JOURNAL OF COMPUTATIONAL PHYSICS 30, 428–451 (1979)

Galerkin Methods Applied to Some Model Equations for Non-Linear Dispersive Waves

M. E. ALEXANDER* AND J. LL. MORRIS

Computer Science Department, University of Waterloo, Waterloo, Ontario, Canada

Received May 24, 1977; revised March 17, 1978

The application of a dissipative Galerkin scheme to the numerical solution of the Korteweg de Vries (KdV) and Regularised Long Wave (RLW) equations, is investigated. The accuracy and stability of the proposed schemes is derived using a localised Fourier analysis. With cubic splines as basis functions, the errors in the numerical solutions of the KdV equation for different mesh-sizes and different amounts of dissipation is determined. It is shown that the Galerkin scheme for the RLW equation gives rise to much smaller errors (for a given mesh-size), and allows larger steps to be taken for the integrations in time (for a specified error tolerance). Also, the interaction of two solitons is compared for the KdV and RLW equations, and several differences in their behaviour are found.

1. INTRODUCTION

The present study is concerned with the numerical solution, using Galerkin methods, of two equations which represent approximations to a larger class of physical problems in which non-linear waves are present. These are the Korteweg de Vries (KdV) equation

$$K[u] \equiv u_t + uu_x + \epsilon u_{xxx} = 0, \tag{1.1}$$

and the Regularised Long Wave (RLW) equation

$$R[u] \equiv u_t + uu_x - \delta u_{xxt} = 0, \tag{1.2}$$

in which ϵ and δ are "small" positive constants, and the corresponding terms represent the modelling of dispersive behaviour.

Since its formulation by Korteweg and de Vries [8], a wide class of exact solutions to equation (1.1) has been found, notably in recent times using the Inverse Scattering Method (Gardner *et al.* [5]): this method generates the well-known N-soliton solutions possessing the property that amplitudes and velocities, as well as the shapes, of individual solitons are preserved in a (non-linear) interaction. Numerical studies of

* Present address: University Observatory, Buchanan Gardens, St. Andrews, Fife KY16 9LZ, Scotland.

0021-9991/79/030428-24$02.00/0

Copyright © 1979 by Academic Press, Inc.
All rights of reproduction in any form reserved.

this equation using Finite Difference methods have been carried out, notably by Greig [6].

By contrast, for the RLW equation (1.2), originally proposed by Peregrine [9], no wide class of exact solutions is yet known. The numerical experiments of Eilbeck and McGuire [4] seem to indicate that the behaviour of soliton solutions is similar to the KdV case; indeed, as may easily be shown, the equations (1.1) and (1.2) possess the common *single-soliton* solution

$$u(x, t) = 3\, c\, \mathrm{sech}^2(kx - \omega t + d), \tag{1.3}$$

where

$$k = 1/2(c/\epsilon)^{1/2};\ \omega = kc, \tag{1.4}$$

and c and d are parameters, provided that the relation

$$\delta = \epsilon/c \tag{1.5}$$

is satisfied. Thus, although it is not yet known whether there are multiple-soliton solutions to the RLW equation, the existence of a common one-soliton solution nevertheless provides the motivation for examining the numerical solutions with N solitons (we consider the case of $N = 2$ only).

In the present study, we extend previous numerical work on equations (1.1) and (1.2) (where Finite Differences were used), by proposing a generalised (ordinary or dissipative) Galerkin scheme, based on the one given by Wahlbin [12] for the KdV equation, and applying it to one- and two-soliton initial conditions. The linearised error and stability analysis is given in Section 2, and the numerical computation procedure in Section 3. In Section 4, we examine the behaviour of the one-soliton solutions for different (uniform) mesh-sizes and different amounts of dissipation in the KdV scheme, and compare it with the solutions of the RLW equation; furthermore, we compare the behaviour of the two-soliton solutions of both the KdV and RLW equations.

2. Accuracy and Stability Analysis

Let S^μ denote the space of smoothest splines, defined piecewise on intervals of length h (= mesh size) as polynomials of order μ (degree $\mu - 1$), having compact support on an interval of length μh. These spline functions can be constructed in the usual way as a $(\mu - 1)$-fold convolution (Schoenberg [10]): Let

$$\begin{aligned} M_1(X) &= 1, \qquad -1/2 \leqslant X \leqslant 1/2, \\ &= 0, \qquad \text{otherwise.} \end{aligned} \tag{2.1}$$

Then

$$M_\mu(X) = M_1 * M_1 * \cdots * M_1 \qquad (\mu - 1 \text{ times}). \tag{2.2}$$

where

$$(f * g)(X) \equiv \int_{-\infty}^{\infty} f(Y)\, g(X - Y)\, dY \tag{2.3}$$

Using (2.1), we see that, inductively,

$$M_\mu(X) = \int_{X-1/2}^{X+1/2} M_{\mu-1}(Y)\, dY \tag{2.4}$$

The basis functions $\phi_l(x)$ of S^μ are then defined by

$$\phi_l(x) = M_\mu\left(\frac{x}{h} - l\right) \qquad (l \in Z). \tag{2.5}$$

We propose the following generalised Galerkin methods for equations (1.1) and (1.2): if $\chi \in S^\mu$, and U denotes the Galerkin solution, then

$$\text{KdV:} \qquad (K[U],\ \chi + q\epsilon h^3 \chi_{xxx}) = 0, \tag{2.6}$$

$$\text{RLW:} \qquad (R[U],\ \chi + q\,\delta^{1/2} h \chi_x) = 0, \tag{2.7}$$

where

$$(f, g) \equiv \int_{-\infty}^{\infty} f(x)\, g(x)\, dx \tag{2.8}$$

and q is an arbitrary parameter determining the amount of dissipation in the scheme ($q = 0$ corresponds to an unmodified (non-dissipative) scheme).

In order to undertake a Fourier analysis of the accuracy and stability of (2.6) and (2.7) we set $u = c =$ constant in the quadratic term uu_x in the equations (1.1) and (1.2). Next, rescale x, t and u to remove the constants ϵ and δ. We obtain the linearised forms

$$\text{KdV:} \quad (U_t + cU_x + U_{xxx},\ \chi + qh^3 \chi_{xxx}) = 0; \tag{2.9}$$

$$\text{RLW:} \qquad (U_t + cU_x - U_{xxt},\ \chi + qh\chi_x) = 0. \tag{2.10}$$

The Galerkin solution may be expressed as

$$U(x)(t) = \sum_j \alpha_j(t)\, \phi_j(x). \tag{2.11}$$

Substituting into (2.9) and (2.10), setting $\chi = \phi_l$, we obtain the Galerkin method in the forms

$$\begin{aligned} \text{KdV:} \quad & \sum_j \alpha_j'(t)[(\phi_j, \phi_l) + qh^3(\phi_j, \phi_l^{(3)}) \\ & + \alpha_j(t)[c(\phi_j^{(1)}, \phi_l) + (\phi_j^{(3)}, \phi_l) + qh^3\{c(\phi_j^{(1)}, \phi_l^{(\cdot)}) \\ & + (\phi_j^{(3)}, \phi_l^{(3)})\}] = 0; \end{aligned} \tag{2.12}$$

$$\text{RLW:}\quad \sum_j \alpha_j'(t)[(\phi_j, \phi_l) + qh(\phi_j, \phi_l^{(1)}) - (\phi_j^{(2)}, \phi_l) - qh(\phi_j^{(2)}, \phi_l^{(1)})]$$

$$+ c\alpha_j(t)[(\phi_j^{(1)}, \phi_l) + qh(\phi_j^{(1)}, \phi_l^{(1)})] = 0. \tag{2.13}$$

The values of the integrals $(\phi_j^{(\alpha)}, \phi_l^{(\beta)})$ clearly depend only on the difference $j - l$, as follows from (2.5); hence (2.12) and (2.13) are of the form of (discrete) convolutions. Now define

$$\begin{aligned}
\hat{\alpha}(\theta, t) &= \sum_l \alpha_l(t)\, e^{-il\theta};\\
g_{\mu,0}(\theta) &= h^{-1} \sum_l (\phi_0, \phi_l)\, e^{-il\theta};\\
g_{\mu,1}(\theta) &= -i \sum_l (\phi_0^{(1)}, \phi_l)\, e^{-il\theta};\\
g_{\mu,2}(\theta) &= h \sum_l (\phi_0^{(1)}, \phi_l^{(1)})\, e^{-il\theta};\\
g_{\mu,3}(\theta) &= ih^2 \sum_l (\phi_0^{(1)}, \phi_l^{(2)})\, e^{-il\theta};\\
g_{\mu,4}(\theta) &= h^3 \sum_l (\phi_0^{(2)}, \phi_l^{(2)})\, e^{-il\theta};\\
g_{\mu,6}(\theta) &= h^5 \sum_l (\phi_0^{(3)}, \phi_l^{(3)})\, e^{-il\theta}.
\end{aligned} \tag{2.14}$$

The various powers of h appearing in the definitions of $g_{\mu,\sigma}$ have been chosen to render these quantities independent of h. Multiplying (2.12) and (2.13) by $e^{-il\theta}$, summing over l, and using the multiplicative property of convolutions:

$$\widehat{(f * g)}(\theta) = \hat{f}(\theta)\, \hat{g}(\theta), \tag{2.15}$$

where

$$\hat{f}(\theta) = \sum_l f_l e^{-il\theta}; \qquad (f * g)_j = \sum_l f_l g_{j-l}, \tag{2.16}$$

we obtain

$$\text{KdV:}\quad \frac{\partial \hat{\alpha}}{\partial t}(\theta, t)[g_{\mu,0} - iqg_{\mu,3}]$$

$$+ h^{-1}\hat{\alpha}(\theta, t)[icg_{\mu,1} - ih^{-2}g_{\mu\,3} - qcg_{\mu,4} + qh^{-2}g_{\mu,6}] = 0; \tag{2.17}$$

$$\text{RLW:}\quad \frac{\partial \hat{\alpha}}{\partial t}(\theta, t)[g_{\mu,0} - iqg_{\mu,1} + h^{-2}g_{\mu,2} - iqh^{-2}g_{\mu,3}]$$

$$+ ch^{-1}\hat{\alpha}(\theta, t)[ig_{\mu,1} + qg_{\mu,2}] = 0. \tag{2.18}$$

Define

$$\Pi_K(\theta, h) = -\frac{icg_{\mu,1} - ih^{-2}g_{\mu,3} - qcg_{\mu,4} + qh^{-2}g_{\mu,6}}{g_{\mu,0} - iqg_{\mu,3}}; \tag{2.19}$$

$$\Pi_R(\theta, h) = -c\frac{ig_{\mu,1} + qg_{\mu,2}}{g_{\mu,0} - iqg_{\mu,1} + h^{-2}g_{\mu,2} - iqh^{-2}g_{\mu,3}}. \tag{2.20}$$

We may, therefore, solve the ordinary differential equations (2.17) and (2.18) to give

$$\hat{\alpha}(\theta, t) = \exp\left[\frac{t}{h}\Pi(\theta, h)\right]\hat{\alpha}(\theta, 0), \tag{2.21}$$

where $\Pi = \Pi_K$ or Π_R. Writing exp [ ] in terms of its Fourier components

$$\exp\left[\frac{t}{h}\Pi(\theta, h)\right] = \sum_l \Phi_l(t, h)\, e^{-il\theta}, \tag{2.22}$$

we may express the final solution, at the mesh-points, in the form (Wahlbin [12])

$$U(jh)(t) = \sum_l \Phi_l(t, h)\, f_0(jh - lh), \tag{2.23}$$

where $f_0(x) = u(x, 0)$ is the initial condition. Thus, the behaviour of the solution is determined by the quantity exp [ ] in (2.21). Recall the following definitions:

The finite difference operators (2.12) and (2.13) are

(i) accurate of order r if

$$\begin{aligned} &\text{KdV:} \quad \Pi_K(\theta, h) = i\theta(-c + h^{-2}\theta^2) + O(\theta^{r+3}); \\ &\text{RLW:} \quad \Pi_R(\theta, h) = -\frac{ic\theta}{1 + h^{-2}\theta^2} + O(\theta^{r+3}), \end{aligned} \qquad \text{as} \quad \theta \to 0; \tag{2.24}$$

(ii) dissipative of order $2s$, if there exists a $\gamma > 0$ such that

$$\mathrm{Re}(\Pi(\theta, h)) \leqslant -\gamma\theta^{2s}, \quad |\theta| \leqslant \pi, \tag{2.25}$$

independently of h.

In order to determine the accuracy and dissipativity of the schemes (2.9) and (2.10), we require the result (e.g., Thomée and Wendroff [11]):

$$\sum_l (\phi_0^{(\alpha)}, \phi_l^{(\beta)})\, e^{-il\theta} = i^{\alpha-\beta}\sum_l (2\pi l + \theta)^{\alpha+\beta}(\hat{\phi}(2\pi l + \theta))^2, \tag{2.26}$$

where

$$\begin{aligned}\hat{\phi}(\theta) &\equiv \int_{-\infty}^{\infty} \phi_0(x)\, e^{-ix\theta}\, dx \\ &= \left(\frac{2 \sin \tfrac{1}{2}\theta}{\theta}\right)^{\mu}.\end{aligned} \tag{2.27}$$

We may express the $g_{\mu,\sigma}$ in the following alternative forms, to be used later.

(i) $$g_{\mu,\sigma}(\theta) = \theta^{\sigma}\hat{\phi}^2(\theta) + R_{\mu,\sigma}(\theta), \tag{2.28}$$

where, for $\theta \to 0$,

$$\begin{aligned} R_{\mu,\sigma} &= O(\theta^{2\mu}), \qquad \sigma \text{ even}, \\ &= O(\theta^{2\mu+1}), \qquad \sigma \text{ odd}. \end{aligned} \tag{2.29}$$

(ii) $$g_{\mu,\sigma}(\theta) = (2 \sin \tfrac{1}{2}\theta)^{2\mu}\, Q_{\sigma}(\theta), \tag{2.30}$$

$$Q_{\sigma}(\theta) = \sum_{l} (2\pi l + \theta)^{-(2\mu-\sigma)}. \tag{2.31}$$

We may rewrite (2.19) and (2.20) in the forms

$$\begin{aligned} \Pi_K(\theta, h) = &-q\left[\frac{-c(Q_0Q_4 - Q_1Q_3) + h^{-2}(Q_0Q_6 - Q_3^2)}{Q_0^2 + q^2Q_3^2}\right] \\ &+ i\left[\frac{-c(g_{\mu,0}g_{\mu,1} + q^2g_{\mu,4}g_{\mu,3}) + h^{-2}(g_{\mu,0} + q^2g_{\mu,6})\, g_{\mu,3}}{g_{\mu,0}^2 + q^2g_{\mu,3}^2}\right], \end{aligned} \tag{2.32}$$

$$\begin{aligned} \Pi_R(\theta, h) = &-qc\left[\frac{(Q_0Q_2 - Q_1^2) + h^{-2}(Q_2^2 - Q_1Q_3)}{(Q_0 + h^{-2}Q_2)^2 + q^2(Q_1 + h^{-2}Q_3)^2}\right] \\ &+ ic\left[\frac{(g_{\mu,0} + q^2g_{\mu,2})\, g_{\mu,1} + h^{-2}(g_{\mu,1} + q^2g_{\mu,3})\, g_{\mu,2}}{(g_{\mu,0} + h^{-2}g_{\mu,2})^2 + q^2(g_{\mu,1} + h^{-2}g_{\mu,3})^2}\right]. \end{aligned} \tag{2.33}$$

The accuracy of the schemes (2.9) and (2.10) may be found by substituting (2.28) and (2.29) into the imaginary parts of (2.32) and (2.33):

$$\operatorname{Im}(\Pi_K) = -\theta\left(c - \frac{\theta^2}{h^2}\right) + O(\theta^{2\mu+1});$$

$$\operatorname{Im}(\Pi_R) = -\frac{c\theta}{1 + \theta^2/h^2} + O(\theta^{2\mu+1}).$$

Hence, by (2.24) both schemes have accuracy of order $2(\mu - 1)$, independently of q.

Next, in order to find the dissipativity, we need to prove some preliminary results for the Q_{σ}, defined in (2.31) (Wahlbin [12] proves a special case of the following).

Since $\mathrm{Re}(\Pi)$ is quadratic in Q_σ, we need consider only the range $0 < \theta \leqslant \pi$ in (2.25). Then, since $2\pi - \theta \geqslant \theta > 0$, we have from (2.31) (since $\sigma \leqslant 2\mu$)

$$Q_\sigma(\theta) = \sum_{l=0}^{\infty} \{(\theta + 2\pi l)^{\sigma-2\mu} + (\theta - 2\pi - 2\pi l)^{\sigma-2\mu}\} \geqslant 0.$$

Also, for $0 < \theta \leqslant \pi$,

$$Q_\sigma(\theta) = \theta^{-(2\mu-\sigma)} + \sum_{l=1}^{\infty} \frac{(2\pi l - \theta)^{2\mu-\sigma} + (-1)^\sigma (2\pi l + \theta)^{2\mu-\sigma}}{((2\pi l)^2 - \theta^2)^{2\mu-\sigma}}. \tag{2.34}$$

If σ is even, then the summation in (2.34) is positive; if σ is odd, it may be written as $-r_\sigma(\theta)$, where $r_\sigma(\theta) \geqslant 0$, so that

$$\theta^{-(2\mu-\sigma)} \geqslant Q_\sigma(\theta) = \theta^{-(2\mu-\sigma)} - r_\sigma(\theta) \geqslant 0 \qquad (\sigma \text{ odd}). \tag{2.35}$$

If σ is even, then there exist positive constants $c_1{}^\sigma$, $c_2{}^\sigma$ such that

$$\theta^{-(2\mu-\sigma)} + c_1{}^\sigma \geqslant Q_\sigma(\theta) \geqslant \theta^{-(2\mu-\sigma)} + c_2{}^\sigma \qquad (\sigma \text{ even}). \tag{2.36}$$

for $0 < \theta \leqslant \pi$.

Now consider the expressions $Q_rQ_s - Q_tQ_u$, where r and s are even, t and u are odd, and $r + s = t + u$, which occur in the real parts of (2.32) and (2.33). From (2.35) and (2.36) we have

$$\begin{aligned}(\theta^{-(2\mu-r)} + c_1{}^r)(\theta^{-(2\mu-s)} + c_1{}^s) - (\theta^{-(2\mu-t)} - r_t)(\theta^{-(2\mu-u)} - r_u) \\ \geqslant Q_rQ_s - Q_tQ_u \geqslant (\theta^{-(2\mu-r)} + c_1{}^r)(\theta^{-(2\mu-s)} + c_2{}^s) - \theta^{-(4\mu-t-u)}.\end{aligned}$$

Hence, for positive $\tilde{c}_1$, $\tilde{c}_2$, and for $\bar{s} = \min\{r, s, t, u\}$,

$$\tilde{c}_1\theta^{-(2\mu-\bar{s})} \geqslant Q_rQ_s - Q_tQ_u \geqslant \tilde{c}_2\theta^{-(2\mu-\bar{s})}. \tag{2.37}$$

This shows that $Q_rQ_s - Q_tQ_u$ is positive. We may conclude, from the real parts in (2.32) and (2.33), that, for $q > 0$:

(i) The KdV scheme (2.6) or (2.9) is *conditionally stable*, and dissipative of order 2μ, for $c > 0$; and *unconditionally stable* for $c < 0$. In the former case, we may make the scheme stable by choosing the value of h sufficiently small.

(ii) The RLW scheme (2.7) or (2.10) is *stable*, and dissipative of order 2μ, for all h and for $c > 0$; but *unstable* for $c < 0$.

Finally, we note that optimal L_2 error bounds have been proved for the KdV scheme (2.6) (Wahlbin [12]). Also, optimal L_2, H_1 and L_∞ error bounds exist for the non-dissipative RLW scheme (2.10) with $q = 0$, in which a second-order accurate time-discretisation is used (Wahlbin [13]).

3. Numerical Computations

In this section, we shall give details of the computational procedure used for implementing the schemes

$$\text{KdV:} \quad (U_t + UU_x + \epsilon U_{xxx}\,,\ \chi + q\epsilon h^3 \chi_{xxx}) = 0 \qquad (\epsilon > 0); \tag{3.1}$$

$$\text{RLW:} \quad (U_t + UU_x - \delta U_{xxt}\,,\ \chi + q\delta^{1/2} h \chi_x) = 0 \qquad (\delta > 0), \tag{3.2}$$

where, in Section 4, the results with $q = 0$ (non-dissipative scheme) only are presented for the RLW equation.

The interval $I_2 = [0, 2]$ of the x-axis is uniformly partitioned into N sub-intervals, each of which has length $h = 2/N$. The basis functions $\phi_i \in S^\mu$, defined by (2.5), and which have support in the interval I_2, are those with $-\mu/2 < i < N + \mu/2$. We shall be dealing with solutions which have their zeroth, first and second derivatives equal to zero on the boundaries: $x = 0,2$ (the solition solutions possess this property). We shall construct special "boundary functions," which do *not* belong to S^μ, and possess the property that the first two derivatives vanish at the boundaries. The numerical computations were carried out for $\mu = 4$ (i.e., the cubic splines); and, for this case, one may choose two such boundary functions, to replace ϕ_1 and ϕ_{N-1}, together with the set $\{\phi_2\,,\ \phi_3\,,\ldots,\ \phi_{N-2}\}$, in order to provide a basis B in S^μ for the solution U. Note that $\phi_j(x)$ has support on the interval $[(j - \mu/2)h,\ (j + \mu/2)h]$.

If we augment the space S^μ by these special boundary functions, and denote the set of functions B by $\{v_1\,,\ v_2\,,\ldots,\ v_{N-1}\}$ ($v_i = \phi_i\,,\ 2 \leqslant i \leqslant N - 2$; $v_1\,,\ v_{N-1}$ being the boundary functions), then the solution U may be expressed as

$$U(x)(t) = \sum_{j=1}^{N-1} \alpha_j(t)\, v_j(x) \tag{3.3}$$

3.1. *Calculation of the Galerkin Constants*

Define the following set of (constant) quantities:

$$\begin{aligned}
A_{ij} &= h^{-1}(v_i\,,\ v_j);\\
B_{ijk} &= (v_i\,,\ v_j v_k^{(1)});\\
C_{ij} &= h^2(v_i^{(1)},\ v_j^{(2)});\\
D_{ijk} &= h^3(v_i^{(3)},\ v_j v_k^{(1)});\\
E_{ij} &= h^5(v_i^{(3)},\ v_j^{(3)});\\
F_{ij} &= h(v_i^{(1)},\ v_j^{(1)});\\
G_{ijk} &= h(v_i^{(1)},\ v_j v_k^{(1)});\\
H_{ij} &= (v_i^{(1)},\ v_j),
\end{aligned} \tag{3.4}$$

where $\mu \geqslant 4$. Thus, A, B,..., H are independent of h. In terms of these definitions, substitution of (3.3) into (3.1) and (3.2) yields the following sets of ordinary differential equations for α_j (summation over the repeated indices from 1 to $N-1$ is implied):

$$\text{KdV:} \quad (A_{ij} + q\epsilon C_{ij}) \frac{d\alpha_j}{dt} + h^{-1}(B_{ijk} + q\epsilon D_{ijk})\, \alpha_j \alpha_k - \epsilon h^{-3}(C_{ij} - q\epsilon E_{ij})\, \alpha_j = 0; \tag{3.5}$$

$$\text{RLW:} \quad [A_{ij} + \delta h^{-2} F_{ij} + q\delta^{1/2}(H_{ij} - \delta h^{-2} C_{ij})] \frac{d\alpha_j}{dt} + h^{-1}[B_{ijk} + q\delta^{1/2} G_{ijk}]\, \alpha_j \alpha_k = 0. \tag{3.6}$$

Now, if none of i, j, or k is 1 or $(N-1)$, then we may use the translation property of the ϕ_i's, as defined by equation (2.5), to suppress the index i of each of the symbols A,..., H in (3.4), since these quantities are functions only of $(j-i)$ and $(k-i)$. Thus retaining the same symbols and reducing the number of indices by one, we define $A_\beta = A_{i,i+\beta}$ $B_{\beta\gamma} = B_{i,i+\beta,i+\gamma}$, etc., where $|\beta|, |\gamma|, |\beta - \gamma| \leqslant \mu - 1$ (since values outside these ranges are zero). Thus, we may write (3.5) and (3.6) in the equivalent forms:

$$\text{KdV:} \quad (A_\beta + q\epsilon C_\beta) \frac{d}{dt}(\alpha_{i+\beta}) + h^{-1}(B_{\beta\gamma} + q\epsilon D_{\beta\gamma})\, \alpha_{i+\beta}\alpha_{i+\gamma} - \epsilon h^{-3}(C_\beta - q\epsilon E_\beta)\, \alpha_{i+\beta} = 0; \tag{3.7}$$

$$\text{RLW:} \quad [A_\beta + \delta h^{-2} F_\beta + q\delta^{1/2}(H_\beta - \delta h^{-2} C_\beta)] \frac{d}{dt}(\alpha_{i+\beta}) + h^{-1}[B_{\beta\gamma} + q\delta^{1/2} G_{\beta\gamma}]\, \alpha_{i+\beta}\alpha_{i+\gamma} = 0, \tag{3.8}$$

where summation from $-(\mu-1)$ to $(\mu-1)$ is implied by the repeated indices β and γ, subject to the restriction $2 \leqslant i+\beta$, $i+\gamma \leqslant N-2$. In order to avoid the complication of the boundary functions, we may extend the allowed range $1 \leqslant i+\beta$, $i+\gamma \leqslant N-1$, if we replace v_1 and v_{N-1} by ϕ_1 and ϕ_{N-1}, respectively (thus preserving the translation property), and apply "boundary corrections" as follows. Define starred quantities to be those calculated using the "incorrect" boundary functions ϕ_1 and ϕ_{N-1}, and unstarred quantities those calculated using the "correct" ones v_1 and v_{N-1}, whenever one of the indices i, j or k in (3.4) has the value 1 or $(N-1)$. The corrections to be applied in (3.7) and (3.8) will then be of the forms

$$\text{KdV:} \quad \delta_{ijk} = (B_{ijk} - B^*_{ijk}) + q\epsilon(D_{ijk} - D^*_{ijk}); \quad \eta_{ij} = (C_{ij} - C^*_{ij}) - q\epsilon(E_{ij} - E^*_{ij}); \tag{3.9}$$

$$\text{RLW:} \quad \delta_{ijk} = (B_{ijk} - B^*_{ijk}) + q\delta^{1/2}(G_{ijk} - G^*_{ijk}), \tag{3.10}$$

whenever one or more of the indices has value 1 or $(N-1)$. Note that in the actual coding, the matrix (P, say) of coefficients multiplying $d\alpha/dt$ in (3.5) or (3.6) was calculated explicitly and stored (in band mode), since the o.d.e. solver requires an

LU factorization of the matrix in order to solve a system of the form $P d\alpha/dt + \mathbf{g} = 0$, as in (3.5) and (3.6). Hence, boundary corrections were applied only in the subroutine which computes the vector $\mathbf{g} = \mathbf{g}(\alpha)$.

The computations were carried out in the case $\mu = 4$, for which definition (2.1) or (2.3) yields the following cubic splines (Schoenberg [10]):

$$\begin{aligned} M_4(y) &= 0, && y \leqslant -2; \\ &= \tfrac{1}{6}(y+2)^3, && -2 < y \leqslant -1; \\ &= \tfrac{1}{6}[(y+2)^3 - 4(y+1)^3], && -1 < y \leqslant 0; \\ &= \tfrac{1}{6}[(-y+2)^3 - 4(-y+1)^3], && 0 < y \leqslant 1; \\ &= \tfrac{1}{6}(-y+2)^3, && 1 < y \leqslant 2; \\ &= 0, && 2 < y. \end{aligned} \tag{3.11}$$

Then,

$$\begin{aligned} \phi_i(x) &= M_4\left(\frac{x}{h} - i\right) \\ &= v_i(x), \qquad \text{for} \quad 2 \leqslant i \leqslant N-2. \end{aligned}$$

The boundary functions v_1, v_{N-1} must satisfy

$$v_1(0) = v_1^{(1)}(0) = v_1^{(2)}(0) = 0; \qquad v_{N-1}(2) = v_{N-1}^{(1)}(2) = v_{N-1}^{(2)}(2) = 0;$$

and, furthermore, match the zeroth, first and second derivatives at the nodes: $x = h$ and $x = (N-1)h$, respectively. This constraints these functions to be (unique) quintic polynomials in the intervals $[0, h]$ and $[(N-1)h, Nh]$. Thus let

$$\begin{aligned} \Phi(y) &= 0, && y \leqslant -1 \\ &= (y+1)^3\,[3(y+1)^2 - \delta(y+1) + \tfrac{17}{3}], && -1 < y \leqslant 0; \\ &= \tfrac{1}{6}[(-y+2)^3 - 4(-y+1)^3], && 0 < y \leqslant 1; \\ &= \tfrac{1}{6}(-y+2)^3, && 1 < y \leqslant 2; \\ &= 0, && 2 < y. \end{aligned} \tag{3.12}$$

Then,

$$v_1(x) = \Phi\left(\frac{x}{h} - 1\right); \qquad v_{N-1}(x) = \Phi\left(-\frac{x}{h} + (N-1)\right). \tag{3.13}$$

The solution was evolved in time by solving the systems of ordinary differential equations (3.7) using the IMSL library (1975) routine DREBS with the order of the method set equal to 2. An error tolerance for the integration was set to 2×10^{-2}.

Note: The computation of the integrals in (3.4) involving piecewise polynomials

(3.11) and (3.12), and their derivatives, is a tedious, but straightforward process, which could well be automated. Since the polynomials are defined piecewise on unit intervals of (x/h), the integrals can be reduced to the form:

$$\int_0^1 [a_0 + a_1 y + \cdots + a_p y^p]\, dy = \sum_{k=0}^{p} a_k/(k+1)$$

Thus, in general, we would require a subroutine to multiply together two or three polynomials of degree $\leqslant(\mu - 1)$ by storing the coefficients of the polynomials in arrays, and implementing an algorithm for "long multiplication" on these coefficients, to yields $a_0, \ldots, a_p$ above.

3.2. *Computation of* $U(0)$

When solving the systems (3.1) or (3.2) of ordinary, first-order differential equations in time, we require a method for determining the initial values $\alpha_j(0)$, $1 \leqslant j \leqslant N - 1$. Thus, if we substitute the form (3.3) with $t = 0$ into the following expression:

$$\text{KdV:}\quad \xi_1 \equiv \Big\| \sum_j \alpha_j(0)\, v_j(\cdot) - f_0(\cdot) \Big\|_{L_2}^2 = \Big(\sum_j \alpha_j(0)\, v_j - f_0 \,,\, \sum_k \alpha_k(0)\, v_k - f_0 \Big) \tag{3.14}$$

and set $\partial \xi_1/\partial \alpha_i = 0$, then we obtain the following equation (using definitions (3.4)), from which to compute $\boldsymbol{\alpha}(0)$:

$$\sum_j A_{ij} \alpha_j(0) = (f_0 \,,\, v_i)/h; \tag{3.15}$$

(3.15) represents the orthogonal projection of f_0 into S^μ. In order to evaluate the integrals in (3.15) we employed the integration routine DCADRE from the IMSL library (IMSL 1975) using relative and absolute error tolerances equal to 10^{-4}.

4. Discussion of the Numerical Results

In this section we present plots of the 1- and 2-soliton solutions for both the KdV and RLW equations. The KdV equation is more extensively treated: the dissipation parameter q was chosen to assume several values, in order to investigate the effect of introducing different amounts of dissipation into the numerical scheme, and its effects on the propagation, with time, of the errors in the solution. The results of the RLW equation with $q = 0$ and $N = 40$ are so good in comparison with the KdV equation that we have not included other RLW results here.

4.1. *Single-Soliton Solutions*

The initial function for the 1-soliton solution was computed from (cf eq. (1.3))

$$f_0(x) = 3c\ sech^2\ (kx + d) \tag{4.1}$$

in which $k = \frac{1}{2}(c/\epsilon)^{1/2} = \frac{1}{2}(1/\delta)^{1/2}$; d = constant.
The following values of the constants were used

$$c = 0.3;\ \epsilon = 4.84(10^{-4});\ d = -k. \tag{4.2}$$

The results for the KdV equation with the 1-soliton solution are summarized in Table 4.1. The time / 1 real second column represents the time in seconds required to evolve the solution and produce the graphical results for 1 second of the real time for the partial differential equation. These figures indicate what one expects, realistically, to have to pay for in computer time for the complete execution of the computer program. The differences in the entries of this column (of all the tables 4.1–4.5) represent real differences in computer time.

Table 4.2 contains corresponding results for $N = 60$ and in Table 4.3 the results for the single value $q = 0$ are summarized. A comparison of the final column in the tables indicates that the maximum error (the difference between computed solution and theoretical solution) is considerably smaller for the RLW equation than the KdV equation for both $N = 40$ and $N = 60$. Also it is noted that the real cost for the RLW equation is significantly less.

In figures 4.3 through 4.10 the actual solutions obtained are graphed for visual comparison.

Figure 4.3 depicts the initial condition where, as in all of the graphical figures, we have enlarged the vertical scale to emphasize the existence of any spurious oscillations that occur.

Figure 4.4 depicts the numerical solution at $t = 0.3958$ of the KdV equation using $q = 0$ for $N = 40$ grid points. The presence of the downstream oscillations are clearly visible. The effect of refining the mesh is apparent in figure 4.5 where the same parameters are used, except $N = 60$. The downstream oscillations are removed at the expense of introducing (smaller amplitude) oscillations behind the soliton. However, as mentioned above, this improvement is obtained at the cost of considerably increasing the time required. The effect of modifying q is shown in figures 4.6 and 4.7. In the former $q = 20.661$ is used and in the latter $q = 103.31$; for both cases $N = 40$. There is clearly a considerable improvement in these figures over the results depicted in figure 4.4. On the basis of the maximum error given in table 4.1 and the graphical results given in figures 4.6 and 4.7 there would appear to be little to choose between the two values of q. However, on the basis of cost it is apparent (from table 4.1) that the larger q has incurred more computer time.

Figure 4.8 indicates that $q = 1.0$ (Wahlbin's scheme) seems not to improve the results over the case $q = 0.0$. However using this value of $q = 1.0$ and refining the

TABLE 4.1

Figure	q	Time/1 real second (seconds)	Max. error
4.4	0	23.5	0.057
4.6	20.661	14.2	0.059
4.7	103.31	30.6	0.025
4.8	1.0	15.7	0.027

(Legend) KdV equation: single-soliton solution: $N = 40$

TABLE 4.2

Figure	q	Time/1 real second (seconds)	Max. error
4.5	0	63.1	0.0153
4.9	1.0	58.4	0.018

(Legend) KdV equation: single-soliton solution: $N = 60$

TABLE 4.3

Figure	q	Time/1 real second (seconds)	Max. error
4.10	0	7.6	0.0039

(Legend) RLW equation: single-soliton solution: $N = 40$

TABLE 4.4

Figure	q	Time/1 real second (seconds)
4.12	0	20.9
4.13	1	21.5

(Legend) KdV equation: two-soliton solution: $N = 40$

TABLE 4.5

Figure	q	Time/1 real second (seconds)
4.15	0	16.7

(Legend) RLW equation: two-soliton solution: $N = 40$

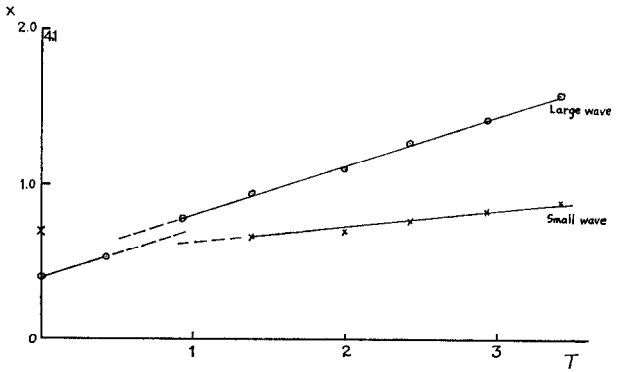

FIG. 4.1. KdV equation, two soliton solution's phase.

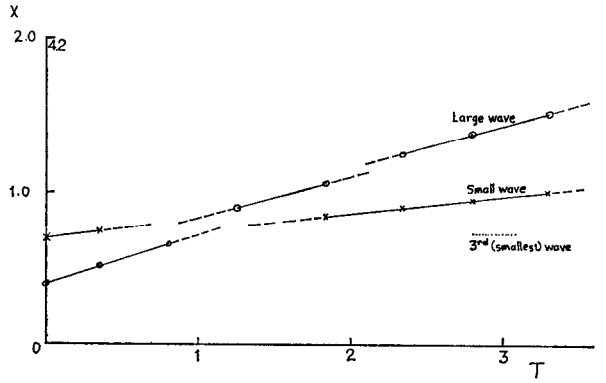

FIG. 4.2. RLW equation two soliton solution's phase.

mesh to $N = 60$ produces much improved results (once again at a much increased cost—see table 4.2).

In computing the solution of the RLW equation with $q = 0.0$, the results depicted in figure 4.10, the absence of oscillations are apparent and the general resolution of the initial soliton, moved to the right with preserved amplitude and shape, is good. As indicated by table 4.3 the cost is a minimum.

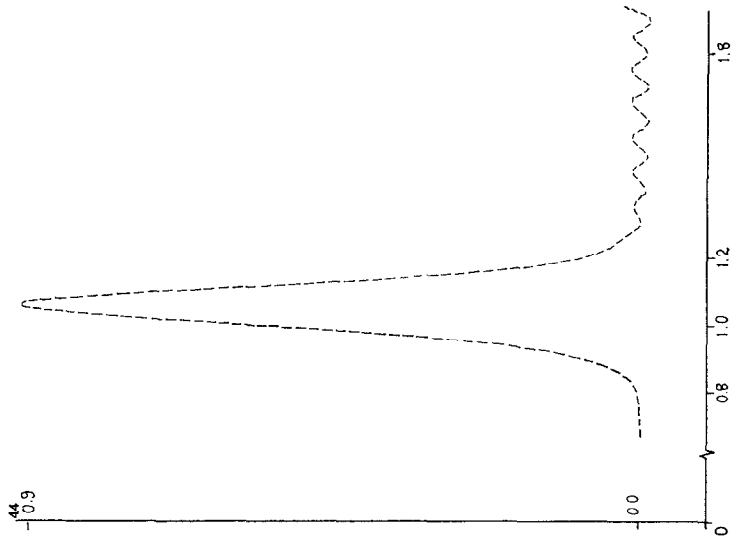

FIG. 4.4. $T = 0.3958$, KdV, $q = 0$, $N = 40$.

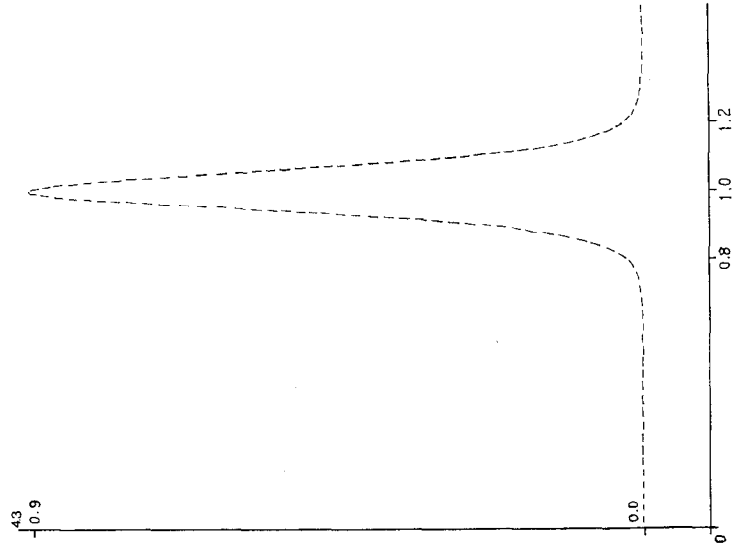

FIG. 4.3. Initial condition 4.1.

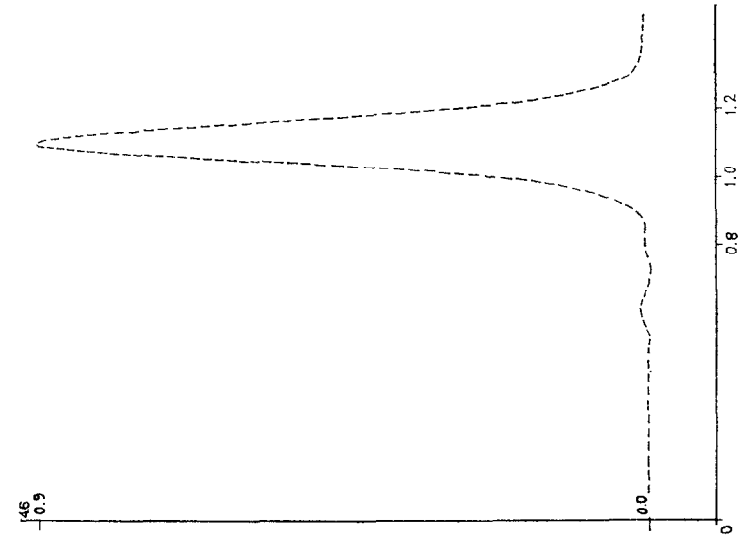

FIG. 4.6. $T = 0.3605$, KdV, $q = 20.661$, $N = 40$.

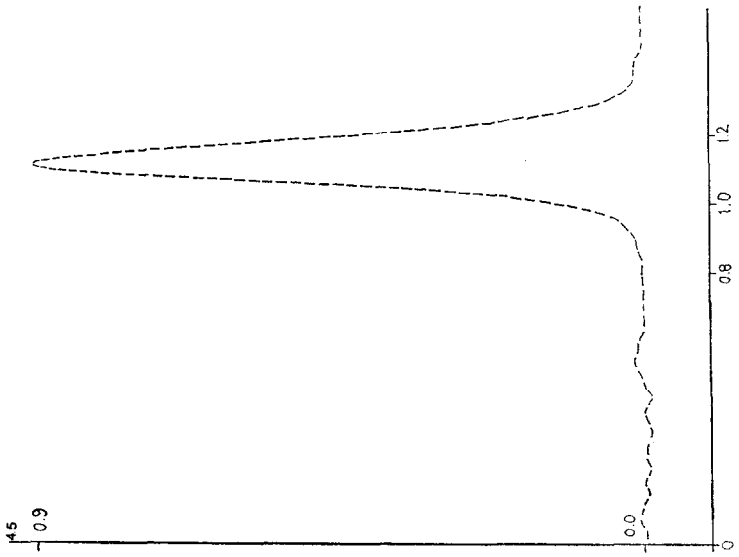

FIG. 4.5. $T = 0.4626$, KdV, $q = 0$, $N = 60$.

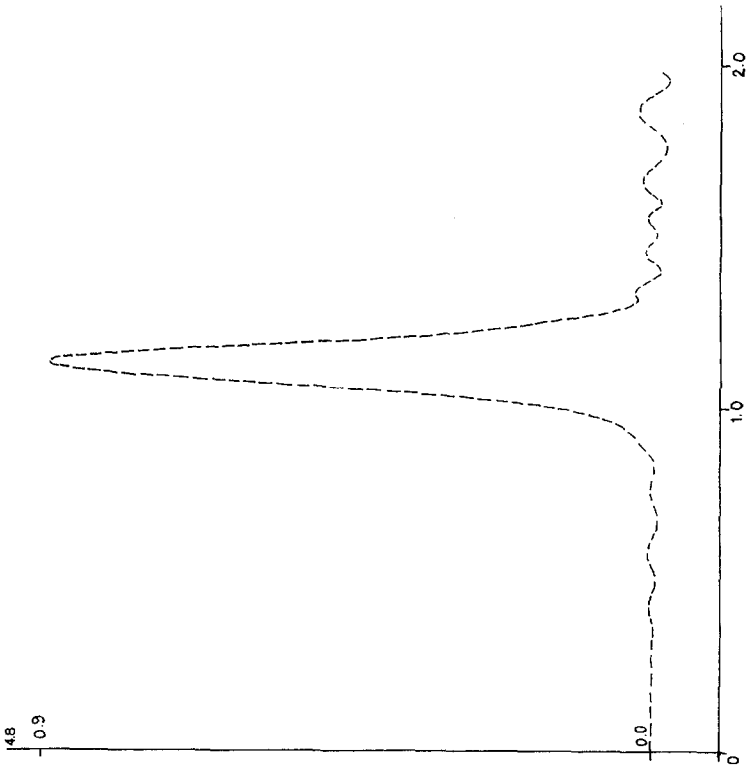

FIG. 4.8. $T = 0.4155$, KdV, $q = 1.0$, $N = 40$.

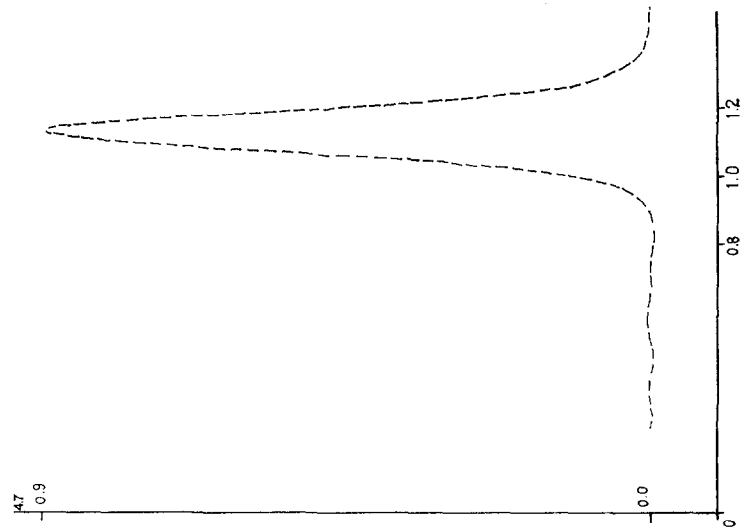

FIG. 4.7. $T = 0.4732$, KdV, $q = 103.31$, $N = 40$.

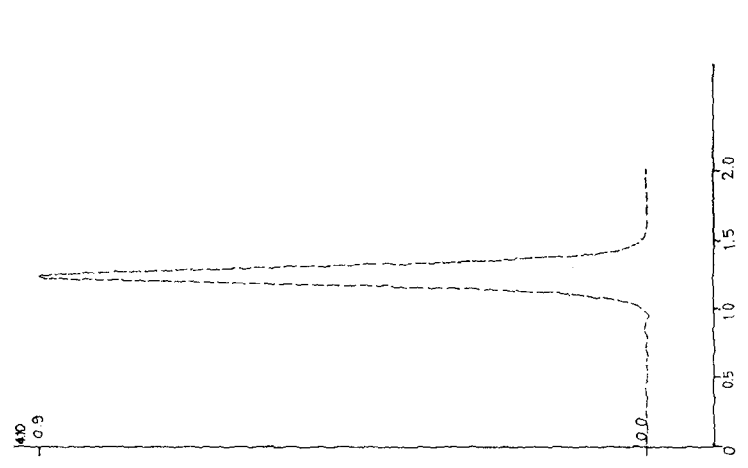

FIG. 4.10. $T = 0.8347$, RLW, $q = 0.0$, $N = 40$.

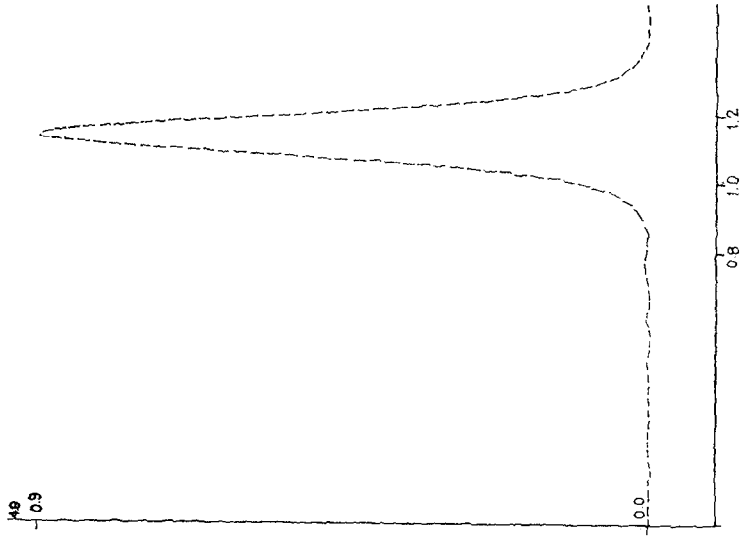

FIG. 4.9. $T = 0.4617$, KdV, $q = 1.0$, $N = 60$.

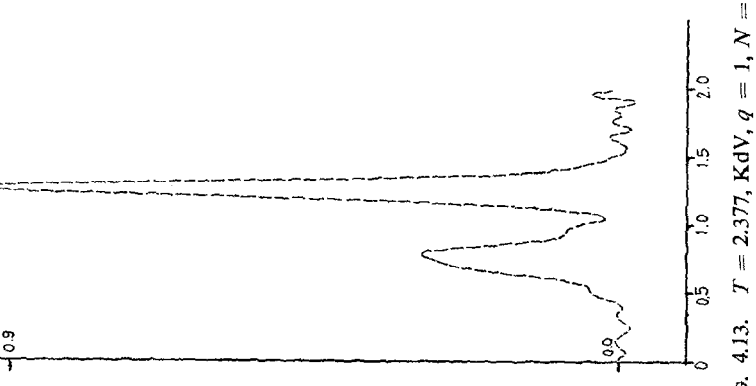

FIG. 4.13. $T = 2.377$, KdV, $q = 1$, $N = 40$.

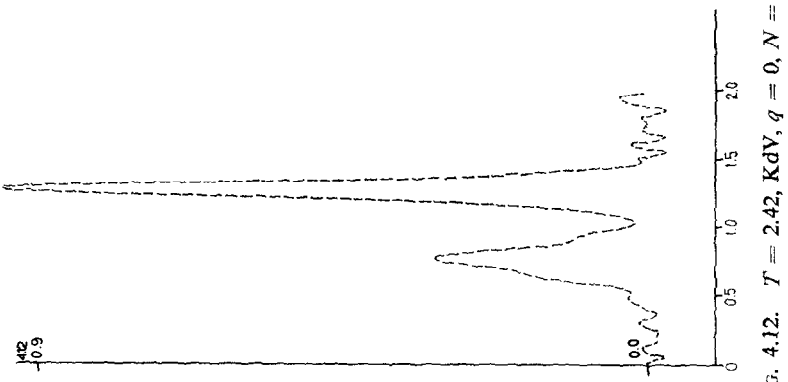

FIG. 4.12. $T = 2.42$, KdV, $q = 0$, $N = 40$.

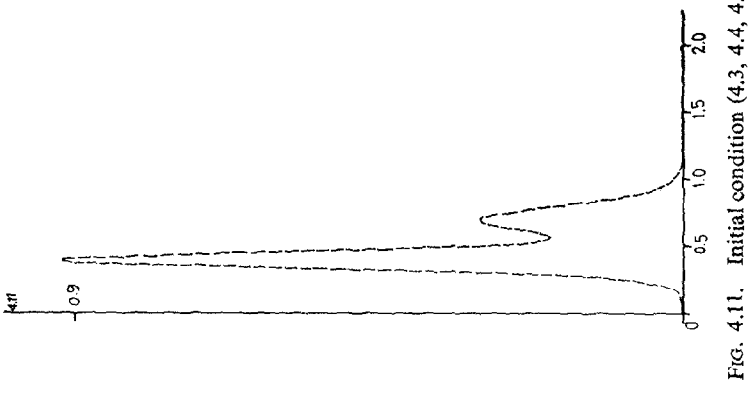

FIG. 4.11. Initial condition (4.3, 4.4, 4.5).

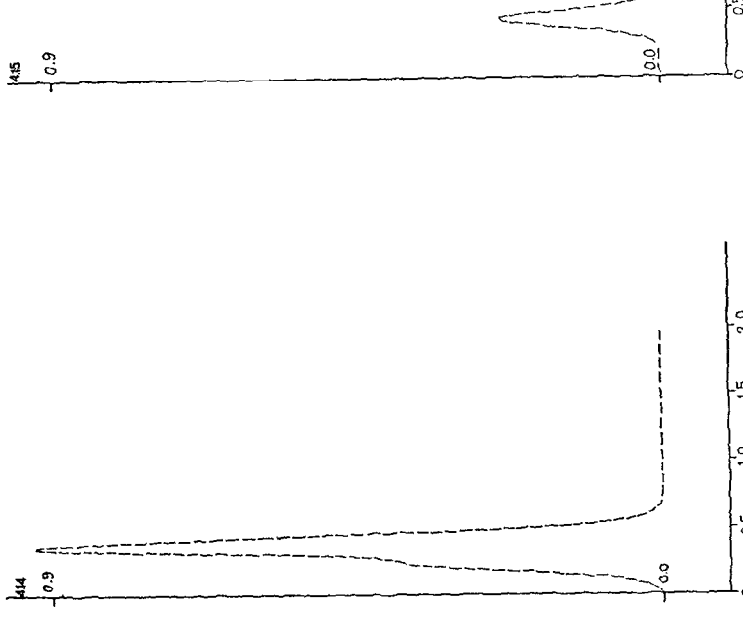

FIG. 4.14. Initial condition (4.3, 4.4, 4.6).

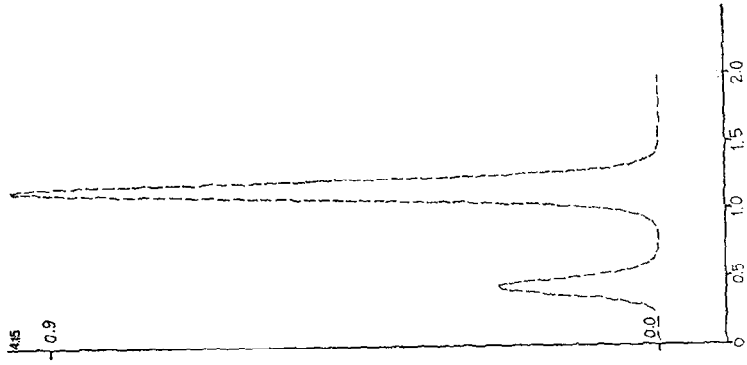

FIG. 4.15. $T = 2.36$, RLW, $q = 0$, $N = 40$.

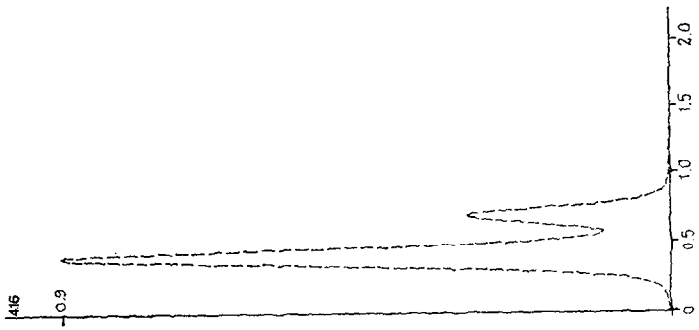

FIG. 4.16. Initial condition (4.3, 4.4, 4.7).

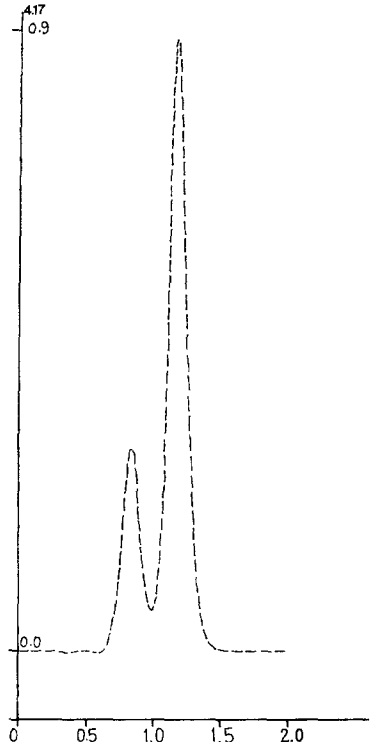

FIG. 4.17. T 2.337, RLW, $q = 0$, $N = 40$.

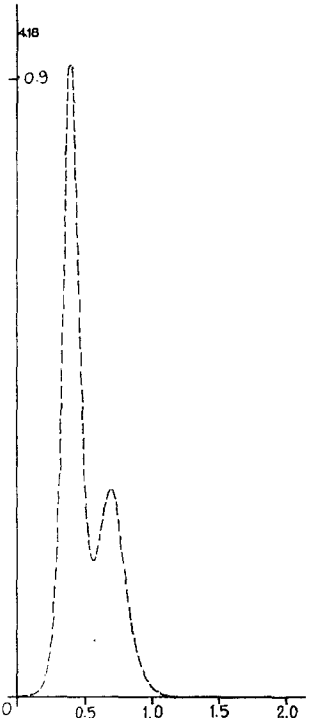

FIG. 4.18. Initial condition (4.3, 4.4, 4.8).

4.2. *Two-Soliton Solutions*

The initial function for the 2-soliton solution was computed from

$$f_0(x) = 3c_1\, sech^2(K_1x + d_1) + 3c_2\, sech^2(K_2x + d_2), \tag{4.3}$$

where the following values were chosen:

$$c_1 = 0.3;\ c_2 = 0.1 \tag{4.4}$$

and

$$\text{KdV:}\quad K_i = \frac{1}{2}\left(\frac{c_i}{\epsilon}\right)^{1/2};\qquad d_i = -5,\quad i = 1, 2. \tag{4.5}$$

These values coincide with those used by Greig [6]. The numerical results are depicted in figures 4.12 for $q = 0$ at $T = 2.42$ seconds and figure 4.13 for $q = 1$ at $T = 2.377$ where $N = 40$ in both cases. The initial condition (4.3) with the parameters given by (4.4) and (4.5) is shown in figure 4.11. The initial profile has been propagated to a stage where the larger soliton has passed through the smaller soliton and emerged

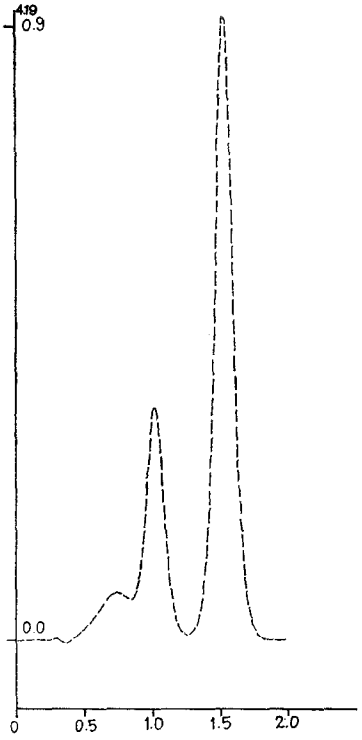

FIG. 4.19. $T = 3.297$, RLW, $q = 0$, $N = 40$.

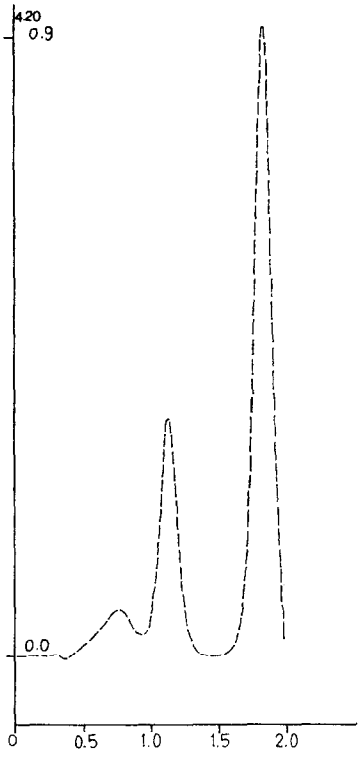

FIG. 4.20. $T = 4.258$, RLW, $q = 0$, $N = 40$.

with their positions interchanged. However, the presence of both upstream and downstream oscillations for $q = 0$ and $q = 1$ is clear. The times for these runs are summarized in table 4.4.

For the RLW equation we tried *three* initial conditions. The first was defined using (4.3) with

$$K_i = \frac{1}{2}\left(\frac{1}{\delta}\right)^{1/2}; \qquad K_2 = K_1; \qquad d_i = -5; \qquad d_2 = -3 \tag{4.6}$$

whose profile is shown in figure 4.14. Using $q = 0$ and $N = 40$ the solution was computed to $T = 2.36$ and is shown in figure 4.15. The expected resolution of the two solitons is seen, without any oscillations.

The second initial condition was (4.3) with the parameters defined by

$$K_i = \frac{1}{2}\left(\frac{1}{\delta}\right)^{1/2}; \qquad K_2 = K_1; \qquad d_1 = -5; \qquad d_2 = \left(\frac{c_1}{c_2}\right)^{1/2} d_1 \tag{4.7}$$

whose profile is shown in figure 4.16.

The solution computed at $T = 2.337$ with $q = 0$ and $N = 40$ is shown in figure 4.17.

A very small oscillation is present behind the solitons. We assume these could be dissipated with a suitable positive q.

Finally we ran the initial condition (4.3, 4.4) and chose

$$K_1 = \frac{1}{2}\left(\frac{1}{\delta}\right)^{1/2} (\delta = \epsilon/c_1); \qquad K_2 = \left(\frac{c_1}{c_2}\right)^{1/2} K_1 ; \qquad d_i = -5 \tag{4.8}$$

This produces exactly the same profile as that used for the KdV equation depicted in figure 4.18 and equivalently in figure 4.11.

Using $q = 0$ and $N = 40$ the solution at $T = 3.297$ and later at $T = 4.258$ are shown in figures 4.19 and 4.20 respectively. In this case we perceive an unexpected feature. Namely the emergence of a third (very) small amplitude soliton. This appears to be a genuine soliton (rather than spurious oscillation) as its emergence causes a phase shift on the largest soliton (and perhaps also on the second soliton although on the scale used this could not be depicted) as the results of figure 4.2 show. The corresponding phase diagram of the 2-soliton solution for the KdV equation is given in figure 4.1. Abdulloev *et al.* [1] have also performed computer experiments on a modified form of the RLW equation, and likewise note the appearance of a third wave subsequent to the interaction of two solitons. Thus, the "inelastic" behaviour of the RLW equation receives further confirmation. On the other hand, the finite difference schemes of Eilbeck and McGuire [4]) do not appear to reproduce this phenomenon.

In figure 4.1 both waves emerge from the interaction with the *same* velocity (and amplitude) as existed before the interaction but are shifted in phase by the amounts $\Delta_1 = +0.09$ and $\Delta_2 = -0.19$, respectively.

In figure 4.2 the two principal waves for the RLW equation do not preserve their velocity after the interaction. The phase shifts after the interaction appear to be $\Delta_1 = +0.09$ for the larger soliton and $\Delta_2 = -0.05$ for the smaller. However, when the (very) small soliton emerges the largest wave appears to undergo yet another phase shift of $\Delta_1' = 0.05$.

5. Conclusions

The purpose of this report has been to explore the numerical properties of the KdV and RLW equations, using the ordinary as well as a class of dissipative Galerkin methods. Although only cubic splines have been used as basis functions, it became clear that, in terms of both economy of run-time and accuracy for a given mesh-size, the RLW equation may be solved much more efficiently and effectively than the KdV equation. (We ascribe this difference to the fact that the RLW equation has only second order spatial derivatives whereas the KdV equation has third order spatial derivatives.) Hence, if the RLW equation is to be used to describe the shallow water wave model, as advocated by Benjamin *et al.* [2], as opposed to the KdV equation, then it appears to be a fact that the Galerkin method treats the RLW equation more

successfully than the KdV equation. However, the final figures act as a warning. There appear to be, for the same initial profile for both KdV and RLW equations, a distinct difference in the behaviour of the 2-soliton solutions with increasing time. A comparison with the Finite Difference Method (Greig [6, table 6.2]) shows that the Galerkin method has the advantages of smaller errors (for the same mesh-size) and of not being restricted as to the size of the time step in the ordinary differential equation solver. (The only constraint imposed here is by virtue of a preassigned choice of accuracy of the solutions—as an alternative, a simple Trapezoidal-like discretization of the ordinary differential equation system would be interesting.) In contrast the finite difference methods of Greig were conditionally stable. For the RLW equation the methods of Eilbeck and McGuire are, however, unconditionally stable.

ACKNOWLEDGMENTS

We would like to thank one of the referees for bringing to our attention the work of Abdulloev *et al.* [1] This work was partially supported by a National Research Council of Canada grant A3597.

REFERENCES

1. KH. O. ABDULLOEV, I. L. BOGOLUBSKY, AND V. B. MAKHANKOV, *Phys. Lett.* **56A** (1976), 427–428.
2. T. B. BENJAMIN, J. L. BONA, AND J. J. MAHONEY, *Philos. Trans. Roy. Soc. London Ser. A* **272** (1972), 47–48.
3. J. C. EILBECK AND G. R. MCGUIRE, *J. Computational Physics* **19** (1975), 43–57.
4. J. C. EILBECK AND G. R. MCGUIRE, *J. Computational Physics* **23** (1977), 63–73.
5. G. S. GARDNER, J. M. GREEN, M. D. KRUSKAL, AND R. M. MIURA, *Phys. Rev. Lett.* **19** (1967), 1095–1097.
6. I. S. GREIG, *J. Computational Physics* **20** (1976), 64–80.
7. IMSL: International Mathematical and Statistical Libraries, Houston, Tex., 1975.
8. D. J. KORTEWEG AND G. DE VRIES, *Philos. Mag.* **39** (1895), 422–443.
9. D. H. PEREGRINE, *J. Fluid Mech.* **25** (1966), 321–330.
10. I. J. SCHOENBERG, *Quart. Appl. Math.* **4** (1946), 45–99.
11. V. THOMÉE AND B. WENDROFF, *SIAM J. Numer. Anal.* **11** (1974), 1059–1068.
12. L. WAHLBIN, (C. de Boor, Ed.), pp. 147–169, Academic Press, New York, 1974.
13. L. WAHLBIN, *Numer. Math.* **23** (1975), 289–303.