SIAM J. NUMER. ANAL.
Vol. 55, No. 6, pp. 2811–2834

GUARANTEED, LOCALLY SPACE-TIME EFFICIENT, AND POLYNOMIAL-DEGREE ROBUST A POSTERIORI ERROR ESTIMATES FOR HIGH-ORDER DISCRETIZATIONS OF PARABOLIC PROBLEMS*

ALEXANDRE ERN†, IAIN SMEARS‡, AND MARTIN VOHRALÍK‡

Abstract. We consider the a posteriori error analysis of approximations of parabolic problems based on arbitrarily high-order conforming Galerkin spatial discretizations and arbitrarily high-order discontinuous Galerkin temporal discretizations. Using equilibrated flux reconstructions, we present a posteriori error estimates for a norm composed of the $L^2(H^1)\cap H^1(H^{-1})$-norm of the error and the temporal jumps of the numerical solution. The estimators provide guaranteed upper bounds for this norm without unknown constants. Furthermore, the efficiency of the estimators with respect to this norm is local in both space and time, with constants that are robust with respect to the mesh-size, time-step size, and the spatial and temporal polynomial degrees. We further show that this norm, which is key for local space-time efficiency, is globally equivalent to the $L^2(H^1)\cap H^1(H^{-1})$-norm of the error, with polynomial-degree robust constants. The proposed estimators also have the practical advantage of being robust with respect to refinement and coarsening between the time steps.

Key words. parabolic partial differential equations, a posteriori error estimates, local space-time efficiency, polynomial-degree robustness, high-order methods

AMS subject classifications. 65M15, 65M60

DOI. 10.1137/16M1097626

1. Introduction.

We consider the heat equation

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

where $\Omega\subset\mathbb{R}^d$, $1\le d\le 3$, is a bounded, connected, polyhedral 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 developing a posteriori error estimates for a class of high-order discretizations of (1.1). In particular, we consider a conforming finite element method (FEM) in space on unstructured shape-regular simplicial meshes, and a discontinuous Galerkin discretization in time, where one is free to vary the approximation orders p in space and q in time, as well as the mesh size h and time-step size τ, leading to what we call an hp-τq method. These methods are highly attractive from the point of view of flexibility, accuracy, and computational efficiency, since it is known from a priori analysis that judicious local adaptation of the discretization parameters can lead to exponential convergence rates with respect to the number of degrees of freedom, even for solutions with singularities near domain corners, edges, and at initial times [37, 39, 47]. In practice, it is desirable to determine the adaptation

*Received by the editors October 6, 2016; accepted for publication (in revised form) August 7, 2017; published electronically November 14, 2017.
http://www.siam.org/journals/sinum/55-6/M109762.html
Funding: The work of the authors was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 647134 GATIPOR).
†Université Paris-Est, CERMICS (ENPC), 77455 Marne-la-Vallée 2, France and Inria Paris, 2 rue Simone Iff, 75589 Paris, France (alexandre.ern@enpc.fr).
‡Inria Paris, 2 rue Simone Iff, 75589 Paris, France and Université Paris-Est, CERMICS (ENPC), 77455 Marne-la-Vallée 2, France (iain.smears@inria.fr, martin.vohralik@inria.fr).

algorithmically, which requires rigorous and high-quality a posteriori error control in order to exploit the potential for high accuracy and efficiency of hp-τq discretizations. We recall that a posteriori error estimates should ideally give *guaranteed* upper bounds on the error, i.e., without unknown constants, should be *locally efficient*, meaning that the local estimators should be bounded from above by the error measured in a local neighborhood, and, moreover, should be *robust*, with all constants in the bounds being independent of the discretization parameters; we refer the reader to [46] for an introduction to these concepts.

In the context of parabolic problems, the a posteriori error analysis for low- and fixed-order methods has received significant attention over the past decade, with efforts mostly concentrated on fixed-order FEM in space coupled with an implicit Euler or Crank–Nicolson time-stepping scheme, leading to estimates for a wide range of norms. These include estimates for the $L^2(H^1)$-norm of the error considered independently by Picasso [35] and Verfürth [44], with efficiency bounds typically requiring restrictions on the relation between the sizes of the time steps and the meshes. Estimates for the $L^2(H^1) \cap H^1(H^{-1})$-norm were first considered by Verfürth in [45], who crucially proved local-in-time yet global-in-space efficiency of estimators without restrictions between time-step and mesh sizes; see also Bergam, Bernardi, and Mghazli [1]. Guaranteed upper bounds for a large family of spatial discretizations were later obtained by Ern and Vohralík in [15], with similar efficiency results as in [45]. There are also upper bounds in $L^2(L^2)$, $L^\infty(L^2)$, and $L^\infty(L^\infty)$ and higher-order norms, based on either duality techniques as in Eriksson and Johnson [12] or the elliptic reconstruction technique originally due to Makridakis and Nochetto [30] and later considered in the fully discrete context by Lakkis and Makridakis [27]; see also [28] and the references therein. Repin [36] studied so-called functional estimates. Finally, a posteriori error estimates developed in the context of the heat equation often serve as a starting point for extensions to diverse applications, including nonlinear problems and spatially nonconforming methods among others [8, 9, 22, 25, 34]. Adaptive algorithms for parabolic problems are studied in [5, 21, 26].

It is apparent from the literature that, even for low- and fixed-order methods, there are remaining outstanding issues, particularly in terms of the efficiency of the estimators. The efficiency of the estimators is significantly influenced by the choice of norm to be estimated, with the strongest available results being attained by Y-norm estimates, where, henceforth, $Y \coloneqq L^2(H^1_0) \cap H^1(H^{-1})$. However, even in this norm, the full space-time local efficiency of the estimators is not known. It is helpful to examine here more closely this issue in order to motivate the approach adopted in this work. For example, let us momentarily consider an implicit Euler discretization in time and a conforming FEM in space, recalling that the implicit Euler method corresponds to the lowest-order discontinuous Galerkin time-stepping method, which uses piecewise constant approximations with respect to time. The resulting numerical solution $u_{h\tau}$ is discontinuous with respect to time, so it is not possible to estimate $\|u - u_{h\tau}\|_Y$. Therefore, it is usual to consider a reconstruction, denoted by $\mathcal{I}u_{h\tau} \in Y$, obtained by piecewise linear interpolation at the time-step nodes, and it is seemingly natural to seek a posteriori error estimates for $\|u - \mathcal{I}u_{h\tau}\|_Y$, where $\|\cdot\|_Y$ is defined in (2.1) below, and where u is the solution of (1.1); for instance, this corresponds to the approach adopted in [45]. However, the main issue in estimates for $\|u - \mathcal{I}u_{h\tau}\|_Y$ is that $\mathcal{I}u_{h\tau}$ fails to satisfy the Galerkin orthogonality property: instead, $\mathcal{I}u_{h\tau}$ satisfies

$$(1.2)\quad \int_{I_n} (f, v_{h\tau}) - (\partial_t \mathcal{I}u_{h\tau}, v_{h\tau}) - (\nabla \mathcal{I}u_{h\tau}, \nabla v_{h\tau})\mathrm{d}t = \int_{I_n} (\nabla(u_{h\tau} - \mathcal{I}u_{h\tau}), \nabla v_{h\tau})\mathrm{d}t$$

for all discrete test functions $v_{h\tau} \in V_{h\tau}$; see section 3 for complete definitions. The lack of Galerkin orthogonality for $\mathcal{I}u_{h\tau}$ is associated with the discrete residual on the right-hand side of (1.2) that involves the $L^2(H^1)$-norm of $u_{h\tau} - \mathcal{I}u_{h\tau}$, and it is this discrete residual that causes the loss of local spatial efficiency in previous analyses. This issue is independent of the specific construction of the error estimators, whether they are residual-type estimators as in [45] or equilibrated flux estimators as considered here. It turns out that the discrete residual in (1.2) is related to the temporal jumps in the numerical solution $u_{h\tau} \notin Y$, which is a form of error in itself since it is tied to the nonconformity of the numerical scheme. This motivates the introduction of a composite norm $\|u-u_{h\tau}\|_{\mathcal{E}_Y}$ that includes both $\|u-\mathcal{I}u_{h\tau}\|_Y$ and the $L^2(H^1)$-norm of $u_{h\tau}-\mathcal{I}u_{h\tau}$. Our analysis is then centered on the error estimation of $\|u-u_{h\tau}\|_{\mathcal{E}_Y}$ instead of $\|u - \mathcal{I}u_{h\tau}\|_Y$, and this allows us to recover the fully space-time local efficiency of the estimators: see (1.4) below, and see Theorem 5.2 of section 5.

For *hp*-FEM discretizations, one of the key issues concerns the robustness of the estimators with respect to the polynomial degree; this issue appears already in the context of elliptic problems, where Melenk and Wohlmuth [33] and Melenk [32] showed that the well-known residual estimators fail to be polynomial-degree robust. In a breakthrough work, Braess, Pillwein, and Schöberl [3] established the polynomial-degree robustness of estimators based on equilibrated fluxes, in the context of elliptic diffusion problems. These estimators are based on a globally $\boldsymbol{H}(\mathrm{div})$-conforming flux computed by solving independent local mixed finite element problems. The polynomial-degree robustness of these estimators was recently generalized to nonconforming and mixed methods in [16], to which we refer the reader for further references on elliptic problems. For parabolic problems, there is the additional question of robustness of the estimators with respect to the temporal polynomial degrees. In comparison to low- and fixed-order methods, there are comparatively few works on a posteriori error estimates for high-order discretizations of parabolic problems. Building on the earlier work of Makridakis and Nochetto [31], Schötzau and Wihler [38] studied the effect of the temporal approximation order of a posteriori estimates for a composite norm of $L^\infty(L^2) \cap L^2(H^1)$-type, in the context of high-order temporal semidiscretizations of abstract evolution equations. Otherwise, a posteriori error estimates for *hp*-τq discretizations of parabolic problems remain essentially untouched.

In this work, we present guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimators for *hp*-τq discretizations of parabolic problems. This is by no means simple, as it requires the treatment of the challenges that have been outlined above. Our main results are the following. Let the spaces $Y \coloneqq L^2(H_0^1) \cap H^1(H^{-1})$ and $X \coloneqq L^2(H_0^1)$ be, respectively, equipped with their standard norms $\|\cdot\|_Y$ and $\|\cdot\|_X$ defined in (2.1) below. Let $Y+V_{h\tau}$ be the sum of the continuous and approximate solution spaces, recalling that $V_{h\tau} \subset X$ and that $u_{h\tau} \in V_{h\tau} \not\subset Y$ due to the temporally discontinuous approximation. Let $\mathcal{I}\colon Y + V_{h\tau} \to Y$ be the reconstruction operator defined in section 3.5 below, where we note that $\mathcal{I}v = v$ if and only if $v \in Y$. Let the norm $\|\cdot\|_{\mathcal{E}_Y}$ be defined by $\|v\|_{\mathcal{E}_Y}^2 \coloneqq \|\mathcal{I}v\|_Y^2 + \|v - \mathcal{I}v\|_X^2$ for all $v \in Y + V_{h\tau}$. In particular, $\mathcal{I}u = u$, so $\|u - u_{h\tau}\|_{\mathcal{E}_Y}^2 = \|u - \mathcal{I}u_{h\tau}\|_Y^2 + \|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X^2$. Hence $\|\cdot\|_{\mathcal{E}_Y}$ has the natural functional interpretation as an extension of the Y-norm to the discrete approximation space.

Guaranteed upper bounds. In Theorem 5.2 of section 5, we show a posteriori estimates in the norm $\|\cdot\|_{\mathcal{E}_Y}$. In the absence of data oscillation, our bound takes the simple form

$$\|u - u_{h\tau}\|_{\mathcal{E}_Y}^2 \leq \sum_{n=1}^{N} \sum_{K\in\mathcal{T}^n} \left\{ \int_{I_n} \|\boldsymbol{\sigma}_{h\tau} + \nabla \mathcal{I}u_{h\tau}\|_K^2 + \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|_K^2 \,\mathrm{d}t \right\}, \tag{1.3}$$

where $\boldsymbol{\sigma}_{h\tau}$ is the $\boldsymbol{H}(\mathrm{div})$-conforming reconstruction; see sections 3 and 4 for full definitions of the notation and construction of the estimators.

Polynomial-degree robustness and local space-time efficiency. We establish local space-time efficiency of our estimators with polynomial-degree robust constants, expressed by the lower bound

$$\int_{I_n} \|\boldsymbol{\sigma}_{h\tau} + \nabla \mathcal{I} u_{h\tau}\|_K^2 + \|\nabla(u_{h\tau} - \mathcal{I} u_{h\tau})\|_K^2 \mathrm{d}t \lesssim \sum_{\mathbf{a} \in \mathcal{V}_K} |u - u_{h\tau}|^2_{\mathcal{E}_Y^{\mathbf{a},n}} + \text{oscillation}, \tag{1.4}$$

where K is an element of the mesh $\mathcal{T}^n$ for time step I_n, where $\mathcal{V}_K$ denotes the set of vertices of K, and where $|u - u_{h\tau}|_{\mathcal{E}_Y^{\mathbf{a},n}}$ is the local component of $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ on the patch associated with the vertex $\mathbf{a}$ and time interval I_n. Here, and in the following, the notation $a \lesssim b$ means that $a \leq Cb$ with a constant C that depends possibly on the shape regularity of the spatial meshes, but is otherwise independent of the mesh size, time-step size, as well as the spatial and temporal polynomial degrees. We stress that this efficiency bound does not require any relation between the sizes of the time step and the mesh. The full bound is stated in Theorem 5.2 below. In addition to the above results, the estimators proposed here are advantageous in terms of flexibility, since they do not require restrictions on coarsening or refinement between time steps that appeared in earlier works, such as the transition condition used in [45, pp. 196, 201]. The main tool to avoid this condition is Lemma 8.1 below.

Relation between $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ and $\|u - \mathcal{I} u_{h\tau}\|_Y$. The a posteriori analysis in this work concerns the estimation of $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$. Independently of the error estimation, we also consider the question of the relation between the new norm $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ and the previously considered norm $\|u - \mathcal{I} u_{h\tau}\|_Y$. Specifically, for arbitrary polynomial degrees, we show the global equivalence result

$$\|u - \mathcal{I} u_{h\tau}\|_Y \leq \|u - u_{h\tau}\|_{\mathcal{E}_Y} \leq 3\|u - \mathcal{I} u_{h\tau}\|_Y + \text{oscillation}. \tag{1.5}$$

The oscillation term in (1.5) is *the minimum* between the source term data oscillation and the coarsening error, as fully detailed in Theorem 5.1 of section 5.1 below. Notice that the constant in the equivalence is therefore robust with respect to all parameters. The proof is based on a simplification and generalization to the higher-order case of a key result of Verfürth [45], namely, that the jumps in the numerical solution can be controlled locally-in-time and globally-in-space by the Y-norm of $u - \mathcal{I} u_{h\tau}$. The key implication of (1.5) is that the global space-time norms $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ and $\|u - \mathcal{I} u_{h\tau}\|_Y$ are essentially equivalent, although we stress that their local (spatial) distributions may differ.

This paper is organized as follows. First, in section 2 we introduce a functional setting for the a posteriori error analysis. We find it worthwhile to provide a complete derivation of the inf-sup analysis of the problem, as we give here quantitatively sharp results that are advantageous for the efficiency of the estimators in practice. Section 3 defines the setting in terms of notation, finite element approximation spaces, and the numerical scheme. Then, in section 4, we define the equilibrated flux reconstruction used in the a posteriori error estimates. In section 5 we gather our main results underlying (1.3), (1.4), and (1.5). The proofs of the main results are treated in the subsequent sections: section 6 establishes the relation between $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ and $\|u - \mathcal{I} u_{h\tau}\|_Y$; the proof of the guaranteed upper bound is given in section 7; and the efficiency of the estimators is the subject of section 8.

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., $\|\cdot\| \coloneqq \|\cdot\|_\Omega$. Following [29, Chap. 3], we consider the function spaces $X \coloneqq L^2(0,T;H_0^1(\Omega))$ and $Y \coloneqq L^2(0,T;H_0^1(\Omega)) \cap H^1(0,T;H^{-1}(\Omega))$, with norms

$$\begin{aligned} \|\varphi\|_Y^2 &\coloneqq \int_0^T \|\partial_t\varphi\|_{H^{-1}(\Omega)}^2 + \|\nabla\varphi\|^2\,\mathrm{d}t + \|\varphi(T)\|^2 \quad &&\forall\,\varphi\in Y,\\ \|v\|_X^2 &\coloneqq \int_0^T \|\nabla v\|^2\,\mathrm{d}t &&\forall\, v\in X. \end{aligned} \tag{2.1}$$

Define the bilinear form $\mathcal{B}_Y\colon Y\times X\to\mathbb{R}$ by

$$\mathcal{B}_Y(\varphi,v) \coloneqq \int_0^T \langle\partial_t\varphi, v\rangle + (\nabla\varphi,\nabla v)\,\mathrm{d}t, \tag{2.2}$$

where $\varphi\in Y$ and $v\in X$ are arbitrary functions, and $\langle\cdot,\cdot\rangle$ denotes here the duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$. Then, the problem (1.1) admits the following weak formulation: find $u\in Y$ such that $u(0)=u_0$ and such that

$$\mathcal{B}_Y(u,v) = \int_0^T (f,v)\,\mathrm{d}t \quad \forall\, v\in X. \tag{2.3}$$

The well-posedness of (2.3) is well known and can be shown by inf-sup arguments in the above functional setting [13]; it can also be shown by Galerkin's method [20, 48]. The inf–sup stability result presented here has the interesting and important property of taking the form of an identity, which is advantageous for the sharpness of a posteriori error analysis. The fact that the constant equals 1 in (2.4) below can also be found in [42, 43]; it can also be seen from [24, p. 249].

THEOREM 2.1 (inf–sup identity). *For every $\varphi\in Y$, we have*

$$\|\varphi\|_Y^2 = \left[\sup_{v\in X\setminus\{0\}} \frac{\mathcal{B}_Y(\varphi,v)}{\|v\|_X}\right]^2 + \|\varphi(0)\|^2. \tag{2.4}$$

Proof. For a fixed $\varphi\in Y$, let $w_*\in X$ be defined by $(\nabla w_*,\nabla v)=\langle\partial_t\varphi,v\rangle$ for all $v\in H_0^1(\Omega)$, a.e. in $(0,T)$, which implies the identity $\|\nabla w_*\|^2=\|\partial_t\varphi\|_{H^{-1}(\Omega)}^2$ a.e. in $(0,T)$. Furthermore, we have $\mathcal{B}_Y(\varphi,v)=\int_0^T(\nabla(w_*+\varphi),\nabla v)\,\mathrm{d}t$, thus implying that $\sup_{v\in X\setminus\{0\}}\mathcal{B}_Y(\varphi,v)/\|v\|_X=\|w_*+\varphi\|_X$. We then obtain the desired identity (2.4) by expanding the square

$$\begin{aligned} \left[\sup_{v\in X\setminus\{0\}} \frac{\mathcal{B}_Y(\varphi,v)}{\|v\|_X}\right]^2 &= \int_0^T \|\nabla(w_*+\varphi)\|^2\,\mathrm{d}t\\ &= \int_0^T \|\nabla w_*\|^2 + 2(\nabla w_*,\nabla\varphi) + \|\nabla\varphi\|^2\,\mathrm{d}t\\ &= \int_0^T \|\partial_t\varphi\|_{H^{-1}(\Omega)}^2 + 2\langle\partial_t\varphi,\varphi\rangle + \|\nabla\varphi\|^2\,\mathrm{d}t\\ &= \|\varphi\|_Y^2 - \|\varphi(0)\|^2, \end{aligned} \tag{2.5}$$

where we note that we have used the identity $\int_0^T 2\langle\partial_t\varphi,\varphi\rangle\,\mathrm{d}t = \|\varphi(T)\|^2-\|\varphi(0)\|^2$. ∎

In order to estimate the error between the solution u of (1.1) and its approximation, we define the residual functional $\mathcal{R}_Y : Y \to X'$ by

$$\langle \mathcal{R}_Y(\varphi), v\rangle \coloneqq \mathcal{B}_Y(u-\varphi, v) = \int_0^T (f,v) - \langle \partial_t \varphi, v\rangle - (\nabla\varphi, \nabla v)\,\mathrm{d}t, \tag{2.6}$$

where $v \in X$ and $\varphi \in Y$. The dual norm of the residual is naturally defined by $\|\mathcal{R}_Y(\varphi)\|_{X'} \coloneqq \sup_{v\in X\setminus\{0\}} \langle \mathcal{R}_Y(\varphi), v\rangle / \|v\|_X$. Theorem 2.1 implies the following equivalence between the error and dual norm of the residual: for all $\varphi \in Y$, we have

$$\|u-\varphi\|_Y^2 = \|\mathcal{R}_Y(\varphi)\|_{X'}^2 + \|u_0 - \varphi(0)\|^2. \tag{2.7}$$

3. Finite element approximation. Consider a partition of the interval $(0,T)$ into time-step intervals $I_n \coloneqq (t_{n-1}, t_n)$ with $1 \le n \le N$, 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 $\tau_n \coloneqq t_n - t_{n-1}$ denote the local time-step size. We will not need any special assumptions about the relative sizes of the time steps to each other. We associate a temporal polynomial degree $q_n \ge 0$ 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. We consider a matching simplicial mesh $\mathcal{T}^n$ of the domain Ω for each $0 \le n \le N$, where we assume shape regularity of the meshes uniformly over all time steps. This allows us to treat many applications where the meshes are obtained by refinement or coarsening between time steps. We consider here only matching simplicial meshes for simplicity, although we indicate that mixed simplicial–parallelepiped meshes, possibly containing hanging nodes, can be also be treated; see [10] 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 $h_K \coloneqq \operatorname{diam} K$ denote the diameter of K. We associate a local spatial polynomial degree $p_K \ge 1$ 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. For a general matching simplicial mesh $\mathcal{T}$ with associated vector of polynomial degrees $\boldsymbol{p} = (p_K)_{K\in\mathcal{T}}$, $p_K \ge 1$ for all $K \in \mathcal{T}$, the $H_0^1(\Omega)$-conforming hp-finite element space $V_h(\mathcal{T}, \boldsymbol{p})$ is defined by

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

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

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

Functions in $V_{h\tau}$ are generally discontinuous with respect to the time variable at the partition points, although we take them to be left-continuous: for all $1 \le n \le N$, we define $v_{h\tau}(t_n)$ as the trace at t_n of the restriction $v_{h\tau}|_{I_n}$. Functions in $V_{h\tau}$ are thus left-continuous; moreover, they 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 $(\!|\cdot|\!)_n$, $0 \le n \le N-1$, are defined on $V_{h\tau}$ by

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

3.3. Refinement and coarsening. Similary to other works, e.g., [45, 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 $V_h^{n-1} + V_h^n \subseteq \widetilde{V_h^n} := V_h(\widetilde{\mathcal{T}^n}, \widetilde{\boldsymbol{p}}_n)$. For a function $v_{h\tau} \in V_{h\tau}$, we observe that $(\!|v_{h\tau}|\!)_{n-1} \in \widetilde{V_h^n}$ for each $1 \le n \le N$ since $v_{h\tau}(t_{n-1}) \in V_h^{n-1}$, $v_{h\tau}(t_{n-1}^+) \in V_h^n$, and $V_h^{n-1} + V_h^n \subseteq \widetilde{V_h^n}$. It is assumed that $\widetilde{\mathcal{T}^n}$ has the same shape regularity as $\mathcal{T}^{n-1}$ 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_h^{n-1} and V_h^n. We note that in the present context, refinement and coarsening can be obtained by modification of the meshes as well as change in the polynomial degrees. Concerning the polynomial degrees, we may choose, for example, $p_{\widetilde{K}} = \max(p_{K'}, p_{K''})$. In the case where V_h^n is obtained from V_h^{n-1} by refinement without coarsening, then we may choose $\widetilde{\mathcal{T}^n} := \mathcal{T}^n$ and $\widetilde{\boldsymbol{p}}_n := \boldsymbol{p}_n$ so that $\widetilde{V_h^n} = V_h^n$. However, we do not need the transition condition assumption from [45, pp. 196, 201], which requires a uniform bound on the ratio of element sizes between $\widetilde{\mathcal{T}^n}$ and $\mathcal{T}^n$.

3.4. Numerical scheme. The numerical scheme for approximating the solution of the parabolic problem (1.1) consists of finding $u_{h\tau} \in V_{h\tau}$ such that $u_{h\tau}(0) = \Pi_h u_0$, and, for each time-step interval I_n,

$$\begin{aligned}\int_{I_n} (\partial_t u_{h\tau}, v_{h\tau}) + (\nabla u_{h\tau}, \nabla v_{h\tau})\,\mathrm{d}t - \left((\!|u_{h\tau}|\!)_{n-1}, v_{h\tau}(t_{n-1}^+)\right)& \\ = \int_{I_n} (f, v_{h\tau})\,\mathrm{d}t \qquad \forall\, v_{h\tau} \in \mathcal{Q}_{q_n}(I_n; V_h^n).&\end{aligned} \tag{3.4}$$

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 lowest-order case; this can be performed efficiently in practice for arbitrary orders; see [40] and the references therein.

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 standard qth 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 $\{L_q^n\}_{q\ge 0}$ 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 $q \ge 0$. We introduce the Radau reconstruction operator $\mathcal{I}$ defined on $V_{h\tau}$ by

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

It is clear that $\mathcal{I}$ is a linear operator on $V_{h\tau}$. It follows from the properties of the Legendre polynomials 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 $1 \leq n \leq N$. Therefore, $\mathcal{I}v_{h\tau}$ is continuous with respect to the temporal variable at the interval partition points $\{t_n\}_{n=0}^{N-1}$, and thus we have

$$\mathcal{I}v_{h\tau} \in H^1\left(0,T;H_0^1(\Omega)\right) \subset Y, \quad \mathcal{I}v_{h\tau}|_{I_n} \in \mathcal{Q}_{q_n+1}\big(I_n;\widetilde{V_h^n}\big) \quad \forall\, v_{h\tau} \in V_{h\tau}, \tag{3.6}$$

where we recall that $V_h^{n-1} + V_h^n \subseteq \widetilde{V_h^n}$. We easily deduce the following property of the reconstruction operator $\mathcal{I}$ from integration by parts and the orthogonality of the polynomials $L_{q_n}^n$ and $L_{q_n+1}^n$ to all polynomials of degree strictly less than q_n on the time-step interval I_n:

$$\int_{I_n} \partial_t \mathcal{I}v_{h\tau}\,\phi\,\mathrm{d}t = \int_{I_n} \partial_t v_{h\tau}\,\phi\,\mathrm{d}t - [\![v_{h\tau}]\!]_{n-1}\phi\left(t_{n-1}^+\right) \quad \forall\,\phi \in \mathcal{Q}_{q_n}(I_n;\mathbb{R}), \tag{3.7}$$

where equality holds in the above equation in the sense of functions in $\widetilde{V_h^n}$. We may therefore use (3.7) to rewrite the numerical scheme (3.4) as

$$\int_{I_n} (\partial_t \mathcal{I}u_{h\tau}, v_{h\tau}) + (\nabla u_{h\tau}, \nabla v_{h\tau})\,\mathrm{d}t = \int_{I_n} (f, v_{h\tau})\,\mathrm{d}t \quad \forall\, v_{h\tau} \in \mathcal{Q}_{q_n}(I_n;V_h^n). \tag{3.8}$$

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

Remark 3.1 (alternative equivalent definitions). The operator $\mathcal{I}$ is the Radau reconstruction operator commonly used in the a posteriori error analysis of the discontinuous Galerkin time-stepping method [31] and in the a priori error analysis of time-dependent first-order PDEs [14]. Several equivalent definitions of $\mathcal{I}$ have appeared in the literature, although it will be particularly advantageous for our purposes to use the definition (3.5) of $\mathcal{I}$; see [23, 40] for further details on the equivalence of the various definitions.

Remark 3.2 (extensions of $\mathcal{I}$ to $Y + V_{h\tau}$). In what follows, it will be helpful to extend $\mathcal{I}$ to a linear operator over $Y + V_{h\tau}$. Note that the definition of the jump operators (3.3) can be naturally extended to $Y + V_{h\tau}$ and, therefore, the definition (3.5) also extends naturally to $Y + V_{h\tau}$. In particular, $\mathcal{I}\colon Y + V_{h\tau} \to Y$, and we have $\mathcal{I}\varphi = \varphi$ if and only if $\varphi \in Y$, since the jumps of any $\varphi \in Y$ vanish identically.

Remark 3.3 (a priori analysis in the Y-norm). Although we focus here on the a posteriori analysis of the error, it is helpful to briefly mention some results from a priori analysis in the current functional framework. An important yet perhaps nonobvious point is that the inclusion of the time derivative in the Y-norm does not necessarily decrease the convergence order with respect to the time-step size in comparison to other norms, such as the X-norm. Indeed, since this is primarily related to the temporal discretization, let us show this by momentarily considering a temporal semidiscretization with solution u_τ obtained by replacing all discrete spaces V_h^n by $H_0^1(\Omega)$ in (3.2) and in (3.4). Then, with f_τ as defined in section 4.2, we have $\int_0^T \langle \partial_t(u - \mathcal{I}u_\tau), v\rangle\,\mathrm{d}t = \int_0^T (f - f_\tau, v) - (\nabla(u-u_\tau), \nabla v)\,\mathrm{d}t$ for all $v \in X$. Therefore, we deduce that

$$\int_0^T \|\partial_t(u - \mathcal{I}u_\tau)\|_{H^{-1}(\Omega)}^2\,\mathrm{d}t \leq \left\{\|u - u_\tau\|_X + \|f - f_\tau\|_{X'}\right\}^2.$$

This explains in a nusthell why the time derivative error is not necessarily worse than the other terms. There are some works on Y-norm-type estimates for the fully discrete case, going back to Dupont [6]. We also mention the recent analysis on quasi-optimality in the spatially semidiscrete case by Tantardini [41] and Tantardini and Veeser [42].

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 a data approximation defined in (4.4) below. We call $\boldsymbol{\sigma}_{h\tau}$ an equilibrated flux. We consider here the natural extension of existing flux reconstructions for elliptic problems [3, 4, 7, 16] to the parabolic setting; see also [11]. In particular, for each time step, $\boldsymbol{\sigma}_{h\tau}$ is obtained as a sum of fluxes computed by solving local mixed finite element problems over the vertex-based patches of the current mesh; see Definition 4.1 of section 4.3 below.

4.1. Local mixed finite element spaces. We now define the mixed finite element spaces that are required for the construction of the equilibrated flux. For each $1 \le n \le N$, let $\mathcal{V}^n$ denote the set of vertices of the mesh $\mathcal{T}^n$, where we distinguish the set of interior vertices $\mathcal{V}^n_{\mathrm{int}}$ and the set of boundary vertices $\mathcal{V}^n_{\mathrm{ext}}$. For each $\mathbf{a} \in \mathcal{V}^n$, let $\psi_{\mathbf{a}}$ denote the hat function associated with $\mathbf{a}$, and let $\omega_{\mathbf{a}}$ denote the interior of the support of $\psi_{\mathbf{a}}$ with associated diameter $h_{\omega_{\mathbf{a}}}$. Furthermore, let $\widetilde{\mathcal{T}^{\mathbf{a},n}}$ denote the restriction of the mesh $\widetilde{\mathcal{T}^n}$ to $\omega_{\mathbf{a}}$. Recalling that the common refinement spaces $\widetilde{V^n_h}$ 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 $\mathbf{a} \in \mathcal{V}^n$ the fixed polynomial degree

$$p_{\mathbf{a}} := \max_{\widetilde{K}\in\widetilde{\mathcal{T}^{\mathbf{a},n}}} (p_{\widetilde{K}} + 1). \tag{4.2}$$

Observe that $\psi_{\mathbf{a}} \partial_t \mathcal{I} u_{h\tau}|_{\widetilde{K}\times I_n}$ is a polynomial function with degree at most q_n in time and at most $p_{\mathbf{a}}$ in space for each $\widetilde{K} \in \widetilde{\mathcal{T}^{\mathbf{a},n}}$, $1 \le n \le N$.

For a polynomial degree $p \ge 0$, let the local spaces $\mathcal{P}_p(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ and $\mathbf{RTN}_p(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ be defined by

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

where $\mathbf{RTN}_p(\widetilde{K}) := \mathcal{P}_p(\widetilde{K};\mathbb{R}^d) + \mathcal{P}_p(\widetilde{K})\boldsymbol{x}$ denotes the Raviart–Thomas–Nédélec space of order p on $\widetilde{K}$. It is important to notice that, whereas the patch $\omega_{\mathbf{a}}$ is subordinate to the vertices of the mesh $\mathcal{T}^n$, the spaces $\mathcal{P}_p(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ and $\mathbf{RTN}_p(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ are subordinate to the submesh $\widetilde{\mathcal{T}^{\mathbf{a},n}}$; of course, in the absence of coarsening, this distinction vanishes.

We now introduce the local spatial mixed finite element spaces $\boldsymbol{V}^{\mathbf{a},n}_h$ and $Q^{\mathbf{a},n}_h$, defined by

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

$$Q^{\mathbf{a},n}_h := \begin{cases} \left\{q_h \in \mathcal{P}_{p_{\mathbf{a}}}(\widetilde{\mathcal{T}^{\mathbf{a},n}}), \quad (q_h,1)_{\omega_{\mathbf{a}}} = 0\right\} & \text{if } \mathbf{a} \in \mathcal{V}^n_{\mathrm{int}}, \\ \mathcal{P}_{p_{\mathbf{a}}}(\widetilde{\mathcal{T}^{\mathbf{a},n}}) & \text{if } \mathbf{a} \in \mathcal{V}^n_{\mathrm{ext}}. \end{cases}$$

We then define the following space-time mixed finite element spaces

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

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. First, we define the semidiscrete approximation f_τ of f by an L^2-orthogonal projection in time. In particular, the approximation $f_\tau \in \mathcal{Q}_{q_n}(I_n; L^2(\Omega))$ is defined on each interval I_n by $\int_{I_n}(f - f_\tau, v)\,\mathrm{d}t = 0$ for all $v \in \mathcal{Q}_{q_n}(I_n; L^2(\Omega))$. Next, for each $1 \le n \le N$ and for each $\mathbf{a} \in \mathcal{V}^n$, let $\Pi_{h\tau}^{\mathbf{a},n}$ be the $L^2_{\psi_{\mathbf{a}}}$-orthogonal projection from $L^2(I_n; L^2_{\psi_{\mathbf{a}}}(\omega_{\mathbf{a}}))$ onto $\mathcal{Q}_{q_n}(I_n; \mathcal{P}_{p_{\mathbf{a}}-1}(\widetilde{\mathcal{T}^{\mathbf{a},n}}))$, where $L^2_{\psi_{\mathbf{a}}}(\omega_{\mathbf{a}})$ is the space of measurable functions v on $\omega_{\mathbf{a}}$ such that $\int_{\omega_{\mathbf{a}}} \psi_{\mathbf{a}} |v|^2\,\mathrm{d}x < \infty$. In other words, the projection operator $\Pi_{h\tau}^{\mathbf{a},n}$ is defined by $\int_{I_n}(\psi_{\mathbf{a}}\Pi_{h\tau}^{\mathbf{a},n} v, q_{h\tau})_{\omega_{\mathbf{a}}}\,\mathrm{d}t = \int_{I_n}(\psi_{\mathbf{a}} v, q_{h\tau})_{\omega_{\mathbf{a}}}\,\mathrm{d}t$ for all $q_{h\tau} \in \mathcal{Q}_{q_n}(I_n; \mathcal{P}_{p_{\mathbf{a}}-1}(\widetilde{\mathcal{T}^{\mathbf{a},n}}))$. We adopt the convention that $\Pi_{h\tau}^{\mathbf{a},n} v$ is extended by zero from $\omega_{\mathbf{a}} \times I_n$ to $\Omega \times (0,T)$ for all $v \in L^2(I_n; L^2_{\psi_{\mathbf{a}}}(\omega_{\mathbf{a}}))$. Then, we define $f_{h\tau}$ by

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

Remark 4.1 (definition of $f_{h\tau}$). The somewhat technical appearance of the definition of $f_{h\tau}$ is due to the possible variation in polynomial degrees across the mesh and the particular requirements of the analysis of efficiency, in particular, the hypotheses of Lemma 8.1 below. Nevertheless, $f_{h\tau}$ has several important approximation properties. First, for any $1 \le n \le N$, any $\widetilde{K} \in \widetilde{\mathcal{T}}^n$, and any real-valued polynomial ϕ of degree at most q_n, we have

$$\int_{I_n} (f - f_{h\tau}, 1)_{\widetilde{K}} \phi \,\mathrm{d}t = \sum_{\mathbf{a}\in\mathcal{V}_K} \int_{I_n} \left(\psi_{\mathbf{a}} \left(f - \Pi_{h\tau}^{\mathbf{a},n} f\right), \phi 1\right)_{\widetilde{K}}\,\mathrm{d}t = 0, \tag{4.5}$$

where $\mathcal{V}_K$ denotes the set of vertices of K, and where we use the fact that the hat functions $\{\psi_{\mathbf{a}}\}_{\mathbf{a}\in\mathcal{V}^n}$ form a partition of unity on Ω. Furthermore, using the orthogonality of the projector $\Pi_{h\tau}^{\mathbf{a},n}$ and the fact that $0 \le \psi_{\mathbf{a}} \le 1$ in Ω, it is straightforward to show that

$$\|f - f_{h\tau}\|_{L^2(I_n;L^2(\widetilde{K}))} \le \sqrt{d+1} \inf_{w_{h\tau}\in\mathcal{Q}_{q_n}(I_n;\mathcal{P}_{p_{\widetilde{K}}}(\widetilde{K}))} \|f - w_{h\tau}\|_{L^2(I_n;L^2(\widetilde{K}))}. \tag{4.6}$$

This shows that $f_{h\tau}$ defines an approximation of f that is at least of the same order as the one associated with the finite element approximation. Furthermore, the approximations are exact, i.e., $f = f_\tau = f_{h\tau}$, if f is a piecewise polynomial of appropriate degrees with respect to the time steps and the common refinement meshes.

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

$$\boldsymbol{\tau}_{h\tau}^{\mathbf{a},n} \coloneqq -\psi_{\mathbf{a}} \nabla u_{h\tau}|_{\omega_{\mathbf{a}}\times I_n}, \tag{4.7a}$$

$$g_{h\tau}^{\mathbf{a},n} \coloneqq \psi_{\mathbf{a}} \left(\Pi_{h\tau}^{\mathbf{a},n} f - \partial_t \mathcal{I} u_{h\tau}\right)|_{\omega_{\mathbf{a}}\times I_n} - \nabla\psi_{\mathbf{a}} \cdot \nabla u_{h\tau}|_{\omega_{\mathbf{a}}\times I_n}. \tag{4.7b}$$

We claim that for all $\mathbf{a} \in \mathcal{V}^n_{\mathrm{int}}$,

$$\left(g_{h\tau}^{\mathbf{a},n}(t), 1\right)_{\omega_{\mathbf{a}}} = 0 \quad \forall\, t \in I_n, \tag{4.8}$$

which is equivalent to showing that $g_{h\tau}^{\mathbf{a},n} \in Q_{h\tau}^{\mathbf{a},n}$ for all $\mathbf{a} \in \mathcal{V}^n$. Indeed, we first observe that the construction of the numerical scheme, in particular, identity (3.8), implies that, for any univariate real-valued polynomial ϕ of degree at most q_n on I_n,

$$\int_{I_n} (g_{h\tau}^{\mathbf{a},n}, \phi 1)_{\omega_{\mathbf{a}}}\,\mathrm{d}t = \int_{I_n} \big(f, \phi\,\psi_{\mathbf{a}}\big)_{\omega_{\mathbf{a}}} - \big(\partial_t \mathcal{I} u_{h\tau}, \phi\,\psi_{\mathbf{a}}\big)_{\omega_{\mathbf{a}}} - \big(\nabla u_{h\tau}, \nabla(\phi\,\psi_{\mathbf{a}})\big)_{\omega_{\mathbf{a}}}\,\mathrm{d}t = 0,$$

where we have used the orthogonality of the projection $\Pi_{h\tau}^{\mathbf{a},n}$ and the fact that $\phi\psi_{\mathbf{a}} \in \mathcal{Q}_{q_n}(I_n; V_h^n)$ is a valid test function in (3.8). Since the function $g_{h\tau}^{\mathbf{a},n}$ is polynomial in time with degree at most q_n, i.e., $g_{h\tau}^{\mathbf{a},n} \in \mathcal{Q}_{q_n}(I_n; \mathcal{P}_{p_{\mathbf{a}}}(\widetilde{\mathcal{T}^{\mathbf{a},n}}))$, we deduce (4.8).

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 $\mathbf{a} \in \mathcal{V}^n$, let the space-time mixed finite element spaces $\boldsymbol{V}_{h\tau}^{\mathbf{a},n}$ and $Q_{h\tau}^{\mathbf{a},n}$ be defined by* (4.3)*. Let $g_{h\tau}^{\mathbf{a},n}$ and $\boldsymbol{\tau}_{h\tau}^{\mathbf{a},n}$ be defined by* (4.7)*. Let $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n}$ be defined by*

$$\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} \coloneqq \operatorname*{argmin}_{\substack{\boldsymbol{v}_{h\tau} \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n} \\ \nabla\cdot\boldsymbol{v}_{h\tau} = g_{h\tau}^{\mathbf{a},n}}} \int_{I_n} \|\boldsymbol{v}_{h\tau} - \boldsymbol{\tau}_{h\tau}^{\mathbf{a},n}\|_{\omega_{\mathbf{a}}}^2\,\mathrm{d}t. \tag{4.9}$$

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

$$\boldsymbol{\sigma}_{h\tau} \coloneqq \sum_{n=1}^{N} \sum_{\mathbf{a} \in \mathcal{V}^n} \boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}. \tag{4.10}$$

Note that $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n}$ is well-defined for all $\mathbf{a} \in \mathcal{V}^n$; in particular, for interior vertices $\mathbf{a} \in \mathcal{V}_{\mathrm{int}}^n$, we use (4.8) to guarantee the compatibility of the datum $g_{h\tau}^{\mathbf{a},n}$ with the constraint $\nabla\cdot\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} = g_{h\tau}^{\mathbf{a},n}$.

The following key result shows that $\boldsymbol{\sigma}_{h\tau}$ from Definition 4.1 leads to an equilibrated flux.

THEOREM 4.2 (equilibration). *Let the flux reconstruction $\boldsymbol{\sigma}_{h\tau}$ be defined by* (4.10) *of Definition* 4.1*. Then $\boldsymbol{\sigma}_{h\tau} \in L^2(0,T; \boldsymbol{H}(\mathrm{div},\Omega))$ and we have* (4.1)*, where the discrete approximation $f_{h\tau}$ is defined in* (4.4)*.*

Proof. After extending each $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}$ by zero from $\omega_{\mathbf{a}} \times I_n$ to $\Omega \times (0,T)$, we have $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} \in L^2(0,T; \boldsymbol{H}(\mathrm{div},\Omega))$ as a consequence of the boundary conditions included in the definition of the space $\boldsymbol{V}_h^{\mathbf{a},n}$. This immediately implies that $\boldsymbol{\sigma}_{h\tau} \in L^2(0,T; \boldsymbol{H}(\mathrm{div},\Omega))$. To show (4.1), the definition of the flux reconstruction $\boldsymbol{\sigma}_{h\tau}$ in (4.10) implies that for any time-step interval I_n and any $K \in \mathcal{T}^n$,

$$\begin{aligned}
\nabla\cdot\boldsymbol{\sigma}_{h\tau}|_{K\times I_n} &= \sum_{\mathbf{a}\in\mathcal{V}_K} \nabla\cdot\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}|_{K\times I_n} = \sum_{\mathbf{a}\in\mathcal{V}_K} g_{h\tau}^{\mathbf{a},n}|_{K\times I_n} \\
&= \sum_{\mathbf{a}\in\mathcal{V}_K} \big(\psi_{\mathbf{a}} \Pi_{h\tau}^{\mathbf{a},n} f - \psi_{\mathbf{a}} \partial_t \mathcal{I} u_{h\tau} - \nabla\psi_{\mathbf{a}} \cdot \nabla u_{h\tau}\big)|_{K\times I_n} \\
&= (f_{h\tau} - \partial_t \mathcal{I} u_{h\tau})|_{K\times I_n},
\end{aligned} \tag{4.11}$$

where $\mathcal{V}_K$ denotes the set of vertices of K, where we use the fact that the hat functions $\{\psi_{\mathbf{a}}\}_{\mathbf{a}\in\mathcal{V}^n}$ form a partition of unity in order to pass to the last line of (4.11), and where we have used the definition of $f_{h\tau}$ in (4.4). This yields (4.1) as required. $\square$

For the purposes of practical implementation, it is easily seen that, for each time-step interval I_n, the fluxes $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}$ can be computed by solving $q_n + 1$ independent spatial mixed finite element problems, provided only that an orthogonal or orthonormal polynomial basis is used in time over I_n. Moreover, the $q_n + 1$ linear systems each share the same matrix, which helps to simplify the implementation and reduce the computational cost.

LEMMA 4.3 (decoupling). *Let $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n}$ be defined by (4.9). Then $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}$ is equivalently uniquely defined as the first component of the pair $(\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}, r_{h\tau}^{\mathbf{a},n}) \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n} \times Q_{h\tau}^{\mathbf{a},n}$ that solves*

$$\int_{I_n} (\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}, \boldsymbol{v}_{h\tau})_{\omega_\mathbf{a}} - (\nabla\cdot \boldsymbol{v}_{h\tau}, r_{h\tau}^{\mathbf{a},n})_{\omega_\mathbf{a}} \,\mathrm{d}t = \int_{I_n} (\boldsymbol{\tau}_{h\tau}^{\mathbf{a},n}, \boldsymbol{v}_{h\tau})_{\omega_\mathbf{a}} \,\mathrm{d}t \quad \forall\, \boldsymbol{v}_{h\tau} \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n}, \tag{4.12a}$$

$$\int_{I_n} (\nabla\cdot \boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}, q_{h\tau})_{\omega_\mathbf{a}} \,\mathrm{d}t = \int_{I_n} (g_{h\tau}^{\mathbf{a},n}, q_{h\tau})_{\omega_\mathbf{a}} \,\mathrm{d}t \quad \forall\, q_{h\tau} \in Q_{h\tau}^{\mathbf{a},n}. \tag{4.12b}$$

Furthermore, for each $1 \le n \le N$, let $\{\phi_j^n\}_{j=0}^{q_n}$ be an $L^2(I_n)$-orthonormal basis for the space of univariate real-valued polynomials of degree at most q_n. For each $\mathbf{a} \in \mathcal{V}^n$, define the functions $\{g_{h,j}^{\mathbf{a},n}\}_{j=0}^{q_n}$ and $\{\boldsymbol{\tau}_{h,j}^{\mathbf{a},n}\}_{j=0}^{q_n}$ over the patch $\omega_\mathbf{a}$ by

$$g_{h,j}^{\mathbf{a},n} := \int_{I_n} g_{h\tau}^{\mathbf{a},n} \phi_j^n \,\mathrm{d}t, \quad \boldsymbol{\tau}_{h,j}^{\mathbf{a},n} := \int_{I_n} \boldsymbol{\tau}_{h\tau}^{\mathbf{a},n} \phi_j^n \,\mathrm{d}t. \tag{4.13}$$

Then, the solution $(\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}, r_{h\tau}^{\mathbf{a},n})$ of (4.12) can be obtained by solving the following spatial problems: for each $0 \le j \le q_n$, find $\boldsymbol{\sigma}_{h,j}^{\mathbf{a},n} \in \boldsymbol{V}_h^{\mathbf{a},n}$ and $r_{h,j}^{\mathbf{a},n}$ in $Q_h^{\mathbf{a},n}$ such that

$$\left(\boldsymbol{\sigma}_{h,j}^{\mathbf{a},n}, \boldsymbol{v}_h\right)_{\omega_\mathbf{a}} - \left(\nabla\cdot \boldsymbol{v}_h, r_{h,j}^{\mathbf{a},n}\right)_{\omega_\mathbf{a}} = \left(\boldsymbol{\tau}_{h,j}^{\mathbf{a},n}, \boldsymbol{v}_h\right)_{\omega_\mathbf{a}} \qquad \forall\, \boldsymbol{v}_h \in \boldsymbol{V}_h^{\mathbf{a},n}, \tag{4.14a}$$

$$\left(\nabla\cdot \boldsymbol{\sigma}_{h,j}^{\mathbf{a},n}, q_h\right)_{\omega_\mathbf{a}} = \left(g_{h,j}^{\mathbf{a},n}, q_h\right)_{\omega_\mathbf{a}} \qquad \forall\, q_h \in Q_h^{\mathbf{a},n}, \tag{4.14b}$$

and then by defining $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} := \sum_{j=0}^{q_n} \boldsymbol{\sigma}_{h,j}^{\mathbf{a},n} \phi_j^n$ and $r_{h\tau}^{\mathbf{a},n} := \sum_{j=0}^{q_n} r_{h,j}^{\mathbf{a},n} \phi_j^n$.

Remark 4.2 (implementation). Several techniques can be used to reduce the computational cost of computing the flux equilibration. First, although the flux $\boldsymbol{\sigma}_{h\tau}$ is defined on the space-time region $\Omega \times (0,T)$ from a theoretical viewpoint, in practice it is only its restriction to the current time step which is required, because values from previous time steps do not need to be reevaluated or even stored. Second, at each time step and at each vertex patch, a single matrix is shared by the decoupled local problems in (4.14), so a single factorization is sufficient. Third, if a patch and its associated polynomial degrees are not changed at the next time step, this factorization can simply be reused. Therefore there are ample opportunities for reuse of previous computations to reduce the total cost. Turning to the cases of refined or coarsened patches, the analysis in the subsequent sections shows that one particular advantage of the equilibrated flux $\boldsymbol{\sigma}_{h\tau}$ of Definition 4.1 is that it leads to estimators that are robust with respect to coarsening (and refinement) between time steps. The price to pay is that the size of the linear systems in (4.14) grows with the size of coarsening between two successive time steps, as (4.14) are defined on the patches $\omega_\mathbf{a}$ partitioned by the common refinement mesh $\widetilde{\mathcal{T}}^n$ for each $1 \le n \le N$. The analysis in [19, section 6], though, shows that this computational cost can be significantly reduced to the solution of two low-order systems over the patches $\omega_\mathbf{a}$, followed by local high-order corrections on the subpatches of $\widetilde{\mathcal{T}^{\mathbf{a},n}}$. We refer the reader to [19, section 6] for the full details of this approach.

5. Main results. In this section, we present the a posteriori error estimate featuring guaranteed upper bounds, local space-time efficiency, and polynomial-degree robustness. Let the norm $\|\cdot\|_{\mathcal{E}_Y} : Y + V_{h\tau} \to \mathbb{R}_{\geq 0}$ be defined by

$$\|v\|_{\mathcal{E}_Y}^2 := \|\mathcal{I}v\|_Y^2 + \|v - \mathcal{I}v\|_X^2 \quad \forall\, v \in Y + V_{h\tau}, \tag{5.1}$$

where we recall Remark 3.2 on the extension of the linear operator $\mathcal{I}$ to $Y + V_{h\tau}$. Since the exact solution $u \in Y$ implies that $\mathcal{I}u = u$, we have the identities

$$\begin{aligned} \|u - u_{h\tau}\|_{\mathcal{E}_Y}^2 &= \|u - \mathcal{I}u_{h\tau}\|_Y^2 + \|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X^2 \\ &= \|u - \mathcal{I}u_{h\tau}\|_Y^2 + \sum_{n=1}^N \tfrac{\tau_n (q_n+1)}{(2q_n+1)\,(2q_n+3)} \|\nabla (\!| u_{h\tau} |\!)_{n-1}\|^2, \end{aligned} \tag{5.2}$$

where we have simplified $\int_{I_n} \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|^2 \,\mathrm{d}t = \frac{\tau_n (q_n+1)}{(2q_n+1)\,(2q_n+3)} \|\nabla (\!| u_{h\tau} |\!)_{n-1}\|^2$, which is an identity easily deduced from (3.5) and from $\int_{I_n} |L_q^n|^2 \,\mathrm{d}t = \frac{\tau_n}{2q+1}$ for all $q \geq 0$; see also [38]. We also introduce the localized seminorms $|\cdot|_{\mathcal{E}_Y^{\mathbf{a},n}}$ for each $1 \leq n \leq N$ and each $\mathbf{a} \in \mathcal{V}^n$, defined by

$$|v|_{\mathcal{E}_Y^{\mathbf{a},n}}^2 := \int_{I_n} \|\partial_t \mathcal{I}v\|_{H^{-1}(\omega_{\mathbf{a}})}^2 + \|\nabla \mathcal{I}v\|_{\omega_{\mathbf{a}}}^2 + \|\nabla(v - \mathcal{I}v)\|_{\omega_{\mathbf{a}}}^2 \,\mathrm{d}t \quad \forall\, v \in Y + V_{h\tau}. \tag{5.3}$$

Similarly to (5.2), we find that

$$\begin{aligned} |u - u_{h\tau}|_{\mathcal{E}_Y^{\mathbf{a},n}}^2 = \int_{I_n} &\|\partial_t (u - \mathcal{I}u_{h\tau})\|_{H^{-1}(\omega_{\mathbf{a}})}^2 + \|\nabla(u - \mathcal{I}u_{h\tau})\|_{\omega_{\mathbf{a}}}^2 \,\mathrm{d}t \\ &+ \tfrac{\tau_n (q_n+1)}{(2q_n+1)\,(2q_n+3)} \|\nabla (\!| u_{h\tau} |\!)_{n-1}\|_{\omega_{\mathbf{a}}}^2. \end{aligned} \tag{5.4}$$

Although it might not be immediately obvious that $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ is equivalent to the Hilbertian sum of the $|u - u_{h\tau}|_{\mathcal{E}_Y^{\mathbf{a},n}}$, up to data oscillation, this will come as a consequence of the results shown here and in section 8.

Remark 5.1 (role of $\|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X$). The error estimators in this work focus on $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$, which is based on the inclusion of the additional term $\|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X$. First, this term allows the extension of the Y-norm to the sum of the continuous and discrete solution spaces $Y + V_{h\tau}$. Thus it measures the lack of conformity of $u_{h\tau} \notin Y$ coming from the jumps between time steps, as shown by (5.2). The second reason to consider this term is that it is naturally connected to $\|u - \mathcal{I}u_{h\tau}\|_Y$. Indeed, recall that $u_{h\tau} - \mathcal{I}u_{h\tau}$ appears in the discrete residual from the right-hand side of (1.2), and that the X' norm of this discrete component of the residual is simply $\|u_{h\tau} - \mathcal{I}u_{h\tau}\|_X$ by Theorem 2.1. Hence this term also has the role of bounding the lack of Galerkin orthogonality of $\mathcal{I}u_{h\tau}$.

We are now ready to state our main results in Theorems 5.1 and 5.2 below.

5.1. Global equivalence of norms. It is helpful to denote the time-localized dual norm of the residual by

$$\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|_{X'} := \sup_{v \in X,\ \|v\|_X = 1} \int_{I_n} (f, v) - \langle \partial_t \mathcal{I}u_{h\tau}, v\rangle - (\nabla \mathcal{I}u_{h\tau}, \nabla v) \,\mathrm{d}t. \tag{5.5}$$

Note that $\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|_{X'}$ can always be bounded from above by the restriction of the Y-norm of the error $u - \mathcal{I}u_{h\tau}$ to the time-step interval I_n.

THEOREM 5.1 (equivalence of norms). *Let the norm $\|\cdot\|_{\mathcal{E}_Y}$ be defined by* (5.1) *and, for each $1 \leq n \leq N$, let the temporal data oscillation $\eta^n_{\mathrm{osc},\tau}$ and the coarsening error η^n_{C} be defined by*

$$\eta^n_{\mathrm{C}} \coloneqq \sqrt{\tfrac{\tau_n(q_n+1)}{(2q_n+1)\,(2q_n+3)}}\|\nabla\left\{u_{h\tau}(t_{n-1}) - P^n_h[u_{h\tau}(t_{n-1})]\right\}\|, \tag{5.6a}$$

$$[\eta^n_{\mathrm{osc},\tau}]^2 \coloneqq \int_{I_n} \|f(t) - f_\tau(t)\|^2_{H^{-1}(\Omega)}\,\mathrm{d}t, \tag{5.6b}$$

where $P^n_h\colon H^1_0(\Omega) \to V^n_h$ denotes the elliptic orthogonal projection onto V^n_h defined by $(\nabla P^n_h w, \nabla v_h) = (\nabla w, \nabla v_h)$ for all $v_h \in V^n_h$. Then, we have

$$\int_{I_n} \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|^2 \mathrm{d}t \leq 8\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|^2_{X'} + \min\left\{[\eta^n_{\mathrm{C}}]^2, 8[\eta^n_{\mathrm{osc},\tau}]^2\right\}, \tag{5.7}$$

where $\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|_{X'}$ is defined in (5.5). *Furthermore, we have*

$$\|u - \mathcal{I}u_{h\tau}\|^2_Y \leq \|u - u_{h\tau}\|^2_{\mathcal{E}_Y} \leq 9\|u - \mathcal{I}u_{h\tau}\|^2_Y + \sum_{n=1}^{N} \min\left\{[\eta^n_{\mathrm{C}}]^2, 8[\eta^n_{\mathrm{osc},\tau}]^2\right\}. \tag{5.8}$$

We delay the proof of Theorem 5.1 until section 6 below. We emphasize that the coarsening term η^n_{C} arises only in the equivalence of norms, and that it does not need to be computed in practice since it does not appear in the a posteriori error estimators below; see, in particular, (5.10).

Remark 5.2 (equivalence). Theorem 5.1 shows that $\|u-u_{h\tau}\|_{\mathcal{E}_Y}$ and $\|u-\mathcal{I}u_{h\tau}\|_Y$ are globally equivalent up to the minimum of temporal data oscillation and coarsening errors. A similar result due to (5.7) actually holds also on each I_n. In particular, one of our key contributions here is to obtain polynomial-degree independent constants in (5.8). It is important to note that although $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ and $\|u - \mathcal{I}u_{h\tau}\|_Y$ are essentially globally equivalent, their local distributions may differ.

Remark 5.3 (relation to [45]). A similar result to (5.7) was previously obtained in the lowest-order case $q_n = 0$ by Verfürth [45]; see, in particular, the bounds of [45, section 7] for what is denoted there $\frac{\tau_n}{3}|u^n_h - u^{n-1}_h|^2_1$, which is equivalent to $\int_{I_n}\|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t$ with $q_n = 0$ in our notation. For higher polynomial degrees, we note that Gaspoz et al. have obtained independently an inequality of a similar kind as (5.7) in [21, Prop. 7], with the difference that (5.7) features a robust constant with respect to the temporal polynomial degree and is sharper with respect to the oscillation term.

5.2. Main a posteriori error estimate. We introduce the following a posteriori error estimators and data oscillation terms:

$$\eta^n_{\mathrm{F},K}(t) \coloneqq \|\boldsymbol{\sigma}_{h\tau}(t) + \nabla\mathcal{I}u_{h\tau}(t)\|_K, \tag{5.9a}$$

$$\eta^n_{\mathrm{J},K} \coloneqq \sqrt{\tfrac{\tau_n(q_n+1)}{(2q_n+1)(2q_n+3)}}\,\|\nabla(\!|u_{h\tau}|\!)_{n-1}\|_K, \tag{5.9b}$$

$$\eta^n_{\mathrm{osc},h,K}(t) \coloneqq \left[\sum_{\widetilde{K}\in\widetilde{\mathcal{T}}^n,\ \widetilde{K}\subseteq K} \frac{h^2_{\widetilde{K}}}{\pi^2}\|f_\tau(t) - f_{h\tau}(t)\|^2_{\widetilde{K}}\right]^{\frac{1}{2}}, \tag{5.9c}$$

$$\eta_{\mathrm{osc},\tau}(t) \coloneqq \|f(t) - f_\tau(t)\|_{H^{-1}(\Omega)}, \tag{5.9d}$$

$$\eta_{\mathrm{osc,init}} \coloneqq \|u_0 - \Pi_h u_0\|, \tag{5.9e}$$

where $t \in I_n$, $K \in \mathcal{T}^n$, the equilibrated flux $\boldsymbol{\sigma}_{h\tau}$ is defined in Definition 4.1, and where the data approximations f_τ and $f_{h\tau}$ are, respectively, defined in section 4.2. The two estimators $\eta^n_{\mathrm{F},K}$ and $\eta^n_{\mathrm{J},K}$ are our principal estimators, where $\eta^n_{\mathrm{F},K}$ measures, respectively, the lack of $\boldsymbol{H}(\mathrm{div})$-conformity of the gradient of the reconstructed solution $\mathcal{I}u_{h\tau}$, and where $\eta^n_{\mathrm{J},K}$ measures the lack of temporal conformity of the numerical solution $u_{h\tau}$. The term $\eta^n_{\mathrm{osc},h,K}$ represents the data oscillation due to the spatial discretization, whereas $\eta_{\mathrm{osc},\tau}$ represents the data oscillation due to the temporal discretization.

We define the global a posteriori error estimators as

$$\eta_Y^2 \coloneqq \sum_{n=1}^N \int_{I_n} \left[\left\{ \sum_{K\in\mathcal{T}^n} [\eta^n_{\mathrm{F},K} + \eta^n_{\mathrm{osc},h,K}]^2 \right\}^{\frac{1}{2}} + \eta_{\mathrm{osc},\tau} \right]^2 \mathrm{d}t + [\eta_{\mathrm{osc,init}}]^2, \tag{5.10a}$$

$$\eta_{\mathcal{E}_Y}^2 \coloneqq \eta_Y^2 + \sum_{n=1}^N \sum_{K\in\mathcal{T}^n} [\eta^n_{\mathrm{J},K}]^2. \tag{5.10b}$$

Notice that in the absence of data oscillation, namely, if $u_0 = \Pi_h u_0$ and $f = f_\tau = f_{h\tau}$ (see Remark 4.1), then η_Y simplifies to $\eta_Y^2 = \int_0^T \|\boldsymbol{\sigma}_{h\tau} + \nabla \mathcal{I}u_{h\tau}\|^2\,\mathrm{d}t$, and $\eta_{\mathcal{E}_Y}$ simplifies to $\eta^2_{\mathcal{E}_Y} = \int_0^T \|\boldsymbol{\sigma}_{h\tau} + \nabla \mathcal{I}u_{h\tau}\|^2 + \|\nabla(u_{h\tau} - \mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t$.

Recall that we write $a \lesssim b$ for two quantities a and b if $a \leq Cb$ with a constant C depending only on the shape regularity of $\mathcal{T}^n$ and $\widetilde{\mathcal{T}^n}$, but otherwise independent of the mesh size, time-step size, and polynomial degrees in space and time.

THEOREM 5.2 ($\mathcal{E}_Y$-norm a posteriori error estimate). *Let $u \in Y$ be the weak solution of* (1.1)*, let $u_{h\tau} \in V_{h\tau}$ denote the solution of the numerical scheme* (3.4)*, and let $\mathcal{I}u_{h\tau}$ denote its temporal reconstruction, where the operator $\mathcal{I}$ is defined in* (3.5)*. Let $\boldsymbol{\sigma}_{h\tau}$ denote the equilibrated flux of Definition* 4.1*. Let $\|\cdot\|_{\mathcal{E}_Y}$ be defined in* (5.1)*, and let the a posteriori error estimators be defined in* (5.9) *with $\eta_{\mathcal{E}_Y}$ defined in* (5.10)*. Then, we have the guaranteed upper bound*

$$\|u - u_{h\tau}\|_{\mathcal{E}_Y} \leq \eta_{\mathcal{E}_Y}. \tag{5.11}$$

Moreover, for each $1 \leq n \leq N$ and for each $K \in \mathcal{T}^n$, the indicators satisfy the following local efficiency bound

$$\int_{I_n} [\eta^n_{\mathrm{F},K}]^2\,\mathrm{d}t + [\eta^n_{\mathrm{J},K}]^2 \lesssim \sum_{\mathbf{a}\in\mathcal{V}_K} \left\{ |u - u_{h\tau}|^2_{\mathcal{E}_Y^{\mathbf{a},n}} + [\eta^{\mathbf{a},n}_{\mathrm{osc}}]^2 \right\}, \tag{5.12}$$

where $|\cdot|_{\mathcal{E}_Y^{\mathbf{a},n}}$ is defined in (5.3)*, $\mathcal{V}_K$ is the set of vertices of the element K, and the local data oscillation term $\eta^{\mathbf{a},n}_{\mathrm{osc}}$ is defined by*

$$[\eta^{\mathbf{a},n}_{\mathrm{osc}}]^2 \coloneqq \int_{I_n} \|f - \Pi^{\mathbf{a},n}_{h\tau} f\|^2_{H^{-1}(\omega_{\mathbf{a}})}\,\mathrm{d}t. \tag{5.13}$$

Furthermore, we have the following global efficiency bound for $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$:

$$\sum_{n=1}^N \sum_{K\in\mathcal{T}^n} \left[\int_{I_n} [\eta^n_{\mathrm{F},K}]^2\,\mathrm{d}t + [\eta^n_{\mathrm{J},K}]^2 \right] \lesssim \|u - u_{h\tau}\|^2_{\mathcal{E}_Y} + \sum_{n=1}^N \sum_{\mathbf{a}\in\mathcal{V}^n} [\eta^{\mathbf{a},n}_{\mathrm{osc}}]^2. \tag{5.14}$$

The proof of Theorem 5.2 is postponed to the following sections: the proof of the upper bound (5.11) is given in section 7, and the proof of the bounds (5.12) and (5.14) is the subject of section 8. Theorem 5.2 shows the local space-time efficiency of the estimators with respect to $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$.

Remark 5.4 (temporal data oscillation). The temporal data oscillation term $\eta_{\mathrm{osc},\tau}$ is defined with respect to a negative norm, as is usual in the literature [15, 45]. Similarly to [15, 45], this temporal data oscillation term can be of the same order as the error in terms of the time-step size. Since this term already appears in the upper bounds of the residual-based estimates of [45, (1.5)], it is seen that this issue is not related to the choice of equilibrated flux a posteriori error estimators, but is rather a part of the error estimation in the Y-norm. In practical computations, it is often advisable to determine a minimal temporal resolution for reducing this term to within a prescribed tolerance, in advance of solving the numerical scheme (3.4). Although the negative norm appearing in the definition of $\eta_{\mathrm{osc},\tau}$ is noncomputable, there are several possibilities for estimating it. First, we mention that $\eta_{\mathrm{osc},\tau}$ is bounded from above by $C_\Omega\|f - f_\tau\|$ with C_Ω the constant of the global Poincaré inequality, although this can be pessimistic in practice. If f is a finite tensorial product of spatial and temporal functions, then sharper bounds can be obtained by solving a set of independent coarse and low-order conforming approximations for elliptic problems, followed by equilibrated flux a posteriori error estimates to achieve guaranteed upper bounds. Finally, we also mention that this issue motivates a posteriori error estimators in other norms; in particular, we show in [18] that X-norm a posteriori estimates benefit from data oscillation terms that are of higher order by an additional factor of $\sqrt{\tau} + h$.

5.3. Extension to $\boldsymbol{Y}$-norm estimates. As a consequence of the proof of Theorem 5.2, we can also show guaranteed upper bounds and local-in-time and global-in-space efficiency of the estimators with respect to $\|u - \mathcal{I}u_{h\tau}\|_Y$, thereby generalizing the results to [45] to higher-order approximations.

COROLLARY 5.3 (Y-norm a posteriori error estimate). *Let the estimator η_Y be defined by* (5.10a). *Then, we have*

$$\|u - \mathcal{I}u_{h\tau}\|_Y \leq \eta_Y. \tag{5.15}$$

Furthermore, for each $1 \leq n \leq N$, we have

$$\sum_{K\in\mathcal{T}^n}\left[\int_{I_n}[\eta^n_{\mathrm{F},K}]^2\,\mathrm{d}t + [\eta^n_{\mathrm{J},K}]^2\right] \lesssim \int_{I_n}\|\partial_t(u - \mathcal{I}u_{h\tau})\|^2_{H^{-1}(\Omega)} + \|\nabla(u - \mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t \\ + \min\left\{[\eta^n_{\mathrm{C}}]^2, 8[\eta^n_{\mathrm{osc},\tau}]^2\right\} + \sum_{\mathbf{a}\in\mathcal{V}^n}[\eta^{\mathbf{a},n}_{\mathrm{osc}}]^2. \tag{5.16}$$

6. Proof of equivalence between $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ and $\|u - \mathcal{I}u_{h\tau}\|_Y$. In this section, we prove Theorem 5.1, along with some corollary results, which relate $\|u - u_{h\tau}\|_{\mathcal{E}_Y}$ to $\|u - \mathcal{I}u_{h\tau}\|_Y$. Our starting point involves the following two original bounds on the norms of the jumps, which generalize one of the key results of Verfürth [45] for the lowest-order case $q_n = 0$. In fact, our result sharpens and simplifies the proof of the result of [45] even in the lowest-order case.

LEMMA 6.1. *For each $1 \leq n \leq N$, let $P^n_h\colon H^1_0(\Omega) \to V^n_h$ denote the elliptic orthogonal projection to V^n_h defined by $(\nabla P^n_h w, \nabla v_h) = (\nabla w, \nabla v_h)$ for all $v_h \in V^n_h$. Then, for each $1 \leq n \leq N$, the jump $(\!| u_{h\tau} |\!)_{n-1}$ satisfies*

$$\frac{\tau_n}{8q_n + 4}\|\nabla (\!| u_{h\tau} |\!)_{n-1}\|^2 \leq \|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|^2_{X'} \\ + \frac{\tau_n}{8q_n + 4}\|\nabla\left\{u_{h\tau}(t_{n-1}) - P^n_h[u_{h\tau}(t_{n-1})]\right\}\|^2, \tag{6.1}$$

where $\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|_{X'}$ *is defined in* (5.5)*. Furthermore, we also have the alternative bound*

$$\frac{\tau_n}{8q_n+12}\|\nabla(\!|u_{h\tau}|\!)_{n-1}\|^2 \le 2\left(\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|_{X'}^2 + [\eta^n_{\mathrm{osc},\tau}]^2\right). \tag{6.2}$$

Proof. First, note that $(u_{h\tau}-\mathcal{I}u_{h\tau})|_{I_n} = \frac{(-1)^{q_n}}{2}(L^n_{q_n+1}-L^n_{q_n})(\!|u_{h\tau}|\!)_{n-1}$ belongs to the space $\mathcal{Q}_{q_n+1}(I_n;\widetilde{V_h^n})$. We define the test function $v_{h\tau} := -\frac{(-1)^{q_n}}{2}L^n_{q_n}P^n_h(\!|u_{h\tau}|\!)_{n-1}$, which belongs to $\mathcal{Q}_{q_n}(I_n;V_h^n)$, and we use it in (3.8) for the numerical scheme, which yields, by orthogonality of the Legendre polynomials and by the definition of the orthogonal projector P_h^n, the identity

$$\begin{aligned}\int_{I_n}\|\nabla v_{h\tau}\|^2\,\mathrm{d}t &= \frac{\tau_n}{8q_n+4}\|\nabla P_h^n(\!|u_{h\tau}|\!)_{n-1}\|^2 = \int_{I_n}(\nabla(u_{h\tau}-\mathcal{I}u_{h\tau}),\nabla v_{h\tau})\,\mathrm{d}t\\ &= \int_{I_n}(f-\partial_t\mathcal{I}u_{h\tau},v_{h\tau})-(\nabla\mathcal{I}u_{h\tau},\nabla v_{h\tau})\,\mathrm{d}t.\end{aligned} \tag{6.3}$$

Therefore, we have $\int_{I_n}\|\nabla v_{h\tau}\|^2\,\mathrm{d}t \le \|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|^2_{X'}$. This bound yields the desired result (6.1) once it is combined with (6.3) and the orthogonality relation

$$\begin{aligned}\|\nabla P_h^n(\!|u_{h\tau}|\!)_{n-1}\|^2 &= \|\nabla(\!|u_{h\tau}|\!)_{n-1}\|^2 - \|\nabla\left\{(\!|u_{h\tau}|\!)_{n-1} - P_h^n(\!|u_{h\tau}|\!)_{n-1}\right\}\|^2\\ &= \|\nabla(\!|u_{h\tau}|\!)_{n-1}\|^2 - \|\nabla\left\{u_{h\tau}(t_{n-1}) - P_h^n[u_{h\tau}(t_{n-1})]\right\}\|^2,\end{aligned}$$

where the last equality above follows from the facts that $(\!|u_{h\tau}|\!)_{n-1} = u_{h\tau}(t_{n-1}) - u_{h\tau}(t_{n-1}^+)$ and that $u_{h\tau}(t^+_{n-1})\in V_h^n$. This completes the proof of the first bound (6.1).

We now turn to the proof of (6.2); the main difference in the proofs of (6.1) and (6.2) is that above we appealed to the numerical scheme using a discrete test function, whereas to establish (6.2), we shall now consider a higher-order polynomial function that is not in the discrete test space. We define v on I_n by $v|_{I_n} := \frac{(-1)^{q_n}}{2}L^n_{q_n+1}(\!|u_{h\tau}|\!)_{n-1}$, and then we extend v by zero outside of I_n, so that $v\in X$. Then, by orthogonality of the Legendre polynomial $L^n_{q_n+1}$ to all polynomials of degree at most q_n on I_n, we have the identities $\int_{I_n}(f_\tau,v)\,\mathrm{d}t = 0$, $\int_{I_n}(\partial_t\mathcal{I}u_{h\tau},v)\,\mathrm{d}t=0$, and $\int_{I_n}(\nabla u_{h\tau},\nabla v)\,\mathrm{d}t = 0$. Therefore, we obtain

$$\begin{aligned}\int_{I_n}\|\nabla v\|^2\,\mathrm{d}t &= \frac{\tau_n}{8q_n+12}\|\nabla(\!|u_{h\tau}|\!)_{n-1}\|^2 = \int_{I_n}(\nabla(u_{h\tau}-\mathcal{I}u_{h\tau}),\nabla v)\,\mathrm{d}t\\ &= \int_{I_n}(f,v)-(\partial_t\mathcal{I}u_{h\tau},v)-(\nabla\mathcal{I}u_{h\tau},\nabla v)+(f_\tau-f,v)\,\mathrm{d}t.\end{aligned}$$

The desired result (6.2) then follows straightforwardly from the above identity. ∎

Proof of Theorem 5.1. The first inequality $\|u-\mathcal{I}u_{h\tau}\|_Y^2 \le \|u-u_{h\tau}\|^2_{\mathcal{E}_Y}$ is obvious from the definition of $\|u-u_{h\tau}\|_{\mathcal{E}_Y}$ in (5.2). Recalling the definitions of $\eta^n_{\mathrm{J},K}$ in (5.9b) and η^n_{C} in (5.6a), we deduce from (6.1) and (6.2) that

$$\sum_{K\in\mathcal{T}^n}[\eta^n_{\mathrm{J},K}]^2 \le \frac{4(q_n+1)}{(2q_n+3)}\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|^2_{X'} + [\eta^n_{\mathrm{C}}]^2 \le 2\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|^2_{X'} + [\eta^n_{\mathrm{C}}]^2, \tag{6.4}$$

and that

$$\begin{aligned}\sum_{K\in\mathcal{T}^n}[\eta^n_{\mathrm{J},K}]^2 &\le \frac{8(q_n+1)}{(2q_n+1)}\left(\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|^2_{X'} + [\eta^n_{\mathrm{osc},\tau}]^2\right)\\ &\le 8\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|^2_{X'} + 8[\eta^n_{\mathrm{osc},\tau}]^2.\end{aligned} \tag{6.5}$$

Therefore, we obtain (5.7) by taking the minimum of the right-hand sides of the above bounds. Finally, we get (5.8) by summing the above inequality over all time steps and noting that $\sum_{n=1}^{N}\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|_{X'}^2 = \|\mathcal{R}_Y(\mathcal{I}u_{h\tau})\|_{X'}^2 \le \|u-\mathcal{I}u_{h\tau}\|_Y^2$ which follows from (2.7). ∎

It is possible to obtain slightly sharper variants of Theorem 5.1 under more specific assumptions. For instance, the following corollary shows that $\|u-u_{h\tau}\|_{\mathcal{E}_Y}$ is equivalent to $\|u-\mathcal{I}u_{h\tau}\|_Y$ without any additional data oscillation, whenever the mesh coarsening error is relatively small compared to the jumps.

COROLLARY 6.2. *Using the notation of Lemma* 6.1*, assume that there exists a constant* $\theta \in [0,1)$ *such that* $\|\nabla[(\!|u_{h\tau}|\!)_{n-1} - P_h^n(\!|u_{h\tau}|\!)_{n-1}]\|^2 \le \theta\|\nabla(\!|u_{h\tau}|\!)_{n-1}\|^2$ *for each* $1 \le n \le N$*. Then, we have*

$$\|u-\mathcal{I}u_{h\tau}\|_Y^2 \le \|u-u_{h\tau}\|_{\mathcal{E}_Y}^2 \le \frac{3-\theta}{1-\theta}\|u-\mathcal{I}u_{h\tau}\|_Y^2. \tag{6.6}$$

Proof. The result is a consequence of Lemma 6.1 and of the identity $(\!|u_{h\tau}|\!)_{n-1} - P_h^n(\!|u_{h\tau}|\!)_{n-1} = u_{h\tau}(t_{n-1}) - P_h^n[u_{h\tau}(t_{n-1})]$, which leads to $\frac{\tau_n}{8q_n+4}\|\nabla(\!|u_{h\tau}|\!)_{n-1}\|^2 \le \frac{1}{1-\theta}\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|_{X'}^2$. Adapting the proof of Theorem 5.1 then yields (6.6). ∎

Note that the case $\theta = 0$ in Corollary 6.2 corresponds to the case of no coarsening.

7. Proof of the guaranteed upper bound. We prove here (5.11) and (5.15). First, it is clear from (5.2) that (5.15) immediately implies (5.11). Therefore, it remains to show (5.15). Keeping in mind the equivalence identity (2.7) between norms of the errors and residuals, we turn our attention to bounds for the residual norm $\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})\|_{X'} = \sup_{v\in X\setminus\{0\}}\mathcal{B}_Y(u-\mathcal{I}u_{h\tau},v)/\|v\|_X$. To this end, consider an arbitrary function $v \in X$ such that $\|v\|_X = 1$. Then, we obtain

$$\langle\mathcal{R}_Y(\mathcal{I}u_{h\tau}),v\rangle = \int_0^T (f - \partial_t\mathcal{I}u_{h\tau} - \nabla\cdot\boldsymbol{\sigma}_{h\tau}, v) - (\boldsymbol{\sigma}_{h\tau} + \nabla\mathcal{I}u_{h\tau}, \nabla v)\,\mathrm{d}t,$$

where we have inserted the flux $\boldsymbol{\sigma}_{h\tau}$ and used integration by parts over Ω. Next, we use (4.1), and we write $f - f_{h\tau} = f - f_\tau + f_\tau - f_{h\tau}$. For any $\widetilde{K} \in \widetilde{\mathcal{T}}^n$, $1 \le n \le N$, we deduce from (4.5) that the function $t \mapsto (f_\tau(t) - f_{h\tau}(t), 1)_{\widetilde{K}}$, which is a real-valued polynomial of degree at most q_n on I_n, vanishes identically on I_n. Therefore, letting $v_{\widetilde{K}}(t)$ denote the mean value of $v(t)$ over the element $\widetilde{K} \in \widetilde{\mathcal{T}}^n$, which is defined for a.e. $t \in I_n$, we deduce from the Poincaré inequality that $|(f_\tau(t) - f_{h\tau}(t), v(t))_{\widetilde{K}}| \le \frac{h_{\widetilde{K}}}{\pi}\|f_\tau(t) - f_{h\tau}(t)\|_{\widetilde{K}}\|\nabla v(t)\|_{\widetilde{K}}$. Therefore, $\langle\mathcal{R}_Y(\mathcal{I}u_{h\tau}),v\rangle$ can be bounded as follows:

$$\begin{aligned}
\langle\mathcal{R}_Y(\mathcal{I}u_{h\tau}),v\rangle &= \sum_{n=1}^{N}\int_{I_n}(f - f_{h\tau}, v) - (\boldsymbol{\sigma}_{h\tau} + \nabla\mathcal{I}u_{h\tau}, \nabla v)\,\mathrm{d}t\\
&\le \sum_{n=1}^{N}\int_{I_n}\sum_{K\in\mathcal{T}^n}\eta_{\mathrm{F},K}^n\|\nabla v\|_K + \left[\sum_{\widetilde{K}\in\widetilde{\mathcal{T}}^n}\frac{h_{\widetilde{K}}}{\pi}\|f_\tau(t) - f_{h\tau}(t)\|_{\widetilde{K}}\|\nabla v(t)\|_{\widetilde{K}}\right]\\
&\quad + \eta_{\mathrm{osc},\tau}\|\nabla v\|\mathrm{d}t\\
&\le \sum_{n=1}^{N}\int_{I_n}\left[\sum_{K\in\mathcal{T}^n}[\eta_{\mathrm{F},K}^n + \eta_{\mathrm{osc},h,K}^n]\|\nabla v\|_K\right] + \eta_{\mathrm{osc},\tau}\|\nabla v\|\,\mathrm{d}t\\
&\le \sum_{n=1}^{N}\int_{I_n}\left[\left\{\sum_{K\in\mathcal{T}^n}[\eta_{\mathrm{F},K}^n + \eta_{\mathrm{osc},h,K}^n]^2\right\}^{\frac{1}{2}} + \eta_{\mathrm{osc},\tau}\right]\|\nabla v\|\,\mathrm{d}t.
\end{aligned}$$

Applying the Cauchy–Schwarz inequality leads to an upper bound for $\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})\|_{X'}$, which we then combine with the identity (2.7) relating errors and residuals to obtain $\|u - \mathcal{I}u_{h\tau}\|_Y \leq \eta_Y$. The corresponding upper bound $\|u - u_{h\tau}\|_{\mathcal{E}_Y} \leq \eta_{\mathcal{E}_Y}$ then follows immediately, as explained above. ∎

8. Proof of local space-time efficiency and robustness.

We prove here the bounds (5.12), (5.14), and (5.16).

8.1. Preliminary result.

The following lemma is a generalization of important results on polynomial-degree robustness of equilibrated flux estimates from [3, Thm. 7], in two space dimensions, and [17, Thm. 2.3] in three space dimensions. In particular, Lemma 8.1 comes from [19, Thm. 1.2] on the existence of a discrete polynomial-degree robust $\boldsymbol{H}(\mathrm{div})$-lifting of data that are piecewise polynomials with respect to the submesh $\widetilde{\mathcal{T}^{\mathbf{a},n}}$. Note that [3, Thm. 7] and [17, Thm. 2.3] only hold for the case where the data are piecewise polynomials on the elements $K \in \mathcal{T}^n$ of the patch $\omega_{\mathbf{a}}$. This generalization is crucial for allowing arbitrary refinement and coarsening between time steps.

LEMMA 8.1 (polynomial-degree-robust stability bound). *For each $1 \leq n \leq N$ and each $\mathbf{a} \in \mathcal{V}^n$, recall that $\widetilde{\mathcal{T}^{\mathbf{a},n}}$ denotes the restriction of $\widetilde{\mathcal{T}^n}$ to $\omega_{\mathbf{a}}$ and that $\psi_{\mathbf{a}} \in H^1(\omega_{\mathbf{a}}) \cap \mathcal{P}_1(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ denotes the hat function associated with $\omega_{\mathbf{a}}$. Let $\Gamma_{\mathbf{a}} = \{x \in \partial\omega_{\mathbf{a}},\ \psi_{\mathbf{a}}(x) = 0\}$. Then, for any $f_h^{\mathbf{a},n} \in \mathcal{P}_{p_{\mathbf{a}}-1}(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ and any $\boldsymbol{\xi}_h^{\mathbf{a},n} \in \mathbf{RTN}_{p_{\mathbf{a}}-1}(\widetilde{\mathcal{T}^{\mathbf{a},n}})$, where it is further supposed that $(f_h^{\mathbf{a}}, \psi_{\mathbf{a}})_{\omega_{\mathbf{a}}} = (\boldsymbol{\xi}_h^{\mathbf{a},n}, \nabla\psi_{\mathbf{a}})_{\omega_{\mathbf{a}}}$ if $\Gamma_{\mathbf{a}} = \partial\omega_{\mathbf{a}}$, we have*

$$\min_{\substack{\boldsymbol{v}_h \in \boldsymbol{H}(\mathrm{div},\omega_{\mathbf{a}}) \cap \mathbf{RTN}_{p_{\mathbf{a}}}(\widetilde{\mathcal{T}^{\mathbf{a},n}}) \\ \nabla\cdot\boldsymbol{v}_h = \psi_{\mathbf{a}} f_h^{\mathbf{a},n} - \nabla\psi_{\mathbf{a}}\cdot\boldsymbol{\xi}_h^{\mathbf{a},n} \\ \boldsymbol{v}_h\cdot\boldsymbol{n} = 0 \text{ on } \Gamma_{\mathbf{a}}}} \|\boldsymbol{v}_h + \psi_{\mathbf{a}}\boldsymbol{\xi}_h^{\mathbf{a},n}\|_{\omega_{\mathbf{a}}} \lesssim \sup_{\varphi \in H_0^1(\omega_{\mathbf{a}})\setminus\{0\}} \frac{(f_h^{\mathbf{a},n}, \varphi)_{\omega_{\mathbf{a}}} - (\boldsymbol{\xi}_h^{\mathbf{a},n}, \nabla\varphi)_{\omega_{\mathbf{a}}}}{\|\nabla\varphi\|_{\omega_{\mathbf{a}}}}.$$

Proof. The result is directly obtained by applying [19, Thm. 1.2], where Ω there stands for $\omega_{\mathbf{a}}$ here, where $\mathcal{T}$ there stands for $\widetilde{\mathcal{T}^{\mathbf{a},n}}$ here, and where $\psi_\dagger$ there stands for $\psi_{\mathbf{a}}$ here. In applying [19, Thm. 1.2], we use the fact that $\Gamma_{\mathbf{a}}$ is the union of the faces of the mesh $\widetilde{\mathcal{T}^{\mathbf{a},n}}$ on which $\psi_{\mathbf{a}}$ vanishes, and we have simplified the constant appearing there by using the fact that $h_{\omega_{\mathbf{a}}}\|\nabla\psi_{\mathbf{a}}\|_\infty \lesssim \|\psi_{\mathbf{a}}\|_\infty = 1$ by shape regularity. ∎

8.2. Stability of the space-time flux equilibration.

For each $1 \leq n \leq N$ and each $\mathbf{a} \in \mathcal{V}^n$, we introduce the patch residual functional $\mathcal{R}_{h\tau}^{\mathbf{a},n} \colon L^2(I_n, H_0^1(\omega_{\mathbf{a}})) \to \mathbb{R}$ that is defined by

$$\langle \mathcal{R}_{h\tau}^{\mathbf{a},n}, v\rangle = \int_{I_n} \left(\Pi_{h\tau}^{\mathbf{a},n} f - \partial_t(\mathcal{I}u_{h\tau}), v\right)_{\omega_{\mathbf{a}}} - \left(\nabla u_{h\tau}, \nabla v\right)_{\omega_{\mathbf{a}}} \mathrm{d}t \tag{8.1}$$

for all $v \in L^2(I_n, H_0^1(\omega_{\mathbf{a}}))$. We are now ready to state the essential result that forms the starting point for our analysis of the efficiency of the error estimators.

LEMMA 8.2 (space-time stability bound). *Let $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n}$ denote the patchwise flux reconstructions of Definition 4.1, and let $\mathcal{R}_{h\tau}^{\mathbf{a},n}$ denote the local patch residual defined by (8.1). Then, we have*

$$\left(\int_{I_n} \|\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} + \psi_{\mathbf{a}}\nabla u_{h\tau}\|_{\omega_{\mathbf{a}}}^2 \,\mathrm{d}t\right)^{\frac{1}{2}} \lesssim \sup_{v \in \mathcal{Q}_{q_n}(I_n; H_0^1(\omega_{\mathbf{a}}))\setminus\{0\}} \frac{\langle \mathcal{R}_{h\tau}^{\mathbf{a},n}, v\rangle}{\left(\int_{I_n} \|\nabla v\|_{\omega_{\mathbf{a}}}^2 \,\mathrm{d}t\right)^{\frac{1}{2}}}, \tag{8.2}$$

where $\mathcal{Q}_{q_n}(I_n; H_0^1(\omega_{\mathbf{a}}))$ denotes the space of $H_0^1(\omega_{\mathbf{a}})$-valued univariate polynomials of degree at most q_n on I_n.

Proof. The definition of $\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n}$ in (4.9) implies that it is enough to show that there exists a $\boldsymbol{v}_{h\tau} \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n}$ such that $\nabla{\cdot}\boldsymbol{v}_{h\tau} = g_{h\tau}^{\mathbf{a},n}$ and such that $\int_{I_n} \|\boldsymbol{v}_{h\tau} - \boldsymbol{\tau}_{h\tau}^{\mathbf{a},n}\|_{\omega_{\mathbf{a}}}^2 \,\mathrm{d}t$ is bounded by the right-hand side of (8.2). Let $\{\phi_j^n\}_{j=0}^{q_n}$ be an L^2-orthonormal basis of polynomials on I_n, and let the functions $\{f_{h,j}^{\mathbf{a},n}\}_{j=0}^{q_n}$ and $\{\boldsymbol{\xi}_{h,j}^{\mathbf{a},n}\}_{j=0}^{q_n}$ be defined by

$$
f_{h,j}^{\mathbf{a},n} \coloneqq \int_{I_n} (\Pi_{h\tau}^{\mathbf{a},n} f - \partial_t \mathcal{I} u_{h\tau}) \phi_j^n \,\mathrm{d}t, \quad \boldsymbol{\xi}_{h,j}^{\mathbf{a},n} \coloneqq \int_{I_n} \nabla u_h \, \phi_j^n \,\mathrm{d}t. \tag{8.3}
$$

It will be useful to keep in mind that $g_{h\tau}^{\mathbf{a},n} = \sum_{j=0}^{q_n} [\psi_{\mathbf{a}} f_{h,j}^{\mathbf{a},n} - \nabla\psi_{\mathbf{a}} \cdot \boldsymbol{\xi}_{h,j}^{\mathbf{a},n}] \phi_j^n$ and that $\boldsymbol{\tau}_{h\tau}^{\mathbf{a},n} = -\sum_{j=0}^{q_n} \psi_{\mathbf{a}} \boldsymbol{\xi}_{h,j}^{\mathbf{a},n} \phi_j^n$. Let $\Gamma_{\mathbf{a}} \coloneqq \{x \in \partial\omega_{\mathbf{a}},\ \psi_{\mathbf{a}}(x) = 0\}$; note that if $\mathbf{a} \in \mathcal{V}_{\mathrm{int}}^n$, then $\Gamma_{\mathbf{a}} = \partial\omega_{\mathbf{a}}$, whereas $\Gamma_{\mathbf{a}}$ is a strict subset of $\partial\omega_{\mathbf{a}}$ if $\mathbf{a} \in \mathcal{V}_{\mathrm{ext}}^n$. We will now use Lemma 8.1 to show that, for each $0 \le j \le q_n$, there exists $\boldsymbol{v}_{h,j}^{\mathbf{a},n} \in \boldsymbol{H}(\mathrm{div}, \omega_{\mathbf{a}}) \cap \mathbf{RTN}_{p_{\mathbf{a}}}(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ such that

$$
\nabla{\cdot}\boldsymbol{v}_{h,j}^{\mathbf{a},n} = \psi_{\mathbf{a}} f_{h,j}^{\mathbf{a},n} - \nabla\psi_{\mathbf{a}} \cdot \boldsymbol{\xi}_{h,j}^{\mathbf{a},n} \quad \text{in } \omega_{\mathbf{a}}, \qquad \boldsymbol{v}_{h,j}^{\mathbf{a},n} \cdot \boldsymbol{n} = 0 \quad \text{on } \Gamma_{\mathbf{a}}, \tag{8.4a}
$$

$$
\|\boldsymbol{v}_{h,j}^{\mathbf{a},n} + \psi_{\mathbf{a}} \boldsymbol{\xi}_{h,j}^{\mathbf{a},n}\|_{\omega_{\mathbf{a}}} \lesssim \|\mathcal{R}_{h,j}^{\mathbf{a},n}\|_{H^{-1}(\omega_{\mathbf{a}})}, \tag{8.4b}
$$

where $\mathcal{R}_{h,j}^{\mathbf{a},n} \in H^{-1}(\omega_{\mathbf{a}})$ is defined by $\langle \mathcal{R}_{h,j}^{\mathbf{a},n}, v \rangle \coloneqq (f_{h,j}^{\mathbf{a},n}, v)_{\omega_{\mathbf{a}}} - (\boldsymbol{\xi}_{h,j}^{\mathbf{a},n}, \nabla v)_{\omega_{\mathbf{a}}}$ for all $v \in H_0^1(\omega_{\mathbf{a}})$. To check the hypotheses of Lemma 8.1, we start by observing that the choice of $p_{\mathbf{a}}$ in (4.2) implies that $f_{h,j}^{\mathbf{a},n} \in \mathcal{P}_{p_{\mathbf{a}}-1}(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ and that $\boldsymbol{\xi}_{h,j}^{\mathbf{a},n} \in \mathbf{RTN}_{p_{\mathbf{a}}-1}(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ for all $0 \le j \le q_n$. For any interior vertex $\mathbf{a} \in \mathcal{V}_{\mathrm{int}}^n$, it is seen from (4.8) that $(f_{h,j}^{\mathbf{a},n}, \psi_{\mathbf{a}})_{\omega_{\mathbf{a}}} = (\boldsymbol{\xi}_{h,j}^{\mathbf{a},n}, \nabla\psi_{\mathbf{a}})_{\omega_{\mathbf{a}}}$ for all $0 \le j \le q_n$. Therefore, the hypotheses of Lemma 8.1 are satisfied, and there exists $\boldsymbol{v}_{h,j}^{\mathbf{a},n} \in \boldsymbol{H}(\mathrm{div}, \omega_{\mathbf{a}}) \cap \mathbf{RTN}_{p_{\mathbf{a}}}(\widetilde{\mathcal{T}^{\mathbf{a},n}})$ satisfying (8.4).

Next, we claim that $\boldsymbol{v}_{h,j}^{\mathbf{a},n} \in \boldsymbol{V}_h^{\mathbf{a},n}$ for all $0 \le j \le q_n$. Indeed, the definition of $\Gamma_{\mathbf{a}}$ implies that $\Gamma_{\mathbf{a}} = \partial\omega_{\mathbf{a}}$ for all $\mathbf{a} \in \mathcal{V}_{\mathrm{int}}^n$ and that $\partial\omega_{\mathbf{a}} \setminus \partial\Omega \subset \Gamma_{\mathbf{a}}$ for all $\mathbf{a} \in \mathcal{V}_{\mathrm{ext}}^n$. Therefore, we have $\boldsymbol{v}_{h,j}^{\mathbf{a},n} \in \boldsymbol{V}_h^{\mathbf{a},n}$ for all $0 \le j \le q_n$. It then follows that the function $\boldsymbol{v}_{h\tau}^{\mathbf{a},n} \coloneqq \sum_{j=0}^{q_n} \boldsymbol{v}_{h,j}^{\mathbf{a},n} \, \phi_j^n \in \boldsymbol{V}_{h\tau}^{\mathbf{a},n}$ and that this function satisfies

$$
\nabla{\cdot}\boldsymbol{v}_{h\tau}^{\mathbf{a},n} = \sum_{j=0}^{q_n} [\psi_{\mathbf{a}} f_{h,j}^{\mathbf{a},n} - \nabla\psi_{\mathbf{a}} \cdot \boldsymbol{\xi}_{h,j}^{\mathbf{a},n}] \phi_j^n = g_{h\tau}^{\mathbf{a},n}, \tag{8.5a}
$$

$$
\int_{I_n} \|\boldsymbol{v}_{h\tau}^{\mathbf{a},n} - \boldsymbol{\tau}_{h\tau}^{\mathbf{a},n}\|_{\omega_{\mathbf{a}}}^2 \,\mathrm{d}t = \sum_{j=0}^{q_n} \|\boldsymbol{v}_{h,j}^{\mathbf{a},n} + \psi_{\mathbf{a}} \boldsymbol{\xi}_{h,j}^{\mathbf{a},n}\|_{\omega_{\mathbf{a}}}^2 \lesssim \sum_{j=0}^{q_n} \|\mathcal{R}_{h,j}^{\mathbf{a},n}\|_{H^{-1}(\omega_{\mathbf{a}})}^2, \tag{8.5b}
$$

where the equality in (8.5b) results from the orthonormality of $\{\phi_j^n\}_{j=0}^{q_n}$. We now claim that

$$
\left\{ \sum_{j=0}^{q_n} \|\mathcal{R}_{h,j}^{\mathbf{a},n}\|_{H^{-1}(\omega_{\mathbf{a}})}^2 \right\}^{\frac{1}{2}} \le \sup_{v \in \mathcal{Q}_{q_n}(I_n; H_0^1(\omega_{\mathbf{a}})) \setminus \{0\}} \frac{\langle \mathcal{R}_{h\tau}^{\mathbf{a},n}, v \rangle}{\left(\int_{I_n} \|\nabla v\|_{\omega_{\mathbf{a}}}^2 \,\mathrm{d}t \right)^{\frac{1}{2}}}. \tag{8.6}
$$

For each $j = 0, \dots, q_n$, we define $z_j \in H_0^1(\omega_{\mathbf{a}})$ by $(\nabla z_j, \nabla v)_{\omega_{\mathbf{a}}} = \langle \mathcal{R}_{h,j}^{\mathbf{a},n}, v \rangle$ for all $v \in H_0^1(\omega_{\mathbf{a}})$. It is then straightforward to show that $\|\nabla z_j\|_{\omega_{\mathbf{a}}}^2 = \langle \mathcal{R}_{h,j}^{\mathbf{a},n}, z_j \rangle = \|\mathcal{R}_{h,j}^{\mathbf{a},n}\|_{H^{-1}(\omega_{\mathbf{a}})}^2$ for each $j = 0, \dots, q_n$. Then, we define $z_* \in \mathcal{Q}_{q_n}(I_n; H_0^1(\omega_{\mathbf{a}}))$ by $z_* \coloneqq \sum_{j=0}^{q_n} z_j \phi_j^n$. It follows from the orthonormality of the temporal basis that

$\int_{I_n} \|\nabla z_*\|^2_{\omega_{\mathbf{a}}}\,\mathrm{d}t = \sum_{j=0}^{q_n} \|\mathcal{R}^{\mathbf{a},n}_{h,j}\|^2_{H^{-1}(\omega_{\mathbf{a}})}$. Fubini's theorem and (8.3) imply that

$$\begin{aligned}
\langle \mathcal{R}^{\mathbf{a},n}_{h\tau}, z_*\rangle &= \sum_{j=0}^{q_n} \int_{I_n} (\Pi^{\mathbf{a},n}_{h\tau} f - \partial_t \mathcal{I} u_{h\tau}, \phi^n_j z_j)_{\omega_{\mathbf{a}}} - (\nabla u_h, \phi^n_j \nabla z_j)_{\omega_{\mathbf{a}}}\,\mathrm{d}t \\
&= \sum_{j=0}^{q_n} \left\{ \left(\textstyle\int_{I_n} (\Pi^{\mathbf{a},n}_{h\tau} f - \partial_t \mathcal{I} u_{h\tau} \phi^n_j)\,\mathrm{d}t, z_j \right)_{\omega_{\mathbf{a}}} - \left(\textstyle\int_{I_n} \nabla u_h \phi^n_j \,\mathrm{d}t, \nabla z_j \right)_{\omega_{\mathbf{a}}} \right\} \\
&= \sum_{j=0}^{q_n} \left\{ (f^{\mathbf{a},n}_{h,j}, z_j)_{\omega_{\mathbf{a}}} - (\boldsymbol{\xi}^{\mathbf{a},n}_{h,j}, \nabla z_j)_{\omega_{\mathbf{a}}} \right\} = \sum_{j=0}^{q_n} \langle \mathcal{R}^{\mathbf{a},n}_{h,j}, z_j\rangle = \sum_{j=0}^{q_n} \|\mathcal{R}^{\mathbf{a},n}_{h,j}\|^2_{H^{-1}(\omega_{\mathbf{a}})}.
\end{aligned}$$

Hence, the above identities immediately imply (8.6). Therefore, we combine (8.5) and (8.6) to deduce that $\boldsymbol{v}^{\mathbf{a},n}_{h\tau} \in \boldsymbol{V}^{\mathbf{a},n}_{h\tau}$ satisfies $\nabla\cdot \boldsymbol{v}^{\mathbf{a},n}_{h\tau} = g^{\mathbf{a},n}_{h\tau}$ and $\int_{I_n} \|\boldsymbol{v}_h - \boldsymbol{\tau}^{\mathbf{a},n}_{h\tau}\|^2_{\omega_{\mathbf{a}}}\,\mathrm{d}t$ is bounded by the right-hand side of (8.2). This implies (8.2) as explained above. ∎

8.3. Local efficiency. We can now prove the local efficiency bound (5.12).

Proof of the local efficiency bound (5.12). Consider a time step I_n and an element $K \in \mathcal{T}^n$. First, note that $[\eta^n_{\mathrm{J},K}]^2 \leq \sum_{\mathbf{a}\in\mathcal{V}_K} |u - u_{h\tau}|^2_{\mathcal{E}^{\mathbf{a},n}_Y}$ trivially, where we recall that $\mathcal{V}_K$ denotes the set of vertices of K. Hence, it remains only to bound $\int_{I_n} [\eta^n_{\mathrm{F},K}]^2\,\mathrm{d}t$. To this end, observe that $\boldsymbol{\sigma}_{h\tau}|_{K\times I_n} = \sum_{\mathbf{a}\in\mathcal{V}_K} \boldsymbol{\sigma}^{\mathbf{a},n}_{h\tau}|_{K\times I_n}$, and that

$$\begin{aligned}
\int_{I_n} [\eta^n_{\mathrm{F},K}]^2\,\mathrm{d}t &= \int_{I_n} \left\| \textstyle\sum_{\mathbf{a}\in\mathcal{V}_K} (\boldsymbol{\sigma}^{\mathbf{a},n}_{h\tau} + \psi_{\mathbf{a}} \nabla \mathcal{I} u_{h\tau}) \right\|^2_K \,\mathrm{d}t \\
&\leq (|\mathcal{V}_K| + 1) \int_{I_n} \sum_{\mathbf{a}\in\mathcal{V}_K} \|\boldsymbol{\sigma}^{\mathbf{a},n}_{h\tau} + \psi_{\mathbf{a}} \nabla u_{h\tau}\|^2_K + \|\nabla(u_{h\tau} - \mathcal{I} u_{h\tau})\|^2_K \,\mathrm{d}t \\
&\leq (|\mathcal{V}_K| + 1) \int_{I_n} \sum_{\mathbf{a}\in\mathcal{V}_K} \|\boldsymbol{\sigma}^{\mathbf{a},n}_{h\tau} + \psi_{\mathbf{a}} \nabla u_{h\tau}\|^2_{\omega_{\mathbf{a}}}\,\mathrm{d}t + [\eta^n_{\mathrm{J},K}]^2, \qquad (8.7)
\end{aligned}$$

where $|\mathcal{V}_K|$ is the number of vertices of the element K, which equals $d+1$ for simplices and where we have used that $\|\cdot\|_K \leq \|\cdot\|_{\omega_{\mathbf{a}}}$ and the definition of $\eta^n_{\mathrm{J},K}$ in the last line.

Keeping in mind Lemma 8.2, we therefore turn our attention to bounding the dual norm of the patchwise residuals $\mathcal{R}^{\mathbf{a},n}_{h\tau}$ for each $\mathbf{a} \in \mathcal{V}_K$. Consider an arbitrary $v \in \mathcal{Q}_{q_n}(I_n, H^1_0(\omega_{\mathbf{a}}))$ such that $\int_{I_n} \|\nabla v\|^2_{\omega_{\mathbf{a}}}\,\mathrm{d}t = 1$; then (2.3) implies that

$$\begin{aligned}
\langle \mathcal{R}^{\mathbf{a},n}_{h\tau}, v\rangle &= \int_{I_n} \big(\Pi^{\mathbf{a},n}_{h\tau} f, v\big)_{\omega_{\mathbf{a}}} - \big(\partial_t(\mathcal{I} u_{h\tau}), v\big)_{\omega_{\mathbf{a}}} - \big(\nabla u_{h\tau}, \nabla v\big)_{\omega_{\mathbf{a}}}\,\mathrm{d}t \\
&= \int_{I_n} \langle \partial_t(u - \mathcal{I} u_{h\tau}), v\rangle + (\nabla(u - \mathcal{I} u_{h\tau}), \nabla v)_{\omega_{\mathbf{a}}}\,\mathrm{d}t \\
&\quad + \int_{I_n} (\nabla(\mathcal{I} u_{h\tau} - u_{h\tau}), \nabla v)_{\omega_{\mathbf{a}}} - (f - \Pi^{\mathbf{a},n}_{h\tau} f, v)_{\omega_{\mathbf{a}}}\,\mathrm{d}t \\
&=: E_1 + E_2 + E_3 + E_4.
\end{aligned}$$

The definition of the $\|\cdot\|_{H^{-1}(\omega_{\mathbf{a}})}$-norm and the Cauchy–Schwarz inequality then yield $|E_1 + E_2 + E_3| \lesssim |u - u_{h\tau}|_{\mathcal{E}^{\mathbf{a},n}_Y}$. Finally, we find that $|E_4| \leq \eta^{\mathbf{a},n}_{\mathrm{osc}}$, where $\eta^{\mathbf{a},n}_{\mathrm{osc}}$ is defined in (5.13). Therefore, we find that

$$\sup_{v\in\mathcal{Q}_{q_n}(I_n,H_0^1(\omega_{\mathbf{a}}))\setminus\{0\}} \frac{\langle \mathcal{R}_{h\tau}^{\mathbf{a},n}, v\rangle}{\left(\int_{I_n}\|\nabla v\|_{\omega_{\mathbf{a}}}^2\,\mathrm{d}t\right)^{\frac{1}{2}}} \lesssim \left(|u-u_{h\tau}|^2_{\mathcal{E}_Y^{\mathbf{a},n}} + [\eta_{\mathrm{osc}}^{\mathbf{a},n}]^2\right)^{\frac{1}{2}}. \tag{8.8}$$

Recalling (8.2) of Lemma 8.2, we deduce that, for each $\mathbf{a}\in\mathcal{V}_K$,

$$\int_{I_n} \|\boldsymbol{\sigma}_{h\tau}^{\mathbf{a},n} + \psi_{\mathbf{a}}\nabla u_{h\tau}\|_{\omega_{\mathbf{a}}}^2\,\mathrm{d}t \lesssim |u-u_{h\tau}|^2_{\mathcal{E}_Y^{\mathbf{a},n}} + [\eta_{\mathrm{osc}}^{\mathbf{a},n}]^2, \tag{8.9}$$

which in combination with (8.7), yields the desired result (5.12). ◻

8.4. Global efficiency. We finally prove the global efficiency bounds (5.14) and (5.16).

Proof of (5.14) *and* (5.16). Recalling the definition (5.3) of the localized seminorms $|\cdot|_{\mathcal{E}_Y^{\mathbf{a},n}}$, we claim that

$$\sum_{\mathbf{a}\in\mathcal{V}^n} |u-u_{h\tau}|^2_{\mathcal{E}_Y^{\mathbf{a},n}} \lesssim \int_{I_n} \|\partial_t(u-\mathcal{I}u_{h\tau})\|^2_{H^{-1}(\Omega)} + \|\nabla(u-\mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t + \int_{I_n}\|\nabla(u_{h\tau}-\mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t. \tag{8.10}$$

The proof is essentially a counting argument after local Riesz mappings are introduced to evaluate the negative norms $\|\partial_t(u-\mathcal{I}u_{h\tau})\|^2_{H^{-1}(\omega_{\mathbf{a}})}$ for all $\mathbf{a}\in\mathcal{V}^n$; see also [2]. Summing (5.12) over $K\in\mathcal{T}^n$ and using (8.10) then leads to

$$\sum_{K\in\mathcal{T}^n}\left[\int_{I_n}[\eta_{\mathrm{F},K}^n]^2\,\mathrm{d}t + [\eta_{\mathrm{J},K}^n]^2\right] \lesssim \int_{I_n} \|\partial_t(u-\mathcal{I}u_{h\tau})\|^2_{H^{-1}(\Omega)} + \|\nabla(u-\mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t + \int_{I_n}\|\nabla(u_{h\tau}-\mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t + \sum_{\mathbf{a}\in\mathcal{V}^n}[\eta_{\mathrm{osc}}^{\mathbf{a},n}]^2. \tag{8.11}$$

Summing the bound (8.11) for all $1\le n\le N$ immediately yields (5.14), whereas (5.16) results from (8.11) after invoking (5.7) and observing that $\|\mathcal{R}_Y(\mathcal{I}u_{h\tau})|_{I_n}\|^2_{X'}$ is bounded from above by $\int_{I_n}\|\partial_t(u-\mathcal{I}u_{h\tau})\|^2_{H^{-1}(\Omega)} + \|\nabla(u-\mathcal{I}u_{h\tau})\|^2\,\mathrm{d}t$. ◻

9. Conclusion and outlook. We have studied a posteriori error estimates for hp-τq discretizations of parabolic problems based on arbitrarily high-order conforming Galerkin spatial discretizations and discontinuous Galerkin temporal discretizations. The equilibrated flux reconstructions lead to guaranteed upper bounds for the norm $\|u-u_{h\tau}\|_{\mathcal{E}_Y}$. Furthermore, the estimators have the key property of being unconditionally locally space-time efficient with respect to the local errors $|u-u_{h\tau}|_{\mathcal{E}_Y^{\mathbf{a},n}}$ with constants that are fully robust with respect to both the spatial and temporal approximation orders. The estimators are flexible in the sense that they do not require restrictive transition conditions on the refinement and coarsening between time steps. We also showed that the composite norm of the error $\|u-u_{h\tau}\|_{\mathcal{E}_Y}$ is globally equivalent to $\|u-\mathcal{I}u_{h\tau}\|_Y$ up to the minimum of coarsening error and data oscillation, with polynomial-degree-robust constants in the equivalence. Finally, the analysis given here can be extended in various directions; in [18], we show that the equilibrated flux reconstruction employed here can also be used for obtaining a posteriori estimates for the X-norm of the error, with guaranteed upper bounds, and local space-time efficiency under the natural parabolic condition that $h^2 \lesssim \tau$. Furthermore, the adaptation of Lemma 8.1 to the case of residual-based estimators is currently under investigation.

REFERENCES

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

[2] J. Blechta, J. Málek, and M. Vohralík, *Localization of the $W^{-1,q}$ Norm for Local A Posteriori Efficiency*, preprint, hal-01332481, 2016.

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

[4] D. Braess and J. Schöberl, *Equilibrated residual error estimator for edge elements*, Math. Comp., 77 (2008), pp. 651–672.

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

[6] T. Dupont, *Mesh modification for evolution equations*, Math. Comp., 39 (1982), pp. 85–107.

[7] P. Destuynder and B. Métivet, *Explicit error bounds in a conforming finite element method*, Math. Comp., 68 (1999), pp. 1379–1396.

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

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

[10] V. Dolejší, A. Ern, and M. Vohralík, *hp-adaptation driven by polynomial-degree-robust a posteriori error estimates for elliptic problems*, SIAM J. Sci. Comput., 38 (2016), pp. 3320–3246.

[11] V. Dolejší, F. Roskovec, and M. Vlasák, *A posteriori error estimates for nonstationary problems*, in Numerical Mathematics and Advanced Applications, Lecture Notes in Computat. Sci. and Engrg., 112, Springer, 2016, pp. 225–233.

[12] 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 (1995), pp. 706–740.

[13] A. Ern and J. L. Guermond, *Theory and Practice of Finite Elements*, Appl. Math. Sci. 159, Springer, New York, 2004.

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

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

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

[17] A. Ern and M. Vohralík, *Stable Broken H^1 and $\boldsymbol{H}(\mathrm{div})$ Polynomial Extensions for Polynomial-Degree-Robust Potential and Flux Reconstruction in Three Space Dimensions*, preprint, arXiv:1701.02161, 2017.

[18] A. Ern, I. Smears, and M. Vohralík, *Equilibrated Flux A Posteriori Error Estimates in $L^2(H^1)$-Norms for High-Order Discretizations of Parabolic Problems*, preprint, arXiv:1703.04987, 2017.

[19] A. Ern, I. Smears, and M. Vohralík, *Discrete p-robust $\boldsymbol{H}(\mathrm{div})$-liftings and a posteriori error analysis of elliptic problems with H^{-1} source terms*, Calcolo, 54 (2017), pp. 1009–1025.

[20] L. C. Evans, *Partial Differential Equations*, Grad. Stud. Math. 19, AMS, Providence, RI, 1998.

[21] F. D. Gaspoz, C. Kreuzer, K. G. Siebert, and D. A. Ziegler, *A Convergent Time-Space Adaptive dG(s) Finite Element Method for Parabolic Problems Motivated by Equal Error Distribution*, preprint, arXiv:1610.06814, 2016.

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

[23] B. Hölm and T. P. Wihler, *Continuous and discontinuous Galerkin time stepping methods for nonlinear initial value problems with application to finite time blow-up*, Numer. Math., to appear.

[24] B. Jovanović and E. Süli, *Analysis of Finite Difference Schemes*, Springer Ser. Comput. Math. 46, Springer, London, 2014.
[25] C. Kreuzer, *Reliable and efficient a posteriori error estimates for finite element approximations of the parabolic p-Laplacian*, Calcolo, 50 (2013), pp. 79–110.
[26] C. Kreuzer, C. A. Möller, A. Schmidt, and K. G. Siebert, *Design and convergence analysis for an adaptive discretization of the heat equation*, IMA J. Numer. Anal., 32 (2012), pp. 1375–1403.
[27] O. Lakkis and C. Makridakis, *Elliptic reconstruction and a posteriori error estimates for fully discrete linear parabolic problems*, Math. Comp., 75 (2006), pp. 1627–1658.
[28] O. Lakkis, C. Makridakis, and T. Pryer, *A comparison of duality and energy a posteriori estimates for* $L_\infty(0,T;L_2(\Omega))$ *in parabolic problems*, Math. Comp., 84 (2015), pp. 1537–1569.
[29] J.-L. Lions and E. Magenes, *Non-homogeneous Boundary Value Problems and Applications, Vol.* I, Grundlehren Math. Wiss., Springer, New York, 1972.
[30] C. Makridakis and R. H. Nochetto, *Elliptic reconstruction and a posteriori error estimates for parabolic problems*, SIAM J. Numer. Anal., 41 (2003), pp. 1585–1594.
[31] C. Makridakis and R. H. Nochetto, *A posteriori error analysis for higher order dissipative methods for evolution problems*, Numer. Math., 104 (2006), pp. 489–514.
[32] J. M. Melenk, *hp-interpolation of nonsmooth functions and an application to hp-a posteriori error estimation*, SIAM J. Numer. Anal., 43 (2005), pp. 127–155.
[33] J. M. Melenk and B. I. Wohlmuth, *On residual-based a posteriori error estimation in hp-FEM*, Adv. Comput. Math., 15 (2001), pp. 311–331 (2002).
[34] S. Nicaise and N. Soualem, *A posteriori error estimates for a nonconforming finite element discretization of the heat equation*, ESAIM Math. Model. Numer. Anal., 39 (2005), pp. 319–348.
[35] M. Picasso, *Adaptive finite elements for a linear parabolic problem*, Comput. Methods Appl. Mech. Engrg., 167 (1998), pp. 223–237.
[36] S. Repin, *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 Mat. Appl. (9), 13 (2002), pp. 121–133.
[37] D. Schötzau and C. Schwab, *Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method*, SIAM J. Numer. Anal., 38 (2000), pp. 837–875.
[38] D. Schötzau and T. P. Wihler, *A posteriori error estimation for hp-version time-stepping methods for parabolic partial differential equations*, Numer. Math., 115 (2010), pp. 475–509.
[39] C. Schwab, *p- and hp-Finite Element Methods*, Numer. Math. Sci. Comput., Clarendon Press, New York, 1998.
[40] I. Smears, *Robust and efficient preconditioners for the discontinuous Galerkin time-stepping method*, IMA J. Numer. Anal., 37 (2016), pp. 1961–1985.
[41] F. Tantardini, *Quasi-Optimality in the Backward Galerkin Method for Linear Parabolic Problems*, Ph.D. thesis, University of Milan, Italy, Milan, 2014.
[42] F. Tantardini and A. Veeser, *The* L^2*-projection and quasi-optimality of Galerkin methods for parabolic equations*, SIAM J. Numer. Anal., 54 (2016), pp. 317–340.
[43] K. Urban and A. Patera, *An improved error bound for reduced basis approximation of linear parabolic problems*, Math. Comp. 83 (2014), pp. 1599–1615.
[44] R. Verfürth, *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 (1998), pp. 487–518.
[45] R. Verfürth, *A posteriori error estimates for finite element discretizations of the heat equation*, Calcolo, 40 (2003), pp. 195–212.
[46] R. Verfürth, *A Posteriori Error Estimation Techniques for Finite Element Methods*, Numer. Math. Sci. Comput., Oxford University Press, Oxford, 2013.
[47] T. Werder, K. Gerdes, D. Schötzau, and C. Schwab, *hp-discontinuous Galerkin time stepping for parabolic problems*, Comput. Methods Appl. Mech. Engrg., 190 (2001), pp. 6685–6708.
[48] J. Wloka, *Partial Differential Equations*, Cambridge University Press, Cambridge, 1987.