_{h\tau} := \left\{ v_{h\tau}|_{(0,T)} \in X,\ v_{h\tau}|_{I_n} \in \mathcal{Q}_{q_n}(I_n; V^n) \quad \forall\, n = 1, \ldots, N,\ v_{h\tau}(0) \in V^0 \right\}. \tag{3.2}$$

Functions in $V_{h\tau}$ are generally discontinuous with respect to the time-variable at the temporal partition points. We take them to be left-continuous: for all $1 \leq n \leq N$, we define $v_{h\tau}(t_n)$ as the trace at t_n of the restriction $v_{h\tau}|_{I_n}$. Moreover, functions in $V_{h\tau}$

also have a well-defined value at $t_0 = 0$. For all $0 \le n < N$, we denote the right-limit of $v_{h\tau} \in V_{h\tau}$ at t_n by $v_{h\tau}(t_n^+)$. Then, the temporal jump operators $(\!|\cdot|\!)_n$ are defined by

$$(\!|v_{h\tau}|\!)_n \coloneqq v_{h\tau}(t_n) - v_{h\tau}(t_n^+), \quad 0 \le n \le N-1. \tag{3.3}$$

3.3 Refinement and coarsening

Similarly to other works, e.g., [Verfürth(2003), p. 196], we assume that we have at our disposal a common refinement mesh $\widetilde{\mathcal{T}^n}$ of $\mathcal{T}^{n-1}$ and $\mathcal{T}^n$ for each $1 \le n \le N$, as well as associated polynomial degrees $\widetilde{\boldsymbol{p}}_n = (p_{\widetilde{K}})_{\widetilde{K}\in\widetilde{\mathcal{T}^n}}$, such that $V^{n-1} + V^n \subset \widetilde{V^n} \coloneqq V_h(\widetilde{\mathcal{T}^n}, \widetilde{\boldsymbol{p}}_n)$. For a function $v_{h\tau} \in V_{h\tau}$, we observe that $(\!|v_{h\tau}|\!)_{n-1} \in \widetilde{V^n}$ for each $1 \le n \le N$ since $v_{h\tau}(t_{n-1}) \in V^{n-1}$, $v_{h\tau}(t_{n-1}^+) \in V^n$, and $V^{n-1} + V^n \subset \widetilde{V^n}$. It is assumed that the shape-regularity of $\widetilde{\mathcal{T}^n}$ is equivalent up to uniform constants to those of $\mathcal{T}^{n-1}$ and $\mathcal{T}^n$, and that every element $\widetilde{K} \in \widetilde{\mathcal{T}^n}$ is wholly contained in a single element $K' \in \mathcal{T}^{n-1}$ and a single element $K'' \in \mathcal{T}^n$. We emphasize that we do not require any assumptions on the relative coarsening or refinement between successive spaces V^{n-1} and V^n. In particular, we do not need the transition condition from [Verfürth(2003), p. 196, 201], which requires a uniform bound on the ratio of element sizes between $\widetilde{\mathcal{T}^n}$ and $\mathcal{T}^n$. In practice, we expect that most adaptive algorithms will obtain each mesh $\mathcal{T}^n$ from an initial coarse mesh, with coarsening/refinements between two successive meshes, using a standard algorithm such as newest vertex bisection. We refer the reader to [Nochetto et al.(2009)] for a discussion in this context. To derive our results, we note that we do not need to restrict ourselves to any particular refinement algorithm.

3.4 Numerical method

The numerical scheme consists of finding $u_{h\tau} \in V_{h\tau}$ such that $u_{h\tau}(0) = \Pi_h u_0$, and such that

$$\int_{I_n} (\partial_t u_{h\tau}, v_{h\tau}) + (\nabla u_{h\tau}, \nabla v_{h\tau})\,\mathrm{d}t - \left((\!|u_{h\tau}|\!)_{n-1}, v_{h\tau}(t_{n-1}^+)\right) = \int_{I_n} (f, v_{h\tau})\,\mathrm{d}t \tag{3.4}$$

for all test functions $v_{h\tau} \in \mathcal{Q}_{q_n}(I_n; V^n)$ and for each time-step interval I_n, $n = 1, \ldots, N$. Here the time derivative $\partial_t u_{h\tau}$ is understood as the piecewise time-derivative on each time-step interval I_n. The numerical solution $u_{h\tau} \in V_{h\tau}$ can thus be obtained by solving the fully discrete problem (3.4) on each successive time-step. At each time-step, this requires solving a linear system that is symmetric only in the case $q_n = 0$; this can be performed efficiently in practice for arbitrary orders following [Smears(2017)]. Note further that the initial condition $u_{h\tau}(0) = \Pi_h u_0$ does not guarantee that the right-limit $u_{h\tau}(0^+)$ should equal $\Pi_h u_0$.

3.5 Reconstruction operator

For each time-step interval I_n and each nonnegative integer q, let L_q^n denote the polynomial on I_n obtained by mapping the q-th Legendre polynomial under an affine transformation of $(-1, 1)$ to I_n. It follows that $L_q^n(t_n) = 1$ for all $q \ge 0$, and $L_q^n(t_{n-1}) = (-1)^q$, and that the mapped Legendre polynomials $\{L_q^n\}_{q\geq 0}$ are L^2-orthogonal on I_n, and satisfy $\int_{I_n}|L_q^n|^2\,\mathrm{d}t = \frac{\tau_n}{2q+1}$ for all $q\geq 0$, see for instance [Schwab(1998), Appendix C]. Following [Makridakis & Nochetto(2006)] (see also [Smears(2017), Remark 2.3]), we introduce the reconstruction operator $\mathcal{I}$ defined on $V_{h\tau}$ by

$$(\mathcal{I}v_{h\tau})|_{I_n} \coloneqq v_{h\tau}|_{I_n} + \frac{(-1)^{q_n}}{2}\left(L_{q_n}^n - L_{q_n+1}^n\right)(\!|v_{h\tau}|\!)_{n-1} \quad \forall\, v_{h\tau}\in V_{h\tau}. \tag{3.5}$$

It is clear that $\mathcal{I}$ is a linear operator on $V_{h\tau}$. Furthermore, the definition ensures that $\mathcal{I}v_{h\tau}|_{I_n}(t_n) = v_{h\tau}(t_n)$, and that $\mathcal{I}v_{h\tau}|_{I_n}(t_{n-1}^+) = v_{h\tau}(t_{n-1})$ for all $1\leq n\leq N$. This implies that $\mathcal{I}v_{h\tau}$ is continuous with respect to the temporal variable at the interval partition points $\{t_n\}_{n=0}^{N-1}$ and hence $\mathcal{I}v_{h\tau}\in H^1(0,T;H_0^1(\Omega))$. Furthermore, $\mathcal{I}v_{h\tau}|_{I_n}\in \mathcal{Q}_{q_n+1}\big(I_n;\widetilde{V^n}\big)$ for any $v_{h\tau}\in V_{h\tau}$, where we recall that $V^{n-1}+V^n\subset \widetilde{V^n}$. It is well-known from [Ern & Schieweck(2016), Makridakis & Nochetto(2006), Smears(2017)] that we may rewrite the numerical scheme (3.4) as

$$\int_{I_n}(\partial_t\mathcal{I}u_{h\tau}, v_{h\tau}) + (\nabla u_{h\tau},\nabla v_{h\tau})\,\mathrm{d}t = \int_{I_n}(f, v_{h\tau})\,\mathrm{d}t \quad \forall\, v_{h\tau}\in\mathcal{Q}_{q_n}(I_n;V^n). \tag{3.6}$$

Note also that $\mathcal{I}u_{h\tau}(0) = \Pi_h u_0$.

4 Construction of the equilibrated flux

The a posteriori error estimates presented in this paper are based on a discrete and locally computable $\boldsymbol{H}(\mathrm{div})$-conforming flux $\boldsymbol{\sigma}_{h\tau}$ that satisfies the key equilibration property

$$\partial_t\mathcal{I}u_{h\tau} + \nabla\cdot\boldsymbol{\sigma}_{h\tau} = f_{h\tau} \quad \text{in } \Omega\times(0,T), \tag{4.1}$$

where $\mathcal{I}u_{h\tau}$ is defined in section 3.5, and $f_{h\tau}\approx f$ is an approximation of the data that is defined in (4.4) below. We call $\boldsymbol{\sigma}_{h\tau}$ an equilibrated flux. The construction of $\boldsymbol{\sigma}_{h\tau}$ given here is exactly the same as in [Ern et al.(2017b)]. This has the practical benefit that a single construction of the equilibrated flux can be used for both a posteriori error estimates in the $L^2(H^1)\cap H^1(H^{-1})$-norm and also in the $L^2(H^1)$-norm.

4.1 Local mixed finite element spaces

For each $1\leq n\leq N$, let $\mathcal{V}^n$ denote the set of vertices of the mesh $\mathcal{T}^n$, where we distinguish the set of interior vertices $\mathcal{V}_{\mathrm{int}}^n$ and the set of boundary vertices $\mathcal{V}_{\mathrm{ext}}^n$. For each $\boldsymbol{a}\in\mathcal{V}^n$, let $\psi_{\boldsymbol{a}}$ denote the hat function associated with $\boldsymbol{a}$, and let $\omega_{\boldsymbol{a}}$ denote the interior of the support of $\psi_{\boldsymbol{a}}$, with associated diameter $h_{\omega_{\boldsymbol{a}}}$. Furthermore, let $\widetilde{\mathcal{T}^{\boldsymbol{a}}}$ denote the restriction of the mesh $\widetilde{\mathcal{T}^n}$ to $\omega_{\boldsymbol{a}}$. Recalling that the common refinement spaces $\widetilde{V^n}$ were obtained with a vector of polynomial degrees $\widetilde{\boldsymbol{p}}_n = (p_{\widetilde{K}})_{\widetilde{K}\in\widetilde{\mathcal{T}^n}}$, we associate with each $\boldsymbol{a}\in\mathcal{V}^n$ the fixed polynomial degree

$$p_{\boldsymbol{a}} \coloneqq \max_{\widetilde{K}\in\widetilde{\mathcal{T}^{\boldsymbol{a}}}}(p_{\widetilde{K}}+1). \tag{4.2}$$

For a polynomial degree $p \geq 0$, let the piecewise polynomial (discontinuous) spaces $\mathcal{P}_p(\widetilde{\mathcal{T}^{\boldsymbol{a}}})$ and $\boldsymbol{RTN}_p(\widetilde{\mathcal{T}^{\boldsymbol{a}}})$ be defined by

$$\begin{aligned}\mathcal{P}_p(\widetilde{\mathcal{T}^{\boldsymbol{a}}}) &:= \{q_h \in L^2(\omega_{\boldsymbol{a}}), \quad q_h|_{\widetilde{K}} \in \mathcal{P}_p(\widetilde{K}) \quad \forall\, \widetilde{K} \in \widetilde{\mathcal{T}^{\boldsymbol{a}}}\},\\ \boldsymbol{RTN}_p(\widetilde{\mathcal{T}^{\boldsymbol{a}}}) &:= \{\boldsymbol{v}_h \in \boldsymbol{L}^2(\omega_{\boldsymbol{a}}), \quad \boldsymbol{v}_h|_{\widetilde{K}} \in \boldsymbol{RTN}_p(\widetilde{K}) \quad \forall \widetilde{K} \in \widetilde{\mathcal{T}^{\boldsymbol{a}}}\},\end{aligned}$$

where $\boldsymbol{RTN}_p(\widetilde{K}) := \boldsymbol{\mathcal{P}}_p(\widetilde{K}) + \mathcal{P}_p(\widetilde{K})\boldsymbol{x}$ denotes the Raviart–Thomas–Nédélec space of order p on the simplex $\widetilde{K}$. It is important to notice that whereas the patch $\omega_{\boldsymbol{a}}$ is subordinate to the elements of the mesh $\mathcal{T}^n$ around the vertex $\boldsymbol{a} \in \mathcal{V}^n$, the spaces $\mathcal{P}_p(\widetilde{\mathcal{T}^{\boldsymbol{a}}})$ and $\boldsymbol{RTN}_p(\widetilde{\mathcal{T}^{\boldsymbol{a}}})$ are subordinate to the submesh elements in $\widetilde{\mathcal{T}^{\boldsymbol{a}}}$; of course, in the absence of coarsening, this distinction vanishes. We now introduce the local spatial mixed finite element space $\boldsymbol{V}_h^{\boldsymbol{a}}$, defined by

$$\boldsymbol{V}_h^{\boldsymbol{a}} := \begin{cases} \left\{\boldsymbol{v}_h \in \boldsymbol{H}(\mathrm{div}, \omega_{\boldsymbol{a}}) \cap \boldsymbol{RTN}_{p_{\boldsymbol{a}}}(\widetilde{\mathcal{T}^{\boldsymbol{a}}}),\ \boldsymbol{v}_h \cdot \boldsymbol{n} = 0 \text{ on } \partial\omega_{\boldsymbol{a}}\right\} & \text{if } \boldsymbol{a} \in \mathcal{V}^n_{\mathrm{int}},\\ \left\{\boldsymbol{v}_h \in \boldsymbol{H}(\mathrm{div}, \omega_{\boldsymbol{a}}) \cap \boldsymbol{RTN}_{p_{\boldsymbol{a}}}(\widetilde{\mathcal{T}^{\boldsymbol{a}}}),\ \boldsymbol{v}_h \cdot \boldsymbol{n} = 0 \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \partial\Omega\right\} & \text{if } \boldsymbol{a} \in \mathcal{V}^n_{\mathrm{ext}}.\end{cases}$$

We then define the space-time mixed finite element space

$$\boldsymbol{V}_{h\tau}^{\boldsymbol{a},n} := \mathcal{Q}_{q_n}(I_n; \boldsymbol{V}_h^{\boldsymbol{a}}), \tag{4.3}$$

where we recall that $\mathcal{Q}_{q_n}(I_n; \boldsymbol{V}_h^{\boldsymbol{a}})$ denotes the space of $\boldsymbol{V}_h^{\boldsymbol{a}}$-valued univariate polynomials of degree at most q_n over the time-step interval I_n.

4.2 Data approximation

Our a posteriori error estimates given in section 5 involve certain approximations of the source term f appearing in (1.1). It is helpful to define these approximations here. For each $1 \leq n \leq N$ and for each $\boldsymbol{a} \in \mathcal{V}^n$, let $\Pi_{h\tau}^{\boldsymbol{a},n}$ be the $L^2_{\psi_{\boldsymbol{a}}}$-orthogonal projection from $L^2(I_n; L^2_{\psi_{\boldsymbol{a}}}(\omega_{\boldsymbol{a}}))$ onto $\mathcal{Q}_{q_n}(I_n; \mathcal{P}_{p_{\boldsymbol{a}}-1}(\widetilde{\mathcal{T}^{\boldsymbol{a}}}))$, where $L^2_{\psi_{\boldsymbol{a}}}(\omega_{\boldsymbol{a}})$ is the space of measurable functions v on $\omega_{\boldsymbol{a}}$ such that $\int_{\omega_{\boldsymbol{a}}} \psi_{\boldsymbol{a}} |v|^2 \,\mathrm{d}x < \infty$. In other words, the projection operator $\Pi_{h\tau}^{\boldsymbol{a},n}$ is defined by $\int_{I_n} (\psi_{\boldsymbol{a}} \Pi_{h\tau}^{\boldsymbol{a},n} v, q_{h\tau})_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t = \int_{I_n} (\psi_{\boldsymbol{a}} v, q_{h\tau})_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t$ for all $q_{h\tau} \in \mathcal{Q}_{q_n}(I_n; \mathcal{P}_{p_{\boldsymbol{a}}-1}(\widetilde{\mathcal{T}^{\boldsymbol{a}}}))$. We adopt the convention that $\Pi_{h\tau}^{\boldsymbol{a},n} v$ is extended by zero from $\omega_{\boldsymbol{a}} \times I_n$ to $\Omega \times (0,T)$ for all $v \in L^2(I_n; L^2_{\psi_{\boldsymbol{a}}}(\omega_{\boldsymbol{a}}))$. Then, we define $f_{h\tau}$ by

$$f_{h\tau} := \sum_{n=1}^{N} \sum_{\boldsymbol{a} \in \mathcal{V}^n} \psi_{\boldsymbol{a}} \, \Pi_{h\tau}^{\boldsymbol{a},n} f. \tag{4.4}$$

See [Ern et al.(2017b)] for further remarks concerning the approximation properties of $f_{h\tau}$. In particular, it is shown there that $f_{h\tau}$ is a data approximation that is at least of the same order as the one used in the numerical scheme (3.4).

4.3 Flux reconstruction

For each $1 \le n \le N$ and each $\boldsymbol{a} \in \mathcal{V}^n$, let the scalar function $g_{h\tau}^{\boldsymbol{a},n} \in \mathcal{Q}_{q_n}(I_n; \mathcal{P}_{p_{\boldsymbol{a}}}(\widetilde{\mathcal{T}^{\boldsymbol{a}}}))$ and vector field $\boldsymbol{\tau}_{h\tau}^{\boldsymbol{a},n} \in \mathcal{Q}_{q_n}(I_n; \boldsymbol{RTN}_{p_{\boldsymbol{a}}}(\widetilde{\mathcal{T}^{\boldsymbol{a}}}))$ be defined by

$$\boldsymbol{\tau}_{h\tau}^{\boldsymbol{a},n} := \psi_{\boldsymbol{a}} \nabla u_{h\tau}|_{\omega_{\boldsymbol{a}} \times I_n}, \tag{4.5a}$$

$$g_{h\tau}^{\boldsymbol{a},n} := \psi_{\boldsymbol{a}} \left(\Pi_{h\tau}^{\boldsymbol{a},n} f - \partial_t \mathcal{I} u_{h\tau} \right)|_{\omega_{\boldsymbol{a}} \times I_n} - \nabla \psi_{\boldsymbol{a}} \cdot \nabla u_{h\tau}|_{\omega_{\boldsymbol{a}} \times I_n}. \tag{4.5b}$$

For interior vertices, the numerical scheme (3.6) implies that

$$(g_{h\tau}^{\boldsymbol{a},n}(t), 1)_{\omega_{\boldsymbol{a}}} = 0 \quad \forall\, t \in I_n. \tag{4.6}$$

Definition 4.1 (Flux reconstruction)**.** *Let $u_{h\tau} \in V_{h\tau}$ be the numerical solution of (3.4). For each time-step interval I_n and for each vertex $\boldsymbol{a} \in \mathcal{V}$, let the space $\boldsymbol{V}_{h\tau}^{\boldsymbol{a},n}$ be defined by (4.3). Let $g_{h\tau}^{\boldsymbol{a},n}$ and $\boldsymbol{\tau}_{h\tau}^{\boldsymbol{a},n}$ be defined by (4.5). Let $\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n} \in \boldsymbol{V}_{h\tau}^{\boldsymbol{a},n}$ be defined by*

$$\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n} := \underset{\substack{\boldsymbol{v}_{h\tau} \in \boldsymbol{V}_{h\tau}^{\boldsymbol{a},n} \\ \nabla \cdot \boldsymbol{v}_{h\tau} = g_{h\tau}^{\boldsymbol{a},n}}}{\operatorname{argmin}} \int_{I_n} \| \boldsymbol{v}_{h\tau} + \boldsymbol{\tau}_{h\tau}^{\boldsymbol{a},n} \|_{\omega_{\boldsymbol{a}}}^2 \,\mathrm{d}t. \tag{4.7}$$

Then, after extending $\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n}$ by zero from $\omega_{\boldsymbol{a}} \times I_n$ to $\Omega \times (0,T)$ for each $\boldsymbol{a} \in \mathcal{V}$ and for each $1 \le n \le N$, we define

$$\boldsymbol{\sigma}_{h\tau} := \sum_{n=1}^{N} \sum_{\boldsymbol{a} \in \mathcal{V}^n} \boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n}. \tag{4.8}$$

Note that $\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n} \in \boldsymbol{V}_{h\tau}^{\boldsymbol{a},n}$ is well-defined for all $\boldsymbol{a} \in \mathcal{V}^n$: in particular, for interior vertices $\boldsymbol{a} \in \mathcal{V}_{\mathrm{int}}^n$, we use (4.6) to guarantee the compatibility of the datum $g_{h\tau}^{\boldsymbol{a},n}$ with the constraint $\nabla \cdot \boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n} = g_{h\tau}^{\boldsymbol{a},n}$. The following key result is quoted from [Ern et al.(2017b)].

Theorem 4.2 (Equilibration)**.** *Let the flux reconstruction $\boldsymbol{\sigma}_{h\tau}$ be given by Definition 4.1, and let $f_{h\tau}$ be defined in (4.4). Then $\boldsymbol{\sigma}_{h\tau} \in L^2(0,T; \boldsymbol{H}(\mathrm{div}))$ and the equilibration identity (4.1) holds.*

Moreover, for the purpose of implementation, it is known that on each patch of the mesh and at each time-step, the solution of the minimization problem (4.7) decouples into q_n+1 independent spatial mixed finite element linear systems, which helps to reduce the cost of computing the flux $\boldsymbol{\sigma}_{h\tau}$.

5 Main results

We introduce the following a posteriori error estimators and data oscillation terms:

$$[\eta^n_{\mathrm{F},K}]^2 := \int_{I_n} \|\boldsymbol{\sigma}_{h\tau} + \nabla u_{h\tau}\|_K^2 \,\mathrm{d}t, \tag{5.1a}$$

$$[\eta^n_{\mathrm{J},K}]^2 := \int_{I_n} \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|_K^2 \,\mathrm{d}t, \tag{5.1b}$$

$$[\eta^n_{\mathrm{osc},h\tau}]^2 := \frac{1+\sqrt{2}}{2} \int_{I_n} \sum_{\widetilde{K}\in\widetilde{\mathcal{T}}^n} \left[\frac{\tau_n}{\pi} + \frac{h_{\widetilde{K}}^2}{\pi^2}\right] \|f - f_{h\tau}\|_{\widetilde{K}}^2 \,\mathrm{d}t, \tag{5.1c}$$

$$\eta_{\mathrm{osc,init}} := \|u_0 - \Pi_h u_0\|, \tag{5.1d}$$

where, $K \in \mathcal{T}^n$, $1 \leq n \leq N$, the equilibrated flux $\boldsymbol{\sigma}_{h\tau}$ is prescribed in Definition 4.1, and where the data approximation $f_{h\tau}$ is defined in section 4.2. The total estimator for the error is defined by

$$[\eta_X]^2 := \sum_{n=1}^{N} \left\{ \left[\sum_{K\in\mathcal{T}^n} \left\{ [\eta^n_{\mathrm{F},K}]^2 + [\eta^n_{\mathrm{J},K}]^2 \right\} \right]^{\frac{1}{2}} + \eta^n_{\mathrm{osc},h\tau} \right\}^2 + [\eta_{\mathrm{osc,init}}]^2. \tag{5.2}$$

The flux estimator $\eta^n_{\mathrm{F},K}$ and the temporal jump estimator $\eta^n_{\mathrm{J},K}$ are the two main estimators. In particular, the flux estimator $\eta^n_{\mathrm{F},K}$ measures the lack of $\boldsymbol{H}(\mathrm{div})$-conformity of $\nabla u_{h\tau}$, and the temporal jump estimator $\eta^n_{\mathrm{J},K}$ measures the lack of temporal conformity of $u_{h\tau}$. Indeed, $\eta^n_{\mathrm{J},K}$ is related to the jump $(\!|u_{h\tau}|\!)_{n-1}$, since it was shown in [Schötzau & Wihler(2010), Ern et al.(2017b)] that $\eta^n_{\mathrm{J},K}$ can be equivalently rewritten as

$$\eta^n_{\mathrm{J},K} = \sqrt{\tfrac{\tau_n(q_n+1)}{(2q_n+1)(2q_n+3)}}\, \|\nabla (\!|u_{h\tau}|\!)_{n-1}\|_K. \tag{5.3}$$

Given that $\eta^n_{\mathrm{F},K}$ and $\eta^n_{\mathrm{J},K}$ respectively measure the lack of spatial and temporal conformity of the approximate solution, it is common in the literature to call $\eta^n_{\mathrm{F},K}$ the spatial estimator and $\eta^n_{\mathrm{J},K}$ the temporal estimator. However, such terminology must not be interpreted as stating that these estimators bound the errors due respectively to the spatial and temporal discretization. In practice, these estimators can be computed by quadrature on the common refinement mesh $\widetilde{\mathcal{T}}^n$. Note that it is possible to split $\eta^n_{\mathrm{J},K}$ into further components, for instance to quantify the effect of coarsening, although this is not strictly necessary for the purposes of the upper and lower bounds on the error to be given below, which is why $\eta^n_{\mathrm{J},K}$ is given in its current form.

Theorem 5.1 (X-norm a posteriori error estimate)**.** *Let $u \in X$ be the weak solution of (1.1), and let $u_{h\tau} \in V_{h\tau}$ denote the solution of the numerical scheme (3.4). Let η_X be defined by (5.2). Then, we have the following X-norm a posteriori error estimate:*

$$\|u - u_{h\tau}\|_X \leq \eta_X. \tag{5.4}$$

If $K \in \mathcal{T}^n$, $1 \leq n \leq N$, is an element such that $h_{\omega_{\boldsymbol{a}}}^2 \leq \gamma_{\boldsymbol{a}} \tau_n$ for each $\boldsymbol{a} \in \mathcal{V}_K$, with $\mathcal{V}_K$ the set of vertices of the element K, with some constant $\gamma_{\boldsymbol{a}} > 0$, where $h_{\omega_{\boldsymbol{a}}}$ denotes the diameter of the patch $\omega_{\boldsymbol{a}}$, then we have the local lower bound for the flux estimator $\eta_{\mathrm{F},K}^n$

$$[\eta_{\mathrm{F},K}^n]^2 \leq \sum_{\boldsymbol{a} \in \mathcal{V}_K} C_{\gamma_{\boldsymbol{a}},q_n}^2 \left\{ \int_{I_n} \|\nabla(u - u_{h\tau})\|_{\omega_{\boldsymbol{a}}}^2 + \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|_{\omega_{\boldsymbol{a}}}^2 \mathrm{d}t + [\eta_{\mathrm{osc}}^{\boldsymbol{a},n}]^2 \right\}, \quad (5.5)$$

where the local data ocillation $\eta_{\mathrm{osc}}^{\boldsymbol{a},n}$ is defined by

$$[\eta_{\mathrm{osc}}^{\boldsymbol{a},n}]^2 := \int_{I_n} \|f - \Pi_{h\tau}^{\boldsymbol{a},n} f\|_{H^{-1}(\omega_{\boldsymbol{a}})}^2 \mathrm{d}t. \quad (5.6)$$

Furthermore, under the hypothesis that there exists $\gamma > 0$ such that $h_{\omega_{\boldsymbol{a}}}^2 \leq \gamma \tau_n$ for every $\boldsymbol{a} \in \mathcal{V}^n$ and every $1 \leq n \leq N$, then we have the global lower bound

$$\sum_{n=1}^N \sum_{K \in \mathcal{T}^n} [\eta_{\mathrm{F},K}^n]^2 \leq C_{\gamma,q_n}^2 \left\{ \|u - u_{h\tau}\|_X^2 + \|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X^2 + \sum_{n=1}^N \sum_{\boldsymbol{a} \in \mathcal{V}^n} [\eta_{\mathrm{osc}}^{\boldsymbol{a},n}]^2 \right\}. \quad (5.7)$$

The constants $C_{\gamma_{\boldsymbol{a}},q_n}$ in (5.5) and C_{γ,q_n} in (5.7) satisfy $C_{\gamma,q_n} \lesssim (q_n+1)^{\frac{1}{2}} + \gamma(q_n+1)^{\frac{5}{2}}$, and may depend on the shape regularity of $\mathcal{T}^n$ and $\widetilde{\mathcal{T}^n}$ and on the dimension d, but otherwise do not depend on the mesh-size, time-step size, spatial polynomial degrees, or on refinement and coarsening between time-steps.

The proof of Theorem 5.1 is given in several stages throughout the following sections. In the first stage, we give the proof of the upper bound (5.4) immediately after the helpful data oscillation estimate of Lemma 6.2 below in section 6. In the second stage, we show the lower bounds (5.5) and (5.7) in section 7.

Remark 5.1 (Bounds for the jump estimator). In the local lower bound (5.5), we have $\int_{I_n} \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|_{\omega_{\boldsymbol{a}}}^2 \mathrm{d}t = \sum_{K \subset \omega_{\boldsymbol{a}}} [\eta_{\mathrm{J},K}^n]^2$, see also (5.3), where the sum is over all elements K of $\mathcal{T}^n$ contained in $\omega_{\boldsymbol{a}}$. Similarly, in the global lower bound (5.7), the term $\|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X^2 = \sum_{n=1}^N \sum_{K \in \mathcal{T}^n} [\eta_{\mathrm{J},K}^n]^2$ appears. Thus our result here is comparable to those in [Picasso(1998)] where the jump estimator also appears on the right-hand side of the local lower bounds. The reason for the appearance of this term can be essentially traced back to the lack of Galerkin orthogonality for the temporal reconstruction $\mathcal{I}u_{h\tau}$, see (3.6). Though a priori error analysis shows that $\|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X$ converges with the same order as $\|u - u_{h\tau}\|_X$ if u is assumed to have some smoothness, a difficulty is that if the jump estimator $\|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X$ becomes too large compared to $\|u - u_{h\tau}\|_X$, then the meaning of the lower bound (5.7) becomes less clear, and similarly for (5.5). However, we note that [Ern et al.(2017b), Theorem 5.1] have shown that the (time-local but space-global) jump estimators are bounded from above by the (time-local space-global) $L^2(H^1) \cap H^1(H^{-1})$-norm of the error in $\mathcal{I}u_{h\tau}$, i.e., $\int_{I_n} \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|^2 \mathrm{d}t \leq 8 \int_{I_n} \|\nabla(u - \mathcal{I}u_{h\tau})\|^2 + \|\partial_t(u - \mathcal{I}u_{h\tau})\|_{H^{-1}(\Omega)}^2 \mathrm{d}t$, up to possible data oscillation.

Remark 5.2 (Comparison with $L^2(H^1) \cap H^1(H^{-1})$-norm estimators). As pointed out by the remark in [Verfürth(2003), p. 198, Par. (5)] concerning the equivalence of residual-based estimators for both $L^2(H^1)$ and $L^2(H^1) \cap H^1(H^{-1})$ norms, it is important to observe that in the absence of data oscillation, the estimator η_X defined above in (5.2) is equivalent (up to the factor $\sqrt{\frac{3+\sqrt{5}}{2}} \leq \sqrt{3}$) to the augmented $L^2(H^1) \cap H^1(H^{-1})$-norm estimator $\eta_{\mathcal{E}_Y}$ defined in [Ern et al.(2017b), Eq. (5.10b)]. However, an important difference between these estimators concerns the data oscillation. Indeed, it is known since [Verfürth(2003)] that $L^2(H^1) \cap H^1(H^{-1})$ estimators generally contain a data oscillation term that can be of same temporal order as the error. By comparison, the data oscillation term (5.1c) features an additional half-order with respect to the time-step size. Therefore we expect that the X-norm estimator given above may be of special use in situations with significant data oscillation in time.

Theorem 5.1 is our main result on a posteriori error estimation of $\|u - u_{h\tau}\|_X$. Due to the challenge of bounding $\|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X$, several authors have also considered various augmented norms and error measures. We refer in particular to [Akrivis et al.(2009), Makridakis & Nochetto(2006), Schötzau & Wihler(2010)] where the norm of the error includes simultaneously norms for $u - u_{h\tau}$ and $u - \mathcal{I}u_{h\tau}$. This can be motivated by a priori error analysis, where it can be shown that $\|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X$ converges with same order as $\|u - u_{h\tau}\|_X$ when u is sufficiently smooth. For instance, we can define the error measure

$$\mathcal{E}_X \coloneqq \max\left\{\|u - u_{h\tau}\|_X, \|u - \mathcal{I}u_{h\tau}\|_X\right\}, \tag{5.8}$$

in an analoguous manner to the norms in, for instance, [Makridakis & Nochetto(2006), eq. (34)] and [Schötzau & Wihler(2010), eq. (28)] without the $L^\infty(L^2)$-norm terms. The choice in (5.8) is only one of many possibilities; for instance we could equally well consider $\|u - u_{h\tau}\|_X + \|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X$. The interest of this approach is that the bounds (5.4), (5.5) and (5.7) immediately yield a global upper bound and local-in-time and local-in-space efficiency with respect to this error measure, see Corollary 5.2 below. However, it is important to note that it does not appear possible to show in general an equivalence between $\mathcal{E}_X$ and $\|u - u_{h\tau}\|_X$, see Remark 5.1.

Corollary 5.2. *Let $\mathcal{E}_X$ be defined by (5.8). Then, we have the guaranteed upper bound*

$$\mathcal{E}_X \leq 2\,\eta_X, \tag{5.9}$$

If $K \in \mathcal{T}^n$, $1 \leq n \leq N$, is an element such that $h_{\omega_{\boldsymbol{a}}}^2 \leq \gamma_{\boldsymbol{a}}\,\tau_n$ for each $\boldsymbol{a} \in \mathcal{V}_K$ with some constant $\gamma_{\boldsymbol{a}} > 0$, where $h_{\omega_{\boldsymbol{a}}}$ denotes the diameter of the patch $\omega_{\boldsymbol{a}}$, then we have the local efficiency bound

$$[\eta_{\mathrm{F},K}^n]^2 + [\eta_{\mathrm{J},K}^n]^2 \leq \sum_{\boldsymbol{a} \in \mathcal{V}_K} C_{\gamma_{\boldsymbol{a}},q_n}^2 \left\{[\mathcal{E}_X^{\boldsymbol{a},n}]^2 + [\eta_{\mathrm{osc}}^{\boldsymbol{a},n}]^2\right\}. \tag{5.10}$$

where the local error measures $\mathcal{E}_X^{\boldsymbol{a},n}$, $\boldsymbol{a} \in \mathcal{V}^n$, are defined by

$$[\mathcal{E}_X^{\boldsymbol{a},n}]^2 \coloneqq \max\left\{\int_{I_n} \|\nabla(u - u_{h\tau})\|_{\omega_{\boldsymbol{a}}}^2\,\mathrm{d}t, \int_{I_n} \|\nabla(u - \mathcal{I}u_{h\tau})\|_{\omega_{\boldsymbol{a}}}^2\,\mathrm{d}t\right\}. \tag{5.11}$$

Furthermore, under the hypothesis that there exists $\gamma > 0$ such that $h_{\omega_{\boldsymbol{a}}}^2 \leq \gamma\, \tau_n$ for every $\boldsymbol{a} \in \mathcal{V}^n$ and every $1 \leq n \leq N$, then we have the global efficiency bound

$$\sum_{n=1}^{N} \sum_{K \in \mathcal{T}^n} [\eta_{\mathrm{F},K}^n]^2 + [\eta_{\mathrm{J},K}^n]^2 \leq C_{\gamma,q_n}^2 \left\{ [\mathcal{E}_X]^2 + \sum_{n=1}^{N} \sum_{\boldsymbol{a} \in \mathcal{V}^n} [\eta_{\mathrm{osc}}^{\boldsymbol{a},n}]^2 \right\}. \tag{5.12}$$

6 Proof of the guaranteed upper bound (5.4)

We will make use of the following preparatory lemmas.

Lemma 6.1. *Let I_n be a given time interval, and let $\varphi \in L^2(I_n; H_0^1(\Omega)) \cap H^1(I_n; H^{-1}(\Omega))$ be an arbitrary function. Let $\varphi^n \in H_0^1(\Omega)$, the time-mean value of φ over I_n, be defined by $\varphi^n := \frac{1}{\tau_n}\int_{I_n} \varphi \,\mathrm{d}t$. Then*

$$\int_{I_n} \|\nabla \varphi^n\|^2 \,\mathrm{d}t \leq \int_{I_n} \|\nabla \varphi\|^2 \,\mathrm{d}t, \tag{6.1a}$$

$$\int_{I_n} \|\varphi - \varphi^n\|^2 \,\mathrm{d}t \leq \frac{\tau_n}{\pi} \left(\int_{I_n} \|\partial_t \varphi\|_{H^{-1}(\Omega)}^2 \,\mathrm{d}t \right)^{\frac{1}{2}} \left(\int_{I_n} \|\nabla \varphi\|^2 \,\mathrm{d}t \right)^{\frac{1}{2}}. \tag{6.1b}$$

Proof. The bound (6.1a) is simply the stability of the L^2-projection with respect to time; thus it remains only to show (6.1b). It is well-known that there exists a maximal sequence $\{\psi_k\}_{k=1}^{\infty}$ that is orthonormal in the $L^2(\Omega)$-inner product and orthogonal in the $H_0^1(\Omega)$ inner product: i.e. $(\psi_k, \psi_j) = \delta_{kj}$ and $(\nabla\psi_k, \nabla\psi_j) = \lambda_k \delta_{kj}$, with $\{\lambda_k\}_{k=1}^{\infty} \subset \mathbb{R}_{>0}$. Then, we have $\varphi = \sum_{k=1}^{\infty} \alpha_k \psi_k$ and $\varphi^n = \sum_{k=1}^{\infty} \alpha_k^n \psi_k$, with real-valued $\alpha_k \in H^1(I_n)$ and $\alpha_k^n = \frac{1}{\tau_n}\int_{I_n} \alpha_k \mathrm{d}t$. Thus we may use the Poincaré inequality for real-valued functions to obtain

$$\begin{aligned}
\int_{I_n} \|\varphi - \varphi^n\|^2 \mathrm{d}t = \sum_{k=1}^{\infty} \|\alpha_k - \alpha_k^n\|_{L^2(I_n)}^2 &\leq \frac{\tau_n}{\pi} \sum_{k=1}^{\infty} |\alpha_k|_{H^1(I_n)} \|\alpha_k\|_{L^2(I_n)} \\
&\leq \frac{\tau_n}{\pi} \left(\sum_{k=1}^{\infty} \frac{1}{\lambda_k} |\alpha_k|_{H^1(I_n)}^2 \right)^{\frac{1}{2}} \left(\sum_{k=1}^{\infty} \lambda_k \|\alpha_k\|_{L^2(I_n)}^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

We then deduce (6.1b) from the identities $\int_{I_n} \|\partial_t \varphi\|_{H^{-1}(\Omega)}^2 \mathrm{d}t = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} |\alpha_k|_{H^1(I_n)}^2$ and $\int_{I_n} \|\nabla\varphi\|^2 \mathrm{d}t = \sum_{k=1}^{\infty} \lambda_k \|\alpha_k\|_{L^2(I_n)}^2$. $\square$

Lemma 6.2. *Let $f \in L^2(0,T; L^2(\Omega))$, let $f_{h\tau}$ be defined by (4.4), and let $\varphi \in L^2(0,T; H_0^1(\Omega)) \cap H^1(0,T; H^{-1}(\Omega))$ be an arbitrary function. Then, for each $1 \leq n \leq N$,*

$$\left| \int_{I_n} (f - f_{h\tau}, \varphi) \,\mathrm{d}t \right| \leq \eta_{\mathrm{osc},h\tau}^n \left(\int_{I_n} \|\partial_t \varphi\|_{H^{-1}(\Omega)}^2 + \|\nabla \varphi\|^2 \,\mathrm{d}t \right)^{\frac{1}{2}}. \tag{6.2}$$

Proof. For a given function $\varphi \in L^2(0,T; H_0^1(\Omega)) \cap H^1(0,T; H^{-1}(\Omega))$, we define the time-mean value of φ over I_n as $\varphi^n := \frac{1}{\tau_n}\int_{I_n} \varphi \,\mathrm{d}t \in H_0^1(\Omega)$, and we define the space-mean value of φ^n over $\widetilde{K}$ as $\varphi^n_{\widetilde{K}}|_{\widetilde{K}} := \frac{1}{|\widetilde{K}|}\int_{\widetilde{K}} \varphi^n \mathrm{d}x$, where $1 \le n \le N$ and $\widetilde{K} \in \widetilde{\mathcal{T}^n}$. Now, we note that the definition of $f_{h\tau}$ in (4.4) implies that $f - f_{h\tau}$ has zero mean value over each space-time element $\widetilde{K} \times I_n$. Therefore, we obtain

$$\int_{I_n} (f - f_{h\tau}, \varphi)\,\mathrm{d}t = \int_{I_n} (f - f_{h\tau}, \varphi - \varphi^n) + \sum_{\widetilde{K}\in\widetilde{\mathcal{T}^n}} (f - f_{h\tau}, \varphi^n - \varphi^n_{\widetilde{K}})_{\widetilde{K}}\,\mathrm{d}t =: A + B.$$

Then, we apply the bounds (6.1a), (6.1b), and the Cauchy–Schwarz inequality $\|\varphi^n - \varphi^n_{\widetilde{K}}\|_{\widetilde{K}} \le \frac{h_{\widetilde{K}}}{\pi}\|\nabla\varphi^n\|_{\widetilde{K}}$ to obtain

$$|A| \le \left(\int_{I_n} \sum_{\widetilde{K}\in\widetilde{\mathcal{T}^n}} \frac{\tau_n}{\pi}\|f - f_{h\tau}\|^2_{\widetilde{K}}\,\mathrm{d}t\right)^{\frac{1}{2}} \left(\int_{I_n} \|\partial_t\varphi\|^2_{H^{-1}(\Omega)}\,\mathrm{d}t\right)^{\frac{1}{4}} \left(\int_{I_n} \|\nabla\varphi\|^2\,\mathrm{d}t\right)^{\frac{1}{4}},$$

$$|B| \le \left(\int_{I_n} \sum_{\widetilde{K}\in\widetilde{\mathcal{T}^n}} \frac{h^2_{\widetilde{K}}}{\pi^2}\|f - f_{h\tau}\|^2_{\widetilde{K}}\,\mathrm{d}t\right)^{\frac{1}{2}} \left(\int_{I_n} \|\nabla\varphi\|^2\,\mathrm{d}t\right)^{\frac{1}{2}}.$$

Then, the Cauchy–Schwarz inequality and the Young inequality $ab + b^2 \le \frac{1+\sqrt{2}}{2}(a^2 + b^2)$ for all $a, b \in \mathbb{R}$, imply that the bound (6.2) holds. □

Proof of the upper bound (5.4). Recall from (2.9) on the equivalence of norms and residuals that $\|u - u_{h\tau}\|_X = \|\mathcal{R}_X(u_{h\tau})\|_{[Y_T]'}$, so we turn our attention to bounding $\langle\mathcal{R}_X(u_{h\tau}), \varphi\rangle$ for an arbitrary test function $\varphi \in Y_T$. By adding and subtracting $\int_0^T (\partial_t\mathcal{I}u_{h\tau} + \nabla\cdot\boldsymbol{\sigma}_{h\tau}, \varphi)\,\mathrm{d}t$ and recalling the flux equilibration identity (4.1), we get

$$\langle\mathcal{R}_X(u_{h\tau}), \varphi\rangle = \int_0^T (f - f_{h\tau}, \varphi) + \langle\partial_t\varphi, u_{h\tau} - \mathcal{I}u_{h\tau}\rangle - (\boldsymbol{\sigma}_{h\tau} + \nabla u_{h\tau}, \nabla\varphi)\,\mathrm{d}t \\ + (u_0 - \Pi_h u_0, \varphi(0)), \quad (6.3)$$

where we have used integration by parts with respect to time for the time derivative $\partial_t\mathcal{I}u_{h\tau}$, noting that $\mathcal{I}u_{h\tau}(0) = \Pi_h u_0$ and that $\varphi(T) = 0$, and also where we have used integration by parts over Ω for the flux $\boldsymbol{\sigma}_{h\tau} \in L^2(0, T; H(\mathrm{div}, \Omega))$. Employing the shorthand notation $\|\varphi\|^2_{Y(I_n)} := \int_{I_n} \|\partial_t\varphi\|^2_{H^{-1}(\Omega)} + \|\nabla\varphi\|^2\,\mathrm{d}t$, we then use Lemma 6.2 and the Cauchy–Schwarz inequality to bound

$$\begin{aligned}
&\int_0^T (f - f_{h\tau}, \varphi) + \langle\partial_t\varphi, u_{h\tau} - \mathcal{I}u_{h\tau}\rangle - (\boldsymbol{\sigma}_{h\tau} + \nabla u_{h\tau}, \nabla\varphi)\,\mathrm{d}t \\
&\le \sum_{n=1}^N \left\{\left[\int_{I_n} \|\boldsymbol{\sigma}_{h\tau} + \nabla u_{h\tau}\|^2 + \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t\right]^{\frac{1}{2}} + \eta^n_{\mathrm{osc},h\tau}\right\} \|\varphi\|_{Y(I_n)} \qquad (6.4) \\
&= \sum_{n=1}^N \left\{\left[\sum_{K\in\mathcal{T}^n} \left\{[\eta^n_{\mathrm{F},K}]^2 + [\eta^n_{\mathrm{J},K}]^2\right\}\right]^{\frac{1}{2}} + \eta^n_{\mathrm{osc},h\tau}\right\} \|\varphi\|_{Y(I_n)}.
\end{aligned}$$

We then combine (6.3) and (6.4) with the Cauchy–Schwarz inequality to find that $\langle \mathcal{R}_X(u_{h\tau}), \varphi\rangle \leq \eta_X \|\varphi\|_{Y_T}$; since $\varphi \in Y_T$ was arbitrary, we obtain $\|u - u_{h\tau}\|_X \leq \eta_X$ as a result of (2.9), thereby completing the proof of (5.4). $\square$

7 Proof of the bounds (5.5) and (5.7)

We start by observing that $\boldsymbol{\sigma}_{h\tau}|_{K\times I_n} = \sum_{\boldsymbol{a}\in\mathcal{V}_K} \boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n}|_{K\times I_n}$, and thus

$$\int_{I_n} [\eta_{\mathrm{F},K}^n]^2 \,\mathrm{d}t = \int_{I_n} \|{\textstyle\sum_{\boldsymbol{a}\in\mathcal{V}_K}} (\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n} + \psi_{\boldsymbol{a}} \nabla u_{h\tau})\|_K^2 \,\mathrm{d}t \leq |\mathcal{V}_K| \sum_{\boldsymbol{a}\in\mathcal{V}_K} \int_{I_n} \|\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n} + \psi_{\boldsymbol{a}}\nabla u_{h\tau}\|_K^2 \,\mathrm{d}t, \tag{7.1}$$

where we recall that $\mathcal{V}_K$ stands for the vertices of the element K and $|\mathcal{V}_K|$ stands for its cardinality. We shall now bound the right-hand side of (7.1). For each $1 \leq n \leq N$ and each $\boldsymbol{a} \in \mathcal{V}^n$, we introduce the patch residual functional $\mathcal{R}_{h\tau}^{\boldsymbol{a},n} : L^2(I_n, H_0^1(\omega_{\boldsymbol{a}})) \to \mathbb{R}$ defined by

$$\langle \mathcal{R}_{h\tau}^{\boldsymbol{a},n}, \varphi\rangle = \int_{I_n} \left(\Pi_{h\tau}^{\boldsymbol{a},n} f - \partial_t \mathcal{I} u_{h\tau}, \varphi\right)_{\omega_{\boldsymbol{a}}} - \left(\nabla u_{h\tau}, \nabla\varphi\right)_{\omega_{\boldsymbol{a}}} \mathrm{d}t \quad \forall\, \varphi \in L^2(I_n; H_0^1(\omega_{\boldsymbol{a}})). \tag{7.2}$$

The essential result that forms the starting point for our analysis is the following abstract efficiency result first shown in [Ern et al.(2017b), Lemma 8.2], which is an application of a more general underlying key result in [Ern et al(2017a), Theorem 1.2] concerning the existence of polynomial-degree robust liftings of piecewise polynomial data into discrete subspaces of $\boldsymbol{H}(\mathrm{div})$, which itself is based on the fundamental results of [Costabel & McIntosh(2010), Braess et al.(2009)].

Lemma 7.1 (Space-time stability bound)**.** *Let $\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n}$ denote the patch-wise flux reconstructions of Definition 4.1, and let $\mathcal{R}_{h\tau}^{\boldsymbol{a},n}$ denote the local patch residual defined by (7.2). Then, we have*

$$\left(\int_{I_n} \|\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n} + \psi_{\boldsymbol{a}} \nabla u_{h\tau}\|_{\omega_{\boldsymbol{a}}}^2 \,\mathrm{d}t\right)^{\frac{1}{2}} \lesssim \sup_{\varphi\in\mathcal{Q}_{q_n}(I_n;H_0^1(\omega_{\boldsymbol{a}}))\setminus\{0\}} \frac{\langle \mathcal{R}_{h\tau}^{\boldsymbol{a},n}, \varphi\rangle}{\left(\int_{I_n} \|\nabla\varphi\|_{\omega_{\boldsymbol{a}}}^2 \,\mathrm{d}t\right)^{\frac{1}{2}}}, \tag{7.3}$$

where $\mathcal{Q}_{q_n}(I_n; H_0^1(\omega_{\boldsymbol{a}}))$ denotes the space of $H_0^1(\omega_{\boldsymbol{a}})$-valued univariate polynomials of degree at most q_n on I_n. In particular, the constant in (7.3) does not depend on the mesh-size, time-step size, spatial and temporal polynomial degrees, or on refinement and coarsening between time-steps.

As explained above in the introduction, our analysis of the efficiency of the equilibrated flux estimator $\eta_{\mathrm{F},K}^n$ relies on two original ideas. We now detail the first one, which is based on the key observation that the set of test functions appearing in (7.3) are polynomials with respect to the time variable. Hence, in order to obtain estimates on the efficiency of the estimators with respect to the X-norm of the error, we shall show that the set of test functions appearing in (7.3) can be restricted to functions vanishing at the end-points of the time interval and thereby lying in the test space Y_T through a bubble-in-time argument, provided that $h_{\omega_{\boldsymbol{a}}}^2 \lesssim \tau_n$.

We start by defining the space $H^1_\dagger(\omega_{\boldsymbol{a}})$ through

$$H^1_\dagger(\omega_{\boldsymbol{a}}) \coloneqq \begin{cases} \{v \in H^1(\omega_{\boldsymbol{a}}), \quad (v, \psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} = 0\} & \text{if } \boldsymbol{a} \in \mathcal{V}^n_{\mathrm{int}}, \\ \{v \in H^1(\omega_{\boldsymbol{a}}), \quad v|_{\partial\omega_{\boldsymbol{a}} \cap \partial\Omega} = 0\} & \text{if } \boldsymbol{a} \in \mathcal{V}^n_{\mathrm{ext}}. \end{cases} \tag{7.4}$$

Recall that the dual norm $\|\cdot\|_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'}$ of $H^1_\dagger(\omega_{\boldsymbol{a}})$ is defined by $\|\Phi\|_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'} = \sup\langle \Phi, v\rangle$, where the supremum is taken among all test functions $v \in H^1_\dagger(\omega_{\boldsymbol{a}})$ such that $\|\nabla v\|_{\omega_{\boldsymbol{a}}} = 1$. The motivation for working with the space $H^1_\dagger(\omega_{\boldsymbol{a}})$ is that the $\psi_{\boldsymbol{a}}$-weighted mean value of the function $u - \mathcal{I}u_{h\tau}$ possesses special properties derived from the numerical scheme; in particular, see Lemma 7.3 and the discussion surrounding (7.13) below.

Lemma 7.2 (Stability with test functions vanishing at both endpoints of I_n)**.** *Let $\boldsymbol{a} \in \mathcal{V}^n$, $1 \le n \le N$, and suppose that there exists a constant $\gamma_{\boldsymbol{a}} > 0$ such that the patch diameter $h_{\omega_{\boldsymbol{a}}}$ and τ_n satisfy $h_{\omega_{\boldsymbol{a}}}^2 \le \gamma_{\boldsymbol{a}} \tau_n$. Then,*

$$\left(\int_{I_n} \|\boldsymbol{\sigma}_{h\tau}^{\boldsymbol{a},n} + \psi_{\boldsymbol{a}} \nabla u_{h\tau}\|_{\omega_{\boldsymbol{a}}}^2 \,\mathrm{d}t \right)^{\frac{1}{2}} \le C_{\gamma_{\boldsymbol{a}}, q_n} \sup_{\substack{\varphi \in \mathcal{Q}_{q_n+2}(I_n; H^1_0(\omega_{\boldsymbol{a}})) \\ \cap H^1_0(I_n; H^1_0(\omega_{\boldsymbol{a}}))}} \frac{\langle \mathcal{R}_{h\tau}^{\boldsymbol{a},n}, \varphi \rangle}{\left(\int_{I_n} \|\partial_t \varphi\|^2_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'} + \|\nabla\varphi\|^2_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t \right)^{\frac{1}{2}}}, \tag{7.5}$$

where $H^1_0(I_n; H^1_0(\omega_{\boldsymbol{a}}))$ denotes the space of functions in $H^1(I_n; H^1_0(\omega_{\boldsymbol{a}}))$ that vanish at both endpoints t_{n-1} and t_n of the time interval I_n. In particular, the constant $C_{\gamma_{\boldsymbol{a}}, q_n}$ in (7.5) satisfies $C_{\gamma_{\boldsymbol{a}}, q_n} \lesssim (q_n+1)^{\frac{1}{2}} + \gamma_{\boldsymbol{a}}(q_n+1)^{\frac{5}{2}}$, and may depend on the shape regularity of $\mathcal{T}^n$ and $\widetilde{\mathcal{T}^n}$ and on the space dimension d, but otherwise does not depend on the mesh-size, time-step size, spatial polynomial degrees, or on refinement and coarsening between time-steps.

Proof. The starting point for the proof is Lemma 7.1. Keeping in mind the right-hand side of (7.3), for each $\varphi \in \mathcal{Q}_{q_n}(I_n; H^1_0(\omega_{\boldsymbol{a}}))$, we shall construct a new function $\varphi_* \in \mathcal{Q}_{q_n+2}(I_n; H^1_0(\omega_{\boldsymbol{a}}))$ defined by

$$\varphi_* \coloneqq \varphi - \varphi(t_{n-1}^+) \frac{(-1)^{q_n+1}}{2} (L^n_{q_n+1} - L^n_{q_n+2}) - \varphi(t_n) \frac{1}{2} (L^n_{q_n+1} + L^n_{q_n+2}).$$

It follows from the fact that $L^n_q(t_{n-1}) = (-1)^q$ and that $L^n_q(t_n) = 1$ for all $q \ge 0$ that $\varphi_*(t_{n-1}^+) = \varphi_*(t_n) = 0$ and hence $\varphi_* \in H^1_0(I_n; H^1_0(\omega_{\boldsymbol{a}}))$. Recalling that the functions $\Pi_{h\tau}^{\boldsymbol{a},n} f$, $\partial_t \mathcal{I} u_{h\tau}$, and $\nabla u_{h\tau}$ appearing in (7.2) are polynomials of degree at most q_n in time, it also follows from the orthogonality of the Legendre polynomials that

$$\langle \mathcal{R}_{h\tau}^{\boldsymbol{a},n}, \varphi_* \rangle = \langle \mathcal{R}_{h\tau}^{\boldsymbol{a},n}, \varphi \rangle.$$

It is then seen that we shall obtain (7.5) as a result of (7.3) provided that we can bound $\int_{I_n}\|\partial_t\varphi_*\|^2_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'}+\|\nabla\varphi_*\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t$ in terms of $\int_{I_n}\|\nabla\varphi\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t$. First, the triangle inequality and the properties of the Legendre polynomials imply that

$$\int_{I_n}\|\nabla\varphi_*\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t \lesssim \int_{I_n}\|\nabla\varphi\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t+\tfrac{\tau_n}{q_n+1}\left(\|\nabla\varphi(t_{n-1})\|^2_{\omega_{\boldsymbol{a}}}+\|\nabla\varphi(t_n)\|^2_{\omega_{\boldsymbol{a}}}\right), \tag{7.6}$$

where the constant is independent of all other quantities. Now, the key point is that we have the inverse inequality

$$\max_{t\in I_n}\|\nabla\varphi(t)\|^2_{\omega_{\boldsymbol{a}}}\lesssim \tfrac{(q_n+1)^2}{\tau_n}\int_{I_n}\|\nabla\varphi\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t, \tag{7.7}$$

where the constant is independent of all other quantities since $\varphi\in\mathcal{Q}_{q_n}(I_n;H^1_0(\omega_{\boldsymbol{a}}))$ is discrete with respect to time. Note in particular that the inverse inequality is valid even though $\mathcal{Q}_{q_n}(I_n;H^1_0(\omega_{\boldsymbol{a}}))$ is itself an infinite dimensional space, see Remark 7.1 below. Therefore, we find from (7.6) and (7.7) that

$$\int_{I_n}\|\nabla\varphi_*\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t\lesssim (q_n+1)\int_{I_n}\|\nabla\varphi\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t. \tag{7.8}$$

To bound $\int_{I_n}\|\partial_t\varphi_*\|^2_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'}\mathrm{d}t$, we recall that $\varphi_*(t)\in H^1_0(\omega_{\boldsymbol{a}})$ for all $t\in I_n$, and therefore satisfies the Poincaré inequality $\|\varphi_*(t)\|_{\omega_{\boldsymbol{a}}}\lesssim h_{\omega_{\boldsymbol{a}}}\|\nabla\varphi_*(t)\|_{\omega_{\boldsymbol{a}}}$ for all $t\in I_n$. Furthermore, we also have a similar Poincaré inequality for all test functions $v\in H^1_\dagger(\omega_{\boldsymbol{a}})$. Therefore, we find that $\|\varphi_*(t)\|_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'}\lesssim h^2_{\omega_{\boldsymbol{a}}}\|\nabla\varphi_*(t)\|_{\omega_{\boldsymbol{a}}}$, for all $t\in I_n$. Thus, we obtain, using an inverse inequality in time (see Remark 7.1 below for details),

$$\begin{aligned}\int_{I_n}\|\partial_t\varphi_*\|^2_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'}\,\mathrm{d}t &\lesssim \tfrac{(q_n+1)^4}{\tau_n^2}\int_{I_n}\|\varphi_*\|^2_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'}\,\mathrm{d}t\\ &\lesssim \tfrac{(q_n+1)^4h^4_{\omega_{\boldsymbol{a}}}}{\tau_n^2}\int_{I_n}\|\nabla\varphi_*\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t\lesssim\gamma^2_{\boldsymbol{a}}(q_n+1)^5\int_{I_n}\|\nabla\varphi\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t,\end{aligned}$$

where we have used the hypothesis that $h^2_{\omega_{\boldsymbol{a}}}/\tau_n\le\gamma_{\boldsymbol{a}}$ in the last inequality. Hence, we have shown that

$$\int_{I_n}\|\partial_t\varphi_*\|^2_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'}+\|\nabla\varphi_*\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t\le C^2_{\gamma_{\boldsymbol{a}},q_n}\int_{I_n}\|\nabla\varphi\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t, \tag{7.9}$$

where the constant $C_{\gamma_{\boldsymbol{a}},q_n}\lesssim(q_n+1)^{\frac12}+\gamma_{\boldsymbol{a}}(q_n+1)^{\frac52}$. The bound (7.5) then follows from (7.9) and the identity $\langle\mathcal{R}^{\boldsymbol{a},n}_{h\tau},\varphi\rangle=\langle\mathcal{R}^{\boldsymbol{a},n}_{h\tau},\varphi_*\rangle$ given above. □

Remark 7.1 (Inverse inequality). The proof of the inverse inequalities appearing above in (7.7) can be found simply by expanding the function φ in any orthogonal basis $\{\psi_k\}_{k=1}^\infty$ of $H^1_0(\omega_{\boldsymbol{a}})$ as $\varphi(t)=\sum_{k=1}^\infty c_k(t)\psi_k$, where the coefficient functions c_k are real-valued polynomials of degree at most q_n, for all $k\ge1$, and then by applying coefficient-wise well known inverse inequalities for real-valued functions, see [Schwab(1998), p. 148].

Lemma 7.2 constitutes the first step towards the local lower bound (5.5). In particular, we see that the test functions in (7.5) are bounded in the $H^1(I_n;[H^1_\dagger(\omega_{\boldsymbol{a}})]')$ norm. In order to exploit this property, we use a second key idea for our analysis, which is to employ the following special property of the time-discretization scheme. Together, these two ingredients allow us to obtain the lower bounds assuming only that $h^2 \lesssim \tau$, rather than the stronger requirements used in [Picasso(1998), Verfürth(1998)].

Lemma 7.3 (Pointwise identity)**.** *For each $1 \le n \le N$ and each interior vertex $\boldsymbol{a} \in \mathcal{V}^n_{\mathrm{int}}$, the functions $\mathcal{I}u_{h\tau}$ and $u_{h\tau}$ satisfy*

$$\langle \partial_t \mathcal{I}u_{h\tau}, \psi_{\boldsymbol{a}}\rangle + (\nabla u_{h\tau}, \nabla\psi_{\boldsymbol{a}}) = (\Pi^{\boldsymbol{a},n}_{h\tau} f, \psi_{\boldsymbol{a}}) \quad \textit{pointwise in } I_n, \tag{7.10}$$

where $\Pi^{\boldsymbol{a},n}_{h\tau} f$ was defined in section 4.2.

Proof. Since $\boldsymbol{a} \in \mathcal{V}^n_{\mathrm{int}}$, it follows that $\phi\psi_{\boldsymbol{a}} \in \mathcal{Q}_{q_n}(I_n;V^n)$ for any polynomial ϕ in time of degree at most q_n over I_n. Therefore, the numerical scheme (3.6) implies that, for any real-valued polynomial ϕ in time of degree at most q_n,

$$\int_{I_n} \phi\,[(f,\psi_{\boldsymbol{a}}) - (\partial_t\mathcal{I}u_{h\tau},\psi_{\boldsymbol{a}}) - (\nabla u_{h\tau},\nabla\psi_{\boldsymbol{a}})]\,\mathrm{d}t = 0.$$

Furthermore, the definition of $\Pi^{\boldsymbol{a},n}_{h\tau}$ implies that $\int_{I_n}\phi(f,\psi_{\boldsymbol{a}})\mathrm{d}t = \int_{I_n}\phi(\Pi^{\boldsymbol{a},n}_{h\tau} f,\psi_{\boldsymbol{a}})\mathrm{d}t$ for any real-valued polynomial ϕ in time of degree at most q_n. Since the function $t \mapsto (\partial_t\mathcal{I}u_{h\tau}(t),\psi_{\boldsymbol{a}}) + (\nabla u_{h\tau}(t),\nabla\psi_{\boldsymbol{a}}) - (\Pi^{\boldsymbol{a},n}_{h\tau} f(t),\psi_{\boldsymbol{a}})$ is a real-valued polynomial of degree at most q_n over I_n, it follows that it vanishes everywhere in I_n. We therefore obtain (7.10). $\square$

We now give the proof of the bounds (5.5) and (5.7) under the hypothesis stated in Theorem 5.1.

Proof of the bounds (5.5) and (5.7) The proof consists in bounding the right-hand side of (7.5) so as to show that, for each $\boldsymbol{a} \in \mathcal{V}^n$, we have the bound

$$\begin{aligned}\int_{I_n} \|\boldsymbol{\sigma}^{\boldsymbol{a},n}_{h\tau} + \psi_{\boldsymbol{a}}\nabla u_{h\tau}\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t \le C^2_{\gamma_{\boldsymbol{a}},q_n}\Bigg\{\int_{I_n}\|\nabla(u-\mathcal{I}u_{h\tau})\|^2_{\omega_{\boldsymbol{a}}} + \|\nabla(u-u_{h\tau})\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t \\ + \int_{I_n}\|f - \Pi^{\boldsymbol{a},n}_{h\tau} f\|^2_{H^{-1}(\omega_{\boldsymbol{a}})}\,\mathrm{d}t\Bigg\},\end{aligned} \tag{7.11}$$

where $C_{\gamma_{\boldsymbol{a}},q_n} \lesssim (q_n+1)^{\frac{1}{2}} + \gamma_{\boldsymbol{a}}(q_n+1)^{\frac{5}{2}}$. Then, once (7.11) is known, it is then straightforward to show (5.5) and (5.7) from (7.1).

To show (7.11), we will treat first the more difficult case where $\boldsymbol{a} \in \mathcal{V}^n_{\mathrm{int}}$ is an interior node. It will be convenient to denote $\overline{\psi}_{\boldsymbol{a}} := \psi_{\boldsymbol{a}}/\|\psi_{\boldsymbol{a}}\|_{L^1(\omega_{\boldsymbol{a}})}$ the renormalized hat function associated with $\boldsymbol{a}$. Let $\varphi \in \mathcal{Q}_{q_n+2}(I_n;H^1_0(\omega_{\boldsymbol{a}})) \cap H^1_0(I_n;H^1_0(\omega_{\boldsymbol{a}}))$ be a fixed but arbitrary test function, such that $\int_{I_n}\|\partial_t\varphi\|^2_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'} + \|\nabla\varphi\|^2_{\omega_{\boldsymbol{a}}}\,\mathrm{d}t = 1$. It follows that the zero-extension of φ to $\Omega \times (0,T)$ belongs to Y_T, and therefore, we may use the weak formulation (2.4) in the definition of $\mathcal{R}_{h\tau}^{\boldsymbol{a},n}$ from (7.2) to find that

$$\langle \mathcal{R}_{h\tau}^{\boldsymbol{a},n}, \varphi \rangle = \int_{I_n} -(u - \mathcal{I}u_{h\tau}, \partial_t \varphi)_{\omega_{\boldsymbol{a}}} + (\nabla(u - u_{h\tau}), \nabla\varphi)_{\omega_{\boldsymbol{a}}} + (\Pi_{h\tau}^{\boldsymbol{a},n} f - f, \varphi)_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t. \quad (7.12)$$

Note that, in general, $u - \mathcal{I}u_{h\tau}$ fails to belong to $H^1_\dagger(\omega_{\boldsymbol{a}})$ when $\boldsymbol{a} \in \mathcal{V}^n_{\mathrm{int}}$ is an interior node because we can not generally guarantee that $(u - \mathcal{I}u_{h\tau}, \psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} = 0$ a.e. in time; thus, $|(u - \mathcal{I}u_{h\tau}, \partial_t\varphi)_{\omega_{\boldsymbol{a}}}| \not\lesssim \|\nabla(u - \mathcal{I}u_{h\tau})\|_{\omega_{\boldsymbol{a}}} \|\partial_t \varphi\|_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'}$ in general. To overcome this obstacle, we introduce the auxiliary function

$$e_{\boldsymbol{a}} \coloneqq u - \mathcal{I}u_{h\tau} - (u - \mathcal{I}u_{h\tau}, \overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}, \quad (7.13)$$

that is, we subtract the $\overline{\psi}_{\boldsymbol{a}}$-weighted average of $u - \mathcal{I}u_{h\tau}$ from $u - \mathcal{I}u_{h\tau}$. It follows from the definition that $e_{\boldsymbol{a}}(t) \in H^1_\dagger(\omega_{\boldsymbol{a}})$ and that $\|\nabla e_{\boldsymbol{a}}(t)\|_{\omega_{\boldsymbol{a}}} = \|\nabla(u - \mathcal{I}u_{h\tau})(t)\|_{\omega_{\boldsymbol{a}}}$ for almost all $t \in I_n$. We now show how to reformulate the patch residual $\langle \mathcal{R}_{h\tau}^{\boldsymbol{a},n}, \varphi\rangle$ in terms of the auxiliary function $e_{\boldsymbol{a}}$. First, we may choose the test function $\overline{\psi}_{\boldsymbol{a}}(\varphi, 1)_{\omega_{\boldsymbol{a}}} \in Y_T$ in (2.4), and use Fubini's theorem and linearity of integration to find that

$$\begin{aligned}\int_{I_n} -((u, \overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}, \partial_t\varphi)_{\omega_{\boldsymbol{a}}} \mathrm{d}t &= \int_{I_n} -\langle u, \partial_t(\overline{\psi}_{\boldsymbol{a}}(\varphi,1)_{\omega_{\boldsymbol{a}}})\rangle \mathrm{d}t \\ &= \int_{I_n} (f, \overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}(\varphi,1)_{\omega_{\boldsymbol{a}}} - (\nabla u, \nabla\overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}(\varphi,1)_{\omega_{\boldsymbol{a}}} \mathrm{d}t.\end{aligned} \quad (7.14)$$

Next, we multiply (7.10) by $(\varphi, 1)_{\omega_{\boldsymbol{a}}}$ and integrate by parts over I_n and obtain

$$\int_{I_n} -((\mathcal{I}u_{h\tau}, \overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}, \partial_t\varphi)_{\omega_{\boldsymbol{a}}} \mathrm{d}t = \int_{I_n} (\Pi_{h\tau}^{\boldsymbol{a},n} f, \overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}(\varphi,1)_{\omega_{\boldsymbol{a}}} - (\nabla u_{h\tau}, \nabla\overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}(\varphi,1)_{\omega_{\boldsymbol{a}}} \mathrm{d}t. \quad (7.15)$$

The combination of (7.12) with (7.14) and (7.15) shows that $\langle \mathcal{R}_{h\tau}^{\boldsymbol{a},n}, \varphi\rangle = \sum_{i=1}^5 R_i$, where the quantities R_i, $1 \leq i \leq 5$, are defined by

$$R_1 \coloneqq \int_{I_n} -(e_{\boldsymbol{a}}, \partial_t\varphi)_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t,$$

$$R_2 \coloneqq \int_{I_n} (\nabla(u - u_{h\tau}), \nabla\varphi)_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t, \qquad R_3 \coloneqq -\int_{I_n} (\nabla(u - u_{h\tau}), \nabla\overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}(\varphi,1)_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t,$$

$$R_4 \coloneqq \int_{I_n} (f - \Pi_{h\tau}^{\boldsymbol{a},n} f, \overline{\psi}_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}(\varphi,1)_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t, \qquad R_5 \coloneqq -\int_{I_n} (f - \Pi_{h\tau}^{\boldsymbol{a},n} f, \varphi)_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t.$$

Using the fact that $\int_{I_n} \|\partial_t\varphi\|^2_{[H^1_\dagger(\omega_{\boldsymbol{a}})]'} \,\mathrm{d}t \leq 1$, where we recall that $H^1_\dagger(\omega_{\boldsymbol{a}})$ is defined in (7.4), and that $\|\nabla e_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} = \|\nabla(u - \mathcal{I}u_{h\tau})\|_{\omega_{\boldsymbol{a}}}$, we find that $|R_1|^2 \leq \int_{I_n} \|\nabla(u - \mathcal{I}u_{h\tau})\|^2_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t$. Next, we find that $|R_2|^2 \leq \int_{I_n} \|\nabla(u - u_{h\tau})\|^2_{\omega_{\boldsymbol{a}}} \,\mathrm{d}t$. To bound R_3 and R_4, we apply the Cauchy–Schwarz inequality and use the Poincaré inequality on $H^1_0(\omega_{\boldsymbol{a}})$ to obtain

$$|R_3|^2 + |R_4|^2 \lesssim \int_{I_n} \frac{h^2_{\omega_{\boldsymbol{a}}} |\omega_{\boldsymbol{a}}| \|\nabla\psi_{\boldsymbol{a}}\|^2_{\omega_{\boldsymbol{a}}}}{\|\psi_{\boldsymbol{a}}\|^2_{L^1(\omega_{\boldsymbol{a}})}} \left[\|\nabla(u - u_{h\tau})\|^2_{\omega_{\boldsymbol{a}}} + \|f - \Pi_{h\tau}^{\boldsymbol{a},n} f\|^2_{H^{-1}(\omega_{\boldsymbol{a}})} \right] \mathrm{d}t,$$

where $|\omega_{\boldsymbol{a}}|$ denotes the measure of $\omega_{\boldsymbol{a}}$. Since there is a constant depending only on the shape-regularity of the elements of the patch $\omega_{\boldsymbol{a}}$ such that $h_{\omega_{\boldsymbol{a}}}|\omega_{\boldsymbol{a}}|^{1/2}\|\nabla\psi_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \lesssim \|\psi_{\boldsymbol{a}}\|_{L^1(\omega_{\boldsymbol{a}})}$, we find that $|R_3|^2 + |R_4|^2 \lesssim \int_{I_n} \|\nabla(u-u_{h\tau})\|^2_{\omega_{\boldsymbol{a}}} + \|f - \Pi^{\boldsymbol{a},n}_{h\tau} f\|^2_{H^{-1}(\omega_{\boldsymbol{a}})}\mathrm{d}t$. Finally, it is straightforward to show that $|R_5|^2 \leq \int_{I_n} \|f - \Pi^{\boldsymbol{a},n}_{h\tau} f\|^2_{H^{-1}(\omega_{\boldsymbol{a}})}\,\mathrm{d}t$. Therefore, the above bounds on the quantities R_i imply (7.11) for the case where $\boldsymbol{a} \in \mathcal{V}^n_{\mathrm{int}}$ is an interior vertex.

The analogous result for the case where $\boldsymbol{a} \in \mathcal{V}^n_{\mathrm{ext}}$ is a boundary vertex poses fewer difficulties than the case of interior vertices, owing to the fact that $u - \mathcal{I}u_{h\tau} \in H^1_\dagger(\omega_{\boldsymbol{a}})$ for a.e. $t \in I_n$, since u and $\mathcal{I}u_{h\tau}$ are both in X and therefore have vanishing trace on $\partial\omega_{\boldsymbol{a}} \cap \partial\Omega$.

Using the triangle inequality $\|\nabla(u-\mathcal{I}u_{h\tau})\|_{\omega_{\boldsymbol{a}}} \leq \|\nabla(u-u_{h\tau})\|_{\omega_{\boldsymbol{a}}} + \|\nabla(u_{h\tau}-\mathcal{I}u_{h\tau})\|_{\omega_{\boldsymbol{a}}}$, it is then straightforward to obtain (5.5) and (5.7) from (7.1) and (7.11). $\square$

References

[Akrivis et al.(2009)] Akrivis, G., Makridakis, C. & Nochetto, R. H. (2009) Optimal order a posteriori error estimates for a class of Runge-Kutta and Galerkin methods. Numer. Math., **114**, 133–160.

[Amrein & Wihler(2016)] Amrein, M. & Wihler, T. P. (2016) An adaptive space-time Newton–Galerkin approach for semilinear singularly perturbed parabolic evolution equations. IMA J. Numer. Anal., **37**, 2004–2019.

[Bergam et al.(2005)] Bergam, A., Bernardi, C. & Mghazli, Z. (2005) A posteriori analysis of the finite element discretization of some parabolic equations. Math. Comp., **74**, 1117–1138.

[Braess et al.(2009)] Braess, D., Pillwein, V. & Schöberl, J. (2009) Equilibrated residual error estimates are p-robust. Comput. Methods Appl. Mech. Engrg., **198**, 1189–1197.

[Chen & Feng(2004)] Chen, Z. & Feng, J. (2004) An adaptive finite element algorithm with reliable and efficient error control for linear parabolic problems. Math. Comp., **73**, 1167–1193.

[Costabel & McIntosh(2010)] Costabel, M. & McIntosh, A. (2010) On Bogovskiĭ and regularized Poincaré integral operators for de Rham complexes on Lipschitz domains. Math. Z., **265**, 297–320.

[Di Pietro et al.(2015)] Di Pietro, D. A., Vohralík, M. & Yousef, S. (2015) Adaptive regularization, linearization, and discretization and a posteriori error control for the two-phase Stefan problem. Math. Comp., **84**, 153–186.

[Dolejší et al.(2013)] Dolejší, V., Ern, A. & Vohralík, M. (2013) A framework for robust a posteriori error control in unsteady nonlinear advection-diffusion problems. SIAM J. Numer. Anal., **51**, 773–793.

[Dolejší et al.(2016)] Dolejší, V., Ern, A. & Vohralík, M. (2016) *hp*-adaptation driven by polynomial-degree-robust a posteriori error estimates for elliptic problems. SIAM J. Sci. Comput., **38**, A3220–A3246.

[Eriksson & Johnson(1995)] Eriksson, K. & Johnson, C. (1995) 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**, 706–740.

[Ern et al(2017a)] Ern, A., Smears, I. & Vohralík, M. (2017a) Discrete p-robust discrete p-robust $\boldsymbol{H}(\mathrm{div})$-liftings and a posteriori error analysis of elliptic problems with H^{-1} source terms. Calcolo, **54**, 1009–1025, doi:10.1007/s10092-017-0217-4.

[Ern et al.(2017b)] Ern, A., Smears, I. & Vohralík, M. (2017b) Guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimates for high-order discretizations of parabolic problems. SIAM J. Numer. Anal., **55**, 2811–2834, doi:10.1137/16M1097626.

[Ern & Schieweck(2016)] Ern, A. & Schieweck, F. (2016) Discontinuous Galerkin method in time combined with a stabilized finite element method in space for linear first-order PDEs. Math. Comp., **85**, 2099–2129.

[Ern & Vohralík(2010)] Ern, A. & Vohralík, M. (2010) A posteriori error estimation based on potential and flux reconstruction for the heat equation. SIAM J. Numer. Anal., **48**, 198–223.

[Ern & Vohralík(2015)] Ern, A. & Vohralík, M. (2015) Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations. SIAM J. Numer. Anal., **53**, 1058–1081.

[Ern & Vohralík(2016)] Ern, A. & Vohralík, M. (2016) Stable broken H^1 and $\boldsymbol{H}(\mathrm{div})$ polynomial extensions for polynomial-degree-robust potential and flux reconstruction in three space dimensions. Submitted for publication. Preprint available at `https://hal.archives-ouvertes.fr/hal-01422204`.

[Gaspoz et al.(2016)] Gaspoz, F., Kreuzer, C., Siebert, K. & Ziegler, D. (2016) A convergent time-space adaptive dG(s) finite element method for parabolic problems motivated by equal error distribution. Submitted for publication. Preprint available at `https://arxiv.org/abs/1610.06814`.

[Georgoulis et al.(2011)] Georgoulis, E. H., Lakkis, O. & Virtanen, J. M. (2011) A posteriori error control for discontinuous Galerkin methods for parabolic problems. SIAM J. Numer. Anal., **49**, 427–458.

[Georgoulis et al.(2017)] Georgoulis, E. H., Lakkis, O. & Wihler, T. P. (2017) A posteriori error bounds for fully-discrete *hp*-discontinuous Galerkin timestepping methods for parabolic problems. ArXiv e-prints 1708.05832.

[Kreuzer et al.(2012)] KREUZER, C., MÖLLER, C. A., SCHMIDT, A. & SIEBERT, K. G. (2012) Design and convergence analysis for an adaptive discretization of the heat equation. IMA J. Numer. Anal., **32**, 1375–1403.

[Kreuzer(2013)] KREUZER, C. (2013) Reliable and efficient a posteriori error estimates for finite element approximations of the parabolic p-Laplacian. Calcolo, **50**, 79–110.

[Lakkis & Makridakis(2006)] LAKKIS, O. & MAKRIDAKIS, C. (2006) Elliptic reconstruction and a posteriori error estimates for fully discrete linear parabolic problems. Math. Comp., **75**, 1627–1658.

[Makridakis & Nochetto(2003)] MAKRIDAKIS, C. & NOCHETTO, R. H. (2003) Elliptic reconstruction and a posteriori error estimates for parabolic problems. SIAM J. Numer. Anal., **41**, 1585–1594.

[Makridakis & Nochetto(2006)] MAKRIDAKIS, C. & NOCHETTO, R. H. (2006) A posteriori error analysis for higher order dissipative methods for evolution problems. Numer. Math., **104**, 489–514.

[Nicaise & Soualem(2005)] NICAISE, S. & SOUALEM, N. (2005) A posteriori error estimates for a nonconforming finite element discretization of the heat equation. M2AN Math. Model. Numer. Anal., **39**, 319–348.

[Nochetto et al.(2009)] NOCHETTO, R. H., SIEBERT, K. G. & VEESER, A. (2009) Theory of adaptive finite element methods: an introduction. Multiscale, nonlinear and adaptive approximation. Springer, Berlin, pp. 409–542.

[Picasso(1998)] PICASSO, M. (1998) Adaptive finite elements for a linear parabolic problem. Comput. Methods Appl. Mech. Engrg., **167**, 223–237.

[Repin(2002)] REPIN, S. (2002) Estimates of deviations from exact solutions of initial-boundary value problem for the heat equation. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl., **13**, 121–133.

[Schötzau & Wihler(2010)] SCHÖTZAU, D. & WIHLER, T. P. (2010) A posteriori error estimation for hp-version time-stepping methods for parabolic partial differential equations. Numer. Math., **115**, 475–509.

[Schwab(1998)] SCHWAB, C. (1998) p- and hp-finite element methods. Numerical Mathematics and Scientific Computation. New York: The Clarendon Press Oxford University Press, pp. xii+374.

[Smears(2017)] SMEARS, I. (2017) Robust and efficient preconditioners for the discontinuous Galerkin time-stepping method. IMA J. Numer. Anal., **37**, 1961–1985, doi:10.1093/imanum/drw050.

[Tantardini & Veeser(2016)] TANTARDINI, F. & VEESER, A. (2016) The L^2-projection and quasi-optimality of Galerkin methods for parabolic equations. SIAM J. Numer. Anal., **54**, 317–340.

[Verfürth(1998)] Verfürth, R. (1998) A posteriori error estimates for nonlinear problems. $L^r(0,T;L^\rho(\Omega))$-error estimates for finite element discretizations of parabolic equations. Math. Comp., **67**, 1335–1360.

[Verfürth(1998)] Verfürth, R. (1998) A posteriori error estimates for nonlinear problems: $L^r(0,T;W^{1,\rho}(\Omega))$-error estimates for finite element discretizations of parabolic equations. Numer. Methods Partial Differential Equations, **14**, 487–518.

[Verfürth(2003)] Verfürth, R. (2003) A posteriori error estimates for finite element discretizations of the heat equation. Calcolo, **40**, 195–212.

[Verfürth(2013)] Verfürth, R. (2013) A posteriori error estimation techniques for finite element methods. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, pp. xx+393.

[Wloka(1987)] Wloka, J. (1987) Partial differential equations. Cambridge: Cambridge University Press, pp. xii+518.