JOURNAL OF COMPUTATIONAL PHYSICS 58, 336–348 (1985)

# The Validity of the Modified Equation for Nonlinear Shock Waves

JONATHAN GOODMAN

*Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, New York 10012*

AND

ANDREW MAJDA*

*Department of Mathematics, University of California, Berkeley, California 94720*

Received April 29, 1983; revised June 15, 1984

The modified (model, equivalent) equation is an important tool in designing and analyzing nonlinear difference schemes. In this note, the validity of this principle is rigorously established for nonlinear shock wave solutions and the upwind scheme in a particular case.


## 1. INTRODUCTION

The use of the modified equation (also called model or equivalent equation in the literature) is an important tool in the design and analysis of difference schemes for time-dependent problems. This method has been applied successfully for a wide variety of practical and theoretical purposes including linear and nonlinear stability theory [10, 17, 21], control of parasitic oscillations [15, 16, 21], design of nonlinear filters [1, 6], and most recently to the design of second-order variation-decreasing schemes [7].

The principle involved in deriving the modified equation requires expanding solutions of the difference equation in Taylor series which results in a higher-order differential equation which formally more closely approximates solutions of the difference scheme than the original differential equation does. An advantage of this principle is that it can be easily applied to nonlinear problems. The reader might suspect that such a procedure involving Taylor expansion has highly questionable validity when applied to discontinuous solutions. Nevertheless, Hedstrom ([9],

* Partially supported by NSF Grant MCS-81-02360 and ARO Grant 483964-25530.



0021-9991/85 $3.00

also see [3]) has proved the validity of the principle of the modified equation for general dissipative difference approximations to scalar linear equations with discontinuous initial data while there is substantial numerical evidence [15, 22] supporting the validity of this procedure for many nonlinear difference schemes. Since the behavior of nonlinear difference approximations for shock wave initial data is completely different than the behavior of approximations to discontinuous initial data in the linear case (see [14] and our discussion in Section II), the reasons for the success of the modified equation in describing nonlinear difference schemes remain obscure.

In this note, we rigorously and explicitly check the validity of the principle of the modified equation for the upwind difference scheme and shock wave solutions of the scalar nonlinear conservation law,

$$\begin{aligned} &U_t + f(U)_x = 0 \\ &\qquad U(x, 0) = u_l, \qquad x < 0 \\ &\qquad\qquad\quad = u_r, \qquad x \geqslant 0 \end{aligned} \tag{1.1}$$

where $f(u)$ is an explicit concave function with $f''(u) < 0$. The Rankine–Hugoniot and entropy conditions for the initial data in (1.1) require (here $a(u) = df/du$)

$$s = \frac{f(u_l) - f(u_r)}{u_l - u_r}$$

and

$$a(u_l) > s > a(u_r) \Leftrightarrow u_l < u_r \tag{1.2}$$

so that the entropy satisfying shock wave solution of (1.1) is given by

$$\begin{aligned} U(x, t) &= u_l, \qquad x < st \\ &= u_r, \qquad x \geqslant st. \end{aligned} \tag{1.3}$$

The starting point for our analysis is the explicit discrete solution formula discovered by Lax in [13] for the upwind difference scheme with the special concave nonlinearity, $f(u) = -\log(\beta + \gamma e^{-u})$, $0 < \beta$, $\gamma < 1$, $\beta + \gamma = 1$.

The upwind difference approximation for (1.1) is the nonlinear difference scheme,

$$\begin{aligned} &u_i^{n+1} = u_i^n - \lambda[f(u_i^n) - f(u_{i-1}^n)] \\ &\quad u_i^0 = u_l, \qquad i < 0 \\ &\qquad\;\; = u_r, \qquad i \geqslant 0 \end{aligned} \tag{1.4}$$

where $\lambda = k/h$ is the time-step to space-step ratio (we set $\lambda \equiv 1$ throughout this paper), and the discrete approximate solution is $u_i^n \cong U(ih, nk)$. The modified equation for (1.4) is defined via Taylor expansion from the identities

$$\frac{u(x,t+h)-u(x,t)}{h}=u_t+\frac{h}{2}u_{tt}+O(h^2)$$

$$=u_t+\frac{h}{2}(a(u)(f(u))_x)_x+O(h^2)$$

$$\frac{f(u(x,t))-f(u(x-h,t))}{h}=(f(u))_x-\frac{h}{2}[f(u)]_{xx}+O(h^2)$$

substituted into (1.4). After dropping terms of order $h^2$, we obtain the modified equation for (1.4) with $\lambda\equiv 1$,

$$w_t^h+(f(w^h))_x=\frac{h}{2}[(1-a(w^h))(f(w^h))_x]_x \tag{1.5}$$

with the discontinuous initial data

$$\begin{aligned} w^h(x,0)&=u_l, && x<0\\ &=u_r, && x\geqslant 0. \end{aligned}$$

The validity of the principle of the modified equation in this context for shock wave initial data requires that

the error $w^h(ih,nh)-u_i^n$
is much smaller than the error (1.6)
$U(ih,nh)-u_i^n$.

Below we set $f(u)=-\log(\beta+\gamma e^{-u})$ and verify this principle for weak shocks through the following three steps which have independent interest:

I. *Approach of the Discrete Solution to an Explicit Discrete Travelling Wave Profile*

For any $u_l$, $u_r$, there is a discrete travelling wave

$$v_i^n=v\left(\frac{ih-snk}{h}\right)=v(i-sn)$$

where $v$ is explicitly given by $([u]=u_l-u_r>0)$

$$v(y)=\log\left(\frac{e^{u_l}+e^{u_r}e^{[u]y}}{1+e^{[u]y}}\right). \tag{1.7}$$

Furthermore, $u$ approaches $v$ exponentially; that is,

$$|u_j^n-v_i^n|\leqslant c_1e^{-c_2n}$$

where $c_1$ and $c_2$ are positive constants depending on $u_l$ and $u_r$.

II. *Approach of Modified Equation Solutions to Continuous Travelling Waves*

For any $u_l < u_r$, the solution $w^h(x, t)$ of the modified equation (1.5) tends exponentially to be a travelling wave $w$:

$$\left| w^h(x, t) - w\left(\frac{x - st}{h}\right) \right| \leqslant c_1 e^{-c_2 t/h}$$

where again $c_1, c_2 > 0$ are constants depending on $u_l, u_r$. The function $w(\xi)$ is the solution of the scalar nonlinear O.D.E. (recall $a(w) = df/dw$),

$$\begin{aligned} w'(\xi) &= (a(w)(I - a(w)))^{-1}(f(w) - sw - c) \\ w(0) &= u_0 \end{aligned} \tag{1.8}$$

with $c = f(u_l) - su_l = f(u_r) - su_r$ and the value $u_0$, $u_l < u_0 < u_r$ is determined uniquely by the condition

$$\int_0^\infty (u_r - w(\xi))\, d\xi + \int_{-\infty}^0 (u_l - w(\xi))\, d\xi = 0. \tag{1.9}$$

The function $w$ is determined from (1.8) by quadrature. This result is due to Illin and Oleinik [11].

With the above facts I and II, we see that a proof of the validity of the modified equation in this case reduces to explicit comparison of the discrete profile (1.7) with the continuous profile (1.8) and with the exact solution in (1.2). The natural scale of comparison for these three functions is given by the shock strength, $\delta = u_r - u_l$. The step size, $h$, is not a natural parameter since both the discrete wave profile, $v$, and the continuous wave profile, $w$, are independent of $h$. By explicit comparison of the wave profiles in (1.7) and (1.8), in Section III we obtain the following.

III. *Comparison of Discrete and Continuous Wave Profiles*

With $\delta = u_r - u_l$, for $\delta < \delta_0$

$$\max_{-\infty < y < \infty} |w(y) - v(y)| \leqslant c\delta^2$$

where $c$ is a fixed constant depending only on $u_l$ and independent of $\delta$.

It is now a simple matter to combine steps I, II, and III to conclude that

$$\max_{-\infty < i < +\infty} |u_i^n - w^h(ih, nh)| \leqslant c\delta^2 + c_1 e^{-c_2 n}. \tag{1.10}$$

On the other hand, (1.3) and (1.7) give

$$\max_{-\infty < i < +\infty} |u_i^n - U(ih, nh)| \geqslant c\delta. \tag{1.11}$$

The estimates in (1.10), (1.11) rigorously prove the validity of the modified equation for any fixed time $T = nh$ for these shock solutions as required in (1.6) provided first $\delta$ is chosen sufficiently small and then $h$ is chosen smaller than a fixed $h_0$ which is uniform for all $T \geqslant T_0 > 0$. Moreover, our explicit comparison in Section III indicates that the agreement of the difference equation and the modified equation is much better in a large region near the shock than the agreement of the solution of the difference equation and $U(x, t)$ from (1.1). Far away from the shock, the differential equation solution $U(x, t)$ agrees exactly with the solution of (1.4); however, from II, the modified equation has exponentially small errors in this region so there is not significant disagreement.

## 2. Approach of the Discrete Evolution Problem to the Discrete Travelling Wave

We study the solution of the difference scheme (1.4) using Lax's [13] discrete analogue of the Hopf–Cole transformation, which linearizes (1.4). The purpose of this section is to prove:

PROPOSITION 1. *The solution to* (1.4) (*with* $\lambda = 1$) *tends exponentially to a travelling wave* $v$, *i.e.*,

$$|u_i^n - v(i - sn)| \leqslant c_1 e^{-c_2 n} \tag{2.1}$$

*for all integers* $n > 0$ *and* $i$, *and for some* $c_1, c_2 > 0$ *depending on* $u_l$ *and* $u_r$.

*Remarks.* (1) Proposition 1 illustrates several differences between linear and nonlinear problems. The linear problem typically does not have discrete travelling waves; on the contrary, the error region typically grows like some fractional power of $n$ [9]. The remarkable exponential convergence in (2.1) is also a distinctly nonlinear phenomenon. See Lax [14] for additional discussion of these differences.

(2) The existence of discrete travelling waves for strictly monotone schemes was proved by Jennings [12]. Jennings also proved convergence of the solution of (1.4) to the discrete travelling wave, but with no rate. (See also [20] for a careful and clear proof.) We believe that our exponential rate (2.1) is very generally true and a proof of this would be extremely interesting. Numerical evidence shows a very rapid convergence rate.

(3) The constants $c_1$ and $c_2$ are not uniform in the shock strength $u_l - u_r$. However, more detailed estimates of the binomial sums below would probably give simple approximate formulas for the discrete solution which go over to the expressions for the linearized problem as $u_l - u_r \to 0$.

(4) Lax [23] has found a family of schemes that can be solved explicitly as below. Some of these schemes are more than first-order accurate. It would be interesting to study discrete shock profiles (or lack of them) for these schemes.

*Proof.* The proof has two parts. First we use Lax's transformation to express the solution of the difference scheme in terms of binomial sums, then we asymptotically evaluate the sums. In this way the solution of (1.4) is equal, except for exponentially small terms, to a simple explicit expression which is the discrete travelling wave solution.

Following Lax [13] we "integrate up" (1.4) as follows. Given $U_i^0$, suppose we compute $U_i^n$, for $n>0$, by

$$U_i^{n+1} = U_i^n - \lambda f(U_i^n - U_{i-1}^n).$$

If we define $u_i^n = U_{i+1}^n - U_i^n$, then $u_i^n$ satisfies (1.4). We can use this to solve (1.4) if we choose $U_i^0$ so that $U_{i+1}^0 - U_i^0 = u_i^0$ for all $i$. Taking $\lambda=1$ and using Lax's flux function $f(u) = -\log(\beta + \gamma e^{-u})$ gives

$$U_i^{n+1} = \log\{\beta \exp(U_i^n) + \gamma \exp(U_{i-1}^n)\}.$$

The transformation

$$V = e^U \tag{2.2}$$

converts this into the desired linear difference equation for $V$,

$$V_i^{n+1} = \beta V_i^n + \gamma V_i^{n-1}$$

which implies that

$$V_i^n = \sum_{k=0}^{n} \binom{n}{k} \beta^{n-k} \gamma^k V_{i-k}^0. \tag{2.3}$$

The initial data in (1.4) is achieved by taking

$$\begin{aligned} U_i^0 &= u_r \cdot i \qquad && i \geqslant 0 \\ &= u_l \cdot i && i < 0. \end{aligned}$$

Using (2.2) this becomes

$$\begin{aligned} V_i^0 &= e^{iu_r} \qquad && i \geqslant 0 \\ &= e^{iu_l} && i < 0 \end{aligned}$$

so that (2.3) becomes

$$\begin{aligned} V_i^n &= \sum_{\substack{i-k<0 \\ 0\leqslant k\leqslant n}} \binom{n}{k} \beta^{n-k}\gamma^k e^{(i-k)u_l} + \sum_{\substack{i-k\geqslant 0 \\ 0\leqslant k\leqslant n}} \binom{n}{k} \beta^{n-k}\gamma^k e^{(i-k)u_r} \\ V_i^n &= e^{iu_l} S_l(i, u_l) + e^{iu_r} S_r(i, u_r) \end{aligned} \tag{2.4}$$

where

$$S_l(i,u)=\sum_{\substack{i-k<0\\0\leqslant k\leqslant n}}\binom{n}{k}\beta^{n-k}\gamma^k e^{-ku}$$
$$S_r(i,u)=\sum_{\substack{i-k\geqslant 0\\0\leqslant k\leqslant n}}\binom{n}{k}\beta^{n-k}\gamma^k e^{-ku}. \tag{2.5}$$

In the case of a shock, $u_l<u_r$ (not $u_r<u_l$ since the flux function is concave, not convex), there is a simple asymptotic expression for (2.4). Either both $S_l$ and $S_r$ are well approximated by the full binomial sum

$$(\beta+\gamma e^{-u_l})^n-S_l(i,u_l)\leqslant c_1e^{-c_2n}S_l(i,u_l) \tag{2.6}$$

$$(\beta+\gamma e^{-u_r})^n-S_r(i,u_r)\leqslant c_1e^{-c_2n}S_r(i,u_r), \tag{2.7}$$

or one of the terms is well approximated by the full binomial sum and the other is exponentially small relative to the first:

$$e^{iu_l}S_l(i,u_l)\leqslant e^{iu_l}(\beta+\gamma e^{-u_l})^n\leqslant c_1e^{-c_2n}S_r(i,u_r) \tag{2.8}$$

or

$$e^{iu_r}S_r(i,u_r)\leqslant e^{iu_r}(\beta+\gamma e^{-u_r})^n\leqslant c_1e^{-c_2n}S_l(i,u_l). \tag{2.9}$$

In other words, there are three regions. In region $A$ (to the left of the shock) (2.6) and (2.9) hold. In region $B$ (near the shock) (2.6) and (2.7) hold. In region $C$ (to the right of the shock) (2.7) and (2.8) hold. We will show that these three regions together cover the whole line; so we have, for some $c_1, c_2>0$, and for all $i$,

$$|V_i^n-\{e^{iu_l}(\beta+\gamma e^{-u_l})^n+e^{iu_r}(\beta+\gamma e^{-u_r})\}|\leqslant c_1e^{-c_2n}V_i^n.$$

Substituting back into (2.2) gives

$$\begin{aligned}u_i^n&=U_{i+1}^n-U_i^n=\log(V_{i+1}^n/V_i^n)\\&=\log\left[\frac{e^{u_l}e^{iu_l}(\beta+\gamma e^{-u_l})^n+e^{u_r}e^{iu_r}(\beta+\gamma e^{-u_r})^n}{e^{iu_l}(a+be^{-u_l})^n+e^{iu_r}(a+be^{-u_r})^n}\right]+\varepsilon_i^n\end{aligned}$$

where $|\varepsilon_i^n|\leqslant c_1e^{-c_2n}$. Using the notation

$$[u]=u_r-u_l>0$$

$$[f]=f(u_r)-f(u_l)=-\log(\beta+\gamma e^{-u_r})+\log(\beta+\gamma e^{-u_l})$$

we have $u_i^n=v_i^n+\varepsilon_i^n$ where

$$v_i^n=\log\left\{\frac{e^{u_l}+e^{u_r}\exp(i[u]-n[f])}{1+\exp(i[u]-n[f])}\right\}. \tag{2.10}$$

This $v$ is the desired discrete travelling shock profile and satisfies the difference scheme as the reader is invited to verify. If one is only interested in discrete wave profiles, then (2.10) may be found in a simpler way. In the case $s=[f]/[u]=\frac{1}{2}$, Eq. (1.4) for $v$ is a three-term recurrence relation which may be explicitly solved. Generalizing this solution to other rational and irrational values of $s$ leads to the ansatz (2.10).

It remains only to verify our claims about the binomial sums $S_l$, $S_r$. Our method, which is due to Feller [5, Chap. 6], is a discrete analogue of the Laplace method (or method of stationary phase) for asymptotic evaluation of Gaussian integrals. It is easy to identify the largest terms in a binomial sum, and most terms are negligibly small compared to the largest terms.

LEMMA 1. *Define partial binomial sums by*

$$S(n,p,q,\rho)=\sum_{0\leqslant k/n\leqslant\rho}\binom{n}{k}p^k q^{n-k};\qquad p,q>0.$$

*If* $\rho>p/(p+q)$ *then there exist positive constants* $c_1, c_2$ *so that*

$$\begin{aligned}(p+q)^n-S(n,p,q,\rho)&\leqslant c_1 e^{-c_2 n}S(n,p,q,\rho)\\&\leqslant c_1 e^{-c_2 n}(p+q)^n.\end{aligned}\tag{2.11}$$

*Proof.* Denote the general term in the sum by

$$t_k=\binom{n}{k}p^k q^{n-k},$$

and let $r_k$ be the ratio of successive terms

$$r_k=\frac{t_{k+1}}{t_k}=\frac{1-\sigma}{\sigma+1/n}\cdot\frac{p}{q},\qquad \text{where}\quad \sigma=\frac{k}{n}.$$

Now, if $r_k=1$ then $\sigma\leqslant p/(p+q)$ and, for all $0\leqslant\sigma\leqslant 1$,

$$\frac{\partial r_k}{\partial\sigma}<-c_0<0\tag{2.12}$$

where $c_0$ is a positive constant. Choose $\varepsilon>0$ so that $\rho-p/(p+q)>2\varepsilon$. Then there exists an integer $n_0>0$ and integers $k_1>k_0>0$ (the latter depending on $n$) so that if $n>n_0$, then

$$\frac{k_0}{n}\geqslant\frac{p}{p+q}+\varepsilon\tag{2.13}$$

$$\rho>\frac{k_1}{n}\geqslant\frac{k_0}{n}+\varepsilon.\tag{2.14}$$

From (2.12) and (2.13) we know that $r_k < 1 - c_0\varepsilon$ when $k > k_0$. This, together with (2.14), gives

$$t_k \leqslant (1 - c_0\varepsilon)^{\varepsilon n}\, t_{k_0} \leqslant (1 - c_0\varepsilon)^{\varepsilon n}\, S(n, p, q, \rho)$$

when $k \geqslant k_1$. This gives (2.11) since

$$\begin{aligned}(p+q)^n - S(n, p, q, \rho) &= \sum_{k/n > \rho} t_k \\ &\leqslant n(1 - c_0\varepsilon)^{\varepsilon n}\, S(n, p, q, \rho) \\ &\leqslant c_1 e^{-c_2 n} S(n, p, q, \rho)\end{aligned}$$

for all $n > n_0$, and also for the finitely many $n < n_0$ after possibly adjusting the constants $c_1, c_2$. This completes the proof of Lemma 1.

To evaluate $S_l$, take $p = \beta$, $q = \gamma e^{-u_l}$ to get $\sigma \leqslant \beta/(\beta + \gamma e^{-u_l})$. This gives (2.6) provided that $i \leqslant n \cdot (\beta/(\beta + \gamma e^{-u_l} - \varepsilon))$ for some $\varepsilon > 0$. On the other hand, if $i \geqslant n([f]/[u] + \varepsilon)$ then

$$e^{iu_l}(\beta + \gamma e^{-u_l})^n \leqslant c_1 e^{-c_2 n} \cdot e^{iu_r}(\beta + \gamma e^{-u_r})^n$$

so (2.7) implies (2.9). Similarly, (2.7) follows from $i \geqslant n(\beta/(\beta + \gamma e^{-u_r}) + \varepsilon)$, and (2.6) implies (2.8) when $i \leqslant n([f]/[u] - \varepsilon)$.

These inequalities have a natural physical interpretation. The sum $S_l$ is well approximated by the full binomial sum, roughly if $i/n = x/t < \beta/(\beta + \gamma e^{-u_l}) = f'(u_l)$, which is the linear sound speed for $u_l$. Also, $S_l$ is small compared to $S_r$, roughly when $i/n = x/t > [f]/[u]$, which is the Rankine–Hugoniot shock speed. Similarly, Lemma 1 applies to $S_r$ when $x/t > f'(u_r)$, and $S_r < S_l$ if $x/t < [f]/[u]$. The fact that regions $A$, $B$, and $C$ cover the whole line is equivalent to

$$f'(u_r) < [f]/[u] < f'(u_l),$$

which is exactly the entropy condition, that characteristics from both sides of the shock wave run into the shock.

## 3. Adjustment of Modified Equation Solutions to Continuous Travelling Waves

The solution of (1.5), with shock initial data, also converges exponentially to a travelling wave solution $w((x - st)/h)$, as we now briefly show. The new variables $\tau = h^{-1}t$, $y = h^{-1}x$ put the modified equation (1.5) into the form

$$u_\tau + (f(u))_y = (b(u)\, u_y)_y \tag{3.1}$$

where $b(u)$ is given by

$$b = \tfrac{1}{2}(1 - a(u))\, a(u).$$

The initial data

$$\begin{aligned} u(y, 0) &= u_l, \qquad y<0 \\ &= u_r, \qquad y>0 \end{aligned} \tag{3.2}$$

are preserved. The exponential convergence in time of order $h$ for the solution of (1.5) therefore follows from exponential (in $\tau$) convergence of solutions of (3.1), (3.2). This, in turn, is contained in results of Illin and Oleinik [11]. Actually, the results in [11] are for $b$ independent of $u$. However, since their proof is based on the comparison principle for parabolic equations, the proof extends to our case, where $b$ is positive, smooth, and bounded away from zero. Therefore we omit a detailed discussion.

PROPOSITION 2. *Let $u$ solve* (3.1) *with* (*for some positive $M$ and $\alpha$*)

$$\int_{-\infty}^{y} |u(y', 0) - u_l|\, dy' \leqslant Me^{\alpha y}, \qquad y<0$$

$$\int_{y}^{\infty} |u(y', 0) - u_r|\, dy' \leqslant Me^{-\alpha y}, \qquad y>0$$

*and let $w$ solve* (1.8) *with $w(y) \to u_l$ as $y \to -\infty$, $w(y) \to u_r$ as $y \to \infty$ and*

$$\int_{-\infty}^{\infty} (u(y, 0) - w(y))\, dy = 0$$

(*this is the condition* (1.9)). *Then, for some positive constants $c_1$ and $c_2$,*

$$|u(y, \tau) - w(y - s\tau)| \leqslant c_1 e^{-c_2 \tau}.$$

*Returning to $x$, $t$ variables, we get the desired estimate*:

$$\left| w^h(x, t) - w\left(\frac{x - st}{h}\right)\right| \leqslant c_1 e^{-c_2 t/h}.$$

## 4. COMPARISON OF CONTINUOUS AND DISCRETE TRAVELLING WAVES

We wish to compare the travelling wave, $w$, for the modified equation (which satisfies (1.8), (1.9)) to the discrete travelling wave (1.7). This is hard to do directly since we cannot solve (1.8) explicitly. Therefore first we compare (1.7) to the travelling wave solutions of Burgers' equation,

$$u_t + (\tfrac{1}{2}u^2)_x = \tfrac{1}{2}u_{xx},$$

which satisfy

$$u'(y)=(u-u_l)(u-u_r) \tag{4.1}$$

$$u(0)=\bar{u}=\frac{u_l+u_r}{2}. \tag{4.2}$$

The solution of (4.1), (4.2) is

$$u_c(y)=\bar{u}+\frac{\delta}{2}\frac{1-e^{\delta y}}{1+e^{\delta y}}; \qquad \delta=u_r-u_l. \tag{4.3}$$

A Taylor series calculation involving (1.7) together with the estimate

$$\frac{1+\varepsilon-(1+\alpha\varepsilon)\,\mu}{1+\mu}=\frac{1-\mu}{1+\mu}+O(\varepsilon) \qquad \text{uniformly for} \quad 0\leqslant\mu<\infty$$

shows that

$$\sup|u_c(y)-v(y)|\leqslant c\cdot\delta^2. \tag{4.4}$$

To complete the argument, we compare $u_c$ with the wave profile $w$. This goes in two steps. A lemma of Caflisch [2] shows that $\sup_y |u(y)-w(y)|=O(\delta^2)$ where $u$ is the solution of (4.1) and

$$u(0)=u_0=w(0). \tag{4.5}$$

Then we show that $u_0-\bar{u}=O(\delta^2)$. But this implies that (4.1), (4.5) is solved by $u(y)=u_c(y+O(\delta))$. Since $\sup_y |u_c'(y)|=O(\delta)$, (4.4) gives

$$|u(y)-u_c(y)|=|u_c(y+O(\delta))-u_c(y)|=O(\delta^2).$$

All together we get

$$|w(y)-v(y)|\leqslant|w(y)-u(y)|+|u(y)-u_c(y)|+|u_c(y)-v(y)|.$$

Having shown that all terms on the right are $O(\delta^2)$ we get the desired estimate

$$\sup_y |w(y)-v(y)|\leqslant c\cdot\delta^2.$$

LEMMA 2. *Suppose $w$ satisfies (for some $\tilde{c}>0$) $w'=B(w)$; $w(0)=u_0$ where, for all $u_l>w>u_r$, we have*

$$B(u_l)=B(u_r)=0 \tag{4.6}$$

$$|1-\tfrac{1}{2}B''(w)|\leqslant\tilde{c}\delta. \tag{4.7}$$

*Let u satisfy* (4.1), (4.4). *Then, for all* $\delta$ *sufficiently small,*

$$\sup_y |u(y)-w(y)| \leqslant c \cdot \delta^2.$$

*Proof.* See [2, Lemma 8]. Caflisch's statement has more hypothesis than we state but these follow from (4.6), (4.7) if $\delta$ is sufficiently small.

To apply the lemma, we show that (1.8) is of the form $w' = B(w)$ where

$$\frac{B(w)}{(w-u_l)(w-u_r)} = 1 + O(\delta)$$

when $u_l \geqslant w \geqslant u_r$, for some $\alpha$. This will verify the hypothesis of the lemma (since (4.8) is equivalent to (4.6), (4.7)). From (1.8) and $f(u) = -\log(\beta + \gamma e^{-u})$ we calculate

$$B(u) = b(u)^{-1}(f(u) - su + c)$$

where

$$b(u) = \frac{1}{2}\frac{\beta\gamma e^{-u}}{(\beta+\gamma e^{-u})^2} > 0.$$

Now, $s$ and $c$ are chosen so that

$$f(u) - su + c = \tfrac{1}{2} f''(\bar{u})(u-u_l)(u-u_r)(1+O(\delta))$$

so (4.8) is satisfied, since $\frac{1}{2}(f''(u)/b(\bar{u})) = 1$.

Finally, to show that $\bar{u} - u_0 = O(\delta^2)$, generalize (1.9) by defining $I(u)$ by the relation

$$I(u) = \int_{-\infty}^{y(u)} (w-u_l)\, dy + \int_{y(u)}^{\infty} (w-u_r)\, dy.$$

Here $y(u)$ is the inverse function of $u(y)$. Since $w$ is monotone, $y(u)$ is unique and we may change variables to write

$$I(u) = \int_{u_l}^{u} (w-u_l)\frac{dw}{B(w)} + \int_{u}^{u_r} (w-u_r)\frac{dw}{B(w)}.$$

From (1.9) we get $I(u_0) = 0$ while (4.8) gives $I(\bar{u}) = O(\delta)$. Since $|(d/du)\, I(u)| \geqslant c \cdot \delta^{-1}$ for all $u_l \geqslant u \geqslant u_r$ we have $\bar{u} - u_0 = O(\delta^2)$ as desired.

*Remarks.* (1) The $O(\delta^2)$ agreement of continuous wave profiles for the model equation and discrete travelling wave profiles established explicitly in this section should be true in much more general situations for first-order and even third-order dissipative schemes (see the conditions in [18] guaranteeing discrete travelling waves and the improved methods and results in [19]). A proof of this fact would provide further evidence for the validity of the modified equation for nonlinear difference schemes. In fact, even a crude version of the maximum norm convergence to discrete wave profiles conjectured in Remark (2) of Section 2 together with the above result would rigorously establish the principle of the modified equation for any strictly monotone scheme and weak shocks by a repeat of the argument given here.

(2) Engquist and Osher [4] have given examples of steady wave profiles with states $u_l$, $u_r$ where the principle of the modified equation fails. However, the schemes that they consider fail to be dissipative when linearized at some (sonic) point $\tilde{u}$ with $u_l > \tilde{u} > u_r$ and $b(u)$ from (3.1) satisfies $b(\tilde{u}) = 0$. Since the principle of the modified equation is not valid for a linear difference scheme which fails to be dissipative [9], these results are not so surprising.

## References

1. J. Boris and D. Book, *J. Comput. Phys.* **11** (1973), 38–69.
2. R. Caflisch, *Comm. Pure Appl. Math.* **32** (1979), 531–548.
3. R. Chin, *J. Comput. Phys.* **18** (1975), 233–247.
4. B. Engquist and S. Osher, *Math. Comp.* **34** (1980), 45–75.
5. W. Feller, "An Introduction to Probability Theory and Its Applications," Vol. 1, Wiley, New York, 1950.
6. A. Harten, *Comm. Pure Appl. Math.* **30** (1977), 611–638.
7. A. Harten, High resolution schemes for hyperbolic conservation laws, to appear.
8. A. Harten, J. Hyman, and P. D. Lax, *Comm. Pure Appl. Math.* **29** (1976), 297–322.
9. G. Hedstrom, *Math. Comp.* **29** (1975), 969–977.
10. C. Hirt, *J. Comput. Phys.* **2** (1968), 339–355.
11. A. M. Illin and O. A. Oleinik, *Math. Sb.* **51** (1960), 191–232.
12. G. Jennings, *Comm. Pure Appl. Math.* **26** (1973), 25–37.
13. P. D. Lax, *Comm. Pure Appl. Math.* **10** (1957), 537–566.
14. P. D. Lax, Accuracy and resolution in the computation of solutions of linear and nonlinear equations, *in* "Recent Advances in Numerical Analysis," pp. 107–117, Academic Press, New York/London, 1978.
15. A. Lerat, Numerical shock structure and nonlinear corrections for difference schemes in conservation form, *in* "Sixth International Conference on Numerical Methods in Fluid Dynamics," Lecture Notes in Physics No. 90, pp. 345–351, Springer-Verlag, New York/Berlin, 1979.
16. A. Lerat and R. Peyret, *C. R. Acad. Sci. Paris Ser. A* **276** (1973), 759–762.
17. A. Majda and S. Osher, *Numer. Math.* **30** (1978), 429–452.
18. A. Majda and J. Ralston, *Comm. Pure Appl. Math.* **32** (1979), 445–482.
19. D. Michelson, Discrete shocks for difference approximations to systems of conservation laws, to appear.
20. S. Osher and J. Ralston, *Comm. Pure Appl. Math.* **35** (1982), 737–749.
21. R. Warming and B. Hyett, *J. Comput. Phys.* **14** (1974), 159–179.
22. G. Zwas and J. Roseman, *J. Comput. Phys.* **12** (1973), 179–186.
23. P. D. Lax, On difference schemes for solving initial value problems for conservation laws, *in* "Proceedings, Rome Symposium on Questions in Numerical Analysis," June, 1958.