

# Guaranteed and robust $L^2$ -norm a posteriori error estimates for 1D linear advection problems

Alexandre Ern, Martin Vohralík, Mohammad Zakerzadeh

► **To cite this version:**

Alexandre Ern, Martin Vohralík, Mohammad Zakerzadeh. Guaranteed and robust  $L^2$ -norm a posteriori error estimates for 1D linear advection problems. 2019. hal-02105418

**HAL Id: hal-02105418**

**<https://hal.archives-ouvertes.fr/hal-02105418>**

Submitted on 21 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Guaranteed and robust $L^2$ -norm a posteriori error estimates for 1D linear advection problems\*

Alexandre Ern<sup>†‡</sup>    Martin Vohralík<sup>†‡</sup>    Mohammad Zakerzadeh<sup>†‡</sup>

April 21, 2019

## Abstract

We propose a reconstruction-based a posteriori error estimate for linear advection problems in one space dimension. In our framework, a stable variational ultra-weak formulation is adopted, and the equivalence of the  $L_2$ -norm of the error with the dual graph norm of the residual is established. This dual norm is showed to be localizable over vertex-based patch subdomains of the computational domain under the condition of the orthogonality of the residual to the piecewise affine hat functions. We show that this condition is valid for some well-known numerical methods including continuous/discontinuous Petrov–Galerkin and discontinuous Galerkin methods. Consequently, a well-posed local problem on each patch is identified, which leads to a global conforming reconstruction of the discrete solution. We prove that this reconstruction provides a guaranteed upper bound on the  $L_2$  error. Moreover, up to a constant, it also gives local lower bounds on the  $L_2$  error, where the generic constant is proven to be independent of mesh-refinement, polynomial degree of the approximation, and the advective velocity. This leads to robustness of our estimates with respect to the advection as well as the polynomial degree. All the above properties are verified in a series of numerical experiments, additionally leading to asymptotic exactness. Motivated by these results, we finally propose a heuristic extension of our methodology to any space dimension, achieved by solving local least-squares problems on vertex-based patches. Though not anymore guaranteed, the resulting error indicator is numerically robust with respect to both advection velocity and polynomial degree, for a collection of two-dimensional test cases including discontinuous solutions.

**Key words:** linear advection problem; discontinuous Galerkin method; Petrov–Galerkin method; a posteriori error estimate; local efficiency; advection robustness; polynomial-degree robustness

## 1 Introduction

This work deals with a linear advection equation of the form: find  $u : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$  such that

$$\mathbf{b} \cdot \nabla u = f, \quad \text{in } \Omega, \quad (1.1a)$$

$$u = 0, \quad \text{on } \partial_- \Omega. \quad (1.1b)$$

---

<sup>†</sup>Inria, 2 rue Simone Iff, 75589 Paris, France

<sup>‡</sup>Université Paris-Est, CERMICS (ENPC), 77455 Marne-la-Vallée, France

alexandre.ern@enpc.fr, martin.vohralik@inria.fr, seyed-mohammad.zakerzadeh@inria.fr

\*This work was funded by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement No 647134 GATIPOR).

The velocity field  $\mathbf{b} \in C^1(\bar{\Omega}; \mathbb{R}^d)$ ,  $\mathbf{b} \neq 0$ , is considered to be divergence-free and we take into account a general source term  $f \in L_2(\Omega)$ . The inflow, outflow, and characteristic parts of the boundary are denoted by  $\partial_-\Omega$ ,  $\partial_+\Omega$ , and  $\partial_0\Omega$ , respectively, with the definitions

$$\partial_{\pm}\Omega := \{x \in \partial\Omega : \pm \mathbf{b}(x) \cdot \mathbf{n}(x) > 0\}, \quad \partial_0\Omega := \{x \in \partial\Omega : \mathbf{b}(x) \cdot \mathbf{n}(x) = 0\}.$$

In the main body of the paper, we focus on the one-dimensional case  $d = 1$ , where  $\Omega \subset \mathbb{R}$  is a bounded simply connected interval; then  $\mathbf{b}$  is a constant scalar. We keep the notations in multi-dimensional form in order to be applicable when we discuss extensions of our results to the multi-dimensional case.

The a posteriori error analysis for problem (1.1) admits a range of functional frameworks and consequently different norms in which the error can be measured. Our goal is to derive an  $L_2$ -norm error estimate of the form

$$\|u - u_h\|_{L_2(\Omega)} \leq \eta, \tag{1.2}$$

where  $u$  is the weak solution of (1.1) in  $L_2(\Omega)$ ,  $u_h$  is its numerical approximation, and  $\eta$  is an a posteriori error estimator *fully computable* from  $u_h$  by some *local procedure*. We seek to have a bound that is *guaranteed*, i.e., featuring no unknown constant, in contrast to *reliability* where a bound up to a generic constant is sufficient. We develop a unified framework treating several classical numerical methods at once. Importantly, we also prove a converse estimate to (1.2) in the form

$$\eta \leq C \|u - u_h\|_{L_2(\Omega)} + \text{data oscillation}. \tag{1.3}$$

This is called *global efficiency* and yields equivalence between the incomputable error  $\|u - u_h\|_{L_2(\Omega)}$  and the computable estimator  $\eta$  up to the data oscillation term that vanishes for piecewise polynomial datum  $f$  and that is of higher order than the error for piecewise smooth datum  $f$ . Crucially, in our developments, the generic constant  $C$  in (1.3) only depends on the mesh shape regularity, requesting for  $d = 1$  each two neighboring elements to be of comparable size. In particular,  $C$  is independent of the problem parameters  $\mathbf{b}$  and  $f$  as well as of the polynomial degree of the approximation  $k$ , yielding both data- and polynomial-degree-*robustness*. We actually also show local efficiency, i.e., a localized version of (1.3), which is highly desirable on the practical side in view of adaptive mesh refinement.

To achieve the above-mentioned goals, we start with the ultra-weak variational formulation at the infinite-dimensional level, where the solution lies in the  $L_2(\Omega)$  trial space and the test space is formed by  $H^1(\Omega)$  functions taking zero value at the outflow boundary. In this setting, we prove the equality of the  $L_2$ -norm of the error with the dual graph norm (relying on  $\|\mathbf{b} \cdot \nabla(\cdot)\|_{L_2(\Omega)}$ ) of the residual. In the one-dimensional case, we are able to prove that the global dual norm can be localized over vertex-based patches of elements under an orthogonality condition against the hat basis functions. Consequently, suitable discrete local problems posed over these patches are identified which lead to local reconstructions  $s_h^a$  combined into a global reconstruction  $s_h$  such that  $\|u_h - s_h\|_{L_2(\Omega)}$  forms the main ingredient of the estimator  $\eta$  satisfying (1.2) and (1.3).

Let us recall some important contributions to a posteriori error estimation for problem (1.1). Bey and Oden in [5] proposed an a posteriori error estimate for a discontinuous Galerkin (dG) formulation of the multi-dimensional advection–reaction problem. In this framework, two infinite-dimensional problems have to be solved on each mesh element; one to obtain the lower bound on the error and one for the upper bound, in two different and inequivalent weighted energy norms. This gives estimates similar to (1.2) and (1.3), but for two different estimators and in two different norms of the error. Additionally, one cannot solve analytically the infinite-dimensional element-wise problems, and, in practice, one needs to approximate them by some higher-order finite

element approximation. Hence, neither simultaneous reliability and efficiency, nor robustness, are granted.

Süli in [29] applied the  $H^1$ -stability result of Tartakoff [31] to the adjoint problem of (1.1) (with the presence of the reaction term and in the multi-dimensional case), and obtained a global reliable upper bound on the  $H^{-1}$ -norm of the error in terms of the  $L_2$ -norm of the residual for a weak formulation of (1.1) with distinct trial and test spaces. He further turned this bound into a reliable  $H^{-1}$ -norm a posteriori error indicator for the streamline-diffusion finite element and the cell-vertex finite volume methods. However, neither the efficiency nor the robustness of this error indicator are discussed. Furthermore, in [29] by Süli and in [21] by Süli and Houston, an analysis of the multi-dimensional advection–reaction problem in the graph space equipped with the full norm  $\|\cdot\|_{L_2(\Omega)} + \|\mathbf{b}\cdot\nabla(\cdot)\|_{L_2(\Omega)}$ , is provided. This functional setting provides the equivalence between the  $L_2$ -norm of the error and the dual graph norm of the residual up to some generic constants, which is a weaker error–residual equivalence result compared to what we establish in the current work, see Theorem 2.1 below, upon replacing the full graph norm by  $\|\mathbf{b}\cdot\nabla(\cdot)\|_{L_2(\Omega)}$  only. Overall, no complete reliability, efficiency, and robustness results of the form (1.2)–(1.3) are obtained.

Becker *et al.* in [4], derived reconstruction-based error estimators for the advection problem (1.1) in two space dimensions. An  $\mathbf{H}(\text{div}, \Omega)$ -conforming reconstruction is proposed for the flux vector  $\mathbf{b}u$  (instead of  $u$  in the present work) which is designed to produce a guaranteed upper bound on the error measured in some dual norm of the advection operator. A unified framework is built, covering the dG and conforming finite element methods with/without stabilization terms. This dual norm is hard to evaluate even for a known exact solution, and, in practice, the authors replace it by the  $L_2$ -norm, so that the guaranteed upper bound property is eventually lost. Proofs of efficiency or robustness are not given, but optimal convergence orders of the estimator are observed in numerical experiments. It is worth mentioning that, restricted to one space dimension, the dual norm of [4] reduces to the weak graph norm we employ. Our contribution in this respect consists in the proofs of (1.2) and (1.3), not given in [4] (where, recall, two space dimensions are treated.)

In a recent result by Georgoulis *et al.* in [19], the authors used the reconstruction proposed by Makridakis and Nochetto in [23] for a dG approximation and provided a reliable upper bound on the error in the energy norm for one-dimensional advection–diffusion–reaction problems, as well as a reliable  $L_2$ -norm estimate for the problem (1.1) in one space dimension. Though a proof of (1.3) is not given, efficiency and robustness are numerically observed. One might also note the earlier work of these authors [18], dedicated to the two-dimensional advection–reaction problem with a similar reconstruction. In that work, a reliable bound on the energy norm of the error is presented, though again without a theoretical elaboration on the efficiency and robustness.

Finally, let us mention the recent result of Dahmen and Stevenson in [12] where the authors provide a posteriori error estimates for the discontinuous Petrov–Galerkin method tailored to the transport equations in two space dimensions. The equivalence of the errors of the bulk and skeleton quantities with the dual norm of the residual is established. This dual norm is later approximated by some equivalent yet computable indicator. The absorbed constants translate into a constant  $C$  in (1.3) which depends on the advective field  $\mathbf{b}$  and the polynomial degree of approximation, therefore precluding the robustness of the error lower bound.

We also mention that in the case of advection–diffusion(–reaction) problems, other approaches were previously considered to obtain robustness with respect to the advective field. Among them, Verfürth [32] proposed to augment the energy norm by a dual norm coming from the skew-symmetric part of the differential operator, and Sangalli [25, 26] used interpolated spaces and a fractional-order norm for the advective term. Extensions of these approaches can be found in [27, 28, 15]. However, the above results are not applicable when the diffusion parameter

vanishes, i.e., as the advection–diffusion problem reduces to (1.1) because the diffusive part of the operator is needed to evaluate the dual norm.

We treat problem (1.1) in one space dimension in Sections 2–8. Section 2 deals with the functional settings, whereby adopting the ultra-weak variational formulation. We prove, in particular, the equality of the  $L_2$ -norm of the error and the dual norm of the residual. Section 3 introduces some numerical schemes for approximating (1.1). Section 4 discusses the localization of the dual norm of the residual over vertex-based patches, showing in particular that this is possible for the schemes discussed in Section 3. In Section 5, we present our local patchwise reconstruction. Sections 6 and 7 then present the proofs for the upper and lower bounds as well as robustness in the form of (1.2)–(1.3). Section 8 then contains results of several numerical experiments to illustrate the developed theory. Finally, in Section 9, we consider the advection problem (1.1) in multiple space dimensions and derive a heuristic extension of our methodology to this case. Although we cannot prove here the guaranteed upper bound, (local) efficiency, and robustness, numerical experiments indicate appreciable properties of the derived estimates also in this case.

## 2 Abstract framework

We start with the presentation of the abstract framework.

### 2.1 Spaces

In the one-dimensional case, the constraint of  $\mathbf{b}$  being a non-zero divergence-free field is translated to  $\mathbf{b}$  being a constant nonzero scalar. Consequently, we are lead to work with the spaces

$$H_-^1(\Omega) = \{w \in H^1(\Omega), w = 0, \text{ on } \partial_-\Omega\}, \quad H_+^1(\Omega) = \{w \in H^1(\Omega), w = 0, \text{ on } \partial_+\Omega\}. \quad (2.1)$$

The trace operator in these spaces is well-defined and the following *integration-by-parts* formula holds:

$$(v, \mathbf{b} \cdot \nabla w)_\Omega + (\mathbf{b} \cdot \nabla v, w)_\Omega = (\mathbf{b} \cdot \mathbf{n} v, w)_{\partial\Omega}, \quad \forall v, w \in H^1(\Omega), \quad (2.2)$$

where the notation  $(v, w)_D := \int_D v w$  is used for an open subdomain  $D \subseteq \Omega$  or its boundary  $\partial D$  and for integrable functions  $v$  and  $w$ . Henceforth,  $\|v\|_D$  denotes the norm  $\|v\|_{L_2(D)} = \sqrt{(v, v)_D}$ . We will drop the subscript when  $D = \Omega$ .

### 2.2 Poincaré inequalities

The Poincaré inequality states that

$$\|v - \bar{v}\|_D \leq h_D C_{P,D} \|\nabla v\|_D, \quad \forall v \in H^1(D), \quad (2.3a)$$

with  $C_{P,D} > 0$  a generic constant, in particular equal to  $1/\pi$  for convex  $D \subset \Omega$ . Here  $\bar{v}$  is the mean value of  $v$  over  $D$  defined as  $\bar{v} = (v, 1)_D / |D|$  and  $h_D$  is the diameter of  $D$ . Similarly, another Poincaré inequality (sometimes called Friedrichs inequality) states that

$$\|v\|_D \leq h_D C_{P,D,\partial_0 D} \|\nabla v\|_D, \quad \forall v \in \{H^1(D), v|_{\partial_0 D} = 0, |\partial_0 D| \neq 0\}, \quad (2.3b)$$

where  $\partial_0 D \subset \partial D$ ; typically  $C_{P,D,\partial_0 D} = 1$ . Henceforth, we will use  $C_{PF,D}$  as a general notation for both  $C_{P,D}$  and  $C_{P,D,\partial_0 D}$ . It follows from the above that for a one-dimensional interval  $D$ ,  $C_{PF,D}$  can be taken as 1.

### 2.3 Variational formulation and residual

The variational framework hinges upon an appropriate choice of the trial and test spaces and their corresponding norms. In particular, it turns out natural to work on spaces well-suited to the non-symmetric structure of the problem. Here we consider Hilbert spaces (non-symmetric formulations in Banach spaces can be found in [8, 24]).

The (usual) weak formulation of (1.1) reads: find  $u \in H_-^1(\Omega)$  such that

$$(\mathbf{b} \cdot \nabla u, v) = (f, v), \quad \forall v \in L_2(\Omega). \quad (2.4)$$

It is classically well-posed as one might confer with [17], [22], and [29, Proposition 6], cf. also [14] and [11, Rem. 2.2]. Here, we rather adopt the so-called *ultra-weak* formulation of problem (1.1) where the bilinear form is obtained by casting the derivatives on the test function, using integration-by-parts. It reads: find  $u \in L_2(\Omega)$  such that

$$-(u, \mathbf{b} \cdot \nabla v) = (f, v), \quad \forall v \in H_+^1(\Omega). \quad (2.5)$$

The well-posedness of (2.5) can be shown by inf-sup arguments (cf. [14, Theorem 2.6] and [11, Theorem 2.4]).

Denote by  $H_+^1(\Omega)'$  the dual space to  $H_+^1(\Omega)$ . For an arbitrary  $u_h \in L_2(\Omega)$ , the formulation (2.5) leads to the definition of the residual  $\mathcal{R}(u_h)$ , a bounded linear functional on  $H_+^1(\Omega)'$ , by

$$\langle \mathcal{R}(u_h), v \rangle := (f, v) + (u_h, \mathbf{b} \cdot \nabla v), \quad \forall v \in H_+^1(\Omega). \quad (2.6)$$

We define its velocity-scaled dual norm by

$$\|\mathcal{R}(u_h)\|_{\mathbf{b}; H_+^1(\Omega)'} := \sup_{v \in H_+^1(\Omega) \setminus \{0\}} \frac{\langle \mathcal{R}(u_h), v \rangle}{\|\mathbf{b} \cdot \nabla v\|}. \quad (2.7)$$

### 2.4 Error-residual equivalence

In this section, we present an important connection between the  $L_2(\Omega)$ -norm of the error and the residual norm (2.7):

**Theorem 2.1** (error-residual equivalence). *Let  $u$  be the ultra-weak solution of (2.5). Then*

$$\|u - u_h\| = \|\mathcal{R}(u_h)\|_{\mathbf{b}; H_+^1(\Omega)'} \quad \forall u_h \in L_2(\Omega).$$

*Proof.* The well-posedness of the weak formulation (2.4), for the velocity field  $-\mathbf{b}$ , implies that for all  $v \in L_2(\Omega)$ , there exists a unique  $z \in H_+^1(\Omega)$  such that

$$-(\mathbf{b} \cdot \nabla z, w) = (v, w) \quad \forall w \in L_2(\Omega).$$

This clearly gives  $\|\mathbf{b} \cdot \nabla z\| = \|v\|$ . Hence, for any  $w \in L_2(\Omega)$ , we have

$$\|w\| = \sup_{v \in L_2(\Omega) \setminus \{0\}} \frac{(w, v)}{\|v\|} = \sup_{z \in H_+^1(\Omega) \setminus \{0\}} \frac{-(w, \mathbf{b} \cdot \nabla z)}{\|\mathbf{b} \cdot \nabla z\|},$$

and the claim follows by the choice  $w = u - u_h$  and using the definitions (2.5) and (2.6).  $\square$

Compared to the similar equivalence provided in [21, Theorem 3.3], Theorem 2.1 shows a form of equality which declares the optimality of the chosen spaces and norms. This is advantageous for the sharpness of the a posteriori error estimation.

### 3 Examples of numerical methods

Let  $\mathcal{T}_h = \{K\}$  be a mesh of  $\Omega$ , i.e., a division of the one-dimensional domain  $\Omega$  into non-overlapping intervals covering  $\Omega$ , shape regular in the sense that two neighboring intervals are of comparable size, up to a constant  $\kappa_{\mathcal{T}_h}$ . Let us denote  $h_K := \text{diam}(K)$  and  $h := \max_{K \in \mathcal{T}_h} h_K$ . We also denote by  $\mathcal{E}_h := \cup_{K \in \mathcal{T}_h} \partial K$  the skeleton of the triangulation  $\mathcal{T}_h$ , coinciding with the set of mesh vertices  $\mathcal{V}_h$  in the present one-dimensional case. Moreover, we need to consider the decompositions  $\mathcal{E}_h = \mathcal{E}_h^{\text{int}} \cup \mathcal{E}_h^{\text{bnd}}$  into internal and boundary faces and  $\mathcal{V}_h = \mathcal{V}_h^{\text{int}} \cup \mathcal{V}_h^{\partial-\Omega} \cup \mathcal{V}_h^{\partial+\Omega}$  into internal, inflow, and outflow vertices, so that in the one-dimensional case  $\mathcal{E}_h = \mathcal{V}_h$  and  $\mathcal{E}_h^{\text{bnd}} = \mathcal{V}_h^{\partial-\Omega} \cup \mathcal{V}_h^{\partial+\Omega}$ . Let  $\mathcal{P}^k(\mathcal{T}_h)$  denote piecewise polynomial functions of at most degree  $k$  on the mesh  $\mathcal{T}_h$ . The following three numerical methods are classical examples of discretizations of (1.1). Please note that in Examples 3.1 and 3.3, we exclude the lowest polynomial degrees. We need to do so to comply with the orthogonality condition in Assumption 4.2, see Lemma 4.4 below.

The first finite element scheme is a finite-dimensional version of the weak formulation (2.4):

**Example 3.1** (continuous trial Petrov–Galerkin (PG1) finite element). *Find  $u_h \in X_h := H_-^1(\Omega) \cap \mathcal{P}^k(\mathcal{T}_h)$ ,  $k \geq 2$ , such that*

$$(\mathbf{b} \cdot \nabla u_h, v_h) = (f, v_h) \quad \forall v_h \in Y_h := \mathcal{P}^{k-1}(\mathcal{T}_h). \quad (3.1)$$

The second finite element scheme stems from the ultra-weak formulation (2.5):

**Example 3.2** (discontinuous trial Petrov–Galerkin (PG2) finite element). *Find  $u_h \in X_h := \mathcal{P}^k(\mathcal{T}_h)$ ,  $k \geq 0$ , such that*

$$-(u_h, \mathbf{b} \cdot \nabla v_h) = (f, v_h) \quad \forall v_h \in Y_h := H_+^1(\Omega) \cap \mathcal{P}^{k+1}(\mathcal{T}_h). \quad (3.2)$$

Finally, the dG method for problem (1.1) (letting  $\nabla$  also denote the broken (elementwise) gradient) reads:

**Example 3.3** (dG finite element). *Find  $u_h \in X_h := \mathcal{P}^k(\mathcal{T}_h)$ ,  $k \geq 1$ , such that*

$$\mathcal{B}_h(u_h, v_h) = (f, v_h) \quad \forall v_h \in Y_h := \mathcal{P}^k(\mathcal{T}_h), \quad (3.3a)$$

where

$$\begin{aligned} \mathcal{B}_h(u_h, v_h) := & - \sum_{K \in \mathcal{T}_h} (u_h, \mathbf{b} \cdot \nabla v_h)_K \\ & - \sum_{e \in \mathcal{E}_h^{\text{int}}} \mathbf{b} \cdot \mathbf{n} \{ \{ u_h \} \} [v_h] + \sum_{e \in \mathcal{E}_h^{\text{int}}} \frac{1}{2} | \mathbf{b} \cdot \mathbf{n} | [ \{ u_h \} ] [v_h] + \sum_{e \in \mathcal{E}_h^{\text{bnd}}} (\mathbf{b} \cdot \mathbf{n})^+ u_h v_h. \end{aligned} \quad (3.3b)$$

Here the notation  $u_h^-$  and  $u_h^+$  stands for the trace value on a vertex from left and from right, respectively, the average is defined as  $\{ \{ u_h \} \} := (u_h^- + u_h^+)/2$ , and the jump is defined as  $[ \{ u_h \} ] := u_h^+ - u_h^-$ . In this formulation, the upwind dG flux is applied on the cell interfaces.

### 4 Error localization

In this section, we show that if the numerical solution  $u_h$  satisfies a first-order orthogonality condition with respect to hat basis functions, one can obtain a two-sided bound on  $\| \mathcal{R}(u_h) \|_{\mathbf{b}; H_+^1(\Omega)}$  by identifying some (infinite-dimensional) problems on patches of elements around vertices.

## 4.1 Patches and partition of unity by the hat functions

Let  $\mathcal{T}_{\mathbf{a}}$  denote the *patch* of all simplices which share the given vertex  $\mathbf{a}$ ,  $\mathcal{T}_{\mathbf{a}} := \{K, \mathbf{a} \in \mathcal{V}_K\}$ . Let  $\omega_{\mathbf{a}}$  be the corresponding open subdomain. Then  $\cup_{\mathbf{a} \in \mathcal{V}_h} \omega_{\mathbf{a}}$  forms an overlapping partition of  $\Omega$ , with  $\mathcal{N} = 2$  maximal overlap in one space dimension. For all  $\mathbf{a} \in \mathcal{V}_h$ , let  $\psi_{\mathbf{a}} \in H^1(\Omega) \cap \mathcal{P}^1(\mathcal{T}_h)$  be the piecewise affine hat function, taking value 1 in vertex  $\mathbf{a}$  and 0 in all other vertices. The hat functions verify  $\text{supp}(\psi_{\mathbf{a}}) = \overline{\omega_{\mathbf{a}}}$  and form a *partition of unity* as

$$\sum_{\mathbf{a} \in \mathcal{V}_h} \psi_{\mathbf{a}} = 1. \quad (4.1)$$

## 4.2 Cut-off estimates

Similarly to (2.1), let  $H_{\pm}^1(\omega_{\mathbf{a}})$  contain those functions from  $H^1(\omega_{\mathbf{a}})$  with zero trace on the outflow boundary of  $\omega_{\mathbf{a}}$ . Define two patchwise spaces

$$H_{\#}^1(\omega_{\mathbf{a}}) := \begin{cases} H_0^1(\omega_{\mathbf{a}}), & \mathbf{a} \notin \mathcal{V}_h^{\partial-\Omega}, \\ H_{+}^1(\omega_{\mathbf{a}}), & \mathbf{a} \in \mathcal{V}_h^{\partial-\Omega}, \end{cases} \quad (4.2)$$

and

$$H_{*}^1(\omega_{\mathbf{a}}) := \begin{cases} \{H^1(\omega_{\mathbf{a}}) : (v, 1)_{\omega_{\mathbf{a}}} = 0\}, & \mathbf{a} \notin \mathcal{V}_h^{\partial+\Omega}, \\ H_{+}^1(\omega_{\mathbf{a}}), & \mathbf{a} \in \mathcal{V}_h^{\partial+\Omega}. \end{cases} \quad (4.3)$$

In the sequel, we will use several times the following fact:

$$v \in H_{*}^1(\omega_{\mathbf{a}}) \implies \psi_{\mathbf{a}} v \in H_{\#}^1(\omega_{\mathbf{a}}). \quad (4.4)$$

Let us define

$$C_{\text{cont,PF}} := \max_{\mathbf{a} \in \mathcal{V}_h} (1 + C_{\text{PF},\omega_{\mathbf{a}}} h_{\omega_{\mathbf{a}}} \|\nabla \psi_{\mathbf{a}}\|_{\infty}).$$

We notice that this constant only depends on the shape-regularity constant  $\kappa_{\mathcal{T}_h}$ . As in the present one-dimensional setting,  $\mathbf{b}$  is a constant scalar, the following cut-off Poincaré estimate follows immediately from [10, Theorem 3.1] or [7, Section 3], cf. also [16, Lemma 3.12].

**Lemma 4.1** (local cut-off estimate). *For any  $\mathbf{a} \in \mathcal{V}_h$ , we have*

$$\|\mathbf{b} \cdot \nabla(\psi_{\mathbf{a}} v)\|_{\omega_{\mathbf{a}}} \leq C_{\text{cont,PF}} \|\mathbf{b} \cdot \nabla v\|_{\omega_{\mathbf{a}}} \quad \forall v \in H_{*}^1(\omega_{\mathbf{a}}).$$

## 4.3 Error localization

The following assumption on the  $\psi_{\mathbf{a}}$ -orthogonality of the residual will be crucial to localize the error:

**Assumption 4.2** ( $\psi_{\mathbf{a}}$ -orthogonality). *The residual  $\mathcal{R}(u_h) \in H_{+}^1(\Omega)'$  defined in (2.6) satisfies*

$$\langle \mathcal{R}(u_h), \psi_{\mathbf{a}} \rangle = (f, \psi_{\mathbf{a}})_{\omega_{\mathbf{a}}} + (u_h, \mathbf{b} \cdot \nabla \psi_{\mathbf{a}})_{\omega_{\mathbf{a}}} = 0 \quad \forall \mathbf{a} \in \mathcal{V}_h^{\text{int}} \cup \mathcal{V}_h^{\partial-\Omega}. \quad (4.5)$$

Since the zero-extension of a function in  $H_{\#}^1(\omega_{\mathbf{a}})$  is in  $H_{+}^1(\Omega)$ , we can define the restriction of  $\mathcal{R}(u_h)$  from (2.6) to the space  $H_{\#}^1(\omega_{\mathbf{a}})$  as

$$\|\mathcal{R}(u_h)\|_{\mathbf{b}; H_{\#}^1(\omega_{\mathbf{a}})'} := \sup_{v \in H_{\#}^1(\omega_{\mathbf{a}}) \setminus \{0\}} \frac{\langle \mathcal{R}(u_h), v \rangle}{\|\mathbf{b} \cdot \nabla v\|_{\omega_{\mathbf{a}}}}.$$

We then have:



**Theorem 4.3** (localizability of residual dual norms with  $\psi_{\mathbf{a}}$ -orthogonality). *Provided  $\mathcal{R}(u_h)$  satisfies Assumption 4.2, we have*

$$\|\mathcal{R}(u_h)\|_{\mathbf{b}; H_{\pm}^1(\Omega)'}^2 \leq 2C_{\text{cont,PF}}^2 \sum_{\mathbf{a} \in \mathcal{V}_h} \|\mathcal{R}(u_h)\|_{\mathbf{b}; H_{\#}^1(\omega_{\mathbf{a}})'}^2. \quad (4.6a)$$

Independently of Assumption 4.2, the following always holds true:

$$\sum_{\mathbf{a} \in \mathcal{V}_h} \|\mathcal{R}(u_h)\|_{\mathbf{b}; H_{\#}^1(\omega_{\mathbf{a}})'}^2 \leq 2\|\mathcal{R}(u_h)\|_{\mathbf{b}; H_{\pm}^1(\Omega)'}^2. \quad (4.6b)$$

*Proof.* The proof proceeds along the lines in [6, 10, 7, 16]. In particular, noting the partition of unity property (4.1) and the  $\psi_{\mathbf{a}}$ -orthogonality of Assumption 4.2, one can use  $v = \sum_{\mathbf{a} \in \mathcal{V}_h} \psi_{\mathbf{a}} v$  as the test function to obtain, for each  $v \in H_{\pm}^1(\Omega)$

$$\langle \mathcal{R}(u_h), v \rangle \stackrel{(4.1),(4.5)}{=} \sum_{\mathbf{a} \in \mathcal{V}_h^{\text{int}} \cup \mathcal{V}_h^{\partial-\Omega}} \langle \mathcal{R}(u_h), \psi_{\mathbf{a}}(v - \bar{v}_{\mathbf{a}}) \rangle + \sum_{\mathbf{a} \in \mathcal{V}_h^{\partial+\Omega}} \langle \mathcal{R}(u_h), \psi_{\mathbf{a}} v \rangle,$$

where  $\bar{v}_{\mathbf{a}}$  is the mean value of  $v$  on  $\omega_{\mathbf{a}}$ . Let  $w_{\mathbf{a}} := v - \bar{v}_{\mathbf{a}}|_{\omega_{\mathbf{a}}}$  if  $\mathbf{a} \in \mathcal{V}_h^{\text{int}} \cup \mathcal{V}_h^{\partial-\Omega}$  and  $w_{\mathbf{a}} := v|_{\omega_{\mathbf{a}}}$  if  $\mathbf{a} \in \mathcal{V}_h^{\partial+\Omega}$ . Then,  $w_{\mathbf{a}} \in H_{*}^1(\omega_{\mathbf{a}})$ , so that  $\psi_{\mathbf{a}} w_{\mathbf{a}} \in H_{\#}^1(\omega_{\mathbf{a}})$  by (4.4). Using the cut-off estimate of Lemma 4.1 for  $v = w_{\mathbf{a}}$ , one can in particular obtain (4.6a) and (4.6b) like in [6, Theorem 3.7].  $\square$

#### 4.4 $\psi_{\mathbf{a}}$ -orthogonality of the residual for the methods of Section 3

The following lemma assesses the validity of Assumption 4.2 for the three methods presented in Section 3:

**Lemma 4.4** ( $\psi_{\mathbf{a}}$ -orthogonality of the residual). *For PG1 of Example 3.1 with  $k \geq 2$ , PG2 of Example 3.2 with  $k \geq 0$ , and dG of Example 3.3 with  $k \geq 1$ , Assumption 4.2 holds true.*

*Proof.* Let  $\mathbf{a} \in \mathcal{V}_h^{\text{int}} \cup \mathcal{V}_h^{\partial-\Omega}$ . We verify the condition for each method:

- From definition (2.6), for the PG1 method (3.1), we have

$$\begin{aligned} \langle \mathcal{R}(u_h), \psi_{\mathbf{a}} \rangle &= \sum_{K \in \mathcal{T}_{\mathbf{a}}} \{ (f, \psi_{\mathbf{a}})_K + (u_h, \mathbf{b} \cdot \nabla \psi_{\mathbf{a}})_K \} \\ &\stackrel{\text{I.B.P.}}{=} \sum_{K \in \mathcal{T}_{\mathbf{a}}} \{ (f, \psi_{\mathbf{a}})_K - (\mathbf{b} \cdot \nabla u_h, \psi_{\mathbf{a}})_K + (\mathbf{b} \cdot \mathbf{n} u_h, \psi_{\mathbf{a}})_{\partial K} \}. \end{aligned} \quad (4.7)$$

For all  $\mathbf{a} \in \mathcal{V}_h^{\text{int}}$ , the jump  $[[\psi_{\mathbf{a}}]]$  vanishes at the vertex  $\mathbf{a}$  and  $\psi_{\mathbf{a}} = 0$  on the boundary edge of the patch. Hence, since  $u_h$  is also continuous in  $\mathbf{a}$  in the PG1 method, the last term in (4.7) disappears and one infers that

$$\langle \mathcal{R}(u_h), \psi_{\mathbf{a}} \rangle = (f, \psi_{\mathbf{a}})_{\Omega} - (\mathbf{b} \cdot \nabla u_h, \psi_{\mathbf{a}})_{\Omega} \stackrel{(3.1)}{=} 0,$$

since we assume  $k \geq 2$ , so that  $\psi_{\mathbf{a}} \in Y_h$ . The same result is valid for  $\mathbf{a} \in \mathcal{V}_h^{\partial-\Omega}$  since  $u_h = 0$  on the inflow as imposed in the definition of  $X_h$ .

- From definition (2.6) and employing the PG2 characterization (3.2), we obtain in a straightforward manner that

$$\langle \mathcal{R}(u_h), \psi_{\mathbf{a}} \rangle = 0$$

for all  $k \geq 0$ .

- For the dG method (3.3), noting that  $(\mathbf{b}\cdot\mathbf{n})^+ = 0$  on the inflow and using the same arguments on the vanishing of the jump  $[[\psi_{\mathbf{a}}]]$  and some  $k \geq 1$  by assumption, we have for any vertex  $\mathbf{a} \in \mathcal{V}_h^{\text{int}} \cup \mathcal{V}_h^{\partial-\Omega}$

$$\sum_{e \in \mathcal{E}_h^{\text{int}}} \left\{ \frac{1}{2} |\mathbf{b}\cdot\mathbf{n}| [[u_h]] - \mathbf{b}\cdot\mathbf{n} \{u_h\} \right\} [[\psi_{\mathbf{a}}]] + \sum_{e \in \mathcal{E}_h^{\text{bnd}}} (\mathbf{b}\cdot\mathbf{n})^+ u_h \psi_{\mathbf{a}} = 0.$$

Hence, also employing definition (2.6), we infer that

$$\langle \mathcal{R}(u_h), \psi_{\mathbf{a}} \rangle = \sum_{K \in \mathcal{T}_{\mathbf{a}}} \{ (f, \psi_{\mathbf{a}})_K + (u_h, \mathbf{b}\cdot\nabla \psi_{\mathbf{a}})_K \} = 0$$

for all  $k \geq 1$  which implies  $\psi_{\mathbf{a}} \in Y_h$ .

□

## 5 Local problems on patches

In this section, we present a local reconstruction technique which provides the key ingredient to evaluate our a posteriori error estimator.

**Definition 5.1** (patchwise problems). *Let  $u_h \in L_2(\Omega)$  satisfy Assumption 4.2. For all vertices  $\mathbf{a} \in \mathcal{V}_h$ , let  $s_h^{\mathbf{a}} \in X_h^{\mathbf{a}}$  be the solution of the following advection–reaction problem on the patch  $\omega_{\mathbf{a}}$*

$$(\mathbf{b}\cdot\nabla(\psi_{\mathbf{a}} s_h^{\mathbf{a}}), v_h)_{\omega_{\mathbf{a}}} = (f \psi_{\mathbf{a}} + (\mathbf{b}\cdot\nabla \psi_{\mathbf{a}}) u_h, v_h)_{\omega_{\mathbf{a}}} \quad \forall v_h \in Y_h^{\mathbf{a}}, \quad (5.1)$$

with the finite-dimensional spaces

$$X_h^{\mathbf{a}} := \mathcal{P}^{k'}(\mathcal{T}_{\mathbf{a}}) \cap H^1(\omega_{\mathbf{a}}), \quad Y_h^{\mathbf{a}} := \mathcal{P}^{k'}(\mathcal{T}_{\mathbf{a}})$$

and  $k' \geq 0$ . Define the global reconstruction  $s_h$  by

$$s_h := \sum_{\mathbf{a} \in \mathcal{V}_h} \psi_{\mathbf{a}} s_h^{\mathbf{a}}. \quad (5.2)$$

**Remark 1** (trial and test spaces). *A priori, the number of degrees of freedom in  $X_h^{\mathbf{a}}$  and  $Y_h^{\mathbf{a}}$  for  $\mathbf{a} \in \mathcal{V}_h^{\text{int}}$  does not match; while there exist  $2(k'+1)$  linearly independent test functions in  $Y_h^{\mathbf{a}}$ , the trial space  $X_h^{\mathbf{a}}$  has only  $2k'+1$  degrees of freedom. For any  $\mathbf{a} \in \mathcal{V}_h^{\text{int}}$ , though, the test function in (5.1) given by  $v_h = 1$  on both  $K \in \mathcal{T}_{\mathbf{a}}$  is actually superfluous. Indeed, on the one hand, we have*

$$(\mathbf{b}\cdot\nabla(\psi_{\mathbf{a}} s_h^{\mathbf{a}}), 1)_{\omega_{\mathbf{a}}} = (\mathbf{b}\cdot\mathbf{n}, \psi_{\mathbf{a}} s_h^{\mathbf{a}})_{\partial\omega_{\mathbf{a}}} = 0, \quad (5.3)$$

according to the definition of  $\psi_{\mathbf{a}}$ . On the other hand, Assumption 4.2 guarantees that the right-hand side vanishes in such a case, hence

$$(f \psi_{\mathbf{a}} + (\mathbf{b}\cdot\nabla \psi_{\mathbf{a}}) u_h, 1)_{\omega_{\mathbf{a}}} = \langle \mathcal{R}(u_h), \psi_{\mathbf{a}} \rangle = 0.$$

We next show that the solution of (5.1) uniquely exists and the proposed reconstruction is well-posed:

**Lemma 5.2** (well-posedness of Definition 5.1). *There exists a unique solution  $s_h^\alpha \in X_h^\alpha$  of problem (5.1). It is stable in the sense that*

$$\|s_h^\alpha\|_{\omega_\alpha} \leq C(h_{\omega_\alpha}\|f\|_{\omega_\alpha} + \|u_h\|_{\omega_\alpha}),$$

for some constant  $C > 0$  only depending on the shape-regularity constant  $\kappa_{\mathcal{T}_h}$ , the polynomial degree  $k'$ , and the advection parameter  $\mathbf{b}$ .

*Proof.* Since  $\psi_\alpha s_h^\alpha \in H_0^1(\omega_\alpha)$  for all  $\alpha \in \mathcal{V}_h^{\text{int}}$ ,  $\psi_\alpha s_h^\alpha \in H_+^1(\omega_\alpha)$  for  $\alpha \in \mathcal{V}_h^{\partial-\Omega}$ , and  $\psi_\alpha s_h^\alpha \in H_-^1(\omega_\alpha)$  for  $\alpha \in \mathcal{V}_h^{\partial+\Omega}$ ,  $\|\mathbf{b} \cdot \nabla(\psi_\alpha \cdot)\|_{\omega_\alpha}$  is a norm on  $X_h^\alpha$ . Noting that  $\mathbf{b}$  is constant and  $\mathbf{b} \cdot \nabla(\psi_\alpha s_h^\alpha) \in \mathcal{P}^{k'}(\mathcal{T}_\alpha) = Y_h^\alpha$ , one can write the inf-sup condition of the bilinear form associated with the left-hand side of (5.1) as

$$\sup_{v_h \in Y_h^\alpha \setminus \{0\}} \frac{(\mathbf{b} \cdot \nabla(\psi_\alpha s_h^\alpha), v_h)_{\omega_\alpha}}{\|v_h\|_{\omega_\alpha}} = \|\mathbf{b} \cdot \nabla(\psi_\alpha s_h^\alpha)\|_{\omega_\alpha},$$

with unit inf-sup constant. Following Remark 1, this injectivity implies the bijectivity of the operator.

To derive a bound on  $s_h^\alpha$ , we observe that

$$\|\mathbf{b} \cdot \nabla(\psi_\alpha s_h^\alpha)\|_{\omega_\alpha} \leq \|f\psi_\alpha + (\mathbf{b} \cdot \nabla\psi_\alpha) u_h\|_{\omega_\alpha}. \quad (5.4)$$

If one wants to check the stability in the  $L_2$ -norm, one can start with the following norm equivalence on  $Y_h^\alpha$ :

$$\|\psi_\alpha s_h^\alpha\|_{\omega_\alpha} \leq \|s_h^\alpha\|_{\omega_\alpha} \leq C(k')\|\psi_\alpha s_h^\alpha\|_{\omega_\alpha},$$

using similar arguments as in [33, Lemma 3.42]. Consequently, one has

$$\|\mathbf{b} \cdot \nabla(\psi_\alpha s_h^\alpha)\|_{\omega_\alpha} \geq \frac{1}{C(k')C_{\text{PF},\omega_\alpha} h_{\omega_\alpha}} |\mathbf{b}| \|s_h^\alpha\|_{\omega_\alpha}, \quad (5.5)$$

employing the Poincaré inequality (2.3b). Using (5.5) and (5.4), we infer that

$$\|s_h^\alpha\|_{\omega_\alpha} \leq \frac{C(k')C_{\text{PF},\omega_\alpha} h_{\omega_\alpha}}{|\mathbf{b}|} \|f\psi_\alpha + (\mathbf{b} \cdot \nabla\psi_\alpha) u_h\|_{\omega_\alpha} \leq C(k')C_{\text{PF},\omega_\alpha} \left[ \frac{h_{\omega_\alpha}}{|\mathbf{b}|} \|f\|_{\omega_\alpha} + C(\kappa_{\mathcal{T}_h}) \|u_h\|_{\omega_\alpha} \right].$$

□

The following lemma presents the main properties of the reconstruction  $s_h$  from Definition 5.1:

**Lemma 5.3** (properties of the reconstruction). *Definition 5.1 yields  $s_h$  such that*

$$s_h \in \mathcal{P}^{k'+1}(\mathcal{T}_h) \cap H_-^1(\Omega), \quad (5.6)$$

*i.e., it lies in a natural finite-dimensional functional space corresponding to the weak formulation (2.4). Moreover, the following orthogonality is satisfied*

$$(f - \mathbf{b} \cdot \nabla s_h, v_h)_K = 0 \quad \forall v_h \in \mathcal{P}^{k'}(K), \quad \forall K \in \mathcal{T}_h. \quad (5.7)$$

*Proof.* For (5.6) is clear that  $s_h \in \mathcal{P}^{k'+1}(\mathcal{T}_h) \cap H_-^1(\Omega)$ , and we only need to show that  $s_h$  satisfies the boundary condition requirement of the space  $H_-^1(\Omega)$ , i.e.,  $s_h|_{\partial-\Omega} = 0$ . We check this by showing that  $s_h^\alpha|_{\partial\omega_\alpha \cap \partial-\Omega} = 0$  for  $\alpha \in \mathcal{V}_h^{\partial-\Omega}$ . We see from (5.1) and Assumption 4.2 that

$$(\mathbf{b} \cdot \nabla(\psi_\alpha s_h^\alpha), 1)_{\omega_\alpha} = (f\psi_\alpha + (\mathbf{b} \cdot \nabla\psi_\alpha) u_h, 1)_{\omega_\alpha} = \langle \mathcal{R}(u_h), \psi_\alpha \rangle = 0,$$

so that the requested equality follows from integration-by-parts similarly to (5.3),

$$(\mathbf{b} \cdot \nabla(\psi_{\mathbf{a}} s_h^{\mathbf{a}}), 1)_{\omega_{\mathbf{a}}} = \mathbf{b} \cdot \mathbf{n} s_h^{\mathbf{a}}|_{\partial_- \Omega},$$

and since  $\mathbf{b} \cdot \mathbf{n} \neq 0$  on  $\partial_- \Omega$  by definition.

To prove (5.7), first note that  $\sum_{\mathbf{a} \in \mathcal{V}_K} \psi_{\mathbf{a}}|_K = 1$  and  $\sum_{\mathbf{a} \in \mathcal{V}_K} (\mathbf{b} \cdot \nabla \psi_{\mathbf{a}}) u_h|_K = 0$ . Thus, since  $Y_h^{\mathbf{a}}|_K = \mathcal{P}^{k'}(K)$ , extending the function  $v_h \in \mathcal{P}^{k'}(K)$  by zero outside  $K$ , and using respectively definitions (5.2) of  $s_h$  and (5.1) of  $s_h^{\mathbf{a}}$ , one has

$$\begin{aligned} (f - \mathbf{b} \cdot \nabla s_h, v_h)_K &= \left( \sum_{\mathbf{a} \in \mathcal{V}_K} \{ \psi_{\mathbf{a}} f + (\mathbf{b} \cdot \nabla \psi_{\mathbf{a}}) u_h - \mathbf{b} \cdot \nabla(\psi_{\mathbf{a}} s_h^{\mathbf{a}}) \}, v_h \right)_K \\ &= \sum_{\mathbf{a} \in \mathcal{V}_K} (\psi_{\mathbf{a}} f + (\mathbf{b} \cdot \nabla \psi_{\mathbf{a}}) u_h - \mathbf{b} \cdot \nabla(\psi_{\mathbf{a}} s_h^{\mathbf{a}}), v_h)_{\omega_{\mathbf{a}}} = 0. \end{aligned}$$

□

**Remark 5.4** (local conservation). *As a special case of (5.7), since  $1|_K \in \mathcal{P}^{k'}(K)$  for all  $k' \geq 0$ , one has the conservation property*

$$(f - \mathbf{b} \cdot \nabla s_h, 1)_K = 0 \quad \forall K \in \mathcal{T}_h. \quad (5.8)$$

## 6 Guaranteed a posteriori estimate

The guaranteed upper bound on the error can be presented as follows:

**Theorem 6.1** (guaranteed a posteriori error estimate). *Let  $u \in L_2(\Omega)$  be the solution of (2.5) and let  $u_h \in L_2(\Omega)$  be arbitrary subject to the  $\psi_{\mathbf{a}}$ -orthogonality in Assumption 4.2. Furthermore, consider  $s_h$  to be the reconstruction from Definition 5.1 with  $k' \geq 0$ . Then*

$$\|u - u_h\| \leq \eta := \left\{ \sum_{K \in \mathcal{T}_h} (\eta_{\text{NC}, K} + \eta_{\text{Osc}, K})^2 \right\}^{1/2},$$

where

$$\eta_{\text{NC}, K} := \|u_h - s_h\|_K$$

and the data oscillation estimator is given as

$$\eta_{\text{Osc}, K} := \frac{h_K}{\pi |\mathbf{b}|} \|(I - \Pi_{\mathcal{P}^{k'}(\mathcal{T}_h)})f\|_K, \quad (6.1)$$

with  $\Pi_{\mathcal{P}^{k'}(\mathcal{T}_h)}$  the  $L_2(\Omega)$ -orthogonal projection onto  $\mathcal{P}^{k'}(\mathcal{T}_h)$ .

*Proof.* Since  $s_h \in H_+^1(\Omega)$  by Lemma 5.3, for any  $v \in H_+^1(\Omega)$  the integration-by-parts formula (2.2) implies that

$$(s_h, \mathbf{b} \cdot \nabla v) + (\mathbf{b} \cdot \nabla s_h, v) = (s_h, v_h \mathbf{b} \cdot \mathbf{n}) = 0. \quad (6.2)$$

By using the error-residual identity of Theorem 2.1, definitions (2.6)–(2.7), and the above equality, one can write

$$\|u - u_h\|_\Omega = \|\mathcal{R}(u_h)\|_{\mathbf{b}; H_+^1(\Omega)'} = \sup_{v \in H_+^1(\Omega) \setminus \{0\}} \frac{(f - \mathbf{b} \cdot \nabla s_h, v) + (u_h - s_h, \mathbf{b} \cdot \nabla v)}{\|\mathbf{b} \cdot \nabla v\|}.$$

Owing to (5.8), denoting by  $\bar{v}_K$  the mean value of  $v$  over the element  $K$ , we infer that

$$\begin{aligned} \|u - u_h\| &\stackrel{(5.8)}{=} \sup_{v \in H_+^1(\Omega) \setminus \{0\}} \frac{\sum_{K \in \mathcal{T}_h} \left[ (u_h - s_h, \mathbf{b} \cdot \nabla v)_K + (f - \mathbf{b} \cdot \nabla s_h, v - \bar{v}_K)_K \right]}{\|\mathbf{b} \cdot \nabla v\|} \\ &\stackrel{(2.3a)}{\leq} \sup_{v \in H_+^1(\Omega) \setminus \{0\}} \frac{\sum_{K \in \mathcal{T}_h} \left[ \|u_h - s_h\|_K \|\mathbf{b} \cdot \nabla v\|_K + \frac{h_K}{\pi |\mathbf{b}|} \|f - \mathbf{b} \cdot \nabla s_h\|_K \|\mathbf{b} \cdot \nabla v\|_K \right]}{\|\mathbf{b} \cdot \nabla v\|} \\ &\leq \left\{ \sum_{K \in \mathcal{T}_h} \left[ \|u_h - s_h\|_K + \frac{h_K}{\pi |\mathbf{b}|} \|f - \mathbf{b} \cdot \nabla s_h\|_K \right]^2 \right\}^{1/2}. \end{aligned}$$

Noting that  $\mathbf{b} \cdot \nabla s_h \in \mathcal{P}^{k'}(\mathcal{T}_h)$ , it follows from (5.7) that  $\mathbf{b} \cdot \nabla s_h = \Pi_{\mathcal{P}^{k'}(\mathcal{T}_h)} f$  so that

$$\|f - \mathbf{b} \cdot \nabla s_h\|_K = \|(I - \Pi_{\mathcal{P}^{k'}(\mathcal{T}_h)})f\|_K,$$

which completes the proof.  $\square$

**Remark 6.2** (data oscillation). *We call the estimator (6.1) “data oscillation” for the following reason: if  $u_h$  is piecewise polynomial of degree  $k \geq 0$ , the error  $\|u - u_h\|$  may converge as  $\mathcal{O}(h^{k+1})$ . By choosing  $k' \geq k$  one obtains, for sufficiently piecewise smooth data  $f$ , the higher convergence order  $\mathcal{O}(h^{k'+2})$  for these terms.*

## 7 Efficiency and robustness

In this section, we show that the error estimate introduced in Theorem 6.1 also gives, up to a constant and up to the data oscillation, a lower bound on the error  $\|u - u_h\|$ . Furthermore, the involved constants are independent of the polynomial degree  $k$  and the velocity  $\mathbf{b}$ . Actually, a local efficiency result also holds true, and we start with it.

### 7.1 Local efficiency and robustness with respect to advection and polynomial degree

Our main theorem on local efficiency and robustness is:

**Theorem 7.1** (local efficiency and robustness). *Let  $u \in L_2(\Omega)$  be the weak solution of (2.5) and let  $u_h \in \mathcal{P}^k(\mathcal{T}_h)$ ,  $k \geq 0$ , be its approximation. Consider  $s_h$  as obtained by Definition 5.1 with  $k' \geq k$  and  $\eta_{\text{NC},K}$  as defined in Theorem 6.1. Then, for all  $K \in \mathcal{T}_h$ , the following holds true*

$$\eta_{\text{NC},K} \leq C_{\text{cont,PF}} \sum_{\mathbf{a} \in \mathcal{V}_K} \|u - u_h\|_{\omega_{\mathbf{a}}} + \sum_{\mathbf{a} \in \mathcal{V}_K} \frac{h_{\omega_{\mathbf{a}}}}{\pi |\mathbf{b}|} \|(I - \Pi_{\mathcal{P}^{k'}(\mathcal{T}_{\mathbf{a}})})(f \psi_{\mathbf{a}})\|_{\omega_{\mathbf{a}}}.$$

*Proof.* Fix an element  $K \in \mathcal{T}_h$ . Noting that  $\sum_{\mathbf{a} \in \mathcal{V}_K} \psi_{\mathbf{a}}|_K = 1$  and using definition (5.2), one has

$$\|u_h - s_h\|_K = \left\| \sum_{\mathbf{a} \in \mathcal{V}_K} \psi_{\mathbf{a}}(u_h - s_h^{\mathbf{a}}) \right\|_K \leq \sum_{\mathbf{a} \in \mathcal{V}_K} \|\psi_{\mathbf{a}}(u_h - s_h^{\mathbf{a}})\|_{\omega_{\mathbf{a}}}. \quad (7.1)$$

Recalling (4.3), we easily see that for any vertex  $\mathbf{a} \in \mathcal{V}_h$ , there is a unique  $v^{\mathbf{a}} \in H_*^1(\omega_{\mathbf{a}})$  such that

$$\mathbf{b} \cdot \nabla v^{\mathbf{a}} = \psi_{\mathbf{a}}(u_h - s_h^{\mathbf{a}})$$

in  $\omega_{\mathbf{a}}$ , and  $v^{\mathbf{a}}$  is nonzero unless  $\psi_{\mathbf{a}}u_h = \psi_{\mathbf{a}}s_h^{\mathbf{a}}$ , in which case  $\|\psi_{\mathbf{a}}(u_h - s_h^{\mathbf{a}})\|_{\omega_{\mathbf{a}}} = 0$ . Moreover, first,  $(\psi_{\mathbf{a}}s_h^{\mathbf{a}})(\mathbf{a}) = s_h(\mathbf{a}) = 0$  when  $\mathbf{a} \in \mathcal{V}_h^{\partial-\Omega}$ , using (5.6), and, second,  $v^{\mathbf{a}}(\mathbf{a}) = 0$  when  $\mathbf{a} \in \mathcal{V}_h^{\partial+\Omega}$ , using (4.3). Thus, similarly to (6.2), for any  $\mathbf{a} \in \mathcal{V}_h$ , we have

$$(\psi_{\mathbf{a}}s_h^{\mathbf{a}}, \mathbf{b} \cdot \nabla v^{\mathbf{a}})_{\omega_{\mathbf{a}}} + (\mathbf{b} \cdot \nabla(\psi_{\mathbf{a}}s_h^{\mathbf{a}}), v^{\mathbf{a}})_{\omega_{\mathbf{a}}} = 0.$$

From the two above identities, we infer that

$$\begin{aligned} \|\psi_{\mathbf{a}}(u_h - s_h^{\mathbf{a}})\|_{\omega_{\mathbf{a}}} &= \frac{(\psi_{\mathbf{a}}(u_h - s_h^{\mathbf{a}}), \mathbf{b} \cdot \nabla v^{\mathbf{a}})_{\omega_{\mathbf{a}}}}{\|\mathbf{b} \cdot \nabla v^{\mathbf{a}}\|_{\omega_{\mathbf{a}}}} \\ &= \frac{(\psi_{\mathbf{a}}u_h, \mathbf{b} \cdot \nabla v^{\mathbf{a}})_{\omega_{\mathbf{a}}} + (f\psi_{\mathbf{a}} + \mathbf{b} \cdot \nabla \psi_{\mathbf{a}}u_h, v^{\mathbf{a}})_{\omega_{\mathbf{a}}}}{\|\mathbf{b} \cdot \nabla v^{\mathbf{a}}\|_{\omega_{\mathbf{a}}}} \\ &\quad + \frac{(\mathbf{b} \cdot \nabla(\psi_{\mathbf{a}}s_h^{\mathbf{a}}), v^{\mathbf{a}})_{\omega_{\mathbf{a}}} - (f\psi_{\mathbf{a}} + \mathbf{b} \cdot \nabla \psi_{\mathbf{a}}u_h, v^{\mathbf{a}})_{\omega_{\mathbf{a}}}}{\|\mathbf{b} \cdot \nabla v^{\mathbf{a}}\|_{\omega_{\mathbf{a}}}} \\ &=: I + II. \end{aligned} \quad (7.2)$$

For the term  $I$ , remark first that from (4.4) and from the definition of  $H_{\#}^1(\omega_{\mathbf{a}})$  in (4.2), we have

$$\psi_{\mathbf{a}}v^{\mathbf{a}} \in H_{\#}^1(\omega_{\mathbf{a}}) \subseteq H_+^1(\omega_{\mathbf{a}}) \subseteq H_+^1(\Omega).$$

Second, recalling the residual definition (2.6) and the ultra-weak formulation (2.5), we have

$$(f\psi_{\mathbf{a}} + \mathbf{b} \cdot \nabla \psi_{\mathbf{a}}u_h, v^{\mathbf{a}})_{\omega_{\mathbf{a}}} + (u_h\psi_{\mathbf{a}}, \mathbf{b} \cdot \nabla v^{\mathbf{a}})_{\omega_{\mathbf{a}}} = \langle \mathcal{R}(u_h), \psi_{\mathbf{a}}v^{\mathbf{a}} \rangle = -(u - u_h, \mathbf{b} \cdot \nabla(\psi_{\mathbf{a}}v^{\mathbf{a}})).$$

Consequently, employing the Cauchy–Schwarz inequality and Lemma 4.1, we infer that

$$I = \frac{-(u - u_h, \mathbf{b} \cdot \nabla(\psi_{\mathbf{a}}v^{\mathbf{a}}))}{\|\mathbf{b} \cdot \nabla(\psi_{\mathbf{a}}v^{\mathbf{a}})\|_{\omega_{\mathbf{a}}}} \frac{\|\mathbf{b} \cdot \nabla(\psi_{\mathbf{a}}v^{\mathbf{a}})\|_{\omega_{\mathbf{a}}}}{\|\mathbf{b} \cdot \nabla v^{\mathbf{a}}\|_{\omega_{\mathbf{a}}}} \leq C_{\text{cont,PF}} \|u - u_h\|_{\omega_{\mathbf{a}}}. \quad (7.3)$$

To bound the term  $II$ , we use the fact that  $(\mathbf{b} \cdot \nabla \psi_{\mathbf{a}})u_h \in Y_h^{\mathbf{a}}$  when  $k' \geq k$  and that  $\mathbf{b} \cdot \nabla(\psi_{\mathbf{a}}s_h^{\mathbf{a}}) \in Y_h^{\mathbf{a}}$ , so that (5.1) actually holds pointwise, in the form

$$\mathbf{b} \cdot \nabla(\psi_{\mathbf{a}}s_h^{\mathbf{a}}) = \Pi_{\mathcal{P}^{k'}(\mathcal{T}_{\mathbf{a}})}(f\psi_{\mathbf{a}}) + (\mathbf{b} \cdot \nabla \psi_{\mathbf{a}})u_h.$$

Hence, denoting  $\bar{v}_K^{\mathbf{a}}$  the mean value of  $v^{\mathbf{a}}$  over the element  $K \in \mathcal{T}_{\mathbf{a}}$  and using (2.3a), we obtain

$$\begin{aligned} II &= \frac{(\Pi_{\mathcal{P}^{k'}(\mathcal{T}_{\mathbf{a}})}(f\psi_{\mathbf{a}}) - f\psi_{\mathbf{a}}, v^{\mathbf{a}})_{\omega_{\mathbf{a}}}}{\|\mathbf{b} \cdot \nabla v^{\mathbf{a}}\|_{\omega_{\mathbf{a}}}} = \frac{\sum_{K \in \mathcal{T}_{\mathbf{a}}} (\Pi_{\mathcal{P}^{k'}(\mathcal{T}_{\mathbf{a}})}(f\psi_{\mathbf{a}}) - f\psi_{\mathbf{a}}, v^{\mathbf{a}} - \bar{v}_K^{\mathbf{a}})_K}{\|\mathbf{b} \cdot \nabla v^{\mathbf{a}}\|_{\omega_{\mathbf{a}}}} \\ &\leq \frac{\max_{K \in \mathcal{T}_{\mathbf{a}}} h_K}{\pi |\mathbf{b}|} \|\Pi_{\mathcal{P}^{k'}(\mathcal{T}_{\mathbf{a}})}(f\psi_{\mathbf{a}}) - f\psi_{\mathbf{a}}\|_{\omega_{\mathbf{a}}}. \end{aligned}$$

The assertion follows by combining the bounds on  $I$  and  $II$  with (7.1).  $\square$

## 7.2 Global efficiency and maximal overestimation

In this section, we show a result on maximal global overestimation, leaving out the data oscillation term for simplicity.

**Lemma 7.2** (global efficiency and maximal overestimation). *Let the assumptions of Theorem 7.1 be valid and assume in addition that  $\psi_{\mathbf{a}} f \in Y_h^{\mathbf{a}}$  for all  $\mathbf{a} \in \mathcal{V}_h$ . Then*

$$\|u_h - s_h\| \leq 2C_{\text{cont,PF}} \|u - u_h\|.$$

*Proof.* Proceeding as in the proof of Theorem 7.1, one has

$$\begin{aligned} \|u_h - s_h\|^2 &= \sum_{K \in \mathcal{T}_h} \left\| \sum_{\mathbf{a} \in \mathcal{V}_K} \psi_{\mathbf{a}} (u_h - s_h^{\mathbf{a}}) \right\|_K^2 \leq 2 \sum_{K \in \mathcal{T}_h} \sum_{\mathbf{a} \in \mathcal{V}_K} \|\psi_{\mathbf{a}} (u_h - s_h^{\mathbf{a}})\|_K^2 \\ &= 2 \sum_{\mathbf{a} \in \mathcal{V}_h} \|\psi_{\mathbf{a}} (u_h - s_h^{\mathbf{a}})\|_{\omega_{\mathbf{a}}}^2 \stackrel{(7.2), (7.3)}{\leq} 2C_{\text{cont,PF}}^2 \sum_{\mathbf{a} \in \mathcal{V}_h} \|u - u_h\|_{\omega_{\mathbf{a}}}^2. \end{aligned}$$

Another estimate for the overlapping of the patches yields  $\sum_{\mathbf{a} \in \mathcal{V}_h} \|u - u_h\|_{\omega_{\mathbf{a}}}^2 \leq 2\|u - u_h\|^2$  and leads to the assertion.  $\square$

## 8 Numerical experiments

We provide in this section a numerical illustration of our results in one space dimension. In the first set of examples in Section 8.1, we consider a polynomial right-hand side function  $f$  and study the robustness of our estimators with respect to the velocity field  $\mathbf{b}$ . Then, in Section 8.2, we consider a more general case to investigate the effect of the increase of the polynomial degree on the quality of the estimators. Henceforth, we consider  $\Omega = (0, 1)$  with the mesh  $\mathcal{T}_h = \{K_i\}_1^n$  with  $K_i = [x_{i-1}, x_i]$ . In the experiments, the numerical solution  $u_h \in \mathcal{P}^k(\mathcal{T}_h)$  will be computed by two methods:

- the PG2 method (3.2) with the finite-dimensional spaces as in Example 3.2,  $k \geq 0$ ,
- the dG method (3.3) with the finite-dimensional spaces as in Example 3.3,  $k \geq 1$ .

The *effectivity index* is defined as  $I_{\text{eff}} := \frac{\eta}{\|u - u_h\|}$ , i.e., as the ratio of the estimated and the actual error from Theorem 6.1.

### 8.1 Robustness with respect to the velocity

Here we consider the advection problem (1.1) with the piecewise quadratic right-hand side defined as

$$f(x) = x^2 + x + \sin(2\pi x_{i-1}), \quad \text{on } K_i, 1 \leq i \leq n,$$

whose exact solution can be easily computed by integration of the right-hand side. The numerical solutions  $u_h$  are obtained by both PG2 and dG methods with  $k = 1, 2$ .

If one sets  $k' = 2$  in Definition 5.1, the oscillation estimators  $\eta_{\text{Osc}, K}$  from (6.1) disappear. In this case, actually, since  $f \in \mathcal{P}^2(\mathcal{T}_h)$ , one has  $s_h \in \mathcal{P}^3(\mathcal{T}_h) \cap H^1_-(\Omega)$ , see (5.6). Moreover, owing to (5.7),  $\mathbf{b} \cdot \nabla s_h = f$  pointwise. Hence,  $s_h$  in this setting coincides with the exact solution  $u$  and  $I_{\text{eff}} = 1$  (up to the machine precision), which is numerically confirmed in Tables 1 and 2.

To assess the behavior in the case where the reconstruction  $s_h$  does not coincide with the exact solution, we also test the choice  $k' = 1$  in Definition 5.1 together with  $k = 1$ . We observe in Tables 1 and 2 that there is still no dependency of the efficiency of our estimates on the magnitude of the velocity  $\mathbf{b}$ , in confirmation of the theory. Actually, one remarks that solely scaling  $\mathbf{b}$  in (1.1) by a factor implies the same scaling of all the exact solution  $u$ , the numerical approximations  $u_h$ , the error  $\|u - u_h\|$ , the reconstruction  $s_h$ , and of all the estimators in Theorem 6.1 by the inverse of this factor, so that the effectivity indices actually remain intact here on each given mesh. Moreover, we numerically observe asymptotic exactness with mesh refinement, for both schemes tested.

Table 1: Effectivity indices  $I_{\text{eff}}$  for different values of the velocity  $\mathbf{b}$  and  $u_h$  obtained by the PG2 method (3.2)

$\mathbf{k} = \mathbf{k}' = 2$		$\mathbf{b}$				
# Elements	DOF( $u_h$ )	$10^{-4}$	$10^{-2}$	$10^0$	$10^2$	$10^4$
4	12	1.0	1.0	1.0	1.0	1.0
16	48	1.0	1.0	1.0	1.0	1.0
64	192	1.0	1.0	1.0	1.0	1.0
256	768	1.0	1.0	1.0	1.0	1.0
$\mathbf{k} = \mathbf{k}' = 1$		$\mathbf{b}$				
# Elements	DOF( $u_h$ )	$10^{-4}$	$10^{-2}$	$10^0$	$10^2$	$10^4$
4	8	1.234	1.234	1.234	1.234	1.234
16	32	1.058	1.058	1.058	1.058	1.058
64	128	1.014	1.014	1.014	1.014	1.014
256	512	1.004	1.004	1.004	1.004	1.004

Table 2: Effectivity indices  $I_{\text{eff}}$  for different values of the velocity  $\mathbf{b}$  and  $u_h$  obtained by the dG method (3.3)

$\mathbf{k} = \mathbf{k}' = 2$		$\mathbf{b}$				
# Elements	DOF( $u_h$ )	$10^{-4}$	$10^{-2}$	$10^0$	$10^2$	$10^4$
4	12	1.0	1.0	1.0	1.0	1.0
16	48	1.0	1.0	1.0	1.0	1.0
64	192	1.0	1.0	1.0	1.0	1.0
256	768	1.0	1.0	1.0	1.0	1.0
$\mathbf{k} = \mathbf{k}' = 1$		$\mathbf{b}$				
# Elements	DOF( $u_h$ )	$10^{-4}$	$10^{-2}$	$10^0$	$10^2$	$10^4$
4	8	1.126	1.126	1.126	1.126	1.126
16	32	1.032	1.032	1.032	1.032	1.032
64	128	1.008	1.008	1.008	1.008	1.008
256	512	1.002	1.002	1.002	1.002	1.002

## 8.2 Robustness with respect to the polynomial degree

We now consider the advection problem (1.1) with a non-polynomial right-hand side  $f(x) = \tan^{-1}(x)$  and  $\mathbf{b} = 1$ , for different polynomial degrees  $0 \leq k \leq 4$ . The results are presented in Table 3 for the PG2 method and in Table 4 for the dG method. We always set  $k' = k$ . We use the notation  $\eta_{\text{NC}} := (\sum_{K \in \mathcal{T}_h} \eta_{\text{NC},K}^2)^{1/2}$  and  $\eta_{\text{Osc}} := (\sum_{K \in \mathcal{T}_h} \eta_{\text{Osc},K}^2)^{1/2}$ . The mesh is refined



uniformly until the error estimator  $\eta \leq 10^{-14}$ ; we encountered some irregularities in  $I_{\text{eff}}$  beyond this point due to machine precision. We observe optimal convergence order of the estimators and the independence of  $I_{\text{eff}}$  from the polynomial degree, in accordance with the theory.

Table 3: Convergence of the error  $\|u - u_h\|$ , the error estimators  $\eta$ ,  $\eta_{\text{NC}}$ , and  $\eta_{\text{Osc}}$ , and the effectivity indices  $I_{\text{eff}}$  for the PG2 method (3.2) with different polynomial degrees  $k$

<b><math>k = 0, k' = 0</math></b>						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	4	3.574e-2	1.446e-2	3.562e-2	4.809e-2	1.35
16	16	8.936e-3	9.040e-4	8.934e-3	9.656e-3	1.08
64	64	2.234e-3	5.650e-5	2.234e-3	2.278e-4	1.02
256	256	5.585e-4	3.531e-6	5.585e-4	5.612e-4	1.01
1024	1024	1.396e-4	2.207e-7	1.396e-4	1.398e-4	1.00
<b><math>k = 1, k' = 1</math></b>						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	8	1.867e-3	3.074e-4	1.868e-3	2.147e-3	1.15
16	32	1.167e-4	4.811e-6	1.167e-4	1.210e-4	1.04
64	128	7.294e-6	7.518e-8	7.294e-6	7.361e-6	1.01
256	512	4.559e-7	1.175e-9	4.559e-7	4.569e-7	1.00
1024	2048	2.849e-8	1.836e-11	2.849e-8	2.851e-8	1.00
<b><math>k = 2, k' = 2</math></b>						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	12	2.598e-5	1.246e-5	2.600e-5	3.467e-5	1.33
16	48	4.066e-7	4.897e-8	4.066e-7	4.356e-7	1.07
64	192	6.354e-9	1.914e-10	6.354e-9	6.462e-9	1.02
256	768	9.928e-11	7.475e-13	9.928e-11	9.97e-11	1.00
1024	3072	1.551e-12	2.920e-15	1.552e-12	1.553e-12	1.00
<b><math>k = 3, k' = 3</math></b>						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	16	7.852e-7	5.666e-7	7.859e-7	1.271e-6	1.62
16	64	3.085e-9	5.579e-10	3.085e-9	3.503e-9	1.14
64	256	1.205e-11	5.451e-13	1.205e-11	1.245e-11	1.03
256	1024	4.718e-14	5.324e-16	4.730e-14	4.756e-14	1.01
<b><math>k = 4, k' = 4</math></b>						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	20	2.847e-8	2.666e-8	2.851e-8	5.133e-8	1.80
8	40	8.957e-10	4.204e-10	8.959e-10	1.217e-9	1.36
16	80	2.804e-11	6.582e-12	2.804e-11	3.282e-11	1.17
32	160	8.765e-13	1.029e-13	8.765e-13	9.482e-13	1.08
64	320	2.742e-14	1.608e-15	2.753e-14	2.852e-14	1.04

Table 4: Convergence of the error  $\|u - u_h\|$ , the error estimators  $\eta$ ,  $\eta_{\text{NC}}$ , and  $\eta_{\text{Osc}}$ , and the effectivity indices  $I_{\text{eff}}$  for the dG method (3.3) with different polynomial degrees  $k$

$k = 1, k' = 1$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	8	3.048e-3	3.074e-4	3.021e-3	3.327e-3	1.10
16	32	1.906e-4	4.811e-6	1.901e-4	1.949e-4	1.03
64	128	1.191e-5	7.518e-8	1.190e-5	1.198e-5	1.01
256	512	7.445e-7	1.175e-9	7.444e-7	7.455e-7	1.00
1024	2048	4.653e-8	1.836e-11	4.653e-8	4.655e-8	1.00
$k = 2, k' = 2$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	12	4.210e-5	1.246e-5	4.450e-5	4.843e-5	1.19
16	48	6.299e-7	4.897e-8	6.307e-7	6.582e-7	1.04
64	192	9.844e-9	1.914e-10	9.847e-9	9.951e-9	1.01
256	768	1.538e-10	7.475e-13	1.538e-10	1.542e-10	1.00
1024	3072	2.403e-12	2.921e-15	2.403e-12	2.405e-12	1.00
$k = 3, k' = 3$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	16	1.186e-6	5.666e-7	1.169e-6	1.658e-6	1.42
16	64	4.664e-9	5.579e-10	4.647e-9	5.075e-9	1.09
64	256	1.822e-11	5.451e-13	1.821e-11	1.861e-11	1.02
256	1024	7.172e-14	5.324e-16	7.181e-14	7.210e-14	1.00
$k = 4, k' = 4$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{Osc}}$	$\ u - u_h\ $	$\eta$	$I_{\text{eff}}$
4	20	4.240e-8	2.666e-8	4.252e-8	6.453e-8	1.52
8	40	1.335e-9	4.204e-10	1.336e-9	1.645e-9	1.23
16	80	4.179e-11	6.582e-12	4.180e-11	4.647e-11	1.11
32	160	1.307e-12	1.029e-13	1.307e-12	1.377e-12	1.05
64	320	4.083e-14	1.608e-15	4.094e-14	4.192e-14	1.02

## 9 Extension to multiple space dimensions

In this section, we investigate a possible extension of the ideas presented so far to the multi-dimensional case. We consider the advection equation (1.1) on a simply-connected Lipschitz polytope  $\Omega \subset \mathbb{R}^d$  for  $d \geq 2$ . The velocity field  $\mathbf{b}(x) \in \mathcal{C}^1(\bar{\Omega}; \mathbb{R}^d)$  is considered to be divergence-free. We also assume that  $\mathbf{b}$  is  $\Omega$ -filling, i.e., its trajectories starting from the inflow boundary  $\partial_- \Omega$  fill  $\bar{\Omega}$  almost everywhere in a finite time. A sufficient condition for the validity of this property is given by [3] (see Lemma 9.1 below). One can find necessary and sufficient conditions in [13, Lemma 2.3], see also [2, 11, 8, 9].

### 9.1 Spaces

We start by introducing proper generalizations of (2.1). Let us define the operator related to (1.1) and its formal adjoint as

$$\mathcal{L}: v \mapsto \mathbf{b} \cdot \nabla v, \quad \mathcal{L}^*: v \mapsto -\nabla \cdot (\mathbf{b}v) = -\mathbf{b} \cdot \nabla v,$$

together with the following graph spaces

$$H(\mathcal{L}, \Omega) := \{v \in L_2(\Omega), \mathcal{L}v \in L_2(\Omega)\}, \quad H(\mathcal{L}^*, \Omega) := \{v \in L_2(\Omega), \mathcal{L}^*v \in L_2(\Omega)\}.$$

Then  $\mathcal{L} : H(\mathcal{L}, \Omega) \rightarrow L_2(\Omega)$  and  $\mathcal{L}^* : H(\mathcal{L}^*, \Omega) \rightarrow L_2(\Omega)$ , and  $H(\mathcal{L}, \Omega) = H(\mathcal{L}^*, \Omega)$ . Moreover, one can define the following subspaces of the graph spaces with incorporated boundary conditions:

$$\begin{aligned} H_0(\mathcal{L}, \Omega) &:= \{v \in H(\mathcal{L}, \Omega), v = 0 \text{ on } \partial_- \Omega\}, \\ H_0(\mathcal{L}^*, \Omega) &:= \{v \in H(\mathcal{L}^*, \Omega), v = 0 \text{ on } \partial_+ \Omega\}. \end{aligned}$$

These definitions are consistent extensions from  $d = 1$  in that the spaces  $H(\mathcal{L}, \Omega)$ ,  $H(\mathcal{L}^*, \Omega)$ ,  $H_0(\mathcal{L}, \Omega)$ ,  $H_0(\mathcal{L}^*, \Omega)$  become respectively  $H^1(\Omega)$ ,  $H^1(\Omega)$ ,  $H_-^1(\Omega)$ , and  $H_+^1(\Omega)$ . One might confer with [29, p. 131] and [21, Theorems 2.1 and 2.2] for the justification of the trace operator which is discussed as an operator from  $H(\mathcal{L}, \Omega)$  to  $H^{-\frac{1}{2}}(\partial_- \Omega)$  (or from  $H(\mathcal{L}^*, \Omega)$  to  $H^{-\frac{1}{2}}(\partial_+ \Omega)$ , respectively). The extension to  $L_2(|\mathbf{b} \cdot \mathbf{n}|; \partial_- \Omega)$  is possible under slightly more restrictive conditions, see [29, page 133], [14, Lemma 3.1] and more recently [11, Proposition 2.3]. Moreover the following integration-by-parts formula holds true:

$$(v, \mathbf{b} \cdot \nabla w) + (\mathbf{b} \cdot \nabla v, w) = (\mathbf{b} \cdot \mathbf{n} v, w) \quad \forall v \in H(\mathcal{L}, \Omega), \quad \forall w \in H^1(\Omega). \quad (9.1)$$

The result (9.1) can be extended to  $w \in H(\mathcal{L}^*, \Omega)$ .

## 9.2 Streamline Poincaré inequality

The following sufficient condition for the field  $\mathbf{b}$  to be  $\Omega$ -filling is given in [3]:

**Lemma 9.1** ( $\Omega$ -filling sufficient condition). *Let  $\mathbf{b} \in C^1(\bar{\Omega}; \mathbb{R}^d)$  and assume that there is a fixed unit vector  $\mathbf{k} \in \mathbb{R}^d$  and a real number  $\alpha > 0$  such that*

$$\forall x \in \bar{\Omega}, \quad \mathbf{b} \cdot \mathbf{k} \geq \alpha. \quad (9.2)$$

*Then  $\mathbf{b}$  is  $\Omega$ -filling.*

For  $\Omega$ -filling  $\mathbf{b}$ , we can extend the inequality (2.3b) along the flow of  $\mathbf{b}$ , cf. [3]:

**Lemma 9.2** (streamline Poincaré inequality). *Let the field  $\mathbf{b} \in C^1(\bar{\Omega}; \mathbb{R}^d)$  be divergence-free and  $\Omega$ -filling. Then there exists a streamline Poincaré constant  $C_{\mathbf{P}, \mathbf{b}, \Omega}$  such that*

$$\|v\| \leq C_{\mathbf{P}, \mathbf{b}, \Omega} \|\mathbf{b} \cdot \nabla v\| \quad \forall v \in H_0(\mathcal{L}, \Omega). \quad (9.3)$$

*The constant  $C_{\mathbf{P}, \mathbf{b}, \Omega}$  is bounded by  $C_{\mathbf{P}, \mathbf{b}, \Omega} \leq 2T$ , where  $T$  is the longest time that trajectories of the field  $\mathbf{b}$  spend in the domain  $\Omega$ . In particular,  $T \leq \text{diam}(\Omega)/\alpha$  under the assumption (9.2).*

A similar result can also be obtained for a non divergence-free field, see [1]. In the case where the field  $\mathbf{b}$  is constant, one can easily set  $\mathbf{k}$  as the direction of the flow and  $\alpha = |\mathbf{b}|$ . A crucial consequence of Lemma 9.2 is that one can equip the spaces  $H_0(\mathcal{L}, \Omega)$  and  $H_0(\mathcal{L}^*, \Omega)$  with the norm  $\|\mathbf{b} \cdot \nabla(\cdot)\|$ .

**Remark 9.3** (functions with mean value zero). *While, following from Lemma 9.2, the streamline Poincaré inequality holds true for functions with zero trace on the inflow of an arbitrary domain  $D$ , such a result is not valid for functions with mean value zero as a variant of the Poincaré inequality (2.3a) in multiple spatial dimensions. This leads to significant differences in the analysis of the multi-dimensional case compared to one-dimensional one, and less rigorous results that we are able to present here.*

### 9.3 Error-residual equivalence

We consider the multi-dimensional extension of the ultra-weak formulation (2.5): find  $u \in L_2(\Omega)$  such that

$$-(u, \mathbf{b} \cdot \nabla v) = (f, v), \quad \forall v \in H_0(\mathcal{L}^*, \Omega). \quad (9.4)$$

Define the residual operator  $\mathcal{R}(u_h) \in H_0(\mathcal{L}^*, \Omega)'$  and its dual norm as in (2.6)–(2.7), upon replacing  $H_+^1(\Omega)$  by  $H_0(\mathcal{L}^*, \Omega)$ . One can extend the equivalence of Theorem 2.1 to the multi-dimensional case as follows:

**Theorem 9.4** (error-residual equivalence). *Let the field  $\mathbf{b} \in \mathcal{C}^1(\bar{\Omega}; \mathbb{R}^d)$  be divergence-free and  $\Omega$ -filling. Let  $u \in L_2(\Omega)$  be the ultra-weak solution of (9.4). Then*

$$\|u - u_h\| = \|\mathcal{R}(u_h)\|_{\mathbf{b}; H_0(\mathcal{L}^*, \Omega)'} \quad \forall u_h \in L_2(\Omega).$$

*Proof.* We use the fact that for all  $v \in L_2(\Omega)$ , there exists a unique  $z \in H_0(\mathcal{L}^*, \Omega)$  such that

$$-(\mathbf{b} \cdot \nabla z, w) = (v, w), \quad \forall w \in L_2(\Omega).$$

The rest of the proof goes along the lines of that of Theorem 2.1.  $\square$

### 9.4 Local problems and the error indicator

In this section, we propose a heuristic approach inspired by the rigorous discussions in the one-dimensional case. First, let us consider the following reconstruction, mimicking Definition 5.1. Here  $\mathcal{T}_h$  is a simplicial mesh of  $\Omega$ ,  $\mathcal{T}_\mathbf{a}$  the patch of all simplices which share the given vertex  $\mathbf{a} \in \mathcal{V}_h$ ,  $\omega_\mathbf{a}$  the corresponding open subdomain, and  $\psi_\mathbf{a}$  the associated hat basis function.

**Definition 9.5** (patchwise problems). *Let  $u_h \in L_2(\Omega)$ . For all vertices  $\mathbf{a} \in \mathcal{V}_h$ , let  $s_h^\mathbf{a} \in X_h^\mathbf{a}$  be the solution of the following least-squares problem on the patch subdomain  $\omega_\mathbf{a}$ :*

$$s_h^\mathbf{a} := \arg \min_{v_h \in X_h^\mathbf{a}} \left\{ \|\psi_\mathbf{a}(u_h - v_h)\|_{\omega_\mathbf{a}}^2 + C_{\text{opt}}^2 \|f\psi_\mathbf{a} + (\mathbf{b} \cdot \nabla \psi_\mathbf{a}) u_h - \mathbf{b} \cdot \nabla (\psi_\mathbf{a} v_h)\|_{\omega_\mathbf{a}}^2 \right\}. \quad (9.5)$$

For  $k' \geq 0$ , we take the finite-dimensional space  $X_h^\mathbf{a} := \mathcal{P}^{k'}(\mathcal{T}_\mathbf{a}) \cap H_0(\mathcal{L}, \omega_\mathbf{a})$  when the vertex  $\mathbf{a}$  lies in the closure of the inflow boundary  $\partial_- \Omega$ , and  $X_h^\mathbf{a} := \mathcal{P}^{k'}(\mathcal{T}_\mathbf{a}) \cap H(\mathcal{L}, \omega_\mathbf{a})$  otherwise. Here  $C_{\text{opt}} > 0$  is a constant to be chosen. The global reconstruction  $s_h$  is defined by

$$s_h := \sum_{\mathbf{a} \in \mathcal{V}_h} \psi_\mathbf{a} s_h^\mathbf{a}, \quad (9.6)$$

leading to  $s_h \in \mathcal{P}^{k'+1}(\mathcal{T}_h) \cap H_0(\mathcal{L}, \Omega)$ .

In order to see the rationale behind the above reconstruction, one might note the following upper bound on the error exploiting Theorem 9.4, the integration-by-parts formula (9.1), the Cauchy–Schwarz inequality, and the streamline Poincaré inequality (9.3): for any  $s_h \in H_0(\mathcal{L}, \Omega)$ , we have

$$\begin{aligned} \|u - u_h\| &= \|\mathcal{R}(u_h)\|_{\mathbf{b}; H_0(\mathcal{L}^*, \Omega)'} = \sup_{v \in H_0(\mathcal{L}^*, \Omega) \setminus \{0\}} \frac{(f - \mathbf{b} \cdot \nabla s_h, v) + (u_h - s_h, \mathbf{b} \cdot \nabla v)}{\|\mathbf{b} \cdot \nabla v\|} \\ &\leq \|u_h - s_h\| + C_{\mathbf{P}, \mathbf{b}, \Omega} \|f - \mathbf{b} \cdot \nabla s_h\|. \end{aligned}$$

Using that almost each point in  $\Omega$  belongs to  $(d+1)$  patch subdomains  $\omega_{\mathbf{a}}$  and the partition of unity (4.1), the construction of  $s_h$  via (9.6) gives the following upper bound:

$$\|u - u_h\| \leq \left\{ 2(d+1) \sum_{\mathbf{a} \in \mathcal{V}_h} \left[ \|\psi_{\mathbf{a}}(u_h - s_h^{\mathbf{a}})\|_{\omega_{\mathbf{a}}}^2 + C_{\mathbb{P}, \mathbf{b}, \Omega}^2 \|f\psi_{\mathbf{a}} + (\mathbf{b} \cdot \nabla \psi_{\mathbf{a}}) u_h - \mathbf{b} \cdot \nabla (\psi_{\mathbf{a}} s_h^{\mathbf{a}})\|_{\omega_{\mathbf{a}}}^2 \right] \right\}^{1/2}.$$

In particular, the idea of adding  $0 = \sum_{\mathbf{a} \in \mathcal{V}_h} \mathbf{b} \cdot \nabla \psi_{\mathbf{a}} u_h$  is inspired by the analysis in the one-dimensional case. By comparison to Definition 9.5 one can see that the least-squares problems (9.5) minimize contributions to the upper bound on the error, and a theoretically-motivated choice for  $C_{\text{opt}}$  would be  $C_{\text{opt}} = C_{\mathbb{P}, \mathbf{b}, \Omega}$ . This in particular leads to the guaranteed estimate  $\|u - u_h\| \leq \eta$  with

$$\eta := \left\{ \sum_{K \in \mathcal{T}_h} \|u_h - s_h\|_K^2 \right\}^{1/2} + C_{\mathbb{P}, \mathbf{b}, \Omega} \left\{ \sum_{K \in \mathcal{T}_h} \|f - \mathbf{b} \cdot \nabla s_h\|_K^2 \right\}^{1/2}. \quad (9.7)$$

Numerical experiments, however, show that the second term in (9.7) does not converge with the right order. This apparently comes from the special structure of the minimization term which cannot be approximated up to the projection error, in contrast to the one-dimensional case, where (5.7) holds true. Congruently, the lack of the Poincaré inequality in the streamline form (see Remark 9.3) implies the loss of the scaling by the mesh element diameters  $h_K$  in the second term in (9.7), compare with  $\eta_{\text{Osc}}$  given by (6.1) in one space dimension.

In order to rectify this issue, we heuristically replace  $C_{\mathbb{P}, \mathbf{b}, \Omega}$  (which typically scales as  $2\text{diam}(\Omega)/\alpha$ , see Lemma 9.2) in the estimate  $\eta$  of (9.7) by local terms  $\frac{C' h_K}{\alpha}$ , where one now needs to choose the constant  $C'$ . Overall, we suggest the following modified error indicator

$$\eta_{\text{mod}} := \left\{ \sum_{K \in \mathcal{T}_h} (\eta_{\text{NC}, K}^2 + \eta_{\text{R}, K}^2) \right\}^{1/2}, \quad \eta_{\text{NC}, K} = \|u_h - s_h\|_K, \quad \eta_{\text{R}, K} = \frac{C' h_K}{\alpha} \|f - \mathbf{b} \cdot \nabla s_h\|_K. \quad (9.8)$$

In this setting, one has two free parameters to choose,  $C_{\text{opt}}$  for the local problems in (9.5) and  $C'$  in (9.8). We set below  $C_{\text{opt}} = 2\text{diam}(\Omega)/\alpha$ , as Lemmas 9.1–9.2 suggest, and  $C' = 2$ . Numerically, our results are actually not sensitive to the choice of the parameter  $C_{\text{opt}}$ .

Let us stress that, unlike  $\eta$  in (9.7), the modified indicator  $\eta_{\text{mod}}$  from (9.8) is not a guaranteed upper bound of the error; however, our numerical results presented below show its applicability.

## 9.5 Numerical experiments

In this section, we provide some numerical tests in two space dimensions to illustrate the properties of the error indicator (9.8), with  $s_h$  constructed following Definition 9.5. The implementation is done in the framework of FreeFEM++ [20] and based on the scripts for the reconstruction-based a posteriori estimation by [30].

One might note that the reconstruction  $s_h$  of Definition 9.5, lying in the space  $H_0(\mathcal{L}, \Omega)$ , possibly allows capturing the discontinuity that may appear in the exact solution  $u$  across the streamlines. This is, however, only in reach if the triangulation is aligned with the streamlines. If this is not the case, the reconstruction  $s_h$  actually lies in the smoother space  $H^1(\Omega)$ . Below, we will consider two test cases, one with both the exact solution  $u$  and the reconstruction  $s_h$  lying in  $H^1(\Omega)$ , and one with a discontinuous solution  $u \in H_0(\mathcal{L}, \Omega)$ , with the triangulation aligned with the streamlines, ideally including also the lines of discontinuity of the exact solution.

Let us consider  $\Omega = (0, 1)^2$  and uniformly refined structured triangulations aligned with the slope  $45^\circ$ . We only test here the dG method (3.3), since it is the only method among those considered in Section 3 which is stable and well-defined in multiple space dimensions.

### 9.5.1 Smooth solution

We apply the right-hand side  $f$  such that the solution of (1.1) is

$$u(x, y) = \sin(\pi x) \sin(\pi y), \quad (9.9)$$

for different velocity fields  $\mathbf{b}$ . The results are presented in Tables 5 and 6 for various polynomial degrees  $k$ , with the choice  $k' = k + 1$  in Definition 9.5. Here  $\eta_{\text{NC}} := (\sum_{K \in \mathcal{T}_h} \eta_{\text{NC}, K}^2)^{1/2}$  and  $\eta_{\text{R}} := (\sum_{K \in \mathcal{T}_h} \eta_{\text{R}, K}^2)^{1/2}$ . The error indicator performs well in actually providing the upper bound of the error and simultaneously not overestimating it excessively. Moreover, the efficiency results appear to be robust with respect to both the velocity field  $\mathbf{b}$  and the polynomial degree  $k$ . Compared to Section 8.1, both  $u$  and  $u_h$ , but actually also  $s_h$  constructed following Definition 9.5, turn out to be insensitive to the scaling of  $\mathbf{b}$  by a constant, so that the estimators in (9.8) do not change either. In Figure 1, the distribution of the errors  $\|u - u_h\|_K$  and of the error estimators  $\eta_{\text{mod}, K} := (\eta_{\text{NC}, K}^2 + \eta_{\text{R}, K}^2)^{1/2}$  is presented. These distributions show a very close behavior, which suggests that the presented indicators should be suitable for adaptive mesh/polynomial degree refinement.

Table 5: Convergence of the error  $\|u - u_h\|$ , the error estimators  $\eta_{\text{mod}}$ ,  $\eta_{\text{NC}}$ , and  $\eta_{\text{R}}$ , and the effectivity indices  $I_{\text{eff}}$  for the dG method (3.3); smooth solution (9.9),  $\mathbf{b} = (1, 1)^t$ , and different polynomial degrees  $k$

$k = 1, k' = 2$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{R}}$	$\ u - u_h\ $	$\eta_{\text{mod}}$	$I_{\text{eff}}$
8	24	9.365e-02	2.083e-01	1.097e-01	2.175e-01	1.98
32	96	2.584e-02	4.156e-02	2.963e-02	4.871e-02	1.64
128	384	6.786e-03	8.666e-03	7.553e-03	1.100e-02	1.46
512	1536	1.727e-03	1.983e-03	1.897e-03	2.630e-03	1.39
2048	6144	4.347e-04	4.773e-04	4.749e-04	6.456e-04	1.35
8192	24576	1.088e-04	1.173e-04	1.187e-04	1.601e-04	1.34
$k = 2, k' = 3$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{R}}$	$\ u - u_h\ $	$\eta_{\text{mod}}$	$I_{\text{eff}}$
8	48	2.271e-02	4.807e-02	1.882e-02	3.360e-02	1.78
32	192	3.106e-03	3.785e-03	2.476e-03	3.495e-03	1.41
128	768	3.972e-04	4.147e-04	3.135e-04	4.254e-04	1.36
512	3072	4.995e-05	5.012e-05	3.929e-05	5.280e-05	1.34
2048	12288	6.253e-06	6.216e-06	4.934e-06	6.592e-06	1.33
8192	49152	7.822e-07	7.843e-07	6.270e-07	8.322e-07	1.32

Table 6: Convergence of the error  $\|u - u_h\|$ , the error estimators  $\eta_{\text{mod}}$ ,  $\eta_{\text{NC}}$ , and  $\eta_{\text{R}}$ , and the effectivity indices  $I_{\text{eff}}$  for the dG method (3.3); smooth solution (9.9), different velocity fields, and  $k = 1$

$k = 1, k' = 2, b = (100, 100)^t$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{R}}$	$\ u - u_h\ $	$\eta_{\text{mod}}$	$I_{\text{eff}}$
8	24	9.365e-02	2.083e-01	1.097e-01	2.175e-01	1.98
32	96	2.584e-02	4.156e-02	2.963e-02	4.871e-02	1.64
128	384	6.786e-03	8.666e-03	7.553e-03	1.100e-02	1.46
512	1536	1.727e-03	1.983e-03	1.897e-03	2.630e-03	1.39
2048	6144	4.347e-04	4.773e-04	4.749e-04	6.456e-04	1.35
8192	24576	1.088e-04	1.173e-04	1.187e-04	1.601e-04	1.34
$k = 1, k' = 2, b = (10, 1)^t$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{R}}$	$\ u - u_h\ $	$\eta_{\text{mod}}$	$I_{\text{eff}}$
8	24	8.299e-02	2.216e-01	1.009e-01	2.307e-01	2.28
32	96	2.057e-02	4.714e-02	2.896e-02	5.137e-02	1.77
128	384	5.325e-03	1.062e-02	7.965e-03	1.188e-02	1.49
512	1536	1.370e-03	2.684e-03	2.069e-03	3.014e-03	1.46
2048	6144	3.459e-04	6.807e-04	5.241e-04	7.636e-04	1.45
8192	24576	8.667e-05	1.711e-04	1.316e-04	1.918e-04	1.45
$k = 1, k' = 2, b = (y, x + 1)^t (\alpha = 1)$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{R}}$	$\ u - u_h\ $	$\eta_{\text{mod}}$	$I_{\text{eff}}$
8	24	9.582e-02	2.239e-01	1.134e-01	2.360e-01	2.08
32	96	2.513e-02	5.212e-02	3.152e-02	5.780e-02	1.83
128	384	6.478e-03	1.233e-02	8.007e-03	1.393e-02	1.74
512	1536	1.636e-03	2.991e-03	2.013e-03	3.409e-03	1.69
2048	6144	4.103e-04	7.379e-04	5.053e-04	8.443e-04	1.67
8192	24576	1.027e-04	1.833e-04	1.267e-04	2.101e-04	1.66

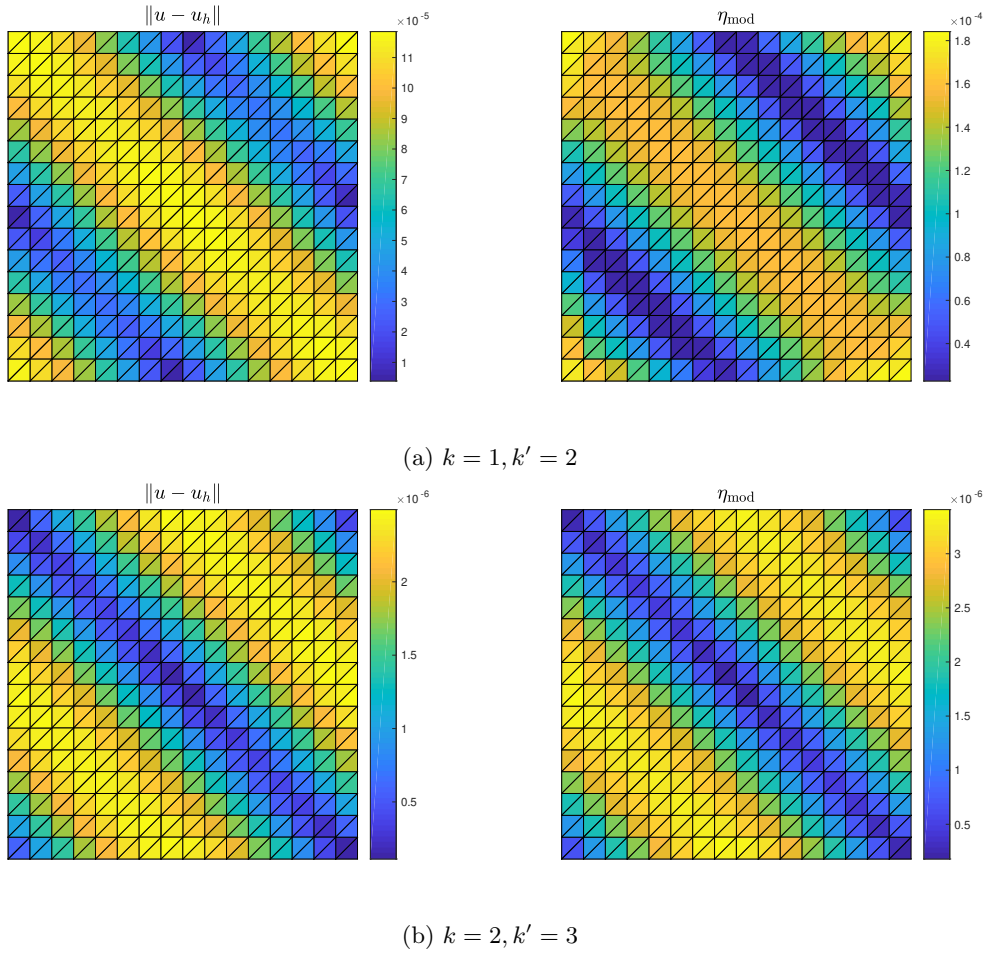


Figure 1: Distribution of the errors  $\|u - u_h\|_K$  (left) and of the local error estimators  $\eta_{\text{mod},K}$  (right) for the dG method (3.3) with 512 elements; smooth solution (9.9),  $\mathbf{b} = (1, 1)^t$ , and different polynomial degrees  $k$

### 9.5.2 Discontinuous solution

In this last example, we consider a discontinuous exact solution. For the velocity field  $\mathbf{b} = (1, 1)^t$ , we set

$$u(x, y) = \begin{cases} 0, & x < y, \\ \sin(\pi x) \sin(\pi y), & x > y, \end{cases} \quad (9.10)$$

and prescribe accordingly the right-hand side  $f$ . The triangulation is set to be aligned with this discontinuity; hence, the reconstruction  $s_h$  is continuous everywhere but not at the discontinuity line of the exact solution. The results are presented in Table 7 for different polynomial degrees  $k$ . They show the robustness with respect to the polynomial degree of approximation. In Figure 2, the distributions of the error and of the error estimators  $\eta_{\text{mod},K}$  are presented, again showing a very similar behavior.



Table 7: Convergence of the error  $\|u - u_h\|$ , the error estimators  $\eta_{\text{mod}}$ ,  $\eta_{\text{NC}}$ , and  $\eta_{\text{R}}$ , and the effectivity indices  $I_{\text{eff}}$  for the dG method (3.3); discontinuous solution (9.10),  $\mathbf{b} = (1, 1)^t$ , and different polynomial degrees  $k$

$k = 1, k' = 2$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{R}}$	$\ u - u_h\ $	$\eta_{\text{mod}}$	$I_{\text{eff}}$
8	24	6.622e-02	1.473e-01	7.758e-02	1.538e-01	1.98
32	96	1.827e-02	2.939e-02	2.095e-02	3.445e-02	1.64
128	384	4.798e-03	6.127e-03	5.341e-03	7.780e-03	1.45
512	1536	1.221e-03	1.402e-03	1.341e-03	1.860e-03	1.38
2048	6144	3.074e-04	3.375e-04	3.358e-04	4.565e-04	1.36
8192	24576	7.705e-05	8.295e-05	8.399e-05	1.132e-04	1.35
$k = 2, k' = 3$						
# Elements	# DOF( $u_h$ )	$\eta_{\text{NC}}$	$\eta_{\text{R}}$	$\ u - u_h\ $	$\eta_{\text{mod}}$	$I_{\text{eff}}$
8	48	1.626e-02	7.463e-02	1.330e-02	3.680e-02	2.76
32	192	2.227e-03	5.102e-03	1.751e-03	3.105e-03	1.77
128	768	2.850e-04	3.544e-04	2.217e-04	3.454e-04	1.56
512	3072	3.583e-05	2.847e-05	2.778e-05	4.181e-05	1.50
2048	12288	4.485e-06	2.849e-06	3.489e-06	5.184e-06	1.48
8192	49152	5.610e-07	3.410e-07	4.433e-07	6.525e-07	1.47

## 10 Conclusions

In this work, we proposed a local reconstruction for numerical approximations of the one-dimensional linear advection equation, easily and independently obtained on each vertex patch. The reconstruction is proved to be well-posed and leads to a guaranteed upper bound of the  $L_2$ -norm error between the actual solution  $u$  and the approximation  $u_h$ . This error estimator is also proved to be locally efficient and robust with respect to both the advective field and the approximation polynomial degree. These results hold in a unified framework that only requires the residual of  $u_h$  to satisfy an orthogonality condition with respect to the hat basis functions. Numerical illustrations support the theory and additionally suggest asymptotic exactness. Motivated by these results, a heuristic extension to any space dimension is presented, with numerical experiments in 2D being in line with those in 1D.

## References

- [1] O. AXELSSON, J. KARÁTSÓN, AND B. KOVÁCS, *Robust preconditioning estimates for convection-dominated elliptic problems via a streamline Poincaré-Friedrichs inequality*, SIAM J. Numer. Anal., 52 (2014), pp. 2957–2976.
- [2] B. AYUSO AND L. D. MARINI, *Discontinuous Galerkin methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 47 (2009), pp. 1391–1420.
- [3] P. AZÉRAD AND J. POUSIN, *Inégalité de Poincaré courbe pour le traitement variationnel de l'équation de transport*, C. R. Acad. Sci. Paris Sér. I Math., 322 (1996), pp. 721–727.

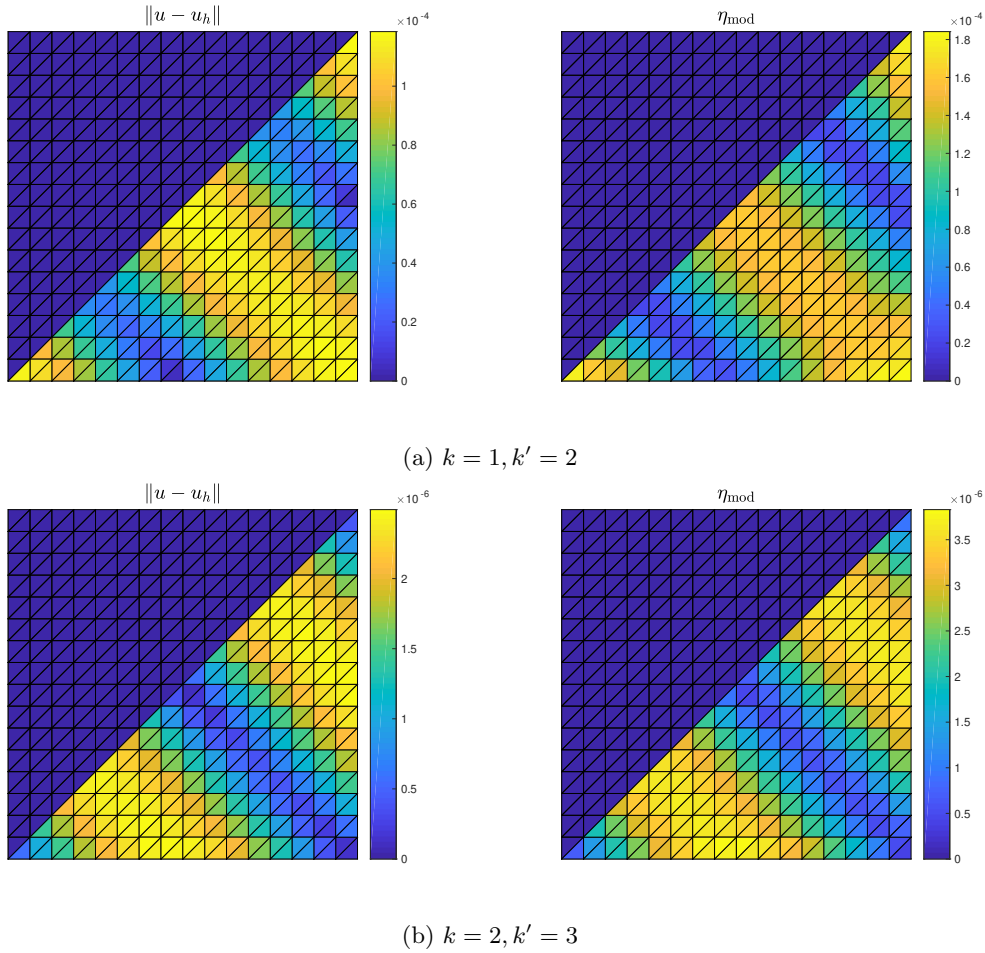


Figure 2: Distribution of the errors  $\|u - u_h\|_K$  (left) and of the local error estimators  $\eta_{\text{mod},K}$  (right) for the dG method (3.3) with 512 elements; discontinuous solution (9.10),  $\mathbf{b} = (1, 1)^t$ , and different polynomial degrees  $k$

- [4] R. BECKER, D. CAPATINA, AND R. LUCE, *Reconstruction-based a posteriori error estimators for the transport equation*, in Numerical mathematics and advanced applications 2011, Springer, Heidelberg, 2013, pp. 13–21.
- [5] K. S. BEY AND J. T. ODEN, *hp-version discontinuous Galerkin methods for hyperbolic conservation laws*, Comput. Methods Appl. Mech. Engrg., 133 (1996), pp. 259–286.
- [6] J. BLECHTA, J. MÁLEK, AND M. VOHRALÍK, *Localization of the  $W^{-1,q}$  norm for local a posteriori efficiency*, IMA J. Numer. Anal., (2019). DOI 10.1093/imanum/drz002.
- [7] D. BRAESS, V. PILLWEIN, AND J. SCHÖBERL, *Equilibrated residual error estimates are p-robust*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 1189–1197.
- [8] P. CANTIN, *Well-posedness of the scalar and the vector advection-reaction problems in Banach graph spaces*, C. R. Math. Acad. Sci. Paris, 355 (2017), pp. 892–902.

- [9] P. CANTIN AND A. ERN, *An edge-based scheme on polyhedral meshes for vector advection-reaction equations*, ESAIM Math. Model. Numer. Anal., 51 (2017), pp. 1561–1581.
- [10] C. CARSTENSEN AND S. A. FUNKEN, *Fully reliable localized error control in the FEM*, SIAM J. Sci. Comput., 21 (1999), pp. 1465–1484.
- [11] W. DAHMEN, C. HUANG, C. SCHWAB, AND G. WELPER, *Adaptive Petrov-Galerkin methods for first order transport equations*, SIAM J. Numer. Anal., 50 (2012), pp. 2420–2445.
- [12] W. DAHMEN AND R. STEVENSON, *Adaptive strategies for transport equations*, arXiv preprint arXiv:1809.02055, (2019).
- [13] A. DEVINATZ, R. ELLIS, AND A. FRIEDMAN, *The asymptotic behavior of the first real eigenvalue of second order elliptic operators with a small parameter in the highest derivatives. II*, Indiana Univ. Math. J., 23 (1973–1974), pp. 991–1011.
- [14] A. ERN AND J.-L. GUERMOND, *Discontinuous Galerkin methods for Friedrichs’ systems. I. General theory*, SIAM J. Numer. Anal., 44 (2006), pp. 753–778.
- [15] A. ERN, A. F. STEPHANSEN, AND M. VOHRALÍK, *Guaranteed and robust discontinuous Galerkin a posteriori error estimates for convection-diffusion-reaction problems*, J. Comput. Appl. Math., 234 (2010), pp. 114–130.
- [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] K. O. FRIEDRICHS, *Symmetric positive linear differential equations*, Comm. Pure Appl. Math., 11 (1958), pp. 333–418.
- [18] E. H. GEORGIOULIS, E. HALL, AND C. MAKRIDAKIS, *Error control for discontinuous Galerkin methods for first order hyperbolic problems*, in Recent developments in discontinuous Galerkin finite element methods for partial differential equations, vol. 157 of IMA Vol. Math. Appl., Springer, Cham, 2014, pp. 195–207.
- [19] ———, *An a posteriori error bound for discontinuous Galerkin approximations of convection-diffusion problems*, IMA J. Numer. Anal., 39 (2019), pp. 34–60.
- [20] F. HECHT, *New development in FreeFEM++*, J. Numer. Math., 20 (2012), pp. 251–265.
- [21] P. HOUSTON, J. A. MACKENZIE, E. SÜLI, AND G. WARNECKE, *A posteriori error analysis for numerical approximations of Friedrichs systems*, Numer. Math., 82 (1999), pp. 433–470.
- [22] P. D. LAX AND R. S. PHILLIPS, *Local boundary conditions for dissipative symmetric linear differential operators*, Comm. Pure Appl. Math., 13 (1960), pp. 427–455.
- [23] 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.
- [24] I. MUGA, M. J. TYLER, AND K. VAN DER ZEE, *The discrete-dual minimal-residual method (DDMRes) for weak advection-reaction problems in Banach spaces*, arXiv preprint arXiv:1808.04542, (2018).
- [25] G. SANGALLI, *Analysis of the advection-diffusion operator using fractional order norms*, Numer. Math., 97 (2004), pp. 779–796.

- [26] ———, *A uniform analysis of nonsymmetric and coercive linear operators*, SIAM J. Math. Anal., 36 (2005), pp. 2033–2048.
- [27] ———, *Robust a-posteriori estimator for advection-diffusion-reaction problems*, Math. Comp., 77 (2008), pp. 41–70.
- [28] D. SCHÖTZAU AND L. ZHU, *A robust a-posteriori error estimator for discontinuous Galerkin methods for convection-diffusion equations*, Appl. Numer. Math., 59 (2009), pp. 2236–2255.
- [29] E. SÜLI, *A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems*, in An introduction to recent developments in theory and numerics for conservation laws (Freiburg/Littenweiler, 1997), vol. 5 of Lect. Notes Comput. Sci. Eng., Springer, Berlin, 1999, pp. 123–194.
- [30] Z. TANG, <https://who.rocq.inria.fr/Zuqi.Tang/freefem++.html>, 2015.
- [31] D. S. TARTAKOFF, *Regularity of solutions to boundary value problems for first order systems*, Indiana Univ. Math. J., 21 (1972), pp. 1113–1129.
- [32] R. VERFÜRTH, *Robust a posteriori error estimates for stationary convection-diffusion equations*, SIAM J. Numer. Anal., 43 (2005), pp. 1766–1782.
- [33] R. VERFÜRTH, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.