J. Fluid Mech. (1969), *vol.* 37, *part* 2, *pp.* 289–306
Printed in Great Britain

On thermohaline convection with linear gradients

By P. G. BAINES AND A. E. GILL
Department of Applied Mathematics and Theoretical Physics,
University of Cambridge

(Received 13 September 1968)

The thermohaline stability problem previously treated by Stern, Walin and Veronis is examined in greater detail. An error in an earlier paper is corrected and some new calculations made. It is shown, for instance, that direct convection can occur for thermal Rayleigh number R much less than 100 R_s when $R_s \gtrsim 0{\cdot}1$, where R_s is the salinity Rayleigh number. A graphical presentation is devised to show the relative importance of the different terms in the equations of motion as a function of R and R_s. The most unstable mode over all wave-numbers for each R, R_s is found and it is shown that where both unstable direct and oscillating modes are present, the most unstable mode is direct in most cases.

1. Introduction and summary

This paper examines the stability of a stationary fluid stratified in both heat and salt content when the linearized equations are applicable. This problem has already been discussed for various horizontal boundary conditions by the authors mentioned below, but for the most part attention has been focused on the determination of criteria for the onset of instability. Notable exceptions here are the finite amplitude studies of Veronis (1965, 1968). In this study the authors, by the usual method of normal modes, have examined the nature of the motion for systems which are well and truly (as distinct from marginally) unstable, and considered the relative importance of the various terms in the governing equations. The motivation for this is twofold: to complete and clarify some aspects of the linear theory and to suggest useful approximations for finite amplitude studies.

Some of the properties of the thermohaline convecting system were first noticed by Stommel, Arons & Blanchard (1956) with the discovery of the phenomenon of the salt fountain, which occurs when hot salty water lies above cold fresh water. Such a system was analyzed by Stern (1960), who noted the general properties of the motion now commonly known as 'salt fingers'. The situation with gradients reversed (i.e. with salt gradient stabilizing and temperature gradient destabilizing) has been studied by Veronis (1965). Both these investigations consider the fluid to lie between two horizontal boundaries, thus permitting the specification of the undisturbed system by two Rayleigh numbers. The situation when no boundaries at all are present has been considered by Walin (1964), and stability criteria for horizontal boundaries of various kinds have been presented by Nield (1967).

The present study restricts attention to fluid lying between two horizontal boundaries which are dynamically free and conducting to both heat and salt. The advantage of this choice is that all the equations and boundary conditions may be satisfied by products of sine and exponential functions. For systems with other boundary conditions (see Nield 1967), many of the same general conclusions obtained here should still apply.

In a dimensionless formulation four parameters are required for description of the motion, the thermal and salinity Rayleigh numbers R and R_s, the Prandtl number $\sigma = \nu/\kappa_T$ and $\tau = \kappa_s/\kappa_T$, where ν, κ_T and κ_s are the viscosity and diffusivities of heat and salt respectively. All these are assumed to be constant. For given σ and τ, results may be presented as contours in a Rayleigh number $(R,\ R_s)$ plane, each point on this plane representing a system with given temperature and salinity stratification. The line $R = R_s$ represents an equilibrium state of constant density, and if $R > R_s$ the density increases upward. This investigation is divided into three sections. First, the nature of the normal modes of the system for a disturbance of given wave-number is considered, for all R, R_s. Secondly, the most unstable (i.e. fastest growing) mode over all wave-numbers is found, and thirdly, with the aid of the preceding sections the balance of terms in the vorticity and energy equations is explored. The main interest in the thermohaline problem arises from those regions of the R, R_s plane where the system behaves in a fundamentally different manner from that of a purely thermally stratified system, and there are two such regions, the 'salt finger' region and the 'overstable' region. This differing behaviour is due to the inequality between the diffusivities of heat and salt, and if these diffusivities were equal the system could be adequately parameterized by a single Rayleigh number.

In the region of the R, R_s plane where

$$R_s/\tau < R - \frac{27\pi^4}{4} < R_s < 0,$$

the system is unstable via the 'salt finger' mechanism. Though the mean density gradient is stable, convective motion is driven by salinity differences between the rising and falling columns of fluid (i.e. fingers), because the temperature is made more nearly uniform by the comparatively rapid diffusion of heat. This phenomenon is well documented experimentally, for example by Stommel & Faller (Stern 1960), Turner & Stommel (1964) and Turner (1967). At the boundary of instability,

$$R_s = \tau\left(R - \frac{27\pi^4}{4}\right), \tag{1.1}$$

the horizontal wavelength of the neutral disturbance is of the order of the depth of the fluid layer; however, as one proceeds a short distance into the unstable region the horizontal wave-number of the most unstable mode increases very rapidly, so that the 'cells' tend to be very tall and thin (in Turner's (1967) experiments they are of the order of 2 mm thick). This property was first indicated by Stern and is verified in detail here. Also, the growth rates of the most unstable modes in the 'salt finger' region are generally much smaller than those for direct modes when $R > R_s$.

The second major phenomenon of interest is the oscillatory instability which occurs when the temperature gradient is unstable and the salt gradient is stable. The mechanism for oscillations is well explained by Veronis (1965) and is another consequence of the fact that $\kappa_T \gg \kappa_s$: the diffusion of heat reverses the local buoyancy gradient each half cycle. For a fixed

$$R_s > R_s^0 = \frac{\sigma+1}{\sigma(1-\tau)}\tau^2\frac{27\pi^4}{4} \ll 1,$$

the behaviour as R increases is as follows. As shown by Veronis, instability is first observed when

$$R = \frac{\sigma+\tau}{\sigma+1}R_s + (1+\tau)\left(1+\frac{\tau}{\sigma}\right)\frac{27\pi^4}{4}, \tag{1.2}$$

with a horizontal cell size of $\sqrt{2}d$, where d is the depth of the fluid layer, and with a frequency which is approximately $[\{1-\tau\}/\{3(\sigma+1)\}]^{\frac{1}{2}}N$ provided $R_s \gg R_s^0$, where N is the Brunt–Väisälä frequency of the stable salt stratification. Oscillations of a type compatible with the theory have been observed by Shirtcliffe (1967) and Turner (1968). For a diagram showing where oscillatory modes occur on the R, R_s plane the reader is referred to figure 3, where $\overline{X}\,\overline{W}$ represents equation (1.2). As R is increased in the unstable region, the line $\overline{X}\,\overline{V}$ is reached where the frequency of an oscillating mode has decreased to zero, and here this mode also has cell width $\sqrt{2}d$. Above this line $\overline{X}\,\overline{V}$ direct modes are possible, and for R_s large $\overline{X}\,\overline{V}$ has the form

$$R = R_s + O(R_s^{\frac{2}{3}}), \tag{1.3}$$

so that direct modes occur for $R \ll 1/\tau R_s - \frac{27}{4}\pi^4$ ($\tau \ll 1$) contrary to a statement by Veronis (1965). The dominant mechanism driving the motion for unstable direct modes in the region $R > R_s$ is pure gravitational ('Taylor') instability, modified by the different diffusive processes.

If we consider the most unstable mode as R increases from $\overline{X}\,\overline{W}$ with R_s constant, the cell width decreases slowly while the frequency of oscillation first increases and then decreases. However, the most unstable mode changes abruptly from oscillatory to direct on a line which is almost indistinguishable from $\overline{X}\,\overline{V}$ and lies slightly above it. On crossing this line in the direction of R increasing, the frequency jumps to zero and there is also a discontinuous increase in the horizontal wavelength (i.e. cell size). This implies that the observed cellular size and motion may be very different for systems which are represented by points near this line but on opposite sides of it.

In connexion with the line $\overline{X}\,\overline{V}$, it is interesting to note some experiments by Goroff (1960) for the closely analogous system of convection restrained by rotation. For various Taylor numbers, Goroff measured heat transport by convection as a function of Rayleigh number, and found that the gradient of the curve Nusselt number *versus* Rayleigh number has a sharp increase at the point where direct modes (as predicted by linear theory) become possible. This suggests that at finite amplitude oscillatory modes tend to be less efficient than direct modes for transporting heat. It is possible that similar behaviour occurs in the heat-salt system, i.e. that for fixed R_s there is a discontinuity near $\overline{X}\,\overline{V}$ in the gradient of the curve Nusselt number *versus* R.

For the thermohaline system with stable salt gradient, the linearized equations are similar to those governing the motion of a stratified fluid where the stabilizing factor is rotation (as has been indicated above), or a vertical magnetic field in the case of a conducting fluid (see Chandrasekhar 1961) rather than a solute. Calcultions for these systems which to some extent parallel the work done here have been made by Danielson (1961) and Weiss (1964).

The work in the final section is equivalent to a scale analysis for thermohaline systems with continuous gradients (albeit without the non-linear terms), and potential applications of this to more practical systems form one of the main justifications for the exhaustive study of the simple linear system.

2. Linear stability analysis

We consider a layer of fluid of depth d with linear temperature and salinity gradients (in the undisturbed state), and lying between two horizontal boundaries which are dynamically free and conducting for both heat and salt. Adopting Veronis's (1965) notation, we have for the density

$$\rho = \rho_m(1-\alpha T^* + \beta S^*), \tag{2.1}$$

where ρ_m is the mean density of the system, T^* and S^* are the temperature and salinity, and

$$\alpha = -\frac{1}{\rho}\left(\frac{\partial \rho}{\partial T^*}\right)_{S,p^*}, \quad \beta = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial S^*}\right)_{T,p^*}, \tag{2.2}$$

where p^* represents pressure. We let the differences in temperature and salinity between the bottom and the top be ΔT and ΔS respectively. Then if the salinity gradient is stabilizing and the temperature gradient destabilizing, we have $\Delta T > 0, \Delta S > 0$. With x and z as horizontal and vertical co-ordinates respectively and t as time variable, the relevant Boussinesq perturbation equations in non-dimensionalized form are

$$\left.\begin{aligned} \left(\frac{1}{\sigma}\frac{\partial}{\partial t} - \nabla^2\right)\nabla^2\psi &= -R\frac{\partial T}{\partial x} + R_s\frac{\partial S}{\partial x}, \\ \left(\frac{\partial}{\partial t} - \nabla^2\right)T &= -\frac{\partial \psi}{\partial x}, \\ \left(\frac{\partial}{\partial t} - \tau\nabla^2\right)S &= -\frac{\partial \psi}{\partial x}, \end{aligned}\right\} \tag{2.3}$$

where ψ is the stream function with velocity given by

$$\mathbf{u} = (u, w) = \left(-\frac{\partial \psi}{\partial z}, \frac{\partial \psi}{\partial x}\right), \tag{2.4}$$

σ is the Prandtl number ν/κ_T and $\tau = \kappa_s/\kappa_T$, where ν is the viscosity and κ_s, κ_T are the diffusivities of salt and heat respectively. R and R_s are the temperature and salinity Rayleigh numbers defined by

$$R = \frac{g\alpha\Delta T d^3}{\nu\kappa_T}, \quad R_s = \frac{g\beta\Delta S d^3}{\nu\kappa_T}, \tag{2.5}$$

where g is the acceleration due to gravity.

The physical (i.e. dimensional) variables, denoted by stars, are given in terms of the non-dimensional ones by

$$\left.\begin{aligned}\mathbf{u}^* = \frac{\kappa_T}{d}\mathbf{u}, \quad t^* = \frac{d^2}{\kappa_T}t, \quad x^* = dx, \quad y^* = dy, \quad z^* = dz,\\ T^* = \Delta T\,.\,T, \quad S^* = \Delta S\,.\,S, \quad \psi^* = \kappa_T\psi.\end{aligned}\right\} \tag{2.6}$$

The boundary conditions at $z = 0$, 1 for the set (2.3) are

$$\psi = \frac{\partial^2\psi}{dz^2} = T = S = 0. \tag{2.7}$$

A set of functions satisfying these equations and boundary conditions is

$$\left.\begin{aligned}\psi &\sim e^{pt}\sin\pi\alpha x.\,\sin n\pi z,\\ T, S &\sim e^{pt}\cos\pi\alpha x.\,\sin n\pi z,\end{aligned}\right\} \tag{2.8}$$

where $\pi\alpha$ is a horizontal wave-number, πn a vertical wave-number where n is necessarily an integer, and p must satisfy (see Walin 1964, Veronis 1965)

$$p^3 + (\sigma+1+\tau)k^2p^2 + [(\sigma+\sigma\tau+\tau)k^4 - (R-R_s)\sigma\pi^2\alpha^2/k^2]p + \sigma\tau k^6 + (R_s - \tau R)\sigma\pi^2\alpha^2 = 0, \tag{2.9}$$

where

$$k^2 = \pi^2(\alpha^2+n^2). \tag{2.10}$$

Writing $p = k^2q$ this equation becomes

$$F(q) \equiv q^3 + (\sigma+1+\tau)q^2 + [\sigma+\sigma\tau+\tau-\sigma(R'-R_s')]q + \sigma(\tau+R_s'-\tau R') = 0, \tag{2.11}$$

where

$$R' = \pi^2\alpha^2R/k^6, \quad R_s' = \pi^2\alpha^2R_s/k^6. \tag{2.12}$$

Hence an investigation of (2.11) will yield the behaviour of the modes for any wave-number k. For any given values of σ, τ, R', R_s' (2.11) has three roots for q. One of these roots is real† in all circumstances, whereas the other two may be real or complex conjugates. For present purposes we will fix σ and τ to the values $\sigma = 10$, $\tau = 0{\cdot}01$, corresponding, roughly speaking, to cold salty water, and consider the behaviour of the roots with variation of R', R_s'. (This choice of values for σ and τ involves some loss of generality in the diagrams, but for different pairs of values they may easily be constructed by the methods used below.)

The nature of the roots in the regions of the (R', R_s') plane is shown in figure 1.‡ The lines

$$\left.\begin{aligned}ZX:&\quad \tau R' - R_s' = \tau,\\ XW:&\quad \sigma(\sigma+1)R' - \sigma(\sigma+\tau)R_s' = (\sigma+1)(\sigma+\tau)(1+\tau),\end{aligned}\right\} \tag{2.13}$$

as found by Veronis, represent the boundary between stability and instability of the system to infinitesimal disturbances of wave-number k (i.e. given α, n),

† As a point of analysis, each single root q_i (regarded as a function of R', R_s') has a branch point (but no other form of singularity) at the point X (figure 1), and its representation in the R', R_s' plane requires a three-sheeted Riemann surface for 1–1 correspondence. q_i takes the value of each of the roots of the cubic equation in turn, on successive sheets.

‡ It has been drawn to the authors' attention that a diagram of the nature of the roots of a similar cubic equation appears in a thesis of D. A. Nield.

and are the lines where a root of (2.11) has a zero real part. These lines meet at the point X given by

$$R' = \frac{\sigma+\tau}{\sigma(1-\tau)}, \quad R'_s = \frac{(\sigma+1)\,\tau^2}{\sigma(1-\tau)}. \tag{2.14}$$

XY is a continuation of the line ZX, and on this straight line one root of (2.11) is zero. The line XV represents part of the boundary between the regions where

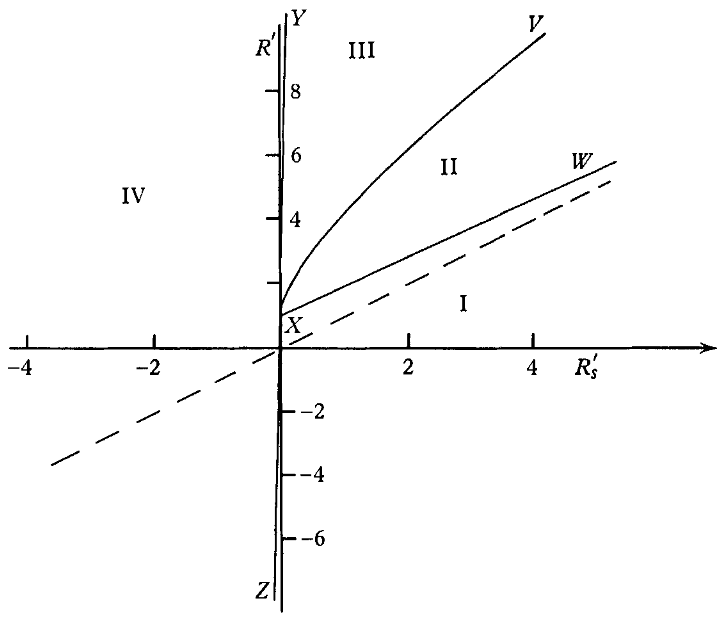

FIGURE 1. Regions of the R', R'_s plane where the roots q of (2.11) have different character. Region I: stable, i.e. no root has a positive real part; region II: unstable, two complex roots have positive real part; region III: unstable, two positive real roots; region IV: unstable, one positive real root. The line shown dashed is the line of neutral buoyancy $R' = R'_s$.

the cubic has three real roots, and one real and two complex conjugate roots, the whole curve being given (see below) by

$$81(\eta+b)^2 = 4(3\zeta+a)^3, \tag{2.15}$$

where

$$\left.\begin{aligned} \zeta &= \sigma(R'-R'_s),\\ \eta &= \sigma(\sigma+1-2\tau)R' - \sigma(\sigma-2+\tau)\,R'_s,\\ a &= \sigma^2+1+\tau^2-(\sigma+\sigma\tau+\tau),\\ qb &= (\sigma+1+\tau)\,[2(\sigma^2+1+\tau^2)-5(\sigma+\sigma\tau+\tau)]+27\sigma\tau. \end{aligned}\right\} \tag{2.16}$$

For small disturbances of wave-number k therefore, systems represented by points in region I (in figure 1) are stable, as in this region no root of (2.11) has a positive real part. In region II the complex roots have a positive real part, and so systems in this region are subject to oscillating instability, or 'overstability', but the direct mode is stable. In regions III and IV, systems have unstable direct modes, one unstable mode in region IV and two in region III. Note that the line representing neutral buoyancy in static conditions is

$$R' = R'_s. \tag{2.17}$$

The lines on which q is a real non-negative constant are shown in figure 2. These are the lines of equal growth rate (for given k) for the direct modes, and by (2.11) are straight lines. To the left of the line ZXY (region IV, as shown in figure 1) only one of these lines passes through each point so there is only one direct

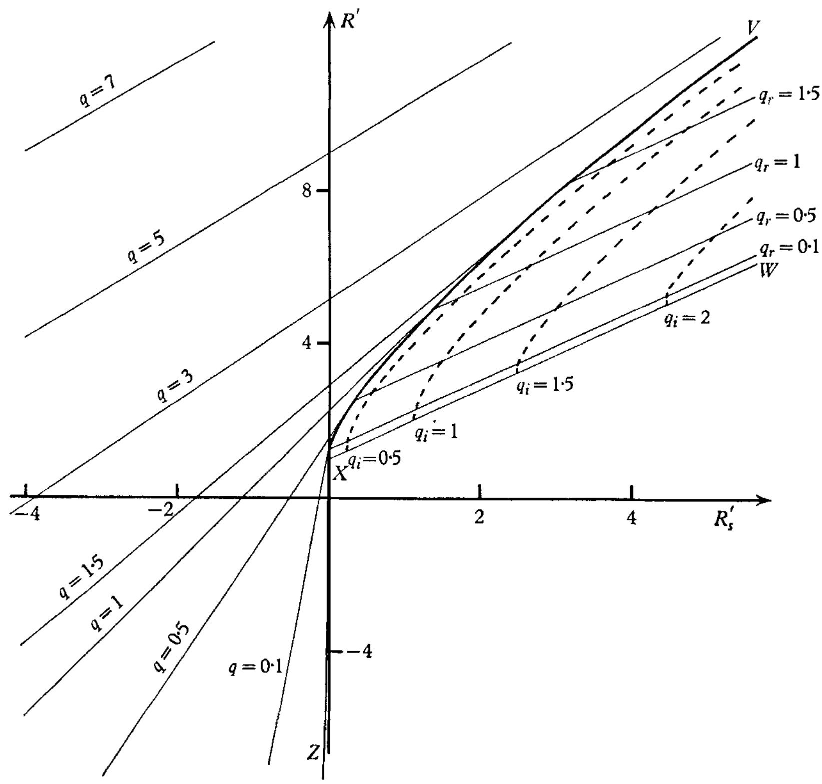

FIGURE 2. Lines of constant growth rate and constant frequency (shown dashed) in the R', R'_s plane.

growing mode. In region III (between XY and XV) two lines pass through each point, so that there are two direct growing modes, and the line XV bordering this region is the envelope of these straight lines. In figure 2 only the line corresponding to the mode of greater growth rate is drawn through a given point, so that each line terminates at the point where it is tangent to the envelope.

In the region II between XV and XW where growing oscillating modes are possible, we write $q = q_r + iq_i$ in (2.11), and by taking real and imaginary parts we obtain, when q_i is non-zero,

$$q_i^2 = F'(q_r) = \frac{F(q_r)}{3q_r+\sigma+1+\tau}. \tag{2.18}$$

Eliminating q_i then gives the following cubic in q_r:

$$8q_r^3 + 8(\sigma+1+\tau)\,q_r^2 + 2[\sigma+\sigma\tau+\tau+(\sigma+1+\tau)^2 - \sigma(R'-R'_s)]\,q_r$$
$$+(\sigma+1+\tau)\,(\sigma+\sigma\tau+\tau) - \sigma\tau - \sigma[(\sigma+1)\,R' - (\sigma+\tau)R'_s] = 0. \tag{2.19}$$

The lines on which q_r is a non-negative constant are drawn in figure 2, and since R', R'_s occur linearly in (2.18) these lines are also straight. Other lines q_i = constant may be obtained from (2.17), and these are also shown in figure 2. The line XW corresponds to the limit $q_r = 0$ and the line XV to the limit $q_i = 0$. Hence from (2.11) and (2.17) we have two parametric equations for the line XV,

$$F(q) = 0, \quad F'(q) = 0, \tag{2.20}$$

and these may be written in the form

$$\left.\begin{aligned}\sigma(R'-R'_s) &= \sigma+\sigma\tau+\tau+2(\sigma+1+\tau)q+3q^2,\\ \sigma(k'-\tau R'_s) &= -\sigma\tau+(\sigma+1+\tau)q^2+2q^3,\end{aligned}\right\} \tag{2.21}$$

the parameter q being the value of the two equal roots at any point on the curve. Near X where q is small the last term in each right-hand side may be ignored, so that the curve is approximated by a parabola tangent to the line XY. This parabola is given approximately by

$$R' - \frac{\sigma+\tau}{\sigma(1-\tau)} = 2\left(\frac{\sigma+1+\tau}{\sigma}\right)^{\frac{1}{2}}\left[R'_s - \frac{(\sigma+1)\tau^2}{\sigma(1-\tau)}\right]^{\frac{1}{2}}. \tag{2.22}$$

At large distances from X, q is large and the last term on the right-hand side of (2.19) dominates. The approximate form is

$$R' = R'_s + \frac{3}{\sigma}\left(\frac{\sigma(1-\tau)R'_s}{2}\right)^{\frac{2}{3}} + O(R'^{\frac{1}{3}}_s). \tag{2.23}$$

The slope of this curve is therefore asymptotic to unity at large distances from the origin.

We note that all results expressed in terms of R' and R'_s may be readily expressed in terms of R, R_s via (2.12), for any given wave-number $k^2 = (\alpha^2+n^2)\pi^2$.

3. The most unstable mode for given Rayleigh numbers

The previous section considers the instability and growth rates of disturbances in a horizontal layer stratified by temperature and salinity where the disturbance has a fixed wave-number. We now consider the behaviour of the instability and growth rates for disturbances of different wave-number, for any given system specified by R and R_s.

Figure 2 shows the function

$$q_r = f(R', R'_s),$$

where f is defined only where q_r is non-negative, and where there are two possible values of q_r for the same R', R'_s, f is defined as the greater. For a given wave-number α, n and given values of R, R_s the growth rate for the quickest growing mode is given by

$$p_r = \pi^2(n^2+\alpha^2)f\left(\frac{\alpha^2 R}{\pi^4(n^2+\alpha^2)^3}, \frac{\alpha^2 R_s}{\pi^4(n^2+\alpha^2)^3}\right). \tag{3.1}$$

If now n is kept fixed and α varied, p_r will change in accordance with (3.1). As α increases from zero to infinity, $\alpha^2/(n^2+\alpha^2)^3$ increases to a maximum of

$\frac{4}{27}n^4$ at $\alpha^2 = \frac{1}{2}n^2$ and then decreases again to zero. The point R', R'_s (in the R', R'_s plane) moves along a straight line from the origin to the point $(4/27\pi^4 n^4)$ (R, R_s) and back again to the origin. From the construction of figure 2, f increases

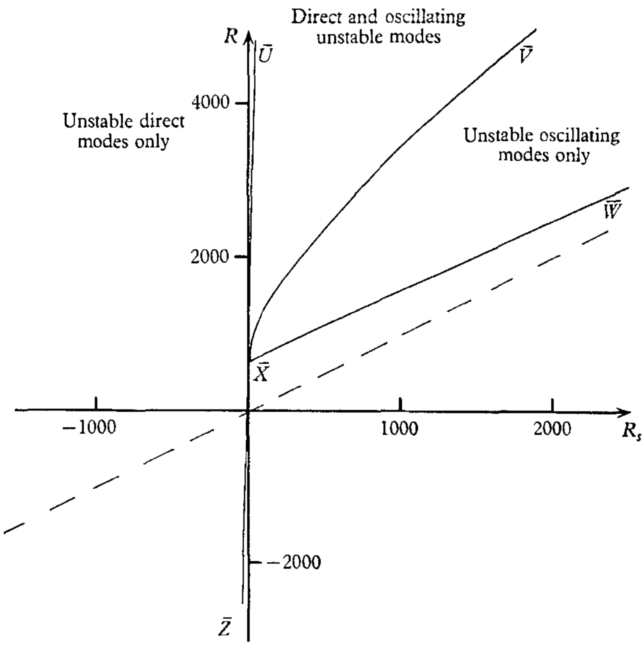

FIGURE 3. Demarcation curves in the R, R_s plane for regions with different types of unstable modes. The line of neutral buoyancy $R = R_s$ is shown dashed.

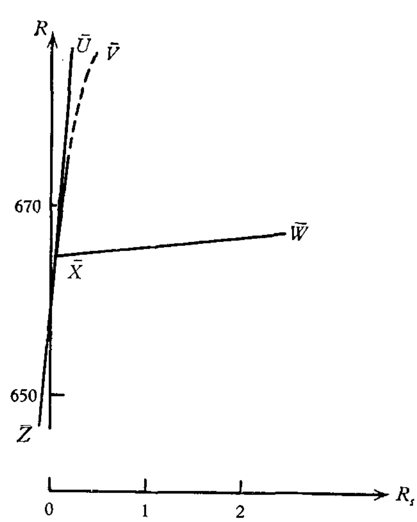

FIGURE 4. Stability curves in the neighbourhood of the triple point $\bar{X}$.

monotonically along such a line. To find the maximum growth rate for given R, R_s one considers how p_r varies with α along the appropriate line segment. Since f is stationary at $\alpha^2 = \frac{1}{2}n^2$ and the factor outside is increasing with α, the maximum value occurs for a value of α^2 which is greater than or equal to

$\frac{1}{2}n^2$. Further, it may be shown that for all R, R_s the most unstable mode has $n = 1$. The details of these calculations are given in the appendix.

Taking $n = 1$, if the line segment in the R', R'_s plane falls short of the line XW and XZ the system will be stable for all wave-numbers. If it intersects XW the system will be unstable for some wave-numbers, but only to oscillating unstable modes if the line segment falls short of XV. Direct modes are also possible if the line segment intersects XV. Thus in the R, R_s plane lines $\overline{X}\overline{Z}$, $\overline{X}\overline{W}$, $\overline{X}\overline{V}$ may be defined as those corresponding to the lines XZ, XW and XV in the R', R'_s plane for $\alpha^2 = \frac{1}{2}$, $n = 1$. These lines are shown in figure 3, together with the line $\overline{X}\overline{U}$, which is the locus of the point X in the R, R_s plane with variation of α^2. To the left of $\overline{Z}\overline{X}\overline{U}$ only direct modes are possible; between $\overline{X}\overline{W}$ and $\overline{X}\overline{V}$ only oscillating (unstable) modes are possible, while between $\overline{X}\overline{U}$ and $\overline{X}\overline{V}$ both oscillating and direct modes are possible. The nature of the intersection of these curves at the 'triple point' $\overline{X}$ is shown in figure 4 on an enlarged scale.

The equations for these lines are

$$\left.\begin{aligned} \overline{X}\overline{Z}:&\quad R = \frac{1}{\tau}\left(R_s - \frac{27\pi^4}{4}\right),\\ \overline{X}\overline{W}:&\quad R = \frac{\sigma+\tau}{\sigma+1}R_s + (1+\tau)\left(1+\frac{\tau}{\sigma}\right)\frac{27\pi^4}{4},\\ \overline{X}\overline{U}:&\quad R = \frac{\sigma+\tau}{(\sigma+1)\,\tau^2}R_s. \end{aligned}\right\} \tag{3.2}$$

The equation for the line $\overline{X}\overline{V}$ is given by (2.12) and (2.15) with $\alpha^2 = \frac{1}{2}$, $n = 1$, and is

$$R = R_s - A_1 + C_1[(\sigma+1+2\tau)R - (\sigma-2+\tau)R_s + B_1]^{\frac{2}{3}}, \tag{3.3}$$

where A_1, B_1, C_1 are constants given by

$$A_1 = \frac{27\pi^4}{12\sigma}a, \quad B_1 = \frac{27\pi^4}{4\sigma}b, \quad C_1 = \frac{3\pi}{2}\left(\frac{3\pi}{2\sigma}\right)^{\frac{1}{3}}, \tag{3.4}$$

and a, b are given by (2.16). For R_s large this has the form

$$R = R_s + O(R_s^{\frac{2}{3}}), \tag{3.5}$$

so that its slope is asymptotic to unity.

The results of the calculations for the most unstable modes are given in figure 5, where lines of constant growth rate, wave-number and frequency are shown, together with the lines $\overline{X}\overline{Z}$, $\overline{X}\overline{W}$ and $\overline{X}\overline{V}$. The curve marking where the most unstable mode changes from oscillatory to direct is denoted $\overline{X}\overline{D}$ and lies above $\overline{X}\overline{V}$, but so close to it that the two curves can hardly be distinguished. Hence in systems where both types of modes are possible it is almost certain that the most unstable mode will be direct rather than oscillatory, the exceptions being those systems represented by points lying between $\overline{X}\overline{V}$ and $\overline{X}\overline{D}$. $p_r = p$ (for the most unstable mode) is continuous across $\overline{X}\overline{D}$ but α^2 is not, generally falling to a value slightly in excess of 0·5 as the curve $\overline{X}\overline{D}$ is crossed in the direction of increasing R. The frequency P_i jumps from a finite value to zero at the same time, and both jumps are larger for larger values of R_s. Thus a small increase in R can result in a sudden change in the character of the most unstable mode and the physical importance of the line $\bar{X}\bar{D}$ is obvious.

The reason for the jump in the most unstable wave-number and for the proximity of the lines $\bar{X}\bar{V}$ and $\bar{X}\bar{D}$ is due to the behaviour near the curve XV of the function $q_r = f(R', R'_s)$ plotted in figure 2. For R'_s/R' fixed and R' close to the value R'_{xv} on XV, from (2.11) and (2.20) $q(R')$ is given approximately by

$$q(R') \simeq q(R'_{xv}) \pm A(R' - R'_{xv})^{\frac{1}{2}},$$

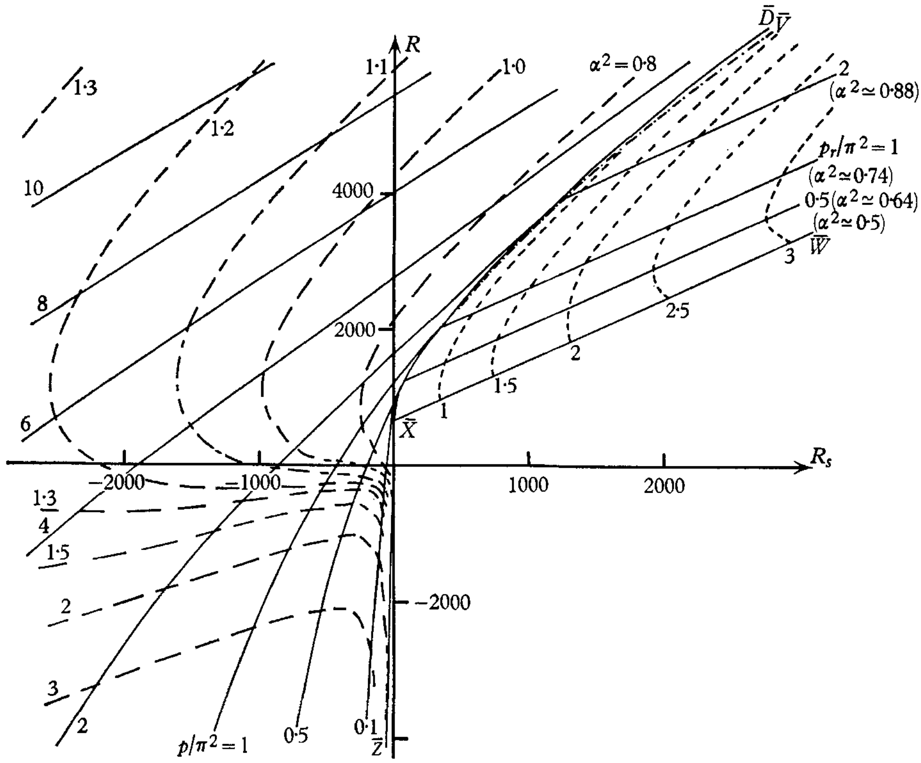

FIGURE 5. Curves showing the nature of the most unstable mode. ——, lines of constant growth rate showing values of p/π^2; – – –, lines of constant wave-number with values of α^2 indicated (with $n = 1$); - - - - -, lines of constant frequency with numbers giving values of p_i/π^2; –·–·–, the line $\bar{X}\bar{V}$, which is almost indistinguishable from the line $\bar{X}\bar{D}$.

where A is a positive constant. For $R' > R'_{xv}$ this gives two direct modes while for $R' < R'_{xv}$ it gives two oscillatory modes. Now the function f is defined as the largest of the possible values of q_r, so

$$\begin{aligned} f(R') &= f(R'_{xv}) + O(R' - R'_{xv}), \quad \text{for} \quad R' < R'_{xv}, \\ &= f(R'_{xv}) + A(R' - R'_{xv})^{\frac{1}{2}} + \dots \quad \text{for} \quad R' > R'_{xv}. \end{aligned}$$

Hence f is singular on the curve XV, the gradient of f being finite at $R' = R'_{xv}-$ but infinite at $R' = R'_{xv}+$. When translated by (3.1) into the behaviour of p_r as a function of R and α^2, the picture shown in figure 6 is obtained. If $R < R_{xv}$ then R' is always less than R'_{xv} and the smooth curve shown in figure 6(a) is obtained. However, if $R > R_{xv}$, R' is greater than R'_{xv} for some values of α^2 near $\frac{1}{2}$ and the spike shown in figure 6(*b*) appears. The reason for the jump in the most unstable wave-number and the proximity of the curves $\overline{XV}$ and $\overline{XD}$ is clear. This behaviour may be expected in quite general circumstances as the argument does not depend on the existence of linear salt and temperature gradients nor in the particular boundary conditions imposed.

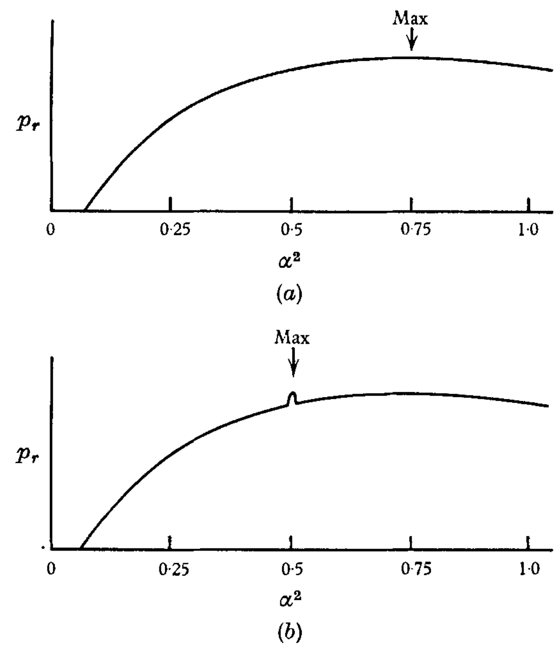

FIGURE 6. (*a*) Form of p_r as a function of α^2 for a point (R, R_s) on $\overline{XV}$ below $\overline{XD}$. (*b*) Form of p_r as a function of α^2 for a point (R, R_s) just above $\overline{XD}$.

Other features of figure 5 will now be described. For the most unstable modes, the curves $p_r =$ constant in the oscillatory-mode region are close to straight lines and have asymptotic gradients for large R_s of $(\sigma+\tau)/(\sigma+1)$, i.e. 0·91. The curves $\alpha^2 =$ constant (not shown in figure 5) are also nearly straight, their asymptotic gradient being unity. The curves $p_i =$ constant are also sketched. These run from the line $\overline{XD}$ to the line $\overline{XW}$. In the direct mode region, the curves $p =$ constant have asymptotic gradient $1/\tau$ (i.e. 100) for large negative R_s. The curves $\alpha^2 =$ constant are S-shaped with asymptotic gradient unity for large positive R and asymptotic gradient $1/\tau$ for large negative R. The latter gradient is not approached until the curves are well off the diagram, except for those with α^2 very close to 0·5.

It is interesting to see in figure 5 that the wave-number of the most unstable mode is comparatively large in the 'salt finger' region so that cells tend to be tall and thin. If we consider the line $R_s = \epsilon R$ where $\tau \ll \epsilon \ll 1$, then it follows from (2.11) that q is given approximately by $q = \epsilon|R'|/(1+|R'|)$, and tends to a constant value ϵ as $|R'| \to \infty$. When translated by (3.1) into the behaviour of p as a function of α^2 for large R it may readily be seen that the maximum value of

p is about $\frac{1}{2}\epsilon|R|^{\frac{1}{2}}$, occurring when $\pi\alpha \simeq |R|^{\frac{1}{4}}$. The corresponding value of $|R'|$ is unity and the value of q is $\frac{1}{2}\epsilon$. Note that in dimensional terms the wave-number α_* of maximum growth is given by

$$\pi\alpha_* \simeq \left(\frac{g\alpha\Delta T/d}{\nu\kappa_T}\right)^{\frac{1}{4}},$$

as found by Stern (1960). The length scale involved is the same as that found for certain boundary layers on vertical walls in thermal convection problems (see, for example, Gill 1966, equation (2.2), Veronis 1967).

4. The dominant factors affecting convection for various configurations of the system

In this section we will calculate, for each R, R_s, the relative importance of the various terms in the equation of motion. In this way we measure the relative importance of the different physical processes involved. It is envisaged that this knowledge will be useful in the study of more complicated problems in suggesting which physical processes will be important and by indicating analytical approximations which will make the finding of solutions more tractable. We will consider disturbances with fixed horizontal wave-number. The horizontal scale may be the one corresponding to the maximum growth rate or possibly one imposed by boundary conditions, and of the three modes which have the given wave-number we will consider only the most unstable. The behaviour of this mode with time is represented by the number $q = p/k^2$, shown as a function of R' and R'_s in figure 2. q is a solution of (2.11) which may be written in the alternative form

$$q+\sigma = \left(\frac{\sigma R'}{q+1} - \frac{\sigma R'_s}{q+\tau}\right). \tag{4.1}$$

This form of the equation is especially significant physically as each term is proportional to a term in the vorticity equation (the first of (2.3)). The first term on the left-hand side corresponds to the rate of change of vorticity, the second term to the rate of loss of vorticity by diffusion and the right-hand side to the rate of generation of vorticity by buoyancy forces. To begin with, we will assume that q is real. The relative importance of the three terms depends on q/σ which is the ratio of the viscous diffusion time to the time scale of the process. Thus the ratio of these terms in the vorticity equation can be ascertained for any given R' and R'_s by reference to figure 2. It will be seen that for most of the points shown in figure 2, q/σ is small so that the first term in (4.1), and hence the first term in the vorticity equation (first of (2.3)) may be neglected.

Similar deductions may be made about the ratio of terms in the other two equations in (2.3), since q is the ratio of the heat diffusion time to the time scale of the problem, and q/τ is the ratio of the salt diffusion time to the time scale of the problem. If one fixes a ratio at which one of the terms is said to become 'negligible', then figure 2 can be divided up into regions in each of which certain approximations to the equations can be made. For instance if it is decided that a term can be neglected if it is less than 1/3 of the biggest term in the same equation, then the boundaries of these regions are given by $q = \frac{1}{2}\tau, 2\tau, \frac{1}{2}, 2, \frac{1}{2}\sigma$ and 2σ. In the region $2\tau < q < \frac{1}{2}$, for example, salt diffusion can be neglected and the time derivative in the heat and vorticity equations can be neglected; in the region $\frac{1}{2} < q < 2$ salt diffusion can be neglected and the time derivative in the vorticity equation can be neglected, but all the terms in the heat equation are important. Similar statements can be made about the other regions.

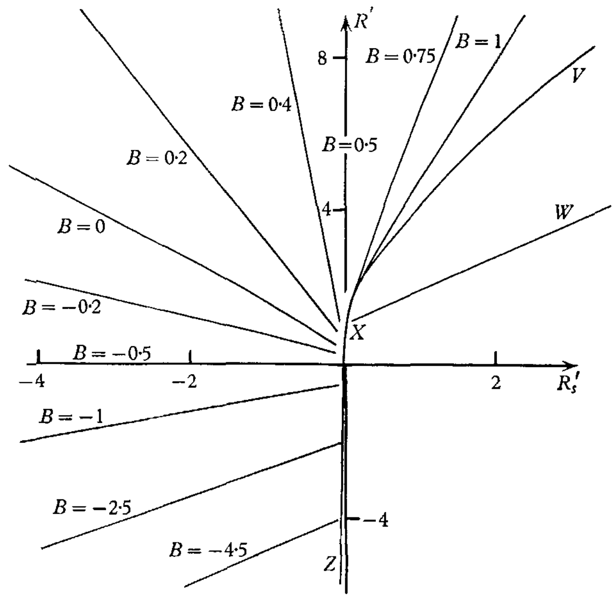

FIGURE 7. Lines of constant B in the R', R_s' plane.

A further refinement can be made by considering the ratio of the two terms in the bracket on the right-hand side of (4.1) to the whole of the left-hand side. The ratio of these terms, in order of their appearance in (4.1) is given by

$$1 : |B+\tfrac{1}{2}| : |B-\tfrac{1}{2}|,$$

where B satisfies simultaneously the equations

$$\left.\begin{aligned} \sigma R' &= (B+\tfrac{1}{2})(q+\sigma)(q+1), \\ \text{and} \qquad \sigma R_s' &= (B-\tfrac{1}{2})(q+\sigma)(q+\tau). \end{aligned}\right\} \qquad (4.2)$$

Contours of B in the R', R_s' plane are shown in figure 7 and, as is readily seen, they are very nearly straight lines. If we define 'negligible' in the same sense as before, figure 7 may be divided up into regions of different physical balances by the lines $B = \pm\frac{1}{6}$, ± 1 and $\pm 2\frac{1}{2}$. If $B > 2\frac{1}{2}$, the two terms on the right-hand side of (4.1) which represent the separate effects of buoyancy due to heat and buoyancy due to salt, dominate over the terms on the left-hand side. If $\frac{1}{6} < |B| < 1$, one of the terms on the right-hand side is small. For other values of B, all terms are comparable.

Similar analyses can be made for equations derived from (2.3) such as the energy equation. The terms in (4.1) are in fact proportional to terms in the energy equation and in suitable units, $q/(q+\sigma)$ is the rate of production of kinetic energy, $\sigma/(q+\sigma)$ is the rate of dissipation, $(B+\frac{1}{2})$ is the rate of release of potential energy from the temperature gradient and $(B-\frac{1}{2})$ is the rate of doing work against the salt gradient. If $B > \frac{1}{2}$, work is done against the salinity gradient and if $B > 1\frac{1}{2}$

most of the energy released from the temperature gradient is used to do work against the salinity gradient. Conversely if $B < +\frac{1}{2}$, energy is released from the salt gradient and if $B < -1\frac{1}{2}$, most of it is used to do work against the temperature gradient. If $q < \sigma$, most of the energy remaining from these two effects is dissipated and only a small fraction is converted into kinetic energy.

Another quantity of interest is the ratio of the average convective density fluxes due to heat and salt at a given level. This ratio is equal to $(\frac{1}{2}+B)/(\frac{1}{2}-B)$ so that if B is large, the two fluxes are nearly equal and opposite.

A similar analysis may be made of the vorticity equation (4.1) in the case where q is complex, viz.

$$q_r + iq_i + \sigma = \frac{\sigma R'(q_{r+1} - iq_i)}{(q_r+1)^2 + q_i^2} - \frac{\sigma R_s'(q_r + \tau - iq_i)}{(q_r+\tau)^2+q_i^2}. \tag{4.3}$$

The real part of this equation (which includes the vorticity diffusion term) corresponds to the terms which are in phase with the velocity fluctuations, while the imaginary part corresponds to the terms which are out of phase. Equating real and imaginary parts gives

$$\left.\begin{aligned} q_r + \sigma &= \frac{\sigma R'(q_r+1)}{(q_r+1)^2+q_i^2} - \frac{\sigma R_s'(q_r+\tau)}{(q_r+\tau)^2+q_i^2}, \\ 1 &= -\frac{\sigma R'}{(q_r+1)^2+q_i^2} + \frac{\sigma R_s'}{(q_r+\tau)^2+q_i^2}, \end{aligned}\right\} \tag{4.4}$$

and it then follows that

$$\left.\begin{aligned} 2q_r + \sigma + \tau &= \frac{\sigma R'(1-\tau)}{(q_r+1)^2+q_i^2}, \\ 2q_r + \sigma + 1 &= \frac{\sigma R_s'(1-\tau)}{(q_r+\tau)^2+q_i^2}. \end{aligned}\right\} \tag{4.5}$$

Hence on the lines $q_r =$ constant the terms in (4.4) are all constant, and it is only necessary to know the one number q_r to calculate the ratios of the various terms. If $\tau \ll q_r \ll \sigma$, as is true for most of the region of oscillatory modes shown in figure 2, the ratio of the terms as they appear in the first part of (4.4) is approximately

$$q_r/\sigma : 1 : q_r + 1 : -q_r.$$

The ratio of the terms in the second part is approximately

$$1/\sigma : -1 : 1.$$

Thus the primary balance in the vorticity equation is between the two buoyancy terms and the diffusion term. The diffusion term is in phase with the velocity fluctuations, so it only appears in the first of (4.4) and becomes relatively less important as q_r becomes large.

As in the case where q is real, when q is complex the terms in (4.3) are also proportional to the terms in the energy equation. However, in this case their relative phases are different. The kinetic energy which has time dependence

$$\exp\left[2(p_r + ip_i)t\right]$$

fluctuates with increasing amplitude with frequency $2p_i$, and this is in phase with the fluctuating viscous dissipation. However, the other terms, namely the rate of increase of kinetic energy, the potential energy released from the temperature buoyancy field, and the work done against the salinity buoyancy field, all have phases differing from that of the kinetic energy. In suitable comparative units (neglecting the exponential time dependence), $(q_r^2+q_i^2)^{\frac{1}{2}}$ is the rate of gain of kinetic energy and σ is the rate of dissipation. In the same units, the rate of release of potential energy from the temperature gradient is

$$(2q_r+\sigma+\tau)\,[(q_r+1)^2+q_i^2]^{\frac{1}{2}}/(1-\tau),$$

and the rate of working against the salinity gradient is

$$(2q_r+\sigma+1)\,[(q_r+\tau)^2+q_i^2]^{\frac{1}{2}}/(1-\tau).$$

Hence for increasing q_r, q_i the relative loss of energy via dissipation becomes smaller, and a greater proportion of energy from the temperature field is used to

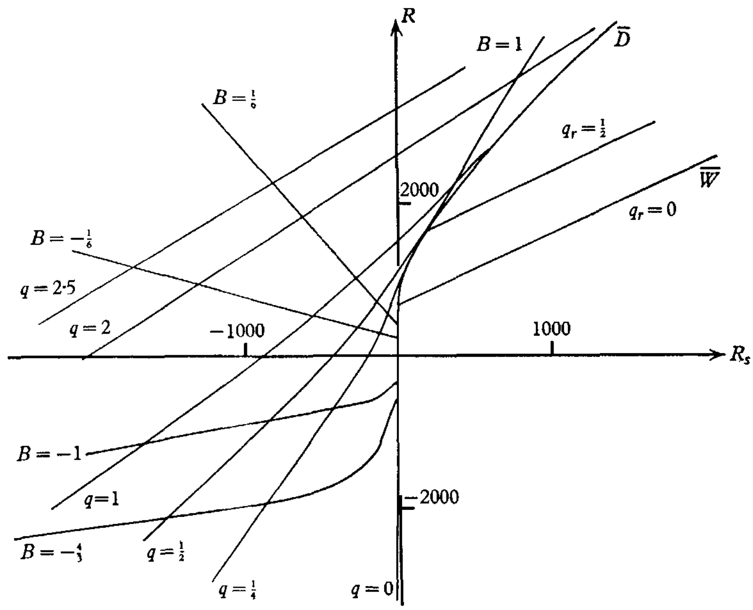

FIGURE 8. Lines of constant q, B and q_r in the R, R_s plane for the quickest growing mode.

do work against the salinity field and also to increase the overall kinetic energy. The ratio of the two above expressions is also the ratio of the rates of convective density transport by heat and by salt at any level.

The above results may be used for any given α, R and R_s to estimate the relative importance of the different terms in (2.3), the ratios of the terms depending on either q and B or on q_r. This is true in particular when the value of α is chosen as the one which, for given R and R_s, gives maximum growth. Hence the values of q, B and q_r for the disturbance of maximum growth rate may be calculated as functions of R and R_s. For the direct modes this may readily be done via (3.5) and (4.2), and for the oscillating modes q_r may be found via (3.7) and (3.8). Figure 8 shows the results of this calculation. The figure shows, for instance, that in the region defined by

$$\tau \ll R_s/R \ll 1$$

and

$$R \ll -2000$$

the time derivatives in the heat and vorticity equations can be neglected, as can the diffusion term in the salt equation. In addition, figure 5 shows that α^2 is large so that z-derivatives can be neglected in comparison with x-derivatives. Such approximations may be useful in examinations of the 'salt finger' phenomenon in situations more complicated than that considered in this paper.

P.G.B. acknowledges the support of a C.S.I.R.O. (Australia) post-graduate studentship.

Appendix

In order to obtain the growth-rate and horizontal wave-number of the most unstable mode for all R, R_s from (2.9), (2.19), we put

$$r = \frac{R}{n^4\pi^4}, \quad r_s = \frac{R_s}{n^4\pi^4}, \quad P = \frac{p}{n^2\pi^2}, \quad y = \frac{\alpha^2}{n^2}. \tag{A 1}$$

Considering first the case when p is real, (2.9) after some re-arrangement yields

$$(1+y)\,(P+\sigma(1+y))/\sigma = \frac{yr}{P+1+y} - \frac{yr_s}{P+\tau(1+y)}. \tag{A 2}$$

Maximum growth (with respect to α^2) then occurs when $\partial p/\partial y = 0$, i.e. when

$$\frac{P+\sigma(1-y^2)}{\sigma} = \frac{y^2r}{(P+1+y)^2} - \frac{\tau y^2 r_s}{(P+\tau(1+y))^2}. \tag{A 3}$$

What is therefore required is to solve (A 2) and (A 3) simultaneously for P and y for given r, r_s. However, since these equations are linear in r and r_s, it is much simpler, for the purposes of obtaining the general picture, to substitute given values of P and y and obtain the corresponding r, r_s (which incidentally will be determined uniquely, so that there is a one-one correspondence between the pairs (P, y) and (r, r_s) corresponding to the most unstable mode). Solving (A 2) and (A 3) for r and r_s yields

$$\left.\begin{aligned} r &= \frac{(P+1+y)^2}{\sigma(1-\tau)y^2P}[P^2+(\sigma+\tau)\,(1-y^2)\,P+\sigma\tau(1+y)^2\,(1-2y)], \\ r_s &= \frac{(P+\tau(1+y))^2}{\sigma(1-\tau)\,y^2P}[P^2+(\sigma+1)\,(1-y^2)\,P+\sigma(1+y)^2\,(1-2y)]. \end{aligned}\right\} \tag{A 4}$$

Lines of constant P and constant α^2 in the R, R_s plane obtained from these formulae are graphed in figure 5 for the case $n = 1$.

The above analysis applies only for the direct modes, and the unstable oscillatory modes need to be treated separately. Writing

$$p_r = k^2q_r = n^2\pi^2 P_r, \tag{A 5}$$

equation (2.19) may be written

$$\frac{(1+y)}{\sigma}[2P_r+(1-\tau)\,(1+y)] = \frac{yr}{2P_r+(\sigma+\tau)\,(1+y)} - \frac{yr_s}{2P_r+(\sigma+1)\,(1+y)}. \tag{A 6}$$

The growth rate for an oscillatory mode is a maximum when $\partial P_r/\partial y = 0$, that is, when

$$\frac{2P_r+(1+\tau)\,(1-y^2)}{\sigma} = \frac{(\sigma+\tau)\,y^2r}{[2P_r+(\sigma+\tau)\,(1+y)]^2} - \frac{(\sigma+1)\,y^2r_s}{[2P_r+(\sigma+1)\,(1+y)]^2}. \tag{A 7}$$

Again we have two equations which are linear in r and r_s. These are readily solved for r and r_s and the resultant formulae may be used to obtain the lines of constant P_r and y. The frequencies of the modes may then be obtained from (2.18) and lines of constant p_r and constant frequency are graphed in figure 5 for $n = 1$.

It remains to show that the most unstable mode always has $n = 1$. From (3.1), on any line radiating from the origin in the R, R_s plane we may write

$$p_r = (n^2+\alpha^2)\, g\left(\frac{\alpha^2 \mathscr{R}}{(n^2+\alpha^2)^3}\right), \tag{A 8}$$

where
$$g\left(\frac{\alpha^2 \mathscr{R}}{(\alpha^2+n^2)^3}\right) = \pi^2 f\left(\frac{\alpha^2 R}{\pi^4(n^2+\alpha^2)^3}, \frac{\alpha^2 R_s}{\pi^4(n^2+\alpha^2)^3}\right), \tag{A 9}$$

and $\mathscr{R} = \text{constant} \times R$. Then

$$\frac{\partial p_r}{\partial n^2} = \frac{\partial p_r}{\partial \alpha^2} - \left(1+\frac{n^2}{\alpha^2}\right) g'\left(\frac{\alpha^2 \mathscr{R}}{(n^2+\alpha^2)^3}\right), \tag{A 10}$$

where g' denotes the derivative of g with respect to its argument. It may be seen from figure 2 that $f(R', R'_s)$ always increases as one proceeds outwards on a line radiating from the origin in the R', R'_s plane, and this implies that $g' > 0$ for all R, R_s. Therefore, when $\partial p_r/\partial \alpha^2 = 0$, $\partial p_r/\partial n^2 < 0$ for all continuous n^2, and p_r will be a maximum for the minimum permissible value of n, which is $n = 1$.

REFERENCES

Chandrasekhar, S. 1961 *Hydrodynamic and Hydromagnetic Stability*. Oxford: Clarendon.

Danielson, R. E. 1961 The structure of sunspot penumbras. *Astrophys. J.* **134**, 289.

Gill, A. E. 1966 The boundary layer régime for convection in a rectangular cavity. *J. Fluid Mech.* **26**, 515.

Goroff, I. R. 1960 An experiment on heat transfer by overstable and ordinary convection. *Proc. Roy. Soc.* A **254**, 537.

Nield, D. A. 1967 The thermohaline Rayleigh–Jeffreys problem. *J. Fluid Mech.* **29**, 545.

Shirtcliffe, T. G. L. 1967 Thermosolutal convection: observation of an overstable mode. *Nature, Lond.* **213**, 489.

Stern, M. E. 1960 The salt fountain and thermohaline convection. *Tellus*, **12**, 172.

Stommel, H., Arons, A. B. & Blanchard, D. 1956 An oceanographical curiosity: the perpetual salt fountain. *Deep Sea Res.* **3**, 152.

Turner, J. S. & Stommel, H. 1964 A new case of convection in the presence of combined vertical salinity and temperature gradients. *Proc. Natn. Acad. Sci. U.S.A.* **52**, 49.

Turner, J. S. 1967 Salt fingers across a density interface. *Deep Sea Res.* **14**, 599.

Turner, J. S. 1968 The behaviour of a stable salinity gradient heated from below. *J. Fluid Mech.* **33**, 183.

Veronis, G. 1965 On finite amplitude instability in thermohaline convection. *J. Marine Res.* **23**, 1.

Veronis, G. 1967 Analogous behaviour of homogeneous, rotating fluids and stratified, non-rotating fluids. *Tellus*, **19**, 326.

Veronis, G. 1968 Effect of a stabilizing gradient of solute on thermal convection. *J. Fluid Mech.* **34**, 315.

Walin, G. 1964 Note on the stability of water stratified by both salt and heat. *Tellus*, **16**, 389.

Weiss, N. O. 1964 Convection in the presence of restraints. *Phil. Trans. Roy. Soc.* A **256**, 99.