Acta Applicandae Mathematicae **39:** 5–37, 1995.

Integrability, Computation and Applications

M. J. ABLOWITZ, S. CHAKRAVARTY
Program in Applied Mathematics, University of Colorado, Boulder, CO 80309, U.S.A.

and

B. M. HERBST
Department of Applied Mathematics, The University of the Orange Free State, Bloemfontein 9300, South Africa

(Received: 13 January 1995)

Abstract. The study of integrable systems and the notion of integrability has been re-energized with the discovery that infinite-dimensional systems such as the Korteweg–de Vries equation are integrable. In this paper, the following novel aspects of integrability are described: (i) solutions of Darboux, Brioschi, Halphen-type systems and their relationships to monodromy problems and automorphic functions, (ii) computational chaos in integrable systems, (iii) we explain why we believe that homoclinic structures and homoclinic chaos associated with nonlinear integrable wave problems, will be observed in appropriate laboratory experiments.

Mathematics Subject Classification (1991): 58F07.

Key words: integrability, KdV equation, nonlinear integrable wave functions, monodromy problems, automorphic functions.

1. Introduction

The study of 'integrable' or exactly solvable systems has a long history. Classical mathematicians such as Euler, Lagrange, Liouville, Riemann, Poincaré, Painlevé amongst many others, investigated nonlinear systems which could either be integrated more or less explicitly or possessed special analytic structures in the complex plane. Perhaps surprisingly, there is still no single adequate definition of 'integrability'. Certainly explicit integration of nonlinear systems in the real domain should be considered as integrable, as should cases where suitable transformations exist to allow an elementary solution, such as occurs in the Hamiltonian case with action-angle variables (often called integrability in the Liouville sense). Less clear is the still developing notion of integrability in the complex plane. For example, if the general solution of an ODE has appropriate analytic properties, such as (i) having poles as the only movable singular points (a movable singular point is one whose location depends on the initial conditions; a fixed singular point is fixed by the coefficients in the equation) or more generally (ii) being everywhere single-valued in its domain of existence, then we consider the equation to be integrable in the complex plane. Many examples of nonlinear

ODEs have shown that when (i) or (ii) are satisfied then the equation falls into a class in which the general solution can be either obtained by explicit integration or can be linearized via an associated Riemann–Hilbert (RH) factorization problem.

The investigation of integrable systems has been an interesting and active field in recent years due to the fact that numerous physically interesting infinite-dimensional systems have been linearized via the method of the Inverse Scattering Transform (IST) and large classes of explicit solutions, such as soliton solutions, have been obtained. The best known example is the Korteweg–de Vries (KdV) equation,

$$u_t + 6uu_x + u_{xxx} = 0, \tag{1}$$

which these proceedings have been named after. The KdV equation arises in many physical problems. It is the canonical equation describing weakly dispersive and weakly nonlinear wave phenomena arising in shallow water waves (this was the original application of Korteweg and de Vries), internal waves in fluids, plasma waves, lattice dynamics, etc.

The IST method yields the general solution to the Cauchy problem on the infinite line corresponding to rapidly decaying initial conditions (cf. [1–3]), a special case of this is the general N-soliton solution. The IST method uses direct and inverse scattering analysis to obtain a matrix RH factorization problem which leads to a Gel'fand–Levitan–Marchenko integral equation as the linearization of KdV. The KdV equation has been investigated in a variety of contexts and we note that there is one other case in which the general initial-value solution is known. This is the periodic boundary-value problem where the solution of the KdV equation can be expressed in terms of Riemann theta functions of arbitrary genus (cf. [1, 2]).

The IST is an effective tool to obtain linearizations and solutions to many other well known nonlinear wave equations in $(1 + 1)$ dimensions, e.g., the nonlinear Schrödinger equation (NLS), sine-Gordon equation, three wave interaction $(1 + 1)$ equations, etc. Moreover, it is significant that there are many $(2 + 1)$-dimensional equations which have solutions via the IST method. The best known are the Kadomtsev–Petviashvili (KP), Davey–Stewartson (DS) and three wave interaction $(2 + 1)$ equations. We note that the IST method needs to be generalized appropriately in order to handle $(2 + 1)$ problems; it was found that nonlocal RH problems and DBAR problems played essential roles (cf. [3]).

In this paper, we shall discuss two issues related to integrability. In the first part, we shall focus on specific differential equations which have novel properties. These equations are obtained from a well known integrable system in four dimensions, the so-called self-dual Yang–Mills (SDYM) system. The other topic we shall discuss involves the computation of integrable systems in the vicinity of what are called homoclinic manifolds. In these regions, the solutions of the non-

linear equation are very sensitive. This allows numerical approximations to the equations to induce a chaotic response. In fact we have observed computational chaos due to (i) truncation effects and (ii) very small errors even on the order of roundoff error. Our investigations and analysis also strongly suggest that analogous experimental investigations (e.g., in water waves and/or nonlinear optics) will be possible and that homoclinic manifolds and related chaotic phenomena will be observable experimentally.

In higher dimensions the SDYM system is one of the few (perhaps the only(?)) known physically important four-dimensional system to which the IST method applies. Significantly, the SDYM system turns out to encompass many of the integrable systems we already know. The self-dual Yang–Mills equations (SDYM) arise in particle physics and relativity (see [3] for many references on SDYM).

Specification of a self-dual system requires the specification of (i) the Lie algebra in which the field variables reside and (ii) the specification of which independent variables will be chosen to keep (or equivalently which variables to ignore). Reductions of the SDYM equations to the well-known soliton equations in $(1 + 1)$ dimensions such as KdV, NLS, sine-Gordon, N-wave and respective hierarchies, etc., all follow by considering two independent variables with the field variables lying in a finite-dimensional Lie algebra (cf. [3]). However, the well-known soliton equations in $(2 + 1)$ dimensions, such as KP, DS, N-wave and respective hierarchies follow from SDYM by assuming two independent variables with the field variables lying in a suitable infinite-dimensional algebra [4].

Important and interesting $(0 + 1)$-dimensional equations (ODEs) are also obtained as reductions of SDYM. By reducing to one independent variable with the field variables lying in a finite-dimensional Lie algebra, it can be shown that the free spinning Euler top, the 'heavy' Kovalevskya top [5], and the six irreducible Painlevé transcendents [6] are special cases. Moreover, a novel system which we call the Darboux–Brioschi–Halphen (DBH) top [5, 7], and is related to self-dual Bianchi IX cosmologies, is also obtained from the SDYM equations with one independent variable, but in this case we take the field variables to lie in an infinite-dimensional algebra of volume preserving diffeomorphisms. The DBH system as the interesting analytic property that it is single-valued everywhere in its domain of existence, but there is a natural boundary beyond which the solution cannot be analytically continued.

The solution of the DBH top has been discussed earlier and there are important connections with modular functions (cf. [7]). However, we have only recently obtained the associated linear pair (by reduction of the associated SDYM linear pair) and now have obtained the solution by investigating the relevant monodromy problem [8]. In this paper, we shall show that the linear problem leads to a monodromy problem whose monodromy *is not* fixed ('nonisomonodromy') in time. Nevertheless, we can 'factorize' the problem into isomonodromy and

nonisomonodromy parts in such a way so as to be able to solve the DBH system.

We have also been able to solve interesting generalizations of the DBH system, of fifth order (we call this the DBH-5 system), which relate to nondiagonal self-dual Bianchi IX cosmologies. Once again there is an associated linear pair which yields solutions. However, it turns out that in this case the solution is not always single-valued, and in fact depending on the initial values, sometimes the solution is densely branched. In the latter case, the equation is not of Painlevé type despite the fact that there is an associated linear pair and that the DBH-5 system is obtained by reduction (with an infinite-dimensional gauge algebra) from the SDYM equations.

In our computational studies, we have been carrying out moderate to long-time numerical calculations associated with integrable nonlinear wave equations with periodic boundary conditions. It is well known that many of these equations have solutions which are linearly unstable and have interesting associated mathematical and physical phenomena. These unstable solutions are in a sense close to homoclinic manifolds. This fact has serious consequences which are reflected in numerical calculations (and we believe physical experments). Prototypical examples are the focusing nonlinear Schrödinger, sine-Gordon and modified KdV equations. (We note in passing that the (real-valued) KdV equation, the defocusing NLS and the sinh-Gordon equations are not amongst this class of equations.) We have observed that corresponding to certain initial/boundary conditions, these equations admit a special class of solutions which we refer to as homoclinic orbits. Alternatively, one can view these solutions geometrically as being part of an infinite-dimensional phase space consisting of an infinite-dimensional torus which has associated homoclinic manifolds ('pinched tori' [10]). In analogy with low-dimensional dynamics, due to perturbations, crossing of these homoclinic regions can drive a chaotic flow. In the numerical problem, the perturbations are due to either truncation or roundoff effects.

We have shown [11, 12, 14] that in the *true* nonlinear wave equation, small perturbations around a particular solution, say of order epsilon, can translate into close proximity to the homoclinic manifolds with a 'distance' of *any* order of epsilon and possibly even BEYOND all orders in epsilon. The choice of these initial conditions is straightforward, depending only on (a) the number of unstable modes that are contained in the initial data, and (b) the frequency of the initial perturbation. In order to demonstrate this situation mathematically, we have found that it is useful to employ the underlying IST formalism. In particular, we have found that homoclinic orbits correspond to suitable double points in the spectra of the associated scattering problem. The 'distance' from these orbits (manifolds) correspond to the order of epsilon in the asymptotic expansion around the double points. We have also found that exhibiting the 'sides' of the homoclinic orbits is straightforward, depending on the choice of phase in the initial conditions.

With this information in hand, we can understand why numerical experiments exhibit chaotic dynamics either due to truncation or roundoff effects. Truncation effects become important when the distance to the homoclinic manifolds are less than the order of truncation. On the other hand, roundoff effects are critical when the distance to the homoclinic manifold is less than the order of roundoff. We have observed these effects in many numerical experiments.

As mentioned earlier, we believe that observations of homoclinic orbits and associated chaotic dynamics should be possible in laboratory experiments. For example, the NLS model is well known to describe weakly nonlinear slow modulations of a wave train in water waves or nonlinear optics. In the physical problem, homoclinic orbits should be 'seen' by observing *different* wave structures on either 'side' of the orbit; since the homoclinic orbit is an unstable wave mode, one does *not* observe it directly. The different 'sides' depend on the phase of the perturbation of the initial conditions, or alternatively, how the double point in the eigenvalue spectrum, splits in the complex plane. We shall discuss this more concretely later in this paper.

On the other hand, we know that the distance to the homoclinic manifold depends on the number of unstable modes in the initial conditions and the frequency of the perturbation to the true solution. With this information we can create data which is arbitrarily close to the homoclinic manifold. Physically this means that the magnitude of the higher-order terms in the perturbation expansion, from which we obtained the NLS equation at leading order, would be large enough so that these (small) terms are, in fact, larger than the distance to the homoclinic orbit. The perturbations should therefore be capable of driving a chaotic response via crossing of homoclinic manifolds. Naturally, the precise conditions for which chaos will be observed will depend on laboratory considerations; e.g., the size of the amplitude of the true periodic wave, the number of unstable modes that can be generated; and other small perturbations which we have omitted from our discussions, such as dissipation. Nevertheless, we believe that the ability to observe homoclinic structures in wave problems is fundamental, and important. It is well understood theoretically and consequently we make the prediction that in the experiments such as we have sketched here they should be observed.

2. Integrability, PDEs and ODEs

2.1. Reduction of self-dual Yang–Mills equations

In this section we give a brief overview of the SDYM equations in a complex Euclidean space $\mathbb{C}^4$, and outline the method of symmetry reductions to integrable systems in lower dimensions. We only outline the reduction procedure in general terms. By necessity, we leave out most of the technical details which can be found in the references cited in the text.

2.1.1. *SDYM in* $\mathbb{C}^4$

We introduce the complex coordinates, $y, \bar{y}, z, \bar{z}$ in $\mathbb{C}^4$ and the differential operators D_1, A_1, D_2 and A_2 defined by

$$D_1 \equiv \partial_y + \lambda\partial_{\bar{z}}, \qquad D_2 \equiv \partial_z - \lambda\partial_{\bar{y}},$$

$$A_1 \equiv A_y + \lambda A_{\bar{z}}, \qquad A_2 \equiv A_z - \lambda A_{\bar{y}}, \tag{2}$$

where $\partial_\alpha \equiv \partial/\partial_\alpha$, $\alpha \in \{y, \bar{y}, z, \bar{z}\}$, $\lambda \in \mathbb{C}P1$ is a complex parameter. $A_\alpha = A_\alpha(x) \in g$, $x \in \mathbb{C}^4$, where g is the Lie algebra corresponding to some fixed gauge group G.

The A_α's are called the Yang–Mills gauge potentials and are defined up to the gauge freedom,

$$A_\alpha \longrightarrow A'_\alpha = fA_\alpha f^{-1} + \partial_\alpha f f^{-1}$$

for $f \in G$. Both A_α and A'_α satisfy the SDYM equations which can be expressed as

$$[D_1 - A_1, D_2 - A_2] = 0, \tag{3}$$

which are the compatibility conditions for the linear system

$$(D_1 - A_1)\psi = (D_2 - A_2)\psi, \tag{4}$$

where $\psi(x, \lambda) \in G$. Equation (3) is quadratic in λ, so by setting each of its coefficients to zero, yields a system of three nonlinear PDEs in $\mathbb{C}^4$ given by

$$F_{yz} = F_{\bar{y}z} = F_{y\bar{y}} + F_{z\bar{z}} = 0 \tag{5}$$

Here $F \equiv \partial_\alpha A_\beta - \partial_\beta A_\alpha - [A_\alpha, A_\beta]$ and $[\,,\,]$ denotes the Lie bracket associated with the Lie algebra g. The linear system (3) is the Lax pair for the SDYM equations given by (5). The above formulation of the SDYM equations has a natural geometric structure: Equation (2) is the zero curvature condition for a connection (with components A_1, A_2) associated with holomorphic sections $\psi(x, \lambda) \in G$ of the Yang–Mills vector bundle. The bundle is restricted to a family of two-planes, called the anti-self-dual (ASD) planes, passing through each $x \in \mathbb{C}^4$ and parametrized by $\lambda \in \mathbb{C}P1$; the vector fields D_1, D_2 are tangent to each ASD plane. The connection is defined by the linear system (4) by requiring that the sections $\psi(x, \lambda)$ be covariantly constant on each ASD plane. The zero curvature condition (2) (equivalently, (5)) guarantees that (4) is integrable on the ASD surfaces and that it admits solutions defined globally over each ASD plane. The space of such solutions considered over the family of two-planes determine a holomorphic vector bundle. Therefore, this geometric construction establishes a one-to-one correspondence between holomorphic vector bundles and solutions of the SDYM equations in $\mathbb{C}^4$. This geometric formulation was given by Ward [15].

Conversely, given such a vector bundle, it is possible to obtain the solutions to (3) that are holomorphic in appropriate neighborhoods of λ (e.g., $\lambda = 0$ and $\lambda = \infty$) via a Riemann–Hilbert factorization problem parametrized by $x \in \mathbb{C}^4$.

2.1.2. *The Reduction Scheme*

The main relationship between the SDYM equations and integrable systems arises from the fact that an enormous number of integrable equations can be obtained from the SDYM equations by imposing appropriate symmetry conditions. This is a very large program of study and a systematic analysis is still lacking. Nevertheless, significant work during the past 15 years or so indicates that most of the well-known equations solvable by IST are embedded in the SDYM equations. The essential steps of the reduction procedure are outlined below:

(i) Coordinate symmetries are imposed such that the SDYM equations are invariant under a subgroup of the conformal group (the full symmetry group of the SDYM equations). In a suitable gauge, the SDYM connection A_α is invariant on the orbits of the symmetry group. Introducing a new set of coordinates $\{x^{K'}, y^K\}$ on $\mathbb{C}^4$, it is then possible to reduce the SDYM equations to a lower-dimensional system of PDEs which are expressed only in terms of the y^K's. Here $x^{K'}, K' = 1, 2, \ldots, 4 - m$, are the coordinates on each orbit, while $y^K, K = 1, \ldots, m$, are the coordinates that label the orbits. For example, when the symmetry group consists of $4 - m$ translations, the SDYM potentials are simply given by $A_\alpha(x) = A_\alpha(y^K)$. Then in (5) all the derivatives involving the ignorable coordinates vanish so that the resulting SDYM equations are reduced to a system of PDEs in the independent variables y^K, $K = 1, \ldots, m$.

(ii) A choice of the gauge group G is necessary such that the gauge potential A_α lie in the corresponding Lie algebra g. For a finite-dimensional gauge group, the elements of g (and hence the A_α's) can be represented by finite rank matrices. Most $(1 + 1)$ integrable equations, e.g., the KdV, NLS and the N-wave equations and many ODEs including the Euler–Arnold and Kovalevskya 'tops', arise from such a choice of gauge group. On the other hand, interesting reductions are also obtained when the A_α's belong to specific representations of *infinite*-dimensional Lie algebras. Examples include the self-dual Einstein equation, the KP, DS and the $(2 + 1)$ N-wave equations. A special ODE reduction of the self-dual Einstein equation, called the DBH system will be discussed in the next section.

(iii) The SDYM equations reduced via steps (i) and (ii) can be further simplified by the residual gauge freedom and by imposing certain autonomy conditions on the dependent variables. After imposing the coordinate symmetries in step (i), the remaining freedom in the gauge transformation, $A_\alpha \to A'_\alpha = fA_\alpha f^{-1} + \partial_\alpha f f^{-1}$, $f(x) \in G$, is essentially restricted to $f(x) = f(y^K)$. Nevertheless, by making judicious choices for f (that takes some of the A_α's

into irreducible ('normal') forms), certain components of the reduced SDYM system can often be trivially solved. The autonomy condition is applied to certain arbitrary functions arising as 'constants' of integration in the reduction process. It requires that if a function is constant with respect to at least one of the independent variables, then it should be set to a constant. Thus this condition guarantees that the symmetry reduced SDYM equations will be an autonomous system.

The following additional remarks are also relevant to the reduction scheme described above.

Remarks. 1. Many equations which are symmetry reductions follow directly from the reduced system of SDYM equations obtained after applying steps (i)–(iii) and most importantly, without any further assumptions. It is a remarkable feature that a large class of integrable equations are naturally embedded in the SDYM equations in four-dimensions. However, sometimes it is necessary to make specific algebraic choices for the SDYM potentials in order to identify the reduced equations with a particular integrable system. The KdV equation is an example of a natural embedding, whereas the N-wave system is derived from the SDYM equations with additional assumptions on the A_α's (i.e. certain components are closen to be skew-symmetric matrices).

2. Another ingredient is to impose an appropriate reality condition on the reduced system in order to derive the soliton-type integrable equations which are usually defined on the real domain. The equations derived from the reduction of SDYM in $\mathbb{C}^4$ are complexified versions of these soliton-type equations. However, it is indeed possible to choose real slices of $\mathbb{C}^4$ with a metric of appropriate signature, and impose suitable reality conditions on the connection components in order to obtain integrable equations defined on the real domain. We will not discuss the question of the reality condition in this article, since our (primary) objective here is to discuss issues related to integrability in the complex domain. Interested readers are referred to [3] for many references on SDYM.

2.1.3. *Reduction to KdV*

We illustrate the reduction scheme outlined in Subsection 2.1.2 with an example. Here we will derive the KdV equation as a symmetry reduction of the SDYM equations. We point out only the key features of the derivation, leaving aside the technical details which can be found in [4, 17]. We begin by imposing the translational symmetries (step (i)) that the A_α's be independent of the variables $x^{K'} = (y - \bar{y}, z) \in \mathbb{C}^4$. Then we redefine the remaining coordinates as $y^K = (y - \bar{y} \equiv x,\ z \equiv t)$ such that each component of the gauge potential is now of the form $A_\alpha = A_\alpha(x, t)$. Here it is convenient to work with the Lax pair, (4),

rather than the SDYM equations (5) themselves. Under these assumptions, we have $D_1 = \partial_x$, $D_2 = \partial_t - \lambda\partial_x$, and (4) becomes

$$\partial_x\psi = (A_y + \lambda A_{\bar{x}})\psi, \qquad (\partial_t - \lambda\partial_x)\psi = (A_z - \lambda A_{\bar{y}})\psi.$$

Substituting the first equation into the second, we obtain a $(1+1)$ linear system for the reduced form of the Lax pair

$$\partial_x\psi = (A_y + \lambda A_{\bar{z}})\psi, \qquad \partial_t\psi = (A_z - \lambda(A_{\bar{y}} + A_y) + \lambda^2 A_{\bar{z}})\psi. \tag{6}$$

Next we assume that the A_α's are elements of the Lie algebra $\mathrm{sl}(2,\mathbb{C})$ (step (ii)) and represent them by the (2×2) trace-free matrices, $A_y \equiv U$, $A_{\bar{z}} \equiv A$, $A_{\bar{y}} \equiv B$, $A_z \equiv V$. The integrability condition of (6) yields the following reduced SDYM equations,

$$\partial_t U - \partial_x V + [U, V] = 0, \tag{7}$$

$$\partial_x(U + B) = [U, B] + [A, V], \tag{8}$$

$$\partial_x A = [A, B]. \tag{9}$$

These equations are simplified considerably by step (iii) of the reduction scheme. It can be shown [4] that there exists a gauge choice in which A, B satisfies $[A, B] = 0$, so that from (9), $\partial_x A = 0$. It follows immediately from the autonomy condition that A must be a constant $\mathrm{sl}(2,\mathbb{C})$ matrix. The remaining gauge freedom allows A to be determined up to conjugation, $A \to fAf^{-1}$ by a constant $\mathrm{sl}(2,\mathbb{C})$ matrix f. This can be used to put A in either of the normal forms for the (2×2) trace-free matrices

(i) $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$,

(ii) $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Cases (i) and (ii) lead to the KdV and NLS equations, respectively. We consider only case (i) here; the treatment of case (ii) is very similar. The residual gauge freedom (that preserves the normal form of A given above as well as leaves the condition $[A, B] = 0$ invariant) is used to choose U to be off-diagonal. Therefore, we have the following matrices,

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad U = \begin{pmatrix} 0 & \alpha \\ u & 0 \end{pmatrix},$$

$$B = \beta\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad V = \begin{pmatrix} v_1 & v_2 \\ v_3 & -v_1 \end{pmatrix},$$

which must satisfy the system given by (7), (8); (9) is identically satisfied in this gauge. From the (1,2) element of (8), it follows (after using the autonomy condition) that α must be a constant. We set $\alpha = 1$, since the case $\alpha = 0$ does not lead to any nontrivial equation. The remaining equations in (8) and (7) can now be solved systematically to determine β and the elements of V in terms of u and its x-derivatives

$$\beta = \tfrac{1}{2}u, \qquad V = \tfrac{1}{4}\begin{pmatrix} \partial_x u & -2u \\ \partial_{xx}u - 2u^2 & -\partial_x u \end{pmatrix}.$$

Now it follows (from the (2,1) element of Equation (7)) that $u(x,t)$ must satisfy the KdV equation

$$4\partial_x u = \partial_{xxx}u - 6u\partial_x u.$$

2.2. The Darboux–Brioschi–Halphen (DBH) system

The symmetry reductions of the SDYM equations gives rise to ODEs in the case where the gauge potentials are restricted to depend on one independent variable. For example, the classical ODEs mentioned in the introduction arise in this way where the underlying gauge group is finite-dimensional. Here we will discuss an interesting system of three coupled ODEs arising from the SDYM equations but in this case the gauge field variables are associated with an infinite-dimensional Lie algebra of volume preserving diffeomorphisms on a 3-sphere, i.e., $g = \mathrm{sdiff}(\mathbf{S}^3)$. This system possesses several remarkable properties which are dramatically different from the other classical ODEs. Although they have been found recently as a SDYM reduction [5], the origin of this system dates back to the last century! It was found by Darboux in his study of orthogonal systems of curvilinear coordinates in $\mathbb{R}^N$. Brioschi and Halphen solved this system in 1881 and we will call it the DBH system. It is interesting that it has appeared only recently in problems in mathematical physics. It was found as the self-dual Einstein equations [19] corresponding to an Euclidian Bianchi-type-IX metric in diagonal form and more recently, it was discovered by Dubrovin [18] as a special similarity solution of the Witten–Dijkgraaf–Verlinde equations in topological field theory. We begin with a brief outline of its derivation from the SDYM equations.

Let the gauge potentials depend on only one coordinate $y+\bar{y} \equiv t$ by imposing three translational symmetries on the SDYM equations. Then fix the gauge by setting $A_y + A_{\bar{y}} = 0$ and introduce the variables

$$A_1 = (A_z + A_{\bar{z}})/2, \quad A_2 = i(A_z - A_{\bar{z}})/2 \quad \text{and} \quad A_3 = i(A_y - A_{\bar{y}})/2,$$

where $A_l = A_l(t)$, $l = 1,2,3$. In terms of these new variables, the SDYM equations (5) are expressed as

$$\frac{\mathrm{d}A_l}{\mathrm{d}t} = [A_j, A_k], \quad l \neq j \neq k \text{ and cyclic.} \tag{10}$$

where the $[\,,\,]$ is the Lie bracket associated with the left-invariant su(2)-generators X_l of the algebra sdiff($\mathbf{S}^3$) satisfying $[X_l, X_j] = X_k$, $l \neq j \neq k$, and cyclic. The vector fields X_l's are the infinitesimal generators for the left SU(2) action on $\mathbf{S}^3$ leaving the standard metric and the volume form on $\mathbf{S}^3$ invariant. The gauge potentials are chosen as follows: Let O: SU(2) $\rightarrow$ SO(3) be a map defined by $O(g) \equiv O_{ij}(g) \in \mathrm{SO}(3)$, $i, j = 1, 2, 3$, $g\sigma_i g^{-1} = \sum_j O_{ij}\sigma_j$, $g \in \mathrm{SU}(2)$ and σ_i's are the standard Pauli matrices. Then the A_i's assume the special form

$$A_i(t) = \sum_j O_{ij}(g)\omega_j(t)X_j.$$

Substituting this in (10) and using the property $2iX_k(g) = g\sigma_k$ for the vector fields (which follows from the definition of left action), yields the DBH system for the ω_i's

$$\partial_t\omega_i = \omega_j\omega_k - \omega_i(\omega_j + \omega_k), \quad i \neq j \neq k, \text{ and cyclic.} \tag{11}$$

The important features of the DBH system which distinguish it from the classical ODEs are as follows:

(i) Its solutions are single-valued in their domain of existence in the complex plane; but they cannot be analytically continued beyond a dense set of essential singularities which form a natural boundary. In contrast, the solutions to classical ODEs which are integrable in the complex plane are single-valued meromorphic functions in the complex plane with simple poles as the only movable singularities.

(ii) The DBH system can be explicitly integrated in terms of the Schwartz triangle function $S(0, 0, 0; t)$ which is the unique inverse of a conformal map from the upper half S-plane to an interior region F in the t-plane bounded by 3 circular arcs intersecting at 0 angles. This fundamental triangle F 'tiles' the interior of a circle C orthogonal to (each side of) F uniformly by a fractional linear transformation $t \rightarrow \gamma(t)$, $\gamma \in \Gamma$; Γ being a discrete subgroup of the modular group PSL$(2, \mathbb{Z})$. This uniformization (or 'tiling') is restricted only to the interior of C and cannot be continued beyond. The Schwartz function is invariant under the transformation $\gamma(t)$: $S(\gamma(t)) = S(t)$ and is the fundamental modular function for the subgroup Γ. $S(t)$ is single-valued within all of C but has no analytic continuation beyond C which forms its natural boundary. The solution to (11) is given by

$$\omega_i = -\tfrac{1}{2}\partial_t \log E_i, \tag{12}$$

where the E_i's are defined by

$$E_1 = \frac{\partial_t S}{S}, \qquad E_2 = \frac{\partial_t S}{S-1}, \qquad E_3 = \frac{\partial_t S}{S(S-1)}, \tag{13}$$

are the (regular) modular forms of weight 2 associated with Γ. The ω_i's also inherit the same natural boundary C whose center and radius can be prescribed

in terms of their initial values. Using (12), (13) the following transformation property can be obtained for the ω_i's: If $\omega_i(t)$ is a solution of (11), then

$$\widetilde{\omega}_i(t) = \frac{\omega_i(\phi(t))}{(ct+d)^2} + \frac{c}{ct+d},$$

is also a solution, where

$$\phi = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}(2,\mathbb{C}), \quad \phi(t) = \frac{at+b}{ct+d}.$$

This can be also verified directly from (11) which is invariant under this transformation. Equation (11) reduces to a single third-order ODE studied by Chazy in 1909 [20],

$$\partial_{ttt} y = 2\partial_t y \partial_{tt} y - 3(\partial_t y)^2, \tag{14}$$

in terms of the variable $y(t) = -2(\omega_1 + \omega_2 + \omega_3)$. Its solution can be obtained by introducing a new variable s such that $t(s) = z_2/z_1, y(t(s)) = 6\partial_t \log z_1$, where z_1 and z_2 are two linearly independent solutions of certain hypergeometric equations. The inverse function $s(t)$ is a different Schwartz triangle function corresponding to the fundamental domain F' with angles $0, \pi/2$, and $\pi/3$. In this case, the transformation group is the full modular group $\mathrm{PSL}(2,\mathbb{Z})$ and the solution $y(t)$ can be expressed as

$$y(t) = \tfrac{1}{2}\partial_t \log\left[\frac{(\partial_t s)^6}{s^4(s-1)^3}\right] = \tfrac{1}{2}\partial_t \log \Delta = \pi i E_2.$$

Here Δ is the unique modular 'cusp' form of weight 12, and E_2 is the normalized Eisenstein series of weight 2 associated with the modular group $\mathrm{PSL}(2,\mathbb{Z})$. It should be clear from above that the solutions of DBH system are quite different from the classical ODEs whose solutions are (a) singly or doubly periodic in the complex plane and are given in terms of the Jacobi elliptic functions (for, e.g., classical 'tops'), (b) expressible in terms of Riemann theta functions of finite genus (associated with periodic solutions of the soliton-type equations), or (c) one of the Painlevé equations.

(iii) Unlike the classical ODEs, the DBH system does not possess any polynomial or rational first integrals, so that it is not algebraically integrable. However, it does admit two transcendental first integrals A and B [21] which can be expressed in terms of complete elliptic integrals. We do not include here the rather complicated formulas for A and B but note that (11) can be expressed in terms of A and B by the (vector product) formula

$$\partial_t \omega_i = M(\omega_i)(\nabla A \times \nabla B)_i,$$

where $(\nabla f)_i \equiv \partial f/\partial \omega_i$ and the integrating factor $M(\omega_i) = (\omega_1 - \omega_2)(\omega_2 - \omega_3)(\omega_3 - \omega_1)$. The equation given above is an example of a Nambu–Hamilton

equation of motion [22]. Such systems are natural generalizations of the canonical Hamiltonian dynamical systems with $2N$ degrees of freedom to those with $3N$ degrees of freedom. The above formula indicates that the phase space (excluding the planes $\omega_i - \omega_j = 0$, $i \neq j$, of the DBH system) admits a nondegenerate Nambu-bracket structure in analogy to the Poisson-bracket structure of the usual Hamiltonian dynamics.

2.2.1. *Lax Pair for the DBH System*

We consider now the Lax pair of the DBH system and discuss how to recover its solution from the linear problem. It will be shown that the Lax pair actually leads to a monodromy problem associated with a special case of the Painlevé VI equations whose solutions are hypergeometric functions. It is then possible to express the solutions of the DBH system in terms of these special P-VI solutions. In this sense, the Lax pair provides an effective 'linearization' of the DBH system. Starting from the linear system (4) for the SDYM equations, it is possible to show that the DBH system (11) is obtained (after a trivial rescaling of the independent variable t) as the integrability conditions for the following pair of ODEs,

$$L\psi \equiv \left(P\frac{\mathrm{d}}{\mathrm{d}\lambda} + \ell\right)\psi = \mu\psi,$$

$$M\psi \equiv \left(\frac{\mathrm{d}}{\mathrm{d}\lambda} + \ell_1 + f_1\frac{\mathrm{d}}{\mathrm{d}\lambda}\right)\psi = \nu\psi \tag{15}$$

for $\psi(\lambda, t) \in \mathbb{C}^2$ provided the function ν satisfies the auxiliary condition

$$\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = \frac{A}{P}\mu, \tag{16}$$

μ being a complex constant. P, ℓ, ℓ_1, f_1 and A are defined below in Equations (18)–(20). Alternatively, (11) also follows from the commutator relation

$$[L, M] = A(\lambda, t)L \tag{17}$$

in terms of the linear differential operators L and M. The relevant quantities in (15), (16) (and (17)) are given as follows: First consider the $\mathrm{sl}(2, \mathbb{C})$ matrices,

$$X = \omega_1 X_1 - \omega_2 X_2, \qquad Y = \omega_1 X_1 + \omega_2 X_2, \qquad Z = \omega_3 X_3,$$

where

$$X_1 = i\sigma_1, \qquad X_2 = i\sigma_2, \qquad X_3 = \sigma_3,$$

σ_i's being the standard Pauli matrices, and $\omega_i = \omega_i(t)$ are the DBH variables. Then, we define

$$\ell = \lambda^2 X + 2\lambda Z + Y, \qquad \ell_1 = Z + \lambda X. \tag{18}$$

Next we have the polynomials (in λ) with the coefficients depending on the ω_i's,

$$P = \alpha(\lambda^4 + 2r\lambda^2 + 1), \qquad f_1 = \alpha\lambda^3 + \beta\lambda,$$
$$A = (\omega_1 + \omega_2 + 2\omega_3) - \alpha\lambda^2, \tag{19}$$

where

$$\alpha = \omega_1 - \omega_2, \qquad \beta = \omega_1 + \omega_2 - 2\omega_3,$$
$$r = \beta/\alpha = (s+1)/(s-1), \qquad s = (\omega_1 - \omega_3)/(\omega_2 - \omega_3). \tag{20}$$

It is a remarkable fact that the original infinite-dimensional Lax pair (recall that the SDYM potentials $A_i \in \text{sdiff}(\mathbf{S}^3)$ can indeed be reduced to a pair of complex vector-valued ODEs. However, the derivation of (15), (16) from (4) is nontrivial and involves technicalities that are outside the scope of the present article. Therefore, they will be presented elsewhere. Equation (15) can be simplified further by eliminating the variables μ and ν via a simple rescaling $\psi = f\psi'$ for some scalar function $f(\lambda, t)$. Then ψ' satisfies

$$\frac{\mathrm{d}\psi'}{\mathrm{d}\lambda} = \frac{\ell}{P}\psi', \tag{21a}$$

$$\frac{\mathrm{d}\psi'}{\mathrm{d}t} + \ell_1\psi' + f_1\frac{\mathrm{d}\psi'}{\mathrm{d}\lambda} = 0, \tag{21b}$$

with the additional equations for f: $\partial_\lambda f = \mu f/P$ and $\partial_t f = \nu f - f_1\partial_\lambda f$. The compatibility condition for the existence of such a function f given by $\partial_{\lambda t} f = \partial_{t\lambda} f$ requires that

$$\partial_t P + f_1\partial_\lambda P - P\partial_\lambda f_1 + AP = 0. \tag{22}$$

Equations (21), (22) provide an equivalent formulation of the DBH Lax pair (15); the DBH system is obtained from the compatibility condition of (21a) and (21b) together with the additional condition given by (22). Note that since P is a polynomial in λ with distinct roots, (21a) is a Fuchsian ODE with regular singular points $\{\lambda_i\}$ and whose fundamental solution ψ' is a multivalued function, analytic in $\lambda \in \mathbb{C}P1 - \{\lambda_i\}$.

2.2.2. *The Monodromy Problem Associated with the DBH System*

In 1857, Riemann formulated the classical problem (the so-called Riemann problem) of constructing a fundamental set of solutions with regular singularities from a prescribed monodromy data. He also studied the deformation of the solution with respect to the singular points when the monodromy remains invariant. The

Riemann problem was subsequently studied extensively by others. Plemelj [23] proved the existence of solutions (under certain assumptions) corresponding to an nth order Fuchsian equation with m regular singular points (for arbitrary m and n), by reducing the Riemann problem to Fredholm-type integral equations. This was also investigated independently by Birkhoff [24] who also generalized the problem to the case of irregular singular points. The monodromy preserving deformation (isomonodromy) problem of such ODEs (with regular as well as irregular singularities) was investigated by Schlesinger, Fuchs, Garnier and more recently by Jimbo *et al.* [25]. For a class of second-order ODEs or (2×2) matrix Fuchsian equations, it was found that the integrability conditions for the associated deformation equations led to the six classical Painlevé transcendents. We show that (21a) belongs to this class of ODEs, namely, the compatibility conditions for (21a) and (21b) is precisely the isomonodromy conditions associated with the Fuchsian ODE (21a) and leads to a Painlevé VI equation with special parameter values. We rewrite (21a) as

$$\frac{\mathrm{d}\psi'}{\mathrm{d}\lambda} = \sum_{i=1}^{4} \left(\frac{\ell_i}{\lambda - \lambda_i}\right) \psi'. \tag{23}$$

Here the singular points $\lambda = \lambda_i$, $i = 1, \ldots, 4$, are the distinct roots of

$$P' = P/\alpha = \lambda^4 + 2r\lambda^2 + 1 = 0,$$

$$\ell_i = \operatorname{Res}(l/P)|_{\lambda=\lambda_i} \in \mathrm{sl}(2,\mathbb{C}), \quad \sum \ell_i = 0$$

(since $\lambda = \infty$ is a regular point for (23)).

The isomonodromy condition for (23) is given by the Schlesinger equations,

$$\mathrm{d}\ell_i = \sum [\ell_i, \ell_j]\, \mathrm{d}(\lambda_i - \lambda_j) + [\mathrm{d}g\, g^{-1}, \ell_i],$$

where $g(t) = \psi'(\lambda = \infty, t)$. It follows from the asymptotic expansion of (21b) near $\lambda = \infty$ that $g(t)$ satisfies $\mathrm{d}g\, g^{-1} = Z\, \mathrm{d}t$.

It can be verified from the definitions in (18)–(20) that the elements of ℓ_i are given only by the ratios $\widehat{\omega}_i \equiv \omega_i/\alpha$ and that the λ_i's are parametrized by the variable s, i.e., $\lambda_i = \lambda_i(s(t))$. Substituting the expressions for the ℓ_i's in the Schlessinger equations (isomonodromy equations) yields

$$\begin{aligned}
\frac{\mathrm{d}\widehat{\omega}_1}{\mathrm{d}s} &= 2\widehat{\omega}_2\widehat{\omega}_3 + (1-r)\widehat{\omega}_1, \\
\frac{\mathrm{d}\widehat{\omega}_2}{\mathrm{d}s} &= 2\widehat{\omega}_1\widehat{\omega}_3 - (1+r)\widehat{\omega}_2, \\
\frac{\mathrm{d}\widehat{\omega}_3}{\mathrm{d}s} &= 2\widehat{\omega}_1\widehat{\omega}_2 - 2r\widehat{\omega}_3,
\end{aligned} \tag{24}$$

together with

$$\alpha = \omega_1 - \omega_2 = \frac{1}{4s}\frac{\mathrm{d}s}{\mathrm{d}t}. \tag{25}$$

It can also be verified in a straightforward manner that the compatibility of Equations (21a) and 21b) give rise to the *same* set of equations as in (24), (25). Furthermore, the Schlessinger equations corresponding to equations of the form (23) with four regular singular points is shown in [25] to yield the Painlevé VI (P-VI) equations, parametrized by the eigenvalues of the ℓ_i's. Therefore, (24) corresponds to a Painlevé VI equation with special values of parameters (the eigenvalues of the ℓ_i's in (23) are $\pm\frac{1}{4}$) in the independent variables s. Note that there is, in fact, only one independent equation in (24) because of the algebraic relations among the $\widehat{\omega}_i$'s, namely, $\widehat{\omega}_1 - \widehat{\omega}_2 = 1, \widehat{\omega}_1 + \widehat{\omega}_2 - 2\widehat{\omega}_3 = r = (s+1)/(s-1)$. This is a Riccatti equation for one of the dependent variables, e.g.,

$$\frac{\mathrm{d}\widehat{\omega}_1}{\mathrm{d}s} = \frac{\widehat{\omega}_1^2}{2s} - \frac{\widehat{\omega}_1}{s-1} + \frac{1}{2s(s-1)}, \tag{26}$$

which is equivalent to a second-order linear ODE. The solution of (26) leads to hypergeometric functions. It is well known that hypergeometric functions are special solutions of P-VI. Equation (26) plays a crucial role in the 'linearization' procedure of the DBH system which is now described.

From (25) and the solutions of (26), it is possible to express the residual variables ω_i's in terms of s and $\mathrm{d}s/\mathrm{d}t$. But in order to obtain the ω_i's explicitly, one still needs to know $s = s(t)$. Also, (24), (25) constitutes only that part of the DBH system, which corresponds to an isomonodromy deformation problem. To obtain the full DBH system, we also need to consider the auxiliary Equation (22) which yields the following equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\ln\alpha = -4\omega_3 \tag{27}$$

in addition to (25). Note that (27) with (24) and (25) are equivalent to the full DBH system (11). From the definition of s in (20) and Equations (25), (27) one can express the ω_i's in terms of s and its time derivatives as follows,

$$\begin{aligned}
\omega_1 &= -\frac{1}{4}\frac{\mathrm{d}}{\mathrm{d}t}\ln\left(\frac{1}{4s(s-1)}\frac{\mathrm{d}s}{\mathrm{d}t}\right),\\
\omega_2 &= -\frac{1}{4}\frac{\mathrm{d}}{\mathrm{d}t}\ln\left(\frac{1}{4(s-1)}\frac{\mathrm{d}s}{\mathrm{d}t}\right),\\
\omega_3 &= -\frac{1}{4}\frac{\mathrm{d}}{\mathrm{d}t}\ln\left(\frac{1}{4s}\frac{\mathrm{d}s}{\mathrm{d}t}\right).
\end{aligned} \tag{28}$$

Note that these are the expressions for the ω_i's which are in terms of the modular function S, given earlier (12), (13), where $s(t)$ is related to S by a Möbius

transformation (and recalling the rescaling of time). Using (27) and (25) and interchanging roles of the independent variables t and s one can solve for the quantity $\widehat{\omega}_3$ as,

$$\widehat{\omega}_3 = 1 + \frac{st''}{t'},$$

where $'$ means $\mathrm{d}/\mathrm{d}s$. Substituting this, using the relationship for $\widehat{\omega}_1, \widehat{\omega}_2$, and $\widehat{\omega}_3$, into the Riccatti equation (26), one obtains the Schwartzian equation for $t = t(s)$,

$$\left(\frac{t''}{t'}\right)' - \frac{1}{2}\left(\frac{t''}{t'}\right)^2 = \frac{1}{2}\left[\frac{1}{s^2} + \frac{1}{(s-1)^2} - \frac{1}{s(s-1)}\right]$$

which can be solved by transforming

$$t(s) = \frac{z_2}{z_1}, \tag{29}$$

where z_1 and z_2 are two linearly independent solutions of the second-order linear equation

$$z'' + v(s)z = 0, \qquad v(s) = \frac{1}{4}\left[\frac{1}{s^2} + \frac{1}{(s-1)^2} - \frac{1}{s(s-1)}\right]. \tag{30}$$

Finally, inverting (29) to obtain $s(t)$, one can express the solution of the DBH system $\omega_i(s(t))$ via (28).

2.2.3. *Remarks*

There are three remarks to be made which are important when considering the monodromy problem and the associated linearization of the DBH system.

1. The compatibility of the Lax pair (15) introduced earlier with $\mu \neq 0, \nu \neq 0$ is *not* an isomorphic deformation problem. This can be seen by expressing the L-equation in (15) as

$$\frac{\mathrm{d}\psi}{\mathrm{d}\lambda} = \left(\frac{\mu I - \ell}{P}\right)\psi = \left(\sum_{i=1}^{4} \frac{R_i}{\lambda - \lambda_i}\right)\psi$$

and noting that here the residue matrices satisfy $\mathrm{Tr}(R_i) = \mathrm{Res}(2\mu/P)|_{\lambda=\lambda_i}$. When $\mu \neq 0$, the eigenvalues of R_i and hence the monodromy is not constant in time. Nevertheless, we have shown that it is possible to 'factorize' this non-isomonodromic problem into two separate parts – the isomonodromic part is given by Equations (21a) and (21b), whereas (22) reflects the departure from isomonodromy. A careful analysis of the direct problem [9] shows that the time evolution equations for the eigenvalues of R_i are equivalent to (22) which leads

to the transformation $s = s(t)$ between the deformation parameter s of the isomonodromy problem and the original independent variable t.

2. In more general terms, the inverse function $s(t)$ obtained from (29), corresponding to more general potentials $v(s)$, may not be single-valued in the complex t-plane. Although the Schwartzian equation can always be solved for $t(s)$ via (29) for arbitrary potentials $v(s)$, the resulting function $s = s(t)$ can be densely branched. Single valuedness in this sense, i.e. integrability in the complex t-plane, is achieved only if $v(s)$ has the form [27]

$$v(s) = \frac{1}{4}\left[\frac{1-\alpha_1^2}{s^2} + \frac{1-\alpha_2^2}{(s-1)^2} - \frac{1+\alpha_3^2-\alpha_1^2-\alpha_2^2}{s(s-1)}\right],$$

where $\alpha_i = 1/n_i$, $n_i \in \mathbb{Z}^+$, or $\alpha_i = 0$, $i = 1,2,3$. The latter case (i.e. $\alpha_i = 0$) corresponds to (30), hence the corresponding function $s(t)$ is single-valued. Consequently, the solutions $w_i(s(t))$ to the DBH system are also single-valued functions. An example where the solutions are not always single-valued (even though the underlying system is 'linearizable' in the same way as the DBH system) is illustrated by the next remark.

3. We now describe a novel fifth-order system which is a generalization of the DBH system (4). It also arises as an ODE reduction of the self-dual Einstein equations associated with a (nondiagonal) Euclidean metric and can be derived as a SDYM reduction in the same manner as the DBH equations via (10). The connection components are given by

$$A_i(t) = \sum_{j=1}^{3} O_{i,j}(g) T_j,$$

where the T_j's are the sdiff(S^3) vector fields defined as

$$T_1 = \omega_1 x_1 + \theta x_2, \qquad T_2 = \phi x_1 + \omega_2 x_2, \qquad T_3 = \omega_3 x_3,$$

where ω_i, θ, and ϕ are functions of t. Substituting these in (10) yields a system of five ODEs for $\omega_i(t)$, $\theta(t)$, and $\phi(t)$ (see [7]),

$$\begin{aligned}
\frac{d\omega_1}{dt} &= \omega_2\omega_3 - \omega_1(\omega_2+\omega_3) + \phi^2,\\
\frac{d\omega_2}{dt} &= \omega_1\omega_3 - \omega_2(\omega_1+\omega_3) + \theta^2,\\
\frac{d\omega_3}{dt} &= \omega_1\omega_2 - \omega_3(\omega_1+\omega_2) - \phi\theta,\\
\frac{d\phi}{dt} &= (\theta-\phi)\omega_1 - (\theta+\phi)\omega_3,\\
\frac{d\theta}{dt} &= -(\theta-\phi)\omega_2 - (\theta+\phi)\omega_3.
\end{aligned} \tag{31}$$

When $\theta = \phi = 0$, (31) reduces to the DBH system. This system can also be solved (in terms of linear equations) in the same way as the DBH system and it is possible to express the dependent variables in terms of a function $s(t)$ and its time derivatives. However, it turns out that $s(t)$ is not single-valued in this case.

We outline the key features of the linearization procedure, space considerations require that we omit most of the details. The Lax pair for (31), also given by (15), (16) (but with different choices for the sl$(2,\mathbb{C})$ matrices and polynomials) can be separated into isomonodromic and nonisomonodromic parts. The isomonodromic deformation problem yields four equations (the Schlessinger equations) involving the ratios of the dependent variables. They also admit two first integrals arising from the (constant) eigenvalues of the corresponding residue matrices ℓ_i. Using the first integrals, it is possible to reduce these equations by quadrature to a single Riccatti equation with respect to the deformation parameter $s(t)$ which in this case is defined by $s(t) = (\omega + R)/(\omega - R)$, where $\omega = \omega_1 = \omega_2 - 2\omega_3$, and $R^2 = (\omega_1 + \omega_2)^2 + (\phi + \theta)^2$. The rest of the system in (31) is recovered from the nonisomonodromic part which, together with the Riccatti equation, provides a Schwartzian equation for $t(s)$. Finally, we arrive at a system where the dependent variables can be expressed in terms of s and its time derivatives in the following form:

$$\omega = \omega_1 + \omega_2 - 2\omega_3 = \tfrac{1}{4}\frac{\mathrm{d}s}{\mathrm{d}t}\left(\frac{s+1}{s(s-1)}\right),$$

$$\omega_3 = -\tfrac{1}{4}\frac{\mathrm{d}}{\mathrm{d}t}\ln\left(\frac{1}{4s}\frac{\mathrm{d}s}{\mathrm{d}t}\right),$$

$$(\omega_1 - \omega_2) \pm i(\phi + \theta) = \frac{1}{4s}\frac{\mathrm{d}s}{\mathrm{d}t}\exp(\pm i\Omega),$$

$$\Omega - \Omega_0 = ia\ln\left(\frac{1+\sqrt{s}}{1-\sqrt{s}}\right),$$

$$\theta - \phi = \left(\frac{ia}{2}\frac{\mathrm{d}s}{\mathrm{d}t}\right)\left(\frac{1}{\sqrt{s}(s-1)}\right),$$

where the constants a and Ω_0 are related to the eigenvalues of ℓ_i, as mentioned above. In addition, $t(s)$ satisfies

$$\left(\frac{t''}{t'}\right)' - \frac{1}{2}\left(\frac{t''}{t'}\right)^2 = \frac{1}{2}\left[\frac{1}{s^2} + \frac{1-a^2}{(s-1)^2} - \frac{1-a^2}{s(s-1)}\right] \equiv 2v(s),$$

which can be solved in the same way as in (29). As mentioned above, here the parameter a is a first integral given explicitly by

$$a^2 = \frac{(\theta-\phi)^2}{(\theta+\phi)^2 + \omega_1 - \omega_2)^2 - (\omega_1 + \omega_2 - 2\omega_3)^2},$$

and is prescribed by the initial values of ω_i, θ, and ϕ. Therefore, a in general, is not zero or the reciprocal of a positive integer which is necessary for the inverse function $s(t)$ to be single-valued, as we mentioned in Remark 2. Consequently, solutions of (31) are in general not single-valued functions in the complex t-plane, although they can essentially be described by a linear second-order equation similar to (30). The system (31) provides a concrete example where linearization does not imply integrability in the sense of single-valuedness in the complex plane! Note that $\theta = \phi$ implies $a = 0$, and in this case the solutions to the resulting fourth-order system are single-valued. (In fact, by suitable transformations of the dependent variables this fourth-order system can be transformed to the DBH system.) But when $a \neq 0$ the solution to (31) is densely branched in the complex domain whenever the initial conditions are such that the value of a differs from being a reciprocal of an integer. Conversely, when a is suitably 'quantized', $a = 1/n$, $n \in \mathbb{Z}$, then the solution of the generalized DBH system is single-valued, and we consider it 'integrable in the complex plane'.

3. Homoclinic Manifolds and Water Waves

Although Stokes published his famous work *On the Theory of Oscillatory Waves* in 1847 [30], it was only during the 1960s that Benjamin [29] realized that the periodic Stokes wave on deep water is, in fact, unstable (see also Whitham [31]). About the same time, Zabusky and Kruskal studied the Korteweg–de Vries (KdV) equation [32], 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. Periodic waves of the KdV equation are stable and, in fact, Benjamin's instabilities are not observed in the KdV equation. Actually, the asymptotic equation governing the one-dimensional slow modulation of periodic wave-trains (such as the Stokes wave) in *deep* water is the nonlinear Schrödinger (NLS) equation. It is the NLS equation where one indeed encounters instabilities similar to those discovered by Benjamin.

Benjamin observed the instability through a linearized analysis of the Euler equations, accordingly the analysis is only valid for the short period of time while the linearized assumption remains valid. The linearized stability analysis does not give qualitative information about the long time evolution of the instabilities. It is perhaps reasonable to conjecture that the end-state is one of complete breakdown with energy equally distributed between all Fourier modes, i.e. thermalization (see, e.g., [35]). However, the remarkable laboratory experiments by Lake *et al.* [34] indicated that thermalization may not occur. They found that the initial instabilities saturate and that the initial state is approximately recovered at some stage of the evolution. These experiments dealt with a small number of unstable modes and relatively short time observations. Since the nonlinear Schrödinger

equation is a useful model for water waves on deep water, they also turned to numerical studies of the (focusing) NLS equation

$$iu_t + u_{xx} + 2|u|^2 u = 0, \tag{32}$$

with periodic boundary conditions, $u(x+L,t) = u(x,t)$. Using up to five unstable modes, their numerical experiments did not indicate any thermalization – the energy remained confined to a relatively small number of Fourier modes. They concluded that the initial-wave profile should be recovered after a sufficiently long time.

From a mathematical point of view it is now known that the instabilities of the periodic waves in the NLS analogous to those of Benjamin and Feir, are closely related to the so-called homoclinic manifolds associated with the NLS equation. We next proceed to explain how this may shed new light on the Benjamin–Feir instabilities and the subsequent time evolution of these instabilities.

3.1. Benjamin–Feir-type instabilities in the NLS equation

Our goal is to develop a deeper understanding of the stability of the Stokes water wave. In the NLS model of water waves it is represented in its simplest form by the solution

$$\tilde{u}(x,t) = a\exp(2i|a|^2 t), \tag{33}$$

where a is any complex constant. In order to investigate its stability, we follow a well-known procedure and add small 'side-bands' to the Stokes wave

$$u(x,t) = \tilde{u}(x,t)(1+\varepsilon(x,t)), \tag{34}$$

where $|\varepsilon| \ll 1$ and

$$\varepsilon(x,t) = \hat{\varepsilon}_{-n}(t)\exp(-i\mu_n x) + \hat{\varepsilon}_n(t)\exp(i\mu_n x) \tag{35}$$

and $\mu_n = 2\pi n/L$. Substituting (34) into (32) and keeping linear terms in ε yield,

$$\varepsilon_t = i\varepsilon_{xx} + 2i|a|^2(\varepsilon + \varepsilon^*).$$

Making use of (35), it follows that the growth rate σ_n of the nth mode, $\hat{\varepsilon}_n(t) = \hat{\varepsilon}_n(0)\exp(\sigma_n t)$, is given by

$$\sigma_{n\pm} = \pm\mu_n\sqrt{4|a|^2 - \mu_n^2}\,. \tag{36}$$

Hence, the side bands are unstable provided that

$$0 < \mu_n^2 < 4|a|^2. \tag{37}$$

The number of unstable modes (which also determines the complexity of the homoclinic structure) is given by the largest integer N, satisfying

$$0 < N < |a|L/\pi.$$

This instability is the analogue (for the NLS model) of the Benjamin–Feir instability of the Stokes water wave described above.

It is possible to associate a specific initial value with the unstable case, namely (see [11])

$$u(x,0) = a + \varepsilon_0 \exp(i\phi)\cos(\mu_n x + \xi), \tag{38}$$

where $0 < |\varepsilon_0| \ll 1$, $\tan(\phi) = \sigma_{n+}/\mu_n^2$ and ξ is an arbitrary real constant. Since the NLS equation admits the symmetry $(x,t,u) \to (x+\xi,t,u)$, it is sufficient to consider the initial condition

$$u(x,0) = a + \varepsilon_0 \exp(i\phi)\cos(\mu_n x). \tag{39}$$

The significance of the phase ϕ will become clear in a moment.

In order to explain how these instabilities are related to homoclinic manifolds, we recall (see, e.g., [11]) the following exact (homoclinic) solution of the NLS equation,

$$u(x,t) = a\exp(2ia^2t)\times \\ \times\frac{1+2\cos(px)\exp(\Omega t+2i\phi+\gamma)+A_{12}\exp(2\Omega t+4i\phi+2\gamma)}{1+2\cos(px)\exp(\Omega t+\gamma)+A_{12}\exp(2\Omega t+2\gamma)}, \tag{40}$$

where

$$\Omega = \pm p\sqrt{4a^2-p^2},$$

$$p = 2a\sin\phi \quad \text{and} \quad A_{12} = \frac{1}{\cos^2\phi}.$$

The periodic boundary conditions are satisfied by imposing $p = 2\pi n/L$.

Thus, for any p satisfying the condition (which is identical to (37))

$$0 < p^2 < 4a^2, \tag{41}$$

the solution (40) represents an orbit homoclinic to the Stokes wave, $u_p(x,t) = a\exp(2ia^2t)$. More specifically, assuming $\Omega > 0$ it is easily verified that $u \to u_p(x,t)$ as $t \to -\infty$ and $u \to u_p(x,t)\exp(4i\phi)$ as $t \to \infty$. Alternatively, it follows that $|u(x,t)| \to |a|$ as $t \to \pm\infty$. In an attempt to follow the homoclinic orbit *numerically*, we choose $L = 2\sqrt{2}\pi$, $a = 0.5$, and γ large negative. Making use of the spatial translational invariance of the NLS equation, an approximate (e.g., for γ large negative) initial value for a homoclinic orbit is obtained from (40),

$$u(x,0) = \tfrac{1}{2}(1+\varepsilon_0\exp(i\phi)\cos(\mu x)), \tag{42}$$

where ε_0 is small, $\mu = 2\pi/L$, and $\sin\phi = \mu$, i.e. $\phi = \pi/4$. (Again note that the initial value for the homoclinic orbit (42) is identical to the initial value (38) obtained from the linearized stability analysis.) A typical numerical solution using $\varepsilon_0 = 10^{-5}$ is shown in Figure 1. Note that the numerical solution stays *close* to, but it is not *on* the homoclinic orbit – instead of returning to a constant modulus, the solution displays quasi-periodic behavior. This is not unexpected since the initial value (42) ignores second-order terms in ε_0 and almost any small perturbation moves the solution away from the homoclinic obit. This also means that one does not expect to observe homoclinic orbits directly in laboratory experiments. However, it is possible to observe homoclinic orbits indirectly numerically by adjusting the phase in the initial value (42). More specifically, Figures 2(a) and (b) show the result for $\phi = \pi/4 \pm 0.01$, respectively, and with $\varepsilon_0 = 0.1$. These figures show that the homoclinic orbit forms a separatrix between two different wave patterns – *we believe that the different sides of the homoclinic orbit should be observable in laboratory experiments*. We remind the reader that similar situations occurs in classical mechanics, for example the pendulum problem. One does not 'observe' the separatrix (homoclinic) solution when the pendulum is in the pure inverted position. Rather, the two 'sides' of the separatrix are observed as the 'oscillating' or 'rotating' states.

The homoclinic orbit (40) is the simplest homoclinic structure (we call this a homoclinic manifold of order one) associated with the focusing NLS equation. More complicated structures (e.g., homoclinic manifolds of order N) can also be

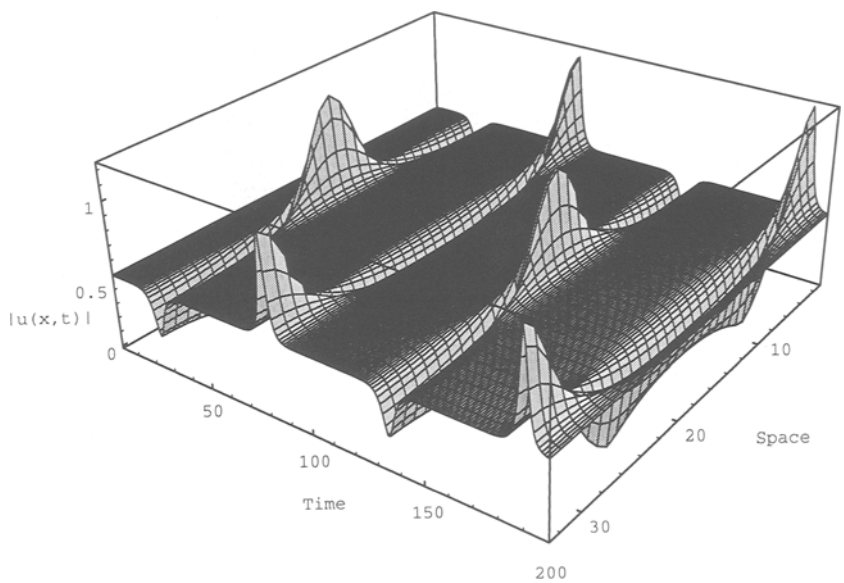

Fig. 1. The homoclinic orbit.

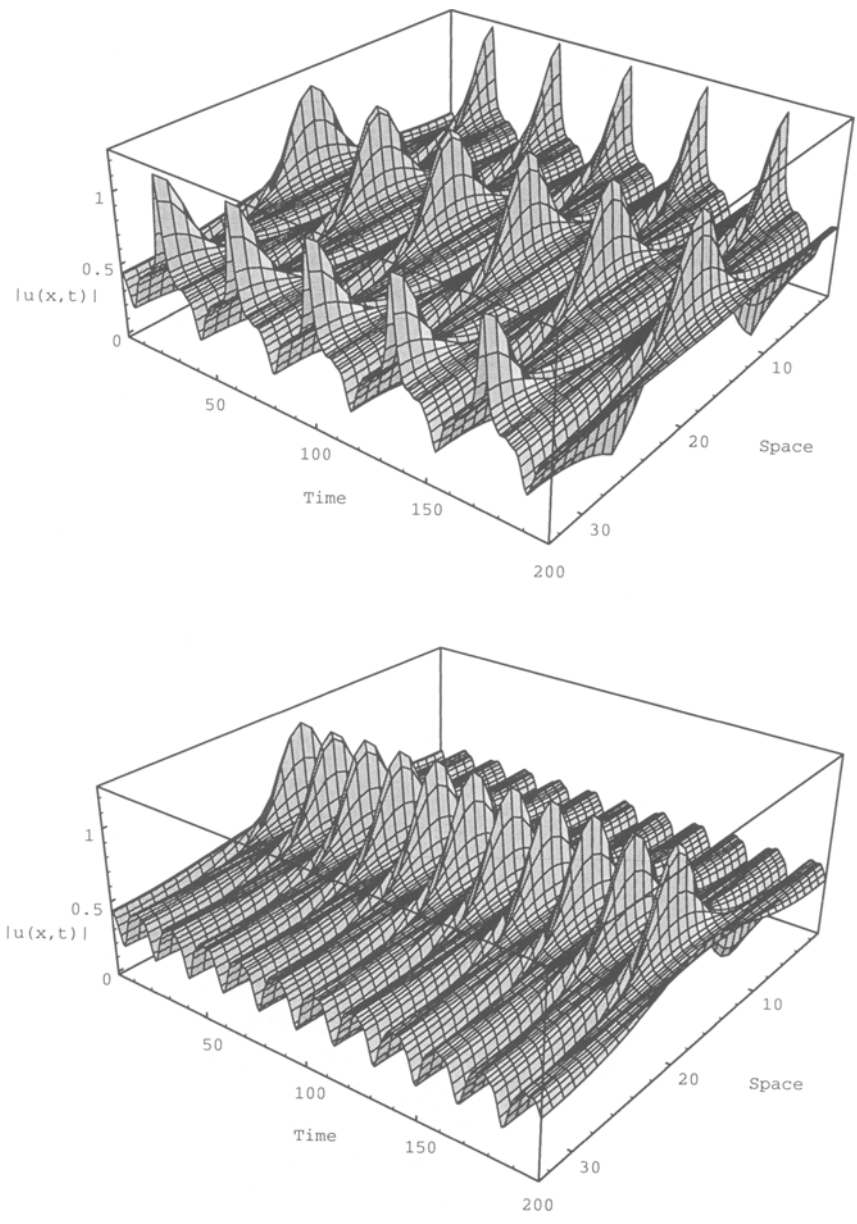

Fig. 2. (a) Outside and (b) inside the homoclinic orbit.

obtained, see, for example, [11]. We now turn our attention to some of the consequences of the presence of homoclinic structures on numerical solutions of the NLS equation. We are particularly interested in the performance of *nonintegrable* vs *integrable* discretizations.

3.2. NUMERICAL SCHEMES

In order to compare the behavior of integrable and nonintegrable discretizations of the NLS equation in the vicinity of homoclinic orbits, we consider two finite difference schemes. The first one is the nonintegrable, standard second-order scheme,

$$i\dot{U}_j + (U_{j-1} - 2U_j + U_{j+1})/h^2 + 2|U_j|^2 U_j = 0, \tag{43}$$

where $h = L/N$ is the grid spacing; periodic boundary conditions are used, $U_{j+N} = U_j$. This scheme is not integrable and the solution may become highly irregular. For example, consider the initial values

$$u(x,0) = a\Big(1 + \tfrac{1}{10}\cos(\mu x)\Big), \tag{44}$$

with $a = \frac{1}{2}$, $L = 4\sqrt{2}\pi$ and $\mu = 2\pi/L$, i.e. a comparison with the instability condition (37) reveals two unstable side bands. A more detailed investigation shows that this initial value is close, but not on a homoclinic orbit (see, for example, [11]). For the time integration the Runge–Kutta–Merson routine, D02BBF, in the NAG library is used with the relative error specified as 10^{-10}. The results using $N = 32$ are shown in Figure 3. One observes the quick and complete breakdown of the solution which has been related to homoclinic crossings (see [33]). This breakdown in the spatial structure may be viewed as thermalization. However, here it is the result of an inadequate spatial resolution and the instability disappears as the mesh is further refined.

The following numerical scheme was carefully designed to be an *integrable* discretization of the NLS equation (see [28]),

$$i\dot{U}_j + (U_{j-1} - 2U_j + U_{j+1})/h^2 + |U_j|^2(U_{j-1} + U_{j+1}) = 0. \tag{45}$$

Note that it is also of second-order accuracy and that the differences with the standard finite difference scheme (43) appear at fourth order in h and only in the *nonlinear* term. However, despite these apparently superficial differences, the two schemes behave very differently. Using the same initial values (44), the Runge–Kutta–Merson scheme and $N = 32$ as before, the solution is shown in Figure 4. There is no indication of any instability and the spatial structure remains smooth, i.e. all the energy are confined to a small number of low modes. The (quasi) periodicity is analogous to the recurrence observed experimentally by Lake *et al.* [34].

Figure 4 shows that the integrable discretization approximates the solutions of the continuous problem at moderate levels of discretization – in the limit as the mesh is further refined, the integrable discretization captures all the features of the continuous problem. However, if the complexity of the homoclinic structure is increased by increasing the number of unstable modes allowed by (37), the integrable discretization may also break down, as we discuss next.

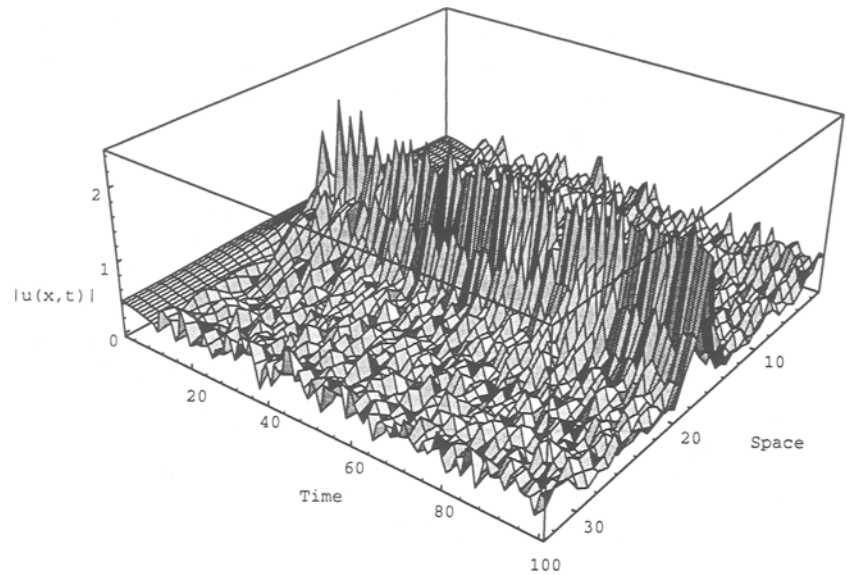

Fig. 3. The finite difference solution.

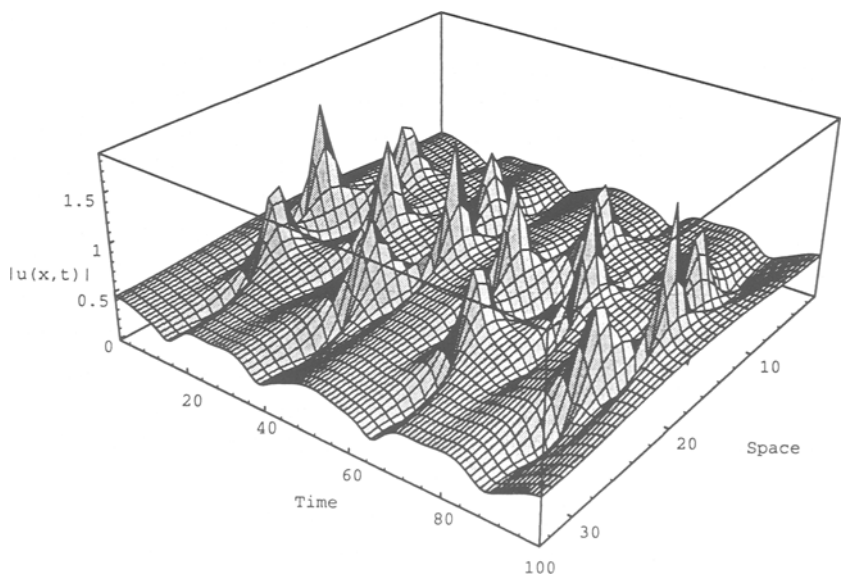

Fig. 4. The integrable discretization.

3.3. MORE UNSTABLE MODES

If we increase the number of unstable modes, we find that the solution may become sensitively dependent on small perturbations. In order to illustrate the sensitivity, we return to the initial value (44) and increase the number of unstable modes to 7 by choosing $a = \frac{5}{2}$ and $L = 2\sqrt{2}\pi$. Consider the integrable discretization (45) as well as the mathematically equivalent one,

$$i\dot{U}_j + (U_{j-1} - 2U_j + U_{j+1})/h^2 + |U_j|^2 U_{j-1} + |U_j|^2 U_{j+1} = 0. \tag{46}$$

The results integrating over 100 time units are shown in Figure 5. In this figure lighter shades of gray indicate larger values of the modulus. This figure, shows that despite their mathematical equivalence, completely different numerical solutions are obtained after a sufficiently long time (about 50 time units)! Calling the two solutions $u_j^{(1)}(t)$ and $u_j^{(2)}(t)$, respectively, Figure 6 shows a plot of $\log_{10} e(t)$ vs t, where $e(t)$ given by

$$e(t) = \sum_{j=0}^{N} \log_{10} \left(|u_j^{(1)}(t) - u_j^{(2)}(t)| \right) / (N+1),$$

is the normalized difference between the two solutions as a function of t. Note that the difference becomes $O(1)$ after about 50 time units.

The two schemes are mathematically equivalent – the only difference computationally is roundoff error (the calculations were done in double precision). Thus, tiny differences due to roundoff error are rapidly amplified to the extent that they have a marked effect on the solution. This sensitive behavior on small perturbations is an indication of an unpredictable, even chaotic, time evolution. However, it should be observed that the spatial structure does not break down completely. Figure 7 shows the Fourier decomposition of the spatial structure (only the lowest 64 modes are shown to facilitate comparison with Figure 8(b)); the energy remains confined to approximately the first 20 modes.

Next we solve the same problem as before (i.e. seven unstable modes), using the Fourier pseudospectral method with $N = 128$ modes and a fourth-order symplectic time integrator (fourth-order split-step) with a time step of 0.001. Since we are dealing with only seven unstable modes in the problem, one may expect the $N = 128$ Fourier modes of the spectral scheme to be large enough to resolve the spatial structure. The solution is shown in Figure 8(a). Note that the solution breaks down reasonably quickly and bears little resemblance to the solutions of Figure 5 after a short period of time – about 10 time units. Its Fourier decomposition is shown in Figure 8(b). As in the case of Figure 7, apart from a few brief excursions to higher modes, the energy is limited to a relatively small number of Fourier modes.

The sensitivity observed in the calculations shown in Figures 5–8 can be explained in terms of the underlying homoclinic manifold (see [12]). The homoclinic manifold is represented by double eigenvalues of the associated linear

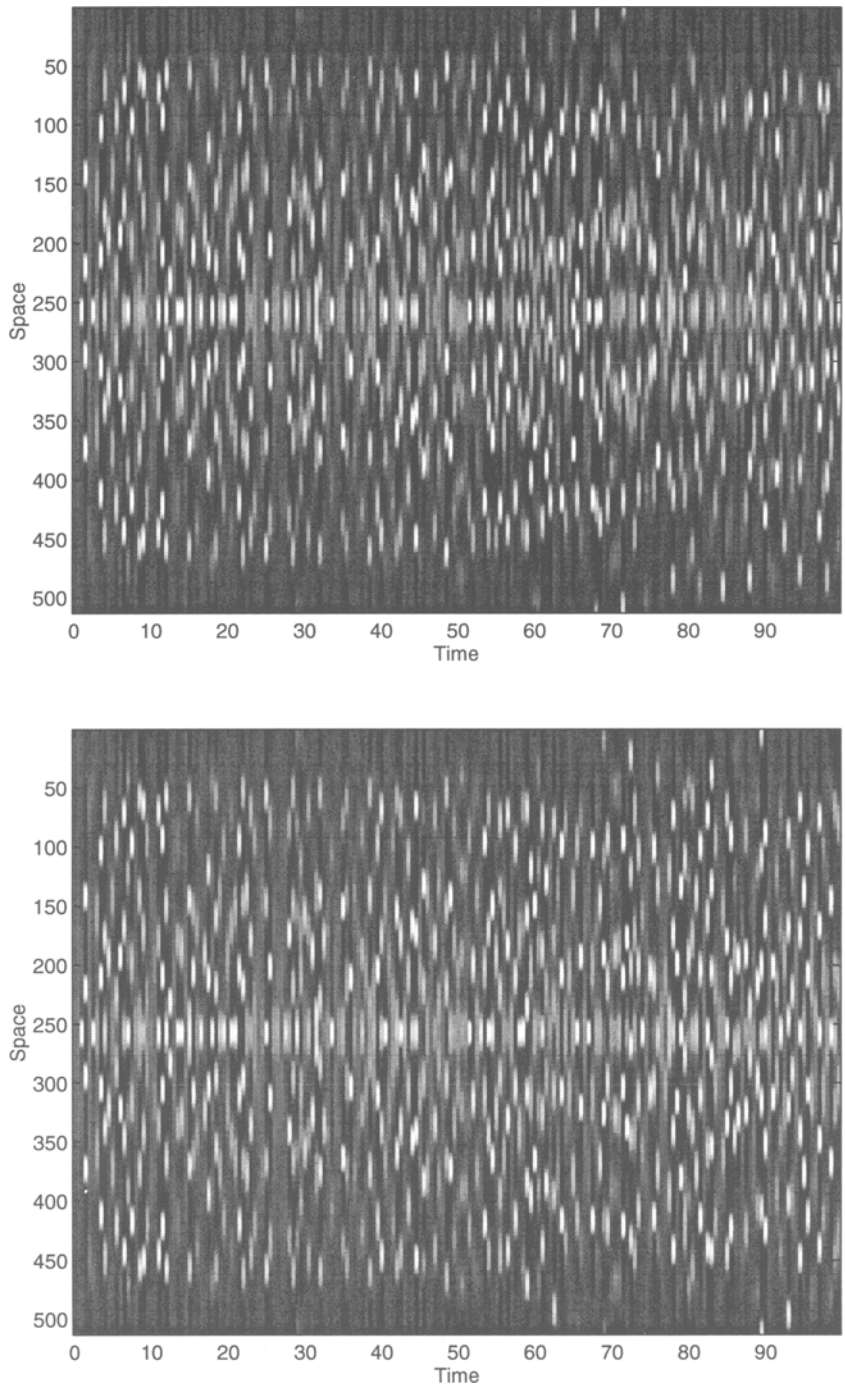

Fig. 5. Two mathematically equivalent forms of the integrable discretization.

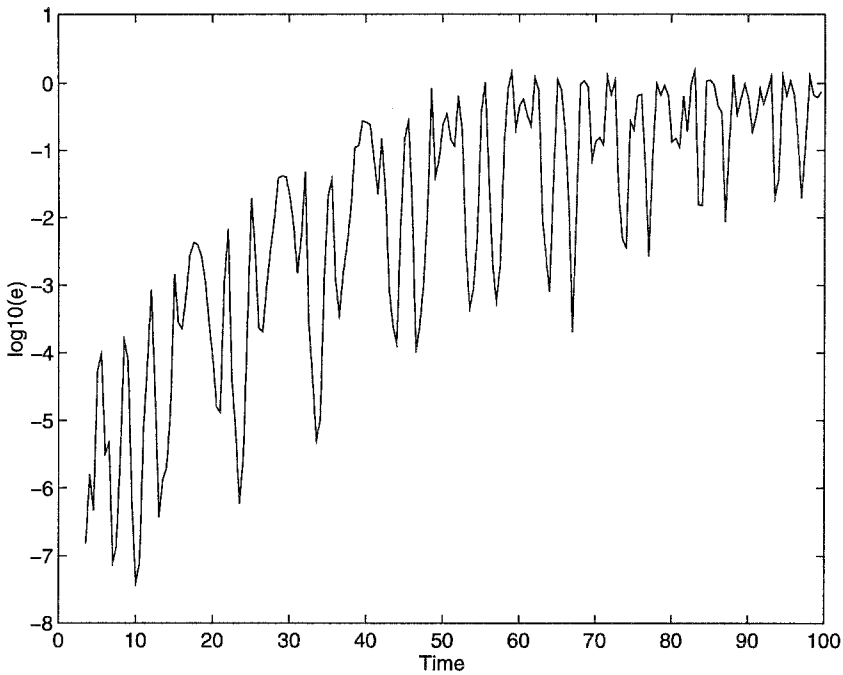

Fig. 6. The difference between the two solutions shown in Figure 5.

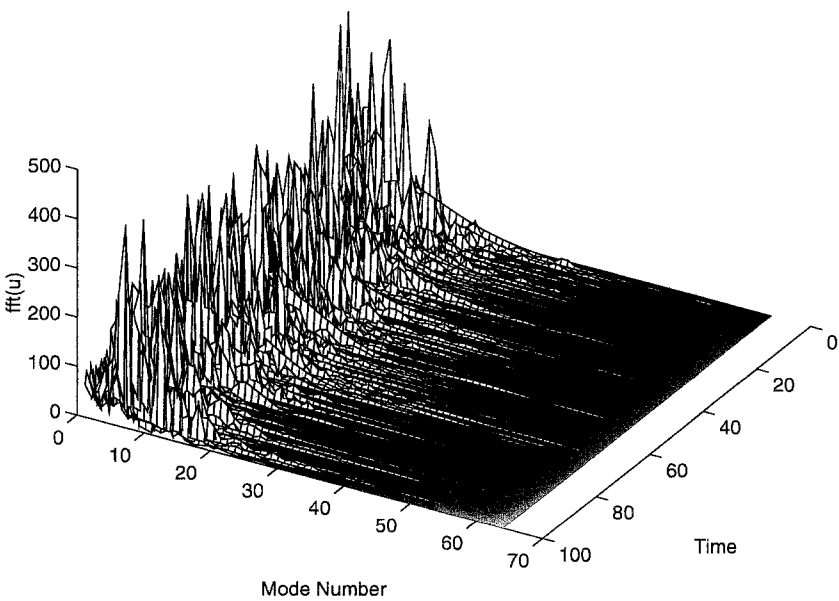

Fig. 7. The Fourier decomposition of the solution of Figure 5 (top).

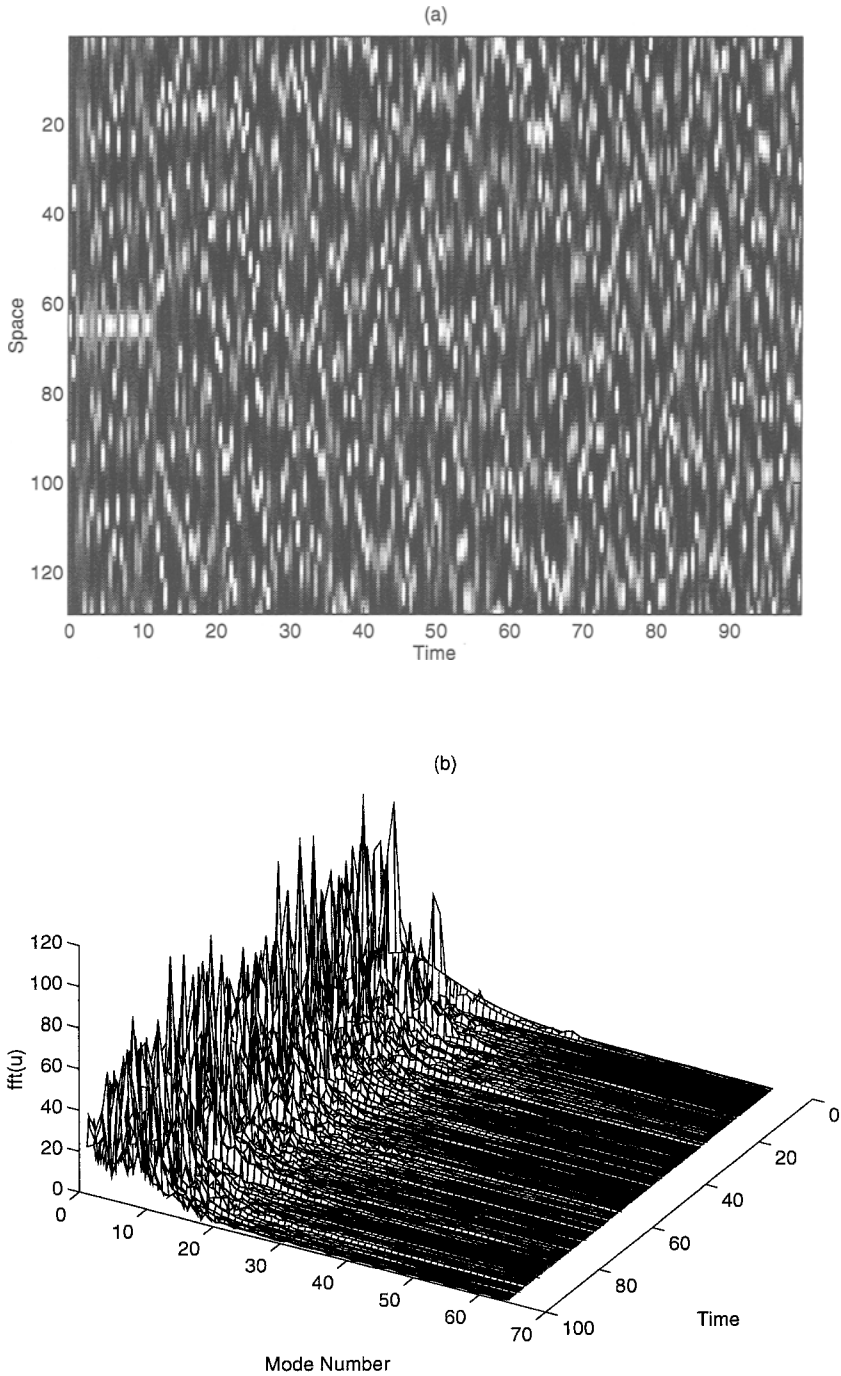

Fig. 8. (a) The Fourier spectral solution and (b) its Fourier decomposition.

spectral problem; each unstable mode represents one double point. If one moves away from the homoclinic manifold, the double eigenvalues are split into gaps and the distance from the homoclinic manifold is given by the 'size' (distance between the eigenvalues) of the gaps. If initial values are chosen on the homoclinic manifold, one expects that small perturbations will cause frequent homoclinic crossings, i.e. the solution will be very sensitive to small perturbations. However, the initial values used in the numerical experiments above (see [12, 14] for a more complete description), are not on the homoclinic manifold. In fact, a perturbation analysis (see [12]) shows that the first double point (unstable mode) is split $O(\varepsilon_0)$, the second one $O(\varepsilon_0^2)$, etc., with the last unstable mode being split $O(\varepsilon_0^n)$ (assuming n unstable modes). Thus, when the numerical errors (including roundoff error) are of $O(\varepsilon_0^n)$, one expects homoclinic crossings and sensitive dependence on the perturbations.

Finally, we turn to a brief discussion of the potential significance of the results obtained from the NLS model for the long-time behavior of the Stokes wave in deep water.

3.4. Conclusions

Our numerical experiments with the NLS equation indicate that the long-time behavior of the Stokes wave depends critically on the number of unstable modes, i.e. the amplitude of the Stokes wave. For a small number of unstable modes (with a perturbation that has the fundamental frequency; this is the most stable situation) our calculations show that the instability saturates and the initial wave is eventually (approximately) recovered. Moreover, our numerical and analytical calculations demonstrate that this situation is fairly robust with respect to small perturbations, i.e. the distance from the homoclinic manifold is larger than the numerical perturbations. However, if the number of unstable modes is increased, the homoclinic manifold becomes more complex, fills a larger part of phase space and the distance from it may become of the same order of magnitude as the numerical perturbations. In this case, the time evolution becomes irregular and no recurrence is observed.

Thermalization and a complete break-down of the spatial structure is only observed if the spatial resolution of the numerical scheme is inadequate, i.e. if the numerical perturbations are large enough.

It is well known that the NLS equation is a good, although idealized model of the actual physical water waves in deep water. However, laboratory experiments of water waves may be interpreted as introducing small perturbations to the NLS equation. For a small number of unstable modes, the NLS equation has been shown to be robust to the small, conservative perturbations introduced by the numerical schemes. One might therefore expect (but it is no less remarkable) to observe the recurrence associated with the NLS equation, also in the laboratory, as did Lake *et al.* [34]. Moreover, for one unstable mode the homoclinic orbit,

more specifically, the two sides of the homoclinic should be observable. If the number of unstable modes are increased, the sensitivity of the NLS solutions prevents any recurrence in numerical simulations – likewise we do *not* expect to observe recurrence in laboratory experiments for a large number of unstable modes. Based on the above arguments, we expect to see a more or less random 'sea of humps' as in Figure 8(a). In the future we will return to the NLS equation and investigate the effect of higher order corrections to the NLS equation on the solutions described above.

Acknowledgements

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. Ablowitz, M. J. and Segur, H.: *Solitons and the Inverse Scattering Transform*, SIAM, Philadelphia, 1981.
2. Novikov, S. P., Manakov, S. V., Pitaevskii, L. P., and Zakharov, V. E.: *Theory of Solitons. The Inverse Scattering Method*, Plenum, New York, 1984.
3. Ablowitz, M. J. and Clarkson, P. A.: *Solitons, Nonlinear Evolution Equations and Inverse Scattering*, Cambridge University Press, Cambridge, 1991.
4. Ablowitz, M. J., Chakravarty, S., and Takhtajan, L. A.: *Comm. Math. Phys.* **158** (1993), 289–314.
5. Chakravarty, S., Ablowitz, M. J., and Clarkson, P. A.: *Phys. Rev. Lett.* **65** (1990), 1085–1087.
6. Mason, L. J. and Woodhouse, N. M. J.: *Nonlinearity* **6** (1993), 569–581.
7. Ablowitz, M. J., Chakravarty, S., and Takhtajan, L. A.: Integrable systems, self-dual Yang–Mills equations and connections with modular forms, in S. Xiao (ed.), *Proc. of Nonlinear Problems in Engineering and Science*, Science Press, Beijing, China, 1992.
8. Takhtajan, L. A.: Modular forms as τ-functions for certain integrable reductions of the Yang–Mills equations, in O. Babelon *et al.* (eds.), *Proc. Verdier Memorial Conference on Integrable Systems*, Birkhauser, Berlin, 1993.
9. Chakravarty, S. and Ablowitz, M. J.: to be published, 1995.
10. Ercolani, N., Forest, M. G., and McLaughlin, D. W.: *Physica* **D43** (1990), 349–384.
11. Ablowitz, M. J. and Herbst, B. M.: *SIAM J. Appl. Math.* **50** (1990), 339–251.
12. Ablowitz, M. J., Schober, C. M., and Herbst, B. M.: *Phys. Rev. Lett.* **71** (1993), 2683–2686.
13. Ablowitz, M. J., Herbst, B. M., and Schober, C. M.: On the numerical solution of the sine-Gordon equation, I. Integrable discretizations and homoclinic manifolds, PAM Report 214, University of Colorado, Boulder, 1994.
14. Ablowitz, M. J. and Schober, C. M.: *Contemp. Math.* **172** (1994), 253–268.
15. Ward, R.S.: *Phys. Lett. A* **61** (1977), 81–82.
    See also, R. S. Ward and R. O. Wells: *Twistor Geometry and Field Theory*, Cambridge University Press, Cambridge, 1990.
16. Chakravarty, S. and Ablowitz, M. J.: On reductions of self-dual Yang–Mills equations, in P. Winternitz and D. Levi (eds.), *Proc. NATO Advanced Research Workshop*, Plenum Press, New York, 1990.
17. Mason, D. and Sparling, G. A. J.: *J. Geom. Phys.* **8** (1991), 263–271.
18. Dubrovin, B.: Private communication.
19. Gibbons, G. W. and Pope, C. N.: *Commun. Math. Phys.* **66** (1979), 267–290.
20. Chazy, J.: *C.R. Acad. Sci. Paris* **149** (1909), 563–565.
21. Chakravarty, S.: To be published.
22. Takhtajan, L. A.: *Commun. Math. Phys.* **160** (1994), 295–315.

23. Plemelj, J.: *Problems in the Sense of Riemann and Klein*, Interscience Publishers, New York, 1964.
24. Birkhoff, G. D.: *Collected Mathematical Papers*, Vol. 1, Dover, New York, 1968.
25. Jimbo, M., Miwa, T., and Ueno, K.: *Physica* **D2** (1981), 306–352 and the references therein.
26. Jimbo, M. and Miwa, T.: *Physica* **D2** (1981), 407–448.
27. Nehari, Z.: *Conformal Mapping*, McGraw-Hill, New York, 1952; (Reprinted by Dover, New York, 1975).
28. Ablowitz, M. J. and Ladik, J. F.: *Stud. Appl. Math.* **55** (1976), 213–229.
29. Benjamin, T. B.: *Proc. Roy. Soc. A* **299** (1967), 59–75.
30. Stokes, G. G.: *Camb. Trans.* **8** (1847), 441–473.
31. Whitham, G. B.: *Linear and Nonlinear Waves*, Wiley, New York, 1974.
32. Zabusky, N. J. and Kruskal, M. D.: *Phys. Rev. Lett.* **15** (1965), 240–243.
33. Schober, C. M. and McLaughlin, D. W.: *Physica* **D57** (1992), 447–465.
34. Lake, B. M., Yuen, H. C., Rungaldier, H., and Ferguson, W. E.: *J. Fluid. Mech.* **83** (1977), 49–74.
35. Hasselmann, K.: Discussion, *Proc. Royal Soc. London (A)*, **299** (1967), 67.