JOURNAL OF COMPUTATIONAL PHYSICS 117, 205–214 (1995)

Numerical Methods for Nonlinear Fourier Analysis, Prediction, and Filtering

D. CANDELA

Laboratory for Low-Temperature Physics, Physics and Astronomy Department, University of Massachusetts, Amherst, Massachusetts 01003

Received January 19, 1993; revised September 7, 1994

The inverse spectral transform is numerically implemented to analyze arbitrary initial data for the Korteweg–de Vries equation with periodic boundary conditions and to predict the wave field at any point in spacetime without integration. Explicit analysis and prediction algorithms are described for arbitrary values of the gap number N. It is found that the method with $N \sim 10$ is computationally inexpensive and provides a useful analysis for a class of problems likely to occur in experimental situations. As an example the method is applied to randomized initial data, in a setting where neither soliton nor linear approximations are accurate. Wave structures occurring with these initial conditions are classified as multi-gap quasiperiodic modes and are simplified by nonlinear filtering. The efficiency of the method is briefly compared with that of direct numerical integration.

1. INTRODUCTION

Fourier analysis solves the initial-condition problem for a linear partial differential equation (PDE) by finding the amplitudes and phases of a set of harmonic modes. The solution may be directly evaluated at any spacetime point (x, t) without integration, as a linear superposition of the modes. In addition to being an efficient method for projecting arbitrary initial data forward in time, Fourier analysis provides a powerful classification of solutions to linear PDEs in terms of their spectral content.

For integrable nonlinear PDEs (Korteweg–de Vries (KdV), nonlinear Schrödinger, sine-Gordon, etc.) the inverse-scattering transform (IST) provides an analogous solution [1]. On an infinite domain these PDEs evolve any localized initial disturbance into weak linear waves plus solitons, which the IST identifies with bound states of an associated eigenvalue problem. However, the IST in this form only solves a limited class of problems, those for which the wave field decays rapidly at large distances.

Dubrovin and Novikov [2] and Its and Matveev [3] generalized the IST to the class of "N-gap" solutions to the KdV equation, which can accurately approximate any well-behaved periodic or quasiperiodic function [4]. This broad class includes all soliton solutions and linear waves as limiting cases, as well as other solutions which are not accurately describable in terms of solitons and linear waves. Briefly, an N-gap solution is one for which the band structure of a Schrödinger equation derived from the wave data has exactly N gaps, or forbidden energy ranges. The method of Refs. [2–3], generally termed the inverse *spectral* transform (ISPT) in the literature, has been extended to a number of other integrable nonlinear equations [2, 5, 6].

The ISPT generalizes the Fourier series to integrable nonlinear systems, in the sense that an analysis of initial-condition data $u(x, t = 0)$ enables prediction of the wave field $u(x, t)$ at any spacetime point without integration. As with linear Fourier analysis, the ISPT enables arbitrary data to be classified in terms of the dominant modes present and permits operations such as filtering to suppress undesired modes (noise). It has been shown [7] that the ISPT becomes ordinary Fourier analysis in the linear (small-amplitude) limit.

Although many interesting results on the ISPT have been obtained, it has not been widely used as a numerical method for analyzing experimental data (Refs. [8, 9] are notable exceptions). This may be because the original work was expressed in the language of algebraic geometry and Reimann-surface theory, and the ISPT has been perceived as a mathematically intricate method unsuitable for numerical work. The central result of the present work is that the full ISPT (analysis and inverse analysis or prediction) is computationally inexpensive and readily applicable to certain common classes of problems. Explicit numerical methods are described here for analyzing and projecting arbitrary periodic data, and a few examples are given.

Some aspects of the analysis described in this paper have already been employed by other workers for different purposes. In particular, there is a substantial literature applying the ISPT to the sine-Gordon equation; one aim has been to investigate the transition to chaos exhibited by the damped, driven version of this equation [10–12]. It should be cautioned that the ISPT is considerably more complicated for PDEs like the sine-Gordon equation and the nonlinear Schrödinger equation, that are associated with non-self-adjoint spectral problems [13]. Furthermore, some of these PDEs (sine-Gordon, and *focussing* nonlinear Schrödinger) posses integrable instabilities. While these instabilities are an interesting subject in their own right [14],

0021-9991/95 $6.00

they render problematic the application of the numerical tools presented here. Both of these topics (non-self-adjoint associated problems and PDEs with instabilities) are beyond the scope of the present paper.

Several groups have used the ISPT to generate families of solutions to integrable equations [5, 6, 14–16]. This corresponds roughly to the predictive part of the present algorithm. Unlike the present work, the nonlinear modes were not related to specific initial-condition data. Conversely, ISPT methods have been used to find the soliton content of data from approximately integrable physical [9] and numerical [12] systems, corresponding to the analysis part of the present work. In contrast to these works, the present analysis is carried further, computing not only the spectrum present in the data but also the frequencies, wavevectors, initial phases, and θ-matrix for the nonlinear modes. This additional information permits rapid projection of the initial conditions to arbitrary points in spacetime, with or without filtering.

Another difference from earlier work concerns the qualitative nature of the solutions considered. Previous authors concentrated on soliton and breather solutions, while the present work is concerned primarily with solutions far from both soliton and linear limits. Osborne and Bergamasco [7] termed the spectrum associated with this sort of mode the intermediate spectrum. The viewpoint advanced here is that the intermediate spectrum will dominate the wave motion in many physical situations, and the ISPT provides the only accurate description of the nonlinear modes present in these cases.

It is hoped that the numerical methods detailed in this paper will be useful for analyzing a wide variety of physical data. The KdV and related integrable PDEs are valid approximations to many physical systems. Whenever such systems are subject to strong periodic forcing which is not particularly localized within the period, nonlinear modes will arise which may be quantitatively analyzed using the ISPT (but not the IST, nor linear theory). The forcing may be spatial or temporal, as the "space" variable x in integrable equations usually maps onto physical space or time to the same order of approximation.

The plan of this paper is as follows. Section 2 introduces notation for nonlinear Fourier series representations of solutions to the KdV equation with periodic boundary conditions. In Section 3 the ISPT formalism is reduced to formulas and algorithms suitable for numerical implementation. Section 4 describes a particular implementation that has been successfully used, and Section 5 details a sample application of this implementation to integration and nonlinear filtering for randomized initial conditions. Section 6 briefly compares the computational efficiency of the ISPT with that of direct numerical integration.

2. THE NONLINEAR FOURIER SERIES

The KdV equation with periodic boundary conditions is

$$u_t + 6uu_x + u_{xxx} = 0, \quad u(x + X, t) = u(x, t). \tag{1}$$

Omitting the second term in (1) gives a linear, dispersive PDE. A Fourier representation of a solution to the linearized equation is

$$\begin{gathered} u(x, t) = u_0 + \sum_{n=0}^{N-1} A_n \cos(\phi_{n0} + k_n x - \omega_n t), \\ k_n = 2\pi(n + 1)/X, \quad \omega_n = -k_n^3. \end{gathered} \tag{2}$$

The quantities $\{u_0, \phi_{n0}, A_n, n = 0 \ldots N - 1\}$ are a linear Fourier series, which may be evaluated using ordinary Fourier analysis from initial data $u(x, 0)$. Accurate approximations to well-behaved initial data are obtained for large, finite N.

Following Dubrovin and Novikov [2], an N-gap solution to (1) has an analogous form

$$\begin{gathered} u(x, t) = u_0 + 2(\partial^2/\partial x^2) \ln \theta(i\phi_0 \ldots i\phi_{N-1}), \\ \phi_n = \phi_{n0} + k_n x - \omega_n t, \quad n = 0 \ldots N - 1. \end{gathered} \tag{3}$$

The Riemann θ-function depends on a real, symmetric, negative-definite matrix B_{nm}, and is defined as an infinite sum over N-vectors of integers $[M_0 \ldots M_{N-1}]$ as

$$\begin{aligned} &\theta(i\phi_0 \ldots i\phi_{N-1}) \\ &\quad = \sum_{M_0 \cdots M_{N-1} = -\infty}^{\infty} \exp\left(\frac{1}{2} \sum_{n,m=0}^{N-1} M_n B_{nm} M_m + \sum_{n=0}^{N-1} i\phi_n M_n\right). \end{aligned} \tag{4}$$

The matrix B_{nm} plays a similar role in the nonlinear theory, as does the vector A_n of Fourier amplitudes in the linear theory. It specifies the modes that are present and their relative strengths. It can be computed numerically from arbitrary wave data, as detailed below. The θ-function is periodic in each of the N phases

$$\begin{aligned} &\theta(i\phi_0 \ldots i[\phi_n + 2\pi] \ldots i\phi_{N-1}) \\ &\quad = \theta(i\phi_0 \ldots i\phi_n \ldots i\phi_{N-1}), \quad n = 0 \ldots N - 1. \end{aligned} \tag{5}$$

The periodicity of (1) therefore forces the wavenumbers k_n to be integer multiples of $2\pi/X$. As discussed below k_n may be chosen to be as in (2) without loss of generality. Conversely, the frequencies ω_n depend upon the nonlinear mode present. Equations (3)–(4) generalize the linear Fourier series to the nonlinear setting, as they include all harmonics and cross-harmonics of the N frequencies and wavenumbers. The linear series is recovered when the θ-matrix is large and nearly diagonal, $B_{nm} \approx -C_n \delta_{nm}$ with $C_n \gg 1$. This may easily be seen by inserting B_{nm} of this form into (4) and then discarding terms in the infinite sum that are relatively smaller by one or more factors of $\exp(-C_n)$. Then by linearizing the logarithm in (3), one obtains the linear solution (2) with $A_n = -4k_n^2 \exp(-C_n/2)$.

The set of quantities $\{u_0, \omega_n, \phi_{n0}, B_{nm}, n, m = 0 \ldots N - 1\}$ will be called a nonlinear Fourier series (NLFS). The ISPT

formalism provides a means to compute the NLFS from arbitrary initial conditions $u(x, 0)$. The wave field at arbitrary (x, t) may be evaluated from the NLFS without integration directly from Eqs. (3)–(4). Unlike the linear Fourier series, the components of the NLFS (apart from the initial phases ϕ_{n0}) may not be chosen arbitrarily or independently of one another [2, 15]. This is a serious technical obstacle to using the ISPT to generate families of solutions to (1) and is related to some problems in the theory of θ-functions and Riemann surfaces [17, 18]. For present purposes this problem does not arise as NLFS calculated from Cauchy data $u(x, 0)$ automatically satisfy the necessary consistency relations.

3. IMPLEMENTATION OF THE ISPT

In this section the ISPT results of Dubrovin and Novikov [2] are expressed in a form suitable for numerical calculations. These results are summarized here only to the extent necessary to introduce the notation used. Although derivations are not repeated from Ref. [2], it should be emphasized that all formulae and quantities needed for numerical work are completely specified below.

The eigenvalues E of the Schrödinger problem

$$-\psi_{xx} - u(x, t)\psi = E\psi \tag{6}$$

are invariant as u evolves according to the KdV equation. In the periodic case the eigenvalues fall in bands separated by gaps, and the band structure determines the characteristics of the nonlinear mode. An N-gap solution $u(x, t)$ is one for which there are N open gaps in the band structure. A general solution to (1) has infinitely many open gaps, but they typically decrease rapidly in width with increasing energy [2]. As for linear Fourier series, approximations to arbitrary solutions are obtained by truncating to finite N.

Label the groundstate energy of (6) E_0, and the bottom and top edges of the gaps $B_0, T_0 \ldots B_{N-1}, T_{N-1}$. These points (the "main spectrum") are independent of x and t. The ISPT uses the analytic structure of the Riemann surface Γ of the root

$$R^{1/2}(E) = \left[(E - E_0)\prod_{n=0}^{N-1}(E - B_n)(E - T_n)\right]^{1/2}. \tag{7}$$

The topological genus of Γ is N, the number of gaps. Figure 1 depicts the Riemann surface Γ for the cases $N = 1, 2$ along with the cycles (integration contours) that are needed for ISPT analysis. The choice of these cycles is discussed below. In addition to the main spectrum one considers the "auxiliary spectrum" $\mu_0 \ldots \mu_{N-1}$ which does vary with x and t; μ_n lies in the gap (B_n, T_n) and is defined as the eigenvalue E for which an eigenfunction $\psi(x)$ of (6) has a node at the point x. As x or t varies, each μ_m traces a cycle on Γ from B_n to T_n along the real axis on one sheet, then back to B_n on the other sheet. The

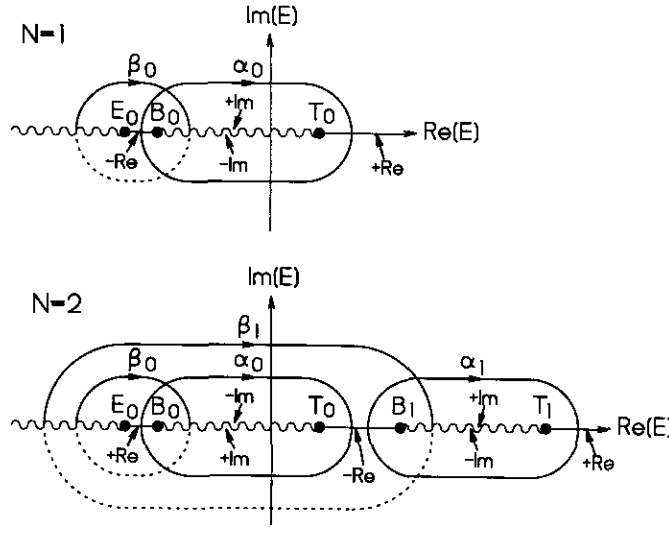

FIG. 1. Basis of cycles $(\alpha_0 \ldots \alpha_{N-1}, \beta_0 \ldots \beta_{N-1})$ on the Riemann surface Γ, for $N = 1$ (top) and $N = 2$ (bottom). The labels $\pm$Re, $\pm$Im specify the sign of $R^{1/2}(E)$ and whether it is real or imaginary, on the indicated segments of the real axis on the top sheet. The sign convention is that $R^{1/2}(E)$ is always positive on the top sheet for $E > T_{N-1}$. With this choice of cycles and signs the wavenumbers k_n are given by Eq. (2) in the text.

wave $u(x, t)$ is given in terms of the auxiliary spectrum by the trace formula

$$u(x, t) = -E_0 + \sum_{n=0}^{N-1} 2\mu_n(x, t) - B_n - T_n. \tag{8}$$

Thus each gap (B_n, T_n) contributes amplitude $\Delta u = 2(T_n - B_n)$ to the nonlinear wave field. If the main spectrum consists of gaps much wider than the bands that separate them ($T_n - B_n > 100[B_n - T_{n-1}]$), the wave motion approximates a multisoliton solution for the nonperiodic (IST) problem [9]. Conversely, narrow gaps separated by wide bands correspond to linear waves. Many possible nonlinear modes are far from either of these limits. These modes have bands and gaps of comparable widths and require the ISPT for quantitative analysis.

The main and auxiliary spectra may be found by computing the real unimodular 2×2 matrix $\mathbf{T}(x, t, E)$ which relates the eigenfunction ψ and its spatial derivative at the point x to these quantities at the point $x + X$,

$$\mathbf{T}(x, t, E) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \begin{bmatrix} \psi_x(x + X) \\ \psi(x + X) \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \psi_x(x) \\ \psi(x) \end{bmatrix}. \tag{9}$$

To calculate $\mathbf{T}(x, t, E)$ the Schrödinger equation (6) is represented as coupled first-order differential equations

$$\frac{d}{dx'}\begin{bmatrix}\psi_x\\ \psi\end{bmatrix}=\begin{bmatrix}0 & -[E+u(x',t)]\\ 1 & 0\end{bmatrix}\begin{bmatrix}\psi_x\\ \psi\end{bmatrix}. \tag{10}$$

Equation (10) is integrated from $x' = x$ to $x' = x + X$, starting with the two initial conditions [1, 0] and [0, 1], yielding the columns $[a, c]$ and $[b, d]$ of $\mathbf{T}(x, t, E)$. The main spectrum is the set of points E for which the Floquet discriminant $\mathcal{F}(x, t, E) = \frac{1}{2}(a + d)$ is 1 or -1, and the auxiliary spectrum is the set of points E within the gaps of the main spectrum for which the off-diagonal element c of $\mathbf{T}(x, t, E)$ vanishes [2]. An equivalent formulation in terms of a 2×2 complex matrix has also been used [2, 19], but the definitions used here appear simpler for numerical work.

Osborne and Segre [19] have numerically integrated the nonlinear ordinary differential equations obeyed by the auxiliary spectrum, which proves difficult for soliton-like $u(x, t)$ because each μ_n is exponentially close to B_n or T_n for much of the wave motion. Here the numerical analysis is carried further to compute the entire NLFS from arbitrary Cauchy data for general N. This requires numerical evaluation of $3N^2$ integrals on Γ and inversion of an $N \times N$ matrix. As discussed in Section 6 small $N \leq 10$ can provide useful, quantitative analysis for a wide class of solutions extending from the soliton limit, through the intermediate-spectrum range, to the linear-wave limit.

To carry out the ISPT analysis, a canonical basis α_n, β_n ($n = 0 \ldots N - 1$) of cycles on Γ must be chosen. Different choices for the basis give different values for the wavenumbers k_n, with corresponding changes to the components of the NLFS [15]. The wave field reconstructed via (3) is unchanged, so the choice of basis is to some extent a matter of convenience. The basis shown in Fig. 1 appears most useful for analyzing arbitrary initial conditions. With this basis, the wavenumbers form a harmonic series $k_n = 2\pi(n + 1)/X$. Also, using this basis each phase ϕ_n is associated dominantly with the gap (B_n, T_n). The wave field is represented as a nonlinear superposition of spatial harmonics with calculable amplitudes $2(T_n - B_n)$ and velocities $v_n = \omega_n/k_n$. The formulae given below are derived from the basis of cycles shown in Fig. 1.

For numerical evaluation, the cycles α_n, β_n may be deformed to lie along the real axis. The following three sets of real integrals will be required for $n, m = 0 \ldots N - 1$:

$$I_n^m = \int_{T_{n-1}}^{B_n} \frac{E^m\,dE}{R^{1/2}(E)}, \quad J_n^m = \int_{B_n}^{T_n} \frac{E^m\,dE}{iR^{1/2}(E)},$$
$$K_n^m = \int_{B_n}^{\mu_n} \frac{E^m\,dE}{iR^{1/2}(E)}. \tag{11}$$

For I_n^m with $n = 0$, the bottom limit of the integral is E_0. Referring to Fig. 1, it is specified that $R^{1/2}(E)$ in these integrals lies on the top sheet of Γ, above the branch cuts. This gives the (real) denominators $R^{1/2}(E)$ or $iR^{1/2}(E)$ in all of the integrands the same sign as -1^{N+n}. In terms of these integrals,

$$\oint_{\alpha_n} \frac{E^m\,dE}{R^{1/2}(E)} = 2iJ_n^m, \quad \oint_{\beta_n} \frac{E^m\,dE}{R^{1/2}(E)} = 2\sum_{p=0}^{n} I_p^m. \tag{12}$$

The ISPT makes use of normalized holomorphic differentials on Γ,

$$\Omega_n(E) = \sum_{m=0}^{N-1} C_{mn}\frac{E^m\,dE}{R^{1/2}(E)}, \quad n = 0 \ldots N-1,$$
$$\oint_{\alpha_m} \Omega_n(E) = 2\pi i\,\delta_{mn}. \tag{13}$$

Combining (12) and (13) the coefficients C_{nm} are computed as the matrix inverse

$$[C_{nm}] = \pi[J_p^q]^{-1}. \tag{14}$$

The wavevectors k_n and frequencies ω_n of the N-gap mode are then obtained as

$$k_n = 2C_{N-1,n}$$
$$\omega_n = -8C_{N-2,n} - 4C_{N-1,n}\left[E_0 + \sum_{m=0}^{N-1}(B_m + T_m)\right]. \tag{15}$$

The first term in the equation for ω_n is absent when $N = 1$. Equations (15) are derived, for example, in Ablowitz and Segur [2]; note, however, that a sign error is corrected here. The B-matrix is given by

$$B_{nm} = \oint_{\beta_m} \Omega_n = 2\sum_{p=0}^{N-1}\left[C_{pn}\sum_{q=0}^{m} I_q^p\right]. \tag{16}$$

Finally, the initial phases are found to be

$$\phi_{n0} = -i\sum_{m=0}^{N-1}\oint_{B_m}^{\mu_m}\Omega_n = \sum_{m,p=0}^{N-1} C_{pn}\sigma_m K_m^p. \tag{17}$$

In this equation $\{\mu_m\}$ is the auxiliary spectrum at $(x, t) = (0, 0)$ (i.e., the point where the phases ϕ_{m0} are to be computed), and $\sigma_m = \pm 1$ depending upon the sheet of Γ, where μ_m is located. With the conventions established here, σ_m is determined by the absolute value of the diagonal element a of $\mathbf{T}(x, t, \mu_m(x, t))$ as follows: $\sigma_m = +1$ when $|a| < 1$, $\sigma_m = -1$ when $|a| > 1$. The first equalities in (16) and (17) are fundamental results of ISPT theory [2], which are not rederived here. The second equalities follow as results of the definitions (11) above.

The one remaining component of the NLFS is the constant u_0, the determination of which is complicated in the absence of initial-condition data [2, 15]. Using the periodicity of the θ-function u_0 is found to be the average of the initial data, $u_0 = (\int_0^X u(x, 0)\,dx)/X$.

Equations (10)–(17) give an explicit prescription for the

nonlinear analysis of an arbitrary initial wave field $u(x, 0)$, yielding the NLFS. For predictive and filtering applications, practical representations are also needed for the reconstruction formulas (3, 4). The logarithmic derivative in (3) is expanded as $\partial^2 \ln \theta/\partial x^2 = (\partial^2\theta/\partial x^2)/\theta - (\partial\theta/\partial x)^2/\theta^2$. Then, the lattice sum (4) is expressed in terms of real functions and differentiated to yield

$$\theta = 1 + \sum_M{}' E_M \cos \Phi_M,$$

$$\frac{\partial\theta}{\partial x} = -\sum_M{}' E_M K_M \sin \Phi_M, \tag{18}$$

$$\frac{\partial^2\theta}{\partial x^2} = -\sum_M{}' E_M K_M^2 \cos \Phi_M,$$

where

$$E_M = 2 \exp\left(\frac{1}{2}\sum_{n,m=0}^{N-1} M_n B_{nm} M_m\right),$$

$$K_M = \sum_{n=0}^{N-1} M_n k_n, \quad \Phi_M = \sum_{n=0}^{N-1} M_n \phi_n. \tag{19}$$

The primed sums are restricted sums over N-vectors of integers $\mathbf{M} = [M_0 \ldots M_{N-1}]$: the zero-vector [0 ... 0] is omitted; and one member of each pair $(\mathbf{M}, -\mathbf{M})$ is omitted. One way to do this is to retain only those $\mathbf{M}$ for which the first non-zero component is positive.

4. NUMERICAL METHODS

In Section 3, the complete ISPT for the KdV equation was reduced to formulas suitable for numerical evaluation. This section details an implementation of this program using standard numerical methods [20], which works well for a variety of initial condition data.

Given initial data $u(x, 0)$, $x \in (0, X)$, the first step is to find the main spectrum $E_0, B_0 \ldots T_{N-1}$ and the auxiliary spectrum $\mu_0, \sigma_0 \ldots \mu_{N-1}, \sigma_{N-1}$ from $\mathbf{T}(x = 0, t = 0, E)$. Anticipating experimental applications the examples shown here were calculated using discretized initial data, specified at P evenly spaced points $x_n = nX/P$, $n = 0 \ldots P - 1$. The integrations of (10) were carried out using the fourth-order Runge–Kutta formula [20] with fixed step size. There is nothing about the ISPT that is particular to discrete initial data, and in general the method used to integrate (10) should be dictated by the form of the data (e.g., analytic or adaptive-step integration for data specified analytically).

Figure 2 shows the behavior of $\mathcal{F}(E) = (a + d)/2$. The main spectrum is found by first finding pairs of E values that bracket each of $E_0, B_0, T_0, B_1, \ldots$, and then using Brent's method [20] to search for the zero within each pair of $\mathcal{F}(E) - 1$ or $\mathcal{F}(E) + 1$. To search efficiently for the bracketing energies,

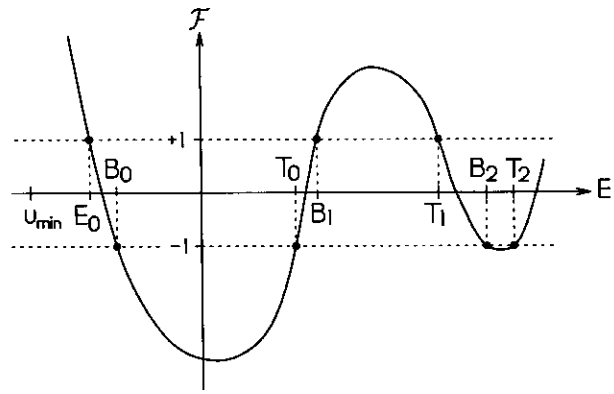

FIG. 2. Schematic dependence of the Floquet discriminant $\mathcal{F}(E)$ on energy. E_0 is the smallest E for which $\mathcal{F}(E) = +1$, and is always greater than the minimum value of the wave field u_{min}.

information about the spectrum of (6) familiar from band-structure theory is used: (a) $E_0 \geq u_{min}$ with the difference equal to the zero-point energy of the groundstate; and (b) for large n a "free-electron" approximation is valid, $B_n \approx T_n \approx u_0 + [(n + 1)\pi/X]^2$. Here u_{min} is the minimum value of $u(x, 0)$. The auxiliary spectrum is found by a Brent's-method search within each gap (B_n, T_n) for the zero of $c(E)$. This E is μ_n and $a(\mu_n)$ is evaluated to find σ_n.

To proceed with the analysis N must be chosen. The relative inaccuracy for any N is given directly by (8) as the fraction of the total gap width omitted. Numerical precision is found to limit the maximum N that may be used in one of two ways. For sufficiently large n, it becomes impossible to find E such that $|\mathcal{F}(E)| > 1$ (i.e., to find the nth gap), or within very narrow gaps the numerical estimate of $c(E)$ is not well behaved. As narrow gaps contribute little to the wave field, these limits on N do not impair the accuracy of the ISPT reconstruction.

The next step is to evaluate the $3N^2$ integrals (11). The integrals are all of the form

$$I = \int_a^{b \text{ or } c} \frac{f(E)\,dE}{+\sqrt{(E-a)(c-E)}}, \quad a < b < c, \tag{20}$$

where $f(E)$ is nonsingular in (a, c). The singularities at a and c are removed by a substitution

$$I = \int_{-1}^{u_b \text{ or } +1} \frac{3f(E(u))\,du}{+\sqrt{4-u^2}}, \quad E(u) = \frac{a+c}{2} + \frac{c-a}{4}u(3-u^2),$$

$$u_b = -\cos\frac{\phi}{3} + \sqrt{3}\sin\frac{\phi}{3}, \quad \phi = \cos^{-1}\left(\frac{a+c-2b}{c-a}\right), \tag{21}$$

$$0 < \phi < \pi.$$

Simple Simpson's-rule [20] evaluations of the integrals in this form suffice for a wide variety of initial conditions. When the

edges of adjacent gaps are close to the limits of integration the integrand is sharply peaked, and an adaptive-step integration [20] would be more efficient. This occurs in the extreme soliton limit, although it is found that accurate analysis for solitons that are not too widely separated may be done using only Simpson's rule.

The final step in the analysis of the initial data is the $N \times N$ matrix inversion (14). Then the NLFS components k_n, ω_n, ϕ_{n0} are directly given by (15)–(17). For sufficiently large N, it is found that $k_n = 2\pi(n + 1)/X$ to good accuracy. For smaller N, the wavenumbers do not meet this condition. The reconstructed wave $u(x, t)$ field with such small N remains a solution to the KdV equation, but it is no longer accurately periodic in x.

The reconstruction of the N-gap wave field $u(x, t)$ from the NLFS is discussed next. A criterion is needed to truncate the infinite lattice sums (18). As cancellation occurs both within these three sums and when they are combined to form $u(x, t)$, it is unclear which are the most important terms. Several criteria were tried, but it seems likely that improvement may be achieved with further work. The most successful criterion on **M** found is

$$\exp\left(\frac{1}{2}\sum_n B_{nm}M_n^2\right) > \varepsilon \exp(B_{00}/2), \tag{22}$$

where $\varepsilon \ll 1$ is an accuracy parameter. While this criterion ignores the off-diagonal elements of B_{nm}, it may be expressed in the form $\Sigma_n C_n M_n^2 < 1$, and it is easy to rapidly generate all **M** meeting such a condition. Another useful criterion is

$$\left(1 + \left[\sum_n M_n(n + 1)\right]^2\right) \exp\left(\frac{1}{2}\sum_{n,m} M_n B_{nm} M_m\right) > \varepsilon \exp(B_{00}/2). \tag{23}$$

It requires more computation to find the set of **M** vectors that satisfy (23), but a smaller set of vectors chosen this way gives $u(x, t)$ to the same accuracy as a larger set chosen according to (22). Choosing smaller ε in (22) or (23) causes more **M** vectors to be used in the sums, with correspondingly greater accuracy and slower execution.

5. A SAMPLE APPLICATION

As an illustration and first application of the method, it was applied to the KdV equation with randomized initial conditions. The goal was to find nonlinear structures that might occur in an experimental or natural situation in which the wave field is driven by a random or noisy process. Of course, many types of random initial conditions are possible, and even a qualitative exploration of all of the possible resulting nonlinear modes is beyond the scope of the present work. The initial conditions chosen were a random superposition of harmonics with a power-law frequency dependence,

$$u(x, 0) = U_0 \sum_{n=1}^{n_{\max}} n^{-\alpha} u_n \cos(2\pi n x/X + b_n) \tag{24}$$

with the amplitudes u_n chosen randomly in (0, 1) and the phases b_n chosen randomly in $(0, 2\pi)$. The examples shown here have $X = 40$, $U_0 = 0.2$, $\alpha = 1$, and $n_{\max} = 7$. The significant parameters are U_0X^2, which is an invariant measure of the nonlinearity, and α, which qualitatively affects the modes generated. Increasing $n_{\max}$ adds high-frequency noise to visualizations but has little effect on the quantitative analysis. $U_0X^2 \approx 300$ was chosen so as to have strong nonlinearity as well as a significant influence of the periodicity on the modes (this influence vanishes as the period $X \to \infty$). This is therefore a different regime than that explored for natural water waves in Ref. [9]. In that work, the period used for the ISPT analysis was not a feature of the physical system, but rather it was a data window similar to that commonly used for linear spectral estimation.

To isolate the effect of nonlinearity, the time evolution generated by the KdV equation was compared with that generated by (1) with the nonlinear term omitted (equivalently, infinitesimal $u(x, t)$ with the full KdV). Figures 3a and b show linear and nonlinear evolutions from a particular set of initial conditions from the ensemble (24). The differences are striking: the linear evolution shows no structure beyond the imposed periodicity in x, while the nonlinear evolution displays a double set of quasiperiodic pulsations with a t-period of about 25.

The tendency of integrable nonlinear PDEs to generate organized structures (even when the soliton description is inaccurate) is well known, and Fig. 3b could have been produced by direct integration of the KdV equation rather than the ISPT. The advantage of the present method is the possibility of understanding and quantifying the nonlinear mode that gives rise to the visible structure and, if desired, to isolate it from other modes present. The main spectrum for these initial conditions, Fig. 4a, shows that the motion is dominated by the three gaps (B_1, T_1), (B_3, T_3), and (B_0, T_0). By associating the gaps with their harmonics and velocities it is possible to plot the locus of points in (x, t), where each phase equals its value at (0, 0), Fig. 4b. This plot immediately explains the pulsations as an interference effect between the two largest gaps, (B_1, T_1) and (B_3, T_3). The dominant gap structure $E_0 \ldots T_3$ is neither solitonic nor nearly linear. Referring to Fig. 4b, the t-period of the pulsations is precisely computed as $T_{\text{pls}} = X/4(\nu_1 - \nu_3) = 23.36$. It is also clear that the unlike appearance of the two pulses in an x-period is due to modulation by the only other large gap, (B_0, T_0). Therefore one predicts a slow alternation between the shapes of the two pulses per x-period, with the very long period $T_{\text{alt}} = X/2(\nu_0 - \nu_1) = 158.6$. It should be emphasized that these characteristics which are evident in visualizations such as Fig. 4b are quantitatively explained by the ISPT.

As the bands separating the first two gaps are very narrow, Fig. 4, it is possible to qualitatively describe the two ridges of

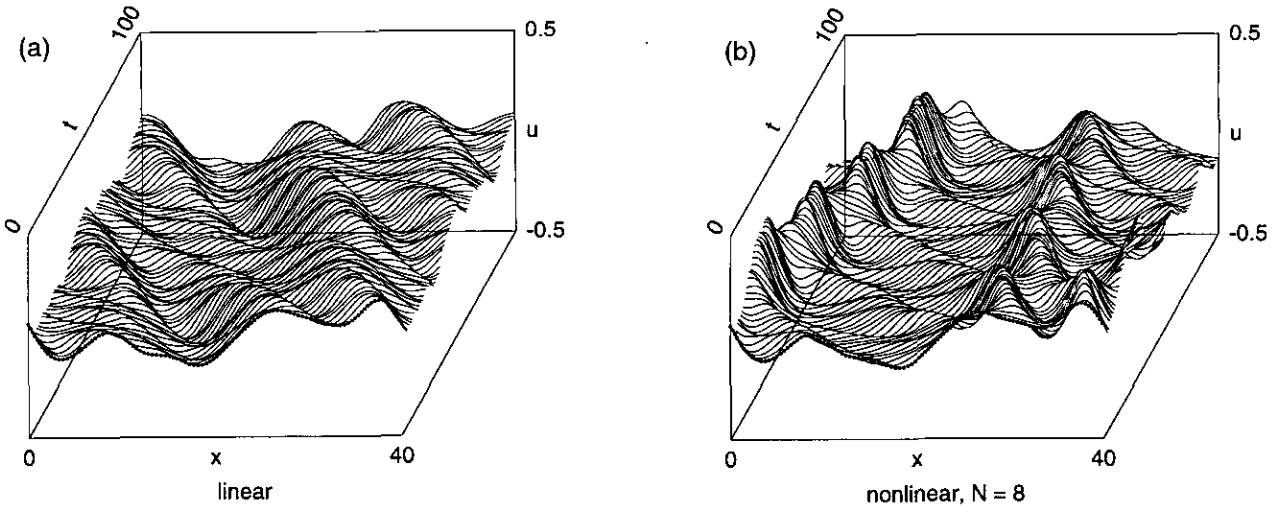

FIG. 3. (a) Linear evolution of the wave field $u(x, t)$ from randomly chosen initial conditions $u(x, 0)$. The initial conditions are shown by the dotted curve at $t = 0$. The spatial period is 40 and one complete period is shown. (b) Nonlinear evolution computed using the ISPT algorithm for the KdV equation, for the same initial conditions as (a). Note the pulsation of the nonlinear evolution with a t-period of approximately 25.

pulsations in Fig. 3b as a pair of solitons [16]. However, the IST prediction for these solitons in isolation is $T_{alt} = 243.7$ (the time between successive collisions of the solitons). The presence of the large gap (B_3, T_3) has drastically modified the properties of the solitons, making the ISPT necessary for quantitative analysis.

Another possibility is to use the ISPT for nonlinear filtering [9]. As for linear filtering, in which undesired Fourier components are omitted, N is reduced to omit gaps superfluous to the main nonlinear mode present. The wave field $u(x, t)$ reconstructed using (3) is then a filtered version of the original wave field. Figure 5 shows the result of nonlinear filtering for the same initial conditions as Fig. 3. The dominant pulsations are now more clearly visualized, but their characteristics (periods, velocities, and amplitudes) are almost unchanged. Excessive reduction of N (below $N = 4$ for this example) wrecks the periodicity of the solution ($k_n \neq 2\pi(n + 1)/X$). Even in this case, it might be useful to reconstruct $u(x, t)$ using (3) for small N, while using the large-N values for k_n, ω_n. This would permit

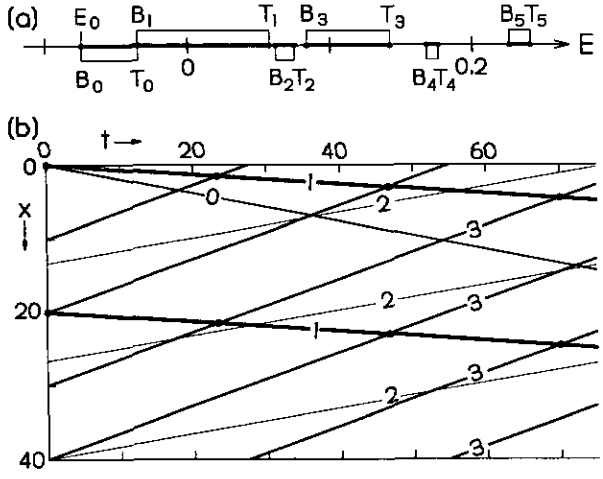

FIG. 4. (a) The main spectrum (band edge energies) computed for the initial conditions of Fig. 1. The first six gaps are shown. The first two bands (E_0, B_0) and (T_0, B_1) are too narrow to see on this figure. (b) Space-time diagram showing the locus of points where each phase ϕ_n equals its value at $(x, t) = (0, 0)$. These lines are labeled with n and the size of the corresponding gap is roughly indicated by the line width. Points of intersection between $n = 3$ and $n = 1$, marked with dots, correspond to the pulsations visible in Fig. 1b.

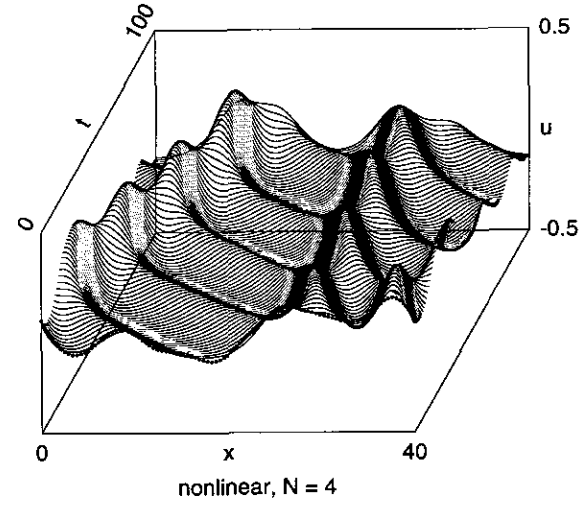

FIG. 5. Result of nonlinear filtering applied to the initial conditions of Fig. 1. The number N of gaps used in the analysis has been reduced to the minimum number necessary to describe the dominant nonlinear mode, four in this case.

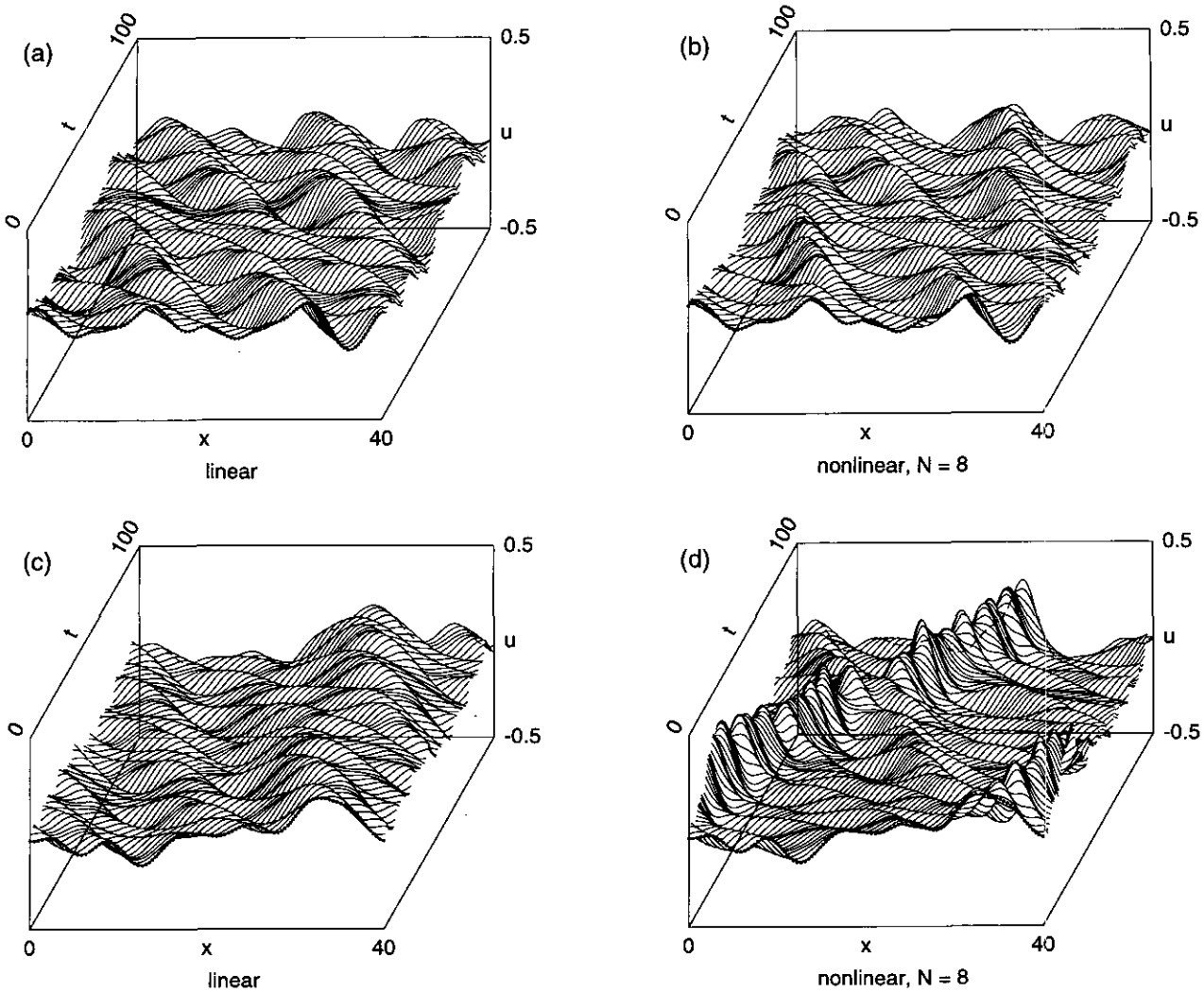

FIG. 6. Linear and nonlinear evolutions from initial conditions drawn from the same random ensemble as for Fig. 3. The random-number seed was s = 2 for (a) and (b), and s = 3 for (c) and (d) (for Fig. 3, s = 1).

the visualization of the part of the wave motion due to the first few gaps.

The ISPT analysis was repeated for several sets of initial conditions drawn from the same random ensemble as for Fig. 3. It is interesting to note which features of the nonlinear evolution are specific to particular realizations of (24), and which are universal. Certainly the pattern of double pulsations, which results from the specific gap structure of Fig. 4a, is not reproduced. However, the majority of initial conditions tried show an organized pulsation, wave or ridge structure, which is never observed in linear evolutions like Fig. 3a. Figure 6 shows linear and nonlinear evolutions as in Fig. 3, for two further sets of initial conditions from the same ensemble. For the first set, Figs. 6a,b, there is little qualitative difference between the linear and nonlinear evolutions. The second set, Figs. 6c,d is more like Fig. 3 in that the nonlinear evolution shows a highly organized structure (a modulated ridge) absent in the linear evolution. As for Fig. 3 this ridge could be interpreted as a soliton, but its properties are greatly altered by the other nonlinear excitations present. The initial conditions for Figs. 3a,b, 6a,b, and 6c,d were generated by choosing successive values for the random-number seed without further selection.

The three examples shown do not exhaust the possibilities, even for the chosen nonlinearity parameter U_0X^2 and exponent α, but they give a good idea of the range of structures encountered. Cataloging nonlinear structures is one possible avenue for future investigation. Another is to search for coherent structures in periodically forced experimental systems, and to investigate the degree to which the ISPT can explain them. The N-gap modes found here are extraordinarily robust against perturbations (compare Figs. 3b and 5). It remains to be seen whether these modes are also robust in the real world. There is already experimental evidence that a 2-gap solution to another inte-

grable equation, the Kadomtsev–Petviashvili equation, persists in water waves well beyond its expected domain of validity [8, 18]. This suggests it is worthwhile to search in physical systems for N-gap modes, using the ISPT as a guide.

6. DOMAIN OF VALIDITY, COMPUTATIONAL EFFICIENCY, AND EXPENSE

This concluding section addresses several issues that determine the feasibility of applying the ISPT in the form described here. One important conclusion from many numerical trials is that the method is useful over a wide range of nonlinearity, from nearly linear to nearly solitonic. It is found that linear evolutions such as Fig. 3a are readily duplicated by using extremely small amplitudes. In this linear regime, the limitation on the method at present is the crudeness of the criteria use to truncate the lattice sums (18). For very large N, which is desirable to model a solution with many Fourier harmonics, the criterion (22) is too unselective and the criterion (23) requires too much computation to implement.

In the other limit, of well separated solitons, neither N nor the number of terms in the lattice sums is excessive (N need only be as large as the number of solitons). A different limitation is encountered when the solitons are very widely separated. In this case the eigenfunctions of the Schrödinger equation (6) vary exponentially over much of the x-domain, and numerical integration of (10) fails. However, it is found that the method is accurate for solitons sufficiently separated to give the IST (infinite domain) velocities and collisional phase shifts to high precision.

It is interesting to compare the amount of computation required for the ISPT with that required for straightforward integration of the nonlinear PDF. The structures of the two calculations are dissimilar: the ISPT requires a large computation to analyze the initial conditions, and then can rapidly predict the solution at any (x, t), while the integration requires computation proportional to the t domain to be explored. Also, the ISPT produces exact solutions to the KdV equation that only approximately meet the specified initial and boundary conditions, while the integration produces a function that exactly meets the initial and boundary conditions but only approximately obeys the KdV equation.

The comparison was carried out for two different initial conditions on the interval $(0, X)$ with $X = 40$: (a) the random initial conditions of Fig. 3, and (b) two solitons with heights 0.25 and 0.5, which interact. The results are listed in Table I. For both the direct integration and the ISPT, the parameters of the calculation were varied so as to minimize the computer time while keeping the maximum error $|\Delta u| < 0.01$. These parameters are the space and time step sizes δx, δt for direct integration, and P, N, and ε for the ISPT (the criterion (22) was used). For direct integration, the simple explicit method of Zabusky and Kruskal [21] was used. As investigated in detail by Taha and Ablowitz [22], this is not generally the fastest way to integrate the KdV equation. The most effective method found by these authors required roughly 10 times as much computation per mesh point, but it was stable and accurate using a larger time step δt (the tabulated results of Ref. [22] suggest a maximum accurate $\delta t \approx \delta x$ for these conditions). It is estimated that the advantage of using this more sophisticated integration method is a only factor of two to seven for the conditions of Table I. The reason the simple explicit method is so successful is the relatively small amplitude of the wave field in these examples. This does not imply that the effects of nonlinearity are small (Fig. 3).

TABLE I

Comparison of Computer Time Required for Direct Integration of the KdV Equation with That Required for the ISPT Analysis with the Same Initial Conditions

Initial conditions	Method	Computer time
Random (Fig. 3)	*Direct integration* from $t = 0$ to $t = 100$, $\delta x = 0.2$, $\delta t = 0.003$	53.0 s
	ISPT with $N = 8$, $\varepsilon = 10^{-6}$	
	Finding main and auxiliary spectra ($P = 100$)	3.1 s
	Finding NLFS and series coefficients,	2.5 s
	Evaluating $u(x, t)$ at 100 points	2.7 s
Colliding solitons	*Direct integration* from $t = 0$ to $t = 100$, $\delta x = 0.4$, $\delta t = 0.024$	3.3 s
	ISPT with $N = 2$, $\varepsilon = 10^{-6}$	
	Finding main and auxiliary spectra ($P = 100$)	1.4 s
	Finding NLFS and series coefficients,	1.7 s
	Evaluating $u(x, t)$ at 100 points	0.09 s

All calculations were carried out on a system with a 33 MHz Intel 80486DX microprocessor, and all real-number calculations used double (15 decimal digit) precision. See text for details.

Table I shows that the ISPT analysis is sometimes competitive with direct integration of the PDE as a method of projecting arbitrary data ahead in time, depending mainly on the number of (x, t) points at which predictions are required. Development of the ISPT as a numerical method is at an early stage, with further optimization likely. A major reason for using the ISPT will remain its ability to provide classification and understanding for the organized structures generated by integrable PDEs. From this point of view the significant result from Table I is that an inexpensive computer requires only a few seconds to complete the ISPT analysis for data sets similar to those generated by experiments.

Note added in proof. Since this work was completed, an experimental study of shallow water waves has been reported that makes use of the ISPT for data analysis [23]. The numerical

methods used in [23] involve the integration of the ordinary differential equations obeyed by the auxiliary spectrum $\{\mu_i\}$, rather than the θ-function approach developed here.

ACKNOWLEDGMENTS

The author thanks R. A. Guyer and J. A. Krumhansl for helpful discussions. This work was supported by NSF Grant DMR 9100044.

REFERENCES

1. C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, *Phys. Rev. Lett.* **19,** 1095 (1967); M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
2. B. A. Dubrovin and S. P. Novikov, *Zh. Eksp. Teor, Fiz.* **67,** 2131 (1974); *Sov. Phys. JETP* **40,** 1058 (1975); B. A. Dubrovin, V. B. Matveev, and S. P. Novikov, *Uspekhi Mat. Nauk* **31**(1), 55 (1976); *Russian Math. Surveys* **31**(1), 59 (1976). M. J. Ablowitz and H. Segur, Ref. [1, Section 2.3].
3. A. R. Its and V. B. Matveev, *Funkt. Anal. Ego Prilozhen.* **9,** 69 (1975); *Funct. Anal. Appl.* **9,** 65 (1975).
4. J. Garnett and E. Trubowitz, *Comment. Math. Helv.* **62,** 18 (1987).
5. E. R. Tracy, H. H. Chen, and Y. C. Lee, *Phys. Rev. Lett.* **53,** 218 (1984) and references therein.
6. L. P. Lévy, *Phys. Rev. B* **31,** 7077 (1985).
7. A. R. Osborne and L. Bergamasco, *Nuovo Cimento B* **85,** 229 (1985); *Physica D* **18,** 26 (1986).
8. J. Hammack, N. Scheffner, and H. Segur, *J. Fluid Mech.* **209,** 567 (1989).
9. A. R. Osborne, E. Segre, G. Boffetta, and L. Cavaleri, *Phys. Rev. Lett.* **67,** 592 (1991).
10. E. A. Overman II, D. W. McLaughlin, and A. R. Bishop, *Physica D* **19,** 1 (1986).
11. N. Ercolani, M. G. Forest, and D. W. McLaughlin, *Physica D* **43,** 349 (1990).
12. M. G. Forest, C. G. Goedde, and A. Sinha, *Phys. Rev. Lett.* **68,** 2722 (1992).
13. M. G. Forest and D. W. McLaughlin, *J. Math. Phys.* **23,** 1248 (1982).
14. R. Flesch, M. G. Forest, and A. Sinha, *Physica D* **48,** 169 (1991).
15. A. I. Bobenko and D. A. Kubenskii, *Teor. i Mat. Fiz.* **72,** 352 (1987); *Theor. Math. Phys.* **72,** 929 (1988).
16. J. P. Boyd, *J. Math. Phys.* **25,** 3390 (1984).
17. B. A. Dubrovin, *Uspekhi Mat. Nauk* **36**(2), 11 (1981); *Russian Math. Surveys* **36**(2), 11 (1981).
18. H. Segur, *Physica D* **51,** 343 (1991).
19. A. R. Osborne and E. Segre, *Physica D* **44,** 575 (1990).
20. Matrix inversion, integration, and root-finding functions were taken from W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, 1988).
21. N. J. Zabusky and M. D. Kruskal, *Phys. Rev. Lett.* **15,** 240 (1965).
22. T. R. Taha and M. J. Ablowitz, *J. Comput. Phys.* **33,** 231 (1984).
23. A. R. Osborne and M. Petti, *Phys. Fluids* **6,** 1727 (1994).