MATHEMATICS OF COMPUTATION
VOLUME 54, NUMBER 189
JANUARY 1990, PAGES 107–129

ON THE CONVERGENCE OF SHOCK-CAPTURING STREAMLINE DIFFUSION FINITE ELEMENT METHODS FOR HYPERBOLIC CONSERVATION LAWS

CLAES JOHNSON, ANDERS SZEPESSY, AND PETER HANSBO

ABSTRACT. We extend our previous analysis of streamline diffusion finite element methods for hyperbolic systems of conservation laws to include a shock-capturing term adding artificial viscosity depending on the local absolute value of the residual of the finite element solution and the mesh size. With this term present, we prove a maximum norm bound for finite element solutions of Burgers' equation and thus complete an earlier convergence proof for this equation. We further prove, using entropy variables, that a strong limit of finite element solutions is a weak solution of the system of conservation laws and satisfies the entropy inequality associated with the entropy variables. Results of some numerical experiments for the time-dependent compressible Euler equations in two dimensions are also reported.

1. INTRODUCTION

In this note we continue our study of streamline diffusion finite element methods (SD methods for short below) for hyperbolic conservation laws started in [12, 13]. SD methods may be viewed as Petrov-Galerkin variants of the usual Galerkin finite element method with certain modifications of the test functions giving added stability without sacrificing accuracy (the error is of order $O(h^{k+1/2})$ for smooth solutions if polynomials of degree k are used). We recall that conventional finite element methods for hyperbolic problems lack in either stability, like the standard Galerkin method, giving spurious oscillations if the exact solution is nonsmooth, or in accuracy, like the classical artificial diffusion method with considerable smearing of sharp fronts and at most first-order accuracy. The basic SD method was proposed by Hughes in 1980/81, and the method has since been developed, theoretically and computationally, into a general finite element technique for hyperbolic-type problems with applications to convection-diffusion equations, the incompressible and compressible Euler and Navier-Stokes equations, see [5–9, 10, 11, 12–14, and 16–18].

The basic modification of the test functions in the SD method is obtained by adding a multiple of (a linearized form of) the hyperbolic operator involved

Received November 10, 1987; revised March 27, 1989.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 65M15.
Key words and phrases. Finite element method, conservation laws, convergence streamline diffusion, shock-capturing, entropy variables.

0025-5718/90 $1.00 + $.25 per page

applied to the test function (in a scalar convection problem, this corresponds to introducing artificial diffusion acting in the direction of the streamlines). The residual of the finite element solution will then be controlled in L_2 to a certain degree, which gives added stability. We recall that the usual Galerkin method is related to a weak formulation of the given hyperbolic equation, and thus we may say that the basic SD method seeks to find an approximate solution which satisfies (approximately) the given hyperbolic equation in both a weak and a strong sense.

Although the basic SD method gives a dramatic improvement over the standard Galerkin method, some over- and undershoots of approximate solutions may still persist at discontinuities or shocks. Recently, in the context of stationary problems, a second modification of the test functions was proposed in [7, 8], consisting of adding a certain 'shock-capturing' term which introduces some 'crosswind' control close to discontinuities. In numerical experiments this eliminated the oscillations at shocks without degrading the accuracy in smooth regions, and thus gave an SD method with very satisfactory properties, see [7] and also [12, 13], where the shock-capturing idea was extended to SD methods for time-dependent problems. However, no theoretical analysis explaining the remarkably improved properties of the shock-capturing SD method is available in the literature.

The main purpose of this note is now to initiate such an analysis. To this end, we shall first give a different interpretation of the shock-capturing term than that used in [7, 9] and [12, 13]. We shall view this term as a certain artificial viscosity with viscosity coefficient depending (locally) on the residual of the finite element solution, where the residual is obtained by inserting the finite element solution into the given hyperbolic differential equation. In fact, from this perspective we are led to somewhat different shock-capturing terms, e.g. for nonhomogeneous problems or time-dependent problems, than those proposed in [7, 9]. Further, by viewing the shock-capturing method in this way, it seems natural to expect the method to add significant artificial viscosity close to a discontinuity where the residual will be large, but only little in smooth regions where the residual may be small. Thus, the shock-capturing term would seem to act qualitatively as the artificial viscosity introduced in many finite difference schemes for hyperbolic conservation laws. Continuing this analogy, it seems as if the shock-capturing SD method, through the streamline diffusion modification, introduces high-order consistent 'streamline' artificial viscosity in the whole region, and through the shock-capturing term, first-order viscosity close to shocks with a continuous transition from first to higher order away from shocks. Thus, with piecewise linear continuous basis functions ($k = 1$) the shock-capturing SD method would seem to have the qualitative properties of a 'quasi second-order' finite difference scheme, which is second-order in smooth regions and first-order close to shocks. Recalling that the main difficulty in constructing such difference schemes is the artificial viscosity term, it seems possible that the shock-capturing SD method could offer a solution to this fundamental problem with wide applicability, since general meshes, variable coefficients and boundary conditions do not pose extra difficulties as in the finite difference case. To give

experimental support of this belief, we present some computational results for the time-dependent compressible Euler equations in two space dimensions.

With the new interpretation of the shock-capturing term, the extra stability introduced through this term becomes visible. As one example of how the shock-capturing term may be used in the theoretical analysis, we shall in this note prove a maximum norm bound for the SD solution of Burgers' equation in the case $k = 1$, thus filling a gap in an earlier convergence result [12]. Further convergence results based on a uniqueness result for measure-valued solutions are given in [16–18] for higher-order accurate shock-capturing SD approximations of a general scalar conservation law in several dimensions, with and without boundary conditions.

We now give an outline of the content of this paper, starting by recalling some of our earlier results. In [12] we proved the following two results for the basic SD method applied to Burgers' equation: (A) If a sequence of finite element solutions converges boundedly a.e. to a function u, then u is an entropy solution of Burgers' equation. (B) If the finite element solutions stay uniformly bounded, then a subsequence will converge a.e. to a function u. In this note we will prove that the hypothesis of (B) holds for the shock-capturing method in the case $k = 1$, that is, we will prove the following result: (C) The finite element solutions given by the shock-capturing SD method with $k = 1$ are uniformly bounded. Combining (A)–(C) we then obtain that a subsequence of the finite element solutions given by the shock-capturing SD method with $k = 1$ will converge to an entropy solution of Burgers' equation. The uniqueness of this solution follows as in [16] by showing that a limit of finite element solutions is an entropy solution in the Kruzhkov sense, so that all convex entropy inequalities are satisfied.

In [13] we proved for general hyperbolic conservation laws written in entropy variables a consistency result of type (A), that is, we proved that limits of finite element solutions given by the basic SD method are entropy solutions of the conservation law. In this note we extend this result to the shock-capturing SD method.

The material is organized as follows: In §2 we introduce the shock-capturing SD method for systems of hyperbolic conservation laws (written in entropy variables) and discuss the choice of the streamline and shock-capturing modifications. In §3 we prove the result of type (A) for systems in several dimensions. In §4 we prove the result (C) in the case of Burgers' equation. Finally, in §5 we briefly discuss some aspects of the numerical implementation of the method, which also includes automatic adaptivity of the finite element meshes, and we give some computational results related to the time-dependent Euler equations for compressible flow in two dimensions.

We denote by C a positive constant independent of the mesh parameter h, not necessarily the same at each occurrence. Further, for $\omega \subset \mathbb{R}^d$ we denote by $H^n(\omega)$ the Sobolev space of functions with derivatives of order $\leq n$ belonging to $L_2(\omega)$, and we use the notation

$$\|\cdot\|_{n,\omega} \equiv \|\cdot\|_{H^n(\omega)}, \qquad \|\cdot\|_{\omega} \equiv \|\cdot\|_{L_2(\omega)}.$$

2. Shock-capturing SD methods for systems of conservation laws

We consider a time-dependent hyperbolic system of conservation laws in $\mathbb{R}^d$, $d \geq 1$,

$$u_t + \sum_{i=1}^{d} f^i(u)_{x_i} = 0, \qquad t > 0, \ x \in \mathbb{R}^d, \tag{2.1a}$$

$$u(0, x) = u_0(x), \qquad x \in \mathbb{R}^d, \tag{2.1b}$$

where $u = (u_1, \dots, u_m)$, $m \geq 2$, $f^i : \mathbb{R}^m \to \mathbb{R}^m$ are given smooth functions, $i = 1, \dots, d$, and $u_0 \in [L_2(\mathbb{R}^d)]^m$ is a given initial function with compact support. Carrying out the differentiation in (2.1a), the system takes the form

$$u_t + \sum_{i=1}^{d} A_i u_{x_i} = 0, \qquad t > 0, \ x \in \mathbb{R}^d, \tag{2.2a}$$

$$u(0, x) = u_0(x), \qquad x \in \mathbb{R}^d, \tag{2.2b}$$

where $A_i = A_i(u) = f^i_u =$ the Jacobian of f^i are $m \times m$ matrices. We shall assume that (2.1) is equipped with a strictly convex entropy $\eta(u)$ with associated entropy flux $q(u) = (q^i(u))$ satisfying the compatibility relation

$$\eta_u f^i_u = q^i_u, \qquad i = 1, \dots, m, \tag{2.3}$$

where η_u denotes the gradient of $\eta(u)$. This assumption is satisfied by the usual systems of gas dynamics [4, 6]. The entropy condition for (2.1) then reads

$$\eta(u)_t + \sum_{i=1}^{d} q^i(u)_{x_i} \leq 0. \tag{2.4}$$

Introducing now the (invertible) change of variables [4, 6, 19] $\bar{u} = \eta_u(u)$, the system (2.2) takes the form

$$\bar{A}_0 \bar{u}_t + \sum_{i=1}^{d} \bar{A}_i \bar{u}_{x_i} = 0, \qquad t > 0, \ x \in \mathbb{R}^d, \tag{2.5a}$$

$$\bar{u}(0, x) = \bar{u}_0(x), \qquad x \in \mathbb{R}^d, \tag{2.5b}$$

where $\bar{A}_0 = \partial u / \partial \bar{u}$ and $\bar{A}_i = A_i \bar{A}_0$. Using the convexity of η and the compatibility (2.3), it follows that the $\bar{A}_i$ are symmetric, with $\bar{A}_0$ positive definite. Note that if the A_i are already symmetric, then η may be taken to be $\eta(u) = \frac{1}{2}|u|^2$, in which case $\bar{u} = u$ and (2.1) and (2.5) coincide; but in general, η is not quadratic and η_u is nonlinear.

The shock-capturing SD method for (2.1) will be based on the formulation (2.5) using the *entropy variables* $\bar{u}$. The advantage of using (2.5) as the starting point for a (Petrov-)Galerkin method may be explained as follows (cf. [5]). We

first note that integrating the entropy inequality (2.4) in x and t gives control of the entropy,

$$\int_{\mathbb{R}^d} \eta(u(t,\cdot))\,dx \le \int_{\mathbb{R}^d} \eta(u_0)\,dx\,. \tag{2.6}$$

Secondly, we recall that (2.4) results from (2.1a) by multiplying with η_u (for smooth solutions, (2.1a) implies (2.4) with equality, and for nonsmooth (entropy) solutions, (2.4) follows through a viscous regularization of (2.1), adding, e.g., a term $-\varepsilon\Delta u$ and letting $\varepsilon \to 0$). Alternatively, (2.6) follows from (2.5a) by multiplication by $\overline{u}$, since $\overline{u}\overline{A}_0\overline{u}_t = \eta_u u_t = \eta_t$. Thus, to obtain the entropy control (2.6), we multiply with η_u in (viscous regularizations of) (2.1a) and with $\overline{u}$ in (2.5a). Now, in a Galerkin method for an equation $A(w) = f$, we typically may multiply by w itself but not easily by nonlinear functions of w (cf. Remarks 2.4 and 4.1 below). Thus, (2.5) may be viewed as a better starting point for a Galerkin method than (2.1), since the entropy control (2.6) is automatically built in, using (2.5). However, to use a standard Galerkin method on (2.5) is not enough; to be able to prove that limits of finite element solutions of (2.5) satisfy the entropy inequality (2.4) locally and not just globally as stated in (2.6), we also need a streamline diffusion modification (cf. the proof of Theorem 3.1 below).

We are now ready to introduce the finite element space to be used in the SD method for (2.1). Let $0 = t_0 < t_1 < t_2 < \cdots$ be a sequence of time levels, set $I_n = (t_n, t_{n+1})$ and introduce the 'slabs' $S_n = \mathbb{R}^d \times I_n$. For $h > 0$ and $n = 0, 1, 2, \ldots$, let T_h^n be a, for simplicity quasi-uniform, triangulation of S_n into triangles K of diameter $h_K \sim h$ with smallest angle uniformly bounded away from zero, and define for a given $k \ge 1$,

$$V_h^n = \{v \in [H^1(S_n)]^m : v|_K \in P_k(K),\ K \in T_h^n\},$$

where $P_k(K)$ denotes the set of polynomials on K of degree at most k. In other words, V^h consists of continuous piecewise polynomials on the slab S_n. Typically, $t_{n+1} - t_n \sim h$, with the slab S_n one element wide. Note that since u_0 has compact support, it follows that also the solution u has compact support in $\mathbb{R}^d \times [0, t]$ for any t. This means that we may restrict the functions in V_h^n to be zero for $|x|$ large.

We seek an approximate solution $\overline{U} = \overline{U}_h$ in the space $V_h = \prod_{n\ge 0} V_h^n$, i.e., for $n = 0, 1, 2, \ldots$, we will have

$$\overline{U}|_{S_n} \in V_h^n\,.$$

Note that the functions in V_h are continuous in x and possibly discontinuous in t at the discrete time levels t_n. The shock-capturing SD method for (2.1), based

on (2.5), can now be formulated: Find $\overline{U} \in V_h$ such that for $n = 0, 1, 2, \ldots$

$$
\begin{aligned}
&\int_{S_n} \left(\overline{A}_0(\overline{U})\overline{U}_t + \sum_i \overline{A}_i(\overline{U})\overline{U}_{x_i} \right) \\
&\qquad \cdot \left(\overline{v} + \delta \left(\overline{A}_0(\overline{U})\overline{v}_t + \sum_i \overline{A}_i(\overline{U})\overline{v}_{x_i} \right) \right) dx\, dt \\
&\qquad + \overline{\delta} \int_{S_n} \frac{|\overline{A}_0(\overline{U})\overline{U}_t + \sum_i \overline{A}_i(\overline{U})\overline{U}_{x_i}|}{\varepsilon + |\nabla \overline{U}|} \nabla \overline{U} \cdot \nabla \overline{v}\, dx\, dt \\
&\qquad + \overline{\overline{\delta}} \int_{S_n} |\widetilde{U}| \nabla_x \overline{U} \cdot \nabla_x \overline{v}\, dx\, dt \\
&\qquad + \int_{\mathbb{R}^d} \left(U_+^n - U_-^n \right) \cdot \overline{v}_+^n\, dx = 0, \quad \forall \overline{v} \in V_h^n,
\end{aligned}
\tag{2.7}
$$

where dot denotes the usual scalar product in $\mathbb{R}^m$, $\mathbb{R}^d$ or $\mathbb{R}^{d+1}$ with corresponding norm $|\cdot|$. Further, $\overline{U}$ and U are related through $\overline{U} = \eta_u(U)$ (note that the original variable U occurs in the last term in (2.7)). We also use the notation

$$
\begin{aligned}
v_\pm^n(t, x) &= \lim_{s \to 0\pm} v(t_n + s, x), \qquad U_-^0 = u_0, \\
\nabla_x v &= (v_{x_1}, \ldots, v_{x_d}), \qquad \nabla v = (v_t, v_{x_1}, \ldots, v_{x_d}), \\
\nabla_x v \cdot \nabla_x w &\equiv \sum_{i=1}^d v_{x_i} \cdot w_{x_i}, \qquad \nabla v \cdot \nabla w = v_t w_t + \nabla_x v \cdot \nabla_x w,
\end{aligned}
$$

and for all $K \in T_h^n$:

$$
\widetilde{U}|_K = \begin{cases} (U_+ - U_-)|_{K \cap (\mathbb{R}^d \times \{t_n\})} & \text{if } \int_{K \cap (\mathbb{R}^d \times \{t_n\})} dx > 0, \\ 0, & \text{otherwise}. \end{cases}
$$

Finally, ε, $\overline{\delta}$ and $\overline{\overline{\delta}}$ are parameters tending to zero as $h \to 0$, and $\delta = \delta(\overline{U})$ is a positive definite $m \times m$ matrix, the choice of which we specify in Remark 2.1. Concerning the choice of ε, $\overline{\delta}$ and $\overline{\overline{\delta}}$, we normally expect to have $\varepsilon, \overline{\delta}, \overline{\overline{\delta}} = \mathcal{O}(h^\alpha)$, with $\alpha \approx 1$ (cf. §4 below). The streamline diffusion modification of the test functions is given by the δ-term, while the shock-capturing is related to $\overline{\delta}$- and $\overline{\overline{\delta}}$-terms, which clearly correspond to artificial viscosity terms, with the viscosity coefficients depending on the two components of the residual, namely $|U_t + \sum_i f^i(U)_{x_i}|$ and $|\widetilde{U}|$.

Remark 2.1. *The choice of* δ. The simplest choice of δ is given by $\delta = ChI$ with I the identity matrix. As pointed out in [7], this choice is not adequate in some situations. To see this, consider a constant-coefficient variant of (2.5) in the case of one space dimension:

$$
A_0 w_t + A w_x = 0, \tag{2.8}
$$

where thus A_0 and A are symmetric, with A_0 positive definite. Let now $E = (A_0)^{-1/2}P$, where P is an orthogonal matrix consisting of eigenvectors of $\tilde{A} \equiv (A_0)^{-1/2}A(A_0)^{-1/2}$. Then E diagonalizes A_0 and A,

$$E^T A_0 E = I, \qquad E^T A E = \Lambda = \operatorname{diag}(\lambda_i), \tag{2.9}$$

with E^T denoting the transpose of E and Λ a diagonal matrix with elements $\lambda_i =$ the eigenvalues of $\tilde{A}$. Introducing the new variable $\overline{w}$ by $w = E\overline{w}$, we have, using (2.9),

$$\begin{aligned}(A_0 w_t + A w_x)\cdot(v + \delta(A_0 v_t + A v_x)) \\ = (A_0 E\overline{w}_t + AE\overline{w}_x)\cdot(E\overline{v} + \delta(A_0 E\overline{v}_t + AE\overline{w}_x)) \\ = (\overline{w}_t + \Lambda\overline{w}_x)\cdot(\overline{v} + E^{-1}\delta E^{-T}(\overline{v}_t + \Lambda\overline{v}_t)) .\end{aligned}$$

Since $\overline{w}_t + \Lambda\overline{w}_x = 0$ is an uncoupled system of m scalar equations, we are led, in analogy with the scalar case [6, 10], to choose in an SD method for (2.8)

$$E^{-1}\delta E^{-T} = h(I + \Lambda^2)^{-1/2} = h\operatorname{diag}(\mu_i), \qquad \mu_i = (1 + \lambda_i^2)^{-1/2},$$

i.e.,

$$\delta = hE(I + \Lambda^2)^{-1/2}E^T = hA_0^{-1/2}(I + \tilde{A}^2)^{-1/2}(A_0)^{-1/2} . \tag{2.10}$$

If now the μ_i vary considerably in size, then $\operatorname{diag}(\mu_i)$ is not close to any multiple of I and thus, if we choose $\delta = ChI$, then some of the components in the corresponding SD method for (2.8) would not get the correct streamline modification. Note that $\tilde{A}$ and A_1 in (2.2) have the same eigenvalues, i.e., λ_i are the eigenvalues of A_1.

In the case $d > 1$, with (2.8) replaced by $A_0 w_t + \sum A_i w_{x_i} = 0$, it is not in general possible to diagonalize all the matrices A_i with the same transformation. A natural generalization of (2.10) to the case $d > 1$ is given by

$$\delta = h(A_0)^{-1/2}\left(I + \sum_{i=1}^{d}\tilde{A}_i^2\right)^{-1/2}(A_0)^{-1/2}, \tag{2.11}$$

where $\tilde{A}_i = (A_0)^{-1/2}A_i(A_0)^{-1/2}$.

In (2.7) we now choose $\delta = \delta(t, x)$ according to (2.11), with the A_i replaced by $\overline{A}_i(\overline{U}(t, x))$.

Remark 2.2. It is possible to generalize the shock-capturing terms by replacing $\nabla\overline{U}\cdot\nabla\overline{v}$ by $M_0\overline{U}_t\overline{v}_t + \sum_i M_i\overline{U}_{x_i}\overline{v}_{x_i}$, where M_i, $i = 0, \ldots, d$, are positive definite $m \times m$ matrices. Various choices of M_i have been proposed in [8]. In the case of one space dimension, a diagonalization as in Remark 2.1 may be used to find suitable M_i. In several dimensions, the choice is less clear. It may be natural to choose $M_i = A_0$, $i = 0, \ldots, d$, corresponding to adding diffusion close to shocks in the form $-h\Delta u$ in the conservation variables. So far, we have used $M =$ Identity in the computations. □

Remark 2.3. Note that (2.7), although expressed in entropy variables, may be considered to have "conservation form", since on each slab

$$U_t + \sum_i f^i(U)_{x_i} \equiv \overline{A}_0(\overline{U})\overline{U}_t + \sum_i \overline{A}_i(\overline{U})\overline{U}_{x_i} .$$

In particular, this means that the correct Rankine-Hugoniot conditions are satisfied by limits of solutions of (2.7), see Theorem 3.1 below. □

Remark 2.4. In the recent work [18] it is proved that the shock-capturing SD method may be applied also in conservation variables with the entropy control and entropy consistency maintained. This is related to the fact that with the shock-capturing term present, in a Galerkin method in conservation variables it is possible to multiply by η_u, even with η_u nonlinear (cf. Remark 4.1). A formulation in conservation variables seems to require less computational work, but more computational experience is needed to evaluate the merits using entropy or conservation variables. □

3. Convergence towards entropy solutions

In this section we prove that limits of finite element solutions given by the SD method (2.7) are entropy solutions of the conservation law (2.1). We recall that $u \in [L_\infty(\Omega)]^m$, $\Omega = (0, \infty) \times \mathbb{R}^d$, is an entropy solution of (2.1) if for all $\varphi \in [C_0^\infty(\overline{\Omega})]^m$, $\overline{\Omega} = [0, \infty) \times \mathbb{R}^d$, we have

$$\int_\Omega \left(u \cdot \varphi_t + \sum_i f^i(u) \cdot \varphi_{x_i} \right) dt\, dx + \int_{\mathbb{R}^d} u_0 \cdot \varphi(0, \cdot)\, dx = 0, \tag{3.1}$$

and for all $\varphi \in C_0^\infty(\Omega)$ with $\varphi \geq 0$,

$$\int_\Omega \left(\eta \varphi_t + \sum_i q^i \varphi_{x_i} \right) dt\, dx \geq 0 . \tag{3.2}$$

We assume that the entropy η is strictly convex, i.e., that there is a compact set $D \subset \mathbb{R}^m$ and positive constants σ, α_1 and α_2 such that for all $v, w \in D \subset \mathbb{R}^m$ with $\delta = \delta(\overline{U}) = \delta(\eta_u(U))$

$$\eta(v) - \eta(w) - \eta_u(w) \cdot (v - w) \geq \sigma |v - w|^2 , \tag{3.3}$$

$$\alpha_1 h \leq x \cdot \delta x , \quad |x \cdot \delta y| \leq \alpha_2 h \qquad \forall x, y \in \mathbb{R}^m , \ |x| = |y| = 1 . \tag{3.4}$$

We have the following result, where η_u^{-1} denotes the inverse of $\eta_u : D \to \eta_u(D)$ and $\overline{u} = \eta_u(u)$. For definiteness we assume here that $\varepsilon = \overline{\delta} = \overline{\overline{\delta}} = h$.

Theorem 3.1. *Suppose that a sequence of finite element solutions* $\{\overline{U}_h\}$ *of* (2.7) *with* Range $(\overline{U}_h) \subset \eta_u(D)$ *converges boundedly a.e. in* Ω *to a function* $\overline{u}$ *as* h *tends to zero. Then* $u = \eta_u^{-1}(\overline{u})$ *satisfies* (3.1) *and* (3.2), *and thus* u *is an entropy solution of* (2.1).

The proof is based on the following stability estimate where σ and α_1 are given in (3.3) and (3.4), respectively.

Lemma 3.1. *For* $N = 1, 2, \dots,$ *we have*

$$\int_{\mathbb{R}^d} \eta(U_-^N)\,dx + \sigma \sum_{n=0}^{N-1} \|U_+^n - U_-^n\|_{\mathbb{R}^d}^2 + \alpha_1 h \sum_{n=0}^{N-1} \left\| U_t + \sum_i f^i(U)_{x_i} \right\|_{S_n}^2$$
$$+ \bar{\delta} \sum_{n=0}^{N-1} \int_{S_n} \frac{|U_t + \sum_i f^i(U)_{x_i}|\,|\nabla \bar{U}|^2}{h + |\nabla \bar{U}|}\,dt\,dx$$
$$+ \bar{\bar{\delta}} \sum_{n=0}^{N-1} \int_{S_n} |\tilde{U}|\,|\nabla_x \bar{U}|^2\,dt\,dx \le \int_{\mathbb{R}^d} \eta(u_0)\,dx\,.$$

Proof. Taking $\bar{v} = \bar{U} = \eta_u(U)$ in (2.7), we get

$$\int_{S_n} \left(\eta_t(U) + \sum_i q^i(U)_{x_i} \right) dt\,dx$$
$$+ \int_{S_n} \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot \delta \left(U_t + \sum_i f^i(U)_{x_i} \right) dt\,dx$$
$$+ \bar{\delta} \int_{S_n} \left| U_t + \sum_i f^i(U)_{x_i} \right| \frac{|\nabla \bar{U}|^2}{h + |\nabla \bar{U}|}\,dt\,dx + \bar{\bar{\delta}} \int_{S_n} |\tilde{U}|\,|\nabla_x \bar{U}|^2\,dt\,dx$$
$$+ \int_{\mathbb{R}^d} \left(U_+^n - U_-^n \right) \cdot \eta_u \left(U_+^n \right) dx = 0,$$

so that, since $U(t, x) = 0$ for $|x|$ large,

$$0 = \sum_{n=0}^{N-1} \int_{\mathbb{R}^d} \left(\eta \left(U_-^{n+1} \right) - \eta \left(U_+^n \right) + \eta_u \left(U_+^n \right) \cdot \left(U_+^n - U_-^n \right) \right) dx$$
$$+ \sum_{n=0}^{N-1} \int_{S_n} \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot \delta \left(U_t + \sum_i f^i(U)_{x_i} \right) dt\,dx$$
$$+ \bar{\delta} \sum_{n=0}^{N-1} \int_{S_n} \left| U_t + \sum_i f^i(U)_{x_i} \right| \frac{|\nabla \bar{U}|^2}{h + |\nabla \bar{U}|}\,dt\,dx + \bar{\bar{\delta}} \sum_{n=0}^{N-1} \int_{S_n} |\tilde{U}|\,|\nabla_x \bar{U}|^2\,dt\,dx$$
$$\ge \int_{\mathbb{R}^d} \eta(U_-^N)\,dx - \int_{\mathbb{R}^d} \eta(u_0)\,dx + \alpha_1 h \sum_{n=0}^{N-1} \int_{S_n} \left| U_t + \sum_i f^i(U)_{x_i} \right|^2 dt\,dx$$
$$+ \bar{\delta} \sum_{n=0}^{N-1} \int_{S_n} \left| U_t + \sum_i f^i(U)_{x_i} \right| \frac{|\nabla \bar{U}|^2}{h + |\nabla \bar{U}|}\,dt\,dx$$
$$+ \bar{\bar{\delta}} \sum_{n=0}^{N-1} \int_{S_n} |\tilde{U}|\,|\nabla_x \bar{U}|^2\,dt\,dx$$
$$+ \sum_{n=0}^{N-1} \int_{\mathbb{R}^d} \left(\eta \left(U_-^n \right) - \eta \left(U_+^n \right) - \eta_u \left(U_+^n \right) \cdot \left(U_-^n - U_+^n \right) \right) dx\,.$$

The lemma now follows from (3.3). □

We also need the following interpolation estimate and "superapproximation" result, where $\pi_h w \in V_h$ is a standard interpolate of a function $w \in \prod_{n\geq 0}[H^1(S_n) \cap \mathscr{C}(S_n)]^m$, $\mathscr{C}(S_n)$ is the space of continuous functions on S_n and $\|\cdot\|_{k,\omega}$ denotes the norm in the Sobolev space $[H^k(\omega)]^m$. A proof of the superapproximation result is given in [18].

Lemma 3.2. *There are constants C such that for $w \in [H^1(S_n) \cap \mathscr{C}(S_n)]^m$, $\varphi \in H^1(S_n) \cap \mathscr{C}(S_n)$, $v \in V_h$, $n = 0, 1, 2, \ldots,$ and $k = 0, 1,$*

$$h^k \|w - \pi_h w\|_{k,S_n} + \sqrt{h}\|w_+^n - (\pi_h w)_+^n\|_{\mathbb{R}^d} \leq Ch^2 \|w\|_{2,S_n},$$
$$h^k \|v\varphi - \pi_h(v\varphi)\|_{k,S_n} + \sqrt{h}\|(v\varphi)_+^n - (\pi_h(v\varphi))_+^n\|_{\mathbb{R}^d}$$
$$\leq Ch\|v\|_{L_\infty(S_n)}(\|\varphi\|_{1,S_n} + h\|\varphi\|_{2,S_n}).$$

Remark 3.1. Note that the superapproximation (2.4) and interpolation estimates (2.3) in [12] are not stated correctly, but should be replaced by Lemma 3.2 above or the following variant thereof: For $\Omega = R \times (0, \infty)$, $\varphi \in C_0^\infty(\Omega)$, $v \in V_h$,

$$\|v\varphi - \pi_h(v\tilde{\varphi})\|_{s,\Omega} \leq Ch^{1-s}\|v\|_{L_\infty(S_n)}\|\varphi\|_{1,\Omega}, \qquad s = 0, 1, \tag{3.5a}$$

$$\sum_{n=0}^{\infty} h\|(v\varphi)_+^n - (\pi_h(v\tilde{\varphi}))_+^n\|_R^2 \leq Ch^2\|v\|_{L_\infty(S_n)}^2\|\varphi\|_{1,\Omega}^2, \tag{3.5b}$$

together with the corresponding estimates for $v \equiv 1$. Here, $\tilde{\varphi} \equiv \varphi * \omega_h$ is a mollification of φ, where ω_h is defined by

$$\omega_h(x, t) = \omega_h^0(x)\omega_h^0(t), \qquad \omega_h^0(s) = h^{-1}\omega^0(s/h),$$
$$0 \leq \omega^0 \in C_0^\infty(R), \quad \int_R \omega^0(s)\,ds = 1, \quad \operatorname{supp}\omega^0 = [-1, 1].$$

The estimates (3.5) are proved in the appendix. The proofs of Theorems 3.1 and 4.1 in [12] should then be modified by replacing φ by $\tilde{\varphi}$. □

We can now give the proof.

Proof of Theorem 3.1. To prove that u satisfies (3.1), we take $\overline{v} = \pi_h \varphi$ in (2.7) where $\varphi \in [C_0^\infty(\overline{\Omega})]^m$ to get

$$\begin{aligned}
&\int_{S_n} \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot \varphi \, dt\, dx + \int_{\mathbb{R}^d} \left(U_+^n - U_-^n \right) \cdot \varphi_+^n \, dx \\
&= \int_{S_n} \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot (\varphi - \pi_h \varphi) \, dt\, dx \\
&\quad + \int_{\mathbb{R}^d} \left(U_+^n - U_-^n \right) \cdot \left(\varphi_+^n - (\pi_h \varphi)_+^n \right) dx \\
&\quad - \int_{S_n} \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot \delta \left(\overline{A}_0(\overline{U}) (\pi_h \varphi)_t + \sum_i \overline{A}_i(\overline{U}) (\pi_h \varphi)_{x_i} \right) dt\, dx \\
&\quad - \overline{\delta} \int_{S_n} \frac{|U_t + \sum_i f^i(U)_{x_i}|}{h + |\nabla \overline{U}|} \nabla \overline{U} \cdot \nabla (\pi_h \varphi) \, dt\, dx \\
&\quad - \overline{\overline{\delta}} \int_{S_n} |\widetilde{U}| \nabla_x \overline{U} \cdot \nabla_x (\pi_h \varphi) \, dt\, dx \\
&\equiv E_n^1 + E_n^2 + E_n^3 + E_n^4 + E_n^5 \, .
\end{aligned}$$

Integrating by parts and summing over n, gives

$$- \int_\Omega \left(U \cdot \varphi_t + \sum_i f^i(U) \cdot \varphi_{x_i} \right) dt\, dx - \int_{\mathbb{R}^d} u_0 \cdot \varphi_+^0 \, dx = \sum_{j=1}^5 \sum_{n \geq 0} E_n^j \equiv \sum_{j=1}^5 R^j \, .$$

By Lemmas 3.1 and 3.2 and using the assumption that $\|\overline{U}\|_{L_\infty(\Omega)}$ is uniformly bounded in h, we easily find that $|R^j| \leq C\sqrt{h}$, $j = 1, \ldots, 5$, and (3.1) follows by letting h tend to zero, using Lebesgue's dominated convergence theorem.

Next, taking $\overline{v} = \pi_h(\overline{U}\varphi)$ in (2.7) with $\varphi \in C_0^\infty(\Omega)$, $\varphi \geq 0$, we get by integrating by parts and summing over $n \geq 0$,

$$\begin{aligned}
&- \int_\Omega \left(\eta(U) \varphi_t + \sum_i q^i(U) \varphi_{x_i} \right) dt\, dx \\
&\quad + \sum_{n \geq 0} \int_{\mathbb{R}^d} \left(\eta\left(U_-^n\right) - \eta\left(U_+^n\right) - \eta_u\left(U_+^n\right) \cdot \left(U_-^n - U_+^n\right) \right) \varphi \, dx \\
&\quad + \int_\Omega \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot \delta \left(U_t + \sum_i f^i(U)_{x_i} \right) \varphi \, dt\, dx \\
&\quad + \overline{\delta} \int_\Omega \left| U_t + \sum_i f^i(U)_{x_i} \right| \frac{|\nabla \overline{U}|^2}{h + |\nabla \overline{U}|} \varphi \, dt\, dx \\
&\quad + \overline{\overline{\delta}} \sum_{n \geq 0} \int_{S_n} |\widetilde{U}| \, |\nabla_x \overline{U}|^2 \varphi \, dt\, dx \equiv \sum_{j=1}^8 \sum_{n \geq 0} F_n^j \equiv \sum_{j=1}^8 G^j \, ,
\end{aligned}$$

where

$$F_n^1 = \int_{S_n} \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot (\overline{U}\varphi - \pi_h(\overline{U}\varphi))\, dt\, dx,$$

$$F_n^2 = \int_{\mathbb{R}^d} (U_+^n - U_-^n) \cdot ((\overline{U}\varphi)_+^n - \pi(\overline{U}\varphi)_+^n)\, dx,$$

$$F_n^3 = \int_{S_n} \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot \delta \left[\overline{A}_0(\overline{U})((\overline{U}\varphi)_t - (\pi_h(\overline{U}\varphi))_t) + \sum_i \overline{A}_i(\overline{U})((\overline{U}\varphi)_{x_i} - (\pi_h(\overline{U}\varphi))_{x_i}) \right] dt\, dx,$$

$$F_n^4 = -\int_{S_n} \left(U_t + \sum_i f^i(U)_{x_i} \right) \cdot \delta \left(\overline{A}_0(\overline{U})\overline{U}\varphi_t + \sum_i \overline{A}_i(\overline{U})\overline{U}\varphi_{x_i} \right) dt\, dx,$$

$$F_n^5 = \overline{\delta} \int_{S_n} \frac{|U_t + \sum_i f^i(U)_{x_i}|}{h + |\nabla \overline{U}|} \nabla \overline{U} \cdot \nabla (\overline{U}\varphi - \pi_h(\overline{U}\varphi))\, dt\, dx,$$

$$F_n^6 = -\overline{\delta} \int_{S_n} \frac{|U_t + \sum_i f^i(U)_{x_i}|}{h + |\nabla \overline{U}|} \left(\overline{U}_t \cdot \overline{U}\varphi_t + \sum_i \overline{U}_{x_i} \cdot \overline{U}\varphi_{x_i} \right) dt\, dx,$$

$$F_n^7 = \overline{\overline{\delta}} \int_{S_n} |\widetilde{U}| \nabla_x \overline{U} \cdot \nabla_x (\overline{U}\varphi - \pi_n(\overline{U}\varphi))\, dt\, dx,$$

$$F_n^8 = -\overline{\overline{\delta}} \int_{S_n} |\widetilde{U}| \left(\sum_i \overline{U}_{x_i} \cdot \overline{U}\varphi_{x_i} \right) dt\, dx.$$

Arguing as above, we see that $|G^j| \le C\sqrt{h}$, $j = 1, \dots, 8$, and (3.2) follows letting h tend to zero, using the convexity of η. □

4. Convergence for Burgers' equation

As an example of the theoretical use of the shock-capturing terms we shall in this section prove uniform boundedness of the finite element solutions of a shock-capturing SD method with $k = 1$ (piecewise linears) applied to Burgers' equation

$$u_t + uu_x = 0, \qquad \text{in } \Omega = \mathbb{R} \times (0, \infty), \tag{4.1a}$$

$$u(x, 0) = u_0(x), \qquad x \in \mathbb{R}, \tag{4.1b}$$

where u_0 is a given bounded function with compact support. In [12] we proved that a subsequence of solutions U_h of an SD method for Burgers' equation without shock-capturing terms converges a.e. to an entropy solution u of (4.1) corresponding to the entropy $\eta = u^2/2$. In the proof we explicitly assumed that $\|U_h\|_{L_\infty(\Omega)}$ remains bounded as h tends to zero. As indicated, we shall now prove, by using the shock-capturing terms, that $\|U_h\|_{L_\infty(\Omega)}$ in fact is uniformly bounded. In this way we thus obtain a complete convergence proof for the

shock-capturing SD method applied to Burgers' equation; cf. [16], where also uniqueness is proved by proving that the limit function u satisfies all entropy inequalities related to convex entropies, and thus is the solution in the Kruzhkov sense.

Using the notation of §3 with $m = k = 1$, we now consider the SD method (2.7) applied to (4.1) (with $\overline{U} = U$ corresponding to $\eta(u) = \frac{1}{2}u^2$): Find $U \equiv U_h \in V_h$ such that for $n = 0, 1, 2, \ldots$

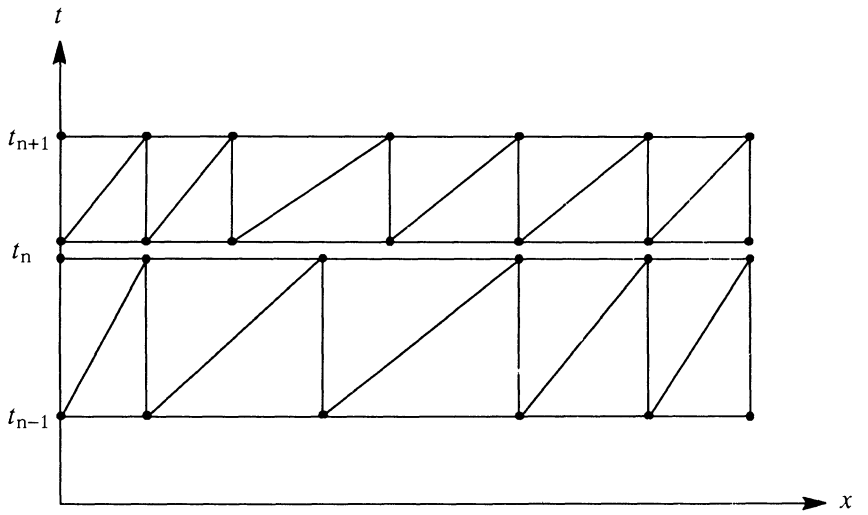
$$
\begin{aligned}
&\int_{S_n} (U_t + UU_x)(v + \delta(v_t + Uv_x))\,dx\,dt \\
&\quad + \overline{\delta} \int_{S_n} \frac{|U_t + UU_x|}{\varepsilon + |\nabla U|}(1 + |U|)\nabla U \cdot \nabla v\,dx\,dt \\
&\quad + \overline{\overline{\delta}} \int_{S_n} |\widetilde{U}| U_x v_x\,dx\,dt + \int_R (U_+^n - U_-^n)v_+^n\,dx = 0 \quad \forall v \in V_h^n,
\end{aligned} \tag{4.2}
$$

where $U_-^0 \equiv u_0$ and

$$
\widetilde{U}|_K \equiv \begin{cases} (U_+ - U_-)|_{K \cap R_n} & \text{if } K \cap R_n \text{ is an edge of } K, \\ 0 & \text{otherwise} \end{cases} \qquad \forall K \in T_h^n,
$$

$$
R_n = R \times \{t_n\}, \qquad n = 0, 1, 2, \ldots .
$$

Further, $\delta, \overline{\delta}$ and $\overline{\overline{\delta}}$ are positive parameters satisfying $\delta = Ch$, $\overline{\delta} = Ch^{\alpha_1}$ and $\overline{\overline{\delta}} = Ch^{\alpha_2}$, where the α_i are constants with $0 < \alpha_1, \alpha_2 < 1$. We also assume that the triangles $K \in T_h^n$ have right angles with two sides parallel to the x- and t-axis, so that the space-time triangulation has the following form (note that the meshes on adjacent slabs do not have to match).



Our main result is now the following, choosing $\varepsilon = 0$ for simplicity:

Theorem 4.1. *Suppose that $u_0 \in L_\infty(R)$ has compact support. Then there is a constant C such that the solutions $U \equiv U_h$ of (4.2) satisfy*

$$\|U_h\|_{L_\infty(\Omega)} \le C\,, \qquad h > 0\,. \tag{4.3}$$

To prove this result, we first state the basic stability estimate for (4.2) obtained by taking $v = U$:

$$\begin{aligned} &\delta\|U_t + UU_x\|^2_{\Omega_M} + \frac{1}{2}\sum_{n=0}^{M}\|U_+^n - U_-^n\|_R^2 \\ &\quad + \frac{1}{2}\|U_-^{M+1}\|_R^2 + \bar{\delta}\int_{\Omega_M}\frac{|U_t + UU_x|}{|\nabla U|}(1+|U|)|\nabla U|^2\,dx\,dt \\ &\quad + \bar{\bar{\delta}}\int_{\Omega_M}|\tilde{U}|(U_x)^2\,dx\,dt \le \frac{1}{2}\|u_0\|_R^2\,, \qquad M = 0, 1, 2, \dots, \end{aligned} \tag{4.4}$$

where $\Omega_M = \bigcup_{n=0}^{M} S_n$. Note that the integrals over Ω_M are to be interpreted as a sum of integrals over the S_n with $n \le M$. We shall further need the following two preliminary results.

Lemma 4.1. *There is a positive constant c independent of p such that for $p = 2m$, $m = 1, 2, \dots$ and $n = 0, 1, 2, \dots$,*

$$\begin{aligned} &ch\sum_{K\in T_h^n}\int_{R_n\cap K}|U_+^n - U_-^n|(U_+^n)_x^2\|U_+^n\|^{p-2}_{L_\infty(K\cap R_n)}\,dx \\ &\qquad \le \int_{S_n}|\tilde{U}|U_x(\pi_h(U^{p-1}))_x\,dx\,dt\,. \end{aligned} \tag{4.5a}$$

Proof. Considering one triangle $K \in T_h^n$ with vertices (x_1, t_n), (x_2, t_n), (x_2, t_{n+1}), where $x_1 < x_2$, we note that $U_x|_K$ and $(\pi_h(U^{p-1}))_x|_K$ are constant and $|\tilde{U}|$ depends only on x. Hence,

$$\begin{aligned} &\int_K|\tilde{U}|U_x(\pi_h(U^{p-1}))_x\,dx\,dt \\ &\quad = \frac{t_{n+1}-t_n}{x_2-x_1}\int_{x_1}^{x_2}|U_+^n - U_-^n|U_x(\pi_h(U^{p-1}))_x(x-x_1)\,dx\,, \end{aligned}$$

and further, by simple computations,

$$
\begin{aligned}
&U_x(\pi_h(U^{p-1}))_x \\
&\quad = \frac{1}{(x_2 - x_1)^2}(U(x_2, t_n) - U(x_1, t_n))(U^{p-1}(x_2, t_n) - U^{p-1}(x_1, t_n)) \\
&\quad = \frac{1}{(x_2 - x_1)^2}(U(x_2, t_n) - U(x_1, t_n))^2(\max(|U(x_2, t_n)|, |U(x_1, t_n)|))^{p-2} \\
&\qquad \cdot \sum_{l=0}^{p-2}[\operatorname{sign}(U(x_2, t_n)U(x_1, t_n)) \\
&\qquad\qquad \cdot \min(|U(x_2, t_n)|, |U(x_1, t_n)|)/\max(|U(x_2, t_n)|, |U(x_1, t_n)|)]^l \\
&\quad \geq \frac{1}{2(x_2 - x_1)^2}(U(x_2, t_n) - U(x_1, t_n))^2 \|U_+^n\|_{L_\infty(K \cap R_n)}^{p-2} \\
&\quad = \frac{1}{2}(U_+^n)_x^2 \|U_+^n\|_{L_\infty(K \cap R_n)}^{p-2},
\end{aligned}
$$

where $\operatorname{sign}(x) = x/|x|$ if $|x| > 0$ and $\operatorname{sign}(0) = 0$.

Now let $f: S^1 \to R$, $S^1 = \{x \in R^2 : |x| = 1\}$, be defined by

$$
f(y_1, y_2) = \frac{\int_0^1 |y_1(1-x) + y_2 x| x \, dx}{\int_0^1 |y_1(1-x) + y_2 x| \, dx}.
$$

For $|\tilde{U}| \neq 0$ we then have

$$
\frac{\int_{x_1}^{x_2} |U_+^n - U_-^n|(x - x_1) \, dx}{\int_{x_1}^{x_2} |U_+^n - U_-^n| \, dx} \geq Ch \inf_{y \in S^1} f(y) \geq Ch,
$$

since f is continuous and strictly positive on S^1. This proves the lemma, since (4.5) is trivially true when $|\tilde{U}| \equiv 0$. □

Lemma 4.2. *There is a constant $c > 0$ independent of p such that for $p = 2m$, $m = 1, 2, 3, \ldots$, $n = 0, 1, 2, \ldots$,*

$$
\begin{aligned}
&\int_{S_n} \frac{|U_t + UU_x|}{|\nabla U|}(1 + |U|)\nabla U \cdot \nabla \pi_h(U^{p-1}) \, dx \, dt \\
&\qquad \geq \frac{c}{p^2} \sum_{K \in T_h^n} \int_K \frac{|U_t + UU_x|}{|\nabla U|}(1 + |U|)|\nabla U|^2 \|U\|_{L_\infty(K)}^{p-2} \, dx \, dt.
\end{aligned}
$$

The proof of Lemma 4.2 is analogous to that of Lemma 4.1 (see [16, 18] for details).

Proof of Theorem 4.1. Taking $v = \pi_h(U^{p-1})$ in (4.2), we get with p an even number ≥ 4,

$$
\begin{aligned}
0 = &\int_{S_n} (U_t + UU_x)U^{p-1}\,dx\,dt + \int_R (U_+^n - U_-^n)(U_+^n)^{p-1}\,dx \\
&- \int_{S_n} (U_t + UU_x)(U^{p-1} - \pi_h(U^{p-1}))\,dx\,dt \\
&- \int_R (U_+^n - U_-^n)((U_+^n)^{p-1} - (\pi_h(U^{p-1}))_+^n)\,dx \\
&+ \delta \int_{S_n} (U_t + UU_x)((U^{p-1})_t + U(U^{p-1})_x)\,dx\,dt \\
&- \delta \int_{S_n} (U_t + UU_x)((U^{p-1})_t - (\pi_h U^{p-1})_t + U((U^{p-1})_x - (\pi_h U^{p-1})_x))\,dx\,dt \\
&+ \overline{\delta} \int_{S_n} \frac{|U_t + UU_x|}{|\nabla U| + \gamma_n}(1 + |U|)\nabla U \cdot \nabla \pi_h U^{p-1}\,dx\,dt \\
&+ \overline{\overline{\delta}} \int_{S_n} |\tilde{U}| U_x (\pi_h(U^{p-1}))_x\,dx\,dt \equiv \sum_{i=1}^{8} E_n^i\,.
\end{aligned}
$$

Using now a standard interpolation estimate, we have

$$
|E_n^3| + |E_n^6| \leq Cp(h + \delta) \sum_{K \in T_h^n} \int_K \frac{|U_t + UU_x|}{|\nabla U|}(1 + |U|)|\nabla U|^2 \|U\|_{L_\infty(K)}^{p-2}\,dx\,dt\,.
$$

Further, using again an interpolation estimate, we get

$$
\begin{aligned}
|E_n^4| &\leq Cp^2h^2 \sum_{K \in T_h^n} \int_{R_n \cap K} |U_+^n - U_-^n|(U_+^n)_x^2 \|U^n\|_{L_\infty(K)}^{p-3}\,dx \\
&\leq Cp^2h^2 \sum_{K \in T_h^n} \int_{R_n \cap \{|U|>1\} \cap K} |U_+^n - U_-^n|(U_+^n)_x^2 \|U^n\|_{L_\infty(K)}^{p-2}\,dx \\
&\quad + Cp^2h^2 \int_{R_n \cap K \cap \{|U| \leq 1\}} |U_+^n - U_-^n|(U_+^n)_x^2\,dx = \mathrm{III}_n + \mathrm{IV}_n\,.
\end{aligned}
$$

By Lemma 4.1 with $p = 2$, and (4.4), we have

$$
\sum_{n \geq 0} \mathrm{IV}_n \leq Cp^2 h \int_\Omega |\tilde{U}| U_x^2\,dx\,dt \leq Cp^2 \frac{h}{\overline{\overline{\delta}}}\,.
$$

Combining these estimates with Lemmas 4.1 and 4.2, we get by summation over $n = 0, 1, 2, \ldots, N$, for $p^3 \leq C \min(\overline{\delta}/h, \overline{\overline{\delta}}/h)$,

$$
\begin{aligned}
&\sum_{n=0}^{N} \int_R ((U_-^{n+1})^p - (U_+^n)^p - (U_-^n - U_+^n)p(U_+^n)^{p-1})\,dx \\
&\qquad + \delta p(p-1) \int_{\Omega_\Lambda} (U_t + UU_x)^2 U^{p-2}\,dx\,dt \leq Cp^3 \frac{h}{\overline{\overline{\delta}}}.
\end{aligned}
$$

Using now the convexity of the function $U \to U^p$ as in (3.3), we have

$$\|U_-^{N+1}\|_{L_p(R)}^p + \delta p(p-1)\int_{\Omega_N}(U_t + UU_x)^2 U^{p-2}\,dx\,dt$$
$$\leq \|u_0\|_{L_p(R)}^p + Cp^3\frac{h}{\delta}, \qquad N \geq 0 .$$

The next step is to obtain L_p-estimates for all $t \in (0, \infty)$. For $t_n \leq t < t_{n+1}$ we have

$$\|U(\cdot, t)\|_{L_p(R)}^p = \|U_-^{n+1}\|_{L_p(R)}^p - \int_t^{t_{n+1}} \frac{\partial}{\partial s}\|U(\cdot, s)\|_{L_p(R)}^p\,ds$$
$$\leq \|U_-^{n+1}\|_{L_p(R)}^p + p\left(\int_{S_n}(U_t + UU_x)^2 U^{p-2}\,dx\,dt \int_t^{t_{n+1}}\int_R U^p\,dx\,dt\right)^{1/2}$$
$$\leq \|U_-^{n+1}\|_{L_p(R)}^p + \delta p(p-1)\int_{S_n}(U_t + UU_x)^2 U^{p-2}\,dx\,dt$$
$$+ \frac{p}{4\delta(p-1)}\int_t^{t_{n+1}}\|U(\cdot, s)\|_{L_p(R)}^p\,ds .$$

Thus, by using Gronwall's lemma, we obtain for $t_N \leq t \leq t_{N+1}$

$$\|U(\cdot, t)\|_{L_p(R)}^p \leq C\|u_0\|_{L_p(R)}^p + Cp^3\frac{h}{\delta} . \tag{4.6}$$

This proves the existence of positive constants C and α_0, independent of p and h, such that

$$\sup_{t\geq 0}\|U(\cdot, t)\|_{L_p(R)} \leq C \quad \text{if } 4 \leq p \leq Ch^{-\alpha_0} .$$

Finally, using an inverse estimate, we have

$$\|U_h\|_{L_\infty(\Omega)} \leq C(ph^{-1})^{1/p}\sup_{t\geq 0}\|U(\cdot, t)\|_{L_p(R)}$$
$$\leq Ce^{C(1+\alpha_0)h^{\alpha_0}\ln 1/h} \leq C$$

for $h \leq C$. It remains to estimate $\|U\|_{L_\infty(\Omega)}$ for $h > C$. By combining (4.4), (4.6) (with $p = 2$) and an inverse estimate, we get

$$\|U\|_{L_\infty(\Omega)} \leq Ch^{-1/2}\|u_0\|_{L_2(R)} \leq C, \qquad h > C . \quad \square$$

Remark 4.1. We note that the proof of Theorem 4.1 is based on choosing the test functions $v = \pi_h(U^{p-1})$ with p large and controlling the difference $U^{p-1} - \pi_h(U^{p-1})$ using the shock-capturing terms. Thus the shock-capturing terms make it possible to use test functions other than the usual choice $v = U$, giving the stability estimate (4.4); cf. the discussion in §2. Note, however, that the coerciveness of the shock-capturing terms with $v = \pi_h(U^{p-1})$ is established directly in Lemmas 4.1 and 4.2, using the fact that $k = 1$. In the case $k > 1$, the shock-capturing terms are still defined using piecewise linears on finer triangulations, which makes it possible to extend the L_∞ bound and convergence proof to methods of arbitrary accuracy, see [17, 18]. $\square$

5. Numerical results

In this section we give some numerical results obtained by applying the shock-capturing SD method (2.7) with $k = 1$ in the case of the time-dependent compressible Euler equations for polytropic gas with adiabatic exponent $\gamma = 1.4$ in a two-dimensional channel flow with a step-up at Mach 3. Our entropy variable formulation (2.5) is based on the physical entropy for the compressible Euler equations given in [4, 6]. A more detailed account of various aspects of the implementation will be presented in [3].

In (2.7) no boundary conditions are needed because the initial function u_0 has compact support and the spatial domain is the whole of $\mathbb{R}^d$. In the present case, the computational domain Ω (see Figure 5.1) is bounded, and then (2.7) was modified as follows so as to include relevant boundary conditions. On the inlet AB (see Figure 5.1) all components of U were prescribed with the free stream values given by

$$\begin{aligned} \text{pressure} &= 1.0, \qquad \text{density} = 1.4, \\ \text{horizontal velocity} &= 3.0, \qquad \text{vertical velocity} = 0.0, \end{aligned}$$

and along the solid walls $BCDE$ and AF the normal velocity was set to zero. Accordingly, the variational formulation (2.7) was then modified in the usual way by restricting the components of the test functions v to be zero where corresponding components of U are prescribed. In particular, at the outlet boundary EF, both U and the test functions v were varying freely. Thus the boundary conditions can easily be handled in the present method (which is one of the advantages of variational techniques).

The initial function u_0 was set equal to the free stream value, in particular with normal velocity different from zero on the step-up CD. The evolution was then abruptly started by forcing the normal velocity to be zero on CD for $t > 0$.

In the implementation, an automatic adaptive procedure was used to construct the finite element mesh T_h on each slab S_n. The mesh T_h was of the form $T_h = \{\tau \times I_n\}$ with $Z_n = \{\tau\}$ a triangulation of the underlying spatial domain Ω. The mesh Z_n was constructed from Z_{n-1} by local mesh refinement, or coarsening, according to the size of the second partial derivatives of the exact solution estimated through the computed solution U_-^n (for details see [3] and [1]). The time step $t_{n+1} - t_n$, i.e., the thickness of the slab, was chosen to be of the order of the smallest diameter of the triangles in Z_n.

The method (2.7) gives a nonlinear system of equations to be solved on each time step. This system was solved iteratively using Gaussian elimination or relaxation methods on linearized forms of the equations (2.7), with $U_+^n - U_-^n$ replaced by $A_0(\overline{U})(\overline{U}_+^n - \overline{U}_-^n)$, and with the 'frozen' coefficients $\overline{A}_i(\overline{U})$ successively updated using the last available approximation of $\overline{U}$.

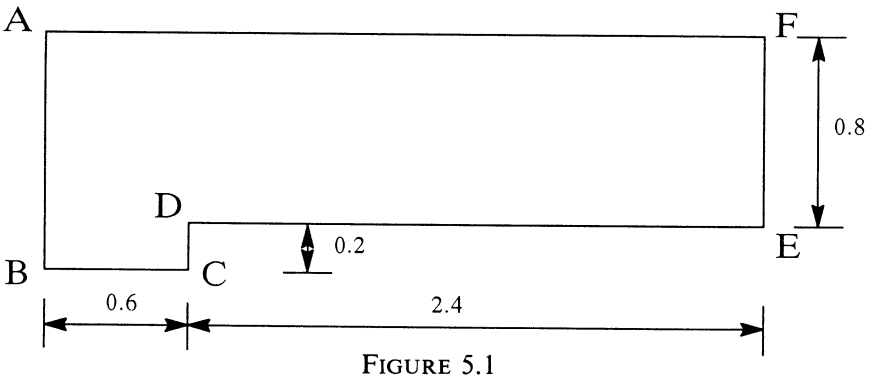


FIGURE 5.1
Computational domain $ABCDEFA$.

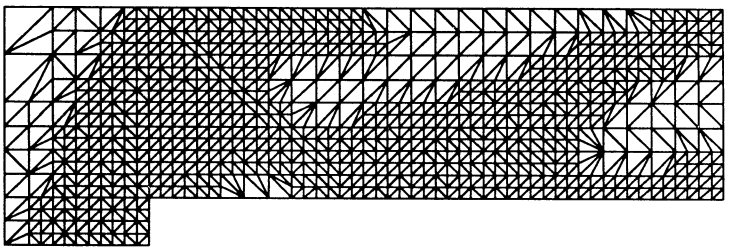


FIGURE 5.2a
Mesh at time 1.95.

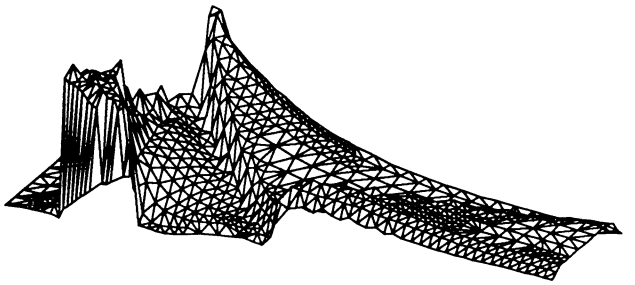


FIGURE 5.2b
Density at time 1.95 with mesh as in Figure 5.2a.

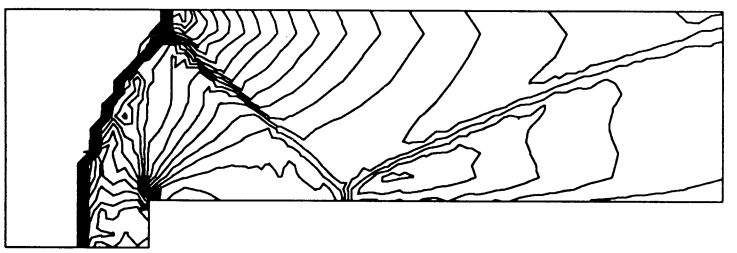


FIGURE 5.2c
Isodensity lines at time 1.95 with mesh as in Figure 5.2a.

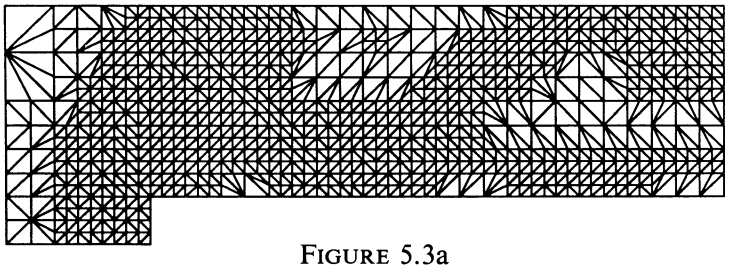


FIGURE 5.3a
Mesh at time 4.

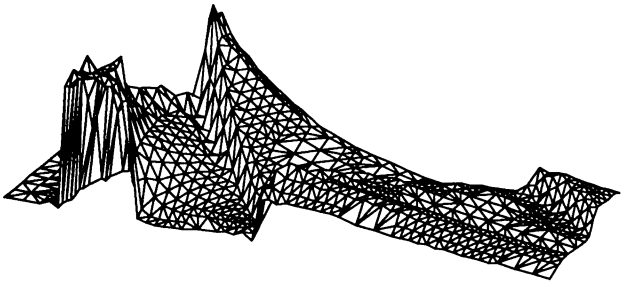


FIGURE 5.3b
Density at time 4 with mesh as in Figure 5.3a.

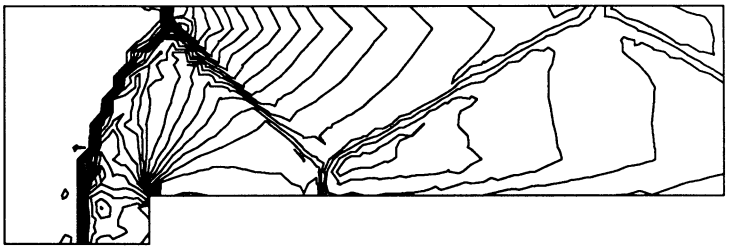


FIGURE 5.3c
Isodensity lines at time 4 with mesh as in Figure 5.3a.

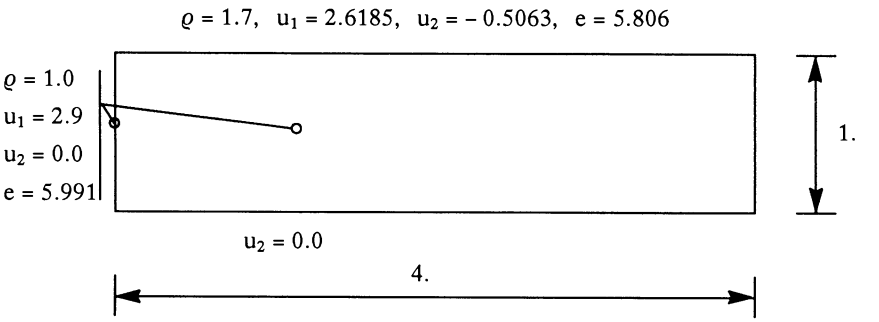


FIGURE 5.4
The computational domain and boundary conditions.

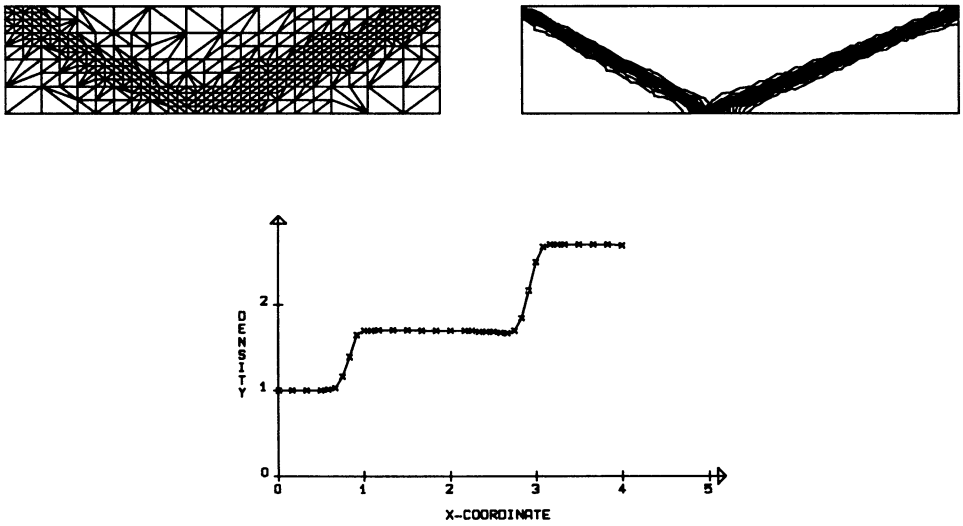


FIGURE 5.5
Mesh, contour lines for density and density profile at $x_2 = 0.5$, with shock-capturing.

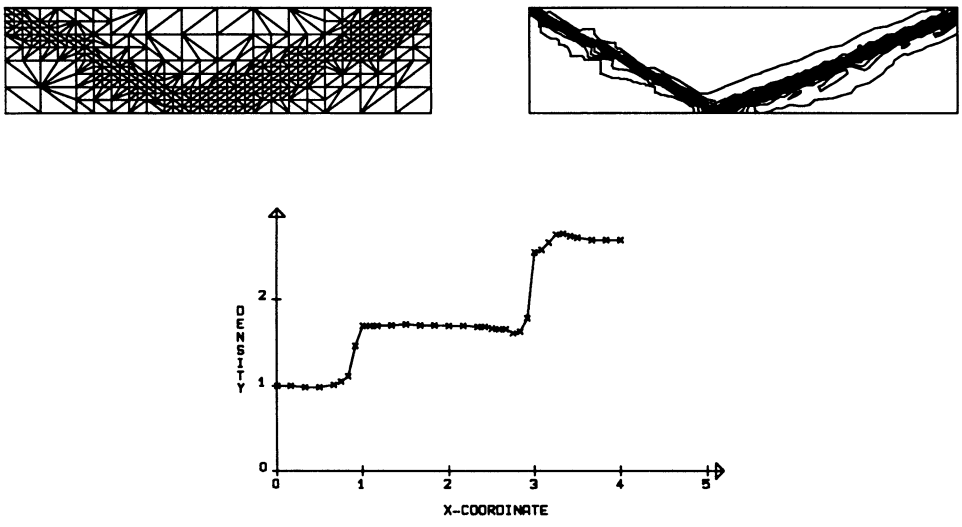


FIGURE 5.6
Mesh, contour lines for density and density profile at $x_2 = 0.5$, without shock-capturing.

In Figures 5.2–5.3 we give the computed velocity, density and mesh at time 1.95 and 4.0 obtained using the following parameter values: δ according to (2.11), $\bar{\delta} = h/20$, $\bar{\bar{\delta}} = 0$, where h is the local element size.

We also present in Figure 5.5 the results obtained with the same SD method for the Euler equations applied to a stationary shock reflection problem (cf. [15]) with geometry and boundary conditions according to Figure 5.4. The numerical solution is obtained after 150 time steps with $k = h_{\min}/2$. We further give in Figure 5.6 the corresponding results with no shock-capturing, i.e., with $\overline{\delta} = \overline{\overline{\delta}} = 0$. We note that in both cases the shock is captured within a couple of elements and that the slight over- and under-shoots in Figure 5.6 are eliminated when shock-capturing is added.

Appendix

Proof of (3.5). It follows by the definition of $\tilde{\varphi}$ that for sufficiently small h,

$$\|\tilde{\varphi}\|_{1,\Omega} \le \|\varphi\|_{1,\Omega}, \tag{A1}$$

$$\|\tilde{\varphi}\|_{2,\Omega} \le Ch^{-1}\|\varphi\|_{1,\Omega}, \tag{A2}$$

$$\|\tilde{\varphi} - \varphi\|_{\Omega} \le \sqrt{2}h\|\varphi\|_{1,\Omega}. \tag{A3}$$

Thus,

$$\|v\varphi - \pi_h(v\tilde{\varphi})\|_{\Omega} \le \|v(\varphi - \tilde{\varphi})\|_{\Omega} + \|v\tilde{\varphi} - \pi_h(v\tilde{\varphi})\|_{\Omega} \equiv \mathrm{I} + \mathrm{II},$$

where by (A3),

$$\mathrm{I} \le \|v\|_{L^\infty}\|\tilde{\varphi} - \varphi\|_{\Omega} \le \sqrt{2}h\|v\|_{L^\infty}\|\varphi\|_{1,\Omega},$$

and by Lemma 3.2, (A1) and (A2),

$$\mathrm{II} \le Ch\|v\|_{L^\infty}(\|\tilde{\varphi}\|_{1,\Omega} + h\|\tilde{\varphi}\|_{2,\Omega}) \le Ch\|v\|_{L^\infty}\|\varphi\|_{1,\Omega}.$$

This proves (3.5a) for $s = 0$. The case $s = 1$ follows similarly by noting that

$$\|v(\varphi - \tilde{\varphi})\|_{1,\Omega} \le \|v\|_{W^{1,\infty}}\|\varphi - \tilde{\varphi}\|_{\Omega} + \|v\|_{L^\infty}\|\varphi - \tilde{\varphi}\|_{1,\Omega} \le C\|v\|_{L^\infty}\|\varphi\|_{1,\Omega},$$

where in the last inequality we used the inverse estimate $\|v\|_{W^{1,\infty}} \le Ch^{-1}\|v\|_{L^\infty}$.

Finally, one can prove, see [18], that for $f \in H^1(S_n)$

$$\|f_+^n\|_R^2 \le 4(h^{-1}\|f\|_{S_n}^2 + h\|f_t\|_{S_n}^2),$$

and thus

$$\sum_{n=0}^{\infty} h\|(v\varphi)_+^n - (\pi_h(v\tilde{\varphi}))_+^n\|_R^2 \le 4(\|v\varphi - \pi_h(v\tilde{\varphi})\|_{\Omega}^2 + h^2\|v\varphi - \pi_h(v\tilde{\varphi})\|_{1,\Omega}^2),$$

which by (3.5a) proves (3.5b). □

Acknowledgment

This work was supported by the Swedish Board for Technical Development (STU).

BIBLIOGRAPHY

1. K. Eriksson and C. Johnson, *An adaptive finite element method for linear elliptic problems*, Math. Comp. **50** (1988), 361–383.
2. P. Hansbo, *Finite element procedures for conduction and convection problems*, Publication 86:7, Dept. of Structural Mechanics, Chalmers Univ. of Technology, S-412 96 Göteborg, 1986.
3. ——, *Streamline diffusion methods and adaptive procedures in finite element methods*, Thesis, Dept. of Structural Mechanics, Chalmers Univ. of Technology, 1989.
4. A. Harten, *On the symmetric form of systems of conservation laws with entropy*, Comput. Phys. **49** (1983), 151–164.
5. T. J. R. Hughes and A. Brook, *Streamline upwind-Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg. **32** (1982), 199-259.
6. T. J. R. Hughes, L. P. Franca, and M. Mallet, *A new finite element formulation for computational fluid dynamics*: I. *Symmetric forms of the compressible Euler and Navier-Stokes equations and the second law of thermodynamics*, Comput. Methods Appl. Mech. Engrg. **54** (1986), 223-234.
7. T. J. R. Hughes, M. Mallet, and A. Mizukami, *A new finite element formulation for computational fluid dynamics*: II. *Beyond SUPG*, Comput. Methods Appl. Mech. Engrg. **54** (1986), 341–355.
8. T. J. R. Hughes and M. Mallet, *A new finite element formulation for computational fluid dynamics*: III. *The general streamline operator for multidimensional advective-diffusive systems*, Comput. Methods Appl. Mech. Engrg. **58** (1986), 305–328.
9. ——, *A new finite element formulation for computational fluid dynamics*: IV. *A discontinuity-capturing operator for multidimensional advective-diffusive systems*, Comput. Methods Appl. Mech. Engrg. **58** (1986), 329–336.
10. C. Johnson, U. Nävert, and J. Pitkäranta, *Finite element methods for linear hyperbolic problems*, Comput. Methods Appl. Mech. Engrg. **45** (1984), 285–312.
11. C. Johnson and J. Saranen, *Streamline diffusion methods for the incompressible Euler and Navier-Stokes equations*, Math. Comp. **47** (1986), 1–18.
12. C. Johnson and A. Szepessy, *On the convergence of a finite element method for a nonlinear hyperbolic conservation law*, Math. Comp. **49** (1987), 427–444.
13. ——, *On the convergence of streamline diffusion finite element methods for hyperbolic conservation laws*, Numerical Methods for Compressible Flow—Finite Difference, Element and Volume Techniques (T. E. Tezduyar and T. J. H. Hughes, eds.), vol. 78, AMD. The American Society of Mechanical Engineers, 1986.
14. ——, *Shock-capturing streamline diffusion finite element methods for nonlinear conservation laws*, in Recent Developments in Computational Fluid Mechanics (T. E. Tezduyar and T. J. R. Hughes, eds.), vol. 95, AMD, The American Society of Mechanical Engineers, 1988.
15. R. Löhner, K. Morgan, and M. Vahdati, *FEM-FCT: Combining unstructured grids with high resolution*, Comm. Appl. Numer. Methods **4** (1988), 717–729.
16. A. Szepessy, *Convergence of a shock-capturing streamline diffusion finite element method for a scalar conservation law in two space dimensions*, Math. Comp. **53** (1989), 527–545.
17. ——, *Measure valued solutions to scalar conservation laws with boundary conditions*, Arch. Rational Mech. Anal. (to appear).
18. ——, *Convergence of the streamline diffusion finite element method for conservation laws*, Thesis, Dept. of Mathematics, Chalmers Univ. of Technology, S-412 96 Göteborg, 1989.
19. E. Tadmor, *Skew-selfadjoint forms for systems of conservation laws*, J. Math. Anal. Appl. **103** (1984), 428–442.

DEPARTMENT OF MATHEMATICS, CHALMERS UNIVERSITY OF TECHNOLOGY AND THE UNIVERSITY OF GÖTEBORG, S-412 96 GÖTEBORG, SWEDEN