CONVERGENCE OF ROE'S SCHEME FOR THE NON-LINEAR SCALAR WAVE EQUATION

P.K. Sweby and M.J. Baines

Numerical Analysis Report 6/81

Abstract

Convergence of the approximation generated by the second order scheme of P. Roe to a weak solution of the non-linear scalar wave equation is proved, when the wave speed does not change sign.

1. Introduction

In a recent paper Le Roux [8] proves convergence of a quasi-second order scheme for the non-linear scalar wave equation. The scheme used is second order except where monotonicity-preservation fails when it is only first order. In particular it is shown that the approximations generated converge towards a weak solution of the Cauchy problem (see §2) in the case where the non-linear term $f(u)$ in the equation is monotonic, i.e. when the wave speed is one-signed. In this paper we shall prove the same result for the more accurate scheme of P. Roe (see §2) which is second order everywhere except at those points where the solution possesses an inflection.

Le Roux also proves convergence for non-monotonic $f(u)$ and stronger convergence criteria. It is proposed to extend the result in the present paper to these cases at a later stage.

The problem and Roe's second order scheme are described in §2. Convergence to a weak solution is proved in §3 and conclusions drawn in §4.

2. The Problem and Difference Scheme

Consider the equation

$$u_t + f(u)_x = 0 \tag{2.1}$$

for (x, t) in $\mathbb{R} \times]0, T[$, $T > 0$ and f in $C'(\mathbb{R})$, with

$$u(x, 0) = u_0(x) \tag{2.2}$$

for x in $\mathbb{R}$ and u_0 in $L^\infty(\mathbb{R})$ assumed to be of locally bounded variation.

The Cauchy problem on $\mathbb{R} \times]0, T[$ associated with (2.1), (2.2) is to find a bounded function u which satisfies (2.1), (2.2). A weak solution to this problem is a function u in $L^\infty(\mathbb{R} \times]0, T[)$ which satisfies

$$\iint_{\mathbb{R}\times]0,T[} \left(u\frac{\partial\phi}{\partial t} + f(u)\frac{\partial\phi}{\partial x}\right) dxdt + \int_{\mathbb{R}} u.(x)\phi(x, 0)dx = 0 \tag{2.3}$$

for all test functions ϕ in $C^2(\mathbb{R} \times [0, T[)$ and compact support in $\mathbb{R} \times [0, T[$.

We discuss the convergence towards such a weak solution of the approximation generated by the second order finite difference scheme of Roe [4].

Let the spatial grid size be $h > 0$ and the time grid size be Δt, related to h via the fixed positive real number q through the relation

$$q = \frac{\Delta t}{h} . \tag{2.4}$$

Around the grid point $[kh, n\Delta t]$ define the rectangle

$$I_k \times J_n =](k - \tfrac{1}{2})h, (k + \tfrac{1}{2})h[\times](n - \tfrac{1}{2})qh, (n + \tfrac{1}{2})qh[\tag{2.5}$$

for $k \in \mathbb{Z}$, $n \in \mathbb{N}$ and $n \leq N = [T/qh] + 1$.

We shall show that a weak solution u to (2.1), (2.2) in the sense of (2.3) is approached by a piecewise constant function u_k defined on $\mathbb{R} \times]0, T[$ by

$$u_h(x, t) = u_k^n \text{ for } (x, t) \in I_k \times J_n, \tag{2.6}$$

where the initial condition (2.2) is projected onto the space of piece-wise constant functions by the restriction

$$u_k^0 = \frac{1}{h}\int_{I_k} u_0(x)dx \tag{2.7}$$

The values u_k^n are calculated by Roe's second order difference scheme. This scheme has several different formulations (see [4], [5], [6], [7]) but the one we shall use here is as follows.

[The convention

$$\begin{aligned} u_k &= u_k^n \\ u^k &= u_k^{n+1} \end{aligned} \tag{2.8}$$

is used to simplify the notation.]

Consider the cell (x_{k-1}, x_k), denoted by $I_{k-\frac{1}{2}}$ (see (2.5)). Let $\nu_{k-\frac{1}{2}}$ be the approximation

$$\nu_{k-\frac{1}{2}} = q\frac{\Delta f_k}{\Delta u_k} \tag{2.9}$$

to the CFL number in $I_{k-\frac{1}{2}}$, where Δf_k denotes $f_k - f_{k-1}$.

Let $$S_{k-\frac{1}{2}} = \mathrm{sgn}(\nu_{k-\frac{1}{2}}) \tag{2.10}$$

be the sign of $\nu_{k-\frac{1}{2}}$ and

$$g_{k-\frac{1}{2}} = -q\Delta f_k = -\nu_{k-\frac{1}{2}}\Delta u_k \tag{2.11}$$

be a flux quantity, proportional to the gradient $(u_k - u_{k-1})/h$.

In Roe's first order scheme the quantity $g_{k-\frac{1}{2}}$, associated with the cell $I_{k-\frac{1}{2}}$, is used to update the values of u at the ends of the cell as follows:

$$\begin{cases} \text{increment } u_k \text{ by } g_{k-\frac{1}{2}} \text{ if } \nu_{k-\frac{1}{2}} \geq 0 \\ \text{increment } u_{k-1} \text{ by } g_{k-\frac{1}{2}} \text{ if } \nu_{k-\frac{1}{2}} \leq 0 . \end{cases} \tag{2.12}$$

(Note that if $\nu_{k-\frac{1}{2}} = 0$ then $g_{k-\frac{1}{2}} = 0$ so no ambiguity arises.)

This is simply the upwind scheme of Godunov. We now describe Roe's device for making the scheme second order.

Define a quantity $a_{k-\frac{1}{2}}$ by

$$|a_{k-\frac{1}{2}}| = \tfrac{1}{2}\min\{|(1-|\nu_{k-\frac{1}{2}}|)g_{k-\frac{1}{2}}|,\ |(1-|\nu_{k-\frac{1}{2}-S_{k-\frac{1}{2}}}|)g_{k-\frac{1}{2}-S_{k-\frac{1}{2}}}|\} \qquad (2.13)$$

where the sign of $a_{k-\frac{1}{2}}$ is that of the term chosen by the min. operator. Thus

$$a_{k-\frac{1}{2}} = \begin{cases} \tfrac{1}{2}(1-|\nu_{k-\frac{1}{2}}|)g_{k-\frac{1}{2}} & \underline{\text{or}} \\ \tfrac{1}{2}(1-|\nu_{k-\frac{1}{2}-S_{k-\frac{1}{2}}}|)g_{k-\frac{1}{2}-S_{k-\frac{1}{2}}} . & \end{cases} \qquad (2.14)$$

Switch the quantity $a_{k-\frac{1}{2}}$ across the cell $I_{k-\frac{1}{2}}$ <u>against</u> the direction of the flow so that the value of u at each end of the cell is increased or decreased by $a_{k-\frac{1}{2}}$ as follows:

$$\left.\begin{aligned} u_k &= u_k - a_{k-\frac{1}{2}} \\ u_{k-1} &= u_{k-1} + a_{k-\frac{1}{2}} \end{aligned}\right\} \quad S_{k-\frac{1}{2}} \geq 0 \qquad (2.15a)$$

$$\left.\begin{aligned} u_k &= u_k + a_{k-\frac{1}{2}} \\ u_{k-1} &= u_{k-1} - a_{k-\frac{1}{2}} \end{aligned}\right\} \quad S_{k-\frac{1}{2}} \leq 0 \qquad (2.15b)$$

The whole process is represented graphically in fig. 1.

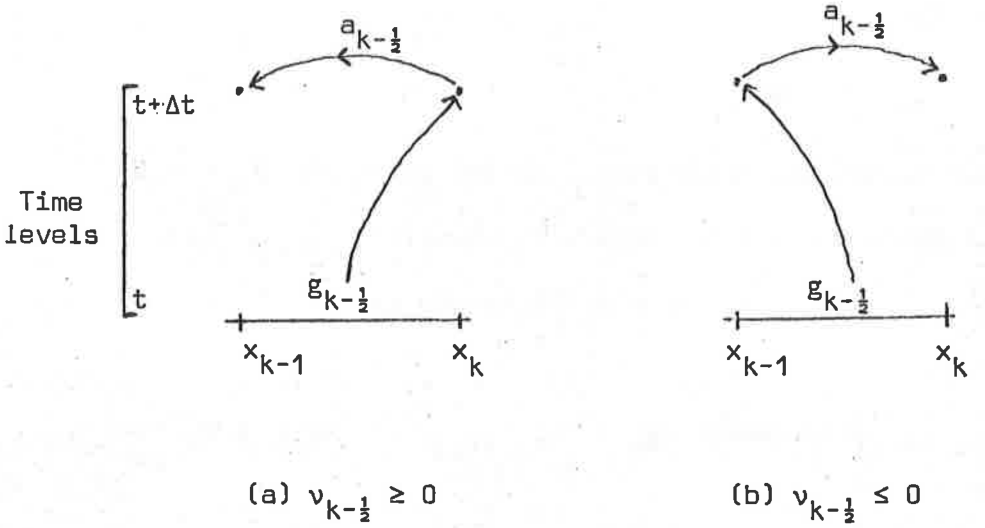

fig. 1

We have

$$\begin{cases} u^k = u_k + g_{k-\frac{1}{2}} + \Delta a_{k+\frac{1}{2}} & (\nu_{k-\frac{1}{2}} \geq 0) & (2.16a) \\ u^k = u_k + g_{k+\frac{1}{2}} - \Delta a_{k+\frac{1}{2}} & (\nu_{k-\frac{1}{2}} \leq 0), & (2.16b) \end{cases}$$

where the terms on the right hand side are the zeroth, first and second order terms, respectively. The scheme can be identified as either Lax-Wendroff or the Warming and Beam upwind scheme ([4], [5]) depending on (2.14).

It may be noted that whereas in some versions of Roe's scheme the second order switch is defined in terms of the $\min\{|g_{k-\frac{1}{2}}|, |g_{k-\frac{1}{2}-S_{k-\frac{1}{2}}}|\}$, the above definition follows the original version [4] and compares $|(1 - |\nu_{k-\frac{1}{2}}|)g_{k-\frac{1}{2}}|$ and $|(1 - |\nu_{k-\frac{1}{2}-S_k}|)g_{k-\frac{1}{2}-S_k}|$ (see (2.13)).

3. Convergence for monotone f

We shall assume throughout that the CFL condition (3.1) below is satisfied.

Consider first $|\nu_{k-\frac{1}{2}}| < 1$ the case of monotone f, i.e. ν one-signed (including zero). We prove

Theorem 1

Suppose f is a monotone function and u_0 lies in $L^\infty(\mathbb{R}) \cap BV_{loc}(\mathbb{R})$, where the condition

$$\sup_k |\nu_k| \le 1 \tag{3.1}$$

is satisfied. Then the family of approximations $\{u_h\}$ generated by Roe's scheme (2.16) from (2.7) contains a sequence $\{u_{h_m}\}_m$ which converges in $L^1_{loc}(\mathbb{R} \times]0, T[)$ towards a weak solution of (2.1), (2.2), as $h_m \to 0$.

Proof

We follow closely the proof of Le Roux [8].

Consider the scheme (2.16a) for $0 \le \nu_k \le 1$, namely,

$$u^k = u_k + g_{k-\frac{1}{2}} + \Delta a_{k+\frac{1}{2}}, \tag{3.2}$$

where, from (2.11) and (2.13),

$$g_{k-\frac{1}{2}} = -\nu_{k-\frac{1}{2}}\Delta u_k = -\frac{\Delta t}{\Delta x}\Delta f_k \tag{3.3a}$$

$$|a_{k-\frac{1}{2}}| = \tfrac{1}{2}\min\{|(1-\nu_{k-\frac{1}{2}})g_{k-\frac{1}{2}}|, |(1-\nu_{k+\frac{1}{2}})g_{k+\frac{1}{2}}|\} \tag{3.3b}$$

the sign of $a_{k+\frac{1}{2}}$ being that of the term chosen by the minimum operator in (3.3b).

Define $\lambda_{k+\frac{1}{2}}$ and $\mu_{k+\frac{1}{2}}$ by

$$2a_{k+\frac{1}{2}} = -\lambda_{k+\frac{1}{2}}\Delta u_k = \mu_{k+\frac{1}{2}}(1-\nu_{k+\frac{1}{2}})g_{k+\frac{1}{2}}. \tag{3.4}$$

Then

$$\lambda_{k+\frac{1}{2}} = \begin{cases} (1-\nu_{k-\frac{1}{2}})\nu_{k-\frac{1}{2}} \quad \underline{\text{or}} \\ (1-\nu_{k-\frac{1}{2}})\nu_{k-\frac{1}{2}}\left\{\dfrac{(1-\nu_{k+\frac{1}{2}})g_{k+\frac{1}{2}}}{(1-\nu_{k-\frac{1}{2}})g_{k-\frac{1}{2}}}\right\} \end{cases} \tag{3.5}$$

so that due to the minimum selecting operator we have the inequality

$$-(1 - \nu_{k-\frac{1}{2}})\nu_{k-\frac{1}{2}} \leq \lambda_{k+\frac{1}{2}} \leq (1 - \nu_{k-\frac{1}{2}})\nu_{k-\frac{1}{2}}. \tag{3.6}$$

Also

$$\mu_{k+\frac{1}{2}} = \begin{cases} \dfrac{(1 - \nu_{k-\frac{1}{2}})g_{k-\frac{1}{2}}}{(1 - \nu_{k+\frac{1}{2}})g_{k+\frac{1}{2}}} & \underline{\text{or}} \\ \\ 1 \end{cases} \tag{3.7}$$

so that due to minimum selecting

$$-1 \leq \mu_{k+\frac{1}{2}} \leq 1 \,. \tag{3.8}$$

Then, from (3.1), (3.2) and (3.3),

$$u^k = u_k\{1 - \tfrac{1}{2}\lambda_{k+\frac{1}{2}} - \tfrac{1}{2}\nu_{k-\frac{1}{2}}[2 - \mu_{k-\frac{1}{2}}(1 - \nu_{k-\frac{1}{2}})]\} +$$

$$u_{k-1}\{\tfrac{1}{2}\lambda_{k+\frac{1}{2}} + \tfrac{1}{2}\nu_{k-\frac{1}{2}}[2 - \mu_{k-\frac{1}{2}}(1 - \nu_{k-\frac{1}{2}})]\} \tag{3.9}$$

Consider the expression

$$\tfrac{1}{2}\lambda_{k+\frac{1}{2}} + \tfrac{1}{2}\nu_{k-\frac{1}{2}}[2 - \mu_{k-\frac{1}{2}}(1 - \nu_{k-\frac{1}{2}})] \tag{3.10}$$

occurring in both brackets in (3.9). This takes a maximum value of

$$\tfrac{1}{2}\{(1 - \nu_{k-\frac{1}{2}})\nu_{k-\frac{1}{2}} + \nu_{k-\frac{1}{2}}(3 - \nu_{k-\frac{1}{2}})\}$$

$$= \tfrac{1}{2}\{\nu_{k-\frac{1}{2}}(4 - 2\nu_{k-\frac{1}{2}})\}$$

$$= \nu_{k-\frac{1}{2}}(2 - \nu_{k-\frac{1}{2}})$$

$$\leq 1 \qquad\qquad \text{for} \quad 0 \leq \nu_{k-\frac{1}{2}} \leq 1. \tag{3.11}$$

Hence the coefficient of u_k in (3.9) is non-negative.

The expression (3.10) also has the minimum value

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

$$= \nu^2_{k-\frac{1}{2}}$$

$$\geq 0. \tag{3.12}$$

Hence the coefficient of u_{k-1} in (3.9) is also non-negative.

From (3.11) and (3.12) we deduce the inequality

$$\text{Min}(u_k, u_{k-1}) \leq u^k \leq \text{Max}(u_k, u_{k-1}) \tag{3.13}$$

known variously as conservation of local stability [3] and compatibility [4]. So by induction we have

$$|u_k|_{L^\infty(\mathbb{R} \times [0, T])} \leq |u_0|_{L^\infty(\mathbb{R})} \quad . \qquad (3.14)$$

We now turn attention to considering u differences in order to establish results on bounded variation. From (3.9),

$$u^{k+1} - u^k = (u_{k+1} - u_k)\{1 - \tfrac{1}{2}\lambda_{k+3/2} - \tfrac{1}{2}\nu_{k+\frac{1}{2}}[2 - \mu_{k+\frac{1}{2}}(1 - \nu_{k+\frac{1}{2}})]\}$$

$$+ (u_k - u_{k-1})\{\tfrac{1}{2}\lambda_{k+\frac{1}{2}} + \tfrac{1}{2}\nu_{k-\frac{1}{2}}[2 - \mu_{k-\frac{1}{2}}(1 - \nu_{k-\frac{1}{2}})]\} \qquad (3.15)$$

in which, from the same inequalities (3.11), (3.12) above, it is seen that the coefficients of both $(u_{k+1} - u_k)$ and $(u_k - u_{k-1})$ are non-negative.

Hence, for any $K \in N$, taking absolute values and summing we obtain

$$\sum_{|k|\leq K} |u^{k+1} - u^k| \leq \sum_{|k|\leq K} |u_{k+1} - u_k|$$

$$+ (u_{k+3/2} - u_{k+\frac{1}{2}})\{1 - \tfrac{1}{2}\lambda_{k+5/2} - \tfrac{1}{2}\nu_{k+3/2}[2 - \mu_{k+3/2}(1 - \nu_{k+3/2})]\}$$

$$+ (u_{-k-1} - u_{-k-2})\{\tfrac{1}{2}\lambda_{-k-\frac{1}{2}} + \tfrac{1}{2}\nu_{-k-3/2}[2 - \mu_{-k-5/2}(1 - \nu_{-k-3/2})]\}$$

$$\leq \sum_{|k|\leq K+1} |u_{k+1} - u_k| \qquad (3.16)$$

In the notation u_k^n of (2.6) we can deduce that

$$\sum_{|k|\leq K} |u_{k+1}^n - u_k^n| \leq \sum_{|k|\leq K+n} |u_{k+1}^0 - u_k^0| . \qquad (3.17)$$

Then for any $R > 0$, we can set $K = [R/h]$ and it follows that (3.17) is bounded by the variation of the initial data u_0 on $]-R, R[$, which is finite. Moreover from (3.13) we have

$$|u^k - u_k| \leq |u_k - u_{k-1}| \qquad (3.18)$$

so that, in the notation of (2.6),

$$|u_k^{n+1} - u_k^n| \leq \sum_{|i|\leq K} |u_k^n - u_{k-1}^n| \leq \sum_{|i|\leq K+n} |u_k^0 - u_{k-1}^0| \qquad (3.19)$$

which for all $n \leq N$ is again bounded by the variation of u_0 on $]-R, R[$, which is finite.

Hence we have a family $\{u_h\}$, each member of which satisfies

$$|u_h|_{L^\infty(R \times [0,\ T])} \le |u_0|_{L^\infty(R)}$$

(3.14) and is of uniform bounded variation in space and time [(3.17), (3.19)].

Following Le Roux [8] we now apply Helly's Theorem to extract from $\{u_h\}$ a subsequence $\{u_{h_m}\}_m$ which converges to a function $u \in L^1_{loc}(\mathbb{R} \times]0,\ T[)$ as $h \to 0$. From (3.14), $u \in L^\infty(\mathbb{R} \times]0,\ T[)$.

It remains to show that u is a weak solution of (2.1), (2.2), i.e. to show that it satisfies (2.3). We introduce a test function $\phi \in C^2(\mathbb{R} \times [0,\ T[)$ with compact support, whose L^2 projection on the space of constant functions on each set $I_k \times J_n$ is

$$\phi_h(x,\ y) = \phi_k^n = \frac{1}{gh^2} \iint_{I_k \times J_n} \phi(x,\ t)dxdt, \qquad (3.20)$$

where $(x,\ t) \in I_k \times J_n$.

Multiplying (3.2) by ϕ_k^n and summing over k and n gives

$$\Big|\sum_k \sum_n [u_k^n(\phi_k^n - \phi_k^{n-1}) + qf(u_k^n)(\phi_{k+1}^n - \phi_k^n)]h + \sum_k u_k^0 \phi_k h\Big|$$

$$\le \sum_k \sum_n |a_{k+\frac{1}{2}}^n||\phi_{k+1}^n - \phi_k^n|h, \qquad (3.21)$$

carrying out summation by parts over k. Now, from (3.1) and (3.3a and b) we have

$$|a_{k+\frac{1}{2}}| \le \frac{1}{8} \max\{|u_{k+1} - u_k|,\ |u_k - u_{k-1}|\}.$$

Hence using the mean value theorem and (3.17), the right hand side of (3.21) becomes

$$\sum_k \sum_n |a_{k+\frac{1}{2}}^n||\phi_{k+1}^n - \phi_k^n|h \le \frac{T}{8q}\left|\frac{\partial \phi}{\partial x}\right|_{L^\infty} \sum_{|k| \le I+n} |u_{k+1}^0 - u_k^0|h \text{ which } \to 0 \text{ as } h \to 0.$$

Hence as $h \to 0$ the inequality (3.21) becomes equation (2.3) and u is a weak solution of the problem (2.1), (2.2).

4. Conclusions

We have proved that the approximation generated by the 2nd order scheme of §2 converges to a weak solution of the Cauchy problem for the non-linear scalar wave equation when the wave speed has constant sign. The weak solution may be discontinuous and in this case uniqueness may fail. To overcome this difficulty the convergence criterion of Kruskov [9] may be used (see [8]). Le Roux shows convergence of the scheme in [8] in the sense of Kruskov under more restrictive conditions but we have not considered it here. It is proposed to prove weak convergence for non-monotone $f(u)$ in a further paper of this series.

Acknowledgement

The above work was carried out while one of us (PKS) was in receipt of a SERC CASE award held in conjunction with RAE, Farnborough.

References

[1] Harten, Hymen & Lax — On finite-difference approximations and entropy conditions for shocks.
Comm. Pure Appl. Maths. 29 1976, 297-322.

[2] P.D. Lax — Hyperbolic systems of conservation laws and the mathematical theory of shock waves.
SIAM Regional Conference Series in Appl. Maths. 11.

[3] A.Y. Le Roux — A numerical conception of entropy for quasi-linear equations.
Maths. Comp. 31 1977, 848-872.

[4] P.L. Roe — Numerical algorithms for the linear wave equation.
RAE Technical Report 81047 1981.

[5] P.K. Sweby — A high order monotonicity preserving algorithm on an irregular grid for non-linear conservation laws.
Reading Numerical Analysis Report 1/81.

[6] M.J. Baines — A numerical algorithm for the solution of systems of conservation laws in two dimensions.
Reading Numerical Analysis Report 2/80.

[7] M.J. Baines — Numerical algorithms for the solution of conservation laws in two and three dimensions.
Reading Numerical Analysis Report 4/81.

[8] A.Y. Le Roux — Convergence of an accurate scheme for first order quasi-linear equations.
R.A.I.R.O. 15 1981, 151-170.

[9] S.N. Kruskov — Math. USSR Sb, 10 1970, 217-243.