ERRING AND BEING CONSERVATIVE

J. de Frutos & J. M. Sanz-Serna

in Numerical Analysis 1993
D F Griffiths & G A Watson
editors
Longman Scientific & Technical
1994
pages 75-88

1 Introduction

The classical analysis of numerical methods for time-dependent, ordinary or partial differential equations is based on the ideas of stability, consistency and convergence. Roughly speaking, consistency means small local errors and stability means that errors do not build up catastrophically. Together, consistency and stability yield convergence: small (global) errors. However it is clear that there are useful theoretical properties of a method beyond its consistency, stability and convergence. Here we are interested in conserved quantities: the differential equations being integrated may possess one or several quantities (mass, energy, etc.) that are conserved in the true evolution and it is reasonable to demand that the numerical scheme also preserves those quantities. Several reasons are usually invoked for using schemes with such conservation properties. In a recent paper [6], C.W. Gear writes "In some cases, failure to maintain certain invariants leads to physically impossible solutions". In other cases conservation quantities are deemed important to avoid spurious blow-up of the numerical solution. In a classical paper [1], Arakawa writes "If we can find a finite difference scheme which has constraints analogous to the integral constraints of the differential form, the solution will not show the false 'noodling', followed by computational instability".

Since, in general, a quantitatively accurate solution cannot be unphysical, the preceding and similar arguments in favour of conservative schemes only apply to long-time integrations. Here the term long-time refers to integrations on time intervals $0 < t < t_{max}$ that are so long, relatively to the step-size Δt being used, that errors become large and the numerical solutions deviate significantly from the theoretical solutions. Therefore the preceding arguments *seem* to imply that *while errors are small* conservation properties are not too important; they become important when t_{max} is so large that the integration goes very wrong quantitatively. In the latter, long-time regime, conservative methods are quantitatively wrong but qualitatively acceptable, while nonconservative numerical solutions may be unacceptable from both the quantitative and the qualitative viewpoints.

However such an assessment of the merits of conservative schemes is too severe. In actual fact, in many cases, conservative schemes have better error propagation mechanisms that render them superior from a quantitative point of view and they should be preferred even for computations where the numerical solution remains close to the theoretical solution. An instance is presented in [3]. It is shown there that, when integrating the two-body problem with some conservative schemes, the leading term of the global error grows linearly with t, while for 'general' schemes the growth is quadratic. This makes conservative methods more efficient than general methods when accurate solutions are needed.

In the present paper we use the Korteweg-de Vries (KdV) equation as a model case. Only soliton solutions are considered, but this particular solution is particularly relevant because other solutions asymptotically give rise to solitons. After presenting the differential equation (Section 2) and the numerical methods considered (Section 3), we describe the behaviour of the numerical solutions by means of soliton perturbation theory (Section 4). It turns out that schemes that preserve the integrals of the solution and the solution squared behave much better than 'general' schemes. Numerical illustrations are presented in Section 5. In the final Section 6 we consider the same issue by using functional analytic techniques. It is shown rigorourly that, for conservative schemes, the leading term of the global error consists of an error in the soliton phase, plus errors that are uniformly bounded for all positive times.

The main observation in the paper is that, if we look at the local error of a numerical method as a vector in a suitable phase space, then conservation properties imply constraints for the direction of the local error. When local errors build up to give rise to the global error, their directions are not irrelevant: there are harmful directions that lead to faster error accumulation. In many instances, the local error of a conservative scheme has a *direction* that renders it relatively harmless and this gives the scheme an advantage. These features are not captured by standard convergence analyses, which just take into account the *size* of the local error.

2 The Korteweg-de Vries equation

We write the KdV equation in the form

$$u_t - 6uu_x + u_{xxx} = 0, \quad -\infty < x < \infty, \quad t > 0. \tag{2.1}$$

The symbol Φ_t, $t \geq 0$, will denote the t-flow of the equation: if $u_0 = u_0(x)$ is a function of the spatial variable x, then $\Phi_t(u_0)$ is the function of x given by $u(\cdot, t)$, where $u(\cdot,\cdot)$ is the solution of (2.1) with initial condition $u(x,0) = u_0(x)$, $-\infty < x < \infty$.

Among the many remarkable properties of (2.1) we focus on two: the existence of solitons and the existence of conservation laws.

A function $\phi(\xi)$, $\xi = x - ct$ provides a solution of (2.1) if ($' \equiv d/d\xi$)

$$-c\phi' - 6\phi\phi' + \phi''' = 0,$$

or, if ϕ and its derivatives vanish at ∞,

$$-c\phi - 3\phi^2 + \phi'' = 0. \tag{2.2}$$

After an elementary integration of this second-order ordinary differential equation, it is found that for $c > 0$ and real d, the function

$$\phi(\xi; c, d) = -\frac{c}{2}\operatorname{sech}^2\frac{\sqrt{c}}{2}(\xi + d), \tag{2.3}$$

provides, through the recipe

$$u(x,t) = \phi(x - ct; c, d), \tag{2.4}$$

a soliton (solitary travelling wave solution) of (2.1) moving at velocity c and initially located at $x = -d$. Obviously, varying the value of d results in a translation of (2.4) along the x-axis. Note that the amplitude or depth of (2.3) is $-c/2$; thus deeper solitons travel faster than shallower solitons.

We now turn our attention to conservation laws for (2.1). There is an infinite number of these, but we only need the first two. For smooth solutions of (2.1) the following quantities do not vary with t:

$$I_1(u) = \int_{-\infty}^{\infty} u(x,t)\,dx, \qquad (2.5)$$

$$I_2(u) = \int_{-\infty}^{\infty} u^2(x,t)\,dx. \qquad (2.6)$$

3 Numerical methods

We consider semidiscrete (discrete t, continuous x) numerical methods for (2.1). Fully discrete and continuous t, discrete x algorithms may also be treated by the techniques in this paper; for brevity they are not included. If Δt denotes the time step and U^n is the numerical solution at time level $t_n = n\Delta t$, then we write the (one-step) method in the symbolic form

$$U^{n+1} = \Psi_{\Delta t}(U^n). \qquad (3.1)$$

The local error (at a function $u_0 = u_0(x)$) is, by definition,

$$L_{\Delta t}(u_0) = \Psi_{\Delta t}(u_0) - \Phi_{\Delta t}(u_0). \qquad (3.2)$$

If p denotes the order of the method, then $L_{\Delta t}(u_0) = O(\Delta t^{p+1})$ for suitably smooth u_0. We also assume the existence, for smooth u_0, of an asymptotic expansion

$$L_{\Delta t}(u_0) = \Delta t^{p+1}\ell_{p+1}(u_0) + o(\Delta t^{p+1}), \qquad (3.3)$$

where ℓ_{p+1} is an operator independent of Δt and involving differentiation with respect to x. This requirement is fulfilled for all methods of practical interest.

If I is a conserved quantity of (2.1), then $\Psi_{\Delta t}$ conserves exactly I if, for all Δt and u_0,

$$I(\Psi_{\Delta t}(u_0)) = I(u_0).$$

From here and $I(\Phi_{\Delta t}(u_0)) = I(u_0)$, it follows, via Taylor expansion, that

$$0 = I(\Psi_{\Delta t}(u_0)) - I(\Phi_{\Delta t}(u_0)) = \Delta t^{p+1} I'_{u_0}(\ell_{p+1}(u_0)) + o(\Delta t^{p+1}),$$

where I'_{u_0} is the first variation of I evaluated at u_0 and acting on $\ell_{p+1}(u_0)$. Hence

$$I'_{u_0}(\ell_{p+1}(u_0)) = 0. \qquad (3.4)$$

For instance, for the functional I_2 in (2.6)

$$I(u_0 + \epsilon v) = I(u_0) + \epsilon \int_{-\infty}^{\infty} 2u_0(x)v(x)\,dx + o(\epsilon),$$

so that

$$I'_{u_0}(v) = \int_{-\infty}^{\infty} 2u_0(x)v(x)\,dx$$

and (3.4) means that for methods that conserve I_2 exactly

$$\forall u_0, \quad \int_{-\infty}^{\infty} 2u_0\ell_{p+1}(u_0)\,dx = 0. \tag{3.5}$$

As pointed out in the introduction it is useful to think of this as an orthogonality relation between $\ell_{p+1}(u_0)$ and u_0.

For I_1 in (2.5) obviously $I'_{u_0} = I_1$ and conservation implies

$$\forall u_0, \quad \int_{-\infty}^{\infty} \ell_{p+1}(u_0)\,dx = 0. \tag{3.6}$$

This relation implies geometrically that $\ell_{p+1}(u_0)$ is contained in the hyperplane of functions with vanishing integral.

It is clear that (3.4) holds not only for methods that conserve I exactly, but for all methods for which

$$I(\Psi_{\Delta t}(u_0)) - I(u_0) = o(\Delta t^{p+1}). \tag{3.7}$$

In what follows, it should be understood that the properties that we prove by using (3.4) for methods that exactly conserve I also hold for methods that satisfy the weaker requirement (3.7).

4 Perturbation theory

Following the well-known methodology of modified equations [14], [7], we now introduce the (modified) equation

$$u_t - 6uu_x + u_{xxx} = \Delta t^p \ell_{p+1}(u), \quad -\infty < x < \infty, \quad t > 0. \tag{4.1}$$

The flow $\tilde{\Phi}_{\Delta t}$ of this equation differs from the method mapping $\Psi_{\Delta t}$ in terms $o(\Delta t^{p+1})$; the right hand side of (4.1) counterbalances the leading term of the difference $\Psi_{\Delta t} - \Phi_{\Delta t}$. Since $\tilde{\Phi}$ and Ψ coincide to higher order than Φ and Ψ, it is expected that solutions of (4.1) describe the behaviour of the numerical solution U with higher accuracy than solutions of the original KdV equation (2.1) being integrated.

Perturbation theory [10], [11] can be used to describe the solution of (4.1) when the initial condition is a soliton profile $\phi(x; c, d)$. It turns out that, to leading order in the perturbation parameter Δt^p, the modified solution is of the form

$$\sigma(x,t) + w(x,t) \tag{4.2}$$

where w is a function to be discussed later and

$$\sigma(x,t) = \phi(x - \mu(t); 4\kappa^2(t), d) = -2\kappa^2(t)\,\mathrm{sech}^2\,\kappa(t)[x - \mu(t) + d],$$

with κ and μ given by the differential equations

$$\frac{d\kappa}{dt} = \frac{M(\kappa)\Delta t^p}{8\kappa^2}, \tag{4.3}$$

$$\frac{d\mu}{dt} = 4\kappa^2 - \frac{N(\kappa)\Delta t^p}{8\kappa^2}, \tag{4.4}$$

with initial conditions $\kappa(0) = \sqrt{c}/2$, $\mu(0) = 0$. In turn, M, N in (4.3)–(4.4) are functions of κ given by

$$\begin{aligned} M(\kappa) &= \int_{-\infty}^{\infty} \ell_{p+1}(\phi(x; 4\kappa^2, 0))\, \phi(x; 4\kappa^2, 0)\, dx, \\ N(\kappa) &= \int_{-\infty}^{\infty} \ell_{p+1}(\phi(x; 4\kappa^2, 0)) \left[\kappa x \operatorname{sech}^2 \kappa x + \tanh\ \kappa x + \tanh^2 \kappa x\right] dx. \end{aligned}$$

Note that σ in (4.2) has, for each fixed value of t, the shape of the soliton of depth $-2\kappa^2(t)$. For the unperturbed problem with $\Delta t = 0$, (4.3) shows that $-2\kappa^2$ remains equal to its initial value $-c/2$. Furthermore, in the unperturbed case, $d\mu/dt = c$, $\mu(t) = ct$, so that σ in (4.2) reproduces the soliton solution $\phi(x - ct; c, d)$. When the perturbation is present, the perturbed soliton depth $-2\kappa^2$ evolves according to (4.3) and furthermore the phase velocity $d\mu/dt$ also undergoes changes.

For a scheme that satisfies (3.5), $M \equiv 0$ and the depth of σ does not vary with t. Furthermore, from (4.4) we see that the phase velocity of σ is of the form $c + m\Delta t^p$, with $m = -N(\sqrt{c}/2)/(2c)$, a constant. Hence, for schemes that conserve I_2 exactly, the σ contribution to (4.2) is a soliton of the correct depth travelling at a perturbed, constant velocity.

On the other hand when (3.5) does not hold, then, in general, $M \neq 0$, and solving the equations (4.3)–(4.4) to leading order in Δt, we find that the perturbed soliton depth is given by

$$-2\kappa^2(t) = -\left(\frac{c}{2} + \frac{M(\sqrt{c}/2)\Delta t^p}{\sqrt{c}} t\right) + O(\Delta t^{2p}),$$

with a linear variation with t at leading order. Substitution in (4.4) reveals that then the perturbed phase velocity differs from the unperturbed c in terms $t\Delta t^p$, which in turn implies that the phase $x - \mu(t)$ in (4.2) differs from the unperturbed $x - ct$ in terms $t^2\Delta t$. Thus, *in the 'general' case and to leading order in Δt, depth errors grow linearly with t and phase errors grow quadratically with t. This is to be compared with the case of conserved I_2 where there is no depth error and phase errors grow linearly.*

Note that in the perturbation formulae, the *direction* of ℓ_{p+1} is important; while in conventional error analysis only the size of the local error is taken into account.

The function w in (4.2) has not been discussed so far. This is also $O(\Delta t^p)$ and describes, on the one hand, the change in the soliton shape due to the perturbation and, on the other, a *tail* induced by the perturbation. Since the tail gets in general longer and longer as t increases (see the numerical experiments in the next section), w grows with t. However, perturbation theory shows that there is no tail formation when

$$\int_{-\infty}^{\infty} \ell_{p+1}(\phi(x; 4\kappa^2, 0)) \tanh^2 \kappa x\, dx = 0. \tag{4.5}$$

We now observe that in view of the relation $\tanh^2 z = 1 - \operatorname{sech}^2 z$, (4.5) holds provided that $M \equiv 0$ and

$$\int_{-\infty}^{\infty} \ell_{p+1}(\phi(x; 4\kappa^2, 0))\, dx = 0,$$

i.e. if the numerical method exactly preserves I_1 and I_2. In this case, w merely represents an $O(\Delta t^p)$ distortion in the soliton shape.

5 Numerical experiments

We first consider the standard backward Euler rule. For an equation $du/dt = A(u)$ this is given by

$$U^{n+1} - U^n = \Delta t A(U^{n+1}).$$

It is well known that the leading term of the truncation error for this rule is $(1/2)\Delta t^2 u_{tt}$. After replacing time derivatives by x-derivatives, we find that (3.3) holds with $p = 1$ and

$$\ell_2(u_0) = \partial_x \left[18u_0^2 \partial_x u_0 - 6u_0 \partial_x^3 u_0 - 9\partial_x u_0 \partial_x^2 u_0 + \frac{1}{2}\partial_x^5 u_0 \right].$$

To implement the method, we introduce a computational domain $-20 < x < 20$ and replace ∂_x by its standard pseudospectral approximation on a grid consisting of 128 equally spaced points. The errors introduced by this spatial discretization are negligible when compared with the time integration errors. For implementation details see [4].

The initial condition is chosen to be $\phi(x; 4, 10)$, i.e. we are dealing with a soliton of velocity $c = 4$ (depth -2) initially located at $x = -10$. The integration is carried out for $0 < t < t_{max} = 6$, so that the soliton travels a distance of 24 units.

For the dissipative backward Euler rule I_2 is not conserved. Actually $M < 0$, and perturbation theory predicts a decrease in the soliton depth as t increases. Furthermore (4.5) does not hold and a tail is expected.

Fig. 5.1 shows the solution at t_{max} when $\Delta t = 1/160$. The continuous line depicts the true position of the soliton ϕ and the crosses give the numerical solution U. It is clear that, in agreement with (4.2) the computed solution consists of a soliton profile plus a tail. The dotted line is the modified soliton σ computed by the the formulae (4.3)–(4.4): the agreement with the numerical solution is excellent. This shows both the ability of the modified equation to approximate the numerical solution and the success of the perturbation theory in describing the behaviour of the solutions of the modified equation (4.1).

Fig. 5.2 gives, for $\Delta t = 1/160, 1/320, 1/640$ the maximum norm of the error $u(\cdot, t_n) - U^n$ as a function of t. While the errors are not too large, say below 0.6, they grow quadratically, as predicted by the perturbation theory. When the phase of the soliton is completely wrong, so that the computed and true soliton do not overlap, the maximum norm of the error is 2. Thus as $t \to \infty$ the errors saturate.

We consider next the nondissipative implicit midpoint rule ($p = 2$)

$$U^{n+1} - U^n = \Delta t A\left(\frac{1}{2}[U^n + U^{n+1}]\right),$$

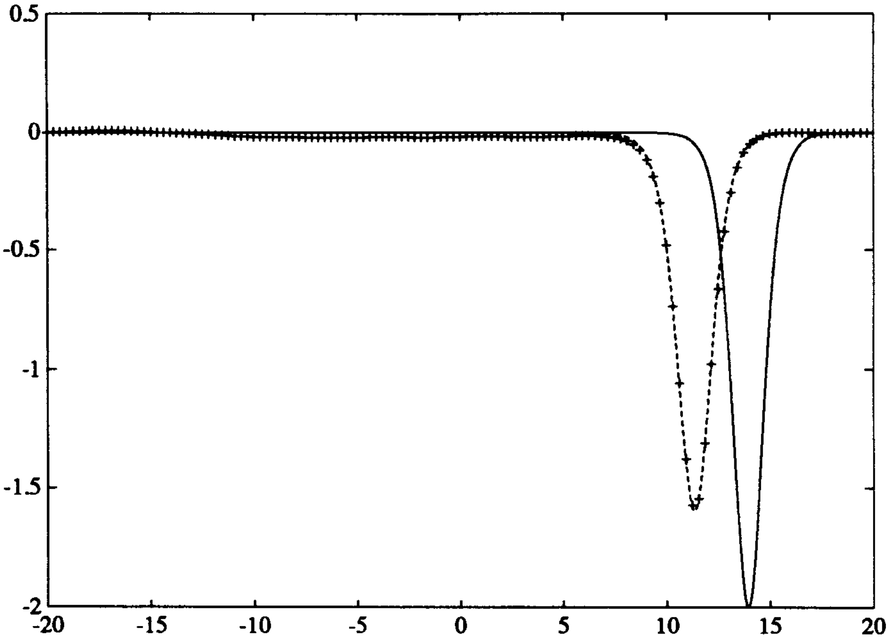

Figure 5.1: True soliton (solid line), modified soliton σ (dotted line) and numerical solution (crosses) for the backward Euler rule, $t = 6$, $\Delta t = 1/160$

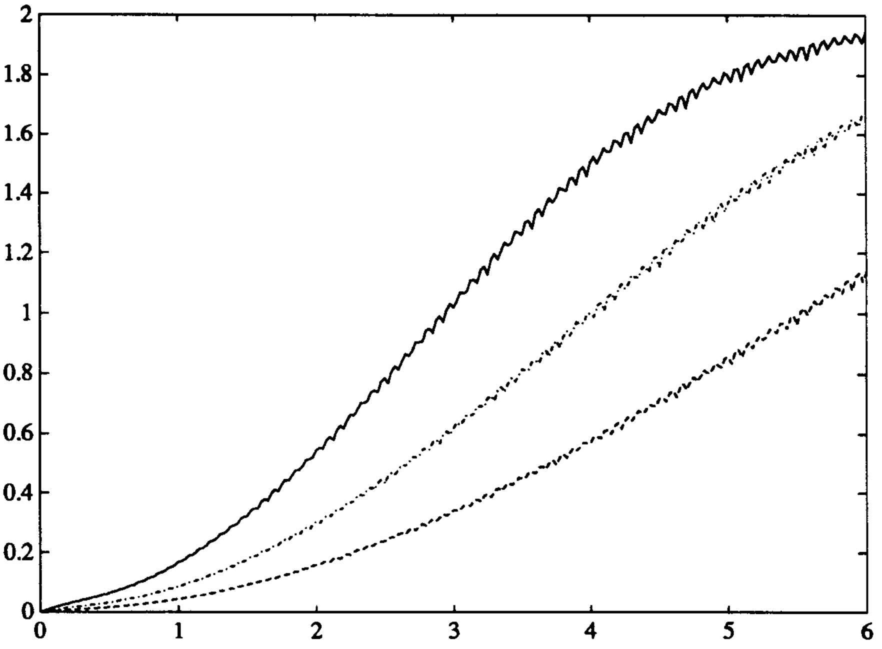

Figure 5.2: Maximum norm of the error as a function of time for the backward Euler rule, $\Delta t = 1/160, 1/320, 1/640$

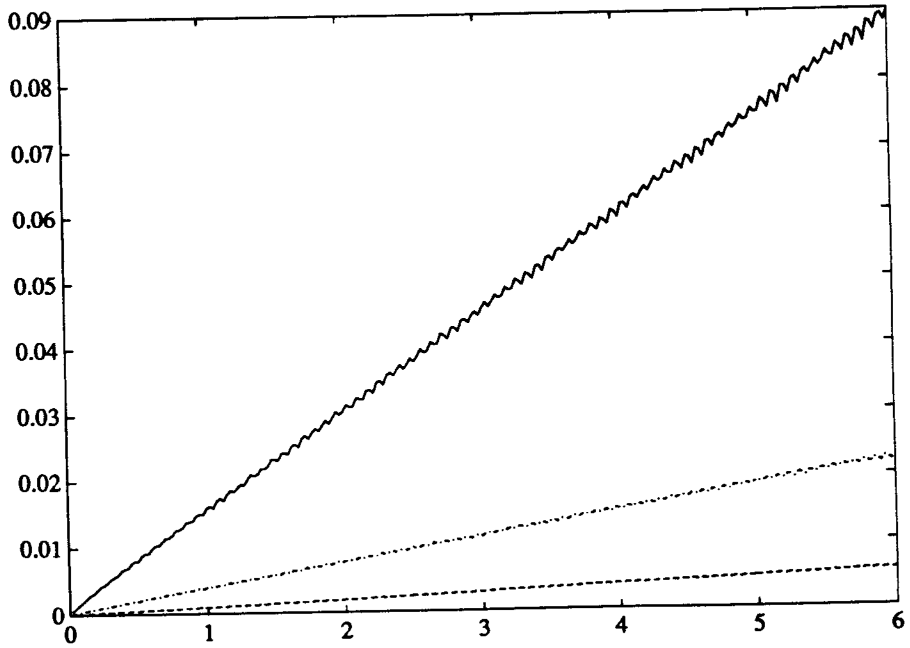

Figure 5.3: Maximum norm of the error as a function of time for the midpoint rule, $\Delta t = 1/40, 1/80, 1/160$

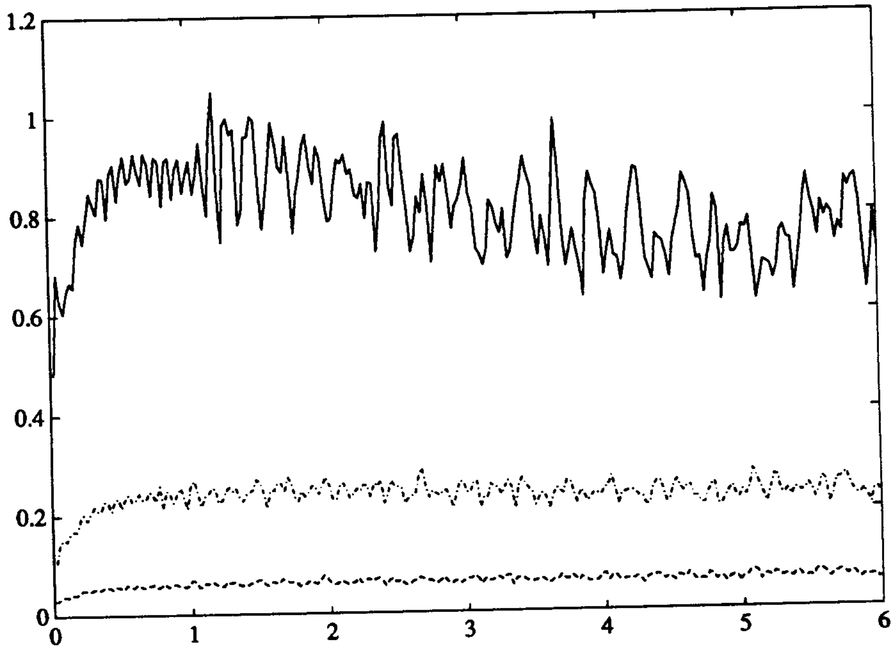

Figure 5.4: Maximum norm $\times 10^3$ of the error with respect to the modified soliton as a function of time for the midpoint rule, $\Delta t = 1/40, 1/80, 1/160$

Table 5.1: Midpoint rule errors with respect to the true solution u and modified soliton σ

	$\|U^n - u(\cdot, t_n)\|_\infty$		$\|U^n - \sigma(\cdot, t_n)\|_\infty$	
Δt	$t_n = 3$	$t_n = 6$	$t_n = 3$	$t_n = 6$
1/40	4.46E-2	8.98E-2	8.60E-4	7.22E-4
1/80	1.12E-2	2.27E-2	2.39E-4	2.22E-4
1/160	2.79E-3	5.69E-3	6.67E-5	5.68E-5

which conserves linear and quadratic invariants [12], [13] and in particular I_1 and I_2. Perturbation theory predicts no tail formation, correct depth and linear growth of phase errors. Experiments (not all of them are shown here) confirm these expectations. Fig. 5.3 gives the maximum norm of the error as a function of t when $\Delta t = 1/40, 1/80, 1/160$: errors clearly grow linearly with t. For the same runs, we have given in Fig. 5.4, the maximum norm of the difference between the computed U^n and the perturbed soliton σ in (4.2). Table 5.1 provides at $t = 3$ and $t = 6$ the maximum norm of the (true) error $U^n - \phi(\cdot - ct_n; c, d)$ with respect the true solution and of the error $U^n - \sigma(\cdot, t_n)$ with respect the perturbed soliton. In the table we see that the errors with respect to σ are much smaller than the true errors with respect to ϕ. In other words, the bulk of the true error consists of the phase error that is removed by comparing with the relocated $\sigma = \phi(x - \mu(t_n); c, d)$ rather than with the KdV solution $\phi(x - ct; c, d)$. However the error with respect to σ still shows an $O(\Delta t^2)$ behaviour: this is due to the w contribution to (4.2). On the other hand note, both in the table and in Fig. 5.4, that errors with respect to σ do not grow with t: with no tail formation w only accounts for a distortion in soliton shape and does not experiment secular growth. To sum up, in this case, experiments show that the true error consists of a phase error, with a $t\Delta t^2$ behaviour, plus further $O(\Delta t^2)$ errors that remain bounded as t increases.

For another application of soliton perturbation theory to numerical methods see [9].

Of course, the aim of the experiments above was not to show that the dissipative, first-order backward Euler rule is not suitable for the problem at hand. This fact is universally appreciated. Our goal was to point out that methods that behave differently with respect to conservation laws also possess different error propagation mechanisms.

6 Analytic results

6.1 Preliminaries

The results mentioned so far are not rigorous. To begin with, asymptotic expansions have been used, without paying attention to the norms in which they may be valid. The combinations of values of t and Δt for which the perturbation formulae are valid have not been spelled out. Furthermore, even though the technique of modified equations may be used *in some cases* in a rigourous way [7], this would require, if at all feasible, a detailed mathematical analysis of the (method dependent) modified equation (4.1). In this section we present a different, but related approach. The key idea is to note that for most practical methods the global error $U^n - u(\cdot, t_n)$ possesses an asymptotic expansion, whose leading term is a solution of a variational equation, see e.g. [8]. We therefore start by studying the variational equation of the KdV equation.

We use the standard Sobolev spaces H^k, k an integer, with norm $\|\cdot\|_k$.

6.2 The homogeneous variational equation around a soliton solution

In what follows, we fix a velocity $c_0 > 0$ and consider the moving coordinates $X = x - c_0 t$, $T = t$, in which the KdV equation becomes

$$u_T - c_0 u_X - 6uu_X + u_{XXX} = 0. \tag{6.1}$$

In the new variables the soliton (2.4) with velocity c_0 and $d = 0$ becomes a T-independent (equilibrium) solution $\psi(X) := \phi(X; c_0, 0)$ of (6.1). Similarly, the soliton of velocity $c \neq c_0$ of (2.1) becomes a soliton of velocity $c - c_0$ of (6.1).

We study perturbations of the equilibrium ψ. If $e(X,0)$ is a small perturbation, then, to the first order of small quantities, the solution of (6.1) whith initial condition $\psi(X) + e(X,0)$ is $\psi(X) + e(X,T)$, where $e(X,T)$ satisfies the (linear) variational equation

$$e_T - c_0 e_X - 6\psi(X)e_X - 6\psi'(X)e + e_{XXX} = 0, \tag{6.2}$$

or

$$e_T = \frac{\partial}{\partial X} L_2 e,$$

with the second-order operator L_2 given by

$$L_2 e = (c_0 + 6\psi)e - e_{XX}. \tag{6.3}$$

Alternatively, the same variational equation may be obtained by first linearizing in (2.1) and then changing variables $(x,t) \to (X,T)$.

Let us find some particular solutions of (6.2). We start from the easily proved relations

$$L_2 \psi' = 0 \tag{6.4}$$

and

$$L_2 \chi = -\psi, \tag{6.5}$$

where $\chi(X) = \phi_c(X; c_0, 0)$.

From (6.4) we see that $\psi'(X)$ is an equilibrium (i.e. T-independent) solution of (6.2). This is readily interpreted: $\psi(X) + \epsilon\psi'(X)$ coincides except for $o(\epsilon)$ terms with $\psi(X+\epsilon)$, i.e. with the soliton of velocity 0 located at $-\epsilon$, a solution of (6.1). Hence $\epsilon\psi'(X)$ is a solution of the variational equation.

On the other hand, by using (6.5)-(6.4), it is a trivial task to prove that,

$$\chi(X) - T\psi'(X) \tag{6.6}$$

also provides a solution of (6.2). The interpretation of this fact is as follows: Consider the initial condition $\psi(X) + \epsilon\chi(X)$ for (6.1). This is, except for order $o(\epsilon)$ terms, the initial profile of the soliton of velocity ϵ located at the origin. At time $T > 0$, this has travelled a distance ϵT so that the solution of (6.1) with initial condition $\psi(X) + \epsilon\chi(X)$ is $\psi(X - \epsilon T) + \epsilon\chi(X - \epsilon T) + o(\epsilon)$. Hence

$$\epsilon(\chi(X) - T\psi'(X)) = [\psi(X - \epsilon T) + \epsilon\chi(X - \epsilon T)] - \psi(X) + o(\epsilon)$$

has to satisfy the variational equation (6.2).

6.3 Stability of the homogeneous variational equation

The expression (6.6) shows that the variational equation (6.2) possesses solutions that *grow unboundedly* with T. This corresponds to the fact that ψ is not a Lyapunov-stable equilibrium of (6.1): if the initial profile is changed from ψ into that of a soliton with a small velocity the difference between ψ and the new soliton solution grows unboundedly. However, Benjamin [2] proved that ψ is stable in the sense that if $u(X,T)$ is a solution of (6.1) with $u(X,0) - \psi(X)$ small, then, for $T > 0$, a suitable translation $u(X - \mu(T), T)$ of $u(X,T)$ remains close to $\psi(X)$ (uniformly in T). Thus the Lyapunov instability only manifests itself as a phase error. The techniques used by Benjamin can be applied to prove the following result [5].

Theorem 6.1 *There exists a constant $C > 0$ such that if $e(X,T)$ is a solution of (6.2) then, for all $T > 0$,*

$$\|e(\cdot,T) - \mu(T)\psi'(\cdot)\|_1 \leq C\,\|e(\cdot,0)\|_1 \qquad (6.7)$$

with $\mu(T) = (e(\cdot,T),\psi')/(\psi',\psi')$.

The theorem shows that, even though the solutions $e(X,T)$ may grow unboundedly with T, they only do so in the direction of ψ', cf. (6.6). This of course corresponds to secular growth of *phase errors* in the KdV equation (6.1).

6.4 The nonhomogeneus variational equation

We now look at the nonhomogeneus initial-value problem

$$e_T = \frac{\partial}{\partial X}L_2 e + s, \quad T > 0, \qquad (6.8)$$
$$e(T = 0) = 0, \qquad (6.9)$$

where the source s is assumed to be independent of T. The following result holds.

Theorem 6.2 *Assume that the (T-independent) source term s is in L^2, is the (distributional) derivative of an L^2 function S and satisfies*

$$(s,\psi) = 0. \qquad (6.10)$$

Then, for a constant C independent of s and T,

$$\left\| e(\cdot,T) - \frac{(e(\cdot,T),\psi')}{(\psi',\psi')}\psi' \right\|_1 \leq C\,\|s\|_{-1}, \quad T > 0. \qquad (6.11)$$

Proof. We first look for a T-independent solution e_1 of (6.8). This satisfies

$$0 = e_{1,T} = \frac{\partial}{\partial X} L_2 e_1 + s = \frac{\partial}{\partial X} L_2 e_1 + \frac{\partial}{\partial X} S,$$

or

$$L_2 e_1 = -S. \tag{6.12}$$

The kernel of the operator L_2 mapping H^2 in L^2 is spanned by ψ'. Hence (Fredholm's alternative) (6.12) has a solution e_1 provided that $(-S, \psi') = 0$, a condition that follows from (6.10). Furthermore e_1 is uniquely defined under the additional condition $(e_1, \psi') = 0$ and with e_1 defined in this way

$$\|e_1\|_2 \leq C \|S\|_0 = C \|s\|_{-1}, \tag{6.13}$$

with C a constant associated with L_2

We now set $e_2 = e - e_1$, where e_2 has to solve

$$e_{2,T} = \frac{\partial}{\partial X} L_2 e_2,$$
$$e_2(T = 0) = -e_1.$$

By (6.7) and (6.13)

$$\left\| e_2(\cdot, T) - \frac{(e_2(\cdot, T), \psi')}{(\psi', \psi')} \psi' \right\|_1 \leq \|e_1\|_1 \leq C \|s\|_{-1}.$$

From this bound and after noticing that $(e_2(\cdot, T), \psi') = (e(\cdot, T), \psi')$, we conclude

$$\left\| e(\cdot, T) - \frac{(e(\cdot, T), \psi')}{\psi', \psi')} \psi' \right\|_1 \leq \|e_1\|_1 + C \|s\|_{-1}$$

and a new application of (6.13) concludes the proof. □

We see that, once more, the theorem ensures that the component of e orthogonal to ψ' remains uniformly bounded for all $T > 0$. However the source term has to satisfy some qualifications: to possess an antiderivative in L^2 and to satisfy (6.10). The solution

$$T\chi - \frac{T^2}{2}\psi'$$

corresponding to the source $s = \chi$ reveals that when $(s, \psi) \neq 0$, the growth with T is not confined to the direction of ψ'. (This particular solution clearly corresponds to quadratic growth of phase errors along with linearly growing amplitude errors, as we found in our discussion in Section 4.)

Before closing this subsection it is important to observe that if s is in $L^2 \cap L^1$ then it has an antiderivative in L^2,

$$S(x) = \int_{-\infty}^{x} s(\bar{x})\, d\bar{x},$$

if and only if $I_1(s) = 0$ (so that the integral decays as $x \to \infty$). Hence for reasonably smooth sources s, the hypotheses of the theorem read $I_1(s) = 0$, $(s, \psi) = 0$; these are essentially the geometric constraints we found in Section 4. These constraints again identify directions in phase-space that are relatively harmless.

6.5 Conclusion

We now revert to the original (x, t) variables. The nonhomogeneous variational equation (6.8) becomes

$$e_t - 6\psi(x - c_0 t)e_x - 6\psi'(x - c_0 t)e + e_{xxx} = s(x - c_0 t), \tag{6.14}$$

with s a real function of a real variable ξ. The bound (6.11) becomes

$$\left\| e(\cdot, t) - \frac{(e(\cdot, t), \psi')}{(\psi', \psi')}\psi' \right\|_1 \leq C \|s\|_{-1}, \quad t > 0.$$

where now ψ' is to be interpreted as the function $\psi'(\cdot - c_0 t)$.

Assume that for a numerical method $\Psi_{\Delta t}$ satisfying (3.3) it holds that the global error when integrating $\phi(x - c_0 t; c_0, 0)$ has an asymptotic expansion

$$U^n = u(\cdot, t_n) + \Delta t^p e(\cdot, t_n) + R(\cdot, t),$$

where R is a residual with $\|R(\cdot, t)\|_1 = o(\Delta t^p)$ uniformly in bounded t intervals and e is the solution of the variational equation (6.14) with source term

$$s(\xi) = \ell_{p+1}(\psi(\xi)).$$

We can conclude from Theorem 2 that *the leading error term $\Delta t^p e$ consists of a phase error in the direction of ψ' plus errors that are uniformly bounded for all t, provided* that the local error satisfies $(\ell_{p+1}(\psi), \psi) = 0$ and $I_1(\ell_{p+1}(\psi)) = 0$. These conditions on the local error are once more satisfied for methods that exactly preserve I_1 and I_2.

Of course to apply rigorously the last result it is necessary to first check that the numerical method converges in H^1 and has the indicated asymptotic expansion for the global error. While this has not been done for the methods considered in the preceding section, we strongly feel that it is feasible by using routine techniques. The verification of the existence of such an asymptotic expansion would however take us too far away from the main purpose of this article: to show additional evidence for the fact that numerically preserving conservation laws not only implies better qualitative behaviour, but may also lead to better error bounds.

References

[1] A. Arakawa, *Computational design for long-term numerical integration of the equations of fluid motion: two-dimensional incompressible flow. Part I*, J. Comput. Phys., 1 (1966), 119–143.

[2] T. B. Benjamin, *The stability of solitary waves*, Proc. R. Soc. Lond. A., 328 (1972), 153–183.

[3] M. P. Calvo and J. M. Sanz-Serna, *The development of variable-step symplectic integrators, with application to the two-body problem*, SIAM J. Sci. Comput., (1993), to appear.

[4] J. de Frutos and J. M. Sanz-Serna, *An easily implementable fourth-order method for the time integration of wave problems,* J. Comput. Phys., 103 (1992), 160–168.

[5] J. de Frutos and J. M. Sanz-Serna, *Error growth and invariant quantities in numerical methods: a case study,* Applied Mathematics and Computation Reports, Report 1993/4, Universidad de Valladolid.

[6] C. W. Gear, *Invariants and numerical methods for ODEs,* Physica D, 60 (1990), 303–310.

[7] D. F. Griffiths and J. M. Sanz-Serna, *On the scope of the method of modified equations,* SIAM J. Sci. Comput., 7 (1986), 994–1008.

[8] E. Hairer, S. P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I, Nonstiff problems,* Springer, Berlin, 1987.

[9] R. L. Herman and C. J. Knickerbocker, *Numerically induced phase shift in the KdV soliton,* J. Comput. Phys., 104 (1993), 50–55.

[10] V. I. Karpman and E. M. Maslov, *Perturbation theory for solitons,* Sov. Phys. JETP, 46 (1977), 281–291.

[11] V. I. Karpman and E. M. Maslov, *Structure of tails produced under the action of perturbations of solitons,* 48 (1978), 252–259.

[12] J. M. Sanz-Serna, *Runge-Kutta schemes for Hamiltonian systems,* BIT, 28 (1988), 877-883.

[13] J. M. Sanz-Serna and M. P. Calvo, *Numerical Hamiltonian Problems,* Chapman and Hall, London, 1993.

[14] R. F. Warming and B. J. Hyett, *The modified equation approach to the stability and accuracy analysis of finite difference methods,* J. Comput. Phys., 14 (1974), 159–179.

Acknowledgement

This research has been supported by project DGICYT PB89-0351.

J. de Frutos and J.M. Sanz-Serna
Departamento de Matemática Aplicada y Computación
Facultad de Ciencias
Universidad de Valladolid
Valladolid, Spain