Report no. 97/21

A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems[a]

Endre Süli

The aim of this article is to present an overview of recent developments in the area of *a posteriori* error estimation for finite element approximations of hyperbolic problems. The approach pursued here rests on the systematic use of hyperbolic duality arguments. We also discuss the question of computational implementation of the *a posteriori* error bounds into adaptive finite element algorithms.

I am grateful to Bernardo Cockburn, Mike Giles, Claes Johnson, John Mackenzie, Rolf Rannacher, Thomas Sonar and Gerald Warnecke for helpful discussions on various aspects of *a posteriori* error analysis and adaptivity. I am particularly indebted to my colleague Paul Houston for performing the numerical experiments which appear in this paper.

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

December, 1997

[a]This work was presented in a lecture series at the International School on Conservation Laws in Freiburg, Germany, October 1997: I wish to express my gratitude to Dietmar Kröner for his kind hospitality.

Contents

1 Introduction

The numerical solution of hyperbolic conservation laws is of fundamental importance in several areas of applied science, particularly fluid dynamics and electromagnetics. Solutions to these partial differential equations frequently exhibit localised structures, such as propagating discontinuities and sharp transition layers whose reliable numerical approximation presents a challenging computational task. Indeed, in order to resolve such localised phenomena in an accurate and efficient way one has to use locally refined computational meshes. In computational fluid dynamics, at least, traditional approaches for constructing such locally adapted meshes resort to *ad hoc* criteria, usually justified on physical grounds, whose impact on the accuracy of the numerical solution is difficult to assess.

In contrast with such heuristic devices, in these notes we shall be concerned with the question of quantitative error control for hyperbolic partial differential equations with the aim to achieve reliability, either in the sense that the numerical solution approximates the analytical solution in a given norm to within a given tolerance, or in the sense that physically relevant derived quantities, which can be thought of as functionals of the solution, are approximated to within a given tolerance. Reliability in the latter sense is particularly important in engineering applications; e.g. in fluid dynamics one may be concerned with calculating the lift and the drag coefficients of a body immersed into a viscous fluid whose flow is governed by the Navier-Stokes equations. The lift and drag coefficients are defined as integrals, over the boundary of the body, of the stress tensor components normal and tangential to the flow, respectively. Similarly, in elasticity theory, the quantities of prime interest, such as the stress intensity factor or the moments of a shell or plate, are derived quantities. To achieve reliability in one sense or the other, we shall derive computable *a posteriori* error bounds in terms of the finite element residual which is obtained by inserting the computed solution into the partial differential equation under consideration. Such bounds represent the key ingredients of reliable adaptive finite element algorithms for hyperbolic problems: error control to within a given tolerance is achieved through a feed-back process where the *a posteriori* error bound plays the rôle of a stopping criterion.

The aim of the present paper is to discuss the construction and the practical implementation of *a posteriori* error bounds for finite element approximations of first-order hyperbolic equations. The derivation of the bounds rests on hyperbolic duality arguments; these will be exploited in a systematic manner throughout. In fact, following the paradigm of *a posteriori* error estimation outlined in [29] by Johnson, our analysis has two basic ingredients: the application of Galerkin orthogonality and the use of the strong stability of the dual (adjoint) problem; the rôle of these concepts in the error analysis will be highlighted below.

Over the last decade the *a posteriori* error analysis of finite element methods for partial differential equations has been an area of active research (see Ainsworth and Oden [2] Szabó and Babuška [59], and Verfürth [60]). Unfortunately, much of the interest has focussed on elliptic and parabolic equations and relatively little progress has been made on the *a posteriori* error analysis of finite element and finite volume approximations

to hyperbolic and nearly-hyperbolic problems. For an overview of current activities in the latter area we refer to the articles [29] and [15]; see also, [28], [30], and references therein. The approach to *a posteriori* error estimation for hyperbolic problems pursued in those papers, particularly in the work of Johnson and Szepessy [31], and reviewed at the beginning of Section 4, rests on performing an elliptic or parabolic regularisation of the hyperbolic problem and exploiting the smoothing properties of the resulting adjoint problem in conjunction with Galerkin orthogonality to derive an *a posteriori* error bound in the L_2 norm; in Section 4 we present an alternative, more direct, process which avoids the need for regularisation of the hyperbolic operator, at the price of arriving at error bounds in weaker (negative Sobolev) norms. In the course of our analysis we shall require some basic results from the theory of function spaces; these are summarised in the next section, followed, in Section 3, by a brief overview of the theory of well-posedness of steady linear hyperbolic equations. Section 5 is devoted to the problem of error generation, error propagation and local error estimation in the context of *a posteriori* error analysis. We have already noted that *a posteriori* error estimation of linear functionals is of great practical importance; this topic is the subject of Section 6. Section 7 concerns the *a posteriori* error analysis of finite element approximations to unsteady hyperbolic problems; we shall also comment on the implementation of our error bounds into an adaptive algorithm. The final section discusses the theory for scalar nonlinear hyperbolic conservation laws; here we rely on the work of Tadmor [57] concerning the strong stability of the linearised dual problem (a backward linear transport equation with discontinuous coefficients) associated with the conservation law, in Lipschitz spaces. Thus we arrive at an *a posteriori* error bound in the dual Lipschitz (Lip') norm.

2 Basic function spaces

In this section, we recall the definitions of some familiar function spaces, including those of continuously differentiable and Lebesgue integrable functions, and Sobolev spaces. For proofs and further details, we refer the reader to the monographs [1], [34] and [46].

2.1 Spaces of continuous functions

Let $\mathbb{N}$ denote the set of non-negative integers. An n-tuple $\alpha = (\alpha_1, \ldots, \alpha_n)$ in $\mathbb{N}^n$ is called a *multi-index*. The non-negative integer $|\alpha| = |\alpha_1| + \ldots + |\alpha_n|$ is called the length of α. We define $\partial^\alpha = \partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n}$ where $\partial_j = \partial/\partial x_j$ for $j = 1, \ldots, n$.

Let Ω be an open set in $\mathbb{R}^n$. For $k \in \mathbb{N}$, we denote by $C^k(\Omega)$ the set of all continuous real-valued functions u, defined on Ω, such that $\partial^\alpha u$ is continuous on Ω for every multi-index α, $|\alpha| \leq k$. Further, we define $C^\infty(\Omega)$ as the intersection $\bigcap_{k \geq 0} C^k(\Omega)$. The notation $C^0(\Omega)$ is abbreviated to $C(\Omega)$.

For $k \in \mathbb{N}$, we denote by $C^k(\bar{\Omega})$ the set of all $u \in C^k(\Omega)$ such that $\partial^\alpha u$ can be continuously extended from Ω onto $\bar{\Omega}$, for every multi-index α, $|\alpha| \leq k$. Further, we define $C^\infty(\bar{\Omega})$ as the intersection $\bigcap_{k \geq 0} C^k(\bar{\Omega})$, and we write $C(\bar{\Omega})$ in place of $C^0(\bar{\Omega})$.

Assuming that Ω is a bounded open set in $\mathbb{R}^n$ and $k \in \mathbb{N}$, the linear space $C^k(\bar{\Omega})$ is a Banach space equipped with the norm

$$\|u\|_{C^k(\bar{\Omega})} = \max_{|\alpha| \le k} \sup_{x \in \Omega} |\partial^\alpha u(x)|.$$

For $k \in \mathbb{N}$ and $0 < \lambda \le 1$, we denote by $C^{k,\lambda}(\bar{\Omega})$ the set of all $u \in C^k(\bar{\Omega})$ such that the quantity

$$|u|_{C^{k,\lambda}(\bar{\Omega})} = \max_{|\alpha|=k} \sup_{x \ne y,\ x,y \in \Omega} \frac{|\partial^\alpha u(x) - \partial^\alpha u(y)|}{|x-y|^\lambda}$$

is finite. $C^{k,\lambda}(\bar{\Omega})$ is a Banach space with the norm

$$\|u\|_{C^{k,\lambda}(\bar{\Omega})} = \|u\|_{C^k(\bar{\Omega})} + |u|_{C^{k,\lambda}(\bar{\Omega})}.$$

When u belongs to $C^{0,1}(\bar{\Omega})$ it is said to be *Lipschitz continuous* on $\bar{\Omega}$.

The *support*, $\operatorname{supp} u$, of a continuous function u defined on an open set Ω is the closure in Ω of the set $\{x \in \Omega \ : \ u(x) \ne 0\}$; in other words, $\operatorname{supp} u$ is the smallest closed subset of Ω such that $u = 0$ on $\Omega \setminus \operatorname{supp} u$. For $k = 0, 1, \ldots, \infty$, $C_0^k(\Omega)$ denotes the set of all $u \in C^k(\Omega)$ whose support is a compact subset of Ω.

2.2 Spaces of integrable functions

For $p \ge 1$ and an open set $\Omega \subset \mathbb{R}^n$, let $L_p(\Omega)$ denote the set of all real-valued Lebesgue measurable functions u defined on Ω such that $|u|^p$ is integrable on Ω with respect to the Lebesgue measure $\mathrm{d}x = \mathrm{d}x_1 \ldots \mathrm{d}x_n$; we assume here that any two functions which are equal almost everywhere (i.e. equal, except maybe on a set of measure zero) are identified. $L_p(\Omega)$ is a Banach space with norm

$$\|u\|_{L_p(\Omega)} = \left(\int_\Omega |u(x)|^p \, \mathrm{d}x \right)^{1/p}.$$

In particular, for $p = 2$, $L_2(\Omega)$ is a Hilbert space with the inner product

$$(u, v) = \int_\Omega u(x)\, v(x) \, \mathrm{d}x.$$

$L_\infty(\Omega)$ denotes the set of all real-valued Lebesgue measurable functions u defined on Ω such that $|u|$ has finite essential supremum; the essential supremum of $|u|$ is defined as the infimum of the set of all positive real numbers M such that $|u| \le M$ almost everywhere on Ω. Again, any two functions which are equal almost everywhere on Ω are identified. $L_\infty(\Omega)$ is a Banach space with norm

$$\|u\|_{L_\infty(\Omega)} = \mathrm{ess.sup}_{x \in \Omega} |u(x)|.$$

Hölder's Inequality: Let $u \in L_p(\Omega)$ and $v \in L_q(\Omega)$, where $1/p + 1/q = 1$, $1 \le p, q \le \infty$; then $uv \in L_1(\Omega)$ and

$$\left| \int_\Omega u(x)\, v(x) \, \mathrm{d}x \right| \le \|u\|_{L_p(\Omega)} \|v\|_{L_q(\Omega)}.$$

For $p = q = 2$, this is referred to as the *Cauchy-Schwarz Inequality.*

2.3 Sobolev spaces

Suppose that Ω is an open set in $\mathbb{R}^n$. For a non-negative integer k and $1 \leq p \leq \infty$, we define

$$W_p^k(\Omega) = \{u \in L_p(\Omega) \,:\, \partial^\alpha u \in L_p(\Omega), \quad |\alpha| \leq k\}.$$

We equip $W_p^k(\Omega)$ with the Sobolev norm defined by

$$\|u\|_{W_p^k(\Omega)} = \left(\sum_{|\alpha| \leq k} \|\partial^\alpha u\|_{L_p(\Omega)}^p \right)^{1/p}$$

when $1 \leq p < \infty$, and by

$$\|u\|_{W_\infty^k(\Omega)} = \max_{|\alpha| \leq k} \|\partial^\alpha u\|_{L_\infty(\Omega)}$$

when $p = \infty$. The associated Sobolev seminorm is defined by

$$|u|_{W_p^k(\Omega)} = \left(\sum_{|\alpha| = k} \|\partial^\alpha u\|_{L_p(\Omega)}^p \right)^{1/p}$$

when $1 \leq p < \infty$, and

$$|u|_{W_\infty^k(\Omega)} = \max_{|\alpha| = k} \|\partial^\alpha u\|_{L_\infty(\Omega)}$$

when $p = \infty$. In these definitions the derivatives are to be understood in the sense of distributions.

The Sobolev space $W_p^k(\Omega)$ can be shown to be a Banach space with the norm $\|\cdot\|_{W_p^k(\Omega)}$, $1 \leq p \leq \infty$, $k \geq 0$. A particularly important case occurs when $p = 2$; the normed linear space $W_2^k(\Omega)$ is a Hilbert space with the inner product

$$(u, v)_{W_2^k(\Omega)} = \sum_{|\alpha| \leq k} (\partial^\alpha u, \partial^\alpha v),$$

where $(\cdot, \cdot)$ is the inner product in $L_2(\Omega)$.

In order to capture finer smoothness properties of integrable functions, we consider fractional-order Sobolev spaces defined in the following way: given that s is a positive real number, $s \notin \mathbb{N}$, let us write $s = m + \sigma$, where $0 < \sigma < 1$ and $m = [s]$ is the integer part of s. The fractional-order Sobolev space $W_p^s(\Omega)$, $1 \leq p < \infty$, is the set of all $u \in W_p^m(\Omega)$ such that

$$|u|_{W_p^s(\Omega)} = \left\{ \sum_{|\alpha| = m} \int_\Omega \int_\Omega \frac{|D^\alpha u(x) - D^\alpha u(y)|^p}{|x - y|^{n + \sigma p}} \, \mathrm{d}x \, \mathrm{d}y \right\}^{1/p} < \infty,$$

with the usual modification when $p = \infty$. When equipped with the norm

$$\|u\|_{W_p^s(\Omega)} = \left\{ \|u\|_{W_p^m(\Omega)}^p + |u|_{W_p^s(\Omega)}^p \right\}^{1/p}, \qquad \text{if } 1 \leq p < \infty,$$

or the norm

$$\|u\|_{W^s_\infty(\Omega)} = \|u\|_{W^m_\infty(\Omega)} + |u|_{W^s_\infty(\Omega)}, \qquad \text{if } p = \infty,$$

the Sobolev space $W^s_p(\Omega)$ is a Banach space.

When a boundary-value problem is considered for a partial differential equation on an open set Ω, it is convenient to incorporate the boundary condition on $\partial\Omega$, the boundary of Ω, into the definition of the function space in which a solution is sought. First, we characterise the smoothness of $\partial\Omega$.

Definition 1 *Suppose that Ω is an open set in $\mathbb{R}^n$. The boundary $\partial\Omega$ of Ω is said to be Lipschitz continuous if, for every $x \in \partial\Omega$, there is an open set $\mathcal{O} \subset \mathbb{R}^n$ with $x \in \mathcal{O}$ and a local orthogonal coordinate system with coordinate $\zeta = (\zeta_1, \ldots, \zeta_n) \equiv (\zeta', \zeta_n)$ and $a \in \mathbb{R}^n$, such that*

$$\mathcal{O} = \{\zeta \,:\, -a_j < \zeta_j < a_j, \quad 1 \le j \le n\},$$

and there is a Lipschitz continuous function φ defined on

$$\mathcal{O}' = \{\zeta' \in \mathbb{R}^{n-1} \,:\, -a_j < \zeta_j < a_j, \quad 1 \le j \le n-1\},$$

with

$$|\varphi(\zeta')| \le a_n/2, \qquad \zeta' \in \mathcal{O}',$$

$$\Omega \cap \mathcal{O} = \{\zeta : \zeta_n < \varphi(\zeta'),\ \zeta' \in \mathcal{O}'\} \text{ and } \partial\Omega \cap \mathcal{O} = \{\zeta : \zeta_n = \varphi(\zeta'),\ \zeta' \in \mathcal{O}'\}.$$

A bounded open set with Lipschitz continuous boundary is called a Lipschitz domain.

An important property of a Lipschitz domain Ω is that the unit outward normal to $\partial\Omega$ is defined almost everywhere with respect to the $(n-1)$-dimensional measure on $\partial\Omega$. A simple example of a Lipschitz domain is a bounded polyhedron in $\mathbb{R}^n$, $n \ge 2$.

Proposition 1 *Suppose that Ω is a Lipschitz domain contained in $\mathbb{R}^n$ and let $1 \le p < \infty$. Then $C^\infty(\bar\Omega)$ is dense in $W^s_p(\Omega)$ for $s \ge 0$.*

We note that while $C^\infty(\bar\Omega)$ is dense in $W^s_p(\Omega)$ for $s \ge 0$ and $1 \le p < \infty$, $C^\infty_0(\Omega)$ is not dense in $W^s_p(\Omega)$ for $s > 1/p$ (although it is dense in $W^s_p(\Omega)$ for $0 \le s < 1/p$, $1 \le p < \infty$).

We conclude this section with a brief discussion about Sobolev spaces on the boundary $\partial\Omega$ of a Lipschitz domain Ω. Let us begin by recalling from Definition 1 that for every x on $\partial\Omega$ there exists a Lipschitz continuous function $\varphi : \mathcal{O}' \subset \mathbb{R}^{n-1} \to \mathbb{R}$ such that, using the notation introduced in Definition 1,

$$\partial\Omega \cap \mathcal{O} = \{\zeta = (\zeta', \varphi(\zeta')) \,:\, \zeta' \in \mathcal{O}'\},$$

so that, locally, $\partial\Omega$ is an $(n-1)$-dimensional hypersurface in $\mathbb{R}^n$. We define the mapping ϕ by

$$\phi(\zeta') = (\zeta', \varphi(\zeta')).$$

Clearly ϕ^{-1} exists and it is Lipschitz continuous on $\phi(\mathcal{O}')$; this leads us to the following definition.

Definition 2 *Let Ω be a Lipschitz domain in $\mathbb{R}^n$. For $0 \le s \le 1$ and $1 \le p < \infty$ we denote by $W_p^s(\partial\Omega)$ the set of all $u \in L_p(\partial\Omega)$ such that the composition $u \circ \phi$ belongs to $W_p^s(\mathcal{O}' \cap \phi^{-1}(\partial\Omega \cap \mathcal{O}))$ for all possible $\mathcal{O}'$ and φ satisfying the conditions of Definition 1, where $\phi(\zeta') = (\zeta', \varphi(\zeta'))$ for $\zeta' \in \mathcal{O}'$.*

In order to equip $W_p^s(\partial\Omega)$ with a norm, we consider any *atlas* $(\mathcal{O}_j, \varphi_j)_{j=1}^J$ for $\partial\Omega$ such that $\mathcal{O}_j$ and φ_j, $j = 1, \dots, J$, satisfy the conditions of Definition 1. We define $\|\cdot\|_{W_p^s(\partial\Omega)}$ by

$$\|u\|_{W_p^s(\partial\Omega)} = \left(\sum_{j=1}^{J} \|u \circ \phi_j\|^p_{W_p^s(\mathcal{O}'_j \cap \phi_j^{-1}(\partial\Omega \cap \mathcal{O}_j))} \right)^{1/p}$$

where $\phi_j(\zeta') = (\zeta', \varphi_j(\zeta'))$ for $\zeta' \in \mathcal{O}'_j$, $j = 1, \dots, J$.

In fact, for $0 < s < 1$ it can be shown that this is equivalent to the following norm

$$\|u\|_{W_p^s(\partial\Omega)} = \left(\int_{\partial\Omega} |u|^p \, \mathrm{d}\sigma + \int_{\partial\Omega} \int_{\partial\Omega} \frac{|u(x) - u(y)|^p}{|x - y|^{n-1+sp}} \, \mathrm{d}\sigma(x) \, \mathrm{d}\sigma(y) \right)^{1/p},$$

where $\mathrm{d}\sigma$ denotes the $(n-1)$-dimensional surface measure on $\partial\Omega$.

Finally, we recall the notion of *trace* of a function on the boundary $\partial\Omega$ of a Lipschitz domain $\Omega \subset \mathbb{R}^n$. If ψ belongs to $C^\infty(\bar{\Omega})$ then we put

$$\gamma_0(\psi) = \psi|_{\partial\Omega} \, . \tag{2.1}$$

The trace of a function u in $W_p^s(\Omega)$ is then defined by extending the operator γ_0 from the dense subspace $C^\infty(\bar{\Omega})$ to the whole of $W_p^s(\Omega)$.

Proposition 2 *Suppose that Ω is a Lipschitz domain in $\mathbb{R}^n$, and let $1 < p < \infty$. Assuming that $1/p < s \le 1$, the mapping γ_0 defined on $C^\infty(\bar{\Omega})$ by (2.1) has a unique continuous extension to a linear operator, still denoted γ_0, from $W_p^s(\Omega)$ onto $W_p^{s-(1/p)}(\partial\Omega)$.*

We adopt the following notational convention: when $p = 2$ we shall write H^s in place of W_2^s to signify the fact that we are dealing with a Hilbert space. We define $H_0^1(\Omega)$ as the closure of $C_0^\infty(\Omega)$ in the norm of the Sobolev space $H^1(\Omega)$; when Ω is a Lipschitz domain, it can be shown that

$$H_0^1(\Omega) = \{u \in H^1(\Omega) \, : \, \gamma_0(u) = 0\}.$$

3 Steady hyperbolic problems

In this section we present a brief review of the theory of steady hyperbolic equations. In the first part of the section we focus on scalar hyperbolic equations, while the second part is concerned with symmetric hyperbolic systems.

3.1 Scalar hyperbolic equations

We consider the question of well-posedness of the hyperbolic boundary-value problem $\mathcal{P}(f)$:

$$\begin{aligned} \operatorname{div}(\mathbf{a}u) + c\,u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial_-\Omega, \end{aligned}$$

where Ω is a Lipschitz domain in $\mathbb{R}^n$, with *inflow boundary*

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

here $\nu(x)$ denotes the unit outward normal vector at $x \in \partial\Omega$ (whenever it is defined). The complement of $\partial_-\Omega$ with respect to $\partial\Omega$ will be denoted $\partial_+\Omega$ and will be referred to as the *outflow boundary*. For the sake of simplicity we shall suppose that $\Omega = (0,1)^n$, that $\mathbf{a} = (a_1, \ldots, a_n)$ is a real-valued continuously differentiable n-component vector function defined on $\bar{\Omega}$, c is a continuous real-valued function on $\bar{\Omega}$ and f is a real-valued square-integrable function on Ω.

In order to set up the variational formulation of $\mathcal{P}(f)$, with $\mathbf{a}$ and c we associate the function space

$$H_-(\Omega) = \{v \in L_2(\Omega) \,:\, \operatorname{div}(\mathbf{a}v) + c\,v \in L_2(\Omega), \quad \gamma_\nu(\mathbf{a}v) = 0 \quad \text{on } \partial_-\Omega\}$$

in which the solution to the problem is sought. We note that the boundary condition is included into the definition of the space $H_-(\Omega)$; here $\gamma_\nu(\mathbf{a}v) = (\mathbf{a}v)\cdot\nu|_{\partial_-\Omega}$ signifies the normal trace of the vector field $\mathbf{a}v$ on $\partial_-\Omega$.

At this stage, the boundary condition should be understood formally: below we shall justify that the definition of $H_-(\Omega)$ is meaningful. To do so, we first recall from [20] (Chapter I, Theorem 2.5 and Corollary 2.8) that the normal trace operator, $\gamma_\nu(\cdot)$, is a continuous surjection of

$$H(\operatorname{div},\Omega) = \{\mathbf{v} \in [L_2(\Omega)]^n \,:\, \operatorname{div}\mathbf{v} \in L_2(\Omega)\}$$

onto $H^{-1/2}(\partial\Omega)$, the latter being the dual space of the fractional-order Sobolev space $H^{1/2}(\partial\Omega) = W_2^{1/2}(\partial\Omega)$. Now, suppose that Γ is a connected relatively open subset of $\partial\Omega$ of positive $(n-1)$-dimensional measure (for our purposes, $\Gamma = \partial_-\Omega$). We denote by $H_0^1(\Gamma)$ the closure of $C_0^\infty(\Gamma)$ in the norm of the Sobolev space $H^1(\Gamma)$. Further, we define $H_{00}^{1/2}(\Gamma)$, using the K-method of function space interpolation (see Bergh and Löfström [8], for example), as the interpolation space 'halfway' between $L_2(\Gamma)$ and $H_0^1(\Gamma)$. Finally, we let $(H_{00}^{1/2}(\Gamma))'$ denote the dual space of $H_{00}^{1/2}(\Gamma)$. Since the trivial extension $\mathcal{E}_0$ is a continuous linear operator from $L_2(\Gamma)$ into $L_2(\partial\Omega)$ and from $H_0^1(\Gamma)$ into $H^1(\partial\Omega)$, we deduce by function space interpolation that it is also a continuous linear operator from $H_{00}^{1/2}(\Gamma)$ into $H^{1/2}(\partial\Omega)$. Thus, by applying the Transposition Theorem (see Theorem 4.1 in Baiocchi and Capelo [4]), we conclude that the transpose of the linear operator $\mathcal{E}_0 : H_{00}^{1/2}(\Gamma) \to H^{1/2}(\partial\Omega)$ is a continuous linear operator ${}^t\mathcal{E}_0$ from $(H^{1/2}(\partial\Omega))' = H^{-1/2}(\partial\Omega)$ into $(H_{00}^{1/2}(\Gamma))'$; ${}^t\mathcal{E}_0$ is called the *restriction* from $\partial\Omega$ to Γ.

Suppose that $v \in L_2(\Omega)$ and $\mathrm{div}(\mathbf{a}v) + c\,v \in L_2(\Omega)$. Then $\mathbf{a}v \in H(\mathrm{div};\Omega)$, and it follows that $\gamma_\nu(\mathbf{a}v) \in H^{-1/2}(\partial\Omega)$. Hence the restriction of $\gamma_\nu(\mathbf{a}v)$ to $\partial_-\Omega$ belongs to $(H_{00}^{1/2}(\partial_-\Omega))'$. The definition of $H_-(\Omega)$ is, therefore, meaningful; in fact, $H_-(\Omega)$ is a Hilbert space with the norm

$$\|v\|_{H_-(\Omega)} = (\|v\|^2_{L_2(\Omega)} + \|\mathcal{L}v\|^2_{L_2(\Omega)})^{1/2},$$

where $\mathcal{L} : H_-(\Omega) \to L_2(\Omega)$ denotes the linear operator defined by $\mathcal{L}v = \mathrm{div}(\mathbf{a}v) + c\,v$, $v \in H_-(\Omega)$.

With this notation, the boundary-value problem $\mathcal{P}(f)$, for $f \in L_2(\Omega)$, can be expressed as follows: find u in $H_-(\Omega)$ such that $\mathcal{L}u = f$. Alternatively, the variational formulation of $\mathcal{P}(f)$ is: find $u \in H_-(\Omega)$ satisfying

$$(\mathrm{div}(\mathbf{a}u) + c\,u, q) = (f, q) \qquad \forall q \in L_2(\Omega). \tag{3.1}$$

A solution of (3.1) can be thought of as a generalised solution of $\mathcal{P}(f)$, with the differential equation satisfied as an equality in $L_2(\Omega)$ and the boundary condition obeyed as an equality in $(H_{00}^{1/2}(\partial_-\Omega))'$.

We adopt the following additional hypothesis on $\mathbf{a}$.

Hypothesis 1 *The components $a_1, \ldots, a_n$ of the vector field $\mathbf{a}$ belong to $C^1(\bar\Omega)$ and are strictly positive functions on $\bar\Omega$.*

This assumption ensures that $\partial_-\Omega$ is a non-characteristic hypersurface of dimension $(n-1)$ for the differential operator $v \mapsto \mathrm{div}(\mathbf{a}v) + c\,v$.

Proposition 3 *Under Hypothesis 1 and given that $f \in L_2(\Omega)$ and $c \in C(\bar\Omega)$, problem (3.1) has a unique solution u in $H_-(\Omega)$. In addition, the linear operator $\mathcal{L}$ is a continuous bijection of $H_-(\Omega)$ onto $L_2(\Omega)$ with a continuous inverse $\mathcal{L}^{-1} : L_2(\Omega) \to H_-(\Omega)$.*

Proof The proof is based on Banach's Closed Range Theorem. The non-trivial step is to verify (3.3) below; to do so, let $C^1_-(\bar\Omega)$ denote the set of all functions in $C^1(\bar\Omega)$ which vanish on $\partial_-\Omega$. It is easily seen that, for an n-component real vector ξ,

$$\begin{aligned}(\mathrm{div}(\mathbf{a}v) + c\,v, \mathrm{e}^{-2\xi\cdot x}v) \;&=\; \left(c + \frac{1}{2}(\mathrm{div}\,\mathbf{a}) + \mathbf{a}\cdot\xi, |\mathrm{e}^{-\xi\cdot x}v|^2\right)\\ &\quad + \frac{1}{2}\int_{\partial_+\Omega} (\mathbf{a}\cdot\nu)|\mathrm{e}^{-\xi\cdot x}v|^2\,\mathrm{d}s \quad \forall v \in C^1_-(\overline{\Omega}).\end{aligned} \tag{3.2}$$

Let ξ be such that the constant

$$M_0 = \inf_\Omega \left(c + \frac{1}{2}(\mathrm{div}\,\mathbf{a}) + \mathbf{a}\cdot\xi\right)$$

is positive, and let M_1 and M_2 be two positive real numbers such that

$$M_1 \le \exp(-2\xi\cdot x) \le M_2 \qquad \forall x \in \bar\Omega.$$

Omitting the second term on the right-hand side of (3.2) and noting that $C^1_-(\bar{\Omega})$ is dense in $H_-(\Omega)$, it follows that

$$(\mathcal{L}v, \mathrm{e}^{-2\xi\cdot x}v) \geq M_0 \|\mathrm{e}^{-\xi\cdot x}v\|^2_{L_2(\Omega)} \qquad \forall v \in H_-(\Omega),$$

and hence

$$\left(1 + \left(\frac{M_2}{M_1 M_0}\right)^2\right)^{1/2} \|\mathcal{L}v\|_{L_2(\Omega)} \geq \|v\|_{H_-(\Omega)} \qquad \forall v \in H_-(\Omega). \tag{3.3}$$

Inequality (3.3) implies that $\mathcal{L}$ is an injective operator from $H_-(\Omega)$ onto its range space $R(\mathcal{L})$, and that the inverse of $\mathcal{L}$ is continuous. Hence $\mathcal{L}$ is an isomorphism from $H_-(\Omega)$ onto $R(\mathcal{L})$; therefore $R(\mathcal{L})$ is a closed subspace of $L_2(\Omega)$. Exploiting the positivity assumption on the components of $\mathbf{a}$, it is easy to prove (using the method of characteristics, for example) that the transpose ${}^t\mathcal{L}$ of $\mathcal{L}$ has trivial kernel, i.e. $\mathrm{Ker}({}^t\mathcal{L}) = \{0\}$; the Closed Range Theorem then implies that $R(\mathcal{L}) = L_2(\Omega)$. Hence $\mathcal{L}$ is an isomorphism from $H_-(\Omega)$ onto $L_2(\Omega)$. In addition, (3.3) implies that

$$1 \leq \|\mathcal{L}^{-1}\|_{L_2 \to H_-} \leq \left(1 + \left(\frac{M_2}{M_1 M_0}\right)^2\right)^{1/2}.$$

That completes the proof. ■

We note that this existence and uniqueness result can be extended in several directions:

a) First, a theorem analogous to Proposition 3 holds for the adjoint boundary-value problem:

$$-\mathbf{a}\cdot\nabla u + c\,u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial_+\Omega,$$

where $\mathbf{a}$, c, f and Ω are as in Proposition 3.

b) Second, a result analogous to Proposition 3 can be developed more generally, in L_p norms, $1 \leq p < \infty$. More precisely, suppose that in the definition of the solution space $H_-(\Omega)$, $L_2(\Omega)$ is replaced by $L_p(\Omega)$; then Proposition 3 holds with $L_2(\Omega)$ replaced by $L_p(\Omega)$ throughout.

c) Third, Hypothesis 1 can be relaxed by supposing that Ω is a Lipschitz domain in $\mathbb{R}^n$ such that $\partial\Omega$ is a non-characteristic hypersuface for $\mathcal{L}$, and that there exists a constant vector $\xi \in \mathbb{R}^n$ such that

$$M_0 = \inf_{\Omega} \left(c + \frac{1}{2}(\mathrm{div}\ \mathbf{a}) + \mathbf{a}\cdot\xi\right) > 0.$$

d) Finally, under Hypothesis 1 the normal trace, $\gamma_\nu(\mathbf{a}v) \in H^{-1/2}(\partial_-\Omega)$ can be shown to belong to $L_2(\partial_-\Omega)$.

We conclude this subsection with the following regularity result which is a special case of the general Differentiability Theorem stated on p.272 of the work of Rauch [49][1].

[1] There is a typographic error in line one of the theorem: u should be replaced by f.

Proposition 4 *Under Hypotheses 1 and assuming that* $\mathbf{a} \in [C^2(\bar{\Omega})]^n$, $c \in C^1(\bar{\Omega})$ *and* $f \in H_0^1(\Omega)$, *the unique weak solution* u *to problem* $\mathcal{P}$ *belongs to* $H^1(\Omega) \cap H_-(\Omega)$. *Moreover,*

$$\|u\|_{H^1(\Omega)} \leq C_1 \|f\|_{H^1(\Omega)},$$

where C_1 *is a positive constant independent of* f.

An identical result holds for the adjoint boundary value problem stated in a) above. The next section extends the well-posedness results discussed here to symmetric positive hyperbolic systems on Lipschitz domains in $\mathbb{R}^n$.

3.2 Symmetric hyperbolic systems

The aim of this section is to give a brief account of the theory of well-posedness for a class of first-order systems of partial differential equations, usually referred to as Friedrichs systems or symmetric positive systems. These represent a natural generalisation of the scalar hyperbolic equation whose properties were discussed in the previous section. In fact, symmetric positive systems embrace a large class of partial differential equations irrespective of their type, allowing a unified treatment of certain elliptic and hyperbolic equations by means of a common tool, *energy analysis*. Historically, the main motivation for introducing this class of equations "was not the desire for a unified treatment of elliptic and hyperbolic equations, but the desire to handle equations which are partly elliptic, partly hyperbolic, such as the Tricomi equation" (see [16]):

$$\left(y \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} \right) u = 0,$$

which plays a basic rôle in the theory of transonic flow. Our presentation of the theory will closely follow the functional-analytic approach adopted in the work of Mackenzie, Süli and Warnecke [40].

Suppose that Ω is a Lipschitz domain in $\mathbb{R}^n$. Let A_i, $i = 1, ..., n$, and C be (matrix-valued) mappings from $\bar{\Omega}$ 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 $\bar{\Omega}$ and the components of C are continuous on $\bar{\Omega}$. We consider the linear first-order system of partial differential equations

$$\mathcal{L}\mathbf{u} \equiv \sum_{i=1}^{n} \frac{\partial}{\partial x_i}(A_i \mathbf{u}) + C\mathbf{u} = \mathbf{f} \qquad \text{in } \Omega. \tag{3.4}$$

The system (3.4) is called *symmetric positive* if the following conditions hold:

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

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

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

is positive definite, uniformly on $\bar{\Omega}$; namely, there exists a positive constant $C_0 = C_0(\Omega)$ such that

$$\frac{1}{2}(K_\xi(x) + K_\xi^*(x)) \geq C_0 I \tag{3.5}$$

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

The positivity hypothesis (b) can be seen as a direct generalisation of the positivity condition imposed on the constant M_0 in the proof of Proposition 3; see also remark c) at the end of the previous section.

In practice a system of partial differential equations, such as (3.4), is rarely considered in isolation and will be accompanied with a boundary condition. In order to motivate the imposition of the boundary condition for symmetric positive systems, we note that for the scalar hyperbolic equation discussed in the previous section a well-posed boundary-value problem is arrived at by prescribing a boundary condition on the inflow part of $\partial\Omega$ only. By formal analogy, for the system, only the 'incoming components' of the solution vector should be imposed on the boundary, in accordance with the physical notion of causality. This will be made more precise below. To begin, let us consider the matrix

$$B = \nu_1 A_1 + \ldots + \nu_n A_n,$$

where $\nu = (\nu_1, \ldots, \nu_n)$ is the unit outward normal vector field on $\partial\Omega$. In order to simplify the exposition, we shall suppose that B is non-singular almost everywhere on $\partial\Omega$ (with respect to the $(n-1)$-dimensional measure on $\partial\Omega$); this is equivalent to requiring that the boundary of Ω is almost everywhere non-characteristic for $\mathcal{L}$. Since the matrix 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.

Let us suppose that $\mathbf{g}$ is a (sufficiently smooth) real-valued vector function defined on $\bar{\Omega}$; then an 'admissible' boundary condition for (3.4) is of the form

$$B^-(\mathbf{u} - \mathbf{g})\big|_{\partial\Omega} = \mathbf{0}. \tag{3.6}$$

We shall also consider the homogeneous counterpart of this boundary condition, corresponding to $\mathbf{g} = \mathbf{0}$; namely,

$$B^-\mathbf{u}\big|_{\partial\Omega} = \mathbf{0}. \tag{3.7}$$

For the time being, these boundary conditions are to be understood formally. Below, following a similar approach as in the scalar case described in the previous section, we shall state a trace theorem which assigns a precise meaning to $B^-\mathbf{u}|_{\partial\Omega}$. First, however, we single out the class of functions for which such a trace will be considered. For this purpose, we introduce the *graph space* of the operator $\mathcal{L}$ as the linear space

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

It is a simple matter to verify that, when equipped with the norm

$$|||\mathbf{v}|||_{\Omega,\xi} = \left(\|\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v}\|^2_{[L_2(\Omega)]^m} + \|\mathrm{e}^{-\alpha(\xi\cdot x)}\mathcal{L}\mathbf{v}\|^2_{[L_2(\Omega)]^m}\right)^{\frac{1}{2}},$$

the graph space $H(\mathcal{L},\Omega)$ is a Banach space (in fact, it is a Hilbert space with the natural inner product associated with this norm). Many of the arguments that we shall use throughout this paper rely on the concept of duality, and we shall also require $\mathcal{L}^*$, the formal adjoint of $\mathcal{L}$, defined by

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

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

Let $\gamma_0 : [H^1(\Omega)]^m \to [H^{1/2}(\partial\Omega)]^m$ signify the usual trace operator, defined in Section 2.3, which to each element of $[H^1(\Omega)]^m$ assigns its restriction to $\partial\Omega$. We denote by $[H^{-1/2}(\partial\Omega)]^m$ the dual space of $[H^{1/2}(\partial\Omega)]^m$; the duality pairing between these two spaces will be labelled $\langle\cdot,\cdot\rangle$. The next proposition, stated and proved in [41], will play an important rôle in the rest of the paper.

Proposition 5 *Assuming that Ω is a Lipschitz domain in $\mathbb{R}^n$, the mapping $\gamma_B : \mathbf{v} \mapsto B\gamma_0(\mathbf{v})$ defined on $[H^1(\Omega)]^m$ can be extended by continuity to a linear and continuous mapping, still denoted γ_B (and referred to as the* conormal trace operator*), from $H(\mathcal{L},\Omega)$ into $[H^{-1/2}(\partial\Omega)]^m$. Moreover, for any $\mathbf{u} \in H(\mathcal{L},\Omega)$ and $\mathbf{v} \in [H^1(\Omega)]^m$ the following Green's formula holds*

$$(\mathcal{L}\mathbf{u},\mathbf{v}) - (\mathbf{u},\mathcal{L}^*\mathbf{v}) = \langle \gamma_B(\mathbf{u}), \gamma_0(\mathbf{v})\rangle.$$

An analogous result holds for $H(\mathcal{L}^*,\Omega)$.

The proof of this result is identical to that of Theorem 18.6 in the monograph of Baiocchi and Capelo [4]. The interested reader may also wish to consult [41] for a detailed proof.

Proposition 5 assigns a precise meaning to the conormal trace operator, by extending $B\gamma_0(\cdot)$ from $[H^1(\Omega)]^m$ to the graph space $H(\mathcal{L},\Omega)$. However the 'admissible' boundary condition (3.6) involves B^- rather than B, so we need to introduce a trace operator based on B^-. To do so, we note that the splitting $B = B^+ + B^-$ induces a natural decomposition of γ_B which leads to the definition of the partial conormal trace operators $\gamma_{B^\pm}$. This can be seen by defining

$$\gamma_{B^\pm}(\mathbf{u}) = B^\pm\gamma_0(\mathbf{u}) \qquad \forall \mathbf{u} \in [H^1(\Omega)]^m,$$

and extending this definition from the dense subspace $[H^1(\Omega)]^m$ to $H(\mathcal{L},\Omega)$ to arrive at continuous linear operators

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

with $\gamma_B = \gamma_{B^+} + \gamma_{B^-}$. One can proceed in the same way for $H(\mathcal{L}^*,\Omega)$.

Equipped with these definitions, in the case of the homogeneous boundary-value problem (3.4), (3.7) we can define the domains of the operators $\mathcal{L}$ and $\mathcal{L}^*$ as, respectively,

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

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

When supplied with the associated graph-norms $|||\cdot|||_{\xi,\Omega}$, $|||\cdot|||_{*,\xi,\Omega}$ and the corresponding natural inner products, $D(\mathcal{L},\Omega)$ and $D(\mathcal{L}^*,\Omega)$ are Hilbert subspaces of $H(\mathcal{L},\Omega)$ and $H(\mathcal{L}^*,\Omega)$, respectively.

Now we are ready to define the concept of *solution*. Suppose that $\mathbf{f}\in[L_2(\Omega)]^m$; a function $\mathbf{u}\in[L_2(\Omega)]^m$ such that

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

will be referred to as *weak solution* of the homogeneous boundary value problem (3.4), (3.7). If $\mathbf{u}$ is a weak solution of (3.4), (3.7) and $\mathbf{u}$ belongs to $H(\mathcal{L},\Omega)$, we shall say that $\mathbf{u}$ is 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 $\mathcal{L}\mathbf{u}\in[L_2(\Omega)]^m$ is required for $\mathbf{u}\in[L_2(\Omega)]^m$ to be a strong solution (rather than $\mathbf{u}\in[H^1(\Omega)]^m$), discontinuities in a strong solution $\mathbf{u}$ may arise across characteristic hypersurfaces.

Proposition 6 *Let $\partial\Omega$ be a non-characteristic hypersurface for $\mathcal{L}$, and assume that $\mathbf{f}\in[L_2(\Omega)]^m$; then the homogeneous boundary-value problem (3.4), (3.7) has a unique strong solution $\mathbf{u}\in D(\mathcal{L},\Omega)$. Further, the linear operator $\mathcal{L}$ is a continuous bijection from $D(\mathcal{L},\Omega)$ onto $[L_2(\Omega)]^m$ with a continuous inverse $\mathcal{L}^{-1}:[L_2(\Omega)]^m\to D(\mathcal{L},\Omega)$. An analogous result holds for $\mathcal{L}^*$.*

Proof As in the scalar case discussed in the previous section, the proof is based on Banach's Closed Range Theorem. Let us suppose that $\mathbf{v}\in[H^1(\Omega)]^m$ and take the inner product of $\mathcal{L}\mathbf{v}$ with $\mathrm{e}^{-2\alpha(\xi\cdot x)}\mathbf{v}$; upon integrating by parts using Proposition 5 and splitting the conormal trace operator γ_B as $\gamma_{B^+}+\gamma_{B^-}$, we deduce that

$$\langle\mathrm{e}^{-\alpha(\xi\cdot x)}\gamma_{B^+}(\mathbf{v}),\mathrm{e}^{-\alpha(\xi\cdot x)}\gamma_0(\mathbf{v})\rangle+\left(\frac{1}{2}(K_\xi+K_\xi^*)\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v},\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v}\right)$$

$$=-\langle\mathrm{e}^{-\alpha(\xi\cdot x)}\gamma_{B^-}(\mathbf{v}),\mathrm{e}^{-\alpha(\xi\cdot x)}\gamma_0(\mathbf{v})\rangle+(\mathrm{e}^{-\alpha(\xi\cdot x)}\mathcal{L}\mathbf{v},\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v}).$$

Noting (3.5) of hypothesis (b), we arrive at the following Gårding inequality:

$$C_0\|\,\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{v}\|_{[L_2(\Omega)]^m}\le\|\,\mathrm{e}^{-\alpha(\xi\cdot x)}\mathcal{L}\mathbf{v}\|_{[L_2(\Omega)]^m}\qquad\forall\mathbf{v}\in[H^1(\Omega)]^m\cap D(\mathcal{L},\Omega).\tag{3.8}$$

As $[H^1(\Omega)]^m\cap D(\mathcal{L},\Omega)$ is dense in $D(\mathcal{L},\Omega)$, it follows that

$$C_0'\|\mathrm{e}^{-\alpha(\xi\cdot x)}\mathcal{L}\mathbf{v}\|_{[L_2(\Omega)]^m}\ge|||\mathbf{v}|||_{\xi,\Omega}\qquad\forall\mathbf{v}\in D(\mathcal{L},\Omega),\tag{3.9}$$

where $C_0'=(1+C_0^{-2})^{1/2}$. The rest of the proof is identical to the final part of the proof of Proposition 3, with $D(\mathcal{L},\Omega)$ and $[L_2(\Omega)]^m$ replacing $H_-(\Omega)$ and $L_2(\Omega)$, respectively. ∎

We deduce from the proof of this theorem that the strong solution to the homogeneous boundary-value problem (3.4), (3.7) obeys the *stability estimate*

$$\|\mathbf{u}\|_{[L_2(\Omega)]^m}\le\frac{1}{C_0}\mathrm{e}^{2\alpha D}\|\mathbf{f}\|_{[L_2(\Omega)]^m},$$

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

More generally, consider the non-homogeneous boundary-value problem (3.4), (3.6), where $\mathbf{f} \in [L_2(\Omega)]^m$ and $\mathbf{g} \in H(\mathcal{L}, \Omega)$. A function $\mathbf{u} \in [L_2(\Omega)]^m$ satisfying

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

is called a *weak solution* of the boundary-value problem. A weak solution $\mathbf{u}$ to the non-homogeneous problem (3.4), (3.6) which belongs to $H(\mathcal{L}, \Omega)$ is called a *strong solution*. Lax and Phillips [36] proved, under the assumption that $\partial\Omega$ is a non-characteristic hypersurface for $\mathcal{L}$, that every weak solution is a strong solution.

In the *a posteriori* error analysis discussed in the next section we shall require a regularity result, analogous to that stated in the scalar case in Proposition 4, for the following adjoint (dual) problem:

$$\begin{aligned} \mathcal{L}^*\mathbf{z} &= \mu \qquad \text{in } \Omega, \\ \gamma_{B^+}(\phi - \chi) &= \mathbf{0} \qquad \text{on } \partial\Omega, \end{aligned} \tag{3.10}$$

where $\mu \in [H_0^1(\Omega)]^m$ and $\chi|_{\partial\Omega} \in [H^1(\partial\Omega)]^m$. Instead of imposing technical assumptions on the data and describing the various instances when such a regularity result holds, for the sake of simplicity of presentation, we shall adopt the following hypothesis.

Hypothesis 2 *Suppose that (a) and (b) hold, and that for $\mu \in [H_0^1(\Omega)]^m$ and $\chi|_{\partial\Omega} \in [H^1(\partial\Omega)]^m$ (with $\partial\Omega$, the boundary of the Lipschitz domain $\Omega \subset \mathbb{R}^n$, representing a non-characteristic hypersurface for $\mathcal{L}$), the solution to (3.10) is in $[H^1(\Omega)]^m$ and satisfies the bound*

$$\|\phi\|_{[H^1(\Omega)]^m} \leq C_1' \left(\|\mu\|_{[H^1(\Omega)]^m} + \| \chi|_{\partial\Omega} \|_{[H^1(\partial\Omega)]^m} \right), \tag{3.11}$$

where C_1' is a positive constant independent of μ and χ.

We shall refer to inequality (3.11) as *strong stability of the dual problem*. General circumstances when such a bound holds are discussed in the paper of Tartakoff [58] (particularly in Theorem 3, p.1118, where $\mu \in [H^1(\Omega)]^m$, in fact, but $\partial\Omega$ is a C^∞ hypersurface); see also Theorem 2, in Section 1 of the work of Kohn and Nirenberg [32]; in the case of a symmetric positive system subject to periodic boundary conditions a bound of this kind was proved by Lax [35].

4 A posteriori error analysis for steady problems

In this section we present the basic theory of *a posteriori* error estimation for linear hyperbolic problems.

4.1 A posteriori error analysis á la Johnson

In order to motivate the approach followed in this paper, we present a brief review of the general theoretical framework of *a posteriori* error analysis, in the context of linear first-order hyperbolic systems, pursued by Johnson and his co-workers; for a detailed account, see [29] and [15].

Let us suppose that Y is a Hilbert space with inner product $(\cdot,\cdot)$ and norm $\|\cdot\|$, and let $\mathcal{L} : Y \to Y$ be a linear operator on Y with domain $D(\mathcal{L}) \subset Y$; in our case, $\mathcal{L}$ is a symmetric positive system of linear first-order hyperbolic differential operators on $Y = [L_2(\Omega)]^m$ with domain $D(\mathcal{L}) = D(\mathcal{L},\Omega)$. Given that $\mathbf{f} \in Y$, we consider the problem of finding $\mathbf{u} \in D(\mathcal{L})$ such that

$$\mathcal{L}\mathbf{u} = \mathbf{f}.$$

Next we consider a Galerkin finite element approximation to this problem. We select a sequence of finite-dimensional spaces $\{X_h\}$, parametrised by the positive discretisation parameter h; for the sake of simplicity we shall suppose that we are dealing with a conforming approximation in the sense that $X_h \subset D(\mathcal{L})$ for each h. Simultaneously, we consider a sequence of finite-dimensional spaces $\{Y_h\}$, with Y_h contained in Y for each h. For the present purposes, X_h and Y_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 Ω; X_h is called the *trial space* while Y_h is referred to as the *test space*. Let Π_h denote the orthogonal projector in Y onto Y_h. The Galerkin finite element method can then be formulated as follows: find an approximation $\mathbf{u}_h$ to $\mathbf{u}$ in X_h such that

$$\Pi_h \mathcal{L}\mathbf{u}_h = \Pi_h \mathbf{f}.$$

Equivalently, we can write this as follows: find $\mathbf{u}_h$ in X_h such that

$$(\mathcal{L}\mathbf{u}_h, \mathbf{v}_h) = (\mathbf{f}, \mathbf{v}_h) \qquad \forall \mathbf{v}_h \in Y_h.$$

In order to obtain a computable bound on the *global error* $\mathbf{e}_h = \mathbf{u} - \mathbf{u}_h$ in terms of the *finite element residual* $\mathbf{r}_h$, defined by

$$\mathbf{r}_h = \mathbf{f} - \mathcal{L}\mathbf{u}_h,$$

we note the *Galerkin orthogonality* property

$$(\mathbf{r}_h, \mathbf{v}_h) = 0 \qquad \forall \mathbf{v}_h \in Y_h$$

which will play a crucial rôle in the analysis. Further, denoting by $\mathcal{L}^*$ the adjoint of $\mathcal{L}$, we consider the following auxiliary problem, referred to as the *dual problem*: find $\mathbf{z} \in D(\mathcal{L}^*) = D(\mathcal{L}^*,\Omega)$ such that

$$\mathcal{L}^*\mathbf{z} = \mathbf{u} - \mathbf{u}_h.$$

As already indicated in the introduction, the *a posteriori* error analysis is based on a duality argument. The first step is to derive a representation of the global error in terms of the residual; this is achieved as follows:

$$\begin{aligned} \|\mathbf{u}-\mathbf{u}_h\|^2 &= (\mathbf{u}-\mathbf{u}_h, \mathbf{u}-\mathbf{u}_h) = (\mathbf{u}-\mathbf{u}_h, \mathcal{L}^*\mathbf{z}) \\ &= (\mathcal{L}(\mathbf{u}-\mathbf{u}_h), \mathbf{z}) = (\mathcal{L}\mathbf{u}-\mathcal{L}\mathbf{u}_h, \mathbf{z}) \\ &= (\mathbf{f}-\mathcal{L}\mathbf{u}_h, \mathbf{z}) = (\mathbf{r}_h, \mathbf{z}), \end{aligned}$$

where the Green's identity stated in Proposition 5 has been used in the transition from line one to line two. Exploiting the Galerkin orthogonality property, namely that $(\mathbf{r}_h, \mathbf{z}_h) = 0$ for any $\mathbf{z}_h \in Y_h$, we deduce that

$$\|\mathbf{u}-\mathbf{u}_h\|^2 = (\mathbf{r}_h, \mathbf{z}-\mathbf{z}_h) = (h^s\mathbf{r}_h, h^{-s}(\mathbf{z}-\mathbf{z}_h)),$$

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

$$\|\mathbf{u}-\mathbf{u}_h\|^2 \leq \|h^s\mathbf{r}_h\|\,\|h^{-s}(\mathbf{z}-\mathbf{z}_h)\|.$$

The first term on the right-hand side of this inequality is of the desired form, involving the (computable) residual $\mathbf{r}_h$ multiplied by an appropriate power of the discretisation parameter, while the second term incorporates $\mathbf{z}$, the solution to the dual problem. Since the dual-problem has the (unknown) global error as data, $\mathbf{z}$ is unknown and has to be eliminated from the analysis by relating it to $\mathbf{u}-\mathbf{u}_h$; moreover, an appropriate choice of $\mathbf{z}_h$ has to be made. The details of these steps are described below.

Let us suppose that $\{W_\sigma\}_{\sigma\geq 0}$ is a scale of Hilbert spaces, with corresponding norms $\|\cdot\|_\sigma$, such that $W_0 = Y$ and W_{σ_2} is continuously embedded into W_{σ_1} when $\sigma_2 \geq \sigma_1$. We hypothesise the following approximation property: for each $\mathbf{z} \in W_s$ there exist $\mathbf{z}_h \in Y_h$ and a positive constant C_{appr} (independent of z and h) such that

$$\|h^{-s}(\mathbf{z}-\mathbf{z}_h)\| \leq C_{appr}\|\mathbf{z}\|_s.$$

For finite element methods, this hypothesis is easily fulfilled by choosing $W_s = [H^s(\Omega)]^m$, for an appropriate $s = s_{appr} > 0$, and referring to standard approximation properties of piecewise polynomial functions in Sobolev spaces, with $\mathbf{z}_h \in Y_h$ taken as the interpolant, the quasi-interpolant or the projection of $\mathbf{z}$. Thus we arrive at the bound

$$\|\mathbf{u}-\mathbf{u}_h\|^2 \leq C_{appr}\|h^s\mathbf{r}_h\|\,\|\mathbf{z}\|_s.$$

Now we have reached the final and most subtle step in the *a posteriori* error analysis. The norm $\|\mathbf{z}\|_s$ appearing on the right-hand side of the last inequality has to be eliminated in terms of $\mathbf{u}-\mathbf{u}_h$ by noting the relationship between $\mathbf{z}$ and $\mathbf{u}-\mathbf{u}_h$, namely that $\mathcal{L}^*\mathbf{z} = \mathbf{u}-\mathbf{u}_h$. In order to proceed, we shall suppose that $\mathcal{L}^*$ is invertible and that $(\mathcal{L}^*)^{-1}$ is a bounded linear operator from Y into W_s for some $s \in [0, s_{appr}]$; thus,

$$\|\mathbf{z}\|_s = \|(\mathcal{L}^*)^{-1}(\mathbf{u}-\mathbf{u}_h)\|_s \leq C_{stab}\|\mathbf{u}-\mathbf{u}_h\|,$$

where C_{stab} is a positive constant (referred to as the stability constant of the dual problem), greater than or equal to the norm of $(\mathcal{L}^*)^{-1}$. Upon combining the last two bounds we deduce the desired *a posteriori* bound on the global error $\mathbf{e}_h = \mathbf{u} - \mathbf{u}_h$ in terms of the finite element residual $\mathbf{r}_h$:

$$\|\mathbf{u} - \mathbf{u}_h\| \leq C_{appr}\, C_{stab} \|h^s \mathbf{r}_h\|.$$

Once the numerical solution $\mathbf{u}_h$ has been determined, the finite element residual $\mathbf{r}_h$ and $\|h^s \mathbf{r}_h\|$ are easy to compute. However the last inequality will only be of practical use if the constants C_{appr} and C_{stab} are also available. Estimating C_{appr} is a relatively simple task, using readily available results from approximation theory (see, for example, Exercise 3.1.2 in Ciarlet's monograph [12] for an explicit formula for C_{appr} in the case of standard finite element spaces consisting of continuous piecewise polynomials on simplices, or the work of Handscomb [23] for sharper estimates of C_{appr} for piecewise linear finite elements on triangles). On the other hand, providing a numerical value for C_{stab} is much harder, involving the study of the well-posedness of the dual problem. Since any value of C_{stab} which is arrived at through the use of general analytical (worst-case-scenario) arguments is bound to be a considerable overestimate of the ratio $\|\mathbf{z}\|_s / \|\mathbf{u} - \mathbf{u}_h\|$, in practice the stability constant C_{stab} is determined computationally for the specific problem at hand, as part of the process of *a posteriori* error estimation or by other computational means (see, for example, the Thesis of Sandboge [50], where strong stability constants of dual problems are predicted by means of statistical analysis).

Finally, we have to determine s, the exponent of h in the error bound. Ideally, one would like the *a posteriori* bound to reflect the approximation property of the test space Y_h to its full extent; consequently, one would wish to choose s as large as possible, and pick $s = s_{appr}$. Unfortunately, for first-order hyperbolic systems the weak smoothing properties of $(\mathcal{L}^*)^{-1}$ (which is a bounded linear operator from $Y = [L_2(\Omega)]^m$ into the anisotropic space $H(\mathcal{L}^*, \Omega)$, but not into $W_1 = [H^1(\Omega)]^m$) pose an unsurmountable limitation on the choice of s. Indeed, since we have restricted ourselves to operating within the realm of standard isotropic Sobolev spaces, such as as $W_s = [H^s(\Omega)]^m$, where approximation theory by piecewise polynomial functions is well developed, it follows that the strongest statement that we can make (in terms of these spaces) is that $(\mathcal{L}^*)^{-1}$ is a bounded operator from Y into $Y = W_0$, only; consequently, s cannot exceed 0 and we end up with the error bound

$$\|\mathbf{u} - \mathbf{u}_h\| \leq C_{appr}\, C_{stab} \|\mathbf{r}_h\|.$$

In fact, we note that when $s = 0$ we do not benefit from the application of Galerkin orthogonality, and we may simply take $\mathbf{z}_h = 0$ in our argument to simplify this bound to

$$\|\mathbf{u} - \mathbf{u}_h\| \leq C_{stab} \|\mathbf{r}_h\|.$$

Either way, we see that 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 explicitly, in terms of powers of h, the approximation

properties of the test space Y_h. Worse still, when the data are discontinuous, linear hyperbolic equations may possess solutions that are discontinuous across characteristic hypersurfaces and, under mesh refinement, the associated residual norm $\|\mathbf{r}_h\|$ 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.

A possible approach to rectifying the problem is based on perturbing the first-order hyperbolic operator $\mathcal{L}^*$ (or, indeed, both $\mathcal{L}$ and $\mathcal{L}^*$) through the addition of a second-order elliptic term with a small coefficient (see [31]), which then provides additional isotropic regularity; in favourable circumstances the perturbed adjoint operator has bounded inverse from $W_0 = [L_2(\Omega)]^m$ into $W_2 \subset [H^2(\Omega)]^m$ which then allows one to take $s = 2$. This approach, however, is associated with undesirable complications when applied in bounded domains, 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, particularly in the vicinity of the boundary; a further difficulty with elliptic regularisation of non-dissipative hyperbolic problems, such as the Maxwell system of electro-magnetism, is that it may introduce a physically unacceptable level of damping into the model. Our aim in these notes is to pursue a direct approach to the *a posteriori* error analysis of finite element methods for first-order hyperbolic problems, namely one that does not require elliptic regularisation of the hyperbolic operator. The next section is devoted to some simple examples which illustrate the technique.

4.2 Petrov-Galerkin finite element methods

In this section we shall present a general framework of *a posteriori* error estimation for Galerkin finite element approximations of hyperbolic problems. Suppose that X and Y are two real Banach spaces, equipped with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. In the context of the problems discussed here, we can think of two natural choices:

α) When the boundary condition $\gamma_{B^-}(\mathbf{u}) = \mathbf{0}$ is enforced *strongly*, we take $X = D(\mathcal{L}, \Omega)$ equipped with the graph norm $\|\cdot\|_X = |||\cdot|||_{\xi,\Omega}$ and $Y = [L_2(\Omega)]^m$ with the norm $\|\cdot\|_Y = \|\cdot\|_{[L_2(\Omega)]^m}$;

β) When the boundary condition $\gamma_{B^-}(\mathbf{u} - \mathbf{g}) = \mathbf{0}$ is enforced *weakly* (through the definition of the bilinear functional in the variational formulation, rather than through the definition of the solution space), then we take $X = H(\mathcal{L}, \Omega)$ equipped with the graph norm $\|\cdot\|_X = |||\cdot|||_{\xi,\Omega}$ and $Y = [H^1(\Omega)]^m$ with the norm $\|\cdot\|_Y = \|\cdot\|_{[H^1(\Omega)]^m}$.

Suppose further that $a(\cdot,\cdot)$ is a bilinear functional on $X \times Y$, and let $l(\cdot)$ be a linear functional on Y; in the context of symmetric positive systems we adopt the following definitions, corresponding to cases α) and β).

α) In the case of a strongly imposed (homogeneous) boundary condition,

$$a(\mathbf{w}, \mathbf{v}) = (\mathcal{L}\mathbf{w}, \mathbf{v}), \qquad \mathbf{w} \in X = D(\mathcal{L}, \Omega), \quad \mathbf{v} \in Y = [L_2(\Omega)]^m$$

and

$$l(\mathbf{v}) = (\mathbf{f}, \mathbf{v}), \qquad \mathbf{v} \in Y = [L_2(\Omega)]^m.$$

β) In the case of a weakly imposed (non-homogeneous) boundary condition,

$$a(\mathbf{w}, \mathbf{v}) = (\mathcal{L}\mathbf{w}, \mathbf{v}) - \langle \gamma_{B^-}(\mathbf{w}), \gamma_0(\mathbf{v}) \rangle,$$
$$\mathbf{w} \in X = H(\mathcal{L}, \Omega), \mathbf{v} \in Y = [H^1(\Omega)]^m$$

and

$$l(\mathbf{v}) = (\mathbf{f}, \mathbf{v}) - \langle \gamma_{B^-}(\mathbf{g}), \gamma_0(\mathbf{v}) \rangle, \quad \mathbf{v} \in Y = [H^1(\Omega)]^m.$$

Either way, we arrive at a variational problem of the following form: find $\mathbf{u}$ in X such that

$$a(\mathbf{u}, \mathbf{v}) = l(\mathbf{v}) \quad \forall \mathbf{v} \in Y. \tag{4.1}$$

The existence of a unique solution to this problem in the case of a strongly imposed homogeneous boundary condition has been shown in the previous section (see Proposition 6). The problem with the weakly imposed non-homogeneous boundary condition is easily seen to be equivalent to the problem with strongly imposed non-homogeneous boundary condition, considered at the end of Section 3.2: $\mathbf{u}$ is a solution to one problem if and only if it is a solution to the other. While in the case of $\mathbf{g} = \mathbf{0}$ the formulations α) and β) are entirely equivalent, this is not necessarily true of the associated Galerkin discretisations; this will be discussed in more detail below.

Let us consider the Galerkin finite element discretisation of problem (4.1). Given that the computational domain Ω is a Lipschitz domain in $\mathbb{R}^n$, we consider a *partition* of Ω; namely, we select a finite collection $\mathcal{T}_h = \{\kappa_i\}$ of Lipschitz subdomains κ_i of Ω such that:

(1) $\kappa_i \cap \kappa_j$ is an empty set if $i \neq j$, and

(2) $\cup_i \overline{\kappa}_i = \bar{\Omega}$.

Furthermore, in order to avoid the presence of "hanging nodes", a partition will be assumed to have an additional property which, for the sake of simplicity, we only formulate in the two-dimensional case using triangular elements; an analogous assumption is adopted for higher dimensions and other types of elements:

(3) No vertex of any triangle lies in the interior of an edge of another triangle.

We select a family of finite element spaces X_h, parametrised by h, $0 < h \leq h_0$ (typically, h is taken to be a piecewise constant function whose value on element κ is equal to the diameter of κ), consisting of piecewise polynomial functions on the partition $\mathcal{T}_h = \{\kappa_i\}$ of Ω, such that $X_h \subset X$ for all $h > 0$. Analogously, we suppose that Y_h is a finite element space contained in Y. It will be assumed that the spaces X_h and Y_h are equipped with the norms of X and Y, respectively. We consider the following approximation of problem (4.1): find $\mathbf{u}_h$ in X_h such that

$$a(\mathbf{u}_h, \mathbf{v}_h) = l(\mathbf{v}_h) \quad \forall \mathbf{v}_h \in Y_h. \tag{4.2}$$

The only hypothesis that we shall adopt throughout is that the spaces X_h and Y_h have been chosen so that *(4.2) has a unique solution* $\mathbf{u}_h$ *for each* $h \in (0, h_0]$*;* we note in passing that this can be ensured by satisfying the conditions of the next proposition (see [3]).

Proposition 7 *Suppose that the bilinear functional* $a(\cdot,\cdot)$ *is bounded on* $X_h \times Y_h$ *and that the linear functional* $l(\cdot)$ *is bounded on* Y_h*; namely, there exists a positive constant* M_1 *such that*

$$|a(\mathbf{w}_h, \mathbf{v}_h)| \leq M_1 \|\mathbf{w}_h\|_X \|\mathbf{v}_h\|_Y$$

for all $\mathbf{w}_h$ *in* X_h *and all* $\mathbf{v}_h$ *in* Y_h*, and a positive constant* M_2 *such that*

$$|l(\mathbf{v}_h)| \leq M_2 \|\mathbf{v}_h\|_Y$$

for all $\mathbf{v}_h$ *in* Y_h*. Suppose further that* $a(\cdot,\cdot)$ *satisfies the following* inf-sup condition*: there exists a positive constant* M_0 *such that*

$$\inf_{0 \neq \mathbf{w}_h \in X_h} \sup_{0 \neq \mathbf{v}_h \in Y_h} \frac{a(\mathbf{w}_h, \mathbf{v}_h)}{\|\mathbf{w}_h\|_X \|\mathbf{v}_h\|_Y} \geq M_0;$$

and

$$\sup_{\mathbf{w}_h \in X_h} a(\mathbf{w}_h, \mathbf{v}_h) > 0 \quad \forall \mathbf{v}_h \in Y_h.$$

Then there exists a unique $\mathbf{u}_h$ *in* X_h *such that*

$$a(\mathbf{u}_h, \mathbf{v}_h) = l(\mathbf{v}_h) \quad \forall \mathbf{v}_h \in Y_h.$$

Furthermore,

$$\|\mathbf{u}_h\|_X \leq \frac{1}{M_0} \|l\|_{Y'}.$$

Of the conditions listed in Theorem 7 the boundedness the bilinear functional $a(\cdot,\cdot)$ on $X_h \times Y_h$ and the boundedness of the functional $l(\cdot)$ on Y_h follow automatically from the boundedness of these functionals on $X \times Y$ and Y, respectively. On the other hand, the verification of the inf-sup condition on $X_h \times Y_h$ can be a non-trivial exercise, depending on the choice of the spaces X_h and Y_h. As the precise structure of the conditions which guarantee the existence and uniqueness of $\mathbf{u}_h$ is of no relevance in the *a posteriori* error analysis that we wish to pursue, we shall simply suppose that (4.2) possesses a unique solution for each $h \in (0, h_0]$; no structural conditions on $a(\cdot,\cdot)$ and $l(\cdot)$ of the kind appearing in Theorem 7 will be made.

Next we derive an *a posteriori* bound on the global error $\mathbf{e}_h = \mathbf{u} - \mathbf{u}_h$. Suppose that $\psi \in [C_0^\infty(\Omega)]^m$, and consider the auxiliary (dual) problem: find $\mathbf{z}$ in Y such that

$$a(\mathbf{w}, \mathbf{z}) = (\mathbf{w}, \psi) \quad \forall \mathbf{w} \in X, \tag{4.3}$$

where $(\cdot,\cdot)$ denotes the inner product of $[L_2(\Omega)]^m$. Thus,

$$(\mathbf{e}_h, \psi) = a(\mathbf{e}_h, \mathbf{z}) = a(\mathbf{u} - \mathbf{u}_h, \mathbf{z}) = a(\mathbf{u} - \mathbf{u}_h, \mathbf{z} - \mathbf{z}_h),$$

where $\mathbf{z}_h$ is any element in Y_h. Hence

$$(\mathbf{e}_h, \psi) = l(\mathbf{z} - \mathbf{z}_h) - a(\mathbf{u}_h, \mathbf{z} - \mathbf{z}_h).$$

We write the right-hand side as $\langle \mathbf{r}_h, \mathbf{z} - \mathbf{z}_h \rangle$, where $\langle \cdot, \cdot \rangle$ is the duality pairing between Y', the dual space of Y, and Y. The quantity $\mathbf{r}_h \in Y'$ is referred to as the finite element residual. Consequently,

$$(\mathbf{u} - \mathbf{u}_h, \psi) = \langle \mathbf{r}_h, \mathbf{z} - \mathbf{z}_h \rangle.$$

Our aim is to derive a bound on the global error in terms of the finite element residual, based on this error representation formula.

From here on, we shall distinguish between the two cases, labelled α) and β), which were formulated earlier on, corresponding to strongly imposed and weakly imposed boundary conditions, respectively. We begin by developing an error analysis for case α).

α) In this case $Y = [L_2(\Omega)]^m$, and therefore

$$(\mathbf{u} - \mathbf{u}_h, \psi) = (\mathbf{r}_h, \mathbf{z} - \mathbf{z}_h). \tag{4.4}$$

We shall suppose that $\mathbf{f} \in [L_2(\Omega)]^m$, that $\mathcal{T}_h$ is a finite element partition of Ω into elements κ, and we adopt the following standard approximation property for the test space Y_h:

(c) There exists a positive constant C_2, independent of h, such that for each $\mathbf{v} \in [H^1(\Omega)]^m$ there is $\mathbf{v}_h \in Y_h$ with

$$\|h^{-1}(\mathbf{v} - \mathbf{v}_h)\|_{[L_2(\Omega)]^m} \leq C_2 \|\mathbf{v}\|_{[H^1(\Omega)]^m}.$$

Theorem 8 *Suppose that Hypothesis 2 and condition (c) hold; then,*

$$\|\mathbf{u} - \mathbf{u}_h\|_{[H^{-1}(\Omega)]^m} \leq C_1' C_2 \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m},$$

where $\|\cdot\|_{[H^{-1}(\Omega)]^m}$ *denotes the norm of the dual space of* $[H_0^1(\Omega)]^m$.

Proof Given that $\psi \in [C_0^\infty(\Omega)]^m$, it follows from (4.4) that

$$\begin{aligned} (\mathbf{u} - \mathbf{u}_h, \psi) &= (\mathbf{r}_h, \mathbf{z} - \mathbf{z}_h) \\ &\leq \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} \|h^{-1}(\mathbf{z} - \mathbf{z}_h)\|_{[L_2(\Omega)]^m} \\ &\leq C_2 \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} \|\mathbf{z}\|_{[H^1(\Omega)]^m}. \end{aligned} \tag{4.5}$$

According to (3.11) with $\mu = \psi$ and $\chi = \mathbf{0}$,

$$\|\mathbf{z}\|_{[H^1(\Omega)]^m} \leq C_1' \|\psi\|_{[H^1(\Omega)]^m}.$$

Substituting this into (4.5), dividing both sides by $\|\psi\|_{[H^1(\Omega)]^m}$, taking the supremum over all $\psi \in [C_0^\infty(\Omega)]^m$ and noting that $[C_0^\infty(\Omega)]^m$ is dense in $[H_0^1(\Omega)]^m$, we obtain the desired error bound. ■

Next, we carry out a similar analysis for case β) corresponding to weakly imposed non-homogeneous boundary condition, with $\mathbf{g} \in [H^1(\Omega)]^m$.

β) Arguing in the same way as in case α), we deduce that

$$\begin{aligned}(\mathbf{u}-\mathbf{u}_h,\psi) &= (\mathbf{r}_h,\mathbf{z}-\mathbf{z}_h)\\ &\quad +\langle\gamma_{B^-}(\mathbf{u}_h-\mathbf{g}),\gamma_0(\mathbf{z}-\mathbf{z}_h)\rangle \equiv I+II.\end{aligned}$$

To proceed, replace c) by the following hypothesis:

(c') There exists a positive constant C_2, independent of h, such that for each $\mathbf{v} \in [H^1(\Omega)]^m$ there is $\mathbf{v}_h \in Y_h$ with

$$\|h^{-1}(\mathbf{v}-\mathbf{v}_h)\|_{[L_2(\Omega)]^m} + \|h^{-1/2}\gamma_0(\mathbf{v}-\mathbf{v}_h)\|_{[L_2(\partial\Omega)]^m} \leq C_2\|\mathbf{v}\|_{[H^1(\Omega)]^m}.$$

Term I is dealt with as in case α) to deduce that

$$I \leq C_2\|h\mathbf{r}_h\|_{[L_2(\Omega)]^m}\|\mathbf{z}\|_{[H^1(\Omega)]^m}.$$

To estimate II, we note that

$$\gamma_{B^-}(\mathbf{u}_h-\mathbf{g}) = B^-\gamma_0(\mathbf{u}_h-\mathbf{g}) \in [L_2(\partial\Omega)]^m,$$

so that

$$II \leq \|h^{1/2}B^-\gamma_0(\mathbf{u}_h-\mathbf{g})\|_{[L_2(\partial\Omega)]^m}\|h^{-1/2}(\mathbf{z}-\mathbf{z}_h)\|_{[L_2(\partial\Omega)]^m}$$

and therefore

$$II \leq C_2\|h^{1/2}B^-\gamma_0(\mathbf{u}_h-\mathbf{g})\|_{[L_2(\partial\Omega)]^m}\|\mathbf{z}\|_{[H^1(\Omega)]^m}.$$

Thus, appealing to the strong stability bound (3.11) for the adjoint (dual) problem, we deduce the following theorem.

Theorem 9 *Suppose that Hypothesis 2 and condition (c') hold; then,*

$$\|\mathbf{u}-\mathbf{u}_h\|_{[H^{-1}(\Omega)]^m} \leq C_1'C_2\left(\|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} + \|h^{1/2}\mathbf{r}_h^-\|_{[L_2(\partial\Omega)]^m}\right),$$

where $\|\cdot\|_{[H^{-1}(\Omega)]^m}$ *denotes the norm of the dual space of* $[H_0^1(\Omega)]^m$, $\mathbf{r}_h = \mathbf{f}-\mathcal{L}\mathbf{u}_h$ *denotes the* interior residual, *and* $\mathbf{r}_h^- = B^-\gamma_0(\mathbf{g}-\mathbf{u}_h)$ *signifies the* boundary residual.

The interior and boundary residuals measure the extent to which $\mathbf{u}_h$ fails to satisfy the partial differential equation and the boundary condition, respectively. In Theorem 8 the boundary residual term did not arise since there we had $X_h \subset X = D(\mathcal{L},\Omega)$, so the boundary condition was satisfied exactly by all elements of the finite element trial space, including $\mathbf{u}_h$.

4.3 The streamline diffusion method

For Petrov-Galerkin approximations of symmetric positive systems ensuring stability is a non-trivial matter. We have already touched on this issue in the previous section when we commented on Proposition 7: to prove stability, one has to verify the inf-sup condition, with a constant M_0 (preferably, independent of h), for the particular choice of test and trial space. For examples of stable Petrov-Galerkin methods for hyperbolic systems, we refer to [37], [38] and [61].

As an alternative to these techniques, in this section we consider a family of methods which use the same test and trial space and a bilinear functional $a_\delta(\cdot,\cdot)$ which is a consistent perturbation of the bilinear functional $a(\cdot,\cdot)$ such that the resulting Galerkin method is stable. The stabilising perturbation term acts along the characteristic hyperplanes of the differential operator $\mathcal{L}$ and can be thought of physically as a numerical diffusion term in the direction of the streamlines; hence the name of the resulting discretisation technique: the *streamline diffusion finite element method*.

In order to highlight the key issues concerning the *a posteriori* error analysis of the streamline diffusion finite element method we consider the symmetric positive system

$$\mathcal{L}\mathbf{u} = \mathbf{f} \quad \text{in } \Omega, \qquad \gamma_{B^-}(\mathbf{u}-\mathbf{g}) = \mathbf{0} \quad \text{on } \partial\Omega, \tag{4.6}$$

where $\mathbf{f} \in [L_2(\Omega)]^m$ and $\mathbf{g} \in [H^1(\Omega)]^m$.

Let us suppose that $\bar{\Omega}$ has been subdivided by a finite element partition $\mathcal{T}_h = \{\kappa_i\}$; here h is a piecewise constant mesh function with $h(x) = \operatorname{diam}(\kappa_i)$ when x is in element κ_i. On this partition we consider the finite element space X_h, $X_h \subset [H^1(\Omega)]^m$, consisting of continuous piecewise polynomials of fixed degree k, $k \geq 1$. It will be assumed that X_h possesses the following standard approximation property:

(c") Given that $\mathbf{v} \in [H^1(\Omega)]^m$, there exists $\mathbf{v}_h \in X_h$ and a constant C_2, independent of $\mathbf{v}$ and h, such that

$$\begin{aligned}\|h^{-1}(\mathbf{v}-\mathbf{v}_h)\|_{[L_2(\Omega)]^m} + \|h^{-1/2}\gamma_0(\mathbf{v}-\mathbf{v}_h)\|_{[L_2(\partial\Omega)]^m} \\ +\|\mathbf{v}_h\|_{[H^1(\Omega)]^m} \leq C_2\|\mathbf{v}\|_{[H^1(\Omega)]^m}.\end{aligned} \tag{4.7}$$

Given a function $\mathbf{v}$ and an associated $\mathbf{v}_h$ satisfying (4.7), we shall write $\mathbf{v}_h = P_h\mathbf{v}$ to denote that $\mathbf{v}_h$ is assigned to $\mathbf{v}$.

Next, we introduce the streamline diffusion finite element approximation of (4.6); to do so, we define the *streamline diffusion parameter* δ as a piecewise constant function on $\bar{\Omega}$ whose value on $\kappa \in \mathcal{T}_h$ is

$$\delta|_\kappa = K_0 \operatorname{diam}(\kappa), \quad \kappa \in \mathcal{T}_h, \tag{4.8}$$

where K_0 is a fixed positive constant. Further, we consider the bilinear form $a_\delta(\cdot,\cdot)$ defined by

$$a_\delta(\mathbf{w},\mathbf{v}) = (\mathcal{L}\mathbf{w}, \mathbf{v}+\delta\mathcal{L}\mathbf{v}) - \langle\gamma_{B^-}(\mathbf{w}), \gamma_0(\mathbf{v})\rangle,$$

and the linear functional

$$l_\delta(\mathbf{v}) = (\mathbf{f}, \mathbf{v} + \delta\mathcal{L}\mathbf{v}) - \langle \gamma_{B^-}(\mathbf{g}), \gamma_0(\mathbf{v})\rangle.$$

In these definitions $(\cdot,\cdot)$ denotes the inner product of $[L_2(\Omega)]^m$.

Streamline diffusion method: Find $\mathbf{u}_h \in X_h$ such that

$$a_\delta(\mathbf{u}_h, \mathbf{v}_h) = l_\delta(\mathbf{v}_h) \quad \forall \mathbf{v}_h \in X_h. \tag{4.9}$$

Assuming that X_h has been equipped with the norm of $H(\mathcal{L},\Omega)$, it is a simple matter to show that a_δ and l_δ satisfy the hypotheses of Theorem 7 on X_h with $X = Y = H(\mathcal{L},\Omega)$ and $Y_h = X_h$, and thereby (4.9) has a unique solution $\mathbf{u}_h$ in X_h. Formally, (4.9) can be thought of as a perturbation of the standard Galerkin method corresponding to $\delta \equiv 0$.

In order to illustrate the qualitative improvement over the standard Galerkin finite element method offered by the streamline diffusion method, we show in Figure 1 the results of a numerical experiment. Our model problem is

$$\nabla \cdot (\mathbf{a}u) = 0 \quad \text{on } \Omega = (0,1)^2, \qquad u = g \quad \text{on } \partial_-\Omega, \tag{4.10}$$

where $\mathbf{a} = (2,1)$, $g(0,y) = 1$ for $0 \le y \le 1$ and $g(x,0) = 0$ for $0 < x \le 1$. On Ω we considered a triangulation which arises from a 21×21 uniform mesh by connecting the bottom-left corner of each mesh-square with its top-right corner; Figure 1 shows the numerical solution obtained by using: (a) the standard Galerkin finite element method with continuous piecewise linear trial and test functions, and (b) the numerical solution given by the streamline diffusion finite element method with the same trial and test spaces, and stabilisation parameter $K_0 = 0.5/\|\mathbf{a}\|$, where $\|\mathbf{a}\|$ denotes the Euclidean norm of $\mathbf{a}$.

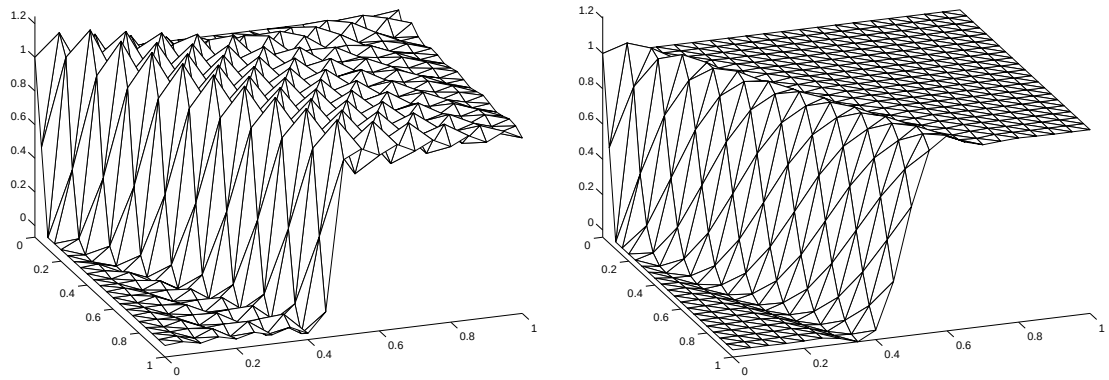


Figure 1: Corner discontinuity problem: (a) Standard Galerkin finite element method; (b) Streamline diffusion method with $K_0 = 0.5/\|\mathbf{a}\|$.

Here we shall be concerned with the *a posteriori* error analysis of the streamline diffusion method. The analysis relies on the following *Galerkin orthogonality* property:

$$a_\delta(\mathbf{u} - \mathbf{u}_h, \mathbf{v}_h) = 0 \quad \forall \mathbf{v}_h \in X_h. \tag{4.11}$$

The equation (4.11) is easily seen to hold by noting (4.9) and that

$$a_\delta(\mathbf{u}, \mathbf{v}_h) = l_\delta(\mathbf{v}_h) \quad \forall \mathbf{v}_h \in X_h.$$

The starting point in the argument is the following *dual problem*: given that $\psi \in [C_0^\infty(\Omega)]^m$, find $\mathbf{z}$ in $H(\mathcal{L}^*, \Omega)$ such that

$$\mathcal{L}^*\mathbf{z} = \psi \quad \text{in } \Omega, \qquad \gamma_{B^+}(\mathbf{z}) = \mathbf{0} \quad \text{on } \partial\Omega. \tag{4.12}$$

The *a posteriori* error bound that we shall state below will be expressed in terms of the finite element residual

$$\mathbf{r}_h = \mathbf{f} - \mathcal{L}\mathbf{u}_h$$

which measures the extent to which $\mathbf{u}_h$ fails to satisfy the differential equation in Ω; thus, as in the previous section, we shall refer to it as the *internal residual*. Also, since $\mathbf{u}_h$ satisfies the boundary condition only approximately rather then in the pointwise sense, the difference $B^-\gamma_0(\mathbf{g} - \mathbf{u}_h)$ is not necessarily zero and will be seen to enter the *a posteriori* error bound; thus, we also define the *boundary residual*

$$\mathbf{r}_h^- = B^-\gamma_0(\mathbf{g} - \mathbf{u}_h).$$

With these definitions we have the following result.

Theorem 10 *Assuming Hypothesis 2 and (c"), the following a posteriori error bound holds:*

$$\|\mathbf{u} - \mathbf{u}_h\|_{[H^{-1}(\Omega)]^m} \leq C_3 \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} + C_1' C_2 \|h^{1/2}\mathbf{r}_h^-\|_{[L_2(\partial\Omega)]^m},$$

where $C_3 = C_1'(C_2 + K_0 C_4)$ with C_1' the strong stability constant for the dual problem appearing in (3.11), C_2 and K_0 the constants from conditions (4.7) and (4.8) and C_4 defined in (4.14) below.

Proof Let $\psi \in [C_0^\infty(\Omega)]^m$. Recalling the dual problem (4.12), integrating by parts, and appealing to the Galerkin orthogonality property (4.11), we deduce that, for any $\mathbf{z}_h \in X_h$,

$$\begin{aligned}(\mathbf{u} - \mathbf{u}_h, \psi) &= (\mathbf{r}_h, \mathbf{z} - \mathbf{z}_h) - \langle \gamma_{B^-}(\mathbf{g} - \mathbf{u}_h), \gamma_0(\mathbf{z} - \mathbf{z}_h)\rangle - (\delta\mathbf{r}_h, \mathcal{L}\mathbf{z}_h) \\ &\equiv I + II + III. \end{aligned} \tag{4.13}$$

Next, we make a specific choice of $\mathbf{z}_h$: we take $\mathbf{z}_h = P_h\mathbf{z}$, where P_h is defined in hypothesis (c"). Terms I and II are dealt with as in case β) of the previous section; thus, noting (4.7), we have that

$$\begin{aligned} |\, I\,| &\leq \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} \|h^{-1}(\mathbf{z} - \mathbf{z}_h)\|_{[L_2(\Omega)]^m} \\ &\leq C_2 \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} \|\mathbf{z}\|_{[H^1(\Omega)]^m}, \end{aligned}$$

and applying (4.7) and noting that $\gamma_{B^-}(\mathbf{g} - \mathbf{u}_h) = B^-\gamma_0(\mathbf{g} - \mathbf{u}_h)$ yields

$$|\, II\,| \leq C_2 \|h^{1/2}\mathbf{r}_h^-\|_{[L_2(\partial\Omega)]^m} \|\mathbf{z}\|_{[H^1(\Omega)]^m}.$$

Further,

$$\begin{aligned} |\,III\,| &\leq \|\delta \mathbf{r}_h\|_{[L_2(\Omega)]^m} \|\mathcal{L}\mathbf{z}_h\|_{[L_2(\Omega)]^m} \\ &\leq K_0 C_4 \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} \|\mathbf{z}\|_{[H^1(\Omega)]^m}, \end{aligned}$$

where

$$C_4 = C_2 \left(\sum_{i=1}^{n} \|A_i\|^2_{[C(\bar{\Omega})]^{m\times m}} + \|C\|^2_{[C(\bar{\Omega})]^{m\times m}} \right)^{1/2}. \tag{4.14}$$

Upon collecting the bounds on I, II and III, and inserting them into (4.13), we deduce that

$$\begin{aligned} |(\mathbf{u}-\mathbf{u}_h, \psi)| \leq\ & C_2 \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} \|\mathbf{z}\|_{[H^1(\Omega)]^m} \\ & + C_2 \|h^{1/2}\mathbf{r}_h^-\|_{[L_2(\partial\Omega)]^m} \|\mathbf{z}\|_{[H^1(\Omega)]^m} \\ & + K_0 C_4 \|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} \|\mathbf{z}\|_{[H^1(\Omega)]^m}. \end{aligned} \tag{4.15}$$

Finally, recalling the strong stability estimate for the dual problem, the last inequality implies the desired *a posteriori* error bound. ∎

When we formally set $K_0 = 0$, the streamline diffusion finite element method reduces to the standard Galerkin finite element method; similarly, the bound derived in Theorem 10 collapses to that in Theorem 9, as expected.

4.4 The cell vertex finite volume method

In the previous section we outlined the error analysis of the streamline diffusion method and we saw that perturbing the bilinear form $a(\cdot,\cdot)$ to $a_\delta(\cdot,\cdot)$ did not affect the *a posteriori* error bound in the H^{-1} norm (except, perhaps, altering the constant in the error bound). In this section, we consider a different perturbation of the basic Galerkin framework by applying numerical quadrature to a Petrov-Galerkin finite element method; as an illustration of the effects of such a "non-Galerkin" perturbation, we discuss the *a posteriori* error analysis of the cell vertex finite volume method.

For the sake of simplicity, we suppose in this subsection that Ω is the unit square $(0,1)^2$. We consider the symmetric positive system in conservation form subject to a non-homogeneous boundary condition:

$$\mathcal{L}\mathbf{u} \equiv \sum_{i=1}^{2} \frac{\partial}{\partial x_i}(A_i\mathbf{u}) + C\mathbf{u} = \mathbf{f} \quad \text{in } \Omega, \tag{4.16}$$

$$\gamma_{B^-}(\mathbf{u}-\mathbf{g}) = \mathbf{0} \quad \text{on } \partial\Omega, \tag{4.17}$$

where $\mathbf{f} \in [L_2(\Omega)]^m$ and $\mathbf{g} \in [H^1(\Omega)]^m$. In order to formulate the cell vertex finite volume discretisation of this problem, we subdivide Ω by a structured mesh consisting of convex quadrilaterals. Here the word *structured* signifies the fact that the partition is topologically equivalent to a uniform square mesh on Ω. The cell vertex finite volume approximation to this boundary value problem is obtained by integrating the system of partial differential equations (4.16) over each quadrilateral in the partition and exploiting the fact that the equations are in divergence form: Gauss' theorem is applied to convert

integrals over quadrilaterals into contour integrals over the boundaries of quadrilaterals; these contour integrals are then approximated by means of the trapezium rule. This process provides a four-point finite difference scheme referred to as the cell vertex finite volume method, given that the unknowns are carried at the vertices of the cells – the quadrilateral elements in the partition.

For the purposes of the present paper it is useful to note that the construction of the cell vertex scheme can be also described in the language of finite element methods. Thus, let $\mathcal{F} = \{\mathcal{T}_h\}$, $h > 0$, be a regular family of structured partitions $\mathcal{T}_h$ of $\Omega = (0,1)^2$ into convex quadrilateral elements κ_{ij}. 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

$$\begin{aligned} Y_h &= \left\{ \mathbf{v} \in [L_2(\Omega)]^m \,:\, \mathbf{v} = \hat{\mathbf{v}} \circ F_{\kappa_{ij}}^{-1},\ \hat{\mathbf{v}} \in [\mathcal{Q}_0(\hat{\kappa})]^m,\ \kappa_{ij} \in \mathcal{T}_h \right\}, \\ X_h &= \left\{ \mathbf{w} \in [H^1(\Omega)]^m \,:\, \mathbf{w} = \hat{\mathbf{w}} \circ F_{\kappa_{ij}}^{-1},\ \hat{\mathbf{w}} \in [\mathcal{Q}_1(\hat{\kappa})]^m,\ \kappa_{ij} \in \mathcal{T}_h \right\}. \end{aligned}$$

Let us denote by $\Pi_h : [L_2(\Omega)]^m \to Y_h$ the orthogonal projector in $[L_2(\Omega)]^m$ onto the linear subspace Y_h, and by

$$\mathcal{I}_h : (H(L,\Omega) \cap [C(\bar{\Omega})]^m)^2 \to X_h \times X_h$$

the interpolation projector onto $X_h \times X_h$. With this notation, we put

$$a_h(\mathbf{w}_h, \mathbf{v}_h) = (\operatorname{div} \mathcal{I}_h(\mathcal{A}\mathbf{w}_h), \mathbf{v}_h) + (C\mathbf{w}_h, \mathbf{v}_h) - \langle \gamma_{B^-}(\mathbf{w}_h), \gamma_0(\mathbf{v}_h) \rangle,$$

and

$$l(\mathbf{v}_h) = (\mathbf{f}, \mathbf{v}_h) - \langle \gamma_{B^-}(\mathbf{g}), \gamma_0(\mathbf{v}_h) \rangle.$$

The cell vertex finite volume approximation of the boundary-value problem (4.16), (4.17) is now defined as follows: find $\mathbf{u}_h \in X_h$ such that

$$a_h(\mathbf{u}_h, \mathbf{v}_h) = l(\mathbf{v}_h) \qquad \forall \mathbf{v}_h \in Y_h. \tag{4.18}$$

In order to proceed we shall suppose that (4.18) has a unique solution $\mathbf{u}_h$; for the discussion of conditions which are sufficient to ensure that this is the case, and for theoretical results concerning the stability and convergence of the cell vertex scheme we refer to [5], [43], [44], [52], [53] and [54].

Theorem 11 *Suppose that Hypothesis 2 and (c') hold; then, the cell vertex scheme obeys the following* a posteriori *error bound:*

$$\|\mathbf{u} - \mathbf{u}_h\|_{[H^{-1}(\Omega)]^m} \leq C_1' \big[C_2 \big(\|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} + \|h^{1/2}\mathbf{r}_h^-\|_{[L_2(\partial\Omega)]^m} \big) + \|\hat{\mathbf{r}}_h\|_{[L_2(\Omega)]^m} \big]$$

where C_1' is the constant from the strong stability estimate (3.11), C_2 is the constant from the approximation property (c'), $\mathbf{r}_h$ and $\mathbf{r}_h^-$ are the interior and boundary residual, respectively, and $\hat{\mathbf{r}}_h = \Pi_h \operatorname{div} (\mathcal{I}_h(\mathcal{A}\mathbf{u}_h) - \mathcal{A}\mathbf{u}_h)$.

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

$$\mathcal{L}^*\mathbf{z} = \psi \qquad \text{on } \Omega, \qquad\qquad \gamma_{B^+}(\phi) = \mathbf{0}.$$

Then,

$$\begin{aligned}(\mathbf{u}-\mathbf{u}_h,\psi) &= (\mathbf{u}-\mathbf{u}_h,\mathcal{L}^*\mathbf{z})\\ &= (\mathcal{L}(\mathbf{u}-\mathbf{u}_h),\mathbf{z}) - \langle\gamma_{B^-}(\mathbf{u}-\mathbf{u}_h),\mathbf{z}\rangle\\ &= (\mathbf{r}_h,\mathbf{z}) - \langle\mathbf{r}_h^-,\mathbf{z}\rangle\\ &= (\mathbf{r}_h,\mathbf{z}-\mathbf{z}_h) - \langle\mathbf{r}_h^-,\mathbf{z}-\mathbf{z}_h\rangle + (\hat{\mathbf{r}}_h,\mathbf{z}_h)\\ &= I + II + III.\end{aligned}$$

Thus, choosing $\mathbf{z}_h = \Pi_h\mathbf{z}$ and exploiting the fact that with this choice of $\mathbf{z}_h$ the approximation property (c') holds, we have that

$$|III| \le \|\hat{\mathbf{r}}_h\|_{[L_2(\Omega)]^m}\,\|\mathbf{z}\|_{[L_2(\Omega)]^m},$$

and

$$|I+II| \le C_2\left(\|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} + \|h^{1/2}\mathbf{r}_h^-\|_{[L_2(\partial\Omega)]^m}\right)\|\mathbf{z}\|_{[H^1(\Omega)]^m},$$

as in the proof of Theorem 9. Adding up the bounds on I, II and III, and recalling the strong stability estimate

$$\|\mathbf{z}\|_{[H^1(\Omega)]^m} \le C_1'\|\psi\|_{[H^1(\Omega)]^m},$$

we obtain the desired result. ■

The third term on the right-hand side of this *a posteriori* error bound can be rewritten as

$$\sup_{\mathbf{v}_h\in Y_h}\frac{|a(\mathbf{u}_h,\mathbf{v}_h) - a_h(\mathbf{u}_h,\mathbf{v}_h)|}{\|\mathbf{v}_h\|_{[L_2(\Omega)]^m}},$$

and can be thought of as the consistency error between the bilinear functional $a(\cdot,\cdot)$ and its discretisation $a_h(\cdot,\cdot)$, resulting from the non-Galerkin-type error committed by applying a numerical integration rule.

4.5 Reliable quantitative error control and adaptivity

We have established a number of *a posteriori* error bounds on the global error $\mathbf{u}-\mathbf{u}_h$ for finite element and finite volume approximations of symmetric positive systems. These bounds are of the following generic form:

$$\|\mathbf{u}-\mathbf{u}_h\|_{[H^{-1}(\Omega)]^m} \le C_*\left(\sum_{\kappa\in\mathcal{T}_h}|\eta_\kappa(\mathbf{u}_h)|^2\right)^{1/2}, \tag{4.19}$$

where C_* is a 'computable' constant and $\eta_\kappa(\mathbf{u}_h)$ is a *local error indicator* on element κ involving the numerical solution $\mathbf{u}_h$; in particular, in Theorem 8

$$\eta_\kappa(\mathbf{u}_h) = \|h\mathbf{r}_h\|_{[L_2(\kappa)]^m},$$

in Theorems 9 and 10

$$\eta_\kappa(\mathbf{u}_h) = \left(\|h\mathbf{r}_h\|^2_{[L_2(\kappa)]^m} + \|h^{1/2}\mathbf{r}_h^-\|^2_{[L_2(\partial\kappa\cap\partial\Omega)]^m} \right)^{1/2},$$

and in Theorem 11 we have that

$$\eta_\kappa(\mathbf{u}_h) = \left(\|h\mathbf{r}_h\|^2_{[L_2(\kappa)]^m} + \|h^{1/2}\mathbf{r}_h^-\|^2_{[L_2(\partial\kappa\cap\partial\Omega)]^m} + \|\hat{\mathbf{r}}_h\|^2_{[L_2(\kappa)]^m} \right)^{1/2}.$$

The right-hand side in the error bound (4.19) can be evaluated once the finite element solution $\mathbf{u}_h$ has been computed and can be used to estimate the size of the global error in the norm of $[H^{-1}(\Omega)]^m$. Moreover, exploiting the *a posteriori* error bound it is possible to adaptively control the global error to a desired tolerance level by suitably refining the partition. In order to achieve *reliability* in the sense that

$$\|\mathbf{u} - \mathbf{u}_h\|_{[H^{-1}(\Omega)]^m} \leq \mathrm{TOL},$$

where TOL is the prescribed *error tolerance*, it suffices to ensure that

$$\eta_\Omega(\mathbf{u}_h) \equiv C_* \left(\sum_{\kappa\in\mathcal{T}_h} |\eta_\kappa(\mathbf{u}_h)|^2 \right)^{1/2} \leq \mathrm{TOL}.$$

In addition to reliability we are also concerned with *efficiency*, which means that among all possible partitions which yield an approximation with this accuracy we want to determine (the) one that has the smallest number of degrees of freedom. Constructing a partition that is optimal in this sense is a difficult nonlinear optimisation problem whose solution is rarely attempted in practice. The usual approach to constructing a partition which does not contain an excessively large number of elements is to proceed iteratively: we start with a coarse mesh and refine it successively based on the size of the *a posteriori* error estimate, and in the course of doing so we try to keep the number of elements as small as possible. The last inequality can be thought of as a stopping criterion in this iterative processes. In fact, one can adopt various strategies to generate a sequence of partitions from an initial coarse mesh; here we mention only three of the most popular approaches, following Rannacher and Suttmeier [48].

Let an error tolerance TOL or a maximal number of elements $N_{\max}$ be given. Starting from some initial coarse partition, the refinement criteria are chosen in terms of the local error indicators $\eta_\kappa(\mathbf{u}_h)$.

1. *Error-per-cell strategy.* In this approach the mesh generation aims to equilibrate the local error indicators by refining or coarsening the elements κ in the current partition $\mathcal{T}_h$ according to the criterion

$$\eta_\kappa(\mathbf{u}_h) \approx \frac{TOL}{C_*\sqrt{N}},$$

where N is the (predicted) number of elements in the resulting new partition. Since N depends on the result of the refinement decision, this strategy is implicit and

requires an iterative implementation. It is common practice to work with a varying value of N on each refinement level, with N successively updated according to the outcome of the refinement process. This strategy will deliver a partition on which $\eta_\Omega(\mathbf{u}_h) \approx TOL$, provided that N_{max} is not exceeded.

2. *Fixed-fraction strategy.* In each refinement step, the elements are ordered according to the size of the local error indicator $\eta_\kappa(\mathbf{u}_h)$, and then a fixed partition (in two dimensions, typically 30%) of the elements κ with largest $\eta_\kappa(\mathbf{u}_h)$ is refined (resulting in about doubling the number of elements). This process is repeated until the stopping criterion $\eta_\Omega(\mathbf{u}_h) \leq TOL$ is satisfied, or $N_{\max}$ is exceeded.

3. *Fixed-reduction strategy.* Here one works with a variable tolerance TOL_{var}. Supposing that on a partition the approximate solution $\mathbf{u}_h$ has been obtained, the tolerance is set to $TOL_{var} = \sigma\,\eta(\mathbf{u}_h)$, where $\sigma \in (0,1)$ is a fixed reduction factor (e.g. $\sigma = 0.5$). In the next step one (or several) cycles of the *error-per-cell* strategy are performed with tolerance TOL_{var}; this provides a new mesh $\mathcal{T}_h^{new}$ and a new solution $\mathbf{u}_h^{new}$ with associated error estimator $\eta(\mathbf{u}_h^{new})$. Then the tolerance is reduced again by setting $TOL_{var} = \sigma\,\eta(\mathbf{u}_h^{new})$ and a new refinement cycle begins. This iterative process is repeated until $TOL_{var} \leq TOL$, or N_{max} is exceeded.

In each of the three strategies we repeat mesh modification followed by solution on the new partition until the tolerance is satisfied, or the prescribed maximum number of elements is exceeded.

We conclude this section by showing that the adaptive algorithms outlined above will terminate in a finite number of steps. We shall suppose, for simplicity, that only mesh refinements are carried out and no derefinements are done. It is clear that if reaching a prescribed maximum number is taken as stopping criterion, then the mesh refinement algorithm will terminate after a finite number of steps. If an error tolerance is given instead as stopping criterion, then termination of the refinement algorithm can be ensured by proving that the finite element method satisfies the *a priori* error bound

$$\|\mathbf{u} - \mathbf{u}_h\|_{[L_2(\Omega)]^m} + |h|\,|\mathbf{u} - \mathbf{u}_h|_{[H^1(\Omega)]^m} \leq C(\mathbf{u})|h|^{1-\epsilon}, \tag{4.20}$$

where $C(\mathbf{u})$ depends on $\mathbf{u}$ (and its Sobolev smoothness),

$$|h| = \max\{h_\kappa \,:\, \kappa \in \mathcal{T}_h\},$$

and ϵ is a fixed real number in the interval $[0,1)$.

Concerning finite element approximations of the kind mentioned in Theorems 1 – 3, an *a priori* error bound of the type (4.20) can be derived under suitable assumptions on the smoothness of $\mathbf{u}$, the choice of the trial and test space, and the regularity of the partition. Then, noting that $\mathbf{r}_h = \mathcal{L}(\mathbf{u} - \mathbf{u}_h)$, it follows that

$$\|h\mathbf{r}_h\|_{[L_2(\Omega)]^m} \leq Const.|h|\|\mathbf{u} - \mathbf{u}_h\|_{[H^1(\Omega)]^m} \leq Const.|h|^{1-\epsilon},$$

and similarly,

$$\begin{aligned}\|h^{1/2}\mathbf{r}_h^-\|_{[L_2(\partial\Omega)]^m} &\leq Const.|h|^{1/2}\|\gamma_0(\mathbf{u}-\mathbf{u}_h)\|_{[L_2(\partial\Omega)]^m} \\ &\leq Const.|h|^{1/2}\|\mathbf{u}-\mathbf{u}_h\|^{1/2}_{[L_2(\Omega)]^m}\|\mathbf{u}-\mathbf{u}_h\|^{1/2}_{[H^1(\Omega)]^m} \\ &\leq Const.|h|^{1-\epsilon}.\end{aligned}$$

Thus, considering Theorems 1 – 3, it is a simple matter to show that, under the same assumptions as are required to ensure that the *a priori* error bound (4.20) holds, we have that

$$\left(\sum_{\kappa\in\mathcal{T}_h}|\eta_\kappa(\mathbf{u}_h)|^2\right)^{1/2} \to 0 \qquad \text{as } |h|\to 0,$$

and, therefore, the stopping criterion will be satisfied eventually as the mesh is refined.

Concerning Theorem 4, it can be shown that the cell vertex finite volume method satisfies an *a priori* error bound of the type (4.20). More precisely, suppose that $\mathbf{u}\in[H^s(\Omega)]^m$ with $s>1$, that the components of the matrices A_i, $i=1,\ldots,n$, and C belong to $C^{[s]+1}(\bar{\Omega})$, that the matrices A_i are positive definite, uniformly on $\bar{\Omega}$, and that the family of structured partitions $\{\mathcal{T}_h\}$ is quasi-parallel (namely, there exists a fixed positive constant c_* independent of $|h|$ such that, for each κ in $\mathcal{T}_h$, the distance between the midpoints of the two diagonals is bounded by $c_*|h|^2$); then

$$\begin{aligned}&\|\mathbf{u}-\mathbf{u}_h\|_{[L_2(\Omega)]^m}+\|\Pi_h\mathrm{div}(\mathcal{A}(\mathbf{u}-\mathbf{u}_h))\|_{[L_2(\Omega)]^m} \\ &\qquad +|h||\mathbf{u}-\mathbf{u}_h|_{[H^1(\Omega)]^m}\leq Const.|h|^{r-1}\|\mathbf{u}\|_{[H^r(\Omega)]^m},\end{aligned}$$

for $1<r\leq\min(s,3)$. In the scalar case ($m=1$) and uniform square meshes this has been proved in [5]; the extension of the error analysis presented in [5] to the case of $m>1$ is straightforward, while quasi-parallel meshes can be dealt with by following the analysis in [53]. At any rate, under these hypotheses and taking $r=2-\epsilon$, $\epsilon\in[0,1)$, it follows that

$$\left(\sum_{\kappa\in\mathcal{T}_h}|\eta_\kappa(\mathbf{u}_h)|^2\right)^{1/2} \to 0 \qquad \text{as } |h|\to 0,$$

and, therefore, the stopping criterion will be satisfied eventually as the mesh is refined.

The use of an *a priori* error bound to prove that the refinement algorithm terminates after a finite number of steps presupposes that the hypotheses under which the *a priori* error bound has been established are valid; in practice, this may be a restrictive requirement, and it is likely that termination will occur in circumstances which are less demanding than those in *a priori* error analysis.

5 Local considerations for steady problems

In the previous section we derived various *a posteriori* error bounds of the general form

$$\|\mathbf{u}-\mathbf{u}_h\|_{[H^{-1}(\Omega)]^m}\leq \text{Const.}\left(\sum_{\kappa\in\mathcal{T}_h}|\eta_\kappa(\mathbf{u}_h)|^2\right)^{1/2},$$

and we showed how reliable quantitative error control, to within a given tolerance, can be achieved through a feed-back process based on mesh adaptation. We also showed that, under mesh refinement, the local error indicator $\eta_\kappa(\mathbf{u}_h)$ must converge to zero. However, it is not clear from these global considerations *to what extent the reduction of the local error indicator on element κ contributes to the reduction of the global error $\mathbf{e}_h = \mathbf{u} - \mathbf{u}_h$ restricted to element κ?* In this section we shall approach this question from two different viewpoints:

1. We argue that, due to error propagation phenomena, the local error indicator $\eta_\kappa(\mathbf{u}_h)$ on a particular element κ controls only a portion of $\mathbf{e}_h|_\kappa$, namely a local quantity $\mathbf{e}_\kappa^{cell}$ called the *cell error*; error propagation is a non-local process and the complementary portion of the error, $\mathbf{e}_\kappa^{trans} = \mathbf{e}_h - \mathbf{e}_\kappa^{cell}$, called the *transmitted error*, does not obey a local bound.

2. Having shown that the local error indicator puts a bound only on part of the global error on each element, we prove that, at least for scalar problems, $\mathbf{e}_h$ restricted to element κ_0 can be bounded by the sum of local error indicators $\eta_\kappa(\mathbf{u}_h)$ over all κ that intersect the *domain of dependence* of κ_0.

The first viewpoint, analysed in Section 5.1 below, highlights the fact that by reducing the local residual $\mathbf{r}_h|_\kappa$ we reduce only the part of the global error which has been 'created' in κ, but not the part which has been 'transported' into κ. The second viewpoint, discussed in Section 5.2, shows that in order to reduce the whole of the global error in an element κ_0, we have to reduce the residual in each element κ whose *domain of influence* intersects κ_0.

5.1 What is controlled by the local residual?

This section is based on the papers [41] and [55]. Here we shall restrict ourselves to an overview of the main results; the reader is referred to these papers for further details.

Let us suppose that κ is a Lipschitz subdomain of Ω whose boundary $\partial\kappa$ is a non-characteristic hypersurface for the operator $\mathcal{L}$. The domain κ can be an element in the finite element partition of Ω or a union of neighbouring elements; we shall refer to κ as a *cell*. Throughout this section $(\cdot,\cdot)_\kappa$ will denote the inner product of the Hilbert space $[L_2(\kappa)]^m$, and $(\cdot,\cdot)_{\partial\kappa}$ will signify the inner product of $[L_2(\partial\kappa)]^m$.

On κ we consider the local boundary-value problem

$$\mathcal{L}\tilde{\mathbf{u}}_h = \mathbf{f} \quad \text{on } \kappa, \qquad \gamma_{B^-}(\tilde{\mathbf{u}}_h - \mathbf{u}_h) = \mathbf{0} \quad \text{on } \partial\kappa.$$

According to the theory outlined in Section 3.2, this problem has a unique strong solution $\tilde{\mathbf{u}}_h$; in fact, $\tilde{\mathbf{u}}_h$ can be thought of as a local solution of the partial differential equation (4.1) subject to a boundary condition whose data is a distortion of the correct local boundary data $B^-\mathbf{u}|_{\partial\kappa}$, due to the numerical error that has been 'created' outside the cell and is being advected into κ through the boundary $\partial\kappa$. We shall refer to the quantity

$$\mathbf{e}_\kappa^{cell} = \tilde{\mathbf{u}}_h - \mathbf{u}_h$$

as the *cell error*; clearly, $\mathbf{e}_\kappa^{cell}$ belongs to $D(\mathcal{L},\kappa)$ and it is the solution of the local boundary-value problem:

$$\mathcal{L}\mathbf{e}_\kappa^{cell} = \mathbf{r}_h \quad \text{on } \kappa, \qquad \gamma_{B^-}(\mathbf{e}_\kappa^{cell}) = \mathbf{0} \quad \text{on } \partial\kappa. \tag{5.1}$$

Thus $\mathbf{e}_\kappa^{cell}$ is governed by $\mathbf{r}_h|_\kappa$ and is not influenced by numerical effects which occur outside κ. The complementary quantity

$$\mathbf{e}_\kappa^{trans} = \mathbf{u} - \tilde{\mathbf{u}}_h,$$

called the *transmitted error*, represents the component of the global error $\mathbf{e}_h$ which has been created upwind of the cell κ and is merely advected into it. The transmitted error restricted to κ is the solution of the local problem

$$\mathcal{L}\mathbf{e}_\kappa^{trans} = \mathbf{0} \quad \text{on } \kappa, \qquad \gamma_{B^-}(\mathbf{e}_\kappa^{trans} - \mathbf{e}_h) = \mathbf{0} \quad \text{on } \partial\kappa.$$

We note that the concept of cell error is analogous to the notion of *local error* arising in the theory of numerical approximations to ordinary differential equations (see Hairer, Norsett and Wanner [22]).

With these definitions, we have the following decomposition of the global error:

$$\mathbf{e}_h|_\kappa = \mathbf{e}_\kappa^{cell} + \mathbf{e}_\kappa^{trans}.$$

Equation (5.1) shows that the local residual $\mathbf{r}_h|_\kappa$ is directly related to the 'locally created' part of the global error, $\mathbf{e}_\kappa^{cell}$, on cell κ.

Next we shall state sharp two-sided bounds on the cell error in terms of the cell residual; these will show that it is reasonable to attempt to improve the accuracy of the numerical method by reducing the size of the residual on those cells where it is largest. In order to simplify the presentation, we shall suppose that the centre of the coordinate system is the centroid (centre of gravity) of cell κ; if this is not the case, then the local exponential weight function $\exp\{-\alpha(\xi\cdot x)\}$ in Theorems 12 and 13 below should be replaced by $\exp\{-\alpha(\xi\cdot(x-x_c))\}$, where x_c is the centroid of cell κ; this ensures that the local weight function is close to 1 when $h = \mathrm{diam}(\kappa) << 1$.

Theorem 12 *We have the two-sided local error bound:*

$$\min_{x\in\kappa} w(x)\|\mathbf{r}_h\|_{[L_2(\kappa)]^m} \le |||\mathbf{e}_\kappa^{cell}|||_{\kappa,\xi} \le c_0'(\kappa)\max_{x\in\kappa} w(x)\|\mathbf{r}_h\|_{[L_2(\kappa)]^m}, \tag{5.2}$$

where $w(x) = \exp\{-\alpha(\xi\cdot x)\}$, $c_0'(\kappa) = (1+1/c_0(\kappa)^2)^{1/2}$, *and* $c_0(\kappa)$ *is the constant from condition (b) applied on the cell* κ *(clearly,* $c_0(\kappa)\ge C_0(\Omega)$ *for all* $\kappa\subset\Omega$*).*

Proof Recalling that $\mathbf{r}_h = \mathcal{L}\mathbf{e}_\kappa^{cell}$ on cell κ, the first inequality is a straightforward consequence of the definition of the norm $|||\cdot|||_{\kappa,\xi}$. The second inequality follows from the Gårding inequality (3.9) with Ω replaced by κ and $\mathbf{v} = \mathbf{e}_\kappa^{cell}$. ∎

It is intuitively clear that the global error $\mathbf{e}_h$ is non-local in character, and an error committed in certain part of the computational domain (say, near an inflow boundary,

for a scalar hyperbolic boundary-value problem) will be also felt in other parts of the domain. Thus, it is unreasonable to expect that the restriction of the global error to a cell is controllable merely in terms of the residual on that cell. This is, indeed, the case: in contrast with the local two-sided estimate obeyed by the cell error, the transmitted error satisfies only a *non-local* one-sided error bound; namely,

$$c_0(\kappa)\|w\mathbf{e}_\kappa^{trans}\|^2_{[L_2(\kappa)]^m} + (B^+ w\mathbf{e}_\kappa^{trans}, w\mathbf{e}_\kappa^{trans})_{\partial\kappa} \leq (-B^- w\mathbf{e}_\kappa^{trans}, w\mathbf{e}_\kappa^{trans})_{\partial\kappa},$$

where $w(x)$ is as in the previous theorem. We say that the bound is non-local because it involves $B^-\mathbf{e}_\kappa^{trans}|_{\partial\kappa}$, the 'incoming components' of $\mathbf{e}_\kappa^{trans}$ which have been created outside κ and are being transported into κ; from the point of view of an observer sitting in cell κ these are pollution effects from outside κ. The proof of this error bound is based on taking the L_2 inner product on κ of the equality $\mathcal{L}\mathbf{e}_\kappa^{trans} = \mathbf{0}$ with $w^2\mathbf{e}_\kappa^{trans}$, integrating by parts, and splitting the conormal trace operator into partial conormal traces.

Using a duality argument, it is possible to derive a local two-sided bound on the L_2 norm of the cell error in terms of the dual graph norm of the residual, the dual graph norm $|||\cdot|||'_{*,\kappa,\xi}$ being defined by

$$|||\mathbf{w}|||'_{*,\kappa,\xi} = \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{(\mathbf{w},\phi)_\kappa}{|||\phi|||_{*,\kappa,\xi}}.$$

Theorem 13 *Suppose that $\mathbf{e}_\kappa^{cell} \in [H^1(\kappa)]^m$; then, we have the two-sided local error bound:*

$$\min_{x\in\kappa} \hat{w}(x)|||\mathbf{r}_h|||'_{*,\kappa,\xi} \leq \|\mathbf{e}_\kappa^{cell}\|_{[L_2(\kappa)]^m} \leq c_0'(\kappa) \max_{x\in\kappa} \hat{w}(x)|||\mathbf{r}_h|||'_{*,\kappa,\xi}, \tag{5.3}$$

where $\hat{w}(x) = 1/w(x)$, and $w(x)$ and $c_0'(\kappa)$ are as in the previous theorem.

Proof Recalling the definition of the dual graph norm $|||\cdot|||'_{*,\kappa}$, we have that

$$\begin{aligned}
|||\mathbf{r}_h|||'_{*,\kappa,\xi} &= |||\mathcal{L}\mathbf{e}_\kappa^{cell}|||'_{*,\kappa,\xi} = \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{(\mathcal{L}\mathbf{e}_\kappa^{cell},\phi)_\kappa}{|||\phi|||_{*,\kappa,\xi}} = \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{(\mathbf{e}_\kappa^{cell},\mathcal{L}^*\phi)_\kappa}{|||\phi|||_{*,\kappa,\xi}} \\
&\leq \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{\|\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{e}_\kappa^{cell}\|_{[L_2(\kappa)]^m}\|\mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\phi\|_{[L_2(\kappa)]^m}}{|||\phi|||_{*,\kappa,\xi}} \\
&\leq \|\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{e}_\kappa^{cell}\|_{[L_2(\kappa)]^m} \leq \max_{x\in\kappa} w(x)\|\mathbf{e}_\kappa^{cell}\|_{[L_2(\kappa)]^m}.
\end{aligned}$$

Hence the first inequality. To prove the second inequality, we consider the local adjoint boundary-value problem

$$\mathcal{L}^*\varphi = \mathrm{e}^{-2\alpha(\xi\cdot x)}\mathbf{e}_\kappa^{cell} \quad \text{on } \kappa, \qquad B^+\varphi = \mathbf{0} \quad \text{on } \partial\kappa,$$

and we note that the corresponding (unique) solution φ belongs to $D(L^*,\kappa)$. Now since $\mathbf{e}_\kappa^{cell} \in D(L,\kappa)\cap[H^1(\kappa)]^m$, upon integration by parts we have that

$$|||\mathbf{r}_h|||'_{*,\kappa,\xi} = \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{(\mathcal{L}\mathbf{e}_\kappa^{cell},\phi)_\kappa}{|||\phi|||_{*,\kappa,\xi}} \geq \frac{(\mathcal{L}\mathbf{e}_\kappa^{cell},\varphi)_\kappa}{|||\varphi|||_{*,\kappa,\xi}} = \frac{(\mathbf{e}_\kappa^{cell},\mathcal{L}^*\varphi)_\kappa}{|||\varphi|||_{*,\kappa,\xi}}$$

$$= \frac{(\mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\varphi, \mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\varphi)_\kappa}{|||\varphi|||_{*,\kappa,\xi}} = \frac{\|\mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\,\varphi\|^2_{[L_2(\kappa)]^m}}{|||\varphi|||_{*,\kappa},\,\xi}$$

$$\geq \frac{1}{c_0'(\kappa)}\|\mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\,\varphi\|_{[L_2(\kappa)]^m} = \frac{1}{c_0'(\kappa)}\|\mathrm{e}^{-\alpha(\xi\cdot x)}\mathbf{e}_\kappa^{cell}\|_{[L_2(\kappa)]^m}.$$

That completes the proof. ■

The *a posteriori* bounds (5.2) and (5.3) provide sharp estimates of the cell error; unfortunately, the dual graph norm of the residual is difficult to compute in practice since its definition involves a supremum over the infinite set $D(\mathcal{L}^*,\kappa)$ (although, in [51] the dual graph norm $|||\cdot|||'_{*,\kappa,\xi}$ was approximated by partitioning the cell κ and considering the supremum over a finite dimensional subspace of $D(\mathcal{L}^*,\kappa)$ consisting of piecewise linear functions on such micro-partitions; this approximation was then successfully implemented into an adaptive finite volume algorithm for the numerical solution of the Euler equations of compressible gas dynamics in two space dimensions). In addition to the fact that the dual graph norm is unattractive from the computational point of view, it is not clear at this stage how (5.3) relates to the *a posteriori* error bounds established in the previous section which involved $\|h\mathbf{r}_h\|_{[L_2(\kappa)]^m}$ instead of $|||\mathbf{r}_h|||'_{*,\kappa,\xi}$. Our aim now is to resolve these issues by showing that $|||\mathbf{r}_h|||'_{*,\kappa,\xi}$ can be further bounded above by a constant multiple of $\|h\mathbf{r}_h\|_{[L_2(\kappa)]^m}$, a quantity that is simple and cheap to compute; we shall also prove that there is a similar lower bound. Thus we shall obtain a local two-sided bound on the L_2 norm of the cell error in terms of the L_2 norm of the finite element residual $\mathbf{r}_h$ scaled by the local mesh size. These results will establish a connection between the global *a posteriori* bounds of Section 4 and the local estimates on the cell error described earlier in this section.

The theory of error estimation that we described so far is valid for any symmetric positive system, irrespective of its type. In order to proceed, we shall replace the positivity condition (b) by a stronger hypothesis, thereby restricting ourselves to symmetric hyperbolic systems. Namely, we shall suppose the following:

(b') There exists $\xi \in \mathbb{R}^n$ such that the matrix $\sum_{i=1}^n \xi_i A_i$ is positive definite, uniformly on $\bar{\Omega}$; i.e. there is a positive constant $c_0 = c_0(\Omega)$, such that

$$\sum_{i=1}^n \xi_i A_i(x) \geq c_0 I \qquad \text{for all } x \in \bar{\Omega}.$$

This condition is referred to as *hyperbolicity in the sense of Lax* (see [35]). In the rest of this section we shall assume that (b') holds instead of (b).

Theorem 14 *Suppose that* $\mathbf{e}_\kappa^{cell} \in [H^1(\kappa)]^m$*; then we have the following one-sided* a posteriori *error bound on the cell error:*

$$\|\mathbf{e}_\kappa^{cell}\|_{[L_2(\kappa)]^m} \leq c_3(\kappa)\|h\mathbf{r}_h\|_{[L_2(\kappa)]^m},$$

where

$$\begin{aligned} c_3(\kappa) &= c_2(\kappa)(1+1/c_0(\kappa)^2)^{1/2}\exp(\alpha(1+h)|\xi|), \\ c_2(\kappa) &= (h^2+c_0(\kappa)^2/4)^{-1/2}\exp(\alpha|\xi|), \end{aligned}$$

$c_0(\kappa)$ is the constant from condition (b') applied on the cell κ and h denotes the diameter of κ.

Proof Let us choose $\zeta \in \mathbb{R}^n$ (to be fixed later on) and let κ be any element in the partition of Ω. In order to simplify the presentation we shall assume that the origin of the coordinate system is the centroid (centre of gravity) of κ; if this is not the case then the local exponential weight-function $\exp(\alpha(\zeta\cdot x))$ in the expressions below should be replaced by $\exp(\alpha(\zeta\cdot(x-x_c)))$ where x_c is the centroid of κ, so that the local weight function remains bounded on κ as $h=\mathrm{diam}(\kappa)$ converges to zero. The proof consists of two parts. First we prove the local Gårding inequality stated in (5.8) below. In the second part of the proof, we use this inequality to show that the dual graph norm of the residual, $|||\mathbf{r}_h|||'_{*,\kappa,\xi}$, is bounded above by a constant multiple of $||h\mathbf{r}_h||_{[L_2(\kappa)]^m}$; this, together with the second inequality in (5.3) will yield the desired result.

Part 1: Given that ϕ is an element of $D(\mathcal{L}^*,\kappa)\cap[H^1(\kappa)]^m$, we have that

$$\begin{aligned} \int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}\left(-\sum_{i=1}^n A_i\frac{\partial\phi}{\partial x_i}+C^*\phi\right)\cdot\phi\,\mathrm{d}x &= -\frac{1}{2}\int_{\partial\kappa}\mathrm{e}^{2\alpha(\zeta\cdot x)}\phi\cdot(B\phi)\,\mathrm{d}s \\ &\quad+\int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}\phi\cdot\left(C+\alpha\sum_{i=1}^n\zeta_iA_i+\frac{1}{2}\sum_{i=1}^n\frac{\partial A_i}{\partial x_i}\right)\phi\,\mathrm{d}x. \end{aligned} \tag{5.4}$$

Since we are dealing with real-valued functions,

$$(\mathrm{e}^{2\alpha(\zeta\cdot x)}\phi, C\phi)_\kappa = (\mathrm{e}^{2\alpha(\zeta\cdot x)}C^*\phi,\phi)_\kappa = (\mathrm{e}^{2\alpha(\zeta\cdot x)}\phi, C^*\phi)_\kappa.$$

Applying this identity in the second integral on the right-hand side of (5.4) gives

$$\begin{aligned} \int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}\left(-\sum_{i=1}^n A_i\frac{\partial\phi}{\partial x_i}+C^*\phi\right)\cdot\phi\,\mathrm{d}x &= -\frac{1}{2}\int_{\partial\kappa}\mathrm{e}^{2\alpha(\zeta\cdot x)}\phi\cdot(B\phi)\,\mathrm{d}s \\ &\quad+\int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}\phi\cdot\left(C^*+\alpha\sum_{i=1}^n\zeta_iA_i+\frac{1}{2}\sum_{i=1}^n\frac{\partial A_i}{\partial x_i}\right)\phi\,\mathrm{d}x. \end{aligned} \tag{5.5}$$

Adding (5.4) and (5.5), and noting that $B=B^++B^-$ with $B^+\phi=\mathbf{0}$ on $\partial\kappa$ and $-\phi\cdot(B^-\phi)\geq 0$ on $\partial\kappa$, we deduce that

$$\int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}\mathcal{L}^*\phi\cdot\phi\,\mathrm{d}x \geq \int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}\phi\cdot\frac{1}{2}(K_\zeta(x)+K_\zeta^*(x))\phi\,\mathrm{d}x. \tag{5.6}$$

The remainder of Part 1 of the proof is devoted to showing that the matrix $K_\zeta(x)+K_\zeta^*(x)$ is positive definite, uniformly on κ, and that the inequality (5.7) below holds. Substituting (5.7) into (5.6) will then yield the Gårding inequality (5.8). Let us therefore consider

$$\frac{1}{2}(K_\zeta(x)+K_\zeta^*(x)) = \frac{1}{2}(C+C^*)+\frac{1}{2}\sum_{i=1}^n\frac{\partial A_i}{\partial x_i}+\alpha\sum_{i=1}^n\zeta_iA_i.$$

So far ζ has been an arbitrary vector from $\mathbb{R}^n$; now (as promised at the beginning of the proof) we fix its value and take

$$\zeta_i = h^{-1}\xi_i, \quad i = 1, \ldots, n, \qquad \text{where } h = \operatorname{diam}(\kappa).$$

Applying hypothesis (b') we have that

$$\frac{1}{2}(K_\zeta(x) + K_\zeta^*(x)) \geq h^{-1}\left(\alpha c_0(\kappa) I + h\left(\frac{1}{2}(C + C^*) + \frac{1}{2}\sum_{i=1}^{n}\frac{\partial A_i}{\partial x_i}\right)\right).$$

Since, by assumption, the entries of C and $\partial A_i/\partial x_i$ belong to $C(\overline{\Omega})$, it follows that, for h sufficiently small and all $x \in \kappa$,

$$\frac{1}{2}(K_\zeta(x) + K_\zeta^*(x)) \geq \frac{\alpha c_0(\kappa)}{2h} I. \tag{5.7}$$

Substituting (5.7) into (5.6) gives the local Gårding inequality

$$\int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}\mathcal{L}^*\phi \cdot \phi \,\mathrm{d}x \geq \frac{\alpha c_0(\kappa)}{2h}\int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}|\phi|^2 \,\mathrm{d}x \tag{5.8}$$

for all $\phi \in D(\mathcal{L}^*, \kappa) \cap [H^1(\Omega)]^m$, and by density also for all $\phi \in D(\mathcal{L}^*, \kappa)$. Further, applying the Cauchy-Schwarz inequality to the left-hand side of (5.8), it follows that

$$\left(\int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}|\phi|^2 \,\mathrm{d}x\right)^{\frac{1}{2}} \leq \frac{2h}{\alpha c_0(\kappa)}\left(\int_\kappa \mathrm{e}^{2\alpha(\zeta\cdot x)}|\mathcal{L}^*\phi|^2 \,\mathrm{d}x\right)^{\frac{1}{2}} \tag{5.9}$$

for all ϕ in $D(\mathcal{L}^*, \kappa)$. This completes the first part of the proof.

Part 2: Now we use (5.9) to show that the dual graph norm of $\mathbf{r}_h$ is bounded above by a constant multiple of $\|h\mathbf{r}_h\|_{[L_2(\kappa)]^m}$; indeed,

$$\begin{aligned}
|||\mathbf{r}_h|||'_{*,\kappa,\zeta} &= \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{|(\mathbf{r}_h, \phi)_\kappa|}{(\|\mathrm{e}^{\alpha(\zeta\cdot x)}\phi\|^2_{[L_2(\kappa)]^m} + \|\mathrm{e}^{\alpha(\zeta\cdot x)}\mathcal{L}^*\phi\|^2_{[L_2(\kappa)]^m})^{\frac{1}{2}}} \\
&\leq \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{\|\mathrm{e}^{-\alpha(\zeta\cdot x)}\mathbf{r}_h\|_{[L_2(\kappa)]^m}\|\mathrm{e}^{\alpha(\zeta\cdot x)}\phi\|_{[L_2(\kappa)]^m}}{(\|\mathrm{e}^{\alpha(\zeta\cdot x)}\phi\|^2_{[L_2(\kappa)]^m} + \|\mathrm{e}^{\alpha(\zeta\cdot x)}\mathcal{L}^*\phi\|^2_{[L_2(\kappa)]^m})^{\frac{1}{2}}} \\
&\leq h(h^2 + \alpha^2 c_0(\kappa)^2/4)^{-1/2}\|\mathrm{e}^{-\alpha(\zeta\cdot x)}\mathbf{r}_h\|_{[L_2(\kappa)]^m}.
\end{aligned} \tag{5.10}$$

This is essentially the desired bound on the dual graph norm of the residual, except that the left-hand side includes $|||\mathbf{r}_h|||'_{*,\kappa,\zeta}$ instead of $|||\mathbf{r}_h|||'_{*,\kappa,\xi}$, and an exponential term appears under the norm sign on the right. The concluding part of the proof shows that this exponential term is bounded independent of h and that the norms $|||\cdot|||'_{*,\kappa,\zeta}$ and $|||\cdot|||'_{*,\kappa,\xi}$ are equivalent.

As $|\alpha(\zeta\cdot x)| = h^{-1}\alpha|\xi\cdot x|$ with $h(<1)$ denoting the diameter of κ, upon recalling that the origin of the coordinate system is at the centroid of κ, it follows that

$$|\alpha(\zeta\cdot x)| \leq \alpha|\xi|. \tag{5.11}$$

Substituting (5.11) into (5.10) gives

$$|||\mathbf{r}_h|||'_{*,\kappa,\zeta} \leq c_2(\kappa)\|h\mathbf{r}_h\|_{[L_2(\kappa)]^m}, \tag{5.12}$$

where $c_2(\kappa) = \mathrm{e}^{\alpha|\xi|}(h^2 + \alpha^2 c_0(\kappa)^2/4)^{-1/2}$. Now let us note that, with

$$|||\phi|||_{*,\kappa,\xi} = (\|\mathrm{e}^{\alpha(\xi\cdot x)}\phi\|^2_{[L_2(\kappa)]^m} + \|\mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\phi\|^2_{[L_2(\kappa)]^m})^{\frac{1}{2}}$$

and

$$\mathrm{e}^{\alpha(\xi\cdot x)} = \mathrm{e}^{\alpha(\zeta\cdot x)}\,\mathrm{e}^{-\alpha(\zeta-\xi)\cdot x} = \mathrm{e}^{\alpha(\zeta\cdot x)}\,\mathrm{e}^{-\alpha\xi(1-h)\cdot(x/h)},$$

we have that

$$\mathrm{e}^{-\alpha|\xi|}|||\phi|||_{*,\kappa,\zeta} \le |||\phi|||_{*,\kappa,\xi}.$$

Consequently,

$$|||\mathbf{r}_h|||'_{*,\kappa,\xi} = \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{|(\mathbf{r}_h,\phi)_\kappa|}{|||\phi|||_{*,\kappa,\xi}} \le \mathrm{e}^{\alpha|\xi|}|||\mathbf{r}_h|||'_{*,\kappa,\zeta}. \tag{5.13}$$

Finally, we recall the second inequality in (5.3),

$$\|\mathbf{e}^{cell}_\kappa\|_{[L_2(\kappa)]^m} \le c_0'(\kappa)\max_{x\in\kappa}\mathrm{e}^{\alpha(\xi\cdot x)}|||\mathbf{r}_h|||'_{*,\kappa,\xi},$$

and combine this with (5.12) and (5.13) to deduce that

$$\|\mathbf{e}^{cell}_\kappa\|_{[L_2(\kappa)]^m} \le c_0'(\kappa)c_2(\kappa)\mathrm{e}^{\alpha(1+h)|\xi|}\|h\mathbf{r}_h\|_{[L_2(\kappa)]^m},$$

and hence the desired bound. ∎

Theorem 14 provides an upper bound on the L_2 norm of the cell error, analogous to the second inequality in (5.3). Now we prove a lower bound on the L_2 norm of the cell error, similar to the first inequality in (5.3). To do so, we consider a uniformly regular family of micro-partitions of the cell κ, and let $S_{\hat{h}} = S_{\hat{h}}(\kappa)$ be a finite element subspace of $D(\mathcal{L}^*,\kappa)$ on such a micro-partition; here $\hat{h} = \hat{h}(\kappa)$ denotes the maximum diameter of elements in the micro-partition. We denote by $P_{\hat{h}}$ the orthogonal projector in $[L_2(\kappa)]^m$ onto the finite element space $S_{\hat{h}}$.

Theorem 15 *We have the following* a posteriori *lower bound on the cell error:*

$$c_4(\kappa)\|\hat{h}P_{\hat{h}}\mathbf{r}_h\|_{[L_2(\kappa)]^m} \le \|e^{cell}_\kappa\|_{[L_2(\kappa)]^m},$$

where $c_4(\kappa)$ is a computable constant.

Proof According to the definition of the dual graph norm, we have that

$$\begin{aligned}\|\mathbf{r}_h\|'_{*,\kappa,\xi} &= \sup_{\phi\in D(\mathcal{L}^*,\kappa)} \frac{|(\mathbf{r}_h,\phi)_\kappa|}{(\|\mathrm{e}^{\alpha(\xi\cdot x)}\phi\|^2_{[L_2(\kappa)]^m} + \|\mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\phi\|^2_{[L_2(\kappa)]^m})^{\frac{1}{2}}} \\ &\ge \sup_{\phi_{\hat{h}}\in S_{\hat{h}}} \frac{|(\mathbf{r}_h,\phi_{\hat{h}})_\kappa|}{(\|\mathrm{e}^{\alpha(\xi\cdot x)}\phi_{\hat{h}}\|^2_{[L_2(\kappa)]^m} + \|\mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\phi_{\hat{h}}\|^2_{[L_2(\kappa)]^m})^{\frac{1}{2}}},\end{aligned} \tag{5.14}$$

where we made use of the fact that $D(\mathcal{L}^*,\kappa) \supset S_{\hat{h}}(\kappa)$. Recalling that the family of micro-partitions of κ has been assumed uniformly regular, we can apply the standard inverse inequality (see [12])

$$\left(\sum_{i=1}^n \|\mathrm{e}^{\alpha(\xi\cdot x)}\frac{\partial\phi_{\hat{h}}}{\partial x_i}\|^2_{[L_2(\kappa)]^m}\right)^{\frac{1}{2}} \le \frac{c_5(\kappa)}{\hat{h}}\|\mathrm{e}^{\alpha(\xi\cdot x)}\phi_{\hat{h}}\|_{[L_2(\kappa)]^m}, \qquad \phi_{\hat{h}}\in S_{\hat{h}},$$

to deduce that

$$\|\mathrm{e}^{\alpha(\xi\cdot x)}\mathcal{L}^*\phi_{\hat{h}}\|_{[L_2(\kappa)]^m} \leq \hat{h}^{-1}c_6(\kappa)\|\mathrm{e}^{\alpha(\xi\cdot x)}\phi_{\hat{h}}\|_{[L_2(\kappa)]^m},$$

where

$$c_6(\kappa) = \hat{h}\|C\|_{[L_\infty(\kappa)]^{m\times m}} + c_5(\kappa)\left(\sum_{i=1}^{n}\|A_i\|^2_{[L_\infty(\kappa)]^{m\times m}}\right)^{\frac{1}{2}}.$$

Therefore,

$$|||\phi_{\hat{h}}|||_{*,\kappa,\xi} \leq \hat{h}^{-1}(\hat{h}^2 + c_6(\kappa)^2)^{1/2}\|\mathrm{e}^{\alpha(\xi\cdot x)}\phi_{\hat{h}}\|_{[L_2(\kappa)]^m}$$

for all $\phi_{\hat{h}} \in S_{\hat{h}}$. Substituting this inequality into (5.14) gives

$$\begin{aligned}\|\mathbf{r}_h\|'_{*,\kappa,\xi} &\geq \hat{h}(\hat{h}^2 + c_6(\kappa)^2)^{-1/2}\sup_{\phi_{\hat{h}}\in S_{\hat{h}}}\frac{|(\mathbf{r}_h,\phi_{\hat{h}})_\kappa|}{\|\mathrm{e}^{\alpha(\xi\cdot x)}\phi_{\hat{h}}\|_{[L_2(\kappa)]^m}}\\ &= \mathrm{e}^{-\alpha h|\xi|}(\hat{h}^2 + c_6(\kappa)^2)^{-\frac{1}{2}}\|\hat{h}P_{\hat{h}}\mathbf{r}_h\|_{[L_2(\kappa)]^m}.\end{aligned}$$

Combining this result with the first inequality of (5.3) we obtain the desired lower bound on the cell error. ■

The upper bound stated in Theorem 14 and the lower bound from Theorem 15 can be coupled into a single two-sided bound on the L_2 norm of the cell error; namely,

$$c_4(\kappa)\|\hat{h}P_{\hat{h}}\mathbf{r}_h\|_{[L_2(\kappa)]^m} \leq \|\mathbf{e}_\kappa^{cell}\|_{[L_2(\kappa)]^m} \leq c_3(\kappa)\|h\mathbf{r}_h\|_{[L_2(\kappa)]^m}, \tag{5.15}$$

where $c_3(\kappa)$ and $c_4(\kappa)$ are computable constants. We note here that unlike the sharp two-sided bound on the L_2 norm of the cell error in terms of the dual graph norm of the finite element residual given in Theorem 13, the two-sided bound (5.15) is not asymptotically sharp because of the mismatch between the expressions under the norm signs in the lower and the upper estimate. Nevertheless, (5.15) is 'almost sharp' in the following sense: in practice $\hat{h}$ can be chosen to be a *fixed fraction* of h such that

$$\|\mathbf{r}_h - P_{\hat{h}}\mathbf{r}_h\|_{[L_2(\kappa)]^m} \leq \epsilon\|\mathbf{r}_h\|_{[L_2(\kappa)]^m},$$

where $\epsilon \in (0,1)$ is a fixed real number; then, by Pythagoras' Theorem,

$$\|\hat{h}P_{\hat{h}}\mathbf{r}_h\|_{[L_2(\kappa)]^m} \geq (1-\epsilon^2)^{1/2}\|\hat{h}\mathbf{r}_h\|_{[L_2(\kappa)]^m}.$$

This indicates that the lower bound in inequality (5.15) is at least 'comparable' in form, if not in size, with the upper bound.

To conclude, inequality (5.15) shows that in adaptive mesh refinement processes driven by residual-based error indicators, such as the ones listed at the beginning of Section 4.5, only cells with large cell error will be flagged for refinement (ignoring boundary conditions); the complementary part of the global error on a cell cannot be detected by measuring the residual on that cell only; to achieve local error control, a more global bound is required. This is the theme of the next section.

5.2 What controls the local size of the global error?

In the previous section we showed that due to pollution effects the finite element residual restricted to a cell puts a bound only on part of the global error restricted to that cell, and we derived local bounds on this part of the global error in terms of the residual. Here we show that in order to bound the whole of the global error restricted to cell κ we have to involve residuals over all the cells which intersect the domain of dependence of cell κ: the resulting *a posteriori* error bound is non-local in nature. The results presented in this section are based on the paper [27]. For the sake of simplicity we focus on scalar hyperbolic equations, corresponding to $m = 1$, and consider the following boundary-value problem:

$$\mathcal{L}u \equiv \nabla \cdot (\mathbf{a}u) + cu = f \quad \text{in } \Omega, \qquad\qquad (\nu \cdot \mathbf{a})^- u\big|_{\partial\Omega} = 0,$$

in a convex Lipschitz domain $\Omega \subset \mathbb{R}^n$, where $(x)^- = \min(x, 0)$ denotes the negative part of the number x. We shall suppose that $\mathbf{a} = (a_1, \ldots, a_n)$ has its components in $C^2(\bar{\Omega})$ and that $c \in C^1(\bar{\Omega})$. The function f will be assumed to be in $L_2(\Omega)$. Recalling the definition of the inflow boundary $\partial_-\Omega$ from Section 3.1, we can restate the problem as follows:

$$\nabla \cdot (\mathbf{a}u) + cu = f \quad \text{in } \Omega, \qquad\qquad u = 0 \quad \text{on } \partial_-\Omega.$$

The adjoint operator $\mathcal{L}^*$ of $\mathcal{L}$ on Ω is defined by

$$\mathcal{L}^* z \equiv -\mathbf{a} \cdot \nabla z + cz.$$

Suppose that κ is an element in the partition of Ω. Let $D(\kappa)$ denote the union of all forward (with respect to $\partial_+\Omega$) characteristics of $\mathcal{L}^*$ contained in Ω which emanate from κ. Equivalently, $D(\kappa)$ can be described as the union of all backward (with respect to $\partial_-\Omega$) characteristics of $\mathcal{L}$ contained in Ω which emanate from κ. Further, we denote by $D_h(\kappa)$ the set of all elements in the partition which intersect $D(\kappa)$.

Similarly as in Section 4.2, we shall consider two situations, referred to symbolically as α) and β), corresponding to a Galerkin finite element method with strongly and weakly imposed boundary conditions, respectively.

α) In this case, we have the following result (note that since we are dealing with the scalar case hypothesis (a) is redundant).

Theorem 16 *Suppose that (b) and (c) hold with $m = 1$, that the entries of $\mathbf{a}$ are in $C^2(\bar{\Omega})$ and that $c \in C^1(\bar{\Omega})$; then*

$$\|u - u_h\|_{H^{-1}(\kappa)} \leq C_1' C_2 \|h r_h\|_{L_2(D_h(\kappa))},$$

where $r_h = f - \mathcal{L}u_h$ denotes the residual corresponding to the finite element approximation u_h to u.

Proof For $\psi \in C_0^\infty(\kappa)$ consider the adjoint (dual) problem

$$\mathcal{L}^* z = \psi \qquad \text{on } \Omega, \qquad\qquad (\nu \cdot \mathbf{a})^+ z\big|_{\partial\Omega} = 0,$$

where $(x)^+ = \max(x, 0)$ denotes the positive part of the number x. Since ψ has compact support in κ, characteristic theory implies that the support of z is contained in $D(\kappa)$. Let $z_h \in Y_h$ denote the quasi-interpolant of z (see [9], [11]); then, by Galerkin orthogonality, we have that

$$(u - u_h, \psi) = (r_h, z - z_h).$$

Further, since the support of z_h is contained in $D_h(\kappa)$ and $D_h(\kappa) \supset D(\kappa)$, it follows that

$$(u - u_h, \psi) = (r_h, z - z_h)_{D_h(\kappa)}.$$

Applying the approximation property (c) (with $m = 1$) and noting that $D_h(\kappa) \subset \Omega$ gives

$$|(u - u_h, \psi)| \leq C_2 \|hr_h\|_{L_2(D_h(\kappa))} \|z\|_{H^1(\Omega)}.$$

Now, recalling the strong stability of the dual problem,

$$\|z\|_{H^1(\Omega)} \leq C_1' \|\psi\|_{H^1(\Omega)} = C_1' \|\psi\|_{H^1(\kappa)}.$$

Hence, for any $\psi \in C_0^\infty(\kappa)$,

$$|(u - u_h, \psi)| \leq C_1' C_2 \|hr_h\|_{L_2(D_h(\kappa))} \|\psi\|_{H^1(\kappa)}.$$

Dividing both sides of this inequality by $\|\psi\|_{H^1(\kappa)}$ and taking the supremum over all ψ in $C_0^\infty(\kappa)$, upon noting that $C_0^\infty(\kappa)$ is dense in $H_0^1(\kappa)$ and recalling the definition of the negative Sobolev norm $\|\cdot\|_{H^{-1}(\kappa)}$, we arrive at the desired *a posteriori* error bound. ∎

In the previous section we decomposed the global error e, restricted to cell κ, as

$$e|_\kappa = e_\kappa^{cell} + e_\kappa^{trans},$$

and we gave, in Theorem 14, a bound on the L_2 norm of the cell error in terms of the local finite element residual, which in the case of a scalar hyperbolic equation ($m = 1$) has the following form:

$$\|e_\kappa^{cell}\|_{L_2(\kappa)} \leq c_3(\kappa) \|hr_h\|_{L_2(\kappa)}.$$

We also noted that the L_2 norm of the transmitted error on κ can be bounded by the L_2 norm of the incoming component of the transmitted error on $\partial\kappa$. Here we show a further bound on the transmitted error which is closer in spirit to that in Theorem 16.

Theorem 17 *Suppose that (b) and (c) hold with $m = 1$, that the entries of* $\mathbf{a}$ *belong to $C^2(\bar{\Omega})$ and that $c \in C^1(\bar{\Omega})$; also suppose that $e_\kappa^{cell} \in H^1(\kappa)$. Then*

$$\|e_\kappa^{trans}\|_{H^{-1}(\kappa)} \leq (C_1' C_2 + c_3(\kappa)) \|hr_h\|_{L_2(D_h(\kappa))}.$$

Proof Writing $e_\kappa^{trans} = e|_\kappa - e_\kappa^{cell}$ and applying the triangle inequality for the $\|\cdot\|_{H^{-1}(\kappa)}$ norm, we have that

$$\|e_\kappa^{trans}\|_{H^{-1}(\kappa)} \leq \|e\|_{H^{-1}(\kappa)} + \|e_\kappa^{cell}\|_{H^{-1}(\kappa)}. \tag{5.16}$$

According to Theorem 16,

$$\|e\|_{H^{-1}(\kappa)} \leq C_1' C_2 \|hr_h\|_{L_2(D_h(\kappa))}. \tag{5.17}$$

Further, by the definition of the $H^{-1}(\kappa)$ norm and recalling Theorem 14,

$$\|e_\kappa^{cell}\|_{H^{-1}(\kappa)} \leq \|e_\kappa^{cell}\|_{L_2(\kappa)} \leq c_3(\kappa)\|hr_h\|_{L_2(\kappa)}. \tag{5.18}$$

Substituting (5.17) and (5.18) into (5.16) and noting that $\kappa \subset D_h(\kappa)$ we complete the proof. ■

β) We consider the scalar hyperbolic equation

$$\nabla \cdot (\mathbf{a}u) + cu = f \qquad \text{in } \Omega,$$

subject to the non-homogeneous boundary condition

$$(\mathbf{a} \cdot \nu)^-(u - g) = 0 \qquad \text{on } \partial\Omega,$$

with the same hypotheses on $\mathbf{a}$, c, f and Ω as in case α) above; further, we suppose that $g \in L_2(\partial\Omega)$. In the case of a Petrov-Galerkin finite element approximation with weakly imposed boundary condition we have the following *a posteriori* error bound on the global error restricted to element κ in terms of the internal and boundary residual.

Theorem 18 *Suppose that hypotheses (b) and (c') hold, that the entries of* $\mathbf{a}$ *are in* $C^2(\bar{\Omega})$ *and that* c *belongs to* $C^1(\bar{\Omega})$. *Then,*

$$\|u - u_h\|_{H^{-1}(\kappa)} \leq C_1' C_2 \left(\|hr_h\|_{L_2(D_h(\kappa))} + \|h^{1/2} r_h^-\|_{L_2(\partial\Omega \cap D_h(\kappa))} \right),$$

where $r_h = f - \mathcal{L}u_h$ *denotes the interior residual, and* $r_h^- = (\mathbf{a} \cdot \nu)^-(g - u_h)|_{\partial\Omega}$ *signifies the boundary residual.*

We shall omit the proof, as it can be easily reconstructed from the proofs of Theorems 9 and 16. A similar result can be shown for the streamline diffusion finite element approximation of the scalar hyperbolic problem, and a bound on the transmitted error akin to that in Theorem 17 can be established.

Figure 2 shows the qualitative behaviour of the different error components in the numerical solution of the model problem (4.10) using the streamline diffusion method on an unstructured triangular mesh consisting of 504 nodes and 926 elements. In the figure caption $\mathcal{E}_1(u_h, h)$ and $\mathcal{E}_2(u_h, h)$ denote the right-hand sides of the error bounds in Theorems 14 and 18, respectively.

6 A posteriori error estimation for functionals

It is frequently the case in engineering problems that the main quantity of concern is not the solution of a partial differential equation, but a derived quantity which can be though of as a functional of the solution. In such instances it is unlikely that *a posteriori* error bounds of the kind stated in Section 4 will be of use in the design of efficient adaptive algorithms.

Our aim in this section is to propose an approach to the derivation of *a posteriori* bounds on the error in linear functionals directly, without attempting to obtain an upper bound on the error in a *norm* in which the functional is bounded. In order to illustrate

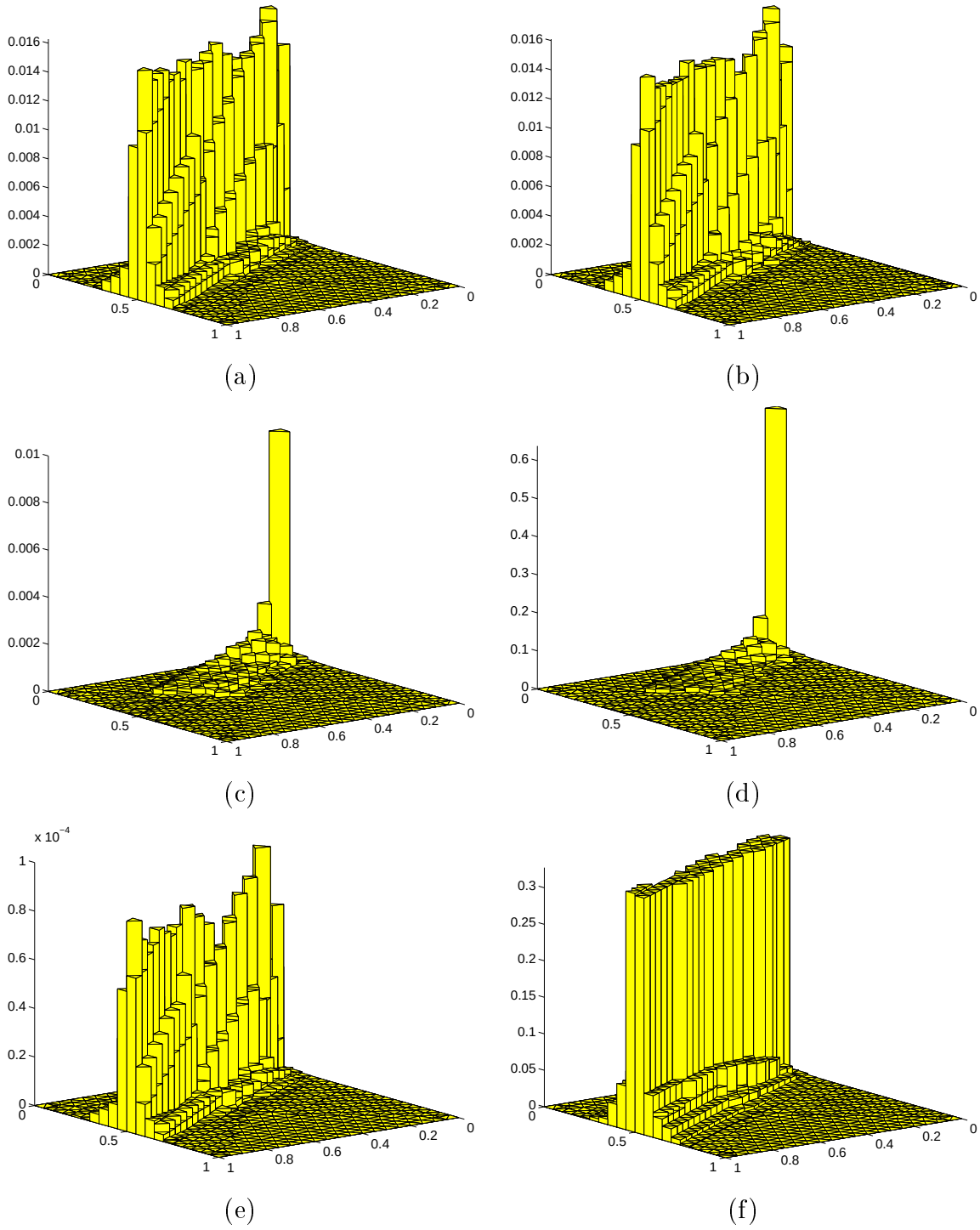


Figure 2: Corner discontinuity problem: (a) $\|e\|_{L^2(\kappa)}$; (b) $\|e^{trans}\|_{L^2(\kappa)}$; (c) $\|e^{cell}\|_{L^2(\kappa)}$; (d) $\mathcal{E}_1(u_h, h)$; (e) $\|e\|_{H^{-1}(\kappa)}$; (f) $\mathcal{E}_2(u_h, h)$.

the key ideas we begin by discussing some specific examples: in the next subsection we consider the *a posteriori* error analysis for a particular linear functional, the normal flux through the boundary of the computational domain; in the second subsection, we present a similar analysis for the local weighted average of the solution. In the third subsection, we approach the problem of *a posteriori* error estimation for linear functionals from a general viewpoint, following the ideas of Becker and Rannacher [7].

6.1 Estimation of the normal flux through the boundary

Let us consider the non-homogeneous boundary-value problem (4.1), (4.3), where $\mathbf{f} \in [L_2(\Omega)]^m$ and $\mathbf{g} \in [H^1(\Omega)]^m$. For the sake of simplicity, we assume that the restriction of $\mathbf{g}$ to $\partial\Omega$ belongs to the restriction of the trial space $X_h \subset H(\mathcal{L}, \Omega)$ to $\partial\Omega$, and the test space Y_h is contained in $[L_2(\Omega)]^m$. We approximate the non-homogeneous problem by the following method: find $\mathbf{u}_h$ in X_h such that

$$\begin{aligned} \Pi_h \mathcal{L}\mathbf{u}_h &= \Pi_h \mathbf{f} \qquad \text{in } \Omega, \\ B^-(\mathbf{u}_h - \mathbf{g})\big|_{\partial\Omega} &= \mathbf{0} \qquad \text{on } \partial\Omega. \end{aligned}$$

Here Π_h denotes the orthogonal projector in $[L_2(\Omega)]^m$ onto Y_h.

Next we define the functional that represents the normal outflow flux through the boundary of the domain Ω. Given any $\psi \in [H^1(\partial\Omega)]^m$, we consider the linear functional

$$N_\psi(\mathbf{v}) = \int_{\partial\Omega} (B^+\mathbf{v}) \cdot \psi \, \mathrm{d}s, \qquad \mathbf{v} \in H(\mathcal{L}, \Omega);$$

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

Theorem 19 *Suppose that Hypothesis 2 and condition (c) hold; then, for each $\psi \in [H^1(\partial\Omega)]^m$ 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

$$\mathcal{L}^*\mathbf{z} = \mathbf{0} \quad \text{in } \Omega, \qquad \gamma_{B^+}\mathbf{z} = B^+\psi \quad \text{on } \partial\Omega. \tag{6.1}$$

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

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

Consequently, for any $\mathbf{z}_h \in Y_h$,

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

Exploiting hypothesis (c), it follows that

$$|N_\psi(\mathbf{u}) - N_\psi(\mathbf{u}_h)| \leq C_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)} \right)^{1/2} \|\mathbf{z}\|_{H^1(\Omega)},$$

and therefore, by Hypothesis 2 with $\mu = \mathbf{0}$ and $\chi|_{\partial\Omega} = \psi$,

$$|N_\psi(\mathbf{u}) - N_\psi(\mathbf{u}_h)| \leq C_1' C_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)} \right)^{1/2} \|\psi\|_{H^1(\partial\Omega)}.$$

■

From the practical point of view, a particularly relevant situation concerns the estimation of the flux through a relatively open subset $\gamma \subset \partial\Omega$. In this case it is tempting to choose ψ equal to the characteristic function of γ; unfortunately such ψ does not belong to $[H^1(\partial\Omega)]^m$ (since the characteristic function of γ is in $H^{1/2-\varepsilon}(\partial\Omega)$ for all $\varepsilon > 0$, but not in $H^s(\partial\Omega)$ for $s \geq 1/2$), so Theorem 19 does not apply. Nevertheless, we note that the choice of a smoother cut-off function $\psi \in [H^1(\partial\Omega)]^m$ with support in γ is covered by Theorem 19, and this may suffice in practice.

6.2 Estimation of the local mean value

In this section we consider the *a posteriori* error analysis of finite element approximations to the local mean value

$$M_\psi(\mathbf{u}) = \int_\Omega \mathbf{u}(x) \cdot \psi(x) \, \mathrm{d}x,$$

where $\psi \in [H_0^1(\Omega)]^m$, and $\mathbf{u}$ is the solution of the non-homogeneous boundary value problem (4.1), (4.3) with $\mathbf{f} \in [L_2(\Omega)]^m$ and $\mathbf{g} \in [H^1(\Omega)]^m$. Again, for the sake of simplicity, we assume that the restriction of $\mathbf{g}$ to $\partial\Omega$ belongs to the restriction of the trial space $X_h \subset H(\mathcal{L}, \Omega)$ to the boundary, and we approximate the non-homogeneous problem by the following method: find $\mathbf{u}_h$ in X_h such that

$$\begin{aligned} \Pi_h \mathcal{L} \mathbf{u}_h &= \Pi_h \mathbf{f} \qquad \text{in } \Omega, \\ B^-(\mathbf{u}_h - \mathbf{g})\big|_{\partial\Omega} &= \mathbf{0} \qquad \text{on } \partial\Omega. \end{aligned}$$

Here, as in the previous section, Π_h denotes the orthogonal projection in $[L_2(\Omega)]^m$ onto the finite element test space $Y_h \subset [L_2(\Omega)]^m$. We have the following *a posteriori* bound on the error between $M_\psi(\mathbf{u})$ and its approximation $M_\psi(\mathbf{u}_h)$.

Theorem 20 *Suppose that Hypothesis 2 and condition (c) hold; then, for each* $\psi \in [H_0^1(\Omega)]^m$ *we have that*

$$|M_\psi(\mathbf{u}) - M_\psi(\mathbf{u}_h)| \leq C_1' C_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)} \right)^{1/2} \|\psi\|_{H^1(\Omega)}.$$

Proof In contrast with the previous section, here the appropriate dual problem is of the form

$$\begin{aligned} \mathcal{L}^*\mathbf{z} &= \psi \qquad \text{in } \Omega, \\ \gamma_{B^+}(\mathbf{z}) &= \mathbf{0} \qquad \text{on } \partial\Omega. \end{aligned}$$

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

$$M_\psi(\mathbf{u}) - M_\psi(\mathbf{u}_h) = (\mathbf{u} - \mathbf{u}_h, \psi) = (\mathbf{u} - \mathbf{u}_h, \mathcal{L}^*\mathbf{z}) = (\mathcal{L}(\mathbf{u} - \mathbf{u}_h), \mathbf{z}).$$

Consequently, for any $\mathbf{z}_h \in Y_h$,

$$M_\psi(\mathbf{u}) - M_\psi(\mathbf{u}_h) = (\mathbf{r}_h, \mathbf{z} - \mathbf{z}_h).$$

Exploiting hypothesis (c), it follows that

$$|M_\psi(\mathbf{u}) - M_\psi(\mathbf{u}_h)| \leq C_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)} \right)^{1/2} \|\mathbf{z}\|_{H^1(\Omega)},$$

and therefore, by the strong stability of the dual problem,

$$|M_\psi(\mathbf{u}) - M_\psi(\mathbf{u}_h)| \leq C_1' C_2 \left(\sum_\kappa \|h\mathbf{r}_h\|^2_{L_2(\kappa)} \right)^{1/2} \|\psi\|_{H^1(\Omega)}.$$

■

Comparing this analysis with the one performed in the previous subsection for the boundary flux, one quickly recognises the similarities. Indeed, the question arises, whether it is possible to provide a general approach to the *a posteriori* error analysis of functionals. This is the theme of the next subsection.

6.3 A general duality argument

We give a brief overview of a general duality argument due to Becker and Rannacher (see [7]) for the *a posteriori* error estimation of functionals. For an alternative perspective on duality arguments, we refer to the work of Giles [18] and, in a slightly different context, the articles of Peraire, Paraschivoiu and Patera [47] and Giles, Larson, Levenstam and Süli [19].

Suppose that X and Y are two reflexive Banach spaces equipped with their norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, that $a(\cdot,\cdot)$ is a continuous bilinear functional on $X \times Y$ and $l(\cdot)$ is a continuous linear functional on Y. In the framework of symmetric positive systems, appropriate choices of X and Y and of $a(\cdot,\cdot)$ and $l(\cdot)$ are given at the beginning of Section 4.2, in α) and β).

We consider the variational problem: find $\mathbf{u}$ in X such that

$$a(\mathbf{u}, \mathbf{v}) = l(\mathbf{v}) \qquad \text{for all } \mathbf{v} \text{ in } Y.$$

This problem is approximated by a Galerkin finite element method using a sequence of trial spaces $X_h \subset X$ and test spaces $Y_h \subset Y$ parametrised by a discretisation parameter h. The discrete problem reads: find $\mathbf{u}_h$ in X_h such that

$$a(\mathbf{u}_h, \mathbf{v}_h) = l(\mathbf{v}_h) \qquad \text{for all } \mathbf{v}_h \text{ in } Y_h.$$

Letting $\mathbf{e}_h = \mathbf{u} - \mathbf{u}_h$ denote the global error, we observe the Galerkin orthogonality property

$$a(\mathbf{e}_h, \mathbf{v}_h) = l(\mathbf{v}_h) \qquad \text{for all } \mathbf{v}_h \text{ in } Y_h.$$

Now suppose that $M(\cdot)$ is a continuous linear functional on X. In order to derive an *a posteriori* bound on the error between $M(\mathbf{u})$ and its approximation $M(\mathbf{u}_h)$, we introduce the following dual problem: find $\mathbf{z}$ in Y such that

$$a(\mathbf{w}, \mathbf{z}) = M(\mathbf{w}) \qquad \text{for all } \mathbf{w} \text{ in } X.$$

Assuming that $a(\cdot, \cdot)$ satisfies the hypotheses of Theorem 7 on $X \times Y$ (rather than $X_h \times Y_h$) with X and Y interchanged[2], we deduce that the dual problem has a unique solution $\mathbf{z}$ in Y. Next, we see that

$$M(\mathbf{u}) - M(\mathbf{u}_h) = M(\mathbf{e}_h) = a(\mathbf{e}_h, \mathbf{z}) = a(\mathbf{e}_h, \mathbf{z} - \mathbf{z}_h)$$

for all $\mathbf{z}_h$ in Y_h. Equivalently, we can write this identity as

$$M(\mathbf{u}) - M(\mathbf{u}_h) = l(\mathbf{z} - \mathbf{z}_h) - a(\mathbf{u}_h, \mathbf{z} - \mathbf{z}_h).$$

Further, noting that $l(\cdot) - a(\mathbf{u}_h, \cdot)$ is a continuous linear functional on Y, we can rewrite the right-hand side as $\langle \mathbf{r}_h, \mathbf{z} - \mathbf{z}_h \rangle$ where $\mathbf{r}_h$ is the residual and $\langle \cdot, \cdot \rangle$ denotes the duality pairing between Y', the dual space of Y, and Y. This leads to the error representation

$$M(\mathbf{u}) - M(\mathbf{u}_h) = \langle \mathbf{r}_h, \mathbf{z} - \mathbf{z}_h \rangle,$$

for all $\mathbf{z}_h$ in Y_h. At this point we are at the same stage in the analysis as in Section 4.2 in the paragraph preceding Theorem 1. Following the same reasoning as there, one can derive an *a posteriori* bound on the error in the functional, $M(\mathbf{u}) - M(\mathbf{u}_h)$. We refer to [7] for further details, in the context of elliptic boundary-value problems.

To conclude this section, we note that this approach allows one to derive *a posteriori* bounds on the global error in norms stronger than $\|\cdot\|_{[H^{-1}(\Omega)]^m}$. Indeed, to obtain an *a posteriori* bound on the global error $\mathbf{e}_h = \mathbf{u} - \mathbf{u}_h$ in the norm of $[L_p(\Omega)]^m$, $1 \leq p < \infty$, we consider the functional

$$M_p(\mathbf{w}) = \int_\Omega \frac{|\mathbf{e}_h|^{p-1}}{\|\mathbf{e}_h\|^{p-1}_{[L_p(\Omega)]^m}} \operatorname{sgn}(\mathbf{e}_h) \cdot \mathbf{w} \, \mathrm{d}x,$$

where $\operatorname{sgn}(\mathbf{v})$ is a vector whose jth entry is the sign of the jth entry of $\mathbf{v}$. Clearly (remembering that we are dealing with real-valued functions!),

$$\|\mathbf{u} - \mathbf{u}_h\|_{[L_p(\Omega)]^m} = M_p(\mathbf{u}) - M_p(\mathbf{u}_h),$$

and thereby,

$$\|\mathbf{u} - \mathbf{u}_h\|_{[L_p(\Omega)]^m} = \langle \mathbf{r}_h, \mathbf{z} - \mathbf{z}_h \rangle, \tag{6.2}$$

[2]It can be shown that the inf-sup condition with X and Y interchanged is, in fact, equivalent to the inf-sup condition in its usual form; for a proof, we refer to Proposition A.2 in the paper of Melenk and Schwab [42].

for any $\mathbf{z}_h$ in the test space; here $\mathbf{z}$ denotes the solution to the dual problem

$$a(\mathbf{w}, \mathbf{z}) = M_p(\mathbf{w}) \qquad \text{for all } \mathbf{w} \text{ in } X. \tag{6.3}$$

We see that the right-hand side of (6.2) is of the usual form; so, at least formally, one can proceed with the *a posteriori* error analysis as before. However, there is a fundamental difference between (6.3) and the dual problems which occurred in Subsections 6.1 and 6.2: while in those problems ψ was a given function, so the data for the dual was known, here the dual problem involves the (unknown) global error $\mathbf{e}_h$. From the point of view of implementation a possible approach might be to compute the numerical solution of two (or more) successively refined meshes, and use their difference as an approximation to $\mathbf{e}_h$ in the functional $M_p(\cdot)$ to fix the right-hand side of the dual problem, and repeat this process in the course of the adaptive mesh refinement driven by the resulting *a posteriori* error bound. More analysis is required to quantify the effects of this additional approximation.

7 A posteriori analysis for unsteady problems

So far, we have been concerned with the *a posteriori* error analysis of finite element approximations to steady hyperbolic problems. In the present section we consider similar questions for unsteady problems.

In the next subsection we discuss a general class of (semi-discrete in time) Petrov-Galerkin methods for strictly hyperbolic systems, while in the second subsection we restrict ourselves to (fully-discrete) evolution Galerkin methods for scalar hyperbolic equations, although the basic steps in the error analysis would be identical for fully-discrete finite element approximations multi-dimensional hyperbolic systems.

7.1 A posteriori error analysis for strictly hyperbolic systems

In this section we shall consider the *a posteriori* error analysis of finite element approximations to the system of partial differential equations

$$\frac{\partial \mathbf{u}}{\partial t} = \sum_{i=1}^{n} A_i(x,t)\frac{\partial \mathbf{u}}{\partial x_i} + C(x,t)\mathbf{u} + \mathbf{f}(x,t), \tag{7.1}$$

where A_i, $i = 1, \ldots, n$, and C are smooth $m \times m$ matrix-valued functions, constant outside a compact subset of $\Omega \times \mathbb{R}$ with

$$\begin{aligned} \Omega &= \{x \in \mathbb{R}^n \,:\, x_1 > 0\}, \\ \partial\Omega &= \{x \in \mathbb{R}^n \,:\, x_1 = 0\}, \\ x &= (x_1, \ldots, x_n) = (x_1, x'). \end{aligned}$$

In physical applications modelled by this system, the variable t plays the rôle of time, and $x = (x_1, \ldots, x_n)$ represent the spatial independent variables.

It will be assumed that the differential operator is strictly hyperbolic; in other words, we shall suppose that the matrix $\sum_{i=1}^{n} A_i(x,t)\xi_i$ has m distinct real eigenvalues for all $\xi \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $(x,t) \in \bar{\Omega} \times \mathbb{R}$. Furthermore, we shall require that the boundary $\partial\Omega \times \mathbb{R}$ is a non-characteristic hypersurface for the differential operator, namely, $\det(A_1) \neq 0$ when $x_1 = 0$.

Equation (7.1) is solved in tandem with an initial condition at $t = 0$, and a boundary condition on $\partial\Omega \times [0,T]$ which is imposed by considering the boundary operator $B(x',t)$, a smooth $l \times m$ matrix-valued function, independent of (x',t) for $|x'| + t$ large, such that $\mathrm{rank}(B) = l$, where l is the number of negative eigenvalues of A_1. Then it is known that, for any $T > 0$, $\mathbf{f} \in L_2([0,T] \times \Omega)$ and $\mathbf{u}_0 \in L_2(\Omega)$, there is a unique strong solution $\mathbf{u}$ of (7.1) subject to the initial condition

$$\mathbf{u}(x,0) = \mathbf{u}_0(x) \qquad \text{for } x \in \Omega \tag{7.2}$$

and the boundary condition

$$B\mathbf{u} = \mathbf{0} \qquad \text{on } \partial\Omega \times [0,T], \tag{7.3}$$

provided that the latter is admissible in a sense that will be made precise below.

By admissibility of the boundary condition (7.3) we mean the following. Let us note that, upon a smooth change of coordinates, A_1 can be written in the form

$$A_1 = \begin{bmatrix} A_1^I & 0 \\ 0 & A_2^{II} \end{bmatrix},$$

where A_1^I is negative definite and A_2^{II} is positive definite; by splitting the vector $\mathbf{u}$ in a similar manner into $\mathbf{u}^I = (u_1, \ldots, u_l)$ and $\mathbf{u}^{II} = (u_{l+1}, \ldots, u_m)$, the boundary condition $B\mathbf{u} = \mathbf{0}$, upon decomposing B correspondingly, can be restated as $S_I\mathbf{u}^I - S_{II}\mathbf{u}^{II} = \mathbf{0}$. We shall say that the boundary condition is *admissible* if S_I is invertible; in that case, (7.3) can be written in the equivalent form

$$\mathbf{u}^I - S\mathbf{u}^{II} = \mathbf{0},$$

where $S = S_I^{-1} S_{II}$.

Now we consider a finite element approximation to this problem. Suppose that we have chosen a finite element trial space $X_h \subset [H^1(\Omega)]^m$ consisting of continuous piecewise polynomial m-component vector functions on a partition $\mathcal{T}_h = \{\kappa\}$ of Ω which satisfy the boundary condition (7.3), and a finite element test space $Y_h \subset [L_2(\Omega)]^m$ on the same partition. Then, we approximate (7.1) – (7.3) by a semi-discrete Petrov-Galerkin finite element method of the following form: find $\mathbf{u}_h(t) \in X_h$, $0 < t \leq T$, such that, for all $\mathbf{q}_h \in Y_h$,

$$\begin{aligned} \left(\frac{\partial \mathbf{u}_h}{\partial t}, \mathbf{q}_h\right) &= \sum_{i=1}^{n} \left(A_i(\cdot,t)\frac{\partial \mathbf{u}_h}{\partial x_i}, \mathbf{q}_h\right) + (C(\cdot,t)\mathbf{u}_h, \mathbf{q}_h) + (\mathbf{f}(\cdot,t), \mathbf{q}_h)\,, \\ (\mathbf{u}_h(\cdot,0), \mathbf{q}_h) &= (\mathbf{u}_0(\cdot), \mathbf{q}_h). \end{aligned}$$

We note in passing that the inclusion of the boundary condition into the definition of the trial space is unreasonable from the practical point of view (and implausible from the theoretical point of view, unless S is a constant matrix); we have adopted this assumption only to simplify the error analysis. A practical method would involve a weakly imposed boundary condition, in the same spirit as in the steady case discussed earlier on. Moreover, in practice, a time-discretisation is required; as long as the latter is also a Galerkin-type method, the error analysis that we provide below in the semi-discrete case is easily modified to include the effects of the time discretisation.

We define the finite element residual

$$\mathbf{r}_h(x,t) = \mathbf{f}(x,t) - \frac{\partial \mathbf{u}_h}{\partial t} + \sum_{i=1}^{n} A_i(x,t)\frac{\partial \mathbf{u}_h}{\partial x_i} + C(x,t)\mathbf{u}_h.$$

It is in terms of this quantity that we wish to obtain a bound on the discretisation error in the spatial H^{-1} norm at time T.

Theorem 21 *Suppose that hypothesis (c) holds. Then, there exists a computable constant C_5 such that*

$$\|\mathbf{u}(\cdot,T) - \mathbf{u}_h(\cdot,T)\|_{H^{-1}(\Omega)} \leq C_5 \left(\|h\mathbf{r}_h\|^2_{L_2(0,T;L_2(\Omega))} + \|h\mathbf{r}_h^0\|^2_{L_2(\Omega)} \right)^{\frac{1}{2}},$$

where $\mathbf{r}_h^0(x) = \mathbf{u}_0(x) - \mathbf{u}_h(x,0)$.

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

$$\begin{aligned} \frac{\partial \mathbf{z}}{\partial t} &= \sum_{i=1}^{n} \frac{\partial}{\partial x_i}(A_i^* \mathbf{z}) - C^* \mathbf{z} \qquad \text{in } \Omega \times [0,T), \\ \mathbf{z}(x,T) &= \psi(x) \qquad \text{for } x \in \Omega, \end{aligned}$$

subject to the boundary condition

$$\mathbf{z}^{II} - \hat{S}\mathbf{z}^{I} = \mathbf{0} \qquad \text{on } \partial\Omega \times [0,T],$$

where $\hat{S} = -A_1^{II\,-*} S^* A_1^{I*}$. Then

$$\begin{aligned} (\mathbf{e}_h, \mathbf{z})|_0^T &= \int_0^T \frac{d}{dt}(\mathbf{e}_h, \mathbf{z})\, \mathrm{d}t = \int_0^T (\frac{\partial \mathbf{e}_h}{\partial t}, \mathbf{z}) + (\mathbf{e}_h, \frac{\partial \mathbf{z}}{\partial t})\, \mathrm{d}t \\ &= \int_0^T (\mathbf{r}_h, \mathbf{z})\, \mathrm{d}t = \int_0^T (\mathbf{r}_h, \mathbf{z} - \mathbf{z}_h)\, \mathrm{d}t \end{aligned}$$

for any $\mathbf{z}_h \in Y_h$. Noting the approximation property (c), we deduce that

$$\begin{aligned} |(\mathbf{e}_h(\cdot,T), \psi)| \leq\ & C_6 \left(\|h\mathbf{e}_h(\cdot,0)\|_{L_2(\Omega)} \|\mathbf{z}(\cdot,0)\|_{H^1(\Omega)} \right. \\ & \left. + \|h\mathbf{r}_h\|_{L_2(0,T;L_2(\Omega))} \|\mathbf{z}\|_{L_2(0,T;H^1(\Omega))} \right). \end{aligned}$$

According to a hyperbolic regularity theorem due to Rauch [49],

$$\left(\|\mathbf{z}(\cdot,0)\|^2_{H^1(\Omega)} + \|\mathbf{z}\|^2_{L_2(0,T;H^1(\Omega))} \right)^{\frac{1}{2}} \leq C_7 \|\psi\|_{H^1(\Omega)},$$

and hence the required result with $C_5 = C_6 C_7$. ∎

In the next section, we consider the *a posteriori* error analysis of a fully discrete method for an unsteady hyperbolic equation.

7.2 A posteriori analysis of evolution-Galerkin methods

Here, we develop the *a posteriori* error analysis of evolution-Galerkin finite element methods for unsteady scalar hyperbolic problems. The presentation closely follows the article of Süli and Houston [56], albeit with some abbreviations. For a detailed study of the (unconditional) stability and accuracy properties of evolution-Galerkin methods we refer to [45].

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 hyperbolic initial-value problem

$$\frac{\partial u}{\partial t} + \mathbf{a} \cdot \nabla u = f, \qquad \mathbf{x} \in \Omega,\ t \in I, \tag{7.4}$$

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{7.5}$$

where $\Omega = \mathbb{R}^n$. For the sake of simplicity we assume that the velocity field $\mathbf{a}$ is in $C([0, T]; C_0^1(\Omega)^n)$, 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 $L_2(I; H^1(\Omega))$.

We consider a subdivision (not necessary uniform) of the time interval $I = [0, T]$ given by $0 = t^0 < t^1 < \ldots < t^M < t^{M+1} = T$; we define time intervals $I^m = (t^{m-1}, t^m]$ and time steps $k_m = t^m - t^{m-1}$. For each m, let $\mathcal{T}^m = \{\kappa\}$ be a partition of Ω into closed simplices κ, with corresponding mesh function h_m, piecewise continuous on Ω, satisfying

$$C_\dagger h_\kappa^n \leq \mathrm{meas}(\kappa) \quad \forall \kappa \in \mathcal{T}^m, \tag{7.6}$$

$$C_* h_\kappa \leq h_m(\mathbf{x}) \leq h_\kappa \quad \forall \mathbf{x} \in \kappa \quad \forall \kappa \in \mathcal{T}^m, \tag{7.7}$$

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 $h(\mathbf{x}, t) = h_m(\mathbf{x})$ for $(\mathbf{x}, t)$ in $\Omega \times I^m$, and we define the corresponding time step function $k = k(t)$ by $k(t) = k_m,\ t \in I^m$. Let $\Lambda^m = \Omega \times I^m$; for $p, q \in \mathbb{N}$, let

$$\begin{aligned}
S^{h_m} &= \{v \in H_0^1(\Omega) : v|_\kappa \in \mathcal{P}_p(\kappa)\ \forall \kappa \in \mathcal{T}^m\},\\
V^{h_m} &= \{v : v(\mathbf{x}, t)|_{\Lambda^m} = \sum_{j=0}^{q} t^j v_j,\ v_j \in S^{h_m}\},\\
V^h &= \{v : v(\mathbf{x}, t)|_{\Lambda^m} \in V^{h_m},\ m = 1, \ldots, M+1\},
\end{aligned}$$

where $\mathcal{P}_p(\kappa)$ 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 $v \in V^{h_m}$ for $m = 1, \ldots, M+1$, then v is continuous in space at any time, but may be discontinuous in time at the discrete time levels t^m. To account for this, we introduce the notation $v_\pm^m := \lim_{s \to 0^+} v(t^m \pm s)$, and $[v^m] := v_+^m - v_-^m$.

The evolution-Galerkin approximation of (7.4), (7.5) makes use of the particle trajectories (or characteristics) associated with equation (7.4): 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

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{X}(\mathbf{x},s;t) &= \mathbf{a}(\mathbf{X}(\mathbf{x},s;t),t), & (7.8)\\ \mathbf{X}(\mathbf{x},s;s) &= \mathbf{x}. & (7.9) \end{aligned}$$

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

$$\begin{aligned} D_t u(\mathbf{x},s) &:= \frac{\mathrm{d}}{\mathrm{d}t} u(\mathbf{X}(\mathbf{x},s;t),t)\,|_{t=s} \\ &= \frac{\partial}{\partial t} u(\mathbf{x},s) + \mathbf{a}(\mathbf{x},s)\cdot\nabla u(\mathbf{x},s) \quad \forall \mathbf{x}\in\Omega,\ s\in I. \end{aligned} \tag{7.10}$$

The evolution-Galerkin time-discretisation is based on approximating the material derivative by a divided difference operator along particle trajectories. The simplest appropriate scheme arises from using Euler's method, giving, for $m = 0,\ldots,M$,

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

Suppose that u_h^m denotes the approximation to $u(\cdot,t^m)$ at time t^m; then, applying the finite element method in space results in what is known as the evolution-Galerkin discretisation of the scalar linear hyperbolic equation (7.4): find $u_h^{m+1} \in S^{h_{m+1}}$, for $m = 0,\ldots,M$, such that

$$\begin{aligned} \left(\frac{u_h^{m+1} - u_h^m(\mathbf{X}(\cdot,t^{m+1};t^m))}{k_{m+1}}, v\right) &= (\bar{f},v) \quad \forall v \in S^{h_{m+1}}, & (7.11)\\ (u_h^0, v) &= (u_0,v) \quad \forall v \in S^{h_0}, & (7.12) \end{aligned}$$

where $\bar{f}|_{\Lambda^{m+1}} := f(\cdot,t^{m+1})$. Alternatively, by integrating (7.11) with respect to t over I^{m+1}, we obtain the following equivalent formulation: find u_h such that, for $m = 0,1,\ldots,M$, $u_h|_{\Lambda^{m+1}} \in V^{h_{m+1}}$ and satisfies

$$\begin{aligned} (D_t^h u_h, v)_{m+1} &= (\bar{f},v)_{m+1} \quad \forall v \in V^{h_{m+1}}, & (7.13)\\ (u_{h-}^0, v) &= (u_0,v) \quad \forall v \in V^{h_0}, & (7.14) \end{aligned}$$

where

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

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

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

Before stating the relevant *a posteriori* error bound for this method, we note that in (7.13), (7.14) the space discretisation may vary in both space and time, but the time steps are only variable in time and not in space, so the corresponding space-time mesh will not be fully optimal. The method obeys the following *a posteriori* error bound.

Theorem 22 *Let u and u_h be solutions of (7.4), (7.5) and (7.13), (7.14), respectively. Then*

$$\|u - u_h\|_{L_\infty(0,T;H^{-1}(\Omega))} \leq \overset{\circ}{\mathcal{E}}(u_h, h, k, f), \tag{7.15}$$

where

$$\begin{aligned}
\overset{\circ}{\mathcal{E}}(u_h, h, k, f) &= \mathcal{E}(u_h, h, k, f) + \mathcal{E}_0(u_0, u^0_{h-}, h), \\
\mathcal{E}(u_h, h, k, f) &= C_1\|hR_1\|_Q + C_2\|kR_1\|_Q \\
&\quad + C_3\|kR_2\|_Q + C_4\|kR_3\|_Q + C_5\|kR_4\|_Q, \qquad (7.16) \\
\mathcal{E}_0(u_0, u^0_{h-}, h) &= C_6\|u_0 - u^0_{h-}\|, \qquad (7.17)
\end{aligned}$$

and

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

and C_i, $i = 1, \ldots, 6$, are (computable) positive constants.

For a proof of this result the reader is referred to Süli and Houston [56], where the precise values of the constants $C_1, \ldots, C_6$ appearing in the error bound are also specified.

In the remainder of this section we consider the computational implementation of the *a posteriori* error bound stated in the last theorem. Much of what will be said, however, applies in a more general setting.

For a given tolerance, TOL, we consider the problem of finding a discretisation in space and time $\mathcal{S}^h = \{(\mathcal{T}^m, t^m)\}_{n\geq 0}$ 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 (7.15) to choose $\mathcal{S}^h$ such that:

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

The term $\mathcal{E}_0(u_0, u^0_{h-}, h)$ 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 $\text{TOL} = \text{TOL}' + \mathcal{E}_0(u_0, u^0_{h-}, h)$. 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

$$\mathcal{E}(u_h, h, k, f) \equiv C'_1\|hR'_1\|_Q + C'_2\|kR'_2\|_Q. \tag{7.18}$$

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:

$$C_1'\|hR_1'\|_Q \leq \mathrm{TOL}_h, \tag{7.19}$$
$$C_2'\|kR_2'\|_Q \leq \mathrm{TOL}_k. \tag{7.20}$$

To design the space-time mesh $\mathcal{S}^h$, at each time level t^m we decompose the norm in (7.19) into norms over elements $\kappa \in \mathcal{T}^m$, and the norm in (7.20) into norms over time slabs as follows:

$$\begin{aligned} C_1'\|hR_1'\|_Q &\leq C_1'\sqrt{T}\max_{1\leq m\leq M+1}\|h_m R_1'(u_h^m)\| \\ &\leq C_1'\sqrt{NT}\max_{1\leq m\leq M+1}\left(\max_{\kappa\in\mathcal{T}^m}\|h_m R_1'(u_h^m)\|_{L_2(\kappa)}\right), \\ C_2'\|kR_2'\|_Q &\leq C_2'\sqrt{T}\max_{1\leq m\leq M+1}\|k_m R_2'(u_h^m)\|, \end{aligned}$$

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

$$\begin{aligned} C_1'\sqrt{NT}\|h_m R_1'(u_h^m)\|_{L_2(\kappa)} &\leq \mathrm{TOL}_h \quad \forall\kappa\in\mathcal{T}^m, \text{ for } m=1,\ldots,M+1, \\ C_2'\sqrt{T}\|k_m R_2'(u_h^m)\| &\leq \mathrm{TOL}_k, \quad \text{for } m=1,\ldots,M+1, \end{aligned}$$

then (7.19) and (7.20) 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 $m = 1, 2, \ldots, M+1$, with $\mathcal{T}_0^m$ a given initial mesh and $k_{m,0}$ an initial time step, determine meshes $\mathcal{T}_j^m$ with N_j elements of size $h_{m,j}(\mathbf{x})$ and time steps $k_{m,j}$ and the corresponding approximate solution $u_{h,j}$ defined on I_j^m such that, for $j = 0, 1, \ldots, \hat{m}-1$,

$$C_1\|h_{m,j+1}R_1(u_{h,j}^m)\|_{L_2(\kappa)} = \frac{\mathrm{TOL}_h}{\sqrt{N_j T}} \quad \forall\kappa\in\mathcal{T}_j^m, \tag{7.21}$$

$$\begin{aligned} &C_2\|k_{m,j+1}R_1(u_{h,j}^m)\| + C_3\|k_{m,j+1}R_2(u_{h,j}^m)\| \\ &+C_4\|k_{m,j+1}R_3(u_{h,j}^m)\| + C_5\|k_{m,j+1}R_4(u_{h,j}^m)\| = \frac{\mathrm{TOL}_k}{\sqrt{T}}, \end{aligned} \tag{7.22}$$

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

$$C_1\|h_{m,\hat{m}}R_1(u_{h,\hat{m}}^m)\|_{L_2(\kappa)} \leq \frac{\mathrm{TOL}_h}{\sqrt{N_{\hat{m}} T}} \quad \forall\kappa\in\mathcal{T}_{\hat{m}}^m, \tag{7.23}$$

$$\begin{aligned} &C_2\|k_{m,\hat{m}}R_1(u_{h,\hat{m}}^m)\| + C_3\|k_{m,\hat{m}}R_2(u_{h,\hat{m}}^m)\| \\ &+C_4\|k_{m,\hat{m}}R_3(u_{h,\hat{m}}^m)\| + C_5\|k_{m,\hat{m}}R_4(u_{h,\hat{m}}^m)\| \leq \frac{\mathrm{TOL}_k}{\sqrt{T}}. \end{aligned} \tag{7.24}$$

By construction, this stopping condition will guarantee reliability of the adaptive algorithm; for efficiency, we try to ensure that (7.23) and (7.24) are satisfied with near

equality. Since the final time T is fixed, the time step given by (7.22) may need to be limited to ensure that $t^M + k_{M+1,\hat{m}} = T$. For the implementation of this adaptive algorithm, we shall assume that $\mathcal{T}_0^m = \mathcal{T}^{m-1}$ for $m = 2, 3, \ldots$.

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

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

we note that, in order to generate the desired mesh we need 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 (7.22). For constructing the spatial mesh in two space dimensions we use the red-green isotropic refinement strategy of Bank [6]. 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 Kenneth Eriksson (Chalmers University).

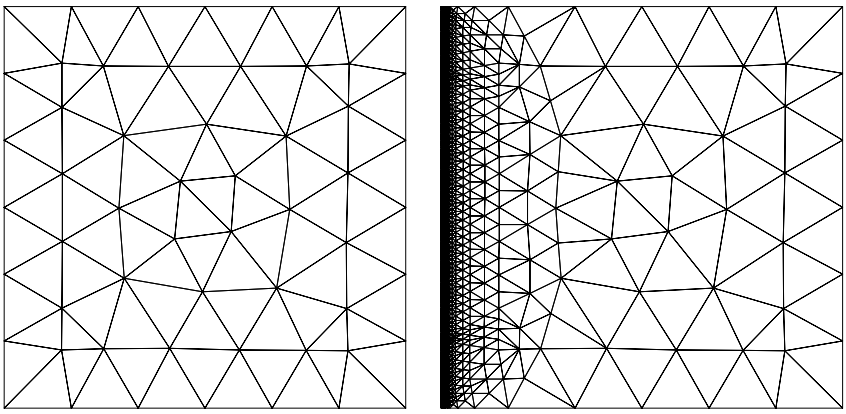
7.3 Numerical experiments

In this section, we present some numerical experiments to illustrate the performance of the adaptive algorithm (7.21) on the model hyperbolic test problem:

$$\frac{\partial u}{\partial t} + \mathbf{a} \cdot \nabla u = f, \qquad \mathbf{x} \in \Omega,\ t \in I, \tag{7.25}$$

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{7.26}$$

where $\Omega = (0,1)^2$, $f = 0$, $\mathbf{a} = (2,1)$, 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 the domain Ω, and eventually exits through $x = 1$. In the following, we shall let $\delta = 7.8125 \times 10^{-3}$ and $T = 0.6$. First, we specify the background mesh as the one shown in Figure 3(a); this is initially refined in order



(a) (b)

Figure 3: (a) Background mesh, with 56 nodes and 86 elements; (b) Background mesh adapted to resolve the initial condition, with 8335 nodes and 15860 elements.

to resolve the boundary layer along $x = 0$ at time $t = 0$, as shown in Figure 3(b). Numerical results are presented in Figure 4 for $\mathrm{TOL}_h = 0.007$ and $\mathrm{TOL}_k = 0.25$.

In Figures 4(a), 4(b) and 4(c), 4(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 4(e), 4(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 [26] was employed in this experiment with $C_1^{\hat{\epsilon}} = C_2^{\hat{\epsilon}} = 0.2$ and $\hat{\epsilon}_{\max} = 7.0 \times 10^{-4}$.

8 Nonlinear conservation laws

In this concluding section we discuss briefly the extension of the approach to *a posteriori* error analysis described earlier in the case of linear problems to nonlinear hyperbolic equations; for a survey of the theory and numerical analysis of conservation laws, we refer to the recent monographs of Godlewski and Raviart [21], and Kröner [33]. For the sake of simplicity we shall restrict ourselves to scalar nonlinear conservation laws of the form

$$\frac{\partial}{\partial t}u(x,t) + \frac{\partial}{\partial x}f(u(x,t)) = 0, \qquad -\infty < x < \infty, \quad 0 < t \leq T, \tag{8.1}$$

with strictly convex flux function f (i.e. $f'' \geq \alpha > 0$), subject to the initial condition

$$u(x,0) = u_0(x), \qquad -\infty < x < \infty, \tag{8.2}$$

where u_0 is a compactly supported Lip^+ bounded function. We say that a function $x \mapsto w(x)$ is Lip^+ bounded if

$$\|w\|_{Lip^+(\mathbb{R})} \equiv \mathrm{ess.sup}_{x \neq y} \left(\frac{w(x) - w(y)}{x - y} \right)^+ < \infty,$$

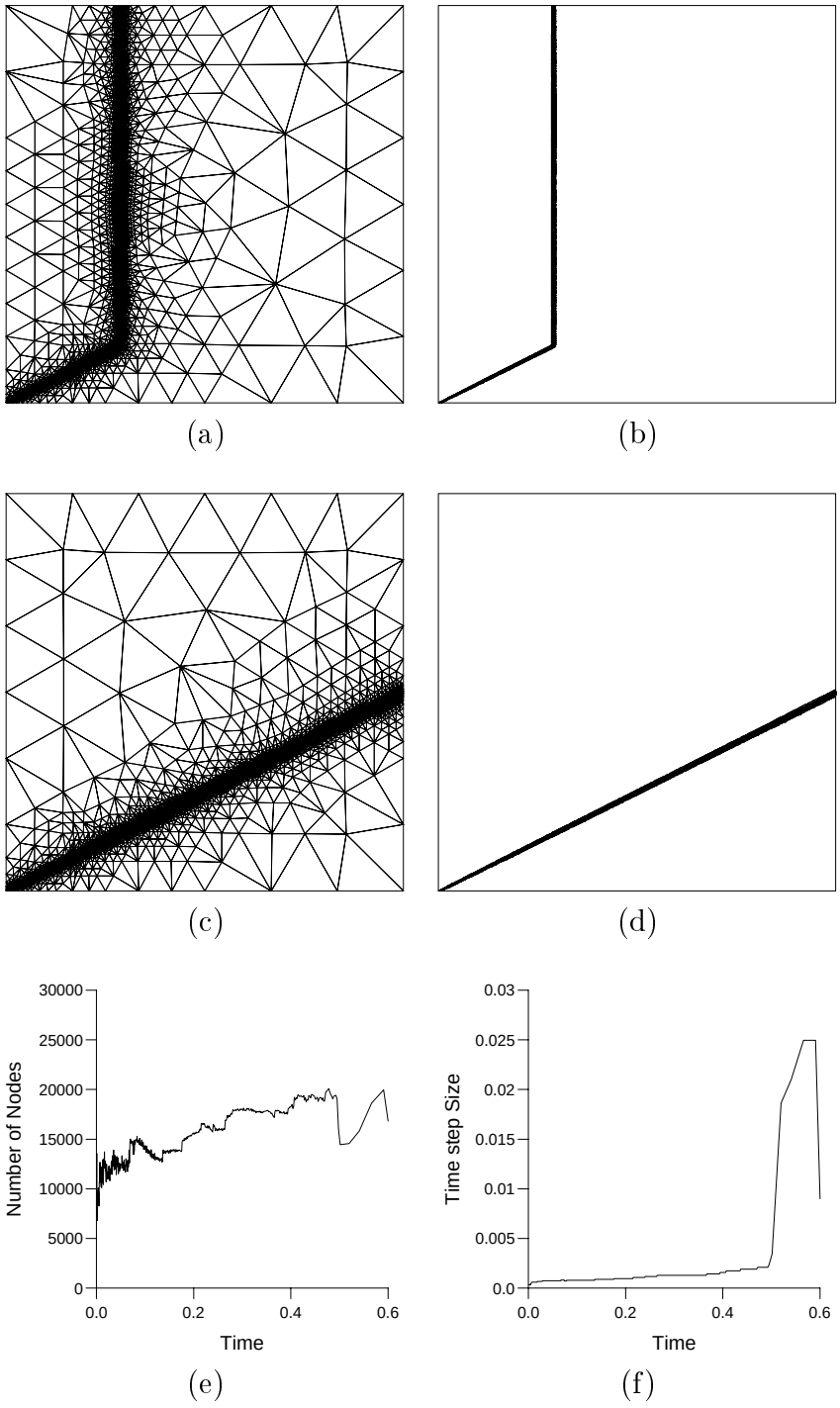


Figure 4: Layer problem for $\mathrm{TOL}_h = 0.007$ and $\mathrm{TOL}_k = 0.25$ with $T = 0.6$: (a) & (b) Mesh and solution (resp.) at time, $t = 0.1428$, with 13596 nodes and 27109 elements; (c) & (d) Mesh and solution (resp.) at final time, $t = 0.6$, with 16829 nodes and 33566 elements; (e) History of nodes against time; (f) History of time step size against time.

where $(\cdot)^+ \equiv \max(\cdot, 0)$. We note that the hypotheses imposed on u_0 imply that it belongs to $L_\infty(\mathbb{R})$. It can be supposed, without restricting generality, that $f(0) = 0$.

A weak solution to this initial-value problem obeys the identity

$$(u(\cdot,T), v(\cdot,T)) = (u_0, v(\cdot,0)) + \int_0^T (u(\cdot,t), v_t(\cdot,t)) + (f(u(\cdot,t)), v_x(\cdot,t))\,\mathrm{d}t$$
$$v \in C^1([0,T], C_0^1(\mathbb{R})), \tag{8.3}$$

where $(\cdot,\cdot)$ denotes the inner product of $L_2(\mathbb{R})$.

We recall that the entropy solution of the nonlinear conservation law (8.1), (8.2) satisfies the estimate (see [10], [57])

$$\|u(\cdot,t)\|_{Lip^+(\mathbb{R})} \leq \frac{1}{\|u_0\|^{-1}_{Lip^+(\mathbb{R})} + \alpha t}, \qquad t \geq 0; \tag{8.4}$$

the case of $\|u_0\|_{Lip^+(\mathbb{R})} = \infty$ is included in this estimate and it corresponds to the exact t^{-1} decay rate of an initial rarefaction. We also note that since u_0 has compact support the same is true of the function $x \mapsto u(x,t)$ for each $t \in (0,T]$.

Identity (8.3) is the starting point for the construction of a finite element approximation to the initial value problem (8.1), (8.2). We consider the (non-uniform) mesh $-\infty < \ldots < x_{-l} < \ldots < x_0 < \ldots < x_l < \ldots < \infty$ on the real line, and the (non-uniform) mesh $0 = t_0 < \ldots < t_m < \ldots < t_M = T$ on the interval $[0,T]$. We define the piecewise constant mesh function $x \mapsto h(x)$ whose value on $(x_{l-1}, x_l]$ is $x_l - x_{l-1}$; similarly, we define the piecewise constant mesh function $t \mapsto k(t)$ whose value on $(t_{m-1}, t_m]$ is $t_m - t_{m-1}$. On the associated partitions of $\mathbb{R}$ and $[0,T]$ we consider a pair of finite element spaces $U_h \subset L_\infty(\mathbb{R})$ and $V_k \subset L_\infty(0,T)$, respectively, and define the finite element trial space as

$$X_{hk} = U_h \otimes V_k.$$

We shall suppose that each element of U_h has compact support in $\mathbb{R}$, which is a reasonable assumption given that $u(\cdot,t)$ has compact support for each $t \in [0,T]$. We adopt the convention that elements of U_h and V_k are chosen to be continuous from the left. We then consider a (possibly different) pair of finite element spaces $S_h \subset W_\infty^1(\mathbb{R})$ and $W_k \subset W_\infty^1(0,T)$ on these two partitions, respectively, and define the finite element test space

$$Y_{hk} = S_h \otimes W_k;$$

we shall suppose that each element of S_h has compact support in $\mathbb{R}$.

The finite element approximation to (8.1), (8.2) is defined as follows: find $u_{hk} \in X_{hk}$ such that, for each $v_{hk} \in Y_{hk}$,

$$(u_{hk}(\cdot,T), v_{hk}(\cdot,T)) = (u_0, v_{hk}(\cdot,0))$$
$$+ \int_0^T (u_{hk}(\cdot,t), v_{hk,t}(\cdot,t)) + (f(u_{hk}(\cdot,t)), v_{hk,x}(\cdot,t))\,\mathrm{d}t. \tag{8.5}$$

In what follows, we shall suppose that a numerical solution u_{hk} exists and that there is a constant L_{hk} (possibly dependent on h and k) such that

$$\text{ess.sup}_{t\in[0,T]}\|u_{hk}(\cdot,t)\|_{Lip^+(\mathbb{R})} \leq L_{hk}. \tag{8.6}$$

We remark that if U_h is a finite element space consisting of continuous piecewise polynomials then this condition is trivially satisfied. The *a posteriori* error analysis of this method relies on considering a dual problem defined as follows:

$$\begin{aligned} -\frac{\partial z}{\partial t} - a(x,t)\frac{\partial z}{\partial x} &= 0, \qquad -\infty < x < \infty,\ 0 \leq t < T, \\ z(x,T) &= \psi(x), \qquad -\infty < x < \infty, \end{aligned} \tag{8.7}$$

where $\psi \in W^1_\infty(\mathbb{R})$ and

$$a(x,t) = \begin{cases} \frac{f(u(x,t))-f(u_{hk}(x,t))}{u(x,t)-u_{hk}(x,t)} & \text{if } u(x,t) \neq u_{hk}(x,t), \\ 0 & \text{otherwise.} \end{cases}$$

We note that, despite being linear, this is a non-standard problem because $a(\cdot,\cdot)$ may be discontinuous and then the classical theory of well-posedness of linear hyperbolic equations does not apply. Nevertheless, under certain hypotheses on a, it can be shown that this problem is meaningful; the next theorem, due to Tadmor (see [57], Theorem 2.2) will make this more precise.

Theorem 23 *Suppose that:*

i) $a \in L_\infty(Q)$ *where* $Q = \mathbb{R}\times(0,T)$*;*

ii) a satisfies the following one-sided Lipschitz condition:

$$\|a(\cdot,t)\|_{Lip^+(\mathbb{R})} \leq m(t), \qquad m \in L_1(0,T). \tag{8.8}$$

Then there exists a unique Lipschitz continuous function $(x,t) \mapsto z(x,t)$ *defined on* $\mathbb{R}\times[0,T]$ *which solves the backward transport equation (8.7); moreover, z obeys the estimate*

$$\|z(\cdot,t)\|_{W^1_\infty(\mathbb{R})} \leq \|\psi\|_{W^1_\infty(\mathbb{R})} \mathrm{e}^{\mu(t)}, \qquad \mu(t) \equiv \int_t^T m(\tau)\,d\tau, \ \ 0 \leq t \leq T. \tag{8.9}$$

Next we verify that the hypotheses of this theorem are satisfied with our choice of a. First note that we can write

$$a(x,t) = \int_0^1 f'((1-\xi)u(x,t) + \xi u_{hk}(x,t))\,\mathrm{d}\xi. \tag{8.10}$$

Recalling that $f''(w) \geq \alpha > 0$ for all real w, a simple calculation shows that

$$\|a(\cdot,t)\|_{Lip^+(\mathbb{R})} \leq A_{hk} \max\left(\|u(\cdot,t)\|_{Lip^+(\mathbb{R})}, \|u_{hk}(\cdot,t)\|_{Lip^+(\mathbb{R})}\right) \equiv m_{hk}(t),$$

where

$$A_{hk} = \max_{|w|\leq K_{hk}} f''(w), \qquad K_{hk} \equiv \max(\|u\|_{L_\infty(Q)}, \|u_{hk}\|_{L_\infty(Q)}).$$

Now (8.4) and (8.6) imply that (8.8) is satisfied; further, since both u and u_{hk} are in $L_\infty(Q)$, hypothesis i) of Theorem 23 is a trivial consequence of (8.10).

Now we turn to the error analysis. First, we derive a representation formula for the error. Given that $\psi \in W^1_\infty(\mathbb{R})$, let z denote the solution of the backward transport problem (8.7) with final data ψ. Then, letting $e_{hk} = u - u_{hk}$, we have that

$$\begin{aligned}
(e_{hk}(\cdot,T),\psi) &= (e_{hk}(\cdot,T), z(\cdot,T)) - \int_0^T (e_{hk}(\cdot,t), z_t + a z_x)\,\mathrm{d}t \\
&= (u(\cdot,T), z(\cdot,T)) - \int_0^T (u, z_t) + (f(u), z_x)\,\mathrm{d}t \\
&\quad -(u_{hk}(\cdot,T), z(\cdot,T)) + \int_0^T (u_{hk}, z_t) + (f(u_{hk}), z_x)\,\mathrm{d}t;
\end{aligned}$$

we remark here that since $u(\cdot,t)$ and $u_{hk}(\cdot,t)$ have compact supports in $\mathbb{R}$ for each $t \in [0,T]$, the same is true of $e_{hk}(\cdot,t)$, so the inner product which appear in this sequence of equalities are all meaningful. Noting that u obeys (8.3), we deduce that

$$\begin{aligned}
(e_{hk}(\cdot,T),\psi) &= (u_0, z(\cdot,0)) - (u_{hk}(\cdot,T), z(\cdot,T)) \\
&\quad + \int_0^T (u_{hk}, z_t) + (f(u_{hk}), z_x)\,\mathrm{d}t \\
&= (u_0, z(\cdot,0) - z_{hk}(\cdot,0)) - (u_{hk}(\cdot,T), z(\cdot,T) - z_{hk}(\cdot,T)) \\
&\quad + \int_0^T (u_{hk}, z_t - z_{hk,t}) + (f(u_{hk}), z_x - z_{hk,x})\,\mathrm{d}t \\
&\quad + (u_0, z_{hk}(\cdot,0)) - (u_{hk}(\cdot,T), z_{hk}(\cdot,T)) \\
&\quad + \int_0^T (u_{hk}, z_{hk,t}) + (f(u_{hk}), z_{hk,x})\,\mathrm{d}t,
\end{aligned}$$

for any z_{hk} in Y_{hk}. By virtue of (8.5), the expression on the right can be further reduced to give

$$\begin{aligned}
(e_{hk}(\cdot,T),\psi) &= (u_0, z(\cdot,0) - z_{hk}(\cdot,0)) - (u_{hk}(\cdot,T), z(\cdot,T) - z_{hk}(\cdot,T)) \\
&\quad + \int_0^T (u_{hk}, z_t - z_{hk,t}) + (f(u_{hk}), z_x - z_{hk,x})\,\mathrm{d}t.
\end{aligned}$$

Next we integrate by parts in order to recover the residual on the right-hand side: thus, noting that z and z_{hk} are continuous functions of x and t, that $u_{hk}(x,\cdot)$ is continuous

from the left for each $x \in \mathbb{R}$ and that $u_{hk}(\cdot, t)$ is continuous from the left at each $t \in (0, T]$, we have that

$$\begin{aligned}(e_{hk}(\cdot,T),\psi) &= (u_0 - u_{hk}(\cdot,0+), z(\cdot,0) - z_{hk}(\cdot,0)) \\ &\quad - \sum_{m=1}^{M-1} ([u_{hk}(\cdot,t_m)], z(\cdot,t_m) - z_{hk}(\cdot,t_m)) \\ &\quad - \sum_{l=-\infty}^{\infty} \int_0^T [u_{hk}(x_l,t)](z(x_l,t) - z_{hk}(x_l,t))\,\mathrm{d}t \\ &\quad - \sum_{l=-\infty}^{\infty} \sum_{m=1}^{M} \int_{x_{l-1}}^{x_l} \int_{t_{m-1}}^{t_m} (u_{hk,t} + f(u_{hk})_x)(z - z_{hk})\,\mathrm{d}x\,\mathrm{d}t,\end{aligned}$$

where

$$[u_{hk}(\cdot,t_m)] = u_{hk}(\cdot,t_m+) - u_{hk}(\cdot,t_m)$$

and

$$[u_{hk}(x_l,\cdot)] = u_{hk}(x_l+,\cdot) - u_{hk}(x_l,\cdot).$$

Thus, letting

$$\begin{aligned}r_{hk}(x,0) &= u_0(x) - u_{hk}(x,0+), \\ r_{hk}(x,t_m) &= [u_{hk}(x,t_m)], \quad x \in \mathbb{R}, \quad m = 1,\ldots,M-1, \\ r_{hk}(x_l,t) &= [u_{hk}(x_l,t)], \quad l = \ldots,-1,0,1,\ldots, \quad t \in \bigcup_{m=1}^{M} (t_{m-1},t_m), \\ r_{hk}(x,t) &= u_{hk,t}(x,t) + f(u_{hk}(x,t))_x, \quad (x,t) \in \bigcup_{l,m} (x_{l-1},x_l) \times (t_{m-1},t_m),\end{aligned}$$

we have that

$$\begin{aligned}|(e_{hk}(\cdot,T),\psi)| &\le \sum_{m=0}^{M-1} \|hr_{hk}(\cdot,t_m)\|_{L_1(\mathbb{R})} \|h^{-1}(z(\cdot,t_m) - z_{hk}(\cdot,t_m)\|_{L_\infty(\mathbb{R})} \\ &\quad + \sum_{l=-\infty}^{\infty} \|kr_{hk}(x_l,\cdot)\|_{L_1(0,T)} \|k^{-1}(z(x_l,\cdot) - z_{hk}(x_l,\cdot))\|_{L_\infty(0,T)} \\ &\quad + \sum_{l=-\infty}^{\infty} \sum_{m=1}^{M} \|r_{hk}\|_{L_1(Q_{lm})} \|z - z_{hk}\|_{L_\infty(Q_{lm})},\end{aligned}$$

where $Q_{lm} = (x_{l-1}, x_l) \times (t_{m-1}, t_m)$.

In order to proceed, we make some weak assumptions on the approximation properties of the test space: we suppose that there exists z_{hk} in Y_{hk} and positive constants C_8

and C_9, independent of h, k, z and z_{hk} such that

$$\begin{aligned}
\|h^{-1}(z(\cdot,t_m)-z_{hk}(\cdot,t_m)\|_{L_\infty(\mathbb{R})} &\leq C_8\|z_x(\cdot,t_m)\|_{L_\infty(\mathbb{R})},\\
\|k^{-1}(z(x_l,\cdot)-z_{hk}(x_l,\cdot)\|_{L_\infty(0,T)} &\leq C_9\|z_t(x_l,\cdot)\|_{L_\infty(0,T)},\\
\|z-z_{hk}\|_{L_\infty(Q_{lm})} &\leq C_8\|hz_x\|_{L_\infty(Q_{lm})}+C_9\|kz_t\|_{L_\infty(Q_{lm})}.
\end{aligned}$$

A standard finite element space consisting of continuous piecewise polynomials will certainly satisfy these inequalities (see [12] or [11]). Hence,

$$\begin{aligned}
|(e_{hk}(\cdot,T),\psi)| \leq C_8 &\sum_{m=0}^{M-1}\|hr_{hk}(\cdot,t_m)\|_{L_1(\mathbb{R})}\|z_x(\cdot,t_m)\|_{L_\infty(\mathbb{R})}\\
&+C_9\sum_{l=-\infty}^{\infty}\|kr_{hk}(x_l,\cdot)\|_{L_1(0,T)}\|z_t(x_l,\cdot)\|_{L_\infty(0,T)}\\
&+\sum_{l=-\infty}^{\infty}\sum_{m=1}^{M}\|r_{hk}\|_{L_1(Q_{lm})}\left(C_8\|hz_x\|_{L_\infty(Q_{lm})}+C_9\|kz_t\|_{L_\infty(Q_{lm})}\right).
\end{aligned}$$

Recalling the strong stability result (8.9), we have that

$$\begin{aligned}
|(e_{hk}(\cdot,T),\psi)| \leq C_8\mathrm{e}^{\mu_{hk}}&\|\psi\|_{W^1_\infty(\mathbb{R})}\sum_{m=0}^{M-1}\|hr_{hk}(\cdot,t_m)\|_{L_1(\mathbb{R})}\\
&+C_9\mathrm{e}^{\mu_{hk}}\|\psi\|_{W^1_\infty(\mathbb{R})}\sum_{l=-\infty}^{\infty}\|a(x_l,\cdot)\|_{L_\infty(0,T)}\|kr_{hk}(x_l,\cdot)\|_{L_1(0,T)}\\
&+\mathrm{e}^{\mu_{hk}}\|\psi\|_{W^1_\infty(\mathbb{R})}\sum_{l=-\infty}^{\infty}\sum_{m=1}^{M}\left(C_8\|hr_{hk}\|_{L_1(Q_{lm})}\right.\\
&\left.+C_9\|a\|_{L_\infty(Q_{lk})}\|kr_{hk}\|_{L_1(Q_{lm})}\right),
\end{aligned}$$

where

$$\mu_{hk}=\int_0^T m_{hk}(t)\,\mathrm{d}t.$$

Upon dividing by $\|\psi\|_{W^1_\infty(\mathbb{R})}$ and taking the supremum over all ψ, we deduce that

$$\begin{aligned}
\|u(\cdot,T)-u_{hk}(\cdot,T)\|_{Lip'(\mathbb{R})} \leq C_8\mathrm{e}^{\mu_{hk}} &\sum_{m=0}^{M-1}\|hr_{hk}(\cdot,t_m)\|_{L_1(\mathbb{R})}\\
&+C_9\mathrm{e}^{\mu_{hk}}\sum_{l=-\infty}^{\infty}\|a(x_l,\cdot)\|_{L_\infty(0,T)}\|kr_{hk}(x_l,\cdot)\|_{L_1(0,T)}\\
&+C_8\mathrm{e}^{\mu_{hk}}\sum_{l=-\infty}^{\infty}\sum_{m=1}^{M}\|hr_{hk}\|_{L_1(Q_{lm})}\\
&+C_9\mathrm{e}^{\mu_{hk}}\sum_{l=-\infty}^{\infty}\sum_{m=1}^{M}\|a\|_{L_\infty(Q_{lk})}\|kr_{hk}\|_{L_1(Q_{lm})},
\end{aligned}\tag{8.11}$$

where we used the notation

$$\|w\|_{Lip'(\mathbb{R})} = \sup_{\psi \in W^1_\infty(\mathbb{R})} \frac{|(w,\psi)|}{\|\psi\|_{W^1_\infty(\mathbb{R})}}$$

for the dual Lipschitz (Lip') norm.

Thus we have proved an *a posteriori* bound on the error between u and its finite element approximation u_{hk} of the form

$$\|u(\cdot,T) - u_{hk}(\cdot,T)\|_{Lip'(\mathbb{R})} \leq \text{Const.}\, \eta_Q(u_{hk}),$$

where $\eta_Q(u_{hk})$ is the *a posteriori* error estimator on $Q = \mathbb{R} \times (0,T)$, appearing on the right-hand side of inequality (8.11). We note that since $u_{hk}(\cdot,t)$ has compact support in $\mathbb{R}$ for each $t \in [0,T]$, the same is true of r_{hk}, so each of the infinite sums appearing in (8.11) collapses to a summation over a finite number of terms.

Instead of using, as we have, a tensor-product grid on Q, we could have considered an unstructured space-time triangulation $\mathcal{T}_h$ of Q with associated finite element trial space $X_h \subset L_\infty(Q)$ and test space $Y_h \subset W^1_\infty(Q)$ instead of X_{hk} and Y_{hk}, respectively. Thus, repeating the same argument as above, we would have arrived at the following *a posteriori* error bound for the numerical solution $u_h \in X_h$ (defined similarly as u_{hk} before):

$$\|u(\cdot,T) - u_h(\cdot,T)\|_{Lip'(\mathbb{R})} \leq \text{Const.}\, \eta_Q(u_h),$$

where

$$\eta_Q(u_h) = \|hr_h^0\|_{L_1(\mathbb{R})} + \sum_{\kappa \in \mathcal{T}_h} \left(\|hr_h\|_{L_1(\kappa)} + \sum_{e \subset \partial\kappa \setminus \partial Q} \|h\hat{r}_h\|_{L_1(e)} \right),$$

with e denoting an edge of a triangle κ in $\mathcal{T}_h$,

$$\begin{aligned} r_h^0(x) &= u_0(x) - u_h(x,0+), \quad x \in \mathbb{R}, \\ r_h(x,t) &= u_{h,t}(x,t) + f(u_h(x,t))_x, \quad (x,t) \in \kappa, \\ \hat{r}_h(x,t) &= [u_h(x,t)\nu_t(x,t) + f(u_h(x,t))\nu_x(x,t)], \quad (x,t) \in e \subset \partial\kappa \setminus \partial Q, \end{aligned}$$

where (ν_x, ν_t) is the unit outward normal to edge e (with respect to the triangle κ), and $[w]$ signifies the jump of w across e. In the case of $X_h = Y_h$, a second-order numerical dissipation term could have also been included into the finite element method; the effects of this on the *a posteriori* error bound can be analysed similarly to the streamline diffusion stabilisation discussed earlier on.

To conclude this section, we note that a direct approach, different from ours, to the *a posteriori* error analysis of numerical approximations to scalar nonlinear conservation laws was pursued in the work of Cockburn and Gau; we refer to [13], [14] for further details.

9 Conclusions

We have presented an overview of recent developments which concern the *a posteriori* error analysis of finite element approximations to linear and nonlinear hyperbolic partial differential equations of first order. We derived various global and local bounds on the discretisation error, and investigated the question of error localisation and error propagation.

While for elliptic equations there is already a well-established theoretical framework of *a posteriori* error estimation which has been successfully implemented into working adaptive algorithms, very much less is known about these issues in the context of hyperbolic and nearly-hyperbolic problems. However, this is now a field of active research, and the outcome of these investigations can make a large impact on the design of reliable numerical algorithms for large-scale computations in engineering applications.

References

[1] Adams, R.A. (1975). *Sobolev Spaces.* Academic Press.

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

[3] Babuška, I. and Aziz, A.K. (1972). *Survey lectures on the mathematical foundation of the finite element method.* In: The Mathematical Foundations of the Finite Element Method, A.K. Aziz and I. Babuška, (Eds.), Academic Press.

[4] Baiocchi, C. and Capelo, A. (1984). *Variational and Quasi-Variational Inequalities: Applications to Free Boundary Problems.* John Wiley & Sons.

[5] Balland, P. and Süli, E. (1997). Analysis of the cell vertex scheme for hyperbolic problems with variable coefficients. *SIAM J. Numer. Anal.,* **34**, 1127–1151.

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

[7] Becker, R. and Rannacher, R. (1996). Weighted a posteriori error control in finite element methods. *Technical Report, Universität Heidelberg*, Preprint No. 96-01.

[8] Bergh, I. and Löfström, J. (1976). *Interpolation Spaces.* Springer-Verlag, Grundlehren der Mathematischen Wissenschaften 223.

[9] Bernardi, C. (1989). Optimal finite-element interpolation on curved domains. *SIAM J. Numer. Anal.*, **26**, 1212–1240.

[10] Brenier, Y. and Osher, S. (1988). The discrete one-sided Lipschitz condition for convex scalar conservation laws. *SIAM J. Numer. Anal.,* **25**, 8–23.

[11] Brenner, S.C. and Scott, L.R. (1997). *The Mathematical Theory of Finite Element Methods.* 2nd corr. ed. Springer-Verlag. Texts in Applied Mathematics 15.

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

[13] Cockburn, B. and Gau, H. (1995). A posteriori error estimates for general numerical methods for scalar conservation laws. *Mat. Aplic. Comp.*, **14**, No. 1, 37–47.

[14] Cockburn, B. and Gremaud, P.-A. (1996). Error estimates for finite element methods for scalar conservation laws. *SIAM J. Numer. Anal.*, **33**, 522–554.

[15] Eriksson, K., Estep, D., Hansbo, P., and Johnson, C. (1995). Introduction to Adaptive Methods for Differential Equations. *Acta Numerica.* Cambridge University Press. 105–158.

[16] Friedrichs, K.O. (1958). Symmetric positive linear differential equations. *Comm. Pure Appl. Math.*, **11**, 333–418.

[17] Führer, C. (1997). *A posteriori error control for nonlinear hyperbolic problems.* Ph.D. Thesis, SFB 359, Universität Heidelberg.

[18] Giles, M.B. (1997). On adjoint equations for error analysis and optimal grid adaptation in CFD. *Oxford University Computing Laboratory Technical Report, NA97/11.*

[19] Giles, M.B., Larson, M.G., Levenstam, M., and Süli, E. (1997). Adaptive error control for finite element approximations of the lift and drag in a viscous flow. *Oxford University Computing Laboratory Technical Report, NA97/06.*

[20] Girault, V. and Raviart, P.-A. (1979). *Finite Element Approximation of the Navier-Stokes Equations.* Lecture Notes in Mathematics 749. Springer-Verlag.

[21] Godlewski, E. and Raviart, P.-A. (1996). *Numerical Approximation of Hyperbolic Systems of Conservation Laws.* Series in Applied Mathematical Sciences 118. Springer-Verlag.

[22] Hairer, E., Norsett, S., and Wanner, G. (1993). *Solving ordinary differential equations.* 2nd rev. ed. Series in Computational Mathematics 8. Springer-Verlag.

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

[24] Hebeker, F.-K, Führer, C., and Rannacher, R. (1997). An adaptive finite element method for unsteady convection-dominated flows with stiff source terms. *Preprint (SFB 359), Universität Heidelberg.*

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

[26] 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.*

[27] Houston, P. and Süli, E. (1997). Local *a posteriori* error analysis for hyperbolic problems. *Oxford University Computing Laboratory Technical Report, NA97/14.*

[28] Johnson, C. (1990). Adaptive finite element methods for diffusion and convection problems. *Computer Methods in Applied Mechanics and Engineering*, **82**, 301–322.

[29] 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, 105–120.

[30] Johnson, C. and Hansbo, P. (1992). Adaptive finite element methods in computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, **101**, 143–181.

[31] Johnson, C. and Szepessy, A. (1995). Adaptive finite element methods for conservation laws based on a posteriori estimates. *Comm. Pure Appl. Math.*, **48**, 199–243.

[32] Kohn, J.J. and Nirenberg, L. (1965). Non-coercive boundary value problems. *Comm. Pure Appl. Math.* **18**, 443–492.

[33] Kröner, D. (1997). *Numerical Schemes for Conservation Laws.* John Wiley & Sons and B.G. Taubner Publishers.

[34] Kufner, A., John, O., and Fučik, S. (1977) *Function Spaces.* Noordhoff International Publishing.

[35] Lax, P.D. (1955). On the Cauchy problem for hyperbolic equations and the differentiability of solutions of elliptic equations. *Comm. Pure Appl. Math.*, **8**, 615–633.

[36] Lax, P.D. and Phillips, R.S. (1960). Local boundary conditions for dissipative symmetric linear differential operators. *Comm. Pure Appl. Math.*, **13**, 427–455.

[37] Lesaint, P. (1973). Finite element methods for symmetric hyperbolic equations. *Numer. Math.,* **21**, 244–255.

[38] Lesaint, P. and Raviart, P.-A. (1979). Finite element collocation methods for first order systems. *Math. Comput.*, **33**, 891–918.

[39] 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, 289–298.

[40] 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**, 221–235.

[41] 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).

[42] Melenk, J.M., and Schwab, C. (1997). An *hp* finite element method for convection-diffusion problems. Reseach Report No 97-05, Seminar für Angewandte Mathematik, ETH, Zürich.

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

[44] 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, 267–288.

[45] Morton, K.W. and Süli, E. (1995). Evolution Galerkin methods and their supra-convergence. *Numerische Mathematik*, **71**, 331–355.

[46] Nečas, J. (1967). *Les méthodes directes en théorie des équations elliptiques.* Masson, Paris.

[47] Peraire, J., Paraschivoiu, M., and Patera, A. (1996). A posteriori finite element bounds for linear functional outputs of elliptic partial differential equations. *Symposium on Advances in Computational Mechanics.* Submitted to Comp. Meth. Appl. Engnrg.

[48] Rannacher, R. and Suttmeier F.-T. (1996). A feed-back approach to error control in finite element methods: application to linear elasticity. *Preprint 96-42 (SFB 359), University of Heidelberg.*

[49] Rauch, J. (1972) $\mathcal{L}_2$ is a continuable initial condition for Kreiss' mixed problems. *Comm. Pure Appl. Math.,* **25**, 265–285.

[50] Sandboge, R. (1996). *Adaptive Finite Element Methods for Reactive Flow Problems.* Ph.D. Thesis. Department of Mathematics. Göteborg.

[51] 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.* (To appear in Numerische Mathematik.)

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

[53] Süli, E. (1992). The accuracy of cell vertex finite volume methods on quadrilateral meshes. *Math. Comput.*, **59**, 359–382.

[54] 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, 253–260.

[55] Süli, E. (1996). A posteriori error analysis and global error control for adaptive finite element approximations of hyperbolic problems. In: D.F. Griffiths and G.A. Watson, eds. *Numerical Analysis 1995*, Pitman Lecture Notes in Mathematics Series 344, 169–190.

[56] Süli, E. and Houston, P. (1997). Finite element methods for hyperbolic problems: a posteriori error analysis and adaptivity. In: I.S. Duff and G.A. Watson, eds. *The State of the Art in Numerical Analysis*, Clarendon Press, Oxford, 441–471.

[57] Tadmor, E. (1991). Local error estimates for discontinuous solutions of nonlinear hyperbolic equations. *SIAM J. Numer. Anal.*, **28**, 891–906.

[58] Tartakoff, D. (1972). Regularity of solutions to boundary value problmes for first order systems. *Indiana University Mathematics Journal*, **21**, No. 12, 1113–1129.

[59] Szabó, B. and Babuška, I. (1991). *Finite Element Analysis*. J. Wiley & Sons, New York.

[60] Verfürth, R. (1996). *A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. B.G. Teubner, Stuttgart.

[61] Winther, R. (1981). A stable finite element method for initial boundary value problems for first-order hyperbolic systems. *Math. Comput.*, **36**, 65–86.