Report no. 96/09

Finite element methods for hyperbolic problems: a posteriori error analysis and adaptivity [a]

Endre Süli Paul Houston

This paper presents an overview of recent developments in the area of *a posteriori* error estimation for first-order hyperbolic partial differential equations. Global *a posteriori* error bounds are derived in the H^{-1} norm for steady and unsteady finite element and finite volume approximations of hyperbolic systems and scalar hyperbolic equations. We also consider the problem of *a posteriori* error estimation for linear functionals of the solution. The *a posteriori* error bounds are implemented into an adaptive finite element algorithm.

Subject classifications: AMS(MOS): 65M15, 65M50, 65M60

Key words and phrases: a posteriori error analysis, adaptivity, hyperbolic problems

The first author would like to acknowledge the financial support of the EPSRC.

The work reported here forms part of the research programme of the Oxford–Reading Institute for Computational Fluid Dynamics.

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

May, 1996

[a] Paper presented as Invited Lecture at the State of the Art in Numerical Analysis Conference, in York, 1-4 April 1996.

Contents

1 Introduction **3**

2 Johnson's paradigm **4**

3 A posteriori error analysis for steady hyperbolic PDEs **7**
3.1 Symmetric hyperbolic systems 7
3.2 A posteriori bound on the global error 11
3.3 Application to the cell vertex finite volume method 12

4 A posteriori error estimation for linear functionals **14**

5 Unsteady problems **19**
5.1 A posteriori error analysis 21
5.1.1 Error representation 21
5.1.2 Interpolation/projection estimates for the dual problem . . 22
5.1.3 Strong stability of the dual problem 23
5.1.4 Completion of the proof of the a posteriori error estimate . 24
5.2 Computational implementation 25
5.2.1 Adaptive algorithm 26
5.2.2 Mesh modification strategy 28
5.3 Numerical experiments 28

6 Conclusions **29**

References **31**

1 Introduction

Partial differential equations of hyperbolic type are of fundamental importance in many areas of applied mathematics, particularly fluid dynamics and electromagnetics. Solutions to these equations exhibit localised phenomena, such as propagating discontinuities and sharp transition layers, and their reliable numerical approximation presents a challenging computational task; indeed, in order to resolve such localised features in an accurate and efficient manner it is essential to use adaptively refined computational meshes whose construction is governed by sharp *a posteriori* error bounds. Over the last decade the *a posteriori* error analysis of finite element methods for partial differential equations of hyperbolic and nearly-hyperbolic character has been a subject of intensive research. For an overview of current activity in this area we refer to the article of Johnson [11]; see also, [10], [12], [13], and references therein. The approach proposed in those papers, and reviewed in the next section, rests on performing an elliptic regularisation of the hyperbolic problem and exploiting the smoothing properties of the resulting adjoint problem in conjunction with Galerkin orthogonality.

Subsequent sections of the paper are devoted to discussing an alternative, more direct, approach to the *a posteriori* error analysis of finite element approximations of hyperbolic problems which avoids the need for regularisation. Because of the inherent lack of strong smoothing properties in hyperbolic problems, particular care has then to be taken about existence of traces and integration-by-parts formulae which the *a posteriori* error analysis relies on; the necessary functional-analytic prerequisites that concern these issues are reviewed in the third section. By exploiting this theory, we derive *a posteriori* bounds on the global error in the H^{-1} norm for a general class of finite element methods (including certain finite volume schemes, interpreted as Petrov-Galerkin methods) for symmetric hyperbolic systems. For related work, see [15], [16], [17], [21], [25].

Controlling the computational error in a given *norm*, however, is not always the ultimate goal in practice. Indeed, in many engineering problems the accurate approximation of particular linear functionals of the solution (such as the lift, the drag, and the flux across the boundary of the domain) is at least as important as the reliable approximation of the solution in a given norm over the entire computational domain. We show, by considering the example of estimating the normal flux through the boundary, how problems of this kind may be approximated in an efficient way by exploiting a hyperbolic duality argument in conjunction with theoretical results concerning the propagation of singularities.

In the final part of the paper we discuss unsteady hyperbolic problems. In particular, we consider the *a posteriori* error analysis of a class of evolution-Galerkin methods for multi-dimensional scalar hyperbolic equations on a spatial domain Ω and finite time interval $[0, T]$. Here, we derive an *a posteriori* bound on the global error in the norm of the space $L_\infty(0, T; H^{-1}(\Omega))$. We conclude by indicating the practical relevance of this work and by illustrating the implementation

of the *a posteriori* analysis into an adaptive evolution-Galerkin algorithm.

2 Johnson's paradigm

In this brief section we review, in the context of first-order hyperbolic systems, the general theoretical framework of *a posteriori* error estimation pursued by Johnson and his co-workers; for a comprehensive account, see [11] and [13].

Suppose that $\mathcal{M}$ is a Hilbert space with inner product $(\cdot,\cdot)$ and let $A : \mathcal{M} \to \mathcal{M}$ be a linear operator on $\mathcal{M}$ with domain $D(A) \subset \mathcal{M}$ and range $R(A) = \mathcal{M}$ (in our case a system of first-order hyperbolic differential operators). Given that $f \in \mathcal{M}$, consider the problem of finding $u \in D(A)$ such that

$$Au = f.$$

In order to construct a Galerkin approximation to this problem, we consider a sequence of finite-dimensional spaces $\{\mathcal{U}^h\}$, parameterised by the positive discretisation parameter h; for the sake of simplicity we shall suppose that $\mathcal{U}^h \subset D(A)$ for each h. Simultaneously, consider a sequence of finite-dimensional spaces $\{\mathcal{M}^h\}$, with $\mathcal{M}^h$ contained in $\mathcal{M}$ for each h. For the purposes of this paper, $\mathcal{U}^h$ and $\mathcal{M}^h$ can be thought of as standard finite element spaces consisting of piecewise polynomial functions on a partition, of granularity h, of the computational domain. Let Π_h denote the orthogonal projector in $\mathcal{M}$ onto $\mathcal{M}^h$. The Galerkin approximation u_h of u is then sought in $\mathcal{U}^h$ as the solution of the finite-dimensional problem

$$\Pi_h A u_h = \Pi_h f.$$

In order to obtain a computable bound on the *global error* $e_h = u - u_h$ in terms of the *residual* r_h, defined by

$$r_h = f - Au_h,$$

we begin by observing the *Galerkin orthogonality* property

$$(r_h, v_h) = 0 \qquad \forall v_h \in \mathcal{M}^h.$$

This will be a key ingredient in the analysis. In addition, denoting by A^* the adjoint of A, we consider the following auxiliary problem, referred to as the dual problem:

$$A^*\phi = u - u_h.$$

The *a posteriori* error analysis is based on a duality argument, and it proceeds as follows:

$$\begin{aligned} \|u - u_h\|^2 &= (u - u_h, u - u_h) = (u - u_h, A^*\phi) \\ &= (A(u - u_h), \phi) = (Au - Au_h, \phi) \\ &= (f - Au_h, \phi) = (r_h, \phi). \end{aligned}$$

By virtue of the Galerkin orthogonality property $(r_h, \phi_h) = 0$ for any $\phi_h \in \mathcal{M}^h$, and therefore

$$\|u - u_h\|^2 = (r_h, \phi - \phi_h) = (h^s r_h, h^{-s}(\phi - \phi_h)),$$

where s is a non-negative real number, to be chosen below. By virtue of the Cauchy-Schwarz inequality,

$$\|u - u_h\|^2 \leq \|h^s r_h\| \, \|h^{-s}(\phi - \phi_h))\|.$$

While the first term on the right-hand side is of the desired form, involving the (computable) residual r_h multiplied by an appropriate power of the discretisation parameter, the second term incorporates ϕ, the solution to the dual problem. Since the dual-problem has the global error as data, ϕ is unknown and has to be eliminated from the analysis by relating it to $u - u_h$; furthermore a suitable choice of ϕ_h has to be made. This is achieved as follows: suppose that $\{\mathcal{W}_\sigma\}_{\sigma \geq 0}$ is a scale of Hilbert spaces, with associated norms $|||\cdot|||_\sigma$, such that $\mathcal{W}_0 = \mathcal{M}$ and $\mathcal{W}_{\sigma_2}$ is continuously embedded into $\mathcal{W}_{\sigma_1}$ whenever $\sigma_2 \geq \sigma_1$. Furthermore, we assume that there exist $\phi_h \in \mathcal{M}^h$ and a positive constant $C_\ddagger$ such that, for $\phi \in \mathcal{W}_s$,

$$\|h^{-s}(\phi - \phi_h)\| \leq C_\ddagger |||\phi|||_s.$$

For finite element methods, this hypothesis is easily fulfilled by choosing $\mathcal{W}_s$ as a hilbertian Sobolev space of index s and referring to standard approximation properties of piecewise polynomial functions in Sobolev spaces, with ϕ_h taken as the projection, the interpolant or the quasi-interpolant of ϕ from $\mathcal{M}^h$. Thereby we arrive at the bound

$$\|u - u_h\|^2 \leq C_\ddagger \|h^s r_h\| \, |||\phi|||_s.$$

This brings us to the final, and most important, step in the *a posteriori* error analysis. The norm $|||\phi|||_s$ appearing on the right-hand side of the last inequality has to be eliminated in terms of $u - u_h$ by recalling the relationship between ϕ and $u - u_h$, namely that $A^*\phi = u - u_h$. Thus to proceed, we assume that A^* is invertible and that $(A^*)^{-1}$ is a bounded linear operator from $\mathcal{M}$ to $\mathcal{W}_s$; hence

$$|||\phi|||_s = |||(A^*)^{-1}(u - u_h)|||_s \leq C_\star \|u - u_h\|,$$

where $C_\star$ is a positive constant, greater than or equal to the norm of $(A^*)^{-1}$. Finally, combining the last two bounds we arrive at the desired *a posteriori* bound on the global error $e_h = u - u_h$ in terms of the computable residual r_h:

$$\|u - u_h\| \leq C_\ddagger C_\star \|h^s r_h\|.$$

In order to turn this into an error bound that can be used in practice, the constants $C_\ddagger$ and $C_\star$ have to be determined. Estimating $C_\ddagger$ is a relatively simple matter using readily available results from approximation theory (see, for

example, Exercise 3.1.2 in Ciarlet's monograph [5] for an explicit formula for $C_\ddagger$ in the case of standard finite element spaces consisting of continuous piecewise polynomials on simplices, or the work of Handscomb [7] for very much sharper estimates of $C_\ddagger$ for piecewise linear finite elements on triangles). On the other hand, supplying $C_\star$ is much harder, involving an analytical study of the well-posedness of the dual problem. Since any value of $C_\star$ that is arrived at through such general analytical arguments is necessarily a considerable overestimate of the ratio $|||\phi|||_s/\|u-u_h\|$, in practice the stability constant $C_\star$ is determined computationally for the problem at hand, as part of the process of *a posteriori* error estimation.

Finally, we have to fix s, the exponent of h in the error bound. As one would like the *a posteriori* bound to reflect the approximation property of the space $\mathcal{M}^h$ to its full extent, one would wish to choose s as large as possible; unfortunately, for linear first-order hyperbolic systems consisting of m equations on a domain $\Omega \subset \mathbb{R}^d$, $(A^*)^{-1}$ is a bounded operator from $\mathcal{M} = [L_2(\Omega)]^m$ to $\mathcal{M} = \mathcal{W}_0 = [L_2(\Omega)]^m$ only, so s cannot exceed 0; thus we end up with the bound

$$\|u-u_h\| \leq C_\ddagger C_\star \|r_h\|.$$

In fact, we note that when $s = 0$ there is no benefit in exploiting Galerkin orthogonality, and we may simply take $\phi_h = 0$ in our argument to improve this bound to

$$\|u-u_h\| \leq C_\star \|r_h\|.$$

One way or the other, in the case of a first-order hyperbolic system, the *a posteriori* error bound that we arrive at on the basis of the reasoning outlined above is unsatisfactory in that it fails to display the approximation properties of the test space $\mathcal{M}^h$. Worse still, when the data are discontinuous, linear hyperbolic systems may possess solutions that are discontinuous across characteristic hypersurfaces and, under mesh refinement, $\|r_h\|$ will then converge to 0 very slowly, if at all; consequently, in the absence of the compensating factor h^s, any adaptive algorithm driven by this error bound is likely to be inefficient.

The problem may be rectified by perturbing the first-order hyperbolic operator A^* (or, indeed, both A and A^*) through the addition of a second-order elliptic term with small coefficient; this then provides additional regularity which, in favourable circumstances, allow one to take $s = 2$. This approach, however, is associated with undesirable complications related to the fact that artificial boundary conditions have to be supplied for the resulting second-order operator in such a way that the features of the solution to the hyperbolic system are retained in the vicinity of the boundary; a further difficulty with elliptic regularisation of non-dissipative hyperbolic systems, such as the Maxwell system of electro-magnetism, is that it may introduce a physically unacceptable level of damping into the model.

Thus, in the rest of the paper, we consider an alternative approach.

3 A posteriori error analysis for steady hyperbolic PDEs

In this section we develop the *a posteriori* error analysis of a general class of finite element methods for steady linear multi-dimensional hyperbolic systems. We begin with a review of basic results from the theory of symmetric hyperbolic systems in the sense of Friedrichs.

3.1 Symmetric hyperbolic systems

Throughout this section Ω will denote a bounded open set in $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$. Suppose that A_i, $i = 1, ..., d$, and C are (matrix-valued) mappings from $\overline{\Omega}$ into $\mathbb{R}^{m\times m}$, $m \geq 1$; we shall assume that, for each i, the entries of A_i are continuously differentiable on $\overline{\Omega}$ and the components of C are continuous on $\overline{\Omega}$. We consider the hyperbolic system of linear first-order partial differential equations

$$L\mathbf{u} \equiv \sum_{i=1}^{d} \frac{\partial}{\partial x_i}(A_i\mathbf{u}) + C\mathbf{u} = \mathbf{f} \qquad \text{in } \Omega. \tag{3.1}$$

The system (3.1) is said to be *symmetric positive* if the following conditions hold:

(a) the matrices A_i, for $i = 1, ..., d$, are symmetric, i.e. $A_i = A_i^*$;

(b) there exists $\alpha \geq 0$ and a unit vector $\xi \in \mathbb{R}^d$, such that the symmetric part of the matrix

$$K_\xi = C + \frac{1}{2}\sum_{i=1}^{d} \frac{\partial A_i}{\partial x_i} + \alpha \sum_{i=1}^{d} \xi_i A_i$$

is positive definite, uniformly on $\overline{\Omega}$, i.e. there exists a positive constant $c_0 = c_0(\Omega)$ such that

$$\frac{1}{2}(K_\xi(x) + K_\xi^*(x)) \geq c_0 I \tag{3.2}$$

for all x in $\overline{\Omega}$.

Since $\partial\Omega$ is Lipschitz continuous, the outer unit normal vector field $\hat{\nu} = (\hat{\nu}_1, ..., \hat{\nu}_d)$ is defined almost everywhere on $\partial\Omega$ with respect to the $(d-1)$-dimensional measure on $\partial\Omega$. Consider the symmetric matrix

$$B = \hat{\nu}_1 A_1 + ... + \hat{\nu}_d A_d.$$

We shall suppose that B is non-singular almost everywhere on $\partial\Omega$, that is, the boundary of Ω is non-characteristic for L. Since B is symmetric and of full rank it can be decomposed as $B = B^+ + B^-$, where B^+ is positive semi-definite and B^-

is negative semi-definite. Given that $\mathbf{g}$ is a sufficiently smooth function defined on $\partial\Omega$, an admissible boundary condition for (3.1) is given by

$$B^-\mathbf{u}\Big|_{\partial\Omega} = B^-\mathbf{g} \qquad \text{on } \partial\Omega; \tag{3.3}$$

we shall also consider the homogeneous counterpart of this boundary condition:

$$B^-\mathbf{u}\Big|_{\partial\Omega} = \mathbf{0} \qquad \text{on } \partial\Omega. \tag{3.4}$$

At this stage these boundary conditions are to be understood formally; below we shall state a trace theorem which assigns a precise meaning to $B^-\mathbf{u}|_{\partial\Omega}$. First, however, we define the *graph space* of the operator L as the linear space

$$H(L,\Omega) = \{\mathbf{v} \in [L_2(\Omega)]^m \,:\, L\mathbf{u} \in [L_2(\Omega)]^m\}.$$

When equipped with the norm

$$|||\mathbf{v}|||_{\Omega,\xi} = \left(\|\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v}\|^2 + \|\mathrm{e}^{-\alpha(\xi\cdot x)}L\mathbf{v}\|^2\right)^{\frac{1}{2}}$$

the graph space $H(L,\Omega)$ is a Hilbert space. Here and throughout the paper $\|\cdot\|$ will stand for the L_2 norm over Ω induced by the $L_2(\Omega)$ inner product $(\cdot,\cdot)$. Consider L^*, the formal adjoint of L, defined by

$$L^*\mathbf{v} = -\sum_{i=1}^{d} A_i\frac{\partial\mathbf{v}}{\partial x_i} + C^*\mathbf{v};$$

the associated graph space $H(L^*,\Omega)$ and graph norm $|||\cdot|||_{*,\Omega,\xi}$ are defined analogously as for L, but with the weight-function $\mathrm{e}^{-\alpha(\xi\cdot x)}$ replaced by $\mathrm{e}^{\alpha(\xi\cdot x)}$.

Let $\Gamma_{0,\partial\Omega} : [H^1(\Omega)]^m \to [H^{1/2}(\partial\Omega)]^m$ signify the usual trace operator (which to each element of $[H^1(\Omega)]^m$ assigns its restriction to $\partial\Omega$). Let $[H^{-1/2}(\partial\Omega)]^m$ denote the dual space of $[H^{1/2}(\partial\Omega)]^m$. The following trace theorem will play a key rôle in our error analysis (see [17]).

Theorem 3.1 *Assuming that $\partial\Omega$ is non-characteristic for the operator L, the mapping $\Gamma_{B,\partial\Omega} : \mathbf{v} \mapsto B(\Gamma_{0,\partial\Omega}\mathbf{v})$ defined on $[H^1(\Omega)]^m$ can be extended by continuity to a linear and continuous mapping, still denoted $\Gamma_{B,\partial\Omega}$, from $H(L,\Omega)$ into $[H^{-1/2}(\partial\Omega)]^m$. Moreover, for any $\mathbf{u} \in H(L,\Omega)$ and $\mathbf{v} \in [H^1(\Omega)]^m$ the following Green's formula holds*

$$(L\mathbf{u},\mathbf{v}) - (\mathbf{u},L^*\mathbf{v}) = \langle\Gamma_{B,\partial\Omega}\mathbf{u},\Gamma_{0,\partial\Omega}\mathbf{v}\rangle.$$

An analogous result is valid for $H(L^*,\Omega)$.

Here and throughout the rest of this paper, $\langle\cdot,\cdot\rangle$ denotes the duality pairing between the function spaces $[H^{-1/2}(\partial\Omega)]^m$ and $[H^{1/2}(\partial\Omega)]^m$.

The splitting $B = B^+ + B^-$ induces a natural decomposition of the trace operator $\Gamma_{B,\partial\Omega}$ which leads to the definition of the partial trace operators

$$\Gamma_{B^\pm,\partial\Omega} \,:\, H(L,\Omega) \to [H^{-1/2}(\partial\Omega)]^m$$

with

$$\Gamma_{B,\partial\Omega} = \Gamma_{B^+,\partial\Omega} + \Gamma_{B^-,\partial\Omega}.$$

Indeed, letting

$$\Gamma_{B^\pm,\Omega}\mathbf{u} = B^\pm(\Gamma_{0,\partial\Omega}\mathbf{u}) \qquad \forall \mathbf{u} \in [H^1(\Omega)]^m,$$

on the dense subspace $[H^1(\Omega)]^m$ of $H(L,\Omega)$ this decomposition can be constructed by splitting the trace of $\mathbf{u} \in [H^1(\Omega)]^m$; the linear operators $\Gamma_{B^\pm,\partial\Omega} : H(L,\Omega) \to [H^{-1/2}(\partial\Omega)]^m$ are then defined by continuous extension from the dense subspace $[H^1(\Omega)]^m$ to all of $H(L,\Omega)$. One can proceed in the same way for $H(L^*,\Omega)$. Thus, in the case of the homogeneous boundary value problem (3.1), (3.4), we can define precisely the domains of the operators L and L^*:

$$D(L,\Omega) = \{\mathbf{u} \in H(L,\Omega) \,:\, \Gamma_{B^-,\partial\Omega}\mathbf{u} = \mathbf{0} \quad \text{on } \partial\Omega\},$$

$$D(L^*,\Omega) = \{\mathbf{u} \in H(L^*,\Omega) \,:\, \Gamma_{B^+,\partial\Omega}\mathbf{u} = \mathbf{0} \quad \text{on } \partial\Omega\};$$

when equipped with the associated graph-norms $|||\cdot|||_{\xi,\Omega}$, $|||\cdot|||_{*,\xi,\Omega}$, $D(L,\Omega)$ and $D(L^*,\Omega)$ are Hilbert subspaces of $H(L,\Omega)$ and $H(L^*,\Omega)$, respectively.

Given that $\mathbf{f} \in [L_2(\Omega)]^m$, a function $\mathbf{u} \in [L_2(\Omega)]^m$ satisfying

$$(\mathbf{u}, L^*\phi) = (\mathbf{f}, \phi) \qquad \forall \phi \in D(L^*,\Omega) \cap [H^1(\Omega)]^m$$

is called a *weak solution* of the homogeneous boundary value problem (3.1), (3.4). A weak solution belonging to $H(L,\Omega)$ is called a *strong solution*. We note that the requirement that $\mathbf{u}$ be a strong solution does not preclude the possibility of $\mathbf{u}$ being discontinuous; indeed, since only $L\mathbf{u} \in [L_2(\Omega)]^m$, discontinuities in $\mathbf{u}$ may occur across characteristic hypersurfaces.

Theorem 3.2 *Assuming that $\partial\Omega$ is a non-characteristic hypersurface for L, and given that $\mathbf{f} \in [L_2(\Omega)]^m$, the homogeneous boundary value problem (3.1), (3.4) has a unique strong solution $\mathbf{u} \in D(L,\Omega)$. In addition the linear operator L is a continuous bijection from $D(L,\Omega)$ onto $[L_2(\Omega)]^m$ with a continuous inverse $L^{-1} : [L_2(\Omega)]^m \to D(L,\Omega)$. An analogous result holds for L^*.*

Proof The proof is based on Banach's Closed Range Theorem. Given that $\mathbf{v} \in [H^1(\Omega)]^m$, upon taking the inner product of $L\mathbf{v}$ with $\mathrm{e}^{-2\alpha(\xi\cdot x)}\mathbf{v}$, integrating by parts using Theorem 3.1 and splitting the trace operator $\Gamma_{B,\partial\Omega}$, we have that

$$\begin{aligned}&\langle \mathrm{e}^{-\alpha(\xi\cdot x)}\Gamma_{B^+,\partial\Omega}\mathbf{v}, \mathrm{e}^{-\alpha(\xi\cdot x)}\Gamma_{0,\partial\Omega}\mathbf{v}\rangle + (\frac{1}{2}(K_\xi + K_\xi^*)\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v}, \mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v})\\ &\quad = -\langle \mathrm{e}^{-\alpha(\xi\cdot x)}\Gamma_{B^-,\partial\Omega}\mathbf{v}, \mathrm{e}^{-\alpha(\xi\cdot x)}\Gamma_{0,\partial\Omega}\mathbf{v}\rangle + (\mathrm{e}^{-\alpha(\xi\cdot x)}L\mathbf{v}, \mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v}).\end{aligned}$$

Thence, recalling (3.2) of hypothesis (b), we deduce the following Gårding inequality:

$$c_0 \| \mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v}\| \le \| \mathrm{e}^{-\alpha(\xi\cdot x)} L\mathbf{v}\| \qquad \forall \mathbf{v} \in [H^1(\Omega)]^m \cap D(L,\Omega). \tag{3.5}$$

Noting that $[H^1(\Omega)]^m$ is dense in $H(L,\Omega)$, it follows that $[H^1(\Omega)]^m \cap D(L,\Omega)$ is dense in $D(L,\Omega)$, so that

$$\|\mathrm{e}^{-\alpha(\xi\cdot x)} L\mathbf{v}\| \ge c_1 |||\mathbf{v}|||_{\xi,\Omega} \qquad \forall \mathbf{v} \in D(L,\Omega), \tag{3.6}$$

where $c_1 = (1 + c_0^{-2})^{-1/2}$. Now (3.6) implies that L is an injective operator from $D(L,\Omega)$ onto its range space $\mathcal{R}(L)$, and that the inverse of L is continuous. Hence L is an isomorphism from $D(L,\Omega)$ onto $\mathcal{R}(L)$; this implies that $\mathcal{R}(L)$ is a closed linear subspace of $[L_2(\Omega)]^m$. By virtue of hypothesis (b), the transpose ${}^tL$ of L has trivial kernel, i.e. $\mathrm{Ker}({}^tL) = \{0\}$. The Closed Range Theorem then implies that $\mathcal{R}(L) = \mathrm{Ker}({}^tL)^\perp = [L_2(\Omega)]^m$. Hence L is an isomorphism from $H(L,\Omega)$ onto $[L_2(\Omega)]^m$. $\square$

We deduce from this theorem that (weak and strong) solutions of the homogeneous boundary value problem (3.1), (3.4) satisfy the stability estimate

$$\|\mathbf{u}\| \le \frac{1}{c_0} \mathrm{e}^{2\alpha L} \|\mathbf{f}\|,$$

where $L = \mathrm{diam}(\Omega)$.

More generally, consider the non-homogeneous boundary value problem (3.1), (3.3), where $\mathbf{f} \in [L_2(\Omega)]^m$ and $\mathbf{g} \in [L_2(\partial\Omega)]^m$. A function $\mathbf{u} \in [L_2(\Omega)]^m$ obeying

$$(\mathbf{u}, L^*\phi) + \langle B^-\mathbf{g}, \phi\rangle = (\mathbf{f}, \phi) \qquad \forall \phi \in D(L^*,\Omega) \cap [H^1(\Omega)]^m$$

is called a *weak solution* of the boundary value problem. A weak solution $\mathbf{u}$ of the non-homogeneous problem (3.1), (3.3) which belongs to $H(L,\Omega)$ is called a *strong solution*. Lax and Phillips [14] proved, under the assumption that $\partial\Omega$ is non-characteristic for L, that every weak solution is a strong solution. Furthermore, assuming that $\mathbf{f} \in [H^1(\Omega)]^m$, $\mathbf{g} \in [H^1(\partial\Omega)]^m$ and the entries of C are continuously differentiable on $\overline{\Omega}$, the following regularity result holds:

$$\|\mathbf{u}\|_{H^1(\Omega)} \le c_1 \left(\|\mathbf{f}\|_{H^1(\Omega)} + \|\mathbf{g}\|_{H^1(\partial\Omega)} \right)$$

(see Lax and Phillips [14]). An analogous bound holds for the problem

$$\begin{aligned} L^*\phi &= \mu && \text{in } \Omega, \\ B^+\phi\big|_{\partial\Omega} &= B^+\zeta && \text{on } \partial\Omega, \end{aligned}$$

with corresponding stability constant c_1'; namely,

$$\|\phi\|_{H^1(\Omega)} \le c_1'(\|\mu\|_{H^1(\Omega)} + \|\zeta\|_{H^1(\partial\Omega)}). \tag{3.7}$$

3.2 A posteriori bound on the global error

Suppose that $\mathcal{U}^h \subset H(L,\Omega)$ is a finite element trial space consisting of piecewise polynomial functions on a partition $(\kappa_i)_{i=1,\ldots,N_h}$ of $\overline{\Omega}$; here h is the mesh parameter. Further, let $\mathcal{M}^h \subset [L_2(\Omega)]^m$ be a finite element test space on this partition, and let Π_h denote the orthogonal projector in $[L_2(\Omega)]^m$ onto $\mathcal{M}^h$. We consider the following general class of Galerkin approximations of the boundary value problem (3.1), (3.3): find $\mathbf{u} \in \mathcal{U}^h$ such that

$$\Pi_h L\mathbf{u}_h = \Pi_h \mathbf{f} \quad \text{on } \Omega, \qquad B^-\mathbf{u}_h\big|_{\partial\Omega} = B^-\mathbf{g}, \tag{3.8}$$

where, for the sake of simplicity, we have assumed that the function $\mathbf{g}$ belongs to the restriction of the finite element trial space $\mathcal{U}^h$ to the boundary. We shall suppose that (3.8) has a unique solution. Particular examples of such discretisations, including finite element and finite volume methods, will be considered below.

We denote by $\mathbf{e}_h = \mathbf{u} - \mathbf{u}_h$ the *global error* and define the *residual* $\mathbf{r}_h = \mathbf{f} - L\mathbf{u}_h$. Since L is a linear operator, it follows that $\mathbf{e}_h$ is the unique solution of the boundary value problem

$$L\mathbf{e}_h = \mathbf{r}_h \quad \text{on } \Omega, \qquad B^-\mathbf{e}_h\big|_{\partial\Omega} = \mathbf{0}.$$

This is a key relationship between the (computable) residual $\mathbf{r}_h$ and the global error $\mathbf{e}_h$ which allows us to obtain computable bounds on $\mathbf{e}_h$ in terms of $\mathbf{r}_h$ by bounding the inverse of the operator L.

We shall further suppose that the finite element test space $\mathcal{M}^h$ possesses the following standard approximation property:

(c) There exists a positive constant c_2, independent of h, such that for each $\phi \in [H^1(\Omega)]^m$ there is $\phi_h \in \mathcal{M}^h$ with

$$\left(\sum_\kappa \|h^{-1}(\phi - \phi_h)\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}} \leq c_2 |\phi|_{H^1(\Omega)}.$$

Theorem 3.3 *Suppose that hypotheses (a), (b) and (c) hold, and that the entries of C are continuously differentiable on $\overline{\Omega}$. Then,*

$$\|\mathbf{u} - \mathbf{u}_h\|_{H^{-1}(\Omega)} \leq c_1' c_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}}.$$

Proof Given $\psi \in [C_0^\infty(\Omega)]^m$, consider the dual problem

$$L^*\phi = \psi \qquad \text{on } \Omega, \qquad\qquad B^+\phi|_{\partial\Omega} = \mathbf{0}.$$

Then,

$$(\mathbf{u} - \mathbf{u}_h, \psi) = (\mathbf{u} - \mathbf{u}_h, L^*\phi) = (L(\mathbf{u} - \mathbf{u}_h), \phi) = (r_h, \phi).$$

Recalling (3.8), it follows that

$$(r_h, \phi_h) = 0;$$

so, for each $\phi_h \in \mathcal{M}^h$, we have that

$$\begin{aligned}(\mathbf{u}-\mathbf{u}_h, \psi) &= (\mathbf{r}_h, \phi - \phi_h) \\ &\leq \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}} \left(\sum_\kappa \|h^{-1}(\phi-\phi_h)\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}} \\ &\leq c_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}} |\phi|_{H^1(\Omega)}. \qquad (3.9)\end{aligned}$$

According to the regularity result (3.7), with $\mu = \psi$,

$$|\phi|_{H^1(\Omega)} \leq \|\phi\|_{H^1(\Omega)} \leq c_1' \|\psi\|_{H^1(\Omega)}.$$

Substituting this into (3.9), dividing both sides by $\|\psi\|_{H^1(\Omega)}$, and taking the supremum over all $\psi \in [C_0^\infty(\Omega)]^m$, we obtain the desired error bound. □

3.3 Application to the cell vertex finite volume method

In this section we present an *a posteriori* error bound for a cell vertex finite volume approximation of (3.1), (3.3); this is a generalisation of a similar result derived, in the case of homogeneous boundary conditions, in [25].

For the sake of simplicity we suppose that $\Omega = (0,1)^2$, and consider the Friedrichs system in conservation form:

$$\begin{aligned}L\mathbf{u} \equiv \sum_{i=1}^{2} \frac{\partial}{\partial x_i}(A_i\mathbf{u}) + C\mathbf{u} &= \mathbf{f} \quad \text{in } \Omega, \\ B^-\mathbf{u}|_{\partial\Omega} &= B^-\mathbf{g} \quad \text{on } \partial\Omega.\end{aligned}$$

The cell vertex finite volume approximation of this boundary value problem is obtained by subdividing Ω using a structured mesh that consists of convex quadrilateral elements, and integrating the differential equation over each element; using Gauss' theorem, integrals over cells are converted into contour integrals which are then approximated by the trapezium rule. This construction gives rise to a conservative four-point finite difference scheme, called the cell vertex scheme. Equivalently, the cell vertex scheme can be formulated as a Petrov-Galerkin finite element method based on continuous piecewise bilinear trial functions and piecewise constant test functions (see [22], [23], [24], [2]). It is this latter interpretation that is adopted here.

Let $\mathcal{F} = \{\mathcal{T}_h\}$, $h > 0$, be a family of structured partitions $\mathcal{T}_h$ of $\Omega = (0,1)^2$ into convex quadrilateral elements. We shall assume that the partition is *quasi-parallel* in the sense that, for each element $\kappa_{ij} \in \mathcal{T}_h$, the distance between the midpoints of the diagonals is bounded by $c_\star \text{meas}(\kappa_{ij})$, where $c_\star$ is a fixed positive

constant. In order to introduce the relevant finite element spaces, we define the reference square $\hat{\kappa} = (-1,1)^2$, and denote by $F_{\kappa_{ij}}$ the bilinear function that maps $\hat{\kappa}$ onto the 'finite volume' κ_{ij}. Let $\mathcal{Q}_1(\hat{\kappa})$ be the set of bilinear functions on $\hat{\kappa}$, and $\mathcal{Q}_0(\hat{\kappa})$ the set of constant functions on $\hat{\kappa}$. We define

$$\begin{aligned} \mathcal{M}^h &= \left\{ \mathbf{q} \in [L_2(\Omega)]^m \,:\, \mathbf{q} = \hat{\mathbf{q}} \circ F^{-1}_{\kappa_{ij}},\ \hat{\mathbf{q}} \in [\mathcal{Q}_0(\hat{\kappa})]^m,\ \kappa_{ij} \in \mathcal{T}_h \right\}, \\ \mathcal{U}^h &= \left\{ \mathbf{v} \in [H^1(\Omega)]^m \,:\, \mathbf{v} = \hat{\mathbf{v}} \circ F^{-1}_{\kappa_{ij}},\ \hat{\mathbf{v}} \in [\mathcal{Q}_1(\hat{\kappa})]^m,\ \kappa_{ij} \in \mathcal{T}_h \right\}, \end{aligned}$$

as well as $\mathcal{U}^h_-$, the set of all functions $\mathbf{v}$ in $\mathcal{U}^h$ such that $B^-\mathbf{v}|_{\partial\Omega} = B^-\mathbf{g}$ on $\partial\Omega$; here we have assumed, for the sake of simplicity, that $\mathbf{g}$ belongs to the restriction of the finite element trial space $\mathcal{U}^h$ to the boundary of the domain. Let us denote by $\Pi_h : [L_2(\Omega)]^m \to \mathcal{M}^h$ the L_2-orthogonal projector onto $\mathcal{M}^h$, and by $\mathcal{I}_h : (H(L,\Omega) \cap [C^0(\bar{\Omega})]^m)^2 \to \mathcal{U}^h \times \mathcal{U}^h$ the interpolation projector onto $\mathcal{U}^h \times \mathcal{U}^h$.

The cell vertex finite volume approximation is defined as follows: find $\mathbf{u}_h \in \mathcal{U}^h_-$ satisfying

$$(\operatorname{div} \mathcal{I}_h(\mathcal{A}\mathbf{u}_h), \mathbf{q}_h) + (C\mathbf{u}_h, \mathbf{q}_h) = (\mathbf{f}, \mathbf{q}_h) \qquad \forall \mathbf{q}_h \in \mathcal{M}^h, \tag{3.10}$$

where $\mathcal{A} = (A_1, A_2)$. We define the *cell vertex operator* $L_h : \mathcal{U}^h \longrightarrow \mathcal{M}^h$ by

$$L_h \,:\, \mathbf{v}_h \longmapsto \Pi_h \left(\operatorname{div} \mathcal{I}_h(\mathcal{A}\mathbf{v}_h)\right) + \Pi_h(C\mathbf{v}_h).$$

With this definition, we can reformulate (3.10) as follows: find $\mathbf{u}_h \in \mathcal{U}^h_-$ satisfying

$$L_h \mathbf{u}_h = \Pi_h \mathbf{f}.$$

We shall adopt the following hypothesis:

(b') The matrices A_i, $i = 1, 2$, are positive-definite, uniformly on $\overline{\Omega}$.

Clearly (b') is sufficient for (b) when $d = 2$. Under this assumption $\mathcal{U}^h_-$ and $\mathcal{M}^h$ have the same dimension and L_h is then easily seen to be bijective. For theoretical results concerning the stability and convergence of the cell vertex scheme we refer to [2],[18], [19], [22], [23] and [24]. Here we derive an *a posteriori* error bound for the method.

Theorem 3.4 *Suppose that hypothesis (b') holds, that $\mathcal{F} = \{\mathcal{T}_h\}$ is a family of structured quasi-parallel partitions of Ω, and that, given s, $1 < s \leq 3$, the coefficients of the matrices A_i, $i = 1, 2$, belong to $H^s(\Omega) \cap C^1(\overline{\Omega})$ and those of C to $C^1(\overline{\Omega})$. Then, the cell vertex scheme obeys the following a posteriori error bound:*

$$\|\mathbf{u} - \mathbf{u}_h\|_{H^{-1}(\Omega)} \leq c_3 \left(\sum_{\kappa} \|h\mathbf{r}_h\|^2_{L_2(\kappa)} + \sum_{\kappa} |h^{s-1}\mathcal{A}\mathbf{u}_h|^2_{H^s(\kappa)} \right)^{\frac{1}{2}},$$

where c_3 is a computable constant.

Proof Suppose that $\psi \in [C_0^\infty(\Omega)]^m$ and consider the dual problem

$$L^*\phi = \psi \qquad \text{on } \Omega, \qquad\qquad B^+\phi|_{\partial\Omega} = \mathbf{0}.$$

Then,

$$\begin{aligned}(\mathbf{u}-\mathbf{u}_h,\psi) &= (\mathbf{u}-\mathbf{u}_h, L^*\phi) = (L(\mathbf{u}-\mathbf{u}_h),\phi)\\ &= (\mathbf{f}-L\mathbf{u}_h, \phi-\phi_h) + (\mathbf{f}-L\mathbf{u}_h,\phi_h)\\ &= (\mathbf{r}_h,\phi-\phi_h) + (\Pi_h(\operatorname{div}\mathcal{I}_h(\mathcal{A}\mathbf{u}_h) - \operatorname{div}(\mathcal{A}\mathbf{u}_h)),\phi_h).\end{aligned}$$

Thus, choosing $\phi_h = \Pi_h\phi$ and exploiting the approximation property (c), we have that

$$\begin{aligned}|(\mathbf{u}-\mathbf{u}_h,\psi)| \;\le\; &\left(\sum_\kappa \|\Pi_h(\operatorname{div}\mathcal{I}_h(\mathcal{A}\mathbf{u}_h) - \operatorname{div}(\mathcal{A}\mathbf{u}_h))\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}} \|\phi\|\\ &+c_2\left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}} |\phi|_{H^1(\Omega)}. \qquad (3.11)\end{aligned}$$

By virtue of a superconvergence result stated in [2] (for the special case of a uniform finite difference mesh; the extension to quasi-parallel meshes is straightforward),

$$\|\Pi_h(\operatorname{div}\mathcal{I}_h(\mathcal{A}\mathbf{u}_h) - \operatorname{div}(\mathcal{A}\mathbf{u}_h))\|_{L_2(\kappa)} \le c_4|h^{s-1}\mathcal{A}\mathbf{u}_h|_{H^s(\kappa)}, \qquad 1 < s \le 3.$$

Inserting this into (3.11), recalling the hyperbolic regularity estimate (3.7),

$$|\phi|_{H^1(\Omega)} \le \|\phi\|_{H^1(\Omega)} \le c_1'\|\psi\|_{H^1(\Omega)},$$

dividing both sides of the inequality by $\|\psi\|_{H^1(\Omega)}$ and taking the supremum over all ψ in $[C_0^\infty(\Omega)]^m$, we obtain the desired result, with $c_3 = c_1'\max(c_2, c_4)$. □

4 A posteriori error estimation for linear functionals

In engineering applications it is frequently the case that linear functionals of the solution (such as the lift, the drag, or the flux of the solution across the boundary) are of prime concern. In such instances it is unlikely that *a posteriori* error bounds of the kind stated in the last two theorems will be of use in the design of adaptive algorithms. Our aim in this section is to propose an approach to deriving *a posteriori* error bounds on linear functionals, without attempting to obtain an upper bound on the error in a *norm* in which the functional is bounded. In order to illustrate the key ideas we discuss a specific example: the *a posteriori* estimation of the error in the outflow flux across the boundary (or part of the boundary) for a symmetric hyperbolic system.

Let us consider the non-homogeneous boundary value problem (3.1), (3.3), where $\mathbf{f} \in [L_2(\Omega)]^m$ and $\mathbf{g} \in [L_2(\partial\Omega)]^m$. For the sake of simplicity, we assume

that $\mathbf{g}$ belongs to the restriction of the trial space $\mathcal{U}^h$ to the boundary, and we approximate the non-homogeneous problem by the following method: find $\mathbf{u}_h$ in $\mathcal{U}^h$ such that

$$\begin{aligned} \Pi_h L\mathbf{u}_h &= \Pi_h \mathbf{f} \qquad \text{in } \Omega, \\ B^-\mathbf{u}_h\big|_{\partial\Omega} &= B^-\mathbf{g} \qquad \text{on } \partial\Omega. \end{aligned}$$

Given any $\psi \in [L_2(\Omega)]^m$, consider the linear functional

$$N_\psi(\mathbf{v}) = \int_{\partial\Omega} B^+\mathbf{v}\,\psi \mathrm{d}s, \qquad \mathbf{v} \in [H^1(\Omega)]^m;$$

here ψ plays the rôle of a weight function that can be chosen freely (e.g. ψ can be taken to have its support contained in a subset of Ω, etc.). We seek to approximate the (weighted) outflow flux $N_\psi(\mathbf{u})$ by $N_\psi(\mathbf{u}_h)$.

Theorem 4.1 *Suppose that hypotheses (a), (b) and (c) hold, and that the entries of C are continuously differentiable on $\overline{\Omega}$. Then, for each $\psi \in [H^1(\partial\Omega)]^m$, assuming that $\mathbf{u}$ and $\mathbf{u}_h$ belong to $[H^1(\Omega)]^m$, we have that*

$$|N_\psi(\mathbf{u}) - N_\psi(\mathbf{u}_h)| \le c_1' c_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)} \right)^{1/2} \|\psi\|_{H^1(\partial\Omega)}.$$

Proof Consider the dual problem

$$\begin{aligned} L^*\phi &= 0 \qquad \text{in } \Omega, \\ B^+\phi|_{\partial\Omega} &= B^+\psi \qquad \text{on } \partial\Omega. \end{aligned}$$

Since $\mathbf{u} - \mathbf{u}_h$ is in $D(L,\Omega) \cap [H^1(\Omega)]^m$, Green's formula from Theorem 3.1 gives

$$0 = (\mathbf{u} - \mathbf{u}_h, L^*\phi) = (L(\mathbf{u} - \mathbf{u}_h), \phi) - N_\psi(\mathbf{u} - \mathbf{u}_h).$$

Consequently, for any $\phi_h \in \mathcal{M}_h$,

$$N_\psi(\mathbf{u}) - N_\psi(\mathbf{u}_h) = (\mathbf{r}_h, \phi - \phi_h).$$

Exploiting hypothesis (c), it follows that

$$|N_\psi(\mathbf{u}) - N_\psi(\mathbf{u}_h)| \le c_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)} \right)^{1/2} |\phi|_{H^1(\Omega)},$$

and therefore, by the hyperbolic regularity theorem

$$|N_\psi(\mathbf{u}) - N_\psi(\mathbf{u}_h)| \le c_1' c_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)} \right)^{1/2} \|\psi\|_{H^1(\partial\Omega)}.$$

□

A particularly relevant case concerns the estimation of the flux through an open subset $\gamma \subset \partial\Omega$. In this case it is tempting to take $\psi = \chi_\gamma$ where χ_γ is the characteristic function of γ; unfortunately such ψ does not belong to $[H^1(\partial\Omega)]^m$ so Theorem 4.1 does not apply.

We shall illustrate on a simple example of a scalar hyperbolic problem how a refined analysis may be pursued to obtain a bound on the flux across $\gamma \subset \Omega$ by recalling that singularities propagate along (bi-)characteristics.

Consider the scalar hyperbolic equation

$$Lu \equiv \nabla \cdot (\mathbf{a}u) + cu = f \qquad \text{in } \Omega \tag{4.1}$$

for Ω the unit square $(0,1)^2$, where $\mathbf{a} = (a_1, a_2)$ is a vector function with positive, continuously differentiable entries defined on $\overline{\Omega}$, c is a continuous function of $\overline{\Omega}$, and $f \in L_2(\Omega)$.

Let $b = \hat{\nu}_1 a_1 + \hat{\nu}_2 a_2$, where $\hat{\nu} = (\hat{\nu}_1, \hat{\nu}_2)$ is the unit outward normal vector field to the boundary of Ω; at the corners of the square we define $\hat{\nu}$ to be the outward pointing unit vector collinear with the diagonal passing through that corner. Clearly b is non-zero almost everywhere on $\partial\Omega$, i.e. $\partial\Omega$ is non-characteristic for L. We consider the partial differential equation (4.1) subject to the homogeneous boundary condition

$$b^- u = 0 \qquad \text{on } \partial\Omega; \tag{4.2}$$

here $b^- = \min(b, 0)$ denotes the negative part of b. The boundary condition (4.2) is equivalent to requiring that $u = 0$ along the *inflow* part of the boundary,

$$\partial_-\Omega = \{x = (x_1, x_2) \in \partial\Omega \ : \ \mathbf{a}(x) \cdot \hat{\nu}(x) < 0\};$$

The complement of the inflow boundary,

$$\partial_+\Omega = \{x = (x_1, x_2) \in \partial\Omega \ : \ \mathbf{a}(x) \cdot \hat{\nu}(x) \geq 0\},$$

is called the *outflow* part of the boundary. Let us suppose that γ is a connected, relatively open measurable subset of the outflow boundary $\partial_+\Omega$, and that we wish to approximate the normal outflow flux along γ, defined as

$$N_\psi(u) = \int_{\partial_+\Omega} (\mathbf{a} \cdot \hat{\nu}) u \, \psi \, \mathrm{d}s, \qquad u \in H^1(\Omega) \cap D(L, \Omega),$$

where $\psi = \chi_\gamma \zeta$, and ζ is a fixed element in $H^1(\gamma)$ which plays the rôle of a weight function (e.g. $\zeta \equiv 1$ on γ). Now Sobolev's embedding theorem implies that ζ is continuous on $\overline{\gamma}$, and therefore ψ has possible discontinuities only at the points α and β, say, that comprise the boundary of γ. Consider the characteristic curves C_α and C_β of L in $\overline{\Omega}$ that contain the points α and β, respectively; these two curves divide Ω into three disjoint regions, one of which, denoted Ω_0, has γ as

its outflow boundary, and two other regions over each of which ϕ, the solution of the dual problem,

$$\begin{aligned} L^*\phi \equiv -\mathbf{a}\cdot\nabla\phi + c\phi &= 0 \qquad \text{in } \Omega \\ \phi|_{\partial_+\Omega} &= \psi \qquad \text{on } \partial_+\Omega, \end{aligned}$$

is identically zero. Let $\Omega_{\alpha,h}$ be a characteristic strip of cross-section $c_7^2 h/2$ that contains the characteristic C_α in its interior, with $\Omega_{\beta,h}$ defined analogously, so that the elements in the partition of Ω that are in contact with C_α (respectively C_β) are contained in $\Omega_{\alpha,h}$ (respectively $\Omega_{\beta,h}$). We define $\gamma_{\alpha,h}$ (respectively $\gamma_{\beta,h}$) as the intersection of the closure of $\Omega_{\alpha,h}$ (respectively $\Omega_{\beta,h}$) with $\partial_+\Omega$.

Theorem 4.2 *Let u and u_h belong to $H^1(\Omega)\cap D(L,\Omega)$, and assume that the finite element test space possesses the following approximation property:*

(c') for each $\phi\in L_2(\Omega)$ that is piecewise H^1 on Ω, there exists $\phi_h\in\mathcal{M}^h$ and a positive constant c_5, independent of h, such that

$$\|h^{-k}(\phi-\phi_h)\|_{L_2(\kappa)} \le c_5|\phi|_{H^k(\kappa)},$$

where $k=0$ if $\phi\in L_2(\kappa)$ and $k=1$ if $\phi\in H^1(\kappa)$.

Then we have the following error bound:

$$\begin{aligned} |N_\psi(u)-N_\psi(u_h)| \le\ & c_5c_6c_7\left(\sum_{\kappa\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|h^{1/2}r_h\|^2_{L_2(\kappa)}\right)^{\frac12}\|\psi\|_{L_\infty(\gamma)} \\ & + c_5c_6\left(\sum_{\kappa\not\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|hr_h\|^2_{L_2(\kappa)}\right)^{\frac12}\|\psi\|_{H^1(\gamma)}, \end{aligned}$$

where c_5, c_6 and c_7 are positive constants, independent of h, and κ signifies any finite element in the partition of Ω whose closure has nonempty intersection with $\overline{\Omega}_0$, the domain of influence for $\overline{\gamma}$.

Proof Since $\phi\equiv 0$ in the complement of $\overline{\Omega}_0$, it will only be necessary to consider those elements κ in the partition of the domain Ω whose closure has nonempty intersection with $\overline{\Omega}_0$. In order to simplify the notation we shall implicitly assume throughout the proof that, for each element κ that we refer to, $\overline{\kappa}\cap\overline{\Omega}_0$ is non-empty. Arguing in the same manner as in Theorem 4.1, we deduce that

$$\begin{aligned} |N_\psi(u)-N_\psi(u_h)| &= |(r_h,\phi-\phi_h)| \le \sum_\kappa |(r_h,\phi-\phi_h)_\kappa| \\ &= \sum_{\kappa\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}|(r_h,\phi-\phi_h)_\kappa| + \sum_{\kappa\not\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}|(r_h,\phi-\phi_h)_\kappa| \\ &\le \sum_{\kappa\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|r_h\|_{L_2(\kappa)}\|\phi-\phi_h\|_{L_2(\kappa)} \end{aligned}$$

$$
\begin{aligned}
&+\sum_{\kappa\not\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|hr_h\|_{L_2(\kappa)}\|h^{-1}(\phi-\phi_h)\|_{L_2(\kappa)}\\
\leq\ & c_5\left(\sum_{\kappa\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|r_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}}\|\phi\|_{L_2(\Omega_{\alpha,h}\cup\Omega_{\beta,h})}\\
&+c_5\left(\sum_{\kappa\not\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|hr_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}}|\phi|_{H^1(\Omega\setminus(\Omega_{\alpha,h}\cup\Omega_{\beta,h}))}.
\end{aligned}
$$

Thus, exploiting hyperbolic regularity on $\Omega_{\alpha,h}\cup\Omega_{\beta,h}$ and its complement, we deduce that

$$
\begin{aligned}
|N_\psi(u)-N_\psi(u_h)| &\leq c_5c_6\left(\sum_{\kappa\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|r_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}}\|\psi\|_{L_2(\gamma_{\alpha,h}\cup\gamma_{\beta,h})}\\
&+c_5c_6\left(\sum_{\kappa\not\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|hr_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}}\|\psi\|_{H^1(\gamma\setminus(\gamma_{\alpha,h}\cup\gamma_{\beta,h}))}\\
&\leq c_5c_6c_7\left(\sum_{\kappa\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|h^{1/2}r_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}}\|\psi\|_{L_\infty(\gamma)}\\
&+c_5c_6\left(\sum_{\kappa\not\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|hr_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}}\|\psi\|_{H^1(\gamma)},
\end{aligned}
$$

and hence the desired result. □

Now choosing $\psi=\chi_\gamma$, the characteristic function of γ, we deduce from Theorem 4.2 that

$$
\begin{aligned}
\left|\int_\gamma(\mathbf{a}\cdot\hat{\nu})u\,\mathrm{d}s-\int_\gamma(\mathbf{a}\cdot\hat{\nu})u_h\,\mathrm{d}s\right| \leq\ & c_5c_6c_7\left(\sum_{\kappa\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|h^{1/2}r_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}}\\
&+c_5c_6|\gamma|^{1/2}\left(\sum_{\kappa\not\subset\Omega_{\alpha,h}\cup\Omega_{\beta,h}}\|hr_h\|^2_{L_2(\kappa)}\right)^{\frac{1}{2}},
\end{aligned}
$$

where $|\gamma|$ denotes the one-dimensional measure of the set γ. This error bound indicates that in order to reduce the size of the error in the normal flux across γ one has to refine only those elements κ that have non-empty intersection with the domain of influence for γ. Note also, that since, at least on uniformly regular partitions, the number of entries in the two sums on the right are, respectively, $O(h^{-1})$ and $O(h^{-2})$, the two terms on the right-hand side are commensurable.

Due to theoretical results about finite speed of propagation of singularities along bi-characteristics and the microlocal regularity theorem of Hörmander, one would expect that a bound akin to Theorem 4.2 holds more generally for a large class of hyperbolic systems.

5 Unsteady problems

In this section we develop the *a posteriori* error analysis of a class of evolution-Galerkin finite element methods for unsteady scalar hyperbolic problems. For a detailed overview of the (unconditional) stability and accuracy properties of evolution-Galerkin methods we refer to [20].

Given a final time $T > 0$, a function $f \in L_2(I; L_2(\Omega))$ with $I = (0, T]$, and $u_0 \in L_2(\Omega)$, we consider the problem

$$\begin{aligned} \frac{\partial u}{\partial t} + \mathbf{a} \cdot \nabla u &= f, \qquad \mathbf{x} \in \Omega,\ t \in I, && (5.1) \\ u(\mathbf{x}, 0) &= u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega, && (5.2) \end{aligned}$$

where $\Omega = \mathbb{R}^d$. For the sake of simplicity we assume that the velocity field $\mathbf{a}$ is in $C([0,T]; [C_0^1(\Omega)]^d)$, that it is incompressible, i.e. $\nabla \cdot \mathbf{a} = 0$ on $\Omega \times [0,T]$, and that the supports of u_0 and f are compact subsets in Ω and $\Omega \times [0,T]$, respectively. This problem has a unique weak solution u in $L_\infty(I; L_2(\Omega))$; moreover, if $u_0 \in H^1(\Omega)$ and $f \in L_2(I; H^1(\Omega))$ then u belongs to $L_2(I; H^1(\Omega))$.

Let $0 = t^0 < t^1 < \ldots < t^M < t^{M+1} = T$ be a subdivision (not necessarily uniform) of I, with corresponding time intervals $I^n = (t^{n-1}, t^n]$ and time steps $k_n = t^n - t^{n-1}$. For each n, let $T^n = \{\kappa\}$ be a partition of Ω into closed simplices κ, with corresponding mesh function h_n, satisfying

$$\begin{aligned} C_\dagger h_\kappa^2 &\le \mathrm{meas}(\kappa) \quad \forall \kappa \in T^n, && (5.3) \\ C_* h_\kappa \le h_n(\mathbf{x}) &\le h_\kappa \qquad \forall \mathbf{x} \in \kappa \quad \forall \kappa \in T^n, && (5.4) \end{aligned}$$

where h_κ is the diameter of κ, and $C_\dagger$ and C_* are positive constants. Further, h will denote the global mesh function given by $h(\mathbf{x}, t) = h_n(\mathbf{x})$ for $(\mathbf{x}, t)$ in $\Omega \times I^n$, and we define the corresponding time step function $k = k(t)$ by $k(t) = k_n,\ t \in I^n$.

Let $\Lambda^n = \Omega \times I^n$; for $p, q \in \mathbb{N}_0$, let

$$\begin{aligned} S^{h_n} &= \{v \in H^1(\Omega) : v|_\kappa \in \mathbf{P}_p(\kappa)\ \forall \kappa \in T^n\}, \\ V^{h_n} &= \{v : v(\mathbf{x}, t)|_{\Lambda^n} = \sum_{j=0}^{q} t^j v_j,\ v_j \in S^{h_n}\}, \\ V^h &= \{v : v(\mathbf{x}, t)|_{\Lambda^n} \in V^{h_n},\ n = 1, \ldots, M+1\}, \end{aligned}$$

where $\mathbf{P}_p(\kappa)$ denotes the set of polynomials of degree at most p over κ.

In the following, we shall assume that $p = 1$ and $q = 0$. We note that if $v \in V^{h_n}$ for $n = 1, \ldots, M+1$, then v is continuous in space at any time, but may be discontinuous in time at the discrete time levels t^n. To account for this, we introduce the notation $v_\pm^n := \lim_{s \to 0^+} v(t^n \pm s)$, and $[v^n] := v_+^n - v_-^n$.

The evolution-Galerkin approximation of (5.1), (5.2) makes use of the particle trajectories (or characteristics) associated with equation (5.1): the path $\mathbf{X}(\mathbf{x}, s; \cdot)$

of a particle located at position $\mathbf{x} \in \Omega$ at time $s \in [0,T]$ is defined as the solution of the initial value problem

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{X}(\mathbf{x},s;t) &= \mathbf{a}(\mathbf{X}(\mathbf{x},s;t),t), & (5.5)\\ \mathbf{X}(\mathbf{x},s;s) &= \mathbf{x}. & (5.6) \end{aligned}$$

For u smooth enough, the *material derivative* $D_t u$ is then defined by

$$\begin{aligned} D_t u(\mathbf{x},s) &:= \frac{\mathrm{d}}{\mathrm{d}t} u(\mathbf{X}(\mathbf{x},s;t),t)\mid_{t=s} \\ &= \frac{\partial}{\partial t}u(\mathbf{x},s) + \mathbf{a}(\mathbf{x},s)\cdot\nabla u(\mathbf{x},s) \quad \forall \mathbf{x}\in\Omega,\ s\in I. \qquad (5.7) \end{aligned}$$

The evolution-Galerkin time-discretisation involves approximating the material derivative by a divided difference operator. The simplest appropriate discretisation is the backward Euler method, giving, for $n = 0,\ldots,M$,

$$D_t u(\cdot,t^{n+1}) \approx \frac{u(\cdot,t^{n+1}) - u(\mathbf{X}(\cdot,t^{n+1};t^n),t^n)}{k_{n+1}}.$$

If we now let u_h^n denote the Galerkin finite element approximation to $u(\cdot,t^n)$ at time t^n, then applying the finite element method in space yields what is known as the evolution-Galerkin discretisation of (5.1): find $u_h^{n+1} \in S^{h_{n+1}}$, for $n = 0,...,M$, such that

$$\begin{aligned} \left(\frac{u_h^{n+1} - u_h^n(\mathbf{X}(\cdot,t^{n+1};t^n))}{k_{n+1}}, v\right) &= (\bar{f},v) \quad \forall v\in S^{h_{n+1}}, & (5.8)\\ (u_h^0, v) &= (u_0,v) \quad \forall v\in S^{h_0}, & (5.9) \end{aligned}$$

where $\bar{f}|_{\Lambda^{n+1}} := f(\cdot,t^{n+1})$. Alternatively, by integrating (5.8) with respect to t over I^{n+1}, we obtain the following equivalent formulation: find u_h such that, for $n = 0,1,\ldots,M$, $u_h|_{\Lambda^{n+1}} \in V^{h_{n+1}}$ and satisfies

$$\begin{aligned} (D_t^h u_h, v)_{n+1} &= (\bar{f},v)_{n+1} \quad \forall v\in V^{h_{n+1}}, & (5.10)\\ (u_{h-}^0, v) &= (u_0,v) \quad \forall v\in V^{h_0}, & (5.11) \end{aligned}$$

where

$$D_t^h u_h|_{\Lambda^{n+1}} = (u_{h-}(\mathbf{X}(\mathbf{x},t^{n+1};t^{n+1}),t^{n+1}) - u_{h-}(\mathbf{X}(\mathbf{x},t^{n+1};t^n),t^n))/k_{n+1};$$

here, for $v,w \in L_2(I^{n+1};L_2(\Omega))$, we have used the notation

$$(v,w)_{n+1} = \int_{t^n}^{t^{n+1}} (v,w)\mathrm{d}t.$$

We remark that in (5.10), (5.11) the space discretisation may vary in both space and time, but the time steps are only variable in time and not in space. Hence, the corresponding mesh will not be fully optimal.

5.1 A posteriori error analysis

In this subsection we derive an *a posteriori* estimate for the error, $e = u - u_h$, in the $L_\infty(0,T;H^{-1}(\Omega))$ norm, where u and u_h are the solutions of (5.1), (5.2) and (5.10), (5.11), respectively. We shall first introduce the following notation: for v and w in $L_2(0,T;L_2(\Omega))$ and $Q := \Omega \times I$, we define

$$\begin{aligned}(v,w)_Q &:= \sum_{n=0}^{M} \int_{t^n}^{t^{n+1}} (v,w)\mathrm{d}t,\\ \|v\|_Q &:= \left((v,v)_Q\right)^{1/2},\end{aligned}$$

where $(\cdot,\cdot)$ is the inner product in $L_2(\Omega)$; throughout this section $\|\cdot\|$ denotes the $L_2(\Omega)$ norm.

5.1.1 Error representation

Given that $\psi \in C_0^\infty(\Omega)$, consider the dual problem: find ϕ such that

$$-\phi_t - \nabla\cdot(\mathbf{a}\phi) = 0, \qquad \mathbf{x}\in\Omega,\ t\in[0,T), \tag{5.12}$$

$$\phi(\mathbf{x},T) = \psi(\mathbf{x}), \qquad \mathbf{x}\in\Omega. \tag{5.13}$$

Multiplying (5.12) by e and integrating by parts in both space and time gives

$$\begin{aligned}&(e(T),\psi) = (e(T),\phi(T))\\ &= \int_0^T (e,\phi_t)\mathrm{d}t + \sum_{n=0}^{M}\int_{t^n}^{t^{n+1}} (e_t,\phi)\mathrm{d}t - \sum_{n=0}^{M}([u_h^n],\phi(t^n)) + (u_0 - u_{h-}^0,\phi(0))\\ &= \sum_{n=0}^{M}\int_{t^n}^{t^{n+1}} (f - \mathbf{a}\cdot\nabla u_h,\phi)\mathrm{d}t - \sum_{n=0}^{M}([u_h^n],\phi(t^n)) + (u_0 - u_{h-}^0,\phi(0)).\end{aligned}$$

If we now let $\Phi \in V^h$, then using (5.10) we have

$$\begin{aligned}(e(T),\psi) &= \sum_{n=0}^{M}\int_{t^n}^{t^{n+1}} ([u_h^n]/k_{n+1} + \mathbf{a}\cdot\nabla u_h - f, \Phi-\phi)\mathrm{d}t\\ &\quad + \sum_{n=0}^{M}\int_{t^n}^{t^{n+1}} (D_t^h u_h - ([u_h^n]/k_{n+1} + \mathbf{a}\cdot\nabla u_h),\Phi)\mathrm{d}t\\ &\quad + \sum_{n=0}^{M}\int_{t^n}^{t^{n+1}} ([u_h^n]/k_{n+1}, \phi-\phi(t^n))\mathrm{d}t + \sum_{n=0}^{M}\int_{t^n}^{t^{n+1}} (f-\bar f,\Phi)\mathrm{d}t\\ &\quad + (u_0 - u_{h-}^0,\phi(0))\\ &\equiv \mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}+\mathrm{V} \quad \forall\Phi\in V^h.\end{aligned} \tag{5.14}$$

It remains to estimate each of the terms I to V.

5.1.2 Interpolation/projection estimates for the dual problem

We shall now choose $\Phi \in V^h$ in (5.14) as a suitable approximation to ϕ. To do so, consider the quasi-interpolation operator $P_n : L_2(\Omega) \to S^{h_n}$ in the spatial variables (see, for example, Bernardi [3]), and the L_2 orthogonal projector $\pi_n : L_2(I^n) \to \mathbf{P}_0(I^n)$ in time defined by

$$\int_{t^{n-1}}^{t^n} (\pi_n \phi - \phi) v \mathrm{d}t \;=\; 0 \quad \forall v \in \mathbf{P}_0(I^n); \tag{5.15}$$

then, we define (locally) $\Phi|_{\Lambda^n} \in V^{h_n}$ by letting

$$\Phi|_{\Lambda^n} = P_n \pi_n \phi = \pi_n P_n \phi \in V^{h_n},$$

where $\phi = \phi|_{\Lambda^n}$. Further, we introduce P and π by

$$(P\phi)|_{\Lambda^n} \;=\; P_n(\phi|_{\Lambda^n}), \tag{5.16}$$

$$(\pi\phi)|_{\Lambda^n} \;=\; \pi_n(\phi|_{\Lambda^n}), \tag{5.17}$$

and define $\Phi \in V^h$ as

$$\Phi = P\pi\phi = \pi P\phi \in V^h. \tag{5.18}$$

It follows from the properties of the quasi-interpolant that

$$\|\Phi\|_Q \le C_2^i \|\phi\|_Q, \tag{5.19}$$

with C_2^i a positive constant, and that the following theorem holds.

Theorem 5.1 *Let $n \in \mathbb{N}_0$ and suppose that T^n satisfies conditions (5.3), (5.4). Let $w \in H^1(\Omega)$ have compact support; then, with $C_1^i > 0$ a constant,*

$$\left(\sum_\kappa \|h_n^{-1}(P_n w - w)\|^2_{L_2(\kappa)} \right)^{1/2} \;\le\; C_1^i C_*^{-1} |w|_{H^1(\Omega)}, \tag{5.20}$$

$$\left(\sum_\kappa \|P_n w\|^2_{L_2(\kappa)} \right)^{1/2} \;\le\; C_2^i \|w\|. \tag{5.21}$$

In the next two lemmas we provide bounds for the operators P and π, in order to estimate $\phi - \Phi = \phi - P\pi\phi$.

Lemma 5.2 *Suppose that $R \in L_2(I; L_2(\Omega))$; then*

$$|(R, P\phi - \phi)_Q| \;\le\; C_1^i C_*^{-1} \|hR\|_Q \|\nabla\phi\|_Q. \tag{5.22}$$

Proof Using the Cauchy-Schwarz inequality gives

$$|(R, P\phi - \phi)_Q| \leq \|hR\|_Q \|h^{-1}(P\phi - \phi)\|_Q.$$

Now

$$\|h^{-1}(P\phi - \phi)\|_Q = \left(\sum_{n=0}^{M} \int_{t^n}^{t^{n+1}} \|h_{n+1}^{-1}(P_{n+1}\phi - \phi)\|^2 \right)^{1/2}, \tag{5.23}$$

and

$$\|h_{n+1}^{-1}(P_{n+1}\phi - \phi)\| \leq C_1^i C_*^{-1} \|\nabla\phi\|.$$

Therefore

$$\|h^{-1}(P\phi - \phi)\|_Q \leq C_1^i C_*^{-1} \|\nabla\phi\|_Q.$$

□

The proof of the next Lemma is given in [8].

Lemma 5.3 *Suppose that $R \in L_2(I; L_2(\Omega))$; then*

$$|(R, P(\pi\phi - \phi))_Q| \leq C_2^i C_3^i \|kR\|_Q \|\phi_t\|_Q, \tag{5.24}$$

$$|\sum_{n=0}^{M} \int_{t^n}^{t^{n+1}} (R, \phi^n - \phi) \mathrm{d}t| \leq \|kR\|_Q \|\phi_t\|_Q, \tag{5.25}$$

where $C_3^i = 1/\sqrt{2}$ and $\phi^n = \phi(\mathbf{x}, t^n)$.

5.1.3 Strong stability of the dual problem

This brief subsection summarises the stability estimates for the solution ϕ of the dual problem (5.12), (5.13) in terms of the corresponding final datum ψ.

Lemma 5.4 *There exists a constant C_1^s such that*

$$\|\phi\|_{L_\infty(I;L_2(\Omega))} \leq C_1^s \|\psi\|.$$

Lemma 5.5 *There exists a constant $C_2^s = C_2^s(T, \mathbf{a})$ such that*

$$\|\nabla\phi\|_Q \leq C_2^s \|\nabla\psi\|.$$

Lemma 5.6 *There exists a constant $C_3^s = C_3^s(T, \mathbf{a})$ such that*

$$\|\phi_t\|_Q \leq C_3^s \|\psi\|_{H^1(\Omega)}.$$

Recalling that $\nabla \cdot \mathbf{a} = 0$, a straightforward energy analysis shows that $C_1^s = 1$, $C_2^s = \max(1, \mathrm{e}^{-\alpha T})$ where α is any real number such that $\nabla\mathbf{a} \geq \alpha I$, and $C_3^s = C_2^s \|\mathbf{a}\|_{L_\infty(\Omega)}$.

5.1.4 Completion of the proof of the a posteriori error estimate

We shall now proceed to estimate the terms I-V on the right-hand side of (5.14). For the first term I, we have

$$\begin{aligned} \mathrm{I} &= \sum_{n=0}^{M} \int_{t^n}^{t^{n+1}} ([u_h^n]/k_{n+1} + \mathbf{a}\cdot\nabla u_h - f, \Phi - \phi)\mathrm{d}t \\ &= (R_1, \Phi - \phi)_Q = (R_1, P\phi - \phi)_Q + (R_1, P(\pi\phi - \phi))_Q \\ &\equiv \mathrm{I}_1 + \mathrm{I}_2, \end{aligned}$$

where P and π are as defined by (5.16) and (5.17), and

$$R_1|_{\Lambda^{n+1}} = [u_h^n]/k_{n+1} + \mathbf{a}\cdot\nabla u_h - f.$$

By Lemma 5.2 we have

$$|\mathrm{I}_1| \le C_1^i C_*^{-1} \|hR_1\|_Q \|\nabla\phi\|_Q \le C_1^i C_2^s C_*^{-1} \|hR_1\|_Q \|\psi\|_{H^1(\Omega)}.$$

Similarly, using Lemma 5.3, we have

$$|\mathrm{I}_2| \le C_2^i C_3^i \|kR_1\|_Q \|\phi_t\|_Q \le C_2^i C_3^s C_3^i \|kR_1\|_Q \|\psi\|_{H^1(\Omega)}.$$

Hence,

$$|\mathrm{I}| \le \left(C_1^i C_2^s C_*^{-1} \|hR_1\|_Q + C_2^i C_3^s C_3^i \|kR_1\|_Q\right) \|\psi\|_{H^1(\Omega)}.$$

Next we consider term II:

$$\begin{aligned} \mathrm{II} &\le \|kR_2\|_Q \|\Phi\|_Q \le C_2^i \|kR_2\|_Q \|\phi\|_Q \le C_2^i \sqrt{T} \|kR_2\|_Q \|\phi\|_{L_\infty(I;L_2(\Omega))} \\ &\le C_1^s C_2^i \sqrt{T} \|kR_2\|_Q \|\psi\|, \end{aligned} \tag{5.26}$$

where

$$R_2|_{\Lambda^{n+1}} = (D_t^h u_h - ([u_h^n]/k_{n+1} + \mathbf{a}\cdot\nabla u_h))/k_{n+1}.$$

For the term III, using Lemma 5.3 and Lemma 5.6, we have

$$|\mathrm{III}| \le \|kR_3\|_Q \|\phi_t\|_Q \le C_3^s \|kR_3\|_Q \|\psi\|_{H^1(\Omega)},$$

where

$$R_3|_{\Lambda^{n+1}} = [u_h^n]/k_{n+1} = (u_h^{n+1} - u_h^n)/k_{n+1}.$$

Now consider the term IV:

$$\begin{aligned} |\mathrm{IV}| &\le \|f - \bar{f}\|_Q \|\Phi\|_Q \le C_2^i \|f - \bar{f}\|_Q \|\phi\|_Q \le C_2^i \sqrt{T} \|f - \bar{f}\|_Q \|\phi\|_{L_\infty(I;L_2(\Omega))} \\ &\le C_1^s C_2^i \sqrt{T} \|f - \bar{f}\|_Q \|\psi\|. \end{aligned}$$

Finally, for term V, the Cauchy-Schwarz inequality and Lemma 5.4 give the following chain of inequalities:

$$|\mathrm{V}| \leq \|u_0 - u_{h-}^0\| \, \|\phi(0)\| \leq \|u_0 - u_{h-}^0\| \, \|\phi\|_{L_\infty(I;L_2(\Omega))} \leq C_1^s \|u_0 - u_{h-}^0\| \, \|\psi\|.$$

Substituting the bounds for the terms I to V into (5.14), dividing both sides by $\|\psi\|_{H^1(\Omega)}$ and taking the supremum over all ψ in $C_0^\infty(\Omega)$, we obtain the following *a posteriori* error bound.

Theorem 5.7 *Let u and u_h be solutions of (5.1), (5.2) and (5.10), (5.11), respectively. Then*

$$\|u - u_h\|_{L_\infty(0,T;H^{-1}(\Omega))} \leq \overset{\circ}{\mathcal{E}}(u_h, h, k, f), \tag{5.27}$$

where

$$\begin{aligned}
\overset{\circ}{\mathcal{E}}(u_h, h, k, f) &= \mathcal{E}(u_h, h, k, f) + \mathcal{E}_0(u_0, u_{h-}^0, h), \\
\mathcal{E}(u_h, h, k, f) &= C_1 \|hR_1\|_Q + C_2 \|kR_1\|_Q \\
&\quad + C_3 \|kR_2\|_Q + C_4 \|kR_3\|_Q + C_5 \|kR_4\|_Q, \qquad (5.28) \\
\mathcal{E}_0(u_0, u_{h-}^0, h) &= C_6 \|u_0 - u_{h-}^0\|, \qquad (5.29)
\end{aligned}$$

and

$$\begin{aligned}
R_1|_{\Lambda^{n+1}} &= [u_h^n]/k_{n+1} + \mathbf{a} \cdot \nabla u_h - f, \\
R_2|_{\Lambda^{n+1}} &= (D_t^h u_h - ([u_h^n]/k_{n+1} + \mathbf{a} \cdot \nabla u_h))/k_{n+1}, \\
R_3|_{\Lambda^{n+1}} &= [u_h^n]/k_{n+1}, \\
R_4 &= (f - \bar{f})/k,
\end{aligned}$$

and C_i, $i = 1, ..., 6$, are (computable) positive constants; namely, $C_1 = C_1^i C_2^s C_*^{-1}$, $C_2 = C_2^i C_3^s C_3^i$, $C_3 = C_1^s C_2^i \sqrt{T}$, $C_4 = C_3^s$, $C_5 = C_1^s C_2^i \sqrt{T}$, and $C_6 = C_1^s$.

5.2 Computational implementation

In Theorem 5.7 we stated an *a posteriori* error estimate in the $L_\infty(H^{-1})$ norm, for the evolution-Galerkin approximation of the unsteady hyperbolic problem (5.1), (5.2), of the form:

$$\|u - u_h\|_{L_\infty(I;H^{-1}(\Omega))} \leq \overset{\circ}{\mathcal{E}}(u_h, h, k, f), \tag{5.30}$$

where $\overset{\circ}{\mathcal{E}}(u_h, h, k, f)$ is a computable global error estimator.

We shall now consider the problem of implementing the error bound (5.30) into an adaptive evolution-Galerkin finite element algorithm to ensure global

error control (and thereby reliability) with respect to a prescribed tolerance, TOL, i.e.

$$\|u - u_h\|_{L_\infty(I;H^{-1}(\Omega))} \leq \text{TOL}, \tag{5.31}$$

with the minimum amount of computational effort.

We begin, in Subsection 5.2.1 by designing an adaptive algorithm to determine the mesh parameters h and k in such a way that (5.31) holds. Then, in Subsection 5.2.2, we describe how the space-time mesh may be locally adapted so that the actual mesh parameters h and k are 'close' to those predicted by the adaptive algorithm.

5.2.1 Adaptive algorithm

For a given tolerance, TOL, we consider the problem of finding a discretisation in space and time $\mathcal{S}^h = \{(T^n, t^n)\}_{n\geq 0}$ such that:

1. $\|u - u_h\|_{L_\infty(I;H^{-1}(\Omega))} \leq \text{TOL}$;
2. $\mathcal{S}^h$ is optimal in the sense that the number of degrees of freedom is minimal.

In order to satisfy these criteria we shall use the *a posteriori* error estimate (5.30) to choose $\mathcal{S}^h$ such that:

1. $\mathring{\mathcal{E}}(u_h, h, k, f) \leq \text{TOL}$;
2. The number of degrees of freedom in $\mathcal{S}^h$ is minimal.

The term $\mathcal{E}_0(u_0, u_{h-}^0, h)$ is easily controlled at the start of a computation; so here we shall only consider the problem of constructing $\mathcal{S}_h$ in an efficient way to ensure that

$$\mathcal{E}(u_h, h, k, f) \leq \text{TOL}',$$

where $\text{TOL} = \text{TOL}' + \mathcal{E}_0(u_0, u_{h-}^0, h)$. To do so, we first write $\mathcal{E}$ symbolically in terms of two residual terms: one that controls the spatial mesh and one that controls the temporal mesh, i.e. we let

$$\mathcal{E}(u_h, h, k, f) \equiv C_1'\|hR_1'\|_Q + C_2'\|kR_2'\|_Q. \tag{5.32}$$

Simultaneously, we split the tolerance TOL' into a spatial part, TOL_h, and a temporal part, TOL_k. Thus, for reliability we require that the following conditions hold:

$$C_1'\|hR_1'\|_Q \leq \text{TOL}_h, \tag{5.33}$$
$$C_2'\|kR_2'\|_Q \leq \text{TOL}_k. \tag{5.34}$$

To design the space-time mesh $\mathcal{S}^h$, at each time level t^n we decompose the norm in (5.33) into norms over elements $\kappa \in T^n$, and the norm in (5.34) into norms over time slabs as follows:

$$\begin{aligned} C_1'\|hR_1'\|_Q &\leq C_1'\sqrt{T}\max_{1\leq n\leq M+1}\|h_nR_1'(u_h^n)\| \\ &\leq C_1'\sqrt{NT}\max_{1\leq n\leq M+1}\left(\max_{\kappa\in T^n}\|h_nR_1'(u_h^n)\|_{L_2(\kappa)}\right), \\ C_2'\|kR_2'\|_Q &\leq C_2'\sqrt{T}\max_{1\leq n\leq M+1}\|k_nR_2'(u_h^n)\|, \end{aligned}$$

where N is the number of elements in the spatial mesh at time t^n. Thus, if

$$\begin{aligned} C_1'\sqrt{NT}\|h_nR_1'(u_h^n)\|_{L_2(\kappa)} &\leq \mathrm{TOL}_h \quad \forall\kappa\in T^n, \text{ for } n=1,\ldots,M+1, \\ C_2'\sqrt{T}\|k_nR_2'(u_h^n)\| &\leq \mathrm{TOL}_k, \quad \text{for } n=1,\ldots,M+1, \end{aligned}$$

then (5.33) and (5.34) will automatically hold.

For the practical implementation of this method, we consider the following adaptive algorithm for constructing $\mathcal{S}^h$, under the assumption that the final time T is fixed: for each $n = 1, 2, \ldots, M+1$, with T_0^n a given initial mesh and $k_{n,0}$ an initial time step, determine meshes T_j^n with N_j elements of size $h_{n,j}(\mathbf{x})$ and time steps $k_{n,j}$ and the corresponding approximate solution $u_{h,j}$ defined on I_j^n such that, for $j = 0, 1, \ldots, \hat{n}-1$,

$$C_1\|h_{n,j+1}R_1(u_{h,j}^n)\|_{L_2(\kappa)} = \frac{\mathrm{TOL}_h}{\sqrt{N_jT}} \quad \forall\kappa\in T_j^n, \tag{5.35}$$

$$\begin{aligned} &C_2\|k_{n,j+1}R_1(u_{h,j}^n)\| + C_3\|k_{n,j+1}R_2(u_{h,j}^n)\| \\ &+C_4\|k_{n,j+1}R_3(u_{h,j}^n)\| + C_5\|k_{n,j+1}R_4(u_{h,j}^n)\| = \frac{\mathrm{TOL}_k}{\sqrt{T}}, \end{aligned} \tag{5.36}$$

where $I_j^n = (t^{n-1}, t^{n-1}+k_{n,j}]$ and $\mathrm{TOL}' = \mathrm{TOL}_h + \mathrm{TOL}_k$. We define $T^n = T_{\hat{n}}^n$, $k_n = k_{n,\hat{n}}$ and $h_n = h_{n,\hat{n}}$, where for each n, the number of trials $\hat{n}$ is the smallest integer such that for $j = \hat{n}$, the following stopping condition is satisfied:

$$C_1\|h_{n,\hat{n}}R_1(u_{h,\hat{n}}^n)\|_{L_2(\kappa)} \leq \frac{\mathrm{TOL}_h}{\sqrt{N_{\hat{n}}T}} \quad \forall\kappa\in T_{\hat{n}}^n, \tag{5.37}$$

$$\begin{aligned} &C_2\|k_{n,\hat{n}}R_1(u_{h,\hat{n}}^n)\| + C_3\|k_{n,\hat{n}}R_2(u_{h,\hat{n}}^n)\| \\ &+C_4\|k_{n,\hat{n}}R_3(u_{h,\hat{n}}^n)\| + C_5\|k_{n,\hat{n}}R_4(u_{h,\hat{n}}^n)\| \leq \frac{\mathrm{TOL}_k}{\sqrt{T}}. \end{aligned} \tag{5.38}$$

By construction, this stopping condition will guarantee reliability of the adaptive algorithm; for efficiency, we try to ensure that (5.37) and (5.38) are satisfied with near equality. Since the final time T is fixed, the time step given by (5.36) may need to be limited to ensure that $t^M + k_{M+1,\hat{n}} = T$. For the implementation of this adaptive algorithm, we shall assume that $T_0^n = T^{n-1}$ for $n = 2, 3, \ldots$.

5.2.2 Mesh modification strategy

In Subsection 5.2.1 we described the construction of the space-time mesh $\mathcal{S}^h$ to achieve the required error control,

$$\|u - u_h\|_{L_\infty(I;H^{-1}(\Omega))} \leq \text{TOL}.$$

However, to obtain the desired mesh we must employ a suitable mesh modification technique. Temporal adaptation is quite straightforward, since the time step can just be set equal to $k_{n,j+1}$ given by (5.36). For constructing the spatial mesh in two space dimensions we use the red-green isotropic refinement strategy of Bank [4]. Here, the user must first specify a (coarse) *background* mesh upon which any future refinement will be based. Red refinement corresponds to dividing a certain triangle (father) into four similar triangles (sons) by connecting the midpoints of the sides. Green refinement is only temporary and is used to remove any hanging nodes caused by a red refinement. We note that green refinement is only used on elements which have one hanging node. For elements with two or more hanging nodes a red refinement is performed. The advantage of this refinement strategy is that the degradation of the 'quality' of the mesh is limited since red refinement is obviously harmless and green triangles can *never* be further refined.

Within this mesh modification strategy it is also possible to de-refine the mesh by removing redundant elements, provided that these do not belong to the original background mesh. Thus, to prevent an overly refined mesh in regions where the solution is smooth the background mesh should be chosen suitably coarse. For the practical implementation of this mesh modification strategy we have used the FEMLAB package developed by K. Eriksson.

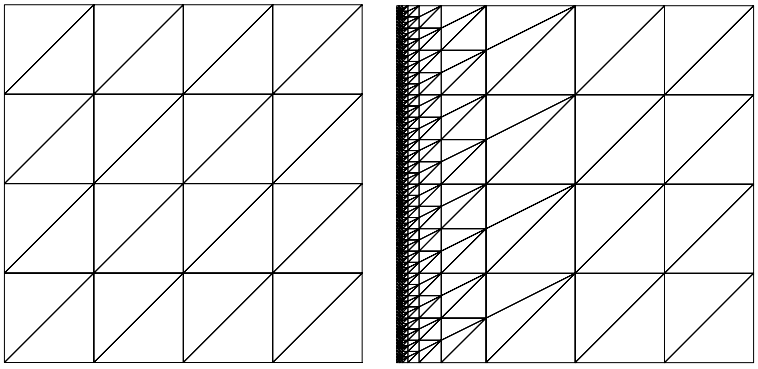
5.3 Numerical experiments

In this section, we present some numerical experiments to illustrate the performance of the adaptive algorithm (5.35), (5.36) on the model hyperbolic test problem:

$$\frac{\partial u}{\partial t} + \mathbf{a} \cdot \nabla u = f, \qquad \mathbf{x} \in \Omega,\ t \in I, \tag{5.39}$$

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{5.40}$$

where $\Omega = (0,1)^2$, $f = 0$, $\mathbf{a} = (2,1)^T$, subject to the boundary condition $u(0,y) = 1$ for $0 \leq y \leq 1$, $u(x,0) = (\delta - x)^+/\delta$ for $0 \leq x \leq 1$. The function u_0 appearing in the initial condition is defined as follows: $u_0(\mathbf{x}) = 0$ for $\mathbf{x} \in \Omega_\delta = (\delta, 1) \times (0,1)$; and for $\mathbf{x} \in \Omega \backslash \Omega_\delta$, $u_0(\mathbf{x})$ is chosen to be the linear function that satisfies the boundary conditions at inflow. We note that initially, for δ small, the solution to this problem has a boundary layer along $x = 0$; this layer then propagates into



(a) (b)

Figure 1: (a) Background mesh, with 25 nodes and 32 elements; (b) Background mesh adapted to resolve the initial condition, with 526 nodes and 900 elements.

the domain Ω, and eventually exits through $x = 1$. In the following, we shall let $\delta = 7.8125 \times 10^{-3}$ and $T = 0.55$.

First, we specify the background mesh to be the 5×5 triangular mesh shown in Figure 1(a). This mesh is initially refined in order to properly resolve the boundary layer along $x = 0$ at time $t = 0$, i.e. so that $\mathcal{E}_0(u_0, u_{h-}^0, h) = 0$ (see Figure 1(b)). Numerical results are presented in Figure 2 for $\mathrm{TOL}_h = 0.01$ and $\mathrm{TOL}_k = 0.25$. The estimated error is $\overset{\circ}{\mathcal{E}}(u_h, h, k, f) = 2.4742 \times 10^{-1}$.

In Figures 2(a), 2(b) and 2(c), 2(d) we see that the adaptive algorithm has refined the spatial mesh in parts of the domain where the solution has a steep layer, and has kept the mesh coarse elsewhere. Figures 2(e), 2(f) show the history of the number of nodes in the spatial mesh against time, and the size of the time step against time, respectively.

We note that the artificial diffusion model introduced in [9] was used in the above experiment with $C_1^{\hat{\epsilon}} = C_2^{\hat{\epsilon}} = 0.2$ and $\hat{\epsilon}_{\max} = 5.0 \times 10^{-4}$.

6 Conclusions

In this paper we have presented an overview of recent developments that concern the *a posteriori* error analysis of finite element approximations to first-order hyperbolic problems. While for elliptic equations there is a well-established theoretical framework of *a posteriori* error estimation that has been successfully implemented into working adaptive algorithms (see, for example, the monograph of Ainsworth and Oden [1] for an extensive list of literature in this area), very much less is known about hyperbolic and nearly-hyperbolic problems. The main problem both from the theoretical and the practical point of view is the estimation of the stability constant of the dual problem that enters the final error

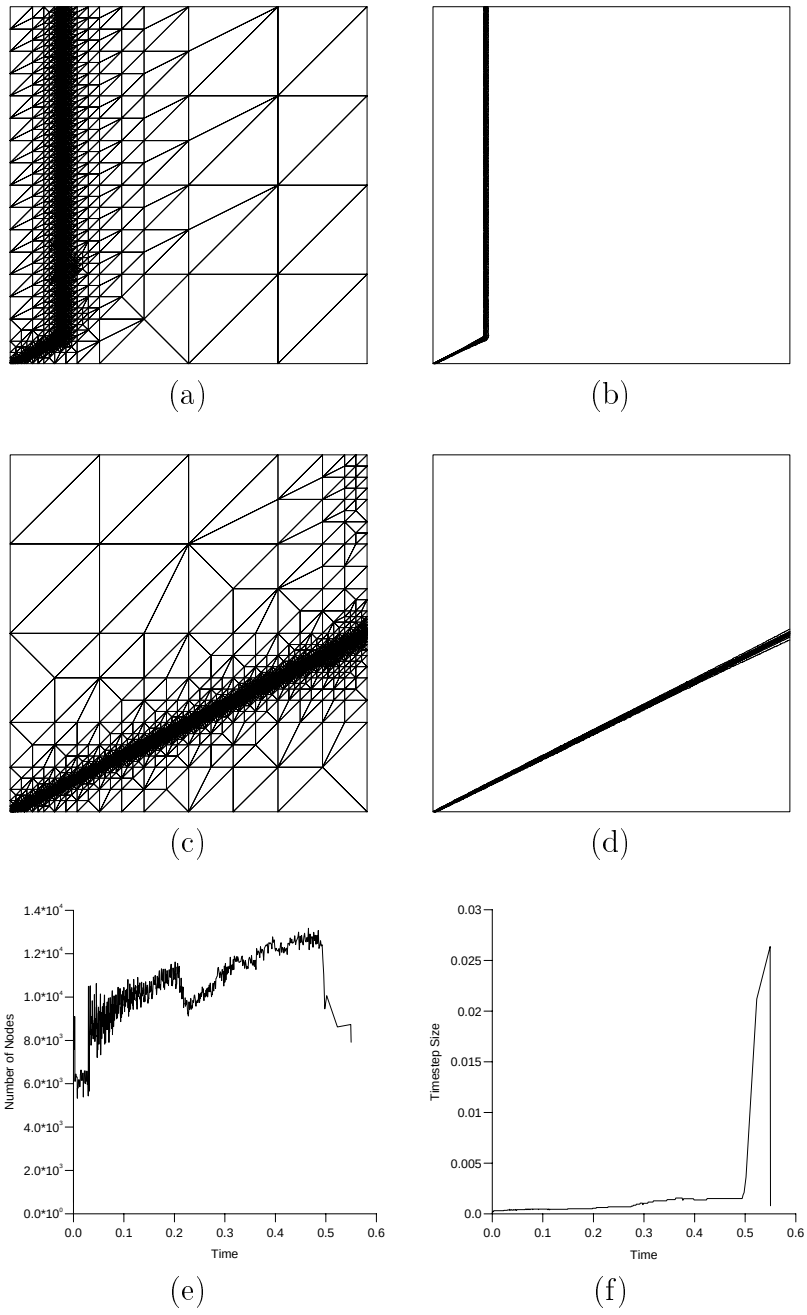


Figure 2: Layer problem for $\mathrm{TOL}_h = 0.01$ and $\mathrm{TOL}_k = 0.25$ with $T = 0.55$: (a) & (b) Mesh and solution (resp.) at time, $t = 0.0714$, with 8425 nodes and 16769 elements; (c) & (d) Mesh and solution (resp.) at final time, $t = 0.55$, with 7914 nodes and 15736 elements; (e) History of nodes against time; (f) History of time step size against time.

bound. Current work in this area focuses on calculating the stability constant during the course of the computation, rather than using general analytical results which tend to overestimate its size by several orders of magnitude.

Another important area where little progress has been made concerns the derivation of two-sided *a posteriori* error bounds for finite element approximations of hyperbolic problems (see, however, [17] where two-sided *a posteriori* bounds have been established for the locally created part of the global error).

Predictably, the *a posteriori* error analysis of finite element and finite volume methods will be an important and fruitful area of research over the next decade; theoretical work in this field is already making impact on large-scale engineering computations, and it is clear that this trend will continue. Indeed, guaranteed error control for the numerical solution of partial differential equations will be the norm rather than the exception for, to quote Babuška, one must have the confidence "to sign the blueprint".

This article is merely a tiny scratch on the surface of a large body of theory that will, we hope, soon be unearthed.

References

[1] Ainsworth, M. and Oden, J.T. (1996) *A Posteriori Error Estimation in Finite Element Analysis.* Series in Computational and Applied Mathematics, Elsevier.

[2] Balland, P. and Süli, E. (1994). Analysis of the cell vertex scheme for hyperbolic problems with variable coefficients. *Oxford University Computing Laboratory Technical Report, NA94/01.* (To appear in SIAM Journal on Numerical Analysis).

[3] Bernardi, C. (1989). Optimal finite-element interpolation on curved domains. *SIAM Journal on Numerical Analysis, 26,* 1212-40.

[4] Bank, R. (1985). *PLTMG user's guide.* Technical Report Edition 4, University of California, San Diego.

[5] Ciarlet, P.G. (1978). *The Finite Element Method for Elliptic Problems,* North Holland, Amsterdam.

[6] Friedrichs, K.O. (1958). Symmetric positive linear differential equations. *Communications in Pure and Applied Mathematics, 11,* 333–418.

[7] Handscomb, D.C. (1995). Error of linear interpolation on a triangle. *Oxford University Computing Laboratory Technical Report, NA95/09.*

[8] Houston, P. and Süli, E. (1995). Adaptive Lagrange-Galerkin methods for unsteady convection–dominated diffusion problems. *Oxford University Computing Laboratory Technical Report, NA95/24.*

[9] Houston, P. and Süli, E. (1996). On the design of an artificial diffusion model for the Lagrange-Galerkin method on unstructured triangular grids. *Oxford University Computing Laboratory Technical Report, NA96/07.*

[10] Johnson, C. (1990). Adaptive finite element methods for diffusion and convection problems. *Computer Methods in Applied Mechanics and Engineering*, *82*, 301–22.

[11] Johnson, C. (1994). A new paradigm for adaptive finite element methods. In: Whiteman, J.R., ed., *The Mathematics of Finite Elements and Applications. Highlights 1993*. John Wiley & Sons, 105–20.

[12] Johnson, C. and Hansbo, P. (1992). Adaptive finite element methods in computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, *101*, 143–81.

[13] Estep, D., Eriksson, K., Hansbo, P. and Johnson, C. (1995). Introduction to adaptive methods for differential equations. *Acta Numerica*. Cambridge University Press, 105–158.

[14] Lax, P.D. and Phillips, R.S. (1960). Local boundary conditions for dissipative symmetric linear differential operators. *Communications in Pure and Applied Mathematics*, **13**, 427–55.

[15] Mackenzie, J., Sonar, T and Süli, E. (1994). Adaptive finite volume methods for hyperbolic problems. In: Whiteman, J.R., ed.,*The Mathematics of Finite Elements and Applications. Highlights 1993*. John Wiley & Sons, 289–98.

[16] Mackenzie, J., Süli, E. and Warnecke, G. (1994). A posteriori error estimates for the cell-vertex finite volume method. In: Hackbusch, W. and Wittum, G., eds., *Adaptive Methods: Algorithms, Theory and Applications*. Vieweg, Braunschweig, *44*, 221–35.

[17] Mackenzie, J., Süli, E. and Warnecke, G. (1995). A posteriori error analysis of Petrov-Galerkin approximations of Friedrichs systems. *Oxford University Computing Laboratory Technical Report. NA95/01.* (Submitted for publication).

[18] Morton, K.W. and Süli, E. (1991). Finite volume methods and their analysis. *IMA Journal of Numerical Analysis*, *11*, 241–60.

[19] Morton, K.W. and Süli, E. (1994). A posteriori and a priori error analysis of finite volume methods. In: Whiteman, J.R., ed., *The Mathematics of Finite Elements and Applications. Highlights 1993*. John Wiley & Sons, 267–88.

[20] Morton, K.W. and Süli, E. (1995). Evolution Galerkin methods and their supraconvergence. *Numerische Mathematik*, *71*, 331–55.

[21] Sonar, T. and Süli, E. (1994). A dual graph-norm refinement indicator for finite volume approximations of the Euler equations. *Oxford University Computing Laboratory Technical Report*, *NA94/09*. (Submitted for publication).

[22] Süli, E. (1989). Finite volume methods on distorted meshes: stability, accuracy, adaptivity. *Oxford University Computing Laboratory Technical Report*, *NA89/06*.

[23] Süli, E. (1992). The accuracy of cell vertex finite volume methods on quadrilateral meshes. *Mathematics of Computation*, *59*, 359–82.

[24] Süli, E. (1991). The accuracy of finite volume methods on distorted partitions. In: Whiteman, J.R., ed., *The Mathematics of Finite Elements and Applications VII*, Academic Press, London, 253–60.

[25] Süli, E. (1996). A posteriori error analysis and global error control for adaptive finite element approximations of hyperbolic problems. In: Griffiths, D.F. and Watson G.A., eds., *Proceedings of the 16th International Conference on Numerical Analysis, Dundee, 1995*, Pitman Research Notes in Mathematics, Longman Scientific and Technical, Harlow.