Non-linear dispersive waves

By G. B. Whitham
California Institute of Technology

(*Communicated by M. J. Lighthill, F.R.S.—Received* 23 *June* 1964)

A general theory is developed for studying changes of a wave train governed by non-linear partial differential equations. The technique is to average over the local oscillations in the medium and so obtain differential equations for the variations in amplitude, wave number, etc. It corresponds to the Krylov–Bogoliubov averaging technique for the ordinary differential equations of non-linear vibrations. The equations obtained in this way are hyperbolic and can be handled by the usual theory of quasi-linear hyperbolic systems, involving the theory of characteristics and shock waves. In this case the 'shocks' are abrupt changes in the amplitude, wave number, etc. They do not involve dissipation, but it turns out that frequency plays the role corresponding to entropy in ordinary gas dynamic shocks. It is not clear whether these shocks will really occur in practice. However, they have a number of interesting properties and seem to be relevant to the discussion of so-called collisionless shocks in plasma dynamics. The main applications envisaged are to water waves and plasma dynamics, and the theory is developed using typical equations from these areas.

If the original equations are linear, this theory predicts the usual description of dispersive waves in terms of group velocity, so it may be considered as an extension of the group velocity concept to non-linear problems. Mathematically, the theory may be considered as an extension of some of the methods and ideas for the non-linear ordinary differential equations of vibration theory to partial differential equations.

1. General discussion

There is a fairly complete theory of linear dispersive systems. In one-dimensional problems, the equations have elementary solutions of the form

$$\alpha \cos(\kappa x - \omega t), \quad \omega = \omega(\kappa), \tag{1}$$

for arbitrary wave number κ and amplitude α. The general solution is then given by the corresponding Fourier integrals of the form

$$\phi(x,t) = \int_0^\infty F(\kappa)\cos(\kappa x - \omega t)\,\mathrm{d}\kappa, \tag{2}$$

where $F(\kappa)$ is chosen to satisfy appropriate initial or boundary conditions. The dependence of the phase velocity ω/κ on κ leads to the 'dispersion' of the different components. Here, the term 'dispersive' will be limited to those cases where $\omega(\kappa)$ is a *real* function with $\omega''(\kappa) \neq 0$. If $\alpha\,\mathrm{e}^{\mathrm{i}(\kappa x - \omega t)}$ is a solution and $\omega(\kappa)$ has an imaginary part, the solution will exhibit damping or diffusion as well as dispersion. On the other hand, if $\omega \propto \kappa$ as in the simple wave equation, the dispersion effects are absent because the phase velocity is independent of κ.

The equations of a linear dispersive system are often obtained by linearization of governing equations which are originally non-linear. In contrast to the linear case, there seems to be neither a precise classification nor a general method of treatment for the corresponding non-linear equations. In this paper, typical examples are considered and a general method of treating them is provided, but the classification is not attempted. The attitude is that any system is open for consideration if upon linearization it leads to dispersive waves in the sense defined above. The proposed treatment of such equations is an approximate one motivated by the following considerations in the linear theory.

Although (2) is an exact solution it is nearly always a complicated function of (x, t) which does not make the main features and important physical consequences immediately obvious. It is usual, therefore, to look at the asymptotic expansion of (2) for large x and t. Indeed this is frequently sufficient for all the purposes at hand. The asymptotic expansion may be obtained by the method of stationary phase, or the saddle-point method, and is

$$\phi \sim F(k)\sqrt{\left\{\frac{2\pi}{t\,|\omega''(k)|}\right\}}\cos\left(kx-\omega t-\tfrac{1}{4}\pi\operatorname{sgn}\omega''\right), \tag{3}$$

where $k(x, t)$ is the solution of

$$x = \omega'(k)\,t. \tag{4}$$

The derivation of this result, together with a good discussion of linear dispersive waves, is given by Jeffreys & Jeffreys (1956), § 17·08.

The solution (3) represents an oscillatory wave train, but in contrast to (1) it is not a uniform wave train because k depends upon x and t. However, when x and t are large compared with a typical wavelength and period, respectively, equation (4) shows that the change of k in a few wavelengths or periods is small. For example, k_x is given by

$$\frac{k_x}{k} = \frac{\omega'(k)}{k\omega''(k)}\frac{1}{x} = O\left(\frac{1}{x}\right), \tag{5}$$

provided $\omega''(k) \neq 0$. The equations of the problem always contain parameters λ_0, τ_0 with dimensions of length and time, and these become the typical orders of the wavelength and period for a wave train. Then, the relative change in k in one wavelength is $O(\lambda_0/x)$ which is small for $x \gg \lambda_0$.

Apart from a phase change which is unimportant for this work, (3) can be written

$$\phi = a\cos(kx-\omega t),$$

where a, k, $\omega = \omega(k)$ are slowly varying functions of x and t in this sense. This expression for ϕ is the elementary solution (1) with constant parameters α, κ replaced by slowly varying functions of x and t. In other words, the wave train is given *locally* in both space and time by the elementary solution, but the amplitude and wave number vary slowly over distances large compared with the wavelength and over times large compared with the period. There are then two scales involved: the smaller scale determined by the wavelength λ_0 and period τ_0 of the oscillations, during which a, k remain approximately constant, and the larger scale, given by x and t themselves, over which a, k change appreciably. In other problems, for example a uniform wave train entering a medium whose properties vary slowly with length scale $L \gg \lambda_0$, the larger scale would be the parameter L rather than x itself.

The determination of the functions $a(x, t)$, $k(x, t)$ in (3) and (4) gives the physical interpretation of the solution in terms of the 'group velocity' $\omega'(k)$. For, equation (3) shows that in order to follow waves of a certain wave number k, an observer must move with the constant velocity $\omega'(k)$. That is, each value of k propagates with the corresponding group velocity $\omega'(k)$. The expression for a in (3) shows that energy also propagates with the group velocity in a certain sense (see Jeffreys & Jeffreys, or § 3 below).

For non-linear problems, it is assumed that exact analytic solutions corresponding to (2) are out of the question and the possibility of finding approximate solutions corresponding to (3) and (4) is explored. In the initial value problem one might argue that the disturbances will be small enough at large x and t for a linearized theory to apply. However, the non-uniform validity of linearized theories over large distances is well known and is again found here; moreover, the initial value problem merely motivates a general discussion of non-uniform wave trains which leads to a number of basic concepts; it is intended to use the same ideas for problems like the wave train entering a non-uniform medium mentioned above.

For the linear case various direct derivations of (3) and (4), without using the exact solution (2), are well known (Rossby 1945; Landau & Lifshitz 1959; Whitham 1961), but it is necessary to find one that can be generalized to non-linear problems. The key to the approach given here, and probably to any approach, is the fact that the non-linear equations concerned do have elementary solutions corresponding to (1). The solutions are no longer sinusoidal but are easily found from the property that they represent uniform wave trains; i.e. the dependent variables are functions of $x-Ut$ alone, where U is the velocity of translation, i.e. the phase velocity.

For example, consider
$$\phi_{tt}-\phi_{xx}+V'(\phi) = 0, \tag{6}$$
which is the simplest equation of the appropriate type. (One may think of this as the string equation with a non-linear restoring force derived from a potential energy $V(\phi)$. But it should be noted that it does not apply strictly to the string problem since other non-linear terms have been neglected in linearizing the derivative terms. In physical variables the length and time parameters λ_0, τ_0 referred to above appear in the equation and have been eliminated by replacing the original x and t by x/λ_0 and t/τ_0.) The elementary steady profile wave is obtained by taking $\phi = \Phi(X)$, $X = x-Ut$. Then,
$$\left.\begin{aligned}(U^2-1)\,\Phi_{XX}+V'(\Phi) &= 0,\\ \tfrac{1}{2}(U^2-1)\,\Phi_X^2+V(\Phi) &= A,\end{aligned}\right\} \tag{7}$$
where A is a constant of integration. In the linearized problem, $V = \frac{1}{2}\Phi^2$ and the solution of (7) is
$$\left.\begin{aligned}&\Phi = \alpha\cos\kappa X,\\ &\kappa^2 = (U^2-1)^{-1}, \quad \alpha^2 = 2A;\end{aligned}\right\} \tag{8}$$
we have an example of (1) with $\omega = U\kappa = \sqrt{(\kappa^2+1)}$. The solution of (7) for other functions $V(\phi)$ is the appropriate extension to the non-linear case. The solution of (7) can be written explicitly as
$$X = \sqrt{\{\tfrac{1}{2}(U^2-1)\}}\int\frac{d\Phi}{\sqrt{\{A-V(\Phi)\}}};$$
if $V(\Phi)$ is a cubic or quartic in Φ, for example, this reduces to elliptic functions.

In general it is found that the non-linear equations considered always have a solution of this type with

$$\left.\begin{aligned} \phi &= \Phi(X;\, U, A_i), \quad X = x - Ut, \\ \text{and} \qquad \Phi_X^2 &= F(\Phi;\, U, A_i), \end{aligned}\right\} \tag{9}$$

where the A_i are constants of integration, the number depending upon the order of the original equation. The solution is oscillatory with Φ oscillating between a pair of zeros of the function F. The amplitude, wave number, frequency, etc., can all be determined in terms of the parameters U, A_i.

In direct analogy with the linear theory discussed above, it is assumed that more general solutions of the non-linear equations will be given *locally* by (9) with the constant parameters U, A_i replaced by slowly varying functions of x and t. The problem is, then, to provide a method for the determination of these functions.

Similar problems occur in vibration theory where time is the only independent variable and the equations are ordinary differential equations. 'Two time' problems arise there and have been much discussed (see, for example, Bogoliubov & Mitropolsky 1961; Minorsky 1962). Typical instances are the slow damping of a vibrating system or the theory of 'adiabatic invariants' for a vibrating system in which an external parameter is slowly varied. In the wave problems, we have an infinite number of vibrating systems and we are studying the slow transfer of energy, etc., between them, rather than the loss of energy, etc., from the system. In vibration theory, various methods yielding various degrees of detail have been developed for these two-time problems. The one that asks least detail, and so is a natural first choice for the more difficult problems of partial differential equations, is the Krylov–Bogoliubov averaging technique. The equations are averaged over the smaller time scale in order to obtain equations involving the large-scale quantities (amplitude, frequency, etc.) alone. It is shown in this paper that a similar technique can be developed for the partial differential equations of dispersive waves. It turns out also that the function

$$W(U, A_i) = \oint \Phi_X \,\mathrm{d}\Phi = \oint \surd\{F(\Phi;\, U, A_i)\}\,\mathrm{d}\Phi, \tag{10}$$

defined from the solution (9) plays an important role and this is similar to the adiabatic invariants, $\oint p\,\mathrm{d}q$, for ordinary Hamiltonian systems.

The averaged equations are derived from conservation equations, such as the energy equation, and they constitute a first-order system of differential equations for the slowly varying functions $U(x,t)$, $A_i(x,t)$. The details are given in the next section. In all the cases considered, the averaged equations are hyperbolic showing that changes in U, A_i *propagate*. The propagation velocities provide a generalization of group velocity to non-linear problems. For linear systems, it will be shown that the group velocity corresponds to a double characteristic of the averaged equations; as noted earlier, this characteristic carries both changes in energy and wave number. In non-linear systems, the propagation velocities depend on the amplitude, and the double characteristic splits into two separate characteristics with different velocities;

the velocities become equal only in the limit as the amplitude tends to zero. This means that the solutions of the non-linear equations will have very different properties from the linear case even when the non-linear terms are small.

In a linear dispersive system, the energy in each element of wave number space remains constant in time, and, consequently, wave number and energy propagate together. In a non-linear system, there is continual energy transfer between different wave numbers. The splitting of the double characteristic velocity into two separate propagation velocities must be associated with this difference.

It is particularly interesting that the equations for U and A_i are hyperbolic and homogeneous in the derivatives even when the original equations for ϕ are not.† The relationship between the mathematics of hyperbolic equations (particularly homogeneous ones) and the physical idea of waves is so strong that people have wondered whether some formulation for deep water waves, for example, could be obtained which would involve hyperbolic equations. This theory gives a possible answer: the detailed 'microscopic' structure of ϕ is not necessarily governed by hyperbolic equations, but the equations for the 'macroscopic' variables such as amplitude, wave number, etc., are hyperbolic and it is the propagation of these important physical quantities that is described by the usual theory of characteristics. After all, the Boltzmann equation for a gas is not hyperbolic either and sound waves only appear from appropriate averaged equations. Indeed, the present work is in much the same spirit as the derivation of continuum fluid mechanics from kinetic theory. The Maxwell distribution describes a uniform state in terms of density ρ, temperature T, mean velocity $\mathbf{u}$. This is a solution for constant values of ρ, T, $\mathbf{u}$, but it is used to derive continuum equations for $\rho, T, \mathbf{u}$ when they are slowly varying functions relative to the mean free path and the collision time. The Maxwell distribution corresponds to (9) and ρ, T, $\mathbf{u}$ correspond to U, A_i.

One of the main features of non-linear hyperbolic equations is the possibility of shock waves. In the general sense used here, shocks are propagating discontinuities in the dependent variables. For the averaged equations, this means discontinuities in U, A_i and, therefore, in amplitude, wave number, etc. The most intriguing result, which arises naturally in the treatment of these shock waves, is that the frequency plays a role analogous to entropy even though there is no dissipation of energy in the system. Among other things, the frequency always increases as waves cross a shock. This concept is perhaps relevant to the discussion of so-called collisionless shocks in plasma physics. In very general terms, the point of view of this paper is similar to that of Camac, Kantrowitz, Litvak, Patrick & Petschek (1961) in their search for a collisionless shock. However, they try to develop a theory of *effective* dissipation in a random wave field with collisions between waves replacing collisions between particles.

The theory is developed first by using the non-linear wave equation (6) as an illustration, since it is the simplest equation of the appropriate type. Then the same

† The example in (6) is hyperbolic, but examples of non-hyperbolic equations are discussed later. Even for (6), the point can be made that the averaged equations are homogeneous in the derivatives. This is a great simplification and leads to the more important propagation speeds.

methods are applied to more interesting cases arising in water waves and plasma dynamics.

The discussion is restricted to one-dimensional waves, but it seems clear that the same approach could be used for more dimensions; a further comment is made in § 6.

2. Conservation equations and averaging

Various conservation equations of the form

$$\frac{\partial P}{\partial t}+\frac{\partial Q}{\partial x}=0 \tag{11}$$

can be deduced from the governing equations. For example, the conservation equations

$$\frac{\partial}{\partial t}\left(\tfrac{1}{2}\phi_t^2+\tfrac{1}{2}\phi_x^2+V(\phi)\right)+\frac{\partial}{\partial x}\left(-\phi_t\phi_x\right)=0, \tag{12}$$

$$\frac{\partial}{\partial t}\left(-\phi_x\phi_t\right)+\frac{\partial}{\partial x}\left(\tfrac{1}{2}\phi_t^2+\tfrac{1}{2}\phi_x^2-V(\phi)\right)=0, \tag{13}$$

can be deduced from equation (6). These conservation equations are averaged as follows to give the equations for the slowly varying functions U, A_i. Define the averaged value of any quantity as

$$\tilde{F}(x,t)=\frac{1}{2\xi}\int_{x-\xi}^{x+\xi}F(x',t)\,\mathrm{d}x'.$$

Then

$$\widetilde{\left(\frac{\partial P}{\partial t}\right)}=\frac{\partial\tilde{P}}{\partial t},\quad \widetilde{\left(\frac{\partial Q}{\partial x}\right)}=\frac{\partial\tilde{Q}}{\partial x},$$

the first of these being trivial and the second proved by

$$\widetilde{\left(\frac{\partial Q}{\partial x}\right)}=\frac{1}{2\xi}\int_{x-\xi}^{x+\xi}\frac{\partial Q}{\partial x'}(x',t)\,\mathrm{d}x'=\frac{1}{2\xi}[Q(x+\xi,t)-Q(x-\xi,t)]$$
$$=\frac{\partial}{\partial x}\frac{1}{2\xi}\int_{x-\xi}^{x+\xi}Q(x',t)\,\mathrm{d}x'=\frac{\partial\tilde{Q}}{\partial x}.$$

Equation (11) is averaged in this way to give

$$\frac{\partial\tilde{P}}{\partial t}+\frac{\partial\tilde{Q}}{\partial x}=0. \tag{14}$$

If the wave train is approximately uniform in the distance 2ξ, the mean quantities $\tilde{P}$ and $\tilde{Q}$ may be calculated from the uniform solution (9), holding U and A_i constant. Moreover, if the interval $x-\xi<x'<x+\xi$ includes a large number of waves these mean values will be functions of U, A_i alone: $\bar{P}(U,A_i)$, $\bar{Q}(U,A_i)$ say. The exact equation (14) is then replaced by the approximate equation

$$\frac{\partial}{\partial t}\bar{P}(U,A_i)+\frac{\partial}{\partial x}\bar{Q}(U,A_i)=0. \tag{15}$$

If λ_0 is the length scale for the wave length and L is the length scale for changes in U and A_i, i.e. U_x/U and A_{ix}/A_i are $O(L^{-1})$ then we choose

$$L \gg \xi \gg \lambda_0,$$

and the relative errors in replacing $\tilde{P}$ by $\overline{P}$ are $O(\lambda_0/\xi)$ and $O(\xi/L)$. As noted in § 1, L is x itself in the initial value problem. Since the uniform wave train is periodic with wavelength $\lambda = \lambda(U, A_i)$, the mean values $\overline{P}$, $\overline{Q}$ can be calculated over one wavelength, i.e.

$$\overline{P}(U, A) = \frac{1}{\lambda}\int_x^{x+\lambda} P(x', t)\,\mathrm{d}x'$$
$$= \frac{1}{\lambda}\int_0^{\lambda} \mathscr{P}\{\Phi(X;\; U, A_i)\}\,\mathrm{d}X, \tag{16}$$

where $\mathscr{P}\{\phi\}$ denotes the functional dependence of P on ϕ and the uniform solution (9) has been substituted for ϕ.

The choice of conservation equations as the equations to average is a natural one for physical reasons: important physical quantities such as energy, momentum, etc., satisfy conservation equations. We now see that it is natural from a mathematical point of view. Both terms in (14) are of the same order $O(\lambda_0/L)$. If an undifferentiated term R, say, were also present, a detailed solution more accurate than (9) would be needed to calculate R to order λ_0/L.

A crucial test of the soundness of the whole method is whether the number of independent conservation equations is equal to the number of variables U, A_i in the wave train solution (9). This turns out to be the case in all the examples considered and indeed some of the conservation equations were not at all obvious but were found only because of faith in this method! It is also easy to give cases where more conservation equations can be found than required, but, when averaged, only the requisite number prove to be independent. The proof of this equivalence between the number of conservation laws and the number of 'integration constants' U, A_i in (9) lies presumably in the transformation properties of the system. It is well known, for example, that conservation equations can be deduced from the invarience properties of a Lagrangian system (Noether 1918 or the review in Courant & Hilbert 1953, § 12·8). Equations (12) and (13) can be derived from the invariance of the Lagrangian density

$$\mathscr{L} = \tfrac{1}{2}\phi_t^2 - \tfrac{1}{2}\phi_x^2 - V(\phi)$$

with respect to translations of x and t. This kind of argument would, perhaps, lead to the proposed theorem.

Another remarkable result of a similar nature is that, in all the examples considered, the mean values $\overline{P}(U, A_i)$, $\overline{Q}(U, A_i)$ for all the conservation equations of the system can be expressed in terms of a *single* function $W(U, A_i)$ and its derivatives $\partial W/\partial U$, $\partial W/\partial A_i$. The function W is defined from (9) by

$$W(U, A_i) = \oint \Phi_X\,\mathrm{d}\Phi$$
$$= \oint \sqrt{\{F(\Phi;\; U, A_i)\}}\,\mathrm{d}\Phi, \tag{17}$$

where the integration is over one complete cycle of Φ. This is reminiscent of and was suggested by the adiabatic invariants† $\oint p\,\mathrm{d}q$ for Hamiltonian systems. The properties are also similar: for a Hamiltonian system with period τ, energy, E, and $I(E) = \oint p\,\mathrm{d}q$, it may be proved that $\tau = \mathrm{d}I/\mathrm{d}E$. For the wave problems it will be shown that frequency, wavelength, as well as the other important physical quantities, are given by the derivatives of W.

In the actual use of the set of averaged equations (15) in non-linear problems, the introduction of this function W is crucial because the averaged quantities $\overline{P}$, $\overline{Q}$ are complicated functions whose properties and relations to each other are not obvious otherwise. But again a deeper formulation leading logically to the introduction of this function has not yet been found. Some general Lagrangian–Hamiltonian approach is indicated.

[*Note added in proof*, 27 *October* 1964. Further progress has been made in this direction and will be described in a subsequent paper.]

3. Linear example $\phi_{tt} - \phi_{xx} + \phi = 0$

The uniform wave train is

$$\phi = a\cos(kx - \omega t), \quad \omega = \sqrt{(k^2+1)},$$

and a, k are a pair of basic parameters (equivalent to U and A). The conservation equations are

$$\frac{\partial}{\partial t}\left(\tfrac{1}{2}\phi_t^2 + \tfrac{1}{2}\phi_x^2 + \tfrac{1}{2}\phi^2\right) + \frac{\partial}{\partial x}\left(-\phi_x\phi_t\right) = 0, \tag{18}$$

$$\frac{\partial}{\partial t}\left(-\phi_x\phi_t\right) + \frac{\partial}{\partial x}\left(\tfrac{1}{2}\phi_t^2 + \tfrac{1}{2}\phi_x^2 - \tfrac{1}{2}\phi^2\right) = 0. \tag{19}$$

The mean values are

$$\overline{\tfrac{1}{2}\phi_t^2 + \tfrac{1}{2}\phi_x^2} = \tfrac{1}{4}(\omega^2 + k^2)\,a^2, \quad \overline{\tfrac{1}{2}\phi^2} = \tfrac{1}{4}a^2,$$
$$\overline{-\phi_x\phi_t} = \tfrac{1}{2}k\omega a^2.$$

Hence the slowly varying functions $a(x,t)$, $k(x,t)$ satisfy

$$\frac{\partial}{\partial t}\left\{\tfrac{1}{2}(k^2+1)\,a^2\right\} + \frac{\partial}{\partial x}\left(\tfrac{1}{2}k\omega a^2\right) = 0,$$

$$\frac{\partial}{\partial t}\left(\tfrac{1}{2}k\omega a^2\right) + \frac{\partial}{\partial x}\left(\tfrac{1}{2}k^2 a^2\right) = 0.$$

These reduce to

$$k_t + C(k)\,k_x = 0, \tag{20}$$

$$a_t + C(k)\,a_x + \tfrac{1}{2}aC'(k)\,k_x = 0, \tag{21}$$

where

$$C(k) = \frac{\mathrm{d}\omega}{\mathrm{d}k} = \frac{k}{\sqrt{(k^2+1)}}$$

† This is of course a standard topic in many books of dynamics, but the writer found the point of view in Landau & Lifshitz (1960) particularly relevant to the present method.

is the group velocity. The characteristic velocities of the equation for ϕ are ± 1, the phase speed $U > 1$, the group speed $C < 1$ and $UC = 1$. It should be noted that (20) and (21) are non-linear even though the original equation for ϕ is linear.

A curve $\mathrm{d}x/\mathrm{d}t = C(k)$ is a double characteristic for the pair of equations (20), (21) and

$$\frac{\mathrm{d}k}{\mathrm{d}t} = 0, \quad \frac{\mathrm{d}a}{\mathrm{d}t} = -\tfrac{1}{2}C'(k)\,k_x\,a \quad \text{on} \quad \frac{\mathrm{d}x}{\mathrm{d}t} = C(k). \tag{22}$$

Thus k is constant on a characteristic $\mathrm{d}x/\mathrm{d}t = C(k)$ or, in other words, k remains constant for an observer moving with velocity $C(k)$. Moreover, if Δx is the distance between two neighbouring characteristics,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(a^2\Delta x) &= 2a\frac{\mathrm{d}a}{\mathrm{d}t}\Delta x + a^2\Delta C \\ &= \left\{2a\frac{\mathrm{d}a}{\mathrm{d}t} + a^2C'(k)\,k_x\right\}\Delta x + O(\Delta x)^2 \\ &= 0,\end{aligned}$$

from (22). Since a^2 is proportional to the energy density, the total energy between two neighbouring characteristics remains constant. In this sense energy propagates with the group velocity.

If the initial disturbance is limited to a finite interval around $x = 0$, then, for sufficiently large x and t,

$$x/t = C(k);$$

hence

$$\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{a}{2t}, \quad a = \frac{a_0(k)}{t^{\frac{1}{2}}}.$$

These agree with the asymptotic results (4) and (3) obtained from the exact solution. The expression of $a_0(k)$ in terms of $F(k)$ and $\omega(k)$ has been discussed in previous papers (Whitham 1961; Broer 1950).

It may be noted that (20) can also be written as

$$k_t + \omega_x = 0, \tag{23}$$

which is a statement of the conservation of waves, k being the density of waves and ω the flux of waves. This is often assumed and made the basis for the derivation of the group velocity. Here it has been deduced from the averaged form of the equations governing the system. To some extent conservation of waves is tacitly assumed in the whole idea of the averaging technique, but the fact that it is obtained only as a consequence of the other basic conservation equations is important in the discussion of shocks in § 6.

In this problem there are an infinite number of possible conservation equations because any derivative of ϕ satisfies the same equation as ϕ and, hence, (18) and (19) also hold with ϕ replaced by ϕ_t, ϕ_x or any higher derivative of ϕ. However, when averaged, there are only two independent equations for a and k and these are (20) and (21). The existence of an infinite set of conservation equations is clearly a feature common to all linear systems.

4. Non-linear example $\phi_{tt} - \phi_{xx} + V'(\phi) = 0$

The uniform wave train is $\phi = \Phi(X)$ where

$$\tfrac{1}{2}(U^2-1)\,\Phi_X^2 + V(\Phi) = A.$$

The conservation equations are (12) and (13), and the mean values of the quantities appearing in them are calculated essentially from the function $W(U,A)$ of (17). In this case, it is convenient to take

$$\begin{aligned} W(U,A) &= (U^2-1)\oint \Phi_X\,\mathrm{d}\Phi \\ &= \sqrt{\{2(U^2-1)\}}\oint \sqrt{\{A-V(\Phi)\}}\,\mathrm{d}\Phi \\ &= \sqrt{(U^2-1)}\,G(A), \quad \text{say.} \end{aligned} \tag{24}$$

In the uniform solution, Φ oscillates between simple zeros $\Phi_1(A)$, $\Phi_2(A)$ of $A - V(\Phi)$, and the integral may be written

$$G(A) = 2^{\frac{3}{2}}\int_{\Phi_1}^{\Phi_2} \sqrt{\{A-V(\Phi)\}}\,\mathrm{d}\Phi.$$

For the linear case, which is also the approximation for small amplitude,

$$V = \tfrac{1}{2}\Phi^2, \quad \Phi_1 = -\sqrt{(2A)}, \quad \Phi_2 = \sqrt{(2A)} \quad \text{and} \quad G(A) = 2\pi A. \tag{25}$$

The wavelength λ is calculated in terms of W as follows:

$$\begin{aligned} \lambda &= \int_0^{\lambda} \mathrm{d}X = \oint \frac{\mathrm{d}\Phi}{\Phi_X} = \sqrt{\{\tfrac{1}{2}(U^2-1)\}}\oint \frac{\mathrm{d}\Phi}{\sqrt{\{A-V(\Phi)\}}} \\ &= \partial W/\partial A. \end{aligned}$$

Then the mean values of the various quantities $\mathscr{P}\{\phi\}$ are calculated from

$$\begin{aligned} \overline{\mathscr{P}\{\phi\}} &= \frac{1}{\lambda}\int_0^{\lambda} \mathscr{P}\{\Phi\}\,\mathrm{d}X \\ &= \kappa \oint \frac{\mathscr{P}\{\Phi\}\,\mathrm{d}\Phi}{\Phi_X}, \end{aligned}$$

where $\kappa = 1/\lambda$. Therefore

$$\begin{aligned} \overline{\tfrac{1}{2}\phi_t^2 + \tfrac{1}{2}\phi_x^2} &= \tfrac{1}{2}\kappa(U^2+1)\oint \Phi_X\,\mathrm{d}\Phi \\ &= \tfrac{1}{2}\kappa\frac{U^2+1}{U^2-1}W, \end{aligned}$$

and, similarly

$$\overline{V(\phi)} = A - \tfrac{1}{2}\kappa W, \quad \overline{-\phi_t\phi_x} = \frac{\kappa U}{U^2-1}W.$$

The averaged equations for the various examples fit a certain canonical form, which may be obtained in this case by judicious use of the identities

$$W_U = \frac{U}{U^2-1}W, \quad \kappa W_A = 1.$$

The choice is

$$\overline{\tfrac{1}{2}\phi_t^2+\tfrac{1}{2}\phi_x^2+V(\phi)} = \kappa(UW_U+AW_A-W),$$

$$\overline{-\phi_x\phi_t} = \kappa W_U \quad\text{or}\quad \kappa U(UW_U+AW_A-W)-UA,$$

$$\overline{\tfrac{1}{2}\phi_t^2+\tfrac{1}{2}\phi_x^2-V(\phi)} = \kappa UW_U-A,$$

and the averaged forms of (12) and (13) are

$$\frac{\partial}{\partial t}\{\kappa(UW_U+AW_A-W)\}+\frac{\partial}{\partial x}\{\kappa U(UW_U+AW_A-W)-UA\} = 0, \tag{26}$$

$$\frac{\partial}{\partial t}(\kappa W_U)+\frac{\partial}{\partial x}(\kappa UW_U-A) = 0, \tag{27}$$

where
$$\kappa = 1/W_A. \tag{28}$$

Equation (26) may be expanded as

$$U\left\{\frac{\partial}{\partial t}(\kappa W_U)+\frac{\partial}{\partial x}(\kappa UW_U-A)\right\}+A\left\{\frac{\partial}{\partial t}(\kappa W_A)+\frac{\partial}{\partial x}(\kappa UW_A-U)\right\}-W\left\{\frac{\partial \kappa}{\partial t}+\frac{\partial}{\partial x}(\kappa U)\right\} = 0,$$

and from (27) and (28) this reduces to

$$\frac{\partial \kappa}{\partial t}+\frac{\partial}{\partial x}(\kappa U) = 0. \tag{29}$$

Since κU is the frequency, the equation for conservation of waves is again obtained as a consequence of the basic conservation equations. Equations (27) and (29) may then be simplified to

$$\left.\begin{aligned}\frac{DW_A}{Dt}-W_A\frac{\partial U}{\partial x} &= 0,\\ \frac{DW_U}{Dt}-W_A\frac{\partial A}{\partial x} &= 0,\end{aligned}\right\} \tag{30}$$

where
$$\frac{D}{Dt} = \frac{\partial}{\partial t}+U\frac{\partial}{\partial x}.$$

It turns out that (30) is a hyperbolic system, so the characteristic form is important. After substitution for W from (24) and some manipulation, the characteristic form is found to be

$$\frac{\mathrm{d}U}{U^2-1}\mp\sqrt{\left(\frac{-G''}{G}\right)}\,\mathrm{d}A = 0 \tag{31}$$

on the characteristic curves

$$C_\pm:\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{1\pm U\sqrt{(-GG''/G'^2)}}{U\pm\sqrt{(-GG''/G'^2)}}. \tag{32}$$

For functions $V(\phi)$ with appropriate properties corresponding to stability of the system, $G''\leqslant 0$, and $G''\to 0$ in the limit of small amplitudes (for which $A\to 0$).

5. Non-linear generalization of group velocity

For a linear system the characteristic velocities for the two averaged equations coincide and are equal to the group velocity. For a non-linear system, as the above example shows, the characteristic velocities are in general unequal. These characteristic velocities are considered to be the non-linear generalization of the group velocities. They are the velocities of propagation of changes in a wave train.

Another possible generalization would be to define the group velocity as the velocity of energy flow, i.e. ratio of energy flux to energy density. For the above example, the ratio would be (see (26))

$$U - \frac{UAW_A}{UW_U + AW_A - W} = \frac{UG}{(U^2-1)\,AG' + G}. \tag{33}$$

For the linearized limit of small amplitude, $G = 2\pi A$ (see (25)), and the three velocities in (32) and (33) coincide in the group velocity $C = 1/U$. However, in the non-linear case, the three velocities are not equal, and the characteristic velocities (32) are the basic ones from a mathematical point of view.

The result that the characteristic velocities of the averaged equations are not equal in the non-linear theory has important consequences in the solutions. The solutions may be of a completely different nature from the linear case. Consider, for example, a wave train which is initially uniform with $A = A_0$, $U = U_0$ outside some finite region. Then after some interaction period the disturbance will separate into two simple waves.†

One simple wave is on the characteristics C_+ of (32) and the Riemann invariant from (31) is

$$\int_{U_0}^{U} \frac{\mathrm{d}U}{U^2-1} + \int_{A_0}^{A} \sqrt{\left(\frac{-G''}{G}\right)}\,\mathrm{d}A = 0;$$

the values of U and A are constant along the C_+ characteristics in this simple wave. The other simple wave is on the characteristics C_- and the Riemann invariant has the other sign. Between the two simple waves, U and A take again their undisturbed values U_0 and A_0. The wave number κ and the amplitude a can be calculated in terms of U and A so the same qualitative statements apply to κ and a.

In contrast, the corresponding equations for the linear system, (20) and (21), have only one family of characteristics and there will be no such separation. Moreover, in the simple waves of the non-linear theory the amplitude remains constant along the characteristics, whereas it decays like $t^{-\frac{1}{2}}$ in the linear theory.

A wave packet with no waves outside the disturbed region, i.e. $A_0 = 0$, would not show this separation because the linear theory holds in the limit $A \to 0$. However, there would be non-linear distortion in the wave packet.

6. Shock waves

The dependence of the characteristic velocities on the amplitude leads to multivalued solutions in exact analogy with the breaking of waves in gas dynamics. In gas dynamics multivalued solutions are meaningless and correspond to a breakdown

† The terminology and results are taken over from the corresponding problem in gas dynamics (see Courant & Friedrichs 1948).

in the inviscid equations. But the solution can be saved without recourse to the more realistic viscous equations by fitting in discontinuous shock waves. We can follow the same steps here: argue that the multivalued solution corresponds to a breakdown of the approximate equations (26), (27), but avoid recourse to the original equation (6) and complete the solution of (26) and (27) by fitting in discontinuities in U and A according to appropriate jump conditions. The jump conditions will be derived below. We must first note that in the present case a multivalued solution may well correspond to a perfectly genuine overlap with two parts of the wave train superimposed and this may be what really occurs. Certainly, in linear theory, if two wave trains moving at different speeds meet, they just pass through each other. However, if the non-linear effects can be strong enough and the speeds differ only slightly so that the interaction takes place at a slow rate over a long time (this being the case for shock formation out of a continuous solution), it is just conceivable that the interaction is confined to a narrow region between the wave trains. In analogy with gas dynamics we call such a region a shock and treat the changes of U and A across it as discontinuities in the solution of the averaged equations (26) and (27). A more detailed analysis would require more accurate approximations to the original equation (6).

It is suggested, then, that discontinuous shocks of this type may possibly occur in non-linear wave trains if the changes are relatively small, but that large changes would lead to overlap and superposition of different parts of the wave train. This is frankly speculative. Some support for this view comes from recent work by D. J. Benney & J. C. Luke (1963, private communication) on non-linear water waves. They consider the two-dimensional problem of the oblique interaction between two cnoidal wave trains incident at angle χ, say, If χ is not too small, the waves pass through each other and the non-linear interaction can be handled by perturbation methods. However, as $\chi \to 0$, the perturbations become infinite and the method breaks down. The present theory, suitably extended to two dimensions and curved wave trains, would predict 'shocks' with a discontinuous change in the direction of the wave crests. It may be that the wave systems interpenetrate and overlap, as found by Benney & Luke, if the change in direction is large, but be separated by a relatively narrow shock of the type discussed here when the angle is small. This could perhaps be tested experimentally.

These shocks also seem to have some bearing on the following question which has arisen in water waves and plasmas: Can a shock-like change of state occur in a reversible system which has absolutely no dissipation? In water waves, the problem is the existence of a bore-like increase in depth *without* appeal to turbulent dissipation or friction. In a plasma it is the possibility of an increase in density over a distance small compared with the mean free path, so that one can not appeal to the mechanism of dissipation in collisions. The shocks proposed here have no dissipation and (as will be shown) an increase of frequency plays the role of a kind of irreversibility.

For all these reasons, it seemed worth while to include a brief discussion of these peculiar shocks even though some alternative solution—overlap or dissipation—may always appear in practice.

There is a standard mathematical theory for the treatment of shock waves for a system of conservation equations (see, for example, Courant–Hilbert 1962, § V. 9). For each conservation equation

$$\frac{\partial f}{\partial t}+\frac{\partial g}{\partial x}=0,$$

there is a corresponding shock condition

$$[g]=V[f], \tag{34}$$

connecting the discontinuities $[f]$, $[g]$ in f, g and the shock velocity V. This jump condition expresses conservation of the quantity with density f and flux g across the shock wave. There is, however, a question of uniqueness that is crucial. A physical system in n unknowns usually has at least $(n+1)$ conservation laws and only n shock conditions must be chosen. Mathematically, any choice of n conditions leads to a satisfactory 'weak solution', but only one choice corresponds to the real physical situation. The correct choice is decided from which of the n quantities are actually conserved across the shock. For example, in the prototype of gas dynamics, the equations of mass, momentum and energy may be written as

$$\left.\begin{aligned}\frac{\partial}{\partial t}\rho+\frac{\partial(\rho u)}{\partial x}&=0,\\ \frac{\partial(\rho u)}{\partial t}+\frac{\partial}{\partial x}(\rho u^2+p)&=0,\\ \frac{\partial}{\partial t}\left(\tfrac{1}{2}\rho u^2+\frac{1}{\gamma-1}\frac{p}{\rho}\right)+\frac{\partial}{\partial x}\left(\tfrac{1}{2}\rho u^3+\frac{\gamma}{\gamma-1}\frac{p}{\rho}\right)&=0.\end{aligned}\right\} \tag{35}$$

But there is an additional conservation equation that can be deduced from these:

$$\frac{\partial}{\partial t}(\rho S)+\frac{\partial}{\partial x}(\rho u S)=0, \tag{36}$$

where the entropy $S(p,\rho)=\ln p/\rho^{\gamma}$. Mathematically, the shock condition (34) corresponding to any three of these four equations may be chosen. Physically, we know that mass, momentum and energy are conserved across the shock but entropy is not. Hence shock conditions are chosen corresponding to the three equations in (35). It is then shown that

$$[\rho u S]\neq V[\rho S].$$

The differential equations (35) imply (36) but the jump conditions for (35) do not imply the jump condition for (36). In fact the condition that the shock be formed by the breaking of a continuous solution, i.e. that its velocity is supersonic relative to the flow ahead and subsonic relative to the flow behind, leads to the result that the entropy always increases across the shock in accord with the second law of thermodynamics.

In the present case with two variables, U and A, there are two conservation equations (26) and (27) corresponding to (35), and an extra one (29) can be deduced from them, corresponding to (36). In a shock wave, the averaging breaks down because of the rapid changes but the original equation still holds and the detailed conservation equations (12) and (13) still hold across the shock. Therefore the shock conditions corresponding to (26) and (27) should be chosen. But (29) was a consequence of the averaging and it could only be deduced from the averaged conservation equations. Hence, it should not be chosen for one of the shock conditions. Thus, provided the original equation $\phi_{tt}-\phi_{xx}+V(\phi)=0$ still holds through the shock, the correct jump conditions for the averaged equations (26), (27) and (29) are

$$[\kappa U(UW_U+AW_A-W)-UA] = V[\kappa(UW_U+AW_A-W)-W], \tag{37}$$

$$[\kappa UW_U-A] = V[\kappa W_U] \tag{38}$$

and

$$[\kappa U] \neq V[\kappa]. \tag{39}$$

It may then be shown from (37) and (38) and from the condition that the shock must be formed by the breaking of a continuous solution that *the frequency of the waves relative to the shock always increase as the waves cross the shock*. In this example, the shock is backwards facing and the waves overtake the shock from behind and waves appear ahead at a higher frequency. For a linear system, of course, these shocks are not required and it is verified that the jump in frequency reduces to zero in that case. Otherwise the jump in frequency is positive and tends to zero as the third power of the shock strength $[A]$. Thus, it is seen that the frequency plays a role analagous to entropy flux in gas dynamics. There is an irreversibility—the discontinuous increase in the frequency of waves—even though the original equation for ϕ is conservative and reversible.

After substitution for κ and W into (37) and (38), we have

$$\left[\frac{U}{U^2-1}\frac{G}{G'}\right] = V\left[\frac{1}{U^2-1}\frac{G}{G'}+A\right],$$

$$\left[\frac{U^2}{U^2-1}\frac{G}{G'}-A\right] = V\left[\frac{U}{U^2-1}\frac{G}{G'}\right].$$

These may be solved to express U_2 and V in terms of A_2, A_1, U_1 where subscripts 1 and 2 refer to values ahead and behind the shock, respectively.

7. The Korteweg–de Vries equation for water waves

In this and the two succeeding sections, the corresponding results for other examples are quoted briefly in order to show the generality of the ideas and to give further support to them. The consequences of the equations will be explored fully in a later paper.

Korteweg & de Vries (1895) studied water waves in the case of relatively long waves and derived the equation

$$\eta_t+\sqrt{(gh_0)}\,(1+\tfrac{3}{2}(\eta/h_0))\,\eta_x+\tfrac{1}{6}\sqrt{(gh_0)}\,h_0^2\,\eta_{xxx}=0,$$

for the elevation η of the water surface above the undisturbed depth h_0. This equation holds when a/h_0, h_0^2/λ^2 are comparable small quantities, where a is a typical amplitude and λ is a typical wavelength. The equation may be transformed into the form

$$\eta_t+6\eta\eta_x+\eta_{xxx}=0. \tag{40}$$

The uniform wave train, discussed by Korteweg & de Vries, is obtained by substituting $\eta = \eta(X)$, $X = x - Ut$. This leads to

$$\eta_{XXX} = U\eta_X - 6\eta\eta_X,$$

which can be integrated twice to give

$$\begin{aligned} \eta_{XX} &= B + U\eta - 3\eta^2, \\ \tfrac{1}{2}\eta_X^2 &= -A + B\eta + \tfrac{1}{2}U\eta^2 - \eta^3, \end{aligned} \tag{41}$$

where A, B are constants of integration. The solution can then be expressed in terms of the Jacobian elliptic function cn θ, and Korteweg & de Vries described the the waves as 'cnoidal waves'.

For the treatment of a slowly varying wave train given locally by (41), three conservation equations are required since the solution involves three parameters U, A, B. The following conservation equations can be derived from (40):

$$\left.\begin{aligned} \frac{\partial}{\partial t}\eta + \frac{\partial}{\partial x}(3\eta^2 + \eta_{xx}) &= 0, \\ \frac{\partial}{\partial t}(\tfrac{1}{2}\eta^2) + \frac{\partial}{\partial x}(2\eta^3 + \eta\eta_{xx} - \tfrac{1}{2}\eta_x^2) &= 0, \\ \frac{\partial}{\partial t}(\eta^3 - \tfrac{1}{2}\eta_x^2) + \frac{\partial}{\partial x}(\tfrac{9}{2}\eta^4 + 3\eta^2\eta_{xx} + \tfrac{1}{2}\eta_{xx}^2 + \eta_x\eta_t) &= 0. \end{aligned}\right\} \tag{42}$$

These do not have the simplicity of form one expects in physical conservation equations because (40) is only an approximation to the full equations of water waves. In fact, equations (42) are essentially the corresponding approximations to the equations for conservation of mass, momentum and energy.

The calculation of the average values of the quantities appearing in (42) from the solution (41), is again simplified by introduction of an auxiliary function. It is convenient to define

$$\begin{aligned} W(A,B,U) &= -\oint \eta_X \, d\eta \\ &= -\sqrt{2}\oint \sqrt{(-A + B\eta + \tfrac{1}{2}U\eta^2 - \eta^3)}\, d\eta. \end{aligned} \tag{43}$$

Then, the wave number is given by

$$\kappa = \frac{1}{\lambda} = \frac{1}{W_A}, \tag{44}$$

and the averaged equations corresponding to (42) become

$$\left.\begin{aligned} \frac{\partial}{\partial t}(\kappa W_B) + \frac{\partial}{\partial x}(\kappa U W_B - B) &= 0, \\ \frac{\partial}{\partial t}(\kappa W_U) + \frac{\partial}{\partial x}(\kappa U W_U - A) &= 0, \\ \frac{\partial}{\partial t}\{\kappa(AW_A + BW_B + UW_U - W)\} & \\ + \frac{\partial}{\partial x}\{\kappa U(AW_A + BW_B + UW_U - W) - \tfrac{1}{2}B^2 - AU\} &= 0. \end{aligned}\right\} \tag{45}$$

As before, the equation for conservation of waves,

$$\frac{\partial \kappa}{\partial t}+\frac{\partial}{\partial x}(U\kappa) = 0, \tag{46}$$

follows from (45). Then the equations may be simplified to

$$\left.\begin{aligned}\frac{\mathrm{D}W_A}{\mathrm{D}t}-W_A\frac{\partial U}{\partial x} &= 0,\\ \frac{\mathrm{D}W_B}{\mathrm{D}t}-W_A\frac{\partial B}{\partial x} &= 0,\\ \frac{\mathrm{D}W_U}{\mathrm{D}t}-W_A\frac{\partial A}{\partial x} &= 0,\end{aligned}\right\} \tag{47}$$

where
$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t}+U\frac{\partial}{\partial x}.$$

In spite of the simple appearance of equations (47), it is a complicated calculation to find the characteristic velocities. In this calculation the zeros (α, β, γ) of the cubic

$$\eta^3-\tfrac{1}{2}U\eta^2-B\eta+A,$$

appearing in (43), are used as variables in place of (A, B, U). Then,

$$W = -2^{\frac{3}{2}}\int_\beta^\alpha \surd\{(\alpha-\eta)(\eta-\beta)(\eta-\gamma)\}\,\mathrm{d}\eta \quad (\alpha>\beta>\gamma), \tag{48}$$

and equations (47) can be written in terms of W_α, W_β, W_γ. After considerable manipulation it may be shown that

$$\frac{\mathrm{D}}{\mathrm{D}t}(\beta+\gamma)+2W_A\frac{(\gamma-\alpha)(\alpha-\beta)(\beta-\gamma)}{W_\beta-W_\gamma}\frac{\partial}{\partial x}(\beta+\gamma) = 0$$

with two similar equations obtained by cyclic permutation. Hence, $\beta+\gamma$ = constant on the characteristic

$$\frac{\mathrm{d}x}{\mathrm{d}t} = U+2W_A\frac{(\gamma-\alpha)(\alpha-\beta)(\beta-\gamma)}{W_\beta-W_\gamma}.$$

The three characteristic velocities may be evaluated as

$$U-\frac{4aK}{K-E}, \quad U-\frac{4a(1-s^2)K}{E-(1-s^2)K}, \quad U+4a\frac{(1-s^2)K}{s^2E}, \tag{49}$$

where
$$a = \frac{\alpha-\beta}{2}, \quad s^2 = \frac{\alpha-\beta}{\alpha-\gamma}, \quad U = 2(\alpha+\beta+\gamma),$$

and $K(s^2)$, $E(s^2)$ are the complete elliptic integrals of the first and second kind. The wave number is

$$\kappa = W_A^{-1} = \frac{a^{\frac{1}{2}}}{2sK},$$

and the mean height is given by

$$\overline{\eta} = \alpha-\frac{2a}{s^2}\frac{K-E}{K}.$$

Therefore, s is a function of a/κ^2 and the propagation velocities in (49) are of the form

$$U+f(a/\kappa^2).$$

The limit $a/\kappa^2 \to 0$ should reproduce the results for the linear equation

$$\eta_t+\eta_{xxx} = 0. \tag{50}$$

It is easily shown that the velocities (49) reduce in this limit to

$$-3(2\pi\kappa)^2, \quad -3(2\pi\kappa)^2, \quad 0. \tag{51}$$

Thus the first two reduce correctly to the group velocity for (50). The appearance of the third velocity can be seen directly in the linear theory. The uniform solution of (50) is

$$\eta = b+a\sin\{2\pi\kappa x-(2\pi\kappa)^3 t\},$$

and the averaged equations of the linear problems give $\partial b/\partial t = 0$ in accordance with the zero velocity in (51); of course, the addition of a constant mean value to the solution is usually neglected as trivial in linear theory.

The shock conditions for the propagation of discontinuous jumps in A, B, U (or equivalently a, κ, $\overline{\eta}$) are obtained in the standard way from (45):

$$\left.\begin{aligned} [\kappa U W_B - B] &= V[\kappa W_B], \\ [\kappa U W_U - A] &= V[\kappa W_U], \\ [\kappa U(A W_A + B W_B + U W_U - W) - \tfrac{1}{2}B^2 - AU] &= V[\kappa(A W_A + B W_B + U W_U - W)]. \end{aligned}\right\} \tag{52}$$

In the special case of no waves ahead of the shock, the amplitude is taken to be zero and the wave number ahead, which is arbitrary, drops out of these conditions.

Conservation of waves, equation (46), is not used across such a shock and there will be a jump in relative frequency across the shock. This is assuming that the original equation (40) and the conservation laws (42) still hold inside the shock, only the averaging breaks down. An alternative is to assume that the original equations are invalid, and, for example, that there is loss of energy into some process neglected in the original formulation of the equations. Then it may be appropriate to choose conservation of waves across the shock, i.e.

$$[\kappa U] = V[\kappa], \tag{53}$$

and drop one of the conditions in (52). This would correspond to the discussion of bores given by Benjamin & Lighthill (1954). They assume steady flow throughout with $U = V$ and allow an arbitrary loss of energy into turbulent motion at the bore. They originally hoped to find *steady* smooth transitions in level without any loss of mass, momentum and energy. From the point of view of the approach given here it was equivalent to looking for shocks satisfying all four relations in (52) and (53). This is over determined, of course; one of the relations must be dropped since the system is only third order. They dropped energy conservation, but one could also drop (53). In practice, weak bores certainly look to be steady flows so that the smooth transition is probably linked with loss of both momentum and energy by frictional forces.

8. The Boussinesq equations for water waves

Before the work of Korteweg & de Vries, Boussinesq (1877) had derived the equations

$$\left.\begin{aligned} h_t + uh_x + hu_x &= 0, \\ u_t + uu_x + gh_x + \nu h_{xtt} &= 0, \end{aligned}\right\} \tag{54}$$

where h is the depth, u is the velocity, $\nu = \frac{1}{3}h_0$, where h_0 is the initial undisturbed depth, and g is the acceleration of gravity. Again these hold when a/h_0 and h_0^2/λ^2 are comparable small quantities. When $1 \gg a/h_0 \gg h_0^2/\lambda^2$, the third derivative term can be neglected and the equations reduce to those of the familiar 'shallow water theory'. Roughly speaking, equations (54) describe waves moving in the positive and negative x directions with speed $\sqrt{(gh)}$. The Korteweg–de Vries equation is the further approximation (with one approximate integration) to study waves moving with velocity $\sqrt{(gh)}$ in the positive direction only.

The uniform wave train involves four parameters, and four conservation equations can be derived from (54). The conservation equations are

$$\left.\begin{aligned} h_t + (uh)_x &= 0, \\ u_t + (\tfrac{1}{2}u^2 + gh + \nu h_{tt})_x &= 0, \\ (hu - \nu h_x h_t)_t + (hu^2 + \tfrac{1}{2}gh^2 + \nu h h_{tt} + \tfrac{1}{2}\nu h_t^2)_x &= 0, \\ (\tfrac{1}{2}hu^2 + \tfrac{1}{2}gh^2 + \tfrac{1}{2}\nu h_t^2)_t + (\tfrac{1}{2}hu^3 + gh^2u + \nu uh h_{tt})_x &= 0. \end{aligned}\right\} \tag{55}$$

The first, third and fourth of these equations are conservation of mass, momentum and energy, respectively. The second one has, apparently, no such general physical interpretation. A fourth one is required by this approach and is, indeed, found.

The uniform wave train is found by substituting

$$u = u(X), \quad h = h(X), \quad X = x - Ut,$$

in (54), but the constants of integration are most conveniently introduced to fit (55). For the uniform solution, equations (55) integrate to

$$\left.\begin{aligned} hu - hU &= C, \\ \tfrac{1}{2}u^2 + gh + \nu U^2 h_{XX} - Uu &= B, \\ hu^2 + \tfrac{1}{2}gh^2 + \nu U^2(h h_{XX} + \tfrac{1}{2}h_X^2) - U(hu + \nu U h_X^2) &= A \\ \text{and} \quad \tfrac{1}{2}hu^3 + gh^2u + \nu U^2 uh h_{XX} - U(\tfrac{1}{2}hu^2 + \tfrac{1}{2}gh^2 + \tfrac{1}{2}\nu U^2 h_X^2) &= AU + BC, \end{aligned}\right\} \tag{56}$$

where A, B, C are constants of integration. Elimination of u and h_{XX} leads to

$$\tfrac{1}{2}\nu U^2 h_X^2 = \tfrac{1}{2}C^2/h + UC - A + (B + \tfrac{1}{2}U^2)\,h - \tfrac{1}{2}gh^2. \tag{57}$$

In this case the average values of the quantities appearing in (55) are calculated in terms of

$$\begin{aligned} W(A,B,C,U) &= -\nu U^2 \oint h_X \,\mathrm{d}h \\ &= -\sqrt{(\nu U^2)} \oint \sqrt{\{C^2/h + 2(UC - A) + (2B + U^2)\,h - gh^2\}}\,\mathrm{d}h. \end{aligned} \tag{58}$$

The averaged equations (55) are found to be

$$\left.\begin{aligned}\frac{\partial}{\partial t}(\kappa W_B)+\frac{\partial}{\partial x}(\kappa U W_B-C) &= 0,\\ \frac{\partial}{\partial t}(\kappa W_C)+\frac{\partial}{\partial x}(\kappa U W_C-B) &= 0,\\ \frac{\partial}{\partial t}(\kappa W_U)+\frac{\partial}{\partial x}(\kappa U W_U-A) &= 0,\\ \frac{\partial}{\partial t}\{\kappa(AW_A+BW_B+CW_C+UW_U-W)\}&\\ +\frac{\partial}{\partial x}\{\kappa U(AW_A+BW_B+CW_C+UW_U-W)-AU-BC\} &= 0.\end{aligned}\right\}\tag{59}$$

where $\kappa = 1/W_A$.

As in the previous case, the equation for the conservation of waves

$$\kappa_t+(\kappa U)_x = 0$$

is deduced, and the set can be simplified to

$$\left.\begin{aligned}\frac{\mathrm{D}W_A}{\mathrm{D}t}-W_A\frac{\partial U}{\partial x} &= 0,\\ \frac{\mathrm{D}W_B}{\mathrm{D}t}-W_A\frac{\partial C}{\partial x} &= 0,\\ \frac{\mathrm{D}W_C}{\mathrm{D}t}-W_A\frac{\partial B}{\partial x} &= 0,\\ \frac{\mathrm{D}W_U}{\mathrm{D}t}-W_A\frac{\partial A}{\partial x} &= 0.\end{aligned}\right\}\tag{60}$$

The shock conditions are deduced from (59) by replacing each $\partial Q/\partial x$ by $[Q]$ and each $\partial P/\partial t$ by $-V[P]$.

The consequences of these equations have not yet been investigated further. It is noted that the general ideas go through in a consistent way and that the averaged equations (59) and (60) follow the standard pattern laid down by the previous two examples.

9. An example from plasma dynamics

Dispersive waves occur in various problems of plasma dynamics, governed by various mathematical formulations. Theoretical interest has centred particularly around collisionless plasmas because of the novelty of the effects. A relatively simple formulation is obtained by neglecting, in addition, the spread of the electron velocity distribution about the mean electron velocity and the spread of the ion velocity distribution about the mean ion velocity; in this sense the plasma is 'cold'. The equations will not be derived here since the object in this paper is to show the generality of the mathematical treatment rather than to discuss the applications. The equations have been derived by Davis, Lüst & Schlüter (1958) and Adlam &

Allen (1958). For propagation across a magnetic field the equations may be reduced to

$$\left.\begin{aligned} \rho_t + u\rho_x + \rho u_x &= 0,\\ u_t + uu_x + vH &= 0,\\ v_t + uv_x + E - uH &= 0,\\ H_t + E_x &= 0,\\ H_x - \rho v &= 0. \end{aligned}\right\} \tag{61}$$

These are in non-dimensional form. The total density of particles is ρ, the electron velocity is $\{u, v(m_i/m_e)^{\frac{1}{2}}, 0\}$, the ion velocity is $\{u, -v(m_e/m_i)^{\frac{1}{2}}, 0\}$, the magnetic and electric fields are $(0, 0, H)$ and $\{-vH(m_i - m_e)/(m_i m_e)^{\frac{1}{2}}, E, 0\}$, where m_e, m_i are the electron and ion masses. There is no net mass motion in the y direction, but there is a net transfer of charge, i.e. a current, measured by v. The first equation in (61) is conservation of mass; the second and third are momentum equations including the Lorenz forces but pressure terms are absent because the plasma is cold; the last two equations are Maxwell's equations with current density ρv and with the displacement current neglected.

In a first attempt to apply the theory to the system (61), five parameters appeared in the uniform wave train but only four conservation equations were found. This inconsistency led to the realization that (61) is not a purely dispersive system in the sense used in this paper. A characteristic equation, namely

$$\frac{\partial}{\partial t}\left(\frac{H - v_x}{\rho}\right) + u\frac{\partial}{\partial x}\left(\frac{H - v_x}{\rho}\right) = 0$$

can be derived from (61). This shows that $(H - v_x)/\rho$ is constant along particle paths. If the flow started from uniform conditions at $t = 0$ or at infinity, this quantity will be constant throughout. By suitable scaling, this constant can be taken equal to unity, the set of equations (61) reduces to the fourth-order system:

$$\left.\begin{aligned} \rho_t + u\rho_x + \rho u_x &= 0,\\ u_t + uu_x + vH &= 0,\\ H_x - \rho v &= 0,\\ H &= \rho + v_x, \end{aligned}\right\} \tag{62}$$

and the electric field is given by

$$E = \rho u - v_t. \tag{63}$$

There are now only four parameters in the uniform solution and four conservation equations for the system (62) are found to be

$$\left.\begin{aligned} \rho_t + (\rho u)_x &= 0,\\ u_t + (\tfrac{1}{2}u^2 + \tfrac{1}{2}v^2 + H)_x &= 0,\\ (\rho u)_t + (\rho u^2 + \tfrac{1}{2}H^2)_x &= 0,\\ (\tfrac{1}{2}\rho u^2 + \tfrac{1}{2}\rho v^2 + \tfrac{1}{2}H^2)_t + (\tfrac{1}{2}\rho u^3 + \tfrac{1}{2}\rho u v^2 + EH)_x &= 0. \end{aligned}\right\} \tag{64}$$

The first, third and fourth of these are conservation of mass, momentum and energy but the second one is somewhat unexpected as was the corresponding one in (55).

In the solution for the uniform wave train all the variables are functions of $X = x - Ut$, and it is convenient to introduce the three parameters A, B, C, required in addition to U as the constants of integration of the first three equations in (64). Then

$$\left.\begin{aligned} \rho(u-U) &= C, \\ \tfrac{1}{2}u^2+\tfrac{1}{2}v^2+H-Uu &= B, \\ \rho u^2+\tfrac{1}{2}H^2-U\rho u &= A, \\ \tfrac{1}{2}\rho u^3+\tfrac{1}{2}\rho uv^2+EH-U(\tfrac{1}{2}\rho u^2+\tfrac{1}{2}\rho v^2+\tfrac{1}{2}H^2) &= UA+BC. \end{aligned}\right\} \tag{65}$$

Substituting $v = H_X/\rho$ from (62) and eliminating ρ and u, we obtain a single equation for H:

$$H_X^2 = C^2\left\{\frac{(2B+U^2-2H)\,C^2}{(A-UC-\frac{1}{2}H^2)^2}-1\right\}. \tag{66}$$

From this expression, the function W is defined as

$$\begin{aligned} W(A,B,C,U) &= -\oint \frac{H_X}{\rho}\,\mathrm{d}H \\ &= -\oint \surd\{2B+U^2-2H-C^{-2}(A-UC-\tfrac{1}{2}H^2)^2\}\,\mathrm{d}H. \end{aligned} \tag{67}$$

The conservation equations (64) are then averaged, evaluating the mean quantities in terms of W. The resulting equations are identical with (59), the only difference between the averaged problems is the difference in the choice of W. This coincidence is explained in the next section.

The shocks are of special interest in this particular problem in view of the considerable discussion as to whether any kind of shock transition is possible at all in a collisionless plasma. Here the shock would be a discontinuous change in the functions A, B, C, U (hence in mean density, amplitude, etc.) without any loss of the conserved quantities in (64) and (65). It would be essentially unsteady—previous attempts have concentrated on steady solutions—since the relative frequency of the waves increases across the shock and the waves behind the shock are moving away from the discontinuity, i.e. the phase velocity U is not equal to the shock velocity V. As remarked earlier in § 7, the case with no wave train ahead of the shock is included. From (59), the shock conditions are

$$\begin{aligned} [\kappa UW_B - C] &= V[\kappa W_B], \\ [\kappa UW_C - B] &= V[\kappa W_C], \\ [\kappa UW_U - A] &= V[\kappa W_U], \\ [\kappa U(AW_A+BW_B+CW_C+UW_U-W)-AU-BC] &\\ &= V[\kappa(AW_A+BW_B+CW_C+UW_U-W)]. \end{aligned}$$

The analysis of these will be complicated and has not yet been carried out. A final judgement on their relevance to the collisionless shock problem can not be given until that has been done.

10. Systems of conservation equations

The symmetric form of the averaged equations (26), (27), (45) and (59), and especially the appearance of the same set (59) for two apparently widely different problems, can be explained as follows. In each case we obtain eventually $n+1$ conservation laws for n quantities, the extra one being the conservation of waves (29). Such systems have special properties.

Let the n dependent variables be $u_1, \ldots, u_n$ and let the $n+1$ conservation laws be written

$$\frac{\partial f_i}{\partial t}+\frac{\partial g_i}{\partial x} = 0 \quad (i = 1, \ldots, n)$$

and

$$\frac{\partial f}{\partial t}+\frac{\partial g}{\partial x} = 0,$$

where f_i, g_i, f, g, are functions of the u_i. Since the last equation can be derived from an appropriate combination of the first n, there exist factors $a_i(u_1, \ldots, u_n)$ such that

$$\mathrm{d}f = a_i\,\mathrm{d}f_i, \quad \mathrm{d}g = a_i\,\mathrm{d}g_i.$$

Let $F = a_i f_i - f$, $G = a_i g_i - g$, then

$$\mathrm{d}F = f_i\,\mathrm{d}a_i, \quad \mathrm{d}G = g_i\,\mathrm{d}a_i$$

and

$$f_i = \frac{\partial F}{\partial a_i}, \quad g_i = \frac{\partial G}{\partial a_i}, \quad f = a_i\frac{\partial F}{\partial a_i} - F, \quad g = a_i\frac{\partial G}{\partial a_i} - G.$$

Therefore, the system of conservation equations can be written

$$\frac{\partial}{\partial t}\left(\frac{\partial F}{\partial a_i}\right)+\frac{\partial}{\partial x}\left(\frac{\partial G}{\partial a_i}\right) = 0 \quad (i = 1, \ldots, n)$$

and

$$\frac{\partial}{\partial t}\left(a_i\frac{\partial F}{\partial a_i} - F\right)+\frac{\partial}{\partial x}\left(a_i\frac{\partial G}{\partial a_i} - G\right) = 0.$$

If we solve the relations

$$F = F(a_1, \ldots, a_n), \quad G = G(a_1, \ldots, a_n)$$

as

$$a_1 = W(F, a_2, \ldots, a_n), \quad a_1 = V(G, a_2, \ldots, a_n)$$

respectively, the above equation may be written

$$\frac{\partial}{\partial t}\kappa+\frac{\partial}{\partial x}\omega = 0,$$

$$\frac{\partial}{\partial t}\left(\kappa\frac{\partial W}{\partial a_\alpha}\right)+\frac{\partial}{\partial x}\left(\omega\frac{\partial V}{\partial a_\alpha}\right) = 0 \quad (\alpha = 2, \ldots, n),$$

$$\frac{\partial}{\partial t}\left\{\kappa\left(F\frac{\partial W}{\partial F}+a_\alpha\frac{\partial W}{\partial a_\alpha} - W\right)\right\}+\frac{\partial}{\partial x}\left\{\omega\left(G\frac{\partial V}{\partial G}+a_\alpha\frac{\partial V}{\partial a_\alpha} - V\right)\right\} = 0,$$

where

$$\kappa = \left(\frac{\partial W}{\partial F}\right)^{-1} = \frac{\partial F}{\partial a_1}, \quad \omega = \left(\frac{\partial V}{\partial G}\right)^{-1} = \frac{\partial G}{\partial a_1}.$$

The equations found in the previous sections are all special cases in which $a_n = U$ and $G = UF$ plus a quadratic expression in $a_1, \ldots, a_{n-1}$. Then $\omega = U\kappa$ and the other expressions simplify. Moreover, it becomes clear why the two fourth-order problems of §§ 8 and 9 *must* lead to essentially the same set of final equations.

References

Adlam, J. H. & Allen, J. E. 1958 *Phil. Mag.* **3**, 448.
Benjamin, T. B. & Lighthill, M. J. 1954 *Proc. Roy. Soc.* A, **224**, 448.
Bogoliubov, N. N. & Mitropolsky, Y. A. 1961 *Asymptotic methods in the theory of non-linear oscillations.* Delhi: Hindustan Publishing Co.
Boussinesq, J. 1877 Essai sur la theorie des eaux courantes. *Mém. prés. Acad. Sci., Paris.*
Broer, L. J. F. 1950 *Appl. Sci. Res.* A, **2**, 329.
Camac, M., Kantrowitz, A. R., Litvak, M. M., Patrick, R. M. & Petschek, H. E. 1961 Shock waves in collision-free plasmas. *Rep.* no. 107. AVCO Corporation, Everett, Mass.
Courant, R. & Friedrichs, K. O. 1948 *Supersonic flow and shock waves.* New York: Interscience.
Courant, R. & Hilbert, D. 1953 *Methods of mathematical physics.* I. New York: Interscience.
Courant, R. & Hilbert, D. 1962 *Methods of mathematical physics.* II. New York: Interscience.
Davis, L., Lüst, R. & Schlüter, A. 1958 *Z. Naturforsch.* **13***a*, 916.
Jeffreys, H. & Jeffreys, B. S. 1956 *Methods of mathematical physics,* (3rd ed.). Cambridge University Press.
Korteweg, D. J. & de Vries, G. 1895 *Phil. Mag.* (5), **39**, 422.
Landau, L. D. & Lifshitz, E. M. 1959 *Fluid mechanics.* London: Pergamon Press.
Landau, L. D. & Lifshitz, E. M. 1960 *Mechanics.* London: Pergamon Press.
Minorsky, N. 1962 *Non-linear oscillations.* Princeton, N.J.: Van Nostrand Inc.
Noether, E. 1918 Invariante Variationsprobleme. *Nachr. Ges. Göttingen,* p. 235.
Rossby, C. G. 1945 *J. Meteorol.* **2**, 187.
Whitham, G. B. 1961 *Commun. Pure Appl. Math.* **14**, 675.