J. Fluid Mech. (1979), *vol.* 92, *part* 4, *pp.* 691–715
Printed in Great Britain

On the evolution of packets of water waves

By MARK J. ABLOWITZ
Applied Mathematics Program, Princeton University,
Princeton, New Jersey 08540†

AND HARVEY SEGUR
Aeronautical Research Associates of Princeton, Inc., 50 Washington Road,
P.O. Box 2229, Princeton, New Jersey 08540

(Received 15 May 1978 and in revised form 29 September 1978)

We consider the evolution of packets of water waves that travel predominantly in one direction, but in which the wave amplitudes are modulated slowly in both horizontal directions. Two separate models are discussed, depending on whether or not the waves are long in comparison with the fluid depth. These models are two-dimensional generalizations of the Korteweg–de Vries equation (for long waves) and the cubic nonlinear Schrödinger equation (for short waves). In either case, we find that the two-dimensional evolution of the wave packets depends fundamentally on the dimensionless surface tension and fluid depth. In particular, for the long waves, one-dimensional (KdV) solitons become unstable with respect to even longer transverse perturbations when the surface-tension parameter becomes large enough, i.e. in very thin sheets of water. Two-dimensional long waves ('lumps') that decay algebraically in all horizontal directions and interact like solitons exist only when the one-dimensional solitons are found to be unstable.

The most dramatic consequence of surface tension and depth, however, occurs for capillary-type waves in sufficiently deep water. Here a packet of waves that are everywhere small (but not infinitesimal) and modulated in both horizontal dimensions can 'focus' in a finite time, producing a region in which the wave amplitudes are finite. This nonlinear instability should be stronger and more apparent than the linear instabilities examined to date; it should be readily observable.

Another feature of the evolution of short wave packets in two dimensions is that all one-dimensional solitons are unstable with respect to long transverse perturbations. Finally, we identify some exact similarity solutions to the evolution equations.

1. Introduction

Our understanding of the evolution of surface water waves of moderate amplitude has increased significantly within the last decade or so. The evolution in one spatial dimension of a packet of inviscid waves of sufficiently small amplitude is governed by linear equations on a short time scale, and by either the Korteweg–de Vries (KdV) equation

$$u_t + uu_x + u_{xxx} = 0 \tag{1.1}$$

† Permanent address: Mathematics Department, Clarkson College, Potsdam, New York.

or the cubic nonlinear Schrödinger equation

$$iA_t + A_{xx} + \sigma|A|^2A = 0 \tag{1.2}$$

on longer time scales, depending on whether or not the typical wavelengths are large in comparison with the fluid depth. In (1.2) and throughout this paper, $\sigma = \pm 1$, and represents an irreducible choice of signs. Both of these equations can be solved exactly as initial value problems, using inverse scattering transforms (IST; an account of IST can be found in Ablowitz *et al.* 1978). In situations in which viscous effects are felt on an even longer time scale, these theories (or viscously corrected versions of them) predict with very reasonable accuracy the evolution of waves over quite long distances in wave tanks (Hammack & Segur 1974, 1978; Yuen & Lake 1975).

Outside of specially designed tanks, surface waves ordinarily evolve in two spatial dimensions and here the theory is much less complete. A two-dimensional generalization of (1.1) for nearly one-dimensional long waves was given by Kadomtsev & Petviashvili (1970) in the form:

$$u_t + uu_x + \sigma u_{xxx} - \int_x^\infty u_{yy}\, d\tilde{x} = 0. \tag{1.3}$$

Results by several authors indicate that (1.3) is of IST-type, but a complete method of inverse scattering, analogous to that in one spatial dimension, has not yet been developed.

Two-dimensional generalizations of (1.2) were derived by Zaharov (1968), Benney & Roskes (1969), Davey & Stewartson (1974), and Djordjevic & Redekopp (1977). All of these analyses followed approximately the same lines. The problem was also studied by Hayes (1973), using somewhat different methods. The most general analysis was by Djordjevic & Redekopp, who included the effects of gravity, surface tension and arbitrary depth to get a system that can be reduced to

$$\left.\begin{aligned} iA_t + \sigma_1 A_{xx} + A_{yy} &= \sigma_2|A|^2A + \Phi_x A, \\ a\Phi_{xx} + \Phi_{yy} &= -b(|A|^2)_{,c}, \end{aligned}\right\} \tag{1.4}$$

where $(a, b, \sigma_1, \sigma_2)$ depend on the (dimensionless) fluid depth and surface tension. In the long wave limit (1.4) reduces to one of the problems that Ablowitz & Haberman (1975) had shown were of IST-type. As with (1.3), beyond identifying the appropriate linear scattering problem and obtaining special solutions, no general inverse scattering theory has yet been developed.

In these two cases, (1.3) and the *long wave limit* of (1.4), one can reasonably anticipate that the necessary inverse scattering theory eventually will be developed, and that the general solutions of (1.3) and (1.4), as initial value problems, will become available. In these cases, the two-dimensional problem should eventually be solved to the extent that the one-dimensional problem is now. However, as discussed in § 5, we conjecture that (1.4) *cannot* be solved by inverse scattering transforms over the entire range of parameters and that the general two-dimensional problem cannot be solved in a manner analogous to that in one dimension.

The purpose of this paper is to identify some important results regarding (1.3) and (1.4), and to suggest the role that they play in the solution of initial value problems. A major result of this study is the dramatic effect that surface tension can have upon the dynamics of the wave motion. A summary of these results, and an outline of the paper is as follows.

The derivations of (1.3) and (1.4) from the physical problem of water waves are discussed in § 2. These equations are well established in the literature, but the question of what boundary conditions and other constraints are required to make the problems well posed is still open. We show that the original problem selects cer ain side conditions as 'natural'. Which conditions are appropriate depends on the dimensionless surface tension and depth. In this section we also consider the physical interpretation of an infinite set of conservation laws.

The role that one-dimensional soliton solutions can play in the two-dimensional problems is examined in § 3 (i.e. stability of solitons). KdV solitons are unstable in (1.3) when $\sigma = -1$, which occurs in sufficiently thin sheets of water (i.e. large enough surface-tension coefficient). For zero surface tension $\sigma = +1$, and the argument does not yield instability. When solitons are unstable, they cannot be viewed as the asymptotic ($t \rightarrow \infty$) states towards which the solution evolves, as they are in the one-dimensional problem. In this case, 'lump' solutions exist and may play an asymptotic role analogous to that of one-dimensional solitons.

Zakharov & Rubenchik (1974) showed that for the one-dimensional cubic nonlinear Schrödinger equation all one-dimensional solitons are unstable. These results apply to the deep water limit of (1.4). We extend their analysis to demonstrate the equivalent results in the case of finite depth.

The most dramatic effect of strong surface tension is *focusing* (§ 4). A wave that is large enough (in a certain integral sense) focuses at a particular point in space after a finite time. Here there is no asymptotic ($t \rightarrow \infty$) state, because the solution of (1.4) develops a singularity in a finite time. Focusing provides a mechanism by which a field of relatively small amplitude waves produces a local region in which the amplitudes are large. Focusing is a potentially important mechanism in the redistribution of energy within the spectrum; it should be readily measurable.

We consider the question of the complete integrability of (1.4) in § 5. Moreover, we exhibit some special solutions that are not one-dimensional, and are candidates for asymptotic states in the two-dimensional problem.

2. Relevant evolution equations

The classical problem of water waves is to find the irrotational motion of an inviscid, incompressible, homogeneous fluid, subject to the forces of gravity and surface tension. The fluid rests on a horizontal and impermeable bed of infinite extent at $z = -h$ (h may be finite or infinite), and has a free surface at $z = \zeta(x, y, t)$.

The fluid has a velocity potential, ϕ, which satisfies

$$\nabla^2\phi = 0, \quad -h < z < \zeta(x, y, t). \tag{2.1}$$

It is subject to boundary conditions on the bottom, $z = -h$:

$$\phi_z = 0; \tag{2.2}$$

and along the free surface, $z = \zeta$:

$$\left.\begin{aligned} D\zeta/Dt = \zeta_t + \phi_x\zeta_x + \phi_y\zeta_y = \phi_z; \\ g\zeta + \phi_t + \tfrac{1}{2}|\nabla\phi|^2 = T\frac{\zeta_{xx}(1+\zeta_y^2) + \zeta_{yy}(1+\zeta_x^2) - 2\zeta_{xy}\zeta_x\zeta_y}{(1+\zeta_x^2+\zeta_y^2)^{\frac{3}{2}}}. \end{aligned}\right\} \tag{2.3}$$

Here g is the gravitational acceleration, and T is the ratio of surface-tension coefficient to fluid density. We note that the linearized dispersion relation for this system is

$$\omega^2 = (g\kappa + \kappa^3 T)\tanh \kappa h. \tag{2.4}$$

In two dimensions, one should interpret $\kappa = (k^2+l^2)^{\frac{1}{2}}$ in (2.4).

2.1. *KdV limit*

The solution of (1.3) provides an approximate solution to these equations that is valid when the initial disturbance consists primarily of nearly one-dimensional long waves of small amplitude. To be precise, let $\boldsymbol{\kappa} = (k, l)$ be the horizontal wavenumber characteristic of the disturbance. Orient the horizontal co-ordinate system such that the x direction is the principal direction of wave propagation. Let a denote the characteristic amplitude of the disturbance. Then we need:

(i) small amplitudes, $$\epsilon \equiv a/h \ll 1; \tag{2.5a}$$

(ii) long waves, $$(\kappa h)^2 \ll 1; \tag{2.5b}$$

(iii) nearly one-dimensional waves,

$$(l/\kappa)^2 \ll 1. \tag{2.5c}$$

The KdV equation (1.1) results when the first two effects balance in truly one-dimensional problems, and (1.3) results when all three effects balance:

$$(\kappa h)^2 = O(\epsilon); \tag{2.5d}$$

$$(l/\kappa)^2 = O(\epsilon). \tag{2.5e}$$

Under the assumptions of (2.5) a first approximation of (2.1)–(2.3) reduces to

$$\frac{\partial^2 \zeta}{\partial t^2} - gh\frac{\partial^2 \zeta}{\partial x^2} = O(\epsilon). \tag{2.6}$$

Thus, to lowest order, the solution of (2.1)–(2.3) may be approximated by

$$\zeta \sim \epsilon h\{f_1[x-(gh)^{\frac{1}{2}}t; y] + f_2[x+(gh)^{\frac{1}{2}}t; y]\}, \tag{2.7}$$

where f_1 and f_2 are known in terms of the initial data. Throughout this paper, we are interested in problems where the initial disturbances are localized, and it is then convenient to assume *a fortiori* that the physical quantities have compact support initially. In this case, it is easy to show that f_1 and f_2 in (2.7) have compact support as well.

To go to higher order, we define scaled, dimensionless variables:

$$\left.\begin{aligned} r &= \epsilon^{\frac{1}{2}}[x-(gh)^{\frac{1}{2}}t]/h, & s &= \epsilon^{\frac{1}{2}}[x+(gh)^{\frac{1}{2}}t]/h;\\ \eta &= \epsilon y/h, & \tau &= \epsilon(gh)^{\frac{1}{2}}t/h;\\ u &= f_1, & v &= f_2;\\ \hat{T} &= T/gh^2. \end{aligned}\right\} \tag{2.8}$$

Now we look for solutions of the form $\zeta \sim \epsilon h[u(r,\tau,\eta) + v(s,\tau,\eta)]$; i.e. we use the method of multiple scales. To eliminate secular terms at the next order, we find

$$(2u_\tau + 3uu_r + (\tfrac{1}{3}-\hat{T})\,u_{rrr})_r + u_{\eta\eta} = 0 \tag{2.9a}$$

and

$$(2v_\tau - 3vv_s - (\tfrac{1}{3}-\hat{T})\,v_{sss})_s - v_{\eta\eta} = 0. \tag{2.9b}$$

The equation given by Kadomtsev & Petviashvili (1970) is in this form.

For most circumstances of interest in water waves,

$$\tfrac{1}{3}-\hat{T} > 0, \tag{2.10}$$

and it follows from (2.4) that the linearized phase speed is a (local) maximum at $\kappa = 0$. Thus, the waves governed by (2.9) travel faster than their neighbours (in κ space) and there should be no disturbance as $r \to +\infty$, or $s \to -\infty$. Consequently, (2.9*a*) may be integrated to

$$2u_\tau + 3uu_r + (\tfrac{1}{3}-\hat{T})\,u_{rrr} - \int_r^\infty u_{\eta\eta}\,dz = 0, \tag{2.11}$$

with a similar equation for (2.9*b*). This is now in the form of an evolution equation for u, as in (1.3). For very thin sheets of water (i.e. $\hat{T}$ large enough) (2.10) is false, the long waves travel slower than their neighbours, and the integral in (2.11) should be over $(-\infty, r)$.

Given (2.10), there is no apparent difficulty in requiring that u should vanish, along with its derivatives, as $r \to +\infty$. However, even if u and all of its derivatives vanish initially as $r \to -\infty$, it is evident from (2.11) that u will not remain zero there unless

$$\int_{-\infty}^{\infty} u_{\eta\eta}\,dr = 0. \tag{2.12}$$

Since u is the derivative of a velocity potential, (2.12) is automatically satisfied at the initial instant. Indeed, for the linearized form of (2.11), (2.12) is a constant of the motion, and it is sufficient to know it initially.

The constraint in (2.12) has a simple physical interpretation. One can identify $\int u(r,\eta,\tau)\,dr$ as the total mass of the wave in a thin strip at η. Then (2.12) assures that the transverse derivative of mass is constant, and this prevents a net flow of mass to (or from) any particular strip.

There are several indications that (2.11), or (1.3), is of IST-type. Dryuma (1974) has identified an appropriate linear scattering problem for (1.3); Zakharov & Shabat (1974) have related special solutions to a linear integral equation; Chen (1975) found a Bäcklund transformation; Satsuma (1976) has obtained 'N soliton', but non-localized, solutions by direct methods. In § 3 we discuss localized lump solutions. However, as mentioned earlier, no complete IST method has been developed for (1.3) to date.

2.2. *The nonlinear Schrödinger limit*

Let us now consider the derivation of (1.4) from (2.1)–(2.3). Here we are following a packet of nearly one-dimensional waves, travelling in the x direction, with an identifiable (mean) wavenumber, $\boldsymbol{\kappa} = (k, l)$. We denote the maximum variation in k within the packet by δk. To derive (1.4) we need:

(i) small amplitudes, $\epsilon \equiv \kappa a \ll 1;$ (2.13*a*)

(ii) slowly varying modulations, $\delta k/\kappa \ll 1;$ (2.13*b*)

(iii) nearly one-dimensional waves, $|l|/\kappa \ll 1;$ (2.13*c*)

(iv) a balance of all three effects, $\delta k/\kappa = O(\epsilon),$ (2.13*d*)

$$|l|/\kappa = O(\epsilon). \tag{2.13e}$$

The dimensionless depth, kh, can be finite or infinite, but to avoid the shallow water limit (and KdV), we need

$$(kh)^2 \gg \epsilon. \tag{2.14}$$

In this limit, the solution of the lowest order (linear) problem is

$$\phi \sim \epsilon \left(\frac{\cosh k(z+h)}{\cosh kh} [\tilde{A} \exp(i\theta) + *] + \text{const.} \right), \tag{2.15a}$$

where $*$ denotes the complex conjugate,

$$\theta = kx - \omega(k) t, \tag{2.15b}$$

and $\omega(k)$ is given by (2.4). To go to higher order, we introduce slow (dimensional) variables (again, using the method of multiple scales),

$$x_1 = \epsilon x, \quad y_1 = \epsilon y, \quad t_1 = \epsilon t, \quad t_2 = \epsilon^2 t, \tag{2.16}$$

and expand ϕ and ζ:

$$\left.\begin{aligned} \phi &\sim \epsilon \left\{ \tilde{\Phi}(x_1, y_1, t_1, t_2) + \frac{\cosh k(z+h)}{\cosh kh} [\tilde{A}(x_1, y_1, t_1, t_2) \exp(i\theta) + *] \right\} + O(\epsilon^2), \\ \zeta &= \epsilon\{\tilde{\zeta}_{11} \exp(i\theta + *\} + O(\epsilon^2), \quad \tilde{\zeta}_{11} = \frac{i\omega}{g + k^2 T} \tilde{A}. \end{aligned}\right\} \tag{2.17}$$

In order to derive (1.4), these expansions must be carried out to $O(\epsilon^3)$. The variations allowed in $\tilde{A}$ reflect the fact that this is a wave packet, rather than a uniform wave-train, and $\tilde{\Phi}$ provides a mean motion generated by the packet. In what follows we shall only discuss the secular effects that the higher order terms have on $\tilde{\Phi}$, and $\tilde{A}$; details can be found in Benney & Roskes (1969), Davey & Stewartson (1974) and Djordjevic & Redekopp (1977).

At the next order of approximation, a secular condition requires that the wave packet travel with its linear group velocity,

$$\partial \tilde{A}/\partial t_1 + C_g(k)\, \partial \tilde{A}/\partial x_1 = 0, \tag{2.18}$$

where $C_g = d\omega/d\kappa$. On this same time scale, $\tilde{\Phi}$ satisfies a forced wave equation,

$$\frac{\partial^2 \tilde{\Phi}}{\partial t_1^2} - gh \left\{ \frac{\partial^2 \tilde{\Phi}}{\partial x_1^2} + \frac{\partial^2 \tilde{\Phi}}{\partial y_1^2} \right\} = k\omega\beta_1 \frac{\partial}{\partial x_1} |\tilde{A}|^2, \tag{2.19}$$

where

$$\beta_1 = \frac{kC_g}{\omega} \operatorname{sech}^2 kh + 2/(1 + \tilde{T}),$$

$$\tilde{T} = k^2 T/g = (kh)^2 \hat{T}.$$

The solution of (2.19) changes dramatically, depending on whether or not

$$gh > C_g^2. \tag{2.20}$$

If the ratio $C_g/(gh)^{\frac{1}{2}}$ is interpreted as the 'Mach number' of the wave packet, then (2.20) is the condition for 'subsonic' flow. In this case, if $\tilde{A}$ has compact support, then $\tilde{\Phi}$ has a forced component that travels with speed C_g [i.e. it satisfies (2.18)], and a free component that radiates outward with speed $(gh)^{\frac{1}{2}}$, and is $O(t_1^{-\frac{1}{2}})$ as $t_1 \to \infty$. Hence with (2.20), as $t_1 \to \infty$, we find that $\tilde{\Phi}$ satisfies both (2.18) and

$$\alpha \frac{\partial^2 \tilde{\Phi}}{\partial x_1^2} + \frac{\partial^2 \tilde{\Phi}}{\partial y_1^2} = -\frac{k\omega}{gh} \beta_1 \frac{\partial}{\partial x_1} |\tilde{A}|^2, \tag{2.21}$$

where

$$\alpha = (gh - C_g^2)/gh,$$

with the boundary condition that $\tilde{\Phi}$ vanishes as $(x_1^2+y_1^2)\to\infty$. These are the boundary conditions prescribed by Davey & Stewartson (1974), and they are correct without surface tension.

If the effects of surface tension are strong enough, (2.20) fails and the flow is 'supersonic'. Now even if $\tilde{A}$ has compact support, $\tilde{\Phi}$ and its derivatives are non-zero along 'Mach lines' that emanate from the support of $\tilde{A}$. In the limit $t_1\to\infty$, $\tilde{\Phi}$ satisfies both (2.18) and (2.21) as before. However, the appropriate boundary conditions for (2.21) now are that $\tilde{\Phi}$ and its derivatives vanish ahead of the support of $\tilde{A}$ (e.g. as $x_1\to\infty$), and *no* conditions as $x_1\to-\infty$. Hence, in general, we cannot expect that global integrals involving $\tilde{\Phi}$ will converge.

The limit $t_1\to\infty$ is of interest because (1.4) appears when one eliminates secular terms on the *next* time scale, $t = O(\epsilon^{-2})$. Carrying this out, and putting the result in dimensionless form, we define

$$\left.\begin{aligned} &\xi = \epsilon k(x-C_g t), \quad \eta = \epsilon k y;\\ &\tau = \epsilon^2 (gk)^{\frac{1}{2}} t;\\ &A = k^2 (gk)^{-\frac{1}{2}} \tilde{A}, \quad \Phi = k^2 (gk)^{-\frac{1}{2}} \tilde{\Phi};\end{aligned}\right\} \tag{2.22}$$

and find that A and Φ satisfy

$$\left.\begin{aligned} iA_\tau + \lambda A_{\xi\xi} + \mu A_{\eta\eta} &= \chi |A|^2 A + \chi_1 A \Phi_\xi,\\ \alpha \Phi_{\xi\xi} + \Phi_{\eta\eta} &= -\beta(|A|^2)_\xi,\end{aligned}\right\} \tag{2.23}$$

where
$$\sigma = \tanh kh, \quad \tilde{T} = k^2 T/g, \quad \kappa = (k^2+l^2)^{\frac{1}{2}}, \tag{2.24a}$$

$$\omega^2 = gk\sigma(1+\tilde{T}) \geqslant 0, \quad \omega_0^2 = g\kappa, \quad \lambda = \kappa^2 \left(\frac{\partial^2 \omega}{\partial \kappa^2}\right) \Big/ 2\omega_0, \tag{2.24b}$$

$$\mu = \kappa^2 \left(\frac{\partial^2 \omega}{\partial l^2}\right) \Big/ 2\omega_0 = \frac{\kappa C_g}{2\omega_0} \geqslant 0, \tag{2.24c}$$

$$\chi = \left(\frac{\omega_0}{4\omega}\right) \left\{ \frac{(1-\sigma^2)(9-\sigma^2) + \tilde{T}(2-\sigma^2)(7-\sigma^2)}{\sigma^2 - \tilde{T}(3-\sigma^2)} + 8\sigma^2 - 2(1-\sigma^2)^2(1+\tilde{T}) - \frac{3\sigma^2 \tilde{T}}{1+\tilde{T}} \right\}, \tag{2.24d}$$

$$\chi_1 = 1 + \frac{\kappa C_g}{2\omega}(1-\sigma^2)(1+\tilde{T}) \geqslant 0, \tag{2.24e}$$

$$\alpha = (gh - C_g^2)/gh, \tag{2.24f}$$

$$\beta = \left(\frac{\omega}{\omega_0 kh}\right) \left\{ \frac{\kappa C_g}{\omega}(1-\sigma^2) + \frac{2}{1+\tilde{T}} \right\} \geqslant 0, \tag{2.24g}$$

$$\nu = \chi - \chi_1 \beta/\alpha. \tag{2.24h}$$

In the above formulae, all functions are evaluated at $l = 0$, since we are considering our underlying wavetrain to be propagating purely in the x direction. It should be noted that (2.23) can be easily scaled to (1.4) where $\sigma_1 = \operatorname{sgn}\lambda$, $\sigma_2 = \operatorname{sgn}\chi$, $a = \alpha\mu/\lambda^2$ and $b = \beta\mu\chi_1/\lambda^2|\chi|$ in (1.4).

Equations (2.23)–(2.24) are equivalent to those of Djordjevic & Redekopp [1977, their equations (2.12)–(2.13)] except for the correction of a misprint. If the initial

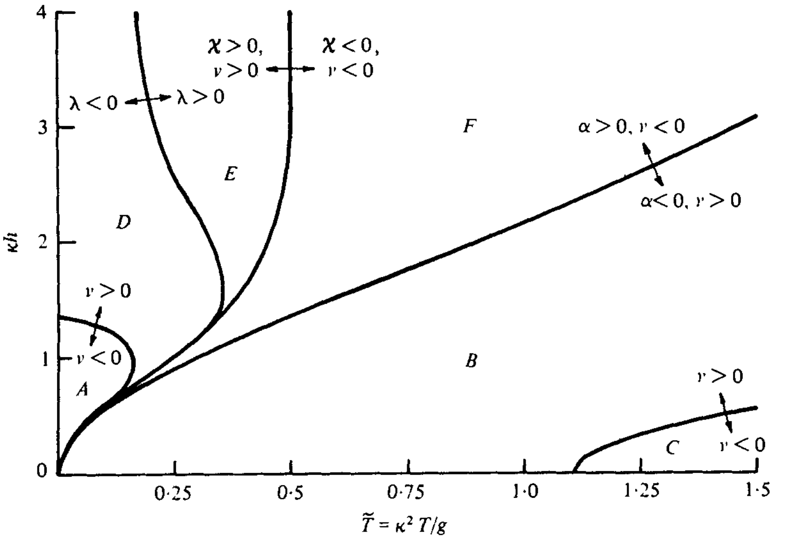


FIGURE 1. Map of parameter space, showing where the coefficients in (2.25) change sign. The dynamics of wave evolution is different in each region.

wave packet is local, it is appropriate to require that A vanishes as $\xi^2+\eta^2\to\infty$. As discussed above, the appropriate boundary conditions for Φ depend on the sign of α.

In the deep water limit, (2.23) reduces to

$$iA_\tau+\lambda_\infty A_{\xi\xi}+\mu_\infty A_{\eta\eta} = \chi_\infty|A|^2A, \tag{2.25}$$

where

$$\lambda_\infty = -\frac{\omega_0}{8\omega}\left(\frac{1-6\tilde{T}-3\tilde{T}^2}{1+\tilde{T}}\right),$$

$$\mu_\infty = \frac{\omega_0}{4\omega}(1+3\tilde{T}),$$

$$\chi_\infty = \frac{\omega_0}{4\omega}\frac{8+\tilde{T}+2\tilde{T}^2}{(1-2\tilde{T})(1+\tilde{T})}.$$

The appropriate boundary conditions for localized initial data are that A vanishes as $\xi^2+\eta^2\to\infty$.

The character of the solution of (2.23) depends fundamentally on the signs of the coefficients in the equations. Figure 1 is a map of parameter space, showing where these signs change. The figure is that of Djordjevic & Redekopp (1977), who used it to explain the various regions of stability/instability of the Stokes wave. Each boundary line corresponds to a simple zero of a coefficient, as shown, except for the two curves bounding region F. These two curves denote singularities of ν. In a neighbourhood of each of these two curves, phenomena occur on a shorter time scale than the $O(\epsilon^{-2})$ scale required elsewhere; cf. Djordjevic & Redekopp (1977).

If we take the long wave limit, $kh\to 0$, of (2.23) but keeping $\epsilon \ll (kh)^2$, we find equations which are of IST-type. We discuss this further in § 5. Alternatively, the long wave limit in which $\epsilon = O((kh)^2)$, where (2.11) applies, corresponds to the lower left-hand corner of this figure ($kh\to 0$, $\tilde{T}\to 0$, $\hat{T}$ fixed). The only parameter that changes sign in this limit is $(\frac{1}{3}-\hat{T})$, which is positive in region A, and negative in region B.

The uniformity of the limits $kh \to 0$, $\epsilon \to 0$ has been discussed by Ursell (1953), Hasimoto & Ono (1972) and Freeman & Davey (1975).

2.3. *Conservation laws*

Our final objective in this section is to give a simple physical interpretation for an infinite set of conservation laws. It is well known that the equations of water waves conserve mass, horizontal momentum and energy. If we interpret 'mass' as the mass associated with the wave, etc. then these conserved quantities may be represented as integrals. In one dimension (which is sufficient for the purpose of this discussion) we have:

$$M = \rho \int \zeta \, dx \quad \text{(mass)}; \tag{2.26}$$

$$m_x = \rho \int \left[\int_{-h}^{\zeta} \phi_x \, dz \right] dx \quad \text{(momentum)}; \tag{2.27}$$

$$\left.\begin{aligned} \text{K.E.} &= \frac{\rho}{2} \int\int_{-h}^{\zeta} |\nabla \phi|^2 \, dz \, dx \\ \text{P.E.} &= \frac{\rho}{2} g \int \zeta^2 \, dx \\ E &= \text{K.E.} + \text{P.E.} \end{aligned}\right\} \quad \text{(energy)}. \tag{2.28}$$

On the other hand, problems that have been solved exactly by IST possess an infinite set of conservation laws. For example, the first few quantities conserved by (1.2) are

$$\left.\begin{aligned} I_1 &= \int |A|^2 \, dx, \\ I_2 &= \int (A^* A_x - A_x^* A) \, dx, \\ I_3 &= \int \left(|A_x|^2 - \frac{\sigma}{2} |A|^4 \right) dx. \end{aligned}\right\} \tag{2.29}$$

There has been some speculation about the proper physical interpretation of this infinite set of conserved quantities. We offer here a very simple explanation. We have seen that (1.1)–(1.4) all are obtained *via* expansions in wave amplitude, ϵ. From this viewpoint, one might also expand (for example) the expression for the mass of the wave in powers of ϵ, to obtain a series of the form

$$M = \rho \sum_{1}^{\infty} \epsilon^n C_n. \tag{2.30}$$

Because M is constant in time, it follows that each coefficient, C_n, is also constant.

Because one generates the complete series for ϕ, ζ through $O(\epsilon^3)$ in deriving (1.4), it is then straightforward to compute the series in (2.30) to this order. In (2.17), any terms involving $\exp(i\theta)$ can be shown to contribute only at higher order, using integration by parts:

$$\begin{aligned} \int \zeta_{11}(x_1, t_1, t_2) \, e^{i\theta} \, dx &= -\frac{1}{ik} \int \frac{\partial \zeta_{11}}{\partial x} e^{i\theta} \, dx, \\ &= -\frac{\epsilon}{ik} \int \frac{\partial \zeta_{11}}{\partial x_1} e^{i\theta} \, dx. \end{aligned}$$

This process can be repeated as many times as ζ_{11} can be differentiated. The result of explicit computation is

$$\left.\begin{aligned} M &= \epsilon a_1 I_1 + \epsilon^2 a_2 I_2 + O(\epsilon^3),\\ m_x &= \epsilon^2 b_2 I_2 + O(\epsilon^3),\\ \text{K.E.} &= \epsilon c_1 I_1 + \epsilon^2 c_2 I_2 + O(\epsilon^3),\\ \text{P.E.} &= \epsilon c_1 I_1 + \epsilon^3 c_2 I_2 + O(\epsilon^3). \end{aligned}\right\} \tag{2.31}$$

I_3 enters at $O(\epsilon^3)$. The coefficients (a_i, b_i, c_i) are unimportant for our purpose. The momentum starts at higher order because it is referred to a co-ordinate system travelling with the group velocity of the wave. The identity of the last two series is a statement of the equipartition of the averaged energy, to this order. It is *not* true that I_1, I_2 and I_3 represent respectively the mass, momentum and energy of the water waves. (Similarly, the first three conserved quantities for KdV are not respectively the leading terms of the expansions of the mass, momentum and energy of the water waves.)

3. Stability of solitons

The primary purpose of this section is to discuss the stability of solitons with respect to transverse perturbations.

3.1. *The KdV limit*

Let us first consider the long wave problem, and (2.11). The one-dimensional limit, $\partial/\partial\eta = 0$, yields KdV. Here initial data on compact support evolve into a finite number of solitons, ordered by amplitude, followed by decaying oscillations that can be described in terms of a modulated similarity solution. The decay rate of the oscillations is not uniform in space, but it is of algebraic order (Ablowitz & Segur 1977*a*). The solitons are (theoretical) waves of permanent form when separated spatially from other waves. They represent water waves that decay only due to viscous effects. A KdV soliton is shown in figure 2*a*. Both the solitons and the decaying oscillations have been observed experimentally (Hammack & Segur 1974, 1978).

Kadomtsev & Petviashvili (1970) analysed the stability of a KdV soliton with respect to long transverse perturbations in (2.11). They found that the soliton is unstable with respect to such perturbations when (2.10) fails (i.e. in the lower left corner of region B in figure 1). The usual situation is region A, where (2.10) applies. Here they did not find that the soliton is unstable.

In region B, where the solitons are unstable, the KdV theory is of limited value. Here the solitons cannot represent asymptotic states, as they did in the one-dimensional problem. Thus, the question arises as to whether (2.11) has any other special solutions that might act as asymptotic states when the solitons are unstable. The answer is not known definitively at this time, but the work by Novikov† and Ablowitz & Satsuma (1978) is suggestive. In region B, but *not* in region A, (2.11) possesses 'lump' solutions. Lumps share many of the important properties of solitons:

(i) Each is a permanent wave whose speed, relative to the linearized speed, $(gh)^{\frac{1}{2}}$, can be made proportional to its amplitude.

† Lecture by S. P. Novikov for V. E. Zakharov on work by L. A. Bordag, A. R. Its, S. V. Manakov, V. B. Matreev and V. E. Zakharov in Rome, June 1977.

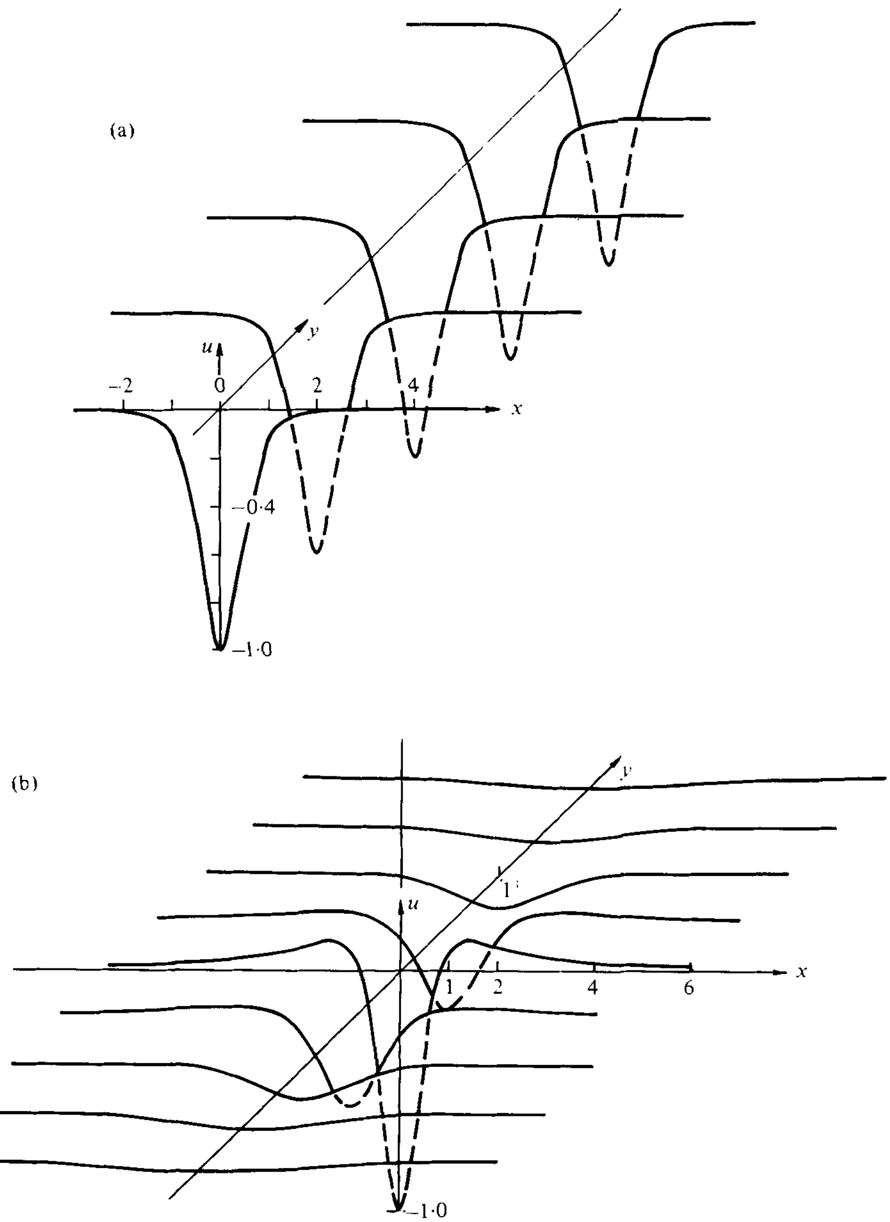


FIGURE 2 (*a*). KdV soliton, as seen in two space dimensions at a fixed time; $\kappa^2 = \frac{1}{12}$ in (3.1), with $\sigma = -1$. (*b*) Lump solution of (3.2) as seen in two dimensions at a fixed time; $p = 0$, $q^2 = \frac{1}{8}$, $\sigma = -1$.

(ii) Solitons are localized waves, with exponential tails in one dimension; lumps are localized waves, with algebraic tails, in two dimensions.

(iii) Two solitons regain their original amplitudes and speeds after a collision; the final effect of the collision is a phase shift of each soliton. Two lumps regain their original amplitudes and speeds after a collision, and suffer *no* phase shift.

(iv) Explicit formulae are available for N solitons, and for N lumps. The formulae for the one soliton and one lump solutions of (1.3), with $\sigma = -1$, are:

$$\left.\begin{aligned} u &= -12\frac{d^2}{dx^2}\ln\{1+\exp(-2\kappa x')\} \\ x' &= x+4\kappa^2 t \end{aligned}\right\} \quad \text{(soliton);} \tag{3.1}$$

$$\left.\begin{aligned} u &= -12\frac{d^2}{dx^2}\ln\{(x'+py')^2+(qy')^2+3/q^2\} \\ x' &= x+(p^2+q^2)\,t \\ y' &= y-2pt \end{aligned}\right\} \quad \text{(lump).} \tag{3.2}$$

These two solutions are drawn in figure 2 for a particular choice of the constants. (The soliton is a negative wave in region B, as shown. In region A, solitons are positive.)

These stability results suggest that, whereas the one-dimensional KdV solution may play an important role in (1.3) with $\sigma = +1$ (region A), no such situation is envisaged when $\sigma = -1$ (region B).

3.2. *The nonlinear Schrödinger limit*

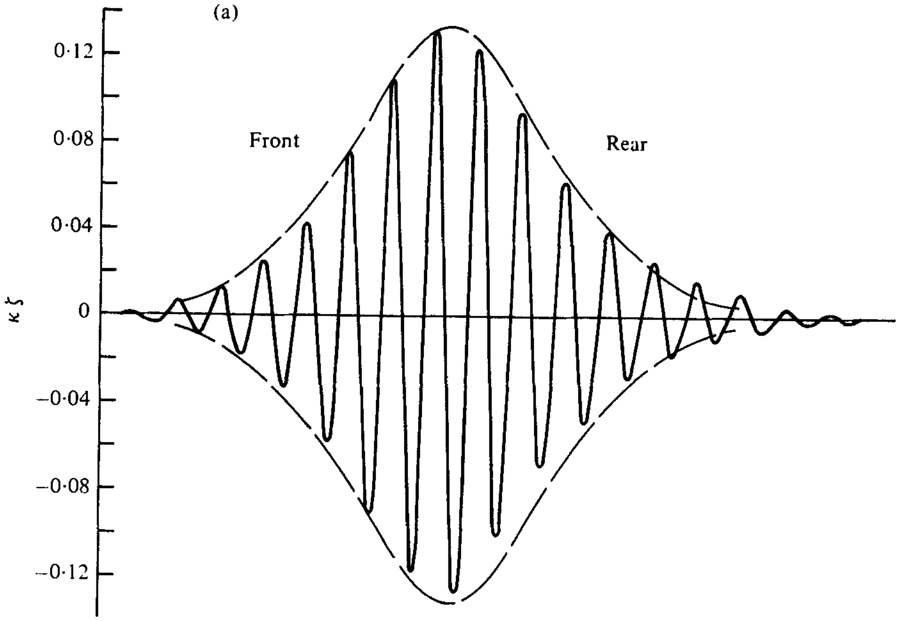
Next, we consider the nonlinear Schrödinger equation (2.23). Observe that (2.23) admits one-dimensional solitons travelling at almost any acute angle relative to the group velocity of the packet. The extreme cases are found by setting either $\partial/\partial\eta = 0$ or $\partial/\partial\xi = 0$. If $\partial/\partial\eta = 0$, the second equation in (2.23) can be integrated once, and the system reduces to

$$\begin{aligned} iA_\tau+\lambda A_{\xi\xi} &= \nu|A|^2A, \\ \Phi_\xi &= -\beta/\alpha|A|^2, \end{aligned} \tag{3.3}$$

where $\nu = \chi-\chi_1\beta/\alpha$, and the coefficients λ, α, β, χ, χ_1 are defined in (2.26). (Throughout this discussion, it should be borne in mind that the amplitude A represents the envelope of a train of plane waves.) Initial data can be created experimentally by modulating (in time) the stroke of an oscillating paddle at the end of a one-dimensional wave tank. If $\lambda\nu > 0$, as it is in regions A, B and E, of figure 1, there are no solitons. The initial data evolve into a field of decaying oscillations that we shall refer to as 'radiation'. This radiation can be described in terms of a modulated similarity solution, and it decays as $\tau^{-\frac{1}{2}}$ (Segur & Ablowitz 1976). In regions C, D and F, $\lambda\nu < 0$, and the same initial data now produce a finite set of *envelope* solitons in addition to the radiation. [For appropriate initial data, multi-soliton states are also possible (Ablowitz *et al.* 1974).] The one-soliton solution of (3.3) is

$$A = a|2\lambda/\nu|^{\frac{1}{2}}\operatorname{sech}\{a(\xi-2b\tau)\}\exp\{ib\xi+i\lambda(a^2-b^2)\,\tau\}. \tag{3.4}$$

The constant b in (3.4) represents an $O(\epsilon)$ correction to the basic wavenumber, k; without loss of generality we take $b = 0$. It is evident from (3.4) that the amplitude of the envelope soliton is of permanent form, and represents a physical wave that decays only due to viscous effects. Figure 3 shows the experimental measurements of such a wave, and we have superposed on the measurements the soliton solution with the same peak amplitude. [This experiment was conducted by Professor J. L. Hammack while at the University of Florida, and we are grateful to him for allowing us to use



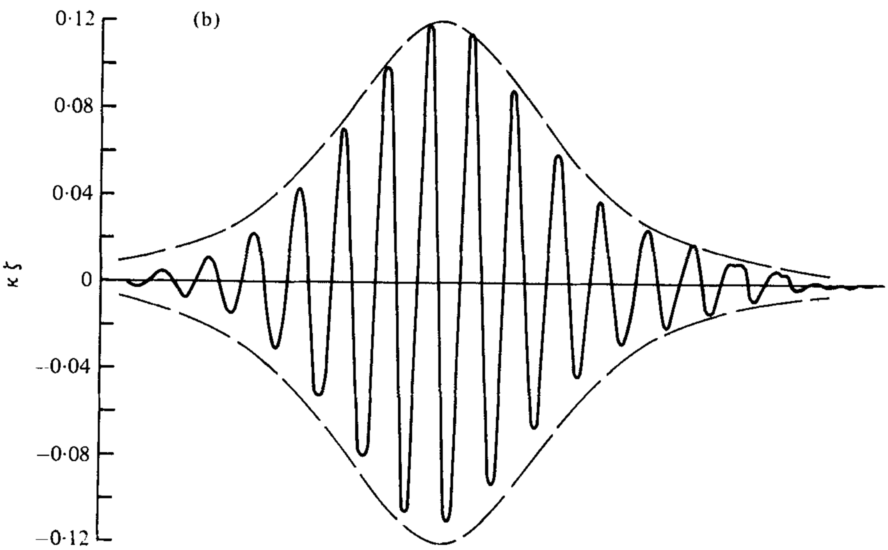


FIGURE 3. Measured surface displacement, showing evolution of envelope soliton at two downstream locations; $h = 1$ m, $kh = 4{\cdot}0$, $\omega = 1$ Hz, $\tilde{T} = 1{\cdot}0 \times 10^{-4}$; ——, measured history of surface displacement; – – –, theoretical envelope shape;

$$\kappa\zeta = \kappa a \operatorname{sech}(z),$$

$$z = [ag/\omega]\,(\nu/8\lambda)^{\frac{1}{2}}\,(C_g t - x);$$

(*a*) 6 m downstream of wave maker, $\kappa a = 0{\cdot}132$. (*b*) 30 m downstream of wave maker, $\kappa a = 0{\cdot}116$.

his unpublished data.] It is clear from this comparison that, at least in some aspects, the model represented by (3.3) is remarkably accurate. For more detailed comparisons, see Yuen & Lake (1975) or Hammack (to be published).

At the other extreme, if $\partial/\partial\xi = 0$ in (2.23), the system reduces to

$$iA_\tau + \mu A_{\eta\eta} = \chi|A|^2 A, \tag{3.5}$$

which is mathematically equivalent to (3.3) but represents a very different situation physically. Here wave crests move in the x (or ξ) direction, but they are modulated in the η direction. These modulations can move only in the η direction. To our knowledge, this configuration has not been explored experimentally in water waves, although it is common in nonlinear optics, where $A_{\eta\eta}$ represents diffraction of the light. In optics, initial data is produced experimentally with a diffraction grating, and the solution of (3.5) provides a nonlinear description of Fraunhofer diffraction (cf. Manakov 1974). Solitons exist where $\chi < 0$ (since $\mu \geqslant 0$) in regions B, C and F. To distinguish them from the soliton solutions of (3.3), we will refer to the solitons in (3.3) as 'envelope solitons', and the solitons in (3.5) as 'waveguides'.

Between these two extremes, $\partial/\partial\eta = 0$ and $\partial/\partial\xi = 0$, is a one-parameter family of other one-dimensional restrictions of (2.23) corresponding to one-dimensional waves (of the envelope) travelling at various angles relative to the group velocity of the carrier wave. Each of these one-dimensional problems is governed by an equation of the form (1.2), except at one angle that corresponds to crossing from region B to F, and another that corresponds to crossing from F to D.

Again, the question arises of the physical relevance of the one-dimensional soliton in the two-dimensional problem. For the nonlinear Schrödinger equation, (2.23), the answer seems to be that except for specially contrived one-dimensional geometries (like laboratory wave tanks), they are unlikely to persist. We show next that every one-dimensional soliton solution of (2.23), envelope soliton or waveguide, is unstable with respect to a long-wave transverse perturbation. Apparently, this instability has not been observed in wave tanks only because the tanks are too narrow to admit the long-wave perturbations required. The instability was discovered first by Zakharov & Rubenchik (1974) for (2.25). Our analysis is a generalization of theirs to the case of finite depth.

Consider first the envelope solitons, which are solutions of (3.3) and can exist in regions C, D and F in figure 1. As remarked above, it is sufficient to demonstrate the instability of the stationary soliton,

$$\left.\begin{aligned} A &= \exp(i\lambda a^2\tau)\,\psi(\xi), \\ \Phi_\xi &= -\beta/\alpha\psi^2(\xi), \end{aligned}\right\} \tag{3.6}$$

where $\psi(\xi)$ is real and satisfies

$$-\lambda\psi_{\xi\xi} + \lambda a^2 + \nu\psi^3 = 0. \tag{3.7}$$

Perturbations about this soliton can be put in the form

$$\left.\begin{aligned} A &= \exp(i\lambda a^2\tau)\,[\psi + u + iv], \\ \Phi &= (\beta/\alpha)\int_\xi^\infty (\psi^2 + 2\psi u)\,dz + w, \end{aligned}\right\} \tag{3.8}$$

where u, v, and w are real, $\quad |u|, |v| \ll \psi, \quad |w| \ll \Phi,$

and $\quad u, v, w \sim \exp(ip\eta \pm i\Omega\tau).$

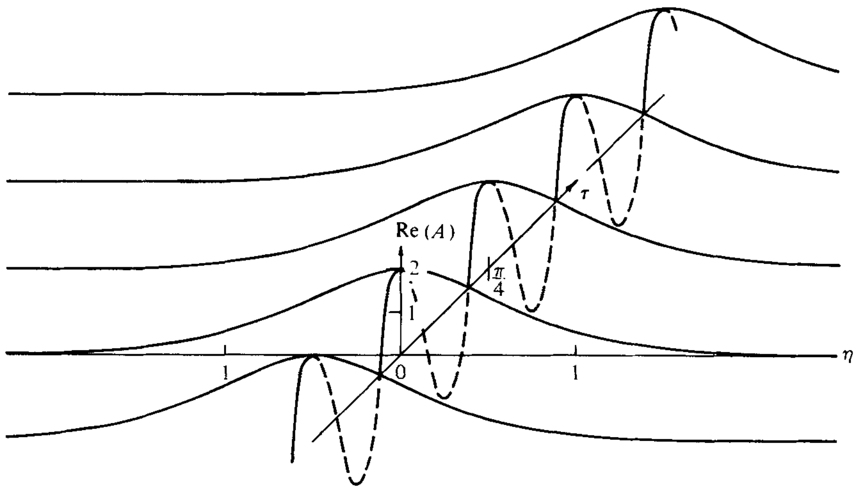


FIGURE 4. Stationary waveguide, as seen in time at a fixed location. In (3.5), $\mu = 2$, $\chi = -4$ and $\mathrm{Re}\,(A) = 2\,\mathrm{sech}\,2\eta\cos 8\tau$ is plotted. The displacement of the free surface, $\kappa\zeta$, is similar.

The question of stability now comes down to determining whether Ω^2 is positive. Substituting (3.8) into (2.23), linearizing and eliminating v yields

$$\left.\begin{aligned} \Omega^2 u &= (L_0+\mu p^2)(L_1+\mu p^2)\,u + \chi_1(L_0+\mu p^2)\,\psi w_\xi, \\ \alpha w_{\xi\xi} &= p^2\left[\frac{2\beta}{\alpha}\int_\xi^\infty (\psi u)\,dz + w\right], \end{aligned}\right\} \tag{3.9}$$

where L_0 and L_1 are the self-adjoint operators defined by

$$L_0 = -\lambda\frac{\partial^2}{\partial\xi^2} + \lambda a^2 + \nu\psi^2,$$

$$L_1 = -\lambda\frac{\partial^2}{\partial\xi^2} + \lambda a^2 + 3\nu\psi^2.$$

In the short-wave limit $(p^2\to\infty)$, (3.9) reduces to

$$\left.\begin{aligned} \Omega^2 u &= \mu^2 p^4 u + O(p^2), \\ \frac{2\beta}{\alpha}\int_\xi^\infty (\psi u)\,dz + w &\sim 0. \end{aligned}\right\} \tag{3.10}$$

Clearly Ω^2 is positive in this limit, and short waves are not unstable. Indeed, if they were unstable, it would be difficult to observe envelope solitons even in narrow wave tanks.

In order to analyse the long wave limit $(p^2\to 0)$, we expand the unknowns in (3.9) as

$$\left.\begin{aligned} u &\sim u_0 + p^2 u_1, \\ w &\sim p^2 w_1, \\ \Omega^2 &\sim p^2\Omega_1^2. \end{aligned}\right\} \tag{3.11}$$

Then to leading order, (3.9) becomes

$$L_0 L_1 u_0 = 0. \tag{3.12}$$

In order to solve (3.12), we define certain odd ($-$) and even ($+$) functions of ξ:

$$u_0^- = \psi_\xi, \quad u_0^+ = -\frac{1}{\lambda}\frac{\partial\psi}{\partial a^2}, \quad v_0^- = -\xi\psi/2\lambda, \quad v_0^+ = \psi. \tag{3.13}$$

The following relations can be obtained from (3.7):

$$\left.\begin{aligned} L_0 v_0^+ &= 0,\\ L_0 v_0^- &= u_0^-,\\ L_1 u_0^- &= 0,\\ L_1 u_0^+ &= v_0^+.\end{aligned}\right\} \tag{3.14}$$

It follows that u_0^- and u_0^+ both satisfy (3.12), and that v_0^- and v_0^+ satisfy the adjoint equation,

$$L_1 L_0 v_0 = 0. \tag{3.15}$$

In each case, there are two other solutions that do not vanish as $|\xi| \to \infty$. We will also need certain scalar products of these functions. Using the notation

$$(\phi_1, \phi_2) = \int_{-\infty}^{\infty} \phi_1 \phi_2 \, d\xi,$$

one computes

$$\left.\begin{aligned} (v_0^+, v_0^+) &= I,\\ (v_0^-, u_0^-) &= I/4\lambda,\\ (v_0^+, u_0^+) &= -\frac{1}{\lambda}\frac{dI}{da^2},\\ (v_0^+, u_0^-) &= (v_0^-, u_0^+) = 0;\\ (u_0^-, u_0^-) &= \int \psi_\xi^2 \, d\xi,\end{aligned}\right\} \tag{3.16}$$

where

$$I = \int |A|^2 d\xi = 4\left|\frac{\lambda a}{\nu}\right|.$$

At $O(p^2)$, (3.9) reduces to

$$L_0 L_1 u_1 = \Omega_1^2 u_0 - \mu(L_0 + L_1) u_0 - \chi_1 L_0 \psi(w_1)_\xi, \tag{3.17a}$$

$$\alpha(w_1)_{\xi\xi} = \frac{2\beta}{\alpha}\int_\xi^\infty (\psi u_0)\, dz. \tag{3.17b}$$

For u_1 to decay as $|\xi| \to \infty$, it is necessary that the non-homogeneous terms in (3.17) be orthogonal to the decaying solutions of the homogeneous adjoint equation (3.15). Because the equations are linear, it is sufficient to consider the odd and even modes separately. Thus, if u_0 in (3.17) is u_0^+, we multiply (3.17a) by v_0^+, integrate over ξ, and use integration by parts to obtain

$$0 = (\Omega_1^2)^+\left(-\frac{1}{\lambda}\frac{dI}{da^2}\right) - \mu I,$$

or

$$(\Omega_1^2)^+ = -\frac{\lambda\mu I}{dI/da^2} = -2\lambda\mu a^2. \tag{3.18}$$

For the odd mode, u_0^-, we multiply (3.17*a*) by v_0^- and use (3.17*b*). The result is

$$(\Omega_1^2)^- = \frac{4\lambda}{I}\left[\mu\int\psi_\xi^2\,d\xi+\frac{\chi_1\beta}{2\alpha^2}\int\psi^4\,d\xi\right]$$
$$= \tfrac{4}{3}\lambda a^2\left[\mu+\frac{2\chi_1\beta}{\alpha^2}|\lambda/\nu|\right]. \tag{3.19}$$

The question of stability of envelope solitons depends only on the sign of λ (the other factors in (3.18) and (3.19) are intrinsically positive). Using $\Omega^2 \sim p^2(\Omega_1^2)$, we summarize the result as follows.

(i) In region D, where $\lambda < 0$, an envelope soliton with amplitude a is unstable with respect to long disturbances that are antisymmetric ($-$) in ξ. The growth rate ($i\Omega$) of the disturbance with wavenumber p is found from

$$\Omega^2 = -\tfrac{4}{3}p^2a^2|\lambda|\left(\mu+\frac{2\chi_1\beta}{\alpha^2}|\lambda/\nu|\right)+O(p^4). \tag{3.20}$$

In the deep water limit, this simplifies to

$$\Omega^2 = -\tfrac{4}{3}p^2a^2\mu|\lambda|+O(p^4), \tag{3.21}$$

as found by Zakharov & Rubenchik (1974). Thus, for an inviscid fluid, the effect of finite depth is to enhance the growth rate of the instability. Zakharov & Rubenchik found the $O(p^4)$ correction to (3.21), and argued qualitatively that the most unstable wave satisfies

$$\mu p^2 = O(|\lambda|a^2) \tag{3.22a}$$

and that the maximum growth rate is on the order of

$$|\Omega| = O(|\lambda|a^2). \tag{3.22b}$$

Moreover, they noted that the growth of a mode that is antisymmetric in ξ and sinusoidal in η tends to bend the wave crest, producing a 'snake' effect; i.e. the crest of the perturbed wave oscillates back and forth in the ξ, η plane about its unperturbed position. Recent numerical computations in Saffman & Yuen (1978) have made (3.22) more precise.

(ii) In regions C and F, where $\lambda > 0$, an envelope soliton with amplitude a is unstable with respect to long symmetric ($+$) disturbances. The growth rate of the disturbances with wavenumber p is found from

$$\Omega^2 = -2p^2a^2\lambda\mu+O(p^4), \tag{3.23}$$

and this result also holds in the deep water limit. Again, qualitative considerations yield (3.22*b*). Growth of a symmetric mode tends to modulate the wave amplitude periodically in η.

Analysis of the stability of waveguides (in regions B, C and F) follows similar lines, and it is necessary only to indicate the main points of the analysis. A stationary waveguide has the form

$$A = \exp(i\mu a^2\tau)\,\psi(\eta), \qquad \Phi = 0, \tag{3.24}$$

where ψ is real and
$$-\mu\psi_{\eta\eta}+\mu a^2\psi-\chi\psi^3 = 0.$$

Perturbations take the form

$$A = \exp(i\mu a^2\tau)\,[\psi+u+iv], \qquad (3.25)$$

$$\Phi = -\int_{\xi}^{\infty} w\,dz.$$

The linearized equations for u and w, derived from (2.25), are

$$\left.\begin{aligned} \Omega^2 u &= (L_0+\lambda p^2)(L_1+\lambda p^2)\,u+\chi_1(L_0+\lambda p^2)\,\psi w,\\ w_{\eta\eta} &= p^2[2\beta\psi u+\alpha w],\\ \text{where}\qquad L_0 &= -\mu\frac{\partial^2}{\partial\eta^2}+\mu a^2+\chi\psi^2,\\ L_1 &= -\mu\frac{\partial^2}{\partial\eta^2}+\mu a^2+3\chi\psi^2. \end{aligned}\right\} \qquad (3.26)$$

These equations are very similar to those in (3.9) and we simply state the final result. Throughout regions B, C and F, (λ,μ) are positive. Anywhere in these regions, a stationary waveguide with amplitude a is unstable with respect to long symmetric (in η) disturbances. The growth rate $(i\Omega)$ of the disturbance with wavenumber p (in ξ) is found from (3.23) and qualitative considerations give (3.22*a*, *b*) with λ, μ interchanged.

We conclude this section by summarizing our results for the nonlinear Schrödinger equation, (2.23). There are many one-dimensional limits, including (3.3) and (3.5). These two limits admit envelope solitons and waveguides, respectively, in various regions of figure 1. However, all possible solitons are unstable with respect to some long-wave transverse perturbation. This instability does not appear in experiments in one-dimensional wave tanks, provided the tank width is small in comparison with the soliton length, because the unstable modes are excluded by the geometry. If this constraint is removed, however, the instability should occur, and neither kind of soliton is a stable asymptotic state that can be achieved from initial data in (2.23).

4. Focusing

In one-dimensional problems, like (1.2), the most dramatic nonlinear effect is that smooth initial data can 'focus' into a localized soliton, or into a set of solitons, which then persist forever. In this section, we show that focusing is even more dramatic in two dimensions and that a solution of (2.23) that evolves from smooth initial data can become singular at a point in space after a finite time. This is known as the 'self-focusing singularity', or simply as 'focusing'. In such a case the water wave equations must be re-examined in the neighbourhood of the focus.

To our knowledge, the phenomenon of focusing has not yet been observed as such in water waves, although it has been known for some time in nonlinear optics (e.g. Vlasov, Petrishchev & Talanov 1974). Some of the analysis discussed here uses the ideas presented by Zakharov & Synakh (1976) who studied what amounts to the two-dimensional version of (1.2) [i.e. (2.25)] in the context of the optics problem.

4.1. *Necessity of focusing*

Our first objective is to identify circumstances under which the solution of (2.23) must focus in a finite time. Consider any point in region F of figure 1, where

$$(\lambda, \mu, (-\chi), \chi_1, \alpha, \beta)$$

are all positive; i.e. consider capillary-type waves in sufficiently deep water. Consider initial data for (2.23) which are infinitely differentiable and which decay rapidly as $(\xi^2+\eta^2)\to\infty$; e.g. $A(\xi,\eta,0)$ might have compact support. If a solution of (2.23) exists and vanishes rapidly enough as $(\xi^2+\eta^2)\to\infty$, then the following integrals are constants of the motion:

$$I_1 = \iint |A|^2 d\xi\, d\eta; \tag{4.1a}$$

$$I_2 = \iint \left(A\frac{\partial A^*}{\partial \xi} - A^*\frac{\partial A}{\partial \xi}\right) d\xi\, d\eta; \tag{4.1b}$$

$$I_3 = \iint \left(A\frac{\partial A^*}{\partial \eta} - A^*\frac{\partial A}{\partial \eta}\right) d\xi\, d\eta; \tag{4.1c}$$

$$I_4 = \iint \left[\left\{\lambda\left|\frac{\partial A}{\partial \xi}\right|^2 + \mu\left|\frac{\partial A}{\partial \eta}\right|^2\right\} - \frac{1}{2}\left\{(-\chi)\,|A|^4 + \frac{\alpha\chi_1^3}{\beta}(\Phi_\xi)^2 + \frac{\chi_1^3}{\beta}(\Phi_\eta)^2\right\}\right] d\xi\, d\eta. \tag{4.1d}$$

Each bracket, $\{\,\}$, in I_4 is positive definite, and the second bracket vanishes in the linear limit of (2.23). Clearly $I_4 < 0$ is possible (e.g. if the initial data has sufficiently large amplitude).

It also follows from (2.23) that

$$\frac{\partial^2}{\partial \tau^2}\iint \left\{\frac{\xi^2}{\lambda} + \frac{\eta^2}{\mu}\right\} |A|^2 d\xi\, d\eta = 8I_4. \tag{4.2}$$

As noted in § 2, one may interpret I_1 as the mass of the wave (to leading order in ϵ). Then the integral in (4.2) may be interpreted as the moment of inertia, and (4.2) is an example of the virial theorem (e.g. Chandrasekhar 1961, p. 581). Equation (4.2) is easily integrated, and we see that, if $I_4 < 0$, then the moment of inertia vanishes at a finite time. Clearly, no global solution exists after this time, because the (positive definite) moment of inertia would become negative! Since the mass of the wave is conserved, (4.2) suggests that prior to this time the radius of gyration is vanishing as the mass accumulates at a single point. The rapid development of this singularity is what we mean by focusing.

Before examining the nature of the singularity that develops, let us consider the implications of this argument outside of region F. In regions B and C, where $\alpha < 0$, global integrals involving Φ are generally unbounded (cf. § 2) and no global information about the solution is available by this approach. Whether focusing exists in these regions is open. In region E there is no focusing in the deep water limit, since the parameters are such that $I_4 > 0$. In arbitrary depth the question of focusing is still open.

In regions A and D, the integral in (4.2) is not of definite sign, and provides no contradiction. Both because of the breakdown of this argument and because the type of instability of solitons is different than in region F, we expect that if singularities develop in these regions, they will be qualitatively different from those of the self-focusing type.

4.2. *Nature of the singularity*

Next, we examine the possible behaviour of the singularity that develops at the focus. Zakharov & Synakh (1976) studied the radially symmetric case of (2.25). They investigated this equation both by numerical computations and an approximate analytic procedure. From these they concluded that as $\tau \to \tau_0$ (τ_0 being the time of focus) the wave amplitude grows as $(\tau_0-\tau)^{-p}$, $p = \frac{2}{3}$. In this section we show that there are a number of quasi-self-similar solutions to the generalized nonlinear Schrödinger equation, (1.4), including one with $p = \frac{2}{3}$, but we have found no convincing argument that this local behaviour is necessarily of the $p = \frac{2}{3}$ type.

For convenience, we consider the scaled form of (2.23), namely (1.4). In region F of figure 1, where focusing can occur, $\sigma_1 = +1$, $\sigma_2 = -1$.

Let $A = B\exp(i\Psi)$ in (1.4), with B, Ψ real and find:

$$(\tfrac{1}{2}B^2)_t + (\Psi_x B^2)_x + (\Psi_y B^2)_y = 0, \tag{4.3a}$$

$$-\Psi_t B + B_{xx} + B_{yy} - B(\Psi_x^2+\Psi_y^2) = -B^3 + \Phi_x B, \tag{4.3b}$$

$$a\Phi_{xx} + \Phi_{yy} = -b(B^2)_x. \tag{4.3c}$$

We seek quasi-self-similar solutions of (4.3) in the neighbourhood of the point of focus in the form

$$B \sim \frac{1}{f}R(\bar{x},\bar{y}) + R_0(\bar{x},\bar{y},t), \tag{4.4a}$$

$$\Phi \sim \frac{1}{f}Q(\bar{x},\bar{y}) + Q_0(\bar{x},\bar{y},t), \tag{4.4b}$$

where $\bar{x} = x/f$, $\bar{y} = y/f$, $f(t) = (t_0-t)^p$, so that $f \to 0$ as $t \to t_0$. This expansion is asymptotic near the focus provided $R_0 \ll R$, $Q_0 \ll Q$ in this region. Zakharov & Synakh (1976) also assumed

$$R_0 = O(fR), \tag{4.5}$$

but this assumption seems to be unnecessary. In any case, the dominant terms in (4.3a) as $t \to t_0$ are

$$(\Psi_x R^2 - \tfrac{1}{2}\bar{x}R^2 f')_x + (\Psi_y R^2 - \tfrac{1}{2}\bar{y}R^2 f')_y = 0. \tag{4.6}$$

A special solution of (4.6) is

$$\begin{aligned}\Psi_x &= \tfrac{1}{2}\bar{x}f' + G_1(\bar{y},t)/R^2,\\ \Psi_y &= \tfrac{1}{2}\bar{y}f' + G_2(\bar{x},t)/R^2.\end{aligned} \tag{4.7}$$

Taking $G_1 = G_2 = 0$ (for which some motivation is provided below) yields

$$\Psi = \tfrac{1}{4}ff'(\bar{x}^2+\bar{y}^2) + g(t), \tag{4.8}$$

and with this we have from (4.3)–(4.4), as $f \to 0$,

$$R_{\bar{x}\bar{x}} + R_{\bar{y}\bar{y}} + R^3 - RQ_{\bar{x}} - g'(t)f^2R - \tfrac{1}{4}f^3f''(\bar{x}^2+\bar{y}^2)R \sim 0, \tag{4.9a}$$

$$aQ_{\bar{x}\bar{x}} + Q_{\bar{y}\bar{y}} + b(R^2)_{\bar{x}} \sim 0. \tag{4.9b}$$

There are various possibilities; e.g.

(*a*) $f''f^3 \ll 1, \quad g' = \kappa/f^2$;

(*b*) $f''f^3 = O(1), \quad g' = \kappa/f^2$;

others are obtained similarly. Case (*a*) implies that to leading order (4.9) reduces to

$$R_{\bar{x}\bar{x}} + R_{\bar{y}\bar{y}} + R^3 - RQ_{\bar{x}} - \kappa R = 0, \tag{4.10a}$$

$$aQ_{\bar{x}\bar{x}} + Q_{\bar{y}\bar{y}} = -b(R^2)_{\bar{x}}. \tag{4.10b}$$

If one also assumes (4.5), then (*a*) becomes $f^3f'' = O(f)$, from which it follows that $p = \frac{2}{3}$. However, the spatial structure defined by (4.10) does not depend on (4.5), or on $p = \frac{2}{3}$. In the deep water case with radial symmetry (see Zakharov & Synakh 1976), $b = Q = 0$, $\bar{r}^2 = \bar{x}^2 + \bar{y}^2$, and (4.10*a*) reduces to

$$R_{\bar{r}\bar{r}} + \frac{1}{\bar{r}} R_{\bar{r}} + R^3 - \kappa R = 0. \tag{4.11}$$

Chiao, Garmire & Townes (1964) first studied (4.11) as a model of cylindrical optical beams, and showed that its bounded solutions decay exponentially for large $\bar{r}$. The equation also arises as an exact reduction of (1.4) if we take

$$\left.\begin{aligned} Q &= b = 0, \\ \bar{r} &= \lambda(x^2+y^2)^{\frac{1}{2}}, \\ A &= \lambda R(\bar{r}) \exp(i\kappa\lambda^2 t). \end{aligned}\right\} \tag{4.12}$$

The fact that (4.11) is exact has important consequences, which we discuss in § 5.

In case (*b*), $p = \frac{1}{2}$ and the solution is exactly self-similar. Here (4.8)–(4.9) yield

$$R_{\bar{x}\bar{x}} + R_{\bar{y}\bar{y}} + R^3 - RQ_{\bar{x}} - \kappa R + \tfrac{1}{16}(\bar{x}^2+\bar{y}^2) R = 0, \tag{4.13a}$$

$$aQ_{\bar{x}\bar{x}} + Q_{\bar{y}\bar{y}} = -b(R^2)_{\bar{x}}, \tag{4.13b}$$

$$\Psi = -\tfrac{1}{8}(\bar{x}^2+\bar{y}^2) - \kappa \ln(t_0 - t) + \Psi_0. \tag{4.13c}$$

In the radially symmetric case, (4.13*a*) becomes

$$R_{\bar{r}\bar{r}} + \frac{1}{\bar{r}} R_{\bar{r}} + (\tfrac{1}{16}\bar{r}^2 - \kappa) R + R^3 = 0, \tag{4.14}$$

and for large $\bar{r}$ all bounded solutions decay as $(\bar{r})^{-1}$. We also note that a somewhat more general equation than (4.14), obtained by retaining G in (4.7), can be found in the symmetric case:

$$R_{\bar{r}\bar{r}} + \frac{1}{\bar{r}} R_{\bar{r}} + (\tfrac{1}{16}\bar{r}^2 - \kappa) R + R^3 - \frac{C^2}{\bar{r}^2 R^3} = 0, \tag{4.15a}$$

$$\Psi = -\tfrac{1}{8}\bar{r}^2 - \kappa \ln(t_0 - t) + C \int^{\bar{r}} \frac{d\rho}{\rho R^2(\rho)} + \Psi_0. \tag{4.15b}$$

However, one can show from (4.15*a*) that R has a finite value at the origin only if $C = 0$. This result provides some justification for neglecting G in (4.7).

Using any of these similarity solutions, going back to the full water wave equations and rescaling, we find that the focusing instability produces a finite [i.e. $O(1)$] region of space in which the wave amplitudes are potentially large enough to break [$O(1)$].

Next, we present an argument which suggests that the self-focusing singularity cannot be of the $p = \frac{1}{2}$ type, as we have described it here. For convenience, we consider the case of radial symmetry. The expansion in (4.4) is valid in a region near the focus where the first terms are dominant. If we assume that this 'inner solution' matches to an outer solution that is $O(1)$, then the inner expansion breaks down where

$$\frac{1}{f}R(\bar{r}) = O(1). \tag{4.16}$$

But for large $\bar{r}$, R decays as $(\bar{r})^{-1}$, so that $(1/f)R$ decays as r^{-1}; i.e. there is no time-dependence. It follows that (4.16) defines a boundary for the focal region, denoted by $r = O(L)$, where L is time-independent. The mass within this region is proportional to

$$M = \int_0^L r\frac{1}{f^2}R^2(\bar{r})\,dr = \int_0^{L/f} \bar{r}R^2(\bar{r})\,d\bar{r}, \tag{4.17}$$

and (because $R \sim \bar{r}^{-1}$) this grows logarithmically as $t \to t_0$. But the total mass is finite, and this is a contradiction.

In case (*a*) $R(\bar{r})$ decays exponentially and no such contradiction appears. Moreover, if the nonlinearity in (1.4) were slightly stronger, no contradiction appears in the purely self-similar case. To be precise, if the nonlinear term in (1.4) were replaced with $|A|^{2a}A$, $a > 1$, then the radially symmetric similarity solution becomes

$$A = (t_0 - t)^{-1/2a}B(\bar{r}), \quad \bar{r} = r/(t_0 - t)^{\frac{1}{2}}. \tag{4.18}$$

In this case the radial decay is $B(\bar{r}) \sim (\bar{r})^{-1/a}$, and again L is finite. However, in this situation, when $a > 1$ the mass remains finite as $\tau \to \tau_0$, and the pure similarity solution is a likely candidate for describing the dynamics of the focus regime.

Finally, we remark that a natural generalization of (1.4) is:

$$iA_t + \sigma_1 A_{xx} + A_{yy} + \sigma_2 A_{zz} = \sigma_3|A|^{2a}A + \Phi_x A, \tag{4.19a}$$

$$a_1\Phi_{xx} + \Phi_{yy} + a_2\Phi_{zz} = -b(|A|^{2a})_x, \tag{4.19b}$$

where the $\sigma_i = \pm 1$ $(i = 1, 2, 3)$ and a_1, a_2, b are constant. The spherically symmetric limit is obtained by taking $b = \Phi = 0$, $\sigma_1 = \sigma_2 = +1$. Since the spherically symmetric equation has wide applicability, and (1.4) is itself physically relevant, we expect that (4.23) will also arise in physical problems.

5. Other solutions of the nonlinear Schrödinger equation

The purpose of this section is to identify other features of the solution of (2.23) that may play a role in its asymptotic $(r \to \infty)$ solution.

5.1. *Complete integrability*

Perhaps the fundamental question to answer about (2.23) is whether it is completely integrable; i.e. whether it can be solved exactly by relating it to an appropriate linear scattering problem. The question is natural in light of the fact that the one-dimensional problem can be solved in this way.

Consider first the long-wave limit of (2.23), subject to the constraint in (2.14). Here, (2.23) becomes (after rescaling of variables)

$$\left.\begin{aligned} iA_\tau - \sigma_1 A_{\xi\xi} + A_{\eta\eta} &= \sigma_1|A|^2A + A\Phi_\xi, \\ \sigma_1\Phi_{\xi\xi} + \Phi_{\eta\eta} &= -2(|A|^2)_\xi, \quad \sigma_1 = \operatorname{sgn}(\tfrac{1}{3} - \hat{T}). \end{aligned}\right\} \tag{5.1}$$

This system is of I.S.T. type (Ablowitz & Haberman 1975; Anker & Freeman 1978). Special N soliton solutions can be constructed either by a direct (Hirota type) method or via the Zakharov–Shabat approach (Anker & Freeman 1978).

The situation seems to be much different in the deep water limit. Here we have already seen that (4.11) is an exact reduction of (2.23) to an ordinary differential equation; i.e. every solution of (4.11) provides an exact solution of (2.23) in this limit. Let us consider those partial differential equations (PDE) which have been solved exactly by IST methods. We have found that every reduction of one of these PDE's to an ordinary differential equation (ODE) results (perhaps after a transformation of dependent variables) in an ODE without moveable critical points (Ablowitz & Segur 1977*b*; Ablowitz, Ramani & Segur 1978).

We expect that if (2.23) can be solved by some IST, then (4.11) should have no moveable critical points. But Ince (1944, especially p. 344) provides a complete list of all such second-order equations; (4.11) is not on this list and cannot be transformed to any equation that is present. Therefore, the solution of (4.11) *has* moveable critical points. Moreover, one can show that (4.11) has logarithmic singularities in addition to poles. On this basis, we conjecture that (2.23) cannot be solved exactly by IST in the deep-water limit.

Although (2.23) can be solved by IST in the shallow-water limit (i.e. lower-left corner of figure 1), it apparently cannot be solved in this way in the deep-water limit. Wherever IST methods fail, one is forced to piece together special solutions of the problem to describe the general solution.

5.2. *Decaying oscillations*

The special solutions discussed so far in this paper have been localized: either solitons (or soliton-like) or self-focusing singular solutions. However, in the one-dimensional limit of (2.23), solitons make up only part of the asymptotic solution of the initial value problem. That part of the solution associated with the continuous spectrum spreads over large regions of space, while it decays as $t^{-\frac{1}{2}}$. In particular, an exact solution of (1.2) is

$$A = t^{-\frac{1}{2}}\Lambda \exp\{i(x^2/4t + \sigma\Lambda^2 \ln t + \phi)\}, \tag{5.2}$$

where Λ and ϕ are real constants; the solution of (1.2) associated with the continuous spectrum tends to a slowly varying modulation of this, where Λ and ϕ depend on (x/t) (Segur & Ablowitz 1976; Segur 1976).

In the two-dimensional problem, (1.4), there is an analogous exact solution:

$$\left.\begin{aligned} A &= t^{-1}\Lambda \exp\left\{i\left(\frac{\sigma_1 x^2 + y^2}{4t} + \sigma_2\Lambda^2/t + B(t) + \phi\right)\right\}, \\ \Phi &= -B'(t)\,x + C(t)\,y + D(t). \end{aligned}\right\} \tag{5.3}$$

Similar solutions in the deep water limit of (2.23) were found by Talanov (1967). On the basis of the one-dimensional theory, we anticipate that the part of the solution of (1.4) that decays in time can be described in terms of a slowly varying modulation of this exact solution.

Moreover, this behaviour would be consistent with the results of Lin and Strauss (to appear) who studied the three-dimensional problem

$$iu_t - \Delta u + |u|^2 u = 0, \tag{5.4}$$

where Δ is the Laplacian in three dimensions. They found that the solution exists for all time and decays as $t^{-\frac{3}{2}}$. The appropriate similarity solution here is

$$u = t^{-\frac{3}{2}}\Lambda \exp\{-i[(x^2+y^2+z^2)/4t+\Lambda^2/2t^2+\phi]\}. \tag{5.5}$$

Without solitons or focusing, the decay rate of the solution of the nonlinear Schrödinger equation seems to be

$$u = O(t^{-\frac{1}{2}n}) \tag{5.6}$$

where n is the number of spatial dimensions. This decay rate is the same as in the linearized problem.

We are grateful to L. G. Redekopp and to J. L. Hammack for lending us the material in figures 1 and 3, respectively, and to Martin Kruskal for useful comments. This work was supported by the Office of Naval Research and by the National Science Foundation.

REFERENCES

Ablowitz, M. J. & Haberman, R. 1975 *Phys. Rev. Lett.* **35**, 1185.
Ablowitz, M. J., Kaup, D. J., Newell, A. C. & Segur, H. 1974 *Stud. Appl. Math.* **53**, 249.
Ablowitz, M. J., Ramani, A. & Segur, H. 1978 *Lett. Nuovo cimento*, **23**, 333.
Ablowitz, M. J. & Satsuma, J. 1978 *J. Math. Phys.* **19**, 2180.
Ablowitz, M. J. & Segur, H. 1977*a* *Stud. Appl. Math.* **57**, 13.
Ablowitz, M. J. & Segur, H. 1977*b* *Phys. Rev. Lett.* **38**, 1103.
Anker, D. & Freeman, N. E. 1978 *Proc. Roy. Soc.* A **360**, 529.
Bennfy, D. J. & Roskes, G. J. 1969 *Stud. Appl. Math.* **48**, 377.
Chandrasekhar, S. 1961 *Hydrodynamic and Hydromagnetic Stability*. Oxford.
Chen, H. 1975 *J. Math. Phys.* **16**, 2382.
Chiao, R. Y., Garmire, E. & Townes, C. H. 1964 *Phys. Rev. Lett.* **13**, 479.
Davey, A. & Stewartson, K. 1974 *Proc. Roy. Soc.* A **388**, 191.
Djordjevic, V. D. & Redekopp, L. G. 1977 *J. Fluid Mech.* **79**, 703.
Dryuma, V. 1974 *Sov. Phys. J. Exp. Theor. Phys.* **19**, 387.
Freeman, N. C. & Davey, A. 1975 *Proc. Roy. Soc.* A **344**, 427.
Hammack, J. Preprint.
Hammack, J. & Segur, H. 1974 *J. Fluid Mech.* **65**, 209.
Hammack, J. & Segur, H. 1978 *J. Fluid Mech.* **84**, 337.
Hasimoto, H. & Ono, H. 1972 *J. Phys. Soc. Japan* **33**, 805.
Hayes, W. D. 1973 *Proc. Roy. Soc.* A **332**, 199.
Ince, E. L. 1944 *Ordinary Differential Equations*. Dover.
Kadomtsev, B. & Petviashvili, V. 1970 *Sov. Phys. Dokl.* **15**, 539.
Lin, J. E. & Strauss, W. A. 1978 *J. Funct. Anal.* **30**, 245.
Manakov, S. V. 1974 *Sov. Phys. J. Exp. Theor. Phys.* **38**, 693.
Saffman, P. G. & Yuen, H. C. 1978 *Phys. Fluids* **21**, 1450.
Satsuma, J. 1976 *J. Phys. Soc. Japan* **40**, 286.
Segur, H. 1976 *J. Math. Phys.* **17**, 714.
Segur, H. & Ablowitz, M. J. 1976 *J. Math. Phys.* **17**, 710.
Talanov, V. I. 1967 *Radiophys.* **9**, 260.
Ursell, F. 1953 *Proc. Camb. Phil. Soc.* **49**, 685.

VLASOV, V. N., PETRISHCHEV, I. A. & TALANOV, V. I. 1974 *Quant. Electron., Radiophys.* **14**, 1062.

YUEN, H. C. & LAKE, B. M. 1975 *Phys. Fluids* **18**, 956.

ZAKHAROV, V. E. 1968 *Sov. Phys. J. Appl. Mech. Tech. Phys.* **4**, 86.

ZAKHAROV, V. E. & RUBENCHIK, A. M. 1974 *Sov. Phys. J. Exp. Theor. Phys.* **38**, 494.

ZAKHAROV, V. E. & SHABAT, A. B. 1974 *Func. Anal. Appl.* **8**, 226.

ZAKHAROV, V. E. & SYNAKH, V. S. 1976 *Sov. Phys. J. Exp. Theor. Phys.* **41**, 465.