JOURNAL OF COMPUTATIONAL PHYSICS 12, 179–186 (1973)

The Effect of Nonlinear Transformations on the Computation of Weak Solutions

GIDEON ZWAS*

Courant Institute of Mathematical Sciences, New York University, New York, New York 10003

AND

JOSEPH ROSEMAN†

Department of Mathematics, Polytechnic Institute of Brooklyn, Brooklyn, New York 11201

Received August 14, 1972

For a nonlinear hyperbolic system, computational methods yield different weak solutions for different forms of the system. An explanation is given of the numerical mechanism by which a scheme selects a particular weak solution and why this mechanism depends not only on the scheme but also on the form of the equations. For the Lax–Friedrichs and Lax–Wendroff schemes, it is shown how a correction term can be added to a transformed system so as to preserve the weak solution. This analysis is illustrated by numerical shock-like solutions of the equations of shallow fluid flow over a ridge.

INTRODUCTION

Frequently, for a nonlinear hyperbolic initial value problem, no solution exists which is smooth for all positive time, even if the initial data are smooth; however, there exist more than one weak solution, and some solutions may be discontinuous. For such problems, in conservation form, it is conjectured that by requiring that specific jump relations (the Rankine–Hugoniot conditions) and an entropy-like condition, be satisfied at discontinuities, there exists a unique weak solution. This has been proved for a single equation by Quinn [3] using a generalized entropy condition and the definitions of these special weak solutions, known as shock solutions, given by Lax [2].

Unfortunately, unlike the classical situation, this uniqueness depends on the form of the given system of partial differential equations. In other words, a nonlinear transformation of the original dependent variables leads to a different

[*] On leave from Tel Aviv University, Israel.

[†] Supported by NSF Grant GP-27960.

shock solution. In the other shock solution the speed of propagation of the discontinuities for example may differ from the original shock speed. Of course, in cases with smooth solutions, such transformations are permissible and all the classical results hold.

Extra care must, therefore, be taken when a system of equations of this type is formulated, especially when numerical computations are to be performed. A computer program including a nonlinear system solver that gave satisfactory smooth results for some physical problems, and even good agreement with observations, might fail when discontinuities arise. This phenomenon has been demonstrated by Lax [2] for a single equation, and lately by Kasahara and Houghton [1] for a system describing shallow flows over a ridge. In [1] it is also shown that when computing with schemes which can handle shocks, the numerical procedure is "loyal" to the form of the equations used, and gives in every case the corresponding, but different, weak solutions.

Two questions immediately arise: (a) which is the preferable form of the equations? (b) knowing the preferable form, is it possible to make a change in the numerical scheme so that a nonlinear transformation would leave the weak solution unchanged?

Our work here will deal with the second question, but we shall first remark that the first question *cannot* be answered within mathematics since it is not a strictly mathematical problem. The question is a physical one and can be answered only by a careful examination of the way in which the equations were obtained from the physical model. These points will be dealt with in our example which is the one presented by Kasahara and Houghton [1].

We begin by writing the partial differential equations governing the motion of an incompressible, homogeneous inviscid, hydrostatic fluid. The equations we write down are those derived directly from the *integral conservation laws* of mass and momentum. This important point was emphasized by Rubin and Preiser [5] who even suggested that numerical schemes should be constructed from the integral form of the physical laws*. Our equations, written in what will be referred to as the "momentum form" [1], are

$$W_t + F_x + K = 0, \tag{1}$$

where

$$W = \begin{pmatrix} m \\ \phi \end{pmatrix}, \quad F = \begin{pmatrix} m^2/\phi + g\phi^2/2 \\ m \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} g\phi H_x \\ 0 \end{pmatrix}.$$

Here, ϕ is the height of the fluid above the lower boundary surface and m is the momentum per unit volume. $m = u \cdot \phi$, where u is the horizontal fluid velocity.

* Preiser and Rubin, private communication.

g is the gravitational acceleration and $H = H(x)$ is the function describing the lower boundary. We shall assume that

$$H(x) = \begin{cases} h(x), & |x| \leqslant a \\ 0, & |x| \geqslant a \end{cases}, \tag{2}$$

namely that we have a finite isoltated ridge of width 2a and elsewhere the lower surface is horizontal. We shall also choose $h(x)$ such that H, H_x and H_{xx} will be everywhere continuous.

If in (1) we multiply the second equation by u, subtract it from the first, divide by ϕ and take the result together with the second equation of (1), then we have our system in "velocity form," namely

$$\tilde{W}_t + \tilde{F}_x + \tilde{K} = 0, \tag{3}$$

where

$$\tilde{W} = \begin{pmatrix} u \\ \phi \end{pmatrix}, \quad \tilde{F} = \begin{pmatrix} u^2/2 + g\phi \\ u\phi \end{pmatrix} \quad \text{and} \quad \tilde{K} = \begin{pmatrix} gH_x \\ 0 \end{pmatrix}.$$

Note that (3), which we obtained after our nonlinear transformation, is also in conservation form. The system (3) has often been used in various applications and produces nature-like results in smooth cases but nonphysical shocks when discontinuities arise.

This brings brings us to the second question: If we intend to use appropriate schemes like the Lax–Wendroff (LW) second order scheme, or the first order Lax–Friedrichs (LF) method [2], can something be done so that (3) can still be used? What is the numerical mechanism that chooses the correct physical discontinuities, how is this mechanism damaged by our nonlinear transformation, and how can it be corrected?

The Hidden Dissipative Term

We start our analysis with the first order Lax–Friedrichs (LF) scheme, known sometimes also as the Lax staggered method. Let us first discretize our equations and denote as usual $f_j^n = f(x_j, t_n)$ and $\lambda = \Delta t/\Delta x$. The ridge function $h(x)$ in (2) will be taken as $h_c \circ (1 - x^2/a^2)$, where usually we chose $a = 1$ and $h_c = 0.5$ and at $x = \pm a$ we interpolated to insure the continuity of H, H_x, and H_{xx}.

The initial conditions at $t = 0$ are $\phi(x, 0) = \phi_0 - H(x)$ and $u(x, 0) = F_0(g\phi_0)^{1/2}$, where usually we chose $\phi_0 = 1$, and $F_0 = 0.7$, a case known to include shocks (see [1] and references therein).

Let us also write down the following facts

$$F_x = AW_x \qquad \text{where} \quad A = \begin{pmatrix} 2m/\phi & g\phi - m^2/\phi^2 \\ 1 & 0 \end{pmatrix} \tag{4}$$

and similarly

$$\tilde{F}_x = \tilde{A}\tilde{W}_x \qquad \text{where} \quad \tilde{A} = \begin{pmatrix} u & g \\ \phi & u \end{pmatrix}. \tag{5}$$

The eigenvalues, in both cases, are $[u \pm (g\phi)^{1/2}]$, and we denote $[| u | + (g\phi)^{1/2}]$ by σ. The known linear stability condition for the LF scheme, as well as for the LW scheme, is $\lambda\sigma \leqslant 1$, and our time steps will be chosen accordingly. We also note, for later purposes that

$$W_{tt} = [A(F_x + K)]_x - \dot{K} \qquad \text{where} \quad \dot{K} = K_t = \begin{pmatrix} -gH_x m_x \\ 0 \end{pmatrix}, \tag{6}$$

and similarly

$$\tilde{W}_{tt} = [\tilde{A}(\tilde{F}_x + \tilde{K})]_x, \qquad \text{since} \quad \dot{\tilde{K}} = 0. \tag{7}$$

Now, the LF scheme for the system (1) (and similarly for (3)) is given by

$$W_j^{n+1} = (W_{j+1}^n + W_{j-1}^n)/2 - (\lambda/2) \cdot (F_{j+1}^n - F_{j-1}^n) - \Delta t \cdot K_j^{\,n}. \tag{8}$$

This scheme is of first order accuracy which means (see [4]) that substituting the solution $W(x, t)$ of the differential system into (8) yields

$$\begin{aligned} &W(x_j, t_{n+1}) - [W(x_{j+1}, t_n) + W(x_{j-1}, t_n)]/2 \\ &\qquad + (\lambda/2) \cdot [F(_{j+1}, t_n) - F(x_{j-1}, t_n)] + \Delta t \cdot K(x_j, t_n) \\ &\quad = 0((\Delta t)^r), \qquad \text{where } r = 2. \end{aligned} \tag{9}$$

We now claim that what (8) is approximating even more closely than (1), is the system

$$W_t + F_x + K = Q \cdot W, \tag{10}$$

where Q is a differential operator to be specified shortly, such that in smooth regions $Q \cdot W = O(\Delta t)$. Using (10) we have

$$\begin{aligned} W(x_j, t_{n+1}) = {} & W(x_j, t_n) + \Delta t(-F_x - K + QW)_j^n \\ & + [(\Delta t)^2/2] \cdot \{[A(F_x + K)]_x - \dot{K}\}_j^n + O((\Delta t)^3). \end{aligned} \tag{11}$$

We now take W in (9) to be the solution of (10) and choose Q such that in (9), r

is at least three. By substituting (11) and other necessary Taylor expansions into (9), it is immediately found that r is indeed three if

$$Q \cdot W = (\Delta t/2)\{(W_{xx}/\lambda^2) - [A(F_x + K)]_x + \dot{K}\} \tag{12}$$

and a similar expression is obtained for the system (3), except that there $\dot{\tilde{K}} = 0$.

The linear version (A and K taken as constants) of this hidden term Q is $(\Delta t/2) \cdot (\lambda^{-2}I - A^2)\, W_{xx}$ and it is precisely our numerical stability requirement which makes the matrix $(\lambda^{-2}I - A^2)$ positive definite, and, therefore, (12) is essentially a parabolic dissipative term. The hidden dissipative term $Q \cdot W$ depends on the specific scheme used.

A similar analysis, but much more cumbersome, can be carried out for the Lax–Wendroff scheme. There it turns out that one must search for a term $Q \cdot W$ in (10) such that the left-hand side of the equation analogous to (9) should be $O((\Delta t)^5)$. This is needed in order to reveal the dissipative feature of the $Q \cdot W$ term which plays an important role in computations involving discontinuities. The linear version of the hidden LW dissipative term (see [4, pp. 331–332]) is

$$Q_{LW} \cdot W = -[(\Delta t)^2/24](\lambda^{-2}I - A^2)(4A \cdot W_{xxx} + 3\Delta t \cdot A^2 \cdot W_{xxx}). \tag{13}$$

Note the negative sign which makes the fourth derivative term a proper dissipative one. Returning to the LF scheme; since we are acutally solving (10), we claim that the dissipative term must be included in any transformation of the system, if the same weak solution is to be obtained numerically.

A Correction to the Velocity Form System

We continue by taking the system (1) and substituting it in (11). After replacing m by $u \cdot \phi$ we get

$$QW = (\Delta t/2) \cdot \binom{q_1}{q_2}, \tag{14}$$

where

$$\begin{aligned} q_1 = {} & (u\phi_{xx} + 2u_x\phi_x + \phi u_{xx})/\lambda^2 - (3u^2 + g\phi)\,\phi u_{xx} - (u^2 + 3g\phi)\,u\phi_{xx} \\ & - 6u\phi u_x^2 - 6u^2u_x\phi_x - 5g\phi u_x\phi_x - 3gu\phi_x^2 - 2g\phi uH_{xx} \\ & - 3g\phi u_xH_x - 3gu\phi_xH_x \end{aligned} \tag{15}$$

and

$$\begin{aligned} q_2 = {} & \phi_{xx}/\lambda^2 - 2u\phi u_{xx} - (u^2 + g\phi)\,\phi_{xx} - 2\phi u_x^2 - g\phi_x^2 \\ & - 4u \cdot u_x\phi_x - g\phi H_{xx} - g\phi_xH_x\,. \end{aligned} \tag{16}$$

The same procedure applied to system (3) leads to

$$\tilde{Q} \cdot \tilde{W} = (\Delta t/2) \cdot \binom{\tilde{q}_1}{\tilde{q}_2}, \tag{17}$$

where

$$\tilde{q}_1 = u_{xx}/\lambda^2 - (u^2 + g\phi)\, u_{xx} - 2gu\phi_{xx} - 2uu_x{}^2 \\ - 3gu_x\phi_x - gu_xH_x - guH_{xx},$$

and

$$\tilde{q}_2 = q_2 .$$

Next, we take the system (1) with (13) on the right-hand side and apply the transformation to it, that is, we multiply the second equation by u, subtsact from the first, and divide by ϕ. A comparison of the result with (3) including (17) on the right-hand side, reveals that the second equations exactly match, but the first corresponding equations differ by

$$s = (\Delta t/\phi) \cdot [u_x\phi_x/\lambda^2 - (u\phi)_x \cdot (u^2/2 + g\phi + gH)_x]. \tag{18}$$

We, therefore, claim that the difference in the numerical solutions obtained from the two sets of equations is caused by the failure to correctly take into account the effect of the transformation on the hidden dissipative term. We further claim that solution of system (3) with the correction term (18) included, will give the same physical shocks that are obtained when system (1) is solved.

This was confirmed by actually computing numerical solutions, a few graphs of which are shown in Fig. 1 to illustrate this phenomenon.

After 750 time steps the main shock, which is travelling to the right, is at $x_s = 3.2$ when the equations in momentum form were used (Fig. 1b). At this time for the system in volocity form without any correction term, the shock is at $x_s = 5.2$ (Fig. 1c). Addition of the correction term (18), again puts the shock at $x_s = 3.2$ for the same time, as seen in Fig. 1a. This clearly illustrates the effect of different dissipative terms on the computation of weak solutions, even when these terms are hidden. In smooth cases, the correction is of the order $\Delta t \cdot s = O((\Delta t)^2)$ and, thus, is negligible for a first order scheme.

In fact, comparison of Figs. 1a and 1b shows the solutions obtained from the momentum form system and the velocity form system with the correction term (18), to be identical. By contrast, comparison of Figs. 1b and 1c shows that the solutions obtained from the momentum and velocity form equations are different *everywhere* to the right of the point at which the momentum form shock first appears. At $t = 18.75$ and $x = 6.0$, for example, we have in Figs. 1a and 1b, $\phi = 0.924$ but in Fig. 1c, $\phi = 0.986$.

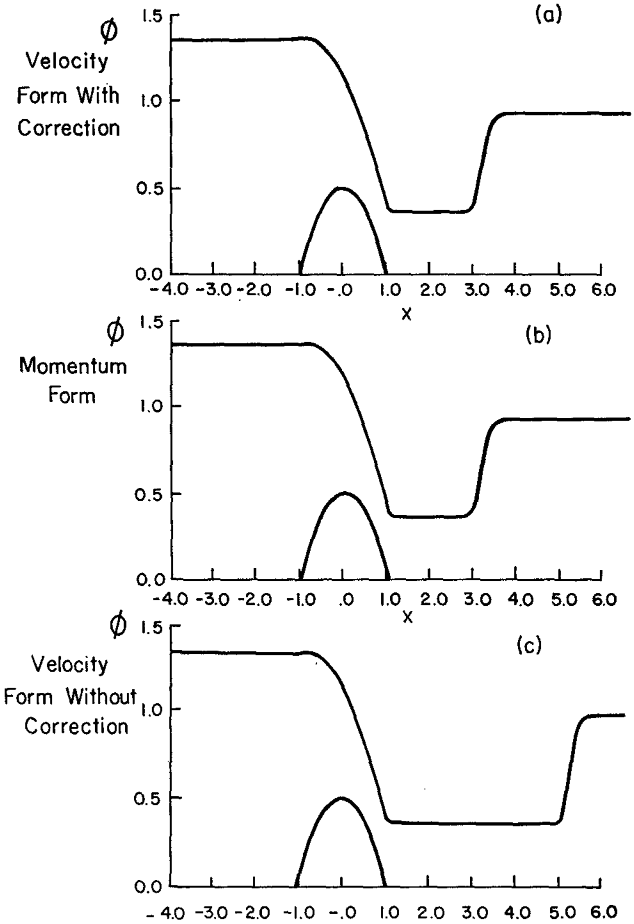

FIG. 1. The fluid's height after 750 Δt's ($t = 18.75$) for $F_0 = 0.7$, $H = (1 - X^2)/2$, and $\Delta X = 0.05$.

Finally, we present an example of our analysis applied to a simple single equation. If one solves the equation

$$u_t + (u^2/2)_x = 0 \tag{19}$$

by the LF scheme, he solves with even higher accuracy the equation

$$u_t + (u^2/2)_x = (\Delta t/2) \cdot [u_{xx}/\lambda^2 - (u^2 u_x)_x]. \tag{20}$$

Now, under the transformation $v = u^2$, (19) becomes

$$v_t + ((2/3) \cdot v^{3/2})_x = 0, \tag{21}$$

and the LF scheme solve with greater accuracy

$$v_t + ((2/3) \cdot v^{3/2})_x = (\Delta t/2) \cdot [v_{xx}/\lambda^2 - (vv_x)_x]. \tag{22}$$

However, under this transformation, (20) becomes

$$v_t + ((2/3) \cdot v^{3/2})_x = (\Delta t/2)[v_{xx}/\lambda^2 - (vv_x)_x + (v_x{}^2/2) \cdot (1 - 1/(\lambda^2 v))]. \tag{23}$$

The term $(\Delta t/4)\, v_x{}^2(1 - 1/(\lambda^2 v))$ is, therefore, the correction term which should be taken into account if one wants to compute a weak solution of (21) which is equal to the weak solution of (19). Note that since these equations are not motivated by a physical problem it is not clear if (19) or (21) is the natural form and, therefore, which should be augmented with a correction term.

Acknowledgments

The authors wish to thank Professors E. Isaacson, S. Preiser, and E. Rubin for helpful discussions and useful comments.

References

1. A. Kasahara and D. Houghton, *Comp. Phys.* **4** (1969), 377.
2. P. Lax, *Comm. Pure Appl. Math.* **7** (1954), 159.
3. B. Quinn, *Comm. Pure Appl. Math.* **24** (1971), 125.
4. R. Richtmyer and W. Morton, "Finite Difference Methods for Initial Value Problems," Interscience-Wiley, New York, 1967.
5. E. Rubin and S. Preiser, *Math. Comp.* **24** (1970), 57.