

# ADAPTIVE DISCONTINUOUS GALERKIN FINITE ELEMENT METHODS FOR NONLINEAR HYPERBOLIC CONSERVATION LAWS\*

RALF HARTMANN<sup>†</sup> AND PAUL HOUSTON<sup>‡</sup>

**Abstract.** We consider the a posteriori error analysis and adaptive mesh design for discontinuous Galerkin finite element approximations to systems of nonlinear hyperbolic conservation laws. In particular, we discuss the question of error estimation for general target functionals of the solution; typical examples include the outflow flux, local average and pointwise value, as well as the lift and drag coefficients of a body immersed in an inviscid fluid. By employing a duality argument, we derive so-called weighted or Type I a posteriori error bounds; these error estimates include the product of the finite element residuals with local weighting terms involving the solution of a certain dual or adjoint problem that must be numerically approximated. Based on the resulting approximate Type I bound, we design and implement an adaptive algorithm that produces meshes specifically tailored to the efficient computation of the given target functional of practical interest. The performance of the proposed adaptive strategy and the quality of the approximate Type I a posteriori error bound is illustrated by a series of numerical experiments. In particular, we demonstrate the superiority of this approach over standard mesh refinement algorithms which employ Type II error indicators.

**Key words.** discontinuous Galerkin, a posteriori error analysis, adaptivity, nonlinear hyperbolic conservation laws

**AMS subject classifications.** 65N12, 65N15, 65N30

**PII.** S1064827501389084

**1. Introduction.** The ever-increasing range of applications of compressible fluid dynamics is a fertile source of difficult and challenging problems with important implications in engineering design. Multidimensional compressible flows are modeled by nonlinear conservation laws whose solutions exhibit a wide range of localized structures, such as shock waves, contact discontinuities, and rarefaction waves. The accurate numerical resolution of these features necessitates the use of locally refined, adaptive computational meshes. The majority of adaptive algorithms used in computational fluid dynamics applications will simply refine or adjust the mesh according to an ad hoc criterion, such as a large gradient in a physical quantity or the size of the local residual, for example. Although this intuitive approach has had some success, it may not provide the most economical mesh design for the control of a given error quantity of interest; cf. [15, 17, 18, 25, 26].

In this paper we consider the a posteriori error analysis and adaptive mesh design for discontinuous Galerkin finite element approximations to systems of nonlinear hyperbolic conservation laws. In particular, we derive computable error estimates for general target functionals of the solution of practical interest; relevant examples include the lift and drag coefficients of a body immersed in an inviscid fluid, the local

\*Received by the editors May 8, 2001; accepted for publication (in revised form) May 28, 2002; published electronically December 19, 2002.

<http://www.siam.org/journals/sisc/24-3/38908.html>

<sup>†</sup>Institute of Applied Mathematics, University of Heidelberg, Im Neuenheimer Feld 293, D-69120 Heidelberg, Germany (Ralf.Hartmann@iwr.uni-heidelberg.de). The research of this author was supported by the DFG Priority Research Program and the SFB 359 at the IWR, University of Heidelberg.

<sup>‡</sup>Department of Mathematics and Computer Science, University of Leicester, Leicester LE1 7RH, UK (Paul.Houston@mcs.le.ac.uk). The research of this author was supported by the EPSRC under grant GR/N24230.

mean value of the field or its flux through the outflow boundary of the computational domain  $\Omega$ , and the point evaluation of the solution at a given point in  $\Omega$ .

By employing a duality argument, we derive an error representation formula which equates the error in an output functional of interest to the inner product of the finite element residuals with weighting terms involving the solution of a dual problem whose data is the density function of the target functional. On the basis of this error representation formula, Type I (cf. [21]) a posteriori estimates are derived which, unlike Type II bounds, retain the weighting terms involving the dual solution. These weights represent the sensitivity of the relevant error quantity with respect to variations of the local element residuals. Since the solution to the dual problem is, usually, analytically unknown, it has to be approximated numerically as part of the error estimation process. The cost of this additional calculation is, in general, a relatively small fraction of the total cost for the solution process, since the dual problem comprises a system of *linear* partial differential equations. Moreover, the additional costs involved in solving linearized dual problems also must be compared to the usually much higher cost of working with nonoptimized meshes designed by traditional Type II error indicators, which do not require the solution of an auxiliary problem. On the basis of the resulting a posteriori error estimate, the mesh is locally adapted and a new approximation to the primal and dual solutions is computed. This algorithm is repeated until the error in the prescribed functional is below some user-defined tolerance, for example. This procedure leads to optimal finite element meshes specifically tailored to the accurate computation of the selected target functional of physical interest. For related work in this area, we refer to [4, 24, 25, 26].

The paper is structured as follows. After introducing, in section 2, our model system of nonlinear hyperbolic conservation laws, in section 3 we formulate its discontinuous Galerkin finite element approximation. Then, in section 4 we derive an error representation formula together with Type I and Type II a posteriori error bounds for general target functionals of the solution. The error representation formula stems from a duality argument and includes computable residual terms multiplied by local weights involving the dual solution. Thereby, in section 5 we discuss how the dual problem may be numerically determined so that approximations to the error representation formula and the Type I a posteriori bound may be computed. On the basis of the approximate Type I error bound, in section 6 we design and implement an adaptive algorithm that produces meshes specifically tailored to the efficient computation of the target functional of practical interest. The performance of the proposed adaptive strategy, and the quality of the approximate error representation formula and approximate Type I a posteriori bound, are then studied in section 7 through a series of numerical experiments. Here, we observe that, even though the dual problem is computed numerically, the approximate error representation formula provides a sharp estimate of the true error in the computed target functional and also that the resulting approximate Type I estimate provides an asymptotic upper bound on the error, in the sense that error control is achieved when the granularity  $h$  of the computational mesh is sufficiently fine. Furthermore, we demonstrate the superiority of using Type I a posteriori error indicators over traditional Type II indicators, which do not require the solution of an auxiliary problem. Finally, in section 8 we summarize the work presented in this paper and draw some conclusions. The work presented in this paper is a complete and improved account of our recent work announced in [15, 17].

**2. Model problem.** Given a final time  $T > 0$ , we consider the system

$$(2.1) \quad \partial_t \mathbf{u} + \sum_{i=1}^d \frac{\partial}{\partial x_i} \mathbf{f}_i(\mathbf{u}) = 0 \quad \text{in } (0, T] \times \Omega, \quad \mathbf{u}(0, \cdot) = \mathbf{u}_0(\cdot) \quad \text{in } \Omega,$$

where  $\Omega$  is an open connected subset of  $\mathbb{R}^d$ ,  $d \geq 1$ ,  $\mathbf{u} = (u_1, \dots, u_m)^T$ , and  $\mathbf{f}_i : \mathbb{R}^m \rightarrow \mathbb{R}^m$ ,  $i = 1, \dots, d$ , are continuously differentiable. We note that (2.1) is hyperbolic when  $A(\mathbf{u}, \nu) := \sum_{i=1}^d \nu_i A_i(\mathbf{u})$  has  $m$  real eigenvalues and a complete set of linearly independent eigenvectors for all vectors  $\nu = (\nu_1, \dots, \nu_d) \in \mathbb{R}^d$ . Here, the  $A_i(\mathbf{u})$  denote the Jacobi matrices of the fluxes  $\mathbf{f}_i(\mathbf{u})$ ,  $i = 1, \dots, d$ , i.e.,  $A_i(\mathbf{u}) := \nabla_{\mathbf{u}} \mathbf{f}_i(\mathbf{u})$ . An important example of such a model problem is represented by the Euler equations of compressible gas dynamics. In two space dimensions, the state vector  $\mathbf{u} = (\rho, \rho u, \rho v, \rho E)$ , where  $\rho$ ,  $(u, v)$ , and  $E$  represent the density, Cartesian velocity, and total energy per unit mass, respectively. Furthermore, the fluxes are defined by  $\mathbf{f}_1 = (\rho u, \rho u^2 + p, \rho uv, \rho H u)$  and  $\mathbf{f}_2 = (\rho v, \rho uv, \rho v^2 + p, \rho H v)$ , respectively. Here,  $p$  denotes the pressure, while  $H$  is the total enthalpy defined by  $H = E + p/\rho$ . The equation of state of an ideal gas is given by  $p = (\gamma - 1)\rho(E - (u^2 + v^2)/2)$ , where  $\gamma$  is the ratio of specific heats which, for dry air, is  $\gamma = 1.4$ .

In the following, we assume that  $\Omega$  is a bounded region in  $\mathbb{R}^d$ , with boundary  $\Gamma$ ; in this case the system of conservation laws (2.1) must be supplemented by appropriate boundary conditions; for example, at inflow/outflow boundaries, we require that  $A^-(\mathbf{u}, \mathbf{n})(\mathbf{u} - \mathbf{g}) = \mathbf{0}$ , where  $\mathbf{n}$  denotes the unit outward normal vector to the boundary  $\Gamma$  and  $\mathbf{g}$  is a (given) real-valued vector function. Here,  $A^-(\mathbf{u}, \mathbf{n})$  denotes the negative part of  $A(\mathbf{u}, \mathbf{n})$ , i.e.,  $A^-(\mathbf{u}, \mathbf{n}) = R\Lambda^-L$ , where  $L$  and  $R$  denote the  $m \times m$  matrices of left and right eigenvectors of  $A(\mathbf{u}, \mathbf{n})$ , respectively, and  $\Lambda^- = \text{diag}(\min(\lambda_i, 0))$  denotes the  $m \times m$  diagonal matrix of the negative eigenvalues of  $A(\mathbf{u}, \mathbf{n})$ .

For notational convenience, we denote by  $\Omega_T$  the set  $(0, T] \times \Omega$ , by  $\Gamma_T$  the boundary of  $\Omega_T$ , and by  $\text{div}$  the divergence with respect to the variable  $\mathbf{x} = (x_0, x_1, \dots, x_d)$ , where  $x_0$  denotes time  $t$ . In addition, we introduce the flux function  $\mathcal{F}(\mathbf{u}) = (\mathbf{f}_0(\mathbf{u}), \mathbf{f}_1(\mathbf{u}), \dots, \mathbf{f}_d(\mathbf{u}))$ , with  $\mathbf{f}_0 = \mathbf{u}$ . Thereby, we may write the system of conservation laws (2.1) in the following compact form: Find  $\mathbf{u} : \Omega_T \rightarrow \mathbb{R}^m$ , such that

$$(2.2) \quad \text{div} \mathcal{F}(\mathbf{u}) = 0 \quad \text{in } \Omega_T,$$

subject to the initial condition in (2.1) and appropriate boundary conditions.

**3. Discontinuous Galerkin method.** We begin by introducing the necessary notation. Suppose that  $\mathcal{T}_h$  is a (space-time) subdivision of  $\Omega_T$  into open element domains  $\kappa$  such that  $\bar{\Omega}_T = \cup_{\kappa \in \mathcal{T}_h} \bar{\kappa}$ . Here,  $h$  denotes the piecewise constant mesh function defined by  $h|_{\kappa} \equiv h_{\kappa} = \text{diam}(\kappa)$  for all  $\kappa \in \mathcal{T}_h$ . Let us assume that the family of subdivisions  $\mathcal{T}_h$  is shape regular and that each  $\kappa \in \mathcal{T}_h$  is a smooth bijective image of a fixed reference element  $\hat{\kappa}$ , that is,  $\kappa = F_{\kappa}(\hat{\kappa})$  for all  $\kappa \in \mathcal{T}_h$ . On the reference element  $\hat{\kappa}$  we define spaces of polynomials of degree  $p \geq 0$  as follows:

$$\mathcal{Q}_p = \text{span} \{ \hat{\mathbf{x}}^{\alpha} : 0 \leq \alpha_i \leq p, 0 \leq i \leq d \}, \quad \mathcal{P}_p = \text{span} \{ \hat{\mathbf{x}}^{\alpha} : 0 \leq |\alpha| \leq p \}.$$

We now introduce the finite element space  $S_{h,p}$ , consisting of discontinuous vector (of dimension  $m$ )-valued polynomial functions of degree  $p \geq 0$ , defined by

$$S_{h,p} = \{ \mathbf{v} \in [L_2(\Omega_T)]^m : \mathbf{v}|_{\kappa} \circ F_{\kappa} \in [\mathcal{Q}_p]^m \text{ if } \hat{\kappa} = F_{\kappa}^{-1}(\kappa) \text{ is the unit hypercube,} \\ \text{and } \mathbf{v}|_{\kappa} \circ F_{\kappa} \in [\mathcal{P}_p]^m \text{ if } \hat{\kappa} = F_{\kappa}^{-1}(\kappa) \text{ is the unit simplex; } \kappa \in \mathcal{T}_h \}.$$

Given that  $v \in H^1(\kappa)$  for each  $\kappa \in \mathcal{T}_h$ , we denote by  $v^+$  and  $v^-$  the interior and exterior traces of  $v$ , respectively, defined on  $\partial\kappa$  and  $\partial\kappa \setminus \Gamma_T$ , respectively; cf. [21].

To formulate the discontinuous Galerkin method, we first introduce a weak formulation of (2.2). To this end, we multiply (2.2) by an arbitrary smooth function  $\mathbf{v}$  and integrate by parts over an element  $\kappa$  in the mesh  $\mathcal{T}_h$ ; thereby, we get

$$(3.1) \quad - \int_{\kappa} \mathcal{F}(\mathbf{u}) \cdot \nabla \mathbf{v} \, d\mathbf{x} + \int_{\partial\kappa} \mathcal{F}(\mathbf{u}^+) \cdot \mathbf{n}_{\kappa} \mathbf{v}^+ \, ds = 0,$$

where  $\nabla$  denotes the gradient operator with respect to the variable  $\mathbf{x} = (x_0, x_1, \dots, x_d)$  and  $\mathbf{n}_{\kappa}$  denotes the unit outward normal vector to  $\partial\kappa$ .

To discretize (3.1), we replace the analytical solution  $\mathbf{u}$  by the Galerkin finite element approximation  $\mathbf{u}_h$  and the test function  $\mathbf{v}$  by  $\mathbf{v}_h$ , where both  $\mathbf{u}_h$  and  $\mathbf{v}_h$  belong to the finite element space  $S_{h,p}$ . In addition, since the numerical solution  $\mathbf{u}_h$  is discontinuous between element interfaces, we must replace the flux  $\mathcal{F}(\mathbf{u}^+) \cdot \mathbf{n}_{\kappa}$  by a numerical flux function  $\mathcal{H}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}_{\kappa})$ , which depends on both the inner- and outer-trace of  $\mathbf{u}_h$  on  $\partial\kappa$ ,  $\kappa \in \mathcal{T}_h$ , and the unit outward normal vector  $\mathbf{n}_{\kappa}$  to  $\partial\kappa$ . Thereby, summing over the elements  $\kappa$  in the mesh  $\mathcal{T}_h$  yields the discontinuous Galerkin finite element discretization of (2.2) as follows: Find  $\mathbf{u}_h \in S_{h,p}$  such that

$$(3.2) \quad \sum_{\kappa \in \mathcal{T}_h} \left\{ - \int_{\kappa} \mathcal{F}(\mathbf{u}_h) \cdot \nabla \mathbf{v}_h \, d\mathbf{x} + \int_{\partial\kappa} \mathcal{H}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}_{\kappa}) \mathbf{v}_h^+ \, ds \right\} = 0 \quad \forall \mathbf{v}_h \in S_{h,p};$$

cf. [3, 9, 10]. Here, for elements  $\kappa \in \mathcal{T}_h$  whose boundary intersects that of the computational domain  $\Omega_T$ , we replace  $\mathbf{u}_h^-$  by the appropriate boundary/initial conditions on the portion of  $\partial\kappa$  for which  $\partial\kappa \cap \Gamma_T \neq \emptyset$ .

We emphasize that the choice of the numerical flux function is *independent* of the finite element space employed. Indeed, the numerical flux function  $\mathcal{H}(\cdot, \cdot, \cdot)$  may be chosen to be any two-point monotone Lipschitz function which is both consistent and conservative; see [23, 28], for example. In section 7 we consider two different numerical fluxes: on triangular elements, we employ the (local) Lax–Friedrichs flux, while on quadrilateral elements, we implement the Vijayasundaram flux.

The (local) Lax–Friedrichs flux  $\mathcal{H}_{LF}(\cdot, \cdot, \cdot)$ , is defined by

$$(3.3) \quad \mathcal{H}_{LF}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}_{\kappa})|_{\partial\kappa} = \frac{1}{2} (\mathcal{F}(\mathbf{u}_h^+) \cdot \mathbf{n}_{\kappa} + \mathcal{F}(\mathbf{u}_h^-) \cdot \mathbf{n}_{\kappa} - \alpha (\mathbf{u}_h^- - \mathbf{u}_h^+)) ,$$

for  $\kappa$  in  $\mathcal{T}_h$ , where  $\alpha$  is an estimate of the largest eigenvalue (in absolute value) of the Jacobi matrix  $\mathcal{A}(\mathbf{u}_h, \mathbf{n}_{\kappa}) = \mathcal{F}'(\mathbf{u}_h) \cdot \mathbf{n}_{\kappa} \equiv \sum_{i=0}^d A_i(\mathbf{u}_h) (\mathbf{n}_{\kappa})_i$ , in the neighborhood of  $\partial\kappa$ . Here,  $A_i$ ,  $i = 0, \dots, d$ , are the Jacobi matrices of the fluxes  $\mathbf{f}_i$ , respectively; for  $i = 0$ , we have  $\mathbf{f}_0 = \mathbf{u}$ , and thus  $A_0$  is the identity. Furthermore,  $(\mathbf{n}_{\kappa})_i$  denotes the  $i$ th component of the unit outward normal vector  $\mathbf{n}_{\kappa}$  to element  $\kappa$ ,  $\kappa \in \mathcal{T}_h$ , for  $i = 0, \dots, d$ .

In the case where the fluxes  $\mathbf{f}_i$ ,  $i = 0, \dots, d$ , are homogeneous of degree 1, the Vijayasundaram flux  $\mathcal{H}_V(\cdot, \cdot, \cdot)$  is defined by

$$(3.4) \quad \mathcal{H}_V(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}_{\kappa})|_{\partial\kappa} = \mathcal{A}^+(\hat{\mathbf{u}}_h, \mathbf{n}_{\kappa}) \mathbf{u}_h^+ + \mathcal{A}^-(\hat{\mathbf{u}}_h, \mathbf{n}_{\kappa}) \mathbf{u}_h^- \quad \text{for } \kappa \in \mathcal{T}_h,$$

where  $\mathcal{A}^+(\hat{\mathbf{u}}_h, \mathbf{n}_{\kappa})$  and  $\mathcal{A}^-(\hat{\mathbf{u}}_h, \mathbf{n}_{\kappa})$  denote the positive and negative parts of the Jacobi matrix  $\mathcal{A}(\hat{\mathbf{u}}_h, \mathbf{n}_{\kappa})$ , respectively, evaluated at some average state  $\hat{\mathbf{u}}_h$  between  $\mathbf{u}_h^+$  and  $\mathbf{u}_h^-$ . Extensions to homogeneous fluxes of degree greater than 1 follow analogously.

Finally, we remark that even when the fluxes at interelement boundaries are upwinded through the choice of an appropriate numerical flux function  $\mathcal{H}(\cdot, \cdot, \cdot)$ , spurious oscillations still may be generated by the discontinuous Galerkin finite element method (3.2), in the vicinity of sharp features of the analytical solution, when higher-order degree polynomials are used, i.e., when  $p > 0$ . In order to compute physically relevant solutions that do not exhibit such oscillatory behavior, the basic discontinuous Galerkin scheme (3.2) must be enhanced by the addition of some form of nonlinear dissipation mechanism which does not adversely affect the formal order of accuracy of the scheme. Indeed, much research activity has focused on the development of appropriate stabilization devices; for example, Cockburn and Shu [10] proposed the use of local projection/limiting techniques for the Runge–Kutta discontinuous Galerkin finite element method; see also Biswas, Devine, and Flaherty [5]. Giannakouros and Karniadakis [12] considered the implementation of the flux corrected transport algorithm for spectral element approximations of nonlinear hyperbolic problems. For a review of these and other techniques, we refer to Gottlieb and Hesthaven [14].

The a posteriori error analysis presented in the proceeding section is based on a hyperbolic duality argument employing the Galerkin orthogonality property of the finite element method; cf. (3.7) below. Thereby, any stabilization technique used to enhance the numerical performance of the standard discontinuous Galerkin method (3.2) should not violate this orthogonality condition. To this end, we add to the scheme (3.2) an artificial viscosity term (also referred to as a shock-capturing term) that depends on both the mesh function  $h$  and the finite element residual. The dependence of the shock-capturing term on the residual of the underlying conservation law ensures the consistency of the resulting finite element method for smooth analytical solutions.

Thereby, we now consider the following discontinuous Galerkin finite element discretization of (2.2): Find  $\mathbf{u}_h \in S_{h,p}$  such that

$$\begin{aligned} \mathcal{N}(\mathbf{u}_h, \mathbf{v}_h) \equiv \sum_{\kappa \in \mathcal{T}_h} \left\{ - \int_{\kappa} \mathcal{F}(\mathbf{u}_h) \cdot \nabla \mathbf{v}_h \, d\mathbf{x} + \int_{\partial\kappa} \mathcal{H}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}_\kappa) \mathbf{v}_h^+ \, ds \right. \\ \left. + \int_{\kappa} \varepsilon \nabla \mathbf{u}_h \cdot \nabla \mathbf{v}_h \, d\mathbf{x} \right\} = 0 \quad \forall \mathbf{v}_h \in S_{h,p}. \end{aligned} \quad (3.5)$$

Here,  $\varepsilon$  denotes the artificial viscosity coefficient matrix defined by

$$\varepsilon = C_\varepsilon h^{2-\beta} |\operatorname{div} \mathcal{F}(\mathbf{u}_h)| I_d, \quad (3.6)$$

where  $C_\varepsilon$  and  $0 < \beta < 1/2$  are positive constants and  $I_d = \operatorname{diag}(0, 1, \dots, 1)$ . The definition of  $\varepsilon$  in (3.6) represents a slight modification of the artificial viscosity model introduced and analyzed by Jaffre, Johnson, and Szepessy [22]; here, we have simplified the definition of  $\varepsilon$  to include only the *interior* residual term  $|\operatorname{div} \mathcal{F}(\mathbf{u}_h)|$ ; cf. (4.9). We note that in [22] the discontinuous Galerkin finite element method (3.5), with the addition of streamline-diffusion stabilization and the numerical flux function  $\mathcal{H}(\cdot, \cdot, \cdot)$  defined as the (global) Lax–Friedrichs flux, was shown to be convergent for any approximating polynomial of degree  $p \geq 0$  on general meshes for multidimensional scalar problems and shown to be entropy consistent for multidimensional systems. Furthermore, in [8] Cockburn and Gremaud derived both a priori and a posteriori error estimates for a discontinuous Galerkin method with a similar type of shock-capturing applied to multidimensional scalar conservation laws. These estimates were then employed to deduce that the method converges to the entropy solution in the  $L_\infty(0, T; L_2(\Omega))$  norm with a rate of at least  $h^{1/4}$  as  $h$  tends to zero.

Finally, inserting the exact solution  $\mathbf{u}$  into the semilinear form  $\mathcal{N}(\cdot, \cdot)$  defined in (3.5) and using both the dependence of the shock-capturing term on the residual and the consistency of the numerical flux function  $\mathcal{H}(\cdot, \cdot, \cdot)$ , we deduce that if the analytical solution to (2.2) is sufficiently regular, then  $\mathcal{N}(\mathbf{u}, \mathbf{v}_h) = 0$  for all  $\mathbf{v}_h$  in  $S_{h,p}$ . Thereby, in the nonlinear case, the Galerkin orthogonality property of the finite element method (3.5) is expressed as follows:

$$(3.7) \quad \mathcal{N}(\mathbf{u}, \mathbf{v}_h) - \mathcal{N}(\mathbf{u}_h, \mathbf{v}_h) = 0 \quad \forall \mathbf{v}_h \in S_{h,p}.$$

**4. A posteriori error analysis.** In this section, we are interested in controlling the error of the numerical solution measured in terms of a given target or output functional  $J(\cdot)$ , rather than in terms of some global norm. Here,  $J(\cdot)$  is the physical quantity of interest and may, for example, represent the mean flow across a line, a point value of the solution, or the lift or drag coefficients of a body immersed in an inviscid fluid.

The proceeding error analysis is based on a duality argument using the general approach developed by Rannacher and his collaborators, see [4, 15, 18, 25, 26]; see also [19, 20, 21] for the application of these techniques to  $hp$ -finite element approximations to first-order hyperbolic problems. We begin by first introducing some notation. Assuming that the functional of interest  $J(\cdot)$  is differentiable, we write  $\bar{J}(\cdot; \cdot)$  to denote the mean value linearization of  $J(\cdot)$  defined by

$$(4.1) \quad \bar{J}(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h) = J(\mathbf{u}) - J(\mathbf{u}_h) = \int_0^1 J'[\theta \mathbf{u} + (1 - \theta) \mathbf{u}_h](\mathbf{u} - \mathbf{u}_h) d\theta,$$

where  $J'[\mathbf{w}](\cdot)$  denotes the functional (Fréchet) derivative of  $J(\cdot)$  evaluated at some  $\mathbf{w}$  in  $V$ . Here,  $V$  is some suitably chosen function space such that  $S_{h,p} \subset V$ .

Analogously, we write  $\mathcal{M}(\mathbf{u}, \mathbf{u}_h; \cdot, \cdot)$  to denote the mean-value linearization of the semilinear form  $\mathcal{N}(\cdot, \cdot)$  given by

$$(4.2) \quad \begin{aligned} \mathcal{M}(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h, \mathbf{v}) &= \mathcal{N}(\mathbf{u}, \mathbf{v}) - \mathcal{N}(\mathbf{u}_h, \mathbf{v}) \\ &= \int_0^1 \mathcal{N}'_{\mathbf{u}}[\theta \mathbf{u} + (1 - \theta) \mathbf{u}_h](\mathbf{u} - \mathbf{u}_h, \mathbf{v}) d\theta \end{aligned}$$

for all  $\mathbf{v}$  in  $V$ . Here,  $\mathcal{N}'_{\mathbf{u}}[\mathbf{w}](\cdot, \mathbf{v})$  denotes the Fréchet derivative of  $\mathbf{u} \rightarrow \mathcal{N}(\mathbf{u}, \mathbf{v})$ , for  $\mathbf{v} \in V$  fixed, at some  $\mathbf{w}$  in  $V$ . We remark that the linearization defined in (4.2) is only a *formal* calculation, in the sense that  $\mathcal{N}'_{\mathbf{u}}[\mathbf{w}](\cdot, \mathbf{v})$  may not, in general, exist. Instead, a suitable approximation to  $\mathcal{N}'_{\mathbf{u}}[\mathbf{w}](\cdot, \mathbf{v})$  in the direction  $\mathbf{u} - \mathbf{u}_h$  must be determined, for example, by computing appropriate finite difference quotients of  $\mathcal{N}(\cdot, \mathbf{v})$ . Below we provide some concrete examples when  $\mathcal{N}'_{\mathbf{u}}[\mathbf{w}](\cdot, \mathbf{v})$  may be uniquely determined; further discussion concerning the calculation of the Fréchet derivative of  $\mathcal{N}(\cdot, \mathbf{v})$  with respect to  $\mathbf{u}$ ,  $\mathbf{v} \in V$  fixed, will be presented in section 5.

For the purposes of the proceeding analysis, we *assume* that the linearization performed in (4.2) is well defined. Under this hypothesis, we introduce the following *dual* or *adjoint* problem: Find  $\mathbf{z} \in V$  such that

$$(4.3) \quad \mathcal{M}(\mathbf{u}, \mathbf{u}_h; \mathbf{w}, \mathbf{z}) = \bar{J}(\mathbf{u}, \mathbf{u}_h; \mathbf{w}) \quad \forall \mathbf{w} \in V.$$

We assume that (4.3) possesses a unique solution. First, the validity of this assumption depends on the well-posedness of the primal problem (2.2) and the existence of a unique entropy solution  $\mathbf{u}$ . Furthermore, the existence of a unique solution to the dual

problem (4.3) depends on both the definition of  $\mathcal{M}(\mathbf{u}, \mathbf{u}_h; \cdot, \cdot)$  and the choice of the target functional under consideration. Below, we present some important examples which are covered by our hypothesis.

*Example 4.1. Linear hyperbolic problems.* Let us first consider the case of a linear hyperbolic equation for the scalar ( $m = 1$ ) variable  $\mathbf{u} \equiv u$ ; then we may write the flux function  $\mathcal{F}(u) = \mathbf{a}u$ , where  $\mathbf{a} = (1, a_1, \dots, a_d)$  and  $a_i = a_i(\mathbf{x})$ , for  $i = 1, \dots, d$ . The weak imposition of boundary conditions on each element gives rise to the numerical flux function

$$\mathcal{H}(u_h^+, u_h^-, \mathbf{n}_\kappa)|_{\partial\kappa} = \mathbf{a} \cdot \mathbf{n}_\kappa \lim_{s \rightarrow 0^+} u_h(\mathbf{x} - s\mathbf{a}) \quad \text{for } \kappa \in \mathcal{T}_h;$$

cf. Cockburn, Karniadakis, and Shu [9]. In the absence of shock capturing, i.e., when  $\varepsilon \equiv 0$ , a simple calculation based on the divergence theorem shows that the dual solution  $z$  satisfies the (formal) adjoint equation subject to appropriate data, depending on the choice of the functional  $J(\cdot)$ . In particular, it was shown in [21] that if  $J(\cdot)$  is chosen to represent the mean flow over  $\Omega_T$ , the value of the solution at a given point  $\mathbf{x}_c$  in  $\Omega_T$ , or the normal flux along a portion of the outflow boundary of  $\Gamma_T$ , then the dual problem is indeed well-posed. Analogous well-posedness results for the dual problem also hold for systems of linear hyperbolic conservation laws, provided the numerical flux function is chosen appropriately.

*Example 4.2. Continuous Galerkin approximations to one-dimensional scalar hyperbolic problems.* The question of well-posedness of the dual problem associated with nonlinear hyperbolic conservation laws is rather more delicate. Let us, for example, consider the one-dimensional scalar hyperbolic equation,

$$(4.4) \quad u_t(x, t) + (f(u(x, t)))_x = 0, \quad -\infty < x < \infty, \quad 0 < t \leq T,$$

with strictly convex flux function  $f$ , i.e.,  $f'' > 0$ , subject to the initial condition

$$(4.5) \quad u(x, 0) = u_0(x), \quad -\infty < x < \infty.$$

Here, we suppose that (4.4), (4.5) is approximated by the standard Galerkin finite element method consisting of continuous piecewise polynomials in both space and time; cf. [26]. Furthermore, we suppose that the functional of interest represents the (weighted) mean flow in space; i.e.,  $J(w) \equiv M_\psi(w) = \int_{-\infty}^{\infty} w\psi \, dx$ , where  $\psi \in L_2(-\infty, \infty)$  is a given weight function. According to (4.3), the dual problem (in strong form) takes the following form: Find  $z$  such that

$$(4.6) \quad \begin{aligned} -z_t - a(x, t)z_x &= 0, & -\infty < x < \infty, & \quad 0 \leq t < T, \\ z(x, T) &= \psi(x), & -\infty < x < \infty, \end{aligned}$$

where  $a(x, t) = \int_0^1 f'((1 - \theta)u + \theta u_h) \, d\theta$ . Although the dual problem (4.6) is linear, this is a nonstandard hyperbolic problem since the velocity field  $a$  may be discontinuous. Thereby, in this case the classical theory of well-posedness for linear hyperbolic equations does not apply. However, assuming that  $a \in L_\infty(\mathbb{R} \times (0, T))$  and satisfies a one-sided Lipschitz condition, Tadmor [27] has shown that (4.6) is a meaningful problem. For the verification that  $a$  satisfies the desired hypotheses when  $f'' > 0$ , see [26]. For related work on the linearization of conservation laws and the well-posedness of linear hyperbolic problems with discontinuous data, we refer to [6, 13, 29].

In general, however, it is not known whether the dual problem (4.3) possesses a unique solution when the hyperbolic conservation law (2.2) is nonlinear and when the

discontinuous Galerkin approximation (3.5) is employed, even for scalar problems, corresponding to  $m = 1$ . Of course such analytical results cannot be expected to hold without well-posedness theory for the primal problem (2.2). Indeed, we can only *hypothesize* that when the system of nonlinear hyperbolic conservation laws (2.2) is well-posed, then the computation of a physically correct approximation to (2.2) will yield a uniquely solvable dual problem. In the absence of such analytical results, we must rely solely on numerical simulation to confirm the validity of this hypothesis. Indeed, section 7 presents a series of test problems which provide numerical evidence to indicate that the dual problem is solvable for the target functionals selected.

For the proceeding error analysis, we must therefore *assume* that the dual problem (4.3) is well-posed. Under this assumption, we have the following general result.

**THEOREM 4.1.** *Let  $\mathbf{u}$  and  $\mathbf{u}_h$  denote the solutions of (2.2) and (3.5), respectively, and suppose that the dual problem (4.3) is well-posed. Then, the following error representation formula holds:*

$$(4.7) \quad J(\mathbf{u}) - J(\mathbf{u}_h) = \mathcal{E}_\Omega(\mathbf{u}, \mathbf{u}_h, h, \mathbf{z} - \mathbf{z}_h) \equiv \sum_{\kappa \in \mathcal{T}_h} \eta_\kappa,$$

where

$$(4.8) \quad \eta_\kappa = \int_\kappa R_h(\mathbf{z} - \mathbf{z}_h) \, d\mathbf{x} + \int_{\partial\kappa} r_h(\mathbf{z} - \mathbf{z}_h)^+ \, ds - \int_\kappa \varepsilon \nabla \mathbf{u}_h \cdot \nabla(\mathbf{z} - \mathbf{z}_h) \, d\mathbf{x}$$

for all  $\mathbf{z}_h$  in  $S_{h,p}$ . Here,  $R_h$  and  $r_h$  denote the internal and boundary finite element residuals, respectively, defined on  $\kappa \in \mathcal{T}_h$  by

$$(4.9) \quad R_h|_\kappa = -\operatorname{div} \mathcal{F}(\mathbf{u}_h) \quad \text{and} \quad r_h|_\kappa = \mathcal{F}(\mathbf{u}_h^+) \cdot \mathbf{n}_\kappa - \mathcal{H}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}_\kappa),$$

respectively.

*Proof.* Choosing  $\mathbf{w} = \mathbf{u} - \mathbf{u}_h$  in (4.3), recalling the linearization performed in (4.1), and exploiting the Galerkin orthogonality property (3.7), we get

$$\begin{aligned} J(\mathbf{u}) - J(\mathbf{u}_h) &= \bar{J}(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h) = \mathcal{M}(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h, \mathbf{z}) \\ &= \mathcal{M}(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h, \mathbf{z} - \mathbf{z}_h) = -\mathcal{N}(\mathbf{u}_h, \mathbf{z} - \mathbf{z}_h) \end{aligned}$$

for all  $\mathbf{z}_h$  in the finite element space  $S_{h,p}$ . Employing the divergence theorem gives

$$(4.10) \quad \begin{aligned} J(\mathbf{u}) - J(\mathbf{u}_h) &= \sum_{\kappa \in \mathcal{T}_h} \left\{ - \int_\kappa \operatorname{div} \mathcal{F}(\mathbf{u}_h)(\mathbf{z} - \mathbf{z}_h) \, d\mathbf{x} \right. \\ &\quad + \int_{\partial\kappa} (\mathcal{F}(\mathbf{u}_h^+) \cdot \mathbf{n}_\kappa - \mathcal{H}(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}_\kappa)) (\mathbf{z} - \mathbf{z}_h)^+ \, ds \\ &\quad \left. - \int_\kappa \varepsilon \nabla \mathbf{u}_h \cdot \nabla(\mathbf{z} - \mathbf{z}_h) \, d\mathbf{x} \right\}. \end{aligned}$$

Using the definition of the residuals in (4.9) gives the desired result.  $\square$

Based on the general error representation formula derived in Theorem 4.1, a posteriori error estimates bounding the error in the computed functional  $J(\cdot)$  may be deduced. Before we proceed, let us first recall the classification of Type I and Type II a posteriori error bounds introduced in [21]. Type I a posteriori error bounds, also referred to as *weighted* a posteriori error bounds (cf. Rannacher [25], for example), involve the multiplication of the residual terms  $R_h$  and  $r_h$  and the shock-capturing term



$\varepsilon \nabla \mathbf{u}_h$  by the difference between the dual solution  $\mathbf{z}$  and its projection, interpolant, or quasi interpolant  $\mathbf{z}_h$ . On the other hand, Type II a posteriori bounds are in the spirit of the error estimates derived by Eriksson et al. [11] and do not depend explicitly on the dual solution. These latter error estimates are derived from Type I a posteriori bounds by employing standard results from approximation theory, together with well-posedness results for the dual problem. The resulting Type II error bounds involve only certain norms of the residuals and shock-capturing term, the mesh function  $h$ , interpolation constants, and the stability factor of the dual problem.

We begin by first considering a Type I a posteriori error bound; it is a straightforward consequence of the error representation formula stated in the previous theorem.

**COROLLARY 4.2.** *Under the assumptions of Theorem 4.1, the following Type I a posteriori error bound holds:*

$$(4.11) \quad |J(\mathbf{u}) - J(\mathbf{u}_h)| \leq \mathcal{E}_{|\Omega|}^{(\text{I})}(\mathbf{u}, \mathbf{u}_h, h, \mathbf{z} - \mathbf{z}_h) \equiv \sum_{\kappa \in \mathcal{T}_h} \eta_{\kappa}^{(\text{I})},$$

where  $\eta_{\kappa}^{(\text{I})} = |\eta_{\kappa}|$  and  $\eta_{\kappa}$  is as defined in (4.8).

*Proof.* The a posteriori error bound (4.11) follows trivially from (4.7) by application of the triangle inequality.  $\square$

To derive a Type II bound, we first recall the following approximation result for the finite element space  $S_{h,p}$ ; cf. [1], for example.

**LEMMA 4.3.** *Given  $\kappa \in \mathcal{T}_h$ , suppose that  $v|_{\kappa} \in H^{k_{\kappa}}(\kappa)$ ,  $0 \leq k_{\kappa} \leq p+1$ . Then, there exists  $\Pi v$  in the finite element space  $S_{h,p}$ , a constant  $C_{\text{int}}$  dependent on  $d$ ,  $k_{\kappa}$ ,  $p$ , and the shape-regularity of  $\mathcal{T}_h$ , but independent of  $v$  and  $h_{\kappa}$  such that for  $0 \leq q \leq k_{\kappa}$ ,*

$$(4.12) \quad \|v - \Pi v\|_{H^q(\kappa)} \leq C_{\text{int}} h_{\kappa}^{k_{\kappa}-q} \|v\|_{H^{k_{\kappa}}(\kappa)}.$$

Moreover, by employing Lemma 4.3, together with the multiplicative trace inequality

$$\|v\|_{L_2(\partial\kappa)} \leq \frac{C_{\text{trace}}}{\sqrt{2}} \left( \|v\|_{L_2(\kappa)} \|\nabla v\|_{L_2(\kappa)} + h_{\kappa}^{-1} \|v\|_{L_2(\kappa)}^2 \right)^{1/2},$$

we deduce that, given  $v|_{\kappa} \in H^{k_{\kappa}}(\kappa)$  for some  $1 \leq k_{\kappa} \leq p+1$  and  $\kappa \in \mathcal{T}_h$ ,

$$(4.13) \quad \|v - \Pi v\|_{L_2(\partial\kappa)} \leq C_{\text{trace}} C_{\text{int}} h_{\kappa}^{k_{\kappa}-1/2} \|v\|_{H^{k_{\kappa}}(\kappa)}.$$

Here,  $C_{\text{trace}} > 0$  is a constant, dependent only on  $d$  and the shape-regularity of  $\mathcal{T}_h$ .

Equipped with (4.12) and (4.13), we now prove the following Type II a posteriori error bound.

**THEOREM 4.4.** *Given that the assumptions of Theorem 4.1 hold, suppose that  $\mathbf{z} \in [H^s(\Omega_T)]^m$ ,  $1 \leq s \leq p+1$ , and that we have found a constant  $C_{\text{stab}}$  such that*

$$(4.14) \quad \|\mathbf{z}\|_{H^s(\Omega_T)} \leq C_{\text{stab}}.$$

*Then, the following Type II a posteriori error bound holds:*

$$(4.15) \quad |J(\mathbf{u}) - J(\mathbf{u}_h)| \leq \mathcal{E}_{|\Omega|}^{(\text{II})}(\mathbf{u}, \mathbf{u}_h, h) \equiv C_{\text{trace}} C_{\text{int}} C_{\text{stab}} \left( \sum_{\kappa \in \mathcal{T}_h} (\eta_{\kappa}^{(\text{II})})^2 \right)^{1/2},$$

where

$$(4.16) \quad \eta_{\kappa}^{(\text{II})} = \|h^s R_h\|_{L_2(\kappa)} + \|h^{s-1/2} r_h\|_{L_2(\partial\kappa)} + \|h^{s-1} \varepsilon \nabla \mathbf{u}_h\|_{L_2(\kappa)}.$$

*Proof.* Applying the Cauchy–Schwarz inequality to the right-hand side of (4.11) in Corollary 4.2, together with the approximation results (4.12) and (4.13), we get

$$(4.17) \quad |J(\mathbf{u}) - J(\mathbf{u}_h)| \leq C_{\text{trace}} C_{\text{int}} \sum_{\kappa \in \mathcal{T}_h} \left( \|h^s R_h\|_{L_2(\kappa)} + \|h^{s-1/2} r_h\|_{L_2(\partial\kappa)} + \|h^{s-1} \varepsilon \nabla \mathbf{u}_h\|_{L_2(\kappa)} \right) \|\mathbf{z}\|_{H^s(\kappa)}$$

for  $1 \leq s \leq p+1$ . Upon application of the Cauchy–Schwarz inequality, together with the stability bound (4.14), the statement of the theorem follows immediately.  $\square$

*Remark 4.5.* Typical values of the interpolation and trace constants,  $C_{\text{int}}$  and  $C_{\text{trace}}$ , respectively, are available within the literature, while the stability factor  $C_{\text{stab}}$  must be determined for the problem at hand either analytically or computationally by studying the well-posedness of the dual problem (4.3) in the norm  $\|\cdot\|_{H^s(\Omega_T)}$ .

As we shall see in section 7, the elimination of the “weighting terms” involving the difference between  $\mathbf{z}$  and  $\mathbf{z}_h$  *en route* to deriving the Type II a posteriori error bound given in Theorem 4.4 may adversely affect the performance of the resulting adaptive finite element method. Indeed, mesh refinement strategies based on Type II a posteriori error bounds which do not require the computation of the dual solution may, under mesh refinement, lead to suboptimal convergence of the error  $|J(\mathbf{u}) - J(\mathbf{u}_h)|$  in the computed functional, resulting in uneconomical mesh design and an inefficient adaptive algorithm (see [15, 18] for related work).

An important issue concerning the practical implementation of Type I a posteriori error bounds is that the calculation of local error indicator  $\eta_\kappa^{(1)}$  presupposes knowledge of the dual solution  $\mathbf{z}$ . Moreover, for nonlinear hyperbolic conservation laws and/or nonlinear target functionals  $J(\cdot)$ , the error representation formula (4.7) and the resulting Type I a posteriori error bound (4.11) depend on the unknown analytical solution to the primal problem  $\mathbf{u}$ . Thus, in order to render these error estimates computable, both  $\mathbf{u}$  and  $\mathbf{z}$  must be replaced by suitable approximations which do not adversely affect the quality of the resulting (approximate) error representation formula and (approximate) Type I a posteriori error bound; cf. section 7 for numerical validation. This subject will be the topic of the following section.

**5. Numerical approximation of the dual problem.** To ensure that the error representation formula (4.7) and the subsequent error bound (4.11) are genuinely a posteriori, the noncomputable quantities, i.e.,  $\mathbf{u}$  and  $\mathbf{z}$ , in the right-hand sides of (4.7) and (4.11), respectively, must be replaced by suitable approximations.

Concerning the approximation of  $\mathbf{u}$ , we note that the dependence of the error representation formula  $\mathcal{E}_\Omega$  and the a posteriori error bound  $\mathcal{E}_{|\Omega|}^{(1)}$  on  $\mathbf{u}$  stems from the mean value linearization of  $\mathcal{N}(\cdot, \mathbf{v})$ ,  $\mathbf{v} \in V$  fixed, and  $J(\cdot)$  carried out in (4.2) and (4.1), respectively. Hence, for linear hyperbolic problems and linear target functionals  $J(\cdot)$ , the dependence of the error representation formula (4.7) and the a posteriori error bound (4.11) on  $\mathbf{u}$  does not arise. In practice, for nonlinear problems and/or nonlinear target functionals, the linearization leading to  $\mathcal{M}(\mathbf{u}, \mathbf{u}_h; \cdot, \cdot)$  and  $\bar{J}(\mathbf{u}, \mathbf{u}_h; \cdot)$  is performed about the numerical solution  $\mathbf{u}_h$ ; i.e.,  $\mathcal{M}(\mathbf{u}, \mathbf{u}_h; \cdot, \cdot)$  is approximated by  $\mathcal{M}(\mathbf{u}_h, \mathbf{u}_h; \cdot, \cdot)$  in (4.2) and  $\bar{J}(\mathbf{u}, \mathbf{u}_h; \cdot)$  is approximated by  $\bar{J}(\mathbf{u}_h, \mathbf{u}_h; \cdot)$  in (4.1); cf. [25].

Second,  $\mathbf{z}$  must be numerically approximated; here, there are essentially two further sources of error introduced into the computation of  $\mathcal{E}_\Omega$  and  $\mathcal{E}_{|\Omega|}^{(1)}$ . First, the Fréchet derivative of  $\mathbf{u} \rightarrow \mathcal{N}(\mathbf{u}, \mathbf{v})$ , for  $\mathbf{v} \in V$  fixed, at some  $\mathbf{w}$  in  $V$ , must be replaced by a

suitable approximation  $\hat{\mathcal{N}}'_u[\mathbf{w}](\cdot, \mathbf{v})$ . Here, we define

$$\begin{aligned} \hat{\mathcal{N}}'_u[\mathbf{w}](\phi, \mathbf{v}) = \sum_{\kappa \in \mathcal{T}_h} \left\{ - \int_{\kappa} (\mathcal{F}'(\mathbf{w})\phi) \cdot \nabla \mathbf{v} \, d\mathbf{x} \right. \\ \left. + \int_{\partial\kappa} \left( \hat{\mathcal{H}}'_{u^+}(\mathbf{w}^+, \mathbf{w}^-, \mathbf{n}_{\kappa})\phi^+ + \hat{\mathcal{H}}'_{u^-}(\mathbf{w}^+, \mathbf{w}^-, \mathbf{n}_{\kappa})\phi^- \right) \mathbf{v}^+ \, ds \right. \\ \left. + \int_{\kappa} (\varepsilon(\mathbf{w})\nabla\phi + \hat{\varepsilon}'[\mathbf{w}](\phi)\nabla\mathbf{w}) \cdot \nabla \mathbf{v} \, d\mathbf{x} \right\}, \end{aligned} \quad (5.1)$$

where  $\mathbf{w} \rightarrow \hat{\mathcal{H}}'_{u^+}(\mathbf{w}^+, \mathbf{w}^-, \mathbf{n}_{\kappa})$  and  $\mathbf{w} \rightarrow \hat{\mathcal{H}}'_{u^-}(\mathbf{w}^+, \mathbf{w}^-, \mathbf{n}_{\kappa})$  denote approximations to the derivatives of the flux function  $\mathcal{H}(\cdot, \cdot, \cdot)$  with respect to its first and second arguments, respectively. For the (local) Lax–Friedrichs flux (3.3), we write

$$\hat{\mathcal{H}}'_{u^{\pm}}(\mathbf{w}^+, \mathbf{w}^-, \mathbf{n}_{\kappa}) = (\mathcal{F}'(\mathbf{w}^{\pm}) \cdot \mathbf{n}_{\kappa} \pm \alpha I) / 2,$$

where  $I$  denotes the identity matrix; here, the term involving the derivative of  $\alpha$  with respect to  $\mathbf{u}$  has been neglected. For the Vijayasundaram flux (3.4), we define

$$\hat{\mathcal{H}}'_{u^{\pm}}(\mathbf{w}^+, \mathbf{w}^-, \mathbf{n}_{\kappa}) = \mathcal{A}^{\pm}(\hat{\mathbf{w}}, \mathbf{n}_{\kappa});$$

here, the derivatives of  $\mathbf{u} \rightarrow \mathcal{A}^{\pm}(\hat{\mathbf{u}}, \mathbf{n}_{\kappa})$  have been ignored. Finally,  $\hat{\varepsilon}'[\mathbf{w}](\phi)$  denotes the following approximation to the derivative of  $\mathbf{u} \rightarrow \varepsilon(\mathbf{u})$  (cf. (3.6)):

$$\hat{\varepsilon}'[\mathbf{w}](\phi) = C_{\varepsilon} h^{2-\beta} \text{sgn}(\varepsilon(\mathbf{w})) (\mathcal{F}'(\mathbf{w}) \cdot \nabla\phi + \mathcal{F}''(\mathbf{w}) \cdot \nabla\mathbf{w} \phi) I_d.$$

With the definition of the approximate Fréchet derivative  $\hat{\mathcal{N}}'_u[\mathbf{w}](\phi, \mathbf{v})$  of  $\mathcal{N}(\cdot, \mathbf{v})$ ,  $\mathbf{v} \in V$  fixed, at some  $\mathbf{w}$  in  $V$  in the direction  $\phi \in V$  (cf. (5.1)), we define

$$\hat{\mathcal{M}}(\mathbf{u}, \mathbf{u}_h; \mathbf{u} - \mathbf{u}_h, \mathbf{v}) = \int_0^1 \hat{\mathcal{N}}'_u[\theta\mathbf{u} + (1-\theta)\mathbf{u}_h](\mathbf{u} - \mathbf{u}_h, \mathbf{v}) \, d\theta. \quad (5.2)$$

Calculating the linearization performed in (5.2) at  $\mathbf{u}_h$  (cf. above), we now define the following (approximate) dual problem: Find  $\hat{\mathbf{z}} \in V$  such that

$$\hat{\mathcal{M}}(\mathbf{u}_h, \mathbf{u}_h; \mathbf{w}, \hat{\mathbf{z}}) = \bar{J}(\mathbf{u}_h, \mathbf{u}_h; \mathbf{w}) \quad \forall \mathbf{w} \in V. \quad (5.3)$$

We recall from the previous section that there were open questions concerning the existence and uniqueness of the formal dual problem defined in (4.3). Indeed, for general nonlinear systems of hyperbolic conservation laws, it is not clear whether (4.3) is well-posed for a given target functional  $J(\cdot)$ , even when there exists a unique entropy solution to the original (primal) problem. In practice, we are concerned about the well-posedness of the approximate dual problem (5.3). Given that the dual problem (5.3) serves only as an approximation to the true dual problem (4.3), further approximations may be made in the definition of (5.3) to ensure that the resulting system of partial differential equations is well-posed. For example, the approximate dual problem (5.3) may have additional artificial viscosity included to guarantee the existence and uniqueness of  $\hat{\mathbf{z}}$ . For the purposes of the current paper, this last approximation is not performed, since for the model problems considered here, the dual problem (5.3) was always found to be numerically solvable. For more complicated engineering problems, this may not be the case, and the additional regularization provided by a “tune-up” of an artificial viscosity term may be necessary.

Finally, we note that once a suitable approximate dual problem has been defined, its analytical solution  $\hat{\mathbf{z}}$  must be numerically determined. Writing  $\hat{\mathbf{z}}_h$  to denote this approximation, we remark that  $\hat{\mathbf{z}}_h$  should not be calculated using the same finite element space  $S_{h,p}$  employed for the primal problem; otherwise the resulting error representation formula would be identically zero. In practice, there are essentially three approaches to computing a numerical approximation  $\hat{\mathbf{z}}_h$  of  $\hat{\mathbf{z}}$ . The first approach is to keep the degree  $p$  of the approximating polynomial used to compute  $\mathbf{u}_h$  fixed, but compute  $\hat{\mathbf{z}}_h$  on a sequence of dual finite element meshes  $\hat{\mathcal{T}}_h$  which, in general, differ from the “primal meshes”  $\mathcal{T}_h$ . Alternatively,  $\hat{\mathbf{z}}_h$  may be computed using piecewise discontinuous polynomials of degree  $\hat{p}$ ,  $\hat{p} > p$ , on the same finite element mesh  $\mathcal{T}_h$  employed for the primal problem. A variant of this second approach is to compute the approximate dual problem using the same mesh  $\mathcal{T}_h$  and polynomial degree  $p$  employed for the primal problem and to patchwise extrapolate the resulting approximate dual solution  $\hat{\mathbf{z}}_h \in S_{h,p}$  to a dual solution  $\hat{\mathbf{z}}_h \in S_{2h,\hat{p}}$ ,  $\hat{p} > p$ . While this latter approach is the cheapest of the three methods, and is still capable of producing adaptively refined meshes specifically tailored to the selected target functional, the quality of the resulting approximate error representation formula may be poor. For the purposes of the current paper, we prefer the second approach due to its computational simplicity of implementation; i.e.,  $\hat{\mathbf{z}}_h \equiv \hat{\mathbf{z}}_h \in S_{h,\hat{p}}$  with  $\hat{p} > p$ . We remark that while this approach is not the cheapest of the three options, we shall show in section 7 that the overall cost of solving the dual problem is still relatively small in comparison to the total cost of the solution process. In particular, the efficiency of employing the proposed Type I error indicator will be demonstrated in comparison with traditional Type II error indicators, which do not require the solution of an auxiliary problem. As a final remark, we note that if the mesh refinement parameters are chosen in order that the number of degrees of freedom in the (dual) finite element space  $S_{h,\hat{p}}$  is roughly the same as the number of degrees of freedom in the new (primal) finite element space designed by employing  $\hat{\mathbf{z}}_h \in S_{h,\hat{p}}$ , then the cost of computing the dual solution  $\hat{\mathbf{z}}_h$  will be roughly equivalent to the cost of a single Newton step in the computation of  $\mathbf{u}_h$  on the newly designed adaptive mesh.

Notwithstanding these approximations, we shall show through numerical experimentation in section 7 that the reliability of the Type I a posteriori error bound (4.11) is not compromised, in the sense that

$$(5.4) \quad \hat{\mathcal{E}}_{|\Omega|}^{(1)} \equiv \mathcal{E}_{|\Omega|}^{(1)}(\mathbf{u}_h, \mathbf{u}_h, h, \hat{\mathbf{z}}_h - \mathbf{z}_h) = \sum_{\kappa \in \mathcal{T}_h} \hat{\eta}_{\kappa}^{(1)},$$

where  $\hat{\eta}_{\kappa}^{(1)} = |\hat{\eta}_{\kappa}|$  and  $\hat{\eta}_{\kappa}$  is defined as in (4.8) with  $\mathbf{z}$  replaced by  $\hat{\mathbf{z}}_h$ , remains an (asymptotic) upper bound on the true error in the target functional  $J(\cdot)$ . In particular, we shall show that the ratio of the approximate error representation formula

$$(5.5) \quad \hat{\mathcal{E}}_{\Omega} \equiv \mathcal{E}_{\Omega}(\mathbf{u}_h, \mathbf{u}_h, h, \hat{\mathbf{z}}_h - \mathbf{z}_h) = \sum_{\kappa \in \mathcal{T}_h} \hat{\eta}_{\kappa}$$

and the true value  $J(\mathbf{u}) - J(\mathbf{u}_h)$  is extremely close to one; see also [15, 17, 18, 24].

*Remark 5.1.* We remark that, in contrast to continuous Galerkin finite element methods, discontinuous Galerkin methods satisfy a local as well as global Galerkin orthogonality property. Thereby, here the local error indicators  $\hat{\eta}_{\kappa}^{(1)}$  are independent of the choice of  $\mathbf{z}_h$ , and indeed  $\mathbf{z}_h \equiv \mathbf{0}$  is permitted.

As a final note, we remark that even if the a posteriori error bound (4.11) should fail to remain an upper bound on the true error in the computed functional  $J(\cdot)$  when

$\mathbf{z}$  is replaced by  $\hat{\mathbf{z}}_h$ , then the adaptive mesh refinement algorithm outlined in the following section still provides the necessary local information to ensure that economical meshes, specifically tailored to the approximation of the underlying functional  $J(\cdot)$  of interest, are generated.

**6. Adaptive mesh refinement.** In this section we consider the design of an adaptive algorithm to ensure the efficient computation of the given target functional  $J(\cdot)$  of practical interest. To this end, we employ the approximate a posteriori error bound  $\hat{\mathcal{E}}_{|\Omega|}^{(1)}$  (cf. (5.4)) to determine when the desired level of accuracy has been achieved. For example, suppose that the aim of the computation is to compute  $J(\cdot)$  such that the error  $|J(\mathbf{u}) - J(\mathbf{u}_h)|$  is less than some user-defined tolerance **TOL**, i.e.,  $|J(\mathbf{u}) - J(\mathbf{u}_h)| \leq \text{TOL}$ ; then, in practice we may enforce the stopping criterion  $\hat{\mathcal{E}}_{|\Omega|}^{(1)} \leq \text{TOL}$ . If this condition is not satisfied on the current finite element mesh  $\mathcal{T}_h$ , then the elementwise terms  $\hat{\eta}_\kappa^{(1)}$  are employed as local error indicators to guide mesh refinement and coarsening. The cycle of the adaptive mesh refinement may be outlined as follows:

1. Construct an initial mesh  $\mathcal{T}_h$ .
2. Compute  $\mathbf{u}_h \in S_{h,p}$  on the current mesh  $\mathcal{T}_h$ .
3. Compute  $\hat{\mathbf{z}}_h \in S_{h,\hat{p}}$  on the same mesh employed for  $\mathbf{u}_h$  with  $\hat{p} > p$ .
4. Evaluate the approximate a posteriori error bound  $\hat{\mathcal{E}}_{|\Omega|}^{(1)} = \sum_{\kappa \in \mathcal{T}_h} \hat{\eta}_\kappa^{(1)}$ .
5. If  $\hat{\mathcal{E}}_{|\Omega|}^{(1)} \leq \text{TOL}$ , where **TOL** is a given tolerance, then **STOP**.
6. Otherwise, refine and coarsen a fixed fraction of the total number of elements according to the size of  $\hat{\eta}_\kappa^{(1)}$  and generate a new mesh  $\mathcal{T}_h$ ; **GOTO** 2.

In section 7, the refinement and coarsening fractions are 20% and 10%, respectively.

**7. Numerical experiments.** In this section we present some numerical examples to highlight the advantages of designing an adaptive finite element algorithm based on the Type I error indicator  $\hat{\eta}_\kappa^{(1)}$  in comparison with traditional refinement strategies based on Type II error indicators, which do not require the solution of the dual problem (4.3). To this end, stimulated by Theorem 4.4, we introduce the (approximate) Type II error indicator

$$(7.1) \quad \hat{\eta}_\kappa^{(11)} = \|hR_h\|_{L_2(\kappa)} + \|h^{1/2}r_h\|_{L_2(\partial\kappa)}.$$

Here, (7.1) stems from (4.16) with  $s = 1$  and formally setting all the constants equal to one. Given that the fixed fraction strategy will be employed, the absolute size of  $\hat{\eta}_\kappa^{(11)}$  is insignificant: only the relative sizes of these quantities matter, which justifies setting all constants to unity. Furthermore, we have neglected the term arising from the artificial viscosity model; given  $\varepsilon$  depends on the internal residual  $R_h$ , we expect that meshes designed by both indicators (7.1) and (4.16) will be very similar in character.

In the following computations, both triangular and quadrilateral meshes will be employed; the latter are based on the deal.II library; cf. [2]. Furthermore, we set  $p = 1$  and  $\hat{p} = 2$ , i.e.,  $\mathbf{u}_h \in S_{h,1}$  and  $\hat{\mathbf{z}}_h \in S_{h,2}$ .

*Example 7.1. Burgers' equation.* In this first example we consider the one-dimensional unsteady inviscid Burgers' equation for the scalar variable  $\mathbf{u} \equiv u$ ; i.e., we have

$$(7.2) \quad u_t + (u^2/2)_x = 0,$$

on the space-time domain  $\Omega_T = (0, 2] \times (0, 3)$ , subject to the initial condition

$$(7.3) \quad \mathbf{u}(0, x) = \frac{2}{(1+x^3)} \sin^2(\pi x),$$

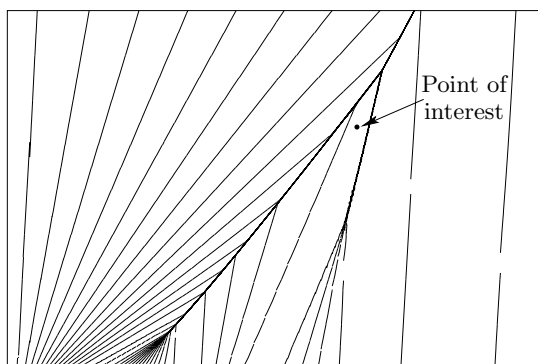


FIG. 7.1. Example 7.1. Solution isolines.

TABLE 7.1  
Example 7.1. Adaptive algorithm for Burgers' equation.

Elements	DOF	$J(u) - J(u_h)$	$\sum_{\kappa} \hat{\eta}_{\kappa}$	$\theta_1$	$\sum_{\kappa} \hat{\eta}_{\kappa}^{(1)}$	$\theta_2$
12	36	2.758e-02	-1.613e-01	-5.85	4.432e-01	16.07
25	75	1.574e-01	1.175e-01	0.75	2.751e-01	1.75
53	159	9.087e-02	-3.190e-02	-0.35	2.244e-01	2.47
110	330	6.231e-02	5.196e-02	0.83	1.206e-01	1.94
234	702	-1.566e-02	-3.274e-03	0.21	3.879e-02	2.48
479	1437	2.746e-04	3.301e-04	1.20	1.084e-02	39.48
1053	3159	1.281e-04	1.213e-04	0.95	3.273e-03	25.54
2256	6768	1.484e-05	1.458e-05	0.98	1.856e-03	125.06
4968	14904	-6.626e-07	-7.163e-07	1.08	7.762e-04	1171.38
11003	33009	-4.755e-07	-4.782e-07	1.01	3.196e-04	672.04
24662	73986	2.934e-08	2.905e-08	0.99	1.279e-04	4357.01

and boundary condition  $u(t, 0) = 0$  for  $t \in [0, 2]$ . The analytical solution to (7.2), (7.3) consists of three smooth “hills” which form shock waves as time increases. These shocks will eventually merge to form a single line of discontinuity in the  $(t, x)$ -plane. Here, we compute the numerical approximation up to a time after which the first two shock waves have merged and select the functional of interest  $J(\cdot)$  to be the value of  $u$  just before these two shocks interact with one another. More precisely, we select  $J(u) = u(1.35, 1.95)$  (cf. Figure 7.1); the true value is given by  $J(u) = 0.451408206331223$ .

In Table 7.1, we demonstrate the performance of the adaptive algorithm with  $C_{\varepsilon} = 1/4$  and  $\beta = 1/10$ ; here, 11 primal and dual problems were solved to obtain the final mesh. In the table, we show the number of elements and degrees of freedom (DOF) in  $S_{h,1}$ , the true error in the functional  $J(u) - J(u_h)$ , the computed error representation formula (5.5), the approximate a posteriori error bound (5.4), and their respective effectivity indices  $\theta_1$  and  $\theta_2$ . We see that initially on very coarse meshes the quality of the computed error representation formula  $\hat{\mathcal{E}}_{\Omega}$  is very poor, in the sense that  $\theta_1 = \hat{\mathcal{E}}_{\Omega}/(J(u) - J(u_h))$  is not close to one; however, as the mesh is refined, the effectivity index  $\theta_1$  approaches unity. On the other hand, we observe that the Type I a posteriori error bound is not sharp, in the sense that the second effectivity index  $\theta_2 = \hat{\mathcal{E}}_{\Omega}^{(1)}/|J(u) - J(u_h)|$  overestimates the true error in the computed functional by three orders of magnitude on the finest mesh; this loss of sharpness is attributed to the

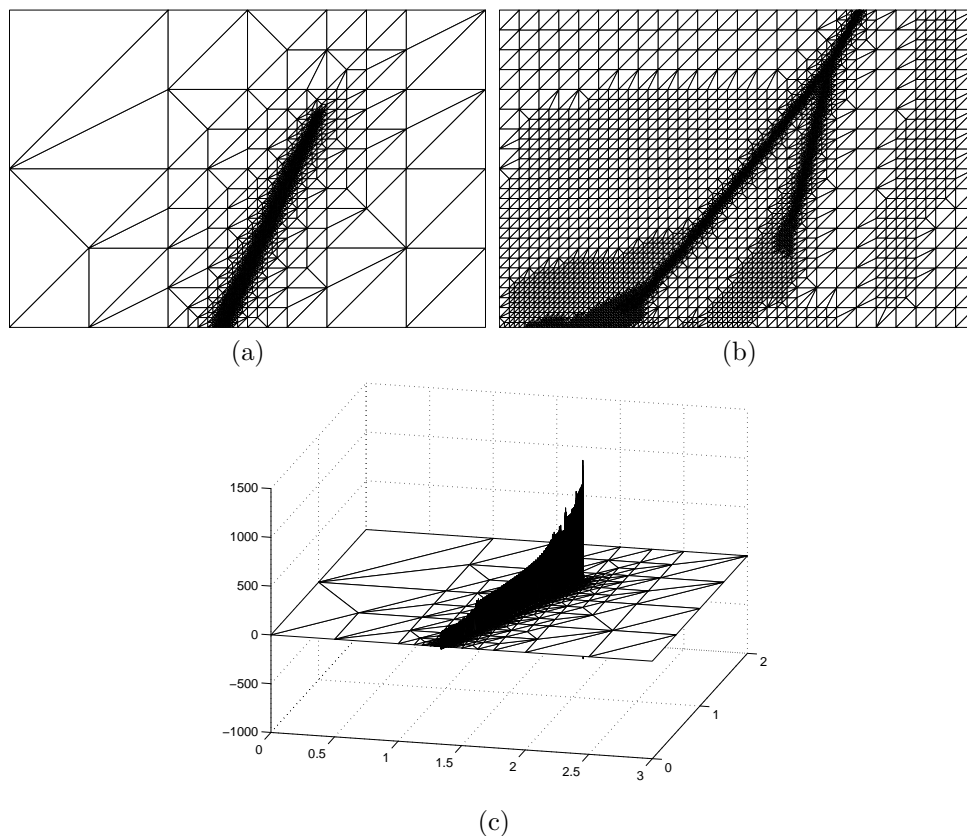


FIG. 7.2. Example 7.1. (a) Mesh constructed using the Type I error indicator with 24662 elements and 73986 degrees of freedom ( $|J(u) - J(u_h)| = 2.934 \times 10^{-8}$ ); (b) mesh constructed using the Type II error indicator with 20186 elements and 60558 degrees of freedom ( $|J(u) - J(u_h)| = 4.015 \times 10^{-4}$ ); (c) dual solution.

loss of interelement cancellation of the local error indicators  $\hat{\eta}_\kappa^{(1)}$ , when the triangle inequality is employed. Thereby, it is clear that any further bounding performed *en route* to deriving a Type II a posteriori estimate will further adversely affect the quality of the computed error bound.

In Figure 7.2 we show the meshes generated using both Type I and Type II error indicators  $\hat{\eta}_\kappa^{(1)}$  and  $\hat{\eta}_\kappa^{(II)}$ , respectively. From Figure 7.2(a), we see that there is virtually no refinement in the regions of the computational domain, where the shocks are located when  $\hat{\eta}_\kappa^{(1)}$  is employed. Indeed, most of the mesh refinement is concentrated in the neighborhood upstream of the point of interest. This is due to the presence of the weighting terms involving the difference between the (approximated) dual solution  $\hat{z}_h$  and  $z_h$ , which multiply the computable residual terms involving the numerical solution  $u_h$  in the definition of the local error indicator  $\hat{\eta}_\kappa^{(1)}$ ; cf. (4.8) with  $z$  replaced by  $\hat{z}_h$ . Indeed, from Figure 7.2(c), we see that the dual solution consists of a single “spike” or delta function located at the point of interest, which is carried backward along the single characteristic passing through  $(t, x) = (1.35, 1.95)$  to the inflow boundary at  $t = 0$ ; the reduction in height of this spike is attributed to the numerical and artificial viscosity inherent in the scheme. In contrast, we see that

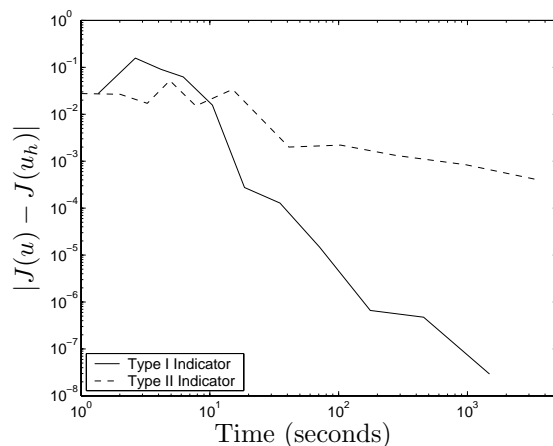


FIG. 7.3. Example 7.1. Convergence of  $|J(u) - J(u_h)|$  using Type I and Type II error indicators.

when these weighting terms are neglected, then the mesh produced using the Type II error indicator is largely concentrated in the vicinity of the first two shock waves, with some refinement of the third shock. In Figure 7.3, we compare the true error in the computed functional  $J(\cdot)$  using the two mesh refinement strategies. Here, we clearly observe the superiority of the Type I a posteriori error indicator; on the final mesh, the true error in the linear functional is over four orders of magnitude smaller than  $|J(u) - J(u_h)|$  computed on the sequence of meshes produced using  $\hat{\eta}_\kappa^{(II)}$ . This clearly indicates that a good numerical resolution of the shocks in this example is irrelevant for the accurate approximation of the functional of interest. Moreover, we see that the additional cost involved in solving the linearized dual problem is relatively small in comparison to the high cost of working with nonoptimized meshes designed by the traditional Type II error indicator  $\hat{\eta}_\kappa^{(II)}$ , which does not require the solution of an auxiliary problem.

**Example 7.2. Ringleb's flow** In this second example we consider the steady two-dimensional compressible Euler equations; here,  $\mathbf{u}$  represents the vector of conserved quantities  $(\rho, \rho u, \rho v, \rho E)$ , where  $\rho$ ,  $(u, v)$ , and  $E$  represent the density, Cartesian velocity, and total energy per unit mass, respectively. We consider Ringleb's flow for which an analytical solution may be obtained using the hodograph method; cf. Chiocchia [7]. This problem represents a transonic flow in a channel; the flow is mostly subsonic, with a small supersonic region near the right-hand-side wall; cf. Figure 7.4.

We take the functional of interest to be the value of the density at the point  $(-0.4, 2)$ , i.e.,  $J(\mathbf{u}) = \rho(-0.4, 2)$ ; consequently, the true value of the functional is given by  $J(\mathbf{u}) = 0.8616065996968034$ . Given that the solution to Ringleb's flow is smooth, no artificial viscosity is required; i.e.,  $C_\varepsilon = 0$ . In Table 7.2 we show the performance of our adaptive algorithm after the numerical solution of eight primal and dual problems; here we see that the quality of the computed error representation formula is extremely good, with  $\theta_1 \approx 1$  even on coarse meshes. Furthermore, the Type I a posteriori error bound (4.11) is sharper for this smooth problem; here  $\hat{\mathcal{E}}_{|\Omega|}^{(I)}$  overestimates the true error in the computed functional by about one order of magnitude.

The meshes produced using both Type I and Type II error indicators  $\hat{\eta}_\kappa^{(I)}$  and  $\hat{\eta}_\kappa^{(II)}$ , respectively, are shown in Figure 7.5. Here we see that, as in the previous example, the mesh constructed using  $\hat{\eta}_\kappa^{(I)}$  is mostly concentrated in the neighborhood



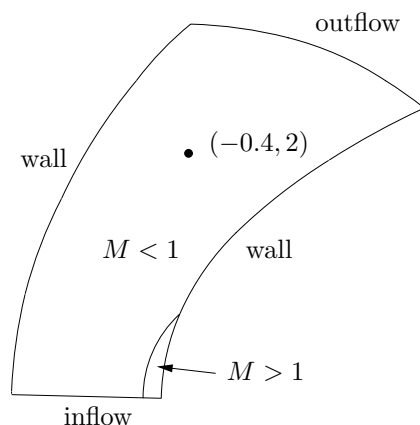


FIG. 7.4. Example 7.2. Geometry for Ringleb's flow; here  $M$  denotes the Mach number.

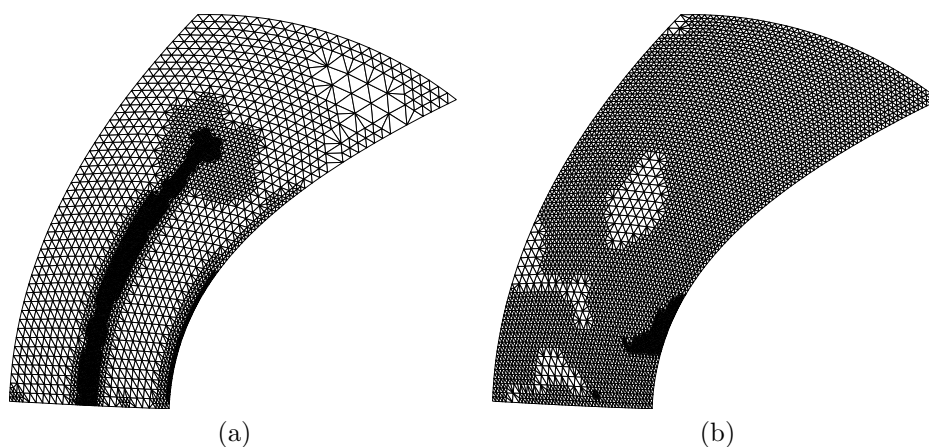


FIG. 7.5. Example 7.2. (a) Mesh constructed using the Type I error indicator with 11066 elements and 132792 degrees of freedom ( $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 5.775 \times 10^{-7}$ ); (b) mesh constructed using the Type II error indicator with 8965 elements and 107580 degrees of freedom ( $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 7.238 \times 10^{-6}$ ).

upstream of the point of interest. However, due to the elliptic nature of the flow in the subsonic region, a circular region containing the point of interest is also refined, together with a strip of cells in the vicinity of the wall on the right-hand side of the domain enclosing the supersonic region of the flow. In Figure 7.6(a) we show the first component of the dual solution; we note that the structure of the other three components is very similar in character. Here, we see that there is a spike centered at the point of interest that is transported toward the inflow boundary. We remark that this spike “diffuses out” slightly into the rest of the computational domain, thereby leading to the refinement of the mesh observed in Figure 7.5(a), when the Type I a posteriori error indicator is employed. In contrast, the mesh constructed using  $\hat{\eta}_\kappa^{(II)}$  is largely concentrated in the region near the right-hand wall, with almost uniform refinement of the rest of the computational domain; we note that subsequent mesh refinement leads to an even greater concentration of elements near the wall. Finally, in Figure 7.6(b) we compare the true error in the computed functional  $J(\cdot)$  using the

TABLE 7.2  
Example 7.2. Adaptive algorithm for Ringleb's flow.

Elements	DOF	$J(\mathbf{u}) - J(\mathbf{u}_h)$	$\sum_{\kappa} \hat{\eta}_{\kappa}$	$\theta_1$	$\sum_{\kappa} \hat{\eta}_{\kappa}^{(1)}$	$\theta_2$
144	1728	-2.995e-04	-3.024e-04	1.01	2.914e-03	9.73
262	3144	8.839e-05	8.703e-05	0.98	1.239e-03	14.02
527	6324	2.937e-05	2.901e-05	0.99	3.793e-04	12.92
992	11904	5.083e-06	5.350e-06	1.05	2.863e-04	56.33
1785	21420	-3.274e-06	-3.225e-06	0.99	1.207e-04	36.87
3382	40584	-1.798e-06	-1.765e-06	0.98	4.274e-05	23.77
6337	76044	-8.327e-07	-8.293e-07	1.00	5.107e-05	61.34
11066	132792	-5.775e-07	-5.808e-07	1.01	1.829e-05	31.67

TABLE 7.3  
Example 7.3. Adaptive algorithm for Ringleb's flow.

Elements	DOF	$J(\mathbf{u}) - J(\mathbf{u}_h)$	$\sum_{\kappa} \hat{\eta}_{\kappa}$	$\theta_1$	$\sum_{\kappa} \hat{\eta}_{\kappa}^{(1)}$	$\theta_2$
32	512	-6.946e-04	-8.609e-04	1.24	3.428e-03	4.94
56	896	9.940e-05	6.123e-05	0.62	1.619e-03	16.29
80	1280	-5.332e-05	-7.316e-05	1.37	9.163e-04	17.19
125	2000	7.571e-05	7.165e-05	0.95	5.488e-04	7.25
212	3392	-6.735e-06	-9.230e-06	1.37	2.412e-04	35.82
353	5648	1.634e-05	1.621e-05	0.99	1.096e-04	6.71
566	9056	5.681e-06	5.541e-06	0.98	5.141e-05	9.05
950	15200	2.778e-06	2.763e-06	0.99	2.001e-05	7.20
1562	24992	1.344e-06	1.367e-06	1.02	9.305e-06	6.92
2534	40544	1.750e-07	1.807e-07	1.03	4.463e-06	25.50

two mesh refinement strategies. Again, as in the previous example, we clearly see that the Type I a posteriori error indicator produces more economical meshes than when the traditional Type II error indicator is employed, in the sense that  $|J(\mathbf{u}) - J(\mathbf{u}_h)|$  is smaller for a given amount of computational effort. Indeed, on the final mesh the true error in the computed functional is over one order of magnitude smaller when the former error indicator is used.

*Example 7.3. Ringleb's flow (revisited).* In this section, we again consider Ringleb's flow; here, we select the target functional to be the horizontal force component exerted on the channel walls. More precisely, we choose

$$J(\mathbf{u}) = \int_{\text{wall}} (\psi \cdot \mathbf{n}) p \, ds,$$

where  $\psi = (1, 0)$ ,  $\mathbf{n}$  denotes the unit outward normal vector to the boundary, and  $p$  is the pressure; this is a slight variant of the target functional selected in [24]. In this case, the true value of the functional on the exact geometry is given by  $J(\mathbf{u}) = -0.5744441759095$ . We note that since the pressure  $p$  is derived from the conserved variables using the equation of state, the selected functional  $J(\cdot)$  is nonlinear.

In Table 7.3 we show the performance of our adaptive algorithm; on each refinement level the computed value of the target functional  $J(\mathbf{u}_h)$  is compared with the true value evaluated on the underlying computational domain. Here, 10 primal and dual problems were solved to obtain the final mesh. We again see that the quality of the computed error representation formula is extremely good, with  $\theta_1 \approx 1$ , while the Type I a posteriori error bound overestimates the true error in the computed functional by about one order of magnitude. Figure 7.7 shows the meshes produced using both Type I and Type II error indicators  $\hat{\eta}_{\kappa}^{(1)}$  and  $\hat{\eta}_{\kappa}^{(II)}$ , respectively. Here, we observe

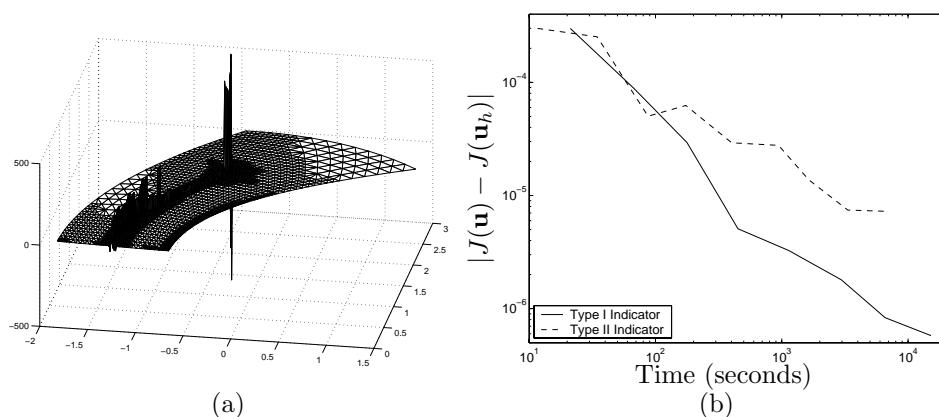


FIG. 7.6. Example 7.2. (a)  $z_1$  component of the dual solution for Ringleb's flow; (b) convergence of  $|J(\mathbf{u}) - J(\mathbf{u}_h)|$  using Type I and Type II error indicators.

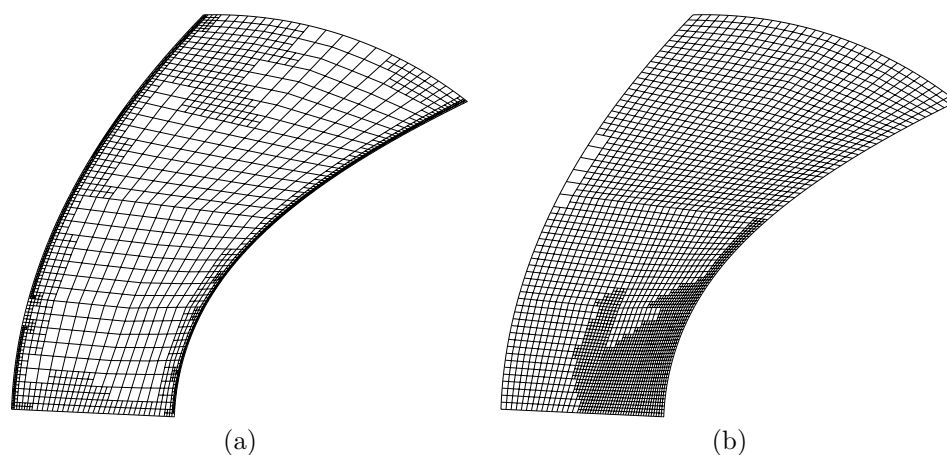


FIG. 7.7. Example 7.3. (a) Mesh constructed using the Type I error indicator with 2534 elements and 40544 degrees of freedom ( $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 1.7504 \times 10^{-7}$ ); (b) mesh constructed using the Type II error indicator with 3056 elements and 48896 degrees of freedom ( $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 2.1622 \times 10^{-5}$ ).

that while the underlying (primal) problem remains the same, the different choice of functional selected leads to a different dual solution, and hence to a completely different mesh, when the error indicator  $\hat{\eta}_K^{(1)}$  is employed. Indeed in Figure 7.8(a), we now see that the dual solution is large in the vicinity of the solid channel walls; thereby this leads to the mesh being concentrated near the walls of the channel. In contrast, the Type II error indicator has no information concerning the target functional of interest and will produce the same mesh irrespective of the choice of  $J(\cdot)$ ; the corresponding mesh employing quadrilateral elements is similar in character to the one produced with triangular elements; cf. Figure 7.5(b).

Finally, in Figure 7.8(b) we compare the true error in the computed functional  $J(\cdot)$  using the two mesh refinement strategies. Again, as before, we clearly see that the Type I a posteriori error indicator produces more economical meshes than when the traditional Type II error indicator is employed; on the final mesh the true error in the computed functional is almost two orders of magnitude smaller when the former error

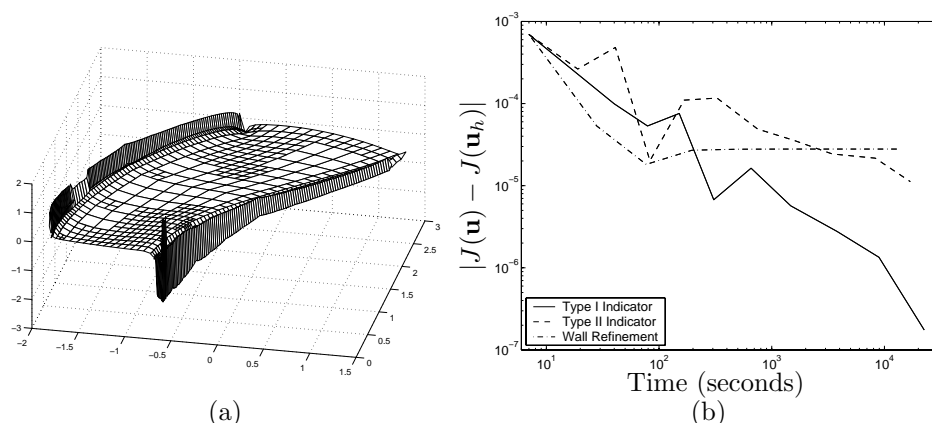


FIG. 7.8. Example 7.3. (a)  $z_1$  component of the dual solution for Ringleb's flow; (b) convergence of  $|J(\mathbf{u}) - J(\mathbf{u}_h)|$  using Type I and Type II error indicators, as well as simply refining elements near the channel walls.

indicator is used. Additionally, in Figure 7.8(b) we also compare the performance of the Type I indicator  $\hat{\eta}_K^{(I)}$  with a mesh refinement algorithm which only refines elements adjacent to the channel walls. Given the choice of the target functional, and the structure of the meshes produced using  $\hat{\eta}_K^{(I)}$ , one may expect that this approach should work quite well without the need to solve the dual problem. However, we now see that after an initial reduction in the error in the computed functional  $J(\cdot)$ ,  $|J(\mathbf{u}) - J(\mathbf{u}_h)|$  becomes  $\mathcal{O}(1)$  under additional mesh refinement. This is attributed to the fact that while it is important to refine the mesh near the channel walls in order to reduce the error in the target functional  $J(\cdot)$ , some mesh refinement is also required in the interior of the computational domain. Here, the Type I a posteriori error indicator  $\hat{\eta}_K^{(I)}$  provides information about not only where to refine the mesh but also by how much; this leads to the consistent reduction in  $|J(\mathbf{u}) - J(\mathbf{u}_h)|$  observed in Figure 7.8(b) as the finite element mesh is adaptively refined.

**Example 7.4. [Supersonic compression corner.]** In this final example we study the formation of an oblique shock when supersonic flow is deflected by a sharp object or wedge. Here, we consider a Mach 3 flow, with inflow density  $\rho = 1$  and pressure  $p = 1$ , over a compression corner of angle  $\alpha$ ; this leads to the development of a shock wave at an angle  $\beta$ ; cf. Figure 7.9. By employing the Rankine-Hugoniot jump conditions, the analytical solution to this problem may be determined for a given  $\alpha$ . Here, we select the wedge angle  $\alpha = 9.5$  degrees; thereby, the angle of the shock is given by  $\beta = 26.9308$  degrees. Furthermore, the true solution on the left- and right-hand sides of the shock, in terms of conserved variables  $(\rho, \rho u, \rho v, \rho E)$ , is given by  $\mathbf{u}_{\text{left}} = [1, 3.5496, 0, 8.8]^T$  and  $\mathbf{u}_{\text{right}} = [1.6180, 5.2933, 0.8858, 13.8692]^T$ , respectively.

Throughout this example, we set  $C_\epsilon = 1/50$  and  $\beta = 1/10$ ; cf. (3.6). In Table 7.4, we show the performance of our adaptive algorithm when the target functional is chosen to be the value of the density just in front of the shock; more precisely, here  $J(\mathbf{u}) = \rho(5, 2.05)$ . As in the previous examples, we again see that the quality of the computed error representation formula (5.5) is extremely good with  $\theta_1$  tending to unity, as the mesh is adaptively refined. Furthermore, the resulting Type I a posteriori bound (5.4) overestimates the true error in the computed target functional by about one to two orders of magnitude. Here, 11 primal and dual problems were solved to obtain the final mesh.

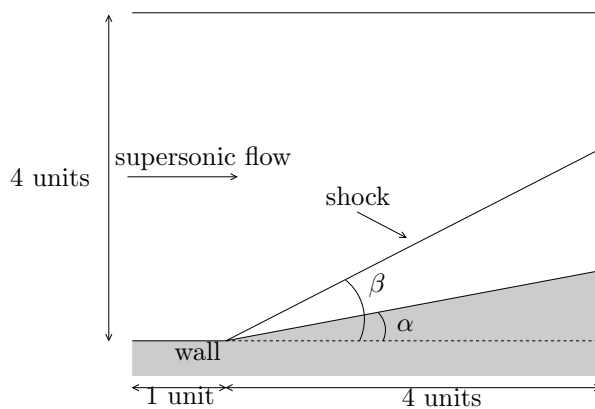


FIG. 7.9. Example 7.4. Geometry for the supersonic compression corner.

TABLE 7.4

Example 7.4. Adaptive algorithm for the supersonic compression corner; point evaluation of the density in front of the shock.

Elements	DOF	$J(\mathbf{u}) - J(\mathbf{u}_h)$	$\sum_{\kappa} \hat{\eta}_{\kappa}$	$\theta_1$	$\sum_{\kappa} \hat{\eta}_{\kappa}^{(1)}$	$\theta_2$
20	320	-3.428e-01	2.771e-02	-0.08	8.565e-02	0.25
35	560	-3.194e-01	-4.315e-02	0.14	8.134e-02	0.25
65	1040	-3.320e-01	-6.235e-02	0.19	7.675e-02	0.23
128	2048	-2.666e-01	-1.356e-01	0.51	1.819e-01	0.68
227	3632	-2.164e-01	-1.665e-01	0.77	2.142e-01	0.99
386	6176	-3.272e-03	-8.919e-02	27.26	2.094e-01	63.99
677	10832	1.686e-02	1.077e-02	0.64	5.179e-02	3.07
1175	18800	-5.475e-03	-5.887e-03	1.08	2.130e-02	3.89
1964	31424	-3.383e-04	-3.635e-04	1.07	3.671e-03	10.85
3293	52688	-4.208e-06	-3.837e-06	0.91	3.000e-04	71.30
5516	88256	4.551e-07	5.056e-07	1.11	2.079e-05	45.67

We end this example by considering the more interesting case of estimating the value of the density just behind the shock; to this end, we select  $J(\mathbf{u}) = \rho(5, 2.01)$ . The performance of the adaptive algorithm in this case is presented in Table 7.5 after 10 primal and dual problems were solved numerically. Here, we now see that while the computed a posteriori estimate still remains a reliable upper bound on the size of the error in the computed functional, the ratio of the (computed) error representation formula  $\hat{\mathcal{E}}_{\Omega}$  and  $J(\mathbf{u}) - J(\mathbf{u}_h)$  no longer tends to one as the finite element mesh is refined; indeed, here we see that (ignoring the first few refinement steps)  $\theta_1$  lies in the interval  $(0.8, 8.35)$ . This degradation in the quality of the computed error representation formula when the point of evaluation lies behind the shock is attributed to the effects of the linearization error and the error in the approximation to the Fréchet derivative of the semilinear form  $\mathcal{N}(\cdot, \cdot)$  (with respect to its first argument), becoming large in the vicinity of the shock. In this case, the weighting terms involving the difference between the dual solution and its projection/interpolant (cf. (5.4)) are poorly approximated in the neighborhood of the shock, where the residual terms  $R_h$  and  $r_h$  and the shock-capturing term  $\varepsilon \nabla \mathbf{u}_h$  are large, thereby leading to a poor approximation of the true error in the target functional  $J(\cdot)$ . We note that this does *not* occur if the target functional is selected upstream of the shock, since then the dual solution no longer interacts with the shock; thereby, in the neighbor-

TABLE 7.5

Example 7.4. Adaptive algorithm for the supersonic compression corner; point evaluation of the density behind the shock.

Elements	DOF	$J(\mathbf{u}) - J(\mathbf{u}_h)$	$\sum_{\kappa} \hat{\eta}_{\kappa}$	$\theta_1$	$\sum_{\kappa} \hat{\eta}_{\kappa}^{(1)}$	$\theta_2$
20	320	2.548e-01	3.952e-02	0.16	9.546e-02	0.37
35	560	2.652e-01	-1.728e-02	-0.07	9.185e-02	0.35
68	1088	2.226e-01	-1.148e-02	-0.05	4.927e-02	0.22
134	2144	1.723e-01	6.407e-02	0.37	1.254e-01	0.73
209	3344	1.259e-01	1.007e-01	0.80	1.369e-01	1.09
389	6224	6.044e-02	9.786e-02	1.62	1.180e-01	1.95
650	10400	-1.453e-02	-1.538e-02	1.06	5.050e-02	3.48
1124	17984	5.425e-04	4.530e-03	8.35	3.220e-02	59.36
1997	31952	4.470e-04	3.006e-03	6.73	2.733e-02	61.15
3395	54320	2.888e-04	1.405e-03	4.86	2.401e-02	83.15

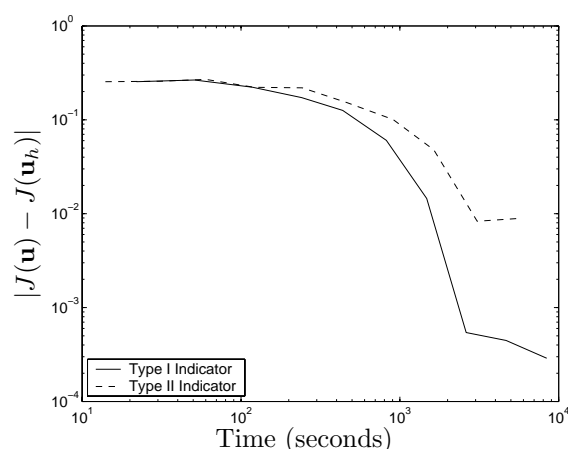


FIG. 7.10. Example 7.4. Convergence of  $|J(\mathbf{u}) - J(\mathbf{u}_h)|$  using Type I and Type II error indicators.

hood of the shock, the weighting terms are negligibly small, and the computed error representation formula is extremely good; cf. above. Thus, in order to guarantee the quality of the computed error representation formula, additional computational work is required to reduce the effects of the linearization error, the error stemming from the approximation of  $\mathcal{N}'_{\mathbf{u}}(\cdot, \mathbf{v})$ ,  $\mathbf{v} \in V$  fixed, and the error generated in the numerical approximation of the (approximate) dual problem; this is beyond the scope of the current paper, but will be considered in [16].

Notwithstanding this reduction in the quality of the computed error representation formula when the point of interest lies behind the shock, the adaptively refined meshes generated by employing the Type I error indicator  $\hat{\eta}_{\kappa}^{(1)}$  are much more economical than those produced using the traditional Type II error indicator  $\hat{\eta}_{\kappa}^{(1)}$ . Indeed, in Figure 7.10 we clearly observe the superiority of the former error indicator; on the final mesh the true error in the computed functional is over one order of magnitude smaller when the weighted error indicator is employed.

Finally, in Figure 7.11 we show the meshes produced using both error indicators. Here, we see that the mesh constructed using  $\hat{\eta}_{\kappa}^{(1)}$  is concentrated in the neighborhood of the shock. In contrast, the mesh produced using  $\hat{\eta}_{\kappa}^{(1)}$  refines the mesh only in the vicinity of the top and bottom parts of the shock. To gain insight into the structure of this mesh, in Figures 7.12(a) and (b) we show the dual solutions corresponding to the

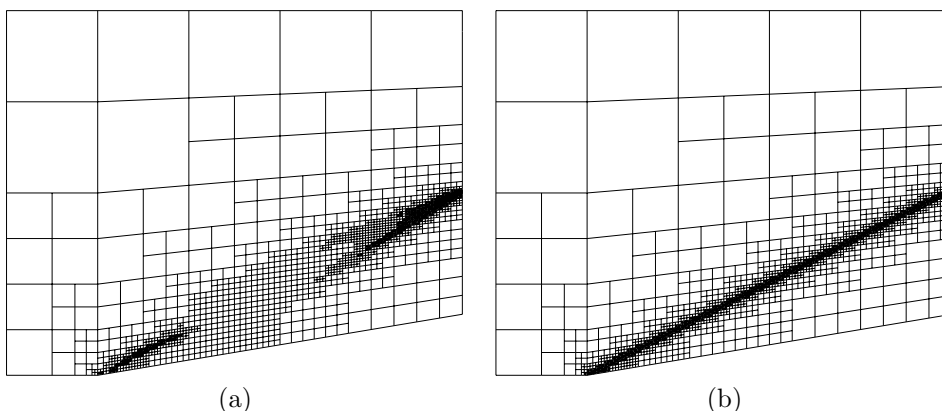


FIG. 7.11. Example 7.4. (a) Mesh constructed using the Type I error indicator with 3395 elements and 54320 degrees of freedom ( $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 2.888 \times 10^{-4}$ ); (b) mesh constructed using the Type II error indicator with 3821 elements and 61136 degrees of freedom ( $|J(\mathbf{u}) - J(\mathbf{u}_h)| = 8.938 \times 10^{-3}$ ).

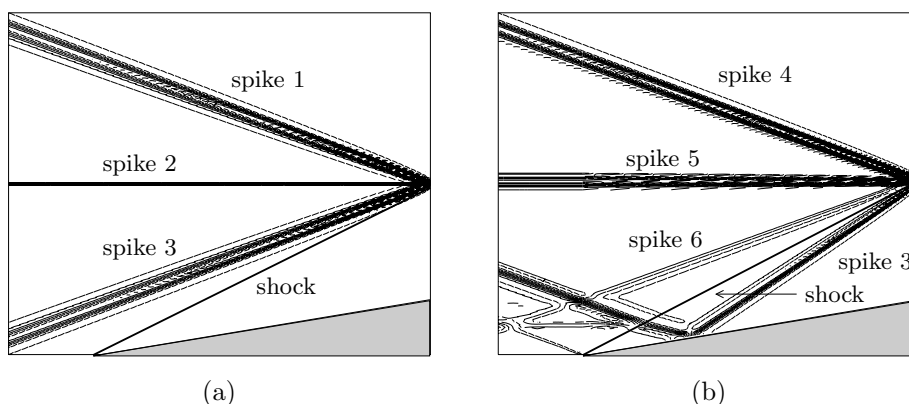


FIG. 7.12. Example 7.4.  $z_1$  component of dual solution for the supersonic compression corner for point evaluation of the density (a) in front of the shock and (b) behind the shock.

two cases when the point of interest is in front of and behind the shock, respectively; for clarity, here we show the  $z_1$  component of the dual solutions computed on globally refined meshes. The dual solution corresponding to the point evaluation of the density in front of the shock consists of three “spikes” labeled 1, 2, and 3 in Figure 7.12(a) and originating from the point of interest. These spikes are transported upstream along the characteristics corresponding to the three eigenvalues  $\bar{v}$  and  $\bar{v} \pm c$ , with  $\bar{v} = \sqrt{u^2 + v^2}$  denoting the velocity of the gas and  $c = \sqrt{\gamma p / \rho}$  the speed of sound. We note that the support of this dual solution does not intersect the region of the computational domain where the shock in the primal solution is located. In contrast, the support of dual solution corresponding to the point evaluation of the density behind the shock (see Figure 7.12(b)) now intersects the region containing the shock; here, we observe that  $z_1$  has a rather complicated structure. The two upper spikes of the dual solution both cross the shock in the neighborhood of the point of evaluation. At their crossing points, again they each split further into three spikes. These six spikes correspond to the three pairs of spikes labeled 4, 5, and 6 in Figure 7.12(b);

the two spikes in each pair cannot be distinguished on the resolution shown, as they are extremely close together. Spike 3 in Figure 7.12(b), corresponding to the same spike in Figure 7.12(a), is reflected off the inclined wall and crosses the shock at its bottom part. A comparison of the dual solution in Figure 7.12(b) and the mesh in Figure 7.11(a) shows that the mesh has been refined only along the support of spikes 3 and 6 in the vicinity of the top part of the shock and in the neighborhood of the point where spike 3 crosses the bottom part of the shock. This demonstrates that it is not necessary to refine the entire shock in this example to gain an accurate evaluation of the target functional under consideration, but only those parts that influence the value of the target functional either by material transport (eigenvalue  $\bar{v}$ ) or by information transported by the sound waves (eigenvalues  $\bar{v} \pm c$ ). Indeed, as seen in Figure 7.10, the meshes produced using the Type I error indicator are much more economical for computing the value of the target functional than meshes produced using the traditional Type II indicator.

**8. Concluding remarks.** In this article, we have developed the a posteriori error analysis of the discontinuous Galerkin finite element method for systems of nonlinear hyperbolic conservation laws. In particular, by employing a duality argument, we have derived (weighted) Type I and (unweighted) Type II a posteriori error bounds for general linear and nonlinear target functionals of the solution. Numerical experiments have been presented to illustrate the quality of the approximate error representation formula, when the (approximate) dual problem is approximated numerically. Moreover, comparisons between Type I and Type II error indicators have clearly demonstrated the superiority of exploiting weighted a posteriori error indicators to guide adaptive mesh refinement; in all examples, the error in the computed target functional is significantly smaller for a given amount of computational effort when the proposed Type I error indicator is employed. Furthermore, Type II a posteriori estimates will in general not be sharp; indeed, the Type I a posteriori bound derived from the error representation formula by simply employing the triangle inequality already overestimates the error in the computed target functional by at least one order of magnitude.

Several open questions still remain, most notably the issue of well-posedness of the dual problem which, although being linear, is a nonstandard hyperbolic problem in the sense that the coefficients may be discontinuous. In addition, further work concerning the quality of the (approximate) error representation formula in the presence of shocks still needs further numerical investigation; cf. [16].

**Acknowledgments.** We wish to express our sincere gratitude to Professor Tim Barth (NASA Ames Research Center) for providing the algorithm used to generate the analytical solution to Ringleb's flow. Finally, we wish to thank Professor Rolf Rannacher (Heidelberg) and Professor Endre Süli (Oxford) for their helpful comments.

#### REFERENCES

- [1] I. BABUŠKA AND M. SURI, *The hp-version of the finite element method with quasiuniform meshes*, M2AN Math. Model. Numer. Anal., 21 (1987), pp. 199–238.
- [2] W. BANGERTH, R. HARTMANN, AND G. KANSCHAT, *deal.II*, Technical Reference Differential Equations Analysis Library, IWR, University of Heidelberg, 2002. Available online at <http://www.dealii.org>.
- [3] F. BASSI AND S. REBAY, *High-order accurate discontinuous finite element solution of the 2D Euler equations*, J. Comput. Phys., 138 (1997), pp. 251–285.



- [4] R. BECKER AND R. RANNACHER, *A feed-back approach to error control in finite element methods: Basic analysis and examples*, East-West J. Numer. Math., 4 (1996), pp. 237–264.
- [5] R. BISWAS, K.D. DEVINE, AND J. FLAHERTY, *Parallel, adaptive finite element methods for conservation laws*, Appl. Numer. Math., 14 (1994), pp. 255–283.
- [6] F. BOUCHUT AND F. JAMES, *One-dimensional transport equations with discontinuous coefficients*, Nonlinear Anal., 32 (1998), pp. 891–933.
- [7] G. CHIOCCHIA, *Exact Solutions to Transonic and Supersonic Flows*, NASA CASI Technical report, Center for Aerospace Information, National Aeronautics and Space Administration, Washington, DC, 1985. Abstract available online at <http://www.sti.nasa.gov/cgi-bin/waisgate>.
- [8] B. COCKBURN AND P.-A. GREMAUD, *Error estimates for finite element methods for scalar conservation laws*, SIAM J. Numer. Anal., 33 (1996), pp. 522–554.
- [9] B. COCKBURN, G.E. KARNIADAKIS, AND C.-W. SHU, *The development of discontinuous Galerkin methods*, in Discontinuous Galerkin Methods, B. Cockburn, G. Karniadakis, and C.-W. Shu, eds., Lecture Notes in Comput. Sci. Engrg. 11. Springer, Berlin, 2000, pp. 3–50.
- [10] B. COCKBURN AND C.-W. SHU, *The Runge–Kutta discontinuous Galerkin finite element method for conservation laws V: Multidimensional systems*, J. Comput. Phys., 141 (1998), pp. 199–224.
- [11] K. ERIKSSON, D. ESTEP, P. HANSBO, AND C. JOHNSON, *Introduction to adaptive methods for differential equations*, Acta Numer., 1995, pp. 105–158.
- [12] J. GIANNAKOULOS AND G.E. KARNIADAKIS, *Spectral element–FCT method for scalar hyperbolic conservation laws*, Internat. J. Numer. Methods Fluids, 14 (1992), p. 707.
- [13] E. GODLEWSKI, M. OLAZABAL, AND P.-A. RAVIART, *On the linearization of systems of conservation laws for fluids at a material contact discontinuity*, J. Math. Pures Appl., 17 (1999), pp. 1013–1042.
- [14] D. GOTTLIEB AND J. S. HESTHAVEN, *Spectral methods for hyperbolic problems*, J. Comput. Appl. Math., 128 (2001), pp. 83–131.
- [15] R. HARTMANN, *Adaptive FE methods for conservation equations*, in Hyperbolic Problems: Theory, Numerics, Applications, Vol. I, II, (Magdeburg, 2000), Internat. Ser. Numer. Math. 140, 141, Birkhäuser, Basel, 2001, pp. 495–503.
- [16] R. HARTMANN AND P. HOUSTON, *A Posteriori Error Estimation for Nonlinear Hyperbolic Conservation Laws Involving Discontinuous Solutions*, in preparation.
- [17] P. HOUSTON, R. HARTMANN, AND E. SÜLI, *Adaptive discontinuous Galerkin finite element methods for compressible fluid flows*, in Numerical Methods for Fluid Dynamics VII, M. Baines, ed., Institute for Computational Fluid Dynamics, University of Oxford, Oxford, UK, 2001, pp. 347–353.
- [18] P. HOUSTON, R. RANNACHER, AND E. SÜLI, *A posteriori error analysis for stabilised finite element approximations of transport problems*, Comput. Methods Appl. Mech. Engrg., 190 (2000), pp. 1483–1508.
- [19] P. HOUSTON, B. SENIOR, AND E. SÜLI, *hp-discontinuous Galerkin finite element methods for hyperbolic problems: Error analysis and adaptivity*, Internat. J. Numer. Methods Fluids, 40 (2002), pp. 153–169.
- [20] P. HOUSTON, B. SENIOR, AND E. SÜLI, *Sobolev Regularity Estimation for hp-Adaptive Finite Element Methods*, Technical report 2002/05, Leicester University, Leicester, UK, 2002; ENUMATH 2001 Conference Proceedings, submitted.
- [21] P. HOUSTON AND E. SÜLI, *hp-adaptive discontinuous Galerkin finite element methods for first-order hyperbolic problems*, SIAM J. Sci. Comput., 23 (2001), pp. 1226–1252.
- [22] J. JAFFRE, C. JOHNSON, AND A. SZEPESSY, *Convergence of the discontinuous Galerkin finite element method for hyperbolic conservation laws*, Math. Models Methods Appl. Sci., 5 (1995), pp. 367–386.
- [23] D. KRÖNER, *Numerical Schemes for Conservation Laws*, Wiley-Teubner, Stuttgart, 1997.
- [24] M.G. LARSON AND T.J. BARTH, *A posteriori error estimation for discontinuous Galerkin approximations of hyperbolic systems*, in Discontinuous Galerkin Methods, B. Cockburn, G. Karniadakis, and C.-W. Shu, eds., Lecture Notes in Comput. Sci. Engrg., 11, Springer, Berlin, 2000, pp. 363–368.
- [25] R. RANNACHER, *Adaptive finite element methods*, in Proc. NATO-Summer School, Error Control and Adaptivity in Scientific Computing, H. Bulgak and C. Zenger, eds., Kluwer, Dordrecht, 1998, pp. 247–278.
- [26] E. SÜLI, *A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems*, in An Introduction to Recent Developments in Theory and Numerics for Conservation Laws, D. Kröner, M. Ohlberger, and C. Rohde, eds., Lecture Notes in Comput. Sci. Engrg. 5, Springer, Berlin, 1998, pp. 123–194.

- [27] E. TADMOR, *Local error estimates for discontinuous solutions of nonlinear hyperbolic equations*, SIAM J. Numer. Anal., 28 (1991), pp. 891–906.
- [28] E. TORO, *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer, Berlin, 1997.
- [29] S. ULBRICH, *Optimal Control of Nonlinear Hyperbolic Conservation Laws with Source Terms*, Habilitation Thesis, Faculty for Mathematics, Munich Technical University, Munich, Germany, 2001.