DISCONTINUOUS GALERKIN METHODS FOR ADVECTION–DIFFUSION–REACTION PROBLEMS ON ANISOTROPICALLY REFINED MESHES

EMMANUIL H. GEORGOULIS *, EDWARD HALL †, AND PAUL HOUSTON ‡

Abstract. In this paper we consider the *a posteriori* and *a priori* error analysis of discontinuous Galerkin interior penalty methods for second–order partial differential equations with nonnegative characteristic form on anisotropically refined computational meshes. In particular, we discuss the question of error estimation for linear target functionals, such as the outflow flux and the local average of the solution. Based on our *a posteriori* error bound we design and implement the corresponding adaptive algorithm to ensure reliable and efficient control of the error in the prescribed functional to within a given tolerance. This involves exploiting both local isotropic and anisotropic mesh refinement. The theoretical results are illustrated by a series of numerical experiments.

Key words. Anisotropic mesh adaptation, discontinuous Galerkin methods, PDEs with non-negative characteristic form

AMS subject classifications. 65N30

1. Introduction.

The mathematical modeling of advection, diffusion, and reaction processes arises in many application areas. Typically, the diffusion is often small (compared to the magnitude of the advection and/or reaction), degenerate, or even vanishes in subregions of the domain of interest. This multi-scale behavior between the diffusion and the advection/reaction creates various challenges in the endeavor of computing numerical approximations to PDE problems of this type in an accurate and efficient manner. In particular, computationally demanding features may appear in the analytical solutions of such problems; these include boundary/interior layers or even discontinuities in the subregions where the problem is of hyperbolic type. When such, essentially lower-dimensional, features are present in the solution, the use of anisotropically refined meshes has been extensively advocated within the literature. Indeed, anisotropically refined meshes aim to be aligned with the domains of definition of these lower-dimensional features of the solution, in order to provide the necessary mesh resolution in the relevant directions, thereby reducing the number of degrees of freedom required to obtain an accurate approximation.

Discontinuous Galerkin finite element methods (DGFEMs) exhibit attractive properties for the numerical approximation of problems of hyperbolic or nearly–hyperbolic type, compared to both (standard) conforming finite element methods (FEMs) and finite volume methods (FVMs). Indeed, in contrast with conforming FEMs, but together with FVMs, DGFEMs are, by construction, locally conservative; moreover, they exhibit enhanced stability properties in the vicinity of boundary/interior layers and discontinuities present in the analytical solution. Additionally, DGFEMs offer advantages in the context of *hp*-adaptivity, such as increased flexibility in the mesh design (irregular grids are admissible) and the freedom to choose the elemental polynomial degrees without the need to enforce any conformity requirements. The

* Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK, email: Emmanuil.Georgoulis@mcs.le.ac.uk.

† Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK, email: ejch1@mcs.le.ac.uk.

‡ School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK, email: Paul.Houston@nottingham.ac.uk. The research of this author was supported by the EPSRC under grant GR/R76615.

implementation of genuinely (locally varying) high-order reconstruction techniques for FVMs still remains a computationally difficult task, particularly on general unstructured hybrid grids. Thereby, the combination of DGFEMs, which produce high–order and stable approximations, even in unresolved regions of the computational domain, with anisotropic mesh refinement, which aims to provide the desired mesh resolution in appropriate spatial directions, is an appealing technique for the numerical approximation of these types of problems.

In this article, we consider the *a priori* and *a posteriori* error analysis of interior penalty discontinuous Galerkin methods for second–order partial differential equations with nonnegative characteristic form on anisotropically refined computational meshes. In particular, we are concerned with the question of error estimation for linear target functionals of the analytical solution, such as the outflow flux and the local average of the solution. The *a priori* error estimation is based on exploiting the analysis developed in [13], which assumed that the underlying computational mesh is shape–regular, together with an extension of the techniques developed in [10] which precisely describe the anisotropy of the mesh; for related anisotropic approximation results, we refer to [1, 22, 21, 6], for example. More specifically, we employ tools from tensor analysis, along with local singular-value decompositions of the Jacobi matrix of the local elemental mappings, to derive directionally-sensitive bounds for arbitrary polynomial degree approximations, thus generalizing the ideas presented in [10], where only the case of approximation with conforming linear elements was considered. These interpolation error bounds are then employed to derive general anisotropic *a priori* error bounds for the DGFEM approximation of linear functionals of the underlying analytical solution.

Additionally, Type I *a posteriori* error bounds are derived based on employing the dual weighted residual approach, cf. [5, 14, 18, 20], for example. On the basis of our *a posteriori* error bound we design and implement two new anisotropic adaptive algorithms to ensure the reliable and efficient control of the error in the prescribed target functional to within a given tolerance. This involves exploiting both local isotropic and anisotropic mesh refinement, based on choosing the most competitive subdivision of a given element κ from a series of trial (Cartesian) refinements. The superiority of the proposed algorithms in comparison with standard isotropic mesh refinement, and a Hessian–based anisotropic mesh refinement strategy, will be illustrated by a series of numerical experiments.

The paper is structured as follows. In Section 2 we introduce the model problem and formulate its discontinuous Galerkin finite element approximation. Then, in Sections 3, 4, and 5 we develop the *a posteriori* and *a priori* analyses of the error measured in terms of certain linear target functionals of practical interest. Guided by our *a posteriori* error analysis, in Section 6 we design two adaptive finite element algorithms to guarantee reliable and efficient control of the error in the computed functional to within a fixed user–defined tolerance based on employing a combination of local isotropic and anisotropic mesh refinement. The performance of the resulting refinement strategies is then studied in Section 7 through a series of numerical experiments. Finally, in Section 8 we summarize the work presented in this paper and draw some conclusions.

Throughout this article we shall assume familiarity with the standard Hilbertian Sobolev spaces $H^k(\omega)$, $k \geq 0$, where ω is a bounded domain in $\mathbb{R}^d$, $d \geq 1$; we also adopt the notational convention $H^0(\omega) \equiv L_2(\omega)$.

2. Model problem and discretization. Let Ω be a bounded open polyhedral domain in $\mathbb{R}^d$, $d = 2, 3$, and let Γ signify the union of its $(d-1)$–dimensional open faces. We consider the advection–diffusion–reaction equation

$$\mathcal{L}u \equiv -\nabla \cdot (a\nabla u) + \nabla \cdot (\mathbf{b}u) + cu = f \,, \tag{2.1}$$

where $f \in L_2(\Omega)$ and $c \in L_\infty(\Omega)$ are real–valued, $\mathbf{b} = \{b_i\}_{i=1}^d$ is a vector function whose entries b_i are Lipschitz continuous real–valued functions on $\bar{\Omega}$, and $a = \{a_{ij}\}_{i,j=1}^d$ is a *symmetric* matrix whose entries a_{ij} are bounded, piecewise continuous real–valued functions defined on $\bar{\Omega}$, with

$$\boldsymbol{\zeta}^\top a(x)\boldsymbol{\zeta} \geq 0 \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^d \,, \quad \text{a.e. } x \in \bar{\Omega} \,. \tag{2.2}$$

Under this hypothesis, (2.1) is termed a *partial differential equation with nonnegative characteristic form*. By $\mathbf{n}(x) = \{n_i(x)\}_{i=1}^d$ we denote the unit outward normal vector to Γ at $x \in \Gamma$. On introducing the so called *Fichera function* $\mathbf{b} \cdot \mathbf{n}$ (cf. [26]), we define

$$\Gamma_0 = \left\{ x \in \Gamma : \ \mathbf{n}(x)^\top a(x)\mathbf{n}(x) > 0 \right\} ,$$
$$\Gamma_- = \{ x \in \Gamma \backslash \Gamma_0 : \ \mathbf{b}(x) \cdot \mathbf{n}(x) < 0 \} \,, \quad \Gamma_+ = \{ x \in \Gamma \backslash \Gamma_0 : \ \mathbf{b}(x) \cdot \mathbf{n}(x) \geq 0 \} \,.$$

The sets Γ_- and Γ_+ will be referred to as the inflow and outflow boundary, respectively. Evidently, $\Gamma = \Gamma_0 \cup \Gamma_- \cup \Gamma_+$. If Γ_0 is nonempty, we shall further divide it into disjoint subsets $\Gamma_{\rm D}$ and $\Gamma_{\rm N}$ whose union is Γ_0, with $\Gamma_{\rm D}$ nonempty and relatively open in Γ. We supplement (2.1) with the boundary conditions

$$u = g_{\rm D} \quad \text{on} \quad \Gamma_{\rm D} \cup \Gamma_- \,, \qquad \mathbf{n} \cdot (a\nabla u) = g_{\rm N} \quad \text{on} \quad \Gamma_{\rm N} \,, \tag{2.3}$$

and adopt the (physically reasonable) hypothesis that $\mathbf{b} \cdot \mathbf{n} \geq 0$ on $\Gamma_{\rm N}$, whenever $\Gamma_{\rm N}$ is nonempty. Additionally, we assume that the following (standard) positivity hypothesis holds: there exists a constant vector $\boldsymbol{\xi} \in \mathbb{R}^d$ such that

$$c(x) + \frac{1}{2} \nabla \cdot \mathbf{b}(x) + \mathbf{b}(x) \cdot \boldsymbol{\xi} > 0 \ \text{ a.e. } \ x \in \Omega \,. \tag{2.4}$$

For simplicity of presentation, we assume throughout that (2.4) is satisfied with $\boldsymbol{\xi} \equiv \mathbf{0}$; we then define the positive function c_0 by

$$(c_0(x))^2 = c(x) + \frac{1}{2} \nabla \cdot \mathbf{b}(x) \ \text{ a.e. } \ x \in \Omega \,. \tag{2.5}$$

For the well-posedness theory (for weak solutions) of the boundary value problem (2.1), (2.3), in the case of homogeneous boundary conditions, we refer to [17, 19].

2.1. Meshes, finite element spaces and traces. Let $\mathcal{T}_h = \{\kappa\}$ be a subdivision of the (polygonal) domain Ω into disjoint open element domains κ constructed through the use of the mappings $Q_\kappa \circ F_\kappa$, where $F_\kappa : \hat{\kappa} \to \tilde{\kappa}$ is an affine mapping from the reference element $\hat{\kappa}$ to $\tilde{\kappa}$, and $Q_\kappa : \tilde{\kappa} \to \kappa$ is a C^1–diffeomorphism from $\tilde{\kappa}$ to the physical element κ. Here, we shall assume that $\hat{\kappa}$ is either the hypercube $(-1, 1)^d$ or the unit d–simplex; in the latter case Q_κ is typically the identity operator, unless curved elements are employed. The mapping F_κ defines the size and orientation of the element κ, while Q_κ defines the shape of κ, without any significant rescaling, or indeed change of orientation, cf. Figure 2.1 for the case when $d = 2$ and $\hat{\kappa} = (-1, 1)^2$.

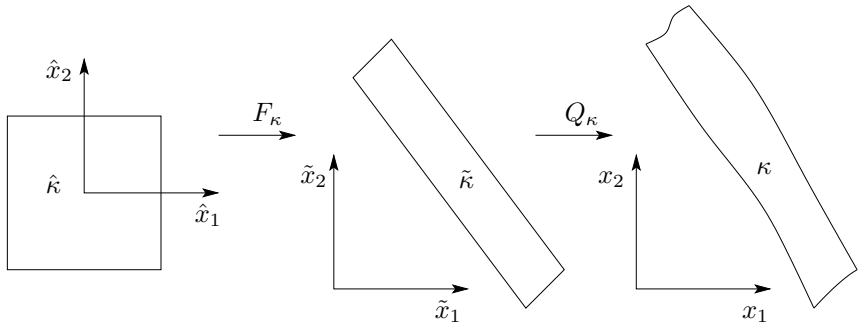

FIG. 2.1. *Construction of the element mapping via the composition of an affine mapping F_κ and a C^1–diffeomorphism Q_κ.*

With this in mind, we assume that the element mapping Q_κ is close to the identity in the following sense: the Jacobi matrix J_{Q_κ} of Q_κ satisfies

$$C_1^{-1} \leq \|\det J_{Q_\kappa}\|_{L_\infty(\kappa)} \leq C_1, \quad \|J_{Q_\kappa}^{-\top}\|_{L_\infty(\kappa)} \leq C_2, \quad \|J_{Q_\kappa}^{-\top}\|_{L_\infty(\partial\kappa)} \leq C_3 \quad (2.6)$$

for all κ in $\mathcal{T}_h$ uniformly throughout the mesh for some positive constants C_1, C_2, and C_3. This will be important as our error estimates will be expressed in terms of Sobolev norms over the element domains $\tilde{\kappa}$, in order to ensure that only the scaling and orientation introduced by the affine element maps F_κ are present in the analysis. Writing m_κ, $m_{\tilde{\kappa}}$, and $m_{\hat{\kappa}}$ to denote the d–dimensional measure of the elements κ, $\tilde{\kappa}$, and $\hat{\kappa}$, respectively, the above condition (2.6) implies that there exists a positive constant C_4 such that

$$C_4^{-1} m_{\tilde{\kappa}} \leq m_\kappa \leq C_4 m_{\tilde{\kappa}} \quad \forall \kappa \in \mathcal{T}_h. \quad (2.7)$$

The above maps are assumed to be constructed in such a manner to ensure that the union of the closure of the disjoint open elements $\kappa \in \mathcal{T}_h$ forms a covering of the closure of Ω, i.e., $\bar{\Omega} = \cup_{\kappa \in \mathcal{T}_h} \bar{\kappa}$. For a function v defined on κ, $\kappa \in \mathcal{T}_h$, we write $\tilde{v} = v \circ Q_\kappa$ and $\hat{v} = \tilde{v} \circ F_\kappa$ to denote the corresponding functions on the elements $\tilde{\kappa}$ and $\hat{\kappa}$, respectively. Thereby, we have that $\hat{v} = v \circ Q_\kappa \circ F_\kappa$.

REMARK 2.1. *We note that a similar construction of the element mappings for general meshes consisting of curved quadrilateral elements has also been employed for both shape-regular and anisotropic meshes in the articles [16] and [11], respectively. The key difference in the current construction to that proposed in [11] is that here the element mapping F_κ contains information about both size and orientation of κ. In contrast, in the construction developed in [11] both orientation and shape information are included in Q_κ, while F_κ only contains information relating to the size of κ.*

REMARK 2.2. *Within this construction we admit meshes with possibly hanging nodes; for simplicity, we shall suppose that the mesh $\mathcal{T}_h$ is 1-irregular, cf. [16].*

Associated with $\mathcal{T}_h$, we introduce the broken Sobolev space of order $s \geq 0$ defined by

$$H^s(\Omega, \mathcal{T}_h) = \{u \in L_2(\Omega) : \ u|_\kappa \in H^s(\kappa) \quad \forall \kappa \in \mathcal{T}_h\},$$

equipped with the usual broken Sobolev seminorm and norm, denoted, respectively, by $|\cdot|_{s,\mathcal{T}_h}$ and $\|\cdot\|_{s,\mathcal{T}_h}$. For $u \in H^1(\Omega, \mathcal{T}_h)$ we define the broken gradient $\nabla_{\mathcal{T}_h} u$ of u by $(\nabla_{\mathcal{T}_h} u)|_\kappa = \nabla(u|_\kappa)$, $\kappa \in \mathcal{T}_h$.

2.2. Interior penalty discontinuous Galerkin method. We introduce the (symmetric) interior penalty DGFEM discretization of the advection–diffusion–reaction problem (2.1), (2.3). To this end, we define the following notation. Given a polynomial degree $p \geq 1$ we define the finite element space $S_{h,p}$ as follows

$$S_{h,p} = \{u \in L_2(\Omega) \,:\, u|_\kappa \circ Q_\kappa \circ F_\kappa \in \mathcal{R}_p(\kappa);\ \kappa \in \mathcal{T}_h\}\,,$$

where $\mathcal{R}_p$ is $\mathcal{P}_p$, when $\hat{\kappa}$ is the unit d–simplex, or $\mathcal{R}_p$ is $\mathcal{Q}_p$, when $\hat{\kappa} = (-1,1)^d$. Here, $\mathcal{P}_p$ denotes the set of polynomials of total degree p on $\hat{\kappa}$ and $\mathcal{Q}_p(\hat{\kappa})$, the set of all tensor-product polynomials on $\hat{\kappa}$ of degree p in each coordinate direction.

An *interior face* of $\mathcal{T}_h$ is defined as the (non-empty) $(d-1)$–dimensional interior of $\partial\kappa_i \cap \partial\kappa_j$, where κ_i and κ_j are two adjacent elements of $\mathcal{T}_h$, not necessarily matching. A *boundary face* of $\mathcal{T}_h$ is defined as the (non-empty) $(d-1)$–dimensional interior of $\partial\kappa \cap \Gamma$, where κ is a boundary element of $\mathcal{T}_h$. We denote by $\Gamma_{\rm int}$ the union of all interior faces of $\mathcal{T}_h$. Given a face $f \subset \Gamma_{\rm int}$, shared by the two elements κ_i and κ_j, where the indices i and j satisfy $i > j$, we write $\mathbf{n}_f$ to denote the (numbering–dependent) unit normal vector which points from κ_i to κ_j; on boundary faces, we put $\mathbf{n}_f = \mathbf{n}$. Further, for $v \in H^1(\Omega, \mathcal{T}_h)$ we define the jump of v across f and the mean value of v on f, respectively, by $[v] = v|_{\partial\kappa_i \cap f} - v|_{\partial\kappa_j \cap f}$ and $\langle v \rangle = \frac{1}{2}\left(v|_{\partial\kappa_i \cap f} + v|_{\partial\kappa_j \cap f}\right)$.

On a boundary face $f \subset \partial\kappa$, we set $[v] = v|_{\partial\kappa \cap f}$ and $\langle v \rangle = v|_{\partial\kappa \cap f}$. Finally, given a function $v \in H^1(\Omega, \mathcal{T}_h)$ and an element $\kappa \in \mathcal{T}_h$, we denote by v_κ^+ (respectively, v_κ^-) the interior (respectively, exterior) trace of v defined on $\partial\kappa$ (respectively, $\partial\kappa \backslash \Gamma$). Since below it will always be clear from the context which element κ in the subdivision $\mathcal{T}_h$ the quantities v_κ^+ and v_κ^- correspond to, for the sake of notational simplicity we shall suppress the letter κ in the subscript and write, respectively, v^+ and v^- instead.

Given that κ is an element in the subdivision $\mathcal{T}_h$, we denote by $\partial\kappa$ the union of $(d-1)$–dimensional open faces of κ. Let $x \in \partial\kappa$ and suppose that $\mathbf{n}_\kappa(x)$ denotes the unit outward normal vector to $\partial\kappa$ at x. With these conventions, we define the inflow and outflow parts of $\partial\kappa$, respectively, by

$$\partial_-\kappa = \{x \in \partial\kappa :\ \mathbf{b}(x) \cdot \mathbf{n}_\kappa(x) < 0\}\,, \quad \partial_+\kappa = \{x \in \partial\kappa :\ \mathbf{b}(x) \cdot \mathbf{n}_\kappa(x) \geq 0\}\,.$$

For simplicity of presentation, we suppose that the entries of the matrix a are constant on each element κ in $\mathcal{T}_h$; *i.e.*,

$$a \in [S_{h,0}]^{d\times d}_{\rm sym}\,. \tag{2.8}$$

We note that, with minor changes only, our results can easily be extended to the case of $\sqrt{a} \in [S_{h,q}]^{d\times d}_{\rm sym}$, $q \geq 0$; moreover, for general $a \in L^\infty(\Omega)^{d\times d}_{\rm sym}$, the analysis proceeds in a similar manner, based on employing the modified DG method proposed in [12]. In the following, we write $\bar{a} = |\sqrt{a}\,|_2^2$, where $|\cdot|_2$ denotes the matrix norm subordinate to the l_2–vector norm on $\mathbb{R}^d$ and $\bar{a}_\kappa = \bar{a}|_\kappa$.

The DGFEM approximation of (2.1), (2.3) is defined as follows: find $u_{\rm DG}$ in $S_{h,p}$ such that

$$B_{\rm DG}(u_{\rm DG}, v) = \ell_{\rm DG}(v) \tag{2.9}$$

for all $v \in S_{h,p}$. Here, the bilinear form $B_{\rm DG}(\cdot,\cdot)$ is defined by

$$B_{\rm DG}(w,v) = B_a(w,v) + B_{\mathbf{b}}(w,v) - B_f(v,w) - B_f(w,v) + B_\vartheta(w,v)\,,$$

where

$$B_a(w,v) = \sum_{\kappa\in\mathcal{T}_h}\int_\kappa a\nabla w\cdot\nabla v\,\mathrm{d}x\,,$$

$$B_{\mathbf{b}}(w,v) = \sum_{\kappa\in\mathcal{T}_h}\left\{-\int_\kappa (w\,\mathbf{b}\cdot\nabla v - cwv)\,\mathrm{d}x \right.$$
$$\left. +\int_{\partial_+\kappa}(\mathbf{b}\cdot\mathbf{n}_\kappa)\,w^+v^+\,\mathrm{d}s + \int_{\partial_-\kappa\setminus\Gamma}(\mathbf{b}\cdot\mathbf{n}_\kappa)\,w^-v^+\,\mathrm{d}s\right\},$$

$$B_f(w,v) = \int_{\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}}\langle(a\nabla w)\cdot\mathbf{n}_f\rangle[v]\,\mathrm{d}s\,,\qquad B_\vartheta(w,v) = \int_{\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}}\vartheta[w][v]\,\mathrm{d}s\,,$$

and the linear functional $\ell_{\mathrm{DG}}(\cdot)$ is given by

$$\ell_{\mathrm{DG}}(v) = \sum_{\kappa\in\mathcal{T}_h}\left(\int_\kappa fv\,\mathrm{d}x - \int_{\partial_-\kappa\cap(\Gamma_{\mathrm{D}}\cup\Gamma_-)}(\mathbf{b}\cdot\mathbf{n}_\kappa)\,g_{\mathrm{D}}\,v^+\,\mathrm{d}s\right.$$
$$\left. -\int_{\partial\kappa\cap\Gamma_{\mathrm{D}}} g_{\mathrm{D}}((a\nabla v^+)\cdot\mathbf{n}_\kappa)\,\mathrm{d}s + \int_{\partial\kappa\cap\Gamma_{\mathrm{N}}} g_{\mathrm{N}}v^+\,\mathrm{d}s + \int_{\partial\kappa\cap\Gamma_{\mathrm{D}}}\vartheta g_{\mathrm{D}}v^+\,\mathrm{d}s\right).$$

Here ϑ is called the *discontinuity-penalization* parameter and is defined by $\vartheta|_f = \vartheta_f$ for $f\subset\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}$, where ϑ_f is a nonnegative constant on face f. The precise choice of ϑ_f, which depends on a and the discretization parameters, will be discussed in detail in the next section. We shall adopt the convention that faces $f\subset\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}$ with $\vartheta|_f = 0$ are omitted from the integrals appearing in the definition of $B_\vartheta(w,v)$ and $\ell_{\mathrm{DG}}(v)$, although we shall not highlight this explicitly in our notation; the same convention is adopted in the case of integrals where the integrand contains the factor $1/\vartheta$. Thus, in particular, the definition of the DG-norm, cf. (3.1) below, is meaningful even if $\vartheta|_f$ happens to be equal to zero on certain faces $f\subset\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}$, given that such faces are understood to be excluded from the region of integration.

3. Stability analysis. Before embarking on the error analysis of the discontinuous Galerkin method (2.9), we first derive some preliminary results. Let us first introduce the DG–norm $|||\cdot|||$ by

$$|||w|||^2 = \sum_{\kappa\in\mathcal{T}_h}\left(\|\sqrt{a}\,\nabla w\|^2_{L_2(\kappa)} + \|c_0w\|^2_{L_2(\kappa)} + \frac{1}{2}\|w^+\|^2_{\partial_-\kappa\cap(\Gamma_{\mathrm{D}}\cup\Gamma_-)}\right.$$
$$\left. +\frac{1}{2}\|w^+ - w^-\|^2_{\partial_-\kappa\setminus\Gamma} + \frac{1}{2}\|w^+\|^2_{\partial_+\kappa\cap\Gamma}\right)$$
$$+\int_{\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}}\vartheta[w]^2\,\mathrm{d}s + \int_{\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}}\frac{1}{\vartheta}\langle(a\,\nabla w)\cdot\mathbf{n}_f\rangle^2\,\mathrm{d}s\,, \tag{3.1}$$

where $\|\cdot\|_\tau$, $\tau\subset\partial\kappa$, denotes the (semi)norm associated with the (semi)inner-product $(v,w)_\tau = \int_\tau|\mathbf{b}\cdot\mathbf{n}_\kappa|vw\,\mathrm{d}s$, and c_0 is as defined in (2.5). We remark that the above definition of $|||\cdot|||$ represents a slight modification of the norm considered in [17]; in the case $\mathbf{b}\equiv\mathbf{0}$, (3.1) corresponds to the norm proposed by Baumann *et al.* [4, 25] and Baker *et al.* [3], cf. [27].

For a given face $f\subset\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}$, such that $f\subset\partial\kappa$, for some $\kappa\in\mathcal{T}_h$, we write $\tilde{f}$ and $\hat{f}$ to denote the respective faces of the mapped elements $\tilde{\kappa}$ and $\hat{\kappa}$, respectively, based on employing the element mappings Q_κ and F_κ. More precisely, we write $\tilde{f} = Q_\kappa^{-1}(f)$

and $\hat{f} = F_\kappa^{-1}(\tilde{f})$. Further, we define m_f, $m_{\tilde{f}}$, and $m_{\hat{f}}$ to denote the $(d-1)$–dimensional measure (volume) of the faces f, $\tilde{f}$, and $\hat{f}$, respectively; clearly, in two–dimensions, i.e., $d = 2$, $m_{\hat{f}}$, the length of the corresponding face on the canonical element, is equal to 2. In view of (2.6), we note that there exists a positive constant C_5, such that

$$C_5^{-1} m_{\tilde{f}} \leq m_f \leq C_5 m_{\tilde{f}} \tag{3.2}$$

for every face $f \subset \Gamma_{\text{int}} \cup \Gamma_{\text{D}}$. Moreover, the surface Jacobian $S_{f,\tilde{f}}$ arising in the transformation of the face f to $\tilde{f}$ may be uniformly bounded in the following manner

$$\|S_{f,\tilde{f}}\|_{L_\infty(\tilde{f})} \leq C_6 \tag{3.3}$$

for all faces $f \subset \Gamma_{\text{int}} \cup \Gamma_{\text{D}}$, where C_6 is a positive constant.

Let us now quote the following inverse inequality.

LEMMA 3.1. *Let κ be an element contained in the mesh $\mathcal{T}_h$ and let f denote one of its faces. Then, the following inverse inequality holds*

$$\|v\|^2_{L_2(f)} \leq C_{\text{inv}} \frac{m_f}{m_\kappa} \|v\|^2_{L_2(\kappa)} \tag{3.4}$$

for all v such that $v \circ Q_\kappa \circ F_\kappa \in \mathcal{Q}_p(\hat{\kappa})$, where C_{inv} is a constant which depends only on the dimension d and the polynomial degree p.

Proof. On the reference element $\hat{\kappa}$, for any function $\hat{v} \in \mathcal{Q}_p(\hat{\kappa})$, there exists a positive constant C'_{inv}, such that

$$\|\hat{v}\|^2_{L_2(\hat{f})} \leq C'_{\text{inv}} \|\hat{v}\|^2_{L_2(\hat{\kappa})}; \tag{3.5}$$

see, for example, [2]. Thereby, employing (3.3) and (3.2) we deduce that

$$\|v\|^2_{L_2(f)} \leq C_6 \|\tilde{v}\|^2_{L_2(\tilde{f})} = C_6 \frac{m_{\tilde{f}}}{m_{\hat{f}}} \|\hat{v}\|^2_{L_2(\hat{f})} \leq \frac{C_6}{C_5} \frac{m_f}{m_{\hat{f}}} \|\hat{v}\|^2_{L_2(\hat{f})}. \tag{3.6}$$

In an analogous manner, by exploiting (2.7) and (2.6) gives

$$\|\hat{v}\|^2_{L_2(\hat{\kappa})} = \det(F_\kappa^{-1}) \|\tilde{v}\|^2_{L_2(\tilde{\kappa})} = \frac{m_{\hat{\kappa}}}{m_{\tilde{\kappa}}} \|\tilde{v}\|^2_{L_2(\tilde{\kappa})} \leq C_4 \frac{m_{\hat{\kappa}}}{m_\kappa} \|\tilde{v}\|^2_{L_2(\tilde{\kappa})} \leq C_1 C_4 \frac{m_{\hat{\kappa}}}{m_\kappa} \|v\|^2_{L_2(\kappa)}. \tag{3.7}$$

Inserting (3.6) and (3.7) into (3.5) gives the desired result. ∎

REMARK 3.2. *The inverse inequality stated in Lemma 3.1 is an extension of the standard result employed on isotropic finite element meshes to the case when anisotropic elements may be present. Indeed, in the isotropic setting, we have that $m_\kappa \approx h_\kappa^d$ and $m_f \approx h_\kappa^{d-1}$, where h_κ denotes the diameter of the element $\kappa \in \mathcal{T}_h$; thereby, the scaling on the right–hand side of the inequality (3.4) is of size $1/h_\kappa$, as expected. Moreover, this result extends the inverse inequality stated in [11] to the case when the affine mapping F_κ includes not only size, but also orientation information, cf. above.*

We now define the function $\mathtt{h}$ in $L_\infty(\Gamma_{\text{int}} \cup \Gamma_{\text{D}})$, as $\mathtt{h}(x) = \min\{m_{\kappa_1}, m_{\kappa_2}\}/m_f$, if x is in the interior of $f = \partial\kappa_1 \cap \partial\kappa_2$ for two neighboring elements in the mesh $\mathcal{T}_h$, and $\mathtt{h}(x) = m_\kappa/m_f$, if x is in the interior of $f = \partial\kappa \cap \Gamma_{\text{D}}$. We note that in the isotropic setting we observe that $\mathtt{h} \sim h$, where h denotes the mesh local mesh size, cf. Remark 3.2 above. Similarly, we define the function $\mathtt{a}$ in $L_\infty(\Gamma_{\text{int}} \cup \Gamma_{\text{D}})$ by $\mathtt{a}(x) = \min\{\bar{a}_{\kappa_1}, \bar{a}_{\kappa_2}\}$ if x is in the interior of $e = \partial\kappa_1 \cap \partial\kappa_2$, and $\mathtt{a}(x) = \bar{a}_\kappa$ if x is in

the interior of $\partial\kappa \cap \Gamma_{\mathrm{D}}$. With this notation, we now provide the following coercivity result for the bilinear form $B_{\mathrm{DG}}(\cdot,\cdot)$ over $S_{h,p} \times S_{h,p}$.

THEOREM 3.3. *Define the discontinuity-penalization parameter ϑ arising in (2.9) by*

$$\vartheta|_f \equiv \vartheta_f = C_\vartheta \frac{\mathtt{a}}{\mathtt{h}} \quad \textit{for} \quad f \subset \Gamma_{\mathrm{int}} \cup \Gamma_{\mathrm{D}}, \tag{3.8}$$

where C_ϑ is a sufficiently large positive constant (see Remark 3.4 below). Then, there exists a positive constant C, which depends only on the dimension d and the polynomial degree p, such that

$$B_{\mathrm{DG}}(v,v) \geq C|||v|||^2 \quad \forall v \in S_{h,p}. \tag{3.9}$$

Proof. This result follows by application of the inverse estimate derived in Lemma 3.1, following the general argument presented by Prudhomme *et al.* [27] in the case when $\mathbf{b} \equiv \mathbf{0}$; cf., also [17]. ∎

REMARK 3.4. *Theorem 3.3 indicates that the DG scheme is coercive over $S_{h,p} \times S_{h,p}$ provided that the constant $C_\vartheta > 0$ arising in the definition of the discontinuity–penalization parameter ϑ, is chosen sufficiently large. More precisely, C_ϑ should be selected to be a positive constant which is greater than $C_f C_{\mathrm{inv}}/2$, where C_{inv} is the constant arising in the inverse inequality stated in Lemma 3.1 and*

$$C_f = \max_{\kappa \in \mathcal{T}_h} \mathit{card}\{f \subset \Gamma_{\mathrm{int}} \cup \Gamma_{\mathrm{D}} : f \subset \partial\kappa\};$$

the restriction to 1–irregular meshes ensures that C_f is uniformly bounded independently of the mesh size.

For the proceeding error analysis, we assume that the solution u to the boundary value problem (2.1), (2.3) is sufficiently smooth: namely, $u \in H^{3/2+\varepsilon}(\Omega,\mathcal{T}_h)$, $\varepsilon > 0$, and the functions u and $(a\nabla u)\cdot\mathbf{n}_f$ are continuous across each face $f \subset \partial\kappa\backslash\Gamma$ that intersects the subdomain of ellipticity, $\Omega_a = \{x \in \bar{\Omega} : \boldsymbol{\zeta}^\top a(x)\boldsymbol{\zeta} > 0 \ \ \forall \boldsymbol{\zeta} \in \mathbb{R}^d\}$. If this smoothness requirement is violated, the discretization method has to be modified accordingly, cf. [17]. We note that under these assumptions, the following Galerkin orthogonality property holds:

$$B_{\mathrm{DG}}(u - u_{\mathrm{DG}}, v) = 0 \quad \forall v \in S_{h,p}\,. \tag{3.10}$$

For simplicity of presentation, it will be assumed in the proceeding analysis, as well as in Section 5.2, that the velocity vector $\mathbf{b}$ satisfies the following assumption:

$$\mathbf{b}\cdot\nabla_{\mathcal{T}_h} v \in S_{h,p} \quad \forall v \in S_{h,p}\,. \tag{3.11}$$

To ensure that (2.1) is then meaningful (*i.e.*, that the characteristic curves of the differential operator $\mathcal{L}$ are correctly defined), we still assume that $\mathbf{b} \in \left[W^1_\infty(\Omega)\right]^d$.

REMARK 3.5. *We note that hypothesis (3.11) is a standard condition assumed for the analysis of the hp–version of the DGFEM; see, for example, [11, 13, 17]. Indeed, this condition is essential for the derivation of* a priori *error bounds which are optimal in both the mesh size h and spectral order p; in the absence of this assumption, optimal h–convergence bounds may still be derived, though a loss of $p^{1/2}$ is observed in the resulting error analysis, unless the scheme (2.9) is supplemented by appropriate streamline–diffusion stabilization, cf. the discussion in [16]. Given that within the current setting, we are only interested in deriving error bounds for the h–version of the DGFEM, hypothesis (3.11) is indeed unnecessary, but for simplicity of presentation, we retain this assumption.*

4. Approximation results. In this section we develop the necessary approximation results needed for the forthcoming *a priori* error estimation developed in Section 5. To this end, on the reference element $\hat{\kappa}$, we define $\hat{\Pi}_p$ to denote the orthogonal projector in $L_2(\hat{\kappa})$ onto the space of polynomials $Q_p(\hat{\kappa})$; *i.e.*, given that $\hat{v} \in L_2(\hat{\kappa})$, we define $\hat{\Pi}_p\hat{v}$ by $(\hat{v}-\hat{\Pi}_p\hat{v},\hat{w})_{\hat{\kappa}} = 0$ for all $\hat{w} \in Q_p(\hat{\kappa})$, where $(\cdot,\cdot)_{\hat{\kappa}}$ denotes the $L_2(\hat{\kappa})$ inner product. Similarly, we define the L_2-projection operators $\tilde{\Pi}_p$ and Π_p on $\tilde{\kappa}$ and κ, respectively, by the relations

$$\tilde{\Pi}_p\tilde{v} := (\hat{\Pi}_p(\tilde{v}\circ F_\kappa))\circ F_\kappa^{-1}, \qquad \Pi_p v := (\tilde{\Pi}_p(v\circ Q_\kappa))\circ Q_\kappa^{-1},$$

for $\tilde{v} \in L_2(\tilde{\kappa})$ and $v \in L_2(\kappa)$, respectively.

We remark that this choice of projector is essential in the following *a priori* error analysis, in order to ensure that

$$(u-\Pi_p u, \mathbf{b}\cdot\nabla_{\mathcal{T}_h} v) = 0 \tag{4.1}$$

for all v in $S_{h,p}$, cf. the proofs of Lemma 5.3 and Theorem 5.4 below. We remark that this same choice of projector is also necessary in the corresponding case when (3.11) fails to hold; in this situation an equality of the form (4.1) with $\mathbf{b}$ replaced by a suitable projection of $\mathbf{b}$ is still necessary for the underlying analysis; cf. [11].

With this notation, we now quote the following approximation result on the reference element $\hat{\kappa}$.

LEMMA 4.1. *Let $\hat{\kappa}$ be the reference element, and let $\hat{f}$ denote one of its faces. Given a function $\hat{v} \in H^k(\hat{\kappa})$, the following error bounds hold for $m = 0, 1$:*

$$|\hat{v}-\hat{\Pi}_p\hat{v}|_{H^m(\hat{\kappa})} \le C|\hat{v}|_{H^s(\hat{\kappa})}, \qquad m \le s \le \min(p+1,k), \tag{4.2}$$

$$|\hat{v}-\hat{\Pi}_p\hat{v}|_{H^m(\hat{f})} \le C|\hat{v}|_{H^s(\hat{\kappa})}, \quad m+1 \le s \le \min(p+1,k), \tag{4.3}$$

where C is a positive constant which depends only on the dimension d and the polynomial order p.

Proof. The proof of (4.2) is standard; see [8], for example. The approximation result (4.3) follows upon application of the multiplicative trace inequality, cf. [16]. ∎

COROLLARY 4.2. *Using the notation of Lemma 4.1, there exists a positive constant C, which depends only on the dimension d and the polynomial order p, such that for $m = 0, 1$:*

$$|v-\Pi_p v|_{H^m(\kappa)} \le C|\det(J_{F_\kappa})|^{1/2}\,\|J_{F_\kappa}^{-\top}\|_2^m\,|\hat{v}|_{H^s(\hat{\kappa})}, \quad m \le s \le \min(p+1,k), \tag{4.4}$$

$$|v-\Pi_p v|_{H^m(f)} \le C|m_f|^{1/2}\,\|J_{F_\kappa}^{-\top}\|_2^m\,|\hat{v}|_{H^s(\hat{\kappa})}, \qquad m+1 \le s \le \min(p+1,k). \tag{4.5}$$

Proof. The proof of the each inequality stated in the corollary is based on exploiting a standard scaling argument to the respective left–hand sides of the approximation results stated in Lemma 4.1, together with (2.6), (3.2), (3.3), and (3.6). Indeed, the proof of (4.4) exploits (4.2), together with the following (scaling) inequality

$$\begin{aligned}|v-\Pi_p v|^2_{H^m(\kappa)} &\le \|\det J_{Q_\kappa}\|_{L_\infty(\kappa)}\,\|J_{Q_\kappa}^{-\top}\|^{2m}_{L_\infty(\kappa)}\,|\tilde{v}-\tilde{\Pi}_p\tilde{v}|^2_{H^m(\tilde{\kappa})}\\ &\le C_1(C_2)^{2m}|\tilde{v}-\tilde{\Pi}_p\tilde{v}|^2_{H^m(\tilde{\kappa})} \le C_1(C_2)^{2m}|\det J_{F_\kappa}|\,\|J_{F_\kappa}^{-\top}\|_2^{2m}\,|\hat{v}-\hat{\Pi}_p\hat{v}|^2_{H^m(\hat{\kappa})};\end{aligned} \tag{4.6}$$

here we have used (2.6). Finally, employing (2.6), (3.3), and (3.2), we deduce that

$$|v-\Pi_p v|^2_{H^m(f)} \le C_3^m C_6|\tilde{v}-\tilde{\Pi}_p\tilde{v}|^2_{H^m(\tilde{f})} \le \frac{C_3^m C_6}{C_5}\frac{m_f}{m_{\hat{f}}}\|J_{F_\kappa}^{-\top}\|_2^{2m}|\hat{v}-\hat{\Pi}_p\hat{v}|^2_{H^m(\hat{f})}. \tag{4.7}$$

Upon substituting (4.7) into (4.3), we deduce (4.5). □

Finally, it remains to scale the $H^s(\hat{\kappa})$, $s \geq 0$, semi-norm defined on the reference element $\hat{\kappa}$ to $\tilde{\kappa}$ based on employing the affine element transformation F_κ. In order to retain the anisotropic mesh information within the Jacobian J_{F_κ}, we first re-write the square of the $H^s(\hat{\kappa})$ semi-norm of a function $\hat{v}$ terms of the integral of the square of the Frobenius norm of an sth–order tensor containing the s–order derivatives of $\hat{v}$. With this definition the transformation of the s–order derivatives of $\hat{v}$ defined over $\hat{\kappa}$ may naturally be transformed to derivatives of the (mapped) function $\tilde{v}$ defined over $\tilde{\kappa}$. Indeed, for the case when $s = 2$, this approach is analogous to the technique employed in [10].

To this end, we now introduce the following tensor notation; here, and in the following we use calligraphic letters $\mathcal{A}, \mathcal{B}, \ldots$ to denote Nth–order tensors, where it is understood that a 0th–order tensor is a scalar, a 1st–order tensor is a vector, a 2nd–order tensor is a matrix, and so on. The following discussion regarding tensors is based on the work presented in the article [24].

DEFINITION 4.3. *For an Nth order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, the matrix unfolding $A_{(n)} \in \mathbb{R}^{I_n \times (I_{n+1} I_{n+2} \ldots I_N I_1 I_2 \ldots I_{n-1})}$, $n = 1, \ldots, N$, contains the element $a_{i_1 i_2 \ldots i_N}$ at the position with row number i_n and column number equal to*

$$(i_{n+1}-1)I_{n+2}I_{n+3}\ldots I_N I_2 \ldots I_{n-1} + (i_{n+2}-1)I_{n+3}I_{n+4}\ldots I_N I_1 I_2 \ldots I_{n-1} + \ldots$$
$$+(i_N - 1)I_1 I_2 \ldots I_{n-1} + (i_1 - 1)I_2 I_3 \ldots I_{n-1} + (i_2 - 1)I_3 I_4 \ldots I_{n-1} + \ldots + i_{n-1}.$$

This definition prompts us to consider a way of multiplying a tensor by a matrix. Clearly if we have a matrix $U \in \mathbb{R}^{J_n \times I_n}$ then we can pre-multiply $A_{(n)}$ by U. Forming an Nth order tensor from $UA_{(n)}$ by reversing the matrix unfolding procedure we have the product of a matrix and a tensor, giving rise to a tensor $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots I_N}$. We formalize this in the following definition:

DEFINITION 4.4. *The n-mode product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ by a matrix $U \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n U$, is an $I_1 \times I_2 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots I_N$-tensor of which the entries are given by*

$$(\mathcal{A} \times_n U)_{i_1 i_2 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} := \sum_{i_n=1}^{I_n} (\mathcal{A})_{i_1 i_2 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} (U)_{j_n i_n}.$$

LEMMA 4.5. *For $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and $U \in \mathbb{R}^{J_n \times I_n}$, we have that*

$$(\mathcal{A} \times_n U)_{(n)} = U\mathcal{A}_{(n)}.$$

Proof. Consider element $(\mathcal{A} \times_n U)_{i_1 i_2 \ldots i_{n-1} j_n i_{n+1} \ldots i_N}$, its position in $(\mathcal{A} \times_n U)_{(n)}$ is at row number j_n and column number k, where

$$k = (i_{n+1}-1)I_{n+2}I_{n+3}\ldots I_N I_2 \ldots I_{n-1} + (i_{n+2}-1)I_{n+3}I_{n+4}\ldots I_N I_1 I_2 \ldots I_{n-1} + \ldots$$
$$+(i_N - 1)I_1 I_2 \ldots I_{n-1} + (i_1 - 1)I_2 I_3 \ldots I_{n-1} + (i_2 - 1)I_3 I_4 \ldots I_{n-1} + \ldots + i_{n-1}.$$

Now,

$$(U\mathcal{A}_{(n)})_{j_n k} = \sum_{i_n=1}^{I_n} (U)_{j_n i_n} (\mathcal{A}_{(n)})_{i_n k} = \sum_{i_n}^{I_n} (\mathcal{A})_{i_1 i_2 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} (U)_{j_n i_n}.$$

Hence, $(\mathcal{A} \times_n U)_{(n)} = U\mathcal{A}_{(n)}$, as required. ∎

By considering a vector $\mathbf{v}$ as an $I_n \times 1$ matrix, then an n-mode product of $\mathbf{v}^\top$ and $\mathcal{A}$ can be formed to produce an $I_1 \times I_2 \times \ldots \times I_{n-1} \times 1 \times I_{n+1} \times \ldots \times I_N$-tensor. This tensor could be viewed as an $N-1$-tensor, but instead we leave it as an N-tensor in order that we can form other m-mode products without the value of m having to change. However, if we have a $1 \times 1 \times \ldots \times 1$-tensor then we simply view this as a scalar. The n-mode produce satisfies the following property:

Property 1. For a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and the matrices $F \in \mathbb{R}^{J_n \times I_n}$ and $G \in \mathbb{R}^{J_m \times I_m}$, $n \neq m$, we have

$$(\mathcal{A} \times_n F) \times_m G = (\mathcal{A} \times_m G) \times_n F = \mathcal{A} \times_n F \times_m G.$$

We also introduce the Frobenius norm of a tensor.

DEFINITION 4.6. *The Frobenius-norm, $\|\cdot\|_F$, of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is given by*

$$\|\mathcal{A}\|_F^2 = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} (\mathcal{A})_{i_1 i_2 \cdots i_N}^2.$$

LEMMA 4.7. *Given a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and an orthonormal matrix $F \in \mathbb{R}^{I_n \times I_n}$, the following holds*

$$\|\mathcal{A} \times_n F\|_F = \|\mathcal{A}\|_F. \tag{4.8}$$

Proof. For a matrix $A \in \mathbb{R}^{I_n \times m}$ we have that

$$\|FA\|_F = \|A\|_F. \tag{4.9}$$

Using the identity in Lemma 4.5, namely, $(\mathcal{A} \times_n F)_{(n)} = F\mathcal{A}_{(n)}$, we deduce that

$$\|\mathcal{A} \times_n F\|_F = \|F\mathcal{A}_{(n)}\|_F.$$

Given that $\mathcal{A}_{(n)} \in \mathbb{R}^{I_n \times I_{n+1} \ldots I_N \ldots I_1 \ldots I_{n-1}}$, exploiting (4.9) gives

$$\|\mathcal{A} \times_n F\|_F = \|F\mathcal{A}_{(n)}\|_F = \|\mathcal{A}_{(n)}\|_F = \|\mathcal{A}\|_F.$$

∎

In order to rescale $|\hat{v}|_{H^s(\hat{\kappa})}$ to the corresponding quantity on $\tilde{\kappa}$, we first note that

$$|\hat{v}|_{H^s(\hat{\kappa})}^2 = \int_{\hat{\kappa}} \|\hat{\mathcal{D}}^s(\hat{v})\|_F^2 \, \mathrm{d}\hat{x},$$

where $\hat{\mathcal{D}}^s(\hat{v}) \in \mathbb{R}^{d \times d \times \cdots \times d}$ is the sth–order tensor containing the sth–order derivatives of $\hat{v}$ with respect to the coordinate system $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_d)$, i.e.,

$$(\hat{\mathcal{D}}^s(\hat{v}))_{i_1, i_2, \ldots, i_s} = \frac{\partial^s \hat{v}}{\partial \hat{x}_{i_1} \cdots \partial \hat{x}_{i_s}}, \quad i_k = 1, \ldots, d, \text{ for } k = 1, \ldots, s.$$

Thereby, for $s = 0$, $\hat{\mathcal{D}}^s(\hat{v}) = \hat{v}$, for $s = 1$, $\hat{\mathcal{D}}^s(\hat{v})$ is the gradient vector, and for $s = 2$, $\hat{\mathcal{D}}^s(\hat{v})$ is the Hessian matrix of second–order derivatives. Writing $\tilde{\mathcal{D}}^s(\tilde{v}) \in \mathbb{R}^{d \times d \times \cdots \times d}$ to denote the sth–order tensor containing the sth–order derivatives of $\tilde{v}$ with respect

to the coordinate system $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_d)$, we now state the following lemma relating $|\hat{v}|^2_{H^s(\hat{\kappa})}$ to $|\tilde{v}|^2_{H^s(\tilde{\kappa})}$.

LEMMA 4.8. *Under the foregoing assumptions, for $\tilde{v} \in H^s(\tilde{\kappa})$, $s \geq 0$, we have that*

$$|\hat{v}|^2_{H^s(\hat{\kappa})} = |\det(J_{F_\kappa}^{-1})| \int_{\tilde{\kappa}} \|\tilde{\mathcal{D}}^s(\tilde{v}) \times_1 J_{F_\kappa}^\top \times_2 J_{F_\kappa}^\top \times_3 \ldots \times_s J_{F_\kappa}^\top\|_F^2 \, \mathrm{d}\tilde{x}.$$

Proof. The case when $s = 0$ follows trivially. For $s \geq 1$, we first note that the entry $(\hat{\mathcal{D}}^s(\hat{v}))_{i_1 i_2 \ldots i_s}$ may be written in the form

$$\frac{\partial^s \hat{v}}{\partial \hat{x}_{i_1} \cdots \partial \hat{x}_{i_s}} = \sum_{j_1=1}^{d} \cdots \sum_{j_s=1}^{d} (J_{F_\kappa})_{j_1 i_1} \cdots (J_{F_\kappa})_{j_s i_s} \frac{\partial^s \tilde{v}}{\partial \tilde{x}_{j_1} \cdots \partial \tilde{x}_{j_s}},$$

for $i_k = 1, \ldots, d$ and $k = 1, \ldots, s$; this follows by employing an induction argument together with the chain rule. Thereby, from Definition 4.4 and Property 1 above, we deduce that

$$\hat{\mathcal{D}}^s(\hat{v}) = \tilde{\mathcal{D}}^s(\tilde{v}) \times_1 J_{F_\kappa}^\top \times_2 J_{F_\kappa}^\top \times_3 \ldots \times_s J_{F_\kappa}^\top. \tag{4.10}$$

The statement of the lemma now follows by a simple change of variables. □

REMARK 4.9. *For the case when $s = 0$, Lemma 4.8 simply states the change of variable formula for the L_2-norm. For $s = 1$ we note that (4.10) gives rise to the usual change of variables for the gradient operator, namely,*

$$\hat{\mathcal{D}}^s(\hat{v}) \equiv \nabla_{\hat{x}} \hat{v} = \tilde{\mathcal{D}}^s(\tilde{v}) \times_1 J_{F_\kappa}^\top = J_{F_\kappa}^\top \nabla_{\tilde{x}} \tilde{v},$$

where $\nabla_{\hat{x}}$ and $\nabla_{\tilde{x}}$ denote the gradient operator with respect to the coordinate systems $\hat{x}$ and $\tilde{x}$, respectively. Similarly, for $s = 2$, (4.10) may be written in the more familiar form $H_{\hat{x}}(\hat{v}) = J_{F_\kappa}^\top H_{\tilde{x}}(\tilde{v}) J_{F_\kappa}$, where $H_{\hat{x}}(\cdot)$ and $H_{\tilde{x}}(\cdot)$ denote the Hessian matrix operators with respect to the coordinate systems $\hat{x}$ and $\tilde{x}$, respectively, cf. [10].

In order to describe the length scales and orientation of the element $\tilde{\kappa}$ we adopt a similar approach to that developed in [10]. Namely, we perform an SVD decomposition of the Jacobi matrix J_{F_κ} of the affine element mapping F_κ. Thereby, we write

$$J_{F_\kappa} = U_\kappa \Sigma_\kappa V_\kappa^\top,$$

where U_κ and V_κ are $d \times d$ orthogonal matrices containing the left and right singular vectors of J_{F_κ}, and $\Sigma_\kappa = \mathrm{diag}(\sigma_{1,\kappa}, \sigma_{2,\kappa}, \ldots, \sigma_{d,\kappa})$ is a $d \times d$ diagonal matrix containing the singular values $\sigma_{i,\kappa}$, $i = 1, \ldots, d$, of J_{F_κ}. By convention, we assume that $\sigma_{1,\kappa} \geq \sigma_{2,\kappa} \geq \ldots \geq \sigma_{d,\kappa} > 0$. Writing $U_\kappa = (\mathbf{u}_{1,\kappa} \ldots \mathbf{u}_{d,\kappa})$, where $\mathbf{u}_{i,\kappa}$, $i = 1, \ldots, d$, denote the left singular vectors of J_{F_κ}, we note that $\mathbf{u}_{i,\kappa}$, $i = 1, \ldots, d$, give the direction of stretching of the element κ, while $\sigma_{i,\kappa}$, $i = 1, \ldots, d$, give the stretching lengths in the respective directions. Indeed, for axiparallel meshes, as considered in [11], for example, then $\mathbf{u}_{i,\kappa}$, $i = 1, \ldots, d$ will be parallel to the coordinates axes and $\sigma_{i,\kappa}$, $i = 1, \ldots, d$, will denote the local mesh lengths within the respective coordinate direction.

With this notation, we make the following observations

$$|\det(J_{F_\kappa})| = \Pi_{i=1}^d \sigma_{i,\kappa}, \quad \|J_{F_\kappa}^{-\top}\|_2 = 1/\sigma_{d,\kappa}, \quad m_f \leq C_7 \, \Pi_{i=1}^{d-1} \sigma_{i,\kappa}, \tag{4.11}$$

where C_7 is a positive constant independent of the element size. Employing Lemma 4.7, we note that

$$\begin{aligned}
&\|\tilde{\mathcal{D}}^s(\tilde{v}) \times_1 J_{F_\kappa}^\top \times_2 J_{F_\kappa}^\top \times_3 \ldots \times_s J_{F_\kappa}^\top\|_F^2 \\
&\quad = \sum_{i_1=1}^{d}\sum_{i_2=1}^{d}\ldots\sum_{i_s=1}^{d}(\sigma_{i_1,\kappa}\sigma_{i_2,\kappa}\ldots\sigma_{i_s,\kappa})^2(\tilde{\mathcal{D}}^s(\tilde{v}) \times_1 \mathbf{u}_{i_1,\kappa}^\top \times_2 \mathbf{u}_{i_2,\kappa}^\top \times_3 \ldots \times_s \mathbf{u}_{i_s,\kappa}^\top)^2 \\
&\quad \equiv D_{\tilde{\kappa}}^s(\tilde{v}, \Sigma_\kappa, U_\kappa). \qquad (4.12)
\end{aligned}$$

Thereby, exploiting (4.11) and (4.12) together with Corollary 4.2, we deduce the following approximation result.

THEOREM 4.10. *Using the notation of Lemma 4.1, there exists a positive constant C, which depends only on the dimension d and the polynomial order p, such that for $m = 0, 1$:*

$$|v - \Pi_p v|_{H^m(\kappa)} \le C|\sigma_{d,\kappa}|^{-m}\left[\int_{\tilde{\kappa}} D_{\tilde{\kappa}}^s(\tilde{v}, \Sigma_\kappa, U_\kappa)\,\mathrm{d}\tilde{x}\right]^{1/2}, \quad m \le s \le \min(p+1, k),$$

$$\|v - \Pi_p v\|_{L_2(f)} \le C|\sigma_{d,\kappa}|^{-1/2}\left[\int_{\tilde{\kappa}} D_{\tilde{\kappa}}^s(\tilde{v}, \Sigma_\kappa, U_\kappa)\,\mathrm{d}\tilde{x}\right]^{1/2}, \quad 1 \le s \le \min(p+1, k),$$

$$|v - \Pi_p v|_{H^1(f)} \le C\left|\frac{m_f}{m_\kappa}\right|^{1/2}|\sigma_{d,\kappa}|^{-1}\left[\int_{\tilde{\kappa}} D_{\tilde{\kappa}}^s(\tilde{v}, \Sigma_\kappa, U_\kappa)\,\mathrm{d}\tilde{x}\right]^{1/2}, \ 2 \le s \le \min(p+1, k).$$

REMARK 4.11. *For the purposes of deriving the forthcoming* a priori *error bound on the error in the computed target functional, cf. Theorem 5.4 below, it is convenient to leave the statement of the third approximation result above in terms of m_f and m_κ, rather than in terms of the stretching factors $\sigma_{i,\kappa}$, $i = 1, \ldots, d$, solely, since these quantities naturally arise within the definition of the discontinuity-penalization parameter σ defined in (3.8).*

In the next section, we consider the *a posteriori* and *a priori* error analysis of the discontinuous Galerkin finite element method (2.9) in terms of certain linear target functionals of practical interest.

5. *A posteriori* and *a priori* error analysis.

Very often in problems of practical importance the quantity of interest is an output or target functional $J(\cdot)$ of the solution. Relevant examples include the lift and drag coefficients for a body immersed into a viscous fluid, the local mean value of the field, or its flux through the outflow boundary of the computational domain. The aim of this section is to develop the *a posteriori* and *a priori* error analysis for general linear target functionals $J(\cdot)$ of the solution; for related work, we refer to [5, 14, 18, 23, 20], for example.

5.1. Type I *a posteriori* error analysis.

In this section we consider the derivation of so-called Type I (cf. [18]) or weighted *a posteriori* error bounds. Following the argument presented in [18, 20] we begin our analysis by considering the following *dual* or *adjoint* problem: find $z \in H^2(\Omega, \mathcal{T}_h)$ such that

$$B_{\mathrm{DG}}(w, z) = J(w) \quad \forall w \in H^2(\Omega, \mathcal{T}_h). \qquad (5.1)$$

Let us assume that (5.1) possesses a unique solution. Clearly, the validity of this assumption depends on the choice of the linear functional under consideration; see the discussion in [18].

For a given linear functional $J(\cdot)$ the proceeding *a posteriori* error bound will be expressed in terms of the finite element residual $R_{\rm int}$ defined on $\kappa \in \mathcal{T}_h$ by $R_{\rm int}|_\kappa = (f - \mathcal{L}u_{\rm DG})|_\kappa$, which measures the extent to which $u_{\rm DG}$ fails to satisfy the differential equation on the union of the elements κ in the mesh $\mathcal{T}_h$; thus we refer to $R_{\rm int}$ as the *internal residual*. Also, since $u_{\rm DG}$ only satisfies the boundary conditions approximately, the differences $g_{\rm D} - u_{\rm DG}$ and $g_{\rm N} - (a\nabla u_{\rm DG})\cdot\mathbf{n}$ are not necessarily zero on $\Gamma_{\rm D}\cup\Gamma_-$ and $\Gamma_{\rm N}$, respectively; thus we define the *boundary residuals* $R_{\rm D}$ and $R_{\rm N}$, respectively, by

$$R_{\rm D}|_{\partial\kappa\cap(\Gamma_{\rm D}\cup\Gamma_-)} = (g_{\rm D} - u^+_{\rm DG})|_{\partial\kappa\cap(\Gamma_{\rm D}\cup\Gamma_-)}\,, \quad R_{\rm N}|_{\partial\kappa\cap\Gamma_{\rm N}} = (g_{\rm N} - (a\nabla u^+_{\rm DG})\cdot\mathbf{n})|_{\partial\kappa\cap\Gamma_{\rm N}}.$$

With this notation, after application of the divergence theorem, the Galerkin orthogonality condition (3.10) may be written in the following equivalent form:

$$\begin{aligned} 0 = B_{\rm DG}(u-u_{\rm DG},v) &= \ell_{\rm DG}(v) - B_{\rm DG}(u_{\rm DG},v) \qquad (5.2)\\ &= \sum_{\kappa\in\mathcal{T}_h}\Bigg(\int_\kappa R_{\rm int}v\,{\rm d}x - \int_{\partial_-\kappa\cap\Gamma}(\mathbf{b}\cdot\mathbf{n}_\kappa)\,R_{\rm D}v^+\,{\rm d}s + \int_{\partial_-\kappa\setminus\Gamma}(\mathbf{b}\cdot\mathbf{n}_\kappa)\,[u_{\rm DG}]v^+\,{\rm d}s\\ &\quad - \int_{\partial\kappa\cap\Gamma_{\rm D}} R_{\rm D}((a\nabla v^+)\cdot\mathbf{n}_\kappa)\,{\rm d}s + \int_{\partial\kappa\cap\Gamma_{\rm D}}\vartheta R_{\rm D}v^+\,{\rm d}s + \int_{\partial\kappa\cap\Gamma_{\rm N}} R_{\rm N}v^+\,{\rm d}s\\ &\quad + \frac{1}{2}\int_{\partial\kappa\setminus\Gamma}\left\{\,[u_{\rm DG}](a\nabla v^+)\cdot\mathbf{n}_\kappa - [(a\nabla u_{\rm DG})\cdot\mathbf{n}_\kappa]v^+\right\}{\rm d}s - \int_{\partial\kappa\setminus\Gamma}\vartheta[u_{\rm DG}]v^+\,{\rm d}s\Bigg) \end{aligned}$$

for all $v \in S_{h,p}$. Here, we have employed the result $\sum_{j=1}^d a_{ij}(x)\mathbf{n}_j = 0$ on $\Gamma\setminus\Gamma_0$, $i = 1,\ldots,d$, cf. [19]. The starting point for the analysis is the following general result.

THEOREM 5.1. *Let u and $u_{\rm DG}$ denote the solutions of (2.1), (2.3) and (2.9), respectively, and suppose that the dual solution z is defined by (5.1). Then, the following error representation formula holds:*

$$J(u) - J(u_{\rm DG}) = \mathcal{E}_\Omega(u_{\rm DG},h,p,z-z_{h,p}) \equiv \sum_{\kappa\in\mathcal{T}_h}\eta_\kappa\,, \qquad (5.3)$$

where

$$\begin{aligned} \eta_\kappa &= \int_\kappa R_{\rm int}(z-z_{h,p})\,{\rm d}x - \int_{\partial_-\kappa\cap\Gamma}(\mathbf{b}\cdot\mathbf{n}_\kappa)\,R_{\rm D}(z-z_{h,p})^+\,{\rm d}s\\ &\quad + \int_{\partial_-\kappa\setminus\Gamma}(\mathbf{b}\cdot\mathbf{n}_\kappa)\,[u_{\rm DG}](z-z_{h,p})^+\,{\rm d}s - \int_{\partial\kappa\cap\Gamma_{\rm D}} R_{\rm D}((a\nabla(z-z_{h,p})^+)\cdot\mathbf{n}_\kappa)\,{\rm d}s\\ &\quad + \int_{\partial\kappa\cap\Gamma_{\rm D}}\vartheta R_{\rm D}(z-z_{h,p})^+\,{\rm d}s + \int_{\partial\kappa\cap\Gamma_{\rm N}} R_{\rm N}(z-z_{h,p})^+\,{\rm d}s - \int_{\partial\kappa\setminus\Gamma}\vartheta[u_{\rm DG}](z-z_{h,p})^+\,{\rm d}s\\ &\quad + \frac{1}{2}\int_{\partial\kappa\setminus\Gamma}\left\{\,[u_{\rm DG}](a\nabla(z-z_{h,p})^+)\cdot\mathbf{n}_\kappa - [(a\nabla u_{\rm DG})\cdot\mathbf{n}_\kappa](z-z_{h,p})^+\right\}{\rm d}s \qquad (5.4) \end{aligned}$$

for all $z_{h,p} \in S_{h,p}$.

Proof. On choosing $w = u - u_{\rm DG}$ in (5.1) and recalling the linearity of $J(\cdot)$ and the Galerkin orthogonality property (5.2), we deduce that

$$J(u) - J(u_{\rm DG}) = J(u-u_{\rm DG}) = B_{\rm DG}(u-u_{\rm DG},z) = B_{\rm DG}(u-u_{\rm DG},z-z_{h,p}), \qquad (5.5)$$

and hence (5.3). ∎

Thereby, on application of the triangle inequality, we deduce the following Type I *a posteriori* error bound.

COROLLARY 5.2. *Under the assumptions of Theorem 5.1, the following Type I* a posteriori *error bound holds:*

$$|J(u) - J(u_{\mathrm{DG}})| \leq \mathcal{E}_{|\Omega|}(u_{\mathrm{DG}}, h, p, z - z_{h,p}) \equiv \sum_{\kappa \in \mathcal{T}_h} |\eta_\kappa| \,, \tag{5.6}$$

where η_κ is defined as in (5.4).

As discussed in [14, 20], the local weighting terms involving the difference between the dual solution z and its projection/interpolant $z_{h,p}$ onto $S_{h,p}$ appearing in the Type I bound (5.6) provide invaluable information concerning the global transport of the error. Thereby, we refrain from eliminating the weighting terms involving the (unknown) dual solution z and approximate z numerically; this will be discussed in Section 6.

5.2. *A priori* error bounds. In this section we derive an *a priori* error bound for the interior penalty DGFEM introduced in Section 2.2. To this end, we decompose the global error $u - u_{\mathrm{DG}}$ as

$$u - u_{\mathrm{DG}} = (u - \Pi_p u) + (\Pi_p u - u_{\mathrm{DG}}) \equiv \eta + \xi \,, \tag{5.7}$$

where Π_p denotes the L_2–projection operator introduced in Section 4. With these definitions we have the following result.

LEMMA 5.3. *Assume that (2.4) and (3.11) hold and let $\gamma_1|_\kappa = \|c/c_0\|^2_{L_\infty(\kappa)}$; then the functions ξ and η defined by (5.7) satisfy the following inequality*

$$|||\xi|||^2 \leq C \left(\sum_{\kappa \in \mathcal{T}_h} \left(\|\sqrt{a}\nabla\eta\|^2_{L_2(\kappa)} + \gamma_1 \|\eta\|^2_{L_2(\kappa)} + \|\eta^+\|^2_{\partial_+\kappa\cap\Gamma} + \|\eta^-\|^2_{\partial_-\kappa\setminus\Gamma} \right) \right.$$
$$\left. + \int_{\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}} \frac{1}{\vartheta} \langle (a\nabla\eta)\cdot\mathbf{n}_f \rangle^2 \,\mathrm{d}s + \int_{\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}} \vartheta [\eta]^2 \,\mathrm{d}s \right) ,$$

where C is a positive constant that depends only on the dimension d and the polynomial degree p.

Proof. From the Galerkin orthogonality condition (3.10), we deduce that $B_{\mathrm{DG}}(\xi, \xi) = -B_{\mathrm{DG}}(\eta, \xi)$, where ξ and η are as defined in (5.7). Thereby, employing the coercivity result stated in Theorem 3.3, gives

$$|||\xi|||^2 \leq -\frac{1}{C} B_{\mathrm{DG}}(\eta, \xi) \,. \tag{5.8}$$

Using the identity (4.1), the right–hand side of (5.8) may be bounded as follows:

$$B_{\mathrm{DG}}(\eta, \xi) \leq C|||\xi||| \left(\sum_{\kappa \in \mathcal{T}_h} \left(\|\sqrt{a}\nabla\eta\|^2_{L_2(\kappa)} + \gamma_1 \|\eta\|^2_{L_2(\kappa)} + \|\eta^+\|^2_{\partial_+\kappa\cap\Gamma} \right.\right.$$
$$\left.\left. + \|\eta^-\|^2_{\partial_-\kappa\setminus\Gamma} \right) + \int_{\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}} \frac{1}{\vartheta} \langle (a\nabla\eta)\cdot\mathbf{n}_f \rangle^2 \,\mathrm{d}s + \int_{\Gamma_{\mathrm{int}}\cup\Gamma_{\mathrm{D}}} \vartheta [\eta]^2 \,\mathrm{d}s \right)^{1/2} ; \tag{5.9}$$

see [9, 17] for details. Substituting (5.9) into (5.8) gives the desired result. □

For the rest of this section, let us now assume that the volume of the elements, denoted by m_κ for each $\kappa \in \mathcal{T}_h$, cf. above, has *bounded local variation*; *i.e.*, there

exists a constant $C_8 \geq 1$ such that, for any pair of elements κ and κ' which share a $(d-1)$–dimensional face,

$$C_8^{-1} \leq m_\kappa / m_{\kappa'} \leq C_8 \,. \tag{5.10}$$

With this hypothesis, we now proceed to prove the main result of this section.

THEOREM 5.4. *Let $\Omega \subset \mathbb{R}^d$ be a bounded polyhedral domain, $\mathcal{T}_h = \{\kappa\}$ a subdivision of Ω, such that the elemental volumes satisfy the bounded local variation condition (5.10). Then, assuming that conditions (2.4), (2.8), and (3.11) on the data hold, and $u \in H^k(\Omega, \mathcal{T}_h)$, $k \geq 2$, $z \in H^l(\Omega, \mathcal{T}_h)$, $l \geq 2$, then the solution $u_{\mathrm{DG}} \in S_{h,p}$ of (2.9) obeys the error bound*

$$\begin{aligned} |J(u) - J(u_{DG})|^2 \leq C \left(\sum_{\kappa \in \mathcal{T}_h} \left\{ \frac{\alpha}{\sigma_{d,\kappa}^2} + \frac{\beta_2}{\sigma_{d,\kappa}} + (\beta_1 + \gamma_1) \right\} \int_{\tilde{\kappa}} D_{\tilde{\kappa}}^s(\tilde{u}, \Sigma_\kappa, U_\kappa) \, \mathrm{d}\tilde{x} \right) \\ \times \left(\sum_{\kappa \in \mathcal{T}_h} \left\{ \frac{\alpha}{\sigma_{d,\kappa}^2} + \frac{\beta_2}{\sigma_{d,\kappa}} + (\beta_1 + \gamma_2) \right\} \int_{\tilde{\kappa}} D_{\tilde{\kappa}}^t(\tilde{z}, \Sigma_\kappa, U_\kappa) \, \mathrm{d}\tilde{x} \right), \end{aligned}$$

for $2 \leq s \leq \min(p+1, k)$ and $2 \leq t \leq \min(p+1, l)$, where $\alpha|_\kappa = \bar{a}_{\tilde{\kappa}}$, $\beta_1|_\kappa = \|c + \nabla \cdot \mathbf{b}\|_{L_\infty(\kappa)}$, $\beta_2|_\kappa = \|\mathbf{b}\|_{L_\infty(\kappa)}$, $\gamma_1|_\kappa = \|c/c_0\|_{L_\infty(\kappa)}^2$, $\gamma_2|_\kappa = \|(c + \nabla \cdot \mathbf{b})/c_0\|_{L_\infty(\kappa)}^2$, for all $\kappa \in \mathcal{T}_h$. Here, C is a constant depending on the dimension d, the polynomial degree p, and the parameters C_i, $i = 1, \ldots, 8$.

Proof. Decomposing the error $u - u_{\mathrm{DG}}$ as in (5.7), we note that the error in the target functional $J(\cdot)$ may be expressed as follows:

$$J(u) - J(u_{\mathrm{DG}}) = B_{\mathrm{DG}}(\eta, z - z_{h,p}) + B_{\mathrm{DG}}(\xi, z - z_{h,p}) \equiv \mathrm{I} + \mathrm{II}. \tag{5.11}$$

Let us first deal with term I. To this end, we define $z_{h,p} = \Pi_p z$ and $w = z - z_{h,p}$; after a lengthy, but straightforward calculation, we deduce that

$$\begin{aligned} \mathrm{I}^2 \leq C \Bigg(\sum_{\kappa \in \mathcal{T}_h} \Big\{ \|\sqrt{a}\nabla\eta\|_{L_2(\kappa)}^2 + \beta_1 \|\eta\|_{L_2(\kappa)}^2 + \beta_2 \epsilon_\kappa^{-1} \|\nabla\eta\|_{L_2(\kappa)}^2 + \|[\eta]\|_{\partial_- \kappa}^2 \\ + \|\vartheta^{-1/2} \langle a \nabla \eta \rangle\|_{L_2(\partial\kappa \cap (\Gamma_{\mathrm{int}} \cup \Gamma_D))}^2 + \|\vartheta^{1/2}[\eta]\|_{L_2(\partial\kappa \cap (\Gamma_{\mathrm{int}} \cup \Gamma_D))}^2 \Big\} \Bigg) \\ \times \Bigg(\sum_{\kappa \in \mathcal{T}_h} \Big\{ \|\sqrt{a}\nabla w\|_{L_2(\kappa)}^2 + \beta_1 \|w\|_{L_2(\kappa)}^2 + \beta_2 \epsilon_\kappa \|w\|_{L_2(\kappa)}^2 + \|w\|_{\partial_- \kappa}^2 \\ + \|\vartheta^{-1/2} \langle a \nabla w \rangle\|_{L_2(\partial\kappa \cap (\Gamma_{\mathrm{int}} \cup \Gamma_D))}^2 + \|\vartheta^{1/2}[w]\|_{L_2(\partial\kappa \cap (\Gamma_{\mathrm{int}} \cup \Gamma_D))}^2 \Big\} \Bigg), \end{aligned} \tag{5.12}$$

for any set of real positive numbers ϵ_κ, $\kappa \in \mathcal{T}_h$. Let us now consider Term II. Here, we note that a bound analogous to (5.9) in the proof of Lemma 5.3 holds with η and ξ replaced by ξ and w in (5.9), respectively. Indeed, in this case we have that

$$\begin{aligned} |B_{DG}(\xi, w)| \leq |||\xi||| \times \Bigg[\sum_{\kappa \in \mathcal{T}_h} \Big(\|\sqrt{a}\nabla w\|_{L_2(\kappa)}^2 + \gamma_2 \|w\|_{L_2(\kappa)}^2 + \|w^+\|_{\partial_- \kappa}^2 \\ + \|\vartheta^{1/2}[w]\|_{L_2(\partial\kappa \cap (\Gamma_{\mathrm{int}} \cup \Gamma_D))}^2 + \|\vartheta^{-1/2} \langle a \nabla w \rangle\|_{L_2(\partial\kappa \cap (\Gamma_{\mathrm{int}} \cup \Gamma_D))}^2 \Big) \Bigg]^{\frac{1}{2}}. \end{aligned} \tag{5.13}$$

Thereby, employing Lemma 5.3 in (5.13) and inserting the result and (5.12) into (5.11) we deduce that

$$
\begin{aligned}
|J(u)-J(u_{DG})|^2 \leq C \Bigg(& \sum_{\kappa\in\mathcal{T}_h} \Big\{ \|\sqrt{a}\nabla\eta\|^2_{L_2(\kappa)} + (\beta_1+\gamma_1)\,\|\eta\|^2_{L_2(\kappa)} + \beta_2\epsilon_\kappa^{-1}\|\nabla\eta\|^2_{L_2(\kappa)} \\
& + \|\eta^+\|^2_{\partial_+\kappa\cap\Gamma} + \|\eta^-\|^2_{\partial_-\kappa\setminus\Gamma} + \|[\eta]\|^2_{\partial_-\kappa} \\
& + \|\vartheta^{-1/2}\langle a\nabla\eta\rangle\|^2_{L_2(\partial\kappa\cap(\Gamma_{\rm int}\cup\Gamma_D))} + \|\vartheta^{1/2}[\eta]\|^2_{L_2(\partial\kappa\cap(\Gamma_{\rm int}\cup\Gamma_D))} \Big\} \Bigg) \\
& \times \Bigg(\sum_{\kappa\in\mathcal{T}_h} \Big\{ \|\sqrt{a}\nabla w\|^2_{L_2(\kappa)} + (\beta_1+\beta_2\epsilon_\kappa+\gamma_2)\,\|w\|^2_{L_2(\kappa)} \\
& + \|w^+\|^2_{\partial_-\kappa} + \|\vartheta^{-1/2}\langle a\nabla w\rangle\|^2_{L_2(\partial\kappa\cap(\Gamma_{\rm int}\cup\Gamma_D))} \\
& + \|\vartheta^{1/2}[w]\|^2_{L_2(\partial\kappa\cap(\Gamma_{\rm int}\cup\Gamma_D))} \Big\} \Bigg)\,. \qquad (5.14)
\end{aligned}
$$

After application of Theorem 4.10 gives

$$
\begin{aligned}
|J(u)-J(u_{DG})|^2 \leq C \Bigg(& \sum_{\kappa\in\mathcal{T}_h} \Bigg\{ \frac{\bar{a}_\kappa}{\sigma^2_{d,\kappa}} \Bigg[1 + \frac{\bar{a}_\kappa}{m_\kappa} \sum_{f\subset\partial\kappa} \frac{m_f}{\vartheta_f} + \frac{\sigma_{d,\kappa}\sum_{f\subset\partial\kappa}\vartheta_f}{\bar{a}_\kappa} \Bigg] \\
& + \frac{\beta_2}{\sigma_{d,\kappa}} \Bigg[1 + \frac{1}{\epsilon_\kappa\sigma_{d,\kappa}} \Bigg] + (\beta_1+\gamma_1) \Bigg\} \int_{\tilde{\kappa}} D^s_{\tilde{\kappa}}(\tilde{u},\Sigma_\kappa,U_\kappa)\,\mathrm{d}\tilde{x} \Bigg) \\
\times \Bigg(& \sum_{\kappa\in\mathcal{T}_h} \Bigg\{ \frac{\bar{a}_\kappa}{\sigma^2_{d,\kappa}} \Bigg[1 + \frac{\bar{a}_\kappa}{m_\kappa} \sum_{f\subset\partial\kappa} \frac{m_f}{\vartheta_f} + \frac{\sigma_{d,\kappa}\sum_{f\subset\partial\kappa}\vartheta_f}{\bar{a}_\kappa} \Bigg] \\
& + \frac{\beta_2}{\sigma_{d,\kappa}} \left[1 + \epsilon_\kappa\sigma_{d,\kappa} \right] + (\beta_1+\gamma_2) \Bigg\} \int_{\tilde{\kappa}} D^t_{\tilde{\kappa}}(\tilde{z},\Sigma_\kappa,U_\kappa)\,\mathrm{d}\tilde{x} \Bigg)\,.
\end{aligned}
$$

The statement of theorem now follows by selecting $\epsilon_\kappa = 1/\sigma_{d,\kappa}$, for each $\kappa\in\mathcal{T}_h$, and employing the definition of the discontinuity-penalization parameter ϑ stated in (3.8), together with the bounded variation of the elemental volumes (5.10) and (4.11). ∎

REMARK 5.5. *The above result represents an extension of the* a priori *error bound derived in the article [13] to the case when general anisotropic computational meshes are employed. We note that although the analysis presented in [13] assumed shape–regular meshes, the explicit dependence of the polynomial degree was retained in the resulting* a priori *error bound; however, following the arguments in [13] an analogous hp–version bound of the form stated in Theorem 5.4 may easily be deduced.*

REMARK 5.6. *The* a priori *bound stated in Theorem 5.4 clearly highlights that in order to minimize the error in the computed target functional $J(\cdot)$, the design of an optimal mesh must exploit anisotropic information emanating from both the primal and dual solutions u and z, respectively. Indeed, a mesh solely optimized for u may be completely inappropriate for z, and vice versa, thus there must me a trade-off between aligning the elements with respect to either solution in order to minimize the overall error in $J(\cdot)$.*

6. Adaptive algorithm. For a user-defined tolerance TOL, we now consider the problem of designing an appropriate finite element mesh $\mathcal{T}_h$ such that

$$
|J(u)-J(u_{\rm DG})| \leq \texttt{TOL}\,,
$$

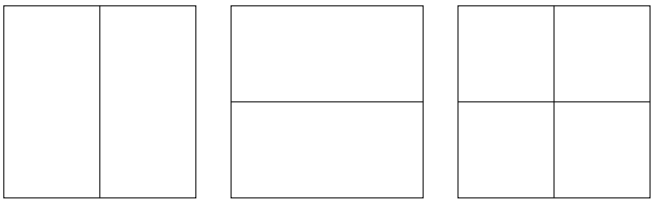

FIG. 6.1. *Cartesian refinement in 2D: (a) & (b) Anisotropic refinement; (c) Isotropic refinement.*

subject to the constraint that the total number of elements in $\mathcal{T}_h$ is minimized; for simplicity of presentation, in this section we only consider the case when $\Omega \subset \mathbb{R}^2$ and $\mathcal{T}_h$ consists of *1–irregular* quadrilateral elements. Following the discussion presented [18], we exploit the *a posteriori* error bound (5.6) with z replaced by a discontinuous Galerkin approximation $\hat{z}$ computed on the same mesh $\mathcal{T}_h$ used for the primal solution u_{DG}, but with a higher degree polynomial, i.e., $\hat{z} \in S_{h,\hat{p}}$, $\hat{p} = p + p_{\mathrm{inc}}$; in Section 7, we set $p_{\mathrm{inc}} = 1$, cf. [14, 20]. Thereby, in practice we enforce the stopping criterion

$$\hat{\mathcal{E}}_{|\Omega|} \equiv \mathcal{E}_{|\Omega|}(u_{\mathrm{DG}}, \hat{z} - z_{h,p}) \leq \mathtt{TOL}\,. \tag{6.1}$$

If (6.1) is not satisfied, then the elements are marked for refinement/derefinement according to the size of the (approximate) error indicators $|\hat{\eta}_\kappa|$; these are defined analogously to $|\eta_\kappa|$ in (5.4) with z replaced by $\hat{z}$. In Section 7 we use the fixed fraction mesh refinement algorithm, with refinement and derefinement fractions set to 20% and 10%, respectively.

To subdivide the elements which have been flagged for refinement, we employ a simple Cartesian refinement strategy; here, elements may be subdivided either anisotropically or isotropically according to the three refinements (in two–dimensions, i.e., $d = 2$) depicted in Figure 6.1. In order to determine the optimal refinement, stimulated by the articles [28, 29], we propose the following two strategies based on choosing the most competitive subdivision of κ from a series of trial refinements, whereby an approximate local error indicator on each trial patch is determined.

`Algorithm 1:` Given an element κ in the computational mesh $\mathcal{T}_h$ (which has been marked for refinement), we first construct the mesh patches $\mathcal{T}_{h,i}$, $i = 1, 2, 3$, based on refining κ according to Figures 6.1(a), (b), & (c), respectively. On each mesh patch, $\mathcal{T}_{h,i}$, $i = 1, 2, 3$, we compute the approximate error estimators

$$\hat{\mathcal{E}}_{\kappa,i}(u_{\mathrm{DG},i}, \hat{z}_i - z_{h,p}) = \sum_{\kappa' \in \mathcal{T}_{h,i}} \eta_{\kappa',i},$$

for $i = 1, 2, 3$, respectively. Here, $u_{\mathrm{DG},i}$, $i = 1, 2, 3$, is the discontinuous Galerkin approximation to (2.1), (2.3) computed on the mesh patch $\mathcal{T}_{h,i}$, $i = 1, 2, 3$, respectively, based on enforcing appropriate boundary conditions on $\partial\kappa$ computed from the original discontinuous Galerkin solution u_{DG} on the portion of the boundary $\partial\kappa$ of κ which is interior to the computational domain Ω, i.e., where $\partial\kappa \cap \Gamma = \emptyset$. Similarly, $\hat{z}_i$ denotes the discontinuous Galerkin approximation to z computed on the local mesh patch $\mathcal{T}_{h,i}$, $i = 1, 2, 3$, respectively, with polynomials of degree $\hat{p}$, based on employing suitable boundary conditions on $\partial\kappa \cap \Gamma = \emptyset$ derived from $\hat{z}$. Finally, $\eta_{\kappa',i}$, $i = 1, 2, 3$, is defined in an analogous manner to η_κ, cf. (5.4) above, with u_{DG} and z replaced by $u_{\mathrm{DG},i}$ and $\hat{z}_i$, respectively.

The element κ is then refined according to the subdivision of κ which satisfies

$$\min_{i=1,2,3} \frac{|\eta_\kappa| - |\hat{\mathcal{E}}_{\kappa,i}(u_{\mathrm{DG},i}, \hat{z}_i - z_{h,p})|}{\#\mathrm{dofs}(\mathcal{T}_{h,i}) - \#\mathrm{dofs}(\kappa)},$$

where $\#\mathrm{dofs}(\kappa)$ and $\#\mathrm{dofs}(\mathcal{T}_{h,i})$, $i = 1, 2, 3$, denote the number of degrees of freedom associated with κ and $\mathcal{T}_{h,i}$, $i = 1, 2, 3$, respectively.

`Algorithm 2:` This is very similar to Algorithm 1; however, here we only construct the mesh patches $\mathcal{T}_{h,i}$, $i = 1, 2$, and compute the approximate local primal and dual solutions on these meshes only. Given an anisotropy parameter $\theta \geq 1$, isotropic refinement is selected when

$$\frac{\max_{i=1,2} |\hat{\mathcal{E}}_{\kappa,i}(u_{\mathrm{DG},i}, \hat{z}_i - z_{h,p})|}{\min_{i=1,2} |\hat{\mathcal{E}}_{\kappa,i}(u_{\mathrm{DG},i}, \hat{z}_i - z_{h,p})|} < \theta;$$

otherwise an anisotropic refinement is performed based on which refinement gives rise to the smallest predicted error indicator, i.e., the subdivision for which $|\hat{\mathcal{E}}_{\kappa,i}(u_{\mathrm{DG},i}, \hat{z}_i - z_{h,p})|$, $i = 1, 2$, is minimal. Based on computational experience, we select $\theta = 2$–3.

For purposes of comparison with standard anisotropic refinement strategies employed within the literature, we also consider the use of a Hessian–based algorithm. More precisely, for each element in the mesh, we construct a metric for the primal and dual problems based on computing the positive part of the Hessian matrix of the computed numerical solutions u_{DG} and $\hat{z}$, respectively. Upon application of the metric intersection algorithm proposed in [7], elements marked for refinement are anisotropically/isotropically subdivided, as in Figure 6.1, according to the relative size of the eigenvalues of the newly constructed metric; see [10] for details.

7. Numerical experiments. In this section we present a number of experiments to numerically to demonstrate the performance of the anisotropic adaptive algorithms outlined in Section 6.

7.1. Example 1. In this first example we consider a linear singularly perturbed advection-diffusion problem on the (unit) square domain $\Omega = (0, 1)^2$, where $a = \varepsilon I$, $0 < \varepsilon \ll 1$, $\mathbf{b} = (1, 1)^\top$, $c = 0$, and f is chosen so that

$$u(x, y) = x + y(1 - x) + [\mathrm{e}^{-1/\varepsilon} - \mathrm{e}^{-(1-x)(1-y)/\varepsilon}]\,[1 - \mathrm{e}^{-1/\varepsilon}]^{-1}, \tag{7.1}$$

cf. [17]. For $0 < \varepsilon \ll 1$, solution (7.1) has boundary layers along $x = 1$ and $y = 1$; throughout this section we set $\varepsilon = 10^{-2}$.

Here, we suppose that the aim of the computation is to calculate the (weighted) mean value of u over Ω, i.e., $J(u) = \int_\Omega u\psi \,\mathrm{d}x$, where $\psi = 100(1 - \tanh(100(r_1 - 0.01)(r_1 + 0.01)))(1 - \tanh(100(r_2 - 0.2)(r_2 + 0.2)))$, $r_1 = x - 1.0$ and $r_2 = y - 0.5$; thereby, $J(u) = 4.409917162888037$.

To demonstrate the versatility of the proposed refinement algorithms, in this section we employ bi-linear, bi-quadratic, and bi-cubic elements, i.e., $p = 1$, $p = 2$, and $p = 3$, respectively. To this end, in Figure 7.1 we plot the error in the computed target functional $J(\cdot)$ using both an isotropic (only) mesh refinement algorithm, together with the three anisotropic refinement strategies outlined in Section 6. Firstly, for each polynomial degree employed, we clearly observe the superiority of employing the anisotropic mesh refinement Algorithms 1 & 2 in comparison with standard isotropic subdivision of the elements. Indeed, the error $|J(u) - J(u_{\mathrm{DG}})|$ computed on the

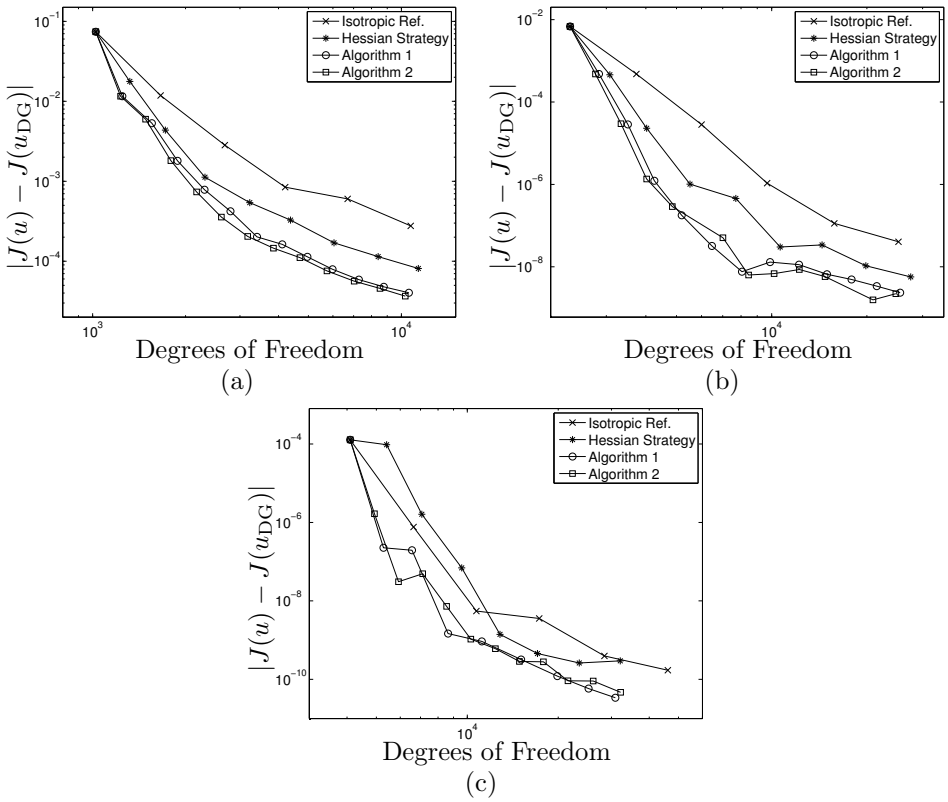

FIG. 7.1. *Example 1: Comparison between adaptive isotropic and anisotropic mesh refinement. (a) $p = 1$; (b) $p = 2$; (c) $p = 3$.*

series of anisotropically refined meshes designed using the two proposed algorithms outlined in Section 6 is always less than the corresponding quantity computed on the isotropic grids. Here, we observe that there is an initial transient whereby the error in the computed target functional decays rapidly using the former refinement algorithms, in comparison with the latter, after which the gradient of the convergence curves become very similar. This type of behavior is indeed expected, since for a fixed order method, i.e. h–version, we can only expect to improve the convergence of the error by a fixed constant, as the mesh is refined. Notwithstanding this, we note that, for each polynomial degree employed, the true error between $J(u)$ and $J(u_{\mathrm{DG}})$ using anisotropic refinement is around an order of magnitude smaller than the corresponding quantity when isotropic refinement is employed alone. Secondly, we observe that for all polynomial degrees employed, the Hessian strategy is inferior to Algorithms 1 & 2, in the sense that the error in the target functional computed using the either of the two latter strategies is always smaller that the corresponding quantity computed using the former strategy, for a fixed number of degrees of freedom. Indeed, even for bi-linear elements, for which the Hessian strategy has been proposed on the basis of interpolation theory, Algorithms 1 & 2 lead to a 35% reduction in the error on the final mesh in comparison with the corresponding quantity computed using the former strategy. Similar behavior is also observed for bi-quadratic and bi-cubic elements, though in the latter case, the Hessian strategy actually generates meshes

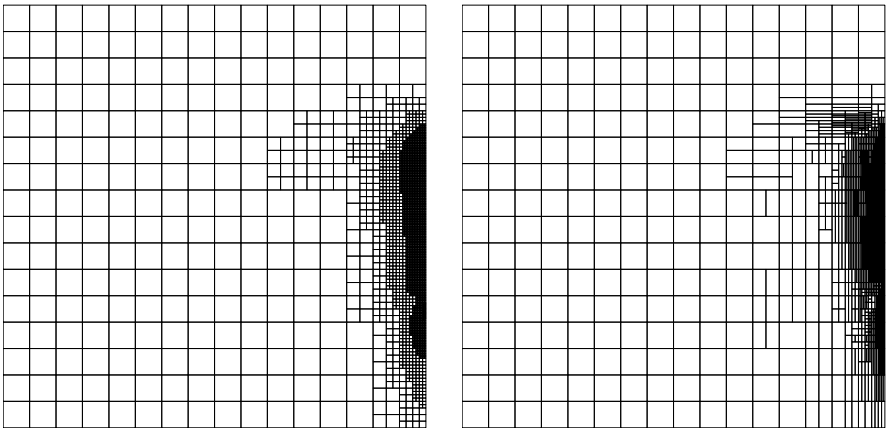

FIG. 7.2. *Example 1: Adaptively refined meshes for $p = 1$. (a) Isotropic mesh after 5 adaptive refinements, with 2680 elements; (b) Anisotropic mesh designed using Algorithm 2 after 7 adaptive refinements, with 963 elements*

which in many cases are inferior to their isotropic counterparts. Finally, we note that, despite the additional work involved in the implementation of Algorithm 1 in comparison to Algorithm 2, we see that both approaches lead to quantitatively very similar reductions in the error in the computed target functional.

In Figure 7.2 we show the meshes generated using both isotropic and anisotropic mesh adaptation. For brevity, we only show the meshes for $p = 1$, and in the latter case employing Algorithm 2. Firstly, we note that in both cases the mesh is primarily concentrated in the vicinity of the boundary layer along $x = 1$, where the support of the weighting function ψ appearing in the definition of the target functional $J(\cdot)$ is non-zero. Indeed, the region of the computational domain where the remainder of the boundary layer along $x = 1$ and moreover where the boundary layer along $y = 1$ are located are not refined, since the resolution of these sharp features present in the analytical solution are not important for the accurate computation of the selected target functional, cf. [14], for example. For Algorithm 2, we observe that the refinement strategy has clearly identified the anisotropy in the underlying primal and dual solutions, and refined the mesh accordingly. Indeed, we observe that the boundary layer along $x = 1$, $0 \leq y \leq 1$, has been significantly refined, as we would expect, with the elements being mostly refined in the direction parallel to the boundary. We note, however, that some anisotropic refinement perpendicular to Γ is performed in the region of the boundary layer in order to accurately capture the anisotropy of the dual solution z.

7.2. Example 2. In this second example we investigate the performance of the proposed anisotropic refinement algorithms applied to a mixed hyperbolic–elliptic problem with discontinuous boundary data. To this end, we let $\Omega = (0, 2) \times (0, 1)$, $a = \varepsilon(x)I$, where $\varepsilon = (1-\tanh(100(r_1-0.12)(r_1+0.12)))(1-\tanh(100(r_2-0.12)(r_2+0.12)))/1000$, $r_1 = x - 1.3$ and $r_2 = y - 0.3$. Furthermore, we set

$$\mathbf{b} = \begin{cases} (y, 1-x)^\top & \text{if } x < 1, \\ (1, 1/10)^\top & \text{if } x \geq 1, \end{cases}$$

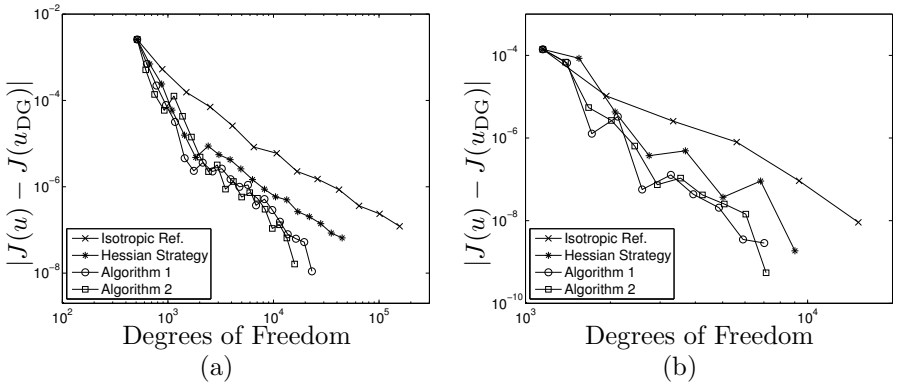

FIG. 7.3. *Example 2: Comparison between adaptive isotropic and anisotropic mesh refinement. (a) $p = 1$; (b) $p = 2$.*

$c = 0$, and $f = 0$. On the inflow boundary Γ_-, we select $u(x, y) = 1$ along $y = 0$, $1/8 < x < 3/4$ and $u(x, y) = 0$, elsewhere. This is a variant of the test problem presented in [15]. We note that the diffusion parameter ε will be approximately equal to 3.6×10^{-3} in the square region $(1.18, 1.42) \times (0.18, 0.42)$, where the underlying partial differential equation is uniformly elliptic. As (x, y) moves outside of this region, ε rapidly decreases through a layer of width $\mathcal{O}(0.1)$; for example, when $x = 1.3$ and $y > 0.7$ we have $\varepsilon < 10^{-15}$, so from the computational point of view ε is zero to within rounding error; in this region, the partial differential equation undergoes a change of type becoming, in effect, hyperbolic. Thus, we shall refer to the part of Ω containing this square region (including a strip of size $\mathcal{O}(0.1)$) as the *elliptic region*, while the remainder of the computational domain will be referred to as the *hyperbolic region*. [Strictly speaking, the partial differential equation is elliptic in the whole of $\bar{\Omega}$.]

Here, we suppose that the aim of the computation is to calculate the value of the (weighted) outflow advective flux along $x = 2$, $0 \leq y \leq 1$, i.e., $J(u) = \int_0^1 (\mathbf{b} \cdot \mathbf{n}) u(2, y) \psi(y) \, \mathrm{d}y$, where the weight function $\psi(y) = \mathrm{e}^{(3/8)^{-2} - ((y-5/8)^2 - 3/8)^{-2}}$. The (approximate) true value of the functional is given by $J(u) = 0.200620167062140$.

In Figure 7.3 we plot the error in the computed target functional $J(\cdot)$ using both an isotropic (only) mesh refinement algorithm, together with the three anisotropic refinement strategies outlined in Section 6 for $p = 1$ and $p = 2$. As in the previous section, we clearly observe the superiority of employing anisotropic mesh refinement in comparison with standard isotropic subdivision of the elements. Indeed, the error $|J(u) - J(u_{\mathrm{DG}})|$ computed on the series of anisotropically refined meshes is (almost) always less than the corresponding quantity computed on isotropic grids. Moreover, we again observe that (apart from an initial transient for $p = 1$), both Algorithms 1 & 2 give rise to an improvement in the error in the computed target functional, for a given number of degrees of freedom, when compared to the Hessian strategy; indeed, on the final mesh, Algorithm 2 leads an improvement in $|J(u) - J(u_{\mathrm{DG}})|$ of around one and two orders of magnitude for $p = 1$ and $p = 2$, respectively. In this example, we again observe that Algorithms 1 & 2 perform very similarly, in the sense that they both lead to approximately the same error in $J(\cdot)$ for a fixed number of degrees of freedom, though Algorithm 2 is still preferred since it is computationally less expensive.

Finally, in Figure 7.4 we show the meshes generated using both isotropic and

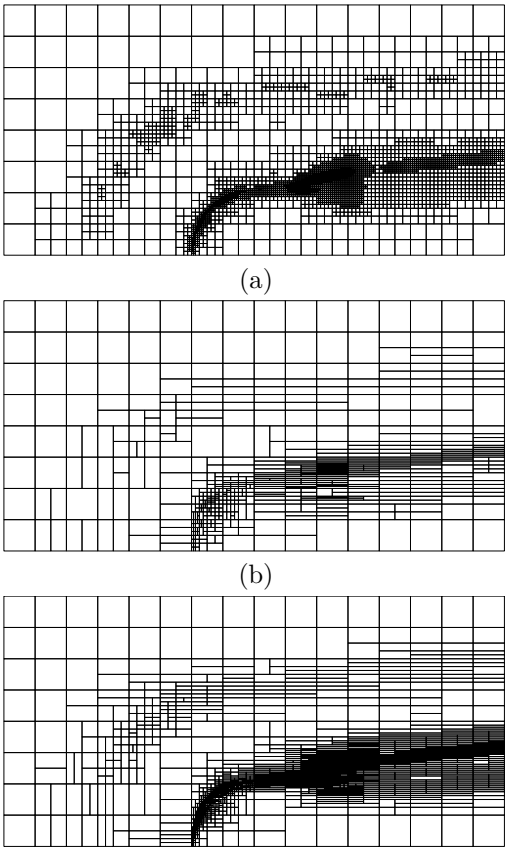

FIG. 7.4. *Example 1: Adaptively refined meshes for $p = 1$. (a) Isotropic mesh after 8 adaptive refinements, with 6539 elements; (b) & (c) Anisotropic meshes designed using Algorithm 2 after: 8 adaptive refinements, with 606 elements, and 14 adaptive refinements, with 1762 elements, respectively.*

anisotropic mesh adaptation (based on Algorithm 2), for bi-linear elements. Firstly, we note that in both cases the grid is primarily concentrated in the vicinity of the discontinuity of the analytical solution u which emanates from the point $(x, y) = (3/4, 0)$ on the inflow boundary; the second discontinuity in u is significantly less refined, as the resolution of this sharp feature in the solution is not essential for the computation of $J(\cdot)$. Additional mesh refinement has also been performed within the elliptic region, as well as the portion of the computational domain downstream of this region, though here we still observe a general concentration of elements within the 'smoothed' discontinuity of the analytical solution. Secondly, we observe that the anisotropic refinement algorithm has clearly identified the anisotropy in the underlying primal and dual solutions, and refined the mesh accordingly. Indeed, here we observe that in regions where the discontinuities/layers in u are well aligned with the mesh lines of the original background mesh, anisotropic refinement has been employed; in other regions of the computational domain, isotropic refinement has been utilized.

8. Concluding remarks. This article has been concerned with the *a priori* and *a posteriori* error analyses of the (symmetric) interior penalty discontinuous Galerkin

finite element discretization of second–order partial differential equations with nonnegative characteristic form, based on employing anisotropically refined computational meshes. Of particular interest has been the approximation of linear output functionals of the analytical solution. To this end, new, sharp directionally-sensitive bounds have been derived for the polynomial approximation on anisotropic elements exploiting the ideas presented in [10], and subsequently generalizing the results of that paper. These new anisotropic polynomial approximation results have been exploited in the proceeding *a priori* analysis of the numerical error for general linear target functionals of the solution on anisotropic meshes. Moreover, Type I (weighted) *a posteriori* error bounds have been derived and implemented within two adaptive mesh refinement algorithms, both employing a combination of local isotropic and anisotropic mesh refinement, where the choice of refinement is based on the solution of (cheap and fully parallelizable) local problems. The performance of the resulting refinement strategies were then studied through a series of numerical experiments, demonstrating the superiority of the proposed algorithms in comparison with both standard isotropic mesh refinement, and an anisotropic Hessian–based refinement strategy.

REFERENCES

[1] T. Apel. *Anisotropic finite elements: Local estimates and applications*. Advances in Numerical Mathematics, Teubner, Stuttgart, 1999.

[2] D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39:1749–1779, 2001.

[3] G.A. Baker, W.N. Jureidini, and O.A. Karakashian. Piecewise solenoidal vector fields and the Stokes problem. *SIAM J. Numer. Anal.*, 27(6):1466–1485, 1990.

[4] C. Baumann. *An hp–adaptive discontinuous Galerkin FEM for computational fluid dynamics*. PhD thesis, TICAM, UT Austin, Texas, 1997.

[5] R. Becker and R. Rannacher. Weighted a posteriori error control in FE methods. Technical report. Preprint 1, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Universität Heidelberg, Heidelberg, Germany, 1996.

[6] W. Cao. On the error of linear interpolation and the orientation, aspect ratio, and internal angles of a triangle. *SIAM J. Numer. Anal.*, 43(1):19–40, 2005.

[7] M.J. Castro-Díaz, F. Hecht, B. Mohammadi, and O. Pironneau. Anisotropic unstructured mesh adaption for flow simulations. *Internat. J. Numer. Methods Fluids*, 25:475–491, 1997.

[8] P.G. Ciarlet. *The finite element method for elliptic problems*. North-Holland, Amsterdam, 1978.

[9] Ch. Schwab E. Süli and P. Houston. *hp*–DGFEM for partial differential equations with nonnegative characteristic form. In B. Cockburn, G.E. Karniadakis, and C.-W. Shu, editors, *Discontinuous Galerkin Methods: Theory, Computation and Applications, Lecture Notes in Computational Science and Engineering, Vol. 11*, pages 221–230. Springer, 2000.

[10] L. Formaggia and S. Perotto. New anisotropic a priori error estimates. *Numer. Math.*, 89:641–667, 2001.

[11] E.H. Georgoulis. *hp*–version interior penalty discontinuous Galerkin finite element methods on anisotropic meshes. *Int. J. Numer. Anal. Model.*, 3:52–79, 2006.

[12] E.H. Georgoulis and A. Lasis. A note on the design of *hp*–version interior penalty discontinuous Galerkin finite element methods for degenerate problems. *IMA J. Numer. Anal.*, 26(2):381–390, 2006.

[13] K. Harriman, P. Houston, B. Senior, and E. Süli. *hp*–Version discontinuous Galerkin methods with interior penalty for partial differential equations with nonnegative characteristic form. In C.-W. Shu, T. Tang, and S.-Y. Cheng, editors, *Recent Advances in Scientific Computing and Partial Differential Equations. Contemporary Mathematics Vol. 330*, pages 89–119. AMS, 2003.

[14] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 24:979–1004, 2002.

[15] P. Houston, E.H. Georgoulis, and E. Hall. Adaptivity and a posteriori error estimation for DG methods on anisotropic meshes. In G. Lube and G. Rapin, editors, *Proceedings of the International Conference on Boundary and Interior Layers (BAIL) - Computational and*

Asymptotic Methods. 2006.

[16] P. Houston, C. Schwab, and E. Süli. Stabilized *hp*–finite element methods for first–order hyperbolic problems. *SIAM J. Numer. Anal.*, 37:1618–1643, 2000.

[17] P. Houston, C. Schwab, and E. Süli. Discontinuous *hp*-finite element methods for advection–diffusion–reaction problems. *SIAM J. Numer. Anal.*, 39:2133–2163, 2002.

[18] P. Houston and E. Süli. *hp*–Adaptive discontinuous Galerkin finite element methods for hyperbolic problems. *SIAM J. Sci. Comput.*, 23:1225–1251, 2001.

[19] P. Houston and E. Süli. Stabilized *hp*-finite element approximation of partial differential equations with non-negative characteristic form. *Computing*, 66:99–119, 2001.

[20] P. Houston and E. Süli. Adaptive finite element approximation of hyperbolic problems. In T. Barth and H. Deconinck, editors, *Error Estimation and Adaptive Discretization Methods in Computational Fluid Dynamics. Lect. Notes Comput. Sci. Engrg.*, volume 25, pages 269–344. Springer, 2002.

[21] W. Huang. Mathematical principles of anisotropic mesh adaptation. *Commun. Comput. Phys.*, 1(2):276–310, 2006.

[22] G. Kunert. *A posteriori error estimation for anisotropic tetrahedral and triangular finite element meshes*. PhD thesis, TU Chemnitz, 1999.

[23] M.G. Larson and T.J. Barth. A posteriori error estimation for discontinuous Galerkin approximations of hyperbolic systems. In B. Cockburn, G.E. Karniadakis, and C.-W. Shu, editors, *Discontinuous Galerkin Methods: Theory, Computation and Applications, Lecture Notes in Computational Science and Engineering, Vol. 11*. Springer, 2000.

[24] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21:1253–1278, 2000.

[25] J.T. Oden, I. Babuška, and C.E. Baumann. A discontinuous *hp*-finite element method for diffusion problems. *J. Comput. Phys.*, 146:491–519, 1998.

[26] O.A. Oleinik and E.V. Radkevič. *Second Order Equations with Nonnegative Characteristic Form*. American Mathematical Society, Providence, R.I., 1973.

[27] S. Prudhomme, F. Pascal, J.T. Oden, and A. Romkes. Review of a priori error estimation for discontinuous Galerkin methods. Technical report, TICAM Report 00–27, Texas Institute for Computational and Applied Mathematics, 2000.

[28] W. Rachowicz, L. Demkowicz, and J.T. Oden. Toward a universal *h–p* adaptive finite element strategy, Part 3. Design of *h–p* meshes. *Comput. Methods Appl. Mech. Engrg.*, 77:181–212, 1989.

[29] R. Schneider and P. Jimack. Toward anisotropic mesh adaptation based upon sensitivity of a posteriori estimates. Technical Report 2005.03, School of Computing, University of Leeds, 2005.