A MODIFICATION OF ROE'S SCHEME FOR ENTROPY SATISFYING SOLUTIONS OF SCALAR NON-LINEAR CONSERVATION LAWS

P. K. SWEBY

NUMERICAL ANALYSIS REPORT 6/82

A MODIFICATION OF ROE'S SCHEME FOR ENTROPY SATISFYING SOLUTIONS OF SCALAR NON-LINEAR CONSERVATION LAWS

P. K. Sweby

Introduction

In Sweby & Baines [2] convergence of Roe's scheme for a scalar non-linear conservation law is proved. However Roe's method suffers from an inability to deal satisfactorily with expansion fans.

In the present report a special procedure is devised for dealing with expansion fans and this is described in detail below. The resulting scheme is then put in a more general framework, which reduces to the scheme in Sweby & Baines (loc cit) away from expansions.

Consider the scalar non-linear equation

$$u_t + f(u)_x = 0 \qquad -\infty < x < \infty,\ 0 < t \le T \tag{1a}$$

with initial data

$$u(x,0) = u_0(x), \tag{1b}$$

which is assumed to be (a) uniformly bounded in the L^∞ norm, and (b) of uniformly bounded variation, namely, for all real δ and $R > 0$

$$\int_{|x|<R} |u_0(x+\delta) - u_0(x)|\,dx \le C(R)\,|\delta|. \tag{2}$$

We divide the x axis uniformly into cells (x_{k-1}, x_k) with

$$x_k = x_{k-1} + h \tag{3}$$

and take a time step Δt, defining the constant q by

$$q = \frac{\Delta t}{h} \tag{4}$$

We now define the piecewise constant projection $u_h(x,t)$ of $u(x,t)$ onto the computational grid by

$$u_h(x,t) = u_k^n \tag{5}$$

for $(x,t) \in (x_{k-1},x_{k+\frac{1}{2}}) \times (\overline{n-\frac{1}{2}}qh,\overline{n+\frac{1}{2}}qh)$, where u_k^n are nodal values generated by a finite difference scheme. The initial data is projected by the restriction

$$u_k^0 = \frac{1}{h}\int_{x_{k-\frac{1}{2}}}^{x_{k+\frac{1}{2}}} u_o(x)\,dx \quad . \tag{6}$$

We now develop the scheme for generating the nodal values u_k^n. For brevity we adopt the notation

$$\left.\begin{aligned} u^k &= u_k^{n+1} \\ u_k &= u_k^n \end{aligned}\right\} \tag{7}$$

whenever there is no danger of confusion.

Due to the piecewise constant nature of $u_h(x,t)$ we can, for sufficiently small Δt, treat the advancement of the solution in each cell from $t = \tau$ to $t = \tau + \Delta t$ as the solution of the local scalar Riemann problem, i.e.

$$\left.\begin{aligned} &u_t + f(u)_x = 0 \qquad (x,t) \in (x_{k-1},x_k) \times (\tau,\tau+\Delta t) \\ &u(x,\tau) = \begin{cases} u_{k-1} & x < x_{k-\frac{1}{2}} \\ u_k & x > x_{k-\frac{1}{2}} \end{cases} \end{aligned}\right\} \tag{8}$$

with Δt chosen such that the characteristics do not intersect.

Let $a(x,t) = a(u(x,b)) = \frac{\partial f}{\partial u}$ be the wave speed and, to ease notation, make the following change of variables:

$$\left.\begin{aligned} X &= x - x_{k-\frac{1}{2}} \\ T &= t - \tau \end{aligned}\right\} \tag{9}$$

The problem (8) has solution (in terms of the new variables) [1]

$$u(X,T) = \begin{cases} u_{k-1} & X/T < a^L \\ u_k & X/T > a^R \end{cases} \tag{10}$$

where a^L, a^R are the left and right wave speeds respectively, i.e.

$$a^L = a(x_{k-1},\tau)$$

$$a^R = a(x_k,\tau)$$

and $$-\infty < a^L \le a^R < \infty \ .$$

If the discontinuity is a shock then $a^L = a^R = a(x_{k-\frac{1}{2}},\tau)$ with

$$a(x_{k-\frac{1}{2}},\tau) = \frac{f(u_k) - f(u_{k-1})}{u_k - u_{k-1}} \tag{11}$$

being the shock speed given by the Rankine-Hugoniot relation (see Fig. 1).

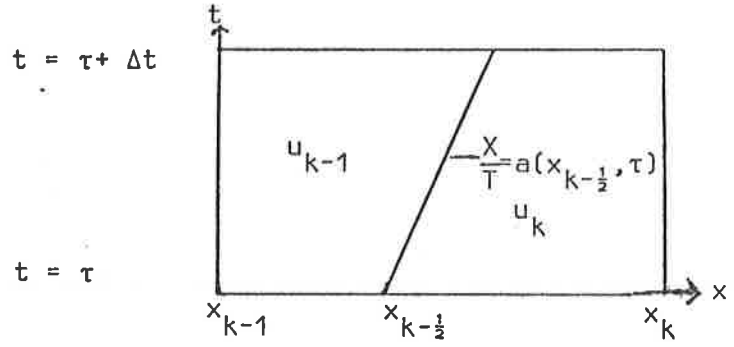


Fig. 1

If, however, $a^L < a(x_{k-\frac{1}{2}},\tau) < a^R$ then the correct physical solution is an expansion wave [1], i.e.

$$u(X,T) = \begin{cases} u_{k-1} & X/T \le a^L \\ u_{k-1} + \theta(u_k - u_{k-1}) & a^L < X/T < a^R \\ u_k & X/T \ge a^R \end{cases} \tag{12}$$

where

$$\theta = \frac{(X/T - a^L)}{(a^R - a^L)} \tag{13}$$

(see Fig. 2).

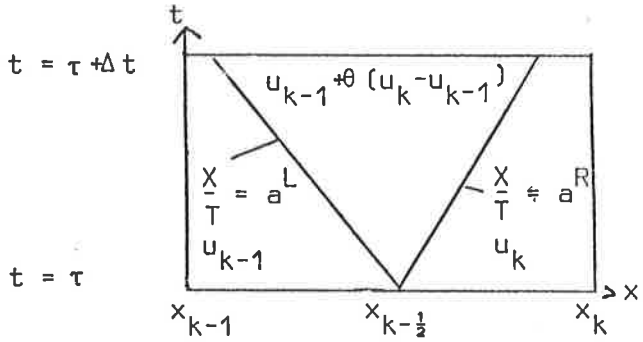


Fig. 2

We now project the piecewise linear solution (12) onto the space of piecewise constant functions.

2. Approximation of the Expansion

The linear solution (12) is approximated by the piecewise constant solution

$$u(X,T) = \begin{cases} u_{k-1} & X/T \le a^L \\ u_{k-\frac{1}{2}} & a^L < X/T < a^R \\ u_k & X/T \ge a^R \end{cases} \tag{14}$$

To obtain $u_{k-\frac{1}{2}}$ we integrate (1a) between x_{k-1} and x_k at time t; i.e.

$$\frac{d}{dt}\int_{x_{k-1}}^{x_k} u(x,t)dx = -\{f(u(x_k,t)) - f(u(x_{k-1},t))\} . \tag{15}$$

Changing variables to X,T gives

$$\frac{d}{dT}\int_{-\frac{h}{2}}^{\frac{h}{2}} u(X,T)dX = -\{f(u_k) - f(u_{k-1})\} \tag{16}$$

Since u is piecewise constant we can evaluate the integral in (16) as

$$\int_{-\frac{h}{2}}^{\frac{h}{2}} u(X,T)dX = \int_{-\frac{h}{2}}^{a^L T} u_{k-1}dX + \int_{a^L T}^{a^R T} u_{k-\frac{1}{2}}dX + \int_{a^R T}^{\frac{h}{2}} u_k dX$$

$$= (a^L T + \tfrac{h}{2})u_{k-1} + (a^R T - a^L T)u_{k-\frac{1}{2}} + (\tfrac{h}{2} - a^R T)u_k . \tag{17}$$

So carrying out the differentiation in (16) we obtain

$$a^L(u_{k-\frac{1}{2}} - u_{k-1}) + a^R(u_k - u_{k-\frac{1}{2}}) = f(u_k) - f(u_{k-1})$$

$$= a\{u_k - u_{k-1}\} , \tag{18}$$

with a as defined in (11). Thus

$$(a^R - a^L)u_{k-\frac{1}{2}} = (a - a^L)u_{k-1} + (a^R - a)u_k . \tag{19}$$

Next we define numerical approximations $a^L_{k-\frac{1}{2}}$, $a^R_{k-\frac{1}{2}}$, $a_{k-\frac{1}{2}}$ for a^L, a^R and a, respectively, by

$$a_{k-\frac{1}{2}} = \begin{cases} \dfrac{f_k - f_{k-1}}{u_k - u_{k-1}} & u_k \ne u_{k-1} \\ \text{approximation for } \left.\dfrac{\partial f}{\partial u}\right|_{x_{k-\frac{1}{2}}} & u_k = u_{k-1} \end{cases} \tag{20}$$

$$a^L_{k-\frac{1}{2}} = \min\{a_{k-1}, a_{k-\frac{1}{2}}\}$$
$$a^R_{k-\frac{1}{2}} = \max\{a_{k-\frac{1}{2}}, a_k\}, \qquad (21)$$

where a_k, a_{k-1} are approximations for $\frac{\partial f}{\partial u}(u_k)$, $\frac{\partial f}{\partial u}(u_{k-1})$, respectively. Note that the minimum and maximum in (21) ensure that $a^L_{k-\frac{1}{2}} \le a_{k-\frac{1}{2}} \le a^R_{k-\frac{1}{2}}$. We now have the situation as shown in Figure 3, and can devise a difference scheme as follows.

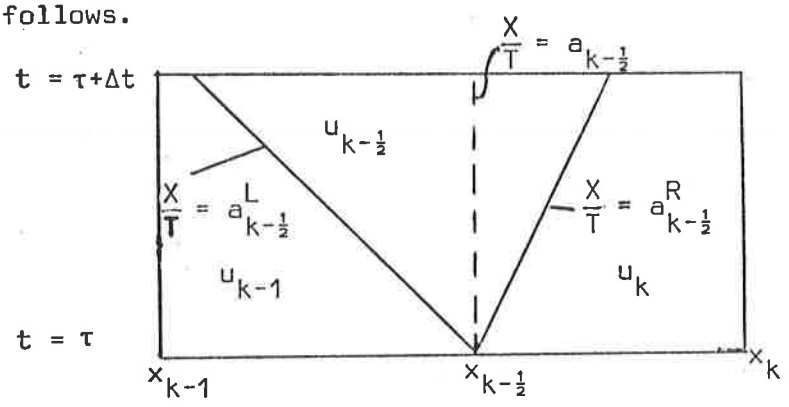


Fig. 3

3. The Finite Difference Scheme

We define left- and right-moving increments $\phi^L_{k-\frac{1}{2}}$ and $\phi^R_{k-\frac{1}{2}}$ as follows;

$$\phi^L_{k-\frac{1}{2}} = \frac{\Delta}{h}\int_{x_{k-1}}^{x_{k-\frac{1}{2}}} u\,dx, \quad \phi^R_{k-\frac{1}{2}} = \frac{\Delta}{h}\int_{x_{k-\frac{1}{2}}}^{x_k} u\,dx, \qquad (22)$$

with a total fluctuation for the cell of

$$\phi_{k-\frac{1}{2}} = \frac{\Delta}{h}\int_{x_{k-1}}^{x_k} u\,dx = -\frac{1}{h}\int_{\tau}^{\tau+\Delta t}\int_{x_{k-1}}^{x_k} fx\,dx\,dt$$

(arising from integrating (1) over the cell between $t = \tau$ on $t = \tau+\Delta t$)

$$= -\frac{1}{h}\int_{\tau}^{\tau+\Delta t}(f(u_k) - f(u_{k-1}))dt$$
$$= -\frac{\Delta t}{h}\, a_{k-\frac{1}{2}}\,(u_k - u_{k-1}) \qquad (23)$$

The assumption $\frac{\Delta t}{h}|f'(u)| \le \frac{1}{2}$ (24a)

and therefore $\frac{\Delta t}{h}|a| \le \frac{1}{2}$ (24b)

for $a = a^L_{k-\frac{1}{2}}, a^R_{k-\frac{1}{2}}, a_{k-\frac{1}{2}}$ is made throughout, ensuring that discontinuities from adjacent cells do not interact.

We now perform the integration for $\phi^L_{k-\frac{1}{2}}, \phi^R_{k-\frac{1}{2}}$ in (22) in a similar manner as for (16).

$$\phi^L_{k-\frac{1}{2}} = \frac{\Delta}{h}\int_{x_{k-1}}^{x_{k-\frac{1}{2}}} u\,dx = \frac{1}{h}\left\{\int_{x_{k-1}}^{x_{k-\frac{1}{2}}} u(x,t+\Delta t)dx - \int_{x_{k-1}}^{x_{k-\frac{1}{2}}} u(x,t)dx\right\}$$

$$= \frac{1}{h}\left\{u_k(-a^R_{k-\frac{1}{2}})^+\Delta t + u_{k-\frac{1}{2}}[(-a^L_{k-\frac{1}{2}})^+ - (-a^R_{k-\frac{1}{2}})^+]\Delta t + u_{k-1}(\frac{h}{2} - (-a^L_{k-\frac{1}{2}})^+\Delta t)\right\} - \frac{1}{h}\left\{u_{k-1}\frac{h}{2}\right\}$$

$$= q\left\{(-a^R_{k-\frac{1}{2}})^+(u_k - u_{k-\frac{1}{2}}) + (-a^L_{k-\frac{1}{2}})^+(u_{k-\frac{1}{2}} - u_{k-1})\right\}, \qquad (25)$$

where $b^+ = \frac{1}{2}(b + |b|)$ is the positive part of b.

Defining

$$\nu_{k-\frac{1}{2}} = q\,a_{k-\frac{1}{2}} \qquad (26)$$

we may now write (25) as

$$\phi^L_{k-\frac{1}{2}} = (-\nu^R_{k-\frac{1}{2}})^+(u_k - u_{k-\frac{1}{2}}) + (-\nu^L_{k-\frac{1}{2}})^+(u_{k-\frac{1}{2}} - u_{k-1}). \qquad (27a)$$

Similarly

$$\phi^R_{k-\frac{1}{2}} = -(\nu^R_{k-\frac{1}{2}})^+(u_k - u_{k-\frac{1}{2}}) - (\nu^L_{k-\frac{1}{2}})^+(u_{k-\frac{1}{2}} - u_{k-1}). \qquad (27b)$$

We observe that

$$\begin{aligned}\phi^L_{k-\frac{1}{2}} + \phi^R_{k-\frac{1}{2}} &= -\nu^R_{k-\frac{1}{2}}(u_k - u_{k-\frac{1}{2}}) - \nu^L(u_{k-\frac{1}{2}} - u_{k-1}) \\ &= -\nu_{k-\frac{1}{2}}(u_k - u_{k-1}) \qquad \text{from (18)} \\ &= \phi_{k-\frac{1}{2}}\,, \end{aligned} \qquad (28)$$

as would be expected due to conservation.

Note that if $\nu^R_{k-\frac{1}{2}} = \nu^L_{k-\frac{1}{2}} = \nu_{k-\frac{1}{2}}$ then

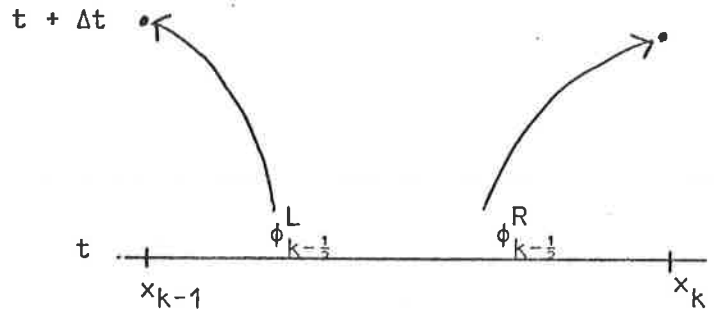
$$\left.\begin{aligned}&\phi^L_{k-\frac{1}{2}} = \phi_{k-\frac{1}{2}},\ \phi^R_{k-\frac{1}{2}} = 0 && \text{if } \nu_{k-\frac{1}{2}} < 0 \\ &\phi^L_{k-\frac{1}{2}} = 0,\ \phi^R_{k-\frac{1}{2}} = \phi_{k-\frac{1}{2}} && \text{if } \nu_{k-\frac{1}{2}} > 0\end{aligned}\right\} \qquad (29)$$

(i) First Order

Incrementing u_{k-1} by $\phi^L_{k-\frac{1}{2}}$ and u_k by $\phi^R_{k-\frac{1}{2}}$ over a time step Δt yields the first order scheme

$$u^k = u_k + \phi^L_{k+\frac{1}{2}} + \phi^R_{k-\frac{1}{2}}, \tag{30}$$

which is shown graphically in Figure 4.



The First Order Scheme

Fig. 4

We note from (29) that this scheme is equivalent to Roe's first order scheme [3] everywhere except at expansions.

(ii) Second Order

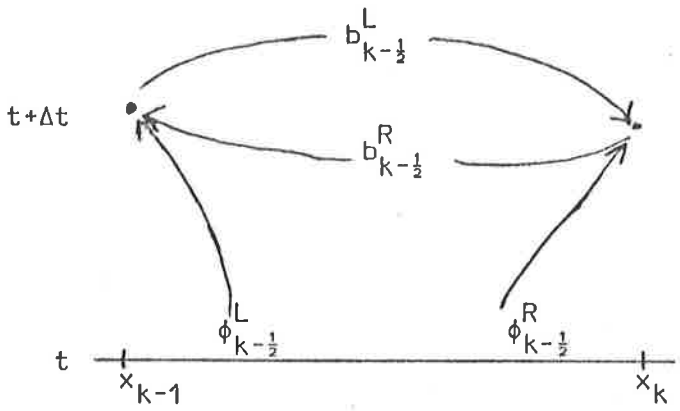
We now introduce an antidiffusion stage by transferring some of the allocated fluctuation [2]. The result is a second order scheme almost everywhere (see below).

Let
$$\alpha_{k-\frac{1}{2}} = \tfrac{1}{2}(1 - |\nu_{k-\frac{1}{2}}|) \tag{31}$$

$$\left.\begin{aligned} b^L_{k-\frac{1}{2}} &= \text{minmod}\,\{\alpha_{k-\frac{1}{2}}\,\phi^L_{k-\frac{1}{2}},\ \alpha_{k+\frac{1}{2}}\,\phi^L_{k+\frac{1}{2}}\} \\ b^R_{k-\frac{1}{2}} &= \text{minmod}\,\{\alpha_{k-\frac{1}{2}}\,\phi^L_{k-\frac{1}{2}},\ \alpha_{k-\frac{3}{2}}\,\phi^R_{k-\frac{3}{2}}\} \end{aligned}\right\} \tag{32}$$

where minmod $\{\cdot,\cdot\}$ selects the argument with minimum absolute value. (In the case of arguments with equal modulus but opposite signs the first argument is chosen).

At the antidiffusion stage the quantity $b^L_{k-\frac{1}{2}}$ is transferred left to right in the cell whilst $b^R_{k-\frac{1}{2}}$ is transferred right to left (see Fig. 5). Again except at expansions either $b^L_{k-\frac{1}{2}}$ or $b^R_{k-\frac{1}{2}}$ will be zero as a consequence of (29) and (32).



The Second Order Scheme

Fig. 5

This scheme can be written

$$u^k = u_k + \xi_{k+\frac{1}{2}} \phi^L_{k+\frac{1}{2}} + \zeta_{k-\frac{1}{2}} \phi^R_{k-\frac{1}{2}}, \tag{33}$$

where

$$\xi_{k+\frac{1}{2}} = \begin{cases} 1 + \dfrac{(b^L_{k-\frac{1}{2}} - b^L_{k+\frac{1}{2}})}{\phi^L_{k+\frac{1}{2}}} & \phi^L_{k+\frac{1}{2}} \neq 0 \\ 1 & \phi^L_{k+\frac{1}{2}} = 0 \end{cases}$$

$$\zeta_{k-\frac{1}{2}} = \begin{cases} 1 - \dfrac{(b^R_{k-\frac{1}{2}} - b^R_{k+\frac{1}{2}})}{\phi^R_{k-\frac{1}{2}}} & \phi^R_{k-\frac{1}{2}} \neq 0 \\ 1 & \phi^R_{k-\frac{1}{2}} = 0 \end{cases} \tag{34}$$

and is second order accurate except at discontinuities of u (see below).

The scheme (33), (34) is equivalent to Roe's second order scheme [3], except at expansions, as studied by Sweby & Baines [2]. If the minmod operators in (32) always choose the first argument then the scheme is the Lax-Wendroff scheme [4]. If the second argument is always chosen it is the Warming and Beam fully upwind scheme [5].

3. Proof of Second Order Accuracy

We now deomonstrate that the scheme (33), (34) is second order accurate away from discontinuities of the solution by comparing it with the Lax-Wendroff scheme.

A necessary and sufficient condition for a scheme formulated as in (33) to be second order accurate is that

$$(\xi_{k+\frac{1}{2}} - \xi^{LW}_{k+\frac{1}{2}})\ \phi^{L}_{k+\frac{1}{2}} + (\zeta_{k-\frac{1}{2}} - \zeta^{LW}_{k-\frac{1}{2}})\phi^{R}_{k-\frac{1}{2}} = O(h^2) \tag{35}$$

where $\xi^{LW}_{k+\frac{1}{2}}$, $\zeta^{LW}_{k-\frac{1}{2}}$ are the appropriate coefficients for the Lax-Wendroff scheme.

Except for the case when $\nu^{L}_{k-\frac{1}{2}} < 0 < \nu^{R}_{k-\frac{1}{2}}$, the Lax-Wendroff scheme is achieved by choosing the first argument in the minmod operators in (32) so it is only necessary to show that the arguments of the minmod operators differ by $O(h^2)$. That is we seek to show that

$$\alpha_{k+\frac{1}{2}}\ \phi^{L}_{k+\frac{1}{2}} - \alpha_{k-\frac{1}{2}}\ \phi^{L}_{k-\frac{1}{2}} = O(h^2) \tag{36}$$

and

$$\alpha_{k-3/2}\ \phi^{R}_{k-3/2} - \alpha_{k-\frac{1}{2}}\ \phi^{R}_{k-\frac{1}{2}} = O(h^2) \tag{37}$$

We shall verify only (36) here, but (37) may be verified in a similar manner. Assuming that u is smooth enough to be expanded in a Taylor series in x (and hence that f can be expanded as a function of x),

$$\begin{aligned}
\alpha_{k+\frac{1}{2}}\ \phi^{L}_{k+\frac{1}{2}} &= -\tfrac{1}{2}(1 - |\nu_{k+\frac{1}{2}}|)\ q\ (f_{k+1} - f_k) \\
&= -\tfrac{1}{2}(1 - |\nu_{k-\frac{1}{2}}| + O(h))\ q\ (h\ \frac{\partial f}{\partial x}\Big|_k + O(h^2)) \\
&= -(\alpha_{k-\frac{1}{2}} + O(h))\ q\ (h\ \frac{\partial f}{\partial x}\Big|_k + O(h^2)) \\
&= -\alpha_{k-\frac{1}{2}}\ q\ h\ \frac{\partial f}{\partial x}\Big|_k + O(h^2)
\end{aligned}$$

But also $\alpha_{k-\frac{1}{2}}\ \phi^{L}_{k-\frac{1}{2}} = -\alpha_{k-\frac{1}{2}}\ qh\ \frac{\partial f}{\partial x}\Big|_k + O(h^2)$

Hence $\alpha_{k+\frac{1}{2}}\ \phi^{L}_{k+\frac{1}{2}} - \alpha_{k-\frac{1}{2}}\ \phi^{L}_{k-\frac{1}{2}} = O(h^2)$ verifying (36).

At discontinuities of u the Taylor expansion is invlaid and hence we cannot claim second order accuracy at such points. Note, however, that the scheme is still second order at points where the minmod selection in (32) changes.

In cells where $\nu^{L}_{k-\frac{1}{2}} < 0 < \nu^{R}_{k-\frac{1}{2}}$ it is well known that the Lax-Wendroff scheme may produce a non-physical solution, and thus we cannot use this comparison

to claim second order accuracy at such points.

4. Convergence of the Scheme

We now consider convergence of the scheme after first giving some useful definitions.

$$\mu^s_{k-\frac{1}{2}} = \left\{ \begin{array}{ll} -\phi^s_{k-\frac{1}{2}}/(u_k - u_{k-1}) & u_k \neq u_{k-1} \\ \left\{ \begin{array}{ll} (-\nu_{k-\frac{1}{2}})^+ & s = L \\ (\nu_{k-\frac{1}{2}})^+ & s = R \end{array} \right\} & u_k = u_{k-1} \end{array} \right\} \quad s = L,R \tag{38}$$

$$\beta_{k-\frac{1}{2}} = \left\{ \begin{array}{ll} b^s_{k-\frac{1}{2}}/\alpha_{k-\frac{1}{2}}\ \phi^s_{k-\frac{1}{2}} & \phi^s_{k-\frac{1}{2}} \neq 0 \\ 1 & \phi^s_{k-\frac{1}{2}} = 0 \end{array} \right\} \quad s = L,R \tag{39}$$

$$\gamma^s_{k-\frac{1}{2}} = \left\{ \begin{array}{ll} b^s_{k-\frac{1}{2}}/\alpha_{k+\frac{1}{2}}\ \phi^s_{k+\frac{1}{2}} & \phi^s_{k+\frac{1}{2}} \neq 0 \\ 1 & \phi^s_{k+\frac{1}{2}} = 0 \end{array} \right\} \quad s = L,R \tag{40}$$

We observe, from (24) and (31), (32), (39), (40), the inequalities

$$\left. \begin{array}{l} 0 \leq \alpha_{k-\frac{1}{2}} \leq \frac{1}{2} \\ |\beta^s_{k-\frac{1}{2}}| \leq 1, \quad |\gamma^s_{k-\frac{1}{2}}| \leq 1 \qquad s = L,R \end{array} \right. \tag{41}$$

and from (24) and (27), (38),

$$\left. \begin{array}{l} 0 \leq \mu^R_{k-\frac{1}{2}} \leq \frac{1}{2} \\ 0 \leq -\mu^L_{k-\frac{1}{2}} \leq \frac{1}{2}. \end{array} \right\} \tag{42}$$

For convergence we require a uniform bound on the variation of the solution and also a uniform bound on the solution itself [2]. Sufficient conditions for the total variation to be non-increasing, and hence the variation to be bounded due to (2) are:

$$\left. \begin{array}{l} 0 \leq -\xi_{k-\frac{1}{2}}\ \mu^L_{k-\frac{1}{2}} \\ 0 \leq \zeta_{k-\frac{1}{2}}\ \mu^R_{k-\frac{1}{2}} \end{array} \right. \tag{43a}$$

$$0 \leq \zeta_{k-\frac{1}{2}}\ \mu^R_{k-\frac{1}{2}} - \xi_{k-\frac{1}{2}}\ \mu^L_{k-\frac{1}{2}} \leq 1, \tag{43b}$$

see [2].

The criterion we use for bounding the solution is

$$\inf\{u_{k-1},u_k,u_{k+1}\} \le u^k \le \sup\{u_{k-1},u_k,u_{k+1}\} \tag{44}$$

which ensures that, as well as preserving monotonicity of the data (also implied by a non-increasing total variation), maxima will not increase nor minima decrease. This implies in turn that the solution is bounded by the initial data in the L^∞ norm [2]. An additional sufficient condition to (43a) for (44) to hold is

$$0 \le \zeta_{k-\frac{1}{2}}\, \mu^R_{k-\frac{1}{2}} - \xi_{k+\frac{1}{2}}\, \mu^L_{k+\frac{1}{2}} \le 1. \tag{45}$$

[If both (45) and (43a) hold the solution will also be bounded in the ℓ_1 norm by the initial data (see [7])]

We note that, using (39), (40), equation (34) may be rewritten as

$$\begin{aligned} \xi_{k+\frac{1}{2}} &= 1 + (\gamma^L_{k-\frac{1}{2}} - \beta^L_{k+\frac{1}{2}})\, \alpha_{k+\frac{1}{2}} \\ \zeta_{k-\frac{1}{2}} &= 1 - (\beta^R_{k-\frac{1}{2}} - \gamma^L_{k+\frac{1}{2}})\, \alpha_{k+\frac{1}{2}} \end{aligned} \tag{46}$$

The inequalities (43a) then follow from (41).

Away from sign changes of $f'(u)$ either ϕ^L or ϕ^R is zero. Consequently either μ^L or μ^R is zero, and the inequalities (43b) and (45) are easily verified [2]. It only remains to verify (43b) and (45) for expansions and shocks. This is done in (i) and (ii) below.

(i) Expansions

Consider first an expansion, the scheme for which is shown graphically in Fig. 6

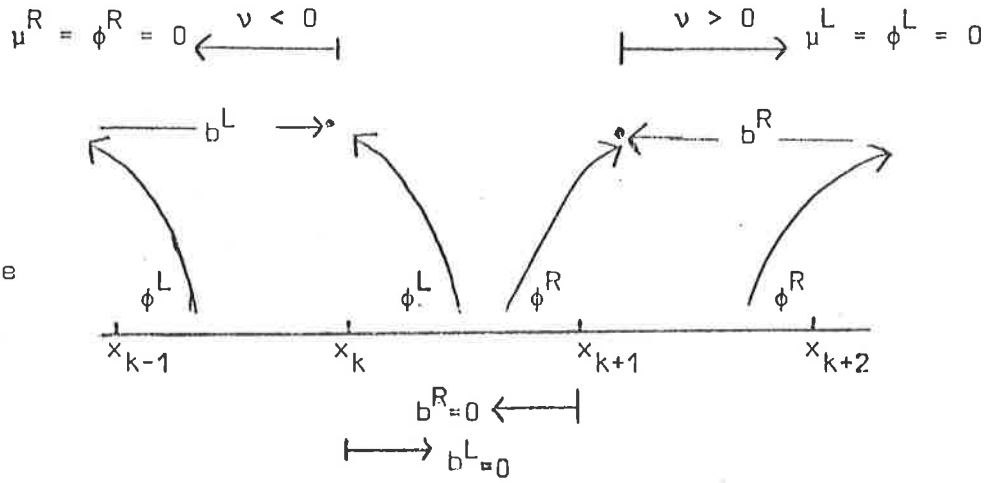


Fig. 6
The Scheme for an expansion

It is easily seen that (45) holds. Now consider

$$\zeta_{k+\frac{1}{2}}\ \mu^{R}_{k+\frac{1}{2}} - \xi_{k+\frac{1}{2}}\ \mu^{L}_{k+\frac{1}{2}} = g(\nu_{k+\frac{1}{2}}), \text{ say.} \tag{47}$$

Clearly, from (43a), $g(\nu_{k+\frac{1}{2}}) \geq 0$.

Since $b^{L}_{k+\frac{1}{2}} = b^{R}_{k+\frac{1}{2}} = 0$ and $u_k \neq u_{k+1}$ for an expansion we may write

$$g(\nu_{k+\frac{1}{2}}) = (1 + \gamma^{L}_{k-\frac{1}{2}}\ \alpha_{k+\frac{1}{2}})\ \frac{\phi^{L}_{k+\frac{1}{2}}}{u_{k+1}-u_k} - (1 + \gamma^{R}_{k+\frac{3}{2}}\ \alpha_{k+\frac{1}{2}})\ \frac{\phi^{R}_{k+\frac{1}{2}}}{u_{k+1}-u_k}$$

$$= -\nu_k(1 + \gamma^{L}_{k-\frac{1}{2}}\ \alpha_{k+\frac{1}{2}})\frac{(u_{k+\frac{1}{2}}-u_k)}{(u_{k+1}-u_k)} + \nu_{k+1}(1 + \gamma^{R}_{k+\frac{3}{2}}\ \alpha_{k+\frac{1}{2}})\frac{(u_{k+1}-u_{k+\frac{1}{2}})}{(u_{k+1}-u_k)}. \tag{48}$$

Using (19) we obtain

$$\frac{u_{k+\frac{1}{2}} - u_k}{u_{k+1} - u_k} = \frac{\nu_{k+1} - \nu_{k+\frac{1}{2}}}{\nu_{k+1} - \nu_k} \tag{49}$$

and

$$\frac{u_{k+1} - u_{k+\frac{1}{2}}}{u_{k+1} - u_k} = \frac{\nu_{k+\frac{1}{2}} - \nu_k}{\nu_{k+1} - \nu_k}.$$

Again for an expansion $\nu_{k+1} \neq \nu_k$, so the denominators are non-zero.

Substituting in (48) yields

$$g(\nu_{k+\frac{1}{2}}) = -\nu_k\ \frac{(\nu_{k+1}-\nu_{k+\frac{1}{2}})}{(\nu_{k+1}-\nu_k)}\ (1+\gamma^{L}_{k-\frac{1}{2}}\ \alpha_{k+\frac{1}{2}}) + \nu_{k+1}\frac{(\nu_{k+\frac{1}{2}}-\nu_k)}{(\nu_{k+1}-\nu_k)}\ (1+\gamma^{R}_{k+\frac{3}{2}}\alpha_{k+\frac{1}{2}})$$

$$\leq \frac{3}{2}\ \left\{\frac{-\nu_k(\nu_{k+1}-\nu_{k+\frac{1}{2}}) + \nu_{k+1}(\nu_{k+\frac{1}{2}}-\nu_k)}{\nu_{k+1} - \nu_k}\right\}, \tag{50}$$

(using (41)), and therefore

$$\leq \max\ (-\tfrac{3}{2}\nu_k, \tfrac{3}{2}\nu_{k+1}) \leq 1, \tag{51}$$

since (50) is linear in $\nu_{k+\frac{1}{2}}$.

Thus (43b) holds, which completes the conditions for expansions.

(ii) Shocks

Next consider a shock, the scheme for which is illustrated in Fig. 7.

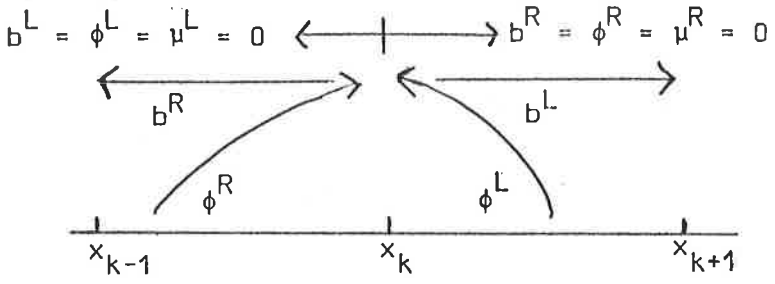


Fig. 7

The Scheme for a Shock

We have immediately that

$$0 \le \zeta_{k\pm\frac{1}{2}}\, \mu^R_{k\pm\frac{1}{2}} - \xi_{k\pm\frac{1}{2}}\, \mu^L_{k\pm\frac{1}{2}} \le 1 ,$$

verifying (43b), and also that

$$0 \le \zeta_{k-\frac{1}{2}}\, \mu^R_{k-\frac{1}{2}} - \xi_{k+\frac{1}{2}}\, \mu^L_{k+\frac{1}{2}} .$$

Moreover

$$\zeta_{k-\frac{1}{2}}\, \mu^R_{k-\frac{1}{2}} - \xi_{k+\frac{1}{2}}\, \mu^L_{k+\frac{1}{2}} = (1 - \beta^R_{k-\frac{1}{2}}\, \alpha_{k-\frac{1}{2}})\nu_{k-\frac{1}{2}} - (1 - \beta^L_{k-\frac{1}{2}}\, \alpha_{k-\frac{1}{2}})\nu_{k+\frac{1}{2}}$$

$$\le (1 + \alpha_{k-\frac{1}{2}})\nu_{k-\frac{1}{2}} - (1 + \alpha_{k+\frac{1}{2}})\nu_{k+\frac{1}{2}}$$

$$\le 1 \qquad \text{using (31), if } \left|\nu_{k\pm\frac{1}{2}}\right| < 0.35 \text{ approximately}$$

This result indicates that the inequality (45) may be violated if $0.35 < \left|\nu_{k\pm\frac{1}{2}}\right| < 0.5$. On close inspection we see that this only happens in the extreme case of oscillatory data with $|\nu| \sim \frac{1}{2}$ adjacent to a shock. Although this is unlikely to occur for most convex f, we may achieve (45) for certain by selecting the sign of the first argument of the minmod operator adjacent to a shock.

Finally from the inequalities in (i), (ii) above we may use compactness arguments [2], [7] to show that we may select from $\{u_h\}$ a convergent subsequence, and have therefore shown the convergence of the scheme (33), (34).

5. Entropy for the Semi-Discrete First Order Scheme

The first order semi-discrete scheme may be written as

$$\frac{\partial u_k^n}{\partial t} + \frac{1}{\Delta x}\Delta_-[f(u_{k+1}^n)(1-\theta_{k+\frac{1}{2}}^n) + \theta_{k+\frac{1}{2}}^n f(u_k^n)] = 0 \tag{52}$$

where
$$\theta_{k+\frac{1}{2}}^n = \begin{cases} 1 & \nu > 0 \text{ in } (u_k, u_{k+1}) \\ 0 & \nu < 0 \text{ in } (u_k, u_{k+1}) \\ \dfrac{\nu_{k+\frac{1}{2}}-\nu_k}{\nu_{k+1}-\nu_k}\dfrac{\nu_{k-1}}{\nu_{k+\frac{1}{2}}} & \nu_k < 0 < \nu_{k+1} \\ \frac{1}{2}(1+\operatorname{sgn}(\nu_{k-\frac{1}{2}})) & \nu_{k+1} < 0 < \nu_k \end{cases} \tag{53}$$

We follow Osher [8] and multiply by $u_k^n \psi_k^n \Delta x$ where ψ_k^n is a positive test function, and sum over k, n, giving

$$\sum_k \{\psi_k^n u_k^n \frac{\partial u_k^n}{\partial t} + \frac{1}{\Delta x}\psi_k^n u_k^n \Delta_-[f(u_{k+1}^n)(1-\theta_{k+\frac{1}{2}}^n) + \theta_{k+\frac{1}{2}}^n f(u_k^n)]\}\Delta x = 0, \tag{54}$$

or

$$\sum_k \{\psi_k^n \frac{\partial}{\partial t}(\tfrac{1}{2}u_k^{n2}) + \frac{1}{\Delta x}\psi_k^n u_k^n [\Delta_+ f_k^n - \Delta_-\{\theta_{k+\frac{1}{2}}^n \Delta_+ f_k^n\}]\}\Delta x = 0 \tag{55}$$

As in Osher's paper we now add and subtract a term

$$\frac{1}{\Delta x}\psi_k^n \int_{u_k^n}^{u_{k+1}^n} sf'(s)ds$$

and sum the Δ_- term by parts over k to give

$$\sum_k \{\psi_k^n \frac{\partial}{\partial t}(\tfrac{1}{2}u_k^{n2}) - \frac{1}{\Delta x}(\Delta_+\psi_k^n)\int_{\bar{u}}^{u_{k+1}^n} sf'(s)\,ds$$

$$+ \frac{1}{\Delta x}(\Delta_+\psi_k^n)u_{k+1}^n\,\theta_{k+\frac{1}{2}}^n\,\Delta_+ f_k^n - \frac{1}{\Delta x}\psi_k^n \int_{u_k^n}^{u_{k+1}^n}[s-u_k^n-(\Delta_+ u_k^n)\theta_{k+\frac{1}{2}}^n]\,f'(s)\,ds\}\Delta x = 0, \tag{56}$$

where $f'(\bar{u}) = 0$.

Note that the third term goes to zero with Δx. After integration by parts we obtain

$$\sum_k \{\psi_k^n \frac{\partial}{\partial t}(\tfrac{1}{2}u_k^{n2}) - \frac{1}{\Delta x}(\Delta_+\psi_k^n)\int_{\bar{u}}^{u_{k+1}^n} sf'(s)\,ds\}\Delta x$$

$$= +\sum_k \frac{1}{\Delta x}\psi_k^n \int_{u_k^n}^{u_{k+1}^n}[s - u_k^n - (\Delta_+ u_k^n)\theta_{k+\frac{1}{2}}]f'(s)\,ds\,\Delta x. \tag{57}$$

As $\Delta x,\ \Delta t \to 0$ the left hand side tends to

$$\int \{\psi \frac{\partial}{\partial t} (\tfrac{1}{2}u^2) - \psi_x \int_{\bar{u}}^{u} sf'(s)\, ds\}dx \quad . \tag{58}$$

We now consider the sign of the remaining term by examining the integral (dropping n superscripts) in (57), namely,

$$I_{k+\frac{1}{2}} = \int_{u_k}^{u_{k+1}} [s - u_k - \theta_{k+\frac{1}{2}} \Delta_+ u_k]f'(s)\, ds, \tag{59}$$

for the possible values of $\theta_{k+\frac{1}{2}}$.

If $f'(s)$ is of constant sign in (u_k, u_{k+1}) then $\theta_{k+\frac{1}{2}} = 0$ or 1. Take the case $f' > 0$, $\theta_{k+\frac{1}{2}} = 1$: then

$$I_{k+\frac{1}{2}} = \int_{u_k}^{u_{k+1}} [s - u_{k+1}]f'(s)\, ds \le 0,$$

and similarly, for $f' < 0$, $\theta_{k+\frac{1}{2}} = 0$,

$$I_{k+\frac{1}{2}} = \int_{u_k}^{u_{k+1}} [s - u_k]\, f'(s)\, ds \le 0.$$

If, however, $f'(u_k) > 0 > f'(u_{k+1})$ we have $u_k > \bar{u} > u_{k+1}$ (since f is convex) and

$$\theta_{k+\frac{1}{2}} = \begin{cases} 1 \text{ if } \dfrac{\Delta_+ f_k}{\Delta_+ u_k} = \dfrac{\int_{u_k}^{u_{k+1}} f'(s)\, ds}{\Delta_+ u_k} \ge 0 \\ 0 \text{ if } \dfrac{\Delta_+ f_k}{\Delta_+ u_k} = \dfrac{\int_{u_k}^{u_{k+1}} f'(s)\, ds}{\Delta_+ u_k} < 0 \end{cases} \tag{60}$$

i.e.

$$\theta_{k+\frac{1}{2}} = \begin{cases} 1 \text{ if } \int_{u_{k+1}}^{u_k} f'(s)\, ds \ge 0 \\ 0 \text{ if } \int_{u_{k+1}}^{u_k} f'(s)\, ds < 0 \end{cases} \tag{61}$$

Taking the first of these as an example:

$$I_{k+\frac{1}{2}} = \int_{u_{k+1}}^{u_k} [u_{k+1} - s]f'(s)\, ds$$

$$= [(u_{k+1} - s)f(s)]_{u_{k+1}}^{u_k} + \int_{u_{k+1}}^{u_k} f(s)\, ds$$

$$= (u_{k+1} - u_k)f(u_k) - (u_{k+1} - u_k)f(\xi) \qquad \xi \in (u_{k+1}, u_k)$$

$$= (u_{k+1} - u_k)(f(u_k) - f(\xi))$$

$$\leq 0 \quad \text{since} \quad \int_{\xi}^{u_k} f'(s)\, ds \geq \int_{u_{k+1}}^{u_k} f'(s)\, ds \geq 0$$

If $\int_{u_{k+1}}^{u_k} f'(s)\, ds < 0$ a similar argument gives $I_{k+\frac{1}{2}} \leq 0$.

The remaining case is $f'(u_{k+1}) > 0 > f'(u_k)$ with $u_{k+1} > \bar{u} > u_{k+1}$.

Here $\theta_{k+\frac{1}{2}} = \dfrac{\nu_{k+\frac{1}{2}} - \nu_k}{\nu_{k+1} - \nu_k} \dfrac{\nu_{k+1}}{\nu_{k+\frac{1}{2}}}$ with $\nu_k \leq \nu_{k+\frac{1}{2}} \leq \nu_{k+1}$

so, if $\int_{u_k}^{u_{k+1}} f'(s)\, ds \geq 0$, $\theta_{k+\frac{1}{2}} \geq 1$:

we have $\int_{u_k}^{u_{k-1}} f'(s)\, ds < 0$, $\theta_{k+\frac{1}{2}} \leq 0$, which gives

$$I_{k+\frac{1}{2}} = \int_{u_k}^{u_{k+1}} [s-c]\, f'(s)\, ds, \quad \text{where} \quad c \geq u_{k+1} \quad \text{if} \quad \int_{u_k}^{u_{k+1}} f'(s)\, ds \geq 0 \qquad (62)$$

$$c \leq u_k \quad \text{if} \quad \int_{u_k}^{u_{k+1}} f'(s)\, ds < 0.$$

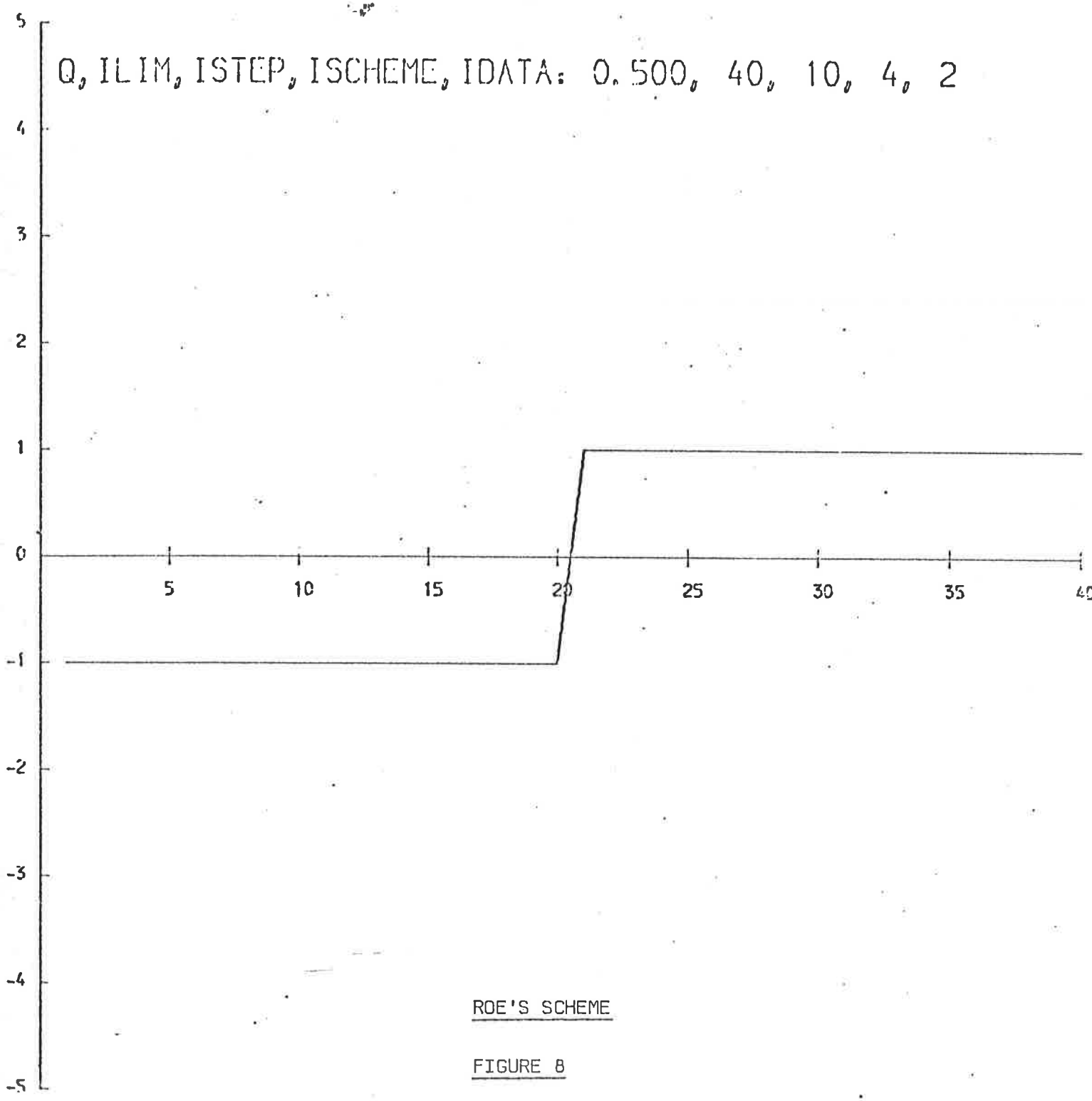
Taking the case $\int_{u_k}^{u_{k+1}} f'(s)\, ds \geq 0$, $c \geq u_{k+1}$, we have

$$I_{k+\frac{1}{2}} = \int_{u_k}^{u_{k+1}} [s-c]f'(s)\, ds$$

$$= [(s-c)f(s)]_{u_k}^{u_{k+1}} - \int_{u_k}^{u_{k+1}} f(s)\, ds$$

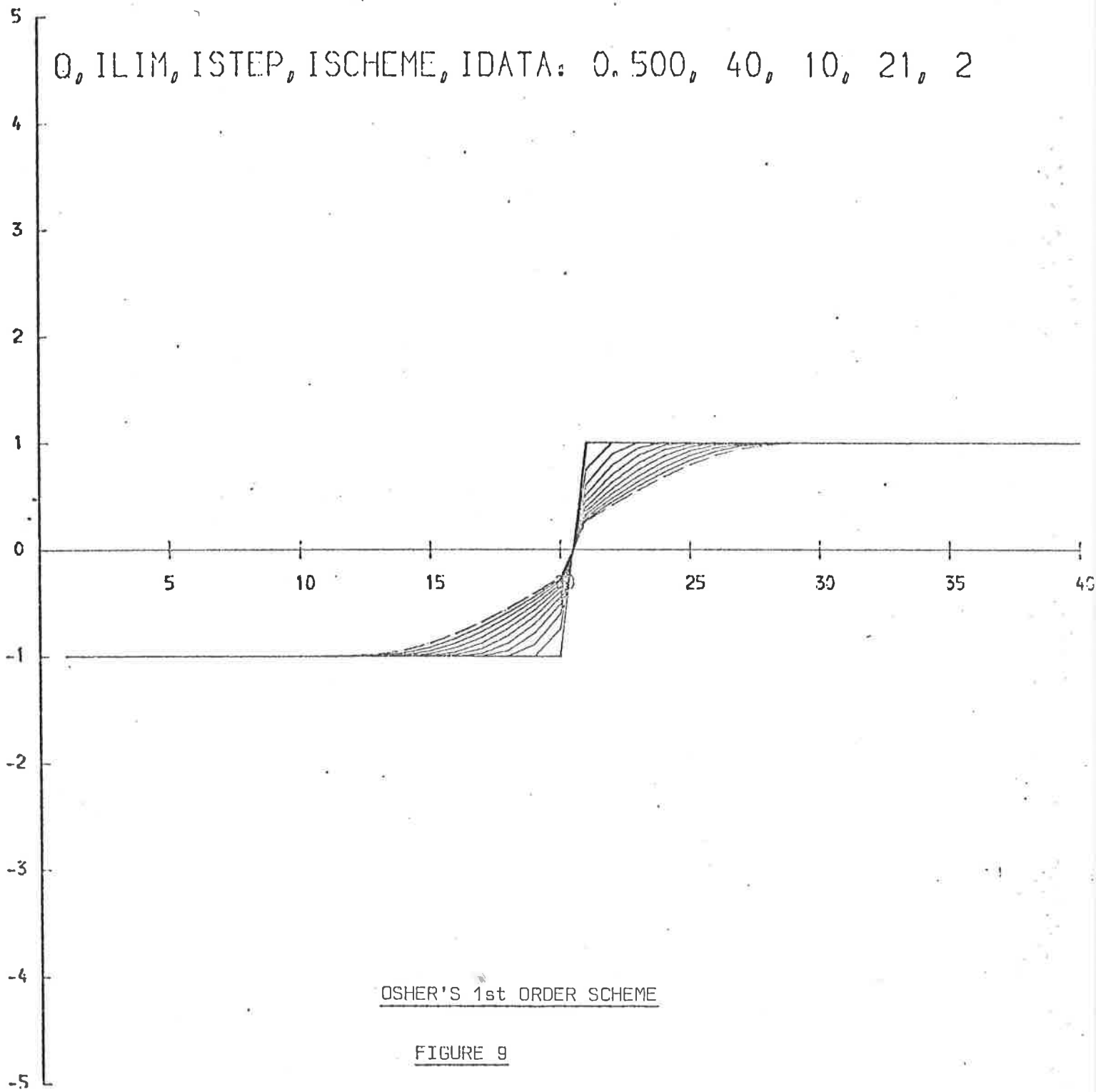
$$= (u_{k+1}-c)f(u_{k+1}) - (u_k-c)f(u_k) - (u_{k+1}-u_k)f(\xi) \quad \xi \in (u_k, u_{k+1})$$

$$= (u_k-c)\{f(u_{k+1})-f(u_k)\} + (u_{k+1}-u_k)(f(u_{k+1}) - f(\xi)). \qquad (63)$$



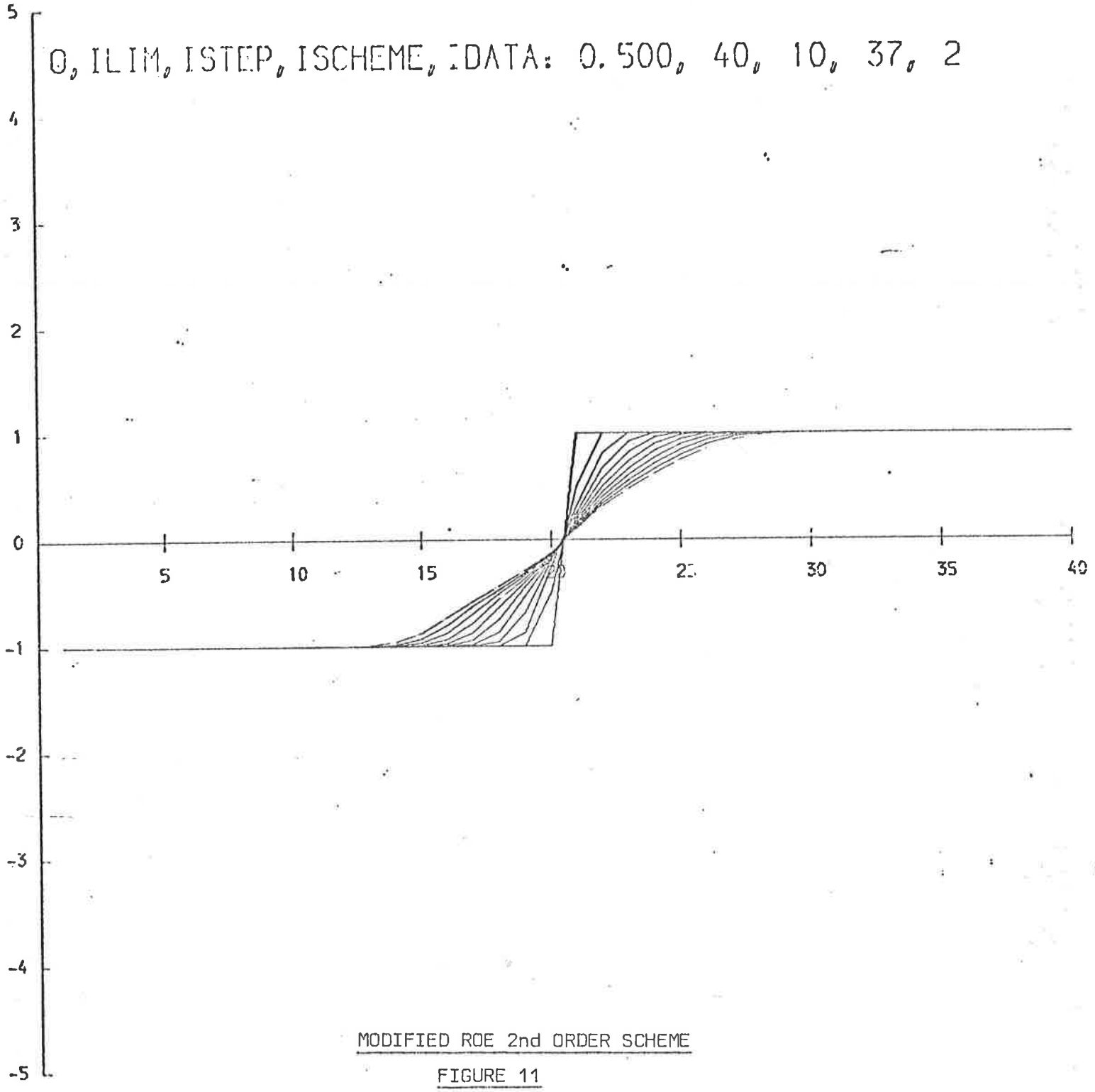
ROE'S SCHEME

FIGURE 8



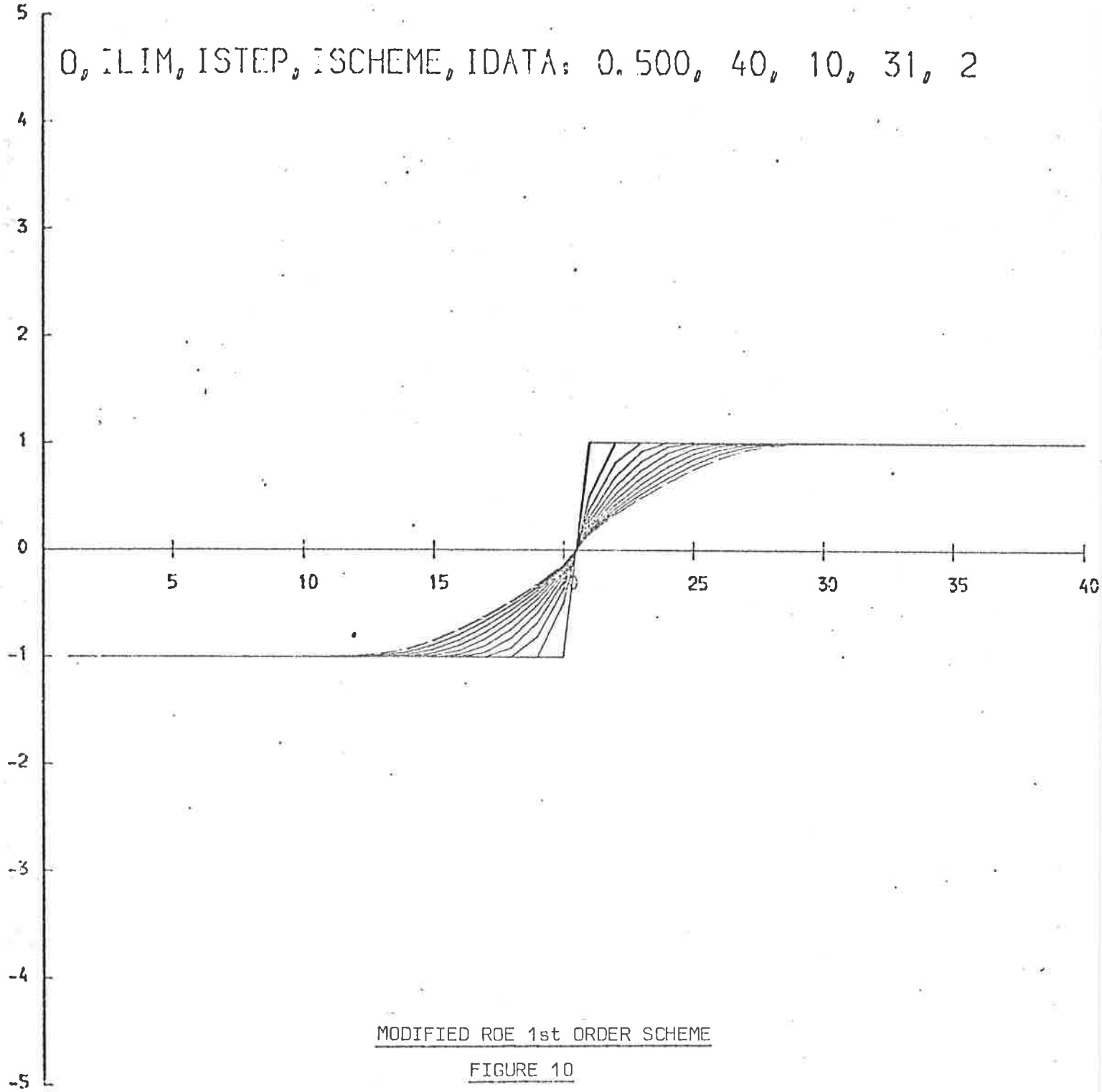
OSHER'S 1st ORDER SCHEME

FIGURE 9



MODIFIED ROE 2nd ORDER SCHEME

FIGURE 11



MODIFIED ROE 1st ORDER SCHEME

FIGURE 10

The first term is negative, whilst the second is positive since $\int_{\xi}^{u_{k+1}} f'(s)\,ds \geq \int_{u_k}^{u_{k+1}} f'(s)\,ds \geq 0$. Comparing absolute values of the terms,

$$\left| \frac{(u_{k+1}-u_k)(f(u_{k+1})-f(\xi))}{(c-u_k)(f(u_{k-1})-f(u_k))} \right| \leq 1, \tag{64}$$

since $c \geq u_{k+1}$: hence the first term is dominant giving $I_{k+\frac{1}{2}} \leq 0$. Similarly if $\int_{u_k}^{u_{k-1}} f'(s)\,ds \leq 0$, $c \leq u_k$ we again have $I_{k+\frac{1}{2}} \leq 0$.

Note that if $\nu_{k+\frac{1}{2}} = 0$ we may consider $\theta_{k+\frac{1}{2}}$ indefinity large, i.e. $c >> u_{k+1}$, for the purposes of this proof: in the scheme itself $\theta_{k+\frac{1}{2}}\,\nu_{k+\frac{1}{2}}$ is finite.

Collecting these results together, we have from (59)

$$\int \{\psi \frac{\partial}{\partial t}(\tfrac{1}{2}u^2) - \psi_x \int_{\bar{u}}^{u} sf'(s)\,ds\}\,dx \leq 0,$$

that is, in the sense of distributions,

$$\frac{\partial V(u)}{\partial t} + \frac{\partial G(u)}{\partial x} \leq 0.$$

where $V(u) = \frac{1}{2}u^2$ is the entropy function, and $G(u) = \int_{\bar{u}}^{u} V'(s)f'(s)\,ds$ is the entropy flux.

It follows that the solutions of the first order scheme converge to the correct entropy satisfying solution of the equation

$$u_t + f(u)_x = 0$$

for convex $f(u)$. (See [6]).

5. Results and Conclusions

Comparisons of Roe's Original Scheme [3], Osher's first order Scheme [8], and the modified Roe 1st and 2nd order schemes described here, are shown for a centred rarefaction in Figs. 8-11, respectively. The test equation used is

$$u_t + (\tfrac{1}{2}u^2)_x = 0.$$

It can be seen that Roe's original scheme produces an entropy violating solution, whilst Osher's, although giving the correct physical solution,

contains a large "dog-leg" compared to the two modified Roe schemes.

Acknowledgements

I would like to thank Dr. M.J. Baines of the University of Reading and Prof. S. Osher of the University of California, Los Angeles for invaluable discussions. I would also like to acknowledge the financial support of the Science & Engineering Research Council and the Royal Aircraft Establishment, Farnborough.

References

[1] Lax, P.D. Hyperbolic Conservation Laws and Mathematical Theory of Shock Waves, 1972, SIAM Regional Conference Series, Lectures in Applied Mathematics, II.

[2] Sweby, P.K. & Baines, M.J. Convergence of Roe's Scheme for the general non-linear scalar wave equation (1981). Numerical Analysis Report 8/81, University of Reading, also submitted to Math. Comp.

[3] Roe, P.L. Numerical algorithms for the linear wave equation, 1981, RAE Technical Report 81047. Also Roe, P.L. & Baines, M.J., 1981. Proceedings of the 4th GAMM Conference: Notes on Numerical Fluid Mechanics, Vol. 5 (Vieweg).

[4] Lax & Wendroff. Comm. Pure Applied Maths, 13, p. 212, 1960.

[5] Warming & Beam. AIAA Jnl. 14, p. 1241, 1976.

[6] Kruzkov, S.N. First order quasilinear equations in several independent variables. Math. USSR Sbornik, 1970, Vol. 10, p. 217-243. Also,

Oleinik, O.A. Discontinuous solutions of non-linear differential equations. Amer. Math. Soc. Trans. Ser. 2, No. 26, p. 95-172.

[7] Sanders. On convergence of Finite Difference Schemes with Variable Spatial Differencing. Ph.D. abstract, Mathematics Dept. University of California, Los Angeles, (1980).

[8] Engquist, B., & Osher, S. Stable and entropy satisfying approximations for transonic flow calculations. Math. Comp. 34 No. 149, Jan. 1980, p. 45-75.