The construction of discretely conservative finite volume schemes that also globally conserve energy or entropy

Antony Jameson

Stanford University
Aerospace Computing Laboratory
Report ACL 2007–1

January 2007

Abstract

This work revisits an idea that dates back to the early days of scientific computing, the energy method for stability analysis. It is shown that if the scalar nonlinear conservation law

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0$$

is approximated by the semi-discrete conservative scheme

$$\frac{du_j}{dt} + \frac{1}{\Delta x}\left(f_{j+\frac{1}{2}} - f_{j-\frac{1}{2}}\right) = 0$$

then the energy of the discrete solution evolves at exactly the same rate as the energy of the true solution, provided that the numerical flux is evaluated by the formula

$$f_{j+\frac{1}{2}} = \int_0^1 f(\hat{u}) d\theta$$

where

$$\hat{u}(\theta) = u_j + \theta(u_{j+1} - u_j).$$

With careful treatment of the boundary conditions, this provides a path to the construction of non-dissipative stable discretizations of the governing equations. If shock waves appear in the solution, the discretization must be augmented by appropriate shock operators to account for the dissipation of energy by the shock waves. These results are extended to systems of conservation laws, including the equations of incompressible flow, and gas dynamics. In the case of viscous flow, it is also shown that shock waves can be fully resolved by non-dissipative discretizations of this type with a fine enough mesh, such that the cell Reynolds number ≤ 2.

1 Introduction

Throughout the history of the development of discrete methods for linear and nonlinear conservation laws, and in particular computational fluid dynamics, there has been an ongoing struggle to find schemes which both assure physically correct non-oscillatory discrete solutions, and also minimize the discretization errors. Standard shock capturing schemes are formulated to satisfy total variation diminishing (TVD) or local extremum diminishing (LED) properties through the addition of sufficient artificial diffusion. On the other hand it seems that if the conservation law has an accompanying energy estimate, it should be possible to construct discrete schemes which satisfy the same estimate, and must therefore be stable without the need to introduce artificial diffusion, at least as long as the solution remains smooth.

The use of energy estimates to establish the stability of discrete approximations to initial value problems has a long history. The energy method is discussed in the classical book by Morton and Richtmyer [1], and it has been emphasized by the Uppsala school under the leadership of Kreiss and Gustafsson. Consider a well posed intitial value problem of the form

$$\frac{du}{dt} = Lu \tag{1.1}$$

where u is a state vector, and L is a linear differential operator in space with approximate boundary conditions. Then forming the inner product with u,

$$\left(u, \frac{du}{dt}\right) = \frac{1}{2}\frac{d}{dt}(u,u) = (u, Lu) \tag{1.2}$$

If L is skew self-adjoint, $L^* = -L$, and the right hand side is

$$\frac{1}{2}(u, Lu) + \frac{1}{2}(u, L^*u) = 0$$

Then the energy $\frac{1}{2}(u,u)$ cannot increase.

If (1.1) is approximated in semi-discrete form on a mesh as

$$\frac{dv}{dt} = Av \tag{1.3}$$

where v is the vector of the solution values of the mesh points, the corresponding energy

balance is

$$\frac{1}{2}\frac{d}{dt}(v^T v) = v^T A v \tag{1.4}$$

and stability is established if

$$v^T A v \leq 0 \tag{1.5}$$

A powerful approach to the formulation of discretizations with this property is to construct A in a manner that allows summation by parts (SBP) of $v^T Av$, annihilating all interior contributions, and leaving only boundary terms. Then one seeks boundary operators such that (1.5) holds. In particular suppose that A is split as

$$A = D + B$$

where D is an interior operator and B is a boundary operator. Then if D is skew-symmetric, $D^T = -D$, the contribution $v^T Dv$ vanishes leaving only the boundary terms.

Skew-symmetric and SBP operators of both second and higher order have been devised for a variety of problems. The benefits of kinetic energy preserving schemes for the treatment of incompressible viscous flow has also been emphasized by Moin and Stanford's Center for Turbulence Research. Honein and Moin have also examined skew-symmetric schemes for compressible flow [2]

SBP operators are typically constructed by splitting the equations into a part in conservation form and a part in quasilinear form. For example, the inviscid Burgers equation is written as

$$\frac{\partial u}{\partial t} + \frac{2}{3}\frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) + \frac{1}{3}u\frac{\partial u}{\partial x} = 0$$

Then the use of second order central difference operators for both parts at every interior point, and one sided operators at each boundary yields an SBP operator.

In nonlinear problems for which the solution may develop shock waves it is generally beneficial to preserve conservation form in the discretization. According to the theorem of Lax and Wendroff [3], this will assure that the discrete solution satisfies the correct shock jump conditions, provided that it converges in the limit as the mesh interval is reduced to zero. In any case it is highly desirable to maintain global conservation properties of the true solution in the discrete solution. Small errors in the global conservation of mass, for example, can lead to large errors in the solution of flows in ducts.

This paper presents a general procedure for constructing semi-discrete schemes for con-

servation laws in conservation form such that the discrete solution also exactly satisfies a discrete global conservation law for a generalized energy or entropy function. To take advantage of this approach it is first necessary to identify an appropriate energy principle for the governing equations. If shock waves appear in the solution the energy principle will need to be modified to allow for their effect on the energy or entropy balance. Moreover, in the light of Godunov's theorem that monotonically varying discrete shocks can only be obtained by locally first-order accurate schemes, the basic discretization scheme will need to be augmented by appropriate shock operators.

The construction of shock operators for the inviscid Burgers equation and for the gas dynamics equations has been discussed by Gustafsson and Olsson [4]. Shock capturing schemes for gas dynamics have been widely studied [Godunov, Boris, Van Leer, Roe, Harten, Liou, Jameson [5, 6, 7, 8, 9, 10, 11, 12]. In general they add artificial diffusion either explicitly or implicitly through the use of upwind operators. The aim of the present work is to devise stable schemes which would require the introduction of artificial diffusion only on the neighborhood of discontinuities and nowhere else.

The next section discusses the Burgers equation and the energy principle. It is shown how to construct a semi-discrete scheme in conservation form, together with boundary operators, such that the discrete energy is bounded from above by the energy of the true solution. When shock waves appear in the solution the energy balance has to be modified to account for the dissipation of energy by the shock waves. It is then shown how to construct a shock operator which enables the discrete solution to track the energy evolution of the true solution very accurately, as verified by numerical experiments.

Section 3 presents a procedure for constructing semi-discrete approximations to general scalar conservation laws in conservation form such that a discret energy principle is satisfied exactly. The conditions for the construction of the numerical flux are given in Theorem 3.1. Section 4 extends the method to the treatment of systems of conservation laws. The conditions for the construction of the numerical flux such that a generalized entropy balance is satisfied exactly are given in Theorem 4.3. The method requires that the system can be symmetrized by a transformation of variables, as described in the work of Godunov, Mock and Harten [13, 14, 15].

Section 5 applies the foregoing method to the treatment of three dimensional incompressible flow. Theorem 5.2 states the discrete energy principle that is exactly satisfied by semi-discrete finite volume schemes with polyhedral control volumes, as has also been veri-

fied numerically. Finally the treatment of the gas dynamics equations is discussed in Section 6. It is shown that a semi-discrete finite volume scheme can be constructed to satisfy exactly a discrete entropy balance as long as the numerical flux is constructed in the proper manner, as stated in Theorem 6.1. The formula, however, for the numerical flux requires the introduction of entropy variables, and is expressed as an integral that does not have a closed form. It is suggested that it should be evaluated by Gauss Lobatto integration.

2 Burgers equation

The Burgers equation is the simplest example of a nonlinear equation which supports wave motion in opposite directions and the formation of shock awaves, and consequently it provides a very useful example for the analysis of the energy method. Expressed in conservation form, the inviscid Burgers equation is

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0, \qquad a \leq x \leq b, \tag{2.1}$$

where

$$f(u) = \frac{u^2}{2} \tag{2.2}$$

and the wave speed is

$$a(u) = \frac{\partial f}{\partial u} = u \tag{2.3}$$

Boundary conditions specifying the value of u at the left or right boundaries should be imposed if the direction of u is towards the interior at the boundary.

Smooth solutions of (2.1) remain constant along characteristics

$$x - ut = \xi$$

If a faster moving wave over-runs a slower moving wave this would indicate a multi-valued solution. Instead a proper weak solution is obtained by assuming the formation of a shock wave across which there is a discontinuous transition between left and right state u_L and u_R. In order to satisfy the conservation law (2.1) in integral form, u_L and u_R must satisfy the jump condition

$$f(u_R) - f(u_L) = s(u_R - u_L) \tag{2.4}$$

where s is the shock speed. For the Burgers equation this gives a shock speed

$$s = \frac{1}{2}(u_R + u_L) \tag{2.5}$$

Provided that the solution remains smooth, (2.1) can be multiplied by u^{k-1} and rearranged to give an infinite set of invariants of the form

$$\frac{\partial}{\partial t}\left(\frac{u^k}{k}\right) + \frac{\partial}{\partial x}\left(\frac{u^{k+1}}{k+1}\right) = 0$$

Here we focus on the first of these

$$\frac{\partial}{\partial t}\left(\frac{u^2}{2}\right)+\frac{\partial}{\partial x}\left(\frac{u^3}{3}\right)=0 \tag{2.6}$$

This may be integrated over x from a to b to determine the rate of change of the energy

$$E=\int_a^b \frac{u^2}{2}\,dx \tag{2.7}$$

in terms of the boundary fluxes as

$$\frac{dE}{dt}=\frac{u_a^3}{3}-\frac{u_b^3}{3} \tag{2.8}$$

This equation fails in the presence of shock waves, as can easily be seen by considering the initial data $u=-x$ in the interval $[-1,1]$. Then a wave moves inwards from each boundary at unit speed toward the center until a stationary shock wave is formed at $t=1$, after which the energy remains constant. Thus

$$E(t)=\begin{cases}\frac{1}{3}+\frac{2t}{3}, & 0\le t\le 1\\ 1, & t>1\end{cases}$$

In order to correct (2.8) in the presence of a shock wave with left and right states u_L and u_R, equation (2.6) should be integrated separately on each side of the shock. If the shock is moving at a speed s there is an additional contribution to $\frac{dE}{dt}$ in the amount

$$s\left(\frac{u_L^2}{2}-\frac{u_R^2}{2}\right)$$

Substituting equation (2.5) for the shock speed

$$\frac{dE}{dt}=\frac{u_a^3}{3}-\frac{u_L^3}{3}+\frac{u_R^3}{3}-\frac{u_b^3}{3}+\frac{1}{2}(u_L+u_R)\left(\frac{u_L^2}{2}-\frac{u_R^2}{2}\right)$$

which can be simplified to

$$\frac{dE}{dt}=\frac{u_a^3}{3}-\frac{u_b^3}{3}-\frac{1}{12}(u_L-u_R)^3 \tag{2.9}$$

In the presence of multiple shocks, each will remove energy at the rate $\frac{1}{12}(u_L-u_R)^3$.

As was already observed by Morton and Richtmyer [1, page 142], a skew-symmetric difference operator consistent with (2.1) for smooth data can be constructed by splitting it between conservation and quasilinear form as

$$\frac{\partial u}{\partial t} + \frac{2}{3}\frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) + \frac{1}{3}u\frac{\partial u}{\partial x} = 0$$

Suppose this is discretized on a uniform mesh $x_j = j\Delta x,\ j = 0, 1, \ldots n$. Central differencing of both spatial derivatives at interior points yields the semi-discrete scheme

$$\frac{du_j}{dt} = \frac{1}{6\Delta x}\left(u_{j+1}^2 - u_{j-1}^2\right) + \frac{1}{6\Delta x}u_j\left(u_{j+1} - u_{j-1}\right) = 0, \qquad j = 1, n-1 \tag{2.10}$$

The skew symmetric operator is completed by the use of one sided schemes at each boundary

$$\begin{aligned}\frac{du_0}{dt} &= \frac{1}{3\Delta x}(u_1^2 - u_0^2) + \frac{1}{3\Delta x}u_0(u_1 - u_0)\\ \frac{du_n}{dt} &= \frac{1}{3\Delta x}(u_n^2 - u_{n-1}^2) + \frac{1}{3\Delta x}u_n(u_n - u_{n-1})\end{aligned} \tag{2.11}$$

Rewriting the quasilinear term as $\frac{1}{6\Delta x}(u_{j+1}u_j - u_j u_{j-1})$ equations (2.10) and (2.11) can be expressed in the conservation form

$$\frac{du_j}{dt} + \frac{1}{\Delta x}\left(f_{j+\frac{1}{2}} - f_{j-\frac{1}{2}}\right) = 0, \qquad j = 1, n-1 \tag{2.12}$$

where

$$f_{j+\frac{1}{2}} = \frac{1}{6}\left(u_{j+1}^2 + u_{j+1}u_j + u_j^2\right) \tag{2.13}$$

and

$$\begin{aligned}\frac{du_0}{dt} + \frac{2}{\Delta x}\left(f_{\frac{1}{2}} - f_0\right) &= 0\\ \frac{du_n}{dt} + \frac{2}{\Delta x}\left(f_n - f_{n-\frac{1}{2}}\right) &= 0\end{aligned} \tag{2.14}$$

where

$$f_0 = \frac{u_0^2}{2}, \qquad f_n = \frac{u_n^2}{2} \tag{2.15}$$

Now let the discrete energy be represented by trapezoidal integration as

$$E = \frac{\Delta x}{2}\left(\frac{u_0^2}{2} + \frac{u_n^2}{2}\right) + \Delta x \sum_{j=1}^{n-1} \frac{u_j^2}{2} \tag{2.16}$$

Multiplying equation (2.12) by u_j and summing by parts

$$\Delta x \sum_{j=1}^{n-1} u_j \frac{du_j}{dt} = -\sum_{j=1}^{n-1} u_j (f_{j+\frac{1}{2}} - f_{j-\frac{1}{2}}) = f_{\frac{1}{2}} u_0 - f_{n+\frac{1}{2}} u_n$$

Hence, including the boundary points, we find that

$$\frac{dE}{dt} = \frac{u_0^3}{3} - \frac{u_n^3}{3} \tag{2.17}$$

which is the exact discrete analog of the continuous energy evolution equation (2.8).

The formulation so far does not include the boundary conditions. Suppose that boundary data $u = g_0$ should be imposed at x_0 if the left boundary x_0 is an inflow boundary, and correspondingly $u = g_n$ should be imposed at x_n if the right boundary x_n is an inflow boundary. It is convenient to introduce the positive and negative wave speeds

$$a^+(u) = \max(a(u), 0), \qquad a^-(u) = \min(a(u), 0)$$

Then we modify the equations at the boundary points by adding simultaneous approximation terms (SAT), so that instead of (2.14) we solve

$$\begin{aligned} \frac{du_0}{dt} + \frac{2}{\Delta x}(f_{\frac{1}{2}} - f_0) + \frac{\tau}{\Delta x} a_0^+ (u_0 - g_0) &= 0 \\ \frac{du_n}{dt} + \frac{2}{\Delta x}(f_n - f_{n-\frac{1}{2}}) - \frac{\tau}{\Delta x} a_0^- (u_n - g_n) &= 0 \end{aligned} \tag{2.18}$$

where the parameter τ determines the amount of the penalty if the boundary condition is not satisfied exactly. The linear case has been analyzed by Mattsson [16]. here we wish to ensure stability in the nonlinear case. It is evident from (2.17) that outflow boundaries ($u_0 < 0$ or $u_n > 0$) promote energy decay. Thus we need only consider the effect of inflow boundary conditions.

For this purpose suppose that $\frac{d}{dt} E_{\text{true}}$ is the rate of change of energy that would result if the boundary conditions were exactly satisfied. Then we wish to choose τ so that $\frac{dE}{dt}$ is

bounded from above by $\frac{d}{dt}E_{\text{true}}$

$$\frac{dE}{dt} \leq \frac{d}{dt}E_{\text{true}} \tag{2.19}$$

Consider the construction at the left boundary assuming it is an inflow boundary. Suppose that a_0^+ is evaluated as $\frac{1}{2}(u_0 + g_0)$. Omitting for the moment the contribution of the right boundary we find that

$$\frac{d}{dt}(E - E_{\text{true}}) = \frac{u_0^3}{3} - \frac{g_0^3}{3} - \frac{\tau}{4}u_0(u_0^2 - g_0^2)$$

Suppose now that u_0 has the value αg_0. Then the rate of change of the energy is modified by the cubic expression $F(\alpha)g_0^3$ where

$$F(\alpha) = \frac{\alpha^3}{3} - \frac{1}{3} - \frac{\tau}{4}\alpha(\alpha^2 - 1)$$

Here g_0 should be positive if it is truly an inflow boundary condition, so we require $F(\alpha)$ to be nonpositive in the range of α corresponding to inflow, $\alpha > -1$. However, $F(\alpha) = 0$ when $\alpha = 1$ and $u_0 = g_0$, so its sign will change at $\alpha = 1$ unless this is a double root. Here

$$F'(\alpha) = \alpha^2 - \frac{\tau}{4}(3\alpha^2 - 1)$$

and the condition $F'(1) = 0$ yields

$$\tau = 2 \tag{2.20}$$

Then

$$F(\alpha) = -\frac{1}{6}(\alpha - 1)^2(\alpha + 2)$$

and is non positive whenever $\alpha > -2$. A similar analysis at the right boundary confirms that condition (2.20) is sufficient to assure the favorable energy comparison (2.19) whenever either boundary is an inflow boundary.

Numerical experiments have been conducted to verify the stability of the semi-discrete scheme (2.12–2.13) with the boundary conditions (2.18–2.20). Shu's total variation diminishing (TVD) scheme [17], was used for time integration. Writing the semi-discrete scheme in the form

$$\frac{du}{dt} + R(u) = 0 \tag{2.21}$$

where $R(u)$ represents the discretized spatial derivative, this advances the solution during

one time step by the three stage scheme

$$\begin{aligned} u^{(1)} &= u^{(0)} - \Delta t\, R(u^{(0)}) \\ u^{(2)} &= \frac{3}{4}u^{(0)} + \frac{1}{4}u^{(1)} - \frac{1}{4}\Delta t\, R(u^{(1)}) \\ u^{(3)} &= \frac{1}{3}u^{(0)} + \frac{2}{3}u^{(2)} - \frac{2}{3}\Delta t\, R(u^{(2)}) \end{aligned}$$

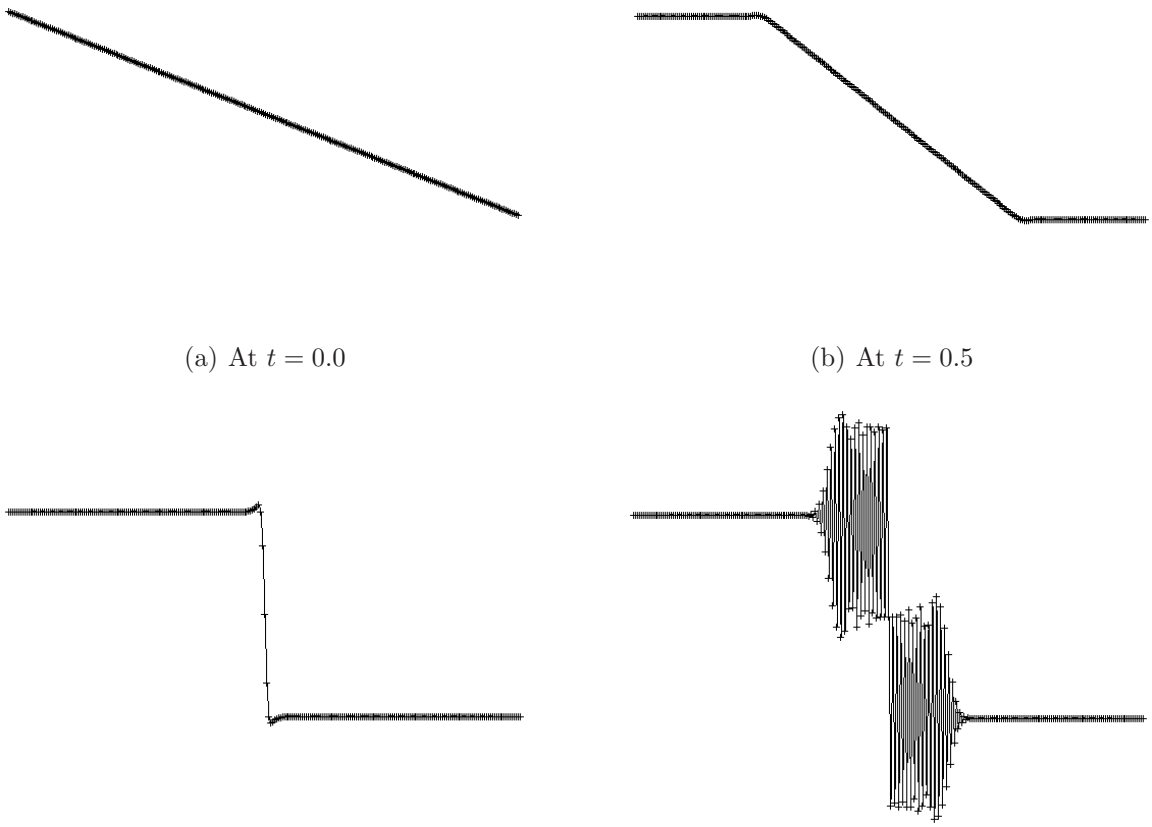
where $u^{(0)}$ and $u^{(3)}$ denote the solution of the beginning and the end of the time step. If (2.21) satisfies the TVD property with forward Euler time stepping, Shu's scheme is also TVD for time steps satisfying the CFL condition $|a|\frac{\Delta t}{\Delta x} \leq 1$. This property has been designated strongly stability preserving (SSP) by Gottlieb [18].

Figure 2.1 displays snap shots of the solution with initial data $u = -x$ in $[-1, 1]$ at times $t = 0, .5, 1, 1.5$ using a grid with 256 intervals. The true solution is a straight line connecting wave fronts moving inwards from both boundaries at unit speed until a shock is formed at $t = 1$. It can be seen that the discrete solution closely tracks the true solution prior to the formation of the shock. After the shock is formed the discrete solution develops strong oscillations in a zone expanding outward from the shock towards both boundaries. This is consistent with the fact that according to equation (2.17), the discrete energy continues to grow at the rate $\frac{2}{3}t$ when $t > 1$ as illustrated in Figure 2.2, and the energy must be absorbed in the solution.

It is evident that the scheme must be modified to preserve stability in the presence of shock waves. It is well known from shock capturing theory [11, 12], that oscillations in the neighborhood of shock waves are eleminated by schemes which are local extremum diminishing (LED) or total variation diminishing (TVD). A semi-discrete scheme is LED if it can be expressed in the form

$$\frac{du_i}{dt} = \sum_j a_{ij}(u_j - u_i) \tag{2.22}$$

where the coefficients $a_{ij} \geq 0$, and the stencil is compact, $a_{ij} \neq 0$ when i and j are not nearest neighbors. This property is satisfied by the upwind scheme in which the numerical



(a) At $t = 0.0$

(b) At $t = 0.5$

(c) At $t = 1.0$

(d) At $t = 1.5$

Figure 2.1: Evolution of the solution of the Burgers equation

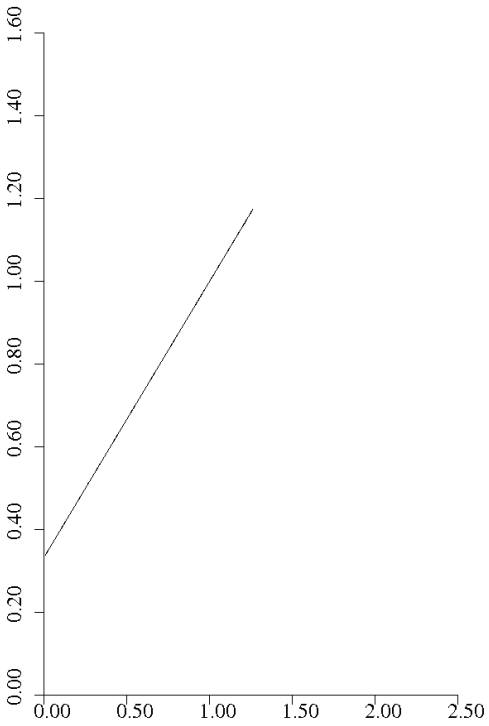


Figure 2.2: Discrete energy growth

flux (2.13) is replaced by

$$f_{j+\frac{1}{2}} = \begin{cases} u_j^2 & \text{if} \quad a_{j+\frac{1}{2}} > 0 \\ u_{j+1}^2 & \text{if} \quad a_{j+\frac{1}{2}} < 0 \\ \frac{1}{2}(u_{j+1}^2 + u_j^2) & \text{if} \quad a_{j+\frac{1}{2}} = 0 \end{cases} \tag{2.23}$$

where the numerical wave speed is evaluated as

$$a_{j+\frac{1}{2}} = \frac{1}{2}(u_{j+1} + u_j) \tag{2.24}$$

Moreover, the upwind scheme (2.23) admits a stationary numerical shock structure with a single interior point.

The LED condition only needs to be satisfied in the neighborhoods of local extrema, which may be detected by a change of sign in the first differences $\Delta u_{j+\frac{1}{2}} = u_{j+1} - u_j$. A shock operator which meets these requirements can be constructed as follows. The numerical flux (2.13) can be converted to the upwind flux (2.23) by the addition of a diffusive term of

the form

$$d_{j+\frac{1}{2}} = \alpha_{j+\frac{1}{2}} \Delta u_{j+\frac{1}{2}}.$$

The required coefficient is

$$\alpha_{j+\frac{1}{2}} = \frac{1}{4} |u_{j+1} + u_j| - \frac{1}{12} (u_{j+1} - u_j) \tag{2.25}$$

In order to detect an extremum introduce the function

$$R(u, v) = \left| \frac{u - v}{|u| + |v|} \right|^q$$

where q is an integer power. $R(u, v) = 1$ whenever u and v have opposite signs. When $u = v = 0$, $R(u, v)$ should be assigned the value zero. Now set

$$s_{j+\frac{1}{2}} = R\left(\Delta u_{j+\frac{3}{2}}, \Delta u_{j-\frac{1}{2}}\right) \tag{2.26}$$

so that $s_{j+\frac{1}{2}} = 1$ when $\Delta u_{j+\frac{3}{2}}$ and $\Delta u_{j-\frac{1}{2}}$ have opposite signs which will generally be the case if either u_{j+1} or u_j is an extremum. In a smooth region where $\Delta u_{j+\frac{3}{2}}$ and $\Delta u_{j-\frac{1}{2}}$ are not both zero, $s_{j+\frac{1}{2}}$ is of the order Δx^q, since $\Delta u_{j+\frac{3}{2}} - \Delta u_{j-\frac{1}{2}}$ is an undivided difference. In order to avoid activating the switch at smooth extrema, and also to protect against division by zero, $R(u, v)$ may be redefined as

$$R(u, v) = \left| \frac{u - v}{\max\{(|u| + |v|), \epsilon\}} \right| \tag{2.27}$$

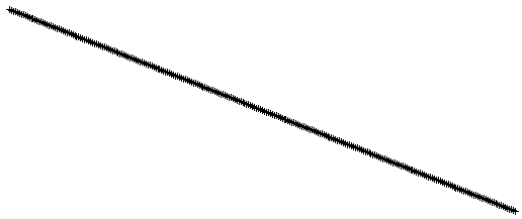
where ϵ is a tolerance [11].

Finally the diffusion term is modified to

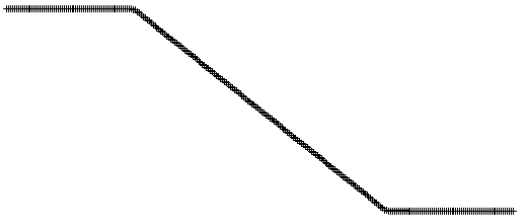
$$d_{j+\frac{1}{2}} = \max(s_{j+\frac{3}{2}}, s_{j+\frac{1}{2}}, s_{j-\frac{1}{2}}) \alpha_{j+\frac{1}{2}} \Delta u_{j+\frac{1}{2}}$$

Thus the coeffcient is reduced to a magnitude of order Δx^q in smooth regions, while it has the value $\alpha_{j+\frac{1}{2}}$ in the neighborhood of a shock. The value $q = 8$ has proved satisfactory in numerical experiments.

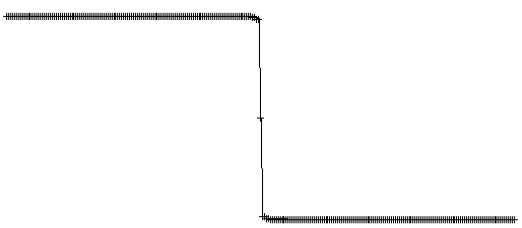
Figure 2.3 shows the evolution of the discrete solution for the same case as Figure 2.1, with initial data $u = -x$ in $[-1, 1]$, using the shock operator defined by equations (2.25-2.27). A stationary shock with a single interior point is formed when $t = 1$ as expected. Figure 2.4

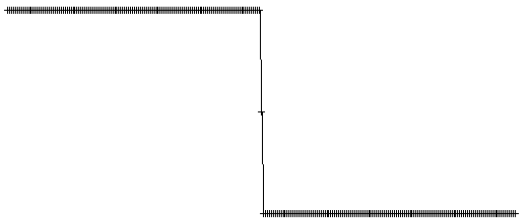


(a) At $t = 0.0$



(b) At $t = 0.5$





(c) At $t = 1.0$

Figure 2.3: Evolution of the solution of the Burgers equation with a switch

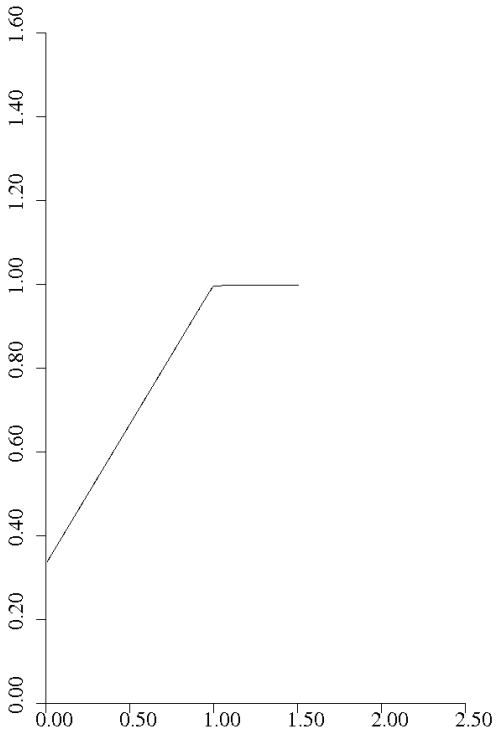


Figure 2.4: Discrete energy growth with a limiter

confirms that the discrete energy grows at the rate $\frac{2t}{3}$ until the shock forms and then remains constant. The difference between the discrete energy and the true energy of the stationary solution is $-\frac{1}{2}\Delta x$ because of the zero value in the middle of the discrete shock. Once the numerical shock structure is established the additional diffusive terms only contribute to $\frac{du_j}{dt}$ at the three points $s-1$, s and $s+1$ comprising the shock, for which

$$u_{s-1} = 1, \qquad u_s = 0, \qquad u_{s+1} = -1$$

The only non-zero values of $\Delta u_{j+\frac{1}{2}}$ are

$$\Delta u_{s-\frac{1}{2}} = -1, \qquad \Delta u_{s+\frac{1}{2}} = -1$$

Also

$$\begin{aligned} a_{s-\frac{1}{2}} &= -\frac{1}{2}, & \alpha_{s-\frac{1}{2}} &= \frac{1}{3} \\ a_{s+\frac{1}{2}} &= -\frac{1}{2}, & \alpha_{s+\frac{1}{2}} &= \frac{1}{3} \end{aligned}$$

Thus the additional contribution to $\frac{dE}{dt}$ due to the shock is

$$\begin{aligned}\sum_{s=-1}^{1} u_s \left(\alpha_{s+\frac{1}{2}}\Delta u_{s+\frac{1}{2}} - \alpha_{s-\frac{1}{2}}\Delta u_{s-\frac{1}{2}}\right) &= \alpha_{s-\frac{1}{2}}u_{s-1}\Delta u_{s-\frac{1}{2}} - \alpha_{s+\frac{1}{2}}u_{s+1}\Delta u_{s+\frac{1}{2}} \\ &= -\frac{2}{3}\end{aligned}$$

This exactly cancels the contribution of $\frac{2}{3}$ from the boundaries, so that the total rate of change of the discrete energy is zero.

In the case of the viscous Burgers equation with the viscosity coefficient ν

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) = \nu\frac{\partial^2 u}{\partial x} \tag{2.28}$$

the energy balance is modified by the viscous dissipation. Multiplying by u, and integrating the right hand side by parts with $\frac{\partial u}{\partial x} = 0$ at each boundary, the energy balance equation assumes the form

$$\frac{dE}{dt} = \frac{u_a^3}{3} - \frac{u_b^3}{3} - \nu\int_a^b \left(\frac{\partial u}{\partial x}\right)^2 dx \tag{2.29}$$

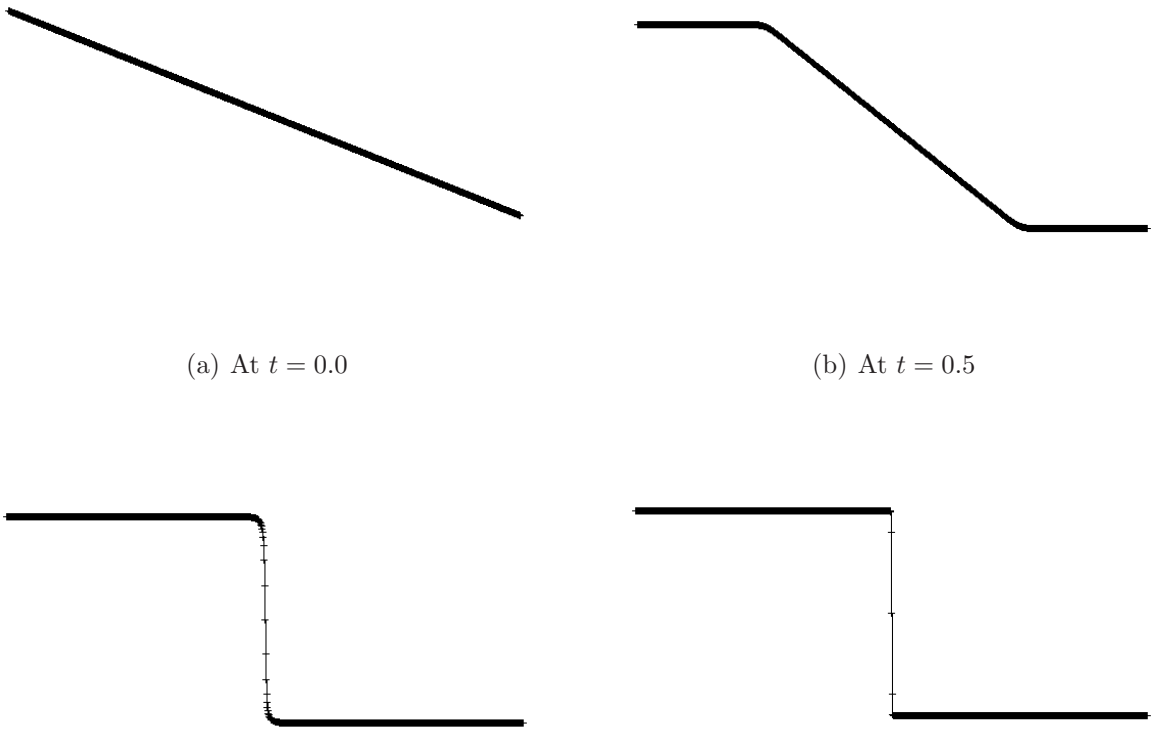
instead of (2.8). Suppose that $\frac{\partial^2 u}{\partial x^2}$ is discretized by a central difference operator at interior points with one sided formulas at the boundaries corresponding to $\frac{\partial u}{\partial x} = 0$,

$$\begin{aligned}&\frac{1}{\Delta x^2}(u_{j+1} - 2u_j + u_{j-1}), \qquad j = 2, n-1 \\ &\frac{1}{\Delta x^2}(u_1 - u_0) \qquad \text{at the left boundary,} \\ &\frac{1}{\Delta x^2}(u_n - u_{n-1}) \qquad \text{at the right boundary}\end{aligned} \tag{2.30}$$

as proposed by Mattsson [16]. Then summing by parts with the convective flux evaluated by (2.13) as before, the discrete energy balance is found to be

$$\frac{dE}{dt} = \frac{u_0^3}{3} - \frac{u_n^3}{3} - \nu\sum_{j=0}^{n-1}(u_{j+1} - u_j)^2 \tag{2.31}$$

This enables the possibility of fully resolving shock waves without the need to add any additional numerical diffusion via shock operators. The convective flux difference $f_{j+\frac{1}{2}} - f_{j-\frac{1}{2}}$



(a) At $t = 0.0$

(b) At $t = 0.5$

(c) At $t = 1.0$

(d) At $t = 1.5$

Figure 2.5: Evolution of the solution of the viscous Burgers equation

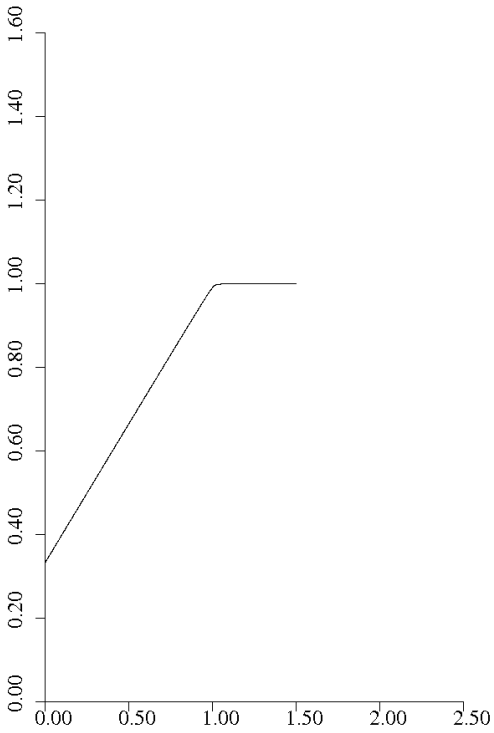


Figure 2.6: Discrete energy growth for the viscous Burgers equation

can be factored as

$$\frac{1}{3\Delta x}(u_{j+1} + u_j + u_{j-1})(u_{j+1} - u_{j-1})$$

Accordingly the semi-discrete approximation to equation (2.28) can written as

$$\frac{du_j}{dt} = a_{j+\frac{1}{2}}(u_{j+1} - u_j) + a_{j-\frac{1}{2}}(u_{j-1} - u_j) \tag{2.32}$$

where

$$a_{j+\frac{1}{2}} = \frac{\nu}{\Delta x^2} - \frac{u_{j+1} + u_j + u_{j-1}}{3\Delta x}$$

and

$$a_{j-\frac{1}{2}} = \frac{\nu}{\Delta x^2} + \frac{u_{j+1} + u_j + u_{j-1}}{3\Delta x}$$

The semi-discrete approximation satisfies condition (2.22) for a local extremum diminishing scheme if $a_{j+\frac{1}{2}} \geq 0$ and $a_{j-\frac{1}{2}} \geq 0$. This establishes theorem 2.1:

Theorem 2.1 *The semi-discrete approximation* (2.12) *using the numerical flux* (2.13) *and the central difference operator* (2.30) *for* $\frac{\partial^2 u}{\partial x^2}$ *is local extremum diminishing if the cell Reynolds*

number satisfies the condition

$$\frac{\bar{u}\Delta x}{\nu} \leq 2 \tag{2.33}$$

where the local speed is evaluated as

$$\bar{u} = \frac{1}{3}\left|u_{j+1} + u_j + u_{j-1}\right| \tag{2.34}$$

It has been confirmed by numerical experiments that shock waves are indeed fully resolved with no oscillation if the cell Reynolds number satisfies condition (2.33). Figure 2.5 shows the evolution of the discrete viscous Burgers equation using this scheme for the same initial data as before, $u = -x$ in $[-1, 1]$. The Reynolds number $\frac{uL}{\nu}$ based on the boundary velocity $u = \pm 1$ and the length of the interval was 2048, and the solution was calculated on a uniform mesh with 1024 intervals, so that the cell Reynolds number condition (2.33) was satisfied in the entire domain. It can be seen that a stationary shock wave is formed at the time $t = 1$, and it is finally resolved with three interior points. Correspondingly the energy becomes constant after the shock wave is formed, as can be seen in Figure 2.6.

3 The one dimensional scalar conservation law

Consider the scalar conservation law

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0 \tag{3.1}$$

$$u(x, 0) = u_0(x),$$

$$u \qquad \text{specified at inflow boundaries.}$$

Correspondingly, smooth solutions of (3.1) also satisfsy

$$\frac{\partial}{\partial t}\left(\frac{u^2}{2}\right) + \frac{\partial}{\partial x} F(u) = 0 \tag{3.2}$$

where

$$F_u = u f_u$$

since multiplying (3.1) by u yields

$$u\frac{\partial u}{\partial t} + u f_u \frac{\partial u}{\partial x} = 0$$

Defining the energy as

$$E = \int_a^b \frac{u^2}{2} dx$$

it follows from (3.2) that smooth solutions of (3.1) satisfy the energy equation

$$\frac{dE}{dt} = F(u_a) - F(u_b) \tag{3.3}$$

Introducing the function $G(u)$ such that

$$G_u = f$$

and multiplying (3.1) by u we obtain

$$\begin{aligned} u\frac{\partial u}{\partial t} + u\frac{\partial f}{\partial x} &= \frac{\partial}{\partial t}\left(\frac{u^2}{2}\right) + \frac{\partial}{\partial x}(uf) - f\frac{\partial u}{\partial x} \\ &= \frac{\partial}{\partial t}\left(\frac{u^2}{2}\right) + \frac{\partial}{\partial x}(uf) - G_u\frac{\partial u}{\partial x} \\ &= \frac{\partial}{\partial t}\left(\frac{u^2}{2}\right) + \frac{\partial}{\partial x}(uf - G) \\ &= 0 \end{aligned}$$

Thus F and G can be identified as

$$F = uf - G, \qquad G = uf - F \tag{3.4}$$

For the inviscid Burgers equation

$$F = \frac{u^3}{3}, \qquad G = \frac{u^3}{6}$$

Suppose now that (3.1) is discretized on a uniform grid over the range $j = 0, n$. Consider a semi-discrete conservative scheme of the form

$$\frac{du_j}{dt} + \frac{1}{\Delta x}(f_{j+\frac{1}{2}} - f_{j-\frac{1}{2}}) = 0 \tag{3.5}$$

at every interior point, where the numerical flux $f_{j+\frac{1}{2}}$ is a function of u_i over a range bracketing u_j such that $f_{j+\frac{1}{2}} = f(u)$ whenever u is substituted for the u_i, thus satisfying Lax's consistency condition. Multiplying (3.1) by u_j and summing by parts over the interior points we obtain

$$\begin{aligned} \Delta x \sum_{j=1}^{n-1} u_j \frac{du_j}{dt} &= -\sum_{j=1}^{n-1} u_j (f_{j+\frac{1}{2}} - f_{j-\frac{1}{2}}) \\ &= -u_1 f_{\frac{3}{2}} - u_2 f_{\frac{5}{2}} \ldots - u_{n-2} f_{n-\frac{3}{2}} - u_{n-1} f_{n-\frac{1}{2}} \\ &\quad + u_1 f_{\frac{1}{2}} + u_2 f_{\frac{3}{2}} + u_3 f_{\frac{5}{2}} \ldots + u_{n-1} f_{n-\frac{3}{2}} \\ &= u_1 f_{\frac{1}{2}} - u_{n-1} f_{n-\frac{1}{2}} + \sum_{j=1}^{n-2} f_{j+\frac{1}{2}} (u_{j+1} - u_j) \end{aligned}$$

Suppose now that

$$f_{j+\frac{1}{2}} = G_{u_{j+\frac{1}{2}}} \tag{3.6}$$

where $G_{u_{j+\frac{1}{2}}}$ is the mean value of G_u in the range from u_j to u_{j+1} such that

$$G_{u_{j+\frac{1}{2}}}(u_{j+1} - u_j) = G(u_{j+1}) - G(u_j) \tag{3.7}$$

Then, denoting $G(u_j)$ by G_j,

$$\begin{aligned} \Delta x \sum_{j=1}^{n-1} u_j \frac{du_j}{dt} &= u_1 f_{\frac{1}{2}} - u_{n-1} f_{n-\frac{1}{2}} + \sum_{j=1}^{n-2} (G_{j+1} - G_j) \\ &= u_1 f_{\frac{1}{2}} - u_{n-1} f_{n-\frac{1}{2}} - G_1 + G_{n-1} \end{aligned}$$

Now let (3.1) be discretized at the end points as

$$\frac{du_0}{dt} + \frac{2}{\Delta x}(f_{\frac{1}{2}} - f_0), \qquad \frac{du_n}{dt} + \frac{2}{\Delta x}(f_n - f_{n-\frac{1}{2}}) \tag{3.8}$$

where

$$f_0 = f(u_0), \qquad f_n = f(u_n)$$

and define the discrete approximation to the energy as

$$E = \frac{\Delta x}{2}\left(\frac{u_0^2}{2} + \frac{u_n^2}{2}\right) + \Delta x \sum_{j=1}^{n-1} \frac{u_j^2}{2}$$

Then

$$\begin{aligned} \frac{dE}{dt} &= u_0 f_0 - u_n f_n - G_0 + G_n \\ &= F_0 - F_n \end{aligned} \tag{3.9}$$

Thus the energy balance (3.3) is exactly recovered by the discrete scheme. Equations (3.6) and (3.7) are satisfied by evaluating the numerical flux as

$$f_{j+\frac{1}{2}} = \int_0^1 f(\hat{u}(\theta)) d\theta \tag{3.10}$$

where

$$\hat{u}(\theta) = u_j + \theta(u_{j+1} - u_j) \tag{3.11}$$

since then

$$\begin{aligned} G_{j+1} - G_j &= \int_0^1 G_u(\hat{u}(\theta)) u_\theta d\theta \\ &= \int_0^1 G_u(\hat{u}(\theta) d\theta (u_{j+1} - u_j) \end{aligned}$$

Thus we have established Theorem 3.1:

Theorem 3.1 *If the scalar conservation law* (3.1) *is approximated by the semi-discrete conservative scheme* (3.5)*, it also satisfies the semi-discrete energy conservation law* (3.8) *if the numerical flux* $f_{j+\frac{1}{2}}$ *is evaluated by equations* (3.10) *and* (3.11)*.*

In the case of Burgers equation formulas (3.10) and (3.11) yields the same numerical flux that was defined in Section 2

$$f_{j+\frac{1}{2}} = \frac{u_{j+1}^2 + u_{j+1}u_j + u_j^2}{6} \tag{3.12}$$

In the case of a general polynomial flux

$$f(u) = \frac{u^q}{q} \tag{3.13}$$

note that

$$G(u) = \int_0^u f(v)dv = \frac{u^{q+1}}{q(q+1)}$$

and

$$\begin{aligned} G_{j+1} - G_j &= \frac{1}{q(q+1)}(u_{j+1}^{q+1} - u_j^{q+1}) \\ &= \frac{u_{j+1} - u_j}{q(q+1)}(u_{j+1}^q + u_{j+1}^{q-1} \ldots + u_j^q). \end{aligned}$$

Thus the numerical flux should be evaluated as

$$f_{j+\frac{1}{2}} = \frac{1}{q+1}(u_{j+1}^q + u_{j+1}^{q-1}u_j \ldots + u_j^q) \tag{3.14}$$

If (3.10) cannot be evaluated in closed form one may approximate it by numerical integration. The Lobatto quadrature rule which uses the end points and $n-2$ interior points is suitable for this purpose, giving an exact result for polynomials of degree up to $2n-3$. Taking $n = 3$

yields Simpson's rule

$$f_{j+\frac{1}{2}} = \frac{1}{6}(f(u_{j+1}) + 4f\left(\frac{1}{2}(u_{j+1} + u_j)\right) + f(u_j))$$

which is exact for the Burgers equation.

In order to enforce appropriate inflow and outflow boundary conditions we introduce simultaneous approximation terms (SAT). Denote the wave speed by

$$a(u) = \frac{\partial f}{\partial u}$$

and let

$$a^+ = \frac{1}{2}(a + |a|), \qquad a^- = \frac{1}{2}(a - |a|)$$

Then we replace (3.8) by

$$\left.\begin{aligned} \frac{du_0}{dt} + \frac{2}{\Delta x}(f_{\frac{1}{2}} - f_0) + \frac{\tau}{\Delta x}a^+(u_0 - g_0) &= 0 \\ \frac{du_n}{dt} + \frac{2}{\Delta x}(f_n - f_{\frac{1}{2}}) - \frac{\tau}{\Delta x}a^-(u_n - g_n) &= 0 \end{aligned}\right\} \tag{3.15}$$

where g_0 and g_n denote the boundary values that should be imposed at inflow boundaries, and the magnitude of the penalty for not exactly satisfying the boundary conditions is determined by the parameter τ.

In the case of the polynomial flux $f(u) = \frac{u^q}{q}$, q even, let $a(u_L, u_R)$ be the average numerical wave speed between the states u_L and u_R such that

$$a(u_L, u_R)(u_R - u_L) = f_R - f_L.$$

Then

$$a(u_L, u_R) = \frac{1}{q}\frac{u_R^q - u_L^q}{u_R - u_L}$$

Now at the left boundary, for example, modify the equation by simultaneous approximation terms so that

$$\begin{aligned} \frac{du_0}{dt} &= -\frac{1}{\Delta x}(f_{\frac{1}{2}} - f_0) - \frac{\tau}{\Delta x}a^+(u_0, g_0)(u_0 - g_0) \\ &= -\frac{1}{\Delta x}(f_{\frac{1}{2}} - f_0) - \frac{\tau}{k\Delta x}(u_0^q - g_0^q) \end{aligned}$$

Then if $u_0 = \alpha g_0$, the penalty term modifies the rate of change of the discrete energy by $F(\alpha) g_0^q$ where

$$F(\alpha) = \frac{\alpha^{q+1} - 1}{q+1} - \frac{\tau}{q}\alpha(\alpha^q - 1)$$

$F(1) = 0$, and $F(\alpha)$ will cross the axis at $\alpha = 1$ unless $F'(1) = 0$. Here

$$F'(\alpha) = \alpha^q - \frac{\tau}{2k}\left((q+1)\alpha^q - 1\right)$$

and $F'(1) = 0$ if $\tau = 2$, giving

$$F'(\alpha) = \frac{1}{q}(1 - \alpha^q)$$

Then $F'(\alpha) > 0$ in the interval $-1 < \alpha < 1$, and $F'(\alpha) < 0$ when $\alpha > 1$. Consequently $F(\alpha) \leq 0$ in the entire range $\alpha > -1$, yielding a favorable modification of the energy growth for all in flow conditions.

4 Discrete conservation of energy or entropy for a one dimensional system

In this section it is shown how to extend the procedure described in Section 3 in order to discretely satisfy an additional invariant for a system of conservation laws. Consider the system

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = 0 \tag{4.1}$$

where u now denotes the state vector, and $f(u)$ is the flux vector. Suppose that $h(u)$ is a convex function of u that can be regarded as an energy density. Multiplying (4.1) by

$$w^T = h_u$$

we obtain

$$h_u \frac{\partial u}{\partial t} = \frac{\partial h}{\partial t} = -w^T \frac{\partial f}{\partial x} = f^T \frac{\partial w}{\partial x} - \frac{\partial}{\partial x}(f^T w)$$

Suppose also that there exists a scalar function $G(u)$ such that

$$G_w = G_u u_w = f^T \tag{4.2}$$

Then

$$\frac{\partial h}{\partial t} = G_w \frac{\partial w}{\partial x} - \frac{\partial}{\partial x}(f^T w) = -\frac{\partial F}{\partial x} \tag{4.3}$$

where the flux for h is

$$F = f^T w - G \tag{4.4}$$

The relation (4.2) implies that

$$f_w = G_{ww}$$

and hence f_w is symmetric. In the case of a system for which f_u is not symmetric, such as the equations of gas dynamics, this precludes the choice

$$h(u) = u^T u$$

The conditions under which a system of conservation laws can be written in symmetric hyperbolic form have been studied in papers by Godunov, Mock and Harten [13, 14, 15].

Suppose that there exists a convex function $h(u)$ such that

$$h_u f_u = F_u \tag{4.5}$$

for some scalar function $F(u)$. Then

$$\frac{\partial h}{\partial t} = h_u \frac{\partial u}{\partial t} = -h_u f_u \frac{\partial u}{\partial x} = -F_u \frac{\partial u}{\partial x}$$

Thus $h(u)$ satisfies the conservation law

$$\frac{\partial}{\partial t} h(u) + \frac{\partial}{\partial x} F(u) = 0 \tag{4.6}$$

Here $h(u)$ is a generalized entropy function and $F(u)$ is the corresponding entropy flux. The main theorems are as follows:

Theorem 4.1 *Suppose* (4.1) *can be symmetrized by a change of variables from u to w, so that it assumes the form*

$$u_w \frac{\partial w}{\partial t} + f_w \frac{\partial w}{\partial x} = 0$$

where u_w and f_w are symmetric and u_w is positive definite. Then

$$u = q_w, \qquad u_w = q_{ww}$$

for some convex function $q(w)$, while

$$f = G_w, \qquad f_w = G_{ww}$$

and (4.1) *has an entropy function*

$$h(u) = u^T w - q(w) \tag{4.7}$$

and entropy flux

$$F(u) = f^T w - G(w) \tag{4.8}$$

Theorem 4.2 *Suppose $h(u)$ is an entropy function for* (4.1)*. Then*

$$w^T = h_u$$

symmetrizes (4.1).

The proofs are given by Harten [15].

Suppose now that (4.1) is approximated in semi-discrete form on a uniform grid over the range $j = 0, n$ as

$$\frac{du_j}{dt} + \frac{1}{\Delta x}(f_{j+\frac{1}{2}} - f_{j-\frac{1}{2}}) = 0 \tag{4.9}$$

where the numerical flux $f_{j+\frac{1}{2}}$ is a function of u_i over a range bracketing u_j. Then we can construct a scheme which discretely satisfies the energy or entropy conservation law in the same manner as for the scalar case. Multiplying (4.9) by $w_j^T = h_{u_j}$ and summing by parts over the interior points

$$\begin{aligned}\sum_{j=1}^{n-1} w_j^T \frac{du_j}{dt} &= \sum_{j=1}^{n-1} h_{u_j} \frac{\partial u_j}{\partial t} = \sum_{j=1}^{n-1} \frac{dh_j}{dt} \\ &= w_1^T f_{\frac{1}{2}} - w_{n-1}^T f_{n-\frac{1}{2}} + \sum_{j=1}^{n-2} f_{j+\frac{1}{2}}^T (w_{j+1}^T - w_j^T)\end{aligned}$$

Now suppose that

$$f_{j+\frac{1}{2}}^T = G_{w_{j+\frac{1}{2}}} \tag{4.10}$$

where $G_{w_{j+\frac{1}{2}}}$ is a mean value of G_w between w_j and w_{j+1} in the sense of Roe, such that

$$G_{w_{j+\frac{1}{2}}}(w_{j+1} - w_j) = G_{j+1} - G_j \tag{4.11}$$

where G_j denotes $G(w_j)$. Then

$$\Delta x \sum_{j=1}^{n-1} \frac{dh_j}{dt} = w_1^T f_{\frac{1}{2}} - w_{n-1}^T f_{n-\frac{1}{2}} - G_1 + G_{n-1}$$

Now let (4.1) be discretized at the end points as

$$\frac{du_0}{dt} + \frac{2}{\Delta x}(f_{\frac{1}{2}} - f_0) = 0, \qquad \frac{du_n}{dt} + \frac{2}{\Delta x}(f_n - f_{n-\frac{1}{2}}) = 0$$

where

$$f_0 = f(u_0), \qquad f_n = f(u_n)$$

Then we obtain the discrete conservation law

$$\begin{aligned} \frac{\Delta x}{2}\left(\frac{dh_0}{dt}+\frac{dh_n}{dt}\right)+\Delta x\sum_{j=1}^{n-1}\frac{dh_j}{dt} &= w_0^T f_n - w_n^T f_n - G_0 + G_n \\ &= F_0 - F_n \end{aligned} \tag{4.12}$$

where F is the entropy flux (4.4).

$G_{w_{j+\frac{1}{2}}}$ can be constructed to satisfy (4.11) exactly in the form

$$G_{w_{j+\frac{1}{2}}} = \int_0^1 G_w\left(\hat{w}(\theta)\right) d\theta$$

where

$$\hat{w}(\theta) = w_j + \theta(w_{j+1} - w_j) \tag{4.13}$$

since then

$$\begin{aligned} G_{j+1} - G_j &= \int_0^1 G_w\left(\hat{w}(\theta)\right) w_\theta \, d\theta \\ &= \int_0^1 G_w\left(\hat{w}(\theta)\right) d\theta (w_{j+1} - w_j) \end{aligned}$$

Thus we can state theorem 4.3:

Theorem 4.3 *The semi-discrete conservation law* (4.9) *satisfies the semi-discrete entropy conservation law* (4.12) *if the numerical flux* $f_{j+\frac{1}{2}}$ *is constructed as*

$$f_{j+\frac{1}{2}} = \int_0^1 f\left(\hat{w}(\theta)\right) d\theta \tag{4.14}$$

where $\hat{w}(\theta)$ *is defined by* (4.13).

For some systems, such as the equations of gas dynamics, it may not be possible to express the integral (4.14) in closed form. Then one may rescale the interval of integration for θ to (-1,1) so that

$$f_{j+\frac{1}{2}} = \frac{1}{2}\int_{-1}^1 f\left(\tilde{w}(\theta)\right) d\theta$$

where

$$\tilde{w}(\theta) = \frac{1}{2}(w_{j+1} + w_j) + \frac{1}{2}\theta(w_{j+1} - w_j)$$

and apply the n point Lobatto rule.

In general neither boundary is necessarily purely inflow or outflow. Consequently, in order to impose proper boundary conditions it is essential to distinguish ingoing and outgoing waves at the boundaries. For this purpose we can split the Jacobian matrix $A = f_u$ into positive and negative parts $A^{\pm}$. Suppose that A is decomposed as

$$A = R\Lambda R^{-1}$$

where the columns of R are the right eigenvectors of A, and Λ is a diagonal matrix comprising the eigenvalues. Then

$$A^{\pm} = R\Lambda^{\pm}R^{-1}$$

where Λ^+ and Λ^- contain the positive and negative eigenvalues respectively. Now the boundary conditions maybe imposed by adding simultaneous approximation terms (SAT) at the boundaries. Accordingly we set

$$\begin{aligned} \frac{du_0}{dt} + \frac{1}{\Delta x}(f_{\frac{1}{2}} - f_0) + \frac{\tau}{\Delta x}A^+(u_0 - g_0) = 0 \\ \frac{du_n}{dt} + \frac{1}{\Delta x}(f_n - f_{n-\frac{1}{2}}) - \frac{\tau}{\Delta x}A^-(u_n - g_n) = 0 \end{aligned} \tag{4.15}$$

where g_0 and g_n define the exterior data and the parameter τ determines the magnitude of the penalty when the solution is not consistent with the incoming waves. Appropriate values of A_0 and A_n may be obtained by taking them to be the mean valued Jacobian matrices in the sense of Roe [8] such that

$$\begin{aligned} A_0(u_0 - g_0) = f(u_0) - f(g_0) \\ A_n(u_n - g_n) = f(u_n) - f(g_n) \end{aligned} \tag{4.16}$$

If shock waves appear in the solution, the scheme needs to be modified by shock operators. A desirable property of a shock capturing scheme is that the numerical shock structure for a stationary shock should contain no more than one interior point [12]. This can be achieved by characteristic upwind schemes or the CUSP scheme [12]. The characteristic scheme can be constructed by adding matrix diffusion. Let f_j denote $f(u_j)$. Following Roe

[8], we introduce the mean Jacobian matrix $A_{j+\frac{1}{2}}$ such that

$$A_{j+\frac{1}{2}}(u_{j+1} - u_j) = f_{j+1} - f_j$$

Decomposing the Jacobian matrix in terms of its eigenvectors and eigenvalues as

$$A_{j+\frac{1}{2}} = R\Lambda R^{-1}$$

as in the treatment above of the boundaries, define the absolute Jacobian matrix as

$$\left|A_{j+\frac{1}{2}}\right| = R|\Lambda|R^{-1}$$

where $|\Lambda|$ is a diagonal matrix containing the absolute values of the eigenvalues. Then the upwind flux can be expressed as

$$f_{U_{j+\frac{1}{2}}} = \frac{1}{2}(f_{j+1} + f_j) - \frac{1}{2}\left|A_{j+\frac{1}{2}}\right|(u_{j+1} - u_j) \tag{4.17}$$

Also let $f_{C_{j+\frac{1}{2}}}$ be the central flux defined by equation (4.14). Then we construct the flux throughout the domain as

$$f_{j+\frac{1}{2}} = \left(1 - S_{j+\frac{1}{2}}\right) f_{C_{j+\frac{1}{2}}} + S_{j+\frac{1}{2}} f_{U_{j+\frac{1}{2}}} \tag{4.18}$$

where $S_{j+\frac{1}{2}}$ is a switching function with values in the range $0 \leq S_{j+\frac{1}{2}} \leq 1$, of the order of a high power of Δx except in the neighborhood of a shock wave, where it should have a value of unity. The switching function can be constructed in a manner similar to the switch used for the Burgers equation, equations (2.26) and (2.27). The same formulas may be used to detect extrema in either the pressure or the entropy. Alternatively we can use these formulas to identify extrema in the the characteristic variables by applying them to

$$\Delta v_{j+\frac{1}{2}} = R^{-1}_{j+\frac{1}{2}} \Delta u_{j+\frac{1}{2}}$$

5 Incompressible fluid flow

The procedure for discretely satisfying the basic conservation laws plus an additional invariant, which has been proposed in Section 4, requires the systems to be expressed in symmetric form so that (4.2) holds. In the case of incompressible flow the pressure is not directly related to the state vector via an equation of state, and must be determined indirectly from the continuity equation. Denoting the velocity components by v_i, and the pressure and density by p and ρ, three dimensional inviscid incompressible flow is described by the continuity equation and three momentum equations

$$\frac{\partial v_i}{\partial x_i} = 0 \tag{5.1}$$

$$\rho\left(\frac{\partial v_i}{\partial t} + v_j\frac{\partial v_i}{\partial x_j}\right) + \frac{\partial p}{\partial x_i} = 0, \qquad i = 1,2,3 \tag{5.2}$$

In order to express the momentum equations in symmetric conservation form introduce the reduced pressure

$$\hat{p} = p - \frac{1}{2}\rho q^2 \tag{5.3}$$

where $q^2 = v_i^2$. For convenience suppose that units are chosen such that $\rho = 1$. Using (5.1), the momentum equations can be expressed as

$$\frac{\partial v_i}{\partial t} + \frac{\partial}{\partial x_j}(v_i v_j) + \frac{\partial}{\partial x_i}\left(\hat{p} + \frac{v_j^2}{2}\right) = 0 \tag{5.4}$$

These can be written in state vector form as

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x_i} f^i(u) + \frac{\partial P^i}{\partial x_i} = 0 \tag{5.5}$$

Here the state and flux vectors are

$$u = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \qquad f^i(u) = v_i u + e_i\frac{q^2}{2} \tag{5.6}$$

where e_i is the unit vector in the i direction and

$$P^i = e_i\,\hat{p} \tag{5.7}$$

Then

$$\frac{\partial v_i}{\partial u} = e_i^T, \qquad \frac{\partial}{\partial u}\left(\frac{q^2}{2}\right) = u^T$$

and the Jacobian matrices are

$$A_i = \frac{\partial f^i}{\partial u} = v_i I + u e_i^T + e_i u^T \tag{5.8}$$

The flux vectors f^i are homogeneous functions of the state variables of degree two, with the consequence that

$$A^i u = 2 f^i(u) \tag{5.9}$$

Moreover

$$f^{i^T}(u) = \frac{\partial G^i}{\partial u} \tag{5.10}$$

where

$$G^i(u) = u_i \frac{q^2}{2} = \frac{1}{3} u^T f^i(u) \tag{5.11}$$

Multiplying (5.4) by v_i

$$v_i \frac{\partial v_i}{\partial t} + v_i \frac{\partial}{\partial x_j}(v_i v_j) + v_i \frac{\partial}{\partial x_i}\left(\hat{p} + \frac{v_j^2}{2}\right) = 0 \tag{5.12}$$

Interchanging the indices in the second term, it may be grouped with the last term to give

$$v_j \frac{\partial}{\partial x_i}(v_i v_j) + v_i v_j \frac{\partial v_j}{\partial x_i} = \frac{\partial}{\partial x_i}(v_i v_j^2)$$

Thus (5.12) reduces to

$$\frac{\partial}{\partial t}\left(\frac{v_i^2}{2}\right) + \frac{\partial}{\partial x_i}(v_i v_j^2 + v_i \hat{p}) - \hat{p}\frac{\partial v_i}{\partial x_i} = 0$$

Defining the energy flux as

$$F^i(u) = v_i(\hat{p} + q^2) = v_i\left(p + \frac{q^2}{2}\right) = v_i p_t \tag{5.13}$$

where p_t is the total pressure, we obtain the energy conservation law

$$\frac{\partial}{\partial t}\left(\frac{v_i^2}{2}\right) + \frac{\partial F^i}{\partial x_i} = \hat{p}\frac{\partial v_i}{\partial x_i} \tag{5.14}$$

where the right hand side vanishes provided that the continuity equation is satisfied. The total kinetic energy in a domain D with volume element dV is

$$E = \int_D \frac{v_i^2}{2} dV$$

Now (5.8) may be integrated to give

$$\frac{dE}{dt} = \int_B n_i F^i dS + \int \hat{p} \frac{\partial v_i}{\partial x_i} dV \tag{5.15}$$

where B is the boundary of D, n_i is the inward normal and dS is the area element.

Suppose now that the domain is covered by a grid, and the equations are discretized in finite volume form with the discrete flow variables defined at the nodes, each of which is contained in a polyhedral control volume. In the case of either a hexahedral or a tetrahedral grid the control volumes may be taken as the dual cells connecting the centers of the primary cells. For example, one could construct the dual cells of a tetahedral mesh by dividing each primary cell about its median into subcells of equal volume which are assembled at the nodes to form the "median" dual cells. Examples of discretization schemes on the median dual mesh include the Airplane code of Jameson, Baker and Weatherill [19], and Stanford University's CDP code [20]. A representative two-dimensional grid is shown in Figure 5.1

The control volumes of the boundary nodes extend into the interior and are closed by faces on the boundary. Now the control volume of each interior node, say node o, consists of faces (not necessarily planar) with a directed face area S_{op} for the face separating nodes o and p. The control volume of a boundary node o is closed by a vector face area S_o which is the negative of the sum $\sum_p S_{op}$ of the face areas between o and each of its neighbors. The semi-discrete finite volume scheme has the form

$$\frac{du_o}{dt} + \frac{1}{vol_o} \sum_p S_{op}^i \left(f_{op}^i + P_{op}^i \right) \tag{5.16}$$

where the sum is over the faces of the control volume containing o, S_{op}^i is the projected area in the i direction of the face separating o and p, f_{op}^i and P_{op}^i are the convective and pressure fluxes between o and p, and the repeated superscript i denotes summation over the coordinate directions.

At a boundary node there is an additional contribution $S_o^i f_o^i$ where $f_o = f(u_o)$ is the

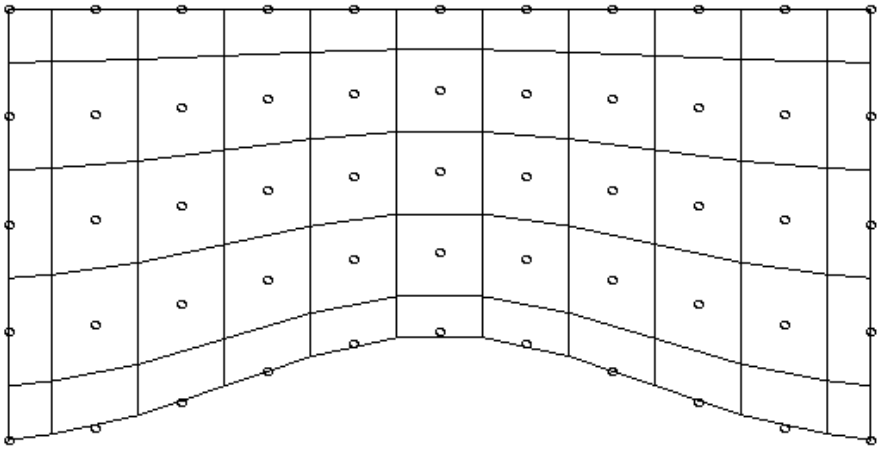


Figure 5.1: Convergent-divergent duct: ∘ = nodes of primary cells, — = dual cells

flux evaluated with the nodal values of the state vectors. The discrete energy is the sum

$$E = \frac{1}{2} \sum_o vol_o \, u_o^T u_o \tag{5.17}$$

over the nodes o, and its rate of change is

$$\frac{dE}{dt} = \sum_o vol_o \, u_o^T \frac{du_o}{dt} \tag{5.18}$$

where $vol_o \, \frac{du_o}{dt}$ is given by the sum (5.16) over the faces for that node. Each interior face appears twice in the resulting double sum because it is contained in the sums (5.16) for the two nodes it separates, while each exterior face appears once. The convective contribution of each interior face to $\frac{dE}{dt}$ is thus

$$\left(u_p^T - u_o^T\right) S_{op}^i \, f_{op}^i$$

where the flux

$$f_{op}^{i^T} = \frac{\partial G_{op}^i}{\partial u} \tag{5.19}$$

Provided f^i_{op} is constructed so that

$$f^{i^T}_{op}(u_p - u_o) = G^i_p - G^i_o \tag{5.20}$$

exactly, the convective contribution of each interior face is thus

$$S^i_{op}\left(G^i_p - G^i_o\right)$$

If one associates all terms containing G_o with the node o, it receives the inner product of the sum of its face areas S_{op} with G_o, but this sum is zero. Thus the only contribution of the convective terms to $\frac{dE}{dt}$ is a sum over the boundary nodes. However, the sum of the vector areas S_{op} for a boundary node is the negative of its external area S_o, so the convective contribution to $\frac{dE}{dt}$ is

$$\sum_{\text{boundary nodes}} S^i_o\, G^i_o$$

There remains in the sum (5.18) the convective contribution

$$\sum_{\text{boundary nodes}} S^i_o\, u^T_o\, f^i_o$$

Thus the total convective contribution to $\frac{dE}{dt}$ is

$$-\sum_{\text{boundary nodes}} S^i_o\left(u^T f^i_o - G^i_o\right) \tag{5.21}$$

where S_o is the outward face area of node o.

It remains to evaluate the contribution of the pressure to the discrete energy balance. For this purpose it is necessary to prove a discrete analog of the Gauss theorem

$$\int_D v_i \frac{\partial p}{\partial x_i} dV = \int_B p v_n dS - \int_D p \frac{\partial v_i}{\partial x_i} dV \tag{5.22}$$

where v_n is the normal velocity through the boundary. Across each face there is a contribution $u^T_o \sum_p S^i_{op} P^i_{op}$ from the pressure. Suppose that the pressure flux is evaluated as

$$P^i_{op} = \frac{1}{2}\left(P^i_o + P^i_p\right) \tag{5.23}$$

The sum (5.18) then contains

$$\frac{1}{2}\sum_o u_o^T\, P_o^i \sum_p S_{op}^i + \frac{1}{2}\sum_o u_o^T \sum_p P_p^i\, S_{op}^i$$

At every interior node o the fluxes $u_p^T S_{op}^i\, P_o^i$ generated by the neighbors can be associated with o, while $u_o^T P_o^i \sum_p S_{op}^i = 0$. Thus one can subtract this quantity to produce

$$-\frac{1}{2}P_o^{i^T}\sum_p (u_p + u_o)\, S_{op}^i$$

while represents $-\hat{p}_o\, vol\ (\bigtriangledown \cdot u)_o$. A boundary node o receives the contributions

$$\frac{1}{2}P_o^{i^T}\sum_p u_p\, S_{po}^i = -\frac{1}{2}P_o^{i^T}\sum_p u_p\, S_{op}^i$$

from its neighbors while it retains the contributions

$$\frac{1}{2}P_o^{i^T} u_o \sum_p S_{op}^i + P_o^{i^T} u_o S_o^i$$

giving the total

$$\begin{aligned}-\frac{1}{2}P_o^{i^T}\sum_p (u_p + u_o)\, S_{op}^i + P_o^{i^T} u_o \sum_p S_{op}^i + P_o^{i^T} u_o\, S_o^i \\ = -\frac{1}{2}P_o^{i^T}\sum_p (u_p + u_o) S_{op}^i - P_o^{i^T}\, u_o\, S_o^i + P_o^{i^T}\, u_o\, S_o^i\end{aligned}$$

The first two terms represent $-\hat{p}_o\, vol_o\, (\bigtriangledown \cdot u)_o$. The third term represents the surface integral on the right hand side of (5.22). This establishes theorem 5.1

Theorem 5.1 *The finite volume discretization of* $v_i \frac{\partial p}{\partial x_i}$ *represented by*

$$\sum_o u_o^T \sum_p S_{op}^i P_{op}^i$$

where P_{op}^i *is evaluated as the arithmatic average exactly satisfies the discrete analog of the Gauss Theorem* (5.22)*.*

Combining the boundary contribution from the pressure with the contribution (5.21) of

the convective terms we obtain

$$-\sum_{\text{boundary nodes}} S_o^i(u^T f_o^i - G_o^i - u^T P_o^i) = -\sum_{\text{boundary nodes}} S_o^i F_o^i$$

where F_o^i is the discrete energy flux

$$F_o^i = v_o^i \, p_{to}$$

evaluated with the nodal values of v^i and p_t. This establishes theorem 5.2

Theorem 5.2 *If the discrete flux vector f_{op}^i is constucted so that it satisfies* (5.20)*, and the pressure flux is evaluated by* (5.23)*, the rate of change of the discrete energy is exactly*

$$\frac{dE}{dt} = -\sum_{boundary\ nodes} S_o^i F_o^i + \sum_o vol\, \hat{p}_o (\bigtriangledown \cdot u)_o$$

where the second sum is over all the nodes and $(\bigtriangledown \cdot u)_o$ is the consistent discretization of $\bigtriangledown \cdot u$ at node o.

Since $f^i(u)$ is a quadratic function of u, equation (5.20) *is satisfied exactly by using Simpson's rule to evaluate*

$$\begin{aligned} f_{op}^i &= \int_o^1 f(u_o + \theta(u_p - u_o))d\theta \\ &= \frac{1}{6}\left(f(u_o) + 4f\left(\frac{1}{2}(u_o + u_p)\right) + f(u_p)\right) \end{aligned}$$

While the proof of theorem 5.2 is somewhat intricate, the author has numerically confirmed that it is satisfied to machine accuracy in test calculations for two different cases, flow past a circular cylinder, and flow through a convergent divergent duct.

6 Gas Dynamics

In this section the theory developed in Section 4 is applied to the three dimensional gas dynamic equations. Because the Jacobians for the conservative variables are not symmetric, the formulation requires a transformation of variables which symmetrizes the equations. In conservation form the equations are

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x_i} f^i(u) = 0. \tag{6.1}$$

Here the state and flux vectors are

$$u = \begin{bmatrix} \rho \\ \rho v_1 \\ \rho v_2 \\ \rho v_3 \\ \rho E \end{bmatrix}, \qquad f^i = \begin{bmatrix} \rho v_i \\ \rho v_i v_1 + \delta_{i1} p \\ \rho v_i v_2 + \delta_{i2} p \\ \rho v_i v_3 + \delta_{i3} p \\ \rho v_i H \end{bmatrix} \tag{6.2}$$

where ρ is the density, v_i are the velocity components, and E and H are the specific energy and enthalpy. Also

$$p = (\gamma - 1)\rho \left(E - \frac{v_i^2}{2} \right), \qquad H = E + \frac{p}{\rho} \tag{6.3}$$

where γ is the ratio of specific heats.

In the absence of shock waves the entropy

$$s = \log \left(\frac{p}{\rho^\gamma} \right) \tag{6.4}$$

is constant along streamlines,

$$\frac{\partial s}{\partial t} + v_i \frac{\partial s}{\partial x_i} = 0$$

Harten shows in his paper [15] that

$$h(s) = \rho g(s), \qquad F^i(s) = \rho v_i h(s) \tag{6.5}$$

constitute a generalized entropy function and entropy fluxes provided that

$$\frac{\ddot{g}(s)}{\dot{g}(s)} < \frac{1}{\gamma} \tag{6.6}$$

This equation ensures the convexity of $h(u)$, which satisfies the entropy conservation law

$$\frac{\partial}{\partial t}h(u) + \frac{\partial}{\partial x_i}F^i(u) = 0 \tag{6.7}$$

Moreover the conservation equations (6.1) are symmetrized by the variables

$$w^T = \frac{\partial h}{\partial u} \tag{6.8}$$

assuming the form

$$u_w \frac{\partial w}{\partial t} + \frac{\partial f^i}{\partial w}\frac{\partial w}{\partial x_i} = 0 \tag{6.9}$$

where u_w and $\frac{\partial f^i}{\partial w}$ are symmetric, and $u_w > 0$.

Hughes, Franca and Mallet [21] and Gerritsen and Olsson [22] have proposed schemes using the entropy variables. Reference [21] presents a finite element formulation for the Navier Stokes equations, choosing

$$h(s) = -\rho s \tag{6.10}$$

in order to obtain a symmetric form for the viscous terms. Gerritsen and Olsson treat the Euler equations, and prefer the form

$$h(s) = \rho e^{\frac{s}{\gamma+\alpha}} = \rho\left(\frac{p}{\rho^\gamma}\right)^{\frac{1}{\gamma+\alpha}} \tag{6.11}$$

where α is a parameter to be chosen, because it leads to a homogeneous function $u(w)$ such that

$$u(\theta w) = \theta^\beta u(w) \tag{6.12}$$

where

$$\beta = -\frac{\gamma+\alpha}{\gamma-1} \tag{6.13}$$

and

$$u_w w = \beta u \tag{6.14}$$

The convexity condition (6.6) is satisfied if $\alpha > 0$ or $\alpha < -\gamma$. The choice $\alpha = 1$ yields the comparatively simple form

$$w = \frac{p^*}{p}\begin{bmatrix} u_5 \\ -u_2 \\ -u_3 \\ -u_4 \\ u_1 \end{bmatrix}, \qquad u = \frac{p}{p^*}\begin{bmatrix} w_5 \\ -w_2 \\ -w_3 \\ -w_4 \\ w_1 \end{bmatrix} \tag{6.15}$$

where

$$p^* = -\frac{1}{\beta}e^{\frac{s}{\gamma+1}} = (\gamma - 1)\left(w_1 - \frac{w_1^2}{2w_5}\right) \tag{6.16}$$

Multiplying (6.1) by w^T we obtain the entropy evolution equation in the form

$$h_u\frac{\partial u}{\partial t} = \frac{\partial}{\partial t}h(u) = -w^T\frac{\partial}{\partial x_i}f^i(u)$$

The left hand side can also be expressed as

$$w^T\frac{\partial u}{\partial t} = w^T u_w\frac{\partial w}{\partial t}$$

Gerritsen and Olsson split this as

$$\begin{aligned}\frac{\beta}{\beta+1}w^T u_w\frac{\partial w}{\partial t} + \frac{1}{\beta+1}w^T\frac{\partial u}{\partial t} &= \frac{1}{\beta+1}\left(u^T\frac{\partial w}{\partial t} + w^T\frac{\partial u}{\partial t}\right)\\ &= \frac{1}{\beta+1}\frac{\partial}{\partial t}(u^T w)\\ &= \frac{1}{\beta+1}\frac{\partial}{\partial t}(w^T u_w w).\end{aligned}$$

in view of the homogeneity relation (6.14). Thus $h(u)$ can be regarded as an energy function for the system (6.1).

Gerritsen and Olsson also show that by splitting the spatial derivatives between the conservation and quasilinear forms as

$$\frac{\beta}{\beta+1}\frac{\partial}{\partial x_i}f^i(w) + \frac{1}{\beta+1}\frac{\partial f^i}{\partial w}\frac{\partial w}{\partial x_i}$$

and using central differencing for both at interior points, one obtains a skew-symmetric

operator which discretely satisfies the entropy conservation law (6.7). Since entropy will be generated by shock waves, artificial diffusion is still required for their capture. The skew-symmetric form, which has been further investigated by Yee, Vinokur and Djomehri [23], gives up conservation form for the basic conservation laws (6.1) in favor of discrete entropy conservation. However, the formulation of Section 4 actually allows discrete entropy conservation while retaining conservation form for the basic equations, as is shown in the following paragraphs.

Consider a finite volume discretization of (6.1) on a mesh similar to that described in Section 5, with u stored at the mesh nodes, each of which is contained in a control volume. The semi-discrete finite volume scheme at node o is

$$\frac{du_o}{dt} + \frac{1}{vol_o} \sum_p S^i_{op} f^i_{op} = 0 \tag{6.17}$$

where the sum is over the faces of the control volume containing o, S^i_{op} is the projected area in the i direction of the face separating o and p, and f^i_{op} is the flux in the i direction between o and p. Then multiplying by $vol_o w_o^T$ and summing over the nodes we obtain the discrete entropy equation

$$\sum_o vol_o w_o^T \frac{du_o}{dt} = \sum_o vol_o \frac{dh_o}{dt} = -\sum_o w_o^T \sum_p S^i_{op} f^i_{op} \tag{6.18}$$

As was shown in Section 5 the contribution of each interior face to this sum is

$$(w_p^T - w_o^T) S^i_{op} f^i_{op}$$

where the flux can be expressed as

$$f^i_{op} = G^i_{w_{op}} \tag{6.19}$$

Provided that $f^i_{op}(w)$ is constructed so that

$$f^{i^T}_{op}(w_p - w_o) = G^i_p - G^i_o \tag{6.20}$$

exactly, the contribution of each interior face is thus

$$S^i_{op}(G^i_p - G^i_o)$$

Associating all terms containing G_o^i with the node o, an interior node recovers a total contribution consisting of the inner product of the sum of the face areas S_{op}^i of its control volume with G_o^i, but the sum is zero. The sum of the face area S_{op}^i separating a boundary node from the neighbors is the negative of its external face area S_o^i, so the total contribution of the interior faces reduces to

$$\sum_{\text{boundary nodes}} S_o^i G_o^i$$

There remains in the sum (6.19) the contribution

$$\sum_{\text{boundary nodes}} S_o^i w_o^T f_o^i$$

Thus the rate of change of the discrete entropy can finally be expressed as

$$\begin{aligned} \frac{d}{dt}\sum_o vol_o h_o &= -\sum_{\text{boundary nodes}} S_o^i (w_o^T f_o^i - G_o^i) \\ &= -\sum_{\text{boundary nodes}} S_o^i F_o^i \end{aligned} \tag{6.21}$$

using the relation (4.8). This establishes

Theorem 6.1 *If the flux vector f_{op}^i is constructed so that it satisfies* (6.20)*, the rate of change of the discrete entropy is exactly*

$$-\sum_{\textit{boundary nodes}} S_o^i F_o^i$$

where F_o^i is the entropy flux evaluated with the values u_o.

The theorem holds for any choice of the entropy function (6.6) which satisfies the convexity requirement. Equation (6.20) is satisfied exactly if the flux f_{op}^i is evaluated in the manner described in Section 4 as

$$f_{op}^i = \int_0^1 f^i(\hat{w}(\theta))d\theta \tag{6.22}$$

where

$$\hat{w}(\theta) = w_o + \theta(w_p - w_o) \tag{6.23}$$

The entropy variables are needed only in the flux evaluation by these formulations. Aside from this equation (6.17) represents a standard finite volume approximation of (6.1) in

conservation form. Because the entropy variables, such as those specified in equation (6.11), contain fractional powers of p and ρ, it appears that one must resort to numerical integration to evaluate (6.22).

For this purpose the interval of integration may be shifted to $[-1, 1]$ as in Section 5, so that f^i_{op} is expressed as

$$f^i_{op} = \frac{1}{2}\int_{-1}^{1} f^i(\tilde{w}(\theta))d\theta$$

where

$$\tilde{w}(\theta) = \frac{1}{2}(w_p + w_o) + \frac{1}{2}\theta(w_p - w_o)$$

Then we may use the n point Lobatto formula. The 3 point formula is Simpson's rule

$$f^i_{op} = \frac{1}{6}\left[f(\tilde{w}(-1)) + 4f(\tilde{w}(\theta)) + f(\tilde{w}(1))\right]$$

The 5 point formula

$$f^i_{op} = \frac{1}{180}\left[9(f(\tilde{w}(-1)) + f(\tilde{w}(1))) + 49(f(\tilde{w}(-\theta_a)) + f(\tilde{w}(\theta_a))) + 64f(\tilde{w}(0))\right]$$

where $\theta_a = \frac{1}{7}\sqrt{21}$ appears to be a good compromise between accuracy and computational cost.

7 Conclusion

The foregoing sections demonstrate that both scalar conservation laws and systems of conservation laws which satisfy an entropy principle can be approximated in semi-discrete conservation form in a manner that also globally satisfies the corresponding discrete energy or entropy principle. This provides a path to the construction of stable non-dissipative discrete operators. However, if the governing equations support weak solutions containing shock waves (Burgers equation, gas dynamics), the energy or entropy principle is no longer valid in its basic form, and must be modified to account for energy dissipation or entropy production by the shock waves. Correspondingly the discrete formulation must be augmented by shock operators which restore the energy or entropy balance.

Since the inception of finite volume methods in computational fluid dynamics [24], there has been an issue of whether it would be better to calculate the interface flux by averaging the flux vectors

$$f_{j+\frac{1}{2}} = \frac{1}{2}(f_{j+1} + f_j)$$

or by calculating the flux from the average of the state vectors

$$f_{j+\frac{1}{2}} = f\left(\frac{1}{2}(u_{j+1} + u_j)\right)$$

Theorem (4.3) suggests that neither is the best choice. Instead, the use of equation (4.14) for the evaluation of the numerical flux is consistent with the discrete entropy principle.

In the course of the original development of the Jameson-Schmidt-Turkel (JST) scheme [25], which has been widely used to calculate transonic and supersonic flows, the scheme which was initially tested added artificial diffusion only in regions with strong pressure gradients, such as the neighborhood of shock waves. However, in numerous numerical experiments it appeared that the use of non-reflecting boundary conditions was not sufficient to assure convergence to a completely steady state, with the residuals reduced to machine zero. This led to the introduction of a higher-order background diffusive terms which were switched off at shock waves to prevent oscillations. It now appears worthwhile to rexamine whether satisfaction of the discrete entropy principle would allow the background dissipation to be substantially reduced or eliminated in simulations of both steady and unsteady flows. Certainly any artificial diffusive terms in a compressible large eddy simulation will need to be very carefully controlled in order to obtain correct energy spectra.

The Kolmogoroff scale of the smallest eddies in a turbulent flow with a Reynolds number Re is $\frac{1}{Re^{\frac{3}{4}}}$. Theorem 2.1 and the supporting numerical experiments indicate that shock waves can be fully resolved without any added numerical diffusion if the mesh interval is reduced to the order of $\frac{1}{Re}$, slightly smaller than the Kolmogoroff scale. Accordingly it can be estimated that direct numerical simulation (DNS) of three dimensional compressible turbulent flow will require a mesh with the order of Re^3 cells in order to resolve the full range of turbulent eddies, and also any shock waves that may appear in the flow. At the time of the introduction of the JST scheme in 1981, high-end computers attained speeds in the range of 100 megaflops (10^8 floating point operations per second). During the past 25 years computer performance has increased by a factor of about a million, with current high-end computers attaining speeds in the range of 100 teraflops. A further increase in performance by a factor of a million to 10^{20} floating point operations per second should enable fully resolved DNS of flows with Reynolds numbers in the range of 1 million on a mesh with 10^{18} cells, using an entropy preserving discretization with no added numerical diffusion. This is still a little short of flight Reynolds numbers of long range transport aircraft which are in the range of 50-100 million, but the eventual use of DNS for large scale compressible turbulent flows can clearly be foreseen.

Remaining questions include the extension of the present approach to higher-order discretizations, and the best choice of a discrete time stepping scheme. Discrete energy conservation could be preserved by an implicit time stepping scheme of Crank-Nicolson type, in which the spatial derivatives are evaluated using the average value of the state vectors between the beginning and the end of the time step,

$$\bar{u}_j = \frac{1}{2}\left(u_j^{n+1} + u_j^n\right)$$

Then, if we approximate the time derivative as

$$\frac{du_j}{dt} = \frac{1}{\Delta t}\left(u_j^{n+1} - u_j^n\right)$$

it follows that

$$\bar{u}_j \frac{du_j}{dt} = \frac{1}{2\Delta t}\left(u_j^{n+1^2} - u_j^{n^2}\right)$$

Accordingly, if the numerical fluxes are calculated by formula (3.10), or (4.14) in the case of a system, the discrete energy or entropy balance will correspond exactly to the continuous energy or entropy conservation law. However, the implementation of such scheme would require computationally expensive inner iterations. Moreover, it is hard to provide a rigorous

estimate of the impact on both accuracy and stability in the case that the inner iterations are not fully converged.

Acknowledgement

The author has benefited tremendously from the continuing support of the Air Force Office of Scientific Research over the last fifteen years, most recently through Grant Number AF–F49620–98–1–2005, under the direction of Dr. Fariba Fahroo. He is also indebted to Nawee Butsuntorn for his help in preparing this work in LaTeX.

References

[1] R. D. Richtmyer and K. W. Morton, "Difference Methods for Initial Value Problems", Interscience, 1967.

[2] A.E. Honein and P. Moin, "Higher Entropy Conservation and Numerical Stability of Compressible Turbulence Simulations", J. Comp. Phys., 201, 2004, 531–545.

[3] P. D. Lax and B. Wendroff, "Systems of Conservation Laws", Comm. Pure. Appl. Math., 13, 1960, 217–137.

[4] Bertil Gustafsson amd Pelle Olsson, "High-Order Centered Difference Schemes with Sharp Shock Resolution", J. Scientific Computing, 11, 1996, 229–260.

[5] S. K. Godunov, "A Difference Method for the Numerical Calculation of Discontinous Solutions of Hydrodynamic Equations", Math. Sbornik, 47, 1959, 271–306.

[6] J. P. Boris and D. L. Book, "Flux Corrected Transport, SHASTA, A Fluid Transport Algorithm that Works", J. Comp. Phys., 11, 1973, 38–69.

[7] Bram Van Leer, "Towards the Ultimate Conservative Difference Scheme, II, Monotomicity and Conservation Combined in a Second Order Scheme", J. Comp. Phys., 14, 1974, 361–370.

[8] P. L. Roe, "Approximate Reimann Solvers, Parameter Vectors and Difference Scheme", J. Comp. Phys., 43, 1981, 357–372.

[9] Amiram Harten, "High Resolution Schemes for Hyperbolic Conservation Laws", J. Comp. Phys., 49, 1983, 357–393.

[10] M. S. Liou and C. J. Steffen, "A New Flux Splitting Scheme", J. Comp. Phys., 107, 1993, 22–39.

[11] Antony Jameson "Analysis and Design of Numerical Schemes for Gas Dynamics 1 Artificial Diffusion, Upwind Biasing, Limiters and Their Effect on Accuracy and Multigrid Convergence", International Journal of Computational Fluid Dynamics, 4, 1995, 171–218.

[12] Antony Jameson "Analysis and Design of Numerical Schemes for Gas Dynamics 2 Artificial Diffusion and Discrete Shock Structure", International Journal of Computational Fluid Dynamics, 5, 1995, 1–38.

[13] S. K. Godunov, DAN USSR 139 1961, 521

[14] M. S. Mock, "Systems of Conservation Laws of Mixed Type", J. Differential Equations, 37, 1980.

[15] Amiram Harten, "On the Symmetric Form of Systems of Conservation Laws with Entropy", J. Comp. Phys., 49, 1983, 151–164.

[16] Ken Mattsson, "Boundary Operators for Summation-by-Parts Operators", J. Scientific Computing, 18, 2003, 133–153.

[17] C. W. Shu, "Total-Variation-Diminishing Time Discretizations", SIAM J. Sci. Statist. Computing, 9, 1988, 1073–1084.

[18] S. Gottlieb, C. W. Shu and E. Tadmor, "Strong Stability-Preserving High-Order Time Discretization Methods", SIAM Review, 2006.

[19] Antony Jameson, T. J. Baker, N.P. Weatherill, "Calculation of Inviscid Transonic Flow Over a Complete Aircraft", AIAA Paper 86–0103, AIAA 24th Aerospace Sciences Meeting, Reno, January 1986.

[20] F. Ham, K. Mattsson, G. Iaccarino and P. Moin, "Towards Time-Stable and Accurate LES on Unstructured Grids", Complex Effects in Large Eddy Simulation, Limassol, September 2005.

[21] T. J. 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", Computer Methods in Applied Mechanics and Engineering, 54, 1986, 223–234.

[22] Margot Gerritsen and Pelle Olsson, "Designing an Efficient Solution Strategy for Fluid Flows", J. Comp. Phys., 129, 1996, 245–262.

[23] H. C. Yee, M. Vinokur and M. J. Djomehri, "Entropy Splitting and Numerical Dissipation", J. Comp. Phys., 162, 2000, 33–81.

[24] R. W. MacCormack and A. J. Paullay, "The Influence of the Computational Mesh on Accuracy for Initial Value Problems with Discontinuous or Non-Unique Solutions", Computers and Fluids, 2, 1974, 339–361.

[25] A. Jameson, W. Schmidt and E. Turkel, "Numerical Solutions of the Euler Equations by Finite Volume Methods Using Runge–Kutta Time-Stepping Schemes", AIAA Paper 81–1259, AIAA 14th Fluid and Plasma Dynamic Conference, Palo Alto, June 1981.