JOURNAL OF COMPUTATIONAL PHYSICS **126,** 299–314 (1996)
ARTICLE NO. 0139

On the Numerical Solution of the Sine–Gordon Equation

I. Integrable Discretizations and Homoclinic Manifolds

M. J. ABLOWITZ,* B. M. HERBST,† AND CONSTANCE SCHOBER*

*Program in Applied Mathematics, University of Colorado, Boulder, Colorado and †Department of Applied Mathematics, The University of the Orange Free State, Bloemfontein 9300, South Africa

Received February 10, 1995; revised November 29, 1995

In this, the first of two papers on the numerical solution of the sine–Gordon equation, we investigate the numerical behavior of a double discrete, completely integrable discretization of the sine–Gordon equation. For certain initial values, in the vicinity of homoclinic manifolds, this discretization admits an instability in the form of grid scale oscillations. We clarify the nature of the instability through an analytical investigation supported by numerical experiments. In particular, a perturbation analysis of the associated linear spectral problem shows that the initial values used for the numerical experiments lie exponentially close to a homoclinic manifold. This paves the way for the second paper where we use the nonlinear spectrum as a basis for comparing different numerical schemes.

1. INTRODUCTION

Although Stokes published his famous work, *On the Theory of Oscillatory Waves,* in 1847 [19], it was only during the 1960s that Benjamin [7] realized that the periodic Stokes wave on deep water is in fact, unstable (see also Whitham [20]). About the same time Zabusky and Kruskal studied the Korteweg de Vries (KdV) equation [21], a pioneering investigation that paved the way for the development of soliton theory. In a suitable asymptotic limit, the KdV equation also describes weakly nonlinear water waves but in *shallow* water; Benjamin's instabilities are therefore not observed in the KdV equation. The asymptotic equation governing the one-dimensional slow modulation of periodic wavetrains (such as the Stokes wave) in *deep* water is the nonlinear Schrödinger (NLS) equation. It is the NLS equation where one indeed encounters the instabilities similar to those discovered by Benjamin.

Benjamin observed the instability through a linearized analysis of the Euler equations; similarly, a linearized analysis of the NLS equation shows that a number of low frequency modes (side bands) may become unstable. The exact number of unstable modes is determined by the amplitude of the carrier wave. Since the NLS equation is completely integrable through the inverse scattering transform (see, for example, [1, 11]) this allows us to study the subsequent *nonlinear* evolution of the instabilities. There are two interesting boundary value problems which can be considered, the ''infinite line'' case with decaying initial values, and the periodic boundary value problem. In the infinite line case the instability results in solitons—the number depending on the amplitude of the initial values. On the other hand, with periodic boundary values, the instabilities are a manifestation of a ''homoclinic'' manifold associated with the NLS equation. The dimension or complexity of the homoclinic manifold depends on the number of unstable modes which can be controlled by the amplitude of the carrier wave. In this paper we consider the periodic problem.

From a numerical point of view these instabilities prove to be troublesome. In earlier work [13, 2] we showed that initial data, which are nearby *low* dimensional homoclinic manifolds, may develop numerically induced spatial and temporal chaos at intermediate levels of discretization. The chaos is not observed with integrable discretizations of the NLS equation and disappears (perhaps very slowly) from the nonintegrable schemes as the discretization is refined and the scheme converges. A more troubling aspect of the phenomenon becomes apparent if the dimension of the homoclinic manifold is increased. In this case *temporal* instabilities and chaos are excited by very small perturbations, even on the order of roundoff error [18]. These instabilities persist as the mesh is refined and cannot be detected by monitoring the conserved quantities; i.e., the temporal evolution of the numerical solution remains unpredictable even if all the conserved quantities are well preserved (by employing a very fine grid). The chaos persists even if a completely integrable (in the spatial variable) scheme is employed. The reason being that for a large number of unstable modes the initial values turn out to be extremely close, e.g., exponentially close, to homoclinic manifolds. Consequently, based on these observations, we

0021-9991/96 $18.00

conclude that numerical, analytical, and laboratory studies of the Stokes wave on deep water require further attention. We also believe that both homoclinic structures (by measuring flows on both sides of the "separatrix") and the behavior we have already observed numerically [13, 2, 18] should be observable in the laboratory.

The difficulties encountered in solving the NLS equation numerically are not unique to this equation. We believe that (i) the instabilities in the modified KdV equation observed by Fornberg and Whitham [12] are of the same nature and (ii) the homoclinic structures associated with the sine–Gordon equation [10] have a marked effect on numerical studies of the sine–Gordon equation. A previous study [14] noted some of the difficulties involved in solving the sine–Gordon equation numerically if initial values are chosen in the proximity of homoclinic manifolds.

The subtleties involved in solving the sine–Gordon equation numerically, are strikingly illustrated by the use of a doubly discrete (space and time) completely integrable discretization of the sine–Gordon equation. One might expect it to be a very efficient numerical scheme—it is local and explicit and it has exact N soliton, quasi-periodic N phase, and homoclinic solutions. Our experience with the integrable discretization of the NLS equation (see [1, 13, 18]) also indicates the usefulness of such integrable discretizations. In view of this, it might be surprising that the doubly discrete integrable discretization of the sine–Gordon equation admits an instability. More specifically, we choose initial values that at first glance appear to be "safe," i.e., close but not too close to a homoclinic manifold. We observe that high frequency oscillations develop on the grid scale. This unexpected behavior has all the characteristics of sensitive dependence on the initial values. A careful perturbation analysis shows that the initial values are exponentially close to a homoclinic submanifold where sensitive dependence on the initial values is expected. We also show that the instability associated with the integrable discretization depends on whether an even or an odd number of grid points are used. Our understanding of the behavior of the double discrete integrable scheme is based upon analysis and is supported by extensive numerical experiments.

2. ANALYTICAL PROPERTIES OF THE SINE–GORDON EQUATION

We study the sine–Gordon equation,

$$u_{tt} - u_{xx} + \sin u = 0, \tag{1}$$

together with periodic boundary conditions, $u(x, t) = u(x + L, t)$.

The sine–Gordon equation is a completely integrable, infinite dimensional Hamiltonian system; see, e.g., [1, 3],

$$q_t = \frac{\delta H}{\delta p}, \quad p_t = \frac{-\delta H}{\delta q}. \tag{2}$$

with Hamiltonian,

$$H(p, q) = \int_0^L [\tfrac{1}{2}p^2 + \tfrac{1}{2}(q_x)^2 + 1 - \cos q]\, dx, \tag{3}$$

where $q := u$ and $p := u_t$ are the conjugate variables. In this context, integrable means that there is a sufficiently large number (infinity) of conserved quantities so that a transformation to action-angle variables can be achieved. However, there are special solutions where the mapping to action-angle variables becomes singular—the so-called homoclinic structures. As mentioned in the Introduction, they are crucial in numerical studies and can be described in elementary terms.

2.1. *A Linearized Analysis*

First we show that there are exponential instabilities associated with some solutions of the sine–Gordon equation. In Section 5 these instabilities are related to double points of the nonlinear spectrum; i.e., the instabilities are associated with homoclinic orbits.

Assuming that

$$u(x, t) = \pi + \varepsilon(x, t), \quad |\varepsilon(x, t)| \ll 1,$$

with

$$\varepsilon(x,t) = \hat{\varepsilon}_n(t) \exp(i\mu_n x) + \hat{\varepsilon}_n^*(t) \exp(-i\mu_n x), \quad \mu_n = 2\pi n/L,$$

n an arbitrary integer, it follows that

$$\frac{d^2}{dt^2}\hat{\varepsilon}_n + \omega_n^2 \hat{\varepsilon}_n = 0 \tag{4}$$

(and similarly for $\hat{\varepsilon}_n^*(t)$), where $\omega_n^2 = \mu_n^2 - 1$. It is clear that the nth mode grows exponentially, if

$$0 \le \mu_n^2 < 1. \tag{5}$$

It is worth pointing out that the zeroth mode, $n = 0$, is the most unstable in the sense that it grows at the fastest rate (it is therefore the dominant mode).

It is convenient to rewrite (4) as a system

$$\frac{d}{dt}\begin{pmatrix} \hat{\varepsilon}_n \\ \hat{\eta}_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 - \mu_n^2 & 0 \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_n \\ \hat{\eta}_n \end{pmatrix}.$$

The eigenvectors, $(1, \lambda)^{\mathrm{T}}$, with $\lambda = \pm\sqrt{1 - \mu_n^2}$ of this system translate into an initial condition given by

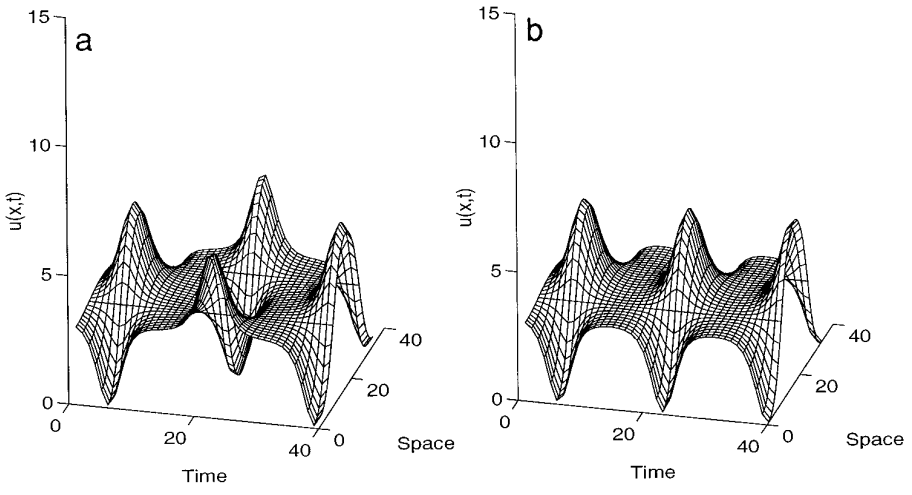


FIG. 1. (a) Outside and (b) inside the homoclinic orbit.

$$u(x,0) = \pi + \varepsilon_0 \cos(\mu_n x)$$
$$u_t(x,0) = \varepsilon_0 \sqrt{1-\mu_n^2}\cos(\mu_n x). \tag{6}$$

The simplest case is obtained from the choice $n = 0$, i.e., no x-dependence. Let us consider the initial condition,

$$u(x,0) = \pi + \varepsilon_0$$
$$u_t(x,0) = \varepsilon_0(1+p),$$

where p is a small number. In fact this is close to the homoclinic orbit of the pendulum equation (an ODE); $p < 0$ leads to an "oscillating," and $p > 0$ to a "rotating," solution (i.e., the pendulum goes "over the top"). The choice $p = 0$ corresponds to the homoclinic orbit $u(t) = \pi + 4 \arctan[\exp(t+\gamma)]$, with $p < 0, p > 0$ on either side of the homoclinic orbit.

A more general situation is illustrated numerically by choosing $L = 2\sqrt{2\pi}$, $\mu = 2\pi/L$, and initial conditions

$$u(x,0) = \pi + 0.1\cos(\mu x)$$
$$u_t(x,0) = (0.1+p)\sqrt{1-\mu^2}\cos(\mu x),$$

where p is again a small number. Figures 1a, b show the solutions with (a) $p = +0.01$ and (b) $p = -0.01$. Different qualitative behavior is readily observed—the temporal period shown in Fig. 1a is about twice that of Fig. 1b—this suggests, which we confirm below, that the orbit initiated by (6) (with $p = 0$) forms a separatrix, i.e., another orbit homoclinic to π, this time with a spatial structure (cf. Fig. 2).

In the next section we find explicit expressions for these homoclinic orbits.

2.2. *An Analytic Expression for the Homoclinic Orbit*

A particularly straightforward way of obtaining exact analytical expressions for the homoclinic orbits suggested above is through Hirota's method [16] (see also [2]). Substituting

$$u = 4\tan^{-1}[g(x,t)/f(x,t)]$$

into the sine–Gordon equation (1), one arrives at the bilinear equations

$$(D_x^2 - D_t^2)g\cdot f = gf$$
$$(D_x^2 - D_t^2)(g\cdot g - f\cdot f) = 0, \tag{7}$$

where

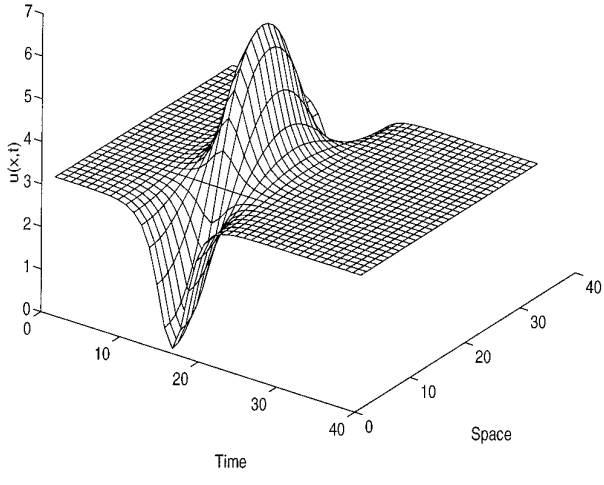


FIG. 2. A homoclinic orbit.

$$D_x^2 g\cdot f := g_{xx}f - 2g_x f_x + g f_{xx}.$$

For example, choosing

$$g(x,t) = b_1\cos(p_1 t + \gamma_1)$$
$$f(x,t) = b_2\cosh(p_2 x + \gamma_2),$$

one obtains the breather solution

$$u(x,t)\,4\tan^{-1}\left[\frac{b_1}{b_2}\cos(p_1 t+\gamma_1)\,\mathrm{sech}(p_2 x+\gamma_2)\right],$$

where

$$p_1^2 + p_2^2 = 1, \quad b_1^2 p_1^2 = b_2^2 p_2^2. \tag{8}$$

The symmetry,

$$x\to t, \quad t\to x, \quad u\to \pi + u,$$

leads directly to the following family of homoclinic orbits (see also [10]):

$$u(x,t) = \pi + 4\tan^{-1}\left[\frac{b_1}{b_2}\cos(p_1 x+\gamma_1)\,\mathrm{sech}(p_2 t+\gamma_2)\right]. \tag{9}$$

One can think of each $p_1 = 2\pi n/L$, n an integer such that $p_1^2 < 1$ (in order to satisfy (8)) as a separate homoclinic orbit. Figure 2 shows the situation for $n = 1$.

Several remarks are in order:

- the case $p_1 = 0$ reduces to the well-known homoclinic orbit of the pendulum equation;
- the expression (9) describe the simplest family of homoclinic orbits to the fixed point $u = \pi$. Much more complicated structures homoclinic to periodic orbits etc. exist (see [10]) and can also be generated from (7) by writing a suitable ansatz (see also (22) below);
- there is a one-to-one correspondence between the number of unstable modes given by (5) and the number of homoclinic orbits given (implicitly) by (8).

3. A DOUBLY DISCRETE, COMPLETELY INTEGRABLE DISCRETIZATION

In this section we consider a doubly discrete, completely integrable discretization due to Hirota [16] and a related one due to Bobenko *et al.* [8, 9] (used in their important and elegant study of discrete geometrical surfaces).

3.1. *An Integrable Discretization*

The basic idea is to discretize the bilinear form, (7). Hirota [16] used

$$\begin{aligned}(\cosh hD_x - \cosh hD_t)g\cdot f &= \tfrac{1}{4}h^2(\cosh hD_x + \cosh hD_t)g\cdot f\\ &\quad (\cosh hD_x - \cosh hD_t)(g\cdot g - f\cdot f) = 0,\end{aligned} \tag{10}$$

where

$$\exp(\varepsilon D_x + \delta D_t)a(x,t)\cdot b(x,t) := a(x+\varepsilon, t+\delta)b(x-\varepsilon, t-\delta).$$

Note that (10) is a discrete analogue of the continuous bilinear form (7). A completely integrable discretization is obtained by substituting

$$f(x,t) = \exp(\tfrac{1}{4}\rho(x,t))\cos(\tfrac{1}{4}\phi(x,t))$$
$$g(x,t) = \exp(\tfrac{1}{4}\rho(x,t))\sin(\tfrac{1}{4}\phi(x,t))$$

into (10) and eliminating ρ from the results leads to

$$\begin{aligned}(1-\tfrac{1}{4}h^2)\tan\tfrac{1}{4}(\phi(x+h,t)+\phi(x-h,t))\\ = (1+\tfrac{1}{4}h^2)\tan\tfrac{1}{4}(\phi(x,t+h)+\phi(x,t-h)).\end{aligned} \tag{11}$$

This scheme is special; it is a completely integrable discretization of the sine–Gordon equation, where $\phi(x,t)$ denotes the approximation of $u(x,t)$.

It is sometimes more convenient to use alternative forms of Eq. (11). Making use of the identity, $\tan^{-1} x = (1/2i)\ln[(1+ix)/(1-ix)]$, it follows after some manipulation that (11) may be written as

$$\tfrac{1}{2}(\phi_m^{n+1} + \phi_m^{n-1}) = \tfrac{1}{2}(\phi_{m-1}^n + \phi_{m+1}^n) + i\ln\left[\frac{1+\frac{1}{4}h^2\exp\frac{1}{2}i(\phi_{m+1}^n+\phi_{m-1}^n)}{1+\frac{1}{4}h^2\exp\frac{1}{2}(-i)(\phi_{m+1}^n+\phi_{m-1}^n)}\right], \tag{12}$$

where $\phi_m^n := \phi(mh, nh)$. The identity $\arg(z) := \theta = (1/2i)\ln(z/z^*)$, where $z = r\exp(i\theta)$, leads to

$$\tfrac{1}{2}(\phi_m^{n+1} + \phi_m^{n-1}) = \tfrac{1}{2}(\phi_{m-1}^n + \phi_{m+1}^n) + 2i\arg[1+\tfrac{1}{4}h^2\exp\tfrac{1}{2}i(\phi_{m+1}^n+\phi_{m-1}^n)]. \tag{13}$$

Alternatively, substituting

$$Q_m^n = \exp(\tfrac{1}{2} i \phi_m^n) \tag{14}$$

into (12), we obtain

$$Q_m^{n+1} Q_m^{n-1} = \frac{Q_{m+1}^n Q_{m-1}^n + \frac{1}{4}h^2}{1 + \frac{1}{4}h^2 Q_{m+1}^n Q_{m-1}^n}. \tag{15}$$

All these different forms represent second-order accurate discretizations on the sine–Gordon equation (1), when used as a numerical scheme.

It is possible to rewrite Hirota's scheme in the form used by Bobenko *et al.* [9] by the following change of variables:

$$\begin{aligned} \phi_{m+1,n}^{(1)} &= -\tfrac{1}{2}(\phi_{m+1}^n + \phi_{m-1}^n) \\ \phi_{n,n-1}^{(2)} &= \tfrac{1}{2}(\phi_m^{n+1} + \phi_m^{n-1}). \end{aligned} \tag{16}$$

In terms of the new variables Hirota's equation (13) becomes

$$\phi_{m,n-1}^{(2)} = -\phi_{m+1,n}^{(1)} + 2i \arg[1 + \tfrac{1}{4}h^2 \exp(-i\phi_{m+1,n}^{(1)})]. \tag{17}$$

If we now use this equation to eliminate $\phi_{m,n}^{(2)}$ from the identity

$$\phi_{m+1,n+1}^{(1)} + \phi_{m+1,n-1}^{(2)} + \phi_{m+1,n-1}^{(1)} + \phi_{m-1,n-1}^{(2)} = 0,$$

and writing ϕ_m^n instead of $\phi_{m,n}^{(1)}$, we arrive at

$$\begin{aligned} \phi_m^{n+1} + \phi_m^{n-1} - \phi_{m-1}^n - \phi_{m+1}^n &= 2i \arg \\ &[1 + \tfrac{1}{4}h^2 \exp(i\phi_{m+1}^n)] \\ &+ 2i \arg[1 + \tfrac{1}{4}h^2 \exp(i\phi_{m-1}^n)]. \end{aligned} \tag{18}$$

Because of the intimate relationship between the schemes (11) and (18), they share the same properties. Thus, from now on we concentrate on Hirota's form of the integrable discretization, (11) or (13).

3.2. *Homoclinic Orbits of the Integrable Discretization*

We proceed to construct several special solutions for Hirota's scheme (11). Following a similar procedure as in the analytical case, first assume that in (10),

$$\begin{aligned} g(x,t) &= b_1 \cos(p_1 t + \gamma_1) \\ f(x,t) &= b_2 \cosh(p_2 x + \gamma_2) \end{aligned} \tag{19}$$

and we obtain the breather solution of the discrete system,

$$\phi(x,t) = 4 \tan^{-1}\left[\frac{b_1}{b_2} \cos(p_1 t + \gamma_1) \operatorname{sech}(p_2 x + \gamma_2)\right], \tag{20}$$

where

$$(1 - \tfrac{1}{4}h^2) \cosh p_2 h = (1 + \tfrac{1}{4}h^2) \cos p_1 h$$

and

$$b_2^2 \sinh^2 p_2 h = b_1^2 \sin^2 p_1 h.$$

Note that in the limit $h \to 0$ these conditions reduce to the analytical ones, (8). Using the symmetry

$$x \to t, \quad t \to x, \quad \phi \to \pi + \phi,$$

one obtains the homoclinic orbit,

$$\begin{aligned} \phi(x,t) = \pi + 4 \tan^{-1} \\ \left[\frac{b_1}{b_2} \cos(p_1 x + \gamma_1) \operatorname{sech}(p_2 t + \gamma_2)\right]. \end{aligned} \tag{21}$$

• It is remarkable that the discrete homoclinic orbit has the same form as the analytical one, apart from a phase shift.

More complicated homoclinic manifolds can also be obtained and are best derived from the N soliton solution of (11). It is not difficult to see that the N soliton solution given by Hirota [16] in light cone coordinates is written in laboratory coordinates,

$$\begin{aligned} f + ig = \sum_{(\mu=0,1)} \exp\left[\sum_{j>k}^{(N)} (A_{jk} + i\pi)\mu_j \mu_k \right. \\ \left. + \sum_{j=1}^{N} \mu_j(\eta_j + i\tfrac{1}{2}\pi)\right], \end{aligned} \tag{22}$$

where

$$\eta_j = P_j x + \Omega_j t + \text{const.}$$

$$(1 - \tfrac{1}{4}h^2) \cosh P_j h = (1 + \tfrac{1}{4}h^2) \cosh \Omega_j h$$

$$\exp A_{jk} = \frac{\cosh(P_j - P_k)h - \cosh(\Omega_j - \Omega_k)h}{\cosh(P_j + P_k)h - \cosh(\Omega_j + \Omega_k)h}$$

and $\sum_{(\mu=0,1)}$ indicates the summation over all possible combinations of $\mu_1 = 0, 1$; $\mu_2 = 0, 1$; ...; $\mu_N = 0, 1$. The N soliton solution is then given by

$$\phi(x,t) = 4 \tan^{-1}(g/f).$$

Next we construct an M breather solution by taking $N = 2M$, $P_{2n-1} = P_{2n} = \tilde{P}_n$, $\Omega_{2n-1} = -\Omega_{2n} = i\tilde{\Omega}_n$, $n = 1, ..., M$. The M homoclinic solution is then obtained from the M breather solution by interchanging space and time variables as before and letting $\phi \to \pi + \phi$.

In order to illustrate the procedure let us choose $N = 2$; we should recover the single homoclinic solution (21). In this case we have

$$\eta_1 = Px + i\Omega t + \gamma_1 + \gamma_2, \quad \eta_2 = Px - i\Omega t + \gamma_1 - \gamma_2,$$

and

$$\exp(A_{21}) = \frac{\sin^2 \Omega h}{\sinh^2 Ph},$$

where γ_1, γ_2 are arbitrary constants. It now follows in a straightforward way that the breather solution is given by

$$\phi = 4 \tan^{-1}[\exp(-A_{21}) \cos(\Omega t + \gamma_2) \operatorname{sech}(Px + \tfrac{1}{2}A_{21} + \gamma_1)],$$

which is the same as what we had in (20). The single homoclinic orbit (21) is obtained by using the symmetry relation above (21) and redefining the constants ($p_1 = \Omega$, $p_2 = P$, etc.).

In the next section we illustrate the qualitative behavior of the integrable discretizations (a more detailed study of its numerical properties is given in [5]).

3.3. *Qualitative Numerical Behavior*

Consider the initial values

$$u(x, 0) = \pi + \varepsilon_0 \cos(\mu x), \quad u_t(x, 0) = 0, \tag{23}$$

with $\mu = 2\pi/L$ and $L = 2\sqrt{2}\pi$. From the earlier discussion (cf. (5)), this corresponds to two unstable modes (one in addition to the unstable $n = 0$ mode). Although these values are close to the homoclinic orbit they are, apparently, not too close; i.e., (6) suggests that the initial values (23) are a distance ε_0 from the homoclinic orbit which is well outside numerical errors (actually we will show that these initial values are, in fact, extremely close to the homoclinic manifold).

The result of integrating (11) over 200 time units using $N = 64$ and $\varepsilon_0 = 0.05$ is shown in Fig. 3. Note the homoclinic crossings when the solution is translated through a multiple of 2π, as well as the oscillations on a grid scale. Neither the homoclinic crossings nor the oscillations are expected for initial values that are close, but not too close, to a homoclinic orbit. Although the oscillations apparently only appear at regular intervals (when the solution is in the vicinity of π), this regularity does not persist—if one integrates for an even longer time the solution becomes saturated by the high frequency oscillations, as shown in Fig. 4 (we now use $\varepsilon_0 = 0.1$).

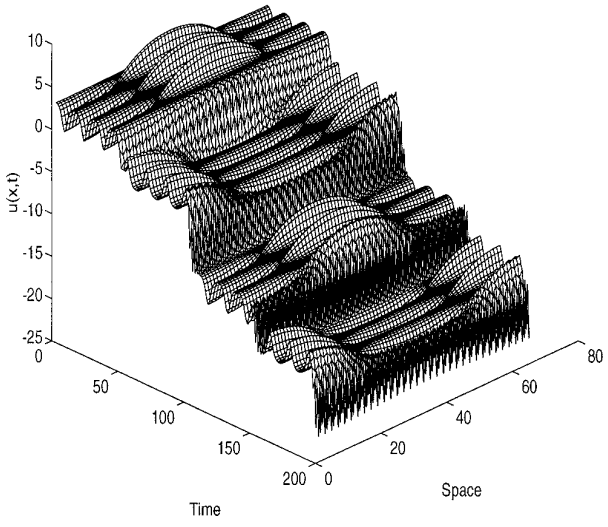


FIG. 3. The integrable discretization.

However, there are brief windows in time when the initial values are approximately recovered. One such window between $t = 450$ and $t = 550$ (with $\varepsilon_0 = 0.1$) is shown in Fig. 5.

For N even, as used in the previous numerical experiments, the presence of the oscillations, as well as windows of stability, are easily explained. Note from (11) that ϕ_m^{n+1} only depends on ϕ_{m-1}^n, ϕ_{m+1}^n, and ϕ_m^{n-1}; there is no contribution from ϕ_m^n. Assuming that N is even, we specify initial values, ϕ_m^0, $m = 1, 3, ..., N - 1$, and ϕ_m^1, $m = 2, 4, ..., N$. Together with (11), these values completely deter-

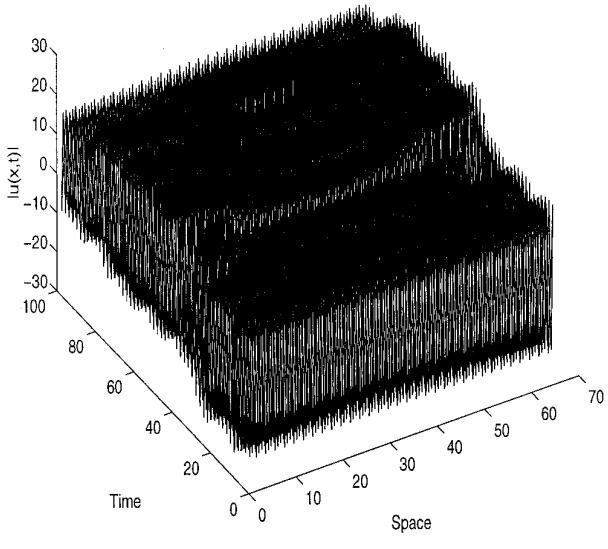


FIG. 4. The "saturation" of the instability between $t = 300$ and $t = 400$ ($\varepsilon_0 = 0.1$).

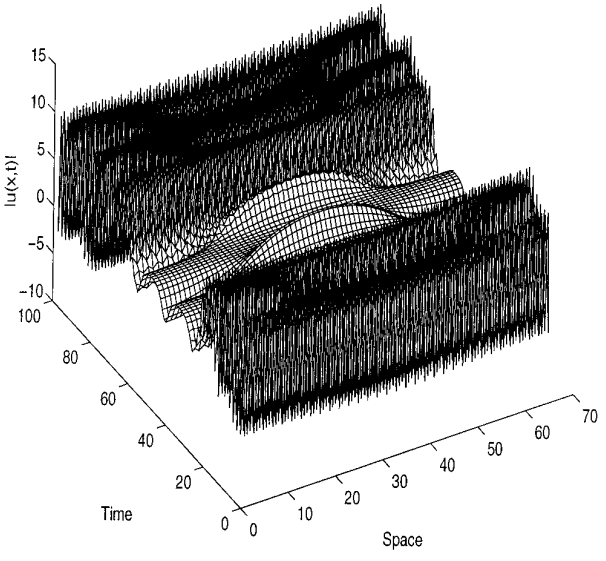


FIG. 5. A window of stability between $t = 450$ and $t = 550$.

mine the solution for all times. However, we may also specify the values, ϕ_m^0, $m = 2, 4, ..., N$, and ϕ_m^1, $m = 1, 3, ..., N - 1$ which, together with (11), again completely determine the solution at all times. We will refer to the two solutions as the even and odd components.

Although the two solutions are computed on the same grid, they evolve completely independently of each other and, if they are plotted on the same graph, any separation of the trajectories followed by the two components, results in the type oscillations we saw in Figs. 3 and 4. The window of stability in Fig. 5 occurred because the two solutions visited the same region of phase space during that time. Figure 6 again shows the solution of Fig. 4, but now with the two components separated. No oscillations are observed and it is clear that the two components follow different trajectories.

Figure 7 shows the modulus of their Fourier transforms, but with the zeroth mode removed. One cannot see any difference in the figure. The differences observed in Fig. 6 are all in the zeroth mode, i.e., the two components differ in the way they translate through 2π. Said differently, they differ in the way they cross the homoclinic orbit.

For the discussion above, we assumed N to be even. For N odd, the situation is different, in this case the two components are coupled through the boundary conditions—we do not have two solutions that develop separately any more. However, the solution again develops rapid oscillations. Using the same initial values as before, the solution for $N = 63$ is shown in Fig. 8 between $t = 400$ and $t = 500$. Note that there is no indication of a saturation by a high frequency oscillation—since the two components are coupled through the boundary conditions, they are not allowed to separate completely and the instability appears and disappears at regular intervals. Figure 9 shows the same separation as in Fig. 6. However, in this case the two components do not satisfy the required boundary conditions—separately they are not solutions of the problem.

3.4. *Linearized Stability*

After the discussion of the previous section it should come as no surprise to learn that the integrable scheme is linearly unstable. Let us briefly return to the case N even. As a consequence of the separation between the even and odd components of the solution it follows that whenever ρ_m^n is a solution of the integrable system (11), then

$$\phi_m^n = (-1)^{m+n}\rho_m^n \tag{24}$$

is also a solution. Assume for a moment that the low modes of ρ_m^n are unstable, i.e., exponentially fast growing, possibly as a consequence of the linearly unstable modes of the continuous sine–Gordon equation which it approximates.

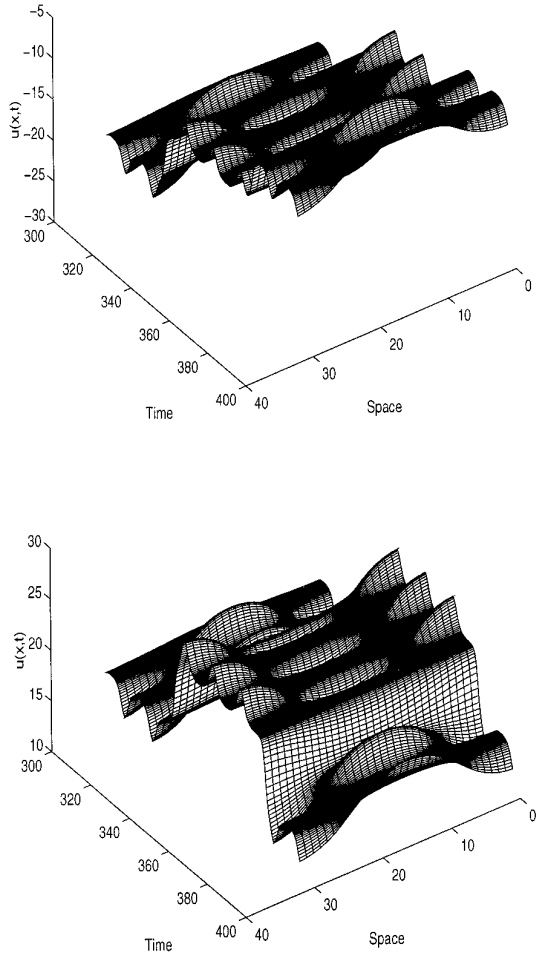


FIG. 6. The even and odd components of the solution shown in Fig. 4.

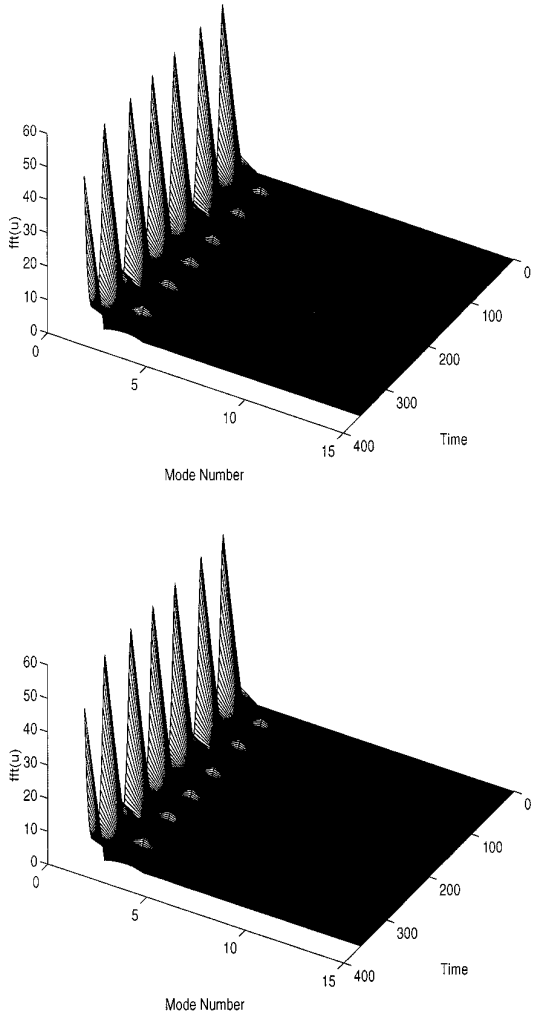


FIG. 7. The Fourier transforms of the two solutions shown in Fig. 6. The zeroth Fourier mode is removed.

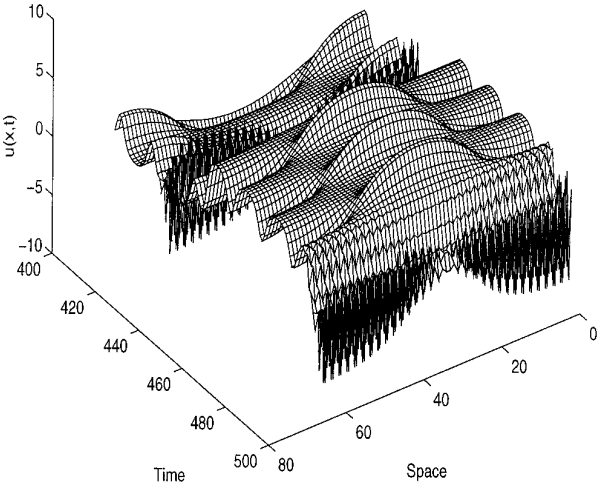


FIG. 8. The integrable discretization: $N = 63$, $400 \le t \le 500$.

Then it is the corresponding high frequency modes of ϕ_m^n in (24) that are unstable. We proceed with a detailed linearized (von Neumann) stability analysis of the discretization (11).

Linearizing around $\phi = \pi$, the following linear partial difference equation is obtained:

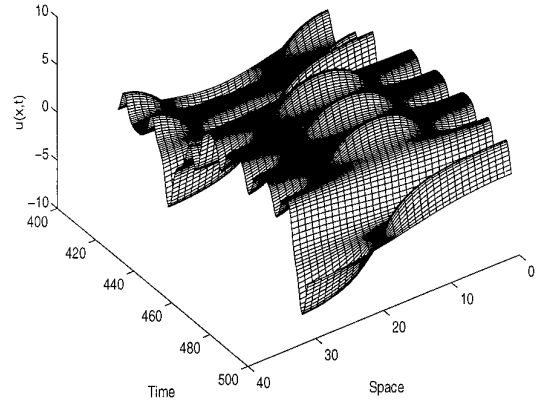
$$(\phi_j^{n+1} - 2\phi_j^n + \phi_j^{n-1})/h^2 - (\phi_{j-1}^n - 2\phi_j^n + \phi_{j+1}^n)/h^2 - \tfrac{1}{4}(\phi_j^{n+1} + \phi_j^{n-1} + \phi_{j+1}^n + \phi_{j-1}^n) = 0. \quad (25)$$

Assuming that $\phi_j^n = r^n \exp(i\mu_l hj)$ and $h = L/N$, where $\mu_l = 2\pi l/L$, $l = -\frac{1}{2}N, ..., \frac{1}{2}N - 1$, and N is the total number of grid points in the interval $(0, L)$, it follows that

$$r^2 - (2/\alpha)\cos(\mu_l h)r + 1 = 0, \quad (26)$$

where

$$\alpha = (1 - \tfrac{1}{4}h^2)/(1 + \tfrac{1}{4}h^2). \quad (27)$$



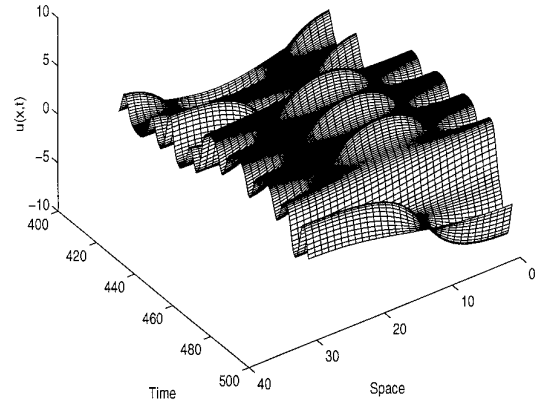


FIG. 9. The even and odd components of the solution shown in Fig. 8.

The stability condition, $|r| \leq 1$ becomes

$$(1/\alpha^2)\cos^2(\mu_l h) \leq 1. \tag{28}$$

Since $\alpha^2 < 1$, it is easy to see that stability condition (28) is violated by low wave numbers near $\mu_l h = 0$ and high wave numbers near $\mu_l h = \pm\pi$. In order to describe the unstable wave numbers more precisely, we note from (28) that the unstable wave numbers satisfy

$$(\cos(\mu_l h) + \alpha)(\cos(\mu_l h) - \alpha) > 0. \tag{29}$$

After some manipulation, where (27) is substituted for α, the *instability* condition becomes

$$(1 - \tfrac{1}{4}h^2 \tan^2(\tfrac{1}{2}\mu_l h))(\tfrac{1}{4}h^2 - \tan^2(\tfrac{1}{2}\mu_l h)) > 0. \tag{30}$$

Thus, all *low* wave numbers are unstable which satisfy the condition,

$$\tan^2 \tfrac{1}{2}\mu_l h < \tfrac{1}{4}h^2, \tag{31}$$

as well as all *high* wave numbers satisfying

$$\tan^2(\tfrac{1}{2}\mu_l h) > 1/(\tfrac{1}{4}h^2). \tag{32}$$

The low wavenumber instability (31) is not surprising—assuming $|\mu_l h| \ll 1$, it becomes

$$(\mu_l)^2 < 1 + O(h^2),$$

which is consistent with the continuous analogue. Therefore, condition (31) is the discrete analogue of the analytical condition (5). On the other hand, condition (32), satisfied by the high wavenumbers near $\mu_l h = \pm\pi$, has no analytical counterpart—it is a discrete artifact. However, there is a direct relationship between the high and low unstable modes. Assuming that μ_l satisfies (31) (i.e., it is any one of the low unstable modes), it follows from $\mu_l h = 2\pi l/N$; i.e., $\pi - \mu_l h = \mu_{(\frac{1}{2})N-l} h$ (similarly for $-\pi + \mu_l h$). This allows us to rewrite the low wavenumber instability condition (31) in the form

$$\tan^2 \tfrac{1}{2}\mu_{\frac{1}{2}N-l} h > 1/(\tfrac{1}{4}h^2). \tag{33}$$

For N even a comparison with (32) shows that every unstable low mode has a corresponding unstable high mode (recall the discussion at the beginning of this section). For N odd the situation is a little different. For instance, assume that the low wavenumber mode μ_l (l small) is just on the edge of the unstable region. Then the high wavenumber mode $\mu_{(1/2)(N-1)}$ lies inside the instability region; i.e., it is unstable.

This analysis shows that the solutions of the integrable discretization is linearly unstable in the vicinity of the homoclinic orbit. For N even the instability may be removed by separating the odd and even components. For N odd this is not possible—the instability cannot be removed.

In this section we have established that in certain regions of phase space, the integrable discretization depends in a sensitive way on the initial values and that the sensitivity can also be characterized as a linear instability, i.e., we have established *that* there is a sensitivity. In the next section we undertake a more detailed investigation of *why* there is a sensitivity.

4. THE SPECTRAL PROBLEM OF THE SINE–GORDON EQUATION

The sensitivity of the numerical solutions on the initial data was clearly illustrated by the ''instabilities'' encountered with the integrable discretization—despite only small differences in the initial values, the even and odd components of the solution follow different trajectories. We now proceed to decribe this sensitivity on the initial values in more detail.

The difficulties encountered by our numerical experiments are caused by peculiarities of the infinite dimensional phase space of the sine–Gordon equation. The underlying homoclinic manifolds have proven to be very sensitive to the perturbations induced by numerical schemes; even the completely integrable, doubly discrete scheme displays a sensitive dependence on initial values, and in a forthcoming paper [5] we investigate the behavior of different numerical schemes in detail.

Exponential instabilities are associated with homoclinic manifolds and one can no longer expect a simple linear drift in the deviations of the numerical solution. Therefore, to assess the timescale on which an accurate numerical solution can be expected, it is necessary to have an a priori estimate of the ''distance'' of the initial data from the homoclinic manifolds. Since the geometry of the sine–Gordon phase space is characterized in terms of the spectrum of an associated linear operator (the $\mathscr{L}^{(x)}(\mathbf{u}, \lambda)$ operator of the Lax pair (34), (35)), we provide a measure of the proximity to homoclinic manifolds in terms of suitable spectral data. Further, for a systematic comparison of the numerical schemes, we have found that the most relevant quantities to monitor are the spectral representation of the ''action'' variables (the ''main'' spectrum). By monitoring the evolution of the spectrum under the perturbed (numerical) flow we can correlate irregularities in the numerical solution to: (i) deviations in the actions, (ii) changes in the number of instabilities present, and (iii) changes in the phase space associated with the initial data. This enables one to determine the ability of the schemes to capture the global dynamics of the system and to more accurately interpret the numerical results.

The sine–Gordon equation (1) arises as the compatibility condition of the following linear operators (i.e., we require that $(d^2/dx\,dt)\mathbf{v} = (d^2/(dt\,dx)\mathbf{v})$ [1, 11],

$$\mathscr{L}^{(x)}(\mathbf{u}, \lambda)\mathbf{v} := \left[A\frac{d}{dx} + \frac{i}{4}B(u_x + u_t) + \frac{1}{16\lambda}C - \lambda I \right]\mathbf{v} = 0, \tag{34}$$

$$\mathscr{L}^{(t)}(\mathbf{u}, \lambda)\mathbf{v} := \left[A\frac{d}{dt} + \frac{i}{4}B(u_x + u_t) - \frac{1}{16\lambda}C - \lambda I \right]\mathbf{v} = 0, \tag{35}$$

where

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$C = \begin{pmatrix} \exp(iu) & 0 \\ 0 & \exp(-iu) \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$\mathbf{u} := (u(x, t), u_t(x, t))$ is the potential, and $\lambda \in \mathbb{C}$ denotes the spectral parameter.

The solutions of the sine–Gordon equation are characterized in terms of the spectrum of $\mathscr{L}^{(x)}$ defined by

$$\sigma(\mathscr{L}^{(x)}) := \{\lambda \in \mathbb{C} | \mathscr{L}^{(x)}\mathbf{v} = 0, |\mathbf{v}| \text{ bounded } \forall x\}. \tag{36}$$

Since the potential $\mathbf{u}$ solves the sine–Gordon equation and is of spatial period L, the spectrum is obtained using Floquet theory. The fundamental matrix, $M(x, x_0; \mathbf{u}, \lambda)$, of the spectral operator (34) is defined by the conditions

$$\mathscr{L}^{(x)}(\mathbf{u}, \lambda)M = 0, \quad M(x_0, x_0; \mathbf{u}, \lambda) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{37}$$

and the Floquet discriminant $\Lambda(\mathbf{u}, \lambda) := \text{tr } M(x_0 + L, x_0; \mathbf{u}, \lambda)$.

In determining the spectrum of $\mathscr{L}^{(x)}(\mathbf{u}, \lambda)$, condition (36) for bounded eigenfunctions yields the following criterion on the discriminant:

$$\sigma(\mathscr{L}^{(x)}) := \{\lambda \in \mathbb{C} | \Lambda(\mathbf{u}, \lambda) \text{ is real and } -2 \le \Lambda(\mathbf{u}, \lambda) \le 2\}. \tag{38}$$

The linear operator is not self-adjoint and the topology of the spectrum in the complex λ plane is determined by the following:

(i) curves of real Λ;

(ii) critical points λ^c,

$$\frac{d}{d\lambda}\Lambda(\mathbf{u}, \lambda)|_{\lambda=\lambda^c} = 0; \tag{39}$$

(iii) periodic (antiperiodic) eigenvalues λ^+ (λ^-),

$$\Lambda(\mathbf{u}, \lambda)|_{\lambda=\lambda^\pm} = \pm 2; \tag{40}$$

(iv) double points λ^d,

$$\Lambda(\mathbf{u}, \lambda^d) = \pm 2, \quad \frac{d}{d\lambda}\Lambda(\mathbf{u}, \lambda)|_{\lambda^d} = 0. \tag{41}$$

Although potentials which have critical points of higher multiplicity in their spectral data exist, in this study we restrict our attention to the case of even potentials which have only simple or double periodic eigenvalues. Without further elaboration, as an example we compute the spectrum for the spatially and temporally uniform solution, $\mathbf{u}(x, t) = (\pi, 0)$. In this case, system (34) has constant coefficients and is readily solved. The Floquet discriminant is given by

$$\Lambda(\mathbf{u}, \lambda) = 2\cos\left(\lambda + \frac{1}{16\lambda}\right)L. \tag{42}$$

Making use of the definitions above, it follows that the continuous spectrum is given by the entire real axis as well as the curve $|\lambda|^2 = \frac{1}{16}$ in the complex plane. The periodic spectrum is given by

$$\lambda_j = \tfrac{1}{2}\left[\frac{j\pi}{L} \pm \sqrt{\frac{j^2\pi^2}{L^2} - \frac{1}{4}}\right], \quad j \text{ integer}. \tag{43}$$

Each of these points is a double point embedded in the continuous spectrum and becomes complex if

$$0 \le \left(\frac{2\pi j}{L}\right)^2 < 1. \tag{44}$$

Comparison with (5) reveals a remarkable fact (which is also true for the NLS and complex modified KdV equations): the linearized instability condition (5) is exactly the same as condition (44) for double points in the complex plane. These double points can split in two ways under perturbations and we associate these with homoclinic or-

bits. In fact, we have arrived at a special case of an important result obtained by [10] which relates homoclinic orbits to complex double points of the spectral problem of the sine–Gordon equation.

Let us briefly digress from the main discussion of this section to point out that not all double points in the nonlinear spectrum are associated with homoclinic orbits and/or instabilities. Consider, for example, the following solution of the sine–Gordon equation on the infinite line (see, for example, [1]),

$$u(x,t) = 4\tan^{-1}[-t\,\mathrm{sech}(x)], \quad -\infty < x < \infty. \tag{45}$$

It represents a limiting situation between a kink, an antikink, and a breather solution which is reflected by a double point in its nonlinear spectrum.

Using the initial values,

$$u(x,0) = 0, \quad u_t(x,0) = 4\,\mathrm{sech}(x) \tag{46}$$

with $-20 \le x \le 20$ and periodic boundary conditions, we solve this using the integrable discretization (11) with $N = 64$ and $N = 128$. The results are shown in Fig. 10a and 10b. The oscillations seen in Fig. 10a are again an indication of a separation of the even and odd components of the solution. However, in this case a very course grid is used (note that the spatial interval is much wider in this case than for the experiments of Section 3.3) and the initial values for the two components differ substantially. It is therefore not surprising that they follow different trajectories. When the grid is refined to $N = 128$ in Fig. 10b the difference in initial values for the two components is reduced. Since there is not the sensitive dependence on the initial values, as was the case near a homoclinic orbit, the trajectories of the two components do not separate on the time scales considered. Thus, although the initial values are represented by a double point in the nonlinear spectrum, there is no exponential growth associated with it and not the sensitivity on initial values as in the case with homoclinic manifolds.

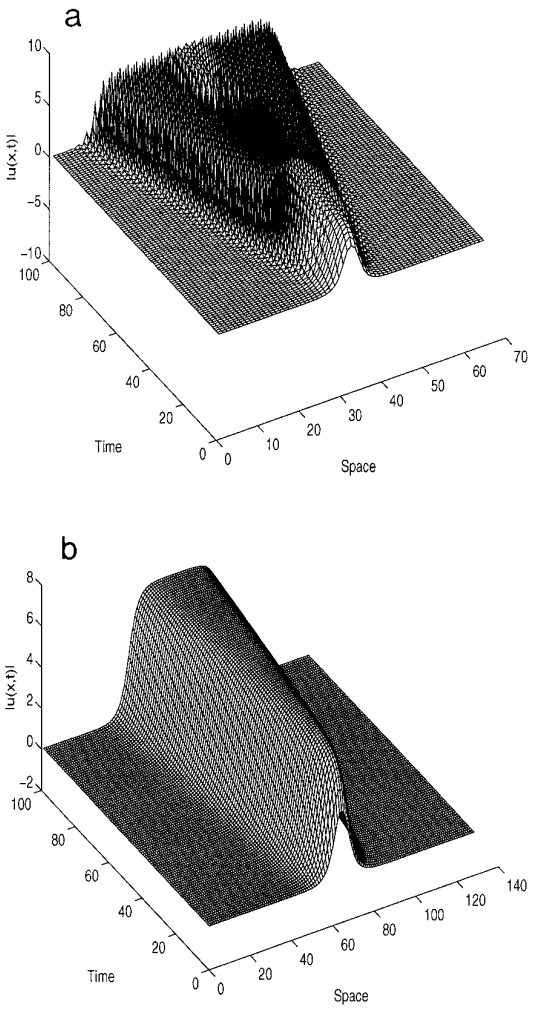


FIG. 10. Soliton initial values: (a) $N = 64$; (b) $N = 128$.

Returning to the main theme of this section, we note that the question of how to determine the proximity of a quasi-periodic solution to a nearby homoclinic manifold can most easily be resolved by examining the spectral data. For a given potential a nearby homoclinic manifold can be obtained by pinching together the pairs of simple eigenvalues along the band of continuous spectrum that they are the endpoints of. Under this procedure, the simple eigenvalues coalesce in double points. If a double point is complex, the associated nonlinear mode may be exponentially unstable. (The reader may find it useful to refer to Fig. 11 below. For instance, note that Fig. 11a is recovered by closing the gap in the spectrum of Fig. 11b.) Let λ_j $(j = 1, N)$ denote the complex double points of the homoclinic manifold associated with the potential $u_0(x, t)$. For potential, $u^\varepsilon(x, t)$, close to $u_0(x, t)$, let the corresponding spectral elements be denoted by $\lambda_j^{(\varepsilon)}$. The proximity of $u^\varepsilon(x, t)$ to a homoclinic submanifold can be measured by

$$\min_{1\le k\le N} |\lambda_k^{(\varepsilon)} - \lambda_k|,$$

where the distance is measured along the continuous spectrum between λ_k^ε and λ_k.

The solutions that we numerically simulate have initial data within an ε neighborhood of $\mathbf{u}(x, 0) = (\pi, 0)$; hence their spectral configuration can be calculated via perturbation analysis. To simplify the analysis we recall a crucial fact—the spectrum is invariant when the potential $\mathbf{u}(x, t)$ evolves under the sine–Gordon flow. Hence we consider $\mathbf{u}(x, 0)$ and suppress the time dependence in the following analysis.

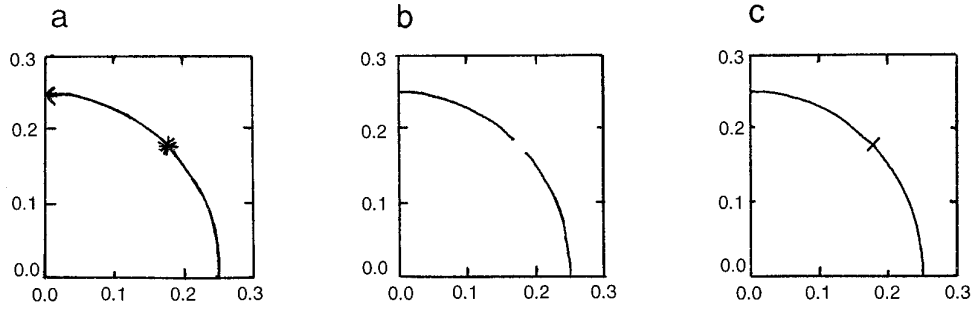


FIG. 11. The nonlinear spectrum: (a) homoclinic orbit; (b) inside the homoclinic orbit; (c) outside the homoclinic orbit.

The initial values (23) we used in our numerical experiments, belong to the class

$$\begin{aligned} u &= \pi + \varepsilon u^{(1)} \\ &= \pi + \varepsilon \cos(\mu_n x), \\ u_t &= \varepsilon\sqrt{1-\mu_n^2}\,(1+p)e^{i\phi}\cos\mu_n x, \end{aligned}$$

where $\mu_n = 2\pi n/L$ (for integer n), $\phi = 0$ for $\mu_n^2 < 1$, $\phi = \pi/2$ for $\mu_n^2 > 1$. At the double points of $\Delta(\lambda, u)$, the discriminant Δ and the eigenfunctions are analytic functions of u, so it is natural to assume the perturbation expansion

$$\begin{aligned} \lambda_j &= \lambda_j^{(0)} + \varepsilon\lambda_j^{(1)} + \varepsilon^2\lambda_j^{(2)} + \ldots \\ \mathbf{v}_j &= \mathbf{v}_j^{(0)} + \varepsilon\mathbf{v}_j^{(1)} + \varepsilon^2\mathbf{v}_j^{(2)} + \ldots, \end{aligned}$$

where

$$\mathbf{v}_j = \begin{pmatrix} v_{j1} \\ v_{j2} \end{pmatrix}.$$

Substituting these expansions into (34) and equating the various orders of ε, we obtain

$$\mathcal{O}(\varepsilon^0) \to \mathscr{L}^{(x)}\mathbf{v}_j^{(0)} = 0, \tag{47}$$

$$\mathcal{O}(\varepsilon) \to \mathscr{L}^{(x)}\mathbf{v}_j^{(1)} = \begin{cases} -s_j^{(1)}v_{j1}^{(0)} - \dfrac{i}{4}q(x)v_{j2}^{(0)} \\ -\tilde{s}_j^{(1)}v_{j2}^{(0)} - \dfrac{i}{4}q(x)v_{j1}^{(0)} \end{cases} \tag{48}$$

$$\mathcal{O}(\varepsilon^2) \to \mathscr{L}^{(x)}\mathbf{v}_j^{(2)} = \begin{cases} -s_j^{(1)}v_{j1}^{(1)} - s_j^{(2)}v_{j1}^{(0)} - \dfrac{i}{4}q(x)v_{j2}^{(1)} \\ -\tilde{s}_j^{(1)}v_{j2}^{(1)} - \tilde{s}_j^{(2)}v_{j2}^{(0)} - \dfrac{i}{4}q(x)v_{j1}^{(1)} \end{cases} \tag{49}$$

$$\vdots$$

$$\mathcal{O}(\varepsilon^m) \to \mathscr{L}^{(x)}\mathbf{v}_j^{(m)} = \begin{cases} -\sum_{k=0}^{m-1} s_j^{(m-k)}v_{j1}^{(k)} - \dfrac{i}{4}q(x)v_{j2}^{(m-1)} \\ -\sum_{k=0}^{m-1} \tilde{s}_j^{(m-k)}v_{j2}^{(k)} - \dfrac{i}{4}q(x)v_{j1}^{(m-1)}, \end{cases} \tag{50}$$

where

$$s_j^{(1)} = \frac{1}{16\lambda_j^{(0)}}\left(\frac{\lambda_j^{(1)}}{\lambda_j^{(0)}} - i\cos\mu_n x\right) - \lambda_j^{(1)}$$

$$\tilde{s}_j^{(1)} = \frac{1}{16\lambda_j^{(0)}}\left(\frac{\lambda_j^{(1)}}{\lambda_j^{(0)}} + i\cos\mu_n x\right) - \lambda_j^{(1)}$$

$$s_j^{(2)} = \frac{1}{16\lambda_j^{(0)}}\left(-\left(\frac{\lambda_j^{(1)}}{\lambda_j^{(0)}}\right)^2 + \frac{\lambda_j^{(2)}}{\lambda_j^{(0)}} + i\frac{\lambda_j^{(1)}}{\lambda_j^{(0)}}\cos\mu_n x + \frac{\cos^2\mu_n x}{2}\right) - \lambda_j^{(2)}$$

$$\tilde{s}_j^{(2)} = \frac{1}{16\lambda_j^{(0)}}\left(-\left(\frac{\lambda_j^{(1)}}{\lambda_j^{(0)}}\right)^2 + \frac{\lambda_j^{(2)}}{\lambda_j^{(0)}} + i\frac{\lambda_j^{(1)}}{\lambda_j^{(0)}}\cos\mu_n x + \frac{\cos^2\mu_n x}{2}\right) - \lambda_j^{(2)}$$

$$\vdots$$

$$s_j^{(m)} = -\frac{1}{16\lambda_j^{(0)}}\sum_{k=0}^{m}\tilde{\lambda}_j^k\frac{(-i\cos\mu_n x)^{m-k}}{(m-k)!} - \lambda_j^{(m)}$$

$$\tilde{s}_j^{(m)} = -\frac{1}{16\lambda_j^{(0)}}\sum_{k=0}^{m}\tilde{\lambda}_j^k\frac{(i\cos\mu_n x)^{m-k}}{(m-k)!} - \lambda_j^{(m)}$$

and

$$q(x) = -\mu_n\sin\mu_n x + \sqrt{1-\mu_n^2}\,(1+p)e^{i\phi}\cos\mu_n x \tag{51}$$

$$\tilde{\lambda}_j^{(m)} = \frac{-\lambda_j^{(m)}}{\lambda_j^{(0)}} + f(\lambda_j^n), \quad n < m. \tag{52}$$

The leading order problem yields the spectrum and eigenfunctions for $\mathbf{u}(x, t) = (\pi, 0)$. As already noted, each element of the periodic spectrum is a double point. We will be concerned with the splitting of each of these double points under the ε perturbation. At the double points, $\lambda_j^{(0)}$ (for simplicity, hereafter we use λ_j, instead of $\lambda_j^{(0)}$), the eigenspace is two-dimensional and is spanned by the eigenfunctions

$$\phi_j^{\pm} = \exp(\pm i k_j x) \begin{pmatrix} 1 \\ \pm i \end{pmatrix}$$

and the general solution is given by

$$\mathbf{v}_j^{(0)} = A^+ \phi_j^+ + A^- \phi_j^-,$$

where $k_j = \lambda_j + 1/16\lambda_j = j\pi/L$.

The eigenvalues are associated with the periodic/antiperiodic eigenfunctions of period L. The solvability condition for the system

$$\mathscr{L}^{(x)}\mathbf{v} = F,$$

with

$$F = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}$$

given by the orthogonality condition

$$\int_0^L (F_1 w_1^* + F_2 w_2^*)\, dx = 0$$

for all w in the nullspace of the Hermitian adjoint operator, $\mathscr{L}^H$, where

$$\mathscr{L}^H = \left[\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \frac{d}{dx} - \frac{i}{4} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} (u_x + u_t) + \frac{1}{16\lambda^*} \begin{pmatrix} \exp(-iu) & 0 \\ 0 & \exp(iu) \end{pmatrix} - \lambda^* I \right].$$

Noting that the nullspace of $\mathscr{L}^H$, at the double points, is spanned by

$$\begin{pmatrix} (\phi_{j1}^{\pm})^* \\ (\phi_{j2}^{\pm})^* \end{pmatrix},$$

the general solvability condition assumes the form

$$\int_0^L (F_1 \phi_{j1}^{\pm} + F_2 \phi_{j2}^{\pm})\, dx = 0. \tag{53}$$

The solvability conditions (53) applied to (48) yield the system of equations

$$\begin{pmatrix} T_+ & T \\ T & T_- \end{pmatrix} \begin{pmatrix} A^+ \\ A^- \end{pmatrix} = 0,$$

where

$$T = 2\lambda_j^{(1)} \left(1 - \frac{1}{16\lambda_j^2} \right)$$

$$T_{\pm} = \begin{cases} i \left(\dfrac{1}{16\lambda_j} - \dfrac{\mu_n \pm i(1+p)e^{i\phi}\sqrt{1-\mu_n^2}}{4} \right), & j = n, \\ 0, & j \neq n. \end{cases}$$

Consequently, the system can be solved nontrivially for $A^{\pm}$ if $T_+ T_- - T^2 = 0$, or

$$(\lambda_j^{(1)})^2 = \begin{cases} -\dfrac{1 - (1+p)^2 e^{2i\phi}}{16} \lambda_j^2, & j = n, \\ 0, & j \neq n. \end{cases}$$

At $\mathscr{O}(\varepsilon)$, there is a correction only to the double point, λ_j ($j = n$). A specific double point is selected in resonance with the perturbation in the eigenfunction. The other double points do not experience an $\mathscr{O}(\varepsilon)$ correction.

The behavior of the correction $\lambda_j^{(1)}$ ($j = n$) depends on whether the double point λ_j is real or complex,

$$(\lambda_j^{(1)})^2 = \begin{cases} -\dfrac{p(2+p)}{16} \lambda_j^2 & \text{for } \lambda_j \text{ complex}, j = n, \\ -\dfrac{1 + (1+p)^2}{16} \lambda_j^2 & \text{for } \lambda_j \text{ real}, j = n, \\ 0, & j \neq n. \end{cases} \tag{54}$$

For complex λ_n, the correction $\lambda_n^{(1)}$ can be real, zero, or pure imaginary, depending on the sign of p in the perturbed potential. For $p = 0$, $\lambda_n^{(1)} = 0$; hence the perturbation is in the direction of the homoclinic orbit. For $p = 1$, the double point splits into a "gap" in the spectrum along the arc of $|\lambda|^2 = \frac{1}{16}$. For $p = -1$ the double point splits into a "cross"

along the radius, as shown in Fig. 11. Recall that Figs. 1 and 2 show the corresponding waveforms.

For real λ_n, regardless of the sign of p, the perturbation is imaginary, thus there is only one way for the eigenvalues to split. "Gaps" cannot appear on the real axis in the spectrum of $\mathscr{L}^{(x)}$ (this can also be established by symmetry arguments). Hence the situation with real double points is very different from that of complex double points. Splitting of the real double points introduces additional degrees of freedom into the spatial structure but does not introduce any exponential instabilities and are therefore not associated with homoclinic manifolds.

To examine the higher order corrections to the double points for which $j \neq n$, we will consider the case of $p = -1$. This corresponds exactly to initial data (23). The analysis for the other cases ($p = 0, 1$) is analogous. At these double points, $\lambda_j^{(1)} = 0$, so the nonhomogeneous term of the linear operator simplifies, i.e.,

$$\mathscr{L}^{(x)}\mathbf{v}_j^{(2)} = \begin{cases} -s_j^{(1)} v_{j1}^{(1)} - s_j^{(2)} v_j^{(0)} - \dfrac{i}{4} q(x) v_{j2}^{(1)} \\ \\ -\tilde{s}_j^{(1)} v_{j2}^{(1)} - \tilde{s}_j^{(2)} v_j^{(0)} - \dfrac{i}{4} q(x) v_{j1}^{(1)} \end{cases}, \qquad (55)$$

where now

$$s_j^{(1)} = \frac{1}{16\lambda_j}(-i\cos\mu_n x)$$

$$\tilde{s}_j^{(1)} = \frac{1}{16\lambda_j}(i\cos\mu_n x)$$

$$s_j^{(2)} = \frac{1}{16\lambda_j}\left(\frac{\lambda_j^{(2)}}{\lambda_j} + \frac{\cos^2\mu_n x}{2}\right) - \lambda_j^{(2)}$$

$$\tilde{s}_j^{(2)} = \frac{1}{16\lambda_j}(i\cos\mu_n x)$$

$$s_j^{(2)} = \frac{1}{16\lambda_j}\left(\frac{\lambda_j^{(2)}}{\lambda_j} + \frac{\cos^2\mu_n x}{2}\right) - \lambda_j^{(2)}.$$

In order to apply the solvability condition to $\mathscr{L}^{(x)}\mathbf{v}_j^{(2)} = F$ we need the $\mathscr{O}(\varepsilon)$ correction to the eigenfunctions, $\mathbf{v}_j^{(1)}$. We find these to be given by

$$\mathbf{v}_j^{(1)} = \mathbf{A}e^{i(k_j+\mu_n)x} + \mathbf{B}e^{i(k_j-\mu_n)x} + \mathbf{C}e^{i(-k_j+\mu_n)x} + \mathbf{D}e^{-i(k_j+\mu_n)x},$$

where

$$\mathbf{A} = \frac{\alpha_+ A^+}{2(-2k_j - \mu_n)}\begin{pmatrix} i \\ 1 \end{pmatrix}$$

$$\mathbf{B} = \frac{\alpha_- A^+}{2(-2k_j + \mu_n)}\begin{pmatrix} i \\ 1 \end{pmatrix}$$

$$\mathbf{C} = \frac{\alpha_- A^-}{2(-2k_j + \mu_n)}\begin{pmatrix} i \\ -1 \end{pmatrix}$$

$$\mathbf{D} = \frac{\alpha_+ A^-}{2(-2k_j - \mu_n)}\begin{pmatrix} i \\ -1 \end{pmatrix}$$

and

$$\alpha_\pm = \frac{1}{16\lambda_j} \pm \frac{\mu_n}{4}.$$

Returning to the solvability condition at $\mathscr{O}(\varepsilon^2)$ we obtain the system

$$\begin{pmatrix} \beta & \lambda_j^{(2)} - \gamma \\ \lambda_j^{(2)} - \gamma & \beta \end{pmatrix}\begin{pmatrix} A^+ \\ A^- \end{pmatrix} = 0, \qquad (56)$$

where

$$\beta = \begin{cases} -\dfrac{\alpha_+\alpha_-}{4(-2k_j + \mu_n)} + \dfrac{1}{128\lambda_j}, & j = 2n, \\ \\ 0, & j \neq 2n, \end{cases}$$

$$\gamma = \frac{-16\lambda_j\left(\dfrac{\alpha_+^2}{-4(2k_j + \mu_n)} + \dfrac{\alpha_-^2}{4(-2k_j + \mu_n)}\right) - \dfrac{1}{4}}{1 - 16\lambda_j}.$$

Having a nontrivial solution, (56) implies

$$\lambda_j^{(2)} = \gamma \pm \beta.$$

Consequently only the double point λ_j ($j = 2n$) experiences an ε^2-splitting of 2α. The other double points, $j \neq n, 2n$, experience just a translation of $\lambda_j^{(2)}$ at $\mathscr{O}(\varepsilon^2)$.

At $\mathscr{O}(\varepsilon^m)$, to examine the splitting of the double points λ_j, $j \neq n, 2n, \ldots, (m-1)n$, note that the general form of the correction to the eigenfunction, $\mathbf{v}_j^{(m-1)}$, is

$$\mathbf{v}_j^{(m-1)} = \mathbf{A}^{(m-1)}e^{i(k_j+(m-1)\mu_n)x} + \mathbf{B}^{(m-1)}e^{i(k_j-(m-1)\mu_n)x} + \mathbf{C}^{(m-1)}e^{i(-k_j+(m-1)\mu_n)x} + \mathbf{D}^{(m-1)}e^{-i(k_j+(m-1)\mu_n)x} + \mathbf{v}_j^{(m-2)},$$

where $\mathbf{A}^{(m-1)}$, $\mathbf{B}^{(m-1)}$, $\mathbf{C}^{(m-1)}$, $\mathbf{D}^{(m-1)}$ are column vectors. The solvability condition at $\mathcal{O}(\varepsilon^m)$ yields a system of the general form

$$\begin{pmatrix} \hat{\alpha}^+ & \lambda_j^{(m)} \\ \lambda_j^{(m)} & \hat{\alpha}^- \end{pmatrix} \begin{pmatrix} A^+ \\ A^- \end{pmatrix} = 0, \tag{57}$$

where

$$\alpha^{\pm} = \begin{cases} \alpha_m^{\pm}, & j = mn, \\ 0, & j \neq mn, \end{cases}$$

where $\alpha_m^{\pm}$ is a function of the coefficients of the eigenfunctions at the previous iterate.

For (57) to have a nontrivial solution we require

$$\lambda_j^m = \begin{cases} \pm \sqrt{\hat{\alpha}_m^+ \hat{\alpha}_m^-}, & j = mn, \\ 0, & j \neq mn. \end{cases}$$

Thus, we arrive at the following general result: Only the double points, λ_j, which are m-fold multiples of the fundamental mode ($j = mn$) will experience an ε^m-splitting. The splitting distance of the remaining double points ($j \neq mn$) is beyond all orders in ε!

When there is more than one complex double point in the initial spectral configuration and the initial condition contains a perturbation of the mode, μ_1, for example, the mode μ_0 is not affected with the result that this initial value is "exponentially close" to a homoclinic submanifold. This is exactly the situation with the initial data (23) used in our numerical experiments—it splits the first double point at $\lambda = i\pi/4$ into an $\mathcal{O}(\varepsilon)$ gap. However, the zeroth double point on the imaginary axis (which corresponds to the most unstable mode) does not split at any order in ε. Thus the initial data used in the numerical experiments actually lies exponentially close to a homoclinic orbit (represented by the complex double point $\lambda = i\pi/2$) and not, as one might naïvely believe, some distance away from it. Due to this exponential proximity we refer to this as an "effectively" chaotic regime [18].

5. CONCLUSIONS

From a numerical point of view the effectiveness of the doubly discrete, integrable discretizations of the sine–Gordon equation can be considered as questionable because of the high frequency oscillations as observed, for instance, in Fig. 3. A close investigation shows that if an even number of gridpoints (N even) is used, the oscillations indicate a separation of two independent solutions computed on the same grid. Since the initial values for the two solutions are near each other, the separation is an indication of a sensitive dependence on initial values. In fact, our perturbation analysis shows that the initial values used in our numerical experiments are exponentially close to a homoclinic submanifold where sensitive dependence on initial values is expected. Therefore we explain this is a nonlinear instability; two solutions, starting from slightly different initial values, follow very different trajectories. Note that due to the sensitivity it is not possible to decide which one is the more acceptable solution.

If N is odd, the two even and odd components of the solution are coupled through the boundary conditions. In this case one can think of a separation of the solution into even and odd components, but now on a grid twice as long, i.e., on a $2N$ grid. However, on the original N grid the two components do not satisfy the prescribed boundary conditions and do not approximate the continuous solution. Consequently, the integrable discretization allows exact solutions that have no counterpart in the continuous sine–Gordon equation.

Let us point out that spurious solutions are also allowed by the Ablowitz–Ladik discretization of the nonlinear Schrödinger equation; see, for example [4, 17]. However, in this case the conservation laws do not permit high frequency oscillations to develop from smooth initial values such as observed for the doubly discrete, integrable discretization of the sine–Gordon equation; see, for example, [6].

Finally we point out that, apart from providing a better understanding of the structure of phase space, the nonlinear spectrum also provides a useful description of the qualitative properties of the different numerical schemes. This can be used as a diagnostic tool to compare the qualitative performance of different numerical schemes in a quantitative manner. This idea will be fully developed in a subsequent paper [5].

ACKNOWLEDGMENTS

This work was partially supported by AFOSR Grant F49620-94-0120, ONR Grant N00014-94-0194, and NSF Grants DMS-9024528, DMS-9404265.

REFERENCES

1. M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
2. M. J. Ablowitz and B. M. Herbst, *SIAM J. Appl. Math.* **50,** 339 (1990).
3. M. J. Ablowitz and P. A. Clarkson, *Solitons, Nonlinear Evolution Equations and Inverse Scattering* (Cambridge Univ. Press, Cambridge, UK, 1991).
4. M. J. Ablowitz, B. M. Herbst, and J. A. C. Weideman, *IMA J. Numer. Anal.* **11,** 539 (1991).
5. M. J. Ablowitz, B. M. Herbst, and C. M. Schober, in preparation.

6. M. J. Ablowitz, B. M. Herbst, and C. M. Schober, preprint.
7. T. B. Benjamin, *Proc. Roy. Soc. A* **299,** 59 (1967).
8. A. Bobenko, N. Kutz, and U. Pinkall, SFB 288, Preprint No. 42, Berlin, 1992 (unpublished).
9. A. Bobenko and U. Pinkall, SFB 288, Preprint No. 127, Berlin, 1994 (unpublished).
10. N. Ercolani, M. G. Forest, and D. W. McLaughlin, *Phys. D* **43,** 349 (1990).
11. L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer-Verlag, Berlin, 1987).
12. B. Fornberg and G. B. Whitham, *Philos. Trans. R. Soc. London* **289,** 373 (1978).
13. B. M. Herbst and M. J. Ablowitz, *Phys. Rev. Lett.* **62,** 2065 (1989).
14. B. M. Herbst and M. J. Ablowitz, *Quaestiones Math.* **15,** 345 (1992).
15. B. M. Herbst and M. J. Ablowitz, *J. Comput. Phys.* **105,** 122 (1993).
16. R. Hirota, *J. Phys. Soc. Japan* **43,** 2079 (1977).
17. P. D. Miller, Ph.D. thesis, Program in Applied Mathematics, University of Arizona, 1994 (unpublished).
18. C. M. Schober, M. J. Ablowitz, and B. M. Herbst, *Phys. Rev. Lett.* **71,** 2683 (1993).
19. G. G. Stokes, *Cambridge Trans.* **8,** 441 (1847).
20. G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
21. N. J. Zabusky and M. D. Kruskal, *Phys. Rev. Lett.* **15,** 240 (1965).