

hp-Finite Element Methods for Hyperbolic Problems ^a

Endre Süli

*Oxford University Computing Laboratory, Wolfson Building,
Parks Road, Oxford OX1 3QD, UK*

Paul Houston

*Oxford University Computing Laboratory, Wolfson Building,
Parks Road, Oxford OX1 3QD, UK*

Christoph Schwab

*Seminar of Applied Mathematics, ETH Zürich,
CH-8092 Zürich, Switzerland*

This paper is devoted to the *a priori* and *a posteriori* error analysis of the *hp*-version of the discontinuous Galerkin finite element method for partial differential equations of hyperbolic and nearly-hyperbolic character. We consider second-order partial differential equations with nonnegative characteristic form, a large class of equations which includes convection-dominated diffusion problems, degenerate elliptic equations and second-order problems of mixed elliptic-hyperbolic-parabolic type. An *a priori* error bound is derived for the method in the so-called DG-norm which is optimal in terms of the mesh size h ; the error bound is either 1 degree or 1/2 degree below optimal in terms of the polynomial degree p , depending on whether the problem is convection-dominated, or diffusion-dominated, respectively. In the case of a first-order hyperbolic equation the error bound is *hp*-optimal in the DG-norm. For first-order hyperbolic problems, we also discuss the *a posteriori* error analysis of the method and implement the resulting bounds into an *hp*-adaptive algorithm. The theoretical findings are illustrated by numerical experiments.

Oxford University Computing Laboratory
Numerical Analysis Group
Wolfson Building
Parks Road
Oxford, England OX1 3QD

June, 1999

^aPresented as Invited Lecture at the 10th Conference on the Mathematics of Finite Elements and Applications, Brunel University, June 1999.

1 Introduction

The discontinuous Galerkin finite element method (DGFEM) has a long and distinguished history. Its roots can be traced back to the work of Pian and collaborators [20] in the early 1960s on hybrid methods for elliptic problems (see also [19] for a historical survey); the mathematical analysis of hybrid methods was initiated by Babuška [4]. In 1971, J. Nitsche [18] considered an alternative scheme where the boundary multipliers present in the hybrid formulation were eliminated in terms of normal fluxes and stabilisation terms were added to recover the optimal convergence rate. In a different context, discontinuous finite element methods were introduced by Reed and Hill [21], and Lesaint and Raviart [17] in order to overcome the stability limitations of conventional continuous finite element approximations to first-order hyperbolic problems. Although subsequently much of the research in the field of numerical analysis of partial differential equations concentrated on the development and the analysis of conforming finite element methods, in recent years there has been an upsurge of interest in discontinuous schemes. This paradigm shift was stimulated by several factors: the desire to handle, within the finite element framework, nonlinear hyperbolic problems (see [10] and [11]) which are known to exhibit discontinuous solutions even when the data are perfectly smooth; the need to treat convection-dominated diffusion problems without excessive numerical stabilisation; the computational convenience of discontinuous finite element methods due to a large degree of locality; and the necessity to accommodate high-order hp -adaptive finite element discretisations in a flexible manner (see [9]). The DGFEM can also be thought of as the higher-order extension of the classical cell centre finite volume method – a popular discretisation technique in the computational aerodynamics community.

In the present paper we develop the error analysis of the hp -DGFEM for partial differential equations of hyperbolic and nearly hyperbolic character. We begin by considering the *a priori* error analysis of the hp -DGFEM for second-order partial differential equations with nonnegative characteristic form; this represents a continuation of our earlier work [14] for first-order hyperbolic equations. In [14] an error bound, optimal both in terms of the local mesh size h and the local polynomial degree p , was derived for the hp -DGFEM supplemented by a streamline-diffusion type stabilisation involving a stabilisation parameter δ of size h/p . Here, we establish a similar result in the case of partial differential equations with nonnegative characteristic form; the resulting error bound is optimal in terms of powers of h , the part of the error bound which arises from the diffusion term is by one power of p below the optimal rate, while the parts which stem from the advection and reaction terms are of optimal order in p . For convection-dominated diffusion equations, suboptimality in p is compensated by the fact that the leading term in the error bound is multiplied by a small number, proportional to the square root of the norm of the diffusion matrix. Indeed, in the case of a first-order hyperbolic equation, our error bound collapses to one that is hp -optimal. On the other hand, when the advective term is absent, the error bound is optimal in terms of powers of h and it is $1/2$ a power below optimal in terms of the polynomial degree p . The hp -DGFEM considered in this paper involves a discontinuity-penalisation device based on the ideas of Nitsche [18], Wheeler [26] and Arnold [3], albeit with a small but significant modification which permits us to pass to the hyperbolic limit with inactive discontinuity-penalisation. The error analysis of the hp -DGFEM discretisation considered

here can also be viewed as an extension of the work of Baumann [6], Oden, Babuška and Baumann [19] and Riviere and Wheeler [22] in the reaction-diffusion case; also, they present an improvement over our earlier results presented in [25] where the error analysis of the hp -DGFEM was considered for partial differential equations with nonnegative characteristic form, albeit without streamline diffusion stabilisation. While in [25] the size of the discontinuity penalisation parameter was required to be $Const.p^2/h$, in the present paper it has been reduced to $Const/h$.

The second part of the paper is concerned with the *a posteriori* error analysis of the hp -DGFEM for hyperbolic problems. Here, we derive an *a posteriori* bound on the error for hp -DGFEM approximations of linear functionals of the analytical solution. The *a posteriori* error bound is based on an error representation formula which stems from a duality argument and the Galerkin orthogonality property of the hp -DGFEM. The error representation formula involves the computable finite element residual and the difference between the dual solution and its projection. We exemplify the relevance of the theoretical results by implementing the *a posteriori* error bound into an hp -adaptive algorithm for calculating the outflow normal flux of the solution to within a prescribed tolerance.

2 Model Problem and Discretisation

Suppose that Ω is a bounded Lipschitz domain in \mathbf{R}^d , $d = 2, 3$, and consider the linear second-order partial differential equation

$$\mathcal{L}u \equiv - \sum_{i,j=1}^d \partial_j (a_{ij}(x) \partial_i u) + \sum_{i=1}^d b_i(x) \partial_i u + c(x)u = f(x) , \quad (2.1)$$

where f is a real-valued function belonging to $L^2(\Omega)$, and the real-valued coefficients a, b, c are such that:

$$\begin{aligned} a(x) &= \{a_{ij}(x)\}_{i,j=1}^d \in L^\infty(\Omega)_{\text{sym}}^{d \times d} , \\ \vec{b}(x) &= \{b_i(x)\}_{i=1}^d \in W^{1,\infty}(\Omega)^d , \quad c(x) \in L^\infty(\Omega) . \end{aligned} \quad (2.2)$$

It will be assumed throughout that the characteristic form associated with the principal part of the partial differential operator \mathcal{L} is nonnegative; that is,

$$\xi^T a(x) \xi \geq 0 \quad \forall \xi \in \mathbf{R}^d \text{ and a.e. } x \in \bar{\Omega} . \quad (2.3)$$

In order to ensure that the restriction of the matrix a to the boundary $\partial\Omega$ of Ω is well defined, we shall assume, for simplicity, that the entries of a are piecewise continuous on $\bar{\Omega}$. This assumption is sufficiently general to cover most cases of practical significance. Now let $\vec{\mu}(x) = \{\mu_i(x)\}_{i=1}^d$ denote the unit outward normal vector to $\Gamma = \partial\Omega$ at $x \in \Gamma$ and define the following subsets of Γ :

$$\begin{aligned} \Gamma_0 &= \{x \in \Gamma : \vec{\mu}^T a(x) \vec{\mu} > 0\} , \\ \Gamma_- &= \{x \in \Gamma \setminus \Gamma_0 : \vec{b} \cdot \vec{\mu} < 0\} \quad \text{and} \quad \Gamma_+ = \{x \in \Gamma \setminus \Gamma_0 : \vec{b} \cdot \vec{\mu} \geq 0\} . \end{aligned}$$

The sets Γ_\mp will be referred to as the inflow and outflow boundary, respectively. With these definitions we have that $\Gamma = \Gamma_0 \cup \Gamma_- \cup \Gamma_+$. We shall further decompose Γ_0 into

two connected parts, Γ_D where a Dirichlet boundary condition is imposed and Γ_N where a Neumann condition is given, and we supplement the partial differential equation (2.1) with the following boundary conditions:

$$u = g_D \quad \text{on } \Gamma_D \cup \Gamma_- \quad \text{and} \quad \vec{\mu}^T a \nabla u = g_N \quad \text{on } \Gamma_N . \quad (2.4)$$

The boundary value problem (2.1), (2.4) includes a range of physically relevant instances, such as the mixed boundary value problem for an elliptic equation corresponding to the case when (2.3) holds with strict inequality, as well as the case of a linear transport problem associated with the choice of $a \equiv 0$ on $\bar{\Omega}$.

2.1 Finite element spaces

Suppose that \mathcal{T} is a subdivision of Ω into open element domains κ such that $\bar{\Omega} = \bigcup_{\kappa \in \mathcal{T}} \bar{\kappa}$. Let us assume that the family of subdivisions \mathcal{T} is shape-regular and that each $\kappa \in \mathcal{T}$ is a smooth bijective image of a fixed master element $\hat{\kappa}$, that is, $\kappa = F_\kappa(\hat{\kappa})$ for all $\kappa \in \mathcal{T}$ where $\hat{\kappa}$ is either the open unit simplex or the open unit hypercube in \mathbf{R}^d . For an integer $r \geq 1$, we denote by $\mathcal{P}_r(\hat{\kappa})$ the set of polynomials of total degree $\leq r$ on $\hat{\kappa}$; when $\hat{\kappa}$ is the unit hypercube, we also consider $\mathcal{Q}_r(\hat{\kappa})$, the set of all tensor-product polynomials of degree $\leq r$ in each coordinate direction. The case of $r = 0$ can be easily incorporated into our analysis, but we have chosen to exclude it for simplicity of presentation so as to ensure that $1/r$ is meaningful for any polynomial degree under consideration. Next, to $\kappa \in \mathcal{T}$ we assign an integer $p_\kappa \geq 1$, collect the p_κ and F_κ in the vectors $\mathbf{p} = \{p_\kappa : \kappa \in \mathcal{T}\}$ and $\mathbf{F} = \{F_\kappa : \kappa \in \mathcal{T}\}$, respectively, and consider the finite element space

$$S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F}) = \{u \in L^2(\Omega) : u|_\kappa \circ F_\kappa \in \mathcal{R}_{p_\kappa}(\hat{\kappa}) \quad \forall \kappa \in \mathcal{T}\} ,$$

where \mathcal{R}_{p_κ} is either \mathcal{P}_{p_κ} or \mathcal{Q}_{p_κ} . Assuming that \mathcal{T} is a subdivision of Ω and $s > 0$, $H^s(\Omega, \mathcal{T})$ will denote the associated broken Sobolev space of index s .

2.2 The numerical method

Discretisation of the Low-Order Terms. Since the emphasis in this paper is on problems of hyperbolic and nearly-hyperbolic character, we begin by considering the hp -DGFEM approximation of the first-order partial differential operator \mathcal{L}_b defined by

$$\mathcal{L}_b w = \vec{b} \cdot \nabla w + c w .$$

Assuming that κ is an element in the subdivision \mathcal{T} , we denote by $\partial\kappa$ the union of open faces of κ . This is non-standard notation in that $\partial\kappa$ is a subset of the boundary of κ ; we have adopted it so as to ensure that the unit outward normal vector $\vec{\mu}(x)$ to $\partial\kappa$ at $x \in \partial\kappa$ is correctly defined. With these conventions, we define the inflow and outflow parts of $\partial\kappa$, respectively, by

$$\partial_- \kappa = \{x \in \partial\kappa : \vec{b}(x) \cdot \vec{\mu}(x) < 0\} , \quad \partial_+ \kappa = \{x \in \partial\kappa : \vec{b}(x) \cdot \vec{\mu}(x) \geq 0\} . \quad (2.5)$$

For each $v \in H^1(\Omega, \mathcal{T})$ and any $\kappa \in \mathcal{T}$, we denote by v^+ the interior trace of v on $\partial\kappa$ (the trace taken from within κ). Let us consider an element κ such that the set $\partial_- \kappa \setminus \Gamma_-$ is

nonempty; then for each $x \in \partial_- \kappa \setminus \Gamma_-$ (with the exception of a set of $(d-1)$ -dimensional measure zero) there exists a unique element κ' , depending on the choice of x , such that $x \in \partial_+ \kappa'$. If $\partial_- \kappa \setminus \Gamma_-$ is nonempty for some $\kappa \in \mathcal{T}$, then we can also define the outer trace v^- of v on $\partial_- \kappa \setminus \Gamma_-$ relative to κ as the inner trace v^+ relative to those elements κ' for which $\partial_+ \kappa'$ has intersection with $\partial_- \kappa \setminus \Gamma_-$ of positive $(d-1)$ -dimensional measure. Furthermore, we introduce the oriented jump of v across $\partial_- \kappa \setminus \Gamma_-$: $[v] = v^+ - v^-$.

Given that $v, w \in H^1(\Omega, \mathcal{T})$, we define, as in [16], for example, the bilinear form

$$\begin{aligned} B_b(w, v) = & \sum_{\kappa \in \mathcal{T}} \int_{\kappa} (\mathcal{L}_b w) v \, dx \\ & - \sum_{\kappa \in \mathcal{T}} \int_{\partial_- \kappa \setminus \Gamma_-} (\vec{b} \cdot \vec{\mu}) [w] v^+ \, ds - \sum_{\kappa \in \mathcal{T}} \int_{\partial_- \kappa \cap \Gamma_-} (\vec{b} \cdot \vec{\mu}) w^+ v^+ \, ds, \end{aligned} \quad (2.6)$$

and the linear functional

$$\ell_b(v) = \sum_{\kappa \in \mathcal{T}} \int_{\kappa} f v \, dx - \sum_{\kappa \in \mathcal{T}} \int_{\partial_- \kappa \cap \Gamma_-} (\vec{b} \cdot \vec{\mu}) g v^+ \, ds. \quad (2.7)$$

Next, we focus on the discretisation of the leading term in the partial differential equation.

Discretisation of the Leading Term. Let us suppose that the elements in the subdivision have been numbered in a certain way, regardless of the direction of the advective velocity vector \vec{b} . We denote by \mathcal{E} the set of element faces (edges for $d=2$ or faces for $d=3$) associated with the subdivision \mathcal{T} . Since hanging nodes are permitted in DGFEM, \mathcal{E} will be understood to consist of the smallest faces in $\partial \kappa$. Also, let \mathcal{E}_{int} , resp. Γ_{int} , denote the set, resp. union, of all faces $e \in \mathcal{E}$ which do not lie on $\partial \Omega$. Given that $e \in \mathcal{E}_{\text{int}}$, there exist indices i and j such that $i > j$ and κ_i and κ_j share the interface e ; we define the (numbering-dependent) jump of $v \in H^1(\Omega, \mathcal{T})$ across e and the mean value of v on e , respectively, by $[v] = v|_{\partial \kappa_i \cap e} - v|_{\partial \kappa_j \cap e}$ and $\langle v \rangle = (v|_{\partial \kappa_i \cap e} + v|_{\partial \kappa_j \cap e})/2$.

It is clear that, in general, $[v]$ will be distinct from $[v]$ in that the latter depends on the direction of the unit outward normal to an element boundary, while the former is only dependent on the element numbering; however, $||[v]|| = |[v]|$. With each face $e \in \mathcal{E}_{\text{int}}$ we associate the normal vector $\vec{\nu}$ which points from κ_i to κ_j ; on boundary faces we define $\vec{\nu} = \vec{\mu}$. Finally, we introduce, following [19], the bilinear form

$$\begin{aligned} B_a(w, v) = & \sum_{\kappa \in \mathcal{T}} \int_{\kappa} a(x) \nabla w \cdot \nabla v \, dx + \int_{\Gamma_D} \{w((a \nabla v) \cdot \vec{\nu}) - ((a \nabla w) \cdot \vec{\nu})v\} \, ds \\ & + \int_{\Gamma_{\text{int}}} \{[w]\langle (a \nabla v) \cdot \vec{\nu} \rangle - \langle (a \nabla w) \cdot \vec{\nu} \rangle [v]\} \, ds, \end{aligned} \quad (2.8)$$

associated with the principal part of the differential operator \mathcal{L} , and the linear functional

$$\ell_a(v) = \int_{\Gamma_D} g_D((a \nabla v) \cdot \vec{\nu}) \, ds + \int_{\Gamma_N} g_N v \, ds.$$

Discontinuity-Penalisation Term. Let $\bar{a} = \|a\|_2$, with $\|\cdot\|_2$ denoting the matrix norm subordinate to the l^2 vector norm on \mathbf{R}^d , and let $\bar{a}_\kappa = \bar{a}|_\kappa$. To each e in \mathcal{E}_{int} which is a common face of elements κ_i and κ_j in \mathcal{T} we assign the nonnegative function $\langle \bar{a} \rangle_e = (\bar{a}_{\kappa_i}|_e + \bar{a}_{\kappa_j}|_e)/2$. Letting \mathcal{E}_{D} denote the set of all faces contained in Γ_{D} , to each $e \in \mathcal{E}_{\text{D}}$ we assign the element $\kappa \in \mathcal{T}$ with that face and define $\langle \bar{a} \rangle_e = \bar{a}_\kappa|_e$. Consider the function σ defined on $\Gamma_{\text{int}} \cup \Gamma_{\text{D}}$ by $\sigma(x) = K \langle \bar{a} \rangle_e / |e|$ for $x \in e$ and $e \in \mathcal{E}_{\text{int}} \cup \mathcal{E}_{\text{D}}$, where $|e| = \text{meas}_{d-1}(e)$ and K is a positive constant (whose value is irrelevant for the present analytical study, so we put $K = 1$) and introduce

$$B_s(w, v) = \int_{\Gamma_{\text{D}}} \sigma w v ds + \int_{\Gamma_{\text{int}}} \sigma [w][v] ds, \quad \ell_s(v) = \int_{\Gamma_{\text{D}}} \sigma g_{\text{D}} v ds. \quad (2.9)$$

We highlight the fact that since the weight-function σ involves the norm of the matrix a , in the hyperbolic limit of $a \equiv 0$ the bilinear form $B_s(\cdot, \cdot)$ and the linear functional ℓ_s both vanish. This is a desirable property, since linear hyperbolic equations may possess solutions that are discontinuous across characteristic hypersurfaces, and penalising discontinuities across faces which belong to these would be unnatural.

Streamline-diffusion stabilisation. Let $\delta \in H^1(\Omega, \mathcal{T})$ be a nonnegative function. In the present context δ will play the role of a *stabilisation parameter*; typically δ is chosen to be constant on each $\kappa \in \mathcal{T}$, although we shall not require this for now. We define the bilinear form and the linear functional, respectively,

$$B_\delta(w, v) = \sum_{\kappa \in \mathcal{T}} \int_{\kappa} \delta (\mathcal{L}w)(\vec{b} \cdot \nabla v) dx, \quad \ell_\delta(v) = \sum_{\kappa \in \mathcal{T}} \int_{\kappa} \delta f(\vec{b} \cdot \nabla v) dx. \quad (2.10)$$

The precise choice of the stabilisation parameter will be given in the next section.

Definition of the Method. Finally, we define the bilinear form $B_{\text{DG}}(\cdot, \cdot)$ and the linear functional $\ell_{\text{DG}}(\cdot)$, respectively, by

$$\begin{aligned} B_{\text{DG}}(w, v) &= B_a(w, v) + B_b(w, v) + B_s(w, v) + B_\delta(w, v), \\ \ell_{\text{DG}}(v) &= \ell_a(v) + \ell_b(v) + \ell_s(v) + \ell_\delta(v). \end{aligned}$$

The hp -DGFEM approximation of (2.1), (2.4) is: find $u_{\text{DG}} \in S^{\text{p}}(\Omega, \mathcal{T}, \mathbf{F})$ such that

$$B_{\text{DG}}(u_{\text{DG}}, v) = \ell_{\text{DG}}(v) \quad \forall v \in S^{\text{p}}(\Omega, \mathcal{T}, \mathbf{F}). \quad (2.11)$$

In the next section we state the key properties of this method. Before we do so, however, we note that in the definitions of the bilinear forms and linear functionals above and in the arguments which follow it has been tacitly assumed that $a \in C(\kappa)$ for each $\kappa \in \mathcal{T}$, that the fluxes $(a \nabla u) \cdot \vec{\nu}$ and $(\vec{b} \cdot \vec{\mu})u$ are continuous across element interfaces, and that u is continuous in an (open) neighbourhood of the subset of Ω where a is not identically equal to zero. If the problem under consideration violates these properties, the scheme and the analysis have to be modified accordingly.

3 Analytical Results

Our first result concerns the positivity of the bilinear form $B_{\text{DG}}(\cdot, \cdot)$ and the existence and uniqueness of a solution to (2.11). In order to prove it, we shall require the following *inverse inequality* (see [23]): there exists a positive constant C_{inv} , dependent only on the constant of the angle condition such that

$$\|\nabla \cdot W\|_{L^2(\kappa)} \leq C_{\text{inv}} \frac{p_\kappa^2}{h_\kappa} \|W\|_{L^2(\kappa)} \quad (3.1)$$

for all $\kappa \in \mathcal{T}$ and all $W = (w_1, \dots, w_d) \in [S^{\mathbf{P}}(\Omega, \mathcal{T}, \mathbf{F})]^d$.

Theorem 1 *Suppose that, in addition to (2.2) and (2.3), there exists a positive constant γ_0 such that $\gamma \equiv c - \frac{1}{2}\nabla \cdot \vec{b} \geq \gamma_0$ on $\bar{\Omega}$. Let us also assume that*

$$0 \leq \delta \leq \frac{1}{2} \min \left(\frac{h_\kappa^2}{C_{\text{inv}}^2 p_\kappa^4 \bar{a}_\kappa}, \frac{\gamma}{\bar{c}_\kappa^2} \right) \quad \forall \kappa \in \mathcal{T} , \quad (3.2)$$

where $\bar{c}_\kappa = \|c\|_{L^\infty(\kappa)}$. Then,

$$B_{\text{DG}}(w, w) \geq |||w|||_{\text{DG}}^2 \equiv D + \frac{1}{2} \sum_{\kappa \in \mathcal{T}} E_\kappa + \frac{1}{2} \sum_{\kappa \in \mathcal{T}} F_\kappa + \frac{1}{2} \sum_{\kappa \in \mathcal{T}} G_\kappa , \quad (3.3)$$

where

$$\begin{aligned} D &\equiv \int_{\Gamma_{\text{D}}} \sigma w^2 \text{d}s + \int_{\Gamma_{\text{int}}} \sigma [w]^2 \text{d}s , & E_\kappa &\equiv \|\sqrt{a} \nabla w\|_{L^2(\kappa)}^2 + \|\sqrt{\gamma} w\|_{L^2(\kappa)}^2 , \\ F_\kappa &\equiv \int_{\partial_- \kappa \cap \Gamma_-} |\vec{b} \cdot \vec{\mu}| w_+^2 \text{d}s + \int_{\partial_+ \kappa \cap \Gamma_+} |\vec{b} \cdot \vec{\mu}| w_+^2 \text{d}s + \int_{\partial_- \kappa \setminus \Gamma_-} |\vec{b} \cdot \vec{\mu}| [w]^2 \text{d}s , \\ G_\kappa &\equiv \|\sqrt{\delta} (\vec{b} \cdot \nabla w)\|_{L^2(\kappa)}^2 , \end{aligned}$$

with \sqrt{a} denoting the (nonnegative) square-root of the matrix a , and σ as in the definition of the discontinuity-penalisation. Furthermore, the hp-DGFEM (2.11) has a unique solution u_{DG} in $S^{\mathbf{P}}(\Omega, \mathcal{T}, \mathbf{F})$.

PROOF: We begin by proving (3.3). First, we note that, trivially,

$$B_s(w, w) = \int_{\Gamma_{\text{D}}} \sigma w^2 \text{d}s + \int_{\Gamma_{\text{int}}} \sigma [w]^2 \text{d}s .$$

Further, as $(\vec{b} \cdot \nabla w)w = \frac{1}{2}\vec{b} \cdot \nabla(w^2)$, after integration by parts we have that

$$B_b(w, w) = \frac{1}{2} \sum_{\kappa \in \mathcal{T}} F_\kappa + \sum_{\kappa \in \mathcal{T}} \int_\kappa |\sqrt{\gamma(x)} w(x)|^2 \text{d}x .$$

Next, we observe that

$$B_a(w, w) = \sum_{\kappa \in \mathcal{T}} \int_\kappa |\sqrt{a(x)} \nabla w(x)|^2 \text{d}x .$$

Finally, by the Cauchy-Schwarz inequality,

$$B_\delta(w, w) \geq \sum_{\kappa \in \mathcal{T}} \left[\frac{1}{2} \|\sqrt{\delta}(\vec{b} \cdot \nabla w)\|_{L^2(\kappa)}^2 - \|\sqrt{\delta}(\nabla \cdot (a \nabla w))\|_{L^2(\kappa)}^2 - \|\sqrt{\delta}cw\|_{L^2(\kappa)}^2 \right] .$$

Noting (3.1) and (3.2), this implies that

$$B_\delta(w, w) \geq \frac{1}{2} \sum_{\kappa \in \mathcal{T}} \left(\|\sqrt{\delta}(\vec{b} \cdot \nabla w)\|_{L^2(\kappa)}^2 - \|\sqrt{a} \nabla w\|_{L^2(\kappa)}^2 - \|\sqrt{\gamma}w\|_{L^2(\kappa)}^2 \right) .$$

Upon recalling the definition of the bilinear form $B_{\text{DG}}(\cdot, \cdot)$, we arrive at (3.3).

To complete the proof of the lemma, we note that since $\gamma > 0$ on each element κ in the subdivision \mathcal{T} , then $B_{\text{DG}}(w, w) > 0$ for all w in $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F}) \setminus \{0\}$, and hence we deduce the uniqueness of the solution u_{DG} . Further, since the linear space $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ is finite-dimensional, the existence of the solution to (2.11) follows from the fact that its homogeneous counterpart has the unique solution $u_{\text{DG}} \equiv 0$. \square

Our second result provides a bound on the global error $e = u - u_{\text{DG}}$. For simplicity, we shall assume that the entries of the matrix a are constant on each element $\kappa \in \mathcal{T}$ (with possible discontinuities across faces $e \in \mathcal{E}$). We quote the following result [5, 23].

Lemma 2 *Suppose that $u \in H^{k_\kappa}(\kappa)$, $k_\kappa \geq 0$, $\kappa \in \mathcal{T}$. Then, there exists $\Pi_{hp}u$ in the finite element space $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$, a constant C dependent on k_κ and the angle condition of κ , but independent of u , $h_\kappa = \text{diam}(\kappa)$ and p_κ , such that*

$$\|u - \Pi_{hp}u\|_{H^s(\kappa)} \leq C \frac{h_\kappa^{\tau_\kappa - s}}{p_\kappa^{k_\kappa - s}} \|u\|_{H^{k_\kappa}(\kappa)} , \quad (3.4)$$

where $0 \leq s \leq \tau_\kappa$ and $\tau_\kappa = \min(p_\kappa + 1, k_\kappa)$ for $\kappa \in \mathcal{T}$.

Our main result concerns the accuracy of the hp -DGFEM (2.11) and is stated in the next theorem. We recall the notation introduced earlier on: given that $\bar{a} = \|a\|_2$, with $\|\cdot\|_2$ denoting the matrix norm subordinate to the l^2 vector norm on \mathbf{R}^d , we let $\bar{a}_\kappa = \bar{a}|_\kappa$. Similarly, we define $\bar{b}_\kappa = \|\vec{b}\|_{L^\infty(\kappa)}$, $\bar{c}_\kappa = \|c\|_{L^\infty(\kappa)}$ and $\bar{\gamma}_\kappa = \|\gamma\|_{L^\infty(\kappa)}$.

Theorem 3 *In addition to the hypotheses of Theorem 1, let us assume that for any $\kappa \in \mathcal{T}$ such that $\bar{b}_\kappa \neq 0$,*

$$\delta(x) = \frac{1}{2} \min \left(\frac{h_\kappa^2}{C_{\text{inv}}^2 p_\kappa^4 \bar{a}_\kappa}, \frac{h_\kappa}{p_\kappa \bar{b}_\kappa}, \frac{\gamma}{\bar{c}_\kappa^2} \right), \quad x \in \kappa . \quad (3.5)$$

Then, the solution $u_{\text{DG}} \in S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ of (2.11) obeys the error bound

$$\begin{aligned} \|u - u_{\text{DG}}\|_{\text{DG}}^2 &\leq C \sum_{\kappa: \bar{b}_\kappa \neq 0} \left(\bar{a}_\kappa \frac{h_\kappa^{2(\tau_\kappa - 1)}}{p_\kappa^{2(k_\kappa - 2)}} + \bar{b}_\kappa \frac{h_\kappa^{2(\tau_\kappa - 1/2)}}{p_\kappa^{2(k_\kappa - 1/2)}} + \bar{\gamma}_\kappa \frac{h_\kappa^{2\tau_\kappa}}{p_\kappa^{2k_\kappa}} \right) \|u\|_{H^{k_\kappa}(\kappa)}^2 \\ &\quad + C \sum_{\kappa: \bar{b}_\kappa = 0} \left(\bar{a}_\kappa \frac{h_\kappa^{2(\tau_\kappa - 1)}}{p_\kappa^{2(k_\kappa - 3/2)}} + \bar{\gamma}_\kappa \frac{h_\kappa^{2\tau_\kappa}}{p_\kappa^{2k_\kappa}} \right) \|u\|_{H^{k_\kappa}(\kappa)}^2 , \end{aligned}$$

where $\tau_\kappa = \min(p_\kappa + 1, k_\kappa)$, and $u \in H^{k_\kappa}(\kappa)$ with $k_\kappa \geq 2$ when $\bar{b}_\kappa \neq 0$, $k_\kappa > 3/2$ when $\bar{b}_\kappa = 0$, for $\kappa \in \mathcal{T}$.

PROOF: Let us decompose $e = u - u_{\text{DG}}$ as $e = \eta + \xi$ where $\eta = u - \Pi_{hp}u$, $\xi = \Pi_{hp}u - u_{\text{DG}}$, and Π_{hp} is as in Lemma 2. Then, by virtue of Theorem 1,

$$|||\xi|||_{\text{DG}}^2 \leq B_{\text{DG}}(\xi, \xi) = B_{\text{DG}}(e - \eta, \xi) = -B_{\text{DG}}(\eta, \xi) ,$$

where we have used the Galerkin orthogonality property $B_{\text{DG}}(u - u_{\text{DG}}, \xi) = 0$ which follows from (2.11) with $v = \xi$ and the definition of the boundary value problem (2.1), (2.4), given the assumed smoothness of u . Thus, we deduce that

$$|||\xi|||_{\text{DG}}^2 \leq |B_a(\eta, \xi)| + |B_b(\eta, \xi)| + |B_s(\eta, \xi)| + |B_\delta(\eta, \xi)| . \quad (3.6)$$

Now, from (2.9) we have that

$$|B_s(\eta, \xi)| \leq |||\xi|||_{\text{DG}} \left(\int_{\Gamma_{\text{D}}} \sigma |\eta|^2 ds + \int_{\Gamma_{\text{int}}} \sigma [\eta]^2 ds \right)^{1/2} . \quad (3.7)$$

Next we consider $B_b(\eta, \xi)$. Upon integration by parts, we obtain

$$\begin{aligned} B_b(\eta, \xi) = & \sum_{\kappa} \int_{\kappa} (c - \nabla \cdot \vec{b}) \eta \xi dx - \sum_{\kappa \in \mathcal{T}} \int_{\kappa} \eta (\vec{b} \cdot \nabla \xi) dx + \sum_{\kappa \in \mathcal{T}} \int_{\partial_+ \kappa \cap \Gamma_+} (\vec{b} \cdot \vec{\mu}) \eta^+ \xi^+ ds \\ & + \sum_{\kappa \in \mathcal{T}} \int_{\partial_+ \kappa \setminus \Gamma_+} (\vec{b} \cdot \vec{\mu}) \eta^+ \xi^+ ds + \sum_{\kappa \in \mathcal{T}} \int_{\partial_- \kappa \setminus \Gamma_-} (\vec{b} \cdot \vec{\mu}) \eta^- \xi^+ ds . \end{aligned} \quad (3.8)$$

Denoting by S_1, \dots, S_5 the five terms on the right-hand side of (3.8), we find, after shifting the ‘indices’ in the summation in S_4 , that

$$|S_4 + S_5| \leq \sum_{\kappa \in \mathcal{T}} \left(\int_{\partial_- \kappa \setminus \Gamma_-} |\vec{b} \cdot \vec{\mu}| |\eta^-|^2 ds \right)^{1/2} \left(\int_{\partial_- \kappa \setminus \Gamma_-} |\vec{b} \cdot \vec{\mu}| [\xi]^2 ds \right)^{1/2} .$$

Also, we note that elements $\kappa \in \mathcal{T}$ with $\bar{b}_\kappa = 0$ can be omitted from the summation in term S_2 . Thus, after multiplying and dividing by $\sqrt{\gamma}$ and $\sqrt{\delta}$ under the integral signs in S_1 and S_2 , respectively, (3.8) yields

$$\begin{aligned} |B_b(\eta, \xi)| \leq & C |||\xi|||_{\text{DG}} \left(\sum_{\kappa : \bar{b}_\kappa \neq 0} \|(1 + \delta^{-1/2}) \eta\|_{L^2(\kappa)}^2 + \sum_{\kappa \in \mathcal{T}} \int_{\partial_+ \kappa \cap \Gamma_+} |\vec{b} \cdot \vec{\mu}| |\eta^+|^2 ds \right. \\ & \left. + \sum_{\kappa \in \mathcal{T}} \int_{\partial_- \kappa \setminus \Gamma_-} |\vec{b} \cdot \vec{\mu}| |\eta^-|^2 ds \right)^{1/2} , \end{aligned} \quad (3.9)$$

where C is a generic positive constant, as in the statement of the theorem.

Next, we consider the term $B_a(\eta, \xi)$:

$$|B_a(\eta, \xi)| \leq I + II + III ,$$

where

$$I \equiv \left| \sum_{\kappa \in \mathcal{T}} \int_{\kappa} a \nabla \eta \cdot \nabla \xi dx \right| , \quad II \equiv \left| \int_{\Gamma_{\text{D}}} \{ \eta ((a \nabla \xi) \cdot \vec{\nu}) - ((a \nabla \eta) \cdot \vec{\nu}) \xi \} ds \right| ,$$

$$III \equiv \left| \int_{\Gamma_{\text{int}}} \{[\eta] \langle (a \nabla \xi) \cdot \vec{\nu} \rangle - \langle (a \nabla \eta) \cdot \vec{\nu} \rangle [\xi] \} ds \right|.$$

Now, we have that

$$\begin{aligned} I^2 &\leq |||\xi|||_{\text{DG}}^2 \sum_{\kappa \in \mathcal{T}} \|\sqrt{a} \nabla \eta\|_{L^2(\kappa)}^2, \\ II^2 &\leq C |||\xi|||_{\text{DG}}^2 \sum_{\kappa : \partial\kappa \cap \Gamma_D \neq \emptyset} \left(\frac{\bar{a}_\kappa p_\kappa^2}{h_\kappa} \|\eta\|_{L^2(\partial\kappa \cap \Gamma_D)}^2 + \bar{a}_\kappa h_\kappa \|\nabla \eta\|_{L^2(\partial\kappa \cap \Gamma_D)}^2 \right), \\ III^2 &\leq C |||\xi|||_{\text{DG}}^2 \sum_{\kappa : \partial\kappa \cap \Gamma = \emptyset} \left(\frac{\bar{a}_\kappa p_\kappa^2}{h_\kappa} \|[\eta]\|_{L^2(\partial\kappa)}^2 + \bar{a}_\kappa h_\kappa \|\nabla \eta\|_{L^2(\partial\kappa)}^2 \right). \end{aligned}$$

Collecting the bounds on the terms I , II and III gives,

$$\begin{aligned} |B_a(\eta, \xi)| &\leq C |||\xi|||_{\text{DG}} \left(\sum_{\kappa \in \mathcal{T}} \|\sqrt{a} \nabla \eta\|_{L^2(\kappa)}^2 \right. \\ &\quad + \sum_{\kappa : \partial\kappa \cap \Gamma_D \neq \emptyset} \left(\frac{\bar{a}_\kappa p_\kappa^2}{h_\kappa} \|\eta\|_{L^2(\partial\kappa \cap \Gamma_D)}^2 + \bar{a}_\kappa h_\kappa \|\nabla \eta\|_{L^2(\partial\kappa \cap \Gamma_D)}^2 \right) \\ &\quad \left. + \sum_{\kappa : \partial\kappa \cap \Gamma = \emptyset} \left(\frac{\bar{a}_\kappa p_\kappa^2}{h_\kappa} \|[\eta]\|_{L^2(\partial\kappa)}^2 + \bar{a}_\kappa h_\kappa \|\nabla \eta\|_{L^2(\partial\kappa)}^2 \right) \right)^{1/2}. \end{aligned} \quad (3.10)$$

Finally, for $B_\delta(\eta, \xi)$ we have the bound

$$|B_\delta(\eta, \xi)| \leq |||\xi|||_{\text{DG}} \left(\sum_{\kappa : \bar{b}_\kappa \neq 0} \|\sqrt{\delta} \mathcal{L} \eta\|_{L^2(\kappa)}^2 \right)^{1/2}. \quad (3.11)$$

The required result now follows by noting that

$$|||u - u_{\text{DG}}|||_{\text{DG}} \leq |||\eta|||_{\text{DG}} + |||\xi|||_{\text{DG}},$$

inserting the estimates (3.7), (3.9), (3.10) and (3.11) into (3.6) to bound $|||u - u_{\text{DG}}|||_{\text{DG}}$ in terms of $|||\eta|||_{\text{DG}}$ and other norms of η , and applying Lemma 2, together with the Trace Inequality to estimate norms over e and $\partial\kappa$ in terms of norms of η over κ , $\kappa \in \mathcal{T}$. \square

We note that in the purely hyperbolic case of $a \equiv 0$ the error bound in Theorem 3 collapses to the hp -optimal error bound $O(h^{\tau-1/2}/p^{k-1/2})$ in the DG-norm established in [14], which represents a generalisation of the optimal h -version bound for the DGFEM (see [16]) to the hp -version. In fact, for $a \equiv 0$ the error bound of Theorem 3 is by $1/2$ a p -order sharper than the corresponding estimate of Bey and Oden [8], except that there the streamline-diffusion parameter was $\delta = h/p^2$, while in the case of $a \equiv 0$ Theorem 3 corresponds to $\delta = h/p$. We note in this respect that the error bound of [8] may be reproduced even with $\delta = 0$, i.e. with less damping in the streamwise direction than required in [8]; see [25]. For further developments in this direction, we refer to [14, 25].

4 A posteriori error analysis

For the second half of this paper, we turn our attention to the subject of *a posteriori* error analysis of first-order hyperbolic problems, corresponding to $a_{ij} \equiv 0$ for $i, j = 1, \dots, d$. In particular, using the approach in [7], we discuss the question of error estimation for linear functionals, such as the outflow flux and the local average of the solution. For simplicity, we restrict ourselves to the case when the streamline-diffusion stabilisation δ is set to zero; the case when $\delta > 0$ may be treated analogously, cf. [13]. Under these assumptions,

$$B_{\text{DG}}(\cdot, \cdot) \equiv B_b(\cdot, \cdot) \quad \text{and} \quad \ell_{\text{DG}}(\cdot) \equiv \ell_b(\cdot) ,$$

where $B_b(\cdot, \cdot)$ and $\ell_b(\cdot)$ are as defined in (2.6) and (2.7), respectively.

Given a linear functional $J(\cdot)$, our aim is to control the discretisation error between the true value $J(u)$, based on the analytical solution u to (2.1), and the actual computed value $J(u_{\text{DG}})$. The proceeding error analysis is based on a hyperbolic duality argument; for full details and further numerical experiments, see [15]. To this end, we introduce the following *dual* or *adjoint* problem: find z in $H(\mathcal{L}^*, \Omega)$ such that

$$B_{\text{DG}}(w, z) = J(w) \quad \forall w \in H(\mathcal{L}, \Omega) , \quad (4.1)$$

where $H(\mathcal{L}, \Omega)$, resp. $H(\mathcal{L}^*, \Omega)$, denotes the graph space of the first-order operator $\mathcal{L} \equiv \mathcal{L}_b$, resp. \mathcal{L}^* . Here, we assume that (4.1) has a unique solution; the validity of this assumption depends on the precise definition of the linear functional $J(\cdot)$ under consideration. In the case when $J(\cdot)$ represents the (weighted) normal flux through the outflow boundary Γ_+ , i.e.

$$J(w) \equiv N_\psi(w) = \int_{\Gamma_+} (\vec{b} \cdot \vec{\mu}) w \psi \, ds , \quad (4.2)$$

z is the (unique) solution to the following partial differential equation: find z in $H(\mathcal{L}^*, \Omega)$ such that

$$\mathcal{L}^* z \equiv -\nabla \cdot (\vec{b}z) + cz = 0 , \quad x \in \Omega , \quad z = \psi , \quad x \in \Gamma_+ . \quad (4.3)$$

Other examples include the mean flow of the field u over the computational domain Ω or some compact subset of Ω .

For each element κ in the mesh \mathcal{T} , we define the *internal residual* $r_{h,p}$ and the *boundary residual* $r_{h,p}^-$ by

$$r_{h,p}|_\kappa = (f - \mathcal{L}u_{\text{DG}})|_\kappa \quad \text{and} \quad r_{h,p}^-|_{\partial_- \kappa \cap \Gamma_-} = (g - u_{\text{DG}}^+)|_{\partial_- \kappa \cap \Gamma_-} , \quad (4.4)$$

respectively. With this notation we have the following general result.

Theorem 4 *Let u and u_{DG} denote the solutions of (2.1) and (2.11), respectively, and suppose that the dual solution z satisfies (4.1). Then, the following a posteriori error bound holds:*

$$|J(u) - J(u_{\text{DG}})| \leq \epsilon(u_{\text{DG}}, h, p, z, z_{h,p}) \equiv \sum_{\kappa \in \mathcal{T}} \eta_\kappa , \quad (4.5)$$

where

$$\eta_\kappa = |(r_{h,p}, z - z_{h,p})_\kappa + ((\vec{b} \cdot \vec{\mu})[u_{\text{DG}}], (z - z_{h,p})^+)_{\partial_- \kappa \setminus \Gamma_-} - ((\vec{b} \cdot \vec{\mu})r_{h,p}^-, (z - z_{h,p})^+)_{\partial_- \kappa \cap \Gamma_-}|$$

and $z_{h,p}$ belongs to the finite element space $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$.

PROOF: Choosing $w = u - u_{\text{DG}}$ in (4.1) and exploiting the Galerkin orthogonality property of the hp -DGFEM (cf. proof of Theorem 3), we deduce that

$$J(u) - J(u_{\text{DG}}) = J(u - u_{\text{DG}}) = B_{\text{DG}}(u - u_{\text{DG}}, z) = B_{\text{DG}}(u - u_{\text{DG}}, z - z_{h,p}) .$$

Recalling the definition of the bilinear form $B_{\text{DG}}(\cdot, \cdot)$ and applying the triangle inequality gives the desired result. \square

While the residual terms $r_{h,p}$ and $r_{h,p}^-$ and the ‘jump’ term $[u_{\text{DG}}]$ are easily evaluated once the numerical solution u_{DG} has been computed, the calculation of the corresponding ‘weights’ involving the dual solution z requires special care. As in [13], z will be estimated by numerically solving the dual problem (4.1); this will be discussed in detail in Section 5.1. First, however, in the case when the functional of interest $J(\cdot)$ is defined to be the mean flow of the field u over Ω , we derive an *a posteriori* bound on the error $u - u_{\text{DG}}$ in negative Sobolev norms, cf. [13, 24], for example.

Theorem 5 *Let u and u_{DG} denote the solutions of (2.1) and (2.11), respectively. Then there exists a positive constant C , dependent only on the dimension d , the shape regularity of \mathcal{T} and m , $m > 0$, such that*

$$\begin{aligned} \|u - u_{\text{DG}}\|_{H^{-m}(\Omega)} \leq C & \left[\left(\sum_{\kappa \in \mathcal{T}} \frac{h_\kappa^{2\tau_\kappa}}{p_\kappa^{2m}} \|r_{h,p}\|_{L_2(\kappa)}^2 \right)^{1/2} + \left(\sum_{\kappa \in \mathcal{T}} \frac{h_\kappa^{2\tau_\kappa-1}}{p_\kappa^{2m-1}} \|[u_{\text{DG}}]\|_{\partial_- \kappa \setminus \Gamma_-}^2 \right)^{1/2} \right. \\ & \left. + \left(\sum_{\kappa \in \mathcal{T}} \frac{h_\kappa^{2\tau_\kappa-1}}{p_\kappa^{2m-1}} \|r_{h,p}^-\|_{\partial_- \kappa \cap \Gamma_-}^2 \right)^{1/2} \right] , \end{aligned}$$

where $\tau_\kappa = \min(p_\kappa + 1, m)$ for all κ in \mathcal{T} .

PROOF: The proof is based on a duality argument using the Galerkin orthogonality of the hp -DGFEM, together with stability bounds for the dual problem, see [15]. \square

We end this section by stating an *a priori* bound on the error in the computed functional in terms of Sobolev norms of the analytical solution u and the dual solution z which indicates the expected rate of convergence for $|J(u) - J(u_{\text{DG}})|$ as the finite element space is enriched, i.e. as $h \rightarrow 0$ and $p \rightarrow \infty$. This will play a crucial role in the design of an hp -adaptive algorithm for automatically controlling the error in the computed functional, see Section 5.2 below. To this end, we assume for the moment that

$$\vec{b} \in [S^1(\Omega, \mathcal{T}, \mathbf{F}) \cap C(\Omega)]^d , \quad c \in S^0(\Omega, \mathcal{T}, \mathbf{F}) , \quad f \in S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F}) . \quad (4.6)$$

Theorem 6 *Let u and u_{DG} denote the solutions of (2.1) and (2.11), respectively. Given that $u|_{\kappa} \in H^{k_{\kappa}}(\kappa)$, $k_{\kappa} \geq 1$, and $z|_{\kappa} \in H^{l_{\kappa}}(\kappa)$, $l_{\kappa} \geq 1$, for all κ in \mathcal{T} , we have*

$$|J(u) - J(u_{\text{DG}})|^2 \leq C \sum_{\kappa \in \mathcal{T}} \frac{h_{\kappa}^{2\tau_{\kappa}-1}}{p_{\kappa}^{2k_{\kappa}-2}} \|u\|_{H^{k_{\kappa}}(\kappa)}^2 \cdot \sum_{\kappa \in \mathcal{T}} \frac{h_{\kappa}^{2\theta_{\kappa}-1}}{p_{\kappa}^{2l_{\kappa}-2}} \|z\|_{H^{l_{\kappa}}(\kappa)}^2, \quad (4.7)$$

where $\tau_{\kappa} = \min(p_{\kappa} + 1, k_{\kappa})$ and $\theta_{\kappa} = \min(p_{\kappa} + 1, l_{\kappa})$ for all $\kappa \in \mathcal{T}$. Here, C is a positive constant, dependent only on d , the shape regularity of \mathcal{T} and k_{κ} and l_{κ} , $\kappa \in \mathcal{T}$.

PROOF: See [15] for details. \square

For uniform orders, $p_{\kappa} = p$, $k_{\kappa} = k \geq 1$, $l_{\kappa} = l \geq 1$, and $h_{\kappa} = h$ for all κ in \mathcal{T} , we have

$$|J(u) - J(u_{\text{DG}})| \leq C \frac{h^{\tau+\theta-1}}{p^{k+l-1}} p \|u\|_{H^k(\Omega)} \|z\|_{H^l(\Omega)}, \quad (4.8)$$

where $\tau = \min(p + 1, k)$ and $\theta = \min(p + 1, l)$. Here, the bound (4.8) is optimal in h and suboptimal in p by one order; in the case of fixed p , (4.8) reduces the optimal h -convergence error bound proved in [13] for a stabilised continuous approximation to u . From (4.8) we may deduce the following *a priori* error bound

$$\|u - u_{\text{DG}}\|_{H^{-m}(\Omega)} \leq C \frac{h^{\tau+\theta-1}}{p^{k+m-1}} p \|u\|_{H^k(\Omega)}, \quad (4.9)$$

where $\tau = \min(p + 1, k)$ and $\theta = \min(p + 1, m)$. In the presence of streamline-diffusion stabilisation, with stabilisation parameter $\delta = h/p$, the bounds (4.8) and (4.9) can be sharpened to ones that are simultaneously optimal in both h and p .

Finally, we note that the dependence of the constant C appearing in the *a priori* bound (4.7) on the regularity of the primal solution u and the dual solution z may be made explicit using the approximation results derived in [23]. In particular, this allows us to deduce that the error in the computed functional $J(\cdot)$ decays exponentially as $p \rightarrow \infty$ if *either* u or z are elementwise analytic, cf. [15]; this will be demonstrated in Section 5.3.

5 Numerical Implementation

5.1 Numerical approximation of the dual solution

To ensure that the *a posteriori* error bound stated in Theorem 4 is fully computable, the dual solution z must be numerically approximated. In this section we describe a DGFEM for this purpose. As stated in Section 4, the particular form of the dual problem is dependent on the functional under consideration. For simplicity, let us suppose that $J(\cdot) = N_{\psi}(\cdot)$, i.e. J represents the outflow normal flux, cf. (4.2). In this case the dual solution z satisfies (4.3) for a given weight function ψ .

As in Section 2.1, we define $\tilde{\mathbf{P}}(\Omega, \tilde{\mathcal{T}}, \tilde{\mathbf{F}})$ to be the finite element space consisting of piecewise polynomials of degree $\tilde{\mathbf{p}}|_{\tilde{\kappa}} = \tilde{p}_{\tilde{\kappa}}$ on a mesh $\tilde{\mathcal{T}}$ consisting of shape regular elements

$\tilde{\kappa}$ of size $\tilde{h}_{\tilde{\kappa}}$. With $\partial_+ \tilde{\kappa}$ defined as in (2.5), we introduce the bilinear form and linear functional

$$\begin{aligned}\tilde{B}_{\text{DG}}(w, v) &= \sum_{\tilde{\kappa} \in \tilde{\mathcal{T}}} \int_{\tilde{\kappa}} \mathcal{L}^* w v dx + \sum_{\tilde{\kappa} \in \tilde{\mathcal{T}}} \int_{\partial_+ \tilde{\kappa} \setminus \Gamma_+} (\vec{b} \cdot \vec{\mu}) [w] v^+ ds \\ &\quad + \sum_{\tilde{\kappa} \in \tilde{\mathcal{T}}} \int_{\partial_+ \tilde{\kappa} \cap \Gamma_+} (\vec{b} \cdot \vec{\mu}) w^+ v^+ ds , \\ \tilde{\ell}_{\text{DG}}(v) &= \sum_{\tilde{\kappa} \in \tilde{\mathcal{T}}} \int_{\partial_+ \tilde{\kappa} \cap \Gamma_+} (\vec{b} \cdot \vec{\mu}) \psi v^+ ds ,\end{aligned}$$

respectively. The hp -DGFEM approximation of (4.3) is defined as follows: find $\tilde{z}_{\text{DG}} \in \tilde{S}^{\mathbf{p}}(\Omega, \tilde{\mathcal{T}}, \tilde{\mathbf{F}})$ such that

$$\tilde{B}_{\text{DG}}(\tilde{z}_{\text{DG}}, v) = \tilde{\ell}_{\text{DG}}(v) \quad \forall v \in \tilde{S}^{\mathbf{p}}(\Omega, \tilde{\mathcal{T}}, \tilde{\mathbf{F}}) . \quad (5.1)$$

5.2 Adaptive algorithm

For a user-defined tolerance TOL, we now consider the problem of designing the hp -finite element space $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ such that

$$|J(u) - J(u_{\text{DG}})| \leq \text{TOL} , \quad (5.2)$$

subject to the constraint that the total number of degrees of freedom in $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ is minimised. To ensure that (5.2) holds, we use the *a posteriori* error bound (4.5) to construct $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ such that

$$\epsilon(u_{\text{DG}}, h, p, z, z_{h,p}) \leq \text{TOL} . \quad (5.3)$$

The stopping criterion (5.3) is enforced by equidistributing $\epsilon|_{\kappa} \equiv \eta_{\kappa}$ over the elements κ in the mesh \mathcal{T} . Thus, we insist that

$$\eta_{\kappa} \approx \text{TOL}/N \quad (5.4)$$

holds for each κ in \mathcal{T} ; here, N denotes the number of elements in the mesh \mathcal{T} .

Thereby, each of the elements in the mesh is flagged for either refinement or derefinement to ensure that the equidistribution principle (5.4) holds. Once an element κ has been flagged a decision must be made whether the local mesh size h_{κ} or the local degree of the approximating polynomial p_{κ} should be adjusted accordingly. Let us first deal with refinement, i.e. when the local error estimator η_{κ} is larger than the ‘localised-tolerance’ TOL/N . Clearly, if the error in the functional is locally ‘smooth’, then p -enrichment will be more effective than h -refinement, since the error will be expected to decay quickly within the current element κ as p_{κ} is increased. However, if the error in the functional has low regularity within the element κ , then h -refinement will be performed. Thus, regions in the computational domain where the error is locally non-smooth are isolated from smooth regions, thereby reducing the influence of singularities/shocks as well as making p -enrichment more effective.

To ensure that the desired level of accuracy is achieved efficiently, an automatic procedure for deciding when to h - or p -refine must be implemented. To this end, we first compute the local error indicator η_κ on each element κ in the mesh \mathcal{T} using both a p_κ and a $p_\kappa - 1$ representation for u_{DG} . Thereby, assuming that $\eta_\kappa(p_\kappa - 1) \neq 0$, the perceived smoothness of the local error may be estimated using the ratio

$$\rho_\kappa = \eta_\kappa(p_\kappa) / \eta_\kappa(p_\kappa - 1) ; \quad (5.5)$$

here, we have written $\eta_\kappa(p_\kappa)$ to emphasise the dependence of the local error indicator η_κ on the local degree p_κ of the approximating polynomial, cf. Adjrid *et al.* [1] and Gui & Babuška [12], for example. If $\rho_\kappa \leq \gamma$, $0 < \gamma < 1$, the error is decreasing as the polynomial degree is increased, indicating that p -enrichment should be performed. On the other hand, $\rho_\kappa > \gamma$ means that the element κ should be locally subdivided. The number γ is referred to as the *type-parameter* [12]. Clearly, the choice of γ is critical to the success of this algorithm and will depend on the asymptotic behaviour of the quantity of interest. Instead of assigning an *ad hoc* value to the type parameter γ , we use ρ_κ together with the *a priori* error bound (4.7) to directly estimate the local regularities k_κ and l_κ of the primal and dual solutions, respectively, on each element κ in \mathcal{T} . More precisely, motivated by (4.8), we assume that on a given element κ in \mathcal{T}

$$\eta_\kappa = \epsilon(u_{\text{DG}}, h_\kappa, p_\kappa, z, z_{h,p})|_\kappa \approx C_\kappa p_\kappa^{-k_\kappa - l_\kappa + 1} .$$

Thus, we have that

$$k_\kappa + l_\kappa = \log(\rho_\kappa) / \log((p_\kappa - 1)/p_\kappa) + 1 .$$

Ideally, we would like to know k_κ and l_κ individually. The dual regularity l_κ may be estimated by calculating the $L^2(\kappa)$ norm of the error between the projection of \tilde{z}_{DG} onto the finite element spaces $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ and $S^{\mathbf{p}-1}(\Omega, \mathcal{T}, \mathbf{F})$, together with the approximation result (3.4). Once, both k_κ and l_κ have been determined on element κ , then κ is p -enriched if either k_κ or l_κ is larger than $p_\kappa + 1$; otherwise the element is subdivided. For computational simplicity, only one hanging node is allowed on each side of a given element κ , though no restriction on the difference between the polynomial degrees on neighbouring elements is imposed. We note that this approach has been developed by Ainsworth & Senior [2] in the context of norm control for second-order elliptic problems.

On the other hand, if an element has been flagged for derefinement, then the strategy implemented here is to coarsen the mesh in smooth low-error-regions and decrease the degree of the approximating polynomial in non-smooth low-error-regions, cf. [1]. To this end, we again compute the local regularities k_κ and l_κ of the primal and dual solutions, respectively, on each element κ in \mathcal{T} as described above. The element κ is then coarsened if either k_κ or l_κ is larger than $p_\kappa + 1$, otherwise the degree p_κ is reduced by one.

For the practical implementation of this adaptive algorithm, the dual solution z will be numerically approximated as outlined in Section 5.1. Here, we write $\hat{\epsilon}$ in lieu of $\epsilon(u_{\text{DG}}, h, p, \tilde{z}_{\text{DG}}, \tilde{z}_{h,p})$, where \tilde{z}_{DG} denotes the numerical approximation to z defined by (5.1) and $\tilde{z}_{h,p}$ denotes the L_2 -projection of \tilde{z}_{DG} onto the finite element space $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ used to calculate u_{DG} . Furthermore, the finite element space $\tilde{S}^{\mathbf{p}}(\Omega, \tilde{\mathcal{T}}, \tilde{\mathbf{F}})$ used to approximate

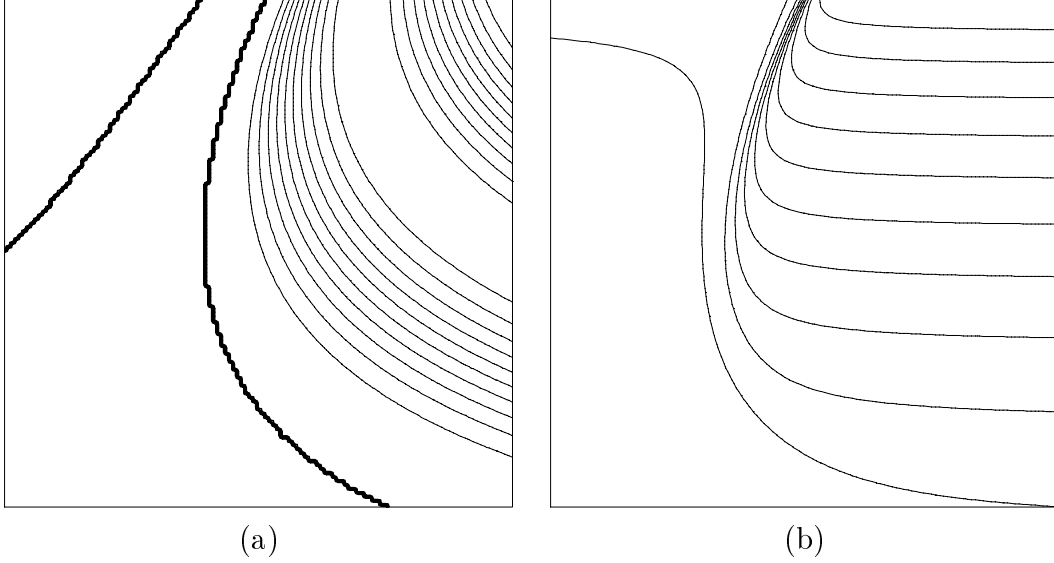


Figure 1: Piecewise bilinear interpolant on a 129×129 mesh of the analytical solution to: (a) Primal problem; (b) Dual Problem.

the dual solution z will be constructed adaptively at the same time as $S^{\mathbf{P}}(\Omega, \mathcal{T}, \mathbf{F})$. For this purpose, we define the following error indicator for the dual approximation

$$\eta_{-1, \tilde{\kappa}} = (\tilde{h}_{\tilde{\kappa}}/\tilde{p}_{\tilde{\kappa}}) \|\mathcal{L}^* \tilde{z}_{\text{DG}}\|_{L_2(\tilde{\kappa})} + (\tilde{h}_{\tilde{\kappa}}/\tilde{p}_{\tilde{\kappa}})^{1/2} (\|\tilde{z}_{\text{DG}}\|_{\partial_+ \tilde{\kappa} \setminus \Gamma_+} + \|\psi - \tilde{z}_{\text{DG}}\|_{\partial_+ \tilde{\kappa} \cap \Gamma_+}) ,$$

which results from controlling the $H^{-1}(\Omega)$ norm of the error $z - \tilde{z}_{\text{DG}}$, cf. Theorem 5. The hp -adaptive algorithm for the dual problem will be based on the *fixed fraction strategy*. Once the elements have been flagged for refinement/derefinement, $\tilde{h}_{\tilde{\kappa}}$ and $\tilde{p}_{\tilde{\kappa}}$ are altered accordingly by estimating the local regularity $\tilde{l}_{\tilde{\kappa}}$ of the dual solution on the dual mesh $\tilde{\mathcal{T}}$ as above by calculating $\eta_{-1, \tilde{\kappa}}$ using a $\tilde{p}_{\tilde{\kappa}}$ and $\tilde{p}_{\tilde{\kappa}} - 1$ representation of \tilde{z}_{DG} , together with the *a priori* error bound (4.9).

5.3 Example

Here we consider a compressible hyperbolic problem subject to discontinuous inflow boundary condition, with $\vec{b} = (2y^2 - 4x + 1, 1 + y)$, $c = 0$ and $f = 0$. The characteristics enter the computational domain Ω from three sides of Γ , namely from $x = 0$, $y = 0$ and $x = 1$, and exit Ω through $y = 1$. Thus, we may prescribe

$$u(x, y) = \begin{cases} 0 & \text{for } x = 0, 0.5 < y \leq 1, \\ 1 & \text{for } x = 0, 0 \leq y \leq 0.5, \\ 1 & \text{for } 0 \leq x \leq 0.75, y = 0, \\ 0 & \text{for } 0.75 < x \leq 1, y = 0, \\ \sin^2(\pi y) & \text{for } x = 1, 0 \leq y \leq 1. \end{cases}$$

We define the weight function ψ in the functional $N_\psi(\cdot)$, cf. (4.2), by

$$\psi = 2 + \tanh((x - 1/2)/\varepsilon) \quad \text{for } 0 \leq x \leq 1, y = 1,$$

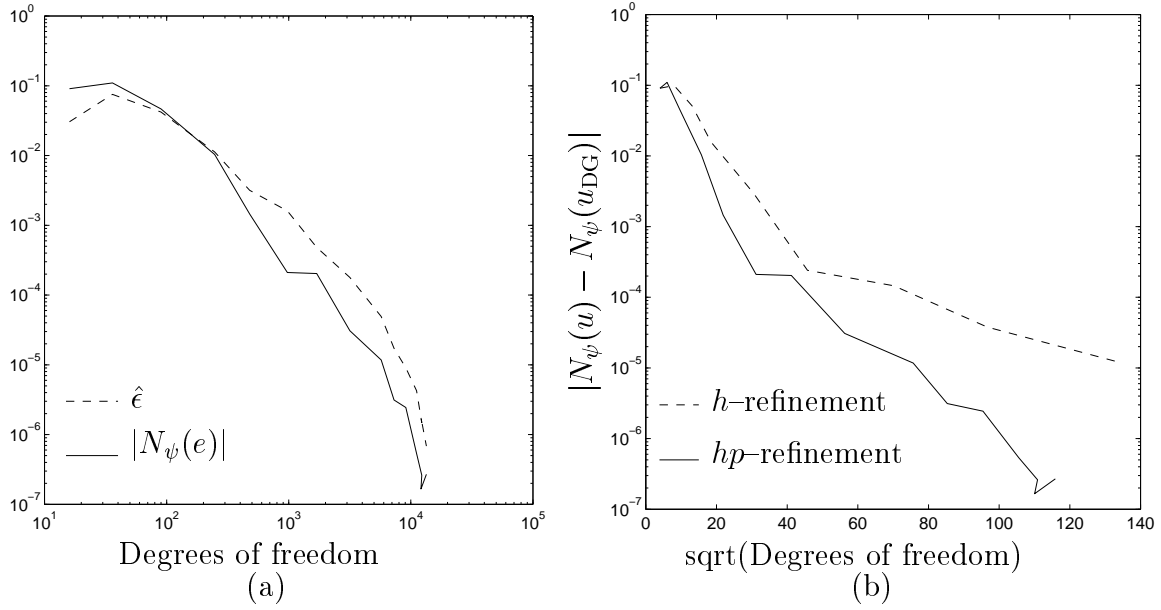


Figure 2: (a) $|N_\psi(u) - N_\psi(u_{DG})|$ and $\hat{\epsilon}$ using hp -refinement; (b) True error in the functional using both h - and hp -refinement.

where $\varepsilon = 10^{-2}$. Thereby, the true value of the outward normal flux is $N_\psi(u) = 2.0115$. The analytical solutions to both the primal and dual problems are shown in Figure 1.

In Figure 2 we show the performance of the adaptive algorithm described in Section 5.2 for $\text{TOL} = 10^{-6}$; we note that this level of accuracy may be far beyond what is of practical importance, but is chosen to illustrate that the true error and the bound $\hat{\epsilon}$ exhibit the same asymptotic behaviour as the finite element space $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ is enriched. In Figure 2(a) we plot the error in the computed functional $N_\psi(\cdot)$, together with the error bound $\hat{\epsilon}$. Here, we see that while on very coarse meshes $\hat{\epsilon}$ slightly underestimates the true error in the functional, as the finite element space is enriched the error bound over-estimates $|N_\psi(u) - N_\psi(u_{DG})|$ by a consistent factor. Furthermore, in Figure 2(b) we compare the true error in the functional using both h - and hp -adaptive refinement. We have plotted the error against the square-root of the number of degrees of freedom on a linear-log scale. While the error $|N_\psi(u) - N_\psi(u_{DG})|$ using h -refinement ‘tails-off’ as $S^{\mathbf{p}}(\Omega, \mathcal{T}, \mathbf{F})$ is enriched, we see that after the initial transient, the error in the computed functional using hp -refinement becomes a straight line, thereby indicating exponential convergence. We note that the slight ‘dip’ and the subsequent rise in the true error in the functional observed at the end of the hp -refinement algorithm, cf. Figure 2, is attributed to the fact that once the desired tolerance has almost been achieved, the last couple of iterations of the adaptive algorithm attempt to equidistribute the local error indicators η_κ over the elements κ in the computational mesh \mathcal{T} .

Finally in Figures 3 and 4 we show primal and dual meshes after 8 and 15 adaptive mesh refinements, respectively. For clarity, in each case we show the h -mesh alone, as well as the corresponding distribution of the polynomial degree and the percentage of elements with

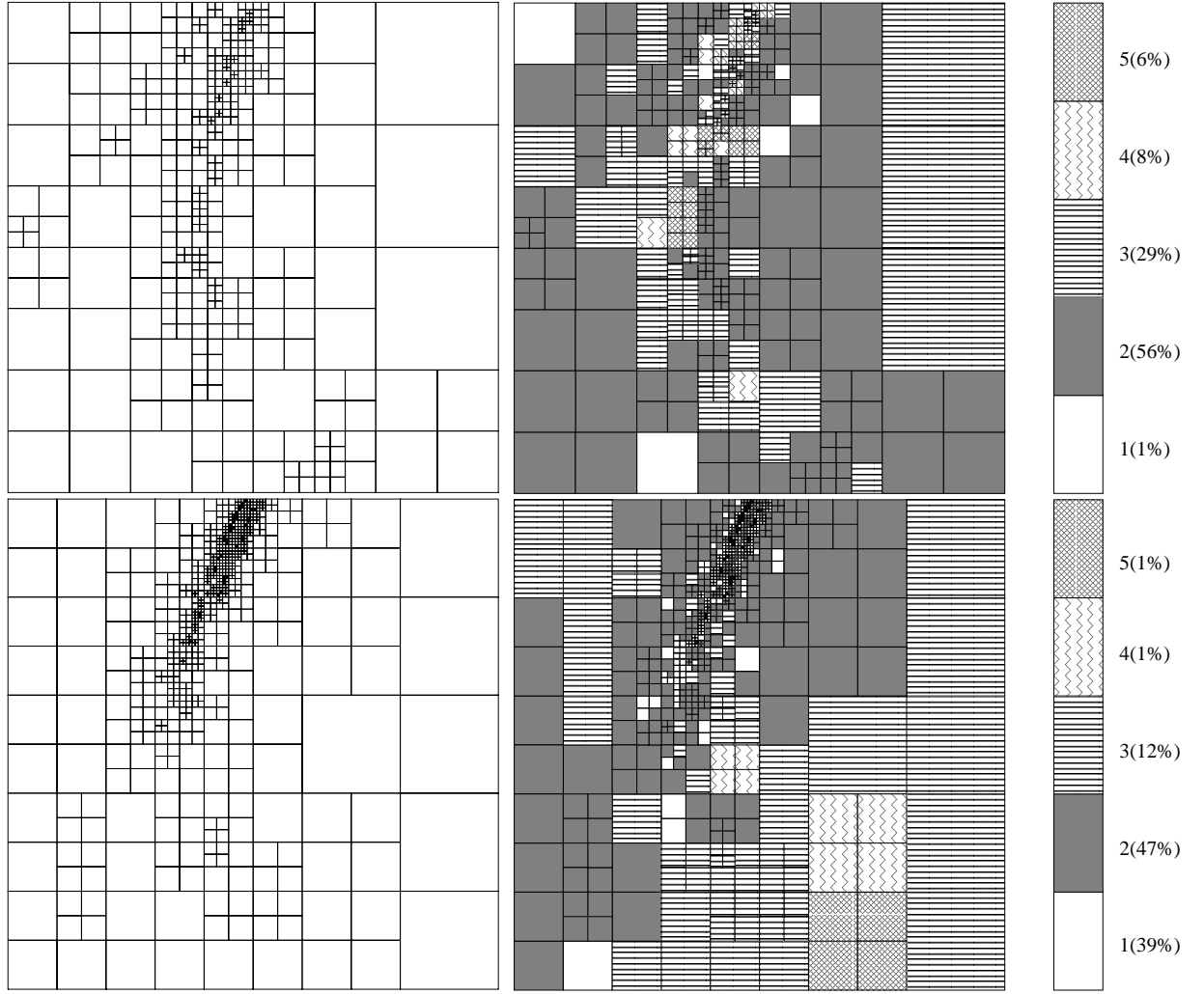


Figure 3: Mesh 9: Primal (top: 412 elements, 531 nodes and 5723 DOF) and Dual (bottom: 865 elements, 1064 nodes and 7037 DOF) h - and hp -meshes

that degree. From Figure 3, we see that the elements in the primal mesh have been refined along the first discontinuity emanating from $(x, y) = (0.75, 0)$, since the dual solution has a layer in this region as well. In contrast, elements lying on the second discontinuity in the primal problem, which emanates from $(x, y) = (0, 0.5)$ have been less refined since the dual solution is smooth here. Furthermore, the mesh for the dual solution is concentrated within the steep layer in the weight function ψ ; the inherent smoothing in the dual problem introduced by the compressible nature of \vec{b} leads to p refinement in this layer as the flow moves away from Γ_+ . The same behaviour is observed in Figure 4 for the primal and dual solutions.

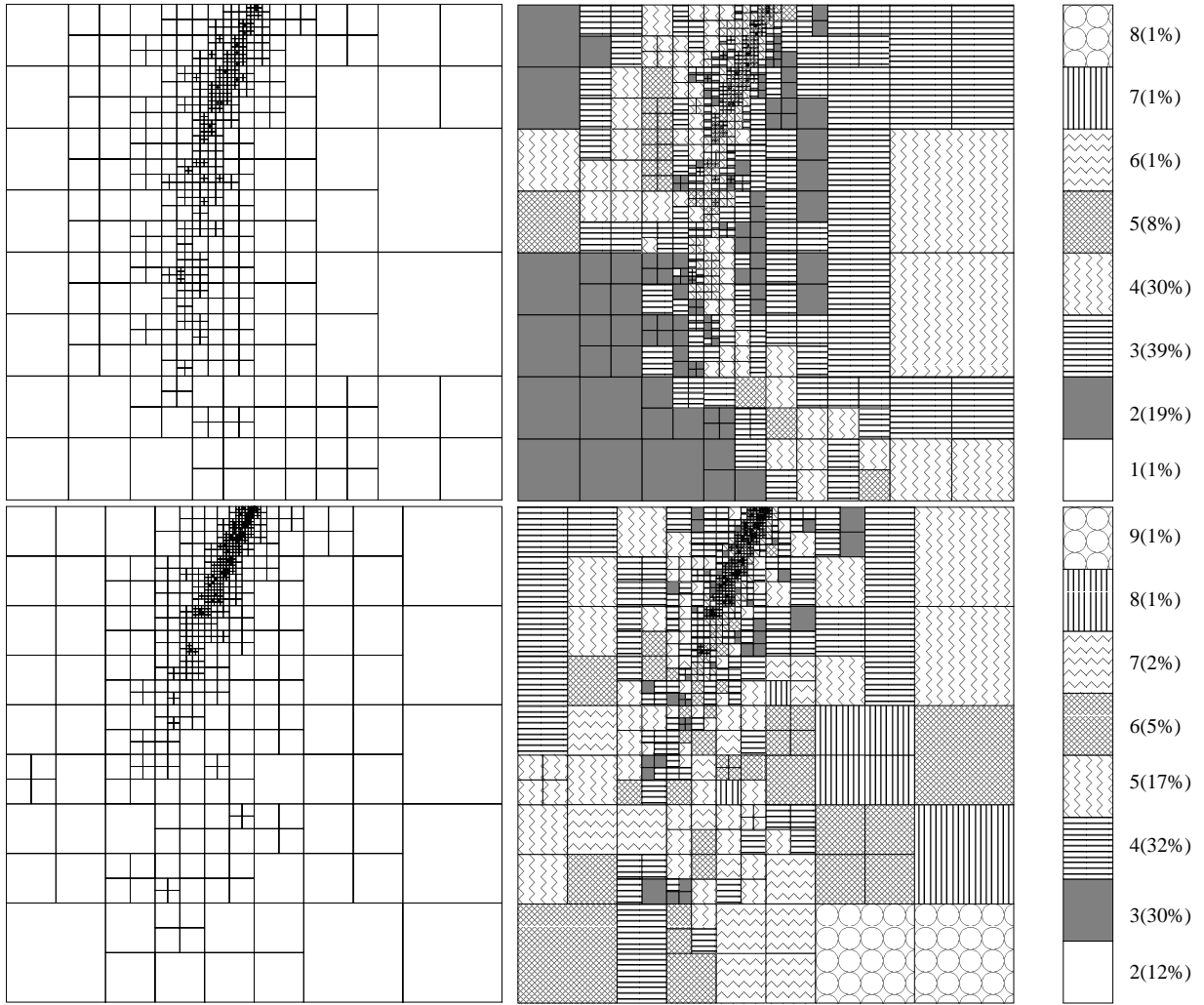


Figure 4: Mesh 16: Primal (top: 700 elements, 868 nodes and 13429 DOF) and Dual (bottom: 961 elements, 1210 nodes and 23601 DOF) h - and hp -meshes

Acknowledgement

E. Süli and P. Houston acknowledge the financial support of the EPSRC (Grant GR/K76221). The authors also wish to express their thanks to Mr. Mark Embree for writing the graphics packages used to display the hp -finite element meshes in this paper.

References

- [1] S. Adjerid, M. Aiffa and J.E. Flaherty (1998). Computational methods for singularly perturbed systems. In: J. Cronin and R.E. O'Malley, editors, *Singular Perturbation Concepts of Differential Equations*, AMS, Providence.

- [2] M. Ainsworth and B. Senior (1998). An adaptive refinement strategy for hp -finite element computations. *Appl. Numer. Maths.* 26:165–178.
- [3] D.N. Arnold (1982) An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19:742–760.
- [4] I. Babuška (1972) The finite element method with lagrangian multipliers. *Numer. Math.*, 20:179–192.
- [5] I. Babuška and M. Suri (1987). The hp -Version of the Finite Element Method with quasiuniform meshes. *M²AN Mathematical Modelling and Numerical Analysis*, 21:199–238.
- [6] C. Baumann (1997). An hp -Adaptive Discontinuous Galerkin FEM for Computational Fluid Dynamics. Doctoral Dissertation. TICAM, UT Austin, Texas.
- [7] R. Becker and R. Rannacher (1998). Weighted a posteriori error control in FE methods, ENUMATH-95, Paris, 18-22 Sept., 1995, in Proc. ENUMATH-97 (H.G. Bock, et al., eds.), pp. 621–637, World Scientific Publishing, Singapore.
- [8] K.S. Bey and J.T. Oden (1996). hp -Version discontinuous Galerkin methods for hyperbolic conservation laws. *Comput. Methods Appl. Mech. Engrg.*, 133:259–286.
- [9] R. Biswas, K. Devine, and J.E. Flaherty (1994). Parallel adaptive finite element methods for conservation laws. *App. Numer. Math.*, 14:255–284.
- [10] B. Cockburn, S. Hou, and C.-W. Shu (1990). TVB Runge-Kutta local projection discontinuous Galerkin finite elements for hyperbolic conservation laws. *Math. Comp.*, 54:545–581.
- [11] B. Cockburn and C.-W. Shu (1998). The local discontinuous Galerkin method for time-dependent reaction-diffusion systems. *SIAM J. Numer. Anal.*, 35:2440–2463.
- [12] W. Gui and I. Babuška (1986). The h , p and h - p versions of the finite element method in 1 Dimension. Part III. The adaptive h - p version. *Numer. Math.* 49:659–683.
- [13] P. Houston, R. Rannacher and E. Süli (1999). A posteriori error analysis for stabilised finite element approximations of transport problems. Oxford University Computing Laboratory Technical Report NA-99/04, 1999 (submitted for publication).
- [14] P. Houston, C. Schwab, and E. Süli (1998). Stabilised hp -finite element methods for first-order hyperbolic problems. Oxford University Computing Laboratory Technical Report NA-98/14 (submitted for publication).
- [15] P. Houston and E. Süli (1999). hp -Adaptive discontinuous Galerkin finite element methods for first-order hyperbolic problems. In preparation.
- [16] C. Johnson and J. Pitkäranta (1986). An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comp.* 46:1–26.

- [17] P. Lesaint and P.-A. Raviart (1974). On a finite element method for solving the neutron transport equation. In: C.A. deBoor, editor, *Mathematical Aspects of Finite Elements in Partial Differential Equations*. Academic Press, New York, pp.89–145.
- [18] J. Nitsche (1971). Über ein Variationsprinzip zur Lösung von Dirichlet Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh. Math. Sem. Univ. Hamburg*, 36:9-15.
- [19] J.T. Oden, I. Babuška, and C. Baumann (1998). A discontinuous *hp*-FEM for diffusion problems. *J. Comp. Phys.*, 146:491–519.
- [20] T.H.H. Pian (1965). Element stiffness matrices for boundary compatibility and for prescribed boundary stiffness. *Proceedings of the Conference on Matrix Methods in Structural Mechanics*, Wright–Patterson Air Force Base, AFFDL-TR-66-80, 457–477.
- [21] W.H. Reed and T.R. Hill (1973). Triangular mesh methods for neutron transport equation. Los Alamos Scientific Laboratory report LA-UR-73-479, Los Alamos, NM.
- [22] B. Riviere and M.-F. Wheeler (1999). Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. Part I. TICAM Technical Report, University of Texas at Austin, Texas.
- [23] C. Schwab (1998). *p*- and *hp*-Finite Element Methods. Theory and Applications to Solid and Fluid Mechanics. Oxford University Press.
- [24] E. Süli (1998). A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems. In: D. Kröner, editor, *Theory and Numerics of Conservation Laws*, pp. 123–194, Springer.
- [25] E. Süli, C. Schwab, and P. Houston (1999). *hp*-DGFEM for partial differential equations with nonnegative characteristic form. In: B. Cockburn, G. Karniadakis, and C.-W. Shu, editors, *Discontinuous Galerkin Finite Element Methods*. Lecture Notes in Computational Science and Engineering. Springer-Verlag (to appear).
- [26] M.F. Wheeler (1978). An elliptic collocation finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15:152–161.