J. Fluid Mech. (1969), *vol.* 39, *part* 3, *pp.* 497–509
Printed in Great Britain

A theory of the propagation of internal gravity waves of finite amplitude

By B. S. H. RARITY
Department of Mathematics, University of Manchester

(Received 2 December 1968)

A non-linear theory of internal gravity waves of finite amplitude is developed in terms of conservation equations averaged with respect to the phase. The theory overcomes the failure of linear ray theory in regions in which waves are trapped and establishes the conditions under which finite amplitude waves may propagate. It gives a geometrical representation of the degeneration of waves into quasi-turbulence and predicts the dependence of the energy density on its parameters.

1. Introduction

The ray theory of small amplitude internal waves in a density stratified liquid fails to cope with the singular reflexion which occurs when waves are trapped by a gradual inhomogeneity of the stratification. The group velocity of waves of zero wavelength degenerates into a vertical vector of zero length, and the theory of waves of arbitrary wavelength predicts waves of ever diminishing wavelength. It has been observed experimentally that small amplitude internal waves are capable of negotiating the trapping region without necessarily acquiring large amplitudes or degenerating into quasi-turbulence and are able to preserve the sense of the outward direction of propagation.

The present paper develops a non-linear theory which predicts a propagation velocity of various average properties of the wave train which is horizontal and non-zero in the vicinity of the (horizontal) plane of trapping. The four average properties of the wave train, the amplitude, the wavelength, the direction of the wave-number vector and the frequency are related by means of four independent differential conservation equations which are hyperbolic for certain ranges of the dependent variables. The conditions under which the equations are hyperbolic are represented by points in two disjoint regions in a three-dimensional space, separated by a region $\mathscr{D}$ in which the equations are elliptic. In the first octant, the disjoint regions can be labelled according to whether or not they contain the points which represent zero amplitude waves which satisfy the dispersion relation. The evolution of a wave train is represented by a curve in either region; a wave of initially zero amplitude will be represented by a curve which lies within the disjoint region of the first octant which contains the zero amplitude dispersion relation curve. If this curve intersects $\mathscr{D}$ so that the equations gradually cease to be hyperbolic, the solution must suffer a discontinuity, analogous to the shock wave at the rear of a transonic bubble in compressible gas dynamics. The post-shock representative points must correspond to elliptic conditions and hence lie in the interior of $\mathscr{D}$, and suggests that the quasi-turbulence into which an unstable internal wave degenerates is non-propagating.

2. The averaged equations

The properties of non-linear internal waves in an inviscid density stratified liquid are studied by recasting the Euler equations of motion in the form of four conservation principles. Three of the relations express the conservation of mass, momentum in the horizontal direction and energy; the fourth is a moment of the continuity equation. These principles are conveniently derived as the invariants of an integral whose variation yields the Euler equations of fluid dynamics; the integral has been given by Seliger & Whitham (1968). An approximate solution of the Euler equations, relevant to stationary periodic disturbances in a density stratified fluid otherwise at rest is well known; the approximation is the Boussinesq approximation in which ω_0^2, the square of the Väisälä–Brunt frequency is retained, whereas ω_0^2/g is put identically zero; the approximate solution is that disturbances are sinusoidal in the phase. This approximate solution which certainly holds locally is substituted into the four conservation relations which are then averaged with respect to the phase, so that the short-time oscillations are smoothed out, leaving a set of equations governing the amplitudes. This set of equations must be hyperbolic for some range of the dependent variables. In particular, in the limit of zero amplitude, the propagation velocity must exist and coincide with the group velocity.

The first task is to investigate the propagation in the vicinity of the line of trapping. The linear theory predicts that a wave of frequency ω propagating into a region of slowly varying ω_0 cannot propagate through the layer in which $\omega_0 = \omega$. In this layer k_2, the vertical component of $\mathbf{k}$ the wave-number vector, is zero; the group velocity is vertical but of zero magnitude. If waves of finite amplitude are to be trapped, and it is confirmed experimentally that they are trapped, then energy cannot propagate vertically through some related layer whose position may have to be specified indirectly in terms of the local properties of the wave train but which must agree with the specification $\omega_0 = \omega$ in the limit of zero amplitude. Since the waves are observed to preserve the sense of outward propagation without reflexion back to the source, the propagation velocity of the average quantities of finite amplitude waves must tend to a horizontal vector of non-zero length. The theory described in the present paper predicts such a velocity in the region in which k_2, the vertical component of $\mathbf{k}$, is zero. It appears that the condition $\omega_0 = \omega$, which is identical to that of $k_2 = 0$ for infinitesimal waves, does not in general correspond to the trapping condition for waves of finite amplitude.

The second task is to investigate the hyperbolicity of the governing equations and to relate the findings to the work of Phillips (1967). Similar complementary approaches were made to the stability of Stokes surface waves by Whitham (1967) and Benjamin (1967). Whitham found that the condition that the governing equations should cease to be hyperbolic corresponded exactly to the condition for the explosive growth of resonant interactions investigated by Benjamin. In particular, we must investigate the possibility of the cascading process postulated by Phillips, by which mechanism internal waves break down.

Consider the propagation of waves in two dimensions conjugate to co-ordinates x and y; the co-ordinate y increases upwards in the vertical direction. Seliger & Whitham (1968) have shown that the Euler equations of motion may be derived from a variational principle

$$\delta \int p \, d\mathbf{x} \, dt = 0,$$

where
$$p = p_0(\alpha) - \rho_0(\alpha)\{\phi_t + \alpha\beta_t + \tfrac{1}{2}(u^2+v^2) + gy\},$$

where p_0 and ρ_0 are the equilibrium distributions of pressure and density with respect to their arguments, and where $\mathbf{v} = \nabla\phi + \alpha\nabla\beta$; equilibrium conditions of uniform rest correspond to $\phi = 0$, $\alpha = y$, $\beta = -gt$. If we put $\alpha = y + \alpha^*$, $\beta = -gt + \beta^*$ and $\phi = \Phi - y\beta^*$ then

$$\nabla^2\alpha^*_{tt} + (\alpha^*_{ytt} - g\alpha^*_{xx})\rho_0'/\rho_0 = 0,$$

$$\Phi_{xx} = \alpha^*_{ty},$$

and
$$\beta^*_y = \nabla^2\Phi.$$

If the Boussinesq approximation is made, that is the term α^*_{ytt} neglected in comparison with $g\alpha^*_{xx}$, then there is an approximate solution

$$\alpha^* = A \sin\theta,$$

$$\Phi = -Ak_2k_1^{-2}\omega \sin\theta,$$

$$\beta^* = -A\omega_0^2\omega^{-1}\cos\theta,$$

where $\theta = \mathbf{k}.\mathbf{x} - \omega t$, $\omega^2 = \omega_0^2 k_1^2(k_1^2+k_2^2)^{-1}$ and $\omega_0^2 = -g\rho_0'/\rho_0$. We use Noether's theorem to derive the invariants of the variational integral. The theorem states that if p is a function of the n independent variables x_i, the m dependent variables y_j and the corresponding derivatives $\partial y_j/\partial x_i$, if $\delta\int p(x, y, \nabla y)\, dx = 0$ and if $\int p\, dx$ remains invariant under a transformation

$$x_i^* = x_i + \epsilon\Psi_i(x, y, \nabla y),$$

$$y_j^* = y_j + \epsilon\psi_j(x, y, \nabla y),$$

then
$$\frac{\partial}{\partial x_i}\left\{\frac{\partial p}{\partial(\partial y_j/\partial x_i)}\theta_j + p\Psi_i\right\} = 0,$$

where
$$\theta_j = \psi_j - \frac{\partial y_j}{\partial x_i}\Psi_i.$$

For example, the transformation $t^* = t + \text{constant}$ leaves $\int p\, dx$ invariant; if we put $x_1 = x$, $x_2 = y$ and $x_3 = t$ with $y_1 = \phi$, $y_2 = \beta$ and $y_3 = \alpha$, then $\Psi_1 = \Psi_2 = 0$, $\Psi_3 = 1$ and $\psi_i = 0$ for all i. Thus $\theta_1 = -\phi_t$, $\theta_2 = -\beta_t$ and $\theta_3 = -\alpha_t$. Substituting these expressions we obtain

$$\frac{\partial}{\partial t}\{\rho_0(\phi_t + \alpha\beta_t) + p\} + \frac{\partial}{\partial x}\{\rho_0 u(\phi_t + \alpha\beta_t)\} + \frac{\partial}{\partial y}\{\rho_0 v(\phi_t + \alpha\beta_t)\} = 0, \tag{1}$$

which is an expression of the conservation of energy. Similarly, the variational integral has an invariant conjugate to translation in the x direction

$$\frac{\partial}{\partial t}(\rho_0 u)+\frac{\partial}{\partial x}(\rho_0 u^2+p)+\frac{\partial}{\partial y}(\rho_0 uv) = 0, \tag{2}$$

which is conservation of momentum in the x direction; it has invariants conjugate to translations in ϕ and β which are respectively

$$\frac{\partial}{\partial t}(\rho_0)+\frac{\partial}{\partial x}(\rho_0 u)+\frac{\partial}{\partial y}(\rho_0 v) = 0, \tag{3}$$

and

$$\frac{\partial}{\partial t}(\alpha\rho_0)+\frac{\partial}{\partial x}(\alpha\rho_0 u)+\frac{\partial}{\partial y}(\alpha\rho_0 v) = 0. \tag{4}$$

We now average the equations with respect to the phase θ so that the short time oscillation corresponding to the periodicity of the waves is smoothed out, leaving a set of equations which govern the slow modulation of the average properties of the waves. We are obliged to average the quantity $\rho_0(\alpha)$ in particular with respect to θ; we are at liberty to specify the functional dependence of ρ_0 on its argument, that is the equilibrium distribution of density. In the interests of algebraic simplicity, we choose $\rho_0(\alpha) = \rho_{00}-b\alpha$, where ρ_0 is a constant and where we confine our attention to a band $-d < \alpha < d$; we specify that b is to be small (in effect the Boussinesq parameter) and that d should be sufficiently small to preclude regions of zero and negative density. We shall show that the waves are trapped within a layer bounded by the lines on which the local vertical component of the wave-number vector is zero. Thus, provided this trapping occurs within the band $-d < \alpha < d$, the functional dependence of ρ_0 on its argument outside this band is immaterial. In any event, we would not expect the precise form of any stable distribution $\rho_0(\alpha)$ to affect the general conclusions. For this particular density distribution, the square of the Väisälä–Brunt frequency

$$\omega_0^2 = -g\rho_0^{-1}\,d\rho_0/d\alpha = gb(\rho_{00}-b\alpha)^{-1}$$

is a slowly decreasing function of increasing depth.

Substituting the Boussinesq solution quoted at the beginning of the section, we have

$$\begin{aligned}\alpha &= y+A\sin\theta = y+\alpha^*,\\ \rho_0 &= \rho_{00}-by-bA\sin\theta,\\ \Phi &= -A\omega k_2 k_1^{-2}\sin\theta,\\ \beta^* &= -A\omega_0^2\omega^{-1}\cos\theta,\end{aligned}$$

and

$$\begin{aligned}\mathbf{v} &= \nabla\Phi+\alpha^*\nabla\beta^*-\beta^*\hat{\mathbf{y}}\\ &= (-A\omega k_2 k_1^{-1}\cos\theta+A^2\omega_0^2 k_1\omega^{-1}\sin^2\theta,\quad A\omega\cos\theta+A^2\omega_0^2 k_2\omega^{-1}\sin^2\theta).\end{aligned}$$

We also need to know the functional dependence of the equilibrium pressure distribution on its argument. Since $dp_0/dy = -g\rho_0(y)$ in the undisturbed medium, we have that

$$p_0(\alpha) = p_{00}-g\rho_{00}\alpha+\tfrac{1}{2}gb\alpha^2,$$

where p_{00} is a constant of integration.

Averaging (4) with respect to θ, we obtain

$$\frac{\partial A^2}{\partial t}+\frac{\partial}{\partial x}\left(\tfrac{3}{4}A^4\omega_0^2\frac{k_1}{\omega}-A^2k_1g\frac{y}{\omega}\right)+\frac{\partial}{\partial y}\left(\tfrac{3}{4}A^4\omega_0^2\frac{k_2}{\omega}-A^2k_2g\frac{y}{\omega}\right)=0. \tag{5}$$

Averaging (3) we obtain

$$\frac{\partial}{\partial x}\left(A^2\frac{k_1}{\omega}\right)+\frac{\partial}{\partial y}\left(A^2\frac{k_2}{\omega}\right)=0. \tag{6}$$

Equation (2) gives

$$\frac{\partial}{\partial t}\left(A^2\omega_0^2\frac{k_1}{\omega}\right)+\frac{\partial}{\partial x}\left\{\tfrac{1}{2}\omega_0^2A^2+\tfrac{1}{2}A^2\omega^2\frac{k_2^2-k_1^2}{k_1^2}+\tfrac{3}{8}A^4\omega_0^4\frac{k_1^2-k_2^2}{\omega^2}\right\}$$
$$+\frac{\partial}{\partial y}\left\{-A^2\omega^2\frac{k_2}{k_1}+\tfrac{3}{4}A^4\omega_0^4\frac{k_1k_2}{\omega^2}\right\}=0, \tag{7}$$

and (1),

$$\frac{\partial}{\partial t}\left\{\tfrac{1}{2}\omega_0^2A^2-\tfrac{1}{2}A^2\omega^2\frac{k_1^2+k_2^2}{k_1^2}-\tfrac{3}{8}A^4\omega_0^4\frac{k_1^2+k_2^2}{\omega^2}\right\}+\frac{\partial}{\partial x}\left\{A^2\omega^3\frac{k_2^2}{k_1^3}+gyA^2\omega_0^2\frac{k_1}{\omega}\right\}$$
$$+\frac{\partial}{\partial y}\left\{A^2\omega^3\frac{k_2}{k_1^2}-gyA^2\omega_0^2\frac{k_2}{\omega}\right\}=0. \tag{8}$$

The group velocity of infinitesimal amplitude waves is recovered as the propagation velocity of (7) in the limit $A^2\to 0$. The system of equations (5)–(8) is expected to be hyperbolic. If the wave speed in any direction $\hat{\mathbf{n}}$ is $c^*\hat{\mathbf{n}}$ then recasting the equations in terms of interior co-ordinates and a co-ordinate η perpendicular to the surface and putting $\partial/\partial t=-c^*\partial/\partial\eta$ we obtain a set of equations

$$Q(\partial S/\partial\eta)+\text{interior derivatives}=0,$$

where S is the column vector $\{A^2\ \ k_1\ \ k_2\ \ \omega\}$. The condition $\det(Q)=0$ determines the wave speeds c^*. If we put $k=(k_1,k_2).\hat{\mathbf{n}}$, $h=(1,0).\hat{\mathbf{n}}$ and $v=(0,1).\hat{\mathbf{n}}$ then the determinantal condition is

$$\begin{vmatrix} -c^*+A_1 & A_2h & A_2v & -A_2k/\omega \\ k & A^2h & A^2v & -A^2k/\omega \\ -c^*+A_3 & -A^2(c^*+A_4)/k_1 & A_5 & A^2(c^*+A_6)/\omega \\ -c^*A_7+A_8 & -A_9c^*+A_{10} & A_{11}c^*+A_{12} & A_{13}c^*+A_{14} \end{vmatrix}=0,$$

where

$$A_1=(\tfrac{3}{2}A^2\omega_0^2-gy)\frac{k}{\omega},$$

$$A_2=\tfrac{3}{4}(A^4\omega_0^2-A^2gy)\frac{h}{\omega},$$

$$A_3=\frac{\omega}{\omega_0^2k_1}\left[\tfrac{1}{2}\omega_0^2+\tfrac{1}{2}\omega^2\left(\frac{k_2^2}{k_1^2}-1\right)+\tfrac{3}{4}A^2\omega_0^4\frac{k_1^2-k_2^2}{\omega^2},\quad -\omega^2\frac{k_2}{k_1}+\tfrac{3}{2}A^2\omega_0^4\frac{k_1k_2}{\omega^2}\right].\hat{\mathbf{n}},$$

$$A_4=\frac{\omega}{\omega_0^2A^2}\left[-A^2\omega^2\frac{k_2^2}{k_1^3}+\tfrac{3}{4}A^4\omega_0^4\frac{k_1}{\omega^2},\quad A^2\omega^2\frac{k_2}{k_1^2}+\tfrac{3}{4}A^4\omega_0^4\frac{k_2}{\omega^2}\right].\hat{\mathbf{n}},$$

$$A_5=\frac{\omega}{\omega_0^2k_1}\left[A^2\omega^2\frac{k_2}{k_1^2}-\tfrac{3}{4}A^4\omega_0^4\frac{k_2}{\omega^2},\quad -A^2\omega^2\frac{1}{k_1}+\tfrac{3}{4}A^4\omega_0^4\frac{k_1}{\omega^2}\right].\hat{\mathbf{n}},$$

$$A_6 = \frac{\omega^2}{\omega_0^2 k_1}\left[\omega\left(\frac{k_2^2}{k_1^2}-1\right)-\tfrac{3}{4}A^2\omega_0^4\frac{k_1^2-k_2^2}{\omega^3},\quad -2\omega\frac{k_2}{k_1}-\tfrac{3}{2}A^2\omega_0^4\frac{k_1k_2}{\omega^3}\right].\hat{\mathbf{n}},$$

$$A_7 = \tfrac{1}{2}\omega_0^2-\tfrac{1}{2}\omega^2\left(\frac{k_2^2}{k_1^2}+1\right)-\tfrac{3}{4}A^2\omega_0^4\frac{k_1^2+k_2^2}{\omega^2},$$

$$A_8 = -\left[\omega^3\frac{k_2^2}{k_1^3}+gy\omega_0^2\frac{k_1}{\omega},\quad -\omega^3\frac{k_2}{k_1^2}+gy\omega_0^2\frac{k_2}{\omega}\right].\hat{\mathbf{n}}.$$

$$A_9 = A^2\omega^2\frac{k_2^2}{k_1^3}-\tfrac{3}{4}A^4\omega_0^4\frac{k_1}{\omega^2},$$

$$A_{10} = \left[3A^2\omega^3\frac{k_2^2}{k_1^4}-gyA^2\frac{\omega_0^2}{\omega},\quad -2A^2\omega^3\frac{k_2}{k_1^3}\right].\hat{\mathbf{n}},$$

$$A_{11} = A^2\omega^2\frac{k_2}{k_1^2}+\tfrac{3}{4}A^4\omega_0^4\frac{k_2}{\omega^2},$$

$$A_{12} = \left[-2A^2\omega^3\frac{k_2}{k_1^3},\quad A^2\omega^3\frac{1}{k_1^2}-gyA^2\omega_0^2\frac{1}{\omega}\right].\hat{\mathbf{n}},$$

$$A_{13} = A^2\omega\left(\frac{k_2^2}{k_1^2}\right)+1-\tfrac{3}{4}A^4\omega_0^4\frac{k_1^2+k_2^2}{\omega^3},$$

$$A_{14} = \left[-3A^2\omega^2\frac{k_2^2}{k_1^3}+gyA^2\omega_0^2\frac{k_1}{\omega^2},\quad 3A^2\omega^2\frac{k_2}{k_1^2}+gyA^2\omega_0^2\frac{k_2}{\omega^2}\right].\hat{\mathbf{n}}.$$

If we write $D = k_2/k_1$, $\epsilon = A^2k_1^2$, $f = \omega_0^2/\omega^2$ and $c = k_1c^*/\omega$, then this determinantal equation has three roots:

$$c = \tfrac{3}{4}\epsilon f(1,D).\hat{\mathbf{n}},$$

and the roots of

$$\begin{vmatrix} -Dcv+A_{15} & cv+A_{16} \\ cA_{17}+A_{18} & cA_{19}+A_{20} \end{vmatrix} = 0, \tag{9}$$

where

$$A_{15} = \frac{k_1}{\omega}[(1,D).\hat{\mathbf{n}}A_4+hA_6],$$

$$A_{16} = \frac{k_1^2}{\omega A^2}(1,D).\hat{\mathbf{n}}A_5+vk_1A_6/\omega,$$

$$A_{17} = \frac{h}{\omega A^2}A_{13}-\frac{k_1}{\omega^2A^2}(1,D).\hat{\mathbf{n}}A_9,$$

$$A_{18} = \frac{k_1}{\omega^2A^2}\left[\frac{k_1}{\omega}(1,D).\hat{\mathbf{n}}A_{10}+hA_{14}\right],$$

$$A_{19} = \frac{1}{\omega A^2}\left[\frac{k_1}{\omega}(1,D).\hat{\mathbf{n}}A_{11}+vA_{13}\right],$$

$$A_{20} = \frac{k_1}{\omega^2A^2}\left[\frac{k_1}{\omega}(1,D).\hat{\mathbf{n}}A_{12}+vA_{14}\right].$$

Clearly, one of the wave speeds is always real and, as can readily be seen from (5) and (6), is associated with the continuity equation and its moment. The hyperbolicity of the system will depend on whether the wave speeds associated with (7) and (8), the energy and the x momentum equations, are real. The equations are partially hyperbolic if at any point on the characteristic surface there is at least one direction in which the determinantal equation has real roots. It is convenient to characterize the nature of the equations by the sign of the discriminant Δ of the equation. The discriminant is a function of the dependent variables D, ϵ and f and of the unit vector $\hat{\mathbf{n}}$. At a point in (D, ϵ, f) space at which Δ is uniformly negative with respect to $\hat{\mathbf{n}}$ so that there are no real wave speeds in any direction, the equations are elliptic. The surface of the region $\mathscr{D}$ in (D, ϵ, f) space in the interior of which Δ is uniformly negative will be $\Delta = 0$ and will

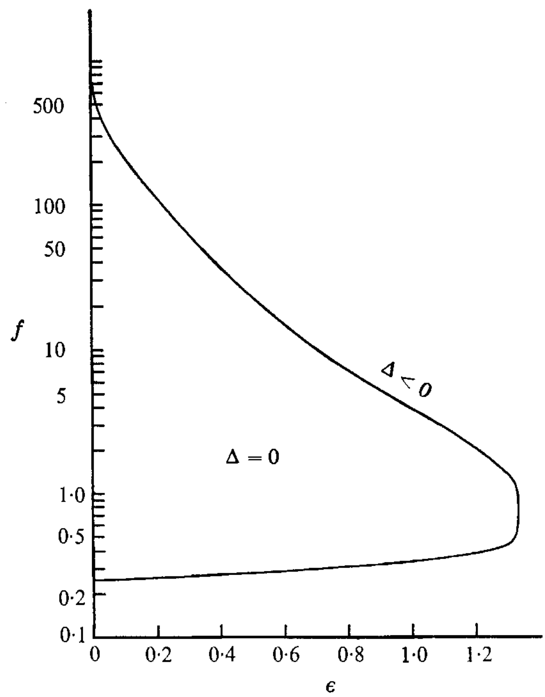

FIGURE 1. The intersection of the surface of $\mathscr{D}$ with the plane $D = 0$.

represent the boundary in the space of the dependent variables between hyperbolic and elliptic equations. If any point on the surface of $\mathscr{D}$ is approached by a sequence of points lying in $\Delta > 0$, that is points representing successive states of the wave train corresponding to which there are real propagation velocities in at least some directions, the point on $\Delta = 0$ will represent the onset of conditions such that the continued outward propagation of average quantities is forbidden. Hence, in the vicinity of such points, the wave train must undergo some drastic transformation. In the appendix the nature of the surface of $\mathscr{D}$ is investigated and it is shown that it is almost everywhere continuous, although it may have a 'crack' which penetrates the whole interior of $\mathscr{D}$. In general, however, an arbitrary path cannot penetrate the surface.

The intersection of $\mathscr{D}$ with the plane $D = 0$ is given by

$$f = \tfrac{1}{4}(\theta+1)^2, \quad \epsilon = \tfrac{16}{3}\theta/(\theta+1)^2,$$

and is shown in figure 1. The intersection of $\mathscr{D}$ with the plane $\epsilon = 0$ with $D \ll 1$ is

$$E \equiv vh(1/f-4) + D[8h^2(1-1/f) - 20v^2 + 4/f] \leqslant 0,$$

with an obvious non-uniformity at $D = 0$. For $D \neq 0$, this is a relation which, for given D and f, must be satisfied uniformly with respect to h and v for points inside $\mathscr{D}$; h and v are the cosine and sine respectively of the angle between the normal $\hat{\mathbf{n}}$ to a characteristic surface and the horizontal direction. The equations are elliptic provided the equation $E = 0$ has no real roots and E is negative. We can always find a pair (h, v) for which $E > 0$ if the equation

$$\tan^2\phi(4/f-20)\,D+\tan\phi(1/f-4)+4D(2-1/f) = 0$$

has real roots; however, if this equation has no real roots then the condition $E \leqslant 0$ is always satisfied provided $f > \frac{1}{5}$. The relation between f and D which corresponds to this latter equation having no real roots is

$$\Delta^* \equiv (1-4f)^2-64D(2f-1)(1-5f) < 0.$$

Note that $\Delta^* < 0$ implies $f > \frac{1}{5}$. The curve $\Delta^* = 0$ is shown in figure 2.

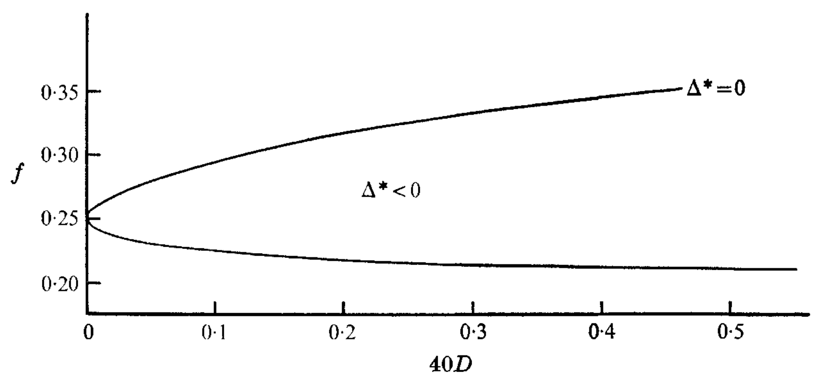

FIGURE 2. The intersection of the surface of $\mathscr{D}$ with the plane $\epsilon = 0$.

The planes $D = 0$ and $\epsilon = 0$ are locations of discontinuities of $\mathscr{D}$. For $D = 0+$, the interior of $\mathscr{D}$ lies close to the curve in figure 1, but is disjoint from the interior of $\mathscr{D}$ on $D = 0$. For $\epsilon = 0+$, the projection of the interior of $\mathscr{D}$ on the plane $\epsilon = 0$ is contained in the region shown in figure 2, but in particular the surface of $\mathscr{D}$ is not continuous at $\epsilon = 0$; the reasons for this behaviour are laid forth in the appendix.

If we investigate the determinantal equation for $\epsilon \rightarrow \infty$ we find that the equations are nowhere elliptic and that this is so for sufficiently large but finite ϵ. Hence the region $\mathscr{D}$ does not extend beyond some plane $\epsilon =$ constant $< \infty$. As $D \rightarrow \infty$, the equations are hyperbolic provided $f/D^2 > 1$ and this holds uniformly in ϵ. Note that if $\epsilon = 0$ and the dispersion relation $f = 1+D^2$ holds, then such points lie in a region of (D, ϵ, f) space in which the equations are always hyperbolic so that the waves are always stable. Further, there is no intersection of the surface of $\mathscr{D}$ with the plane $f = \infty$; likewise, there is no intersection with the plane $f = 0$. The surface of $\mathscr{D}$ is a double surface whose normal section is similar to that of figure 2. For $f = O(1)$, the surfaces are piriform, while for large f the surfaces are asymptotic to $\epsilon = 0$ for all values of D. The interior of $\mathscr{D}$ consists of the region enclosed by these surfaces together with that portion of the plane

$D = 0$ for which $\Delta < 0$. The interior of $\mathscr{D}$ on the plane $D = 0$ is open as we would expect from the properties of infinitesimal waves. For $D = 0+$, the dispersion relation $f = 1 + D^2$ shows that $f = 1+$ and we know that for infinitesimal amplitude waves of frequency less than the Väisälä–Brunt frequency the group velocity is non-zero. However, if $D = 0$ so that $f = 1$, the waves can no longer propagate; that is, the point in (D, ϵ, f) space with co-ordinates $(0+, 0, 1+)$ lies strictly outside $\mathscr{D}$ whereas $(0, 0, 1)$ lies strictly inside $\mathscr{D}$. Strictly speaking, we have classified the point $(0, 0, 1)$ at which one wave speed becomes zero and the other infinite as a non-propagating point on the same footing as ordinary points in the interior of $\mathscr{D}$. In fact, this point is a point of higher order degeneracy of $\mathscr{D}$ than the interior of $\mathscr{D}$ on $D = 0$ itself. However, it serves to illustrate the openness of $\mathscr{D}$ at $D = 0$.

Let us now consider the trapping problem, that is the direction and magnitude of the propagation velocities as $D \to 0$ from points outside $\mathscr{D}$. We shall find that the propagation velocities are real and horizontal. One wave speed is $c = \frac{3}{4}\epsilon f(1, D) . \hat{\mathbf{n}}$ which tends to $c = \frac{3}{4}\epsilon f h$ as $D \to 0$, which is always real and has a maximum in the horizontal direction. The other wave speeds from whose reality the surface and the interior of $\mathscr{D}$ was determined are given by the determinantal equation (9). This equation is of the form

$$c^2 + F(h, v, \epsilon, D, f)\, ch + G(h, v, \epsilon, D, f)\, h^2 = 0 \quad (D \ll 1),$$

where as $D \to 0$ the functions F and G become independent of h and v. Now, we know that at a point in the exterior of $\mathscr{D}$ near $D = 0$ there is at least one direction in which the roots are real. Thus we may conclude that arbitrarily near $D = 0$ but not on $D = 0$, $c = C(\epsilon, f)\, h$ where C is non-zero in general and real. On $D = 0$, the roots are also proportional to h, but for points which are accessible to infinitesimal amplitude waves the function C is not purely real and changes in a discontinuous fashion, in general, as $D = 0$ is attained. The point $f = 1$, $\epsilon = 0$, $D = 0$ is an exceptional point in that $C = 0$ there.

Consider now the wider problem of the stability of the wave train. The first octant of (D, ϵ, f) space not occupied by $\mathscr{D}$ represents states of the wave train in which the propagation is at least partially hyperbolic. So far, we have not assumed that any special relation, for example, a dispersion relation, holds between the dependent variables. Hence there are many points denoting states in which propagation in a hyperbolic manner is possible but which do not satisfy the dispersion relation or its generalization. Nevertheless, we expect from the work of Whitham that one of the differential relations which hold across characteristic surfaces will be the generalization to finite amplitudes of the infinitesimal amplitude dispersion relation, $f = 1 + D^2$. If a wave is generated initially with small amplitude it presumably continues to satisfy everywhere the generalization of the dispersion relation, and is a simple wave. The relation defines a surface in (D, ϵ, f) space and the path of a succession of states of the wave train must lie on this surface. The wave train propagates unimpeded unless the particular path of states on the dispersion relation surface meets the surface of $\mathscr{D}$. The first point to note is that such a path need not approach or intersect the surface of $\mathscr{D}$. Further, an intersection with the interior of $\mathscr{D}$ on $D = 0$ corresponds to trapping of the waves and has been discussed. Thus, we may confine our attention to the surface of $\mathscr{D}$ which lies strictly in the first octant of (D, ϵ, f) space. If a path of states does intersect the surface of $\mathscr{D}$, then outward propagation is no longer possible. An exactly analogous situation occurs in transonic flow. A supersonic bubble embedded in a subsonic flow is in general terminated by a shock wave; the Mach number gradually changes as the rear of the bubble approaches and a discontinuity occurs to effect the transition to elliptic conditions. The succession of states would correspond to the successive values of the Mach number along any fixed streamline. It seems unlikely that the general necessity of having a discontinuity separating supersonic from subsonic conditions along a streamline, that is hyperbolic from elliptic regions in a succession of states, is a phenomenon peculiar to compressible gas flows. Thus, we would expect a discontinuity in D, ϵ and f, the representative post-shock point lying in the interior of $\mathscr{D}$ so that the wave would not propagate. If the post-shock frequencies are *smaller* than those immediately prior to the shock wave, then representative points must lie in a region of $\mathscr{D}$ in which $f\epsilon^2 \sim$ constant for large f. If the frequency *increases* across the shock wave, as Whitham (1967) suggests may well be necessary for 'irreversibility', then the representative points are not quite so closely confined since the surfaces of $\mathscr{D}$ have a finite separation for small f. However, both surfaces behave like $f^{\frac{1}{2}} - \frac{1}{4} \sim \frac{3}{64}\epsilon$ for small ϵ. For both situations $Ak_1 \sim$ constant. Now A is the magnitude of the maximum displacement of a particle from its equilibrium position so that it is itself a measure of the energy in the mode. The corresponding scalar energy density function $E(k)$ therefore satisfies the relation

$$E(k) \sim gk^{-2},$$

replacing k_1^2 by $k^2 = k_1^2(1+D^2)$. Note that we could have produced a result identical to the energy density-wave number relation of Phillips (1967) if we had been prepared to identify $A\omega_0$ as a typical velocity, for in that case

$$E(k) \sim A^2\omega_0^2/k \sim \omega^2 k^{-3}.$$

The value $A\omega_0$ is certainly consistent with the expression for $\mathbf{v}$ which precedes equation (5). The prediction of the energy density spectrum is derived from the result

$$\epsilon \sim F(f)$$

and puts very little constraint on the form of the surface $\mathscr{D}$. In other words, the dependence of E on k alone is very little test of the theory; the postulation of an entirely reasonable characteristic value of velocity can change the k dependence of E from k^{-2} to k^{-3}. However, it also changes the dependence on f.

If the frequency *decreases* across the shock, then $f\epsilon^2 \sim$ constant. This yields

$$E \sim (\omega/\omega_0)^{\frac{1}{2}} gk^{-2}$$

as a function of k and ω/ω_0. If the frequency *increases* across the shock then

$$E \sim \{(\omega_0/\omega) - \tfrac{1}{4}\}^{\frac{1}{2}} gk^{-2}.$$

The result derived by Phillips (1967) is

$$E \sim \omega_0^2 k^{-3}.$$

Further, the discontinuity links points near the surface of $\mathscr{D}$ (not points on $\mathscr{D}$) with points in the interior of $\mathscr{D}$. Postulating that the velocity of the discontinuity is zero, the jump conditions depend on the orientation of the discontinuity surface; since there seems to be no physical condition which would determine the orientation uniquely, there are a variety of jumps co-existing, corresponding to different orientations. This multiplicity of post-shock conditions stemming from a single pre-shock state presumably corresponds to the cascade of Phillips.

The author is not aware of any experimental results which confirm or contradict these results. It is essential for the derivation of the energy density spectrum that the post-shock waves should not propagate. It is indeed possible to have a shock wave which separates two hyperbolic regions, for example, if the initial conditions are concave; however, it seems clear that the analogy with the transonic bubble is the appropriate one here.

Finally, two points are worthy of mention. First, there are, in principle, propagating internal waves whose representative points lie in the disjoint region of (D, ϵ, f) space which does not contain the infinitesimal amplitude dispersion relation curve. Such waves, if they exist, cannot evolve in any continuous manner from infinitesimal amplitude waves, but could conceivably be attained as post-shock states following a discontinuity arising from concave initial conditions, where subsequent simple waves overtake preceding waves of the same family; a convenient analogy would be the formation of the shock wave in one-dimensional unsteady flow of a compressible gas induced by an advancing piston. Secondly, a finite amplitude is *per se* neither necessary nor sufficient for the wave train to break down. From figure 1 one might suppose that ϵ may attain values of about $\frac{4}{3}$ before breakdown need take place. This is based on the assumption that the dispersion relation surface (which has not been obtained) is essentially independent of ϵ. It is perhaps relevant to point out that in the experiments reported by Mowbray & Rarity (1967) values of ϵ, exceptionally as large as 0·7, more typically 0·3 were attained without any sign of instability being observed.

Appendix

The polynomial equation for the wave speeds c of the non-linear partial differential equations has real coefficients which are continuous functions of the dependent variables D, ϵ and f and of the direction of the unit vector $\hat{\mathbf{n}}$ in which a wave speed is sought. If this equation has no real root for any direction $\hat{\mathbf{n}}$ then the equations are elliptic. The discriminant of the equation may be considered to be a polynomial expression P_{2n} in $\tan\phi$, where ϕ is the angle between $\hat{\mathbf{n}}$ and the x axis. Then the equations are elliptic provided this discriminant is uniformly of one sign, say negative, for all ϕ, that is the equation $P_{2n} = 0$ has no real roots and the coefficient of the highest power is negative. The coefficients of the polynomial P_{2n} are likewise continuous real functions of the dependent variables. Hence, the partial differential equations change their character from hyperbolic to elliptic at those points (D, ϵ, f) in the space $\mathscr{S}$ at which $P_{2n} = \delta < 0$, say, where δ is arbitrarily small and negative. The locus of such points is the surface of $\mathscr{D}$ and the interior of $\mathscr{D}$ corresponds to $P_{2n} < \delta$. Since the coefficients of P_{2n} are continuous functions of D, ϵ and f the surface of $\mathscr{D}$ is continuous and dense except possibly in the neighbourhood of points at which the higher coefficients of P_{2n} vanish. Consider now the form of $\mathscr{D}$ and its surface in the neighbourhood of a singular point s in $\mathscr{S}$ at which the two highest coefficients p_{2n} and p_{2n-1} of P_{2n} simultaneously vanish. The region $\mathscr{D}$ is determined by the condition $\mathscr{C}$ that P_{2n} is uniformly negative. Now we can write the polynomial $P_{2n} = \sum_{r=0}^{2n} p_r x^r$ in the form $(\epsilon_1 x^2 + \epsilon_2 x + 1) \sum_{r=0}^{2n-2} q_r x^r$ where the coefficients ϵ_1, ϵ_2 and q_r are defined by the relations

$$p_{2n} = \epsilon_1 q_{2n-2}, \quad p_{2n-1} = \epsilon_1 q_{2n-3} + \epsilon_2 q_{2n-1},$$
$$p_r = \epsilon_1 q_{r-2} + \epsilon_2 q_{r-1} + q_r, \quad 0 \leqslant r \leqslant 2n-2.$$

Then, if $Q_{2n-2} = \sum_{r=0}^{2n-2} q_r x^r$, at any point p in $\mathscr{S}$ we can represent $\mathscr{C}$ as the union of two conditions, $\mathscr{C}_1$ the condition that $\epsilon_1 x^2 + \epsilon_2 x + 1$ should be uniformly positive,

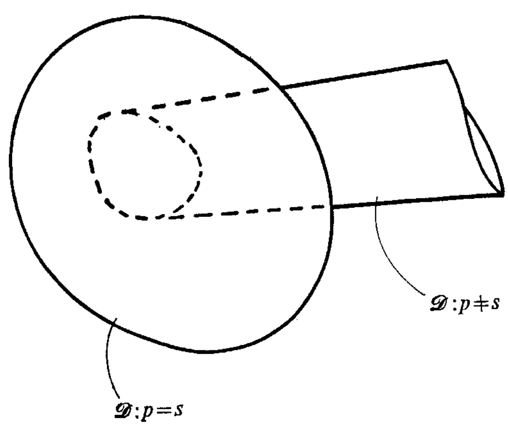

FIGURE 3. A sketch of a region $\mathscr{D}$ near a singular surface.
(Not to be interpreted as depicting the region $\mathscr{D}$ of the main text.)

that is has no real roots and $\epsilon_1 > 0$, and $\mathscr{C}_2$ the condition that Q_{2n-2} should be uniformly less than δ. The condition $\mathscr{C}_1$ is satisfied in a region $V_1^{(p)}$ and $\mathscr{C}_2$ in a region $V_2^{(p)}$. Then the region $V^{(p)}$ in which $\mathscr{C}$ is satisfied is, to an arbitrarily good approximation, $V^{(p)} = V_1^{(p)} \cap V_2^{(p)}$. Now, as $p \to s$, one of the set of points at which p_{2n} and p_{2n-1} are zero, $\mathscr{C}_2^{(p)} \to \mathscr{C}_2^{(s)}$ continuously and since $Q_{2n-2} \to P_{2n-2}$ smoothly as $p \to s$, where $P_{2n-2} = \sum_{r=0}^{2n-2} p_r x^r$, the condition $\mathscr{C}_2^{(s)}$ is also the condition that P_{2n-2} is uniformly negative, and for any point p near s, $\mathscr{C}_2^{(p)}$ is arbitrarily close to the corresponding condition for P_{2n-2}. Hence, $V_2^{(p)} \to V_2^{(s)}$ continuously. However, at $p = s$, $\epsilon_1 = \epsilon_2 = 0$ so that the condition $C_1^{(s)}$ is null, whereas $\mathscr{C}_1^{(p)}$, $p \neq s$, is non-null. Hence, $V_1^{(s)}$ is the whole space $\mathscr{S}$ whereas $V_1^{(p)}$, $p \neq s$, is not the whole space. Hence, $\lim_{p \to s} V_1^{(p)} \cap V_2^{(p)} \subseteq V_2^{(s)}$. Thus the surface of $V_1^{(p)} \cap V_2^{(p)}$, that is the surface of $\mathscr{D}$, need not be continuous at $p = s$; see the sketch in figure 3 (which is not intended to resemble the surface $\mathscr{D}$ of the main text). We can perform this limit process from points on either side of the singular surface $p = s$ with the same result. In general, the projections of the interior of $\mathscr{D}$ at $p = s\pm$ onto $p = s$ will not intersect and the surface of $\mathscr{D}$ will be discontinuous at $p = s$. However, the singular surface $p = s$ does not provide a means of penetrating the interior of $\mathscr{D}$, that is it is not a crack.

However, at points at which the highest coefficient p_{2n} or p_0 or both vanish, then $P_{2n} = 0$ has at least one real root so that P_{2n} cannot be uniformly negative. Such a condition, $p_{2n} = 0$ say, defines a surface in $\mathscr{S}$. The interior of $\mathscr{D}$ is open at its intersection with this surface and the surface does provide a path through the interior of $\mathscr{D}$. However, it remains true that the surface of $\mathscr{D}$ is almost everywhere a barrier to a path of points representing successive states of the system.

For completeness, it should be mentioned that 'non-hyperbolic' is often taken as corresponding to a zero discriminant of the wave speed equation, that is 'parabolic'. The surface corresponding to $P_{2n} = 0$ is arbitrarily close to that for $P_{2n} = \delta < 0$ for arbitrarily small δ even near singular surfaces $p = s$.

REFERENCES

Benjamin, T. B. 1967 *Proc. Roy. Soc. Lond.* A **299**, 59.
Mowbray, D. E. & Rarity, B. S. H. 1967 *J. Fluid Mech.* **28**, 1.
Phillips, O. M. 1967 *Proc. Roy. Soc. Lond.* A **299**, 104.
Seliger, R. L. & Whitham, G. B. 1968 *Proc. Roy. Soc. Lond.* A **305**, 1.
Whitham, G. B. 1967 *Proc. Roy. Soc. Lond.* A **299**, 6.