# SOME FLUID DYNAMICAL PROBLEMS IN ASTROPHYSICS

Luke O'Connor Drury

Trinity College, Cambridge

and

Institute of Astronomy, Cambridge

JUNE 1979



A dissertation submitted for the degree of

Doctor of Philosophy

# SOME FLUID DYNAMICAL PROBLEMS IN ASTROPHYSICS

In the first part of my dissertation I consider certain aspects of the cosmic turbulence theory of galaxy formation (as revived by Ozernoi). Using a generalised form of a transformation due to Kurskov and Ozernoi I exhibit a formal equivalence between the problem of turbulence in an expanding universe containing a coupled matter-radiation fluid and in a non-expanding fluid with a time-dependent viscosity. This enables me to extend the Olson-Sachs formula for vorticity generation in cosmic turbulence to a matter-radiation fluid and to show that, contrary to the hypothesis of Ozernoi, the turbulence can not have an inertial subrange at the epoch of recombination.

In the second part I consider the linear inviscid stability of axisymmetric flows. Using the projective form of the perturbation equations (rather than the linear form) I obtain a simple proof of a generalised Richardson criterion which holds for all boundary conditions which do not actively feed energy to the perturbation. Further analysis shows the uniform density and pressure discs with self-similar rotation laws, $\Omega = r^{-\lambda}$ $1 < \lambda < 2$, are stable to perturbations which are incompressible in character, but that instability is a generic feature of differentially rotating compressible systems; this result is explained in terms of over-reflection of sound waves from the corotation radius. Two families of self-similar discs in which the perturbation equations can be solved exactly are described and used to estimate the magnitude of this effect.

In the third part I consider the problem of numerically solving boundary value problems of the Orr-Sommerfeld type by shooting methods and describe a unifying geometrical interpretation of the principal methods.

In appendix A I apply the methods of section two to plane parallel flows and derive among other results an extension of Rayleigh's inflection point theorem to compressible isobaric and isentropic flows.

In appendix B I describe a program for locating the zeros of a holomorphic function which is very useful in eigenvalue problems of the type considered in section three and give an example of its use.

# PREFACE

Except where explicit reference is made to the work of others this dissertation is the result of my own work and includes nothing which is the result of work done in collaboration. It is not substantially the same as any I have submitted for a degree, diploma or other qualification at any other university, nor has any part been submitted for such a degree, nor is any part being so submitted. Its total length does not exceed 60000 words.

Luke O'C. Drury.

ACKNOWLEDGEMENTS

I am indebted to my supervisor, Dr. J.M.Stewart, for his patient direction and constant encouragement of my research and also for a more general interest and concern of which I am very sensible.

I am grateful to Trinity College Cambridge for supporting me during my first three years with a research studentship and to the electors of the Isaac Newton fund for allowing me an extra year in Cambridge. I owe much to discussions with many people at the Institute of Astronomy, in particular D.Lynden-Bell, S.Tremaine, D.Gough, D.Lin, J.Papaloizau, B.Jones, and also to the excellent but anonymous referees of the Journal of Fluid Mechanics. I must also thank D.Moore, H.Huppert and M.McIntyre for helpful discussions. I acknowledge a debt to the mathematical tradition of my primary university, Dublin, for a sense of the power of geometrical method in mathematical investigation and to that of my secondary, Cambridge, for an apprehension of the meaning of physical investigation. I thank the University of Cambridge Computing service for the excellent facilities provided (and for typing the bulk of the manuscript).

"In an expanding universe gravitational instability would not be sufficient to form subsystems, while turbulence could do it . ."

C.F. von Weizsacker 1951

## 1 Introduction

Modern cosmography shows that the large scale structure of the universe is almost indistinguishable from that described by one of the simplest of cosmological models; a universe containing a homogeneous mixture of matter and electro-magnetic radiation expanding isotropically from an initial singular state of infinite density (the 'hot big bang' model). The simplicity of this model and the strength of the evidence in its favour, in particular the Hubble law, the isothermality and isotropy of the microwave background and the relative abundancies of the light elements, have raised it to the status of a standard model (Peebles 1971a)*. Yet on a small scale it is clear that the Cosmos is not homogeneous; matter has agglomerated into condensations on many different length (and mass) scales. One of the major problems of modern cosmology is to understand the formation and evolution of this structure, in particular that of galaxies and clusters of galaxies.

-----------------

* Though the recent claim that the microwave background shows significant deviations from a pure thermal spectrum at the 5 $\sigma$ level (Woody & Richards,1979) may slightly reduce its standing.

It is natural to suppose that the force of gravity which maintains this structure was also the main agent of its creation. In a homogeneous static fluid any density perturbation on a length scale greater than a scale (the Jeans length) determined by the stabilizing effect of pressure will grow exponentially. However in an expanding universe this instability is much less efficient and perturbations to the cosmic density only grow algebraically in time (as $t^{2/3}$ in the flat Friedmann universe), a result obtained by Lifschitz (1946). (This effect is closely related to the distinction between $t$ and $t^*$ time introduced in section 3).

This slow growth rate was interpreted as showing that galaxies could not have been formed by the gravitational amplification of random fluctuations in the density of the cosmic matter. However Lifschitz also showed that vortical motions in an expanding universe would not decay as long as the universe was radiation dominated. As an alternative to the gravitational instability theory von Weizsacker (1951) suggested that galaxies were formed by turbulent motion in the early universe*, an idea which was

-----------------

* An advantage of this theory is that the galaxies being the fossils of turbulent eddies are expected to possess angular momenta and magnetic fields. The origin of the angular momenta of galaxies used to be considered a serious problem for the gravitational instability theory though it now seems that the mechanism suggested by Hoyle (1951) and investigated in detail by Peebles (1969) whereby galaxies are spun up by the tidal torques of their neighbours affords a satisfactory explanation. A recent numerical investigation is that by Efstathiou & Jones (1979). The generation

supported by Gamow (1952). However apart from some work in Japan by Nariai (1956a,b) no further work was done on the theory until its revival by Ozernoi and his coworkers in the late sixties (Ozernoi and Chernin 1968a,b; Ozernoi and Chibisov 1971,1972; Ozernoi 1972; Kurskov and Ozernoi 1974a,b,c). The review article by Jones (1976) gives a very useful summary of the various theories of galaxy formation (with particular emphasis on the cosmic turbulence theory).

Ozernoi's theory differed from the earlier work in its introduction of the concept of 'freezing', in its use of the improved cosmographic data of the sixties to constrain the background model, and in its emphasis on the theory's relative insensitivity to the initial conditions. This last idea was based on the assumption that the turbulence, no matter how it originated, would evolve a universal Kolmogorov spectrum on certain length scales from which the mass and rotation distribution functions for galaxies could be obtained. In this form the theory contains only one free parameter (characterising the strength of the turbulence) so that it is not necessary to have very special initial conditions to produce the observed structure of the universe. However because this independence requires the existence of a Kolmogorov spectrum at recombination it is important to investigate the evolution of turbulence in a universe which on large scales and in some mean approximates the 'hot big

-----------------

of magnetic fields by cosmic turbulence has been investigated by Harrison (1970).

supported by Gamow (1952). However apart from some work in Japan by Nariai (1956a,b) no further work was done on the theory until its revival by Ozernoi and his coworkers in the late sixties (Ozernoi and Chernin 1968a,b; Ozernoi and Chibisov 1971,1972; Ozernoi 1972; Kurskov and Ozernoi 1974a,b,c). The review article by Jones (1976) gives a very useful summary of the various theories of galaxy formation (with particular emphasis on the cosmic turbulence theory).

Ozernoi's theory differed from the earlier work in its introduction of the concept of 'freezing', in its use of the improved cosmographic data of the sixties to constrain the background model, and in its emphasis on the theory's relative insensitivity to the initial conditions. This last idea was based on the assumption that the turbulence, no matter how it originated, would evolve a universal Kolmogorov spectrum on certain length scales from which the mass and rotation distribution functions for galaxies could be obtained. In this form the theory contains only one free parameter (characterising the strength of the turbulence) so that it is not necessary to have very special initial conditions to produce the observed structure of the universe. However because this independence requires the existence of a Kolmogorov spectrum at recombination it is important to investigate the evolution of turbulence in a universe which on large scales and in some mean approximates the 'hot big

-----------------

of magnetic fields by cosmic turbulence has been investigated by Harrison (1970).

bang' model to see if this can occur.

With our present limited understanding of even the simplest form of turbulence (in a Newtonian fluid) and considering the complications introduced by the expansion of the universe and the relativistic nature of the cosmic fluid (at early epochs the density of the radiation far outweighs that of the matter) one might doubt whether useful results could be obtained in this problem. Remarkably this is not the case. I will show, using a generalisation of a transformation originally found by Kurskov and Ozernoi (1974a), that the hydrodynamic theory of a matter-radiation fluid in an expanding universe is formally identical to the familiar classical hydrodynamic theory. Thus the cosmological problem is no harder (and no easier! ) than the standard problem of turbulence. It follows that the results obtained in the classical theory can, when suitably interpreted, be applied in cosmology. And perhaps more importantly the experimental evidence relating to these classical results may also apply to their cosmological extensions (the approximations made in turbulence theory are usually justified empirically by comparing the experimental data with their predictions rather than by analytic arguments; without the formal equivalence this could not be done for the cosmological problem).

## 2 Basic equations

Let us consider cosmic turbulence against the background of a Friedmann-Robertson-Walker type cosmological model. This is reasonable if the typical scale of the turbulence is much smaller than the horizon size; I exclude the very early universe where this small wavelength approximation is invalid. As the models with k=+1,0,-1 are virtually indistinguishable at early epochs let us adopt the simplest, k=0, and take as the metric (in co-moving coordinates)

$$ds^2 = -dt^2 + a(t)^2\{dx^2 + dy^2 + dz^2\} \qquad [2:1]$$

The evolution of cosmic turbulence can be divided into two phases. In the first, prior to recombination, the matter and radiation are strongly coupled and can be treated as a single fluid. The high sound speed and large Jeans length allow inhomogeneities in the matter and radiation densities to be ignored if the turbulent peculiar velocities are small compared with c.† With these assumptions one can derive the equations (Peebles 1971)



† c = speed of light in vacuo

$$\rho_t\left\{\frac{\partial \underset{\sim}{v}}{\partial t} + \frac{1}{a}\underset{\sim}{v}\cdot\nabla\underset{\sim}{v} + \frac{\dot{a}}{a}\underset{\sim}{v}\right\} + \frac{1}{3}\dot{\rho}_r\underset{\sim}{v} + \frac{1}{a}\nabla p = \frac{\rho_t \nu}{a^2}\nabla^2\underset{\sim}{v}$$

$$\nabla\cdot\underset{\sim}{v} = 0 \qquad [2:2]$$

$$\dot{\rho}_m + \frac{3\dot{a}}{a}\rho_m = 0$$

$$\dot{\rho}_r + \frac{4\dot{a}}{a}\rho_r = 0$$

where $\rho_m$ is the matter density, $\rho_r$ the radiation density and

$$\rho_t = \rho_m + \frac{4}{3}\rho_r \qquad [2:3]$$

the density of inertia. The viscosity arises from the radiative transfer of momentum giving †

$$\nu = \frac{8}{27}\frac{\lambda c \rho_r}{\rho_t} = \frac{8}{27}\frac{m_p c \rho_r}{\sigma \rho_m \rho_t} \qquad [2:4]$$

where $\lambda$ is the photon mean free path (Chan & Jones 1975).

Unfortunately the passage through recombination can not be treated by any simple analysis. For want of a better approximation I assume that recombination occurs instantaneously (though this is an unreasonable assumption and significant damping may occur in the transition (Chibisov (1972)).

In the second phase, after recombination, the



† $m_p$ = mass of the proton, $\sigma$ = Thompson cross section

matter and radiation are decoupled and one need consider only a pure matter fluid. However the much lower sound speed means that this is now compressible and gravitationally unstable. In the Newtonian approximation one can derive the hydrodynamic equations

$$\frac{\partial \underset{\sim}{v}}{\partial t} + \frac{1}{a}\underset{\sim}{v}\cdot\nabla\underset{\sim}{v} + \frac{\dot{a}}{a}\underset{\sim}{v} + \frac{1}{a}\frac{\nabla p}{\rho} = \frac{\nu}{a^2}\nabla^2\underset{\sim}{v} - \frac{1}{a}\nabla\phi$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{a}\nabla\cdot(\rho\underset{\sim}{v}) + \frac{3\dot{a}}{a}\rho = 0 \qquad [2:5]$$

$$\frac{1}{a^2}\nabla^2\phi = 4\pi G(\rho - \bar{\rho}) = 4\pi G\bar{\rho}\,\delta$$

where $\rho$ is the density, $\bar{\rho}$ the spatial mean density and

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}} \qquad [2:6]$$

## 3 Transformation of variables

The dynamical equations, both before and after recombination, contain terms generated by the cosmic expansion. It is remarkable that these can be removed, completely in the pre-recombination and almost completely in the post-recombination equations, by transforming to new variables. These are, in the pre-recombination phase,

$$
\begin{aligned}
\underset{\sim}{v}^* &= \alpha\beta^{-1}(1 + a/a_{eq})\,\underset{\sim}{v} \\
L^* &= \alpha\,(a/a_{eq})\,L \\
dt^* &= \beta(a/a_{eq})^{-1}(1+a/a_{eq})^{-1}\,dt \\
\nu^* &= \alpha^2\beta^{-1}(1 + a_{eq}/a)\,\nu \\
p^* &= \alpha^2\beta^{-2}(a/a_{eq})^4(1+a/a_{eq})\,p
\end{aligned}
\tag{3:1}
$$

and in the post-recombination phase,

$$
\begin{aligned}
\underset{\sim}{v}^* &= \gamma\epsilon^{-1}(a/a_{eq})\,\underset{\sim}{v} \\
L^* &= \gamma\,(a/a_{eq})\,L \\
dt^* &= \epsilon\,(a/a_{eq})^{-2}\,dt \\
\nu^* &= \gamma^2\epsilon^{-1}\,\nu \\
p^* &= \gamma^2\epsilon^{-2}(a/a_{eq})^5\,p \\
\rho^* &= (a/a_{eq})^3\,\rho \\
\phi^* &= \gamma^2\epsilon^{-2}(a/a_{eq})^2\,\phi
\end{aligned}
\tag{3:2}
$$

where $a_{eq}$ is the value of $a$ at the epoch of equality when

$$\rho_m = \frac{4}{3}\,\rho_r = b \tag{3:3}$$

and $\alpha, \beta, \gamma, \epsilon$ are arbitrary constants (this freedom arising from one's ability to independently rescale length and time in the Newtonian limit). In these variables the dynamical equations become

$$\frac{\partial \underset{\sim}{v}^*}{\partial t^*} + \underset{\sim}{v}^* \cdot \nabla \underset{\sim}{v}^* + \frac{1}{b} \nabla p^* = \nu^* \nabla^2 \underset{\sim}{v}^*$$

$$\nabla \cdot \underset{\sim}{v}^* = 0 \qquad [3:4]$$

before recombination and

$$\frac{\partial \underset{\sim}{v}^*}{\partial t^*} + \underset{\sim}{v}^* \cdot \nabla \underset{\sim}{v}^* + \frac{1}{\rho^*} \nabla p^* = \nu^* \nabla^2 \underset{\sim}{v}^* - \nabla \phi^*$$

$$\frac{\partial \rho^*}{\partial t^*} + \nabla \cdot (\rho^* \underset{\sim}{v}^*) = 0 \qquad [3:5]$$

$$\nabla^2 \phi^* = \gamma^2 \epsilon^{-2} \, 4\pi G \frac{a}{a_{eq}} \, b \, \delta$$

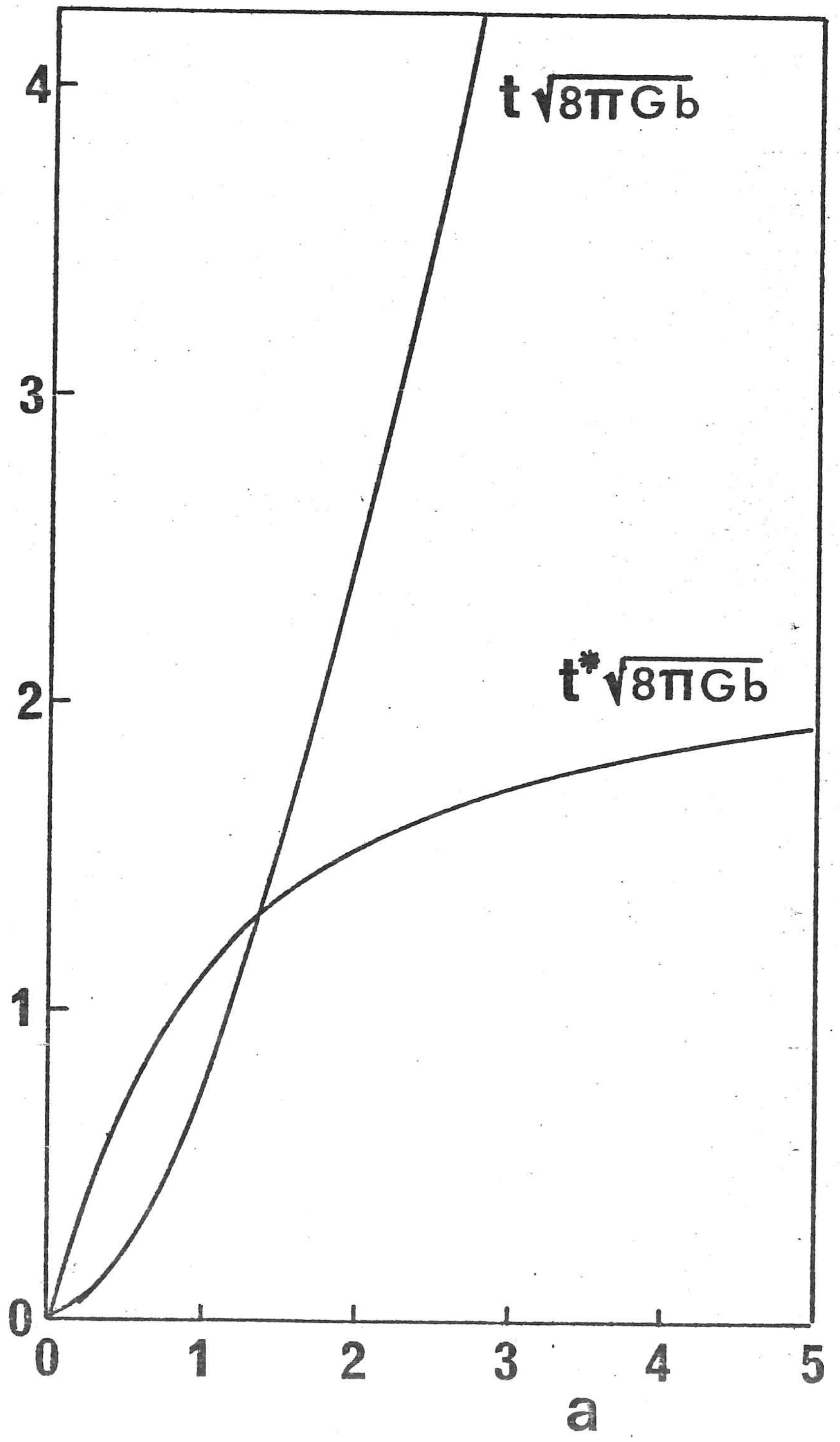
after recombination.

It is convenient to choose

$$\alpha = \beta = \gamma = a_{eq} = 1 \, , \quad \epsilon = \frac{a_{rec}}{1 + a_{rec}} \qquad [3:6]$$

Then both transformations reduce all lengths to their comoving values at the epoch of equality (note that this

Fig. 1 :- t and t* as functions of a ($a_{rec}$ = 1)




Facing page 10



transformation has already been carried out in equations [2:2],[2:5]) and give the same transformation of velocities at recombination.

Combining the defining relations for $t^*$ with the Friedmann equation (see e.g. Peebles 1971)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\left(\rho_m + \rho_r\right) \qquad [3:7]$$

and integrating we get (see fig.1)

$$\sqrt{8\pi G b}\; t = \left(1+\tfrac{4}{3}a\right)^{1/2}\left(a-\tfrac{3}{2}\right) + \tfrac{3}{2}$$

$$\sqrt{8\pi G b}\; t^* = \begin{cases} 4\sqrt{3}\left\{\arctan\sqrt{3+4a} - \frac{\pi}{3}\right\} & a < a_{rec} \\ \text{Const.} - \frac{4\,a_{rec}}{1+a_{rec}} \ln\left\{\sqrt{\frac{3}{a}+4} + \sqrt{\frac{3}{a}}\right\} & a > a_{rec} \end{cases} \qquad [3:8]$$

Perhaps the most interesting feature here is that though the Universe expands for an infinite amount of real time, for its hydrodynamic evolution it only has a finite amount of $t^*$ 'time'. Hence, just as large scale motions in the early universe are frozen because insufficient time has elapsed for them to undergo significant evolution, so in the late Universe large motions are frozen because there is not enough 'time' left for their further evolution. Thus the use of this formalism allows a very natural interpretation of the important concept of 'freezing' introduced by Ozernoi.

## 4 Evolution of turbulence

A further advantage of this transformation is that the reduction of the problem to one in a non-expanding fluid greatly simplifies the theory and assists one's intuition. This is especially true in the pre-recombination phase where the transformed equations differ from those of a conventional fluid only in having a time-dependent viscosity. In the inviscid limit even this difference disappears and we have a body of material which by mechanical substitution will yield cosmological results.

For example it is obvious from energy conservation that the mean value of $v^{*2}$ (corresponding to $(1+a)^2 v^2$ ) is constant. Thus in a radiation dominated Universe ($a \ll 1$) the mean square peculiar velocity is constant whereas in a matter dominated one ($a \gg 1$ ) it falls off as $a^{-2}$. This well known result (Lifschitz (1946)) is important because it shows that a theory which posits observable consequences of primordial turbulence (in particular the formation of galaxies) is only tenable in a Universe which is radiation dominated for a substantial part of the time prior to recombination. And it also shows that the hydrodynamics of matter dominated and radiation dominated Universes can be quite distinct. This last remark acquires more significance when one realises that one of the few analytic results in the theory of cosmic turbulence, the vorticity generation formula of Olson and Sachs (1973), was derived for a pure matter fluid in a Newtonian Universe.

It is therefore of considerable interest that we can, by applying this transformation (or rather its inverse) to the classical formula of Proudman and Reid (1954), obtain not only the formula of Olson and Sachs, but also its analogue in a Universe with arbitrary radiation content, and this without the heavy algebra of their paper. Indeed the very simplicity of the derivation tends to hide the significance of the result. But as Jones (1976) remarks "although the Olson-Sachs analysis contains no detailed spectral information it does provide a criterion for deciding whether or not inertial transfer can win out over the cosmic expansion. It will obviously be of great value to extend the Olson-Sachs type analysis yet further."

Neglecting viscosity and using the Millionshchikov hypothesis of zero fourth-order velocity cumulants (i.e. the assumption that the fourth order correlations can be related to the lower order correlations by expressions which are exact in the case of normally distributed variables, that is if 1,2,3,4 denote four velocity components at four space-time points

$$\overline{1234} = \overline{12}.\overline{34} + \overline{13}.\overline{24} + \overline{14}.\overline{23};$$

the implications of this closure hypothesis are well discussed in Monin & Yaglom) Proudman & Reid obtained the differential equation

$$\frac{d^2}{dt^2}\overline{\omega^2} = \frac{1}{3}\left(\overline{\omega^2}\right)^2 \qquad [4:1]$$

for $\overline{\omega^2}$, the mean square vorticity. This they solved in terms of the Weierstrass elliptic function $\wp$ obtaining

$$\overline{\omega^2} = \overline{\omega_0^2}\,\wp\left(\alpha + \frac{\sqrt{\overline{\omega_0^2}}\,(t-t_0)}{3\sqrt{2}}\,;\,0,4\right) \qquad [4:2]$$

where $\overline{\omega_0^2}$ is the value of $\overline{\omega^2}$ at $t_0$, these being the two constants of integration, and $\alpha$ a constant $\sim 1{\cdot}2$. The equivalent cosmological equation is

$$\overline{\omega^2} = \overline{\omega_0^2}\,\frac{a_0^2(1+a_0)^2}{a^2(1+a)^2}\,\wp\left(\alpha + \frac{\sqrt{\overline{\omega_0^2}}}{3\sqrt{2}}\int_{t_0}^{t}\frac{a_0(1+a_0)}{a(1+a)}\,dt\,;\,0,4\right) \qquad [4:3]$$

which in the matter dominated limit reduces to the formula given by Olson and Sachs. This expression is thus an extension of their result to a fluid containing both matter and radiation*.

-----------------

* This result should be viewed with caution. The neglect of viscosity is serious, the more so since vorticity production is concentrated in the small eddies where viscous effects are largest. Another possible objection is that it

## 5 The Form of the Spectrum

These results do not tell us much about the structure of the turbulence. Partial information about this can be obtained by using some of the standard concepts of classical turbulence theory such as the energy spectrum, its division into ranges, the Kolmogorov theory of the energy cascade, the universal equilibrium and its inertial subrange (Monin and Yaglom 1975).

If we define the two point velocity correlation tensor

$$R_{ij}(\underset{\sim}{r}) = \overline{u_i(\underset{\sim}{x})\, u_j(\underset{\sim}{x}+\underset{\sim}{r})} \qquad [5:1]$$

-----------------

is derived using Millionshchikov's hypothesis which is known to be unreliable. But, as pointed out by Proudman and Reid, the form of equation [4:1] does not actually depend on this hypothesis. It is quite easy to show (see e.g. Monin & Yaglom) that if one ignores viscosity,

$$\frac{d^2}{dt^2}\overline{\omega^2} = \frac{49}{90}\, s^2 \left(\overline{\omega^2}\right)^2 \qquad [4:4]$$

where $s$ is the skewness of the random variable $\partial u/\partial x$. Thus the hypothesis has only been used to assign a definite value to $s$; but whereas when Proudman and Reid wrote their paper it was thought that as one proceeded to larger values of the turbulent Reynolds number $R_e$, $s$ would tend to some asymptotic constant, it has become increasingly clear over the last ten years that in fact $s$ continues to rise with increasing $R_e$. The reason for this is the intermittent or localized nature of high Reynolds number turbulence, an aspect of the theory to which increasing attention is now being paid (see Rosenblath & van Atta 1972).

which is a well defined quantity if the turbulence is statistically homogeneous (the bar denotes an average which in this context could be taken to be spatial, but is probably better thought of as an ensemble average) and its associated Fourier transform

$$\Phi_{ij}(\underset{\sim}{k}) = \frac{1}{(2\pi)^3} \iiint e^{i\underset{\sim}{k}\cdot\underset{\sim}{r}} R_{ij}(\underset{\sim}{r})\, d^3\underset{\sim}{r} \qquad [5:2]$$

it is easy to show that in the isotropic case if the flow is incompressible

$$\Phi_{ij}(\underset{\sim}{k}) = \frac{E(k)}{4\pi k^2}\left[\delta_{ij} - \frac{k_i k_j}{k^2}\right] \qquad [5:3]$$

where the scalar function $E$ has been so defined that

$$\frac{1}{2}\overline{u^2} = \int_0^\infty E(k)\, dk \qquad [5:4]$$

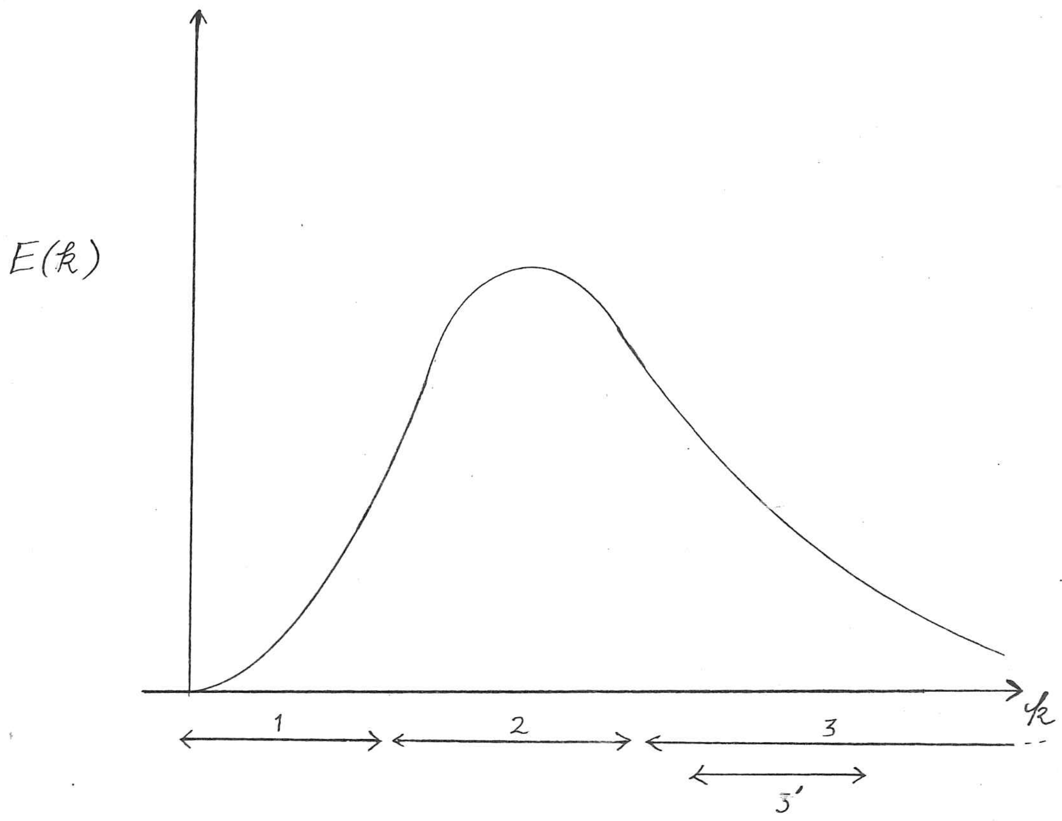
It is clear that the function $E$ provides a natural way to give a precise definition to the intuitive concept of the amount of energy contained in the eddies of a certain size; it is usually called the energy spectrum of the turbulent flow.

To study the time evolution of the turbulence it is natural to seek a dynamical equation for $E$. If we do this, differentiating the definitions given above and using the Navier Stokes equations we obtain

$$\frac{\partial}{\partial t} E(k) + T(k) + 2\nu k^2 E(k) = 0 \qquad [5:5]$$

where $T(k)$ is an expression generated by the non-linear terms in the Navier Stokes equations and depending on the third order velocity correlations. It represents the transfer of energy by inertial effects from one scale to another and is usually called the transfer term. Unfortunately the third order correlations that occur in the expression for $T(k)$ are unknown functions. If we attempt to form a dynamical equation for them it contains fourth order correlations, and in general the evolution of the correlations of a given order depends on correlations of order one higher. Thus we are unable in this way to obtain a closed system of equations. This problem of closure is central to the standard statistical theory of turbulence; although many solutions have been proposed (for example Millionshchikov's hypothesis) none can be said to be very successful. Several approximate expressions have been proposed relating $T(k)$ to $E(k)$, the best known perhaps being that due to Heisenberg:

Fig. 2    A typical energy spectrum and its division into ranges.




Facing page 17



$$T(k) = \frac{d}{dk}\left\{\gamma \int_k^\infty E(k')^{\frac{1}{2}} k'^{-3/2}\, dk' \int_0^k 2E(k')k'^2\, dk'\right\} \qquad [5:6]$$

This and other similar approximations are widely discussed in the literature and in general yield results that are in good agreement with the experimental data over fairly large ranges of wavenumber $k$. However as many of the features of the spectrum can be obtained by quite general dimensional arguments we should not perhaps be too surprised at this.

It is customary in discussions of turbulence to consider certain ranges of wavenumber, corresponding to certain inversely related length scales $L = k^{-1}$. These ranges are in no sense absolute but are defined with respect to a particular turbulent flow. A typical energy spectrum and its associated ranges is illustrated in fig. 2.

The region 1, immediately adjacent to $k = 0$, is the range of the largest eddies. These eddies decay on a time scale which is very slow relative to the general decay of the turbulence and so the form of the spectrum in this range remains nearly constant during the decay. It is this behaviour which gives rise to the so-called invariants of Loitsiansky and Saffman†. If it is assumed that $E(k)$ possesses an analytic expansion about $k = 0$ the invariant is simply the coefficient of the leading term in the expansion. However as these invariants relate to the behaviour of



† (see e.g. Monin & Yaglom 1975)

infinitely large eddies they must be considered somewhat unphysical.

The region 2, where the spectrum peaks, is the range of the energy containing eddies. The dynamical time scale here ( $L/u$ where $L$ is the typical length and $u$ the typical velocity), determines the energy dissipation rate for the entire flow. If we assume that an eddy breaks up into smaller eddies in one eddy turnover time we get

$$\varepsilon \sim u^3/L \qquad [5:7]$$

for the energy dissipation rate.

The region 3, extending from just below the scale of the energy containing eddies down to arbitrarily small scales is called the universal equilibrium range. Here the dynamical time scale is much shorter than the decay time of the energy containing eddies. In Kolmogorov's first theory it is assumed that the motion on these scales is essentially independent of the motion on the large scales and so reaches a quasi-stationary equilibrium in which the flow of energy into this range from the decay of the large eddies is balanced by its dissipation in the small eddies. This equilibrium is universal in that it depends only on the energy dissipation rate $\varepsilon$ and the viscosity $\nu$. Using these parameters one can construct the inner or micro scales of length, time and velocity

$$\eta = \nu^{3/4}\varepsilon^{-1/4} \quad \tau = \nu^{1/2}\varepsilon^{-1/2} \quad v = \nu^{1/4}\varepsilon^{1/4} \tag{5:8}$$

and when expressed in terms of these units this part of the spectrum should be the same in any turbulent flow.

If the inner length scale is very much smaller than the outer scale (of the energy containing eddies) there will be a region of the universal equilibrium range, 3', in which the effects of viscosity can be ignored. If such a region exists it is called the inertial subrange of the universal equilibrium range. Its great importance is due to the fact that in Kolmogorov's theory the spectrum in this subrange is determined by the single parameter $\varepsilon$ and thus the exact functional form can be determined by dimensional analysis. This yields the well known Kolmogorov '5/3' law for the energy spectrum in the inertial subrange.

$$E(k) \propto \varepsilon^{2/3} k^{-5/3} \tag{5:9}$$

There is now experimental evidence that such a spectrum does occur in turbulent flow at high Reynolds number.

Although Kolmogorov's ideas were developed for a conventional fluid there is no difficulty in applying them to one with a time dependent viscosity; the most useful result of his theory, that in the inertial subrange the energy spectrum should have the form [5:9] applies in a region (in wavenumber space, not physical space) where viscous effects are negligible. The main problem is simply to locate the dissipation scale $\eta$ or lower boundary of the inertial subrange. It is readily verified that

$$\nu^* = \frac{2}{9} \frac{m_p c}{\sigma} \frac{a^2}{b} \qquad [5:10]$$

before recombination so that the 'viscosity' is a monotonic increasing function of 'time'. Thus the dissipation scale will be smaller than in a fluid with constant viscosity, but the difference will be insignificant if the dynamical time scale at that length scale is less than or equal to the time scale on which the viscosity varies.

If the typical velocity on a scale $L^*$ at 'time' $t^*$ is $u^*(L^*,t^*)$ then at each 'time' $t^*$ there will be a certain scale $L_0^*(t^*)$ with

$$u^*(L_0^*, t^*) = L_0^* / t^* \qquad [5:11]$$

and such that motions on larger scales are still 'frozen' in their initial forms. On smaller scales however the inertial time scale will be much less than $t^*$ and sufficient evolution will have occured for the spectrum to lose its initial form and approach a Kolmogorov quasi-equilibrium. If the dissipation length scale $\eta^*$ is much smaller than $L^*$, i.e. if there is an inertial subrange, then the spectrum between these two scales will have the characteristic Kolmogorov form

$$E^*(k^*) \sim \varepsilon^{*\,2/3} k^{*\,-5/3} = u_0^{*\,4/3} t^{*\,-2/3} k^{*\,-5/3} \qquad [5:12]$$

as found by Kurskov and Ozernoi (1974a). The question of course is whether $\eta^*$ is sufficiently small for an inertial subrange to form. In normal laboratory flows this requires a ratio $L/\eta$ of at least $10^4$ (Hinze 1975, page 253; this reference also gives a good account of the various subranges of the spectrum). The reason for this large ratio is that the effect of viscosity extends upwards about a decade from $\eta$ and the effect of the initial spectrum down about another decade from $L$, then another couple of decades have to be left for the inertial subrange. To be conservative I will take a ratio of greater than $10^2$ as the criterion for the existence of an inertial subrange. This gives

$$L^*/\eta^* = u_0^{*\,3/2} t^{*\,3/4} \nu^{*\,-3/4} \geqslant 10^2 \qquad [5:13]$$

or

$$u_0^* \geqslant 10^{4/3} (\nu^*/t^*)^{1/2} \qquad [5:14]$$

Taking the present density of the Universe as

$$\rho_m = 1.9 \times 10^{-29} \Omega h^2 \quad [g\,cm^{-3}] \qquad [5:15]$$

(h is the Hubble constant normalized to 100 km $s^{-1}$ $Mpc^{-1}$ and $\Omega$ is the ratio of the density to the closure density) and of the microwave background as

$$\rho_r = 5 \times 10^{-34} \quad [g\,cm^{-3}] \qquad [5:16]$$

(Allen 1973) we find the present-day values

$$a = 2.9 \times 10^{4} \Omega h^2$$

$$[5:17]$$

$$b = 4.4 \times 10^{-16} \Omega^4 h^8 \quad [g\,cm^{-3}]$$

Assuming a redshift at recombination of $10^3$, almost independent of $\Omega h^2$, $a$ at recombination $\sim 29 \Omega h^2$, i.e. probably close to unity. And the viscosity just before recombination would have been

$$\sim 3.2 \times 10^{28} \Omega^{-2} h^{-4} \quad [cm^2 s^{-1}] \qquad [5:18]$$

The extreme upper limit on $t^*$ is

$$\sqrt{\frac{\pi}{6Gb}} = 1.33 \times 10^{11} \Omega^{-2} h^{-4} \quad [s] \qquad [5:19]$$

and $t^*_{rec}$ must lie somewhere between 0.1 and 0.8 times this (corresponding to $\Omega h^2 \sim 0.01$ and $\Omega h^2 \sim 1.0$). Inserting these values in the above inequality we obtain

$$U^*_o(t^*_{rec}) \geqslant 10^{10} \quad [cm\, s^{-1}] \quad (\Omega h^2 \leqslant 1) \qquad [5:20]$$

which is impossible for subsonic (and non-relativistic) turbulence (at recombination the true velocity $U_o$ is at least half $U^*_o$ ). It is easy to check that on a scale $10^{-2} L^*_o(t^*_{rec})$ the inertial response time is smaller than the time scale for variation of $\nu^*$ even at $a_{rec} \sim 30$; thus the above analysis is justified if $\Omega h^2 \leqslant 1$.

From this inequality (which could be strengthened for $\Omega h^2 << 1$) we can deduce that at recombination the dissipation scale will be within two decades of the 'melting' scale $L^*_o$ if $\Omega h^2 \leqslant 1$. In these circumstances a Kolmogorov inertial subrange will not be visible and much of the advantage of Ozernoi's theory disappears*. In particular the mass and angular momentum distributions of the galaxies formed can not be found from a simple power law spectrum with a known exponent but depend on the unknown initial spectrum. Of course it is still possible for sufficient turbulent activity to survive through the radiation era to induce galaxy formation. However there is another serious objection to cosmic turbulence as a

-----------------

* Dr. Matsuda informs me that a student of his has confirmed this result.

progenitor of galaxies; the post-recombination turbulence is highly supersonic and may be expected to produce strong shocks and highly condensed objects whereas the galaxies we observe are fairly diffuse objects (Peebles 1971b, Jones 1977). There is however another interesting possibility. Because only a finite amount of 'time' is left after recombination the motions that survive the pre-recombination and recombination damping may be 'frozen' in the post-recombination phase. If this is the case the residual motions though supersonic will not have enough 'time' to give rise to shocks and large density contrasts. This is termed by Ozernoi the 'calm' mode of evolution of cosmic turbulence to distinguish it from the usual 'rough' unfrozen mode. But whether or not this happens depends critically on the amount of damping that occurs during recombination.

# 6 References


Chan,K.L. & Jones,B.J.T. 1975 Ap.J. 200,454.

Chibisov,G.N. 1972 Sov.Ast.-AJ 18,157.

Efstathiou,G. & Jones,B.J. 1979 M.N. 186,133.

Gamow,G. 1952 Phys.Rev. 86,231.

Harrison,E.R. 1970 M.N. 147,279.

Hinze,J.O. 1975 Turbulence (2nd edn.),McGraw Hill,N.Y.

Hoyle,F. 1951 IAU Symposium on the Motions of Gaseous Masses of Cosmical Dimensions, Problems of Cosmical Aerodynamics, (Central Air Documents Office).

Jones,B.J.T. 1976 Rev.Mod.Phys. 48,107.

Jones,B.J.T. 1977 M.N. 180,151.

Kurskov,A.A. & Ozernoi,L.M. 1974a Sov.Ast.-AJ 18,157.

Kurskov,A.A. & Ozernoi,L.M. 1974b Sov.Ast.-AJ 18,300.

Kurskov,A.A. & Ozernoi,L.M. 1974c Sov.Ast.-AJ 18,700.

Lifschitz,E.M. 1946 J.Phys.USSR 10,116.

Monin,A.S. & Yaglom,A.M. 1975 Statistical Fluid Mechanics, Vol II, MIT Press, Cambridge, Massachussets.

Nariai,H. 1956a Sci.Rep.Tohuku Univ.ser.1 39,213.

Nariai,H. 1956b Sci.Rep.Tohuku Univ.ser.1 40,40.

Olson,D.W. & Sachs,R.K. 1973 Ap.J. 185,91.

Ozernoi,L.M. & Chernin,A.D. 1968a Sov.Ast.-AJ 11,907

Ozernoi,L.M. & Chernin,A.D. 1968b Sov.Ast.-AJ 12,901

Ozernoi,L.M. & Chibisov,G.V. 1971a Sov.Ast.-AJ 14,615.

Ozernoi,L.M. & Chibisov,G.V. 1971b Ap.Lett. 7,201.

Ozernoi,L.M. & Chibisov,G.V. 1972 Sov.Ast.-AJ 15,923.

Peebles,P.J.E. 1969 Ap.J. 155,393

Peebles,P.J.E. 1971a Physical Cosmology, Princeton University Press,Princeton.

Peebles,P.J.E. 1971b Ap.Space Sci. 11,443.

Proudman,I. & Reid,W.H. 1954 Phil.Trans.Roy.Soc.Lond. A 247,163.

Rosenblatt,M. & Van Atta,C. 1972 Statistical Models and Turbulence, Springer-Verlag, Berlin.

Weizsacker,C.F. von 1951 Ap.J. 114,165.

Woody,D.P. & Richards,P.L. 1979 Phys.Rev.Lett. 42,925.

"Thus a flat rotating disc will remain. . . Its rotation will not be uniform and the differences in rotational velocity will go on producing turbulence. . . The remaining turbulence exerts friction and thereby dissipates energy. Therefore the rotation cannot be stable unless it becomes uniform. . . The result will be the contraction of part of the body towards the centre, while the gravitational energy set free by the contraction enables the rest of the mass to return to the surrounding cosmic space, carrying with it most of the angular momentum of the body."

C.F. von Weizsacker 1951

## 1 Introduction

Accretion discs are now widely accepted as models for some energetic astrophysical objects ranging in scale from quasars and giant radio sources to dwarf novae and galactic X-ray sources. The idea is that if a compact object (such as a black hole or white dwarf) accretes matter carrying significant angular momentum the matter will tend to form a flat disc rotating about the compact object (in some binary systems the optical light curves provide quite direct evidence for the existence of such discs; Smak (1971), Bath (1972)). If by some mechanism angular momentum can be transported outwards through the disc and the specific angular momentum of the matter in the disc is an

"Thus a flat rotating disc will remain. . . Its rotation will not be uniform and the differences in rotational velocity will go on producing turbulence. . . The remaining turbulence exerts friction and thereby dissipates energy. Therefore the rotation cannot be stable unless it becomes uniform. . . The result will be the contraction of part of the body towards the centre, while the gravitational energy set free by the contraction enables the rest of the mass to return to the surrounding cosmic space, carrying with it most of the angular momentum of the body."

C.F. von Weizsacker 1951

## 1 Introduction

Accretion discs are now widely accepted as models for some energetic astrophysical objects ranging in scale from quasars and giant radio sources to dwarf novae and galactic X-ray sources. The idea is that if a compact object (such as a black hole or white dwarf) accretes matter carrying significant angular momentum the matter will tend to form a flat disc rotating about the compact object (in some binary systems the optical light curves provide quite direct evidence for the existence of such discs; Smak (1971), Bath (1972)). If by some mechanism angular momentum can be transported outwards through the disc and the specific angular momentum of the matter in the disc is an increasing function of radius there will be a compensating inward flux of matter (except perhaps at the outer edge); as the matter moves deeper into the gravitational potential well of the central object a newtonian calculation suggests that half of its gravitational binding energy can be dissipated (if the object is not a black hole more energy can be released when the matter hits the surface) and in this way objects with very high luminosities and hard spectra can be formed.

The basic structure of such discs has been investigated by many workers. Perhaps the earliest description of the accretion disc mechanism (though in a different context) is that in von Weizsacker (1951) which refers to an investigation being carried out by Lust and Trefftz (Lust 1952). More recent investigations are those of Pringle and Rees (1972), Shakura and Sunyaev (1973), Lynden-Bell and Pringle (1974), and Eardley and Lightman (1975). However in all this work there is a problem with the angular momentum transport process. The molecular viscosity is far too small (at least in conventional models) to transfer significant amounts and it is normal to posit a turbulent or magnetic viscosity the magnitude of which is estimated using some simple phenomenological model (for example the $\alpha$ disc models of Shakura and Sunyaev). The usual argument is that because the Reynolds number of the disc is so large the flow is probably turbulent; but with the exception of a preliminary analysis by Stewart (1975,1976) I know of no previous attempt to prove this.

The problem I wish to consider is this. There certainly exist axisymmetric solutions to the equations of fluid mechanics describing gas discs where the streamlines are nearly circular (the molecular viscosity causing a slight radial drift); but supposing such a disc to be established would this state persist or would the flow develop irregularities and become turbulent?

This question belongs to the theory of hydrodynamic stability which is extensive and in many areas still ill-understood; it has three main branches, the inviscid, the viscous, and the finite amplitude theory. The inviscid theory of hydrodynamic stability originated with a study by Helmholtz (1868) of the stability of the vortex sheet, a problem considered in greater detail by Kelvin (1871), but is largely the creation of Rayleigh (1880). The theory proceeds by a straightforward linearization of the Euler equations about some flow invariant under a three parameter symmetry group and a Fourier decomposition relative to these symmetries (usually translation in time and two spatial directions). In this way a linear second order ordinary differential system is obtained, a mathematically tractable problem but one whose content, both mathematical and physical, is subtle and was previously obscure (consider for example Kelvin's complaint of the 'disturbing infinity' in Rayleigh's work and the many discussions of the completeness of the normal modes, see e.g. Lin (1961)). The fault is not so much in the theory as in the approximation on which it is based. It is well known that there are ambiguities associated with the Euler equations which have to be resolved by going to the Navier-Stokes equations; so too Lin (1945a,b,1946) was able to remove many of the problems of the inviscid theory by relating it to the viscous theory.† This theory, primarily the creation of Orr, Sommerfeld and Heisenberg, is based on an analogous linearization of the Navier Stokes equations and has the formal advantage that the equations are non-singular; however their study, analytic or numerical, is much harder (essentially because there are two distinct scales in the problem, the viscous and the dynamical). The non-linear theory, developed among others by Meksyn & Stuart (1951), is the most complicated and has not yet achieved a standard form.

Examples are known of flows where the transition to turbulence appears to be mediated by instabilities whose description requires the use of each of these theories. In general the flows which exhibit inviscid instability are those with maxima in the modulus of the vorticity (the paradigm being the vortex sheet). However certain flows which are stable in the inviscid approximation are destabilised by viscous effects (the classical example of this is plane Poiseuille flow in which the effect was first demonstrated by



† These ambiguities can also be resolved by consideration of the inviscid initial value problem; see e.g. K.M. Case (1961) JFM 10,420.

Heisenberg). This slightly surprising result can be understood as a consequence of the stabilising effect of vorticity conservation in inviscid flows (at least in two dimensions); with a slight viscosity vorticity generated at the boundaries can diffuse into the flow. My feeling is that this viscous destabilisation is primarily a boundary effect and thus not likely to be important in Astrophysical problems. Finally it is a curious fact that the flow in which the transition to turbulence was first described by Reynolds, cylindrical Poiseuille flow, appears completely stable in both linear theories. However there is good evidence, both theoretical and experimental, that this flow though stable to infinitesimal perturbations is unstable to perturbations of finite amplitude. It seems quite probable that the axisymmetric gas flows I consider are unstable to finite amplitude perturbations (as indicated I do not think viscous destabilisation important), but this would be an exceptionally complicated process to analyze and would in any case require a prior study of the linear inviscid problem. For these reasons I have confined myself to the linearised inviscid theory† (in consequence this investigation is only concerned with dynamical instabilities and does not consider possible secular instabilities, for example of the type investigated by Goldreich & Schubert (1967), and Fricke (1968)).

This is itself no easy problem unless the perturba-



† For a viscous stability analysis of circular Couette flow see the classic paper by G.I. Taylor, Phil.Trans. Roy.Soc.(London) A223,289 (1923).

tions are also axisymmetric. In this special case the problem is well understood the solution being essentially due to Rayleigh (1916). Consider a ring of fluid at some radius and expand it so that it lies at a slightly larger radius; the specific angular momentum, or equivalently the circulation around the ring, is conserved during this process. If the angular velocity of the ring at its new location is now greater than that of the surrounding fluid it will experience a force tending to increase the displacement, if less the force will be a restoring one; thus if the circulation increases (in magnitude) as one moves outwards the rotation has a stabilising effect on axisymmetric perturbations, if it decreases a destabilising one. Another way of expressing this is that the effect is stabilising or destabilising according as the angular velocity (the global rotation) and the vorticity (the local rotation) are parallel or anti-parallel. In its various forms this is known as Rayleigh's criterion for the centrifugal instability of axisymmetric flows. A convenient quantitative measure is that given by

$$\kappa^2 = 2\Omega(2\Omega + r\Omega') = 2\Omega\omega_0 \qquad [1:1]$$

which is the square of the frequency with which the fluid ring would execute radial oscillations in the absence of buoyancy forces (an imaginary frequency indicating as usual an exponential instability); this coincides with the epicyclic frequency of orbital theory. Defining the analogous Brunt-Vaisala frequency $N$ for the buoyancy forces it is reasonable to expect that the condition for an axisymmetric rotating radially stratified system to be stable with respect to axisymmetric perturbations is

$$K^2 + N^2 \geqslant 0 \qquad [1:2]$$

In section 4 I give a formal proof of this result.

It is clear that this argument is inapplicable to non-axisymmetric perturbations where angular momentum can be redistributed by azimuthal pressure gradients, however Chandrasekhar (1960, summarised in the first edition of his book 'Hydrodynamic and Hydromagnetic Stability' (1961) but not the second) claimed that the result did in fact apply to all perturbations, both axisymmetric and non-axisymmetric. This was disputed by Howard and Gupta (1962) who pointed out that a cylinder of fluid at rest surrounded by one in uniform rotation must be Kelvin-Helmholtz unstable even though the circulation increases outwards. In general the stability of inviscid axisymmetric flows to non-axisymmetric perturbations appears to have received very little attention. This perhaps reflects the difficulty of the problem: in cylindrical geometry there is no equivalent of Squire's theorem (which in Cartesian geometry enables one to reduce the problem of the stability of plane parallel flows against three dimensional perturbations to that against two dimensional perturbations) and the equations contain such additional terms that any attempt to apply what may be termed the standard procedure, the manipulation of the equations into forms where the Wronskian relation between the solution and its complex conjugate involves positive definite integrals, requires great ingenuity and patience if non-trivial results are to be obtained.

Most workers have attempted to produce generalisations of results obtained in the theory of plane parallel flows (these are reviewed in Drazin and Howard (1966)), in particular Rayleigh's inflection point theorem, the Richardson number theorem of Miles (1961) and Howard (1961), and the semicircle theorem of Howard (1961).

Of these the first has rather an odd history. At the end of his famous paper of 1880, 'On the Stability or Instability of Certain Fluid Motions' Rayleigh stated

> If the stream lines of the steady motion be concentric circles instead of parallel straight lines, the character of the problem is not greatly changed. It may be proved that, if the fluid move between two rigid concentric circular walls, the motion is stable, provided that in the steady motion the rotation either continually increases or continually decreases in passing outwards from the

axis.

Thus it was known to Rayleigh that the analogue for axisymmetric flows of his inflection point theorem was that there should be an extremum in the vorticity, a result which Kelvin (1880) claimed to have 'nearly reached in the year 1875 by rigid mathematical investigation . . . . but I was anticipated in the publication of it by Lord Rayleigh'. Despite this the result seems to have been forgotten among astrophysicists so that Lebovitz (1967; repeated in Fricke and Kippenhahn (1972)) could suggest that the generalisation might be an inflection in the azimuthal velocity. The correct result is stated without proof or attribution in Spiegel and Zahn (1970) and Zahn (1974) (apparently by private communication from D.Gough to E.Spiegel to J.P.Zahn). However this result has only been demonstrated for two dimensional perturbations of a perfect fluid in which case it is a particular instance of some much more powerful results obtained by Arnol'd (1966) (who shows Liapunov rather than linear stability; a special case of his argument appears to have been discovered by Drazin and Howard and there are suggestions of it in Kelvin's work).

The Richardson criterion has been independently generalised by Sung (1974a) and Lalas (1975) (the later considers a general swirling flow) though Sung's secondary argument (197~~4b~~5) that it is both a necessary and a sufficient condition is false† (if it were true any



† Sung's argument is local and takes no account of the boundary conditions, cf. remarks on page 38.

homogeneous shear flow would be unstable). In section 4 I obtain a much simpler proof of this result and show that if rather artificial boundary conditions are allowed it is necessary and sufficient; I also restrict the frequency of a growing mode to a certain region. A curious point which emerges from the analysis is that allowing the perturbations to have a three dimensional structure produces effects very similar to those of a stable density stratification. This partially explains why there is no analogue of Squire's theorem and the difficulty of proving any results for general three dimensional perturbations.

Several generalisations of the semicircle theorem (which states that the complex phase speed $c_r + i c_i$ of an unstable disturbance in a plane parallel flow with rigid boundaries lies within the semicircle having as diameter the range of velocity of the basic flow;†

$$\left[c_r - \frac{V_{max} + V_{min}}{2}\right]^2 + c_i^2 \leqslant \left[\frac{V_{max} - V_{min}}{2}\right]^2 \qquad [1:3]$$

~~see~~ appendix A for a proof) have been claimed, but all those known to me either involve special restrictions on the basic flow or give a circle too large to be of much use. Thus Warren (1976) proves the natural generalisation of the semicircle theorem, that the pattern speed of any unstable mode lies within the semicircle having as diameter the range of angular velocities of the basic flow, under the assump-



† See also G.T.Kochar and R.K.Jain (1979)JFM,91,489.

tion that the basic flow is that of a perfect fluid with the Rayleigh discriminant $K^2$ nowhere less that the square of the mean angular velocity; Rathy and Chandra (1972) that if the fluid is heterogeneous and incompressible and

$$(\rho_0\Omega^2)' < 0 \;,\; (\rho_0\Omega^2 r^4)' > 0 \;,\; \rho_0' < 0 \qquad [1:4]$$

it lies within a circle centred on the origin of radius the maximum angular velocity; Eckart (1963) that for a general fluid it lies within a semicircle concentric with but larger than the natural one.

A related problem in the theory of rotating stars was considered by Cowling (1951). His primary concern was not whether differential rotation could lead to instability but whether rotation could suppress the Rayleigh- Taylor instability and hence the onset of convection. This can certainly happen for axisymmetric disturbances; a fluid with a negative $N^2$ will still be dynamically stable to axisymmetric perturbations if $K^2 + N^2$ is positive (though from the analogy with two component stratifications of the thermohaline type an axisymmetric double-diffusive secular instability should exist). From a local study of the non-axisymmetric perturbations Cowling concluded that though rotation modified convection it could not prevent it. Indeed he argued that even if the density stratification was statically stable except at one point where it was neutral, differential rotation would produce a dynamical instability. However as with the analyses of Sung (197~~4b~~5) and Stewart (1976) this proves too much; the common fault is the attempt to prove the existence of an instability from a local approximation. This works when the problem is self-adjoint and can be cast into the form of a variational principle because local test functions can then determine global upper bounds on the eigenvalues (and for this reason local criteria can be found for stability to axisymmetric perturbations) but such cases are exceptional.

I should also mention a paper by Hazelhurst (1963) who considers the stability of the spiral flow

$$\underset{\sim}{U}_0 = \frac{A}{r}\underset{\sim}{e}_\vartheta + \frac{B}{r}\underset{\sim}{e}_r \qquad [1:5]$$

of a perfect fluid. He claims that it is unstable if the flow spirals inwards and stable if it spirals outwards; however an elementary integration of his equation [31] shows this to be incompatible with his boundary conditions.

## 2 The Basic Model

Consider the steady axisymmetric flow with pressure $p_0$ and density $\rho_0$ in cylindrical polar coordinates $r, \vartheta, z$

$$\underset{\sim}{U}_0 = r\Omega\,\underset{\sim}{e}_\vartheta \qquad [2:1]$$

supposed maintained by some suitable combination of a pressure gradient and central (gravitational) force. To simplify the analysis a homogeneous vertical structure is assumed (but see discussion below); this translational symmetry is needed to remove the $z$ derivatives and obtain a system of ordinary rather than partial differential equations.

If we linearize about this basic flow and make small (Eulerian) perturbations

$$\underset{\sim}{u} := u \, \underset{\sim}{e}_r + v \, \underset{\sim}{e}_\vartheta + w \, \underset{\sim}{e}_z \qquad [2:2]$$

in velocity, $p$ in pressure and $\rho$ in density, we obtain a system of linear differential equations. Using the symmetries of the basic flow we Fourier analyze in $t$, $\vartheta$, $z$ and consider each Fourier component separately; symbolically if $q$ denotes any perturbation quantity we write

$$q(r,t,\vartheta,z) := q(r)\, e^{i(\omega t - m\vartheta - \beta z)} \qquad [2:3]$$

The linearized Euler equations then become (cf. Lalas 1975)

$$\sigma w = \beta p / \rho_0$$

$$\sigma v = m p / (r \rho_0) + i \omega_0 u \qquad [2:4]$$

$$p' = \rho_0 (2 \Omega v - i \sigma u) - \frac{g}{c^2} p + i \rho_0 \frac{N^2}{\sigma} u$$

and the linearized continuity equation is

$$u' = -\frac{u}{r} + \frac{imv}{r} + i\beta w - \frac{i\sigma p}{\rho_0 c^2} + \frac{g}{c^2} u \qquad [2:5]$$

In these equations

$$\omega_0 = 2\Omega + r\Omega' \qquad [2:6]$$

is the vorticity of the basic flow and

$$\sigma = \omega - m\Omega \qquad [2:7]$$

can be interpreted as $m$ times the angular velocity with which the perturbation appears to rotate as seen in a frame rotating with the basic flow. The quantity $N$ is the Brunt-Väisälä frequency, i.e. the frequency with which a fluid element would execute radial oscillations under the influence of buoyancy forces alone

$$N^2 = \frac{p_0'}{\rho_0}\left(\frac{\rho_0'}{\rho_0} - \frac{1}{c^2}\frac{p_0'}{\rho_0}\right), \qquad [2:8]$$

$c$ is the sound speed (the square root of the ratio of the Lagrangian changes in $p$ and $\rho$ ) and

$$g = -\frac{p_0'}{\rho_0} \qquad [2:9]$$

can be regarded as an effective local gravity.

The full set consists of two algebraic and two differential equations: avoiding differentiation as far as possible we use the two algebraic equations to eliminate $v$ and $w$ whereupon we obtain the differential system

$$\begin{aligned} u' &= \frac{g}{c^2}u - \frac{1}{r}\left[1 + \frac{m\omega_0}{\sigma}\right]u + \frac{i}{\rho_0\sigma}\left[\beta^2 + \frac{m^2}{r^2} - \sigma^2\right]p \\ p' &= -\frac{g}{c^2}p + \frac{2m\Omega}{r\sigma}p + i\rho_0\left[\frac{\kappa^2+N^2}{\sigma} - \sigma\right]u \end{aligned} \qquad [2:10]$$

for $u$ and $p$. This constitutes the fundamental system of equations. Further elimination yielding a single second order equation is possible, but neither necessary nor desirable.

In deriving this system it is necessary to make the physically unrealistic assumption of a homogeneous vertical structure in [2:1]. However the model may not be as bad as it appears on first sight; it can be made to look much more respectable if one notices that (essentially because it has reflection as well as translational symmetry in $z$) by superimposing the solutions for $+\beta$ and $-\beta$ one can produce solutions where $w$ and $p$ are zero on two horizontal surfaces. These represent a slab of fluid confined between two rigid horizontal surfaces or confined by an external medium of negligible density but capable of exerting a pressure, i.e. a free boundary condition on the two horizontal surfaces. A more radical approach is to consider the two dimensional form of the above equations (obtained by setting $\beta = 0$) as phenomenological equations for the vertically averaged structure of a thin disc. This is probably the more satisfactory interpretation.

This system has singular points at $r = 0$, $r = \infty$ and the point (or points) where $\sigma = 0$. While the singularities at 0 and $\infty$ are essentially coordinate singularities and have little physical significance, those which occur when $\sigma = 0$, the corotation singularities (or critical layers), are of great importance. The physical reason for this is not hard to see; at such points the perturbation is stationary in the corotating frame and strong resonance effects are to be expected. It should perhaps be pointed out that the 'Lindblad Resonances' where

$$\sigma^2 = \kappa^2 = 2\Omega\omega_0 \qquad [2:11]$$

are not singular points of the system; this is because we are considering a fluid system and have therefore fewer degrees of freedom than are present in a treatment starting from the full Liouville equation.

The fundamental question is this: does the system [2:10] admit a non-trivial solution satisfying suitable boundary conditions and such that the frequency $\omega$ has a negative imaginary part? If it does we have found an exponentially growing disturbance of the basic flow [2:1] which is therefore unstable. The converse is not true; the flow may be destabilised by non-linear effects and even in the linear theory there may exist disturbances associated with the continuous spectrum of non-analytic modes which grow algebraically in time (Chimonas 1979). However for convenience we will use 'stable' to mean the absence of an exponentially growing normal mode.

In answering this question it turns out to be advantageous to sacrifice linearity (which is of little use if one has no explicit solutions) for a reduction in order. If we define a new variable

$$\zeta = \frac{-ir\rho_0 u}{p} \qquad [2:12]$$

we find that it satisfies the (Riccati) differential equation are not singular points of the system; this is because we are considering a fluid system and have therefore fewer degrees of freedom than are present in a treatment starting from the full Liouville equation.

The fundamental question is this: does the system [2:10] admit a non-trivial solution satisfying suitable boundary conditions and such that the frequency $\omega$ has a negative imaginary part? If it does we have found an exponentially growing disturbance of the basic flow [2:1] which is therefore unstable. The converse is not true; the flow may be destabilised by non-linear effects and even in the linear theory there may exist disturbances associated with the continuous spectrum of non-analytic modes which grow algebraically in time (Chimonas 1979). However for convenience we will use 'stable' to mean the absence of an exponentially growing normal mode.

In answering this question it turns out to be advantageous to sacrifice linearity (which is of little use if one has no explicit solutions) for a reduction in order. If we define a new variable

$$\zeta = \frac{-ir\rho_0 u}{p} \qquad [2:12]$$

we find that it satisfies the (Riccati) differential equation

$$\zeta' = \left(\frac{\rho_0'}{\rho_0} + \frac{2g}{c^2}\right)\zeta + \frac{1}{r\sigma}\left\{ r^2\left(\beta^2 - \frac{\sigma^2}{c^2}\right) + (m - 2\Omega\zeta)(m - \omega_0\zeta) + (N^2 - \sigma^2)\zeta^2 \right\} \qquad [2:13]$$

The significance of this transformation is worth considering. Physically in a linearized theory the absolute magnitudes and phases are irrelevant; what are important are the relative magnitudes and phases of the perturbation components, information contained in the variable $\zeta$. Mathematically the linearity of the system [2:10] means that the evolution it determines in (a copy of) $\mathbb{C}^2$ preserves the ray structure (i.e. two sets of initial data which belong to the same ray, or one dimensional subspace, evolve into sets of final data which also belong to the same ray). Thus it also defines an evolution on the rays, i.e. in the projective space $P_1(\mathbb{C})$. This space can be identified either with the two sphere or $\mathbb{C} \cup \{\infty\}$, the identification being given by the Riemann sphere construction (stereographic projection from one pole to the tangent plane at the other pole). It is probably better to think of $\zeta$ as a point on the sphere because this makes the topological structure more evident and avoids the need for special treatment of the point at infinity.† Of course considerable information is lost in this projection, but it is information which is not needed



† To see this make the substitution $\eta = \zeta^{-1}$; the equation $\zeta' = a + b\zeta + c\zeta^2$ then becomes $-\eta' = a\eta^2 + b\eta + c$ which is of the same type.

for the solution of the eigenvalue problem (at least formally, in numerical computation there are some reasons for retaining it; see chapter 3). Thus the projective equation should display the fundamental behaviour of the system more directly that the original linear system. And if we could solve the projective equation, the solution of the full linear system would be reduced to quadratures (indeed it was by this method that I obtained the exact solutions described in section 11).

## 3 The Boundary Value Problem

For convenience initially suppose the system to be bounded by two rigid concentric cylindrical boundaries at radii $a$ and $b$. Then the appropriate boundary conditions are that the radial velocity perturbation should be zero at the boundaries which implies

$$\zeta = 0 \qquad r = a, b \tag{3:1}$$

In fact most of our results will be obtained under the more general assumption that the boundary conditions require the value of $\zeta$ to be real; the extent to which the results are contingent on this particular choice of boundary condition, its physical significance, and the possible alternative choices will be discussed later (in section 10).

The differential equation [2:13] together with the boundary conditions [3:1] seems to constitute a well defined eigenvalue problem; there is however one residual ambiguity which must be resolved, the path from $a$ to $b$ along which it is to be integrated has not been specified. The question is non-trivial because of the corotation singularities which occur when $\sigma = 0$ (we have tacitly assumed that $\Omega$ has been analytically extended off the real axis into the complex $r$ plane). Although it seems probable that this should be along the real axis the correct answer can only be obtained by a detailed analysis of the effect of viscosity on the problem (see Lin(1945a,b 1946)) or by consideration of the initial value problem. This shows that for growing modes the path of integration should be along the real axis (or homologous to this path). The correct choice for neutral modes can then be found as the limiting case of the choice for growing modes, i.e. along the real axis, but indented at a corotation point in the sense opposite to that in which the singularity would move if the frequency were given a small negative imaginary part.

It is worth noting that this choice of path destroys the symmetry of the boundary value problem under complex conjugation (or more physically an 'arrow of time' has been incorporated by considering the effect of irreversible viscous processes operating in an internal shear layer about the corotation point). Thus an anti-spiral theorem of the type stated by Lynden-Bell and

Ostriker (1967) which depends for its proof on the time reversibility of the system does not apply in this case. (A good example of the danger of time reversibility arguments in fluid dynamics is the paradox that while a candle can easily be blown out it is very hard to suck it out).†

## 4 A Sufficient Condition for Stability

In the fundamental differential equation [2:13] the only non-real quantities are $\sigma$ and $\zeta$. This allows a very simple proof of a sufficient condition for stability. Suppose there exists a growing mode. Then the frequency $\omega$ has a strictly negative imaginary part and the path of integration can be along the real axis. Let us look at the imaginary part of $\zeta'$ for real values of $\zeta$. We find easily

$$\mathrm{Im}(\zeta') = -\frac{\mathrm{Im}(\omega)}{r|\sigma|^2}\left\{\beta^2 r^2 + \frac{|\sigma|^2 r^2}{c^2} + (m - 2\Omega\zeta)(m - \omega_0\zeta) + \zeta^2(N^2 + |\sigma|^2)\right\} \qquad [4:1]$$

Now if the boundary conditions are such as to require the initial and final values of $\zeta$ to be real this expression cannot be of one sign for all $r$ in the range of integration and all real values of $\zeta$. The easiest way to see this is to consider $\zeta$ as a point on the Riemann sphere and the differential equation [2:13] as defining a vector field on the sphere. The real values of $\zeta$ define a great



† cf. G.Birkhoff, Hydrodynamics, Princeton(1950).

circle on this sphere dividing it into two hemispheres. If the vector field is regular and on this circle always points from one hemisphere into the other, it is clear that an integral curve of the vector field which starts on the boundary between the hemispheres can never return to that boundary but is forever trapped in the hemisphere it enters.

It follows that if the boundary conditions require $\zeta$ to assume real values a sufficient condition for stability is

$$\beta^2 r^2 + \frac{|\sigma|^2 r^2}{c^2} + (m - 2\Omega\zeta)(m - \omega_0\zeta) + \zeta^2(N^2 + |\sigma|^2) \geqslant 0 \qquad [4:2]$$

$$\forall \zeta \in \mathbb{R} \qquad \forall r \in [a, b]$$

This can only hold for all values of the frequency $\omega$ if the stronger inequality

$$\beta^2 r^2 + (m - 2\Omega\zeta)(m - \omega_0\zeta) + \zeta^2 N^2 \geqslant 0 \qquad [4:3]$$

is satisfied; this is equivalent to

$$4(\beta^2 r^2 + m^2)(K^2 + N^2) - m^2(2\Omega + \omega_0)^2 \geqslant 0 \qquad [4:4]$$

or

$$4N^2(\beta^2 r^2 + m^2) + 4\beta^2 r^2 K^2 \geqslant (m r \Omega')^2 \qquad [4:5]$$

(incidentally if we set $m=0$ in this we have a proof of the result previously obtained by physical reasoning, that a sufficient condition for stability to axisymmetric disturbances is

$$K^2 + N^2 \geqslant 0 \qquad [4:6]$$

).

If we define a generalised local Richardson number

$$Ri = \frac{\beta^2 K^2 + N^2(\beta^2 + m^2/r^2)}{(m\,\Omega')^2} \qquad [4:7]$$

the above condition becomes

$$Ri \geqslant \frac{1}{4} \qquad [4:8]$$

In this way we find that a sufficient condition for the stability of an arbitrary axisymmetric flow is that the generalised Richardson number should everywhere exceed 1/4. The analogous result for plane parallel flows was suggested by G.I.Taylor and proved (under certain restrictions later removed by Howard(1961)) by Miles (1961). The axisymmetric flow result agrees with that obtained (using a method based on that of Howard) by Chao-Ho Sung (1974). The advantage of the preceding derivation is that it does not require one to guess the (non-trivial) transformations which will eventually lead to a positive definite integral expression. The method can probably be applied to any inviscid hydrodynamic stability problem where there is sufficient symmetry to reduce the problem to a system of two linear ordinary differential equations; the general statement would seem to be that in all such systems a sufficient condition for stability is that the equations should be oscillatory at the critical layer (or corotation point) for all real frequencies. It is easy to show that for the general axisymmetric swirling flow

$$\underset{\sim}{u}_0 = r\Omega\,\underset{\sim}{e}_\theta + W\underset{\sim}{e}_z \qquad [4:9]$$

the appropriate generalised Richardson number is

$$Ri = \frac{\beta^2(N^2+K^2) + m^2N^2/r^2 - 2m\beta\Omega W/r}{(m\Omega' + \beta W')^2} \qquad [4:10]$$

(compare Lalas(1975)). In section 10 I extend these results to more general boundary conditions.

Even if the generalised Richardson number is less than 1/4 at some point this argument can be used to restrict the possible values of $\omega$. For any unstable mode there must exist a value of $r$ in $[a, b]$ for which

$$\left(\beta^2 r^2 + m^2 + \frac{|\sigma|^2 r^2}{c^2}\right)\left(K^2 + N^2 + |\sigma|^2\right) \le \frac{m^2(2\Omega + \omega_0)^2}{4} \qquad [4:11]$$

This implies

$$\frac{|\sigma|^4}{m^2 c^2} + |\sigma|^2 \left\{ \frac{K^2+N^2}{m^2 c^2} + \frac{1}{r^2} + \frac{\beta^2}{m^2} \right\} \leqslant \frac{(\Omega')^2}{4} - \frac{N^2}{r^2} - \frac{\beta^2}{m^2}(K^2+N^2) \qquad [4:12]$$

which confines $\omega$ to a disc centred on $m\Omega$. It follows that $\omega$ lies within the union of these discs as $r$ runs from $a$ to $b$. This result bears some resemblence to the semicircle theorem of Howard (1961) but is quite distinct. It is, as will be shown, the best possible result if the only restriction placed on the boundary points and values is that the boundary values of $\zeta$ be real; the semicircle theorem (for plane parallel flows) requires the special boundary values of 0 or $\infty$.

## 5 Importance of The Neutral Modes

The stability result found in the previous section is quite weak: however it provides the starting point from which stronger results can be derived by studying the neutral modes. The essential tool is a simple result on the persistence of solutions to a boundary value problem when the parameters of the equation are varied. The argument, which will be used several times, is that if a growing mode exists for certain values of the parameters and we can vary these in a continuous manner until they lie in a range where from some stability criterion we know there to be no growing modes, then at some point during the variation the mode must become neutral (the mode can not disappear unless one of the corotation singularities moves onto the path of integration). It can not become an isolated non-singular neutral mode because these when perturbed remain neutral; however it is possible for two such to coalesce and produce one growing and one decaying mode (an exact analogy is the behaviour of the zeros of a polynomial with real coefficients when these are varied; in principle complex zeros can arise from zeros of multiplicity greater than two, but these are not structurally stable). It follows that the limit of an unstable mode is either a singular neutral mode or a double neutral mode. Therefore stability or instability having been demonstrated for a range of parameter values, if it can be shown that in an extension of this range there are no singular neutral modes and no double neutral modes then the stability result holds for the enlarged range.

To do this we must seek characterisations of the singular and double neutral modes that are incompatible with the boundary conditions. In this we are helped by the fact that $\sigma$ is now a real quantity and thus $\zeta$ is the only quantity left in the equation which can assume a complex value. And indeed if the value of $\zeta$ is real at a regular point, it must remain so until one of the singular points is reached. In particular if the boundary conditions require $\zeta$ to assume real values, this property propagates inwards as far as the nearest singularity, and if for simplicity we temporarily assume a monotone rotation law so that there is only one corotation resonance, it follows that $\zeta$ must assume real values on either side of this point.

## 6 Structure at Corotation I

From the preceding discussion it is clear that an analysis of what happens at the corotation singularity is needed for an understanding of the singular neutral modes. Though this can be done in terms of the Riccati equation [2:13] (see section 7) there is some advantage in temporarily reverting to the linear form [2:10]. Examination of this system shows that in the neighbourhood of the corotation point we can write it in the form†

$$y' = \frac{\sigma'}{\sigma}\left[A_0 + A_1\sigma + A_2\sigma^2 \cdots\right] y \qquad [6:1]$$

where

$$y = \begin{pmatrix} u \\ p \end{pmatrix} \qquad [6:2]$$

and the $A_i$ are constant 2x2 matrices. The corotation point is a regular singular point and the solutions about this point can be obtained by the method of Frobenius (see e.g. Hille (1969)). If we write [6:1] as



† Assuming $\sigma$ to have a simple zero; an infinitesimal variation of $\omega$ will remove higher even order zeros and reduce odd order ones to this case.

$$\sigma \frac{d}{d\sigma} y = [A_0 + A_1\sigma + \cdots] y \qquad [6:3]$$

and

$$y = \sigma^c (y_0 + y_1\sigma + y_2\sigma^2 + \cdots) \qquad [6:4]$$

we find

$$(c+n) y_n = \sum_{i+j=n} A_i y_j \qquad [6:5]$$

The indicial equation is

$$|cI - A_0| = 0 \qquad [6:6]$$

which implies that the exponents are the eigenvalues of $A_0$, and provided these do not differ by an integer the recurrence relation [6:5] allows one to to find two independent series solutions of the differential system. In the specific problem we are considering the matrix $A_0$ is

$$-\frac{1}{m\Omega'} \begin{bmatrix} -\frac{m\omega_0}{r} & i(\beta^2 + \frac{m^2}{r^2})/\rho_0 \\ i\rho_0(K^2+N^2) & \frac{2m\Omega}{r} \end{bmatrix} \qquad [6:7]$$

(all quantities being evaluated at corotation) and the indicial equation is

$$c^2 - c + \{\beta^2 K^2 + (\beta^2 + m^2/r^2) N^2\}/(m\Omega')^2 = 0 \qquad [6:8]$$

giving

$$c = \tfrac{1}{2} \mp \sqrt{\tfrac{1}{4} - Ri} = \tfrac{1}{2}(1 \mp \nu) \qquad [6:9]$$

where

$$Ri = \frac{\beta^2 K^2 + (\beta^2 + m^2/r^2) N^2}{(m\Omega')^2} \qquad [6:10]$$

is the generalised 'Richardson Number' previously introduced.

Let us now consider how this may be interpreted in terms of the Riccati variable $\zeta$. Excluding the special cases when $\nu = 0$ and $\nu = 1$ and retaining only the leading terms we find

$$\zeta \sim \frac{A\sigma^{\frac{1+\nu}{2}} + B\sigma^{\frac{1-\nu}{2}}}{C\sigma^{\frac{1+\nu}{2}} + D\sigma^{\frac{1-\nu}{2}}} = \frac{A\sigma^{\nu} + B}{C\sigma^{\nu} + D} \qquad [6:11]$$

where $A, B, C, D$ are constants determined by the particular solution chosen.

Thus as $\sigma$ moves along the (indented) real axis, the path followed by $\zeta$ is obtained (to lowest order in $\sigma$) from that of $\sigma^{\nu}$ by a bilinear transformation. In view of the remark above it is clearly of interest to determine if this can be confined to the real axis.

Now if $0<\nu<1$ the path of $\sigma^{\nu}$ consists of two straight line segments joined at an angle of $\nu\pi$. If the bilinear transformation is non-degenerate the image of this path will also contain a kink which will prevent it coinciding with the real axis. Therefore $\zeta$ can take real values on both sides of the corotation point only if

$$AD = BC \qquad [6:12]$$

But if $f_0$ and $g_0$ are the values taken by $\zeta$ on the eigenvectors of $A_0$,

$$A = f_0 C \quad , \quad B = g_0 D \qquad [6:13]$$

Thus the condition that the bilinear transformation be degenerate becomes

$$CDf_0 = CDg_0 \qquad [6:14]$$

which is only possible in the (excluded) case $\nu = 0$ or when one of $C$, $D$ is zero (in which case all points map onto $f_0$ or $g_0$ which it is easily confirmed are real). In other words if the differential system is non-oscillatory at the corotation point there are two solutions which give real values of $\zeta$ on both sides of the corotation point and they are exactly the series solutions generated by the method of Frobenius. (For an alternative derivation of this result using the mean Reynolds stress of the perturbation see Miles (1961)).

But if the generalised Richardson Number exceeds $1/4$ at the corotation point, $\nu = i\mu$ is a pure imaginary and the behaviour of the differential system at this point is oscillatory. In this case the path followed by $\sigma^{\nu}$ consists of arcs of two disjoint circles, one of radius 1 and the other of radius $e^{\pm\mu\pi}$. It is again clear that no non-degenerate bilinear transformation can map this onto the real axis. However if the transformation is degenerate it maps everything onto $f_o$ or $g_o$ and these are now complex. We conclude that in this case it is impossible for $\zeta$ to take real values on both sides of the corotation point.

## 7 Structure at Corotation II

Let us again restrict ourselves to flows which have only one corotation point and where the boundary conditions require $\zeta$ to take real values. We know from the previous section that if singular neutral modes are to exist the differential system must be non-oscillatory at the corotation point and that if this condition is satisfied there are in general two solutions which take real values on both sides of the corotation point. If these solutions are integrated out to the boundary points they give two sets of real boundary values of $\zeta$ for which singular neutral modes certainly exist. Thus if the generalised Richardson number is less than 1/4 at any point it is certainly possible to have singular neutral modes. It is clear that any sharper stability criterion must depend on using some global property of the differential equation to relate values of $\zeta$ at the boundary points to those near corotation and will thus be more involved than the simple Richardson criterion.

As before we begin by studying the behaviour near corotation only this time we start with the Riccati equation [2:13] and concentrate on the case $0 \leqslant \mathcal{R}i \leqslant 1/4$. Abstracting the essential features we consider the equation

$$\frac{d\zeta}{d\sigma} = \frac{a}{\sigma}(\zeta - f)(\zeta - g) \qquad [7:1]$$

where $a, f, g$ are given real analytic functions of $\sigma$† (throughout this section a subcripted quantity denotes the appropriate coefficient in that quantity's expansion as a power series in $\sigma$, thus e.g.

$$a = a_0 + a_1\sigma + a_2\sigma^2 + \cdots \qquad [7:2]$$

).

If we now seek an analytic solution of [7:1], we find

$$n\zeta_n = \sum_{i+j+k=n} a_i(\zeta_j - f_j)(\zeta_k - g_k) \qquad [7:3]$$



† Since $Ri \leqslant \frac{1}{4}$ and $\sigma \neq 0$, see footnote on page 53.

and in particular, if $n = 0$,

$$0 = a_0(\zeta_0 - f_0)(\zeta_0 - g_0) \qquad [7:4]$$

Now $a_0$ cannot be zero (for if it were the singularity would be removable), therefore $\zeta_0$ must equal $f_0$ or $g_0$ and we may without loss of generality take

$$\zeta_0 = f_0 \qquad [7:5]$$

Then

$$\zeta_1 = \frac{-a_0(f_0 - g_0)}{1 - a_0(f_0 - g_0)} f_1 \qquad [7:6]$$

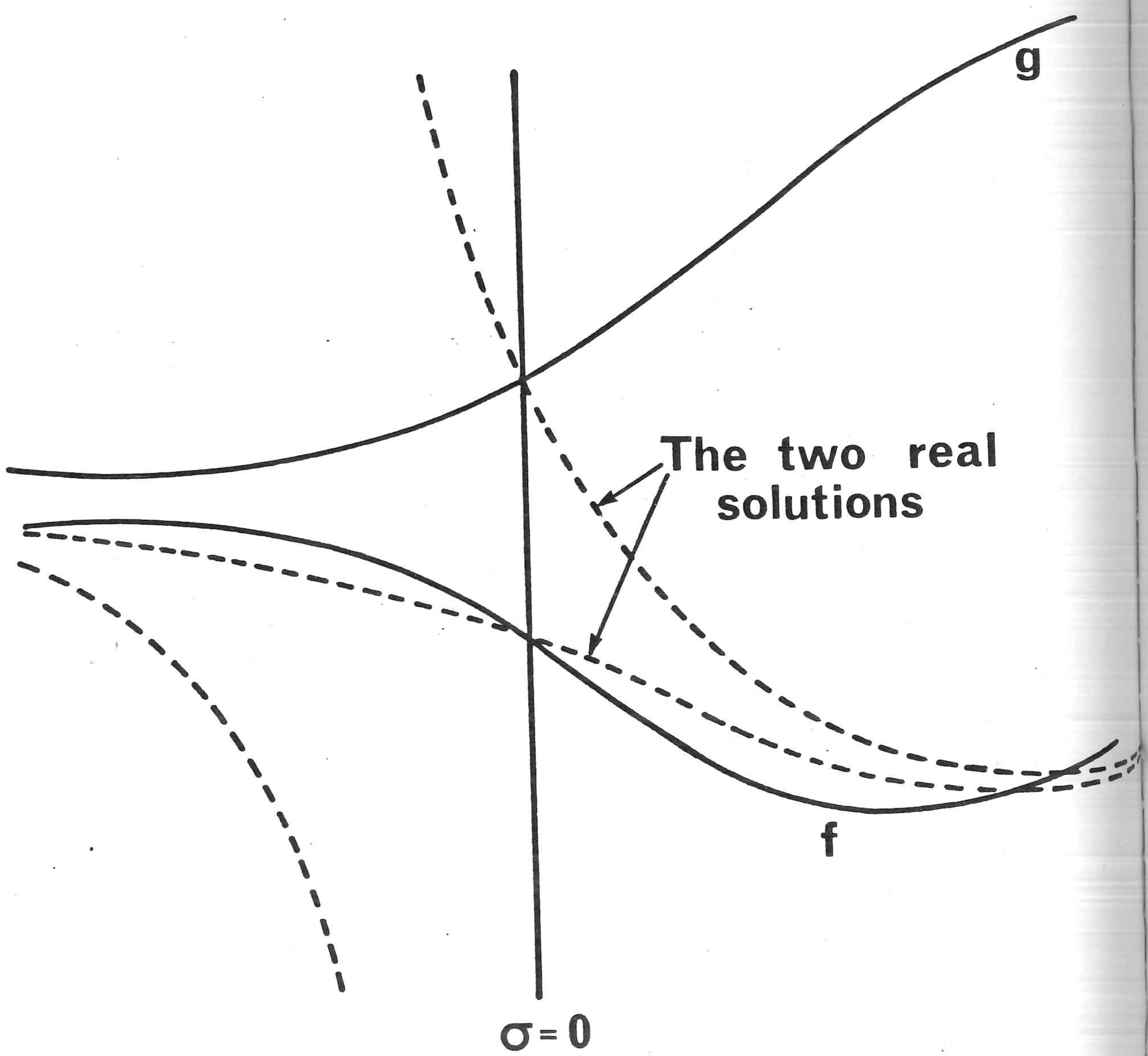
and the higher terms can be readily calculated provided $a_0(f_0 - g_0)$ is not a positive integer.

Either by direct comparison of [7:1] with [2:13], or by correspondence with the results of the previous section, we see that

$$|a_0(f_0 - g_0)| = \nu = \sqrt{1 - 4Ri} \qquad [7:7]$$

and that for $0 < \nu < 1$ there are two analytic solutions of [7:1], one passing through $g_0$ and the other through $f_0$ at

Fig.1 The two analytic solutions of equation 7:1 and their relation to the attractor $g$ and repellor $f$ in the neighbourhood of $\sigma = 0$.




Facing page 60



$\sigma = 0$. This confirms our previous results and in addition yields the slopes of the two solutions at the corotation point. The situation is best illustrated graphically (see fig.1). The vector field $\zeta'(\zeta, \sigma)$ vanishes on two curves $\zeta = f(\sigma)$ and $\zeta = g(\sigma)$. As one integrates towards the corotation radius $\sigma = 0$ one of these (say $g$) will be attractive and one ($f$) repulsive, the motion of $\zeta$ being away from $f(\sigma)$ and towards $g(\sigma)$; examination of [7:1] shows that this implies $a(f-g) < 0$. The expression [7:6] for the slopes of the analytic solutions at $\sigma = 0$ now shows that

$$\zeta_0 = f_0 \Rightarrow \zeta_1 = \frac{\nu}{1+\nu} f_1$$
$$\zeta_0 = g_0 \Rightarrow \zeta_1 = \frac{-\nu}{1-\nu} g_1 \qquad [7:8]$$

in other words (remembering $0<\nu<1$) the analytic solution which intersects the attractor at the critical point does so with slope opposed to that of the attractor, whereas that which intersects the repeller does so with a slope of the same sign. Further, in the limit $\nu \to 0$ the two solutions coincide and in the limit $\nu \to 1$ they coincide except in a neighbourhood of the corotation point unless the slope of the attractor ($g_1$) is zero.

To clarify the structure in the above limiting cases and verify that there are no other real solutions we now derive the complete solution of [7:1]. The procedure given above always enables us to find one analytic solution,

$$\tilde{\zeta} = \tilde{\zeta}_0 + \tilde{\zeta}_1 \sigma + \cdots \qquad [7:9]$$

with

$$\tilde{\zeta}_0 = f_0 , \quad \tilde{\zeta}_1 = \frac{\nu}{1+\nu} f_1 , \quad \ldots . \qquad [7:10]$$

If we now write

$$\zeta = \tilde{\zeta} + w^{-1} \qquad [7:11]$$

we find

$$\sigma \frac{dw}{d\sigma} + bw + a = 0 \qquad [7:12]$$

where

$$b = a(2\tilde{\zeta} - f - g) \qquad [7:13]$$

Now let

$$B = b_0 \ln \sigma + b_1 \sigma + \frac{b_2}{2} \sigma^2 + \cdots \qquad [7:14]$$

so that

$$\sigma \frac{d}{d\sigma} B = b \quad , \quad e^{B} = \sigma^{b_0}(1 + b_1\sigma + \cdots\cdots) \qquad [7:15]$$

then

$$w = Ce^{-B} - e^{-B}\int e^{B}\frac{a}{\sigma}\,d\sigma \qquad [7:16]$$

Now

$$b_0 = a_0(2\tilde{\zeta}_0 - f_0 - g_0) = a_0(f_0 - g_0) = -\nu \qquad [7:17]$$

so that if $0<\nu<1$ $w$ can only be real if $C$ is zero (the complementary function is complex) and we have indeed found all the real solutions. If $\nu = 0$

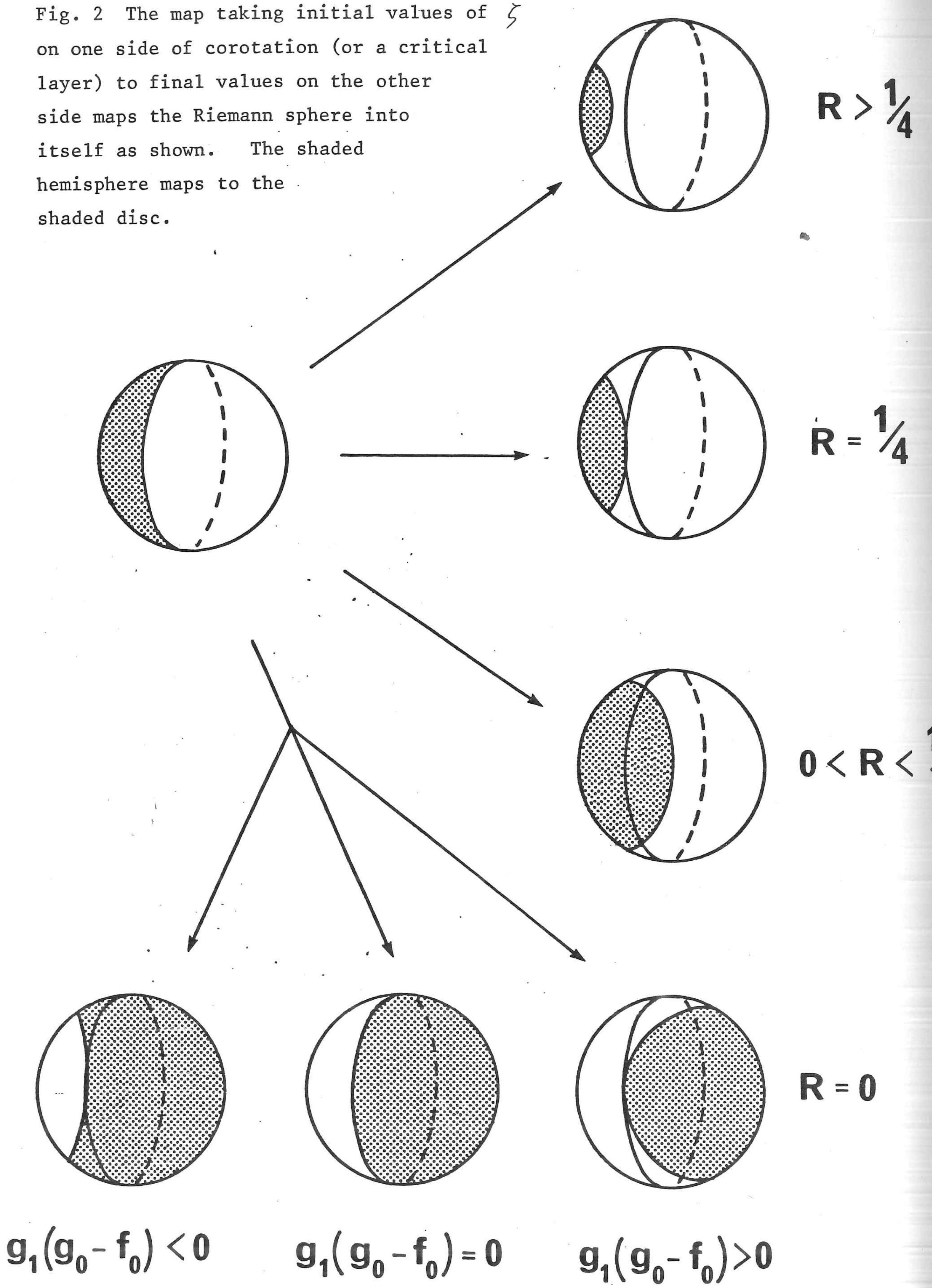
$$w = C(1 - b_1\sigma + \cdots) - (1 - b_1\sigma\cdots)(a_0 \ln\sigma + \cdots\cdots) \qquad [7:18]$$

is always complex and $\tilde{\zeta}$ is the only real solution. And if $\nu = 1$

$$w = C\sigma(1 - b_1\sigma + \cdots) - \sigma(1 - b_1\sigma + \cdots)(\cdots(a_0 b_1 + a_1)\ln\sigma + \cdots) \qquad [7:19]$$

is always complex if $a_0 b_1 + a_1 \neq 0$ implying that $\tilde{\zeta}$ is

Fig. 2 The map taking initial values of $\zeta$ on one side of corotation (or a critical layer) to final values on the other side maps the Riemann sphere into itself as shown. The shaded hemisphere maps to the shaded disc.




Facing page 63

See appendix C



the only real solution, but if $a_0 b_1 + a_1 = 0$, $\zeta$ is real for all real values of $C$ and we deduce that in this special case all solutions real on one side of corotation are real on the other side as well. Because

$$a_0 b_1 + a_1 = a_0[a_0(2\tilde{\zeta}_1 - f_1 - g_1) + a_1(2\tilde{\zeta}_0 - f_0 - g_0)] + a_1$$

$$= -a_0^2 g_1 \qquad [7:20]$$

the condition for this special case to occur is (in accord with the simple discussion of the limit $\nu \rightarrow 1$) that the slope of the attractor vanish at the corotation point.

Some of these results and those of the previous section can be given an interesting and useful interpretation if we consider the map which takes an initial value of $\zeta$ at a point on one side of the corotation point to the corresponding final value of $\zeta$ at a point on the other side. Because this map from $P_1(\mathbb{C})$ to $P_1(\mathbb{C})$ is induced by a linear endomorphism of $\mathbb{C}^2$ it must be a bilinear transformation. Thus the real axis will be mapped into some circle which can intersect the real axis of the target space in 0,1,2 or $\infty$ points. If we consider the above results and appeal to continuity, we see that the situation can be summarised by the following diagram (see fig.2).

However the important point in many applications is that if we know the form of the functions $f$ and $g$ (which is a simple algebraic problem) we can often use elementary methods of the qualitative theory of differential equations to sketch the solutions which are real on both sides of corotation and constrain their possible boundary values.

## 8 Double Neutral Modes

In the previous two sections I have shown that singular neutral modes can occur if $Ri \leqslant 1/4$ and have obtained certain information about them. In this section I examine the other possible limiting form of unstable mode, the double neutral mode, and attempt to obtain similar results.

A double neutral mode is a solution of the differential equation

$$\zeta' = A(r,\omega)\zeta^2 + B(r,\omega)\zeta + C(r,\omega) \qquad [8:1]$$

on the interval $[a, b]$ with suitable boundary conditions which is non-singular and has multiplicity two; thus it also satisfies the variational equation

$$\left(\frac{\partial \zeta}{\partial \omega}\right)' = \frac{\partial A}{\partial \omega}\zeta^2 + \frac{\partial B}{\partial \omega}\zeta + \frac{\partial C}{\partial \omega} + (2A\zeta + B)\frac{\partial \zeta}{\partial \omega} \qquad [8:2]$$

with

$$\frac{\partial \zeta}{\partial \omega} = 0 \qquad r = a, b \qquad [8:3]$$

(assuming the boundary conditions not to depend on the frequency; see section 10 for an extension to more general conditions) solving which we obtain as the necessary and sufficient condition for a mode to be a double neutral mode

$$0 = \int_a^b \left( \frac{\partial A}{\partial \omega} \zeta^2 + \frac{\partial B}{\partial \omega} \zeta + \frac{\partial C}{\partial \omega} \right) e^{-\int_a^r (2A\zeta + B)\,dr'} dr \qquad [8:4]$$

It follows that on a double neutral mode the expression

$$\frac{\partial A}{\partial \omega} \zeta^2 + \frac{\partial B}{\partial \omega} \zeta + \frac{\partial C}{\partial \omega} \qquad [8:5]$$

must change sign. Conversely if this expression is positive for some values of $r$ and $\zeta$ and negative for others there will exist, for suitable boundary points and values, solutions which are double neutral modes. The condition that the above expression change sign is precisely the condition that the vector field defined by the original equation should point into both the upper and the lower hemisphere when the frequency is made slightly complex (the geometrical content of this is simply that one has to be able to move away from and then back to the real axis); in other words it will be satisfied if the generalised Richardson criterion is violated.

We conclude that $Ri < 1/4$ is the necessary and sufficient condition for the existence of both singular neutral modes and double neutral modes† and that any double neutral mode must lie in part within the region

$$\frac{\partial A}{\partial \omega} \zeta^2 + \frac{\partial B}{\partial \omega} \zeta + \frac{\partial C}{\partial \omega} \leq 0 \qquad [8:6]$$

## 9 Application to Selfsimilar Flows

Let us now apply these ideas to the study of the self-similar flows

$$\Omega = r^{-\lambda} \qquad 0 < \lambda < 2 \qquad [9:1]$$

(removing a constant of proportionality by an appropriate choice of length and time scales). These are natural models to consider because, as well as simplifying the analysis, they give good representations of certain astrophysically important cases*. The case of uniform rotation is represented by $\lambda = 0$ and if $\lambda > 2$ the flows are centrifugally unstable. In many spiral galaxies the rotation law is well represented by $\lambda = 1$ over a large range in radius and for a Keplerian disc $\lambda = 1.5$. Let us also suppose that the basic flow is of uniform density and pressure. For boundary conditions we require $\zeta$ to be real

-----------------

* Except of course at the origin where they are singular; for a physically realistic model the rotation law should be modified inside some small core radius so that $\Omega$ is finite at the origin.



† if arbitrary real boundary values of $\zeta$ are allowed (see section 10)

at some small radius (see next section) and the perturbation to vanish as $r \to \infty$. This gives the simplest realistic model in which to study the effects of differential rotation. Although the equations are considerably simplified (both $g$ and $N$ vanish) they remain much more complicated than those of the plane parallel theory . However we can still apply the general methods developed in the previous sections (the application of these methods to plane parallel flows gives some interesting results which are summarised in appendix A).

With these simplifications equation [2:13] becomes

$$\zeta' = \frac{1}{r\sigma}\left\{\beta^2 r^2 - \frac{\sigma^2 r^2}{c^2} + (m - 2\Omega\zeta)(m - \omega_0\zeta) - \sigma^2\zeta^2\right\} \qquad [9:2]$$

but in fact for our purposes it turns out to be better to use not $\zeta$ but $\eta = \Omega\zeta$ ; this gives

$$\eta' = -\frac{\lambda}{r}\eta + \frac{\Omega}{r\sigma}\left\{\beta^2 r^2 - \frac{\sigma^2 r^2}{c^2} + (m - 2\eta)(m - (2-\lambda)\eta) - \frac{\sigma^2}{\Omega^2}\eta^2\right\} \qquad [9:3]$$

One advantage of studying these self-similar models is that because the basic flow has no associated length scale we can always rescale a singular neutral mode so that the corotation point falls at some given radius, say $r = 1$. Thus it is only necessary to study the functional form of the attractor and repeller with the corotation singularity at this one point.

These functions, $f$ and $g$, are the roots of the quadratic equation

$$\eta^2\left[2(2-\lambda)-\frac{\sigma^2}{\Omega^2}\right]+\eta\left[m(\lambda-4)-\frac{\lambda\sigma}{\Omega}\right]+\left[m^2+\beta^2r^2-\frac{\sigma^2r^2}{c^2}\right] \qquad [9:4]$$

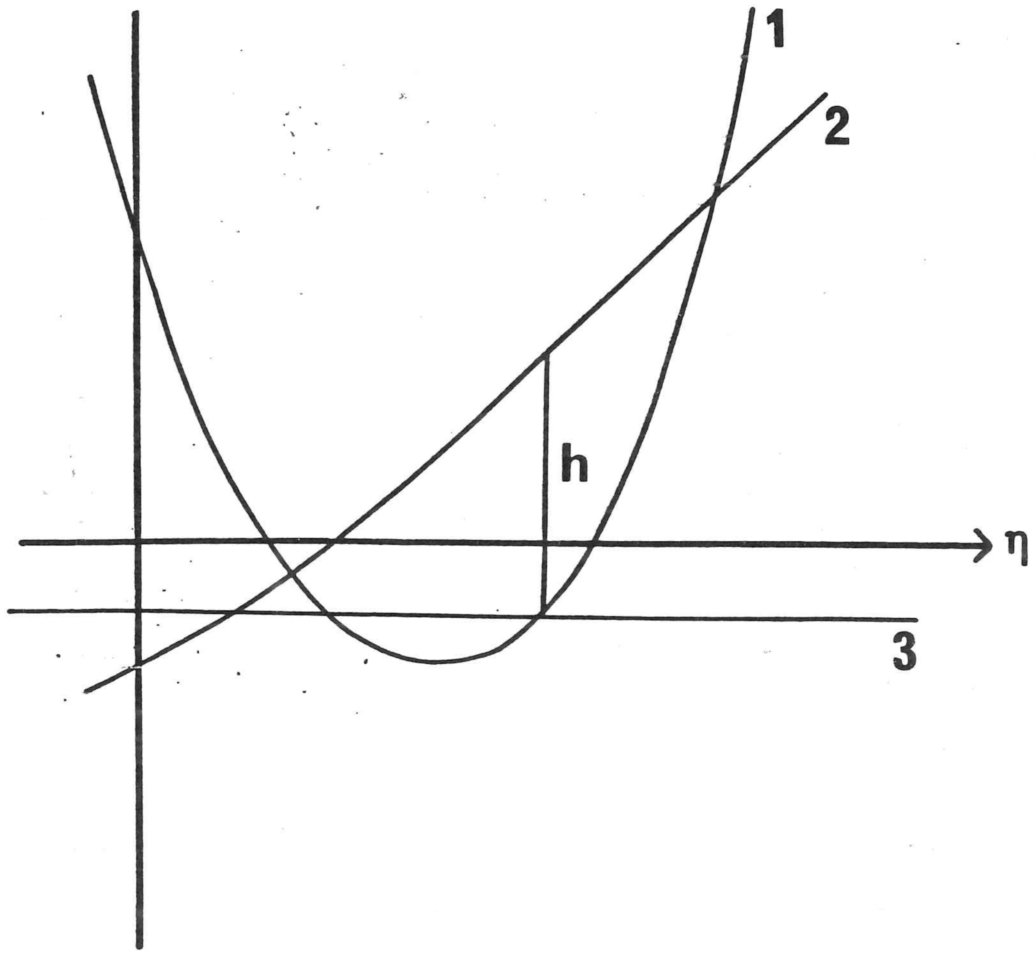
$$= \mathcal{A}\eta^2-\mathcal{B}\eta+\mathcal{C}=0$$

If the corotation point is at $r=1$, $\sigma/\Omega = m(r^{\lambda}-1)$ and is an increasing function. Thus for $r>1$ $\mathcal{A}$ is a decreasing function passing through zero at the outer Lindblad resonance and $\mathcal{B}$ is a positive increasing function. If $\beta c/\omega > 1$, i.e. if the system has exponential behaviour at large radii, $\mathcal{C}$ is positive for $r>1$. At $r=1$ both roots lie between $m/2$ and $m/(2-\lambda)$, $f$ being the lesser and $g$ the greater. At the outer Lindblad resonance $g=\infty$ and $f=\mathcal{C}/\mathcal{B}$. Beyond this point $g$ is negative. When the roots are real $f$ is positive.

<u>Theorem</u>

If $\beta c/\omega > 1$ (no sound waves at large radii) and the roots are real between corotation and the outer Lindblad resonance and in this region $g(r)\geqslant g(1)$ there are no

Fig. 3  The parabola 1 represents the (constant) left hand side of equation [9:6] and parabola 2 its (variable) right hand side. At corotation (when $\sigma = 0$) 2 degenerates into the horizontal line 3.








singular neutral modes which decay exponentially as $r \to \infty$ and for which $\eta$ is real as $r \to 0$.

Proof :-

Exponential decay as $r \to \infty$ requires $\eta \to g$ as $r \to \infty$ but under the hypotheses of the theorem both real solutions are bounded from above by

$$\psi(r) = \max_{1 \leq r' \leq r} \{ g(1) , f(r') \} \qquad [9:5]$$

and from below by 0 (the known form of the two real solutions near r=1 shows this to be initially true and no solution curve can cross 0 from above or $\psi$ from below). Thus both real solutions are positive as $r \to \infty$ whereas $g$ is negative.

---

It is easy to show that the hypotheses of the theorem are satisfied if $2 > \lambda \geq 1$; $g(r)$ is the larger of the two values of $\eta$ for which

$$(m - 2\eta)(m - (2-\lambda)\eta) = \frac{\lambda\sigma}{\Omega}\eta + \frac{\sigma^2 r^2}{c^2} - \beta^2 r^2 + \frac{\sigma^2}{\Omega^2}\eta^2 \qquad [9:6]$$

and it is sufficient to show (see fig.3)

$$h(r) = \frac{\lambda\sigma}{\Omega} g(1) + \frac{\sigma^2 r^2}{c^2} - \beta^2(r^2 - 1) + \frac{\sigma^2}{\Omega^2} g(1)^2 \geq 0 \quad (r \geq 1) \qquad [9:7]$$

or

$$h(r) \geqslant m\lambda(r^{\lambda}-1)g(1) + m^2(r^{\lambda}-1)^2 g(1)^2 - \beta^2(r^2-1) \geqslant 0 \qquad [9:8]$$

But

$$\beta^2 \leqslant \frac{m^2\lambda^2}{8(2-\lambda)}, \quad g(1) \geqslant m\frac{4-\lambda}{4(2-\lambda)}, \quad r^{\lambda} \geqslant r, \quad m \geqslant 1 \qquad [9:9]$$

$$(\lambda \geqslant 1,\ r \geqslant 1,\ Ri \leqslant \tfrac{1}{4})$$

$$\Rightarrow \quad h(r) \geqslant \frac{r-1}{4(2-\lambda)}\left[\frac{(4-\lambda)^2}{4(2-\lambda)} - \frac{\lambda^2}{2}\right]\left\{r-1 + \frac{\lambda(4-\lambda)-\lambda^2}{\left[\frac{(4-\lambda)^2}{4(2-\lambda)} - \frac{\lambda^2}{2}\right]}\right\} \qquad [9:10]$$

and as

$$1 \leqslant \lambda \leqslant 2 \quad \Rightarrow \quad (4-\lambda)^2 \geqslant 2\lambda^2(2-\lambda), \quad \lambda(4-\lambda) \geqslant \lambda^2 \qquad [9:11]$$

$$h(r) \geqslant 0 \qquad (r \geqslant 1) \qquad [9:12]$$

This is by no means the best possible result; numerical calculation of $g$ shows that it is a monotonic function between corotation and the outer Lindblad resonance for $2>\lambda>0.13$ and better results could be obtained with other forms of the equation. I conjecture that the result holds for $2>\lambda>0$; it seems improbable that a singular neutral mode exists in a flow so close to uniform rotation that $\lambda < 0.13$.

It remains to consider the possibility of double neutral modes. Using the result at the end of section 4 for some value of $r$ in the range of integration

$$|\sigma|^2 = |\omega - m\Omega|^2 \leqslant \Omega^2 \frac{m^2\lambda^2/4 - 2(2-\lambda)\beta^2 r^2}{m^2 + \beta^2 r^2} \qquad [9:13]$$

This confines the possible values of $\omega$ to a region of the complex plane which excludes the origin; thus a double neutral mode must corotate at a point inside the inner boundary and can not corotate at infinity (i.e. instability can not set in through a stationary perturbation). But we also know from section 8 that such a double neutral mode must cross the locus

$$(m - 2\eta)(m - (2-\lambda)\eta) = -\beta^2 r^2 - \frac{\sigma^2 r^2}{c^2} - \frac{\sigma^2}{\Omega^2}\eta^2 \qquad [9:14]$$

and it is easy to see that it is in consequence also trapped between $\psi$ and 0.

Thus there can exist neither singular neutral modes nor double neutral modes where $\eta$ is real at the inner boundary and represents an exponentially decaying perturbation at large radii. As the generalised Richardson criterion shows the flow to be stable for sufficiently large $\beta$ we deduce that the flows $\Omega = r^{-\lambda}$, $1 < \lambda < 2$ are stable to perturbations whose character is essentially incompressible.

But in an astrophysical context the sound speed will in general be quite low, perhaps of order a tenth of the azimuthal velocity. Therefore this result, while interesting, is largely irrelevant. If $c$ is small the differential system has oscillatory behaviour outside the region enclosed by the Lindblad resonances. This has certain interesting consequences. Let us suppose that the boundary conditions are

$$\eta = 0 \qquad r = a, b \tag{9:15}$$

Then there are many singular neutral modes; for a general value of $\beta$ we need only choose one of the two real solutions, integrate away from the corotation point until we enter an oscillatory region and then locate $a$ or $b$ at one of the many points where the solution passes through zero. Having found values of $a$ and $b$ for which a singular neutral mode exists we can now ask what happens when we slightly perturb one of the boundary points.

Let us define a function $\Delta$ as follows; starting with $\eta = 0$ at $a$ we integrate [9:3] along a suitable contour to $b$, the value of $\Delta$ is then the terminal value of $\eta$. Plainly $\Delta$ is a holomorphic function of $\omega, \beta, b$ and a mode corresponds to a zero of $\Delta$. The behaviour of a mode when perturbed can be deduced if we know the leading terms in the Taylor series expansion of $\Delta$; we have

$$\Delta \sim \frac{\partial\Delta}{\partial\omega}\delta\omega + \frac{\partial\Delta}{\partial b}\delta b + \frac{\partial\Delta}{\partial\beta}\delta\beta = 0 \qquad [9:16]$$

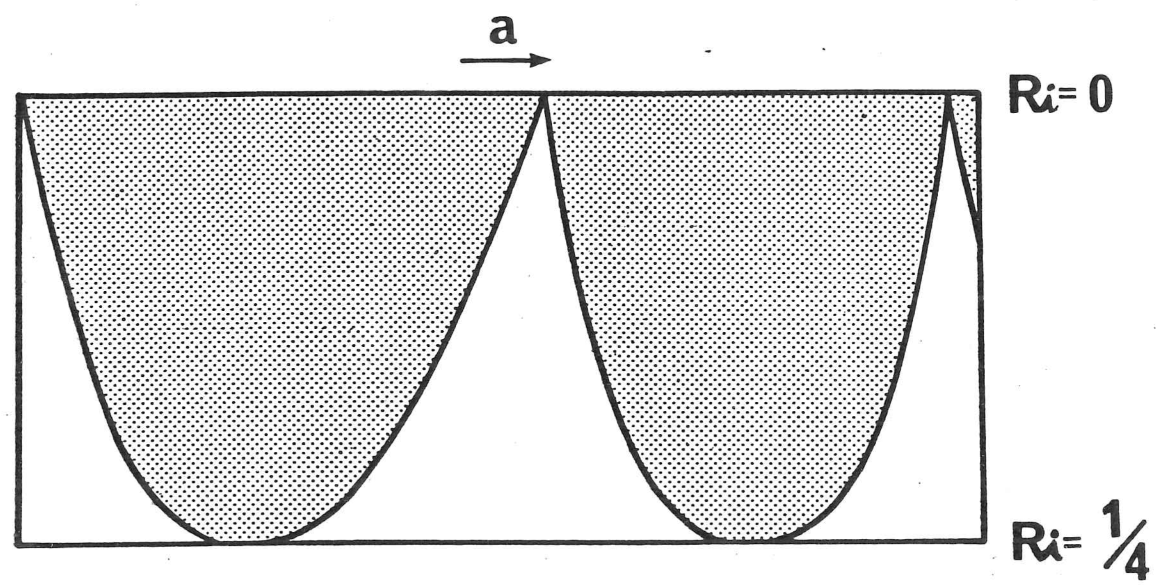
But $\partial\Delta/\partial b$ is the final value of $\eta'$, thus a negative, non-zero real number†. And $\partial\Delta/\partial\omega$ must be complex for almost all modes unless the Richardson number at corotation, $\mathcal{R}i$, is 0 or 1/4. To see this make a small real change in the frequency; this has two effects, firstly the corotation point and the two solutions which are real change slightly, secondly the 'wavelength' in the oscillatory regions is altered. If the inner boundary point $a$ is sufficiently far away the second effect dominates and the change in $\Delta$ is equivalent to that produced by a displacement of $a$. But this shows, using the results of 8, that $\partial\Delta/\partial\omega$ is non-zero and complex except when $\mathcal{R}i$ is 0 or 1/4 when it is real. Thus a positive or negative displacement of the outer boundary will produce a $\delta\omega$ with a negative imaginary part unless $\mathcal{R}i$ =0,1/4.

Now if we leave two degrees of freedom (e.g. by allowing both boundary points to move, or by letting the frequency change and one boundary point move) it is possible to follow a singular neutral mode as the Richardson number $\mathcal{R}i$ is varied from 0 to 1/4. Because each mode at $\mathcal{R}i$ =0,1/4 results from the confluence of two modes it is possible to follow one mode to $\mathcal{R}i$ =0 or 1/4 and return following another. By drawing a few diagrams one sees that by cycling



† See equation [9:3], b is located in an oscillatory region so that the quadratic on the RHS is positive or negative definite, in this case negative.

Fig. 4 The neutral curve and the regions of instability in the 'local Richardson number at corotation' against 'location of the inner boundary point' plane for one mode: note that this is not the stability boundary of the flow.




Facing page 74



*Ri* between 0 and 1/4 any two singular neutral modes which have the same number of nodes (points where $\eta$ = 0) between the corotation point and the outer boundary point can be connected to one another; each cycle changes the number of nodes between the inner boundary point and corotation by one (a pole and a zero of one mode come together at the corotation point when *Ri* = 0)†. Thus there is really only one such mode. If we imagine fixing the outer boundary point and then plotting against *Ri* the locations of the inner boundary points of all the singular neutral modes which can be connected as above the result will have the structure of fig.4. But this diagram can also be interpreted as showing the regions of stability and instability of the given mode; the boundaries between such regions must consist of points at which the mode is neutral and we have seen that except when *Ri* =0,1/4 the neutral modes do separate regions of stability and instability. Because the flow is stable if *Ri* > 1/4 we deduce that the regions of instability are those shaded in fig.4.

The remarkable conclusion is that when *Ri* = 0, i.e. the perturbations are two dimensional and there are no stabilising radial buoyancy forces, the flow is unstable for almost all locations of the boundary points provided these are sufficiently far apart to lie one on each side of the corotation point and in the oscillatory regions. We can also give some classification of the modes and make various



† See appendix C

predictions about their behaviour under perturbations. These have been confirmed by some preliminary numerical work.

Although this argument has been presented for one specific case it is clear that the result is quite general and that we have effectively proved the following result

> A flow is unstable if the boundary conditions are time-symmetric (see below) and the differential equation, for some frequency, has a critical layer with local Richardson number 0 † surrounded by two oscillatory regions containing the boundary points.

## 10 Boundary Conditions

These results have been obtained on the assumption that the boundary conditions require $\eta$ (or $\zeta$) to take real values and it is reasonable to ask if there is any physical reason for distinguishing such boundary conditions. Examining the definition of $\zeta$ we see that if it is real the radial velocity and pressure perturbations are in quadrature; thus the boundary can do no work on, nor extract any energy from, the system. This explains the importance of such conditions in the theory of singular neutral modes; for such modes any physical boundary condition which does not incorporate an 'arrow of time' must be of this form. These time-symmetric boundary conditions occur frequently;



† Indeed for some boundary locations there exist unstable modes for $0 \leq Ri < \frac{1}{4}$.

they include the rigid and free boundary conditions of classical hydrodynamics as well as conditions of exponential decay in some limit or regularity at some point. The one important case not included is that of a radiation boundary condition; clearly a condition of 'no incoming wave energy' is not time-symmetric. This is unfortunate because such boundary conditions are often the most natural ones to impose. In particular for compressible disc models the natural boundary conditions are that the solution be regular at the origin and have no energy flux coming in from infinity.

The above argument explains why in many cases the boundary values of $\zeta$ should be real for neutral modes. However in general this value will depend on the frequency (as a concrete example consider the boundary conditions appropriate to a membrane of given density and elasticity) so that the value of $\zeta$ will be complex for growing modes. The boundary conditions can be regarded as descriptions of the coupling between the perturbations of the fluid system and the hidden internal modes of some other mechanical system (for example the vibrations of the membrane). For most natural systems if the amplitude of the fluid perturbation increases these hidden modes will grow by absorbing energy from the fluid perturbation;† indeed if the boundary were to feed energy into a growing mode the resulting instability would be essentially one of the hidden system



† i.e. the hidden modes of the boundary are assumed to have positive energy, of course the fluid modes may have either positive or negative energy.

rather than of the fluid system (which would now be more naturally regarded as imposing boundary conditions on the other; a situation where this is relevant is the analysis of a model formed by patching together two flows, for example a differentially rotating disc with the centre replaced by a uniformly rotating core). Thus most boundary conditions should absorb energy from growing modes; this determines the sign of the imaginary part of $\zeta$ (which is also that appropriate to a radiation boundary condition).

In section 4 I showed that if the frequency has a negative imaginary part (signifying a growing mode) then it is easy for $\zeta$ to cross the real axis from one hemisphere to the other in one direction, but difficult (and impossible if at all points $Ri > 1/4$) to go in the reverse direction. It is easy to guess (and to verify) that the solutions crossing in the easy direction correspond to increasing perturbations driven by an influx of energy through both boundaries. Thus if the boundary conditions are passive and do not feed energy to the perturbation the boundary values are either real or lie in the hemispheres which it is difficult to join. It follows that with these very general boundary conditions the generalised Richardson criterion provides a sufficient condition for stability and that any double neutral node must still have some points in the region [8:6]. Of course necessity no longer holds; it is intuitively reasonable that allowing energy to leak out of the system should make it more stable.†

While it is not so easy to extend the other results some general conclusions can be drawn. There are two critical parameters which together determine what happens at the corotation point, the generalised Richardson number $\mathcal{R}i$ and the quantity $\mathcal{G} = g_1(g_0 - f_0)$. In the two wave regions we can think of the solutions as waves carrying energy towards and away from the corotation point and define approximate reflection and transmission coefficients (some care is needed here because when $\sigma$ is negative the group and phase velocities of the waves have opposite signs).

If $\mathcal{R}i = 0$ and $\mathcal{G} \geqslant 0$ (as is the case in the self-similar disc models) my results can be interpreted as showing that over-reflection occurs for wave packets incident from outside, i.e. the amplitude of the reflected packet exceeds that of the incident. If $\mathcal{R}i = 0$ and $\mathcal{G} \leqslant 0$ over-reflection occurs for packets incident from inside. Thus the corotation point can act as an amplifier of acoustic waves; and if we introduce sufficient positive feedback by reflecting some of the amplified waves back it can act as an oscillator. This offers a physical explanation of the results of section 9 and indicates that at least one of the boundary conditions can be relaxed to a radiation boundary condition without changing the conclusion.

The phenomenon of over-reflection was first discovered by Miles (1957) and Ribner (1957) while studying



† But only in this weak sense; there can exist 'negative energy modes' which only grow if the boundaries absorb energy (e.g. the solutions of section 11).

the interaction of sound waves with a vortex sheet; it has since been demonstrated for internal gravity waves by Jones (1968), for Rossby waves by Dickinson and Clare (1973) and for magneto-acoustic waves by Fejer (1963). An excellent account is that given by Acheson (1976) which contains an extensive bibliography and a lucid account of the 'negative energy' aspect.

Physically there are two effects involved; firstly the presence of waves inside corotation reduces the total energy of the system whereas outside they increase it, secondly they can cause permanent changes of order the wave amplitude squared in the mean flow near the radius of corotation (the generalised Charney-Drazin theorem, (Andrews and McIntyre (1978)), shows that the waves only induce transient mean flow changes at other radii). In the absence of the second effect energy conservation shows that over-reflection must occur. However in general the critical layer absorption or emission is the dominant effect; this is controlled by the parameters $\mathcal{R}i$ and $\mathcal{G}$.

## 11 Exact Solutions

An estimate of the magnitude of the first effect and a confirmation of the general theory is provided by certain exact solutions. As noted in section 2 the differential system [2:10] has singular points at zero, infinity and the corotation points where $\sigma = 0$. However if at a corotation point $R_i = 0$ and $y = 0$ that singularity is removable; this condition requires that either

$$p_0' = 0 \quad \& \quad (\rho_0^{+1}\omega_0)' = 0 \qquad [11:1]$$

or

$$p_0' = c^2\rho_0' \quad \& \quad (\rho_0^{-1}\omega_0)' = 0 \qquad [11:2]$$

If [11:1] or [11:2] holds globally the corotation singularities are always removable and the only singular points are zero and infinity. In general these are irregular but there are two families of self-similar models for which zero is a regular singular point: in these cases the general solution to the perturbation equations can be found in terms of Whittaker functions.

<u>Case I</u>

If $\Omega = r^{-\lambda}$, $\rho_0 = r^{-\lambda}$ and $c = c_0 r^{1-\lambda}$ then the general solution is ( $x \equiv r^{\lambda}$ )

$$u = \frac{x}{r}\left\{\lambda^2(2-\lambda)\, m \frac{dw}{dx} - \frac{\lambda(2-\lambda)^2}{c_0^2}\,\sigma w\right\}$$

$$p = -\frac{i}{x}\left\{\lambda^2(2-\lambda)^2 \frac{dw}{dx} + \lambda(2-\lambda)\, m \sigma w\right\} \qquad [11:3]$$

where $w$ is any solution of the confluent hypergeometric equation

$$\frac{d^2 w}{dx^2} + \frac{w}{\lambda^2}\left\{\frac{\sigma^2}{C_0^2} - \frac{m^2 + 2(2-\lambda)/C_0^2}{x^2}\right\} \qquad [11:4]$$

<u>Case II</u>

If $\Omega = r^{-\lambda}$, $\rho_0 = r^{+\lambda}$ and $c = c_0 r^{1-\lambda}$ the general solution is

$$u = \frac{1}{r}\left\{\lambda^2(2-\lambda)\, m \frac{dw}{dx} - \frac{\lambda(2-\lambda)^2}{C_0^2}\sigma w\right\}$$

$$[11:5]$$

$$p = -i\left\{\lambda^2(2-\lambda)^2 \frac{dw}{dx} + \lambda(2-\lambda)\, m\, \sigma w\right\}$$

where $w$ and $x$ are as in case I.

Either by a simple WKBJ approximation or by consideration of the known asymptotic forms of Whittaker functions these solutions can be used to estimate the ratio of the amplitude of the reflected to that of the incident wave packet: both methods give results of order

$$1 + e^{-\frac{\pi}{\lambda}\left[m c_0\left(1+\frac{\lambda^2}{4m^2}\right) + \frac{2(2-\lambda)}{m c_0}\right]} \qquad [11:6]$$

so that maximum over-reflection occurs when

$$m^2 c_0^2 \quad = \quad 2(2-\lambda) + \frac{\lambda^2 c^2}{4} \qquad [11:7]$$

and gives an amplification of

$$e^{-\frac{2\pi}{\lambda}\sqrt{2(2-\lambda)+\lambda^2 c^2/4}} \qquad [11:8]$$

The effect vanishes as the discs tend towards uniform rotation and becomes very efficient as they tend towards irrotational ( $\lambda$ =2) discs; for a keplerian disc the maximum effect is about 0.015

This is quite small. However the main reason for this is the width of the 'barrier' through which the perturbation has to tunnel and if this could be reduced by making the disc slightly self-gravitating the effect might be considerably increased; also when the second process operates and energy is extracted from the mean flow at the corotation radius, because the perturbation only has to tunnel half the distance, an order of magnitude estimate is that the amplification will be about the square root of that due to the first process. This suggests that amplifications of a few percent are quite possible in Keplerian discs.

## 12 Conclusions

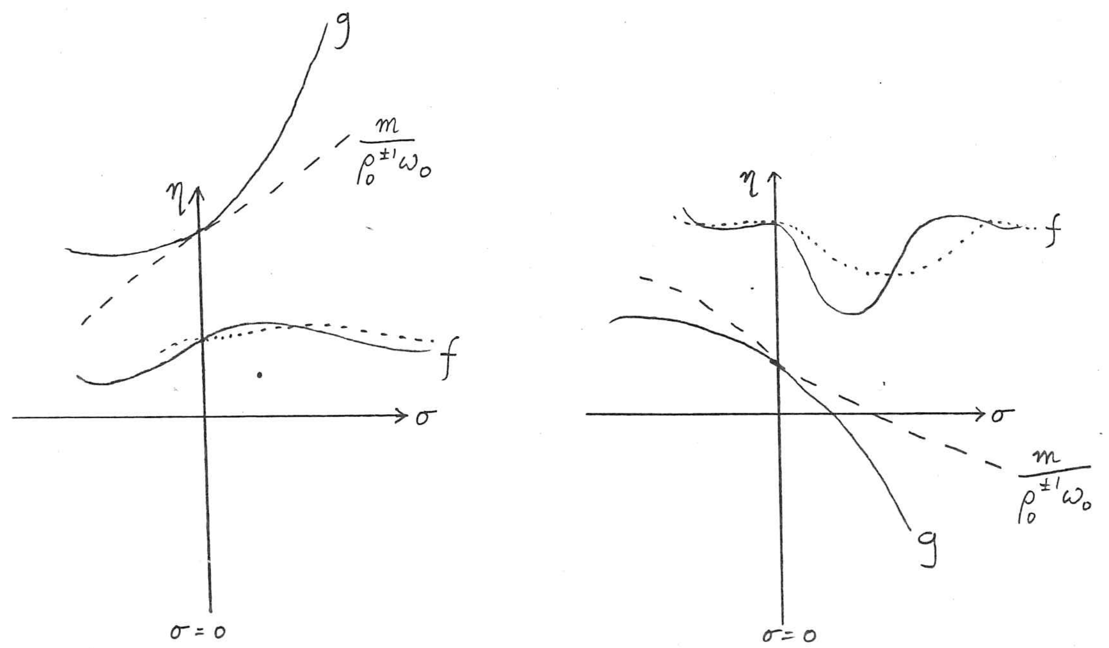
In compressible discs sound waves can be reflected with an amplification of a few percent from regions where the local Richardson number is zero, but in general only from one side. If this side faces the centre of the disc the amplified wave will be reflected back and an instability will result. It is probable that self-gravity of the disc enhances this effect*. The resulting unstable mode probably grows until the non-linear interaction between itself and other modes can feed energy into modes of lower amplitude at a rate equal to that at which it is absorbing energy from the mean flow. The prospect arises of an accretion disc which is not turbulent in the usual sense, but in which the angular momentum is transported by a stochastic spectrum of such weakly non-linear sound waves, rather as in the oceans horizontal momentum is transported vertically by internal gravity waves (Müller (1976)).

In the case of subsonic (or incompressible)

-----------------

* Actually that part of the analysis which depends only on the two-dimensionality and linearity of the problem can be carried out even for a self-gravitiating disc. A highly non-linear integro-differential equation can be found for a projective variable and if the boundary conditions require this variable to be real there are still only 0,1,2 or $\infty$ solutions with this property. However the non-local nature of the relation between the potential and the density, reflected in the fact that the equation is integro-differential rather than differential, appears to prevent any further progress. I intend to investigate the consequences of using a local approximation for the potential (as in spiral density wave theory).

Fig. 5 By the substitution $\eta = \rho_0^{\pm}\zeta$ one obtains an equation in which the attractor $g$ is tangent to the curve $\eta = m/(\rho_0^{\pm1}\omega_0)$ at corotation and thus can not have a vanishing slope unless there is an extremum in $\rho_0^{\pm1}\omega_0$ at that radius. But when the slope of the attractor is non-zero the only possible singular neutral mode (shown dotted) passes through $f$ at corotation. When the flow is subsonic $f$ is bounded away from zero; the required result follows from the fact that $f$ and $g$ are seperated by a monotonic function (namely $m/(\rho_0^{\pm1}\omega_0)$ ).








axisymmetric systems it is easy to show that if the perturbations are two-dimensional and the boundaries are rigid concentric cylinders a necessary condition for the existence of a singular neutral mode is an extremum in $\omega_0/\rho_0$ if the flow is isentropic and an extremum in $\omega_0\rho_0$ if it is isobaric. The proof is virtually identical to that of the corresponding result in the plane parallel case (see appendix A) and should be evident from fig. 5. One can also show that if there is a singular neutral mode in such a flow and it only corotates at one radius, then the extremum must be a maximum in $|\omega_0\rho_0^{\pm1} - 2\omega/m|$ at the corotation radius. In the absence of a semicircle theorem these results do not yield simple stability results, but I think it very probable that as in the case of plane parallel flows the necessary conditions for the existence of singular neutral modes are also necessary conditions for instability. In general dynamical instability in the subsonic case appears to require a fairly sharp maximum in the vorticity (so that the Kelvin-Helmholtz instability could be said to be the generic dynamical instability of subsonic flows). However it seems to be impossible to obtain simple general results for cases where the local Richardson number does not vanish identically. In the study of specific models, as shown by the example of section 2:10, my methods can be quite powerful; that example shows that, at least in a uniform density and pressure subsonic disc with a rotation law of the

Keplerian type, allowing the perturbations to have a three-dimensional structure has no destabilising effect.

I end with a summary of the results established in this chapter.

(1) A sufficient condition for the stability of an axisymmetric flow with passive boundary conditions is that the local Richardson number (defined by equation 4:7) should everywhere exceed 1/4.

(2) Without further restricting the boundary conditions no stronger result can be obtained.

(3) There is a geometric technique which can be used to describe the neutral modes (especially the singular neutral modes).

(4) The flows $\Omega = r^{-\lambda}$ $(1 \leq \lambda < 2)$ with uniform density and pressure, a passive inner boundary and an outer boundary condition of exponential decay as $r \to \infty$ are stable to perturbations which do not represent sound waves at large radii.

(5) These flows, when sufficiently compressible to support sound waves and with sufficiently separated reflecting boundaries, can be unstable if $0 < Ri < 1/4$ and are unstable if $Ri = 0$ (except perhaps for a discrete set of boundary locations).

(6) There exist two families of self-similar flows for which the solutions of the perturbation equations can be expressed in terms of Whittaker functions.

## 13 References

Acheson,D.J. 1976 J.F.M. 77,433.

Andrews,D.G. & McIntyre,M.E. 1978 J.Atmos.Sci. 35,175.

Arnol'd,V.I. 1966 J.Mécan. 5,29.

Bath,G.T. 1972 Ap.J. 173,121.

Chandrasekhar,S. 1960 J.Ind.Math.Soc. 24,211.

Chandrasekhar,S. 1961 Hydrodynamic & Hydromagnetic Stability,Oxford.

Chimonas,G. 1979 J.F.M. 90,1.

Cowling,T.G. 1951 Ap.J. 114,272.

Dickinson,R.E. & Clare,F.J. 1973 J.Atmos.Sci. 30,1035.

Drazin,P.G. & Howard,L.N. 1966 Adv.Appl.Mech. 9,1.

Eardley,D.M. & Lightman,A.P. 1975 Ap.J. 200,187.

Eckart,C. 1963 Phys.Fl. 6,1042.

Fricke,K.J. 1968 Z.Ap. 68,317.

Fricke,K.J. & Kippenhahn,R. 1972 Ann.Rev.Astr.&Ap. 10,45.

Goldreich,P. & Schubert,G. 1967 Ap.J. 150,571.

Hazelhurst,J. 1963 Ap.J. 137,125.

Helmholtz,H. 1868(translation) Phil.Mag. 36,337.

Hille,E. 1969 Lectures on Ordinary Differential Equations, Addison-Wesley.

Howard,L.N. 1961 J.F.M. 10,509.

Jones,W.L. 1968 J.F.M. 34,609.

Kelvin,W.1871 Phil.Mag. 42,368.

Kelvin,W. 1880 Math.&Phys. Papers, vol.4,p.172; Cambridge 1910.

Lalas,D.P. 1975 J.F.M. 69,65.

Lin,C.C. 1945a Quart.Appl.Math. 3,117.

Lin,C.C. 1945b Quart.Appl.Math. 3,218.

Lin,C.C. 1946 Quart.Appl.Math. 4,278.

Lin,C.C. 1961 J.F.M. 10,430.

Lust,R. 1952 Z.Naturforschung 7A,87.

Lynden-Bell,D. & Ostriker,J.P.1967 M.N. 136,293.

Lynden-Bell,D. & Pringle,J.E. 1974 M.N. 168,603.

Meksyn,D. & Stuart,J.T.1951 Proc.Roy.Soc.Lond.A 208,517.

Miles,J.W. 1957 J.Acoust.Soc.Am. 29,226.

Miles,J.W. 1961 J.F.M. 10,496.

Müller,P. 1976 J.F.M. 77,789.

Pringle,J.E. & Rees,M.J.1972 Astr.&Ap. 21,1.

Rathy,R.K. & Chandra,K. 1972 The Math.Stud. 40,129.

Rayleigh,Lord 1880 Proc.Lond.Math.Soc. 11,57.

Rayleigh,Lord 1916 Proc.Roy.Soc.Lond.A 93,148.

Ribner,H.S. 1957 J.Acoust.Soc.Am. 29,435.

Spiegel,E.A. & Zahn,J-P. 1970 Comments Ap.Space Sci. 2,178.

Stewart,J.M. 1975 Astr.&Ap. 42,95.

Stewart,J.M. 1976 Astr.&Ap. 49,39.

Sung,C.-H. 1974a Astr.&Ap. 33,99.

Sung,C.-H. 197~~4b~~5 Astr.&Ap. 33,127.

Shakura,N.I. & Sunyaev,R.A. 1973 Astr.&Ap. 24,337.

Warren, F.W. 1976 Proc.Roy.Soc.Lond.A 350,213.

Weizsacker, C.F. von 1951 Ap.J. 114,165.

## 1 Introduction

While working on the analytic theory described in the last chapter I considered the possibility of numerically integrating the linearized perturbation equations with the viscous terms retained. This required the solution of a problem of the Orr-Sommerfeld type, a boundary value problem in which the differential system has solutions with widely differing growth rates (corresponding to the presence of two scales, the viscous and the dynamical, in the problem). Such problems are notoriously hard to solve numerically (for a general introduction to the earlier literature see the review by J.M.Gersting and D.F.Jankowski (1972)). The methods used can be divided into two categories. Firstly there are those which seek to determine the entire solution at once, either by using matrix methods to solve a finite difference scheme or by determining the coefficients in some expansion of the solution. Secondly there are the shooting methods which attempt to solve the boundary value problem by initial value methods. The methods in the second category have the advantage of simplicity and small storage requirements although on general grounds the methods in the first category should be slightly more efficient. I analysed these shooting methods and found a unifying geometrical interpretation of the principal methods, those of orthonormalisation and invariant imbedding, which indicates a slight superiority for the method of orthonormalisation in difficult cases.

## 2 The Problem

The generic problem consists of a system of ordinary differential equations of order $n$, linear and depending analytically on some parameter $p$ (if there is more than one parameter $p$ should be interpreted as a vector);

$$\frac{d}{dx} y = F(x, p)\, y$$

$$y : x \longmapsto y(x) \in V \qquad [2:1]$$

$$F : (x, p) \longmapsto F(x, p) \in \mathcal{L}(V, V)$$

$$x \in [x_a, x_b]$$

where $V$ is a vector space of dimension $n$ (we will normally consider $V$ as a vector space over $\mathbb{C}$, but the application to real vector spaces is obvious) and $[x_a, x_b]$ is a closed interval of the real line. Integration of this system defines a flow on $V$, i.e. a map

$$R(x_1, x_2) : V \to V \qquad [2:2]$$

such that

$$R(x_1, x_2)\, y(x_1) = y(x_2) \qquad [2:3]$$

Because the system is linear the map $R$ is a linear map.

The system is subject to $k$ independent linear homogeneous boundary conditions at $x_a$ and to $k' = n-k$ at $x_b$. That is there exist linearly independent subsets

$$\{\alpha_i\}_{i=1\cdots k} \qquad \{\beta_i\}_{i=1\cdots k'} \qquad [2:4]$$

of the dual space $V'$ and the boundary conditions are

$$\alpha_i(y(x_a)) = \beta_j(y(x_b)) = 0 \qquad [2:5]$$
$$i = 1\cdots k \qquad j = 1\cdots k'$$

It is convenient to introduce bases for the annihilators of the subspaces spanned by these sets, i.e. linearly independent sets of vectors in $V$

$$\{a_i\}_{i=1\cdots k'} \qquad \{b_i\}_{i=1\cdots k} \qquad [2:6]$$

such that

$$\alpha_j(a_i) = 0 = \beta_i(b_j) \qquad [2:7]$$
$$i=1\cdots k' \qquad j = 1\cdots k$$

Then any solution of the system satisfying both sets of boundary conditions will have for its value at $x_a$ a linear combination of the $a_i$ and at $x_b$ a linear combination of the $b_i$.

## 3 Simple Shooting

This observation forms the basis of the simple shooting method; if the value of the solution at $x_a$ is a linear combination of the $a_i$, then because of the linearity of $R$ its value at $x_b$ will be the same linear combination of the vectors

$$\{ R(x_a, x_b)\, a_i \}_{i=1 \cdots k'} \qquad [3:1]$$

and this will be expressible as a linear combination of the $b_i$ iff the exterior product (essentially a determinant)

$$D = R(x_a, x_b)\, a_1 \wedge \ldots\ldots \wedge R(x_a, x_b) a_{k'} \wedge b_1 \wedge \ldots \wedge b_k \qquad [3:2]$$

is zero. The simple shooting method consists of choosing $k'$ sets of initial values satisfying the boundary conditions at $x_a$ (the $a_i$), advancing them to $x_b$ by any initial value method and then evaluating the determinant $D$. If the differential system is non-singular in the range $[x_a, x_b]$ $D$ is a holomorphic function of the parameter $p$ and its zeros may be found by any of the standard methods.

This simple scheme breaks down when applied to problems of the Orr-Sommerfeld type. Although the initial vectors $a_i$ are linearly independent, the final vectors may well be nearly dependent because the growth of some components of the solution is so extreme as to swamp the less rapidly growing components. A partial cure is to match at some point $x_m$ in the middle of the range and set

$$\mathcal{D} = R(x_a, x_m)a_1 \wedge \dots \wedge R(x_a, x_m)a_{k'} \wedge R(x_b, x_m)b_1 \wedge \dots \\ \dots \wedge R(x_b, x_m)b_k \, . \qquad [3:3]$$

## 4 The Method of Orthonormalisation

A much better solution is the method of orthonormalisation (see e.g. Conte (1966))[†]. In this the range of integration is divided into a number of subranges. As in the simple shooting method a set of $k'$ vectors is advanced from $x_a$ to $x_b$ by integration across each subrange. However at the end of each subrange the $k'$ vectors are orthonormalised (say by the Gram-Schmidt process) with respect to some arbitrary inner product on $V$. This ensures that the vectors remain independent and that the growth of one solution does not swamp the others (or cause overflow problems! ). As before a determinant $\mathcal{D}$ can be formed, either at $x_b$ or at some internal matching point $x_m$, the zeros of which identify non-trivial solutions of the boundary value problem. It may not however be a holomorphic function of the parameter; in particular if the quadratic form used to define the inner product is a positive definite Hermitean one it will not be holomorphic because of the occurence of complex conjugate quantities in the expression



[†] see also Kaplan's method as described in Betchov and Criminale, Stability of Parallel Flows, Academic Press (1967), p83-91.

for the inner product. For this reason the use of an orthogonal form (even though it is indefinite) has been recommended by some authors (starting with Gary and Helgason (1970)). It is certainly true that the determinant found in this way will be locally holomorphic, but because of the different branches of the square root which may be taken when normalising the region of holomorphy tends to be very small*. Thus it is impossible to use global zero finding methods based on the principle of the argument.

By contrast if the inner product is based on a Hermitian form, not only do we have the security of a positive definite inner product, but it is also easy to see that the determinant formed is the product of the determinant that we would have obtained using the simple shooting method with exact arithmetic and a positive real function. Thus the determinant is simply a rescaled version of the determinant produced by the simple method, and indeed if some information is retained from the orthonormalisations it is easy to calculate the scaling. But this is not necessary if we wish to apply the principle of the argument, for the arguments of the two determinants agree and this very powerful and useful method can be applied directly (Appendix B describes a FORTRAN package for just such an application).

-----------------

* In fact it is not hard to see from Jensen's formula that no method which controls the growth can return a globally holomorphic determinant.

## 5 Invariant Imbedding

Consideration of this method shows clearly that the object of importance is not the particular set of vectors being advanced, but the linear subspace that they span. This suggests that it would be better to formulate the problem and its solution in as basis independent a way as possible.

The differential system induces a flow, not only on the space $V$ , but also on all the geometric objects associated with $V$ . In particular it induces a flow on the Grassmanians $\mathcal{G}_k(V)$ ( $\mathcal{G}_k(V)$ is defined as the set of $k$-dimensional linear subspaces of $V$ ; for example $\mathcal{G}_2(\mathbb{R}^3)$ is the set of all planes through the origin of $\mathbb{R}^3$ ). The boundary conditions can be stated as

$$y(x_a) \in \mathcal{A} = \mathrm{Span}(a_i) \in \mathcal{G}_{k'}(V)$$
$$y(x_b) \in \mathcal{B} = \mathrm{Span}(b_j) \in \mathcal{G}_k(V) \qquad [5:1]$$

so that if the induced flow on $\mathcal{G}_k(V)$ is denoted $R_k$ the condition for a non-trivial solution is

$$R_{k'}(x_a, x_b)(\mathcal{A}) \cap \mathcal{B} \neq 0 \qquad [5:2]$$

or if matching at some intermediate point,

$$R_{k'}(x_a, x_m)(\mathcal{A}) \cap R_k(x_b, x_m)(\mathcal{B}) \neq 0 \qquad [5:3]$$

Clearly this is a minimal formulation of the problem, all arbitrary factors having been removed. Actually this minimality is not entirely desirable. The additional information retained in the method of orthonormalisations (the orientation of the basis defining the subspace) greatly assists the location of solutions by enabling one to use the principle of the argument.

For purposes of calculation this method can only be used if we have some representation of elements of $\mathcal{G}_k(V)$ and of the flow $R_k$. But this is not difficult. $\mathcal{G}_k(V)$ is a manifold (of dimension $k(n-k)=kk'$) and there exist certain natural parametrizations of $\mathcal{G}_k(V)$ by linear maps which for purposes of calculation may be identified with matrices.

These natural parametrizations may be defined as follows. Let $\mathcal{U}_1 \in \mathcal{G}_k(V)$ and let $\mathcal{U}_2$ be a subspace complementary to $\mathcal{U}_1$ in $V$, i.e.

$$V = \mathcal{U}_1 \oplus \mathcal{U}_2 \qquad [5:4]$$

$$\dim(\mathcal{U}_1) = k \qquad \dim(\mathcal{U}_2) = k'$$

Now let $\alpha \in \mathcal{L}(\mathcal{U}_1, \mathcal{U}_2)$ and consider the map

$$\alpha \mapsto \mathrm{graph}(\alpha) = \{ u \oplus \alpha u : u \in \mathcal{U}_1 \} \subset V \qquad [5:5]$$

This is easily seen to be a parametrization of an open neighbourhood of $\mathcal{U}_1$ in $\mathcal{G}_k(V)$ by elements of $\mathcal{L}(\mathcal{U}_1, \mathcal{U}_2)$ which can be identified (though non-canonically) with $\mathbb{C}^{kk'}$. And clearly the set of such parametrizations (or rather their inverse charts) constitute an atlas for the manifold $\mathcal{G}_k(V)$.

Having parametrized $\mathcal{G}_k(V)$ (or at least an open subset of it) the flow $R_k$ defines and is defined by a system of differential equations in the parametrization coordinates.

This system is easily determined from the original equation. The splitting of $V$ induces a splitting of $F$ into four linear maps which we denote $A, B, C, D$ to conform with the standard literature.

$$F(u_1 \oplus u_2) = (Au_1 + Bu_2) \oplus (Cu_1 + Du_2) \qquad [5:6]$$

$$\forall\, u_1 \in \mathcal{U}_1 \quad \forall\, u_2 \in \mathcal{U}_2$$

The original differential equation was

$$y' = Fy \qquad [5:7]$$

If

$$y = u \oplus \alpha u \qquad [5:8]$$

then

$$\begin{aligned} u' \oplus (\alpha' u + \alpha u') &= (Au + B\alpha u) \oplus (Cu + D\alpha u) \\ \Rightarrow \qquad u' &= Au + B\alpha u \\ \alpha' u &= Cu + D\alpha u - \alpha u' \\ &= Cu + D\alpha u - \alpha A u - \alpha B \alpha u \end{aligned} \qquad [5:9]$$

If the second equation is to hold for all initial values of $u$,

$$\alpha' = C + D\alpha - \alpha A - \alpha B \alpha \qquad [5:10]$$

which is the required equation determining the flow on $\mathcal{G}_k(V)$. In the particular case when $V$ is of even dimension and $k = n/2$ this is the central equation of those methods generally known as Ricatti or invariant imbedding methods (see e.g. Curl & Graebel (1972), Scott (1973), Davey (1977))†. An advantage of this derivation is that the extension to cases of arbitrary dimensionality is obvious, but equally as important is the conceptual gain from having a clear geometric formulation of the method. For instance it is well known that singularities may be encountered in the integration of the Ricatti equation and that the solution is to change to another set of variables which satisfy a related Ricatti equation. From the geometric point of view it is clear that this behaviour is an essential consequence of the fact that the manifold $\mathcal{G}_k(V)$ is not homeomorphic to an open subset of $\mathbb{C}^{kk'}$ and so cannot be covered by any one



† also D.M.Sloan, J.Comp.Phys.24,320(1977).

parametrization; whenever in the integration we move out of the region covered by our chart we must switch to another chart in the atlas which does cover the region we are entering. This raises the interesting question of how many such charts are needed to form a complete atlas. If it is confined to natural charts (as defined above and used in the standard methods) it is easy to see that at least $k+1$ are required (in contradiction to the impression often given in the literature that two will suffice). For given any $k$ natural charts on $\mathcal{G}_k(V)$ defined by the decompositions of $V$,

$$V = \mathcal{U}_1 \oplus \overline{\mathcal{U}}_1 = \cdots\cdots = \mathcal{U}_k \oplus \overline{\mathcal{U}}_k \qquad [5:11]$$

choose one vector from each of the complementary subspaces $\overline{\mathcal{U}}_1 \ldots\ \ldots\ \ldots \overline{\mathcal{U}}_k$ and if necessary adjoin to this set such additional arbitrary vectors that the subspace of $V$ it generates is of dimension $k$. Then this subspace is an element of $\mathcal{G}_k(V)$ that does not belong to any of the coordinate domains of the given $k$ charts.

## 6 Other Methods

This geometric point of view can also be used to describe many alternative integration techniques and suggests several new ones. By integrating in the dual space $V'$ we get the method of adjoints. The simple shooting method and the method of orthonormalisation can be regarded as an integration in $\oplus^k V$. And the method proposed by Davey in a recent preprint† involves integrating in $\wedge^k V$. A novel method which is not very efficient, but is reliable and easy to program, consists of integrating in the space of sets of $k$ orthogonal vectors.

This is obtained from the simple shooting method by a slight modification of the derivative subroutine; instead of integrating

$$y_\alpha' = F y_\alpha \qquad \alpha = 1 \cdots k \qquad [6:1]$$

we take an initial set of orthonormal vectors and integrate

$$y_\alpha' = \prod_{\beta=1}^{k} (1 - y_\beta y_\beta^\dagger) F y_\alpha \qquad \alpha = 1 \cdots k \qquad [6:2]$$

In other words we only keep that component of each derivative which is perpendicular to the subspace spanned by the vectors. This means that the vectors remain orthonormal and only rotate as much as is necessary for them to stay in the correct subspace.

It should perhaps be pointed out that the non-linear differential equations which arise in these methods



† Now published in J.Comp.Phys.30,137(1979).
$\wedge$ denotes the exterior or wedge product.

are generally quite stiff so that it may in certain cases be more efficient to use an initial value method designed for such equations. In general because of its robustness and economy I would recommend the method of orthonormalisations (using a Hermitian form) combined with an automated zero finder based on the principle of the argument (see Appendix B) as the best shooting method for this problem.

## 7 References

Conte,S.D. 1966 SIAM Rev. 8,309.

Curl,M.L. & Graebel,W.P. 1972 SIAM J.Appl.Math. 23,380.

Davey,A. 1977 J.Comp.Phys. 24,331.

Gersting,J.M. & Jankowski,D.F. 1972 Int.J.Num.Meth.Eng. 4,195.

Gary,J. & Helgason,R. 1970 J.Comp.Phys. 5,169.

Scott,M.R. 1973 J.Comp.Phys. 12,334.

In this appendix I apply the ideas developed in chapter 2 to plane parallel flows. As well as being a good illustration of these ideas the results obtained are in some cases stronger than those obtained by classical means (for an account of which see Drazin and Howard (1966)). The equations for plane parallel flow which correspond to equations [2:10] and [2:13] of chapter 2 must first be obtained.

If the unperturbed basic flow is

$$\underset{\sim}{U}_0 = V(x)\, \underset{\sim}{e}_y \qquad [A:1]$$

(I take the basic flow parallel to $\underset{\sim}{e}_y$ rather than $\underset{\sim}{e}_x$ for consistency with the axisymmetric case) and the velocity perturbation is

$$\underset{\sim}{u} := u\, \underset{\sim}{e}_x + v\, \underset{\sim}{e}_y \qquad [A:2]$$

(by Squire's theorem (Squire (1933)) we need only consider two dimensional perturbations) the equations of the linear system are (Fourier decomposing in $t$ and $y$,

$$q(x,y,t) := q(x)\, e^{i(\omega t - \alpha y)} \qquad [A:3]$$

)

$$u' = \frac{g}{c^2}u - \frac{\alpha V'}{\sigma}u + \frac{ip}{\rho_0}\left(\frac{\alpha^2}{\sigma} - \frac{\sigma}{c^2}\right)$$

[A:4]

$$p' = -\frac{g}{c^2}p + i\rho_0\left(\frac{N^2}{\sigma} - \sigma\right)u$$

$$\left[\sigma = \omega - \alpha V,\ g = -p_0'/\rho_0,\ N^2 = \frac{p_0'}{\rho_0}\left(\frac{\rho_0'}{\rho_0} - \frac{1}{c^2}\frac{p_0'}{\rho_0}\right)\right]$$

If we now define

$$\zeta = -\frac{ip}{\rho_0 u} \qquad [A:5]$$

the equation analogous to [2:13] is †

$$\zeta' = -\left(\frac{\rho_0'}{\rho_0} + \frac{2g}{c^2}\right)\zeta + \frac{1}{\sigma}\left\{N^2 - \sigma^2\left(1 + \frac{\zeta^2}{c^2}\right) + \alpha\zeta(\alpha\zeta + V')\right\} \qquad [A:6]$$

Examining the imaginary part of this for real values of $\zeta$ we find

$$(\zeta')_i = \frac{\omega_i}{|\sigma|^2}\left\{N^2 + |\sigma|^2\left(1 + \frac{\zeta^2}{c^2}\right) + \alpha\zeta(\alpha\zeta + V')\right\} \qquad [A:7]$$

(the subscripts $r$ and $i$ denote real and imaginary parts). Using the argument of section 2:4 it follows that a



† This can also be obtained from 2:13 in the limit $r \to \infty$

sufficient condition for stability with passive boundary conditions is that the Richardson number

$$Ri = N^2/(V')^2 \qquad [A:8]$$

should everywhere exceed 1/4 (it is interesting that this method of proof handles the compressible case as easily as the incompressible (compare Chimonas (1970)).

An important result in the theory of plane parallel flows is the semicircle theorem of Howard (1961); the argument of section 2:4 can also be adapted to prove this theorem.

<u>Theorem</u>

The frequency $\omega = \omega_r + i\,\omega_i$ of any unstable mode of a plane parallel statically stable flow between rigid boundaries lies in the semicircle having as diameter the range of $\alpha V$ in the basic flow.

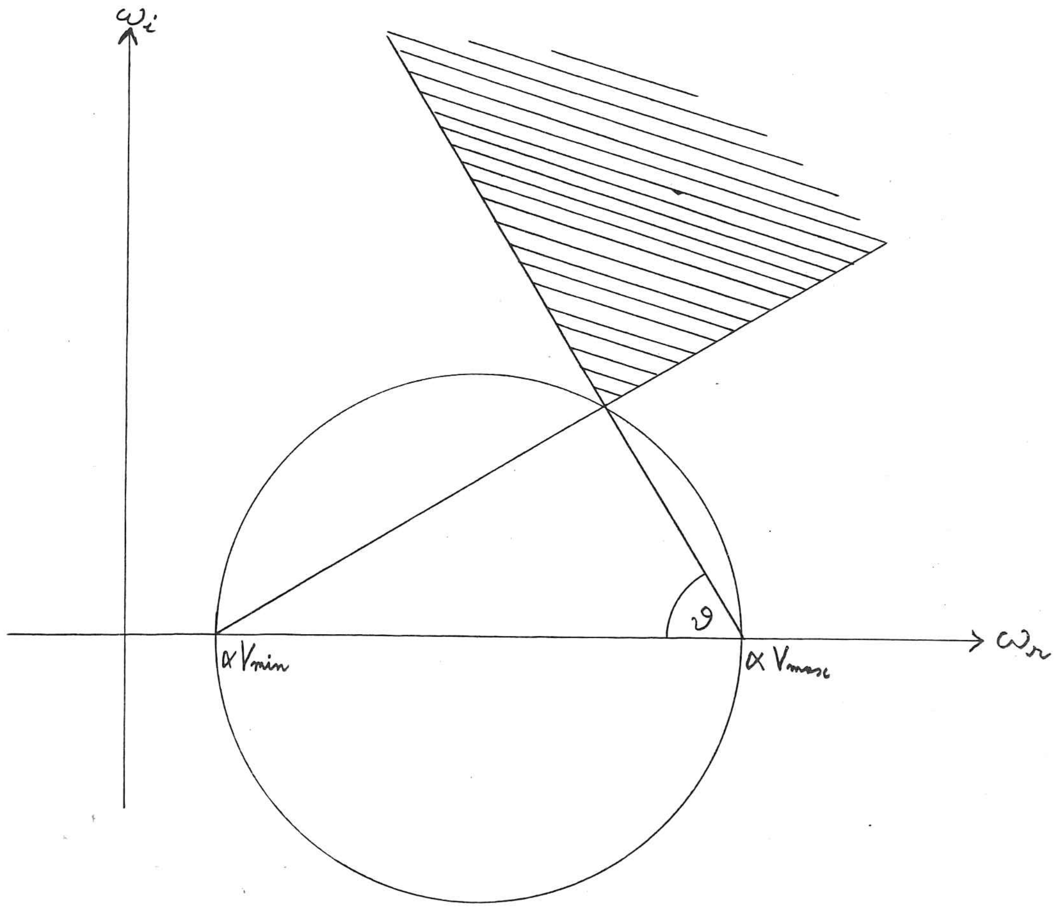
Proof

If in equation [A:6] we substitute $\eta = \sigma\zeta$ we obtain

$$\eta' = -\left(\frac{\rho_0'}{\rho_0} + \frac{2g}{c^2}\right)\eta + N^2 - \sigma^2 + \eta^2\left(\alpha^2/\sigma^2 - \frac{1}{c^2}\right) \qquad [A:9]$$

Examining the imaginary part of this on the great circle defined by the real values of $\eta$ we find

Fig.A1 The complex frequency $\omega$ is excluded from the shaded region




Facing page 106



$$(\eta')_i = -2\sigma_r\sigma_i\left[1+\frac{\alpha^2\eta^2}{|\sigma|^4}\right] \qquad \text{[A:10]}$$

so that if the boundary conditions require $\eta$ to be real a necessary condition for instability is that $\sigma_r$ should change sign (a result known to Rayleigh). However the only natural boundary conditions of this form are those which require $\eta = 0$ or $\infty$; from now on I will assume that the boundaries are rigid and the boundary conditions are $\eta = \zeta = \infty$. If we examine the vector field on all the circles passing through $0$ and $\infty$ we obtain a much stronger result. Multiplying $\eta$ by a pure phase, $\chi = e^{2i\vartheta}\eta$ †, we obtain

$$\chi' = -\left(\frac{\rho_0'}{\rho_0}+\frac{2g}{c^2}\right)\chi + (N^2-\sigma^2)e^{2i\vartheta} + \left(\frac{\alpha^2}{\sigma^2}-\frac{1}{c^2}\right)e^{-2i\vartheta} \qquad \text{[A:11]}$$

and thus for real $\chi$

$$(\chi')_i = -\left[2(\sigma e^{i\vartheta})_r(\sigma e^{i\vartheta})_i - N^2\sin 2\vartheta\right] - \frac{\alpha^2\chi^2}{|\sigma|^4}\left[2(\sigma e^{i\vartheta})_r(\sigma e^{i\vartheta})_i - \frac{|\sigma|^4}{\alpha^2c^2}\sin 2\vartheta\right] \qquad \text{[A:12]}$$

It follows that if $N^2>0$ (so that the flow is statically stable) $\omega$ is excluded from the region indicated in fig. A1; as this holds for all values of $\vartheta$ we see that $\omega$ must lie in the semicircle having as diameter the range of $\alpha V$.

---



† Which does not alter the form of a rigid boundary condition.

This is the semicircle theorem of Howard (for compressible flows); an important consequence is that for plane parallel flow between rigid boundaries instability can only set in through a singular neutral mode (Miles (1961)).

The only condition for stability proved so far is the Richardson criterion; the main purpose of the following lemma is to extend that result (in the sense of section 2:5) to flows without stratification.

Lemma

A plane parallel flow of an incompressible heterogeneous fluid with rigid boundaries has no singular neutral modes if the velocity profile is monotonic ($V' \neq 0$), the product $\rho_0 V'$ increases in modulus and $\rho_0^2 N^2$ decreases ($|\rho_0 V'|' \geqslant 0$ & $(N^2\rho_0^2)' \leqslant 0$) or vice versa, and the stratification is statically stable ($N^2 > 0$).

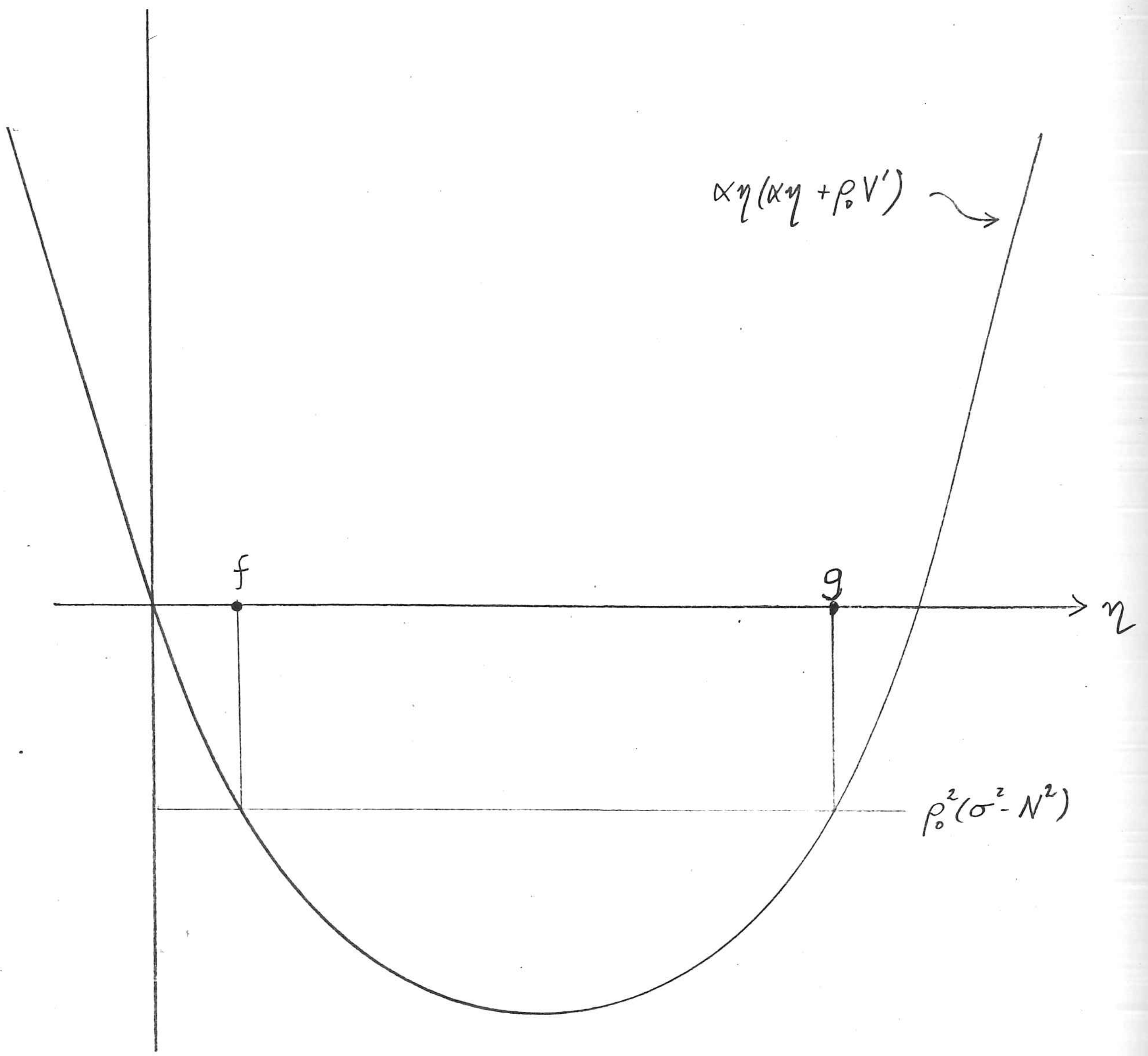
Proof

Suppose $|\rho_0 V'|' \geqslant 0$ and $(N^2\rho_0^2)' \leqslant 0$ (if the inequalities are reversed we simply invert the coordinate system). Because $V' \neq 0$ there can only be one critical layer, say at $x = c$, between the boundary points $x = a$ and $x = b$. The differential equation is

$$\zeta' = -\frac{\rho_0'}{\rho_0}\zeta + \frac{1}{\sigma}\left\{N^2 - \sigma^2 + \alpha\zeta(\alpha\zeta + V')\right\} \qquad [A:13]$$

or setting $\eta = \rho_0 \zeta$

Fig. A2   Graphical solution of equation A;15




Facing page 108



$$\eta' = \frac{1}{\rho_0\sigma}\left\{\rho_0^2(N^2-\sigma^2) + \alpha\eta(\alpha\eta+\rho_0 V')\right\} \qquad [A:14]$$

with boundary conditions $\eta = \infty$ at $x = a, b$.

From the general theory of singular neutral modes developed in sections 2:6 and 2:7 we know that there are only two solutions of [A:14] which could be singular neutral modes and that their behaviour can be partially determined by studying the functions $f(x)$ and $g(x)$. These are the roots of the quadratic equation in $\eta$

$$\alpha\eta(\alpha\eta+\rho_0 V') = -\rho_0^2(N^2-\sigma^2) \qquad [A:15]$$

the identification being that indicated in fig. A2 ($g$ is defined to be attractive as one integrates towards the critical layer).

As one integrates from $c$ towards $b$ the RHS of equation [A:15] never falls below its initial value at $c$ and $|\rho_0 V'|$ increases. Thus $|g(x)|$ also increases initially never falling below its value at $c$ and the roots $f(x)$ and $g(x)$ are real between $c$ and $b$. It follows, using the information on the slopes of the two solutions at the critical layer obtained in section 2:7, that neither solution can reach infinity and satisfy the boundary condition at $b$. Thus there can be no singular neutral mode.

---

Given velocity and density profiles satisfying the hypotheses of the lemma one can construct a sequence of models by increasing the strength of the gravitational field until the flow's stability is assured by the Richardson criterion. It then follows that because there are no singular neutral modes at any point on the sequence all the models are stable no matter how feeble the buoyancy effects; indeed going to the other end of the sequence we find that with no gravity operative a heterogeneous flow with a monotonic velocity profile is stable if $(\rho_0 V')' \neq 0$. The stability of the intermediate models is interesting as the only example known to me of a stability result holding for a fairly general class of stratified flows.† The particular hypotheses of the lemma were chosen to facilitate the proof; in general given a specific velocity profile and some information about the stratification a little sketching of $f$ and $g$ will suggest much sharper results. The effects of compressibility can be included in the analysis without much trouble; as long as the velocity differences in the basic flow remain below the sound speed they tend to be stabilising, but if the flow is supersonic the possibility of sound waves leads to a radical change in the nature of the problem.

I have shown that when there is no gravity a heterogeneous incompresible flow with $(\rho_0 V')' \neq 0$ and a monotonic velocity profile is stable. This result (without



† Other than the Richardson criterion.

the restriction to monotonic velocity profiles) is due to Synge (1933) and is a particular case of a general result on flows where the local Richardson number vanishes everywhere. This can happen for two reasons; either there is no gravity and the flow is isobaric or the flow is isentropic.

$$N^2 = \frac{p_0'}{\rho_0}\left(\frac{\rho_0'}{\rho_0} - \frac{1}{c^2}\frac{p_0'}{\rho_0}\right) = 0$$
$$\Rightarrow p_0' = 0 \quad \text{or} \quad p_0' = c^2 \rho_0' \qquad [A:16]$$

In both cases the vanishing of $Ri$ means that instead of there being two solutions taking real values on both sides of a critical layer there is only one unless the slope of the attractor vanishes at the critical layer (in which case the singularity is removable).

In the isobaric case [A:6] becomes

$$\zeta' = -\frac{\rho_0'}{\rho_0}\zeta + \frac{1}{\sigma}\left\{\alpha\zeta(\alpha\zeta + V') - \sigma^2\left(1 + \frac{\zeta^2}{c^2}\right)\right\} \qquad [A:17]$$

or setting $\eta = \rho_0 \zeta$

$$\eta' = \frac{1}{\rho_0 \sigma}\left\{\alpha\eta(\alpha\eta + \rho_0 V') - \sigma^2\left(\rho_0^2 + \frac{\eta^2}{c^2}\right)\right\} \qquad [A:18]$$

and in the isentropic case

$$\zeta' = +\frac{\rho_0'}{\rho_0}\zeta + \frac{1}{\sigma}\left\{\alpha\zeta(\alpha\zeta + V') - \sigma^2\left(1 + \frac{\zeta^2}{c^2}\right)\right\} \qquad [A:19]$$

or setting $\eta = \zeta \rho_0^{-1}$

$$\eta' = \frac{\rho_0}{\sigma}\left\{\alpha\eta\left(\alpha\eta + \frac{V'}{\rho_0}\right) - \sigma^2\left(\frac{1}{\rho_0^2} + \frac{\eta^2}{c^2}\right)\right\} \qquad [A:20]$$

Examining these equations we see that at the critical layer $f(x)$ is tangent to $\eta = 0$ and $g(x)$ to $\eta = \rho_0^{\pm 1} V'$; thus the slope of the attractor at the critical layer is determined by that of the product or quotient of the vorticity and the density according as the flow is isobaric or isentropic. We can now prove what is probably the most general form of Rayleigh's inflection point theorem,† but first we need the following lemma.

Lemma

If the maximum velocity difference between two parts of the basic flow is less than the sound speed, then a necessary condition for the existence of a singular neutral mode in a plane parallel flow between rigid boundaries of an isobaric fluid is that there be an extremum in $\rho_0 V'$ and of an isentropic fluid is that there be an extremum in $V'/\rho_0$.

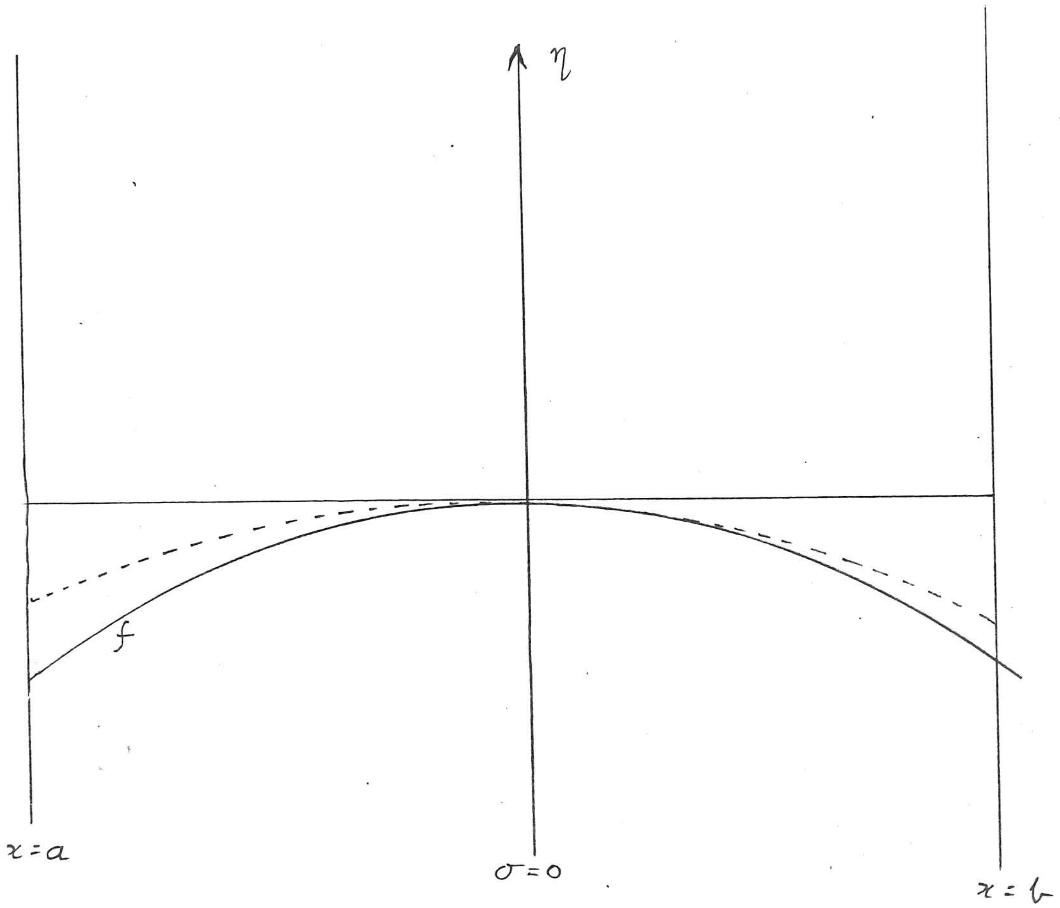
Proof

Let us consider the isobaric case, the proof for the isentropic case being essentially identical, and examine the possibility of a singular neutral mode in a flow where $(\rho_0 V')' \neq 0$. The velocity of the basic flow is the same at each critical layer so that the sign of $V'$ must alternate at successive layers; as the density is a positive definite



† Chimonas (1970) was unable to generalise Rayleigh's criterion to compressible flows.

Fig. A3 The possible solution (dashed) and the function $f$.




Facing page 112



function there can not be more than two critical layers without an extremum in $\rho_0 V'$. The hypothesis that the maximum velocity difference in the basic flow is less than the minimum sound speed implies

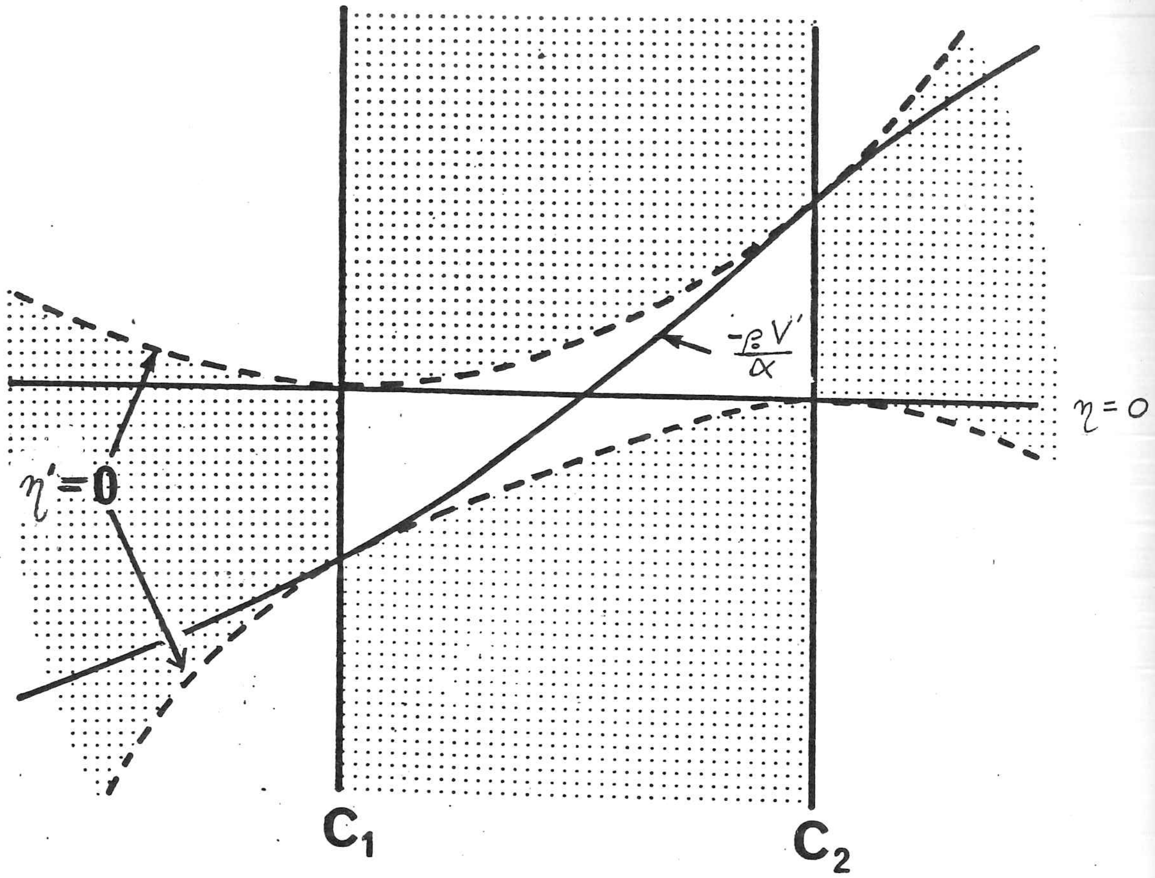
$$\frac{|\sigma|^2}{c^2} < \alpha^2 \qquad [A:21]$$

so that the roots $f(x)$ and $g(x)$ are real and finite at all points in $[a, b]$.

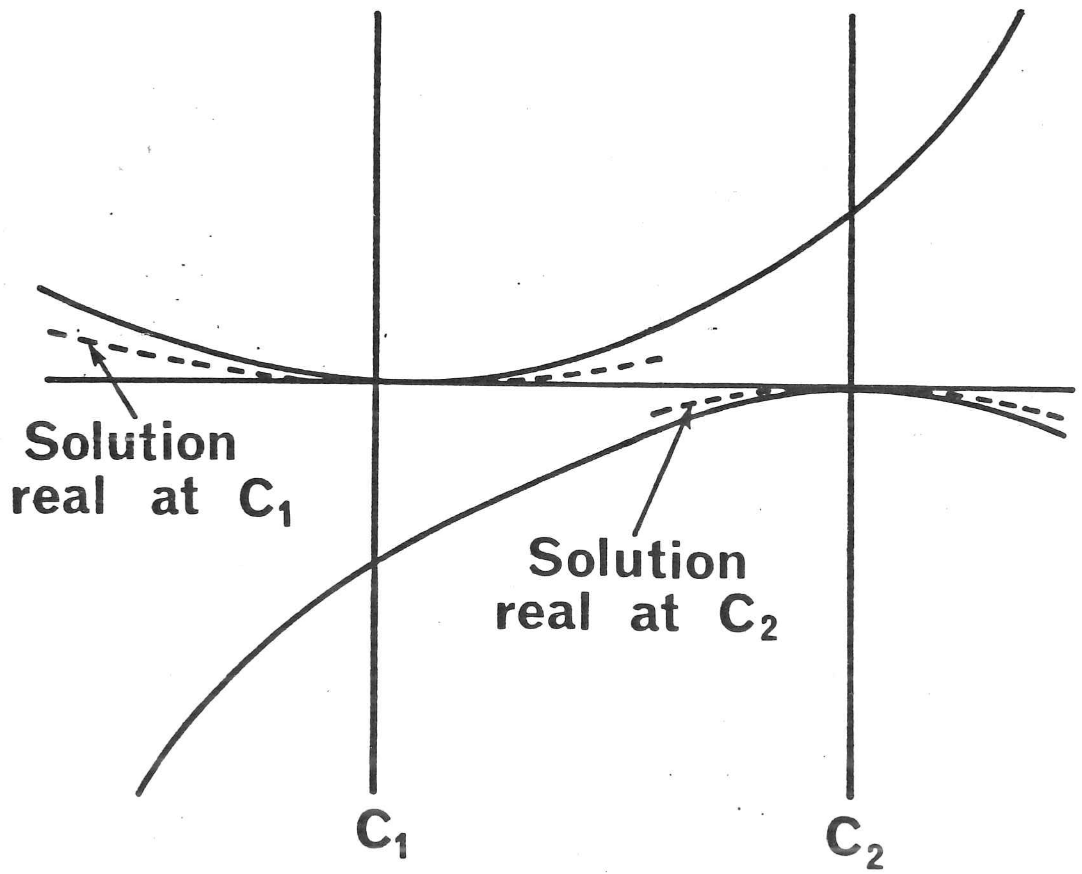
Let us consider the possibility that the singular neutral mode has one critical layer. Then the values of $\eta$ are real on either side of this layer and we know that there is only one solution with this property. However it is easy to see that this solution is always finite and so can not satisfy the boundary condition $\eta = \infty$ at either $a$ or $b$ (see fig. A3). Thus any singular neutral mode must have two critical layers.

Now consider the case of two critical layers in terms of the map defined at the end of section 2:7. As we pass the first critical layer the real axis and one hemisphere of the $\eta$ sphere is mapped into a disc in the opposite hemisphere and tangent to the real axis at one point. Because the sign of $V'$, and hence of $\rho_0 V'$, alternates and the sign of $(\rho_0 V')'$ is unchanged (by hypothesis) that of $\mathcal{G}$ changes; thus at the second critical layer this hemisphere (and its contained disc) is mapped

Fig A4 The case of two critical layers $C_1$ and $C_2$.




In the shaded areas $\eta'$ is negative








into a disc contained in the first hemisphere. Therefore the composite mapping obtained by integrating from $a$ to $b$ past both critical layers takes the real axis and one hemisphere into a subdisc of itself which in general will not touch the real axis; the condition for it to do so is that the solution which is real on both sides of the first critical layer should also be real on both sides of the second. Thus if there is a singular neutral mode with two critical layers the values taken by $\eta$ must be real between the critical layers as well as between the boundaries and the critical layers. But it is easy to see that the solutions real about each critical layer are incompatible both with themselves and with the boundary conditions (see fig. A4).

Thus there can be no singular neutral modes.

---

Theorem

If the maximum velocity difference between two parts of the basic flow is less than the sound speed, then a necessary condition for instability in a plane parallel flow between rigid boundaries of an isobaric fluid is that there be an extremum in $\rho_0 V'$ and of an isentropic fluid is that there be an extremum in $V'/\rho_0$.

Proof

We construct a series of flows none of which can have a singular neutral mode and which link a flow satisfying these conditions to one which we know to be stable. First we connect a heterogeneous flow to a homogeneous one.

Given an isobaric or isentropic flow with density $\rho_0$ and velocity profile $V_0$ such that $(\rho_0^{\pm 1} V_0')' \neq 0$ we define $\rho_\lambda = \rho_0^\lambda$ and $V_\lambda$ as any continuous family of solutions of $V_\lambda' = \rho_0^{\pm(1-\lambda)} V_0'$. Then as $\lambda$ runs from 0 to 1 we obtain a sequence (more exactly a homotopy) of flows, with density $\rho_\lambda$ and velocity $V_\lambda$, such that for all $\lambda$

$$(\rho_\lambda^{\pm 1} V_\lambda')' = (\rho_0^{\pm\lambda} \rho_0^{\pm(1-\lambda)} V_0')' = (\rho_0^{\pm 1} V_0')' \neq 0 \qquad [A:22]$$

which connects the given flow to a homogeneous one.

Then we 'straighten out' the velocity profile. Given a velocity profile $V_0$ with $V_0'' \neq 0$ we define

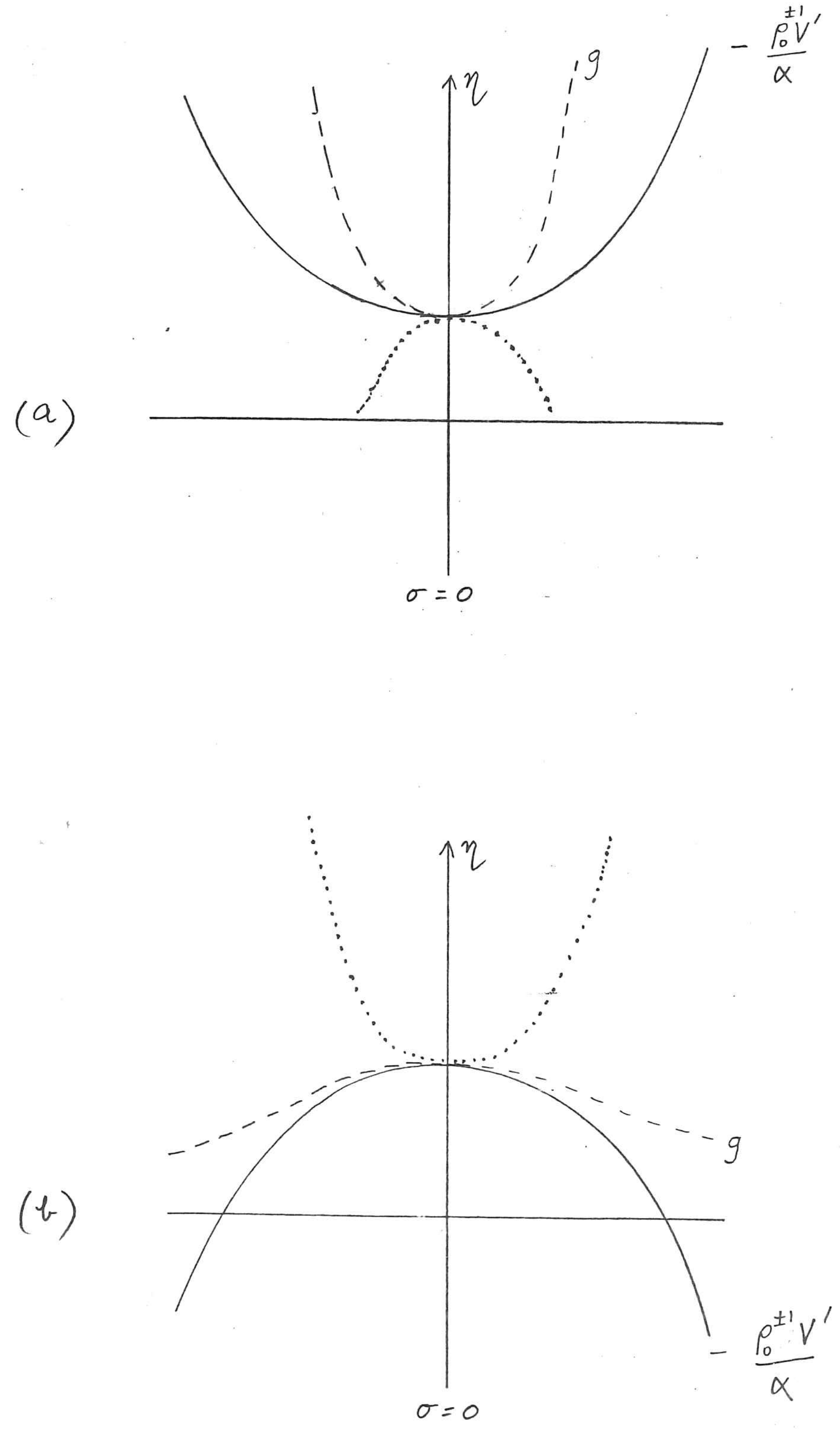
$$V_\lambda = (1-\lambda) V_0 + \lambda x \qquad [A:23]$$

Then

$$V_\lambda' = (1-\lambda) V_0' + \lambda \,, \quad V_\lambda'' = (1-\lambda) V_0'' \qquad [A:24]$$

so that as $\lambda$ runs from 0 to 1 we have a homotopy carrying the flow to one with a linear velocity profile (plane Couette flow). In both sequences the sound speed must be allowed to increase as much as is needed to keep the flows

Fig. A5 The attractor (dashed) and a typical solution (dotted) when the extremum is (a) a minimum in $|\rho_0^{\pm 1}\omega_0|$ and (b) a maximum. Only in the latter case can a solution escape to infinity on both sides of the critical layer.








subsonic. The above lemma will then show that none of the intermediate flows can have a singular neutral mode.

It follows from the particular result established earlier, that a heterogeneous incompressible isobaric flow with a monotonic velocity profile and $(\rho_0 V')' \neq 0$ between rigid boundaries is stable (alternatively one can use a special argument to prove the stability of plane Couette flow), that all these flows are stable.

---

This interesting extension of Rayleigh's inflection point theorem to compressible isentropic and isobaric flows appears to be new (though the result for incompressible isobaric flows was obtained by Synge(1933)). One further result can be obtained. A necessary condition for the existence of a singular neutral mode in a subsonic isentropic or isobaric flow between rigid boundaries is the existence of an extremum in $\rho_0^{\pm 1} V'$ (and the critical layer of the mode will coincide with the extremum); if the flow is such that $\rho_0^{\pm 1} V'$ has only one extremum, the necessary condition is that the extremum be a <u>maximum</u> in $|\rho_0^{\pm 1} V'|$ (this is an analogous extension of Fjortoft's (1950) result). The proof depends on the fact that if the extremum were a minimum then $|g|$ would have a minimum at the critical layer and any solution passing through $g_0$ at that point would, at least on one side, be unable to go to $\infty$ but would cross 0 and be trapped by $f$ (see fig A5). As all solutions pass through $g_0$ except for one which passes through $f_0$ and is also trapped (cf. fig A3) it follows that the boundary conditions can not be satisfied unless the extremum is a maximum.

## References

Chimonas,G. 1970 J.F.M. 43,833.

Drazin,P.G. & Howard,L.N. 1966 Adv.Appl.Mech. 9,1.

Fjortoft,R. 1950 Geofys.Publ. 17,1.

Squire,H.B. 1933 Proc.Roy.Soc.Lond. A142,621.

Synge,J.L.,1933 Trans.Roy.Soc.Can.27,1.

This FORTRAN package was written for use in solving Orr-Sommerfeld type equations by the method of orthonormalisation. It implements the following algorithm for finding the zeros of a holomorphic function (or of the product of a holomorphic function and a positive real function).

Given the function (defined by the external COMPLEX*16 function FCT) and a rectangular region of the complex plane,

$$XMIN < x < XMAX, \; YMIN < y < YMAX$$

it scans the boundary of this region (using an initial step DX) recording the location of the points at which the argument of the function (normalized to the range $+\pi$ to $-\pi$) jumps between $+\pi$ and $-\pi$ and the direction of the jump. From this information and Cauchy's principle of the argument it calculates the number of zeros of the function in the region. If this is non-zero it bisects the region by inserting a line joining the mid-points of the two longer sides and scans this line. This process of subdivision of those regions containing zeros is continued NPASS times so that at the end of the process each zero is located in a small box.

The process is very robust and works as well on clustered or multiple zeros as on isolated zeros. Of course once good approximations have been found to the zeros other methods (such as the secant method or a modified Newton method) with faster convergence can be used, but it is often surprising how good the approximations have to be before the local methods can be made to converge. The other great advantage of the method in eigenvalue calculations is that it is exhaustive; it gives a complete list of all the zeros in the region investigated with their multiplicities.

This particular implementation is designed on the assumption that function evaluations are slow and the general computation expensive. Before each function call it checks (using the local Cambridge routine TMTOGO and the argument ITIME) whether there is enough time left for the evaluation; if there is not it dumps out enough information to restart the search at that point (using an unformatted write to unit IDUMP). This test can be avoided by giving IDUMP the value 0.

The following computer output, as well as containing a complete listing of the BISECT package, is intended to illustrate its use in a typical example. The first part of the program (consisting of the subroutines DET, DERIV2, ORTHON and AB4) is the simplest possible implementation of the method of orthonormalisation to solve the eigenvalue problem for plane Poiseuille flow at a Reynold s number of $10^4$. For this demonstration the largest possible step has been taken in the integration routine; in consequence although the general pattern of the eigenvalues is correct (the three families in Mack's terminology: Mack (1976) J.F.M. 73,497) and the total number is correct the locations obtained are not exact (except for those relatively isolated eigenvalues belonging to the family which includes the single unstable eigenvalue). However this example is intended to demonstrate the use of the BISECT package in locating the eigenvalues of the discretised problem so that the accuracy of the discretisation is irrelevant.

The program listed was compiled and run (using the FORTQ compiler) and produced the appended output in 2mins 4sec of CPU time. The spurious "-1 zero" (which should cancel one of its neighbours) probably results from one zero lying almost exactly on a dividing line. The speed and convenience of the package is evident, particularily when one remembers that all the data obtained is available to other programs through the dump.

```
      IMPLICIT REAL*8 (A-H,O-Z)
      COMPLEX*16 DET
      EXTERNAL DET
      COMMON/PARAMS/CR,CI,RE
      RE=1.D4
      CALL BISECT(0.D0,1.D0,-0.7D0,0.3D0,DET,1.D-2,10,5,1)
      STOP
      END
C
C-----------------------------------------------------------------------
C
      FUNCTION DET(C)
      ***************
C
C
C    A simple implementation of the method of orthonormalisation
C    for plane Poiseuille flow.
C
      IMPLICIT COMPLEX*16 (A-H,O-Z)
      REAL*8 RE,X,NODE
      EXTERNAL DERIV2
C
      DIMENSION Y(8),NODE(11)
C
      DATA NODE/0.D0,0.1D0,0.2D0,0.3D0,0.4D0,
     &          0.5D0,0.6D0,0.7D0,0.8D0,0.9D0,1.D0/
C
      COMMON/PARAMS/CFUDGE,RE
C
C
      CFUDGE=C
C
C    Fudge to get parameter into common .
C    Note also tricky use of COMPLEX*16 as equivalent to
C    two REAL*8's in subroutine calls and common.
C
      DO 100 I=1,8
100   Y(I)=(0.D0,0.D0)
      Y(1)=(1.D0,0.D0)
      Y(7)=(1.D0,0.D0)
C
C    Initial values
C
      DO 200 I=1,9
         CALL AB4(Y,Y,NODE(I),NODE(I+1),30,DERIV2,16)
200      CALL ORTHON(Y)
      CALL AB4(Y,Y,NODE(10),NODE(11),30,DERIV2,16)
C
C
      DET = Y(1)*Y(6)-Y(2)*Y(5)
C
C
      RETURN
      END
C
C-----------------------------------------------------------------------
C
      SUBROUTINE DERIV2(DY,Y,X)
      *************************
C
C
C    This calculates the derivatives for the two solutions.
```

```
C    Real arithmetic is used (rather than complex) for speed.
C
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION Y(16),DY(16)
      COMMON/PARAMS/CR,CI,RE
C
C
      T1R=1-X*X-CR
      T1I=-CI
      T2R=2-T1R
      T2I=CI
C
C
      DO 100 I=1,9,8
         I6=I+5
         DO 101 J=I,I6
101         DY(J)=Y(J+2)
         DY(I+6)=Y(I+4)+Y(I+4)-Y(I)-RE*(T1R*Y(I+5)+T1I*Y(I+4)
     &   +T2R*Y(I+1)+T2I*Y(I))
100      DY(I+7)=Y(I+5)+Y(I+5)-Y(I+1)+RE*(T1R*Y(I+4)-T1I*Y(I+5)
     &   +T2R*Y(I)-T2I*Y(I+1))
C
C
C
      RETURN
      END
C
C-----------------------------------------------------------------------------
C
      SUBROUTINE AB4(YS,YE,XS,XE,NSTEP,DERIV,N)
C     ******************************************
C
C    A simple Adams-Bashforth integrator (4th order) with a RK starter.
C
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION Y(100),YS(N),YE(N),DY(400),W(400)
      EXTERNAL DERIV
C
      X=XS
      DO 100 I=1,N
  100 Y(I)=YS(I)
      H=(XE-XS)/DFLOAT(NSTEP)
      H2=H*.5D0
      H6=H/6.D0
      H24=H6/4.D0
      IP1=0
      IP2=N
      IP3=N+N
      IP4=IP3+N
C
      ITOP=IP3+1
      DO 600 I=1,ITOP,N
      CALL DERIV(DY(I),Y,X)
      X=X+H2
      DO 200 J=1,N
  200 W(IP4+J)=Y(J)+H2*DY(J+I-1)
      CALL DERIV(W(1),W(IP4+1),X)
      DO 300 J=1,N
  300 W(IP4+J)=Y(J)+H2*W(J)
      CALL DERIV(W(N+1),W(IP4+1),X)
```

```
      X=X+H2
      DO 400 J=1,N
  400 W(IP4+J)=Y(J)+H*W(N+J)
      CALL DERIV(W(IP3+1),W(IP4+1),X)
      DO 500 J=1,N
  500 Y(J)=Y(J)+H6*(DY(J+I-1)+W(J)+W(J)+W(N+J)+W(N+J)+W(IP3+J))
600   CONTINUE
C
C   We can now start to use the predictor-corrector method.
C
      NEND=NSTEP-3
      DO 1600 I=1,NEND
      CALL DERIV(DY(IP4+1),Y,X)
C
C   Predict
C
      DO 1100 J=1,N
 1100 W(J)=Y(J)+H24*(5.5D1*DY(J+IP4)-5.9D1*DY(J+IP3)+3.7D1*DY(J+IP2)
     &-9.D0*DY(J+IP1))
C
C
      X=X+H
      IPSV=IP1
      IP1=IP2
      IP2=IP3
      IP3=IP4
      IP4=IPSV
C
C
      CALL DERIV(DY(IP4+1),W(1),X)
C
C  Correct
C
      DO 1400 J=1,N
 1400 Y(J)=Y(J)+H24*(9.D0*DY(J+IP4)+1.9D1*DY(J+IP3)
     &-5.D0*DY(J+IP2)+DY(J+IP1))
C
C
 1600 CONTINUE
C
      DO 1700 I=1,N
 1700 YE(I)=Y(I)
C
      RETURN
      END
C
C-----------------------------------------------------------------------------
C
      SUBROUTINE ORTHON(Y)
      ********************
C
C
C   This orthonormalises the two vectors.
C
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION Y(16)
      ZR=0.D0
      DO 100 I=1,8
100         ZR=ZR+Y(I)*Y(I)
      ZR=1.D0/DSQRT(ZR)
      DO 101 I=1,8
```

```
101       Y(I)=Y(I)*ZR
C
C
      ZR=0.D0
      ZI=0.D0
      DO 200 I=1,7,2
          ZR=ZR+Y(I)*Y(I+8)+Y(I+1)*Y(I+9)
200       ZI=ZI+Y(I)*Y(I+9)-Y(I+1)*Y(I+8)
      DO 201 I=9,15,2
          Y(I)=Y(I)-ZR*Y(I-8)+ZI*Y(I-7)
201       Y(I+1)=Y(I+1)-ZR*Y(I-7)-ZI*Y(I-8)
C
C
      ZR=0.D0
      DO 300 I=9,16
300       ZR=ZR+Y(I)*Y(I)
      ZR=1.D0/DSQRT(ZR)
      DO 301 I=9,16
301       Y(I)=Y(I)*ZR
C
C
      RETURN
      END
C
C-----------------------------------------------------------------------
C
      SUBROUTINE BISECT(XMIN,XMAX,YMIN,YMAX,FCT,DX,NPASS,ITIME,IDUMP)
      ***************************************************************
C
C
C    The region to be searched is specified by XMIN-YMAX; the function
C    by the external COMPLEX*16 function FCT; DX is an estimate of the
C    step needed in scanning; NPASS the number of bisections; if IDUMP
C    is non-zero the data obtained is dumped to unit IDUMP when less
C    than ITIME seconds are left before the job runs out of cpu time.
C
      COMPLEX*16 VERTEX,DCMPLX
      REAL*8 DX,DX0,XMAX,XMIN,YMAX,YMIN,DREAL,DIMAG
      LOGICAL FLAG,HORZ
C
      DIMENSION VERTEX(200),VARG(200),INDEX(200,4),MAP(100,4),STACK(500)
     & ,BLOCK1(1500),BLOCK2(1232)
      EXTERNAL FCT
C
      COMMON/BIS1/ VARG,STACK,VERTEX
      COMMON/BIS2/ SCDATA(16),DX0,ARG0,N0,N1,IT,ID,
     & NVERT,NSTK,NREG,NVERT0,NREG0,
     & IPASS,IREG,FLAG,HORZ,
     & MAP,INDEX
      EQUIVALENCE(VARG(1),BLOCK1(1)),(SCDATA(1),BLOCK2(1))
C
C
      VERTEX(1)=DCMPLX(XMIN,YMIN)
      VERTEX(2)=DCMPLX(XMAX,YMIN)
      VERTEX(3)=DCMPLX(XMIN,YMAX)
      VERTEX(4)=DCMPLX(XMAX,YMAX)
C
      DO 1 I=1,200
          VARG(I)=10.0
          DO 1 J=1,4
              INDEX(I,J)=0
```

```
1       CONTINUE
C
C     Define the first region.
C
        DO 2 I=1,4
      2 MAP(1,I)=I
C
C     Initialize the pointers, counter and flag.
C
        NVERT=5
        NREG=2
        NSTK=1
        IPASS=0
        FLAG=.TRUE.
C
C     Scan the boundary of the first region.
C
        DX0=DX
        IT=ITIME*38400
C   38400 timer units per second!
        ID=IDUMP
        INDEX(1,1)=NSTK
 10     CALL SCAN(1,3,FCT)
        INDEX(1,2)=NSTK
        INDEX(1,3)=NSTK
 20     CALL SCAN(1,2,FCT)
        INDEX(1,4)=NSTK
        INDEX(2,1)=NSTK
 30     CALL SCAN(2,4,FCT)
        INDEX(2,2)=NSTK
        INDEX(3,3)=NSTK
 40     CALL SCAN(3,4,FCT)
        INDEX(3,4)=NSTK
C
C     Iterate.
C
 3      IPASS=IPASS+1
           CALL REDUCE
           IREG=1
           NVERT0=NVERT
           NREG0=NREG
           HORZ=(DIMAG(VERTEX(MAP(1,3)))-DIMAG(VERTEX(MAP(1,1)))
     &         .GT.DREAL(VERTEX(MAP(1,2)))-DREAL(VERTEX(MAP(1,1))))
 50             CALL SUBDIV(IREG,FCT)
                IREG=IREG+1
                IF (IREG.LT.NREG0) GO TO 50
           IF (IPASS.LT.NPASS) GO TO 3
        CALL REDUCE
        IF (ID .NE. 0) WRITE(ID) BLOCK1,BLOCK2
        RETURN
C       ******
        ENTRY RESUME(IDOLD,IDNEW,NPASS,FCT)
C
C     This enables one to read in (on unit IDOLD) the data dumped
C     previously and carry on; having old and new units saves copying
C     or rewinding files; the data is shuffled between two files.
C
        READ(IDOLD) BLOCK1,BLOCK2
        ID=IDNEW
        FLAG=.FALSE.
```

```
      IF (IPASS.GT.0) GO TO 50
      IF (N1 .EQ. 4) GO TO 100
      IF (N1 .EQ. 3) GO TO 10
      GO TO 20
  100 IF (NO .EQ. 2) GO TO 30
      GO TO 40
      END
C-----------------------------------------------------------------------------
      SUBROUTINE COUNT(N,NZERO)
C     ****************
C
C   This subroutine counts the number of zeros in region N
C
      REAL*8 DX
      DIMENSION STACK(500),INDEX(200,4),MAP(100,4)
      COMMON/BIS1/ VARG(200),STACK,VERTEX(800)
      COMMON/BIS2/ SCDATA(23),
     & NVERT,NSTK,NREG,NVERTO,NREGO,
     & IPASS,IREG,FLAG,HORZ,
     & MAP,INDEX
C
      NV1=MAP(N,1)
      NV2=MAP(N,2)
      NV3=MAP(N,3)
      NV4=MAP(N,4)
C
      NA1=INDEX(NV1,1)
      NB1=INDEX(NV1,2)
      NA2=INDEX(NV1,3)
      NB2=INDEX(NV1,4)
      NA3=INDEX(NV2,1)
      NB3=INDEX(NV2,2)
      NA4=INDEX(NV3,3)
      NB4=INDEX(NV3,4)
C
      NZERO=0
C
      IF (NA1.EQ.NB1) GO TO 1
      NEND=NB1-1
      DO 2 I=NA1,NEND
    2 NZERO=NZERO+IFIX(SIGN(1.0,STACK(I)))
    1 IF (NA4.EQ.NB4) GO TO 3
      NEND=NB4-1
      DO 4 I=NA4,NEND
    4 NZERO=NZERO+IFIX(SIGN(1.0,STACK(I)))
    3 IF (NA3.EQ.NB3) GO TO 5
      NEND=NB3-1
      DO 6 I=NA3,NEND
    6 NZERO=NZERO-IFIX(SIGN(1.0,STACK(I)))
    5 IF (NA2.EQ.NB2) GO TO 7
      NEND=NB2-1
      DO 8 I=NA2,NEND
    8 NZERO=NZERO-IFIX(SIGN(1.0,STACK(I)))
    7 RETURN
      END
C-----------------------------------------------------------------------------
      SUBROUTINE INSERT(N1,N2,N3,FCT)
C      ****************
C
C     This subroutine defines a new vertex half way between vertices N1 and
```

```
C     N2 if one does not already exist ; if a new vertex is created the stack
C     and index are updated.
C
      COMPLEX*16 VERTEX,VN,FCT,Z,VTEST,DCMPLX
      REAL*8 DREAL,DIMAG,DBLE,DX,XT,XEND,XHALF,XSUP,XINF,XERR
      LOGICAL HORZ
      EXTERNAL FCT
      DIMENSION VERTEX(200),VARG(200),INDEX(200,4),STACK(500),MAP(100,4)
     & ,BLOCK1(1500),BLOCK2(1232)
      EQUIVALENCE (VARG(1),BLOCK1(1)),(SCDATA(1),BLOCK2(1)),
     &(DX,SCDATA(17))
      COMMON/BIS1/VARG,STACK,VERTEX
      COMMON/BIS2/ SCDATA(21),ITIME,IDUMP,
     & NVERT,NSTK,NREG,NVERTO,NREGO,
     & IPASS,IREG,FLAG,HORZ,
     & MAP,INDEX
C
      ARG(Z)=ATAN2(SNGL(DIMAG(Z)),SNGL(DREAL(Z)))
C
      VN=(VERTEX(N1)+VERTEX(N2))*0.5D0
      IF (NVERT.EQ.NVERTO) GO TO 101
      NVERT1=NVERT-1
      DO 100 I=NVERTO,NVERT1
C     Search to see if vertex already exists.
         IF((DREAL(VERTEX(I)).EQ.DREAL(VN)).AND.(DIMAG(VERTEX
     &      (I)).EQ.DIMAG(VN))) GO TO 800
  100 CONTINUE
  101 N3=NVERT
      IF (IDUMP.EQ.0) GO TO 110
      CALL TMTOGO(IT)
      IF (IT.GT.ITIME) GO TO 110
         WRITE(IDUMP) BLOCK1,BLOCK2
         WRITE(6,9100)
         STOP 3
         ******
C
  110 VARG(N3)=ARG(FCT(VN))
C
      IH=2
      IF (HORZ) IH=0
      NSTART=INDEX(N1,IH+1)
      NSTOP=INDEX(N1,IH+2)
      NEND=NSTOP-1
      N=NSTART
      IF (NSTART.EQ.NSTOP) GO TO 500
C
      IF (HORZ) XEND=DIMAG(VERTEX(N2))-DIMAG(VERTEX(N1))
      IF (.NOT.HORZ) XEND=DREAL(VERTEX(N2))-DREAL(VERTEX(N1))
      XHALF=XEND*0.5D0
      DO 200 N=NSTART,NEND
         AS = ABS(STACK(N))
         IF (AS .GT. XHALF) GO TO 210
200   CONTINUE
      N=NSTOP
210   IF (N.EQ.NSTART) GO TO 211
         XINF = ABS(STACK(N-1))
         GO TO 212
211   XINF=0.D0
212   IF (N.EQ.NSTOP) GO TO 221
         XSUP = AS
         GO TO 222
```

```
221         XSUP = XEND
222      CONTINUE
         XERR=(XSUP-XINF)*0.2D0
         IF (XHALF-XINF.LT.XERR) GO TO 300
         IF (XSUP-XHALF.LT.XERR) GO TO 400
         GO TO 500.
C
300      IF (HORZ) VTEST=VERTEX(N1)+DCMPLX(0.D0,XINF)
         IF (.NOT.HORZ) VTEST=VERTEX(N1)+DCMPLX(XINF,0.D0)
         ARGT=ARG(FCT(VTEST))
         CALL LFIT(ARGT,VARG(N3),XINF,XHALF)
         NSTK=NSTK-1
         STACK(N-1)=SIGN(STACK(NSTK),STACK(N-1))
         IF (ABS(STACK(N-1)).GT.XHALF) N=N-1
         GO TO 500
C
400      IF (HORZ) VTEST=VERTEX(N1)+DCMPLX(0.D0,XSUP)
         IF (.NOT.HORZ) VTEST=VERTEX(N1)+DCMPLX(XSUP,0.D0)
         ARGT=ARG(FCT(VTEST))
         CALL LFIT(VARG(N3),ARGT,XHALF,XSUP)
         NSTK=NSTK-1
         STACK(N)=SIGN(STACK(NSTK),STACK(N))
         IF (ABS(STACK(N)).LT.XHALF) N=N+1
C
500      INDEX(N1,IH+2)=N
         INDEX(NVERT,IH+1)=N
         INDEX(NVERT,IH+2)=NSTOP
         IF (N .EQ. NSTOP) GO TO 601
         XHALFS=SNGL(XHALF)
            DO 600 I=N,NEND
600         STACK(I)=STACK(I)-SIGN(XHALFS,STACK(I))
601      VERTEX(NVERT)=VN
         NVERT=NVERT+1
         IF (NVERT .LT. 200) RETURN
C                            ******
         WRITE (6,9000)
 9000    FORMAT(10X,'WARNING! Attempt to define too many vertices.')
 9100    FORMAT('    THE END IS NIGH; DATA DUMPED FROM ROUTINE INSERT')
         STOP
C        ****
  800    N3=I
         RETURN
C        ******
         END
C-----------------------------------------------------------------------------
         SUBROUTINE SCAN(N0,N1,FCT)
C        ****************
         COMPLEX*16 V0,V1,DIRN,Z,FCT,VERTEX
         REAL*8 X0,X1,XE,XM,DX,DREAL,DIMAG,CDABS
         LOGICAL FLAG
         EXTERNAL FCT
         DIMENSION BLOCK1(1500),BLOCK2(1232)
         EQUIVALENCE (VARG(1),BLOCK1(1)),(V0,BLOCK2(1))
         COMMON/BIS1/ VARG(200),STACK(500),VERTEX(200)
         COMMON/BIS2/ V0,V1,DIRN,XE,X0,DX,ARG0,NN0,NN1,ITIME,IDUMP,
       & NVERT,NSTK,NREG,NVERT0,NREG0,
       & IPASS,IREG,FLAG,HORZ,
       & MAP,INDEX
         DATA PI,PI6,PI2,PI23,PI43,PI32,PI116/3.141593,.5235987,1.570796
       &,2.094395,4.188790,4.712389,5.759587/
```

```
C
      ARG(Z)=ATAN2(SNGL(DIMAG(Z)),SNGL(DREAL(Z)))
C
      IF (FLAG) GO TO 10
         FLAG=.TRUE.
         IF (NO.NE.NNO.AND.N1.NE.NN1) GO TO 10
         ARG1=ARGO
         X1=X0
         GO TO 300
 10   V0=VERTEX(N0)
      V1=VERTEX(N1)
      NN0=N0
      NN1=N1
      XE=CDABS(V1-V0)
      DIRN=(V1-V0)/XE
      X1=0.D0
      DX=DX*0.5D0
      IF (VARG(N0).EQ.10.0) VARG(N0)=ARG(FCT(V0))
      ARG1=VARG(N0)
      GO TO 300
 100  IF (IDUMP.EQ.0) GO TO 101
      CALL TMTOGO(IT)
      IF (IT.GT.ITIME) GO TO 101
         WRITE(IDUMP) BLOCK1,BLOCK2
         WRITE (6,9000)
         STOP 3
         ******
C
 101  ARG1=ARG(FCT(V0+X1*DIRN))
 102  ADIFF=ABS(ARG1-ARG0)
      IF (ADIFF.GT.PI6) GO TO 200
         DX=1.4D0*DX
         GO TO 300
 200  IF (ADIFF.GT.PI2) GO TO 210
         GO TO 300
 210  IF (ADIFF.GT.PI23) GO TO 220
         DX=DX*0.5D0
         GOTO 300
 220  IF (ADIFF.GT.PI43) GO TO 230
         XM=(X0+X1)*0.5D0
         DX=DX*0.5D0
         IF (IDUMP.EQ.0) GO TO 221
         CALL TMTOGO(IT)
         IF (IT .GT. ITIME) GO TO 221
            WRITE (IDUMP) BLOCK1,BLOCK2
            WRITE (6,9000)
            STOP 3
            ******
C
 221     ARGM=ARG(FCT(V0+XM*DIRN))
         IF (ABS(ARG1-ARGM).LT.PI) GO TO 222
            CALL LFIT(ARGM,ARG1,XM,X1)
            GO TO 300
 222     IF (ABS(ARGM-ARG0).LT.PI) GO TO 300
            CALL LFIT(ARG0,ARGM,X0,XM)
            GO TO 300
 230  IF (ADIFF.GT.PI32) GO TO 240
         DX=DX*0.5D0
         GO TO 260
 240  IF (ADIFF.GT.PI116) GO TO 250
         GO TO 260
 250  DX=1.4D0*DX
```

```
  260 CALL LFIT(ARGO,ARG1,XO,X1)
C
  300 IF (X1.EQ.XE) RETURN
C                   ******
      XO=X1
      X1=X1+DX
      ARGO=ARG1
      IF (X1.LT.XE) GO TO 100
      X1=XE
      IF (VARG(N1).EQ.10.0) VARG(N1)=ARG(FCT(V1))
      ARG1=VARG(N1)
      GO TO 102
C
 9000 FORMAT ('    THE END IS NIGH!! DATA DUMPED')
      END
C-----------------------------------------------------------------------
      SUBROUTINE LFIT(ARGO,ARG1,XO,X1)
C     ***************
      REAL*8 XO,X1,DX
      DATA PI/3.141593/
      COMMON /BIS1/ VARG(200),STACK(500),VERTEX(800)
      COMMON/BIS2/ SCDATA(23),
     & NVERT,NSTK,NREG,NVERTO,NREGO,
     & IPASS,IREG,FLAG,HORZ,
     & MAP(100,4),INDEX(200,4)
      YO=SNGL(XO)
      Y1=SNGL(X1)
      TARGO=ARGO-SIGN(PI,ARGO)
      TARG1=ARG1-SIGN(PI,ARG1)
      YZ=(YO*TARG1-Y1*TARGO)/(TARG1-TARGO)
      STACK(NSTK)=SIGN(YZ,ARG1)
      NSTK=NSTK+1
      IF (NSTK.LT.500) RETURN
C                      ******
      WRITE (6,9000)
 9000 FORMAT (10X,'WARNING! Stack about to overflow.')
      STOP
C     ****
      END
C-----------------------------------------------------------------------
      SUBROUTINE REDUCE
C     ***************
C
C  This subroutine counts the number of zeros in each region and deletes
C  from the map those with none
C
C
      REAL*8 DX,DREAL,DIMAG
      COMPLEX*16 VERTEX,V1,V4
      COMMON/BIS1/ VARG(200),STACK(500),VERTEX(200)
      COMMON/BIS2/ SCDATA(23),
     & NVERT,NSTK,NREG,NVERTO,NREGO,
     & IPASS,IREG,FLAG,HORZ,
     & MAP(100,4),INDEX(200,4)
      L=1
      NEND=NREG-1
      DO 1 J=1,NEND
      CALL COUNT(J,NZ)
      IF (NZ.EQ.0) GO TO 1
      NV1=MAP(J,1)
```

```
      NV4=MAP(J,4)
      V1=VERTEX(NV1)
      V4=VERTEX(NV4)
      XMAX=DREAL(V4)
      XMIN=DREAL(V1)
      YMAX=DIMAG(V4)
      YMIN=DIMAG(V1)
      IF (NZ.EQ.1) GO TO 10
       WRITE (6,9001) XMIN,XMAX,YMIN,YMAX,NZ
      GO TO 11
10     WRITE (6,9002) XMIN,XMAX,YMIN,YMAX
11    CONTINUE
 9001 FORMAT(10X,'In the region,',G14.5,'< x <',G14.5,',',
     &G14.5,'< y <',G14.5,'there are',I3,' zeros.')
 9002 FORMAT(10X,'In the region,',G14.5,'< x <',G14.5,','
     &,G14.5,'< y <',G14.5,'there is 1 zero.')
      DO 2 K=1,4
    2 MAP(L,K)=MAP(J,K)
      L=L+1
    1 CONTINUE
      NREG=L
      IF (L.EQ.1) GO TO 3
      WRITE (6,9003)
 9003 FORMAT(//10X,100('+')//)
      RETURN
C     ******
    3 WRITE (6,9004)
 9004 FORMAT (30X,'     THERE ARE NO ZEROS IN THE GIVEN REGION')
      STOP
      END
C-----------------------------------------------------------------------------
      SUBROUTINE SUBDIV(N,FCT)
C     *****************
C
C  This subroutine divides region n into two by inserting a vertical
C  line if the region is broader than it is tall and a horizontal one
C  otherwise.   The map entries are updated and the new line scanned.
C
      COMPLEX*16 VERTEX,V1,V2,V3,V4,V5,V6
      REAL*8 DREAL,DIMAG,DX
      LOGICAL FLAG,HORZ
      DIMENSION VERTEX(200),INDEX(200,4),MAP(100,4)
      EXTERNAL FCT
      COMMON/BIS1/VARG(200),STACK(500),VERTEX
      COMMON/BIS2/ SCDATA(19),N0,N1,IT,ID,
     & NVERT,NSTK,NREG,NVERT0,NREG0,
     & IPASS,IREG,FLAG,HORZ,
     & MAP,INDEX
C
      NV1=MAP(N,1)
      NV2=MAP(N,2)
      NV3=MAP(N,3)
      NV4=MAP(N,4)
      V1=VERTEX(NV1)
      V2=VERTEX(NV2)
      V3=VERTEX(NV3)
C
C  It is necessary to decide whether the region is to be divided horizontally
C  or vertically.
C
```

```
      IF (HORZ) GO TO 10
      CALL INSERT(NV1,NV2,NV5,FCT)
      CALL INSERT(NV3,NV4,NV6,FCT)
C
      INDEX(NV5,1)=NSTK
    2 CALL SCAN(NV5,NV6,FCT)
      INDEX(NV5,2)=NSTK
C
      MAP(N,2)=NV5
      MAP(N,4)=NV6
C
      MAP(NREG,1)=NV5
      MAP(NREG,2)=NV2
      MAP(NREG,3)=NV6
      MAP(NREG,4)=NV4
C
      NREG=NREG+1
      RETURN
C     ******
C
   10 CALL INSERT(NV1,NV3,NV5,FCT)
      CALL INSERT(NV2,NV4,NV6,FCT)
C
      INDEX(NV5,3)=NSTK
   12 CALL SCAN(NV5,NV6,FCT)
      INDEX(NV5,4)=NSTK
C
      MAP(N,3)=NV5
      MAP(N,4)=NV6
C
      MAP(NREG,1)=NV5
      MAP(NREG,2)=NV6
      MAP(NREG,3)=NV3
      MAP(NREG,4)=NV4
C
      NREG=NREG+1
      RETURN
C     ******
      END
```

In the region, .0 < x < 1.0000 , -.70000 < y < .30000 there are 27 zeros.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

In the region, .0 < x < .50000 , -.70000 < y < .30000 there are 5 zeros.
In the region, .50000 < x < 1.0000 , -.70000 < y < .30000 there are 22 zeros.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

In the region, .0 < x < .50000 , -.70000 < y < -.20000 there are 2 zeros.
In the region, .50000 < x < 1.0000 , -.70000 < y < -.20000 there are 16 zeros.
In the region, .0 < x < .50000 , -.20000 < y < .30000 there are 3 zeros.
In the region, .50000 < x < 1.0000 , -.20000 < y < .30000 there are 6 zeros.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

In the region, .50000 < x < .75000 , -.70000 < y < -.20000 there are 12 zeros.
In the region, .0 < x < .25000 , -.20000 < y < .30000 there are 2 zeros.
In the region, .25000 < x < .50000 , -.70000 < y < -.20000 there are 2 zeros.
In the region, .75000 < x < 1.0000 , -.70000 < y < -.20000 there are 4 zeros.
In the region, .25000 < x < .50000 , -.20000 < y < .30000 there is 1 zero.
In the region, .75000 < x < 1.0000 , -.20000 < y < .30000 there are 6 zeros.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

In the region, .50000 < x < .75000 , -.70000 < y < -.45000 there are 6 zeros.
In the region, .0 < x < .25000 , -.20000 < y < .50000E-01there are 2 zeros.
In the region, .25000 < x < .50000 , -.20000 < y < .50000E-01there is 1 zero.
In the region, .75000 < x < 1.0000 , -.20000 < y < .50000E-01there are 6 zeros.
In the region, .50000 < x < .75000 , -.45000 < y < -.20000 there are 6 zeros.
In the region, .25000 < x < .50000 , -.45000 < y < -.20000 there are 2 zeros.
In the region, .75000 < x < 1.0000 , -.45000 < y < -.20000 there are 4 zeros.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

In the region, .50000 < x < .62500 , -.70000 < y < -.45000 there is 1 zero.
In the region, .25000 < x < .37500 , -.20000 < y < .50000E-01there is 1 zero.
In the region, .75000 < x < .87500 , -.20000 < y < .50000E-01there are 2 zeros.
In the region, .50000 < x < .62500 , -.45000 < y < -.20000 there are 4 zeros.
In the region, .25000 < x < .37500 , -.45000 < y < -.20000 there is 1 zero.
In the region, .75000 < x < .87500 , -.45000 < y < -.20000 there are 4 zeros.
In the region, .62500 < x < .75000 , -.70000 < y < -.45000 there are 5 zeros.
In the region, .12500 < x < .25000 , -.20000 < y < .50000E-01there are 2 zeros.
In the region, .87500 < x < 1.0000 , -.20000 < y < .50000E-01there are 4 zeros.
In the region, .62500 < x < .75000 , -.45000 < y < -.20000 there are 2 zeros.
In the region, .37500 < x < .50000 , -.45000 < y < -.20000 there is 1 zero.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

```
In the region,    .25000    < x <    .37500    ,    -.20000       < y <   -.75000E-01there   is   1  zero.
In the region,    .75000    < x <    .87500    ,    -.20000       < y <   -.75000E-01there   are  2  zeros.
In the region,    .50000    < x <    .62500    ,    -.45000       < y <   -.32500     there   is   1  zero.
In the region,    .75000    < x <    .87500    ,    -.45000       < y <   -.32500     there   is   1  zero.
In the region,    .62500    < x <    .75000    ,    -.70000       < y <   -.57500     there   are  3  zeros.
In the region,    .12500    < x <    .25000    ,    -.20000       < y <   -.75000E-01there   is   1  zero.
In the region,    .87500    < x <    1.0000    ,    -.20000       < y <   -.75000E-01there   are  2  zeros.
In the region,    .50000    < x <    .62500    ,    -.57500       < y <   -.45000     there   is   1  zero.
In the region,    .50000    < x <    .62500    ,    -.32500       < y <   -.20000     there   are  3  zeros.
In the region,    .25000    < x <    .37500    ,    -.32500       < y <   -.20000     there   is   1  zero.
In the region,    .75000    < x <    .87500    ,    -.32500       < y <   -.20000     there   are  3  zeros.
In the region,    .62500    < x <    .75000    ,    -.57500       < y <   -.45000     there   are  2  zeros.
In the region,    .12500    < x <    .25000    ,    -.75000E-01   < y <    .50000E-01there   is   1  zero.
In the region,    .87500    < x <    1.0000    ,    -.75000E-01   < y <    .50000E-01there   are  2  zeros.
In the region,    .62500    < x <    .75000    ,    -.32500       < y <   -.20000     there   are  2  zeros.
In the region,    .37500    < x <    .50000    ,    -.32500       < y <   -.20000     there   is   1  zero.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

In the region,    .50000    < x <    .56250    ,    -.45000       < y <   -.32500     there   is   1  zero.
In the region,    .75000    < x <    .81250    ,    -.45000       < y <   -.32500     there   is   1  zero.
In the region,    .62500    < x <    .68750    ,    -.70000       < y <   -.57500     there   is   1  zero.
In the region,    .87500    < x <    .93750    ,    -.20000       < y <   -.75000E-01there   are  2  zeros.
In the region,    .50000    < x <    .56250    ,    -.32500       < y <   -.20000     there   is   1  zero.
In the region,    .75000    < x <    .81250    ,    -.32500       < y <   -.20000     there   are  3  zeros.
In the region,    .87500    < x <    .93750    ,    -.75000E-01   < y <    .50000E-01there   is   1  zero.
In the region,    .62500    < x <    .68750    ,    -.32500       < y <   -.20000     there   is   1  zero.
In the region,    .31250    < x <    .37500    ,    -.20000       < y <   -.75000E-01there   is   1  zero.
In the region,    .81250    < x <    .87500    ,    -.20000       < y <   -.75000E-01there   are  2  zeros.
In the region,    .68750    < x <    .75000    ,    -.70000       < y <   -.57500     there   are  2  zeros.
In the region,    .18750    < x <    .25000    ,    -.20000       < y <   -.75000E-01there   is   1  zero.
In the region,    .56250    < x <    .62500    ,    -.57500       < y <   -.45000     there   is   1  zero.
In the region,    .56250    < x <    .62500    ,    -.32500       < y <   -.20000     there   are  2  zeros.
In the region,    .31250    < x <    .37500    ,    -.32500       < y <   -.20000     there   is   1  zero.
In the region,    .68750    < x <    .75000    ,    -.57500       < y <   -.45000     there   are  2  zeros.
In the region,    .18750    < x <    .25000    ,    -.75000E-01   < y <    .50000E-01there   is   1  zero.
In the region,    .93750    < x <    1.0000    ,    -.75000E-01   < y <    .50000E-01there   is   1  zero.
In the region,    .68750    < x <    .75000    ,    -.32500       < y <   -.20000     there   is   1  zero.
In the region,    .43750    < x <    .50000    ,    -.32500       < y <   -.20000     there   is   1  zero.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

In the region,    .50000    < x <    .56250    ,    -.45000       < y <   -.38750     there   is   1  zero.
In the region,    .62500    < x <    .68750    ,    -.70000       < y <   -.63750     there   are -1  zeros.
In the region,    .50000    < x <    .56250    ,    -.32500       < y <   -.26250     there   is   1  zero.
In the region,    .75000    < x <    .81250    ,    -.32500       < y <   -.26250     there   is   1  zero.
In the region,    .87500    < x <    .93750    ,    -.75000E-01   < y <   -.12500E-01there   is   1  zero.
In the region,    .81250    < x <    .87500    ,    -.20000       < y <   -.13750     there   are  2  zeros.
In the region,    .68750    < x <    .75000    ,    -.70000       < y <   -.63750     there   are  2  zeros.
In the region,    .18750    < x <    .25000    ,    -.20000       < y <   -.13750     there   is   1  zero.
In the region,    .56250    < x <    .62500    ,    -.57500       < y <   -.51250     there   is   1  zero.
In the region,    .56250    < x <    .62500    ,    -.32500       < y <   -.26250     there   is   1  zero.
In the region,    .68750    < x <    .75000    ,    -.57500       < y <   -.51250     there   is   1  zero.
In the region,    .93750    < x <    1.0000    ,    -.75000E-01   < y <   -.12500E-01there   is   1  zero.
In the region,    .75000    < x <    .81250    ,    -.38750       < y <   -.32500     there   is   1  zero.
```

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| In the region. | .62500 | < x < | .68750 | , | -.63750 | < y < | -.57500 | there are 2 zeros. |
| In the region. | .87500 | < x < | .93750 | , | -.13750 | < y < | -.75000E-01 | there are 2 zeros. |
| In the region. | .75000 | < x < | .81250 | , | -.26250 | < y < | -.20000 | there are 2 zeros. |
| In the region. | .62500 | < x < | .68750 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .31250 | < x < | .37500 | , | -.13750 | < y < | -.75000E-01 | there is 1 zero. |
| In the region. | .56250 | < x < | .62500 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .31250 | < x < | .37500 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .68750 | < x < | .75000 | , | -.51250 | < y < | -.45000 | there is 1 zero. |
| In the region. | .18750 | < x < | .25000 | , | -.12500E-01 | < y < | .50000E-01 | there is 1 zero. |
| In the region. | .68750 | < x < | .75000 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .43750 | < x < | .50000 | , | -.26250 | < y < | -.20000 | there is 1 zero. |

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| In the region. | .50000 | < x < | .53125 | , | -.32500 | < y < | -.26250 | there is 1 zero. |
| In the region. | .75000 | < x < | .78125 | , | -.32500 | < y < | -.26250 | there is 1 zero. |
| In the region. | .81250 | < x < | .84375 | , | -.20000 | < y < | -.13750 | there is 1 zero. |
| In the region. | .68750 | < x < | .71875 | , | -.70000 | < y < | -.63750 | there is 1 zero. |
| In the region. | .18750 | < x < | .21875 | , | -.20000 | < y < | -.13750 | there is 1 zero. |
| In the region. | .56250 | < x < | .59375 | , | -.32500 | < y < | -.26250 | there is 1 zero. |
| In the region. | .68750 | < x < | .71875 | , | -.57500 | < y < | -.51250 | there is 1 zero. |
| In the region. | .93750 | < x < | .96875 | , | -.75000E-01 | < y < | -.12500E-01 | there is 1 zero. |
| In the region. | .75000 | < x < | .78125 | , | -.38750 | < y < | -.32500 | there is 1 zero. |
| In the region. | .62500 | < x < | .65625 | , | -.63750 | < y < | -.57500 | there is 1 zero. |
| In the region. | .87500 | < x < | .90625 | , | -.13750 | < y < | -.75000E-01 | there is 1 zero. |
| In the region. | .75000 | < x < | .78125 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .53125 | < x < | .56250 | , | -.45000 | < y < | -.38750 | there is 1 zero. |
| In the region. | .65625 | < x < | .68750 | , | -.70000 | < y < | -.63750 | there are -1 zeros. |
| In the region. | .90625 | < x < | .93750 | , | -.75000E-01 | < y < | -.12500E-01 | there is 1 zero. |
| In the region. | .84375 | < x < | .87500 | , | -.20000 | < y < | -.13750 | there is 1 zero. |
| In the region. | .71875 | < x < | .75000 | , | -.70000 | < y < | -.63750 | there is 1 zero. |
| In the region. | .59375 | < x < | .62500 | , | -.57500 | < y < | -.51250 | there is 1 zero. |
| In the region. | .65625 | < x < | .68750 | , | -.63750 | < y < | -.57500 | there is 1 zero. |
| In the region. | .90625 | < x < | .93750 | , | -.13750 | < y < | -.75000E-01 | there is 1 zero. |
| In the region. | .78125 | < x < | .81250 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .65625 | < x < | .68750 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .34375 | < x < | .37500 | , | -.13750 | < y < | -.75000E-01 | there is 1 zero. |
| In the region. | .59375 | < x < | .62500 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .34375 | < x < | .37500 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .71875 | < x < | .75000 | , | -.51250 | < y < | -.45000 | there is 1 zero. |
| In the region. | .21875 | < x < | .25000 | , | -.12500E-01 | < y < | .50000E-01 | there is 1 zero. |
| In the region. | .71875 | < x < | .75000 | , | -.26250 | < y < | -.20000 | there is 1 zero. |
| In the region. | .46875 | < x < | .50000 | , | -.26250 | < y < | -.20000 | there is 1 zero. |

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| In the region. | .75000 | < x < | .78125 | , | -.32500 | < y < | -.29375 | there is 1 zero. |
| In the region. | .81250 | < x < | .84375 | , | -.20000 | < y < | -.16875 | there is 1 zero. |
| In the region. | .68750 | < x < | .71875 | , | -.70000 | < y < | -.66875 | there is 1 zero. |
| In the region. | .18750 | < x < | .21875 | , | -.20000 | < y < | -.16875 | there is 1 zero. |
| In the region. | .75000 | < x < | .78125 | , | -.38750 | < y < | -.35625 | there is 1 zero. |
| In the region. | .62500 | < x < | .65625 | , | -.63750 | < y < | -.60625 | there is 1 zero. |
| In the region. | .87500 | < x < | .90625 | , | -.13750 | < y < | -.10625 | there is 1 zero. |
| In the region. | .75000 | < x < | .78125 | , | -.26250 | < y < | -.23125 | there is 1 zero. |
| In the region. | .65625 | < x < | .68750 | , | -.70000 | < y < | -.66875 | there are -1 zeros. |
| In the region. | .90625 | < x < | .93750 | , | -.75000E-01 | < y < | -.43750E-01 | there is 1 zero. |

```
In the region.  .71875  < x <  .75000  ,  -.70000      < y <  -.66875     there is 1 zero.
In the region.  .34375  < x <  .37500  ,  -.13750      < y <  -.10625     there is 1 zero.
In the region.  .59375  < x <  .62500  ,  -.26250      < y <  -.23125     there is 1 zero.
In the region.  .21875  < x <  .25000  ,  -.12500E-01  < y <   .18750E-01 there is 1 zero.
In the region.  .71875  < x <  .75000  ,  -.26250      < y <  -.23125     there is 1 zero.
In the region.  .50000  < x <  .53125  ,  -.29375      < y <  -.26250     there is 1 zero.
In the region.  .56250  < x <  .59375  ,  -.29375      < y <  -.26250     there is 1 zero.
In the region.  .68750  < x <  .71875  ,  -.54375      < y <  -.51250     there is 1 zero.
In the region.  .93750  < x <  .96875  ,  -.43750E-01  < y <  -.12500E-01 there is 1 zero.
In the region.  .53125  < x <  .56250  ,  -.41875      < y <  -.38750     there is 1 zero.
In the region.  .84375  < x <  .87500  ,  -.16875      < y <  -.13750     there is 1 zero.
In the region.  .59375  < x <  .62500  ,  -.54375      < y <  -.51250     there is 1 zero.
In the region.  .65625  < x <  .68750  ,  -.60625      < y <  -.57500     there is 1 zero.
In the region.  .90625  < x <  .93750  ,  -.10625      < y <  -.75000E-01 there is 1 zero.
In the region.  .78125  < x <  .81250  ,  -.23125      < y <  -.20000     there is 1 zero.
In the region.  .65625  < x <  .68750  ,  -.23125      < y <  -.20000     there is 1 zero.
In the region.  .34375  < x <  .37500  ,  -.23125      < y <  -.20000     there is 1 zero.
In the region.  .71875  < x <  .75000  ,  -.48125      < y <  -.45000     there is 1 zero.
In the region.  .46875  < x <  .50000  ,  -.23125      < y <  -.20000     there is 1 zero.

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

This appendix expands some parts of sections 2:7 and 2:9. In the first of these (on page 60) it is shown that in a neighbourhood of a simple corotation point there are two real (in fact analytic) solutions of equation [7:1] if the local Richardson number at the corotation point lies between 0 and 1/4, their expansions beginning

$$\begin{aligned} \zeta &= f_0 + \frac{\nu}{1+\nu} f_1 \sigma \cdots \\ \zeta &= g_0 - \frac{\nu}{1-\nu} g_1 \sigma \cdots \\ \nu &= \sqrt{1-4Ri} \end{aligned} \qquad [C:1]$$

When $Ri$ = 1/4 there is only one real solution ($\nu$=0 and $f_0 = g_0$); as $Ri$ is decreased this bifurcates into two real solutions, initially close, but as $Ri$ approaches 0 (and $\nu \to 1$) the slope of that which intersects $g$ at corotation becomes more and more extreme (and opposed to that of $g$) unless $g_1$ =0. When $Ri$ is very close to zero and $g_1 \neq 0$ this causes one complete oscillation of the solution to be squashed into a small neighbourhood of the corotation point with the two solutions being again almost coincident outside this region. If one considers the 'angular seperation' on the Riemann sphere of the values of the two solutions at a fixed point (excluding corotation)

$$\vartheta = \tan^{-1}\zeta_1 - \tan^{-1}\zeta_2 \qquad [C:2]$$

then

$$Ri \to \frac{1}{4} \quad \Rightarrow \quad \vartheta \to 0 \qquad [C:3]$$

but

$$Ri \to 0 \quad \Rightarrow \quad \vartheta \to \begin{cases} 0 & (\mathcal{G}\sigma > 0) \\ 2\pi & (\mathcal{G}\sigma < 0) \end{cases} \qquad [C:4]$$

where $\mathcal{G} = g_1(g_0 - f_0)$

This implies the geometric structure of fig.2 (facing page 63) and is the basis for the remarks at the top of page 74. The structure of fig.2 can also be deduced analytically from the formulae on pages 61,62; when $Ri$=1/4 the general solution is

$$\zeta = \tilde{\zeta} + \frac{1 + b_1\sigma \cdots}{C - a_0 \ln \sigma} \qquad [C:5]$$

and when $Ri$ =0 it is

$$\zeta = \tilde{\zeta} + \frac{\sigma^{-1}(1 + b_1 \sigma \cdots \quad)}{C \cdots \quad - (a_0 b_1 + a_1) \ln \sigma} \qquad [C:6]$$

Thus if we start with real values at negative $\sigma$ and integrate to positive $\sigma$, the sign of the imaginary part of $\zeta$ will in the first case be that of $a_0 \ln \sigma$ and in the second that of $(a_0 b_1 + a_1) \ln \sigma$. Because $\underset{\text{when } Ri=0}{} a_0(g_0 - f_0)=1$ and $a_0 b_1 + a_1 = -a_0^2 g_1$, if $\mathcal{J}$ is positive $a_0 b_1 + a_1$ has the opposite sign to $a_0$; thus the real axis (except for one point) maps to one hemisphere when $Ri$ =1/4 and the other when $Ri$ =0, but to the same when $\mathcal{J}$ is negative.

In the context of compressible flows this process of bifurcation and recombination has the interesting consequences described in section 2:9. If $g_1 \neq 0$ the fact that as $Ri \to 0$ a complete oscillation of one solution is compressed into a neighbourhood of the corotation point means that on recombination of the two solutions at $Ri$ =0 the nodes of one (on one side of corotation) have been shifted by one relative to those of the other solution (the 'vanishing' node has in effect been absorbed by the corotation singularity). The form of the 'neutral curve' depicted in fig.4 (facing page 74) follows at once.

The argument on page 73 then shows that this curve, except perhaps for one mode, must separate regions of stability and instability; the crucial point is the demonstration that $\partial\Delta/\partial\omega$ is complex for most modes which is amplified and improved in the following (it is slightly more convenient to use [9:2] than [9:3]).

Let us suppose that for one mode, i.e. a solution $\zeta$ of [9:2] with $\zeta(a) = \zeta(b)=0$, $\partial\Delta/\partial\omega$ is real (defining $\Delta$

as in section 2:9 to be the final value of $\xi$ at $b$ derived from an initial value of zero at $a$). Then making a first order perturbation in $\omega$ to $\tilde{\omega} = \omega + d\omega$ the imaginary part of $\Delta$ remains zero (to first order). The perturbed frequency $\tilde{\omega}$ has a unique associated perturbed real solution $\tilde{\xi}$ (if $Ri \neq 0, 1/4$) and $\tilde{\xi}(a)$ must satisfy the boundary condition at $a$ to first order. Moving the inner boundary point to one of the other nodes of $\xi$ in the inner oscillatory region one obtains a sequence of associated modes with the same frequency $\omega$.

I now show that on none of these associated modes can $\partial\Delta/\partial\omega$ also be real. For suppose $a_1$ and $a_2$ to be the inner boundary points of two modes with the same frequency $\omega$ such that for both $\partial\Delta/\partial\omega$ is real. Then $\xi(a_1) = \xi(a_2) = 0$ and by the above $\tilde{\xi}(a_1) = \tilde{\xi}(a_2) = 0$ to first order. This implies the existence of a solution to the variational equation

$$\left(\frac{\partial\xi}{\partial\omega}\right)' = -\frac{1}{r\sigma^2}\left\{\beta^2 r^2 + \frac{\sigma^2 r^2}{c^2} + (m - 2\Omega\xi)(m - \omega_0\xi) + \sigma^2\xi^2\right\} - \left\{\frac{2\sigma}{r}\xi + \frac{2m\Omega + m\omega_0}{r\sigma}\right\}\frac{\partial\xi}{\partial\omega} \qquad [C:7]$$

in $[a_1, a_2]$ with $\partial\xi/\partial\omega = 0$ at $a_1, a_2$. As in section 2:8 this implies that

$$\beta^2 r^2 + \frac{\sigma^2 r^2}{c^2} + (m - 2\Omega\xi)(m - \omega_0\xi) + \sigma^2\xi^2 \qquad [C:8]$$

musat change sign in $[a_1, a_2]$, but this interval lies in the oscillatory region where

$$\beta^2 r^2 - \frac{\sigma^2 r^2}{c^2} + (m - 2\Omega\zeta)(m - \omega_0\zeta) - \sigma^2\zeta^2 \qquad [C:9]$$

is a definite quadratic form in $\zeta$ and so

$$(2m\Omega + m\omega_0)^2 < 4(K^2 - \sigma^2)\left(m^2 + \beta^2 r^2 - \frac{\sigma^2 r^2}{c^2}\right) \qquad [C:10]$$

which implies

$$(2m\Omega + m\omega_0)^2 < 4(K^2 + \sigma^2)\left(m^2 + \beta^2 r^2 + \frac{\sigma^2 r^2}{c^2}\right) \qquad [C:11]$$

and thus [C:8] is also definite. This contradiction shows that at most on one out of a large, possibly infinite, set of modes can $\partial\Delta/\partial\omega$ be real. In fact for a neutral curve not to have contiguous unstable modes both $\partial\Delta/\partial\omega$ and $\partial\Delta/\partial\beta$ would have to be real all along it.