

AD A088023

SOLITONS AND THE INVERSE SCATTERING TRANSFORM

by

Harvey/Segur

Senior Consultant

Aeronautical Research Associates of Princeton, Inc.

50 Washington Road, P.O. Box 2229

Princeton, New Jersey 08540 U.S.A.

Contract N00014-80-C-0048
Contract DAAG-29-78-C-0003

Abstract

Under appropriate conditions, ocean waves may be modeled by certain nonlinear evolution equations that admit soliton solutions and can be solved exactly by the inverse scattering transform (IST). The theory of these special equations is developed in five lectures. As physical models, these equations typically govern the evolution of narrow-band packets of small amplitude waves on a long (post-linear) time scale. This is demonstrated in Lecture I, using the Korteweg-deVries equation as an example. Lectures II and III develop the theory of IST on the infinite interval. The close connection of aspects of this theory to Fourier analysis, to canonical transformations of Hamiltonian systems, and to the theory of analytic functions is established. Typical solutions, including solitons and radiation, are discussed as well. With periodic boundary conditions, the Korteweg-deVries equation exhibits recurrence, as discussed in Lecture IV. The fifth lecture emphasizes the deep connection between evolution equations solvable by IST and Painlevé transcendents, with an application to the Lorenz model.

AUG 18 1980

A

DDC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

80 8 4 142

Part I

The Physical Meaning of Equations with Solitons

One of the important recent advances in mathematical physics has been the discovery that certain nonlinear evolution equations can be solved exactly as initial value problems, using a method that may be called the inverse scattering transform (IST). One of the remarkable features of many of these special equations is that they admit as solutions extremely stable objects called solitons. For convenience, we may use "soliton theories" loosely to describe all of these equations solvable by IST.

It may be regarded as something of a miracle that there are nonlinear evolution equations that are completely integrable. It is a second miracle that several of these equations arise naturally as models of various physical systems, including aspects of ocean waves. My lectures at this School will describe some of the mathematical theory that has been developed in the last fifteen years to solve these special equations. Before discussing the theory, however, it may be useful to give some idea of the sense in which these soliton theories model ocean waves. Consequently, in this first lecture, I will try to describe the physical meaning of equations with solitons.

The spectrum of ocean waves is large and diverse, and soliton theories describe a rather small part of it. In order to see the context in which soliton theories arise, we may attempt a crude classification of ocean waves, admitting in advance that the classification probably will be incomplete.

Let us first classify ocean waves on the basis of whether the wave amplitudes are large or small. Large amplitude waves may break, among other things. Fully nonlinear theories are needed to describe them, and will be discussed by other speakers at this School. I will restrict my attention to small amplitude waves, because soliton theories arise in the context of small amplitude waves. These waves

the soliton theory applies, perhaps followed by an even longer time-scale on which something else applies.

What are these soliton theories? Here is a list of some of the equations that can be solved exactly by IST, and also model ocean waves.

1. Small amplitude waves propagating in only one spatial dimension.

(a) The Korteweg-deVries (KdV) equation,

$$u_t + 6uu_x + u_{xxx} = 0, \quad (1.1)$$

governs long surface (or internal) waves. (Here subscripts denote partial derivatives.) See Korteweg and deVries (1895), and Hammack and Segur (1974, 1978).

(b) Under other circumstances, the evolution of "long" internal waves may be governed instead by the modified Korteweg-deVries (mKdV) equation

$$u_t - u^2 u_x + u_{xxx} = 0, \quad (1.2)$$

an equation due to Benjamin (1967) and Ono (1975),

$$u_t + uu_x + \partial_x^2 \frac{1}{\pi} \int \frac{u(y,t)}{x-y} dy = 0, \quad (1.3)$$

or by other models.

(c) Nearly monochromatic (i.e., narrow band) surface waves are governed by the nonlinear Schrödinger equation,

$$i\psi_t + \psi_{xx} + \sigma|\psi|^2\psi = 0, \quad \sigma = \pm 1; \quad (1.4)$$

(Zakharov, 1968; Hasimoto and Ono, 1972; Yuen and Lake, 1975).

(d) The sine-Gordon equation,

$$\phi_{xt} = \sigma \sin \phi, \quad (1.5)$$

governs the waves that are nearly neutrally stable in a baroclinically unstable system on a beta-plane earth (Gibbon, James and Moroz, 1979).

2. Small amplitude waves in more dimensions.

(a) The equation of Kadomtsev and Petviashvili (1970),

$$(u_t + uu_x + \sigma u_{xxx})_x + u_{yy} = 0, \quad (1.6)$$

is a two-dimensional generalization of the KdV equation.

Another generalization of KdV (in a different limit) is the cylindrical KdV equation,

$$q_t + (2t)^{-1} q - 6qq_x + q_{xxx} = 0, \quad (1.7)$$

(b) The resonant interaction of three (narrow-band) packets of nearly resonant internal waves satisfy

$$\begin{aligned} D_1 a_1 &= i\gamma_1 a_2^* a_3^* , \\ D_2 a_2 &= i\gamma_2 a_3^* a_1^* , \\ D_3 a_3 &= i\gamma_3 a_1^* a_2^* , \end{aligned} \quad (1.8)$$

where $D_j = \partial_t + (\vec{C}_j \cdot \nabla)$.

Clearly there is no point in trying to derive all of these equations. Rather, let us derive the KdV equation fairly carefully by a multiple time-scale argument, and I simply will assert that the other equations may be derived by multiple time-scale arguments as well. We begin with the following "exact" problem. We assume that the fluid is homogeneous, incompressible and inviscid. It is subject to a constant,

vertical gravitational force (g), and rests on a horizontal impermeable bed at $z = -h$ (see Figure 1). The pressure vanishes at the free surface, defined by $z = \eta(x, y, t)$, where surface tension may be acting or not. (In this derivation we will omit it, but almost no qualitative changes are needed if it is included.) The motion of the fluid under these forces is assumed to be irrotational, two-dimensional ($\partial y \equiv 0$), and either to vanish as $|x| \rightarrow \infty$ or to be periodic in x for all time.

Fig. 1

The governing equations under these conditions are well known (e.g., Stoker, 1957):

$$\nabla^2 \phi = 0 \quad -h < z < \eta(x, t) \quad (1.9a)$$

$$w = \partial \phi / \partial z = 0 \quad z = -h \quad (1.9b)$$

$$\eta_t + u \eta_x = w \quad (1.9c)$$

$$\phi_t + \frac{1}{2} |\nabla \phi|^2 + g \eta = \text{constant} \quad z = \eta(x, t), \quad (1.9d)$$

and either

$$|\nabla \phi|, \eta \rightarrow 0 \quad |x| \rightarrow \infty \quad (1.9e)$$

or

$$|\nabla \phi|, \eta \text{ periodic in } x \text{ with period } L.$$

These equations, along with appropriate initial conditions, uniquely determine the fluid motion, at least for some finite time. A consequence of these equations is that there are three globally conserved quantities:

$$\begin{aligned} M &= \rho \int \eta \, dx && \text{net mass of wave,} \\ m_x &= \rho \int \left\{ \int_{-h}^{\eta} u \, dz \right\} dx && \text{horizontal momentum,} \\ E &= \rho \int \left\{ \frac{1}{2} g \eta^2 + \int_{-h}^{\eta} \frac{1}{2} |\nabla \phi|^2 \, dz \right\} dx && \text{energy.} \end{aligned} \tag{1.10}$$

To derive the KdV equation from (1.9), we must make more assumptions, to restrict the possible solutions. For this restricted class of solutions, we may replace (1.9) by a simpler problem, whose solutions will approximate some of the solutions of (1.9). In particular, the KdV equation follows by making the following assumptions.

i) Small amplitude waves. If $\bar{\eta}$ represents a "typical" wave amplitude,

$$\epsilon = \bar{\eta}/h \ll 1. \tag{1.11a}$$

ii) Long waves. If k is a typical horizontal wavenumber,

$$(kh)^2 \ll 1. \tag{1.11b}$$

iii) These two effects approximately balance,

$$(kh)^2 = O(\epsilon). \tag{1.11c}$$

Two remarks are in order here. The first is that because (1.9) uniquely determines its solution its not obvious that any additional assumptions are permitted. We must verify a posteriori that our final solution is consistent with the assumptions that led to it. Second, (kh) and ϵ are not well-defined. We may use them as convenient

computational tools only because they eventually drop out of the analysis.

Consistent with (1.11), we may scale the independent variables as

$$z^* = z/h, \quad x^* = \sqrt{\epsilon} x/h, \quad t^* = \sqrt{\epsilon} ct/h, \quad (1.12a)$$

where $c^2 = gh$ (the only speed available in the linearized, long-wave limit). We also introduce a slow time,

$$\tau = \epsilon t^*, \quad (1.12b)$$

so that

$$\frac{\partial}{\partial t} = \sqrt{\epsilon} \frac{c}{h} \left[\frac{\partial}{\partial t^*} + \epsilon \frac{\partial}{\partial \tau} \right]. \quad (1.12c)$$

Rayleigh (1876) noted that if ϕ is analytic at $z = -h$, then we may expand ϕ in a power series there. The final result is that the solution of (1.9a,b) is

$$\phi = \sum_{n=0}^{\infty} \frac{[-\epsilon(1+z^*)^2]^n}{(2n)!} \left(\frac{\partial}{\partial x^*} \right)^{2n} \phi_0(x^*, t^*, \tau) \quad (1.13)$$

If ϕ is analytic at $z = -h$, this series is convergent. If all of the derivatives of ϕ_0 are bounded, it is also asymptotic (in ϵ).

Next we expand the unknowns:

$$\eta(x, t; \epsilon) = h[\epsilon \eta_1(x^*, t^*, \tau) + \epsilon^2 \eta_2 + \dots], \quad (1.14)$$

$$u \Big|_{z=-h} = \frac{\partial \phi}{\partial x} \Big|_{z=-h} = \epsilon \sqrt{gh} [u_1(x^*, t^*, \tau) + \epsilon^2 u_2 + \dots].$$

It follows that at the free surface, $z = \eta$,

$$u = \epsilon \sqrt{gh} \left[u_1 + \epsilon \left(u_2 - \frac{1}{2} \frac{\partial^2 u_1}{\partial x^{*2}} \right) + O(\epsilon^2) \right], \quad (1.15)$$

$$w = -\epsilon \sqrt{g\epsilon h} \left[\frac{\partial u_1}{\partial x^*} + \epsilon \left(\frac{\partial u_2}{\partial x^*} + \eta_1 \frac{\partial u_1}{\partial x^*} - \frac{1}{6} \frac{\partial^3 u_1}{\partial x^{*3}} \right) + O(\epsilon^2) \right].$$

Now we simply substitute this into (1.9c) and the tangential derivative of (1.9d), and collect terms. The result at leading order is:

$$\frac{\partial \eta_1}{\partial t^*} + \frac{\partial u_1}{\partial x^*} = 0, \quad (1.16)$$

$$\frac{\partial u_1}{\partial t^*} + \frac{\partial \eta_1}{\partial x^*} = 0,$$

The unique solution is:

$$\eta_1(x^*, t^*, \tau) = f(r, \tau) + g(l, \tau), \quad (1.17)$$

$$u_1(x^*, t^*, \tau) = f(r, \tau) - g(l, \tau),$$

where $r = x^* - t^*$, $l = x^* + t^*$. The initial data for (η, u) define (f, g) , which inherit their boundedness, smoothness, etc. At this order, the solution consists of a left- and a right-traveling wave. There is no interaction and no evolution, so all waves are permanent waves.

At the next order, we obtain

$$\frac{\partial \eta_2}{\partial t^*} + \frac{\partial u_2}{\partial x^*} + \frac{\partial \eta_1}{\partial \tau} + \frac{\partial}{\partial x^*} (u_1 \eta_1) - \frac{1}{6} \frac{\partial^3 u_1}{\partial x^{*3}} = 0, \quad (1.18)$$

$$\frac{\partial u_2}{\partial t^*} + \frac{\partial \eta_2}{\partial x^*} + \frac{\partial u_1}{\partial \tau} + u_1 \frac{\partial u_1}{\partial x^*} - \frac{1}{2} \frac{\partial^3 u_1}{\partial x^{*2} \partial t^*} = 0.$$

where (u_1, η_1) are defined by (1.17). The next step is more transparent if we eliminate u_2 (or η_2) and write the result in characteristic coordinates:

$$\begin{aligned}
 -4 \frac{\partial^2 \eta_2}{\partial r \partial \ell} - \frac{\partial}{\partial r} \left[2 f_\tau + 3 f f_r + \frac{1}{3} f_{rrr} \right] + \left[2 g_\ell f_r + g f_{rr} + g_{\ell\ell} f \right] \\
 - \frac{\partial}{\partial \ell} \left[-2 g_\tau + 3 g g_\ell + \frac{1}{3} g_{\ell\ell\ell} \right] = 0. \quad (1.19)
 \end{aligned}$$

The dependence of each term on (r, ℓ) is explicit, so (1.19) can be integrated easily. It is apparent that η_2 will grow linearly both in r and in ℓ unless we require

$$2 f_\tau + 3 f f_r + \frac{1}{3} f_{rrr} = 0, \quad (1.20)$$

$$-2 g_\tau + 3 g g_\ell + \frac{1}{3} g_{\ell\ell\ell} = 0,$$

which can be rescaled to the form (1.1) if desired. We must also require that

$$f, \quad f_r, \quad \int_{-\infty}^{\infty} f \, dr, \quad g, \quad g_\ell, \quad \int_{-\infty}^{\infty} g \, d\ell \quad (1.21a)$$

all be bounded if $-\infty < x < \infty$, and that

$$\oint f \, dr = 0 = \oint g \, d\ell \quad (1.21b)$$

in the periodic problem. Then η_2 is bounded;

$$\begin{aligned}
 \eta_2(r, \ell, \tau) = \frac{1}{4} \left[g_\ell \int^r f \, d\hat{r} + f_r \int^\ell g \, d\hat{\ell} + 2fg \right] \\
 + f_2(r, \tau) + g_2(\ell, \tau). \quad (1.22a)
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 u_2 = & -\frac{1}{4} \left[g_{\ell} \int^r f \, d\hat{r} - f_r \int^{\ell} g \, d\hat{\ell} \right] - \frac{1}{4} (f^2 - g^2) \\
 & + \frac{1}{3} (f_{rr} - g_{\ell\ell}) + f_2 - g_2 .
 \end{aligned} \tag{1.22b}$$

The conditions (1.21) guarantee that the left- and right-going waves do not affect each other long enough to interact strongly on this time-scale [$\tau = O(1)$]. However, each of the two wave trains does interact with itself for a long time, and the two KdV equations govern the evolution of each wave train on this longer time-scale. This is the main point of the derivation. The KdV equation governs the evolution on a slow time-scale of a small amplitude wave that satisfied the linear wave equation on a fast time-scale.

To stop at $O(\epsilon^2)$, one may set $f_2 = g_2 = 0$, and (1.22) gives the second order corrections of the unknowns. To go on to third order, it is necessary to use f_2 and g_2 to eliminate secular terms at the next order. In principle, the expansion can be carried to arbitrarily high order, although this is rarely done in practice.

We will discuss how to solve the KdV equation in the second lecture. For the moment, we consider only some of the implications of this derivation. The first is that both the wave equation (1.16) and the KdV equation (1.20) are ϵ -independent. The expansion is intended to be valid in the limit $\epsilon \rightarrow 0$, and we may take this limit without emasculating the governing equations. That is indicative that the KdV equation is a true asymptotic equation.

A second consequence of this derivation concerns the conservation laws. It follows from (1.20) that if f vanishes rapidly as $|r| \rightarrow \infty$, or is periodic, then

$$I_1 = \int f \, dr , \quad I_2 = \int f^2 \, dr , \quad I_3 = \int \left[f^3 - \frac{1}{3} (f_r)^2 \right] dr$$

(1.23)

all are conserved. In fact, the KdV equation has an infinite number of conservation laws. These may be reconciled with the three physical conservation laws in the following way. We obtained (1.20) by expanding the dependent variables, as in (1.14). If we also expand the mass integral (for example) in powers of ϵ , each coefficient in that expansion must also be conserved. The infinity of conserved quantities for KdV are related to these coefficients. For example, we may easily verify that for the right-going waves (i.e., $g \equiv 0$), to $O(\epsilon^2)$,

$$\begin{aligned} M &= \frac{\rho h^2}{\sqrt{\epsilon}} [\epsilon I_1 + O(\epsilon^3)] , \\ m_x &= \rho h^2 \sqrt{gh/\epsilon} [\epsilon I_1 + \frac{3}{4} \epsilon^2 I_2 + O(\epsilon^3)] , \\ KE &= \frac{\rho g h^3}{2\sqrt{\epsilon}} [\epsilon^2 I_2 + \epsilon^3 I_3 + O(\epsilon^4)] , \\ PE &= \frac{\rho g h^3}{2\sqrt{\epsilon}} [\epsilon^2 I_2 + O(\epsilon^4)] , \\ E &= KE + PE . \end{aligned} \tag{1.24}$$

Part II

Introduction to the Inverse Scattering Transform

The first lecture was devoted to the physical derivation of equations that admit solitons. We saw that even though the equations themselves are fully nonlinear, they typically arise in physical problems by eliminating secular terms in a weakly nonlinear theory.

This lecture is devoted to the method of solution of these special equations. This method goes under a variety of names, including the inverse scattering transform. It turns out that this method can be viewed as a generalization of Fourier analysis to certain nonlinear problems. It provides the exact solution to certain nonlinear evolution equations, just as the Fourier transform does for certain linear evolution equations.

The outline of this lecture is first to review briefly the method of Fourier transforms for linear problems, then to sketch how IST works for certain nonlinear problems, and to show that it is a generalization of Fourier analysis. In the process we can say something about the solutions of these equations, and the class of equations to which the method applies. This approach is essentially that of Ablowitz, Kaup, Newell and Segur (1974).

1.

Linear Evolution Equations

Let us consider three examples of linear problems on the infinite interval $(-\infty < x < \infty)$

$$(a) \quad u_t + u_{xxx} = 0, \quad (2.1a)$$

$$(b) \quad i u_t + u_{xx} = 0, \quad (2.1b)$$

$$(c) \quad u_{\tau\tau} - u_{\chi\chi} + u = 0, \text{ or}$$

$$u_{xt} = u. \quad (2.1c)$$

Part (a) may be called the Airy equation; it arises in certain problems in optics. Part (b) is the time-dependent Schrödinger equation, with no potential. Part (c) is the Klein-Gordon equation, written both in laboratory and in characteristic coordinates. In all three cases, the equation holds for all real x , initial data also must be given at $t=0$, and we will require that the initial data be smooth and decay rapidly as $|x| \rightarrow \infty$.

These equations all can be solved by Fourier transform methods. The first step in that approach is to map the initial data into its Fourier transform:

$$\hat{u}(k) = \int_{-\infty}^{\infty} u(x, 0) \exp(-ikx) dx, \quad (2.2)$$

As t changes, $u(x, t)$ evolves according to a partial differential equation, but $\hat{u}(k, t)$ satisfies an ordinary differential equation. (This is precisely the advantage of Fourier transforms.) The equation is so simple that we often skip that step and simply look for solutions in the form

$$u(x, t) = \frac{1}{2\pi} \int \hat{u}(k) \exp(ikx - i\omega t) dk. \quad (2.3)$$

Substituting (2.3) into the (linear) evolution equation yields the (linear) dispersion relation, $\omega(k)$. In particular, for our example problems,

$$\begin{aligned} (a) \quad \omega &= -k^3, \\ (b) \quad \omega &= k^2, \\ (c) \quad \omega &= 1/k. \end{aligned} \quad (2.4)$$

If the problem has a dispersion relation, it contains all of the information that was in the original partial differential equation.

For example,

(i) for a first order (in time) equation, if $\omega(k)$ is real for real k , then the original problem has an "energy" integral that is conserved. In our example problems,

$$(a) \quad \partial_t \int_{-\infty}^{\infty} |u|^2 dx = 0 ,$$

$$(b) \quad \partial_t \int |u|^2 dx = 0 , \quad (2.5)$$

$$(c) \quad \partial_t \int |u_x|^2 dx = 0 ,$$

(ii) If $(d^2\omega/dk^2) \neq 0$, then the problem is "dispersive". Each wave-number travels with its own speed, and the waves sort themselves out in time. [These concepts are discussed in detail by Stoker (1957) and by Whitham (1974), among others.]

The net effect is that in the long time limit, the solutions of each of these equations have rather characteristic features determined largely by the group velocity, $(d\omega/dk)$. For example, a typical solution of (2.1a) for large times is shown in Figure 2. The waves are spreading slowly to the left, because $(d\omega/dk) \leq 0$. Energy is conserved, because $\omega(k)$ is real, so as the waves spread out, the amplitude tends to zero [for (2.1a), as $t^{-1/3}$ or faster].

Fig.2

Thus, the solutions of these conservative, dispersive problems are comparatively simple for large times. They are characterized by the waves dispersing over larger and larger regions of space, with the amplitudes decaying as required by energy conservation. The method to obtain these solutions depends on the existence of two functions: $\hat{u}(k)$ represents the initial data; and $\omega(k)$ represents the evolution equation. This method of solution (Fourier transforms)

may be represented schematically as follows.

$$u(x, 0) \longrightarrow \hat{u}(k, 0) \xrightarrow{\omega(k)} \hat{u}(k, t) \longrightarrow u(x, t) . \quad (2.6)$$

2.

Nonlinear Evolution Equations

Consider next three nonlinear generalizations of the equations in (2.1):

(a) the Korteweg-deVries equation,

$$u_t + 6uu_x + u_{xxx} = 0 ; \quad (2.7)$$

(b) the nonlinear Schrödinger equation,

$$iu_t + u_{xx} + 2|u|^2 u = 0 ; \quad (2.8)$$

(c) the sine-Gordon equation,

$$u_{xt} = \sin u . \quad (2.9)$$

Why these examples? First, each of these equations linearizes to one of the linear equations in (2.1). Second, they have the same energy integrals as their linear counterparts. Third, each arises as a model of some aspect of ocean waves. Fourth, they all possess solitary wave solutions. For KdV,

$$u(x, t) = 2\kappa^2 \operatorname{sech}^2 [\kappa(x - x_0 - 4\kappa^2 t)] , \quad (2.10)$$

where κ and x_0 are free constants; see Figure 3. Finally, and most importantly, each of these equations can be solved by the inverse scattering transform (IST).

Fg.3

Understanding how IST works requires some knowledge of direct and inverse scattering theory. Here is an extremely superficial introduction for the Schrödinger scattering problem. For a very

thorough treatment, see Deift and Trubowitz (1979). Let $q(x)$ be given, satisfying

$$\int_{-\infty}^{\infty} (1+x^2) |q| dx < \infty. \quad (2.11)$$

The direct scattering problem is to find pairs $[\lambda, \psi(x; \lambda)]$ such that

$$\psi_{xx} + [\lambda + q(x)]\psi = 0, \quad -\infty < x < \infty. \quad (2.12)$$

Here λ is a real number, an "eigenvalue", and ψ is required to be a bounded function satisfying certain boundary conditions. For $\lambda > 0$, we may set $\lambda = k^2$, and require that

$$\begin{aligned} \psi &\sim a(k) \exp(-ikx) && \text{as } x \rightarrow -\infty, \\ \psi &\sim \exp(-ikx) + b(k) \exp(ikx) && \text{as } x \rightarrow +\infty. \end{aligned} \quad (2.13)$$

These are related by $|a|^2 + |b|^2 = 1$ (a Wronskian relation). [In the scattering context, these solutions are called "radiation", $a(k)$ is called the "transmission coefficient", and $b(k)$ the "reflection coefficient".] For $\lambda < 0$, there are only a finite number of discrete eigenvalues ["bound states"]. We may set $\lambda = -\kappa_n^2$, and normalize $\int |\psi_n|^2 dx = 1$. The boundary conditions are:

$$\begin{aligned} \psi_n &\sim d_n \exp(\kappa_n x) && \text{as } x \rightarrow -\infty, \\ \psi_n &\sim c_n \exp(-\kappa_n x) && \text{as } x \rightarrow +\infty. \end{aligned} \quad (2.14)$$

A given "potential", $q(x)$, generates certain scattering data $[a(k), b(k); c_n, \kappa_n]$. These may be collected into a single function,

$$B(x) = \frac{1}{2\pi} \int b(k) \exp(ikx) dk + \sum_{n=1}^N c_n^2 \exp(-\kappa_n x). \quad (2.15)$$

The direct scattering problem is to find $B(x)$ for a given $q(x)$.

The inverse scattering problem is given $B(x)$, to find the potential, $q(x)$, that generated it. This is an interesting mathematical question that was solved in a slightly different form in a famous paper by Gel'fand and Levitan (1951). Skipping all intermediate details, the final result is that one must solve a linear integral equation:

$$K(x, y) + B(x+y) + \int_x^\infty K(x, z) B(z+y) dz = 0, \quad y > x. \quad (2.16)$$

(An equation of this form now is called a Gel'fand-Levitan-Marchenko equation.) Then the solution of the inverse scattering problem is

$$q(x) = 2 \frac{d}{dx} K(x, x). \quad (2.17)$$

Thus, we may represent the direct and inverse scattering problems by

$$q(x) \rightarrow B(x+y) \rightarrow K(x, y) \rightarrow K(x, x) \rightarrow q(x). \quad (2.18)$$

After that rather long detour, let us return to nonlinear evolution equations, and come to the main point. In 1967, Gardner, Greene, Kruskal and Miura made a remarkable discovery about the KdV equation, (2.7). Denote its initial data by $u(x, 0)$ and consider

$$\psi_{xx} + [\lambda + u(x, 0)]\psi = 0. \quad (2.19)$$

In this way, $u(x, 0)$ is mapped into scattering data, summarized by $B(2x, 0)$. As t changes, $u(x, t)$ evolves according to (2.7), and of course, the scattering data change as well. The remarkable fact is that if u satisfies (2.7), then

$$\partial_t \lambda = 0. \quad (2.20)$$

In other words, $u(x, t)$ evolves through a family of potentials, all of which have exactly the same spectrum! [For this reason, Calogero and Degasperis (1976) prefer to let IST abbreviate the "Iso-Spectral Transform".] The rest of the scattering data also evolves simply:

$$\begin{aligned} \frac{\partial}{\partial t} a(k, t) &= 0, \\ \frac{\partial}{\partial t} b(k, t) &= 8ik^3 b, \\ \frac{\partial}{\partial t} c_n(t) &= 4\kappa_n^3 c_n, \end{aligned} \quad (2.21)$$

so that

$$\begin{aligned} B(2x, t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} b(k) \exp[2ikx + 8ik^3 t] dk \\ &+ \sum_{n=1}^N c_n^2 \exp[8\kappa_n^3 t - 2\kappa_n x]. \end{aligned} \quad (2.22)$$

Therefore $u(x, t)$ satisfies a nonlinear evolution equation (KdV), but $B(2x, t)$ satisfies a linear evolution equation whose dispersion relation is that of the linearized KdV equation,

$$\omega(2k) = -8k^3. \quad (2.23)$$

But now we know $B(2x, t)$ for any t , and we may reconstruct the new potential, $u(x, t)$, via inverse scattering. To summarize, we have a method to solve the KdV equation:

$$u(x, 0) \rightarrow B(x+y; 0) \xrightarrow{\omega(2k)} B(x+y; t) \rightarrow K(x, y; t) \rightarrow u(x, t). \quad (2.24)$$

In fact, this scheme represents the method of solution for all problems solvable by IST. The steps are:

- i) Map the initial data of the nonlinear evolution equation into the scattering data, using the linear scattering problem;
- ii) Let the scattering data evolve, in accord with the dispersion relation of the linearized evolution equation;
- iii) Reconstruct the solution of the nonlinear evolution equation at a later time by solving a linear integral equation.

Notice that each step in this method is linear, and that the whole procedure parallels the method of Fourier transforms for linear problems. In this sense, IST is a generalization of Fourier analysis to certain nonlinear problems.

What about the solutions of an equation solvable by IST on $-\infty < x < \infty$? The scattering data consists of a discrete spectrum (bound states) and a continuous spectrum (radiation), and these represent different kinds of solutions. Each discrete eigenvalue represents one solitary wave (or "soliton", since they are no longer required to be solitary):

$$B \sim c_n^2 \exp(-\kappa_n x) \longrightarrow u \sim 2\kappa_n^2 \operatorname{sech}^2\{\kappa_n(x - x_n - 4\kappa_n^2 t)\}, \quad (2.25)$$

where $x_n = x_n(c_n)$. The permanence of these waves is insured by the fact that the eigenvalues $(-\kappa_n^2)$ are time-independent. Notice that each of these waves has positive speed, and that the bigger waves move faster.

The continuous spectrum requires a little more care, but in a crude sense it represents a part of the solution that behaves almost as if the problem were linear. More precisely, it represents a dispersive wave-train in which the linearized group-velocity (which is negative for KdV) plays an important role. The net result is that

the long-time solution of KdV is comparatively simple, as shown in Figure 4. It may be worth emphasizing that even though the equation is fully nonlinear, its long time behavior may be predicted to any desired accuracy. The problem is stable, and there are no chaotic solutions.

Fg.4

3.

Generalizations

The reader may have the impression by now that IST is a miracle that works. To a certain extent that is true, but that is no reason not to use it wherever possible. The question is, "To what problems does IST apply?" Next we show that there are infinitely many problems solvable by IST. (So even though it is a miracle, it may not be uncommon.) Let us consider a different scattering problem on $-\infty < x < \infty$:

$$\begin{aligned}\partial_x V_1 + i\zeta V_1 &= qV_2, \\ \partial_x V_2 - i\zeta V_2 &= rV_1,\end{aligned}\tag{2.26}$$

with $q, r \rightarrow 0$ as $|x| \rightarrow \infty$. This problem was first analyzed by Zakharov and Shabat (1972) for $r = -q^*$ (* denotes complex conjugate). Note that if $r = 1$, (2.26) can be reduced to (2.12). Note further that if $r = 0$ and $V_1 \exp(i\zeta x) \rightarrow 1$ as $x \rightarrow -\infty$, then

$\lim_{x \rightarrow +\infty} [V_1 \exp(i\zeta x)]$ is just the Fourier transform of $q(x)$. In this sense (2.26) is already a generalization of Fourier transforms.

To construct an IST out of (2.26), we need $[q(x, t), r(x, t)]$ to evolve so that $\partial_t \zeta = 0$; i.e., we will force the eigenvalue to be time-independent. To do this, we allow the eigenfunctions to evolve according to linear equations:

$$\partial_t V_1 = A V_1 + B V_2 , \quad (2.27)$$

$$\partial_t V_2 = C V_1 - A V_2 ,$$

where $A = A(q, r; \zeta)$, etc. Compatibility of (2.26) and (2.27) requires that $(\vec{V})_{xt} = (\vec{V})_{tx}$. Demanding $\partial_t \zeta = 0$ yields:

$$A_x = q C - r B ,$$

$$B_x + 2i \zeta B = q_t - 2A q , \quad (2.28)$$

$$C_t - 2i \zeta C = r_t + 2A r ,$$

If (A, B, C) satisfy these coupled ordinary differential equations (in x), then the eigenvalue is constant, and we can construct an IST based on (2.26). Boundary conditions for (2.28) are obtained by comparing (2.26), (2.27) as $|x| \rightarrow \infty$. Because $(q, r) \rightarrow 0$, we know that $V_1 \sim C_1 \exp(-i\zeta x)$, $V_2 \sim C_2 \exp(i\zeta x)$ from (2.26). To assure compatibility with (2.27), we require

$$A \rightarrow A_0(\zeta), \quad B, C \rightarrow 0 \quad \text{as} \quad |x| \rightarrow \infty . \quad (2.29)$$

This gives six boundary conditions for (2.28), which is now overdetermined. Therefore (2.28), (2.29) will have no solution except in special cases.

Given (2.26), the choice of nonlinear evolution equations is determined by $A_0(\zeta)$. Here we use the linearized dispersion relation:

$$\omega(2\zeta) = 2i A_0(-\zeta) . \quad (2.30)$$

For example, if $A_0(\zeta) = -2i \zeta^2$, then (2.28) has a solution only if

$$\begin{aligned} i q_t + q_{xx} - 2q^2 r &= 0 , \\ i r_t - r_{xx} + 2q r^2 &= 0 . \end{aligned} \quad (2.31)$$

If $r = -q^*$, each of these reduce to the nonlinear Schrödinger equation, (2.8).

Theorem (Ablowitz, Kaup, Newell, and Segur, 1974)

Let $\omega(k)$ be a ratio of entire functions, and be real for real k . Then (2.26) generates a nonlinear evolution equation that is solvable by IST, and whose linearized dispersion relation is $\omega(k)$. If $\omega(k)$ also is an odd function of k , then (2.12) generates a different nonlinear equation solvable by IST. Its linearized dispersion relation also is $\omega(k)$.

If your objective is to generate nonlinear equations solvable by IST, you need two ingredients:

- i) A scattering problem; and,
- ii) A linearized dispersion relation.

Each such pair generates one nonlinear problem solvable by IST. If your objective is to solve a particular equation, this is almost no help at all. In my fifth lecture, I will discuss how to determine whether a given equation can be solved by IST.

Part III

More Inverse Scattering on the Infinite Interval

In the second lecture we saw that the inverse scattering transform works, and that it may be regarded as a generalization of the Fourier transform to certain nonlinear problems. For the KdV equation (2.7), IST may be represented schematically as follows:

$$\begin{array}{ccc}
 & \text{direct} & \\
 & \text{scattering} & \\
 u(x, 0) & \longrightarrow & \{a(k, 0), b(k, 0), \kappa_n, c_n(0)\} \\
 & & \downarrow \\
 & & \left[\begin{array}{l} \frac{\partial a}{\partial t} = 0, \quad \frac{1}{b} \frac{\partial b}{\partial t} = 8ik^3 \\ \frac{\partial \kappa_n}{\partial t} = 0, \quad \frac{1}{c_n^2} \frac{\partial c_n^2}{\partial t} = 8\kappa_n^3 \end{array} \right] \quad (3.1) \\
 & & \downarrow \\
 u(x, t) & \longleftarrow & \{a(k, t), b(k, t), \kappa_n, c_n(t)\} \\
 & \text{inverse} & \\
 & \text{scattering} &
 \end{array}$$

In this lecture we will examine IST on the infinite line in more detail. The extra information may help to explain why IST works and what sort of solutions it admits.

1.

Hamiltonian Mechanics

Zakharov and Faddeev (1971) pioneered a description of IST as a canonical transformation of a Hamiltonian system to action-angle variables. This description is an alternative to that of IST as Fourier analysis for nonlinear problems. Both are legitimate; which is preferable is a matter of taste.

The reader may recall that Hamiltonian mechanics is simply a

variational description of certain dynamical systems. (Basic references here are Arnold, 1978; or Goldstein, 1950.) In this formulation, one identifies generalized "coordinates" (q) and "momenta" (p), which describe completely the state of the system. If the problem has N degrees of freedom, then p and q are each N-dimensional vectors. For the partial differential equations of interest here, p and q are infinite-dimensional. One introduces a Hamiltonian, H(p, q), which must have the property that Hamilton's equations,

$$\dot{q} = \frac{\delta H}{\delta p}, \quad \dot{p} = -\frac{\delta H}{\delta q}, \quad (3.2)$$

are equivalent to the equations of motion of the system. Here $(\dot{}) = \partial_t()$, and the derivatives of H in (3.2) are functional derivatives. In general H may depend on time, but for any of the conservative systems under consideration it does not. Any system that has such a variational formulation is said to be Hamiltonian.

It happens that problems solvable by IST are Hamiltonian. As examples, consider

$$H_1 = - \int \{ p q_x^2 + p^2 q_x - p_x q_{xx} \} dx, \quad (3.3)$$

$$H_2 = -i \int \{ q_x p_x + p^2 q^2 \} dx.$$

H_1 yields two evolution equations, which admit the identification $p = q_x$. Each equation reduces to KdV under this identification. Similarly, H_2 gives the nonlinear Schrödinger equation if we identify $p = \pm q^*$ (complex conjugate).

Canonical transformations play an important role in Hamiltonian mechanics. Roughly speaking, a canonical transformation is simply a change of variables,

$$(p, q) \longrightarrow (P, Q), \quad (3.4)$$

that preserves the volume of the phase space; (this corresponds to a mapping that is 1-1 and onto). One may check this property either by means of Poisson brackets (old-fashioned) or a symplectic form (new-fangled). Necessarily, a canonical transformation does not affect the form of Hamilton's equations:

$$H(P, Q): \quad \dot{Q} = \frac{\partial H}{\partial P}, \quad \dot{P} = -\frac{\partial H}{\partial Q}. \quad (3.5)$$

Out of all possible canonical transformations, an especially desirable one is one in which the new Hamiltonian is independent of Q. These are called action (P) and angle (Q) variables. Obviously, if $H = H(P)$, then from (3.5),

$$\dot{P} = 0, \quad \dot{Q} = \frac{\partial H}{\partial P} = \text{constant}, \quad (3.6)$$

so that integration of the equations is as simple as possible. Another way to say this is that if a Hamiltonian system has action-angle variables, then its motion is basically very simple, when viewed appropriately. Unfortunately, there is no general method known to determine whether a system has action-angle variables, or to find them if they exist.

The valuable insight of Zakharov and Faddeev (1971) was that (3.1), the equations that describe the evolution of the scattering data for KdV, was in the form of (3.6).

Theorem (Zakharov and Faddeev, 1971)

Let $u(x, t)$ represent a KdV solution on $-\infty < x < \infty$.

The IST mapping

$$u(x, t) \longrightarrow \left\{ \begin{array}{l} P(k) = k \ln |a(k)|^2, \quad Q(k) = \text{Im} \ln b(k) \\ P_n = -2\kappa_n^2, \quad Q_n = \ln c_n \end{array} \right\} \quad (3.7)$$

is a canonical transformation to action-angle variables.

From this standpoint, it is not surprising that the KdV equation has an infinite set of constants of the motion. They are a representation of the (time-independent) action variables. Moreover, given that the equation on $-\infty < x < \infty$ has an asymptotic state, it is not surprising that the asymptotic behavior is relatively simple. The existence of action-angle variables means that motion is basically simple (when viewed appropriately); the simple asymptotic behavior reveals the basically simple motion.

2.

Scattering Theory

Another apparent miracle in IST is the fact that scattering theory works so well. Even after accepting that the potential in the scattering problem [e.g., $u(x)$ in (2.12)] is determined by the scattering data, one is still surprised at the simplicity of the inversion procedure. In fact, scattering theory works as well as it does because it uses the powerful theory of analytic functions of a complex variable.

We may illustrate this close connection to the theory of analytic functions by considering a scattering problem due to Zakharov and Shabat (1972), and generalized by Ablowitz, Kaup, Newell and Segur (1974):

$$\begin{aligned}\partial_x V_1 + i \zeta V_1 &= q(x) V_2, \\ \partial_x V_2 - i \zeta V_2 &= r(x) V_1.\end{aligned}\tag{3.8}$$

Here $-\infty < x < \infty$, and we assume that $q(x), r(x)$ vanish rapidly as $|x| \rightarrow \infty$. This scattering problem is appropriate for the nonlinear Schrödinger equation, the sine-Gordon equation and infinitely many other problems. However, as time-dependence is not germane to scattering theory per se, we will hold time fixed for the current discussion. Thus (q, r) may be considered known functions that are absolutely integrable.

Solutions of (3.8) may be identified by the boundary conditions they satisfy, and we define for real ζ ,

$$\begin{aligned} \dagger &= \begin{pmatrix} \phi_1(x, \zeta) \\ \phi_2(x, \zeta) \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \exp[-i\zeta x] & \text{as } x \rightarrow -\infty, \\ \bar{\phi} &\rightarrow \begin{pmatrix} 0 \\ -1 \end{pmatrix} \exp[i\zeta x] & \text{as } x \rightarrow -\infty, \\ \psi &\rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix} \exp[i\zeta x] & \text{as } x \rightarrow +\infty, \\ \bar{\psi} &\rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \exp[-i\zeta x] & \text{as } x \rightarrow +\infty. \end{aligned} \tag{3.9}$$

For real ζ , ϕ and $\bar{\phi}$ are linearly independent for all real x , as are ψ and $\bar{\psi}$. Because (3.8) has only two linearly independent solutions, we may define $\{a(\zeta), \bar{a}(\zeta), b(\zeta), \bar{b}(\zeta)\}$ by

$$\begin{aligned} \phi &= a\bar{\psi} + b\psi, \\ \bar{\phi} &= \bar{b}\bar{\psi} + \bar{a}\psi. \end{aligned} \tag{3.10}$$

These are related by a Wronskian relation, $a\bar{a} + b\bar{b} = 1$. From (3.9) and (3.10),

$$\phi \rightarrow \begin{pmatrix} a(\zeta) \exp[-i\zeta x] \\ b(\zeta) \exp[i\zeta x] \end{pmatrix} \text{ as } x \rightarrow +\infty. \tag{3.11}$$

The set $\{a, \bar{a}, b, \bar{b}\}$ for real ζ makes up part of the scattering data. These functions may be viewed as representing the asymptotic behavior of certain solutions of (3.8) as in (3.11), or simply as Wronskians of solutions from (3.10).

Once these functions have been defined for real ζ , they may be extended into the complex ζ -plane. Ablowitz, Kaup, Newell and

Segur (1974) proved that:

- i) $[\phi(x, \zeta) \exp(i\zeta x)]$ and $[\psi(x, \zeta) \exp(-i\zeta x)]$ are analytic functions of ζ for all real x if $\text{Im}(\zeta) > 0$;
- ii) $[\bar{\phi} \exp(-i\zeta x)]$ and $[\bar{\psi} \exp(i\zeta x)]$ are analytic in ζ if $\text{Im}(\zeta) < 0$;
- iii) $a(\zeta) = \lim_{x \rightarrow \infty} \phi_1 \exp(i\zeta x)$ is analytic for $\text{Im}(\zeta) > 0$, and $\bar{a}(\zeta)$ is analytic for $\text{Im}(\zeta) < 0$;
- iv) the discrete eigenvalues of (3.8) are zeros of $a(\zeta)$ or $\bar{a}(\zeta)$ in their regions of analyticity;
- v) as $|\zeta| \rightarrow \infty$, $\text{Im}(\zeta) > 0$, $a(\zeta) \rightarrow 1 + O(\zeta^{-1})$,
 $\phi_2 \exp(i\zeta x) \rightarrow -[r(x)/2i\zeta] + O(\zeta^{-2})$,
 $\psi_1 \exp(-i\zeta x) \rightarrow [q(x)/2i\zeta] + O(\zeta^{-2})$, with similar results for $\text{Im}(\zeta) < 0$.

Most of the results of IST are a consequence of these relations and the time-dependence of the scattering data. In particular, one finds that $a(\zeta)$ and $\bar{a}(\zeta)$ are time-independent. Then it follows from (iv) that the discrete eigenvalues are time-independent as well (the iso-spectral property). It follows from (v) that $\ln a(\zeta)$ has an asymptotic expansion as $|\zeta| \rightarrow \infty$ for $\text{Im}(\zeta) > 0$:

$$\ln a(\zeta) = \sum_{n=1}^{\infty} I_n / (2i\zeta)^n. \quad (3.12)$$

Because $a(\zeta)$ is time-independent, the coefficients in this expansion must be time-independent as well. These are the infinite set of conserved integrals of any of the evolution equations solved by (3.8). It also follows from (v) that if we can reconstruct $\phi_2 \exp(i\zeta x)$ and $\psi_1 \exp(-i\zeta x)$ from the scattering data, then $q(x)$ and $r(x)$ may be obtained by taking a limit. We show next that inverse scattering theory follows just that strategy.

To obtain the linear integral equations that are the heart of inverse scattering theory, write (3.10a) as

$$\frac{\phi \exp(i\zeta x)}{a} = \bar{\psi} \exp(i\zeta x) + \frac{b}{a} \psi \exp(i\zeta x), \quad \zeta \text{ real}, \quad (3.13)$$

For simplicity, let us assume that $a(\zeta)$ has no zeros for $\text{Im}(\zeta) \geq 0$. Then $[\phi \exp(i\zeta x)]/a$ is analytic for $\text{Im}(\zeta) > 0$ and vanishes as $|\zeta| \rightarrow \infty$ there, while $\bar{\psi} \exp(i\zeta x)$ has similar properties for $\text{Im}(\zeta) < 0$. Now the problem of reconstructing these analytic functions from (3.13) is very similar to a famous problem posed by Hilbert (cf., Muskhelishvili, 1953, Ch. 5). That problem may be stated as follows. $F_+(z)$ is analytic for $\text{Im}(z) > 0$ and vanishes as $|z| \rightarrow \infty$ there. $F_-(z)$ is analytic for $\text{Im}(z) < 0$ and vanishes as $|z| \rightarrow \infty$ there. They are related on the real axis by:

$$F_+(x) - F_-(x) = f(x) \quad \text{on } z = x, \quad (3.14)$$

where $f(x)$ is a given function. The "Hilbert problem" is to construct both F_+ and F_- from $f(z)$. If $f(x)$ is absolutely integrable, the solution of the problem is given by:

$$F(z) = \frac{1}{2\pi i} \int \frac{f(x) dx}{x-z} = \begin{cases} F_+(z), & \text{Im}(z) > 0, \\ F_-(z), & \text{Im}(z) < 0. \end{cases} \quad (3.15)$$

Comparing (3.13) and (3.14), it is evident that because (b/a) is known for real ζ , if $\psi \exp(i\zeta x)$ were known for real ζ , then (3.13) would be solved by a formula like (3.15). However, because $\psi \exp(i\zeta x)$ is unknown, we obtain instead a linear integral equation, with a Cauchy-type singularity. Another singular integral equation follows from (3.10b). The usual Gel'fand-Levitan-type of integral equations for (3.8) are essentially the Fourier transform of these coupled singular integral equations.

The original work by Gel'fand and Levitan (1951) and others on inverse scattering theory did not make this connection to the Hilbert problem, which has been developed more recently (e.g., Zakharov and

Manakov, 1979). The advantage of this approach is that it emphasizes the fundamental role played by analytic functions in the theory of inverse scattering.

3.

Solutions of the Nonlinear Schrödinger Equation

The practical consequence of all of this remarkable structure is that the solutions of these special nonlinear evolution equations are quite predictable, and can be computed explicitly (especially for large times) once the initial data has been mapped into scattering data. We may illustrate this by focussing our attention on the nonlinear Schrödinger equation,

$$i q_t + q_{xx} + 2|q|^2 q = 0, \quad (3.16)$$

which describes the nonlinear instability of a packet of nearly monochromatic water waves of small amplitude in one dimension.

The appropriate scattering problem is (3.8), with $r = -q^*$. The discrete spectrum in this case is represented by N discrete eigenvalues, $\zeta_j = (\xi + i\eta)_j$ with $\eta_j \geq 0$, along with N parameters, c_j . The continuous spectrum is represented by $b/a(\xi)$, on $\eta = 0$. [The equation

$$i q_t + q_{xx} - 2|q|^2 q = 0, \quad (3.17)$$

also has physical interest. In this case, $r = +q^*$ in (3.8), and there are no discrete eigenvalues.]

A single eigenvalue with $\eta > 0$ corresponds to a single soliton solution of (3.16).

$$q(x, t) = 2\eta \operatorname{sech} \{2\eta(x + 4\xi t + x_0)\} \exp \{-2i[\xi x + 2(\xi^2 - \eta^2)t + \phi_0]\}, \quad (3.18)$$

This is an "envelope soliton" (as opposed to a KdV-type soliton). It

represents a one-dimensional wave packet that is stable with respect to one-dimensional perturbations, and is shown in Figure 5. For a one-dimensional packet of water waves that are nearly monochromatic and of small amplitude, (3.16) describes the Benjamin-Feir (1967) instability, and one could think of an envelope soliton as a stable equilibrium state for that process.

Fig. 5

N discrete eigenvalues generate N solitons. If they have N different speeds ($\xi_j \neq \xi_k$ for $j \neq k$), they separate in space as $t \rightarrow \infty$, so that the long-time solution of (3.16) is simply a sum of N individual solitons. We should emphasize that solitons are intrinsically nonlinear objects, that disappear in the linear limit.

On the other hand the "radiation", which corresponds to the continuous spectrum, is qualitatively quite similar to the solution of the linearized problem. For comparison we note that

$$i y_t + y_{xx} = 0 \quad (3.19)$$

has a family of self-similar solutions, including

$$y = t^{-1/2} A \exp(i x^2/4t + i \phi) . \quad (3.20)$$

One may also solve (3.19) as an initial value problem on $-\infty < x < \infty$, and evaluate the solution in the long time limit. The result is that along $x/t = 2k$ (the group velocity),

$$y(x, t) \sim t^{-1/2} \hat{y}(k) \exp(i x^2/4t - i \pi/4) , \quad (3.21)$$

where $\hat{y}(k)$ is the Fourier transform of the initial data. This has the form of a slowly-varying similarity solution.

There is a nearly identical similarity solution of (3.16),

$$q = t^{-1/2} A \exp[i(x^2/4t + 2A^2 \ln t + \phi)] . \quad (3.22)$$

In the absence of solitons, the asymptotic solution of (3.21) that evolves from appropriate initial data also takes the form of a slowly-varying similarity solution. Along $x/t = -4\xi$,

$$q \sim t^{-1/2} f(x/t) \exp \{i[x^2/4t + 2f^2 \ln t + g(x/t)]\}, \quad (3.23)$$

where

$$f^2(x/t) = -\frac{1}{4\pi} \ln [1 - |b/a(\xi)|^2], \quad (3.24)$$

and $g(x/t)$ also is determined by $b/a(\xi)$. [See Zakharov and Manakov, 1976; Ablowitz and Segur, 1975; and Segur, 1975.]

The general solution of (3.16) on $-\infty < x < \infty$ involves both solitons and radiation. For KdV these two components separate in space, but for (3.16) they coexist. In the long time limit, the solution consists of N envelope solitons riding on a sea of radiation. The solitons are permanent wave packets, while the radiation decays as $t^{-1/2}$.

Part IV
The Korteweg-deVries Equation
with Periodic Boundary Conditions

To this point we have discussed solitons and IST only on the infinite interval. However, the KdV equation with periodic boundary conditions also has applications in water waves. Moreover, the original discovery of solitons by Zabusky and Kruskal (1965) was based on numerical experiments on the periodic KdV problem.

There is a theory of inverse scattering transforms for the KdV equation with periodic boundary conditions. The most complete version is due to McKean and Trubowitz (1976). In this lecture, we will follow the less general but simpler version of Dubrovin and Novikov (1974). In contrast to the theory of IST on the infinite interval, however, the theory for the periodic problem cannot yet be considered a practical tool for use in applications. After presenting the theory in its current form, we will identify some practical questions that have not yet been answered satisfactorily.

The first work on the periodic KdV problem,

$$\begin{aligned}u_t + 6uu_x + u_{xxx} &= 0, \\u(x, t) &= u(x + L, t),\end{aligned}\tag{4.1}$$

was done by Korteweg and deVries (1895), who found a periodic, traveling wave solution of (4.1):

$$u(x, t) = 2p^2 k^2 \operatorname{cn}^2[p(x - ct + \bar{x}); k] + \beta.\tag{4.2}$$

Here $\operatorname{cn}(\phi; k)$ is the Jacobian elliptic function with modulus k ($0 \leq k^2 \leq 1$),

$$\begin{aligned}c &= 6\beta - 4p^2(1 - 2k), \\pL &= 2K(k),\end{aligned}\tag{4.3}$$

where $K(k)$ is the complete elliptic integral of the first kind (cf., Byrd and Friedman, 1971). If we require $\oint u dx = 0$, as required by (1.21), then

$$\beta = -2p^2 \left[\frac{E(k)}{K(k)} - 1 + k^2 \right], \quad (4.4)$$

where $E(k)$ is the complete elliptic integral of the second kind. Korteweg and deVries called these "cnoidal waves", by analogy with sinusoidal waves. A typical solution is shown in Figure 6, for $k = \frac{1}{2}$. These nonlinear, periodic, traveling waves reduce to infinitesimal sinusoidal waves if $k \rightarrow 0$,

Fg. 6

$$u(x, t) \rightarrow p^2 k^2 \cos 2p(x + 4p^2 t + \bar{x}), \quad (4.5)$$

with $pL \rightarrow \pi$. At the other extreme, if $k \rightarrow 1$,

$$u(x, t) \rightarrow 2p^2 \operatorname{sech}^2 p(x - 4p^2 t + \bar{x}). \quad (4.6)$$

Aside from these special solutions, however, almost nothing was known about the periodic KdV problem until the important numerical work of Zabusky and Kruskal (1965), who were motivated by the earlier work of Fermi, Pasta and Ulam (1955). In both of these studies, the authors observed "recurrence of initial states" after a relatively short time. In other words, rather than showing any tendency to an equipartition of energy among all of the degrees of freedom of the system, the solution of the initial value problem almost returned to its initial configuration repeatedly. The time required for this recurrence was short enough to suggest that only a few of the possible degrees of freedom of the system actually were participating in the process described by (4.1). But how could the solution of (4.1) be so constrained unless (4.1) itself carried additional constraints that had not been discovered? These extra constraints turned out to be the infinite set of conservation laws of the KdV equation (Miura, Gardner and Kruskal, 1968). The discovery of these conservation laws led in turn to the development of IST (Gardner,

Greene, Kruskal and Miura, 1967). We should emphasize the historical importance of the careful numerical studies of Zabusky and Kruskal (1965), which indicated that the KdV equation possessed additional mathematical structure.

Let us now outline the theory of IST for the KdV problem. The reader will observe that some of this theory is analogous to that on the infinite interval, but some of it is not. The scattering problem is:

$$V_{xx} + [\lambda + u(x)]V = 0, \quad (4.7)$$

just as it was on the infinite interval, but in this case $u(x)$ is a periodic function of x . Because scattering theory requires no knowledge of the time-dependence of $u(x, t)$, we may consider (4.7) at a fixed time.

One of the difficult conceptual questions about the periodic KdV problem is to find what plays the role of the scattering data here, since there is no "point at infinity" where the solution simplifies. Without such a special point, we simply choose an arbitrary x_0 in order to begin the analysis. With x_0 fixed (and time fixed), we may identify two linearly independent solutions of (4.7), $\phi(x; x_0, \lambda)$ and its complex conjugate, ϕ^* , by imposing boundary conditions for real λ at $x = x_0$:

$$\begin{aligned} \phi(x_0; x_0, \lambda) &= 1 & \phi^*(x_0; x_0, \lambda) &= 1 \\ \phi_x(x_0; x_0, \lambda) &= ik = i\sqrt{\lambda} & \phi_x^*(x_0; x_0, \lambda) &= -ik \end{aligned} \quad (4.8)$$

One period to the right ($x \rightarrow x+L$), these two functions satisfy the same differential equation again, so they must be a linear combination of the ϕ and ϕ^* :

$$\begin{bmatrix} \phi(x+L; x_0, \lambda) \\ \phi^*(x+L; x_0, \lambda) \end{bmatrix} = \begin{bmatrix} a(x_0, \lambda) & b(x_0, \lambda) \\ b^*(x_0, \lambda) & a^*(x_0, \lambda) \end{bmatrix} \begin{bmatrix} \phi(x; x_0, \lambda) \\ \phi^*(x; x_0, \lambda) \end{bmatrix} \quad (4.9)$$

The matrix connecting these two sets of solutions of (4.7) is called the "monodromy matrix". In some ways it plays the role of the scattering data. Its coefficients are related by a Wronskian relation,

$$|a|^2 - |b|^2 = 1. \quad (4.10)$$

The fact that the coefficients of (4.6) are periodic does not necessarily mean that its solutions are periodic. In fact, most of them are not. The solutions of (4.6) that are periodic play a fundamental role in the theory of inverse scattering. We define next the "Bloch eigenfunctions" to be solutions of (4.6) that satisfy

$$\begin{aligned} \psi(x_0; x_0, \lambda) &= 1, \\ \psi(x+L; x_0, \lambda) &= \mu \psi(x; x_0, \lambda). \end{aligned} \quad (4.11)$$

Because these must also be a linear combination of ϕ and ϕ^* , one may show that for each (x_0, λ) ,

$$\mu^2 - 2a_r \mu + 1 = 0, \quad (4.12)$$

where

$$a = a_r + i a_i. \quad (4.13)$$

Equation (4.12) admits three possibilities.

i) If $|a_r| > 1$, one root of (4.12) is larger than one in magnitude, and the other is smaller. From (4.11), these correspond to Bloch eigenfunctions that grow without bound, either as $x \rightarrow +\infty$ or as $x \rightarrow -\infty$. These are said to be "unstable".

ii) If $|a_r| < 1$, we may define $a_r(\lambda) = \cos p(\lambda)$, and show that

$$\mu = \exp(ip). \quad (4.14)$$

These represent "stable" eigenfunctions.

iii) If $|a_r| = 1$, i.e.,

$$a_r^2 = 1, \quad (4.15)$$

then $\mu = \pm 1$ and the eigenfunctions are either periodic or antiperiodic functions of x . Thus, (4.15) identifies the periodic solutions of (4.6).

These possibilities may be summarized on a "Floquet diagram", in which a_r is plotted as a function of λ for fixed x_0 , as shown in Figure 7. The regions in which $|a_r| > 1$ are called "unstable bands". Necessarily, each unstable band lies between two successive roots of (4.15), numbered λ_{2n} and λ_{2n+1} . One may show (by oscillation theorems; cf., Magnus and Winkler, 1979) that $\lambda_{2n} > \lambda_{2n-1}$, and that $\lambda_{2n+1} \geq \lambda_{2n}$. Moreover, one stationary point of a_r occurs in each unstable band, and none occur elsewhere. Unstable bands that consist of a single point (if $\lambda_{2n} = \lambda_{2n+1}$) are called "degenerate".

Fig. 7

Following Dubrovin and Novikov (1974), we now assume that $u(x)$ generates $a_r(x_0, \lambda)$ with only a finite number (N) of non-degenerate bands. (Roughly, this corresponds to approximating a periodic function with a finite Fourier series.)

Next, we define two spectra for (4.6). These may be defined either by attaching boundary conditions to (4.6), or by imposing conditions on the monodromy matrix.

i) The main spectrum is defined by (4.15) or by requiring that (4.6) admit periodic solutions. Points of this spectrum ($\lambda_1, \lambda_2, \dots$) occur at the edges of the unstable bands. Note that periodic solutions of (4.6) are independent of x_0 .

ii) The auxiliary spectrum ($\gamma_1, \gamma_2, \dots$) is defined by

$$a_i + b_i = 1, \quad (4.16)$$

or by requiring that there be a solution of (4.6) that satisfies

$$y(x_0; x_0, \lambda) = 0 = y_0(x_0 + L; x_0, \lambda). \quad (4.17)$$

It follows from (4.10) and (4.16) that $a_r^2 \geq 1$ at each of these points, which must therefore lie in the unstable bands. One may show that each unstable band contains exactly one point of the auxiliary spectrum. By assumption, at most N of these points of the auxiliary spectrum do not coincide with points of the main spectrum.

Given $u(x)$ and x_0 , determination of $\{\lambda_j\}$ and $\{\gamma_j\}$ completes the direct scattering problem. The inverse mapping requires knowledge of the analytic properties of the monodromy matrix, as one might expect by analogy with the inverse problem on the infinite interval. The final result is miraculously simple:

$$u(x_0) = - \frac{2N+1}{\prod_{j=1}^N \lambda_j} + \frac{N}{\prod_{j=1}^N \gamma_j}. \quad (4.17)$$

Now it remains only to find how the two spectra depend on x_0 , and on time.

The boundary conditions that define $\{\lambda_j\}$ and $\{\gamma_j\}$ suggest that the main spectrum should be independent of x_0 , but that the auxiliary spectrum should not be. This turns out to be the case. One shows that

$$\frac{\partial \lambda_j}{\partial x_0} = 0, \quad (4.18a)$$

$$\frac{\partial \gamma_j}{\partial x_0} = \frac{[2i \sigma_j R^{1/2}(\gamma_j)]}{\left[\prod_{k \neq j}^N (\gamma_j - \gamma_k) \right]}, \quad j = 1, \dots, N \quad (4.18b)$$

where $\sigma_j = \pm 1$ and

$$R(\lambda) = \prod_{j=1}^{2N+1} (\lambda - \gamma_j) . \quad (4.18c)$$

The x_0 -dependence of $\{\gamma_j\}$ is as frightening as (4.17) was appealing. One is faced in (4.18b) with the integration of N coupled, nonlinear, ordinary differential equations, even before time-dependence is brought into the picture! Fortunately, another miracle occurs, and there is a change of variables (involving hyperelliptic functions) that permits one to integrate (4.18b) by quadrature. Thus although (4.18b) looks foreboding, it represents a straight-line motion when viewed appropriately. The cost of acquiring this simple picture is that one must introduce hyperelliptic functions.

The time-dependence of the two spectra parallels their x_0 -dependence. For the KdV equation,

$$\frac{\partial \gamma_j}{\partial t} = 0 , \quad (4.19a)$$

$$\frac{\partial \gamma_j}{\partial t} = 8i \hat{\sigma}_j \left[\prod_{k \neq j}^N (\gamma_j - \gamma_k) \right]^{-1} \left[\prod_{k \neq j}^N \gamma_k - \frac{1}{2} \sum_k^{2N+1} \lambda_k \right] R^{\frac{1}{2}}(\gamma_j) ,$$

$$j = 1, \dots, N , \quad (4.19b)$$

where $\hat{\sigma}_j = \pm 1$, and $R(\lambda)$ was defined by (4.18c). Again, the integration in (4.19b) may be reduced to quadrature by introducing hyperelliptic functions.

What are the consequences of introducing hyperelliptic functions? One may show that for an N -band solution of the KdV equation (i.e., whose auxiliary spectrum contains only N nondegenerate points),

$$u(x, t) = 2 \frac{\partial^2}{\partial x^2} \ln \theta_N(\eta_1, \eta_2, \dots, \eta_N) + \text{constant} , \quad (4.20a)$$

where

$$\eta_j = i(k_j x - \omega_j t), \quad (4.20b)$$

each k_j and ω_j is constant, and θ_N is the theta function, an analytic function that is periodic in each of its N variables separately (cf., Siegel, 1971, Vol. II). It follows that the motion described by an N -band solution of the KdV equation may be thought of as uniform translation of a point along a straight-line path on an N -dimensional torus. Because any such solution has only N degrees of freedom, rather than infinitely many as (4.1) suggests a priori, it has a relatively short recurrence time that may be estimated from a knowledge of $\{\omega_j\}_{j=1}^N$ in (4.20b). This is apparently the explanation of the relatively short recurrence time observed numerically by Zabusky and Kruskal (1965). [Recently, Thyagaraja (1979) proposed another explanation of the short recurrence times in the periodic KdV problem. His method is based on counting the number of Fourier modes, rather than the number of unstable bands, required to describe a KdV solution. His method provides a rigorous upper bound on the recurrence time, but it seems not to give an accurate estimate of the true recurrence time.]

In an abstract sense, this theory of the periodic KdV problem (or its generalization by McKean and Trubowitz, 1976) is complete. In a practical sense, it is difficult to extract from the theory the numbers that one needs to make comparisons with experiments. [The difficulty here is simply that hyperelliptic functions have not yet been reduced to a practical engineering tool.] Here are some of the practical problems for which satisfactory answers are not yet available.

- (1) Given an N -band solution of (4.1), what is its recurrence time?
- (2) Given appropriate initial data for (4.1), estimate its recurrence time.
- (3) Zabusky (1969) defined T_b to be the time of breakdown of

the solution of

$$u_t + uu_x = 0 \quad (4.21)$$

that evolves from $u(x, 0) = \sin(2\pi x/L)$. He found empirically that the recurrence time of the corresponding solution of

$$u_t + uu_x + \delta^2 u_{xxx} = 0 \quad (4.22)$$

was

$$T_r = (0.71/\delta) T_b. \quad (4.23)$$

Can this formula be derived theoretically?

- (4) "Solitons" have come to mean special solutions of problems solvable by IST on the infinite interval. Each soliton is associated with a discrete eigenvalue in that formulation. However, the word was coined by Zabusky and Kruskal (1965) to describe phenomena observed in the periodic problem. What were the "solitons" they observed? How does the band structure of a solution of the periodic problem relate to the number of solitons that one would observe numerically?

Part V
Deterministic and Chaotic Models

In the course of these lectures, I have tried to survey the theory of nonlinear evolution equations ("soliton theories") that can be solved by IST. If one were asked to identify the single feature of these equations that identifies them as special, one might choose the predictability of their solutions as their common identifying feature. This property may be stated in several ways.

- i) These equations are deterministic, not only for some finite time, but for all time.
- ii) Their solutions are (neutrally) stable with respect to perturbations in the initial data.
- iii) Their solutions are predictable.

All of these notions are closely related to each other, and they amount to saying that for these special equations, the initial-value problem makes sense in a practical, real-world way. Because of their predictability, these completely integrable equations may be regarded as prototypes of deterministic problems.

As models of physical systems, they contrast sharply with stochastic models, which require a certain amount of unpredictability. This unpredictability may come from a variety of sources; one possibility is that the dynamics themselves may be unpredictable. An example of this inherent unpredictability may be seen in a system of coupled ordinary differential equations known as the "Lorenz model":

$$\begin{aligned}\dot{x} &= \sigma(y - x) , \\ \dot{y} + y &= -x(z - r) , \\ \dot{z} &= xy - bz ,\end{aligned}\tag{5.1}$$

where $(\dot{}) = d()/dt$, and (σ, r, b) are parameters. In their original context these equations modelled a problem in convection, and σ represented the Prandtl number, r the Rayleigh number, and b was a measure of a length scale. For this discussion, however, we may regard them simply as a dynamical system with parameters.

Lorenz (1963) chose certain values of the parameters ($\sigma = 10$, $b = 8/3$, $r = 28$), and integrated the equation numerically. His results are described in detail in his original work, and may be summarized by saying that the solutions he computed were unpredictable. These equations have become a popular model of a dynamical system exhibiting chaotic behavior. Based on a great deal of numerical work following that of Lorenz, the current consensus is that the solutions of these equations are ergodic, although no proof has been found. [The original notion of an ergodic trajectory, due to Boltzmann, was that in the course of time it should wander everywhere in the available phase space. Some technical revisions have been required since then, but the original notion still is close to being correct. For a more complete account, see Arnold and Avez (1968) or Arnold (1978).]

We may regard the Lorenz model with the particular parameters chosen by Lorenz as a prototype of a "chaotic model". Its features may be characterized in the following ways.

- i) The equations are deterministic for a finite time (i.e., given finite initial data, a unique solution exists). However, the solution becomes less and less determined by the initial data as time increases without bound.
- ii) The solution is unstable with respect to perturbations of the initial data.
- iii) Given any initial data, the solution is unpredictable over a long time scale in any practical sense.

Now we have two kinds of dynamical systems. The Lorenz model is chaotic, and one might imagine using statistical methods to analyze the problem further. Soliton theories are deterministic, and statistical methods would only obscure matters there. Thus, a fundamental question is "How does one know whether a given model will be chaotic or deterministic?"

This same question arises in the foundations of statistical mechanics, and of the theory of turbulence. Perhaps the first context in which the question arose was theology. In that context, deterministic theory was called "predestination", while evidence of unpredictable behavior was attributed to "free will". A corresponding debate, with somewhat different names, is now in progress in psychology.

No attempt will be made here to classify all dynamical systems on the basis of how chaotic their solutions are. We will examine a narrower question, which may be regarded as a first step in constructing such a grand classification scheme.

Q: What determines whether a given partial differential equation can be solved by IST?

This question has been the motivation of recent work by Ablowitz, Ramani and Segur (1978, 1980a,b). The answer seems to be related to what we may call the "Painlevé property" (which will be defined shortly).

After this rather long introduction, we may finally outline the lecture. First, we must define the Painlevé property for ordinary differential equations (ODEs), because everything else follows from it. Next we may show that ODEs with the Painlevé property are related to evolution equations solvable by IST. Our conjecture about characterizing these nonlinear partial differential equations (PDEs) then is almost obvious: they must reduce to ODEs of Painlevé type. What is less obvious is how to prove the conjecture, but a partial proof is available. The notion that ODEs of P-type are closely

related to IST and complete integrability may be used in a variety of ways. We will look at two: (i) to test whether a given PDE can be solved by IST; and (ii) to find conditions under which the Lorenz model is completely integrable.

1.

The Painlevé Property

Consider first a linear ordinary differential equation, say of second order:

$$\frac{d^2w}{dz^2} + p(z) \frac{dw}{dz} + q(z)w = 0. \quad (5.2)$$

For suitable $p(z)$ and $q(z)$, this equation may be viewed in the complex plane, and the singularities of the solution of (5.2) are found by examining $p(z)$ and $q(z)$ (e.g., Ince, 1956, Ch. 15). In particular, the general solution has two constants of integration,

$$w(z; A, B) = Aw_1(z) + Bw_2(z), \quad (5.3)$$

and the location in the complex plane of the singularities of $w(z)$ do not depend on A or B . The singularities of a linear differential equation are said to be fixed, because they do not depend on the constants of integration.

Nonlinear differential equations lose this property. A very simple example of a nonlinear ODE is

$$\frac{dw}{dz} + w^2 = 0, \quad (5.4)$$

its general solution is

$$w(z; z_0) = \frac{1}{z - z_0}. \quad (5.5)$$

Here z_0 is the constant of integration, and it also defines the location of the singularity. This singularity is movable, because its

location depends on the constant of integration.

So linear differential equations have only fixed singularities, nonlinear equations can have both fixed and movable singularities. About 100 years ago, mathematicians asked the following question:

Q: Which nonlinear ODEs admit no movable branch points or essential singularities?

Movable poles are allowed, as are fixed singularities of any kind. We will refer to this property as the Painlevé-property, and equations that possess it will be said to be of Painlevé-type, or P-type.

It turns out that the only first order equations with the Painlevé-property are generalized Riccati equations

$$\frac{dw}{dz} = p_0(z) + p_1(z)w + p_2(z)w^2. \quad (5.6)$$

(A complete review of the nineteenth century work in this field may be found in Ince, 1956, Ch. 12-14.)

Painlevé and his coworkers were able to answer the question comprehensively for second-order equations of the form

$$\frac{d^2w}{dz^2} = F\left(\frac{dw}{dz}, w, z\right), \quad (5.7)$$

where F is rational in dw/dz and w , and analytic in z . They showed that out of all possible equations of the form (5.7) only 50 canonical equations have the Painlevé property of no movable branch points or essential singularities. Further, they showed that 44 of these equations can be reduced to something already known, such as elliptic functions. That left six equations that defined new transcendental functions, called the Painlevé transcendents. The first three of these are:

$$\frac{d^2w}{dz^2} = 6w^2 + z \quad P_I$$

$$\frac{d^2w}{dz^2} = 2w^3 + zw + \alpha \quad P_{II}$$

$$\frac{d^2w}{dz^2} = \frac{1}{w} \left(\frac{dw}{dz} \right)^2 - \frac{1}{z} \left(\frac{dw}{dz} \right) + \frac{1}{z} (\alpha w^2 + \beta) + \gamma w^3 + \frac{\delta}{w} \quad P_{III}$$

There are three more.

The question of which equations have the Painlevé-property is appropriate at any order, but comprehensive results are available only at the first and second order.

2.

Relation to IST

A relation between ODEs of P-type and the inverse scattering transform is formulated in the following:

Conjecture (Ablowitz, Ramani, and Segur, 1978)

Every nonlinear ODE obtained by an exact reduction of a nonlinear PDE solvable by some inverse scattering transform has the Painlevé-property.

Here are some examples. The Boussinesq equation

$$u_{tt} = u_{xx} + \left(\frac{u^2}{2} \right)_{xx} + \frac{1}{4} u_{xxxx} \quad (5.8)$$

is a nonlinear PDE solvable by IST (Zakharov, 1973). An exact reduction to an ODE may be obtained by looking for a traveling wave solution:

$$u(x, t) = w(x - ct) = w(z) \quad (5.9)$$

Then (5.8) becomes

$$(1 - c^2) w'' + \left(\frac{w^2}{2} \right)'' + \frac{1}{4} w'''' = 0, \quad (5.10)$$

which can be integrated twice. Depending on the constants of integration, the result after rescaling is either

$$w'' + 2w^2 + a = 0 \quad \text{or} \quad w'' + 2w^2 + z = 0. \quad (5.11)$$

The first possibility defines an elliptic function, whose only singularities are poles. The second possibility is the equation for P_I . In either case, the ODE has the Painlevé property. So the PDE solvable by inverse scattering reduces to an ODE of P-type.

Another example is the modified KdV equation

$$u_t - 6u^2 u_x + u_{xxx} = 0, \quad (5.12)$$

which can be solved by IST (Wadati, 1972). An exact reduction to an ODE may be obtained by looking for a self-similar solution:

$$\begin{aligned} u(x, t) &= (3t)^{-2/3} w(z); & z &= x/(3t)^{1/3}, \\ \Rightarrow w''' - 6w^2 w' - (zw)' &= 0. \end{aligned} \quad (5.13)$$

This can be integrated once

$$w'' = 2w^3 + zw + \alpha. \quad P_{II}$$

Again, the ODE is of P-type.

The sine-Gordon equation

$$u_{xt} = \sin u \quad (5.14)$$

can be solved by IST (Ablowitz, Kaup, Newell and Segur, 1973). It has a self-similar solution

$$u(x, t) = f(z), \quad z = xt \quad (5.15)$$

If we set $w(z) = \exp(if)$, then

$$w'' = \frac{1}{w} (w')^2 - \frac{1}{z} (w') + \frac{1}{2z} (w^2 - 1) \quad P_{III}$$

Again, the ODE is of P-type.

By now the pattern is evident, and I may simply state that there is a nonlinear evolution equation solvable by IST that reduces to P_{IV} , and another that reduces to P_V . The point is not that the evolution equation must reduce to one of the six Painlevé transcendents, which are all of second order, but that it must reduce to an ODE of P-type. We have checked an enormous number of examples. In every case checked, PDEs that can be solved by IST reduce to ODEs of P-type and PDEs that are not solvable by IST (e.g., this may be determined by observing numerically that two solitary waves do not interact like solitons) reduce to ODEs that are not of P-type.

So there is some kind of relation between partial differential equations solvable by IST and ordinary differential equations of P-type. This relation can be used to examine either the ODEs or the PDEs. To see how it helps in the study of the ODEs, consider the mKdV equation and P_{III} . Recall that the last step of IST, the inverse scattering part, goes as follows. $F(x, t)$ satisfies a linear partial differential equation

$$F_t + F_{xxx} = 0 \quad (5.16)$$

subject to some boundary and initial conditions. Then $K(x, y; t)$ satisfies a linear integral equation of the Gel'fand-Levitan-Marchenko type,

$$K(x, y) = F(x+y) + \int_x^\infty \int_x^\infty K(x, z) F(z+s) F(s+y) dz ds \quad y \geq x \quad (5.17)$$

Once K is known, then $q(x, t) = K(x, x; t)$ satisfies mKdV:

$$q_t - 6q^2 q_x + q_{xxx} = 0. \quad (5.18)$$

In the full IST treatment, F depends on the initial data of $q(x, 0)$ through the direct scattering problem. Here we simply start

with F , and force everything to be self-similar:

$$\begin{aligned} \xi &= x/(3t)^{1/3}, & \eta &= y/(3t)^{1/3}, \\ F(x, t) &= (3t)^{-1/3} F(\xi), & K(x, y; t) &= (3t)^{-1/3} K(\xi, \eta). \end{aligned} \quad (5.19)$$

Then (5.16) becomes a linear ODE

$$F'''(\xi) - (\xi F)' = 0 \quad (5.20)$$

and a one-parameter family of solutions is

$$F(\xi + \eta) = r \operatorname{Ai}\left(\frac{\xi + \eta}{2}\right), \quad (5.21)$$

where $\operatorname{Ai}(\xi)$ is the Airy function. The integral equation (5.17) becomes:

$$K(\xi, \eta) = r \operatorname{Ai}\left(\frac{\xi + \eta}{2}\right) + \frac{r^2}{4} \int_{\xi}^{\infty} \int_{\xi}^{\infty} K(\xi, \zeta) \operatorname{Ai}\left(\frac{\zeta + \theta}{2}\right) \operatorname{Ai}\left(\frac{\theta + \eta}{2}\right) d\zeta d\theta \quad \eta \geq \xi \quad (5.22)$$

The Airy function decreases rapidly as its argument becomes large, so the integral term in (5.22) is very well-behaved. Therefore, it is relatively easy to solve (5.22) for $\eta \geq \xi$. On $\eta = \xi$, the solution of (5.22) satisfies the self-similar form of mKdV, viz., P_{II} :

$$\frac{d^2}{d\xi^2} K(\xi, \xi) = 2K^3(\xi, \xi) + \xi K(\xi, \xi). \quad (5.23)$$

(Two different proofs of this fact are given by Ablowitz, Ramani and Segur, 1978, 1980a.) The point here is that (5.22) is an exact linearization of P_{II} : every solution of the linear integral equation also solves P_{II} . The general solution of (5.23) involves two arbitrary constants; the linear integral equation gives a one parameter

(r) family, which includes all of the bounded real solutions of (5.23).

Next let us sketch a partial proof of why this test actually works. Consider a linear equation of the form

$$K(x, y) = F(x+y) + \int_x^\infty K(x, z) N(x; z, y) dz, \quad y \geq x, \quad (5.24)$$

where F vanishes rapidly for large values of the argument and N depends on F . For example, in (2.16), $N(x, z, y) = F(z+y)$. In (5.22), we had

$$N(x, z, y) = \int_x^\infty F(z+s) F(s+y) ds. \quad (5.25)$$

Other choices are also possible. We want to show that every solution of a linear integral equation like (5.24) must have the Painlevé property. Then if K also satisfies an ODE, the family of solutions of the ODE obtained via (5.24) necessarily has the Painlevé property as well. So the Painlevé property is not out of the blue, it is a consequence of the linear integral equation.

Very roughly, the proof goes like this (for details, see Ablowitz, Ramani and Segur, 1980a; also McLeod and Olver, 1980).

- i) F satisfies a linear ODE, and therefore has no movable singularities at all.
- ii) If F vanishes rapidly enough, then the Fredholm theory of integral equations applies. It follows that (5.24) has a unique solution in the form:

$$K(x, y) = F(x+y) + \int_x^\infty F(x+z) \frac{D_1(x, z, y)}{D_2(x)} dz. \quad (5.26)$$

where D_1 and D_2 are entire functions of their arguments. Then the

singularities of K can only come from the fixed singularities of F , or the movable zeros of D_2 . But D_2 is analytic, so these movable singularities must be poles.

3.

Applications

Here are two examples of how the conjecture may be used.

Example:

In (1+1) dimensions, the nonlinear Schrödinger equation is

$$iu_t = u_{xx} + a|u|^2 u. \quad (5.27)$$

It can be solved by IST (Zakharov and Shabat, 1972). A natural generalization to (2+1) dimensions is

$$iu_t = \nabla^2 u + a|u|^2 u. \quad (5.28)$$

We claim this equation cannot be solved by IST, because (5.28) has a similarity solution in the form

$$u(x, y, t) = R(\sqrt{x^2 + y^2}; \lambda) \exp(i\lambda t), \quad (5.29)$$

and the ODE for $R(r)$ is not of P-type. So the nonlinear Schrödinger equation is solvable in (1+1) dimensions, but not in (2+1) dimensions, or in (3+1) dimensions. On the same grounds, we claim that the equation for water waves in deep water,

$$iu_t + u_{xx} - u_{yy} + \sigma|u|^2 u = 0, \quad (5.30)$$

cannot be solved by IST.

Example:

If the Painlevé property is as closely tied to complete integrability as we have claimed, it ought to identify values of the parameters for which the Lorenz model (5.1) is completely integrable.

Thus, we may ask whether the Lorenz model is ever of P-type. The answer is that there are exactly four choices of (σ, r, b) for which (5.1) is of P-type.

i) $\sigma = 0$. In this case the equations are effectively linear, and therefore deterministic. Certainly the solutions exhibit no chaotic behavior.

ii) $\sigma = 1/2$, $b = 1$, $r = 0$. The equations have two exact integrals:

$$y^2 + z^2 = A^2 \exp(-2t), \quad (5.31)$$

$$x^2 - z = B \exp(-t), \quad (5.32)$$

after which the third integration may be obtained by quadrature, or the solution may be expressed in terms of elliptic functions. The solutions may be considered generalizations of the periodic orbits identified by Lorenz.

iii) $\sigma = 1$, $b = 2$, $r = 1/9$. A first integral is

$$x^2 - 2z = C \exp(-2t). \quad (5.33)$$

After an involved change of variables, the resulting second-order equation becomes P_{II} . Again, the problem is deterministic.

iv) $\sigma = 1/3$, $b = 0$, r arbitrary. (The analysis of this case is due to A. Ramani.) We may write $y = 3\dot{x} + x$ from (5.1a), and replace (5.1) with a third-order equation for x . It has a first integral:

$$\ddot{x}x - \dot{x}^2 + \frac{x^4}{4} = C \exp(-4/3 t). \quad (5.34)$$

With the substitution,

$$T = \exp(-t/3), \quad x(t) = TW(T), \quad (5.35)$$

(5.34) reduces to P_{III} with $\alpha = \beta = 0$. Thus the Lorenz model (5.1) has at least one integral, and reduces to a classically known equation of lower order, whenever the coefficients (σ, r, b) are chosen so that (5.1) is of P-type.

These isolated points of parameter space are embedded in larger regions in which the equations have first integrals, although they may not be completely integrable.

v) If $b = 1, r = 0$, then (5.31) obtains for any σ . The existence of this integral precludes ergodic trajectories.

vi) If $b = 2\sigma$, then for any (r, σ) ,

$$x^2 - 2\sigma z = C \exp(-2\sigma t). \quad (5.36)$$

Again, ergodic trajectories are impossible.

These different possibilities are shown in a map of parameter space in Figure 8. The point here is that although the Lorenz model may be chaotic for some values of the parameters, it is completely integrable for others, and looking for the Painlevé property provides an effective means of identifying points of deterministic behavior. Of course, this notion is not restricted to the Lorenz model.

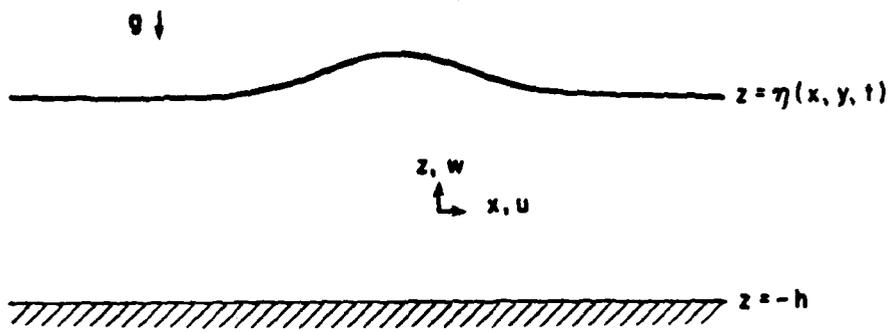
Fig.8

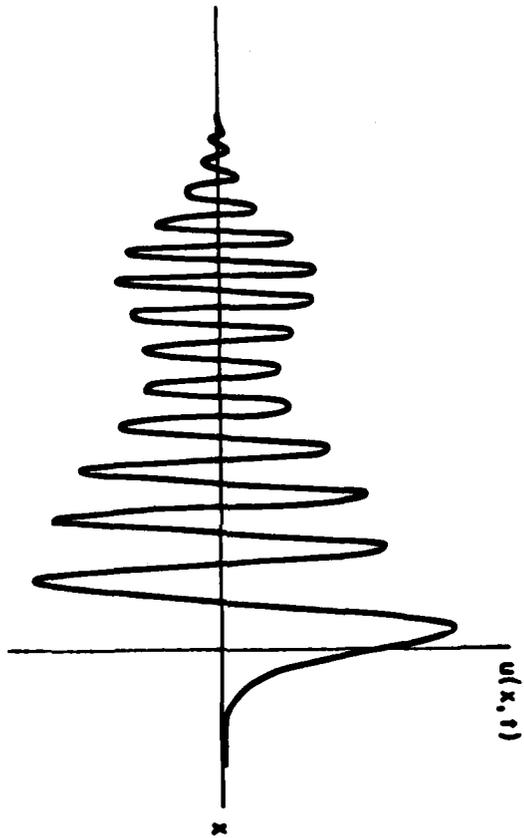
Acknowledgements

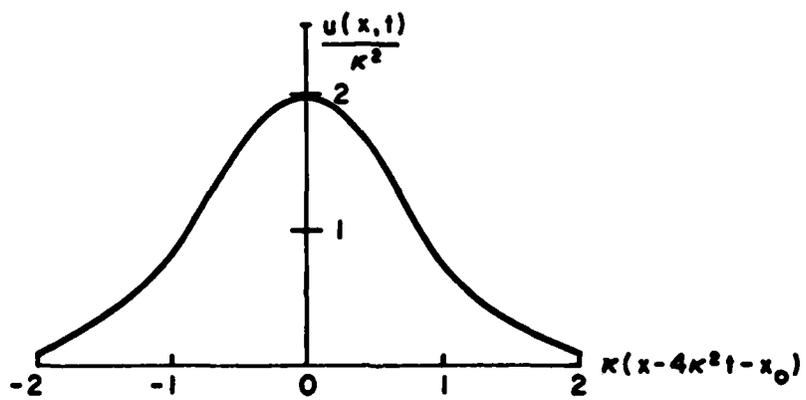
This work was supported by the U.S. Office of Naval Research and by the U.S. Army Research Office.

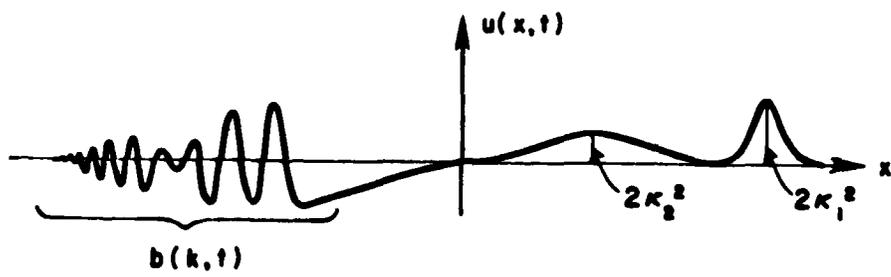
Figure Captions

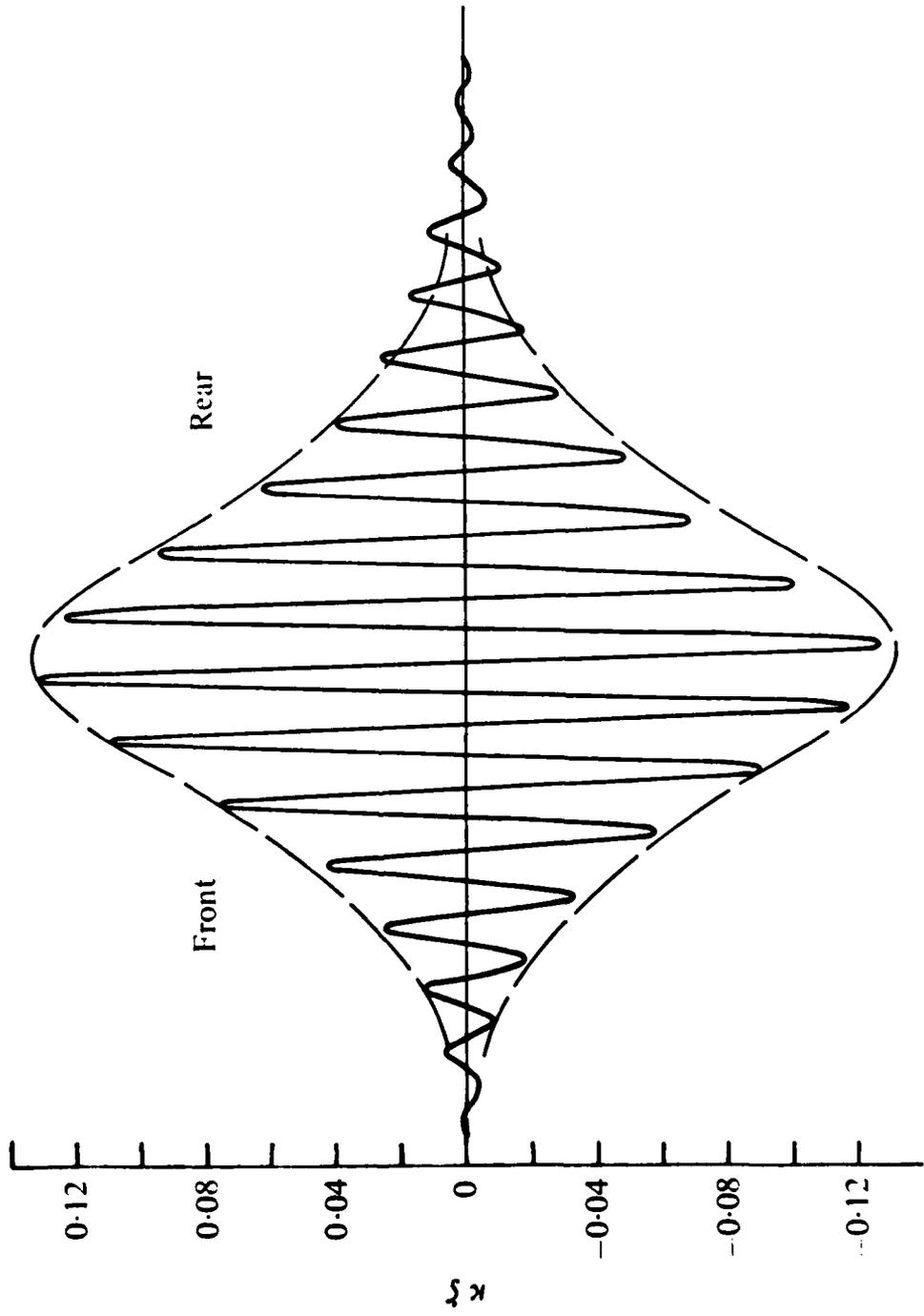
1. Physical configuration, showing notation for (1.9).
2. Typical long-time solution of (2.1), with the wave-train spreading to the left. Modulations of the wave-train are determined by the initial data.
3. Soliton solution of the KdV equation (2.7).
4. Typical long-time KdV solution. The solitons are determined by the discrete spectrum [κ_1^2 and κ_2^2 in this case], while the radiation is associated with the continuous spectrum [$b(k)$].
5. Measured packet of monochromatic (frequency = 1 Hz) surface waves of small amplitude. The theoretical shape of the appropriate envelope soliton solution of (3.16) is given by the dashed line. From Ablowitz and Segur (1979).
6. Cnoidal wave solution of the KdV equation (4.1) with $k = 1/2$ and $\oint u dx = 0$.
7. Floquet diagram for a particular $u(x)$ in (4.7). In this example there are four unstable bands.
8. Map of parameter space for Lorenz model (5.1), showing where equations have Painlevé property ($P_1 - P_4$) and where they admit exact integrals.

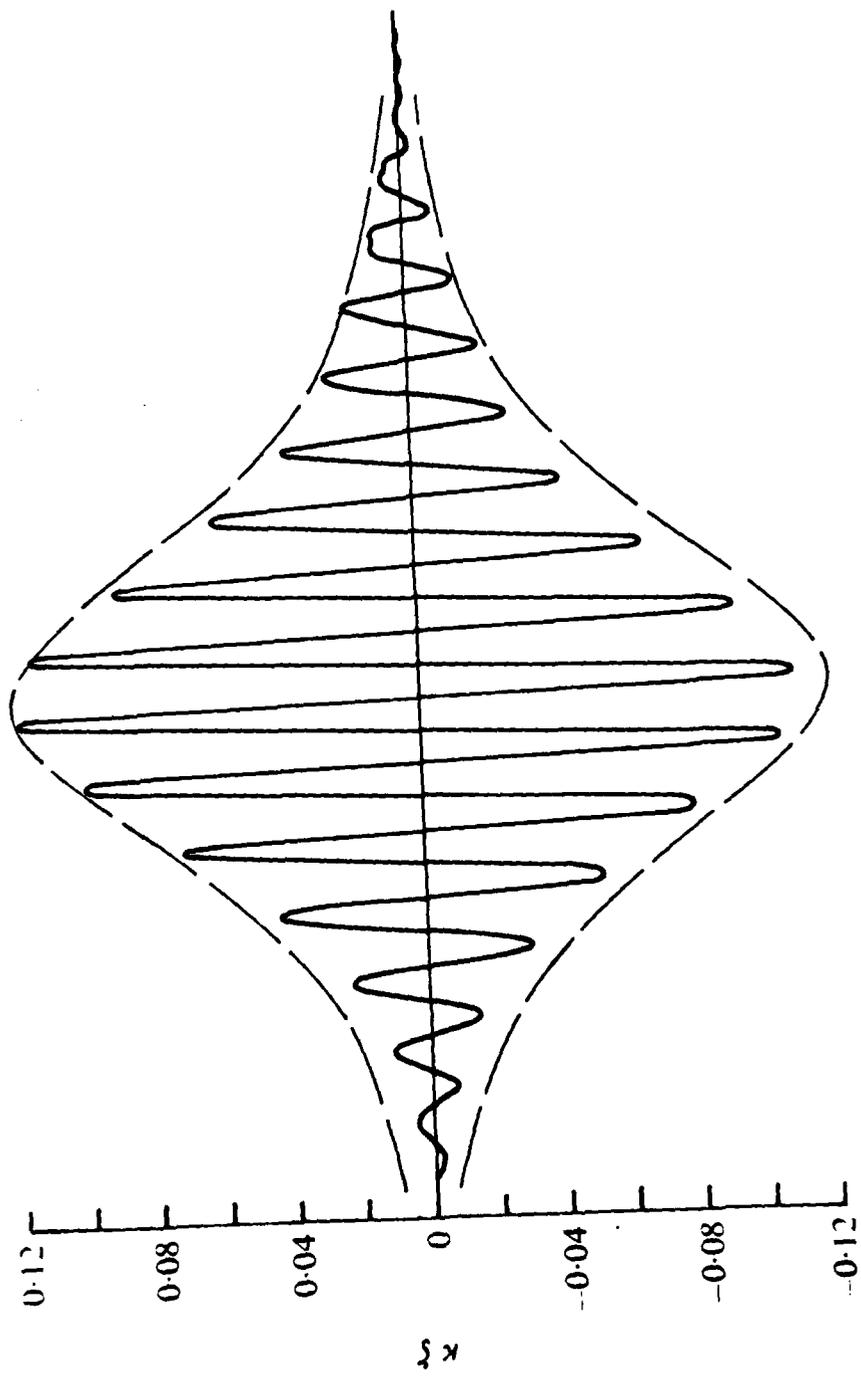


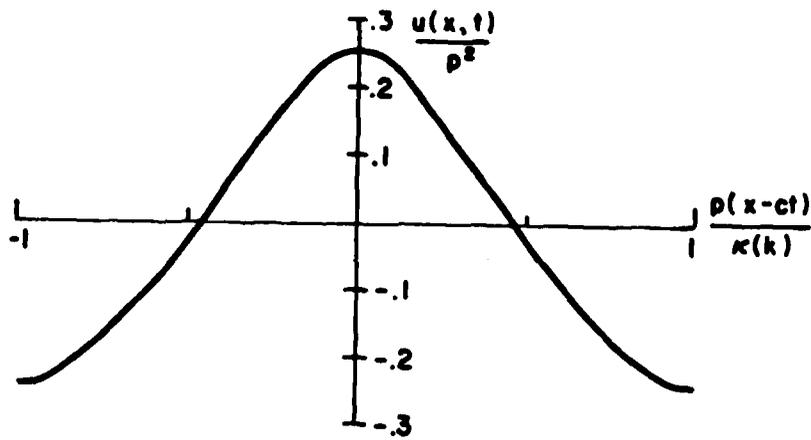


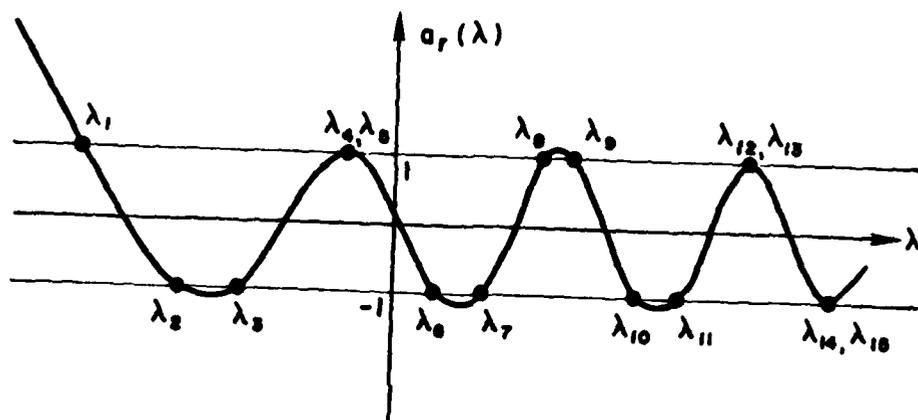


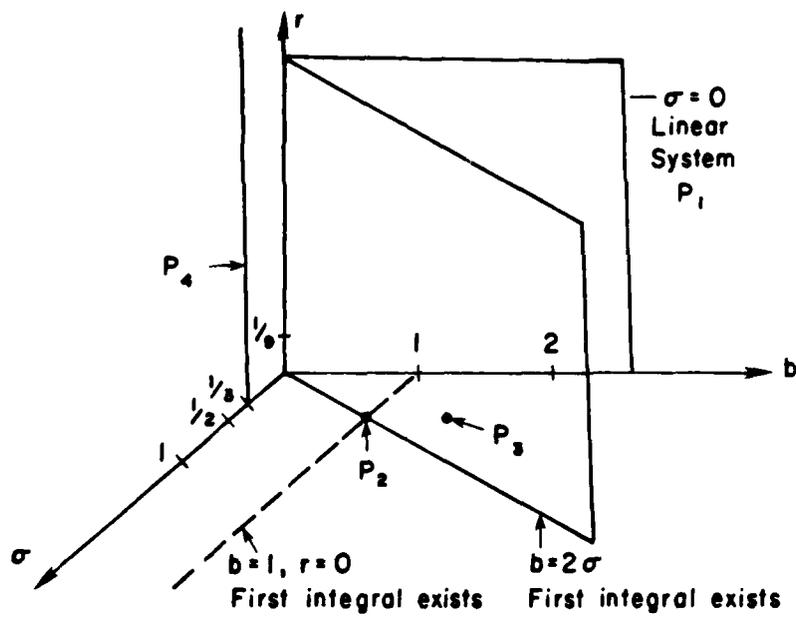












References

General References

By the time the Proceedings of this school appear in print, a number of books on the theory of equations with solitons also will have appeared in print, including:

G. Lamb, Elements of Soliton Theory, Wiley, 1980.

V. E. Zakharov, S. V. Manakov, S. P. Novikov, et al., (title unknown, in Russian), 1980.

M. J. Ablowitz and H. Segur, Solitons and the Inverse Scattering Transform, SIAM, 1981.

Most of the material in these lectures was drawn from this last reference.

I. Physical Meaning of Equations with Solitons

Benjamin, T. B., J. Fluid Mech., vol. 25, p. 241, 1966.

Gibbon, J. D., I. N. James and I. M. Moroz, Proc. Roy. Soc. London A, vol. 367, p. 219, 1979.

Hammack, J. L. and H. Segur, J. Fluid Mech., vol. 65, p. 289, 1974.

—————, J. Fluid Mech., vol. 84, p. 337, 1978.

Hasimoto, H. and H. Ono, J. Phys. Soc. Japan, vol. 33, p. 805, 1972.

Kadomtsev, B. B. and V. I. Petviashvili, Sov. Phys. Doklady, vol. 15, p. 539, 1970.

Korteweg, D. J. and G. deVries, Phil. Mag., ser. 5, vol. 39,
p. 422, 1895.

Lamb, H., Hydrodynamics, Dover, N.Y., 1932.

Ono, H., J. Phys. Soc. Japan, vol. 39, p. 1082, 1975.

Rayleigh, Lord (J. W. Strutt), Phil. Mag., ser. 5, vol. 1,
p. 257, 1876.

Stoker, J. J., Water Waves, Interscience, N.Y., 1957.

Wehausen, J. V. and E. V. Laitone, "Surface Waves" Handbuch der
Physik, vol. 9, Springer-Verlag, 1960.

Yuen, H. C. and B. M. Lake, Phys. Fluids, vol. 18, p. 956, 1975.

Zakharov, V. E., Sov. Phys. J. App. Mech. and Tech. Phys., vol. 4,
p. 190, 1968.

II. Introduction to the Inverse Scattering Transform

Ablowitz, M. J., D. J. Kaup, A. C. Newell and H. Segur, Stud.
App. Math., vol. 53, p. 249, 1974.

Calogero, F. and A. Degasperis, Nuovo Cimento, vol. 32 B, p. 201,
1976.

Deift, P. and E. Trubowitz, Comm. Pure and App. Math., vol. 32,
p. 121, 1979.

Gel'fand, I. M. and B. M. Levitan, Izv. Akad. Nauk SSSR, ser.
mat. 15, p. 309, 1951; translated in Amer. Math. Soc. Trans., ser. 2,
vol. 1, p. 253, 1955.

Gardner, C. S., J. M. Greene, M. D. Kruskal and R. M. Miura,
Phys. Rev. Lett., vol. 19, p. 1095, 1967; also Comm. Pure App. Math.,
vol. 27, p. 97, 1974.

Stoker, J. J., Water Waves, Interscience, N.Y., 1957.

Whitham, G. B., Linear and Nonlinear Waves, Wiley-Interscience, N.Y., 1974.

Zakharov, V. E. and A. B. Shabat, Sov. Phys. JETP, vol. 34, p. 62, 1972.

III. More Inverse Scattering on the Infinite Interval

Ablowitz, M. J., D. J. Kaup, A. C. Newell and H. Segur, Stud. App. Math., vol. 53, p. 249, 1974.

Ablowitz, M. J. and H. Segur, J. Math. Phys., vol. 17, p. 710, 1976.

—————, J. Fluid Mech., vol. 92, p. 691, 1979.

Arnold, V. I., Mathematical Methods in Classical Mechanics, Benjamin, N.Y., 1978.

Benjamin, T. B. and J. E. Feir, J. Fluid Mech., vol. 27, p. 417, 1967.

Gel'fand, I. M. and B. M. Levitan, Izv. Akad. Nauk SSSR, ser. mat. 15, p. 309, 1951; translated in Amer. Math. Soc. Trans., ser. 2, vol. 1, p. 253, 1955.

Goldstein, H., Classical Mechanics, Addison-Wesley, Reading, Mass., 1950.

Muskhelisvili, N. I., Singular Integral Equations, Noordhoff, Groningen, Holland, 1953.

Segur, H., J. Math. Phys., vol. 17, p. 714, 1976.

Zakharov, V. E. and L. D. Faddeev, Funct. Anal. App., vol. 5, p. 280, 1971.

Zakharov, V. E. and S. V. Manakov, Sov. Phys. JETP, vol. 44,
p. 106, 1976.

-----, Sov. Phys. Reviews, vol. 1, p. 133, 1979.

Zakharov, V. E. and A. B. Shabat, Sov. Phys. JETP, vol. 34, p. 62,
1972.

IV. The Korteweg-deVries Equation with Periodic Boundary Conditions

Byrd, P. F. and M. D. Friedman, Handbook of Elliptic Integrals,
Springer-Verlag, N.Y., 1971.

Dubrovin, B. A. and S. P. Novikov, Sov. Phys. JETP, vol. 40,
p. 1058, 1975.

Fermi, E., J. Pasta and S. Ulam, "Studies of Nonlinear Problems",
Los Alamos Rep't LA 1940, 1955; reprinted in The Collected Papers of
Enrico Fermi, U. of Chicago Press, 1965.

Gardner, C. S., J. M. Greene, M. D. Kruskal and R. M. Miura,
Phys. Rev. Lett., vol. 19, p. 1095, 1967; also Comm Pure App. Math.,
vol. 27, p. 97, 1974.

Korteweg, D. J. and G. deVries, Phil. Mag., ser. 5, vol. 39,
p. 422, 1895.

Magnus, W. and S. Winkler, Hill's Equation, Dover, N.Y., 1979.

McKean, H. P. and E. Trubowitz, Comm. Pure App. Math., vol. 29,
p. 143, 1976.

Miura, R. M., C. S. Gardner and M. D. Kruskal, J. Math. Phys.,
vol. 9, p. 1204, 1968.

Siegel, C. L., Topics in Complex Function Theory, vol. II,
Wiley-Interscience, N.Y., 1971.

Thyagaraja, A., Phys. Fluids, vol. 22, p. 2093, 1979.

Zabusky, N. J., J. Phys. Soc. Japan, vol. 26, supplement 196, 1969.

Zabusky, N. J. and M. D. Kruskal, Phys. Rev. Lett., vol. 15, p. 240, 1965.

V. Deterministic and Chaotic Models

Ablowitz, M. J., D. J. Kaup, A. C. Newell and H. Segur, Phys. Rev. Lett., vol. 30, p. 1262, 1973.

Ablowitz, M. J., A. Ramani and H. Segur, Lett. Nuovo Cim., vol. 23, p. 333, 1978.

—————, J. Math. Phys., vol. 21, p. 715, 1980.

—————, J. Math. Phys., vol. 21, p. 1006, 1980.

Arnold, V. I., Mathematical Methods in Classical Mechanics, Benjamin, N.Y., 1978.

Arnold, V. I. and A. A. Avez, Ergodic Problems in Classical Mechanics, Benjamin, N.Y., 1968.

Ince, F. L., Ordinary Differential Equations, Dover, N.Y., 1956.

Lorenz, E. N., J. Atm. Sci., vol. 20, p. 130, 1963.

McLeod, J. B. and P. J. Olver, "The Connection between Completely Integrable Partial Differential Equations and Ordinary Differential Equations of Painlevé-Type", preprint.

Wadati, M., J. Phys. Soc. Japan, vol. 32, p. 1681, 1972.

Zakharov, V. E., Sov. Phys. JETP, vol. 38, p. 108, 1974.

Zakharov, V. E. and A. B. Shabat, Sov. Phys. JETP, vol. 34, p. 62, 1972.